НАЦІОНАЛЬНА АКАДЕМІЯ НАУК УКРАЇНИ
ІНСТИТУТ ПРОБЛЕМ РЕЄСТРАЦІЇ ІНФОРМАЦІЇ

ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ ТА БЕЗПЕКА

МАТЕРІАЛИ XXI МІЖНАРОДНОЇ НАУКОВО-ПРАКТИЧНОЇ КОНФЕРЕНЦІЇ

ВИПУСК 21

Київ – 2021

Рекомендовано до друку Вченою радою
Інституту проблем реєстрації інформації НАН України
(протокол № 12 від 28 грудня 2021 р.)

Інформаційні технології та безпека. Матеріали XXI Міжнародної науково-практичної конференції ІТБ-2021. – Київ: Інжиніринг. – 228 с. ISBN: 978-966-2344-84-4

До збірника увійшли матеріали доповідей, представлених на XXI Міжнародній науково-практичній конференції «Інформаційні технології та безпека» (ІТБ-2021, 9 грудня 2021 року, м. Київ, Україна).

У збірнику представлені матеріали, присвячені питанням безпеки та живучості критичних інфраструктур, комп'ютерного моделювання складних систем, технологій аналітики великих обсягів даних (Big Data), аналітичних систем на основі відкритих джерел інформації (OSINT), моделювання, аналізу та прогнозування процесів мережевої взаємодії, методів і засобів підтримки прийняття рішень.

Для фахівців в області інформаційних технологій, інформаційної і кібернетичної безпеки, а також для аспірантів і студентів старших курсів вищої школи відповідних спеціальностей.

ISBN 978-966-2344-84-4

АВТОМАТИЗОВАНІ СИСТЕМИ ОРГАНІЗАЦІЙНОГО УПРАВЛІННЯ ОБ'ЄКТІВ КРИТИЧНИХ ІНФРАСТРУКТУР: БЕЗПЕКА І ФУНКЦІОНАЛЬНА СТІЙКІСТЬ

О.Г. Додонов, О.С.Горбачик, М.Г.Кузнєцова
Інститут проблем реєстрації інформації Національної академії наук України
Київ, Україна
dodonov@ipri.kiev.ua, ges@ipri.kiev.ua, margle@ipri.kiev.ua

Запропоновано підхід до підвищення безпеки об'єктів критичних інфраструктур на засадах теорії живучості систем. Враховуючи особливості функціонування автоматизованих систем організаційного управління (СОУ) в умовах виникнення і реалізації надзвичайних ситуацій (НС), обґрунтовано можливості забезпечення функціональної стійкості цих систем завдяки застосуванню механізмів підвищення живучості. Визначено поняття функціональної стійкості СОУ і способи його оцінювання. Сформульовано основні задачі СОУ у процесі напрацювання управлінських рішень в умовах НС.

Вступ

Автоматизовані системи організаційного управління об'єктів критичних інфраструктур (ОКІ) являють собою складні соціотехнічні системи, що функціонують в умовах мінливості зовнішнього середовища. Порушення у функціонуванні ОКІ (ядерна енергетика, хімічна промисловість, військова та авіаційна галузі, транспорт тощо) представляють потенційну загрозу людському життю й навколишньому середовищу. Імовірність виникнення надзвичайних ситуацій на об'єктах та у критичних інфраструктурах змушує враховувати наявність таких ситуацій, оцінювати ризики їх реалізації. Автоматизована СОУ має не лише гарантувати нормальне функціонування ОКІ у заданих умовах експлуатації, а й забезпечити адекватну реакцію на потенційну НС, ініціювати виконання відповідних заходів для її подолання. Виникають задачі, які не притаманні звичайному нормальному режиму функціонування СОУ та її складових, зростає навантаження на управлінців, які мають приймати рішення в умовах дії несприятливих впливів та обмеженого часового ресурсу. Для забезпечення керованості ОКІ і критичної інфраструктури в цілому необхідна функціональна стійкість СОУ щодо виконання тих функцій, які дозволять досягти бажаної цілі функціонування і протистояти руйнівним впливам.

Функціональна стійкість СОУ ОКІ та її оцінювання

Критичними об'єктами, або об'єктами критичних інфраструктур, визначено об'єкти, порушення (або припинення) функціонування яких може призвести до втрати управління, руйнування інфраструктури, незворотної негативної зміни (або руйнування) економіки країни, суб'єкта або адміністративно-територіальної одиниці, істотного погіршення безпеки життєдіяльності

населення, яке проживає на даних територіях. Фактично ОКІ являють собою складні багаторівневі ієрархічні системи, контроль їхнього стану і функціонування виконується відповідними багаторівневими керуючими системами з використанням засобів автоматичного контролю і управління за визначеними технічними регламентами [1].

Сучасні автоматизовані системи організаційного управління є складними соціотехнічними системами, у яких відбувається збір, аналіз та опрацювання інформації стосовно об'єкта управління, його внутрішнього середовища і взаємодії із зовнішнім середовищем [2]. Під функціональної стійкістю таких систем розуміється властивість зберігати структуру управління і виконувати основні управлінські функції в межах, установлених нормативними вимогами, в умовах впливу потоків відмов, несправностей і збоїв. Функціональна стійкість забезпечує знаходження СОУ у межах допустимого простору її станів, тобто, крім того, що система має можливість виконувати основні функції цільового призначення, вона не здійснює негативних впливів на оточуюче середовище та не становить загрози для його існування [3].

Вимоги щодо безпеки ОКІ обумовлюють потребу у функціональній стійкості СОУ ОКІ. Для оцінювання функціональної стійкості конкретних СОУ можна застосовувати такі показники:

- кількість виконуваних управлінських функцій,
- виконання визначеної функції з заданим показником якості;
- сумарна вага виконуваних функцій не нижче заданої,
- безперервність функціонування СОУ;
- керованість технологічних процесів;
- зв'язність управлінської структури;
- завершеність управлінських процесів.

Забезпечення стійкості функціонування СОУ в умовах надзвичайних ситуацій

Зміни в оточуючому середовищі і в критичній інфраструктурі здатні викликати структурні і функціональні порушення в СОУ, які можуть викликати збої у її функціонуванні, появу некерованих процесів і розвиток надзвичайних ситуацій на ОКІ, і навіть призвести до катастроф.

Надзвичайна ситуація – це порушення нормальних умов життя і діяльності людей на об'єкті або території, спричинене аварією, катастрофою, стихійним лихом, епідемією, пожежею, застосуванням засобів ураження. Виділяють 5 типових етапів (фаз розвитку НС): накопичення відхилень — ініціювання надзвичайної події — прояв дії основних вражаючих факторів — дія вторинних вражаючих факторів — дія залишкових вражаючих факторів. Заходи, які мають здійснюватись щодо НС: запобігання виникненню НС, реагування на НС та ліквідація наслідків НС.

Виникнення і розвиток НС відбувається у реальному часі, тому вирішення задачі відслідковування і аналізу етапів НС потребує побудови і застосування відповідних динамічних моделей. Системи управління в умовах НС мають зважати на характер і швидкість її розвитку, мають контролювати не лише параметри ОКІ, а й характеристики оточуючого середовища.

Під час дії несприятливих факторів в СОУ ОКІ мають виконуватися наступні функції:

- моніторингу, основною задачею якого є неперервна обробка даних з ОКІ у реальному масштабі часу і сигналізація про вихід тих чи інших параметрів за допустимі межі;
- діагностики місця виникнення НС для локалізації зони несправностей чи помилок;
- прогнозування для передбачення наслідків певних подій або явищ на основі аналізу даних від різних джерел;
- планування дій щодо попередження виникнення НС;
- корегування управлінських рішень у разі діагностування помилок і збоїв;
- забезпечення управління ОКІ і відповідного режиму функціонування самої СОУ;
- забезпечення необхідною інформацією і рекомендаціями всіх рівнів.

В умовах НС у функціонуванні СОУ виникають проблеми, які пов'язані з високим темпом змін стану ОКІ, непередбачуваністю подій, залежністю інформаційних потоків від ситуації, зміною розподілу функцій, розширенням або звуженням області дій, тощо [4]. У таких умовах бажано надати управлінцям певну підтримку для напрацювання рішень, зокрема, заздалегідь напрацювати схеми дій в умовах розвитку НС, створити базу прецедентів, де накопичити попередній досвід подолання наслідків НС. Це потребує впровадження в СОУ аналітичної складової.

При створенні аналітичного ресурсу СОУ щодо НС доцільно розглянути «повільний» і «швидкий» сценарії розвитку аварій. Частину впливів і наслідків їх дій, які можна передбачити, слід заздалегідь проаналізувати і напрацювати ефективні управлінські рішення. При «повільному» сценарії є час на застосування визначених засобів захисту та виконання певних захисних дій для локалізації НС.

У разі реалізації «швидких» сценаріїв події розгортаються стрімко і потрібні такі заходи та засоби, які дозволять зреагувати раніше, ніж станеться руйнування. Управлінці мають визначити заходи реагування і оцінити можливості їх застосування при наявних ресурсах.

Розробляючи засоби аналітичної підтримки в СОУ ОКІ для забезпечення ефективного і безпечного її функціонування в умовах виникнення та розвитку НС, слід приділити увагу розв'язанню наступних задач:

1) оцінювання середовища функціонування, оцінка ризиків безпеки і виникнення НС;

2) виявлення місць в критичній інфраструктурі, які мають найвищі ризики, та оцінка їх впливу на структуру і функціонування ОКІ, оцінювання потенційного збитку;

3) формування динамічних моделей предметної області, придатних для аналізу ймовірної поведінки системи при зміні умов у зовнішньому чи внутрішньому середовищі та напрацювання рішень щодо варіантів реагування СОУ ОКІ;

4) визначення і моніторинг індикаторів функціонування ОКІ, побудова моделей оцінки та прогнозування стану ОКІ, побудова сценаріїв розвитку подій при ймовірних зовнішніх і внутрішніх впливах на ОКІ та СОУ;

5) визначення переліку засобів і заходів протидії виникненню надзвичайних ситуацій, розробка сценаріїв виникнення, розвитку, локалізації та ліквідації наслідків НС;

6) створення і застосування ресурсів аналітичної складової СОУ для моделювання ситуацій, напрацювання управлінських рішень та аналізу їх ефективності.

З еволюцією й зростанням складності ОКІ змінюються його параметри і складові, що веде і до зростання кількості та різноманітності видів і типів ризиків, проявлення яких може призводити до порушення функціонування ОКІ та всієї інфраструктури [2], тому має постійно проводитись аналіз спектру можливих ризиків, моніторинг стану виділених параметрів і оточуючого середовища, моделювання нових ризикових ситуацій, мають впроваджуватись нові технологічні рішення для забезпечення ефективного та безпечного функціонування ОКІ при загрозі НС.

Застосування механізмів підвищення живучості для підтримки функціональної стійкості СОУ

Для забезпечення функціональної стійкості СОУ при проектуванні традиційно вводиться певна надмірність (структурна, програмна, часова, ресурсна); впроваджуються системи вбудованого контролю; формуються контури захисту СОУ та ОКІ від впливу негативних факторів зовнішнього середовища; вибираються компоненти із підвищеним рівнем захищеності і надійності. Ці традиційні рішення мають певні обмеження. Додаткова надмірність, на жаль, веде до погіршення техніко-економічних характеристик систем. Системи контролю відстежують заданий ряд параметрів, але не завжди або зовсім не можуть забезпечити адекватну реакцію на нештатну ситуацію та не можуть впливати на зменшення ймовірності виникнення таких ситуацій. Контур захисту може мінімізувати вплив зовнішніх факторів, але повністю його не виключає. Вибір елементної бази з підвищеним рівнем захищеності й надійності підвищує відмовостійкість СОУ, але не забезпечує функціональної стійкості, коли відмова вже сталася.

В умовах виникнення НС функціональна стійкість СОУ ОКІ має бути підтримана механізмами підвищення живучості, адже живучість є тією

властивістю, яка забезпечує системі можливість адаптуватися до нових умов функціонування при здійсненні несприятливого впливу, зберігати або оперативно відновлювати виконання функцій системи з мінімальними втратами ефективності в разі деградації чи виходу з ладу окремих компонент системи за рахунок використання наявних працездатних ресурсів.

Для підтримки функціональної стійкості СОУ доцільно застосування наступних механізмів підвищення живучості [5]:

- розпізнавання й локалізації – для виявлення атак на інформаційні складові СОУ, успішних вторгнень у систему та її інформаційні ресурси, виникнення ризиків втрати чи викривлення інформації, підвищення ризику і виходу з ладу життєво-важливих (критичних) компонент СОУ, ідентифікацію елемента чи компонента, що вийшов з ладу, фіксацію виходу певних параметрів СОУ, ОКІ та оточуючого середовища за встановлені межі;
- протидії – створення і застосування набору заздалегідь визначених засобів і заходів підтримки заданих умов функціонування і мінімізації збитків, які пов'язані з переходом у нештатний режим функціонування;
- відновлення – розробка і використання сукупності методів і програмно-апаратних засобів для відновлення функціональності і працездатності компонент і системи в цілому, її інформаційних ресурсів і інформаційно-комунікаційних засобів при несприятливих впливах;
- адаптації – розробка сукупності процедур цілеспрямованої зміни параметрів і структури СОУ на основі інформації про зміни в умовах функціонування, виникнення непередбачених ситуацій, про наслідки щодо порушення захищеності інформаційного ресурсу;
- реорганізації – розробка алгоритмів і процедур перерозподілу функцій компонент, які вийшли з ладу, між працездатними компонентами або, у разі неможливості перерозподілу, – організація переходу системи до нової цілі функціонування;
- реконфігурації – виконання автоматичної(автоматизованої) перебудови структури мережі обміну інформацією або зміна алгоритму функціонування для досягнення найбільшої ефективності виконання цілі функціонування на наявних працездатних ресурсах системи;
- реконструкції – застосування редукції цілі функціонування (переліку виконуваних функцій СОУ) та ресурсів системи до визначених базових рівнів, коли система може виконувати тільки чітко окреслену множину функцій, зберегти певний визначений обсяг інформації, забезпечити плавність деградації визначених параметрів.

Механізми підвищення живучості мають застосовуватись комплексно, інтегруючи наявні в системі засоби і проектні рішення щодо підвищення

надійності й відмовостійкості, моніторингу, автоматичного контролю і компенсації, вбудовані алгоритми і засоби захисту компонент і складових СОУ, наявні можливості аналітичної складової СОУ для напрацювання і корегування процесів управління в умовах виникнення несприятливих впливів або розвитку НС.

Доцільність й ефективність запропонованого підходу до забезпечення безпеки ОКІ і критичної інфраструктури в цілому в умовах наявності несприятливих впливів шляхом підвищення функціональної стійкості СОУ апробовано при створенні автоматизованих СОУ спеціального призначення.

Висновки

Створення і впровадження автоматизованих СОУ на засадах теорії живучості систем з розвинутою аналітичною складовою дозволить забезпечити функціональну стійкість таких систем і підвищити безпеку ОКІ та всієї критичної інфраструктури.

Перелік посилань

1. Кузнєцова М.Г. Створення функціонально стійких систем організаційного управління об’єктів критичних інфраструктур. *Реєстрація, зберігання і обробка даних*: зб. наук. праць Щорічної підсумкової наукової конференції (17–18 травня 2018, м. Київ). Київ, 2018. С. 116–119.

2. Горбачик О.С. Технологія реконфігурації у системах організаційного управління критичних інфраструктур. *Реєстрація, зберігання і оброб. даних*: зб. наук. праць за матеріалами Щорічної підсумкової наукової конференції 16–17 травня 2018 року ІПРІ НАН України. Київ, 2018, С. 105–108.

3. Барабаш О.В., Дурняк Б.В., Машков О.А. та ін. Забезпечення функціональної стійкості складних технічних систем // Моделювання та інформаційні технології: Зб. наук. праць. — К.: Ін-т проблем моделювання в енергетиці ім. Г.Є. Пухова НАН України, 2012. — 64. — С. 36–41.

4. Системы управления в чрезвычайных ситуациях.URL: http://risk.keldysh.ru/risk/g112.htm

5. Харченко В.С., Яковлев С.В., Горбачик О.С. та ін. Забезпечення функціональної безпеки критичних інформаційно-керуючих систем. Харків: Константа, 2019. 272 с.

ПОКАЗНИК ЧАСУ РЕЛАКСАЦІЇ ЯК УНІКАЛЬНА ХАРАКТЕРИСТИКА ДЛЯ КЛАСТЕРИЗАЦІЇ МЕРЕЖ

А.О. Снарський [1,2][0000-0002-4468-4542] Д.В. Ланде [1,2][0000-0003-3945-1178]
О.О. Дмитренко [1,2][0000-0001-8501-5313]

[1] Інститут проблем реєстрації інформації НАН України, Київ, Україна
[2] Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського", Київ, Україна

asnarskii@gmail.com, dwlande@gmail.com, dmytrenko.o@gmail.com

В цій роботі досліджується числові характеристики вузлів мережевих структур – показник часу релаксації мережі та індивідуальний показник часу релаксації вузла, які характеризують стійкість складної мережі та, відповідно, кожного вузла окремо до зовнішніх збуджень. Обчислення показників часу релаксації здійснюється за допомогою уповільненого ітераційного алгоритму HITS. Показано, що показник часу релаксації є унікальними числовими характеристиками вузлів мережі, і їх можна використовувати для знаходження центроїдів кластерів та об'єднання вузлів у групи за цими показниками. Апробацію представлених характеристик показника часу релаксації та індивідуального показника часу релаксації було проведено на прикладі кластеризації випадкових мереж з чітко вираженими кластерами.

Вступ

Складні мережі широко поширені у природі. Наприклад, такі мережі, як Всесвітня мережа Інтернет, мережі співавторства [1] та інші є складними [2], [3]. Всі вони володіють нетривіальними топологічними властивостями, а отже, становлять великий інтерес для досліджень. Важливе місце також посідають соціальні мережі, де бінарні зв’язки між людьми у групі можна представити у вигляді мережі, де кожен об'єкт – це точка, а його зв'язок з іншим об'єктом – дуга. Розбиття вибірки об’єктів на підмножини, що не перетинаються, тобто розбиття на кластери й, наприклад, виявлення спільнот у соціальних мережах є актуальним завданням.

Існує безліч методів кластеризації графів складних мереж [4, 5], які відрізняються ідеєю об’єднання схожих вузлів. Оскільки різні моделі використовують різні алгоритми, то розрізняють різні моделі кластеризації, зокрема такі як моделі зв’язності, центроїдні, статистичні, групові, нейронні моделі та інші. Наприклад, існують моделі, які ґрунтується на основі зв’язку, де об’єкти, які знаходяться у просторі ближче, є більш подібними (спорідненими), а також існують моделі, що засновані на знаходженні центроїда – кластери представлені у вигляді центрального вектора.

В цій роботі ми пропонуємо використовувати нову характеристику для кластеризації складних мереж – показник часу релаксації [6].

Показник часу релаксації

Під час дослідження складних мереж було помічено, що значення мережевих характеристик вузлів після зовнішніх збурень, в результаті перерахунку відновлюються до своїх початкових рівноважних значень за деякий індивідуальний для кожного вузла час. Порівнюючи показник часу релаксації із самою мережевою характеристикою, значення якої піддавалося збуренню, простежується нечітка залежність, яка показує, що показник часу релаксації є унікальною та неподібною до інших числовою характеристикою вузлів мережі [6].

У випадку дослідження складних мереж показником часу релаксації називається кількість ітераційних кроків τ відповідного алгоритму, які необхідні, щоб з деякою наперед заданою точністю μ – умовою збіжності (зазвичай μ=10^{-4}) досягти початкових рівноважних числових значень певної характеристики після її збурення.

У випадку, коли досліджується відновлення всієї мережі, то показником часу релаксації мережі після збурення певного m-го вузла $\tau^{(m)}$ називають кількість ітераційних кроків, які необхідні, щоб відновилось значення кожного вузла (відновилась вся мережа), або $\max_k(\tau_k^{(m)})$ (k = 1,….,n, де n – кількість вузлів мережі). А кількість ітерацій $\tau_m^{(m)}$ (або просто τ_m), які необхідні для відновлення саме того вузла, числове значення характеристики якого було збурене, називатиметься індивідуальним показником часу релаксації вузла. Тобто перший показник характеризує вузол в термінах стійкості всієї мережі після наданого йому зовнішнього збурення, останній – характеризує індивідуальну стійкість вузла після його збурення.

В якості числової характеристики, числове значення якої піддається збуренню, в цій роботі використовувалась характеристика, що відповідає ітераційному алгоритму ранжування HITS (Hyperlink Induced Topic Search). Цей алгоритм був запропонований та розроблений в 1998 році Дж. Клейнбергом [7] для вибору із масиву документів кращих «авторів» (першоджерел, на які посилаються інші документи) та «посередників» (документів, які посилаються на ці першоджерела). Для кожного документа j обчислюється його важливість як «автора» a(j) (з англ. Authority) і як «посередника» h(j) (з англ. Hub) відповідно до формул:

$$a(j) = \sum_{i \to j} h(i),\ h(j) = \sum_{j \to i} a(i), \qquad (1)$$

Оскільки вузли складної мережі мають велику кількість взаємозв'язків, то ітераційний процес перерахунку значень вузлів після їх збурення зазвичай є швидким, і достатньо всього декілька ітераційних кроків – часу релаксації, аби досягти початкових рівноважних значень характеристики HITS. Тож після збурення певного вузла й перерахунку числових значень,

для того, щоб уповільнити процес збіжності, у даній роботі пропонується здійснити уповільнення алгоритмів. Тож після надання збурення одному із вузлів мережі застосовується відповідний ітераційний алгоритм HITS з уповільненням:

$$h_0(j) \to h_1(j) \to h_2(j) \to \ldots \to h_{\text{рівноважне}}(i) \quad (2)$$

$$\widehat{\boldsymbol{A}} h_1(i) = h_2(i),\ \widehat{\boldsymbol{A}} h_n(i) = h_{n+1}(i) \quad (3)$$

$$h_{n+1}(i) \longleftarrow \beta h_{n+1}(i), \quad (4)$$

де $0 < \beta < 1$ – коефіцієнт уповільнення, $\widehat{\boldsymbol{A}}$ – оператор алгоритму HITS.

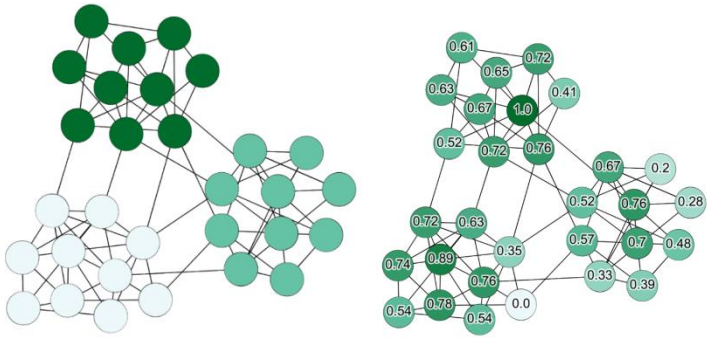
Приклади кластеризації мереж

Апробацію представлених характеристик показника часу релаксації та індивідуального показника часу релаксації було проведено на прикладі кластеризації випадкових мереж різної розмірності з чітко вираженими кластерами. Зокрема, було досліджено випадково згенеровану матрицю розмірності 30×30 з 3-ма кластерами та матрицю 100×100 з 4-ма кластерами.

Для кожної мережі, що описується відповідною матрицею, після збурення значення мережевої характеристики h(j) завдяки уповільненому алгоритму HITS було обчислено показник часу релаксації мережі та індивідуальний показник часу релаксації для кожного вузла. Для кожної мережі коефіцієнт уповільнення β алгоритму HITS та умова збіжності μ числових значень h(j) та a(j) до передзбурених значень обиралася індивідуально.

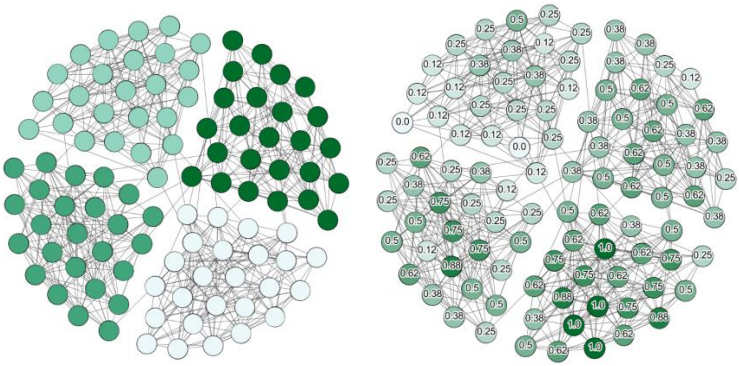
Також для того щоб уникнути великих числових значень, які інколи виникають після застосування уповільнених алгоритмів, всі результуючі значення показників часу релаксації були нормовані в інтервалі [0,1].

На рис. 1 та 2 наведено візуальне представлення мереж. На кожному з рисунків вузли розфарбовані за класом модулярності (рис. 1(a) та 2(a)) та за показником часу релаксації мережі та індивідуальним показником часу релаксації вузла (для представлених прикладів нормовані числові значення цих характеристик співпадають та представлені у вигляді міток вузлів – рис. 1(b) та 2(b)). Показник часу релаксації для мережі, що представлена на рис. 1, був отриманий в результаті уповільнення алгоритму HITS з коефіцієнтом уповільнення $\beta = 0.9$ та умовою збіжності $\mu = 10^{-6}$. Для мережі, що представлена на рис. 2 – $\beta = 0.9$ та $\mu = 10^{-4}$.



a) b)

Рис. 1. Приклад мережі, що відповідає випадково згенерованій матриці розмірності 30×30 з 3-ма кластерами, де групи вузлів об'єднані (розфарбовані) за a) класом модулярності та b) показником часу релаксації (числові значення представлені у вигляді міток вузлів)



a) b)

Рис. 2. Приклад мережі, що відповідає випадково згенерованій матриці розмірності 100×100 з 4-ма кластерами, де групи вузлів об'єднані (розфарбовані) за a) класом модулярності та b) показником часу релаксації (числові значення представлені у вигляді міток вузлів)

Можна помітити, що групи вузлів, об'єднані за класом модулярності, відрізняються від груп, отриманих за допомогою об'єднання вузлів за

показником часу релаксації. Це означає, що показник часу релаксації мережі та індивідуальний показник часу релаксації вузла є унікальними числовими характеристиками вузлів мережі, і їх можна використовувати для знаходження центроїдів кластерів та об'єднання вузлів у групи за цими показниками.

Висновки

В цій роботі було розглянуто нові числові характеристиками вузлів мережевих структур – показник часу релаксації мережі та індивідуальний показник часу релаксації вузла. Апробацію показника часу релаксації було проведено на прикладі кластеризації випадкових мереж з чітко вираженими кластерами. Було встановлено, що показник часу релаксації мережі та індивідуальний показник часу релаксації вузла є унікальними числовими характеристиками вузлів мережі, і їх можна використовувати для знаходження центроїдів кластерів та об'єднання вузлів у групи за цими показниками.

Літературні джерела

1. Barabâsi, A. L., Jeong, H., Néda, Z., Ravasz, E., Schubert, A., & Vicsek, T.: Evolution of the social network of scientific collaborations. Physica A: Statistical mechanics and its applications, 311(3-4), 590-614 (2002).
2. Newman, M. E. J.: The structure and function of complex networks. SIAM Review, vol. 45. pp. 167–256. (2003). doi:10.1137/S003614450342480
3. Strogatz, S. H.: Exploring complex networks. nature, 410(6825), 268-276 (2001).
4. Lancichinetti, A., & Fortunato, S. Community detection algorithms: a comparative analysis. Physical review E, 80(5), 056117 (2009).
5. Estivill-Castro, V.: Why so many clustering algorithms: a position paper. ACM SIGKDD explorations newsletter, 4(1), 65-75 (2002).
6. Lande D., Snarskii A., Dmytrenko O., Subach I.: Relaxation time in complex network. ARES '20: Proceedings of the 15th International Conference on Availability, Reliability and SecurityAugust 2020. Article No.: 99 1-6. (2020) doi: https://doi.org/10.1145/3407023.3409231
7. Kleinberg, J. M.: Authoritative sources in a hyperlinked environment. In Processing of ACM-SIAM Symposium on Discrete Algorithms, 46(5), pp. 604–632 (1998).

IMPROVING THE EFFICIENCY OF TYPICAL SCENARIOS OF ANALYTICAL ACTIVITIES

Oleksandr Koval [1][0000-0003-0991-6405], Valeriy Kuzminykh [1][0000-0002-8258-0816], Iryna Husyeva [1][0000-0002-8762-3918], Xu Beibei [2][0000-0003-1430-5334], and Zhu Shiwei [2][0000-0002-6651-8449]

[1]National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine

[2] Information Institute, Qilu University of Technology (Shandong Academic of Sciences), Jinan, China

avkovalgm@gmail.com, vakuz0202@gmail.com, iguseva@yahoo.com, xubeibei1987@163.com, zhusw@sdas.org

The article considers methods for evaluating the efficiency of modeling scenarios of analytical activities based on ontology using graphs. The algorithm for evaluating the efficiency of scenario modeling by analytical activities is presented, together with the algorithm for performing a sequential check of the obtained scenario for its optimality. The results of simulation experiments of modeling of scenarios with use of the offered methods and the developed software on two hundred test examples are described.

Introduction

Analytical activity, as a rule, means a set of actions of the user (analyst) of the software system for the collection, accumulation, processing and analysis of data in order to justify and prepare decisions or the formation of new knowledge. In the tasks of analytical activities, the scenario is a set of methods for describing and organizing the steps of analytical activities under given constraints. In this case, the ontology of the subject domain, as a basis for building a model in the form of a graph, provides a description of the structure of the subject area and provides definitions for a set of concepts and possible relationships between them [1].

It allows to develop different scenarios for achieving the objectives of analytical research, taking into account a number of factors and limitations. The scenario of analytical activity is a sequence of functional tasks, initial, final and intermediate events aimed at achieving the goals of analytical research. The scenario description consists of a branched directed sequence of elements connected by links that form a structure as a directed graph. Each element of this graph is either an action performed by the scenario executor or an event that affects the subsequent action of the scenario. The event can be caused both by changes in the external environment and by changes in the state of the object of study. Each action can be described as a function, or as a procedure that can implement any sequence of operations with a given level of detail.

Scenario construction is often based only on a visual representation of the data processing model and a fixed set of requirements. But most of the requirements at

the initial stage are fuzzy, without formalizing and taking into account the semantic content of the processes, which leads to the logical errors in the execution of scenarios and the need to return to the beginning of the modelling process. This greatly complicates the modelling process and increases the necessary time. Formally, the analytical process is represented by an directed graph, which reflects the sequence of possible actions of the user to achieve the goal of analytical research [2].

Currently, there are a lot of methods for selecting appropriate scenarios from a set of valid scenarios, based on the relevant ontology, and analysing the nature and composition of factors that may affect the process of planning and execution of the scenario. In practice, planning and implementation of the scenario determines the features of the ways which differentiate a scenario by factors based on values of the criteria to make the decisions and determine the procedures for forming the necessary scenarios. One of the most complicated and relevant tasks today, which requires the construction of effective scenarios for its solution, is the task of building and optimizing the scenarios for analytical activities based on branched information.

The task of evaluating the efficiency of information streams flow collection for analytical activities based on branched information is one of the most important tasks of processing large amounts of data [3].

Efficiency estimates of graph-based scenario modelling

The structural approach to ontology consideration is the most appropriate to the task stated, as the representation of ontology structure in the form of a multilevel hierarchical graph makes it possible to determine its properties using metrics, which can determine its quality and make recommendations for its improvement. An important aspect of the structural approach is the need to assess the efficiency of the ontological model based on the assessment of the reduction of search costs on the graph describing the ontology, compared with the unfocused search for a path. Such graphs typically use cost estimates for both the processing of information-related actions at individual nodes in the graph and the cost estimates for the actions (along the edges of the graph) when moving from one node to another. Thus, it is possible to estimate the total cost of each possible path according to the scenario or a separate fragment of this path to compare and determine the most efficient paths included in the description of the selected scenario (Fig.1) [4].

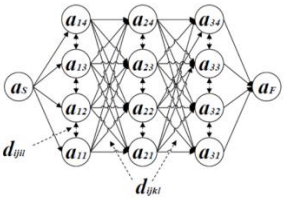


Fig.1. Graph structure for script construction.

The steps of the algorithm for evaluating the efficiency of scenario modelling by analytical activities are as follows.

Step 1. Select on the bottom layer of nodes of the graph the appropriate assessment of the efficiency of finding the optimal path on the graph according to the relevant criteria.

Step 2. Analyse the relationships of the selected nodes with the nodes of the next level, which are located above in the hierarchy of the model.

Step 3. Save the graph edges which correspond to the found links for further analysis and use in search operations.

Step 4. Repeat iterations to the top-level search source.

Step 5. Consolidation of all collected edges and nodes of the graph, which are built on the basis of the appropriate efficiency assessment - the information retrieval request.

Step 6. Arrange all paths on a constant graph according to the evaluation function.

Step 7. Analysis of the graphical model of the scenario, which is based on the choice of the shortest paths with the minimum number of edges on the graph (Fig. 2).

Estimation of process efficiency of analytical activity scenario modelling as efficiency of search of an optimum path on a graph can be defined as aggregate estimation action costs on each of nodes and an estimation of costs of transition to the next node in transition from initial node to final:

$$P_a = \sum_k (t_k + r_k),$$

where t_k – time spent on each of k nodes, corresponding to the selected path of the graph; r_k – resources spent on each of k nodes, corresponding to the selected path of the graph; $k \in K$ – node numbers on the selected path of the graph.

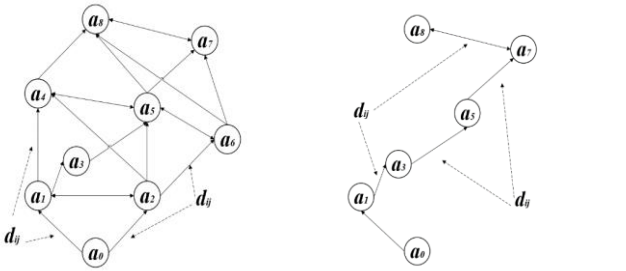


Fig. 2. An example of a model for constructing the optimal path from the initial a_0 to final a_8 node at possible paths (edges) d_{ij}

Accordingly, the evaluation criterion can be defined as the sum of processing costs in the transition edges between nodes:

$$P_d = \sum_{i,j} (T_{ij} + R_{ij}),$$

where T_{ij} – time spent at each ij edge of the graph, corresponding to the selected path of the graph; R_{ij} – resources spent at each ij edge of the graph, corresponding to the selected path of the graph; $i,j \in K$ – node numbers on the selected path of the graph.

Then the general criterion can be defined as the sum P_a and P_d. General criterion

$$P = P_a + P_d.$$

or

$$P = \sum_{k} (t_k + r_k) + \sum_{i,j} (T_{ij} + R_{ij}), \quad k,i,j \in K,$$

where K – a set of indices corresponding to the selected path.

Thus both criterion of efficiency, and an estimation of the received results in certain node according to concrete task is defined by presence of necessary results according to estimations of analysts-experts:

$$P = \sum_{k} (c_t t_k + c_r r_k) + \sum_{i,j} (c_T T_{ij} + c_R R_{ij}), \quad k,i,j \in K,$$

where c_t – coefficient of impact of time spent on each of k nodes;
c_r – coefficient of impact of resources spent on each k nodes;
c_T – coefficient of impact of time spent on each of ij edge of graph;
c_R – coefficient of impact of resources spent on each ij edge of graph.
Additionally, $c_t, c_r, c_T, c_R \geq 0$.

Depending on the values of these coefficients, the analyst determines the type of task for finding the optimal path: by the criterion of time, by the minimum cost of resources or using a combined criterion. The choice of the total direction of the path can be determined by the presence of the most efficient direction in terms of the value of criterion V at each subsequent step of choosing the direction on the graph.

The trigger of the evaluation of results at each step determines whether there is an improvement in results or not.

$Q_i > 0$ – the node is considered as possible for comparison and selection.

$Q_i = 0$ – node a_i is not considered as possible for comparison and selection.

This assessment should be performed by the analyst based on the results obtained.

For each level, the directions (edges) and corresponding nodes are considered in comparison $Q_i > 0$. Other directions (edges) and nodes are removed from further consideration.

Value $Q_i > 0$ can be determined by certain selected by the analyst measures as an estimate of the increase in the results of solving the analytical problem.

In the absence of available and reasonable assessment methods $Q_i = 1$ – node a_i is considered as one that improves the analyst's decision. $Q_i = 0$ – node a_i is not considered possible for comparison and selection, as not providing new knowledge or new results for the analyst.

Without reducing the degree of generality, the model, which considers the costs associated with both the overall cost estimate for each node and the cost estimate for the transition to the next node in the transition from the initial node to the final, can be considered as a model taking into account only the costs on the edges of the graph.

In this case, the costs in the nodes will be included in the costs of the edges of the graph:

$$P_{ij} = t_j + r_j + T_{ij} + R_{ij}, \quad i,j \in K.$$

Based on the general description of the characteristics of the ontology model graph, the formulation of the problem of finding the shortest paths on the graph (Single Source Shortest Path - SSSP) [5,6] can be formulated as follows.

It is necessary to specify a path from the initial state of the graph a_S, to the final state a_F, which has the lowest possible total weight*: P*(a_S,a_F).*

f(aS)=f(P*(a_S,a_F)=min f(P(a_S, a_F)).*

For each level of the graph $j=1...m$

A={a_{ij}} for $i=1...n$ – it is a complete set of actions that have similar in quality, but different in performance, results.

If d_{ijkl} – is the edge from a_{ij} to a_{kl}, and d_{ijil} – is the edge from a_{ij} to $a_{il,}$ in this case, in most cases, we can assume that

$d_{ijil} << d_{ijkl}$ for $i,k = 1...n$ and $j,l = 1...m$ at $i \neq k$.

A typical scenario defines a set of edges d^T_{ii+1} for $i = 1...n-1$

A scenario can be defined as improved if the total score of the scenario which identifies the new scenario for the new path d_{iji+1l} less than the estimates of the sum of the edges of the typical scenario $d^T_{ii+1.}$

The algorithm for performing a sequential check of the obtained scenario for its optimality can be composed of the following steps.

1. For each level of the scenario $i = 1...n-1$ and $j,l = 1...m$ value d_{iji+1l} is compared to d^T_{ii+1}.

2. If the values are such that $d_{iji+1l} < d^T_{ii+1}$, then the new value of the current default scenario is defined as $d^{TN}_{ii+1} = d_{iji+1l}$.

3. Building a new improved scenario sequence from the node a_{i+1l} sequentially on all subsequent levels to level m.

4. If the total value d^{TN}_{ii+1} for the new path is less than that for d^T_{ii+1} initial typical scenario, this scenario is considered as typical.

5. If the conditions of paragraph 4 are not met, the process may be repeated from the final node of the typical scenario, from which the new direction of the scenario was adopted, in another direction to the last level of the graph.

When there is a better path of action according to the chosen criterion (for example, according to the time of receiving and processing information)

compared to the typical scenario, it will be defined to replace the typical scenario with a new one.

Otherwise, it can be argued that the previously obtained scenario is optimal. Thus, the constructed scenario is checked for optimality.

As a result of performing the actions of the described algorithm, we obtain a scenario which meets the specified conditions, in accordance with the basic ontology described in graph [7]. The presented algorithm reflects the sequential process of detecting the sequence of edges connecting the nodes of the graph from the bottom layer to the top, taking into account the parameters of suitability.

Simulation experiments of scenarios modelling with the use of the suggested methods and the developed software were carried out on two hundred test examples of the decision of analytics problems with averaging of test results examples for various problem areas (ontological models). Cost estimation is defined as the estimation of the reduction of analytical procedures.

Analysis of test results

To confirm the efficiency of the above-described approach to scenario building, a number of tests were performed on the generated test cases in a specialized test environment. More than two hundred test cases were considered to estimate the number of calculations and comparisons when searching for the optimal path on graphical structures, ie on graphs which reflect the initial ontology.

Conducting such testing and analysis of the results were necessary to confirm the efficiency of the approach and to identify the dependence of the efficiency of the proposed algorithm for sequential refinement of the scenario compared to a complete analysis of the graph.

In preparing the examples for testing and analysis, two groups of examples were considered, which are based on different in structure, but similar in construction, approaches.

The number of comparisons required to select the optimal path on the graphs, which is determined by the number of nodes at each level of the hierarchical structure of the test example, was used to quantify the efficiency of the algorithm. At the same time, each of the comparisons was evaluated at random from 1 to 10 points. Thus, the possible time of processing at each node was modelled.

Two models of example generation in two directions were considered. First, to analyse the number of nodes at each level of the hierarchical structure, and secondly, to analyse the dependence of the number of comparisons on the number of levels in the hierarchical structure in the generated examples of graphs.

According to the structure, all generated graphs had at least one possible optimal path and a number of paths that did not lead to the final solution of the task of building a scenario.

Fig. 3 presents the results of constructing the optimal scenarios on test examples with averaging the results on one hundred test examples. Examples with the number of levels 4 and 10 and the number of nodes at each level 5, 10, 15, 20, 25 and 30 were considered.

Analysis of the results shows a significant reduction in costs when using standard scenarios (indicated in the chart of the typical scenario). In addition, as can be seen from the figure, the advantage of using an algorithm based on a typical scenario increases significantly with the number of nodes at each level. Thus, the more complicated the structure of the graph, which is the basis for the construction of the scenario, the more efficient the use of consistent refinement of the scenario is.

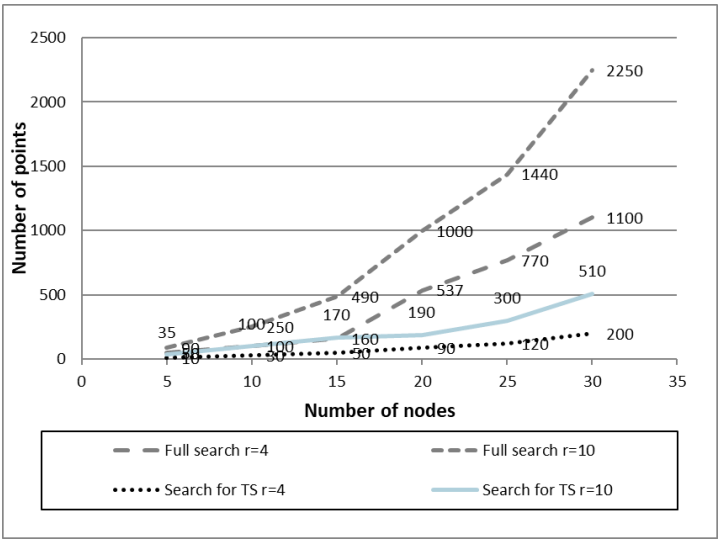


Fig. 3. Comparison of the results of using the algorithm of sequential refinement of the scenario with a complete analysis of the graph by the number of nodes.

According to a similar scheme, the second part of testing was conducted on a group of hundred of test examples, which were built to identify and confirm the relationship between efficiency between the number of levels and scores for the algorithm for sequential refinement of the scenario compared to full graph analysis. The test results are shown at Fig. 4.

Examples with the number of nodes 10 and 20 and the number of nodes at each level 4, 8, 12, 16, 20 and 24 were considered, similarly to the previous testing and analysis. The number of points in the generated examples was also determined randomly from 1 to 10. Only those examples were considered that had at least one path, which makes it possible to build the necessary scenario.

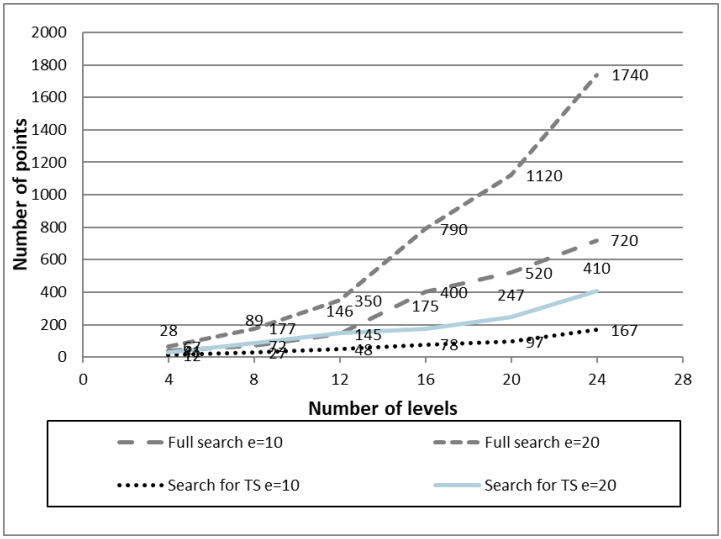


Fig. 4. Comparison of the results of using the algorithm of sequential refinement of the scenario with a complete analysis of the graph by the number of levels.

Conclusions

The analysis of the results shows a significant reduction in costs when using typical scenarios (indicated in the graph of the typical scenario), which gives the advantage of using the algorithm based on the typical scenario with increasing levels. Increasing the complexity of the structure of the graph, which is the basis for building a scenario, shows higher efficiency in the use of consistent refinement of the scenario.

Test results show that using such an approach to improving the quality of scenarios based on the use of typical scenarios can significantly reduce the complexity and cost of building new and more efficient scenarios according to certain criteria. This makes it possible to significantly simplify analytical activities using standard scenarios to improve the efficiency of analytical activities. The developed approach can be used in the development of information and analytical systems.

References

1. Додонов А.Г., Сенченко В.Р., Коваль А.В. Аналитика и знания в компьютерных системах. Киев: ИПРИ НАН Украины, «КПИ имени Игоря Сикорского», 2020. 315 с.
2. Koval O.V., Kuzminykh V.O., Svistunov S.Y., Xu Beibei, Zhu Shiwei. Data collection for analytical activities using adaptive microservice architecture. Реєстрація, зберігання і обробка даних. 2021. Т. 23, № 1. С. 11–31.

3. Черняк, Леонид Большие Данные — новая теория и практика// Открытые системы. СУБД. — М.: Открытые системы, 2011. — № 10. — ISSN 1028-7493
4. Oleksandr Koval, Valeriy Kuzminykh and Maksym Voronko. «Standard Analytic Activity Scenarios Optimization based on Subject Area Analysis», in Selected Papers of the XIX International Scientific and Practical Conference «Information Technologies and Security» (ITS 2019) CEUR Workshop Proceedings. 2019. Vol. 2577. P. 37–46.
5. Moore, Edward F. “The Shortest Path Through a Maze,” International Symposium on the Theory of Switching, 285–92, 1959.
6. Lee, C Y. “An Algorithm for Path Connections and Its Applications.” IEEE Transactions on Electronic Computers 10, no. 3 (September 1961): 346–65.
7. O. Koval, V. Kuzminykh, S. Otrokh and V. Kravchenko. «Optimization of Scenarios for Collecting Information Streaming Wide-Area Network», 3rd International Conference on Advanced Information and Communications Technologies (AICT) (2–6 July, 2019, Lviv). Lvyv, 2019. P. 213–215.

МЕТОД ВИЗНАЧЕННЯ ПОЛІТИЧНОГО РЕЙТИНГУ ЗА ДОПОМОГОЮ СОЦІАЛЬНИХ МЕРЕЖ

Рудник Т. П. [1][0000-0001-9492-0374], Чертов О. Р. [1][0000-0003-0087-1028]

[1] Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського", м. Київ, Україна
{tarasrudnykoleg.r.chertov}@gmail.com

В даних тезах наведено порівняльний аналіз соціальних мереж для дослідження рейтингів українських політиків. Проведено експеримент з визначення змін рейтингу чинного Президента України – Володимира Зеленського та аналіз отриманих результатів.

Вступ

В сучасному світі соціальні мережі набувають все більшої популярності. Вони вже давно вийшли за рамки лише розважального наповнення. На сьогоднішній день соціальні мережі часто використовують для просування бізнесу, маркетингу, популяризації особистого бренду і як засіб впливу на людей. Саме тому майже кожен політичний діяч має сторінку в Facebook, Twitter та Instagram.

Соціальні опитування, зазвичай, використовуються для визначення політичних рейтингів та прогнозування результатів виборів. Не зважаючи на те, що даний підхід позитивно зарекомендував себе за довгі роки, у нього є суттєві недоліки, такі як: вартість, великий обсяг затраченого часу, затримка між початком збору даних та отриманням кінцевих результатів. Крім того, результати соціологічних опитувань можуть бути спотворені зловмисниками. Автоматизація процесу визначення політичних рейтингів могла б нівелювати наведені недоліки та дозволити отримувати результати регулярно і майже в реальному часі. І, звичайно, підкупити алгоритм або комп'ютер неможливо та легко перевірити чи було втручання в його роботу.

В даних тезах пропонується підхід до виявлення політичного рейтингу чинного Президента України – Володимира Зеленського за допомогою аналізу методами машинного навчання його сторінки в соціальній мережі Twitter.

Соціальні мережі в Україні та світі

Згідно з даними за жовтень 2021 року, наданими Statista [1], Facebook налічує 2,895 мільярди активних користувачів у світі, Instagram – 1,393 мільярди, а Twitter – 463 мільйони. Research & Branding Group провела дослідження використання соціальних мереж в Україні з 20 січня по 1 лютого 2021 р., в результаті якого виявлено, що Facebook є найбільшою соціальною мережею в Україні, яку використовують 59% людей, Instagram – 30%, Twitter – 6%. Число користувачів російських соціальних мереж з

кожним роком зменшується і становить 5% для Однокласників та 3% для Вконтакті [2].

Збір даних з Facebook дуже сильно обмежений та отримання більшості персональних даних суворо заборонене. Instagram, перш за все, використовується для поширення світлин та відео, ніж для текстових повідомлень та впливу на людей. Російські Однокласники та Вконтакті, зазвичай, використовуються лише проросійськими кандидатами та виборцями. У той же час Twitter надає багато можливостей для збору даних з використанням відкритого прикладного програмного інтерфейсу. Відомі українські політики (наприклад, Володимир Зеленський та Петро Порошенко) використовують цю соціальну мережу для спілкування як з співвітчизниками, так і з міжнародними партнерами. Саме тому для подальшого дослідження було обрано соціальну мережу Twitter.

Опис алгоритму

Для виявлення політичного рейтингу Володимира Зеленського в період з 07.11.2020 по 28.02.2021 р. збиралися унікальні ідентифікатори всіх передплатників чинного Президента України один раз на тиждень. У період дослідження відписалося зі сторінки Володимира Зеленського 8 289 облікових записів. Наступним кроком було зібрано 639 960 повідомлень даних облікових записів, написані з початку президентської компанії Володимира Зеленського, яка почалася у січні 2019 році. Таким чином, були отримані повідомлення за період з 31.01.2019 по 31.03.2021 р. Проаналізувавши їх за допомогою методів машинного навчання та після агрегації помісячно, вдалося отримати результати близькі до наданих соціологічною групою «Рейтинг» [3]. Отримані результати зображено на рисунку 1, де зеленим кольором відмічені результати соціального опитування, а червоним ―отримані в результаті експерименту.

Висновки

В даному дослідженні запропоновано метод автоматичного виявлення політичного рейтингу за допомогою соціальної мережі Twitter. Підхід базується на аналізі передплатників та тих, хто відписався, а також їх повідомлень. Експериментально отримані результати на сторінці Володимира Зеленського підтверджують, що даний метод може надати значення,близькі до даних соціологічних опитувань. Для подальшого покращення результатів необхідно враховувати додаткові параметри, а саме: географічне розташування, вік, матеріальний стан передплатника. Також має сенс ввести вагові коефіцієнти з поправкою кількості людей, які користуються Twitter відносно всього населення України, оскільки даною соціальною мережею переважно користується молодь та люди з великих

міст. Якщо у політичного діяча є історія участі у попередніх виборах, то теж варто враховувати її при оцінці поточного рейтингу.

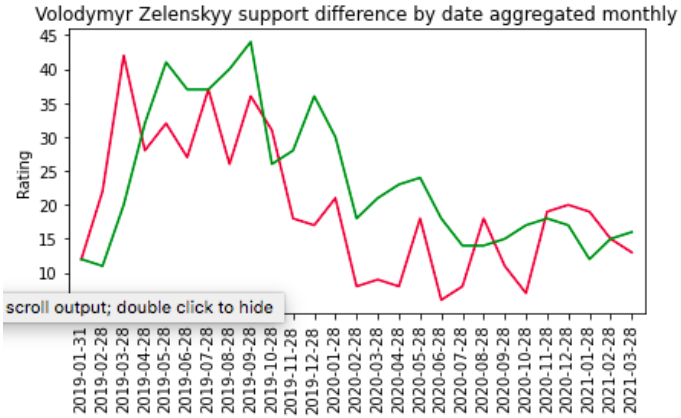


Рис. 1. Підтримка Володимира Зеленського по місяцях. Соціальне опитування (зеленим) та експериментальні результати (червоним)

Перелік посилань

1. Statista. Most popular social networks worldwide as ofJanuary 2021, rankedbynumberofactiveusers. Availableat: https://www.statista.com/statistics/272014/global-social- networks-ranked-by-number-of-users/ (AccessedJune 06, 2021).
2. Research & Branding Group. Social networks andmessengersinUkraine (inUkrainian). Availableat: http://rb.com.ua/uk/blog-uk/omnibus-uk/socialni-merezhi-ta-mesendzheri-v- ukraini/ (AccessedJune 06, 2021).
3. Sociologicalgroup “Rating”. Volodymyr Zelenskyy rating chart. Availableat: http://ratinggroup.ua/files/ratinggroup/reg_files/rg_ukraine_covid_cati_ix_wave_022021_press.pdf (AccessedJune 06, 2021)

РОЗПОДІЛЕНІ ІНТЕЛЕКТУАЛЬНІ АГЕНТИ ДОБУВАННЯ КОНТЕНТУ ІЗ СОЦІАЛЬНИХ МЕРЕЖ

А.М. Соболєв[1], Д.В. Ланде[1,2]
[1] Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського", Київ, Україна
[2] Інститут проблем реєстрації інформації НАН України, Київ, Україна

dwlande@gmail.com

Запропоновано розподілені агенти добування контенту із соціальних мереж, які складається з декількох серверів, що проводять збір інформації і розміщені в різних дата центах, які контролюються єдиною інтелектуальною системою, яка забезпечує належний рівень відмовостійкості та достовірності отриманої інформації.

Останнім часом соціальні мережі стали зручним та ефективним засобом комунікації. Вони надають величезну свободу дій в інформаційному просторі, який є відкритим та доступним для всіх бажаючих. При ефективному їх використанні вони стають потужним інструментом для проведення OSINT (Open-Source Intelligence) та водночас дають можливість отримання стратегічно важливої, але загальнодоступної інформації у питаннях національної безпеки та критичної інформації, що дозволяє проводити оцінку настрою суспільства у визначеному інформаційному полі.

Популярні соціальні мережі зберігають та контролюють величезний об'єм даних про повсякденне життя та соціальні взаємодії людей і тому вони прискіпливо перевіряти кожен запит, що приходить до них та не дозволяють стороннім сервісам без їх згоди копіювати таку інформацію собі. У разі коли ці сервіси помічають нестандартну поведінку від клієнта, що робить запит до їх даних вони негайно заблоковують йому доступ та надалі прискіпливо перевіряють запити, що приходять з IP Address заблокованого клієнта. Ці дії провокують до використання певної технології у таких клієнтів, щоб забезпечити стандартний рівень поведінки, що притаманний звичайному користувачу їх сервісів.

Для забезпечення можливості одночасного процесу добування інформації з соціальних мереж без використання сторонніх платних сервісів та для контролю і управління такою системою з єдиного місця, запропоновано впровадити команди агентів, що дозволяють завантажувати дані та обмінюватись такою інформацією між собою та забезпечує цілісність отриманих даних і розподіляє навантаження між собою.

Основою для контролю, управління, синхронізації навантаження та сумісної роботи агентів використано документо-орієнтовану систему керування базами даних MongoDB, що дозволяє зберігати інформаціє про запуски агентів, список їх джерел та алгоритм їх поведінки. NoSQL база

даних MongoDB дозволяє з легкістю впроваджуватись у найбільш популярні операційні системи, невибаглива до ресурсів та витримує велике навантаження

Оскільки системи управління доступом в популярних соціальних мережах націлені на виявлення нестандартної поведінки користувачів, що тим самим призвело то створення ефективного алгоритму доступу агента до таких мереж, який базується на кількості часу проведеного в цих мережах, об'ємі інформації, що збирається за один запит та кількості інформація, що надається таким агентом. Також слід відмітити, що основною увагою таких сервісів є поведінка клієнтів в нічний час (із регіону з якого відбувається запит), оскільки в такий період кількість запитів від клієнтів має мінімізуватись а в іншому випадку такий клієнт вже є об'єктом для дослідження.

Для розподіленої взаємодії використано 3 сервери, що територіально розміщені в різних дата центрах та знаходяться на великих відстанях між собою:

1. Нідерланди;
2. Україна;
3. Сполучені Штати Америки.

На даних серверах Рис. 1 розгорнуто MongoDB кластер, інтелектуальну систему управління та команди агентів для добування інформації. Протоколом для керування та взаємодії між агентами використовується протокол HTTPS, оскільки він являється найбільш популярним в глобальній мережі Інтернет, дозволяє швидко оптимізувати команди для мережевих агентів та має належний рівень безпеки.

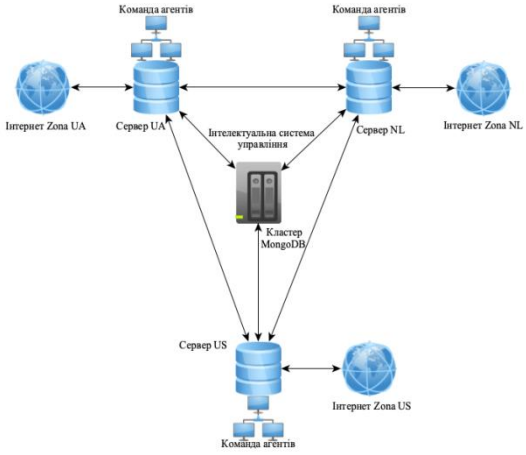


Рис. 1 – Схема розподіленого добування інформації на основі 3 серверів

Також дані команди агентів використовують в основі архітектуру RESTful сервісу, що дозволяє ефективно проводити налаштування та їх контроль роботи. Також, в основі їх взаємодії використовуються системні повідомлення типу Heartbeat, що дозволяють ефективно досліджувати життєвий цикл агентів і у разі відмови будь-якого з них швидко виявити проблему без втрати отриманих ним даних.

Запропоновані команди агентів ∇ являють собою кластер високої доступності, у якому при відмові одного агента його функції переймає інший доступний агент. Таким чином, процес добування інформації із соціальних мереж продовжує працювати без зупинки завдяки інтелектуальній системі контролю цих агентів. Щоб побудувати відмовостійку структуру, потрібно мінімум два фізичні сервери із системами зберігання даних, тому для забезпечення відмовостійкості у двух серверів використовується допоміжний третій сервер, який дозволяє ефективно використовувати ресурси та рівномірно розподіляти навантаження між двума серверами. Також, інтелектувальна система контролю за мережевими агентами працює за таким принципом: якщо один з агентів виходить з ладу, в роботу автоматично включається інший при цьому йде оповіщення про збій агента. Загальна логіка кластера агентів створюється лише на рівні програмних протоколів і дає можливість:

1. Керувати всіма мережевими агентами за допомогою одного інтелектуального модуля;
2. Додавати та вдосконалити програмні та апаратні ресурси, без зупинки системи та масштабних архітектурних перетворень;
3. Забезпечувати безперебійну роботу системи при виході з ладу одного або двух агентів;
4. Синхронізувати дані між кластерами агентів;
5. Ефективно розподіляти запити на кластери агентів.

Один із серверів, що входять до складу кластера, є центральним сервером кластера. Центральний сервер, крім обслуговування клієнтських з'єднань, управляє роботою всього кластера і зберігає при цьому реєстр кластера.

Під час встановлення з'єднання агент звертається до центрального сервера кластера. Центральний сервер, на основі аналізу статистики завантаженості агентів, спрямовує його до конкретного робочого процесу, який йому необхідно виконати.

Отже головним завданням кластера агентів є виключення простою системи та надання звітів, скільки агенти зібрали самі, а скільки взяли в інших агентів. В ідеалі будь-який інцидент, пов'язаний із зовнішнім втручанням або внутрішнім збоєм у роботі ресурсу, повинен дозволяти продовжувати роботу системи.

Висновки

На основі проведеного тестування та оцінки роботи даних мережевих агентів, що добувають контент із соціальних мереж та складаються з 3 серверів, можна зробити висновок, про ефективність запропонованої взаємодіє оскільки це надало системі належний рівень відмовостійкості та достовірності отриманої інформації і забезпечило рівномірність завантаження кластерів агентів. Також представлена система виконує задачі безпеки сервісу моніторингу, що дозволяє забезпечити цілісність отриманих даних, обходячи обмеження у зборі інформації, доступність даних для моніторингу і повноту отриманої інформації. У разі, якщо для якось країни інформація буде змінена, агенти при взаємодії між собою відобразлять ці зміни і збережуть всі копії отриманих даних.

ПРИНЯТИЕ РЕШЕНИЙ ОРГАНИЗАЦИОННОГО УПРАВЛЕНИЯ НА ОСНОВЕ ОНТОЛОГИИ ДЕЯТЕЛЬНОСТИ

А.В. Никифоров[1], А.Г. Додонов[2], В.Г. Путятин[2]
[1] Харьковский национальный университет Воздушных Сил
[2]Институт проблем регистрации информации Национальной академии наук Украины, г. Киев

Конечной целью процесса выработки решения по управлению организацией для проведения какого-либо манёвра является: формирование перечня задач для подразделений; определение времени и очерёдности выполнения поставленных задач; распределение имеющихся ресурсов по задачам (подразделениям); установление порядка функционирования подразделений при решении поставленных перед ними задач. По степени детализации вырабатываемое решение охватывает два нижерасположенных иерархических уровня рассматриваемой системы управления. То есть решение прорабатывается вплоть до цеха (подразделения), отдельного производства, отдельной мастерской (боевого расчёта, экипажа). При этом, как правило, учитывается фактор персонификации (личные интересы (предпочтения) и возможности (компетентность) исполнителей). При этом, перечисленные параметры решения, как правило, определяются на основании единовременной (одношаговой) управленческой процедуры со стороны лица, принимающего решение (ЛПР). Такие приёмы, например, как пошаговое приближение к искомому результату путём решения последовательности (каскада) экстремальных задач, в большинстве случаев не применимы. Все составляющие решения имеют сильную взаимозависимость. Невозможно варьировать какую-либо отдельную часть системы, заморозив её окружение. Решение принимается сразу по всем аспектам, во всей полноте представления процессов управления.

Одним из возможных методов решения этой проблемы выступает автоматизация процессов организационного управления. Это обуславливает актуальность научной задачи по осуществлению формализованного представления процессов организационного управления.

Развитие моделей и методов организационного управления осуществляется по таким трём магистральным направлениям, как: детализация [1] – [3], учёт фактора активности (субъектности) исполнителей [4], [5] и адаптация [6] – [8].

Решения, принимаемые в трёхзвенной системе управления, имеют иерархическую структуру. Верхний (третий) иерархический уровень представлен действиями, A_i, i=1, … , H, которые раскрывают замысел организационного манёвра (операции). В общем случае, замысел представляет собою некоторую структуру, которая разворачивается во времени. Средний (второй) уровень представлен действиями, которые детализируют замысел, $B_{i,j}$, i=1, … , H, j=1, … , N_i. Это задачи отдельных субъектов корпорации, предприятий, подразделений. Нижний (первый)

уровень – это действия отдельных исполнителей, отдельных технических комплексов, этапы выполнения отдельных задач второго уровня, $C_{i,j,k}$, i=1, … , H, j=1, … , N_i, k=1, … , $M_{i,j}$. Здесь H, N_i, $M_{i,j}$, i=1, … , H, j=1, … , N_i – количество действий, составляющих замысел, i-ое действие второго уровня и ij-ое действие первого уровня, соответственно.

Сформированный план действий можно записать как двойной булеан множества бинарных отношений действий замысла A_i, i=1, … , H и дискретных моментов времени t_τ, τ=1, … , T

$$P = \text{B}(\text{B}(A \times A) \times t), \quad (1)$$

где P – сформированное решение (план).

В свою очередь каждое действие замысла есть двойной булеан бинарных отношений действий второго уровня и дискретных моментов времени

$$A = \text{B}(\text{B}(B \times B) \times t). \quad (2)$$

Для действий второго уровня, по аналогии с (1) и (2), можно записать

$$B = \text{B}(\text{B}(C \times C) \times t). \quad (3)$$

С учётом (1) – (3), сформированное решение может быть записано как

$$P = \text{BBB}(\text{B}(C \times C) \times t). \quad (4)$$

Булеан бинарных отношений действий первого уровня (этапы или элементы выполнения задач субъектов корпорации, действия отдельных исполнителей и технических комплексов), □(C×C), есть онтология деятельности организации.

Таким образом, процесс принятия решения выглядит как трёхкратная группировка элементов онтологии и дискретных моментов времени. То есть, говоря о формализованном представлении процессов управления организациями, следует выделить задачи: разработки процедур группировки элементов онтологии, а также – синтеза, собственно, онтологической сети.

На основании полученных результатов по проектированию АСУ организационного управления следует выделить следующие актуальные задачи для дальнейших научных исследований. Это, во-первых, задача трансформации пространства принятия решений в соответствии с меняющимися аспектами управления. Во-вторых, задача конструирования (изменения) системы процедур, обеспечивающих количественную оценку альтернатив в данном пространстве. В-третьих, задача синтеза альтернативных решений из элементов онтологии деятельности.

Трансформация пространства принятия решений в соответствии с аспектами управления

Чётко заданное пространство принятия решений позволяет формализовать, а значит и автоматизировать, такие элементы кризисного управления, как:

– классификация гипотетических состояний для реализуемого управленческого манёвра или, иначе – генерирование вариантов альтернатив для принятия решений;

– количественная оценка альтернатив при принятии решения;

Трудностью, применительно к построению пространства принятия решений, является получение предельно обобщённых характеристик описания гипотетических управленческих ситуаций. Данные характеристики должны быть инвариантны к изменениям аспектов управления, а также должны позволять их разворачивание (детализацию) с учётом особенностей (аспектов) реализуемого управленческого манёвра.

Получение предельно обобщённых характеристик для рассматриваемых процессов управления есть нетривиальная задача, которая требует, для эффективного своего решения, большого практического опыта в конкретной области деятельности. Так, в области военной стратегии, А.А. Свечиным [9] в качестве таких обобщённых параметров выделялись: пространство; время; сила (военная сила). Для социально-экономических систем, П.Г. Кузнецовым в [10], как предельно-обобщённые характеристики, рассматривались: мощность производства, мощность производства энергии; эффективность использования производимого продукта; скорость изменения социально-необходимого времени. В [10] отмечалось, что инвариантом, по отношению к различным системам координат (аспектам управления), является мощность. Показатели, имеющие в своей основе фактор мощности, есть инвариант при полиаспектном управлении, они представляют предельно-обобщённое описание пространства принятия решений. Следовательно, исследования по построению пространства принятия решений должны лежать в русле интерпретации понятий «мощность», «изменение мощности», «скорость передачи мощности» по отношению к автоматизируемой деятельности. Это даст возможность сформировать исходное пространство, инвариантное к учитываемым аспектам управления.

Далее, для конкретизации или детализации обобщённых (инвариантных) показателей для управления по установленному аспекту, целесообразно использовать метод категориального анализа [11].

Конструирование (изменение) системы процедур, обеспечивающих количественную оценку альтернатив в актуальном пространстве принятия решений

Используемые в процессе принятия решений формализованные процедуры должны быть актуальны: содержанию и характеру управляемого процесса, а также цели и учитываемым аспектам управления. Это касается процедур «восстановления» текущего состояния и формирования команд (распоряжений).

Задача синтеза процедурного регламента (ПР) может быть записана как задача отыскания перечня и очерёдности активации формализованных процедур из процедурной библиотеки по критерию минимизации «пути» и при выполнении условия определения значений параметров из установленного перечня. Для решения данной задачи могут быть применены известные алгоритмы нахождения оптимального пути. Также могут быть использованы методы тензорного преобразования сетей [12]. Для приведения задачи синтеза ПР к задаче тензорного преобразования сети,

исходную структуру моделей и параметров пространства принятия решения следует интерпретировать в виде электрической сети. Электрическая сеть удобна для формализованного описания структурных связей различного рода. В отличие от других типов неэлектрических сетей, электрическая сеть всегда окружена динамическим электромагнитным полем, создаваемым ею самой и простирающимся до бесконечности во всех направлениях. Используя модель индукции и самоиндукции в ветвях электрической сети, удобно описывать различного рода внутрисистемные связи между элементами рассматриваемой структуры (организационной топологии).

На этой основе Г. Кроном была предложена методология теоретико-множественного и алгебраического синтеза топологий в форме метода анализа и тензорного преобразования электрических сетей [12].

То есть, если интерпретировать ПР в виде многокатушечных электрических сетей – возможно использовать тензор преобразования Крона как механизм синтеза системы процедур с заданными свойствами (установленным перечнем параметров входа и выхода).

Задача синтеза процедурного регламента (ПР) состоит в исключении из исходной сети (библиотеки процедур) тех её фрагментов, которые не используются для определения актуальных параметров.

Тензорные преобразования, с помощью которых формируется ПР, заключаются в следующем. Сначала определяются параметры геометрических объектов исходной сети (тензор взаимосоединений, векторы узловых напряжений и узловых токов, тензор адмиттанса) [12]. Затем, для полученных характеристик геометрических объектов исходной сети, вводятся формализованные условия, которые определяют потребный ПР. На основании решения образованной системы линейных уравнений формируется искомый ПР.

Для этого необходимо разделить катушки сети (узловые пары) на три условные группы:

- группу узловых пар, которые интерпретируют входные параметры. То, что приходит с внешних источников информации;
- группу узловых пар, которые интерпретируют выходные (целевые) параметры;
- группу промежуточных узловых пар, а также узловых пар, не вошедших в перечень целевых параметров. Промежуточные узловые пары интерпретируют формализованные процедуры или параметры, которые являются промежуточными при преобразовании данных.

Каждой группе узловых пар присваивается порядковый номер в порядке перечисления этих групп. Узловые токи (токи, прикладываемые к узлам при разрыве сети) в группе промежуточных узловых пар равны нулю. На узловых парах остальных групп (входные и целевые параметры) напряжения и токи не равны нулю. Для сгруппированных таким образом узловых пар исходной сети может быть записана следующая система уравнений

$$\begin{cases} \boldsymbol{I_1} = \boldsymbol{Y^{11}}\,\boldsymbol{E_1} + \boldsymbol{Y^{12}}\,\boldsymbol{E_2} + \boldsymbol{Y^{13}}\,\boldsymbol{E_3}\,, \\ \boldsymbol{I_2} = \boldsymbol{Y^{21}}\,\boldsymbol{E_1} + \boldsymbol{Y^{22}}\,\boldsymbol{E_2} + \boldsymbol{Y^{23}}\,\boldsymbol{E_3}\,, \\ \boldsymbol{I_3} = \boldsymbol{Y^{31}}\,\boldsymbol{E_1} + \boldsymbol{Y^{32}}\,\boldsymbol{E_2} + \boldsymbol{Y^{33}}\,\boldsymbol{E_3}\,, \end{cases} \quad (5)$$

или, с учётом равенства нулю приложенных узловых токов в промежуточных узловых парах:

$$\begin{cases} \boldsymbol{I_1} = \boldsymbol{Y^{11}}\,\boldsymbol{E_1} + \boldsymbol{Y^{12}}\,\boldsymbol{E_2} + \boldsymbol{Y^{13}}\,\boldsymbol{E_3}\,, \\ \boldsymbol{I_2} = \boldsymbol{Y^{21}}\,\boldsymbol{E_1} + \boldsymbol{Y^{22}}\,\boldsymbol{E_2} + \boldsymbol{Y^{23}}\,\boldsymbol{E_3}\,, \\ 0 = \boldsymbol{Y^{31}}\,\boldsymbol{E_1} + \boldsymbol{Y^{32}}\,\boldsymbol{E_2} + \boldsymbol{Y^{33}}\,\boldsymbol{E_3}\,, \end{cases} \quad (6)$$

где $\boldsymbol{Y^{ij}}$ – матрица адмиттансов сформированных групп катушек; $\boldsymbol{E_i}, \boldsymbol{I_i}$ – векторы напряжений и токов на катушках *i*-ой группы.

В записанной системе уравнений тензоры $\boldsymbol{Y^{ij}}$ содержат элементы тензора адмиттанса примитивной сети, $\boldsymbol{Y}$, сгруппированные по определённым признакам. Например, $\boldsymbol{Y^{11}}$ содержит элементы тензора адмиттанса примитивной сети, характеризующие взаимовлияние катушек первой группы.

На основании сформированной системы уравнений может быть вычислен вектор напряжений на узловых парах третьей группы. Те узловые пары, для которых будет установлено нулевая разность потенциалов, должны быть исключены из исходной сети формализованных процедур. То есть будет сформирован ПР, удовлетворяющий оговоренным свойствам по определению целевых параметров. Искомый вектор напряжений (признаков включения-исключения) узловых пар третьей группы рассчитывается как

$$\boldsymbol{E_3} = -\boldsymbol{Y^{31}}\,\boldsymbol{Y^{33}}^{-1}\boldsymbol{E_1} - \boldsymbol{Y^{32}}\boldsymbol{Y^{33}}^{-1}\boldsymbol{E_2}\ . \quad (7)$$

Тензор взаимовлияния катушек третьей группы, $\boldsymbol{Y^{33}}$, является квадратной матрицей. Следовательно, существует обратная матрица $\boldsymbol{Y^{33}}^{-1}$. Векторы напряжений входных и выходных параметров, $\boldsymbol{E_1}$ и $\boldsymbol{E_2}$, соответственно, известны из требований, которые предъявляются к синтезируемому ПР. Записанное уравнение имеет решение.

Синтез альтернативных решений из элементов онтологии деятельности

Задача синтеза альтернативных решений из элементов онтологии деятельности также может быть решена с помощью методов тензорного преобразования электрических сетей при условии, что имеется библиотека (онтология) функционирования подчинённых центров нижнего уровня (элементы отдельных технологий, этапы выполнения конкретных задач, режимы работы комплексов (средств) и так далее). Используя подход тензорных преобразований сетей, каждый элемент онтологии может быть представлен как совокупность двух узлов, (C_i; C_j), соединённых проводником с индуктивностью и соответствующим импедансом, $z_{i,j}$. Узлы – это действия (состояния), детализированные до первого (нижнего) иерархического уровня деятельности. Например, узел C_i может интерпретироваться как начало взлёта летательного аппарата. Узел C_j – как завершение набора установленной высоты. Тогда элемент (C_i; C_j) – есть

процесс набора высоты. Импеданс, $z_{i,j}$, характеризует трудоёмкость операции по осуществлению перехода из первого во второе состояние. Импеданс рассматриваемого элемента в общем случае зависит от направления тока. Имеют место случаи, когда течение тока возможно только в одном направлении. Импеданс для запрещённого направления тока принимает бесконечно-большое значение. На один из узлов подаётся мгновенное напряжение. В зависимости от величины электрического напряжения, поданного на первый узел, а также от величины импеданса, в проводнике возникает мгновенный электрический ток определённой силы, $i_{i,j}$. На втором узле возникает определённое мгновенное напряжение.

Задача синтеза вариантов действий подчинённых центров представляется как задача отыскания тензора соединений (преобразования) примитивной электрической сети в сеть с заданными свойствами (вектор напряжений и вектор токов на элементах сети). Заданные свойства определяется на основании формализованного описания целей и задач организации. Такое формализованное описание возможно при условии существования формализованного представления пространства принятия решений.

Таким образом, автоматизация организационного управления имеет ряд проблем, связанных с плохой структурированностью и многоаспектностью задач управления. Это обуславливает актуальность задач разработки научно-методического аппарата проектирования полиаспектных АСУ.

Развитие математических моделей и методов, обеспечивающих формализованное описание процессов принятия решений в условиях изначальной неопределённости (многоаспектности) управления, происходит по трём направлениям: усиление степени детализации до практически значимых уровней; учёт фактора субъектности центров принятия решений нижнего уровня; развитие свойств быстрой адаптации управления к меняющимся управленческим аспектам.

Реализация организационного управления на основе онтологической сети деятельности организации даёт ряд преимуществ в отношении усиления детализации и адаптивности формируемых управленческих решений. Процесс формирования решения при этом выглядит как трёхкратная группировка элементов онтологии и дискретных моментов времени. При этом являются актуальными такие научные задачи, как:

- формирование пространств принятия решений, определяемых на наборах предельно-обобщённых характеристик, инвариантных к различным аспектам управления;
- конкретизация предельно-обобщённых параметров под актуальный аспект управления;
- конструирование (реконфигурация) системы вычислительных процедур под сформированное пространство принятия решения;
- синтез альтернативных решений из элементов онтологии.

Литература

1. Канторович Л.В. Математические методы организации и планирования производства / Л.В. Канторович – Л.: Изд-во ЛГУ. – 1939. – 68 с.
2. Немчинов В.С. Экономико-математические методы и модели / В.С. Немчинов – М., Мысль, 1965. – 253 с.
3. Янг С. Системное управление организацией Пер. с англ. под ред. С.П. Никанорова, С.А. Батасова / С. Янг. – М.: «Советское радио». – 1972. – 456 с.
4. Landauer C. (1994) Theory of national economic planning. Berkeley (Cal.): University of California Press, 274. (1944 first edition)
5. Бурков В.Н. Механизмы функционирования организационных систем / В.Н. Бурков, В.В. Кондратьев. – М.: Наука, Главная редакция физико-математической литературы. – 1981. – 384 с.
6. Beni G. Swarm Intelligence in Cellular Robotic Systems, Proceed. NATO Advanced Workshop on Robots and Biological Systems / G. Beni, J. Wang, Tuscany, Italy, 1989. P. 26–30
7. Lewis F.L. Cooperative Control of Multi-Agent Systems: Optimal and Adaptive Design Approaches (Communications and Control Engineering) // F.L. Lewis, H. Zhang, K. Hengster-Movric, A. Das. – Springer, 2014. – 307 p.
8. Nikiforov A. Modeling Complexes of Organizational Management Automated Systems - a Means to Overcome the Management Crisis / Dodonov A., Nikiforov A., Putyatin V., Dodonov V. // Selected Papers of the XIX International Scientific and Practical Conference "Information Technologies and Security" (ITS 2019) Kyiv, Ukraine, November 28, 2019. ISSN 1613-0073 – Vol-2577 urn:nbn:de:0074-2577-8, p. 100 – 115
9. Свечин А. А. Стратегия. – М.-Л.: Госвоениздат, 1926. – 400 с.
10. Гвардейцев М.И. Математическое обеспечение управления. Меры развития общества / М.И. Гвардейцев, П.Г. Кузнецов, В.Я. Розенберг. – М.: «Радио и связь». – 1996. – 176 с.
11. Nikiforov A. Applying the method of categorical analysis for conceptual design of an automated control system of a group of unmanned aerial vehicles / Nikiforov A., Klushnikov I. // ISAIC 2020. IOP Publishing. Journal of Physics: Conference Series. Ser. 1828 (2021) 012069. doi:10.1088/1742-6596/1828/1/012069, 17 p.
12. Kron G. Tensor analysis of networks. John Wiley and sons, inc. London: Capman and Hall, Limited. New York. – 1966. – 720 p.

ВИКОРИСТАННЯ МЕРЕОЛОГІЧНОГО ПІДХОДУ ДЛЯ ФОРМУВАННЯ СТРУКТУРИ СЕМАНТИЧНИХ ЗВ'ЯЗКІВ ТИПУ «ЧАСТИНА-ЦІЛЕ» МІЖ СТОРІНКАМИ СЕМАНТИЗОВАНОГО WIKI-РЕСУРСУ

Юлія Рогушина[1], Анатолій Гладун[2]
[1]Інститут програмних систем НАНУ, Київ, Україна
[2]Міжнародний науково-навчальний центр інформаційних технологій та систем НАН та МОН України, Київ, Україна

ladamandraka2010@gmail.com, glanat@yahoo.com

Проаналізовано сучасні тенденції розвитку відкритих інформаційних ресурсів Web. Розглянуто Wiki-технології як поширений засіб колективної розробки інформаційних ресурсів. Визначено, що одна з основних проблем, які виникають в процесі створення семантизованого Wiki-ресурсу зі складною гетерогенною структурою, – це побудова несуперечного та узгодженого набору семантичних властивостей, які дозволяють коректно відобразити зміст відношень між окремими Wiki-сторінками інформаційного ресурсу. Запропоновано використовувати елементи онтологічного та мереологічного аналізу, які дозволяють формалізувати характеристики семантичних відношень, що використовуються, і визначити їх підтипи, що є принциповим для колективної розробки ресурсів незалежними та територіально відокремленими користувачами з різними терміносистемами.
Розглянуто найбільш загальні типи відношень між поняттями реальних предметних областей, та визначено методологічний базис для формалізації характеристик цих зв'язків. Мереологія аналізує один з фундаментальних типів зв'язку між об'єктами, а також природним і штучним середовищами, – зв'язок частини і цілого, а таксономія – зв'язок "клас–підклас". Основні підтипи мереологічних зв'язків розглянуто на прикладах відношень між гаслами Великої української енциклопедії, портальна версія якої реалізована на основі семантизованих Wiki технологій.

Відкриті інформаційні ресурси Web та Wiki-технології

Сьогодні забезпечити на практиці спільне, вільне й відкрите використання інформації досить важко. Потрібний новий спеціалізований підхід, який би забезпечив баланс між відкритістю інформації й безконтрольністю її

поширення. Колективне збирання й аналіз інформації «Інтелект Відкритих Джерел» (Open Source Intelligence – OS-INT) [1] є одним з напрямків концепції відкритих джерел Open Source [2], який зі збільшенням кількості користувачів Web і зростанням їхньої кваліфікації стає усе більш значущим джерелом знань. До Open Source Intelligence відносяться соціальні мережі, Wiki-ресурси, блоги тощо, які створюють та поповнюють мільйони користувачів Web. Інформація в таких джерелах має певну структурованість, але вона досить часто буває суперечливою й неповною.
OS-INT базується на таких принципах спільної праці, як експертна оцінка, практика взаємного оцінювання експертами власної роботи, превалювання незалежної репутації над адміністративним контролем, вільний обіг продуктів і гнучкі рівні залежності й відповідальності. Оцінки тих спеціалістів, що мають високу репутацію, мають більшу вагу. Знання здобуваються в процесі уточнень і досягнення консенсусу. Сьогодні одним з найпоширеніших напрямів OS-INT є технологія Wiki [3]. Саме ця технологія нині активно використовується для розробки систем менеджменту знань у різних сферах, приміром, при створенні Вікіпедії [4]. Програмне забезпечення й контент, що міститься на Wiki-сторінках, найчастіше використовують ліцензію вільного поширення документації – GNU Free Documentation License (GFDL), що гарантує сумісність ліцензії з наявною сукупністю текстів GFDL. Це дає можливість мати один Wiki зі сторінками, що мають різну ліцензію.

Wiki-середовище – це гіпертекстове середовище (зазвичай – Web-сайт), яке дозволяє користувачам відносно легко створювати й модифікувати його контент. Контент такого сайту, який складається з окремих Wiki-сторінок та зв'язків між ними, називають *Wiki-ресурсом*. Його створення та редагування є колективним процесом. Оскільки процес перегляду й уточнення є публічним і безперервним, то немає принципової різниці між попередніми й фінальними версіями інформації. Використання принципу експертного перегляду відкритих джерел дає змогу Wiki-ресурсу поступово зростати як за кількістю статей, так і за їх глибиною та якістю внаслідок колективного внеску тих користувачів, що є експертами в окремих питаннях.

Під *Wiki-технологією* зазвичай розуміють таку технологію побудови Wiki-ресурсу, яка дає змогу відвідувачам брати участь у редагуванні його вмісту – виправляти помилки, додавати нові матеріали, не використовуючи спеціальні програми, не реєструючись на сайті й не вивчаючи мову HTML. Одним з поширених движків для створення і підтримки Wiki-ресурсів, що забезпечує семантизацію інформації, є Semantic MediaWiki (SMW). Це семантичне розширення MediaWiki – вільного програмного забезпечення з відкритим кодом.

Одна з основних проблем, що виникають в процесі створення семантизованого Wiki-ресурсу зі складною гетерогенною структурою, – це побудова набору семантичних властивостей, які дозволяють коректно відобразити зміст відношень між окремими Wiki-сторінками. Важливо уникнути як використання різних назв для відношень з однаковим змістом,

так і застосування однакових назв для відображення відношень з різною семантикою – це може бути викликано неоднозначністю термінології різних предметних областей (ПрО). Вбудовані засоби Semantic MediaWiki не дозволяють формально відобразити всі характеристики семантичних властивостей, і тому доцільно використовувати для рішення цієї проблеми елементи онтологічного та мереологічного аналізу.

Постановка проблеми

Мета роботи – розробка методів та засобів коректного та однозначного визначення семантики зв'язків (відношень) між Wiki-сторінками семантизованого інформаційного ресурсу (ІР), які з точки зору семантичних технологій можуть розглядатися як класи ти екземпляри класів онтології. Пропонується використовувати мереологічний аналіз як основу для більш чіткого та інтероперабельного визначення та диференціації різних видів відношень типу «частина-ціле».

Актуальність проблеми

Семантизація Wiki-технологій надає можливість для явного визначення відношень між Wiki-сторінками, але для кожної ПрО (та відповідно для того ІР, що описує цю ПрО) доцільно визначати фіксований набір відношень, специфічний саме для цієї області. Деякі відношення є однаковими для різних ПрО (наприклад, «клас-підклас»), деякі є специфічними для окремих ПрО та не мають визначених характеристик. Але існує група відношень типу «частина-ціле», які характерні для більшості ПрО, але містять різні підтипи з подібними характеристиками, але різною семантикою, визначення якої є важливою складовою опису різних ПрО [5]. Теоретичною основою для визначення таких підтипів є мереологія – науковий напрямок, який стосується формалізації опису відношень між частинами та цілим. У створенні сучасних енциклопедійних ІР, орієнтованих на функціювання у Web та взаємодію з іншими інтелектуальними застосуваннями, доцільно використовувати цей теоретичний базис для розширення функціоналу таких ІР.

Стан вивчення проблеми

Мереологія – це формальна теорія про частини і зв'язані з ними поняття, розроблена С.Лесьневським [6] . Об'єктом мереології є дослідження відношення «частина-ціле». Мереологія є частиною тріади дедуктивних теорій, що включає також прототетику й онтологію (у Лесьневського онтологія, на відміну від сучасного підходу Semantic Web, розглядається тільки як система з єдиним відношенням «є» – «is_a»). Відношення «частина-ціле» є винятково важливим тому, що воно утворює основу поняття системи, яке є центральним у науковому пізнанні. Система являє собою структурне з'єднання своїх елементів. Її базовою формальною характеристикою є те, що елементи не просто входять у систему, а входять у неї в результаті взаємодії з іншими елементами.

Найбільш загальні відношення в реальних ПрО – це зв'язок еквівалентності, таксономічний зв'язок, структурний зв'язок, зв'язок залежності, топологічний зв'язок, зв'язок причини і наслідку, функціональний зв'язок, хронологічний зв'язок, зв'язок подоби, умовний зв'язок і цільовий зв'язок. Однак не усі ці онтологічні відношення рівноцінні, у цьому наборі відношень можна виділити найбільш фундаментальні – таксономію і мереологію. Відношення називають фундаментальним, якщо на базі одного цього відношення може бути побудована формальна система, яка дозволяє виражати основні математичні поняття. Існує чотири фундаментальних відношення, на базі кожного з яких може бути побудована математика: відношення приналежності (теорія множин ZF і NF), відношення між функцією, її аргументом і результатом (теорія множин фон Неймана), відношення іменування (онтологія Лесьневського) і відношення «частина-ціле» (мереологія).

Якщо термін «таксономія» досить широко розповсюджений і не вимагає додаткового визначення, то термін «мереологія» у дослідженнях, пов'язаних з інформаційними технологіями, застосовується значно рідше і має потребу в додатковому поясненні. *Мереологія* (від грецького «частина» і «вивчення») — формальна теорія про частини і пов'язані з ними поняття [7]. Автор цієї теорії – польський математик С.Лесьневський, а її розвитком займалися К. Айдукевич, А. Тарський, Т. Котарбінський (львівсько-варшавська математична школа) [8]. Відношення «частина-ціле» є винятково важливим тому, що воно утворює основу поняття системи, яка часто використовується в сучасному науковому пізнанні. У мереології відношення «є частиною» (part-of) має наступні властивості: 1. Асиметрія: $(x << y) \Rightarrow (y$; 2. Транзитивність: $(x << y) \wedge (y << z) \Rightarrow x << z$.

Система являє собою структурне поєднання своїх елементів. Її базовою формальною характеристикою є те, що елементи не просто входять до системи, а входять до неї також результаті взаємодії з іншими елементами. Мереологія виходить за межі вивчення часткових відносин між елементами загальних систем. Вона займається також тими об'єктами, частини яких є релевантними цілому. Такі об'єкти ідентифікуються як екземпляри. Серед мереологічних відношень можна виділити сім окремих класів: 1. Компонент-Об'єкт; 2. Член-Колекція; 3. Частина-Маса; 5. Матеріал-Об'єкт; 5. Властивість-Діяльність; 6. Стадія-Процес; 7. Місцевість-Область.

Створюючи формальний опис ПрО, яку відображає Wiki-ресурс, необхідно чітко визначати, до якого типу відношень належать ті відношення, які встановлюються між термінами. Помилкова класифікація може негативно вплинути на якість інформації та призвести до її некоректної інтерпретації користувачами.

Основний зміст дослідження

Використання сучасних технологій керування знаннями у Web надає користувачам значно більше можливостей для аналізу та пошуку інформації. На жаль, навіть найрозвиненіші та популярні онлайнові енциклопедії практично не використовують методи та засоби подання знань, які вже стали фактично стандартом для інтелектуальних Web-застосувань. Вибір засобів семантизації Wiki-технологій є окремою науковою проблемою та має враховувати специфіку тієї конкретної задачі, для якої ці технології застосовуються [10]. Портальна версія Великої української енциклопедії (e-ВУЕ) використовувати технологічну платформу MediaWiki [11] з відкритим кодом та її семантичне розширення Semantic MediaWiki (SMW) [12] , що забезпечує формальне визначення та обробку змісту зв'язків між Wiki-сторінками з урахуванням стандартів Semantic Web [13]. Зараз у e-ВУЕ реалізовано велику кількість семантичних відношень різних рівнів та типів.

Онтологічна модель описує семантичні властивості і відношення інформаційних об'єктів (ІО), що зв'язані з результатами навчання, формалізує відношення між цими ІО і встановлює їхню ієрархію. Користувач семантично розмічає контент Wiki-сторінок елементами цієї онтології: визначає категорію (чи набір категорій) будь-якого документу й ідентифікує окремі елементи його контенту.

Важливо відмітити, що внаслідок універсальності в e-ВУЕ для зв'язків між поняттями використовується не єдина ієрархія мереологічних властивостей, а постійно розширюваний набір незалежних ієрархій – приміром, та сама персоналія може бути елементом (частиною) наукової школи, політичної партії, мешканцем країни або міста тощо.

Щоб спростити процес використання семантичних властивостей для визначення змісту відношень між гаслами e-ВУЕ, розроблено спеціальний шаблон «Відношення». Зараз цей шаблон знаходиться на етапі апробації, і надалі до нього будуть додаватися ті семантичні властивості, що мають характеризувати існуючі зв'язки між гаслами енциклопедії.

Висновки

Використання мереології як теоретичної основи для визначення специфічних для енциклопедійного ІР підтипів відношень «частина-ціле» між поняттями дозволяє більш однозначно визначити семантику зв'язків між ними, покращує навігацію у ресурсі та зменшує час, що потрібний користувачам для пошуку необхідних відомостей. За допомогою мереологічного підходу класифікація та диференціація підтипів семантичних відношень «Є частиною» стають більш зрозумілими для тих, хто займається доданням семантичної розмітки до Wiki-сторінок. Побудова конкретних відношень усередині кожного підтипу визначається контентом енциклопедичного ресурсу, але наявність критеріїв класифікації значно спрощують пошук потрібного відношення.

Таблиця 1. Підтипи мереологічних властивостей в е-ВУЕ

	Тип властивості	Властивість	Зворотна властивість	Приклади
1.	Компонент-Об’єкт	«є компонентом»	«має компонентом»	Літак-крило,
2.	Член-Колекція	«є елементом »	«має елемент »	область-країна
3.	Частина-Маса	«входить до»	«містить»	Атрибут-База даних
4.	Матеріал-Об’єкт	«є матеріалом»	«містить матеріал»,	алюміній-літак
5.	Властивість-Діяльність	«потрібно для»	«використовує»	Авторизація – автентифікація
6.	Стадія-Процес	«є частиною процесу»	«процес містить частину»	Персональна ідентичність-Автономіза́ція соціа́льна
7.	Місцевість-Область	«знаходиться у»	«є місцем знаходження»	Закарпаття-Україна

Рис.2. Шаблон «Відношення» та його використання в е-ВУЕ.

Використання онтології для формалізації характеристик мереологічних відношень, забезпечує однозначність їх інтерпретації та зменшує ймовірність помилок їх використання.

Список використаних джерел

1. Open Source Initiative. http://opensource.org.
2. OSINT (Open Source Intelligence) – Online Strategies. http://www.onstrat.com/osint/.
3. *Wagner C.* Wiki: A technology for conversational knowledge management and group collaboration // The Communications of the Association for Information Systems. – 2004. – V. 13(1). – P. 264—289. – Режим доступу : http://aisel.aisnet.org/cgi/viewcontent.cgi? article= 3238& context= cais.
4. WikipediA. https:// www. wikipedia.org.
5. *Рогушина Ю.В., Гладун А.Я.* Мерологические аспекты онтологического анализа интеллектуальных Web-сервисов. *Збірник праць VII Міжнародної конференції «Інтелектуальний аналіз інформації» IAI-2007, 12-14 травня 2007, Київ*. С. 312–321.
6. *Лесьневский С.* Об основаниях математики. – http://www.philosophy.ru/library/logic/lesnewxi.html
7. Халина, Н. В., Валюлина, Е. В. (2018). Мереология событий: измерения средовой семантики. Международный научно-исследовательский журнал, (1-4 (67)).
8. Kotarbiǹsky T. *Elementy teorji poznania, logiki formalnej i metodologji nauk*. Lwow, 1929. — 232p.
9. Домбровский Б. Философская система Т. Котарбинского. http://www.ruthenia.ru/logos/number/1999_07/1999_7_02.htm.
10. *Rogushina, J. V.* Semantic Wiki-resources and their use for the construction of personalized ontologies. Problems in programming, (2-3), 2018, P.188-195.
11. MediaWiki. – https://www.mediawiki.org/wiki/MediaWiki.
12. *Krötzsch M., Vrandečić D., Völkel M.* Semantic MediaWiki// International semantic web conference, 2006, pp. 935-942. - https://link.springer.com/content/pdf/10.1007/11926078_68.pdf.
13. *Davies J.* Towards the Semantic Web: Ontology-driven knowledge management / Davies J., Fensel D., van Harmelen F. , 2002. — 288 p.
14. Рогушина Ю.В., Гладун А.Я., Осадчий В.В. , Прийма С.М. Онтологічний аналіз у Web. Монографія. – Мелітополь: МДУПУ ім. Богдана Хмельницького, 2015.

CASCADING EFFECTS SIMULATION FOR CYBER ATTACKS ON THE POWER SUPPLY NETWORK

Georgy Vedmedenko, Iryna Stopochkina, Oleksii Novikov, Mykola Ilin
National Technical University of Ukraine "Igor Sikorsky KPI", Educational and Scientific Institute of Physics and Technology, Kyiv, Ukraine
hopotion@gmail.com, irst-ipt@lll.kpi.ua, o.novikov@kpi.ua, m.ilin@kpi.ua

The paper substantiates mathematical models that can be computationally suitable for modeling of the cascade effects in the networks of critical infrastrusture objects. An interpretation of SIS and Motter-Lai models in the case of damage caused by cyber attacks in networks of critical infrastructure in the field of energy sector is considered. The improvements have been made to the Motter-Lai algorithm. This improvement lets to take into account the volatile nature of attack success, which depends particularly on the cyber security level of critical infrastructure object and cyber security awareness of the staff. Data on the structure of the energy network of Ukraine and data on a successful attack that caused wide failures (2015, Ivano-Frankivsk region, Ukraine) were used. The load parameters of network nodes were calculated. The conclusions about the nature of the dynamics of the network connectivity coefficient depending on the overload factor were made. The parameters of the models were adjusted using the known data about of Ivano-Frankivsk cyber attack history. The software was developed using Python NetworkX. A computational experiment was performed. An experiment showed the suitability of the selected models for predicting cascade blackouts caused by cyberattacks. Also it was shown that the attack has a main impact at the first stage of its development, then the growth of the damaged nodes number became less active. The software can be used for prediction of the network behavior when implementing harmful effects that lead to the failure of the object, and the subsequent cascading effect in the whole network.

Introduction

The applied relevance of the problem is determined by the need to predict a situations caused by cyber attacks in the networks of critical energy infrastructure. For this purpose, it is expedient to determine the degree of influence of each network node on others, indicators of network resistance to the harmful influences, modeling of cascade effects in the corresponding complex network.

In addition, computer simulation may be a good solution for numerical assessment of a number of criteria for assigning objects to critical infrastructure [1] (in particular, the severity of possible negative consequences of disruptions; duration of elimination of consequences and further negative impact on other sectors) and calculation of relevant network indicators. This paper substantiates the use of models that allow modeling and assessing the effects of harmful effects

on energy facilities, and predict the duration of the elimination of the consequences and the extent of negative impacts, using the results of theoretical and applied research infrastructures of other countries [2-5]. The algorithm, which works on the basis of the Motter-Lai (M-L) model, has been improved, which makes it possible to take into account the probabilistic nature of object failures caused by malicious interference. Software was developed, a computational experiment was performed on the example of the energy network of Ukraine.

Mathematical models

In the article [6] the model of cascade process of blackouts in the critical infrastructure network was proposed. It uses topological facilities of investigated network and allows to simulate blackout process taking into account a possibility of impact on all set of network nodes, but not only on ajacent. The main idea of the model is investigation of resulting coherence of the network at the finish of cascade blackout process.

Let *M* is investigated network. For the network nodes we can obtain the values centrality measures using the formula:

$$B(i) = \sum_{st} \frac{\sigma_{st}(i)}{\sigma_{st}},$$

where i is current network node, $\sigma_{st}(i)$- number of the shortest paths between *s* and *t* nodes, that pass through the selected node *i*, σ_{st} – general number of shortest paths between nodes *s* and *t* in the whole network. For the each network node we introduce the boundary load factor of the node:

$$c_i = (1 + \alpha)b_i,$$

where b_i - the calculated load factor of the node in the network (Betweenness centrality, BC), and value $\alpha \geq 0$ is the coefficient of stability of the node, which plays the role of an indicator of the allowable overload of the *i* node.

The algorithm of cascade shutdowns of Motter-Lai is presented in the following form:

1. Delete the attacked node, or set of network nodes, and the corresponding connections, recalculate b and obtain $b`$.
2. Check the overload condition on all nodes: $b` > c$? If true then delete node. If no nodes deleted then fo to step 4.
3. Return to the step 1.
4. Get a modified network M` as a result of damage. Calculate *G* – coefficient of final network connectivity after cascade effect:

$$G = \frac{N_{after}}{N},$$

where *N* – the general number of nodes in *M*, and N_{after} – the number of nodes in the most connected component of the resulting graph *M'*.

SIS model is a modification of well known epidemiology SIR model. This model differs from the previous one in that instead of three states (susceptible, infectious, recovered) there are only two states (susceptible, infectuous). Such a model has been proposed to describe situations in which, in principle, an object cannot receive "immunity" (this is typical for most APT attacks on critical infrastructure objects). Model equation is:

$$\begin{cases} \frac{ds}{dt} = -\frac{\beta s_i}{N} + \gamma i; \\ \frac{di}{dt} = \frac{\beta s_i}{N} - \gamma i, \end{cases}$$

where variables s and i point to a number of objects in "susceptible" and "indected" states accordingly, and β shows the speed of attack propagation or attack effect cascade propagation, and γ parameter shows the recover speed for typical object (node). The total number of objects remains the same throughout the experiment. An alternative simulation model for this investigation can be a cellular automate, however, for large volumes of the data, the computational complexity of its work algorithm can be an obstacle.

Cyber attack on the electricity network of Ukraine in 2015

On December 23, 2015, the Ukrainian regional electricity distribution company (Kyivoblenerho, Ivano-Frankivsk region) announced the disconnection of services to consumers. The cyberattack affected additional parts of the distribution network and forced operators to switch to manual mode. The failure was due to an attack that allowed attackers to remotely control the SCADA system [7]. The attack was carried out on three independent regions of the country with a short period of time between them: Ivano-Frankivsk, Kyiv and Kyiv region, and Chernivtsi region were affected. The attack resulted in several power outages in different parts of the country, causing about 225,000 consumers to lose electricity in various areas. This situation is used to build software models (in particular, the selection of constants), and verify their adequacy.

Initial data processing

To investigate this cyber attack effect, the map of the electricity grid of Ukraine, which was obtained from official site of Ministry of Energy, was used.

The following approach was developed and used for the Motter-Lai simulation model: 1) the load factors (BC) were calculcated for all network nodes. The software package NetworkX Python for obtaining the factors of the network was used. The nodes for the following steps were selected basing on the ontained factor values (the most important values are shown in Table 1); 2)the distribution of nodes by regions was performed for indetifying node classes that possibly were shutdowned first in a given area; 3) after the classification of nodes for these areas and the calculation of the values of the coefficients BC, in each area were selected those nodes whose BC values are the largest.

A set of the most important nodes has been identified: for I-F region – node1, Burshtyn thermal power plant; for Kyiv and Kyiv region – node 77, Bila Tserkva distribution station; for Chernivtsi region – node 6, Dniester hydroelectric power plant. The initial number of successfully attacked nodes is 3, and the numbers of «susceptible» nodes - 169.

Table 1 – Nodes classification and load factor values (normalized on the segment [0, 1] and non-normalized)

Ivano-Frankivsk (I-F)			Kyiv			Chernivtsi		
Node	**BC**	**Normalized**	**Node**	**BC**	**Normalized**	**Node**	**BC**	**Normalized**
1	1216.4	0.083	12	1019.6	0.070	5	372.7	0.026
40	0	0	13	373.9	0.026	6	471.2	0.032
41	692.8	0.048	14	57.3	0.004	45	71.7	0.005
42	0	0	57	0	0	46	54.7	0.004
43	381.7	0.026	77	1046.4	0.072			
44	0	0	78	791.1	0.054			
			79	171.8	0.012			
			80	61.1	0.004			
			81	0	0			
			86	444.9	0.031			

For SIS epidemiology model possible options in the development of the attack are determined by the specific choice of coefficients β and γ, however, for imitational M-L model there is no possibility to determine such impact on a network, therefore, adjastments were made: in addition to coefficient of acceptable overload α the probability of disconnection of the object in case of exceeding the threshold was also introduced. For example, during an attack, in the case of untrained personnel, and low security level of the object, it can be assumed that the value of the probability of shutdown in this case will be close to 1. Taking into account the introduced adjustments, a software implementation was developed.

The results of simulation of cyber attack cascading effects

Based on the processed initial data, the behavior of the network at different parameter values was investigated. Regarding the M-L model, there were cases when the number of deleted nodes increases and the network connectivity parameter G decreases with increasing overload parameter α (Fig. 1). This behavior is associated with the so-called "islanding" effect [4]. For some sets of parameters in the network, it is possible that it is divided into several separate subgraphs ("islands"), not related to each other. The first and last stages of the evolution of the network, obtained on the basis of the adjusted algorithm, are shown in Fig. 2.

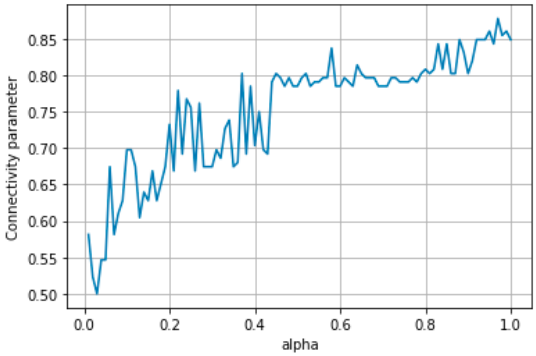


Fig. 1. An example of network connectivity *G* dynamics ($0 < \alpha < 1$, $\Delta\alpha = 0.01$)

The results of modeling based on SIS model are presented in Fig. 3.

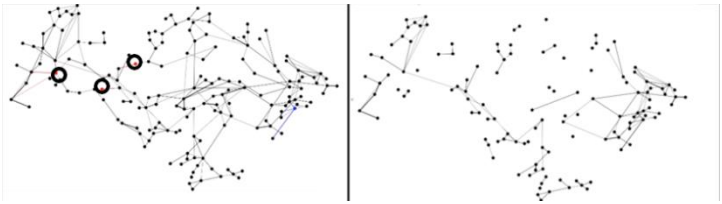


Fig. 2. Evolution of the network in time based on the Motter-Lai algorithm ($\alpha = 0.05$, $P = 0.75$) (the attack sources are marked with circles)

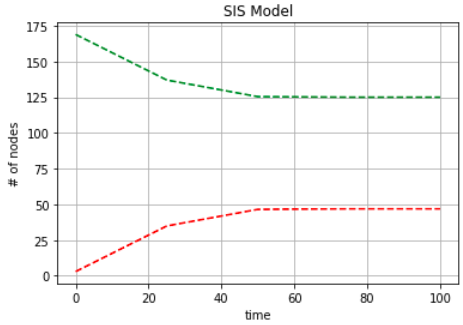


Fig. 3. Number of active (declining) and disabled (growing) nodes in the SIS model ($\beta = 0.55$, $\gamma = 0.40$)

Conclusions

The selected models make it possible to effectively predict the behavior of the network when implementing harmful effects that lead to the failure of the energy supply object, and the subsequent cascading effect of failure in the network. The models agree well in the forecast results, which gives grounds to consider them suitable for an approximate estimate of the number of damaged nodes in the network.

The Motter-Lai model allows us to evaluate the results from the standpoint of simulation modeling of the behavior of objects in the network. The model allows to take into account such factors as the value of the allowable overload of the object and the probabilistic nature of disconnections.

The SIS model makes it possible to estimate the number of active and the number of non-functioning nodes in the network. The experiment showed that the vast majority of failures of child nodes occur in the first time period after the failure of a critical node. Then the effect of failures becomes more stable, which gives reason to believe that with the loss of control over the node in the first hours of the accident, in the future the success of the network does not depend so much on the speed of interventions.

The further research can be related with the analysis of networks of critical infrastructure from the standpoint of negative impacts in the supply chain, which occur when economic and other non-technical relations are disrupted.

Information sources

1. Cabinet of Ministers of Ukraine (October 9, 2020). Resolution No. 1009 — Some Issues of Critical Infrastructure Facilities. [in Ukrainian].

2. Chopade P., Bikdash M. (2011). Critical infrastructure interdependency modeling: Using graph models to assess the vulnerability of smart power grid and SCADA networks.

3. Carreras B. A., Newman D. E., Dobson I. (2014). Thresholds and Complex Dynamics of Interdependent Cascading Infrastructure Systems.

4. Zio E., Sansavini G. (2010). Modeling failure cascades in critical infrastructures with physically-characterized components and interdependencies.

5. Blokus-Roszkowska A. (2019). Safety Analysis of Interdependent Critical Infrastructure Networks.

6. Motter E. A., Lai Y.-C. (2002). Cascade-based attacks on complex networks.

7.Lee R. M., Assante M. J., Conway T. (March 2016). Analysis of the Cyber Attack on the Ukrainian Power Grid.

ДИНАМІЧНЕ МАСКУВАННЯ ДАНИХ ЗІ ЗБЕРЕЖЕННЯМ ФОРМАТУ

Михайло Коломицев, Світлана Носок
Навчально-науковий фізико-технічний інститут
Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського», м. Київ, Україна

box144a@ukr.net, nos.sv.ol@gmail.com

Стаття присвячена актуальній проблемі захисту інформації у базах даних. Автори розглядають метод захисту даних шляхом маскування. Суть маскування даних полягає у незворотній заміні в базі даних конфіденційної інформації (наприклад, даних, що ідентифікують конкретних людей) несекретними даними для запобігання доступу до неї неавторизованих користувачів. Як правило, конфіденційні дані замінюються схожими на реальні значення, щоб їх можна було використовувати в тестових системах з гарантією, що початкові дані не можуть бути отримані, вилучені або відновлені. Маскування дозволяє власникам бази даних самим визначати, скільки конфіденційних даних слід відображати, за мінімального впливу на роботу додатків – дані повинні залишатися функціонально придатними для прикладної обробки (в основному, завдань тестування, навчання тощо). В статті пропонується методика динамічного маскування даних зі збереженням формату.

Вступ

Маскування даних – це процес перетворення вхідних даних (рядків таблиць бази даних), який зберігає частину значення оригіналу, а іншу частину перетворює в форму, яка виглядає як вхідні дані, але які не містять початкових конфіденційних даних. Це, у свою чергу, зменшує масштаби та складність заходів із захисту ІТ. Маскування має працювати із загальними сховищами даних, такими як файли та бази даних, не порушуючи сховища. Маскування повинне зробити неможливим або непрактичним зворотний інженерний аналіз замаскованих значень. Таке перетворення необхідно, наприклад, для виконання вимог Закону України про захист персональних даних [1]. Маскована інформація не повинна ідентифікувати особу, дані якої зберігаються в загальнодоступній базі даних.

Необхідність маскування даних виникає у багатьох областях обробки даних:

- Статичне маскування, відоме як ETL (Extract, Transform і Load) — застосовується для приховування конфіденційної інформації в середовищі, яке відрізняється від виробничого наявністю третіх осіб (програмісти,…) і слабшою політикою безпеки. вихідного репозиторію. Кожна фаза ETL зазвичай виконується на окремих серверах: сховищі вихідних даних,

маскувальному сервері, який організовує перетворення, і цільовий сервер базі даних. Сервер маскування підключається до джерела, отримує копію даних, застосовує маску до вказаних стовпців даних, а потім завантажує результат на цільовий сервер. У деяких випадках потрібно створити замасковану копію безпосередньо в базі даних. Наприклад, коли виробничі дані виявлені в тестовій системі, дані можуть бути замасковані на місці, не переміщаючись взагалі. Таке маскування зменшує ймовірність того, що складні зв'язки, збережені процедури не імпортуються належним чином, таким чином зберігаючи повну цілісність платформи виробничої бази даних. Можливість забезпечити точну копію — з точки зору даних, структури та функціональності — є ключем до того, щоб тестове середовище належним чином імітувало виробниче середовище.

- Ще одним напрямком застосування маскування є впровадження платформи BusinessIntelligence (BI)для обробки значних обсягів інформації (BigData), що надійшла з різних джерел. Така обробка включає аналітичну обробку (OLAP), інструменти візуалізації даних та інтегроване сховище даних (DW). Це – загальна тенденція покращення бізнес-процесів за допомогою використання статистичних методів та методів аналізу даних.Вилучення конфіденційних даних з операційних баз даних, з подальшим збереженням їх у центральному сховищі,забезпечує як конфіденційність інформації так і ефективність аналізу.Методи маскування даних використовуються для мінімізації ризику ненавмисного розкриття конфіденційних даних, а також для збереження базової якості аналізу даних.

Вимоги до процесу маскування наведені в [2]:

1. Маскування не повинно бути зворотним.

2. Результати повинні бути репрезентативними для вихідних даних. Маскування даних полягає в тому, щоб надати інформацію, яка все ще нагадує виробничі дані для цілей розробки та тестування.

3. Необхідно підтримувати цілісність посилань. Маскування не повинні порушувати цілісність посилань (зовнішніх ключів таблиць).

4. Маскувати треба тільки конфіденційні дані.

5. Маскування має бути повторюваним процесом.

Однак відомі реалізації методів маскування як правило не забезпечуютьвиконання п. 2,3 вимог. Формат маскованих даних, належність до того ж домену, що і вхідні дані, цілісність посилань, як правило не підтримуються.

У цій роботі пропонується методика маскування, що зберігає формат даних.При створенні методики використовувались результати авторів, наведені в [3,4]. Пропонованаметодика полягає у необоротному методі маскування, що забезпечує як конфіденційних даних, так і збереження формату маскованих даних. Запропоновано методику збереження цілісності посилань і дозволяє безпечно використовувати конфіденційні дані, відображаючи їх за допомогою належних необоротних методів маскування. Методика не використовує криптографічні перетворення, забезпечуючи можливість динамічного перетворення даних.

Загальні характеристики пропонованої методики:

Збереження формату: маскування створює дані з такою ж структурою, що й вихідні дані. Це означає, що якщо вихідні дані мають довжину 15 символів, масковані дані також довжиною 15 символів. Маскування даних типу дата і час повинне давати числа в правильних діапазонах для дня, місяця та року і часу.

Збереження типу даних: Маскування, що зберігають формат, неявно зберігає і тип даних.

Семантична цілісність: бази даних часто накладають додаткові обмеження на дані, які вони містять, в вигляді умови CHECK. Наприклад, максимальне значення зарплати співробітників. Методика дозволяє виявляти таки обмеження і робити перетворення відповідно до них.

Посилальна цілісність: методика маскування підтримує посилальну цілісність між таблицями. Це гарантує, що завантаження нових даних працює без помилок.

Розподіл частот: в деяких випадках необхідно підтримувати логічне групування значень. Наприклад, якщо вихідні дані описують географічне розташування об'єктів за поштовими індексами, масковані випадкові поштові індекси знищують цінну географічну інформацію. Методика дозволяє зберігати елементи шаблону даних, забороняючи маскувати визначену частину даних.

Унікальність: замасковані значення є унікальними і повторне маскування одного і того ж значення дає інший результат

На рисунку 1 показана загальна модель захисту конфіденційності великих даних з використанням методів маскування даних.

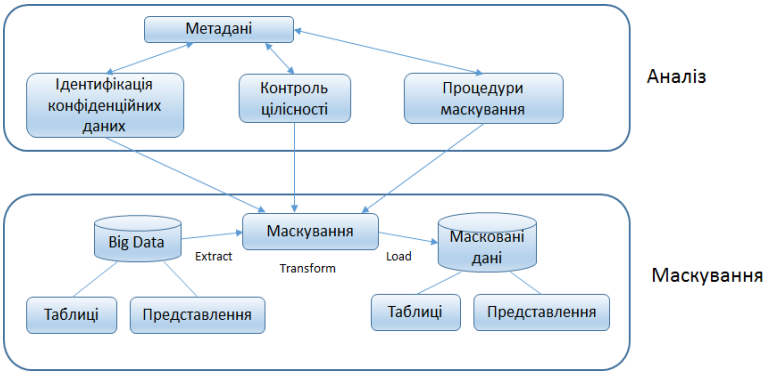


Рисунок 1 - Загальна модель захисту конфіденційності великих даних

Ідентифікація конфіденційних даних. На підставі результатів аналізу метаданих користувачеві надається перелік таблиць і атрибутів даних з яких він позначає ті, що необхідно маскувати.

Контроль цілісності. Шляхом аналізу метаданих визначаються зв'язки між таблицями. Зовнішні ключи маскуються однаково з первинним ключем батьківської таблиці.

Процедури маскування. В залежності від типу і формату даних обирається одна з процедур маскування. Розроблені методики маскування даних типу рядок, число, дата. Процедури маскування дозволяють зберігати реалістичність даних. Для рядкових змінних зберігається кодова сторінка, наявність символів верхнього і нижнього регістру. Для чисел зберігається розрядність і знак. Значення дати перетворюються в коректний діапазон дати і часу. Деякі елементи даних можна позначити як такі, що не перетворюються. Оскільки криптографічні перетворення не використовуються, процедури можна використовувати для динамічного маскування.

Маскування. За результатами аналізу метаданих визначаються тип даних і належність зовнішніх ключів. В першу чергу маскуються виявлені зовнішні ключі разом з відповідними первинними ключами. У подальшому перетворенні вони участі не беруть. В залежності від типу даних стовпця таблиці автоматично обирається і використовується відповідна процедура маскування.

Висновок

Пропоновані методики реалізовані в вигляді фреймворку з графічним інтерфейсом. Фреймворк забезпечує спільну інфраструктуру для розробки, тестування та підтримки, забезпечуючи масштабованість і повторюваність. Повторність використання забезпечується автоматичним аналізом метаданих БД, який визначає спосіб маскування для кожного елемента даних (стовпчика таблиці).

В роботі наведені приклади використання розробленого фреймворка для перетворення даних.

Порівняльний статичний аналіз маскованих і немаскованих даних показав, що масковані дані не дозволяють відновити статистичні характеристики немаскованих даних.

Перелік посилань

1. Закон України «Про захист персональних даних» від 20.12.2012 №2297- VI. (The Law of Ukraine "About personal data protection" from 20.12.2012 №2297- VI)

2. Underst and in gand Selecting Data Masking Solutions: Creating Secure and UsefulData https://securosis.com/assets/library/reports/ UnderstandingMasking_FinalMaster_V3.pdf

3. Маскування таблиць бази даних із використанням технології SQLCLR, Коломицев М.В., Носок С.О., Мазуренко А.Є.Науково-технічний журнал «Захист інформації», №1(19)2017р. – С.16-22.

4. Забезпечення цілісності зовнішніх ключів маскованої бази даних, Коломицев М.В., Носок С.О., Мазуренко А.Є. Науково-технічний журнал «Захист інформації», №4(17)2015р. – С.306-311.

АНАЛІЗ І ПРОГНОЗУВАННЯ ВІДМОВОСТІЙКОСТІ ЗАСОБІВ ЗБЕРІГАННЯ ІНФОРМАЦІЇ

Наталія Кузнєцова, Петро Бідюк
Інститут прикладного системного аналізу
Національного технічного університету України «Київський політехнічний інститут ім. І.Сікорського», Київ, Україна
natalia-kpi@ukr.net, pbidyuke_00@ukr.net

Дослідження присвячене розробці підходу до прогнозування надійності та відмовостійкості технічних систем. У роботі виконано моделювання на основі накопичених історичних даних стосовно стану дисків, що забезпечували зберігання інформації клієнтів-замовників хмарних сервісів, для яких важливими є вимоги щодо безперебійного доступу і високої якості (без втрат) збереженої інформації. Це потребує вчасного оцінювання відмовостійкості й надійності технічного обладнання, яке забезпечує збереження даних. Ідея запропонованого підходу полягає у постійному моніторингу та оцінюванні технічних характеристик (SMART) обладнання, вивченні його глобальної продуктивності, класифікації та прогнозуванні обладнання, для якого можуть настати відмови у наступні моменти часу за допомогою альтернативних методів машинного навчання. Основними етапами дослідження є аналіз відмовостійкості, прогнозування збою протягом визначеного періоду, і, за необхідності, вчасна заміна такого обладнання на нове. Використано різні методи машинного навчання для прогнозування ймовірності події – відмови обладнання, а також прогнозування часу (днів) безвідмовної роботи обладнання і часовий проміжок для їх заміни.

Вступ

Вимоги до роботи різноманітних онлайн веб-сервісів пов’язані з великим навантаженням на обладнання та технічні системи, які забезпечують їх безперебійне функціонування. Одночасний збій та відмова найбільших сервісів соціальних мереж, які сталися у жовтні-листопаді 2021 року в Європі та у всьому світі, були пояснені компаніями-розробниками як одночасне технічне навантаження, з яким не справилось технічне обладнання. Аналогічні виклики постали в Україні перед додатком "Дія", коли велика кількість користувачів змушена була скористатися сервісом, створюючи таким чином надмірне навантаження, що призвело до відмови системи, а в деяких випадках і часткової втрати даних. Саме вимоги до даних, до їх конфіденційності, надійності їх збереження та можливості доступу за вимогою їх власників наразі є найбільш актуальними. Сьогодні спостерігається все більша діджиталізація різних сфер діяльності та перехід на цифрові дані, які регулярно збираються та зберігаються у величезних базах і сховищах даних, щоденно оновлюються, доповнюються

та коригуються, а тому потребують величезного простору для їх розміщення.

Компанія BackBlaze займається хмарним зберіганням і резервним копіюванням даних, наданням масштабних послуг хмарних сховищ, допомагаючи клієнтам розробити і створити власні хмарні рішення, організувати резервні копії існуючих систем і файлів та забезпечити надійну взаємодію з робочими процесами і інструментами. Компанія Backblaze має сотні попередньо створених інтеграцій та партнерів, зокрема і Facebook. Зважаючи на згадані вище технічні збої саме у сервісах компанії Facebook, метою нашого дослідження була саме перевірка надійності збереження даних на дисках різних виробників та визначення найбільш тривалих надійних моделей для надання рекомендації компанії Backblaze.

Аналіз існуючих публікацій з відмовостійкості технічного обладнання і постановка завдання дослідження

У роботі [1] була проведена оцінка ефективності технічних SMART характеристик для передбачення відмов. Автори зробили висновок, що самі параметри SMART навряд чи будуть корисними для прогнозування збоїв окремих дисків, але враховуючи, що технології звітності вдосконалилися, у нашій роботі буде досліджено, чи зможуть ці параметри для жорстких дисків (HDD) з 2015 по 2020 рік бути значущими характеристиками. У роботі [2] Баптіста Марсія застосував комбінацію моделей часових рядів та методів на основі машинного навчання для прогнозування подій збою технічного обладнання авіакомпаній. Він використав статистику характеристик минулих значень як вхідні дані моделі, яка також включає наступну подію передбачення відмов як вхідні дані моделі ARMA, таким чином підвищуючи точність прогнозу моделі у порівнянні зі звичайною Weibull-моделлю [3]. Використовуючи та оцінюючи продуктивність різних класифікаторів машинного навчання для прогнозування відмови в промисловому обладнанні для виробництва анодів, у роботі [4] підкреслено важливість розглянутого часового вікна та його вплив на точність вимірювання відмови. У роботі [5] наголошується на продуктивності LSTM для прогнозування відмов у бойлерах, незважаючи на незбалансований набір даних. Тому у даному дослідженні також було застосовано алгоритм LSTM до набору даних для дисків компанії BackBlaze і оцінено його ефективність. У роботі [6] автори аналізували звіти компанії BackBlaze за 2011-2015 роки. Було використано кластеризацію на основі методу *k*-середнього, щоб вручну виправити величезний дисбаланс класів, який наявний у вхідній вибірці, проте фокусувались лише на 4 різних моделях жорстких дисків (2 Seagate і 2 Hitachi). У нашому дослідженні також зменшено вибірку набору даних, але випадковим чином, і проаналізовано усі моделі виробників жорстких дисків, які використовуються компанією BackBlaze. У попередній роботі [7] було розглянуто ризик відмови жорсткого диска протягом наступного місяця на основі SMART-характеристик за останні 3 місяці, визначивши найбільш

важливі змінні і на основі саме цих SMART-характеристик було спрогнозовано вихід диску з ладу у наступні місяці. Результати, отримані за допомогою цієї моделі, були дійсно точними, але наразі зосередимось на нормалізованих SMART-характеристиках за допомогою аналізу часових рядів. Кожна характеристика SMART буде розглядатися як часовий ряд і будуть розглянуто 2 типи прогнозування: чи вийде жорсткий диск з ладу чи ні за короткий проміжок часу (1,3,10 днів) і прогнозування очікуваного терміну експлуатації жорсткого диска.

Послідовність проведення аналізу

Виконання дослідження проводилось у декілька етапів. На першому етапі здійснювався попередній аналіз – вивчення глобальної продуктивності дисків BackBlaze з використанням моделей часових рядів. На другому етапі виконувався аналіз і класифікація – прогнозування відмов наступного дня на основі SMART характеристик та алгоритмів класифікації машинного навчання. На наступному етапі виковувалвся аналіз відмовостійкості жорсткого диска з використанням нейронних мереж RNN і LSTM та аналізу часових рядів SMART-характеристик.

Попередній аналіз вхідних даних

Основною метою першого етапу є оцінка та прогнозування рівня ризику, з яким BackBlaze стикається з часом, а також глибинне дослідження наявного набору даних. Для оцінювання технічного ризику необхідно передбачити дві його складові: ймовірність відмови (частота відмов) обладнання і оцінити можливі збитки.

Класифікація дисків щодо відмовостійкості методами машинного навчання і прогнозування виходу з ладу на основі SMART- характеристик

У попередній роботі [7] було встановлено кореляцію між деякими SMART- характеристиками та ймовірністю відмови диску. Проведений аналіз враховував лише жорсткі диски до 2015 року, проте за цей час було зроблено значні вдосконалення в технологіях звітності дисків. Тому було вирішено вивчити вплив та важливість особливостей SMART характеристик. Оскільки частина методів на основі нейронних мереж (LSTM і RNN) працюють як чорна скринька, то для вибору ознак використовувався інший алгоритм. У цьому випадку задача визначається як короткострокове прогнозування, яке виконується таким чином: на основі всіх SMART характеристик ми намагаємося передбачити, чи вийде з ладу жорсткий диск протягом наступних 24 годин. Використовуючи різні методи машинного навчання, такі як: градієнтний бустинг, випадковий ліс, метод *k*-найближчих сусідів або логістичну регресію, ми будемо вивчати вплив SMART характеристик на ймовірність відмови.

Корекція дисбалансу класів

Оскільки у вхідному наборі ймовірність відмови за добу низька і не перевищує 0,05%, то під час вирішення задачі класифікації дисків, які можуть відмовити, ми стикаємося з проблемою великого класового дисбалансу. Для вибору важливих ознак під вибірку було виконано перебалансування вибірки для досягнення частки: 95% – хороших, 5% – відмов.

Усі виробники і навіть різні моделі одного виробника використовують по-різному SMART характеристики, тому їх було вирішено пронормувати. Кожна SMART характеристика надає різну інформацію про продуктивність жорсткого диска. Матриця кореляції показує сильну кореляцію між характеристиками SMART 5, 7, 197 та 198, що визначаються як кількість перерозподілених секторів, частота помилок пошуку, поточна кількість секторів в очікуванні, невиправна кількість секторів. Проте для всіх методів класифікації було вирішено використати всі 10 характеристик.

Результати класифікації для короткострокового прогнозування відмови

Було використано різні методи і алгоритми класифікації для класифікації тих дисків, які вийдуть з ладу (відмова) і продовжать працювати у наступні моменти часу. Використано такі основні метрики для класифікації як: точність (accuracy), відгук (recall) та оцінка F1-score.

Таблиця 1. Точність методів класифікації для короткострокового прогнозування відмов дисків

		Клас 0: Жорсткий диск не вийде з ладу в момент d+1				Клас 1: Жорсткий диск вийде з ладу в момент d+1			
Метод	**Вибрка**	**Accuracy**	**Recall**	**F1-score**	**Support**	**Accuracy**	**Recall**	**F1-score**	**Support**
Випадковий ліс	Навчальна	1.00	1.00	1.00	991 271	0.73	0.98	0.84	5 612
	Тестова	1.00	1.00	1.00	251 812	0.28	0.21	0.24	1 110
Градієнтний бустинг	Навчальна	1.00	1.00	1.00	991 271	0.90	0.16	0.28	5 612
	Тестова	1.00	1.00	1.00	251 812	0.46	0.24	0.31	1 110
Логістична регресія	Навчальна	0.99	1.00	1.00	991 271	0.84	0.08	0.14	5 612
	Тестова	1.00	1.00	1.00	251 812	0.56	0.15	0.24	1 110
***k*-найближчих сусідів**	Навчальна	0.99	1.00	1.00	992 594	0.87	0.04	0.07	5 619
	Тестова	1.00	1.00	1.00	251 700	0.60	0.10	0.18	1 108

Аналіз відмовостійкості жорсткого диска з використанням нейронних мереж RNN і LSTM та аналіз часових рядів SMART-характеристик

Використовуючи функцію виживання (у нашому випадку відмовостійкості) і SMART характеристики диску прогнозуємо ймовірність його відмови у наступні N визначених днів. Для цього було виконано додаткове перетворення вхідних даних. Спочатку розраховувався час до відмови диску на основі теорії аналізу виживання, а потім проводилась класифікація шляхом порівняння часу до відмови і параметра – кількість передбачених днів. Завдяки виконаному аналізу виживання жорсткого диску тепер можна визначити, чи вийде він з ладу протягом наступних N днів. Ця інформація

додається як додатковий стовпець Dataframe. Після поділу матриці на дві частини *X* і *Y*, матриця *X* перетворюється у 3D-масив на базі 10 SMART характеристик і 7 значень із лагами. Крім того, різні SMART характеристики масштабуються значеннями від 0 до 1. Наступним кроком було виконано прогнозування параметру кількості днів, які диск ще пропрацює (N_days_forecasted_after), щоб визначити, на який горизот доцільно робити короткострокове прогнозування. Найефективнішим для прогнозування факту відмови було визначено прогнозування збою протягом наступних 10 днів. Вхідна вибірка даних складалась з 118 210 послідовностей, у якій також спостерігався дисбаланс, оскільки кількість відмов була значно меншою, порівняно з нормальними дисками. Всі вхідні дані були розділені на навчальну (80%), тестову (10%) та валідаційну (10%) вибірки. Результати комбінації алгоритмів на основі нейронних мереж і моделей виживання наведено у таблиці 2.

Було виконано класифікацію і прогнозування дисків, які відмовлять у наступний часовий горизонт, на основі нейронних мереж двох типів (LSTM, RNN) та основних і лагових SMART-характеристик дисків, а також на основі комбінованого підходу з використанням нейронних мереж і моделей виживання. Загалом було виявлено різке збільшення оцінки F1-score, порівняно з короткостроковим прогнозуванням, яке було реалізовано раніше завдяки аналізу SMART-характеристик як змінних у часі ознак, що підвищило точність класифікації. Крім того, при використанні парних параметрів введення функції виживання значно підвищується показник F1. Використання значень SMART-характеристик за останні 7 днів виявилось хорошим компромісом для зменшення переобладнання та належного обліку будь-яких істотних змін інформації. При виборі алгоритму нейронних мереж не було помічено великої різниці між алгоритмами LSTM і RNN як за часом обробки, так і за результатами продуктивності.

Таблиця 2. Порівняння моделей

Параметри		Навчальна вибірка				Тестова вибірка			
Метод	Алгоритм	Accuracy	F1-score	Precision	Recall	Accuracy	F1-score	Precision	Recall
TSMART + Survival Function	LSTM	0,76	0,45	0,49	0,48	0,69	0,47	0,40	0,66
TSMARTSF + Survival Function	RNN	0,73	0,38	0,46	0,39	0,68	0,46	0,42	0,64
TSMART	LSTM	0,76	0,44	0,48	0,47	0,67	0,46	0,38	0,66
TSMART	RNN	0,74	0,41	0,44	0,44	0,68	0,45	0,38	0,65

Висновки

Виконане дослідження підтвердило, що використання комбінації моделей виживання для прогнозування кількості днів сталої роботи обладнання дозволило отримати більш точні результати класифікації. Застосований поетапний аналіз до розв’язання задачі дозволив отримати якісніші і точніші моделі класифікації та прогнозування за рахунок додаткової попередньої обробки, препроцесингу даних, перетворення SMART-характеристик у

вигляді часових рядів. Залучення у модель параметрів, що характеризують спрогнозовану ймовірність відмовостійкості та ймовірність виходу з ладу обладнання, а також використання моделей виживання при оцінюванні ризиків дало можливість спрогнозувати кількість днів для безперебійної роботи обладнання. Описаний підхід слід узагальнити на більшу кількість методів класифікації з використанням різних моделей виживання і застосувати до більш широкого класу технічних систем. Це дозволить не лише прогнозувати, а й виконувати вчасну заміну обладнання, яке за оцінками моделей може вийти з ладу на обраному часовому горизонті. Такий підхід дозволяє більш ретельно виконувати профілактичні роботи, зменшуючи негативні наслідки, пов'язані з втратою цінної інформації.

Література

[1] Pinheiro, Eduardo, et al. Failure Trends in a Large Disk Drive Population. 2007. www.usenix.org, https://www.usenix.org/conference/fast-07/failure-trends-large-disk-drive-population.

[2] Baptista, Marcia, et al. «Forecasting Fault Events for Predictive Maintenance Using Data-Driven Techniques and ARMA Modeling». Computers & Industrial Engineering, vol. 115, january 2018, p. 41-53, doi:10.1016/j.cie.2017.10.033.

[3] Upadhyay, S. K., et M. Peshwani. « Choice Between Weibull and Lognormal Models: A Simulation Based Bayesian Study». Communications in Statistics - Theory and Methods, vol. 32, no 2, january 2003, p. 381-405,doi:10.1081/STA-120018191.

[4] Kolokas, N., et al. « Forecasting faults of industrial equipment using machine learning classifiers». 2018 Innovations in Intelligent Systems and Applications (INISTA), 2018, p. 1−6. IEEE Xplore, doi:10.1109/INISTA.2018.8466309.

[5] Fernandes, Sofia, et al. «Forecasting Appliances Failures: A Machine-Learning Approach to Predictive Maintenance». Information, vol. 11, no 4, april 2020, p. 208. doi:10.3390/info11040208.

[6] Botezatu, Mirela Madalina, et al. «Predicting Disk Replacement towards Reliable Data Centers». Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Association for Computing Machinery, 2016, p. 39−48. ACM Digital Library, doi:10.1145/2939672.2939699.

[7] N. Kuznietsova and M. Kuznietsova, "Data Mining Methods Application for Increasing the Data Storage Systems Fault-Tolerance," 2020 IEEE 2nd International Conference on System Analysis & Intelligent Computing (SAIC), Kyiv, Ukraine, 2020, pp. 315-318, doi: 10.1109/SAIC51296.2020.9239222.

РОЗРОБКА ЗАСОБІВ АНАЛІЗУ КОНТЕНТУ СЕМАНТИЗОВАНИХ WIKI-РЕСУРСІВ

Юлія Рогушина
Інститут програмних систем НАНУ, Київ, Україна
Ladamandraka2010@gmail.com

Обґрунтовується можливість та доцільність застосування Wiki-технологій та їх семантичних розширень для побудови енциклопедійного Web-порталу зі складною структурою бази знань. Проаналізовано можливості, які надає семантизація Wiki-технології, та напрямки їх застосування. Розглянуто, які проблеми виникають при розробці бази знань семантичного Wiki-порталу, та запропоновано методи їх розв'язання з використанням машинного навчання та онтологічного аналізу. Принципова відмінність семантизованих Wiki-ресурсів від традиційних (таких, як Вікіпедія) – у наявності формального та однозначного визначення семантики елементів ресурсу та відношень між цими елементами, тому розробка Wiki-ресурсу великого обсягу потребує застосування онтологічної моделі самого ресурсу та засобів її інтеграції з онтологіями відповідних предметних областей. Наведені у роботі дослідження спрямовані на розробку структури розподіленої бази знань, яка відображає різні семантичні аспекти Wiki-порталу, який побудовано з використанням виразних засобів Semantic MediaWiki. Відмінність запропонованого підходу від існуючих методів розробки семантизованих Wiki-ресурсів базується на застосуванні онтологічного аналізу для інтероперабельного подання знань предметної області рівня національної енциклопедії та у створенні засобів для автоматизованої семантичної розмітки контенту порталу. Запропоновано методи та засоби автоматизованого поповнення цієї бази знань семантичними відношеннями між окремими Wiki-сторінками для більш ефективного задоволення інформаційних потреб користувачів порталу. Ці методи використовують елементи машинного навчання та лінгвістичного аналізу. Наявність відношень між сторінками з формально визначеною семантикою забезпечує виконання семантичних запитів з використанням цих структурних елементів в якості умов, інтеграції контенту Wiki-сторінок та логічного виведення значень їх властивостей.

Наведені у роботі дослідження спрямовані на розробку структури розподіленої бази знань семантизованого Wiki-порталу засобами Semantic MediaWiki. Запропоновано методи та засоби автоматизованого поповнення цієї бази знань семантичними відношеннями між окремими Wiki-сторінками елементами для більш ефективного задоволення інформаційних

потреб користувачів порталу. Ці методи використовують елементи машинного навчання та лінгвістичного аналізу.
Використання цих результатів дозволить більш ефективно розробляти семантизовані Web-застосунки на основі Wiki-технологій. Застосування онтологічного аналізу та елементів машинного навчання забезпечує більш коректне створення семантичної розмітки Wiki-сторінок засобами Semantic MediaWiki і дозволяє використовувати такі інформаційні ресурси в якості розподілених баз знань [31].

Список використаних джерел

1. Hendler, J., & Golbeck, J. (2008). Metcalfe's law, Web 2.0, and the Semantic Web. Journal of Web Semantics, 6(1), 14-20. http://www.cs.umd.edu/~golbeck/downloads/Web20-SW-JWS-webVersion.pdf. (accessed 2021, 19 Oct)
2. *Davies J., Fensel D., van Harmelen F.* (2002) Towards the Semantic Web: Ontology-driven knowledge management. – John Wiley & Sons Ltd, , England.
3. Völkel, M., Krötzsch, M., Vrandecic, D., Haller, H., & Studer, R. (2006). Semantic wikipedia. In Proceedings of the 15th international conference on World Wide Web, pp. 585-594. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.103.9471&rep=rep1&type=pdf. (accessed 2021, 23 Oct)
4. Bao, J., Ding, L., Huang, R., Smart, P. R., Braines, D., & Jones, G. (2009). A semantic wiki based light-weight web application model. In Asian Semantic Web Conference (pp. 168-183). Springer, Berlin, Heidelberg.
5. Schaffert, S. (2006). IkeWiki: A semantic wiki for collaborative knowledge management. In 15th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE'06) (pp. 388-396). IEEE. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.232.3232&rep=rep1&type=pdf.
6. Garrett, J. J. (2005). Ajax: A new approach to web applications. https://www.scriptol.fr/ajax/ajax_adaptive_path.pdf.
7. *Wagner C.* (2004) Wiki: A technology for conversational knowledge management and group collaboration *The Communications of the Association for Information Systems*, V. 13(1), P. 264-289.
8. *Bonomi, A., Mosca, A., Palmonari, M., Vizzari, G.* (2008) Integrating a wiki in an ontology driven web site: Approach, architecture and application in the archaeological domain. *Third Workshop on Semantic Wikis–The Wiki Way of Semantics 5 th European Semantic Web Conference*, P. 106-116.
9. OWL 2 Web Ontology Language Document Overview. W3C. 2009 Available at: http://www.w3.org/TR/owl2-overview/ (accessed 2021, 15 Oct).

10. *Krotzsch M., Vrandecic D., Volkel M.* Semantic MediaWiki. http://c.unik.no/images/6/6d/SemanticMW.pdf (accessed 2021, Nov 15).
11. SMW annotations manual. http://semantic-mediawiki.org/wiki/Help:Annotation.
12. Gil, Y., & Ratnakar, V. (2002, May). A Comparison of (Semantic) Markup Languages. In FLAIRS Conference (pp. 413-418).
13. *Li Z. Y., Wang Z. X., Zhang A. H., Xu Y.* (2007) The domain ontology and domain rules based requirements model checking. *International Journal of Software Engineering and Its Applications*, 1(1), P.89-100
14. *Lawson T.* (2012) Ontology and the study of social reality: emergence, organisation, community, power, social relations, corporations, artefacts and money. *Cambridge Journal of Economics*, 36(2), P.345-385.
15. *Jones, C. B., Abdelmoty, A. I., Fu, G.* (2003)Maintaining ontologies for geographical information retrieval on the web. In OTM Confederated International Conferences. *On the Move to Meaningful Internet Systems*, . Springer, Berlin, Heidelberg, P. 934-951.
16. *Athanasiadis, T., Tzouvaras, V., Petridis, K., Precioso, F., Avrithis, Y., Kompatsiaris, Y. (*2005) Using a Multimedia Ontology Infrastructure for Semantic Annotation of Multimedia Content. *SemAnnot ISW.*
17. Gruber, T.R. The role of common ontology in achieving sharable, reusable knowledge bases. Available at: http://www.cin.ufpe.br/~mtcfa/files/ 10.1.1.35.1743.pdf. (accessed 2021, Nov 28).
18. *Resnik P.* (1999) Semantic Similarity in a Taxonomy: An Information-Based Measure and its Application to Problems of Ambiguity in Natural Language. *Journal of Artificial Intelligence Research* 11, P.95-130.
19. *Lee, J. H., Kim, M. H., Lee, Y. J.* (1993) Information retrieval based on conceptual distance in IS-A hierarchies. *Journal of Documentation*, 49(2), P.188-207.
20. Kousetti, C., Millard, D. E., Howard, Y. (2008). A study of ontology convergence in a semantic wiki. Proc. of the 4th International Symposium on Wikis, P. 1-10.
21. Herzig, D. M., & Ell, B. (2010). Semantic mediawiki in operation: Experiences with building a semantic portal. International Semantic Web Conference, P. 114-128. Springer, Berlin, Heidelberg.
22. *Andon P.I., Rogushina J.V., Grishanova I.Y., Reznichenko V.A., Kyrydon A.M., Aristova A.V., Tyschenko A.O.* (2021) Experience of Semantic Technologies Use for Development of Intelligent Web Encyclopedia. Proc. of the 12th International Scientific and Practical Conference of Programming (UkrPROG 2020), *CEUR Workshoop Proceedings*, Vol-2866, P.246-259.
23. Quinlan, J. R. (1986). Induction of decision trees. Machine learning, 1(1), 81-106.https://link.springer.com/content/pdf/10.1007/BF00116251.pdf.
24. Hssina, B., Merbouha, A., Ezzikouri, H., Erritali, M. (2014). A comparative study of decision tree ID3 and C4. 5. International Journal of Advanced Computer Science and Applications, 4(2), P.13-19.

25. Gupta, V., & Lehal, G. S. (2009). A survey of text mining techniques and applications. Journal of emerging technologies in web intelligence, 1(1), 60-76.http://www.jetwi.us/uploadfile/2014/1230/20141230112729939.pdf. (accessed 2021, 21 Oct).
26. Рогушина Ю.В., Гладун А.Я. (2014) Методика розробки термінології інформаційних ресурсів як базису формування онтологій та тезаурусів для семантичного пошуку // Інженерія програмного забезпечення, №1(17), С.41-51.
27. Rogushina J. (2019) Use of Semantic Similarity Estimates for Unstructured Data Analysis // Інформаційні технології та безпека. Матеріали XIX Міжнародної науково-практичної конференції ІТБ-2019. Selected Papers of the XIX International Scientific and Practical Conference "Information Technologies and Security" (ITS 2019). CEUR Vol-2577, P.246-258. – http://ceur-ws.org/Vol-2577/paper20.pdf. (accessed 2021, 11 Oct).
28. *Рогушина Ю.В., Гришанова І.Ю.* (2020) Розробка засобів семантичного пошуку та навігації в онлайн-версії «Великої української енциклопедії» на основі онтологічного подання знань. *Енциклопедичні проекти – чинники національного поступу : монографія* / За ред. Киридон А. М. – Київ : Державна наукова установа «Енциклопедичне видавництво», С.141-160.
29. *Rogushina J.V.* The Use of Ontological Knowledge for Semantic Search of Complex Information Objects. *Open semantic technologies for intelligent systems*, 2017, pp.127-132.
30. *Гладун А.Я., Рогушина Ю.В.* (2008) Основи методології формування тезаурусів з використанням онтологічного та мереологічного аналізу. *Штучний інтелект*, №5, С.112-124.
31. *Rogushina, J., Gladun A., Pryima S., Strokan O.* (2021) Development of advisory system based on semantic technologies. *Открытые семантические технологии проектирования интеллектуальных систем.* Выпуск 5. Минск, С.47-54s.

AN APPROACH TO GREEN FINANCIAL CREDIT RISKS MODELING

Kuznietsova Nataliia[1], Remi Amoroso[2]

[1] "Igor Sikorsky Kyiv Polytechnic Institute" National Technical University of Ukraine, Kyiv, Ukraine
[2] Ecole Centrale de Lyon, Lyon, France
natalia-kpi@ukr.net, remi.amoroso@ecl17.ec-lyon.fr

This research is focused on development systemic approach for evaluation and including green indicators in scoring model for credit ratings. The main idea is to give the credits to those companies who modify their business, production and economic based on the aim to decrease the pollution, gas emission, carbon oil and other environmental influences on the world and climate change.

Introduction

The climate change is a crucial challenge for the society, that even shown by the increase of policies and actions taken by governments nowadays. In particular, the Paris agreement [1] signed in 2015 demonstrates the will of barely all countries of the world to take climate-change as a serious issue and to find solutions in the near future. Whether it is the climate-change itself or the fight against it, it has huge consequences on the companies' social and environmental behavior as well as on the financial sphere. Green finance is indeed a major challenge today and businesses need to adapt their strategies to maintain growth. An essential tool for companies development is credits. So the task of implementation green force for the companies who make production and business at each country is quite important. In global mean the country should stimulate their business and citizens to decrease the quantity of policies each day. For this reason it is proposed special types of green credits for business and individuals, where is evaluated how new projects and improvement or rebuilding existed business will decrease the policies and negative environmental influence in green metrics. Special agencies are focused on development approach for the evaluation of this influence but there is no still properly approach to this.

Overview: green economic risks

The risks related to climate-change include transition risks (such as changing technology, business models and consumer demands from the transition to a low-carbon economy), water-related risks (such as water pollution, water scarcity, and flooding), resource-related risks (including stranded assets and scarcity of certain

minerals), and natural capital-related risks (such as ecosystem degradation, deforestation, air pollution, extreme weather events and soil nutrient loss) [2]. The fundamental problem with assessing climate risk is trying to map the physical risks onto their fiscal and economic indicators. It is difficult to quantify physical risks: climate risk in general is not quantifiable at all. So agencies go through a process of trying to granulate those risks in a way that they can integrate them into the financial measures that they use. Here we must talk about the levels of physical risk and how they actually translate into financial and economic risks.

To limit the environmental impacts, governments and international organizations issue recommendations and take actions. The main international goal is the Paris agreement [1], which was signed in 2015 by 195 delegations and aims to limit global warming to well below 2, preferably to 1.5 degrees Celsius in 2100, compared to pre-industrial levels. But this kind of policy also has consequences on companies and investors businesses.

Several organizations have specialized in studying the impact of the green transition on financial risks. The PACTA tool (Paris Agreement Capital Transition Assessment) (2DII s.d.), developed by 2° Investing Initiative (2DII) with backing from UN Principles for Responsible Investment, enables users to measure the alignment of financial portfolios with climate scenarios as well as to analyze specific companies. The independent financial think tank Carbon Tracker (Carbon-Tracker-Initiative s.d.) carries out in-depth analysis on the impact of the energy transition on capital markets and the potential investment in high-cost, carbon-intensive fossil fuels. It issued reports giving advice about how to minimize the financial risk. In particular, the major oil companies should consider a 2°C scenario with big policies taken by government in the 2020's.

All the presented studies are based on scenarios. It is needed to forecast several scenarios regarding global temperatures, global carbon emissions, national and international policies, etc. Thus, the prediction model is able to assess the assets' value at risk. Several scenarios are given by climate institutions. What is important to assess the finance risk of a portfolio, is to assess the portfolio alignment with the possible scenarios of climate change, given by different institutions.

In addition to the models based on international possible scenarios, we can find articles assessing the consistency of economies with green measures. In article [3] it was explained a method used to assess the development of green finance in China. On the contrary of assessing a risk, it builds an index that measures the level of development of green finance and applies it to China. The assessment was based only on green credit because it represents 95% of green finance volume in China and because data were available.

In the book [4] it was offered a complete guide about scoring models and scorecards for the credits evaluation. The idea of a scorecard is to transform the probabilities of a client paying the loan (creditworthiness of an individual) into a number that could be easily interpreted, guiding business decisions. The idea of scoring models and scoring cards was used in this research as the way how to calculate the green indicators influence on the credit return probability.

Problem statement

The main idea of this article is development the approach for the business credits aimed to receive move sustainable development indicators. It is focused on the companies' credit risk and try to establish a mean to forecast the credit probability of companies default based on given financial, but also green indicators. Thus, the problem statement could be presented as the following: how to build a scoring model which gives the possibility to financial entities to assess a credit risk according to environmental measures?

Research work will be divided into two major parts. The first one supports and demonstrates the hypothesis that companies' green health has indeed a significant impact on the associated credit risk. For that, datasets containing average data about loans, environmental taxes and greenhouse gas emissions of companies in the USA were used. The classifying model was created in order to determine credit risk categories. Thereafter, the second part aims to establish a reliable model to assess companies' credit rating from financial and environmental features of these companies.

Green Finance modelling

While the problem is quite new especially for Ukraine there is no available history, experience and dataset for this measures of green influence on credit scores. That's why in this research first of all it is needed to make a representative dataset and to check both the scoring models without and with green indicators. It was decided to make a dataset for credit loans of big US companies which are making their business. So firstly it was made based on their issued from the FRED (Federal Reserve Bank of Saint Louis) the average loan value for all non-financial companies taking credit from commercial bank in the US and their classification into 3 default risk categories: minimal, low and average (FRED 2021). Data were available quarterly on 20 years for each category so there were in total 243 observations. Next datasets were made from the environmental indicators. The datasets give average data about greenhouse gas (GHG) emissions and environmental taxes related to non-financial companies in the USA. Both datasets were sorted by business sectors. The GHG emissions were calculated in tons of CO2 equivalent and the taxes were evaluated in US Dollars. We take here only data from 1997 to 2017 to be consistent with the financial dataset. We have division into different sectors and units so there are 3921 observations for the GHG emissions and 1479 for the environmental taxes. Now after all preprocessing and preparation we can work through the three datasets in order to find whether there is an actual link between the companies' environmental behavior and the probability of default of their credits. There are a total of 29 features that can be used to forecast the credit risk category. After the financial one which is the loan value, there are 12 features that can be used about GHG

emissions and 16 about environmental taxes. While the company risks evaluation is determined as classification tasks (where based on the defined metrics the categories of companies are made. For this reason the most promising and classical classification machine learning techniques were chosen such as: gradient boosting, random forest, k-nearest neighbors, logistic regression, neural network. Nevertheless, even with the method GridsearchCV that enables to tune the models' parameters and to check the results with cross validation, the accuracies of the classification were not good (the maximum score was 61%).

On the next part we chose another 3 datasets which had already needed indicators: financial data, environmental behavior and credit rating. The first two datasets were from the Carbon Disclosure Project (CDP 2021). The first dataset contained six financial indicators as revenue. Cost of revenue, operating income, operating expenses, depreciation and amortization, EBIDTA on 368 US and Canadian companies for the years 2018, 2019 and 2020. The second dataset was made based on the green indicators: greenhouse gas emissions and water security for the same companies that in previous dataset. The third dataset was made based on the Standard & Poor's forecast of credit ratings for 592 big US firms.

On the new dataset the task was divided into three sub-stages: establishing a predicting model with only green features, then only financials' and finally with both of them using different machine learning methods. The most accurate model received on each step it was a logistic regression.

Table 1 Comparison of models accuracy depending on the indicators and variables

Indicators used for the classification model	**Accuracy**
Green indicators only	35%
Financials indicators only	88%
All features	93%

The results support the statement proven in the first stage: using environmental features allow to predict more accurately (93% against 88%) the companies' credit rating. It also shows that financials are crucial because using only green indicators give a classifier worse than randomization. Moreover, when all features are used to predict credit rating, we notice that they are all approximately equally important.

Conclusions

In made research we can make a sum up here that indeed it is a link between a company credit risk and ratings and their green and environmental-oriented behavior. In a first stage, working on average data of all non-financial US companies it allowed to prove the statement. Nevertheless on the failed attempts to build a proper dataset by own hands it is clear that in future banks should gather

all needed information from the companies before evaluating their credit scores. Moreover, new interest and efforts of all international projects are focused on Green Deal so the importance of the evaluating this factors for the companies is already fixed. Probably, based on this indicators it could be made new world credit risks rating agencies with own ratings and metrics (such as INCRA).

To conclude, environmental behavior of companies impacts the probability of their credit defaults, and we provide here a model which assess this probability, given environmental and financial information about the companies. Moreover, a mathematical classifying model was created in order to classify any company of which the features needed are available by credit rating.

In future we will work on improvement of the accuracy and precision of the best classification models based on the new requirements and indicators. In this research it was proposed a global model and a new approach but in our next research we will focus on developing different scoring models for different business sectors or company sizes and individual green credit needs.

Literature

[1] 2DII. PACTA. s.d. URL: https://www.transitionmonitor.com/.

[2] Green Finance Platform. URL: https://www.greenfinanceplatform.org/themes/climate-change.

[3] Xingyuan W., Hongkai Z., Kexin B. :The measurement of green finance index and the development forecast of green finance in China. Springer, 2021.

[4] Siddiqi, N.: Intelligent Credit Scoring. John Wiley & Sons, Inc, 2017.

РОЗРОБКА СТРУКТУРОВАНОГО ІНФОРМАЦІЙНОГО ПОЛЯ ОБ'ЄКТА ІЗ ЗАДАНИМИ ВЛАСТИВОСТЯМИ ДЛЯ ПОБУДОВИ ЙОГО СЕМАНТИЧНОЇ МОДЕЛІ

Анатолій Гладун[1] [0000-0002-4133-8169],
Родріго Мартінез-Бежар[2] [0000-0002-9677-7396]
[1] Міжнародний науково-навчальний центр інформаційних технологій та систем НАНУ та МОНУ, Київ, пр. Глушкова, 40 Україна
[2] Університет Мурсії, СР 30180 Буллас, Іспанія
glanat@yahoo.com, rodrigo@um.es

Впровадження прикладних систем штучного інтелекту потребує нових підходів до формалізації та моделювання знань предметної області. Збір даних про інформаційні об'єкти, які є складовими компонентами інтелектуальних систем потребує створення достовірного інформаційного поля об'єкта із наперед заданими обмеженнями, які впливають на структуру поля та забезпечення його наповнення із відкритого Web-середовища. У роботі запропоновані засоби і методи попередньої обробки поля даних (класифікації, кластеризації, асоціативних зв'язків) та інтеграції накопичених ресурсів, а також методика поетапного розроблення структури інформаційного поля, яка починається з визначення інформаційного об'єкта дослідження, навколо якого буде будуватися поле. Проблема достовірності інформаційних ресурсів із яких наповнюється інформаційне поле вирішується аналітичними методами та методом семантичного аналізу метаданих, які забезпечують інтерпретацію, розпізнавання та підбір достовірних даних релевантних до вирішуваної задачі. Проблема точності розпізнавання та інтерпретації даних накопичених даних вирішуються на основі методів обчислення семантичної близькості та семантичної відстані між семантичними моделями. Практичним впровадженням отриманих результатів було створена мультиагентної система е-комерції на основі онтологічного підходу.

Вступ

Створення прикладних систем штучного інтелекту потребує формалізації знань про інформаційні об'єкти, що забезпечує успішне вирішення поставлених задач, де семантичні моделі (онтології, таксономії тезауруси) виступають як засоби формалізації інтероперабельних знань. Для створення семантичних моделей інформаційних об'єктів розробник потребує достовірної динамічно-оновлюваної інформації про досліджувані об'єкти. З цією метою використовуються семантично розмічені відкриті інформаційні джерела Web-середовища та технологія OSINT[1, 2]. Накопичена інформація

для створення семантичних моделей формується в систему - інформаційне поле. Це потребує в першу чергу розробки концепції та структури інформаційного поля, методів попередньої обробки зібраної інформації (класифікація, кластеризація, асоціативні зв'язки тощо), збереження, автоматичного оновлення та оцінювання накопичених ресурсів.

Поняття структури інформаційного поля об'єкта

Створення актуальних семантичних моделей інформаційних об'єктів (онтологій, таксономій, тезаурусів), які задіяні у прикладних системах штучного інтелекту потребує побудови достовірного інформаційного поля цього об'єкта, як джерела динамічної постійно-обновлюваної інформації про об'єкт.

«Інформаційне поле» є загальним поняттям і при дослідженні певної предметної області (ПрО) та потребує уточнення виду і дефініції самого поняття та його аспектів. Ми будемо розглядати інформаційне поле, як сукупність інформаційних ресурсів (ІР) або масивів інформації, які висвітлюють інформацію з певним ступенем достовірності, з певних точок зору, певної якості та цінності, орієнтованого на дослідника, розробника ІІС. У загальному випадку **інформаційне поле** це – «сховище» усієї інформації, що надходить від інформаційного об’єкта. Інформація, що надходить до інформаційного поля може бути структурно розподіленою на окремі підполя за певними характеристиками, ознаками і зберігатися в полі, створюючи так званий накопичувальний «банк даних». При цьому інформаційне поле про об’єкт, подію, явище може бути необмеженим у часі і просторі. Категорія **"інформаційне поле", як частина поняття "інформаційний простір", може охоплювати певний обмежений обсяг фактів і подій реального світу.**

Інформаційне поле як таке є необмеженим, але межі щодо його сприйняття залежать від можливостей та цілей системи або дослідника, що його застосовує [3]. Інформаційне поле об'єкта є частиною інформаційного простору, і тому воно наслідує деякі характеристики притаманні простору, а саме поле містить інформаційні відношення і деякі елементи поля (підполя), які є результатом виявлених закономірностей і залежностей. Елементи інформаційного поля можуть бути дискретними та безперервними і не обов'язково числами (це можуть бути аудіо чи відеофайли, зображення, текстові документи тощо)[4].

Термін "інформаційне поле" на сьогодні не має остаточного визначення, тому ми наведемо приклади інших визначень, які є у науковій літературі. "Інформаційне поле" [5] – поле, у кожному точці якого визначено один чи декілька інформаційно визначуваних параметрів. Такий показник може бути дихотомічним (розбитим на дві частини), просторово-параметричним, просторовим, географічним, інтелектуальним. Інформація, що входить до інформаційного поля, може бути охарактеризованою з точки зору трьох аспектів: 1) *синтаксичного* (структурного) – структура, способи організації інформації; 2) *семантичного* (смислового) – прогнозування та моделювання,

стратегії та підходи до моделювання; 3) *прагматичного* – цінність та корисність інформації, її здатність впливати на процеси пов'язані з інформацією [6,7].

Методика розроблення структури інформаційного поля

Першим кроком в методиці розроблення структури інформаційного поля є визначення інформаційного об'єкта дослідження, навколо якого буде будуватися поле. Саме інформаційний об'єкт є першим з обмежень, що накладається на інформаційне поле. Крім того, досить часто початкові знання дослідника щодо обраного об'єкта є мінімальними і з часом вони будуть поповнюватися динамічно в процесі роботи системи. На рисунку 1 наведено методику розроблення структури достовірного інформаційного поля об'єкта.

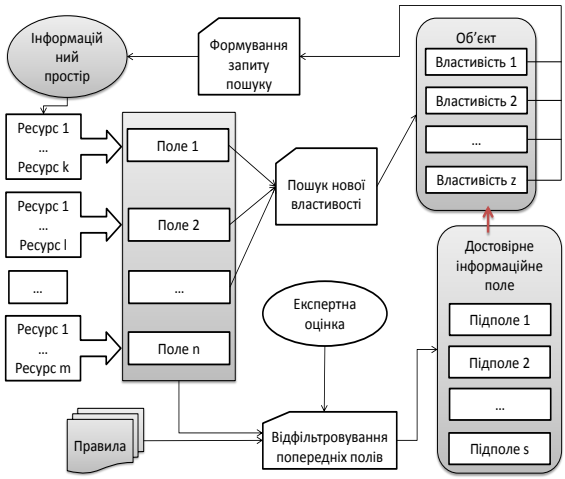


Рис. 1 – Методика розроблення структури достовірного інформаційного поля об'єкта

У лінгвістиці кореферентність – це явище, яке позначає відношення між знаковими одиницями, компонентами висловлювання, коли кілька компонентів відсилають до одного об'єкта позамовної дійсності (референта). Під кореферентними властивостями будуть розумітися властивості, які залежать не стільки від лексичної семантики, а від контексту. Наприклад «пустий» і «чистий» не є семантично близькими виразами, проте для поняття «бланк» ці вирази будуть позначати одне й те саме «незаповнений бланк», тобто бути кореферентними властивостями об'єкта «бланк. Усі властивості можна представити деревом, моделлю класів, таксономією, тезаурусом або словником.

II крок. За обраною властивістю $v_i \in V$ або $w_g \in V$ формується запит пошуку до відкритого інформаційного простору – Web-середовища, а саме Інтернет- ресурси (IP), електронні енциклопедії, електронні бібліотеки, електронні тлумачні словники, Wiki-ресурси, тощо.

III крок. З інформаційного простору за запитом пошуку щодо обраної властивості об'єкта відбираються релевантні ресурси IR_r, які групуються в попереднє поле p_j^p.

IV крок. В отриманому полі p_j^p проводиться пошук нової властивості v_{i+1} або множини кореферентних властивостей w_{g+1}. Нова властивість/множина долучається до блоку властивостей об'єкта V, який перевіряють на кореферентність з уже існуючими властивостями.

V крок. Кроки I-IV повторюються доки не буде отримано необхідна кількість властивостей об'єкта $V_z = \{v_i \mid i \in \{1,..,z'\}, w_g \mid g \in \{1,..,z''\}\}$, $z = z' + z''$, а також необхідна кількість полів $p_j^p \in P^p$, $j = \{1,..,n\}$ для цих властивостей. Тобто множина P^p, умовно поділена на окремі підполя, що відрізняються джерелами генерування інформації та її характером і буде нашим інформаційним полем.

VI крок. Після формування інформаційного поля P^p, застосовуючи правила та експертну оцінку джерела ресурсів до кожного $p_j^p \in P^p$, ми відфільтровуємо інформаційні ресурси IR_r, що є не актуальними та мають низку цінність чи міру достовірності.

VII крок. Відфільтрована множина підполів P^p складає нову множину P^d, що уже є достовірним інформаційним полем з достовірними підполями $p_q^d, q = \{1,..,s\}$, тобто $p_q^d \in P^d$.

Висновки

Розроблено нові підходи до структуризації інформаційного поля об'єкта на основі таксономічного подання вибраних характеристик досліджуваного об'єкта з метою побудови його семантичної моделі для задач розпізнавання та адаптації до специфіки вирішуваної задачі. Адаптація онтологічних моделей задачі і інформаційного об'єкта базується на їх семантичній близькості з застосуванням ваг концептів та відношень використовуваних в задачах розпізнавання інформаційних об'єктів.

Перелік посилань

8. Zgurovsky M., Lande, D., Boldak A., Yefremov K. and Perestyuk M. Linguistic Analysis of Internet Media and Social Network Data in the Problems of Social Transformation Assessment // Cybernetics and Systems Analysis, 2021, Vol. 57, No. 2.-P.228-237.
9. Распознавание информационных операций / А. Г. Додонов, Д. В. Ланде, В. В. Цыганок, О. В. Андрейчук, С. В. Каденко, А. Н. Грайворонская. - К.: Инжиниринг, 2017. — 282 с. ISBN 978-966-2344-60-8.
10. Дьячков В.В. Информационное поле взаимодействия экономических систем: автореф. канд. экон. наук. 08.00.01. (экономическая теория). Тамбов. 2007
11. Цветков В.Я. Естественное и искусственное информационное поле// Межд. журнал прикладных и фундаментальных исследований. – 2014. – No 5, ч. 2. – С. 178–180
12. Иванников А.Д., Цветков В.Я. Получение знаний для формирования информационных образовательных ресурсов. – М.: ФГУ ГНИИ «Информика», 2008 – 440 с.
13. A.Y. Gladun, K.A. Khala Ontology-based semantic similarity to metadata analysis in the information security domain // Scientific journal «Problems of programming». – 2021. – № 2. –PP. 34-41.
14. Gladun A., Khala K., Subach I. Ontological approach to big data analytics in cybersecurity domain. *Inform. Technol. and Security*. 2020. Vol. 8. No. 2 (15). – P. 3—15.

ОЦІНЮВАННЯ МЕТОДІВ МАШИННОГО НАВЧАННЯ ДЛЯ ВИЯВЛЕННЯ DOS/DDOS АТАК В IOT

Гальчинський Л.Ю., Грайворонський М.В., Носок С.О.
НТУУ «КПІ імені Ігоря Сікорського»

ВСТУП

Бурхливий розвиток Інтернету породив цілу множину нових явищ, зокрема появу Інтернету речей (англ. Internet of Things, далі IoT) в різних сферах життя. IoT - це сукупність системи пов'язаних пристроїв, якою керують за допомогою веб-сервісів або інших різного роду типів інтерфейсів [1]. Зовсім не дивно, що ця нова популярна технологія, попала в поле зору кіберзлочинців, які націлюють на IoT - системи різноманітні прийоми злому. Ця проблема погіршується відсутністю необхідних стандартів в системах Інтернету речей, а також специфікою предметної області IoT : використання відносно дешевих та малопотужних пристроїв [2]. Ця специфіка робить пристрої IoT вразливими для класів кібертак, зокрема відмова в обслуговуванні (DoS) та розподілена відмова в обслуговуванні (DDoS) [2]. Такі атаки, як правило, завдають великих збитків юридичним особам, дотичним до сфери IoT, бо легко приводять до блокування штатної роботи пристроїв і , як наслідок, унеможливлюють коректне обслуговування системи. На додачу технологія цих атак відома своєю відносною простотою, а що є ще одним з обтяжливих моментів проблеми вразливості Інтернету речей.

Проблема ще і в тому, що традиційні високоякісні рішення інформаційної безпеки малопридатні для захисту IoT-систем. Тому розвиток цієї галузі потребує релевантної системи протидії проти атак в мережі Інтернету речей. Ще одне джерело знаходиться вразливості витікає високої ціни даних, які збираються і обробляються пристроями Інтернету речей.

Вчасне виявлення різного роду вразливостей та атак на пристрої IoT вкрай важливе, оскільки пристрої в мережі Інтернету речей є в цілому автономними, тобто не вимагають втручання користувача для коректної роботи. З цього витікає, що релевантні мережеві рішення безпеки для системи IoT мають лежати в площині, де відразу має бути забезпечені відразу декілька критеріїв, зокрема достатня швидкість, точність оцінки і ще ряд показників. На наш погляд рішення з машинним навчанням (ML) є одним з найбільш перспективних для розглянутої проблеми.

Хоча рішення інформаційної безпеки з ML набули певної популярності [3],[4],[5], але в питання виявленню атак саме в мережах IoT ще недостатньо вивчене.

Вразливості та атаки в мережах IoT

Спочатку коротко оглянемо різновиди вразливостей та атак в мережах IoT. Як відомо, значне зростання технології IoT було обумовлене завдяки технічному прогресу, який привів до суттєвого зменшення споживання електроенергії у порівнянні зі звичайними вбудованими системами. У свою чергу. Підключення через мережу Інтернет зробило передачу даних від кінцевого користувача до сервера зручним та ефективним. Але зростання мереж IoT спричинило і зростання загроз. Як правило аналіз загроз безпеці та методи їх виявлення обмежуються одним рівнем мережі. Але для мереж IoT реальні атаки можуть бути комбінацією загроз з декількох мережевих рівнів. Стандартна техніка захисту, що розроблена для певного рівня IoT і мереж датчиків, може виявитись неефективною проти атак з інших рівнів. З цього витікає, що для надійної оцінки необхідний аналіз мережі в цілому. Очевидно, також рівень загроз через пристрої IoT, які працюють автономно буде невпинно зростати. Очевидно, пристрої IoT мають розглядатися як найслабша ланка в ланцюжку безпеки мережі IoT. Наразі виробники пристроїв IoT мало дбають про їх безпеку, бо більше заклопотані бажанням захопити якомога більшу частку ринку у конкурентній боротьбі.

Мережі IoT вразливі до відомих кібератак, серед яких на першому місці є відмова в обслуговуванні (DoS) , а також розподілена відмова в обслуговуванні, атаки повторення, людина посередині атаки, маршрутизації та підслуховування[6]. Криптографічні рішення для пристроїв IoT не можуть бути надійними із-за своїх обмежених обчислювальних ресурсів. Система керування ключами може бути скомпрометована через обмеження енергоспоживання пристроїв IoT. Такі мережеві протоколи, як IPv6 і IPv4, є цілями для спроби встановлення віддаленого доступу. Встановлено, що до пристроїв IoT можна отримати віддалений доступ через інтерфейс командного рядка, через використання протоколу Telnet [7]. Мережі IoT часто безперервно обробляють великі дані. Втрата даних або відмова в обслуговуванні може призвести до появи величезного обсягу трафіку в мережі та втрати контролю над роботою системи.

Вразливості IoT можна розподілити на наступні категорії:

Вразливості на рівні процесорів, вразливості апаратного забезпечення, вразливості мережі**,** програмні вразливості.

Для вразливості на на рівні процесорів зловмисники вставляють певне зловмисне вставлення(hardware Trojan - HT) у пристрій IoT, щоб спричинити витік інформацій або щоб маніпулювати нею, чи обійти систему безпеки. Інтегровані мікросхеми, специфічні для застосунків системи IoT та пристрої, що програмуються, є вразливими від цифрових та аналогових HT. Через малу обчислювану спроможність пристроїв їх криптографічний захист значно менший, ніж для високопродуктивних інтегральних схем. Ця вразливість не може бути усунена звичайним програмним забезпеченням та є витратним при виправленні. До того ж через складнощі відновлення, процедура нейтралізації HT знижує якість обслуговування. Цей троян може

бути легко реалізований ненадійними сторонніми постачальниками IP-адресів, ненадійним виробником та стороннім інструментом автоматизації електронного дизайну для виготовлення мікрочіпів для пристроїв Інтернету речей.

Вразливості апаратного забезпечення в основному характеризуються атаками на розширену інфраструктуру інтелектуального вимірювання (AMI) і можуть створити нову дірку для зловмисників, щоб зламати інформацію з електромережі. Основними наслідками атак AMI є крадіжка даних, викрадення електроенергії, локалізована або масова відмова в подачі електроенергії та порушення роботи електромережі.

Вразливості мережі характерні тим, що засоби кібербезпеки для мереж IoT не встигають за швидкістю підключення нижчих граничних вузлів. Будь-який геп в безпеці створить можливість для супротивника здійснити атаки та спричинити витік конфіденційної інформації. Відомі факти, що хакери змогли зламати пристрої IoT, такі як дверний замок та домашня камера відеонагляду. Зловмисники змогли перехопити імена користувачів і паролі через незашифрований протокол передачі гіпертексту (HTTP) у локальній мережі користувача, таким чином отримавши доступ до пристроїв IoT. Експерти галузі вважають, що легко порушити безпеку камери IoT і голосових помічників

Вразливості програмного забезпечення надзвичайно різноманітні, а наслідки атак через ці вразливості можуть мати катастрофічний характер для мережі IoT. Зловмисники можуть отримати контроль пристрої через вразливість програмного забезпечення. Обмежені можливості розміру цього дослідження не дозволять тут хоча б коротко описати всі варіанти вразливостей програмного забезпечення IoT. Тому далі розглянемо тільки одну з найбільших кіберзагроз для мереж IoT- DoS/DDoS-атаки. Головна небезпека переважної кількості DoS/DDoS-атак – в їх абсолютній очевидності і майже «нормальності». Помилки програмного забезпечення відразу ж фіксується і оперативно виправляються. А тут - тільки повна витрата ресурсів – це майже нормальна поведінка для сучасних інформаційних систем. Сервери налаштовують таким чином, щоб вони не простоювали, тому із вичерпанням ресурсів щоденно стикаються тисячі адміністраторів. Стратегія ж обмеження на трафік і ресурси неправильна, бо загрожує втратою частини клієнтів, яким буде відмовлено в обслуговуванні. Як же боротися з такою загрозою?

Реалізація системи виявлення вторгнень

Як показує аналіз: найефективнішими засобами для виявлення та запобігання DoS/DDoS атак є ті, які використовують інтелектуальні методи (IDMS), серед яких можна виділити: експертні системи, статистичний метод, методи на основі використання нейронних мереж, генетичних алгоритмів, та машинне навчання. Методи які базуються на статистиці спроможні змінювати свої оцінки залежно від поведінки суб'єкта, вони не вимагають знання про можливі атаки. Аналізатори з використанням нейронних мереж

можуть «вивчати» характеристики атак і класифікувати елементи, але точність і швидкість виявлення атак залежить від якості навчання цих нейромереж. У випадку використання ML витрачається набагато менше часу для виявлення, але вимагається висока обчислювальна потужність на машині, де аналізується поведінка [1], [2]. Завданням методів виявлення DDoS-атак полягає в тому, що аналізуючи дані, отримані з пристроїв, можна було б зробити найбільш достовірну оцінку відносно того, чи атакують пристрої в даний проміжок часу, чи ні. При цьому найбільш бажаним є швидкодія обраного методу, тому що найменша затримка може привести до значних втрат. Стандартні методи аналізу статистики не дозволяють виявляти невідомі раніше атаки, тому як інструмент для вирішення даної проблеми доцільно використовувати алгоритми машинного навчання [3]. Реалізація такої стратегії виявлення DoS/DDoS-атак відбувається через системи виявлення вторгнень (IDS). Мережеві системи виявлення вторгнень (NIDS) розміщуються як правило, на вузлі, який з'єднує внутрішню мережу з Інтернетом для того, щоб сканувати вхідний і вихідний трафік на предмет відомих моделей атак[8]. В міру виявлення відомих зловмисних шаблонів будуть відображуватися повідомлення та можуть бути запущені алгоритми експертизи, що зменшує час відповіді та потенційно підвищує точність розслідування. От чому IDS мають включати ML у свій процес виявлення аномалій з класифікацією, кластеризацією та іншими методами, що використовуються для виявлення аномального трафіку. Для створення прототипу реальної системи виявлення вторгнень на основі ML треба було розробити програмний застосунок та протестувати на відповідному наборі даних. Для експериментів ми обрали набір даних BotIoT, створеного у лабораторії Австралійського центру кібербезпеки (ACCS) спеціально для того щоб дослідники вивчали можливості машинного навчання у виявленнні атак в мережах IoT.

Причини такого вибору наступні: він відкритий, містить захоплення трафіку саме з мережі IoT, в ньому велика різноманітність атак, включаючи великий обсяг прикладів DoS/DDoS атак з використанням різних протоколів, він включає реальний трафік, та дає можливість генерувати нові ознаки з необробленого набору даних. Розмір опублікованих записаних файлів .pcap становить 69,3 ГБ і містить понад 72 000 000 записів. Проте, ми поки що використали тільки 5% вихідного набору даних, добутих за допомогою запитів MySQL. Ці добуті 5% складаються із чотирьох файлів сумарним розміром приблизно 1,07 ГБ і включають близько 3 мільйонів записів.

Для реалізації та тестування алгоритму виявлення DoS/DDoS атак в мережі IoT на основі ML було використано мову програмування Python. В основі програмного застосунку була покладена модель машинного навчання, яка складалася підготовки та збору даних, вибору ознак (параметрів), вибору алгоритму, вибору параметрів і моделей, навчання (тренування) та оцінки ефективності результатів. Всі ці етапи важливі, проте окремо виділимо етап вибору релевантних алгоритмів для вирішення

конкретної проблеми. На основі аналізу сучасних наукових досліджень [1],[3],[8],[9] було обрано наступні алгоритми:

- Метод k-найближчих сусідів (KNN);
- Наївний Баєсів класифікатор (NB);
- Випадковий ліс(Random Forest);
- Логістична регресія (Logistic Regression);
- Дерево ухвалення рішень (Decision Tree)

Оцінювання алгоритмів машинного навчання

При оцінці ефективності моделей ML, важливо визначити які показники ефективності підходять для вирішення завдання. Для того, щоб оцінити наші результати, ми використовували найважливіші показники ефективності: час передбачення, влучність (precision), повнота (recall), точність(accuracy) та f-міра(f-measure) як показано в визначені наступним чином [10]:

$$\text{влучність} = \frac{TP}{TP+FP} \quad , \quad \text{повнота} = \frac{TP}{TP+FN} \quad ,$$

$$\text{точність} = \frac{TP+TN}{TP+FP+TN+FN} \quad , \quad f - \text{міра} = \frac{2}{\frac{1}{\text{повнота}}+\frac{1}{\text{влучність}}} .$$

TP – істинно негативне, FN – хибно негативне, TN – істинно негативне, FP – хибно позитивне. Зокрема f-міра є найбільш значимою, бо вона є компромісним показником, щодо двох інших основоположних метрик: влучності та повноти.

Результати оцінювання обраних алгоритмів з використанням обраного датасету за цими п’ятьма метрикам не показали очевидної переваги когось з них. Для обгрунтованого вибору кращого треба застосувати методику, яка враховує багатокритеріальну природу такого вибору. В ситуації багатокритеріального вибору неочевидно, яке рішення краще, Тому потрібно шукати компромісне рішення, яке враховує важливість кожного критерія. Це приводить до поняття ефективного (оптимального щодо Парето) рішення. Властивість ефективності (у крайньому випадку слабку ефективність) повинне мати будь-яке рішення, що претендує на те, щоб його назвали найкращим.

Оскільки загального вирішення задачі вибору, оптимального по Парето не існує, то наука підтримки прийняття рішень пропонує багато методів, з множини яких треба обрати той, який би відповідав поставленій задачі. Серед множини методів багатокритеріально вибору можна виділити два класи: Методи , що орієнтуються на функцію корисності та методи, що орієнтуються на попарне порівняння критеріїв з наступною процедурою узгодження. В нашому випадку доцільно опертись на останній підхід, бо загальновизнана функція корисності для цієї задачі ще не існує. Одним з лідерів такого підходу є метод ELImination Et Choix Traduisant la Realite(ELECTRE)[11]. Цей метод характеризує чотири фундаментальні аспекти моделювання переваг Electre методи: моделювання ситуацій коливань

(часткових і всеосяжних), моделювання всебічних непорівнянностей, концепція узгодженості та відносної важливості критеріїв, а також концепція невідповідності та порогів вето.
Відзначимо, що процедури агрегації, для яких методи Electre краще пристосовані, тоді коли в моделі прийняття рішень включають не менше п'яти критеріїв.

У всіх модифікаціях методу ЕЛЕКТРА на першому етапі з допомогою особи, яка приймає рішення(ОПР) визначаються ваги критеріїв – позитивні дійсні числа, які тим більше, чим важливіше для ОПР відповідний критерій. Застосувавши певні експертні оцінки стосовно бінарних переваг та використавши попередні розрахунки значень часткових критеріїв, була проведена процедура вибору найкращого алгоритму ML для виявлення DoS/DDoS-атак. Для розрахунку результатів було використано пакет XLSSTAT, який надає інтерфейс користувача для застосування методів ELECTRE II та ELECTRE III. В результаті отримано наступні місця в рейтингу для алгоритмів:

Random Forest, кращий вибір за 4 інші алгоритми, Decision Tree – випереджає 3 інші альтернативи,

Logistic Regression - кращий за два алгоритми, KNN - кращий за один алгоритм. NB був останнім в рейтингу.

Висновки

Показано, що протидія кіберзагрозам є наразі актуальною задачею для мереж Інтернету речей. Особливу актуальність мають різновиди вразливостей програмного типу, а серед них - на перших місцях DoS/DDoS-атаки. Шляхом аналізу джерел встановлено, що ефективну протидію DoS/DDoS-атакам можна створити тільки на основі ML. Принципи моделі ML були закладені в розробку застосунка. Програма була в протестована на значній множині достовірного датасета, в якому були відображення наявності DoS/DDoS-атак. Були розраховані оцінки п’яти критеріїв для різних алгоритмів машинного навчання На підставі цих оцінок було обрано найкращий з алгоритмів методом багатокритеріального прийняття рішень ELECTRE.

Джерела

1. N. Moustafa, B. Turnbull, K.-K.R. Choo, Towards automation of vulnerability and exploitation identification in iiot networks, in: 2018 IEEE International Conference on Industrial Internet.
2. N. Moustafa, B. Turnbull, K.-K.R. Choo, An ensemble intrusion detection technique based on proposed statistical flow features for protecting network traffic of IoT, IEEE Internet Things J. (2018).
3. M. Tayyab, B. Belaton, and M. Anbar, ‘‘ICMPv6-based DoS and DDoS attacks detection using machine learning techniques, open challenges,

and blockchain applicability: A review,'' IEEE Access, vol. 8, pp. 170529–170547, 2020.
4. Ahmed, Sheikh. (2021). A Study of ML Algorithms for DDoS Detection. International Journal for Research in Applied Science and Engineering Technology.
5. Das, Saikat & Mahfouz, Ahmed & Venugopal, Deepak & Shiva, Sajjan. (2019). DDoS Intrusion Detection through ML Ensemble.
6. Kasinathan P. Denial of service detection in 6LoWPAN based internet of things [Текст]. / P. Kasinathan, C. Pastrone, M.A. Spirito and M. Vinkovits // – 2013. – pp. 600-607.
7. J. Czyz, M. J. Luckie, M. Allman, and M. Bailey, "Don't forget to lock the back door! a characterization of ipv6 network security policy", in NDSS, 2016.
8. K. Nickolaos, N. Moustafa, E. Sitnikova, and B. Turnbull. "Towards the development of realistic botnet dataset in the IoT for network forensic analytics: Bot-iot dataset." Future Generation Comp. Systems 100 (2019).
9. Alsemiri, Jadel & Alsubhi, Khalid. (2019). Internet of Things Cyber Attacks Detection using ML. International Journal of Advanced Comp. Science and Applications.
10. Mehryar Mohri, Afshin Rostamizadeh, & Ameet Talwalkar. (2018). Foundations of Machine Learning (2nd. ed.). The MIT Press.
11. Jose Figueira, Salvatore Greco, Bernard Roy, Roman Slowinski. Electre Methods: Main Features and Recent Developments. 2010. hal-00876980

INFORMATION SECURITY OF PRINTING ORGANIZATIONS

D.P.Kucherov [1], I.V. Ogirko [2], O.I. Ogirko [3]
[1]Faculty of Cybersecurity, Computer and Software Engineering
National Aviation University, Kyiv, Ukraine
[2]Ukrainian Academy of Printing, Lviv, Ukraine
[3]Lviv State University of Internal Affairs
d_kucherov@ukr.net, ogirko@gmail.com, ogirko@gmail.com

The paper discusses the main options for technical protection at the stages of pre-printing and post-printing preparation of products by a printing company using modern cloud technologies. In the study course, insufficient security of the SSL / TLS protocol for transferring data between websites was established, which is confirmed by numerous reports in the press about the theft of personal data and the constant improvement of this protocol. It was also established that data security has a dependency on the user's responsibility for personal information protection. In essence, options for protecting documents of a printing company can be supplemented by the introduction of digital signatures or passwords. Based on the theory of planning and processing the results of experiments, the whole factor experiment was prepared and carried out to analyze the password "strength" based on two signs of password strength: length and different types of categories of characters used in passwords. An experiment execution plan has been built based on an orthogonal matrix. According to the matrix plan, the unknown coefficients of the response function were determined; and we were determined their significance as well; and we have also confirmed the adequacy of the obtained equation. In the paper, we also present a model experiment is, confirming the validity of the hypothesis of the relativity of the importance of the message length concerning the types of categories of symbols used.

Introduction

Recently, there has been a sharp leap in the development of the printing industry due to the widespread introduction of digital content publications into the practice of publishing organizations, based on the consistent distribution of computer technology in the 90s of the last century [1].

The advent of computer publishing systems made it possible to transfer an electronic layout of a publication directly from a computer or to a photographic form, or a photographic plate, and later directly to a digital printing machine, which determined the prospect of obtaining a personalized publication of any required circulation. Digital processes (prepress, print, and post-press) at all manufacturing stages of printed products have created the basis for the transition to full automation of production of a printing company based on online solutions [2]. Innovations in computer technology offer to abandon traditional servers and

switch to Cloud Computing technologies. The problems of automating the task of accounting for the budgetary process to discharge are presented in works [3-15].

The emergence of cloud solutions along with several advantages, namely: an unlimited amount of data storage, independence from the operating system when working with documents, the possibility of constant and shared access to document, the rational use of resources, reduction in software costs, the ability to regularly update it; there was a strict need to protect deleted data.

So, it is possible to intercept confidential and private data when one's working with the "cloud"; and the internet provider can view the client's data if they are not protected by a password; this data can also become the property of hackers who have managed to break into the provider's security systems.

The reliability, timeliness, and availability of data in the "cloud" strongly depend on the intermediate parameters, which include data transmission channels on the way from the client to the "cloud"; the reliability of the last mile, the quality of the client's Internet provider, the availability of the "cloud" itself in the given moment. If the online storage company is liquidated, a computer customer can lose all their papers.

The purpose of this study is to establish an effective system for protecting printed products at all stages of their preparation.

Problem statement

The preparing process of a digital layout of products by some printing company using some publishing system is considered. A feature of the preparation process is its multistage, the essential component of which is prepress preparation. An electronic version of the final document is prepared in the prepress process. The electronic layout development of the document involves the participation of various staff members working with cloud storage that provides collective access.

As you know, the problem of storing data in the "cloud" is the data protection in the case of interception during transmission over a communication channel; there is also a problem associated with devices of the "last mile"; in addition, there is a problem with data integrity during remote storage.

The most common technical solutions for data protection are data encryption, closing documents with a password, and using a digital signature.

Usually, data transmission performs by the SSL/TLS protocol. It should be noted that despite the similarities between SSL and TLS, a series of technical documents (for example, RFC 8446) have recently been promoting TLS 1.3 for browser use. The TLS 1.3 protocol support for the webspace has been provided by Google, Facebook, and Cloudflare since July 2019, in addition, the main branches of the Chrome and Firefox browsers. The protocol supports three services: authentication, encryption, and message integrity checking. The suggested protective measures are considered sufficient by the safety experts. Nevertheless, the interception problem of messages is really, as evidenced by the numerous

confirmations of theft of personal data. Thus, the user himself should be primarily concerned about his data protection.

From the point of view of protecting printed information, one of the most effective measures is to set a password for a printing document. The problem here is the establishment of obvious passwords, which may be available to an attacker, or the creator of the documents does not consider it necessary to use a password here. To secure your information and not use external audit services, it is recommended that you independently acquire a strong password.

We must abide by some rules to avoid the negative consequences of data theft due to the introduction of "weak" passwords. To create a strong password, in [16] is recommended:

- avoid primitive passwords (repetitive words, English terminology, widespread digital combinations, etc.);
- avoid the same passwords for all services and resources;
- store the password in an open, visible place (on a table, on a monitor, in a browser, etc.);
- change the discredited password;
- do not use short passwords;
- a complex password must have a uniform distribution of alphanumeric and other symbols.

The study poses and solves the problem of determining an effective password based on the theory of planning and processing the results of experiments.

Problem solution

Based on the independence of the factors that affect the quality of the password, we will consider them independent, the combination of the same characters is not allowed, and all of the characters have the same distribution. Given the considerations, we assume that the password security function satisfies the regression equation in the form

$$y = b_0 + \sum_{i=1}^{k} b_i x_i + \sum_{i,j=1}^{k} b_{ij} x_i x_j + b_{123} x_1 x_2 x_3 \ldots, \quad (1)$$

According to the results of testing user passwords, it was found that the strength of passwords, which are quantitatively expressed in relative units y, are most significantly influenced by two factors: the length of the password x_1, measured by the number of characters in a password, and the number of categories of characters x_2, measured by various types of characters.

We will assume that at the user's disposal, lowercase and uppercase letters of the Latin alphabet can be used as symbols total of numbers are 26 characters, numbers are total numbers of 10 characters, and special characters are the total number have 33 characters. Thus, we have total of numbers 4 categories of characters.

Based on the theory of planning experiments, we also introduce the levels of variation by factors, the values of which are presented in Table. 1.

Table 1. Levels of factors

Levels of factors	Factors	
	x_1, symbols	x_2, categories
Main (zero)	14	3
Low	2	2
High	26	4
Interval of variation	12	1

The planning matrix looks as shown in Table 2. In this table j is the number of experiments, x_1 is the first sign, which means the number of symbols in the password; x_2 is the number of categories used in the password; y_i is the result of an experiment. In this test, we proposed that for each condition of experiment j we can obtain three results, therefore $\bar{y}_j$ is the average of this results; $\bar{S}_y^2$ is the variance characterizing the set of y_{jv} values under constant experimental conditions (i.e., at one point of the design) is found by the following formula

$$\bar{S}_j^2 = \frac{1}{m-1}\sum_{v=1}^{m}(y_{jv} - \bar{y}_j)^2 . \qquad (2)$$

Table 2 . Planning matrix

j	x_0	x_1	x_2	x_1x_2	y_i	$\bar{y}_j$	$\bar{S}_j^2$	$\hat{y}_j$
1	+	+	+	+	27, 31, 35	31	16	29.5
2	+	–	+	–	15, 20, 25	20	25	20.5
3	+	+	–	–	51, 53, 55	53	4	53.5
4	+	–	–	+	30, 33, 36	33	9	93.5

Based on Table 2 data and using the feature of the orthogonality planning matrix for the full factorial experiment, we can significantly simplify the calculation of the coefficients of the response equation. For the number of factors k, sample estimates b_i are calculated by the formulas

$$b_i = \frac{1}{n}\sum_{j=1}^{n} x_{ij}\bar{y}_j \ , \ \bar{y}_j = \frac{1}{m}\sum_{v=1}^{m} y_{jv} \ . \qquad (3)$$

Using formulas (3), we obtain

$$b_0 = \frac{1}{4}\sum_{j=1}^{4} x_{0j}\bar{y}_j = 34{,}25; \ \ b_1 = \frac{1}{4}\sum_{j=1}^{4} x_{1j}\bar{y}_j = \frac{31-20+53-33}{4} = 7{,}75;$$

$$b_2 = \frac{1}{4}\sum_{j=1}^{4} x_{2j}\bar{y}_j = \frac{31+20-53-33}{4} = -8{,}75;$$

$$b_{12} = \frac{1}{4}\sum_{j=1}^{4} x_{1j}x_{2j}\bar{y}_j = \frac{31-20-53+33}{4} = -2{,}25;$$

Next, we calculate $\bar{S}_y^2$ by using values $\bar{S}_j^2$ (these values are shown in Table 2)

$$S_y^2 = \frac{1}{4}\sum_{j=1}^{4} \bar{S}_j^2 = \frac{16+25+4+9}{4} = 13.5$$

Find the variance of the regression coefficients:

$$S_{b_i}^2 = \frac{1}{n(m-1)} S_y^2 = \frac{13.5}{4\cdot(3-1)} = 1.6875 \ , \ S_{b_i} = 1.299 \ .$$

We choose the level of reliability of calculations $\alpha = 0.95$ and find from the table value $t_{\frac{1+\alpha}{2}} = t_{0.975}$ at $f = 4\cdot(3-1) = 8$ degrees of freedom: $t_{0.975}(8) = 2.31$. Calculate Student Statistics for model coefficients:

$$t_0 = \frac{|b_0|}{S_{b_i}} = \frac{34{,}25}{1.299} = 26.37 \ ; \ t_1 = \frac{|b_1|}{S_{b_i}} = \frac{7.75}{1.299} = 5.966 \ ;$$

$$t_2 = \frac{|b_2|}{S_{b_i}} = \frac{8.75}{1.299} = 6.74 \ ;$$

$$t_{12} = \frac{|b_{12}|}{S_{b_i}} = \frac{2{,}25}{1.299} = 1.732 \ .$$

We see that $t_{12} = 1.54 < t_{0.975}(8) = 2.31$. Therefore, the coefficient b_{12} does not differ significantly from zero. Then the regression equation takes the following form

$$y = 34.25 + 7.75x_1 - 8.75x_2 .$$

Let's calculate now $\sum_{j=1}^{4}(\bar{y}_j - \hat{y}_j)^2 = 20.25$. Following the task, the number of significant coefficients of the model is d = 3; and $S^2 = \frac{3}{4-3}\sum_{j=1}^{n}(\bar{y}_j - \hat{y}_j)^2 = \frac{3 \cdot 20{,}25}{4-3} = 60{,}75$.

Now we find $F = \frac{S^2}{S_y^2} = \frac{60{,}75}{13.5} = 4{,}5$ that, less than the critical value $F < F_{0.05;\, m1=1,\, m2=8} = 5.32$, and the final formula for the response function is:

$$y = 51.46 + 0.62x_1 - 8.75x_2 . \quad (4)$$

Formula (4) allows you to evaluate the effectiveness of the entered password within the framework of the experiment. The analysis of equation (4), where the parameter x_1 varies from 4 to 6 symbols, and the parameter x_2 varies from 1 to 3 types of symbol categories, is presented in Table 3.

Table 3. Response function values for different types of passwords

Password	x_1, symbols	x_2, the numbers of categories	y
a1cd	4	2	36,44
a1Cd	4	3	27,69
a1bcd	5	2	37,06
a1Bcd	5	3	28,31
a1Bcd!	6	1	46,43

This analysis of the results obtained allows us to conclude that password efficiency is primarily affected by the password duration. The effect of mutual influence, based on the initial conditions of the experiment, practically does not

have a significant effect on the put forward condition of reliability, which made it possible to exclude this regressor from the equation.

Conclusions

The article proposes the centralization of technical computing resources, IT specialists, and software in the general budget cloud. Access to the cloud via the Internet, individualized access to data, a set of software functions provided by the bandwidth of network communication channels will be transferred to the sites.

To increase the security of printed products at the stages of pre-print and post-print preparation, it is recommended to use passwords consisting of a large number of characters, including lowercase and uppercase letters of the Latin alphabet, numbers, and auxiliary symbols. Based on a whole factor experiment, a type of response function has been established that allows you to compare passwords and set the strength of a password.

Referances

1. Armbrust M., Fox A., Griffith R., Joseph A. D., Katz R., Konwinski A., Lee G., Patterson D., Rabkin A., Stoica I., and Zaharia M. A View of Cloud Computing / M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia // ACM Communications, vol. 53, pp. 50-58, 2010.
2. Machuga P.I. Virtualizacia i hmarni technologii v obliku: daleke maybutne chi realne cegodennya? / P.I. Machuga // Efektivna ekonomika. – №5. – 2013. (Мачуга, Р. І. Віртуалізація і хмарні технології в обліку: далеке майбутнє чи реальне сьогодення? / Р. І. Мачуга // Ефективна економіка. – 2013. – №5.) – [On-line recourse]. – Access: http://www.economy. nayka.com.ua/?op=1&z=2008. (in Ukranian)
3. Bondar E.S. Hmarni obchislennya ta ih zastosuvannya / E.S. Bodar, M.M. Glubovetc, S.S. Gorohovskii // Visnik KNU im. T. Shevchenka. – Vip. 1. – K.: KNU, 2011. – S. 74-82. (Бондар Є. С. Хмарні обчислення та їх застосування / Є. С. Бондар, М. М. Глибовець, С. С. Гороховський // Вісник КНУ ім. Т. Шевченка. – Вип. № 1. – К.: КНУ, 2011. – С. 74–82.) (in Ukranian)
4. Voloshchuk L.A. Support for the decision making on implementation of applications in the hybrid cloud infrastructure / L.A. Voloshchuk, O.I. Roznovets, D.D. Voloshchuk // Informatics and Mathematical Methods in Simulation. – Vol. 8. – No. 1. – 2018. – P. 86–97.
5. Ramgovind S., Eloff M. M., and Smith E. The Management of Security in Cloud computing / S. Ramgovind, M. M. Eloff, and E. Smith // in Information Security for South Africa (ISSA). – 2010. – P. 1–7.

6. Strategichni napryami rozvitku pidpriemstv vidanichoi galuzi, poligrafichnoi diyalnosti s knigotorgivli // Za red. Yu. S. Ganzhurova. – Electronne naukove vidannya. – K.: NTUU «KPI», 2015. – 252 s. (Стратегічні напрями розвитку підприємств видавничої галузі, поліграфічної діяльності і книготоргівлі // За ред. Ю. С. Ганжурова, - Електронне наукове видання. -К: НТУУ «КПІ», 2015. – 252 с.) (in Ukranian)
7. Zmihov Ya.Yu., Shablyi I.V., Ogirko I.V. Formati danih avtomatizovanoi sustemi keruvannya drukarneu / Ya.Yu. Zmihov, I.V. Shablyi, I.V.Ogirko // Kvalilogia knigi. Pbirnik naukovih pratc. Ukrainska academia drukarctva. – Vip. 2 (34). – 2018. – S. 51–56. (Жмихов Я. Ю., Шаблій І. В., Огірко І. В. Формати даних автоматизованої системи керування друкарнею. Кваліогія книги. Збірник наукових праць українська академія друкарства. – Випуск № 2 (34). – 2018. – С. 51–56.) (in Ukranian)
8. Wyld, D. C. Moving to the Cloud: An Introduction to Cloud Computing in Government / D. C. Wyld. – IBM Center for the Business of Government, 2009. – P. 10.
9. Gudzovata O.O. Hmarni servisi: mozhlivosti, bezpeka, percpektivi: kollektivnaya monografia: u 4 t. / O.O. Gudzovata // Teoretichni ta prikladni aspekti pidvischennya konkurentospromozhnosti pidpriemstv. – Dnipropetrovsk: «Gerda», 2013. – T. 1. – 352 s. – S. 102–110. (Гудзовата О. О. Хмарні сервіси: можливості, безпека, перспективи: колективна монографія: у 4 т. / О. О. Гудзовата // Теоретичні та прикладні аспекти підвищення конкурентоспроможності підприємств. – Дніпропетровськ: «Герда», 2013. – Т. 1 – 352 с. – С. 102–110.) (in Ukranian)
10. Kaufman L. M. Can Public-Cloud Security Meet Its Unique Challenges? / L. M. Kaufman // IEEE Security & Privacy. – Vol. 8. – 2010. – P. 55–57.
11. Anthes G. Security in the Cloud / G. Anthes // ACM Communications. – Vol. 53. – 2010. – P. 16–18.
12. Christodorescu M., Sailer R., Schales D. L., Sgandurra D., and Zamboni D. Cloud Security is not (just) Virtualization Security: A Short Paper / M. Christodorescu, R. Sailer, D. L. Schales, D. Sgandurra, and D. Zamboni // in Proceedings of the 2009 ACM workshop on Cloud Computing Security, Chicago, Illinois, USA, 2009. – P. 97–102.
13. Shaver C. and Acken J. M. Effects of Equipment Variation on Speaker Recognition Error Rates / C. Shaver and J. M. Acken // in Proceeding IEEE International Conference on Acoustics Speech and Signal Processing, Dallas, Texas, 2010.
14. Ristenpart T., Tromer E., Shacham H., and Savage S. Hey, you, get off of my cloud: Exploring Information Leakage in Third-party Compute Clouds / T. Ristenpart, E. Tromer, H. Shacham, and S. Savage // in

Proceedings of the 16th ACM conference on Computer and Communications Security, Chicago, Illinois, USA, 2009. – P. 199–212.
15. Kucherov D.P. Control of computer network overload / D.P. Kucherov // Proceeding of the XVII International Scientific and Practical Conference on Information Technologies and Security (ITS 2017), Kyiv, Ukraine, November 30, 2017, pp. 69 -75, 2017, CEUR-WS.org, online. URL: http://ceur-ws.org/Vol-2067/paper10.pdf
16. Paroli – horoshie, plohie I uzhasnie. (Пароли – хорошие, плохие и ужасные) – [on-line]. - Access: https://itglobal.com/ru-ru/company/blog/paroli-horoshie-plohie-i-uzhasnye/ (in Russian)
17. Kobzar A.I. Prikladnaya matematicheskaya statistika. Dlya inzhnerov I nauchnih rabotnikov / A.I. Kobzar // M.: Fizmatlit, 2006. – 816 c. (Кобзарь А.И. Прикладная математическая статистика. Для инженеров и научных работников / А.И. Кобзарь. – М.: Физматлит, 2006. – 816 с.) (in Russian)

A BRIEF OVERVIEW OF MODELS OF INFORMATION INFLUENCE IN SOCIAL NETWORKS

V. Putyatin[0000-0002-5575-9159]
Institute for Information Recording of NAS of Ukraine

1. Formulation of the problem

Staging aboutRecently, social networks are attracting more and more attention from various researchers around the world. At the same time, there is a significant interest in a wide range of different models related to the network activity of their participants. Analysis of a number of sources has shown that researchers use different classes of models of informational influence in social networks. The number of works devoted to various aspects of impact models is growing rapidly

2. The goal of the work

The purpose of this work is a brief overview of the most common models of influence on social networks.

3. Presentation of the main results of the study

The author's research has shown that at present there are many models of informational influence in social networks such as [1-3]:

1) ***Optimization and simulation models:*** models with thresholds, models of independent cascades, models based on analogies with physics, medicine, other branches of science (models of impregnation and infection; Ising models); models based on cellular automata; models based on Markov chains; models of indirect influence in the social network [4-6];

2) ***Theoretical game models:*** models of mutual awareness; models of concerted collective action; communication models; network stability models; information impact and management models; models of reputation and information management in social networks; models of information confrontation [7-10];

3) ***Other models:*** reputation models; models of thought dynamics; model of distribution of mass media messages on the Internet; Schelling models; model of preferences of groups of people; model of the type of "thermal beetles"; model of organization of collective behavior;

4) ***Models of information influence:*** manipulative model; dialogue model; individual-oriented models; multiagent models.

5) ***Models of political crisis*** (threshold models, conflict models, Epstein model).

Here are some examples of such models.

Model of independent cascades. Belongs to the category of models of so-called "interacting particle systems" [6].

Threshold models of racial segregation by T. Schelling. This is a model of spatial neighborhood and a model of a limited environment [11-13].

Impact models with thresholds. In the model [13], the nodes of the network (graph) can be in the active and inactive state, and the transition is possible only in

one direction: from inactive to active state. A node that is in an inactive state is affected by all its active neighbors. A node becomes active if the sum of the effects of neighboring nodes on it exceeds a certain threshold. A threshold model is any model that has a threshold value or set of threshold values used to change states. Classical models with thresholds were developed by Schelling [11-13], Axelrod [2] and Granovetter [14] to model collective behavior.

Model of preferences of groups of people. According to this model [14] the advantages of groups of people were studied. At the same time, it was initially assumed that the groups differed only in ethnicity. However, the model can also take into account any other types of differences in which individual group membership is visible and stable. In this Axelrod-Hammond model, the agent is the individual. Each agent is "painted in colors," which can be interpreted as his or her ethnicity or other sign of group membership. Each agent also has a two-part strategy. The first part of the strategy determines whether the agent cooperates (or not) with a neighbor who has the same colors. The second part of the agent's strategy determines whether the agent is working with a neighbor who has a different color.

Models of collective action. These are models of concerted collective action; model of organization of collective behavior; model of dynamics of collective behavior; Granovetter's crowd model [14], a model of the "heat beetle" type; Granovetter's threshold model of collective behavior; game-theoretic threshold model of conformal behavior; Mueller's dialogue model; models of mutual awareness.

Reputation models. This is a model of J. Tyrol's reputation; Shapiro-Stiglitz reputation model; models of reputation and information management in social networks [18].

Individual-oriented models. These are multi-agent simulation models in which the integral characteristics of the population are the result of many local interactions of agents (individuals). The model [15] consists of a description of the environment in which interactions take place and a set of individuals for whom rules of conduct and characteristic parameters are defined, which may also change in the process of evolution. One of the most popular models of artificial society in the framework of individual-oriented modeling is the so-called sugar model of J. Epstein and R. Axtel [24]. Despite its simplicity, it is a powerful tool for analyzing social processes.

Ising model. This is a mathematical model [7] that describes the origin of the magnetization of the material (quantum Ising model), proposed in 1924 to describe the magnetic phase transitions). In the original Ising model we consider a system of N identical particles in nodes of a regular two-dimensional lattice with magnetic moments ("spins") $\{si = \pm 1\}$ and energy $Ei = -\Sigma Jsisj - hsi$, where J is the exchange parameter, h is the external magnetic field and summation is performed on particles {j} in adjacent lattice nodes. Schemes based on the Ising model are applied to sociology problems in which the state of the agent is reflected by a discrete variable (choice of several strategies, products, policies of changing agents' opinions under the influence of their immediate environment and

(or) averaged "social field"). in a large social group can be modeled using the Ising model, the influence of close neighbors is decisive, and the analogue of temperature is the group's willingness to think creatively, willingness to accept new ideas.

Impregnation and infection models. The classical (epidemiological) model of the spread of the epidemic [32, 33] is based on the following cycle of the carrier's disease: first, the person is susceptible to the disease; if he comes in contact with an infected person, he becomes infected with some probability of β; later, after some time, the person becomes healthy, acquires immunity, or dies; immunity decreases over time, and the person becomes susceptible to the disease again. In the SIR model (according to the first letters of the three stages of the disease cycle) the convalescent becomes immune to the disease: $S \rightarrow I \rightarrow R$. Accordingly, society is represented by three groups: S (t) - a group of people not yet infected or susceptible to the disease at time t; I (t) - group of infected people; R (t) - a group of recovered people. There are other similar more complex models, in particular, in the SIRS model the recovered person becomes susceptible to the disease after some time. The SIR model was proposed by Kermak and McKendrick in 1927. The entire population is divided into groups: "Susceptible" - healthy people susceptible to the disease (hereinafter S); "Infectious" - infectious patients (hereinafter I); "Recovered" - people who have become ill with the simulated disease are no longer susceptible to it (hereinafter R). Given that the total number of individuals in the population remains unchanged, the increase in the number of people in each of the selected groups can be described using the following system of equations: $dS / dt = -\beta SI$, $dI / dt = \beta SI$, $dR / dt = \gamma I$, where β - intensity of contacts between people, γ - intensity of recovery of agents. Impregnation and infection models are a popular way to study information dissemination and innovation on social media.

Models of public opinion. This is the voter model; majority model; "Schneid model", voting model [17]. At the heart of all models is a random change in the agent's opinion under the influence of the environment. In models, the original opinion of the majority usually wins; on complex networks "metastable" areas with polarization of thoughts are possible. In the voter model, first proposed in evolutionary biology, the dynamics of public opinion M (t) is reproduced by a stochastic process, at each step of which a randomly chosen agent randomly accepts the opinion of one of his neighbors on the playing field. "Discrete" preferences of people in sociophysics are often analyzed using various modifications of the quantum model of Ising. In the voting model, which belongs to the class of models of "interacting particle systems", at each step, each agent can change his mind, randomly selecting one of the neighbors and accept his opinion. This model is similar to the threshold model in the sense that the agent is more likely to change his mind to the opinion that is supported by most of his neighbors. However, in the voting model, in contrast to the threshold model, the agent may become inactive. In the model proposed by Galam [28], a majority of the general structured population randomly identifies a “discussion group” with an odd number of agents in adjacent lattice nodes, in which all agents accept the

opinion of the group majority; then the process is repeated. In the version of the majority model, named after the authors called "Schneid" model, a randomly selected "reference group" of two neighboring agents in the presence of a single opinion induces this opinion on all nearby nodes, and a pair of agents with opposing views does not affect neighbors.

Model of social influence. According to the i-th agent, the whole set of lattice nodes is affected; the force of influence of the j-th agent is proportional to its "convincing ability" αj and decreases with increasing "social distance" dij between agents. In all such models, the lattice or other structure that serves as a "medium" for disseminating ideas reflects the location of agents in some abstract "public space."

Model of information influence in a social network. Agents included in the social network are described in the plural *N = {1, 2, ..., n}.* Agents in the network affect each other, and the degree of influence is given by the matrix of direct influence t dimension $n \times n$, де $t_{ij} \geq 0$ indicates the degree of confidence of the i-th agent of the j-th agent (the influence of the j-th on the i-th). It is considered [5] that the rationing condition is fulfilled: $\forall i \in N \sum_{j=1}^{n} t_{ij}$. If the i-th agent trusts the j-th, and the j-th trusts the k-th, it means the following: the k-th agent indirectly affects the i-th.

Model of measuring the impact on social networks. Agents included in the social network [30] are represented by many *N* = (1, 2, .., *n*). The influence of agents is given by the matrix of direct influence v of dimension n * n, while vij> 0 characterizes the degree of influence of the j-th agent on the i-th agent or the degree of confidence of the i-th agent of the j-th agent. The concepts of "influence" and "trust" are considered in this case opposite. Assuming that each agent reliably knows only their trust parameters, ie the i-th line of the matrix v, which presents data on how much and to whom he trusts.

Models of influence based on Markov chains. Model [10] is a dynamic Bayesian network with a two-level structure: the level of individuals (simulated actions of each agent) and the level of the group (simulated actions of the group as a whole). Processes modeled by Markov circuits are processes with short memory.

Social network dissemination model. The members of the network are agents (users) who perform certain actions (writing a post, writing a comment to the post, etc.) from a fixed finite set at certain times [17]. The set of actions contains a binary relation "action a is the cause of action b" (or, equivalently, "action b is a consequence of action a"), which is denoted as follows: a → b. Based on the determination of the influence of sets of actions and agents, the influence of agents on each other is determined as a fraction of the influence of one agent, which is due to the actions of another agent (user).

Multiagent models of information influence. Models [34] include studies of the interaction of information agents with each other, as well as with the external environment.

1. *Потемкин А.В.* Выявление информационных операций в средствах массовой информации сети интернет. Кандидатская диссертация. Орел, 2015. – 144с.

2. *Axelrod R.* Modeling Security Issues of Central Asia // Gerald R. Ford School of Public Policy University of Michigan (Project on “Security in Central Asia” June 2004. US Govt. Contract # 2003*H513400*000)

3. *Epstein J.M.* Modeling civil violence: an agent-based computational approach, Proc Natl Acad Sci USA. 2002, 99, Suppl. 3, 7243-7250

4. *Борщев А.В.* Практическое агентное моделирование и его место в арсенале аналитика // Exponenta Pro, 2004. – № 3, 4.

5. *Губанов, Д.А., Новиков, Д.А., Чхартишвили, А.Г.* Социальные сети: модели информационного влияния, управления и противоборства. — М.: Издат. физико-математической литературы, 2010. - 228 с.

6. *Губанов Д.А., Новиков Д.А., Чхартишвили А.Г.* Модели информационного влияния и информационного управления в социальных сетях // Проблемы управления. 2009. №5. С. 28-35.

7. *Tarnow, E.* Like Water and Vapor, Conformity and Independence in the Large Group [Electronic resource] / E. Tarnow; Access mode: URL: http:/cogprints.org/4274/l/LargeGroupOrderTarnow.pdf.

8. *Goldenberg, J., Libai, B., Muller, E.* Talk of the Network: A Complex Systems Look at the Underlying Process of Word-of-Mouth [Text] / J. Goldenberg, B. Libai, E. Muller; Marketing Letters - 2001. - № 2. - P. 11-34.

9. *Бреер В.В., Новиков Д.А., Рогаткин А.Д.* Управление толпой: математические модели порогового коллективного поведения. – М.: ЛЕНАНД, 2016. – 168 с.

10. *Губанов Д.А., Новиков Д.А., Чхартишвили А.Г.* Модели влияния в социальных сетях. Управление большими системами. Вып.27. М.: ИПУ РАН. 2009. С.205-281.

11. *Schelling T.*, A process of residential segregation: Neighborhood tipping, in Racial Discrimination, in: A. Pascal (Eds.), Economic Life, Lexington, MA: Lexington Books. 1972.

12. *Schelling T., Hockey Helmets*, Concealed Weapons, and Daylight Saving: A Study of Binary Choices with Externalities. The Journal of Conflict Resolution, Vol. 17, No. 3 (Sep., 1973), pp. 381-428.

13. *Schelling T.*, Dynamic models of segregation J. Math. Sociology, 1, 143–186. 1971

14. *Granovetter M.* Threshold Models of Collective Behavior // American Journal of Sociology. 1978. Vol. 83. No. 6. P. 1420-1443].

15. *Grimm V.* Ten years of individual-based modelling in ecology: what have we learned and what could we learn in the future // Ecological Modelling, 2002, 1999. – Vol. 115 (2-3). – P. 129-148.

16. *Siettos C.I., Russo L.* Mathematical modeling of infectious disease dynamics // Virulence. — 2013. — Vol. 4, № 4. — P. 1–12.

17. *Абрамов К.Г.* Модели угрозы распространения запрещенной информации в информационно-телекоммуникационных сетях. Кандидатская диссертация. Владимир, 2014. – 129с.

18. *Губанов Д.А.* Обзор онлайновых систем репутации/доверия. – М.: ИПУ РАН, 2009 / Интернет-конференция по проблемам управления (www.mtas.ru/forum). – 25 с.

19. *Bailey N.* The Mathematical Theory of Infectious Diseases and Its Applications. – New York: Hafner Press, 1975.

20. *Friedkin N.E.* Structural Cohesion and Equivalence Explanations of Social Homogeneity // Sociological Methods and Research. 1984. No 12. P. 235-261.

21. *Goldenberg J., Libai B., Muller E.* Talk of the Network: A Complex Systems Look at the Underlying Process of Word-of-Mouth // Marketing Letters. 2001. № 2. P. 11-34.

22. *Zhang D., Gatica-Perez D., Bengio S., Roy D.* Learning Influence among Interacting Markov Chains [Text] / D. Zhang, D. Gatica-Perez, S. Bengio, D. Roy; Neural Information Processing Systems (NIPS). -2005. - P. 132-141.

23. *Epstein J.M.* Remarks on the foundations of agent-based generative social science // SFI (Santa Fe Institute) Working Paper. DOI: SFIWP 05-06-024, SFI Working Papers, 2005.

24. *Epstein J. M., Axtell R.* Artificial societies and generative social science // Artificial Life and Robotics. – Vol., № 1 / March, 1997. – P. 33-34.

25. *Горбулін В. П., Додонов О. Г., Ланде Д. В.* Інформаційні операції та безпека суспільства: загрози, протидія, моделювання: монографія. — К.: Інтертехнологія, 2009. — 164 с.

26. *Goldstone, R. L., & Janssen, M. A.* (2005). Computational models of collective behavior. *Trends in Cognitive Sciences*, *9*(9), 424-430.

27. *Новиков Д.А.* Математические модели формирования и функционирования команд. - М.: Издательство физико-математической литературы, 2008. - 184 с.

28. *S.Galam.* Sociophysics: a physicist's modeling of psycho-political phenomena. 2012, Springer, 536 p.

29. *Михайлов А. П., Петров А. П., Маревцева Н. А.* Моделирование распространения информации и информационного противоборства в структурированном социуме // Международная научно-практическая конференция Теория

активных систем (ТАС-2014) Институт проблем управления им. В.А. Трапезникова РАН Москва. — 2014. — С. 209–2010.

30. *Ломакин М.И., А.В. Докукин А.В., Г.А. Соседов Г.А., Н.В. Шинелин Н.В.* Модель измерения влияния в социальных сетях. Компетентность: Научно-практический журнал. - М.: Академия стандартизации, метрологии и сертификации (учебная). - 2014 .- №7 . – С.34-40.

31. *Словохотов Ю.Л.* Физика и социофизика // Проблемы управления. — 2012. — Ч. 1. — № 1. — С. 2—20; Ч. 2. — № 2. — С. 2—31; Ч. 3. — № 3. — С. 2—34.

32. *Кондратьев М.А.* Методы прогнозирования и модели распространения заболеваний. Компьютерные исследования и моделирование 2013 Т. 5 № 5 С. 863–882.

33. *Coburn B.J., Wagner B.G., Blower S.* Modeling influenza epidemics and pandemics: insights into the future of swine flu (H1N1) // BMC Medicine. — 2009. — Vol. 7, № 30.

34. *Додонов В.А., Ландэ Д.В.* Мультиагентные модели информационного влияния // Информационные технологии и безопасность. Материалы XV Международной научно-практической конференции ИТБ-2015. - К.:ИПРИ НАН Украины, 2015. - С. 75-83.

ІНСТРУМЕНТАЛЬНІ ЗАСОБИ ANALYTIC SOLVER DATA MINING В КЕРОВАНІЙ ДАНИМИ БЕЗПЕКОВІЙ АНАЛІТИЦІ: ОГЛЯД ТА ЗАСТОСУВАННЯ

Кузьмичов Анатолій Іванович[1], Kuzmychov Danyl[2]
[1]Інститут проблем реєстрації інформації НАНУ, м. Київ, Україна,
[2]California State University, San Francisco, USA
akuzmychov@gmail.com, kuzmychovd@gmail.com

Практичну цінність науки про дані та засоби дата-аналітики якнайкраще сприймають за навчально-дослідницьким практикумом, здійсненим на звичній інструментальній платформі. Зокрема, світовий освітній досвід посилається на доступне кожному середовище Excel із використанням інструментальної платформи ASP, у складі якої – надбудова Analytic Solver Data Mining (XLMiner), її навчальна версія забезпечує вступний курс дата-аналітики із розділами: передобробка наборів даних великого розміру, їх класифікація та передбачення, за методикою машинного навчання із застосуванням 50 різних інструментів.

Цими досконалими інструментами відшукуються «брудні» записи табличної бази даних, вміст яких викликає підозру й змушує здійснювати їх корегування чи вилучати і детально вивчати.

Передобробка (очищення, перетворення типів) та розвідування (візуалізація, зниження розмірності, відбір ключових змінних, кластеризація) наборів «сирих» даних – характерна властивість і перший етап безпекової аналітики на шляху до отримання корисної інформації про об'єкт, представлений багатовимірною табличною базою даних.

«Стиснення» набору даних методом *PCA*

Набір даних – числова таблиця із n рядків (об'єктів) та p стовпців (властивостей, змінних) X_1, …, X_p із багатовимірним, необов'язково нормальним, сумісним розподілом значень x_{ij}, її «недолік» – велика кількість змінних як ускладнення розрахунків.

Стиснення набору даних за методом PCA (*Principal component analysis*, К. Пірсон, 1901 р.)

здійснюється оптимальним переходом (за визначеною змінною) від p заданих до q нових змінних, значення q (число «головних»), $q < p$, визначає дослідник, вимір n зберігається.

Інструмент: *Data Analysis*→ *Transform* → *Principal Components Analysis*

Файл *nndb-flat.csv* (джерело: *USDA National Nutrient Database*) містить очищені записи про різноманітні харчові продукти: n = 8618, p = 38.

Сформовано і упорядковано 17 нових змінних (компонент) та таблицю шуканих вагових коефіцієнтів заданих змінних розміром (8618 × 17):

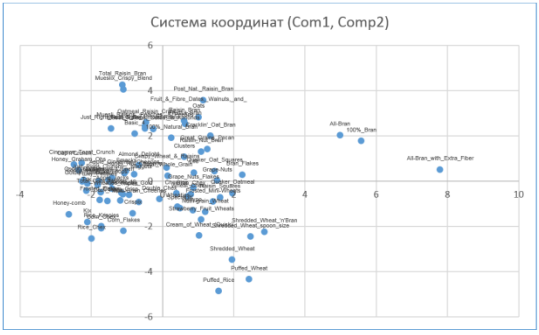
75	Explained Variance			
76	Component	Eigenvalue	Variance, %	Cumulative Variance, %
77	Component 1	9.87	25.97	25.97
78	Component 2	4.11	10.83	36.80
79	Component 3	3.43	9.02	45.82
80	Component 4	2.93	7.72	53.53
81	Component 5	2.38	6.27	59.80
82	Component 6	2.09	5.51	65.31
83	Component 7	1.74	4.59	69.90
84	Component 8	1.63	4.28	74.18

	3	4	5	6	19	20
34	Principal Components					
35	Feature\Component	Component 1	Component 2	Component	Component 16	Component 17
36	Energy_kcal	0,093	0,107	-0,145	-0,0494	-0,0220
37	Protein_g	0,111	-0,122	-0,149	0,1604	-0,1818
38	Fat_g	0,013	0,018	-0,136	-0,0376	0,1701
39	Carb_g	0,099	0,207	-0,009	-0,1200	-0,1961
40	Sugar_g	0,039	0,137	0,038	0,2134	0,1576
41	Fiber_g	0,112	0,142	-0,138	-0,4715	-0,3232
42	VitA_mcg	0,109	-0,338	0,064	-0,0701	-0,1408
43	VitB6_mg	0,240	0,083	0,123	-0,2164	0,3331
44	VitB12_mcg	0,148	-0,359	0,066	-0,1011	0,0569
45	VitC_mg	0,070	0,052	0,086	-0,0468	-0,0638
46	VitE_mg	0,099	0,078	-0,049	0,0420	-0,1248
47	Folate_mcg	0,213	0,131	0,165	0,1056	-0,0509
48	Niacin_mg	0,259	0,059	0,146	0,1666	-0,0617

,

першій й найсильнішій компоненті (*Component 1*) припадає 26% змінюваності, а разом із наступними 3-ма компонентами – більше 50%. Якщо вирішено, що достатньо цих 4-ох «головних» компонент (q = 4), значить, набір даних стиснуто до розміру (8618 × 4) зі змінюваністю 53,53%, яку буде використано іншими інструментами.

Достатньо виокремити перші дві головні компоненти (інший аналогічний приклад, 52% змінюваності) як нові змінні (Com1, Com2), щоби наочно розрізнити об’єкти (продукти) на точковій діаграмі:



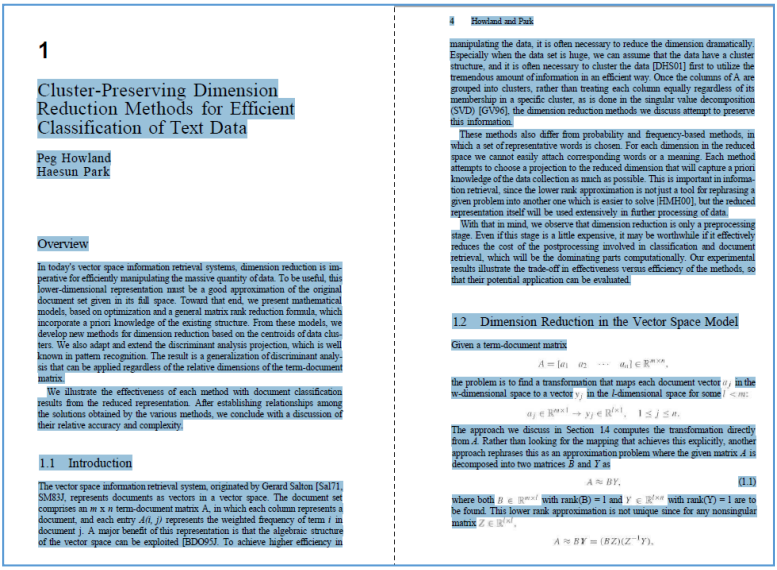
- перша головна компонента ***Com1*** (горизонтальна вісь) визначає корисність продукту, яка зростає зліва направо, кращими, у вигляді викидів, виявилися три продукти на основі висівок (*Bran*)
- друга головна компонента ***Com2*** (вертикальна вісь) визначає зростання знизу догори структурних властивостей продуктів, від подрібнених до цілих зерен.

Щільні скупчення – близькі за сукупними властивостями продукти (калорії, цукор, жири, вітаміни).

Text Mining

Мета – зрозуміти, про що йдеться у текстовому документі, отриманням топ ключових слів (у вигляді термів). Інструментом *Text Miner* отримують шукану інформацію із неструктурованих текстових даних їх попереднім перетворенням в придатний для поглибленої аналітичної роботи набір структурованих (табличних) даних.

Вхідний текст (.PDF) та топ-10 ключових термів:



1

Cluster-Preserving Dimension Reduction Methods for Efficient Classification of Text Data

Peg Howland
Haesun Park

Overview

In today's vector space information retrieval systems, dimension reduction is imperative for efficiently manipulating the massive quantity of data. To be useful, this lower-dimensional representation must be a good approximation of the original document set given in its full space. Toward that end, we present mathematical models, based on optimization and a general matrix rank reduction formula, which incorporate a priori knowledge of the existing structure. From these models, we develop new methods for dimension reduction based on the centroids of data clusters. We also adapt and extend the discriminant analysis projection, which is well known in pattern recognition. The result is a generalization of discriminant analysis that can be applied regardless of the relative dimensions of the term-document matrix.

We illustrate the effectiveness of each method with document classification results from the reduced representation. After establishing relationships among the solutions obtained by the various methods, we conclude with a discussion of their relative accuracy and complexity.

1.1 Introduction

The vector space information retrieval system, originated by Gerard Salton [Sal71, SM83], represents documents as vectors in a vector space. The document set comprises an $m \times n$ term-document matrix A, in which each column represents a document, and each entry $A(i, j)$ represents the weighted frequency of term i in document j. A major benefit of this representation is that the algebraic structure of the vector space can be exploited [BDO95]. To achieve higher efficiency in

4 Howland and Park

manipulating the data, it is often necessary to reduce the dimension dramatically. Especially when the data set is huge, we can assume that the data have a cluster structure, and it is often necessary to cluster the data [DHS01] first to utilize the tremendous amount of information in an efficient way. Once the columns of A are grouped into clusters, rather than treating each column equally regardless of its membership in a specific cluster, as is done in the singular value decomposition (SVD) [GV96], the dimension reduction methods we discuss attempt to preserve this information.

These methods also differ from probability and frequency-based methods, in which a set of representative words is chosen. For each dimension in the reduced space we cannot easily attach corresponding words or a meaning. Each method attempts to choose a projection to the reduced dimension that will capture a priori knowledge of the data collection as much as possible. This is important in information retrieval, since the lower rank approximation is not just a tool for rephrasing a given problem into another one which is easier to solve [HMH00], but the reduced representation itself will be used extensively in further processing of data.

With that in mind, we observe that dimension reduction is only a preprocessing stage. Even if this stage is a little expensive, it may be worthwhile if it effectively reduces the cost of the postprocessing involved in classification and document retrieval, which will be the dominating parts computationally. Our experimental results illustrate the trade-off in effectiveness versus efficiency of the methods, so that their potential application can be evaluated.

1.2 Dimension Reduction in the Vector Space Model

Given a term-document matrix

$$A = [a_1 \quad a_2 \quad \cdots \quad a_n] \in \mathbb{R}^{m \times n},$$

the problem is to find a transformation that maps each document vector a_j in the m-dimensional space to a vector y_j in the l-dimensional space for some $l < m$:

$$a_j \in \mathbb{R}^{m \times 1} \to y_j \in \mathbb{R}^{l \times 1}, \quad 1 \le j \le n.$$

The approach we discuss in Section 1.4 computes the transformation directly from A. Rather than looking for the mapping that achieves this explicitly, another approach rephrases this as an approximation problem where the given matrix A is decomposed into two matrices B and Y as

$$A \approx BY, \tag{1.1}$$

where both $B \in \mathbb{R}^{m \times l}$ with rank(B) = l and $Y \in \mathbb{R}^{l \times n}$ with rank(Y) = l are to be found. This lower rank approximation is not unique since for any nonsingular matrix $Z \in \mathbb{R}^{l \times l}$,

$$A \approx BY = (BZ)(Z^{-1}Y).$$

Term	Collection Frequency
space	9
vector	8
reduct	7
reduc	6
rank	5
term	4
retriev	4
repres	4
represent	4
set	4

Нейромережева класифікація

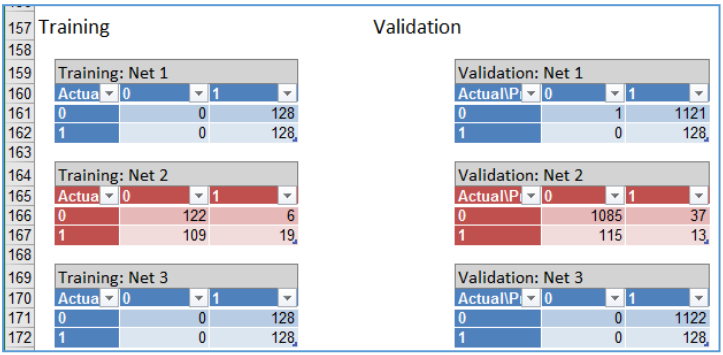
Набір даних (таблична база даних) складається із 2500 записів, кожен має значення 14 змінних й клас, де успішних записів класу 1 усього 256 (10%). За машинним навчанням набір даних попередньо розділяють в режимі Oversamling на дві основні вибірки: *Training set* (256, з них 50% успішних) та *Validation set* (1250), представлені у діапазоні *Partitioned Data*.

Результат
Сформовано 90 нейромереж із різною конфігурацією, кожна із відповідним % помилок класифікації записів (усіх класів):

Architecture Search Error Log

NetID	# Hidden Layers	# Neurons (Layer 1)	# Neurons (Layer	Training # Errors	Training % Error	Training % Sensitiv	Training % Specific	Training % Precisi	Training % F1-Score	Validation # Errors	Validation % Error
Net 1	1	1	0	128	50	100	0	50	66,66666667	1121	89,68
Net 2	1	2	0	115	44,921875	14,84375	95,3125	76	24,83660131	152	12,16
Net 89	2	10	7	134	52,34375	79,6875	15,625	48,5714286	60,35502959	983	78,64
Net 90	2	10	8	128	50	0	100	N/A	N/A	128	10,24

Наводиться список цих мереж із оцінками, вибір кращої мережі із відповідною моделлю входів нейронів – за користувачем:



Training

Validation

Training: Net 1

Actua	0	1
0	0	128
1	0	128

Validation: Net 1

Actual\P	0	1
0	1	1121
1	0	128

Training: Net 2

Actua	0	1
0	122	6
1	109	19

Validation: Net 2

Actual\P	0	1
0	1085	37
1	115	13

Training: Net 3

Actua	0	1
0	0	128
1	0	128

Validation: Net 3

Actual\P	0	1
0	0	1122
1	0	128

Література

1. Analytic Solver Data Mining User Guide. V2021. – Frontline Systems, Inc., 2021. – 176 p.
2. Додонов О.Г., Кузьмичов А.І. Мережеві організаційні структури управління. Моделювання та візуалізація засобами Excel. – К.: Ліра-К, 2021. – 264 с.
3. Camm J. et al. Business Analytics: Descriptive, Predictive, Prescriptive. – Cengage Learn., 2019. – 818 p.
4. Ragsdale C. Spreadsheet Modeling and Decision Analytics. A Practical Introduction to Business Analytics, 8-ed. – Cengage Learn., 2018. – 869 p.
5. Powell S., Baker K. Business Analytics. The Art of Modeling with Spreadsheets, 5-ed. – Wiley, 2017. – 555 p.
6. Shmueli G. et al. Data Mining for Business Analytics: Concepts, Techniques, and Application with XLMiner, 3-ed. – Wiley, 2016. – 549 p.

APPROACH TO DETERMINATION OF INFORMATION RELIABILITY FOR ANALYTICAL ACTIVITY

A.V. Boichenko, V.R.Senchenko
Institute for Information Recording of NAS of Ukraine, Kyiv, Ukraine, 03113

This work based on the analysis of relevant publications, which presented various aspects of management decision support. By collecting them, we were able to identify the main aspects needed to develop a common model. The proposed approach determines the information characteristics of the tested web sites by weighting each of their news, and then comparing the results with the corpus of the testing categorizes.

Introduction

We propose an information reliability of assurance methodology for decision support systems. The reliable information helps understand the problem and make decisions about the problem, but spreading of misinformation represents is a serious issue nowadays. The evaluation of information reliability is a complex task, from the selection of the relevant dataset to the content analysis, what depends from data characteristics such as accuracy, completeness, validity, consistency and timeliness. Reliability of information has become an increasingly important problem, and automating it presents many technical challenges.

The problem of reliability of information and knowledge is of great importance in any field of analytical activity. Because when dealing with facts or events that have not been verified, there is a very high probability of incorrect conclusions, which can lead to catastrophic consequences in decision-making [17].

The creation of fake information is not new and dates back to early thirteenth century BC where Ramasses II spread lies and propaganda on victory of Egyptian Empire over Hittite Empire in the battle of Kadesh [15].

Traditional decision-making approach provides a satisfactory way for formulating and structuring decision scenarios, but disadvantages is that only the expected value used at a decision node.

Related work

This work based on the analysis of compatible publications, which identifies the different aspects of reliability of information sources. The use of machine learning tools opens wide opportunities for application in analytical activities when working with various sources of information, including OSINT.

Technologies of ontology formation based on automatic extraction of concepts from external sources is often used. The categorical representation of word definitions in the dictionary and relationships between them implies using OWL, what supports more accurate reference for meaning and semantics than other description languages developed for metadata [16].

Fake news detection can be achieved using the International Fast Checking Network (ICFN) and other manual fact checking websites like Washington post, Snopes, Fast Checker, FastCheck and TruthOrFiction [4]. There are some Ukrainian fact checking sites like StopFake, БезБрехні, Детектор медіа, Інститут масової інформації (IMI) etc. [5,6].

There is a public Web service Botometer that able to estimate how likely the account is to be a bot. Some organizations use applications like Decodex that help to recognize trustworthy and false information. [7].

Recent research has pursued development of automated fact-checking systems, but these systems do not explicitly model joint interactions between key variables: claims, sources and annotators [10].

Spatial analysis can also help create new spatial parameters as test variables that will be explored in consolidation models. These variables can represent and estimate the complexity of the intersection. Models must take into account the technological features of spatial organization and data processing in relation to the sequence of actions of the analyst in solving functional problems and decision-making based on their analysis [11].

The classification of whether an article title can be supported by content in the body has been interchangeably referred to as both fake news detection, stance classification and incongruent headline detection [14].

Thorne et al. introduced a fact checking dataset containing 185K claims about properties of entities and concepts, which must be verified using articles from Wikipedia. Combined with the labels on the claims, these machine-readable fact checks allow training machine learning-based fact checking models. [10].

The approaches underlying the methods of detecting bots on social networks can vary greatly. However, they can be generally divided into three common methods: schedule-based, crowdsourcing and machine learning. The process of detecting social bots begins with retrieving data from the Twitter stream. Once the data has been collected, the next step involves preparing this data for the selected classifier by extracting and selecting features that can be studied using statistical methods, such as, or those that can be manually labeled using previous work [11 - 14].

Approach

In this study, analytical activity means a set of processes (collection, search, analysis of information and knowledge) that contribute to the emergence of new knowledge and aimed at justifying decisions in various fields (economic, financial, social, managerial, political and other areas).

Metadata is data about data. The Semantic Network Extension provides not only semantic information about the machine, but also conceptual information for the user.

There are various machine-learning algorithms, which can used to solve complex problems. However, it is very important to choose the right one that fits your data well. Usually, bots are trying to avoid disclosure.

We arrange sources of information used in carrying out analytical activities:

- source status - state, official, publication with a good reputation, media, commercial;
- thematic areas - the target function of the source of information:
- thematic breadth - a variety of information;
- significance of information - important unimportant. Information can be important and accurate and at the same time useless because it is incomplete and insufficient for understanding and adequate action. Such information should not be discarded, but try to compare with other information to determine its significance;
- reliability (unreliability) of information - all information obtained should reflect the real state of the problem, phenomenon or process,
- the level of coverage of the problem - (depth of analysis) in comparison with other sources.
- level of qualification of authors of publications - the use of special languages and terms is typical.
- probability of information distortion - mechanisms and degree of information distortion
- main disadvantages - inherent in this type of information source;
- processing capabilities - can and cannot be processed, the ability to quickly copy large amounts of information.

To verify knowledge, it proposed to automatically weigh the concepts of the text in the form of an ontology and establish semantic relationships between them using the method of establishing the coefficients of importance of the concepts of ontology.

Information importance coefficient is a numerical measure that characterizes the significance of a concept and can vary according to certain rules when working with the ontology.

To scrub the data, manipulate it, and generally make it more intelligible to analysts we used KNIME (Konstanz Information Miner) – a free and open-source data analytics, reporting and integration software [8].

First task is collecting enough and relevant data for the analysis. KNIME workflow including the nodes for all of the steps analysis process for information reliability sources shown in Figure 1.

The KNIME Text Processing extension allows the specification of a dictionary of concepts to identify.

The machine-learning algorithm used with 80% of dataset for training and 20% - for testing.

Several machine-learning algorithms has been consider for applying information reliability of assurance research domain. Platform KNIME supports using of very dierent classification algorithms [8] to cluster the ontology of sources of information on the basis of reliability:

- decision trees;
- gradient boosted trees;
- logistic regression;
- naive bayes model;

- support-vector machine (SVM).

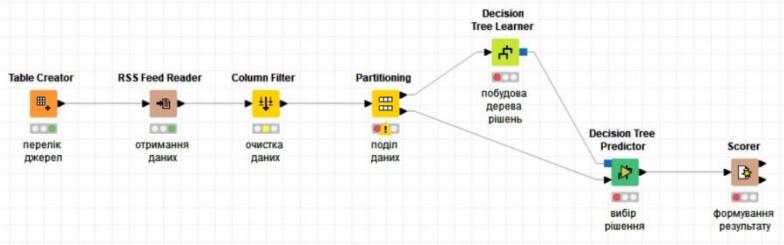


Figure 1. Creating workflow using KNIME Platform for analysis of information reliability sources

The ontology was developed using open world assumption, what helps to handle partial definitions of fuzzy concepts. The ontology model – Ont(IRS) using the mechanism of coefficients of importance of the concepts of the subject area – information reliability sources (IRS) and their relationships can be represented as a mathematical expression [17]:

$$\mathrm{Ont(IRS)} = \left\{ \mathrm{Cl}_k^{\mathrm{IRS}}, \mathrm{Rel}_j, \mathrm{W}\left(\mathrm{Cl}_k^{\mathrm{IRS}}\right), \mathrm{L}_{j,k}, \mathrm{Ax} \;\middle|\; k = \overline{1,N};\ j = \overline{1,M} \right\}$$

where $\{\mathrm{Cl}_k^{\mathrm{IRS}}\}$ – set of concepts – classes/subclasses ($k = \overline{1,N}$) of ontology of information reliability sources;

$\{\mathrm{Rel}_j\}$ – set of semantic relationship between concepts $\{\mathrm{Cl}_k^{\mathrm{IRS}}\}$ of IRS;

$\{\mathrm{W}\left(\mathrm{Cl}_k^{\mathrm{IRS}}\right)\}$ – set of concepts weighting factor of IRS;

$\{\mathrm{L}_{j,k}\}$ – set of concept relationship coefficient of IRS;

{Ax} – set of ontology axioms.

To calculate the Concepts weighting factor of IRS, the following rules are used:

The total weight $\mathrm{W}\left(\mathrm{Cl}_k^{\mathrm{IRS}}\right)_k^i$ of the ontology class is equal to the sum of its own weight – $\mathrm{W}\left(\mathrm{Cl}_k^{\mathrm{OWN}}\right)_k^i$, the weight of the subclasses – $\mathrm{W}\left(\mathrm{Cl}_k^{\mathrm{SUB}}\right)_k^i$ and the weight of the adjacent classes – $\mathrm{W}\left(\mathrm{Cl}_k^{\mathrm{ASS}}\right)_k^i$ (classes associated with this class is not "is - as" relationship):

$$\mathrm{W}\left(\mathrm{Cl}_k^{\mathrm{IRS}}\right)_k^i = \mathrm{W}\left(\mathrm{Cl}_k^{\mathrm{OWN}}\right)_k^i + \mathrm{W}\left(\mathrm{Cl}_k^{\mathrm{SUB}}\right)_k^i + \mathrm{W}\left(\mathrm{Cl}_k^{\mathrm{ASS}}\right)_k^i$$

$$\mathrm{W}\left(\mathrm{Cl}_k^{\mathrm{ASS}}\right)_k^i = \sum_k \mathrm{W}\left(\mathrm{Cl}_k^{\mathrm{ASS}}\right)_k^{i+1} \mathrm{L}_{j,k}$$

Where $\mathrm{L}_{j,k}$ – weight of relationship connection between own classes $\mathrm{Cl}_k^{\mathrm{OWN}^i}$ and associated classes – $\mathrm{Cl}_k^{\mathrm{ASS}^j}$. As the IRS ontology configured using training samples that determine ontology properties, a new edge (between the corresponding vertices of the ontology graph) may appear, and the class weight added of the new adjacent class weight.

The weight of an instance in the knowledge base $Ax\left(Cl_k^{IRS}\right)^i$ is equal to the total weight of its class $W\left(Cl_k^{IRS}\right)_k^i$ in the ontology IRS.

The proposed rules create the basis for automatic recalculation of the weight of ontology classes and instances of the knowledge base during its filling and adjustment to a given software.

The low-level formatting of web-documents involves spelling correction, normalization of punctuation, and removal extra-data (various styles of punctuation are usually used in web documents) from files to randomize the corpus. Excluding each item from a single document and follow randomizing them helps make the corpus feasible for using in models.

Different styles of punctuation marks have been used in the documents or articles.

The input to model is news documents.

Each text file Wd has a sequence of words: $Wd = \{wd_1, wd_2, \ldots wd_n\}$, where n is number of words. For each file Wd, the meta reference exists: $Rm = \{r_1, r_2, \ldots r_m\}$.

The reliable texts contain true description of events, while the untruthful ones contain claims that are not aligned with facts. Next issues used to evaluate of data sources:

- validity;
- credibility marker;
- source reputation;
- language characteristics;
- headline quality;
- parameters for predicting stances.

To check out the productivity a confusion matrix was built. Confusion matrix represents of a classification model performance on the test set.

Using proposed approach, suspected publications are automatically selected while monitoring web sites and social networks, and it is very easy to analyze the original sources and propagation of fake news.

Conclusion and future work

The purpose of this publication has demonstrate the possibilities and benefits of using machine learning Technologies in conjunction with the ontology model, which more fully describes the hidden links between the concepts of the subject area, especially when researching the reliability of OSINT sources.

More other, the use of modern analytical platforms (such as KNIME) significantly reduces the process of modeling analytics scenarios, especially when working with various sources of information OSINT.

We believe that our research to support management decisions through artificial intelligence and ontology can be used in future research and applications in knowledge management.

References

[1] Crawford, M., Khoshgoftaar, T.M., Prusa, J.D., Richter, A.N., & Al Najada, H. (2015). Survey of review spam detection using machine-learning techniques. Journal of Big Data, 2(1), 1–24.

[2] Deepak, G., Ahmed, A., Skanda, B.: An intelligent inventive system for personalized webpage recommendation based on ontology semantics. Int. J. Intell. Syst. Technol. Appl. 18(1–2), 115–132 (2019).

[3] Du, S., Wang, J., Zhang, H., Cui, W., Kang, Z., Yang, T., Yuan, Q.: Predicting COVID-19 using hybrid AI model. Available at (2020) http://dx.doi.org/10.2139/ssrn.3555202

[4] G. Gravanis, A. Vakali, K. Diamantaras, P. Karadais, behind the cues: A benchmarking study for fake news detection, Expert Syst. Appl. 128 (2019) p201–213.

[5] https://detector.media.

[6] https://imi.org.ua.

[7] https://www.stopfake.org.

[8] https://docs.knime.com.

[9] James Thorne, Andreas Vlachos, Christos Christodoulopoulos, and Arpit Mittal. 2018. FEVER: a large-scale dataset for fact extraction and variation. In NAACL-HLT.

[10] Mullainathan, S., & Spiess, J. (2017). Machine learning: An applied econometric approach. The Journal of Economic Perspectives, 31(2), 87–106.

[11] Popat, K.; Mukherjee, S.; Str¨otgen, J.; and Weikum, G. 2017. Where the truth lies: Explaining the credibility of emerging claims on the web and social media. In WWW.

[12] Sandhu, T.H. (2018). Machine learning and natural language processing—A review. International Journal of Advanced Research in Computer Science, 9(2), 582–584.

[13] Santia, Giovanni C., Munif Ishad Mujib, and Jake Ryland Williams. Detecting social bots on facebook in an information veracity context. Proceedings of the International AAAI Conference on Web and Social Media. Vol. 13. 2019.

[14] Sophie Chesney, Maria Liakata, Massimo Poesio, and Matthew Purver. 2017. Incongruent Headlines: Yet Another Way to Mislead Your Readers. In Natural Language Processing meets Journalism workshop at EMNLP 2017, number 1, pages 56–61.

[15] Weir, W.: History's greatest lies. Fair Winds, Beverly, MA (2009)

[16] Бойченко А.В., Бойченко О.А. Розширення можливостей дистанційної освіти засобами штучного інтелекту. Штучний інтелект. 2020, Т. 26, № 2. – с. 22-29.

[17] Додонов А.Г., Сенченко В.Р., Коваль А.В. Аналитика и знания в компьютерных системах. Киев: ИПРИ НАН Украины, «КПИ имени Игоря Сикорского», 2020. 315 с.

СТРУКТУРНИЙ АНАЛІЗ ВЗАЄМОДІЇ РЕФЛЕКСИВНИХ СУБ'ЄКТІВ

Конторчук Н. І., Смирнов С. А.
КПІ ім. Ігоря Сікорського, Київ, Україна
nikontorchuk@gmail.com, sergsmr@gmail.com

На основі моделі Лефевра групової рефлексії та Q-аналізу розроблено і розглянуто модель колективної свідомості для трьох і більше суб'єктів з урахуванням структурних особливостей взаємодії в групі. За допомогою Q-аналізу було визначено складну структуру і виявлено емерджентні властивості системи, а за допомогою рефлексивного підходу вдалося виявити механізми функціонування колективної свідомості та проаналізувати можливі впливи на групу рефлексивних індивідів. Також розглянуто частковий випадок для n суб'єктів.

Вступ

Питання про моделювання свідомості групи суб'єктів не нове та вже достатньо багато досліджувалось з різних сторін. Більшість досліджень на дану тематику базуються на емпіричних даних і описують лише конкретні випадки без фундаментального підґрунтя й якісного узагальнення. Одним із серйозних досліджень, щодо аналізу подібного, було проведено Лефевром у своїх роботах. Він більше часу присвятив опису, що таке рефлексивний суб'єкт і яким він буває в залежності від обставин виховання, що великій мірі залежить від етичної системи мислення суб'єкта. Концепція Лефевра пояснює багато речей, але питання про рефлексивного суб'єкта в групі розкрито не так широко, як потрібно для розв'язання реальних задач, адже він обмежується простими механізмами з однаковим баченням ситуації всіма індивідами, що далеко від реальності. Подальші роботи, які дозволяють описати це іншим способом, наприклад, за допомогою графів, також зустрічаються з певними обмеженнями. Вони ніяк не враховують структурних особливостей взаємодії групи. Наступним логічним крок в даному напрямку, який ліг в основу даної роботи, була ідея зберегти структуру колективної взаємодії, скориставшись Q-аналізом, який дозволяє розглядати більш складну структуру свідомості, а також виявляти різні рівні впливу й аналізувати зв'язність системи. Подальший опис буде спиратись на два ефективних інструменти: модель рефлексивної взаємодії суб'єктів в групі та Q-аналіз. Це дозволить зберегти структуру відношень і виявити неявні зв'язності та механізми впливу одних суб'єктів на інших.

Рефлексивний підхід Лефевра

Розглянемо загальну ситуацію рефлексивної взаємодії трьох суб'єктів. В початкових роботах Лефевра, всі суб'єкти мають однакове бачення ситуації, як розширення його ідеї - це коли кожен суб'єкт, має власне бачення ситуації. Наприклад, розглянемо ситуацію, коли є 3 суб'єкти: a, b, c. На

нульовому рівні це реальна ситуація, тобто реальна взаємодія між суб'єктами в групі. Кожен наступний рівень - це бачення кожного суб'єкта ситуації, тобто образ реальної ситуації в його свідомості. Таким чином, кожен з членів групи має своє бачення і внаслідок цього виникає перший рівень рефлексії. У кожного рефлексивного суб'єкта також існує ще один рефлексивний рівень - це як суб'єкт спостерігає бачення інших суб'єктів у своїй картині світу. Це і є другий рівень рефлексії. Це все можна представити у вигляді 3-дерева (рис. 1)

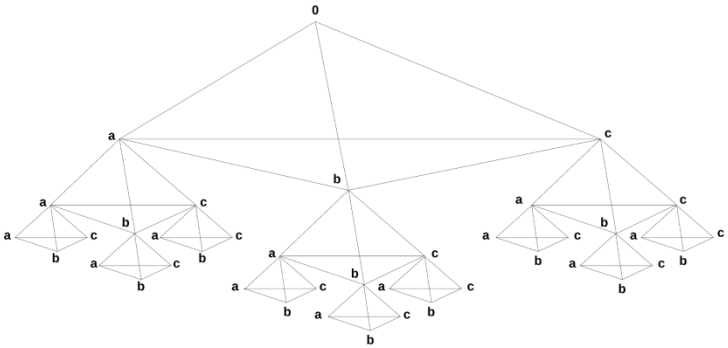


Рис. 1. Взаємодія суб'єктів у вигляді 3-дерева

Для більш реального випадку можна представляти ребра графів у вигляді класичних відрізків конфлікту і союзу і позначати відповідними операціями в залежності від різних етичних систем.

Модифікація моделі Лефевра з використанням Q-аналізу

Розглянувши модель Лефевра і представивши кожен рефлексивний рівень не як сукупність бінарних відношень чи графів, а як симплеціальний комплекс, отримаємо узагальнену модель. Яка має більше практичного значення, адже дозволить зберегти структурні взаємозв'язки і дозволяє розглядати випадки не лише статичну ситуацію, а й за допомогою Q-аналізу вивчати зміни на кожному рівні. Шляхом добавлення або вилучення симплексів, зміною структури взаємозв'язків між суб'єктами в групі.

Використання модифікованої моделі на реальній задачі

Розглянемо тривіальну задачу, група з 7 суб'єктів. Цю ситуацію можна представити у вигляді 3-рівневого дерева, по аналогії з базовою концепцією Лефевра. Для простоти не будем візуалізовувати все дерево, а розглянемо лише один образ одного з рівнів.

Наприклад:

- суб'єкт А1 знайомий з А2
- суб'єкт А2 знайомий з А1, А3, А4
- суб'єкт А3 знайомий з А2, А4

- суб'єкт А4 знайомий з А1, А2, А5, А6, А7
- суб'єкт А5 знайомий з А4, А6, А7
- суб'єкт А6 знайомий з А4, А5, А7
- суб'єкт А7 знайомий з А4, А5, А6

Це можна візуалізувати у вигляді графа (рис. 2) та симплеціального комплексу:

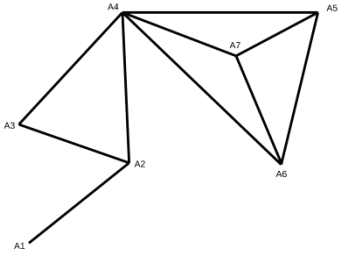


Рис. 2. Представлення взаємодії групи рефлексивних суб'єктів одного рівня у вигляді графу.

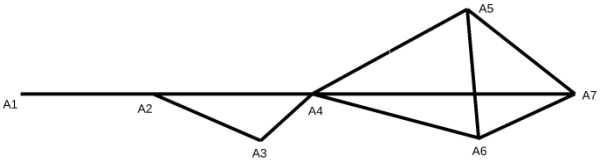


Рис. 3. Представлення взаємодії групи рефлексивних суб'єктів одного рівня у вигляді симплеціального комплексу.

Наступний крок, виділити всі симплекси, побудова матриці зв'язності для всього комплексу, знаходження значень зв'язності, побудова локальних карт та Q-дерева. Пропустимо деякі викладки і розглянемо лише важливі моменти. На рис. 4 бачимо представлення структури заданої групи у вигляді комплекса симплексів.

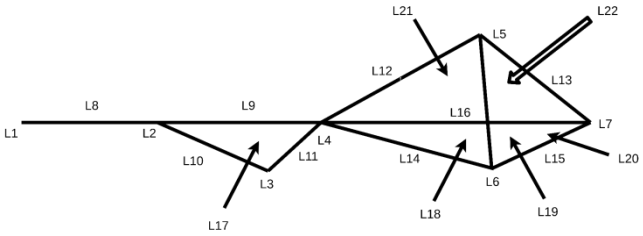


Рис. 4. Симплеціальний комплекс K(A; λ)

Використовуючи Q-аналіз, отримаємо наступні значення зв’язності:

Табл. 1. Q-аналіз для K(A; λ)

q = 3	*Q4 = 1*	{L22}
q = 2	*Q3 = 2*	{L17}, {L18- L22}
q = 1	*Q2 = 2*	{L9-L11, L17}, {L12-L16, L18-L22}
q = 0	*Q0 = 1*	{L1 - L22}

Після проведення аналізу отримаємо структурний вектор зв'язності Q = (1, 2, 2, 1).

За допомогою цієї таблиці легко побудувати локальні карти та Q-дерево і проводити подальший аналіз для отримання відповідей на ряд запитань, що з себе представляє структура групи? Яка зв'язність між компонентами? Які суб'єкти є кореневими в групі? Як зміна складу групи може вплинути на прийняття колективного рішення? І т.д.

Для переходу між рівнями потрібно використовувати аналіз Лефевра, тобто можна, привести отриманий граф до декомпозованих бінарних відношень, а далі класичні обчислення.

В результаті попереднього опису видно, що якщо ми розглядали би менш повну ситуацію, могла б зменшитись n-арность системи і змінилася б структура зв'язності, послабились би впливи окремих компонент на систему. Зазвичай в реальних групах рівень розрідженості залежить від наявності додаткових взаємодій між суб'єктами. Класична модель Лефевра добре працює в середовищі незнайомих людей без спільних інтересів і кореляцій в міркуваннях. Що не властиво людям, бо в будь якій групі, яка приймає рішення, після кількох ітерацій з'являється певна гармонія і передбачуваність(самоорганізованість). Тому ця модель має практичну цінність, бо дозволяє моделювати реальні системи. В ситуації з усіма симплеками ми спостерігали ніщо інше, як колективну свідомість. Це є тим що описує самоорганізованість в рефлексивних групах, чим не можна нехтувати.

Подивимось на табл. 1, і побачимо, що між суб'єктами прямої взаємодії немає. Є лише неявні взаємозв'язки через інші симплекси(де симплекси вище 0-симплексів - є колективними свідомостями). Тобто якісь зміни в групі можуть з'являтись і поширюватись лише на цих рівнях. І окрім саморефлексії на суб'єкт може здійснюватись вплив ще й через зміну колективної свідомості, яка сама по собі є незалежною від кожного суб'єкта по окремості. Вона з'являється непідконтрольною суб'єктами і функціонує сама по собі. В реальному житті це є найпростіший спосіб впливу на суб'єктів і на групу в цілому. Як ми бачимо, зникнення попарної свідомості може призвести до зменшення n-арності, що призводить до індивідуалізму і сприяє погіршенню рівня самоорганізованості групи. І навпаки, поява нових

взаємозв'язків може змінити якісно організацію групи і призвести до появи нових властивостей такої системи.

Висновки

На основі моделі взаємодії рефлексивних суб'єктів Лефевра і □-аналізу отримано та описано розширену модель колективної свідомості з урахуванням структурних особливостей групи суб'єктів. За допомогою □-аналізу формалізовані компоненти свідомості, які забезпечують неявний вплив на колективну поведінку. Вони відповідають свідомості окремих підгруп, взаємодія яких визначає структуру свідомості групи в цілому, яка є емерджентною властивістю групи та здійснює значний вплив на її поведінку. Хоча окремим суб'єктом групи вона може повною мірою не усвідомлюватись і розглядатись як щось стороннє, проявляючись для нього лише на поведінковому рівні, що зазвичай складно виявити й описати. За допомогою структурного дерева і локальних карт відповідного комплексу видно, що взаємодія між компонентами здійснюється через симплекси, які відповідають колективній свідомості підгруп, а не шляхом прямого впливу одного суб'єкта на іншого. Це знов підтверджує, що позиція суб'єкта у групі залежить не тільки від його власного сприйняття, але і від того впливу, який суб'єкт зазнає через колективну свідомість підгруп, до яких він відноситься. І рефлексивний вплив на нього може бути здійснений скритим, непрямим чином, через вплив на такі підгрупи. Це дозволяє краще зрозуміти, що таке колективна та індивідуальна свідомість, як вони взаємодіють і де шукати джерела небажаного впливу та захист від нього.

Перелік використаних джерел

1. В. А. Лефевр. Алгебра совести. — 2003.
2. В. А. Лефевр. Лекции по теории рефлексивных игр. — 2009.
3. H. Atkin R. Combinatorial Connectivies in Social Systems. — 1997.
4. В. І. Полуциганова, С. А. Смирнов. Методологія побудови основних метрик Q-аналізу та їх застосування // Системні дослідження та інформаційні технології. - 2019. - № 3.

THE INVERSE PROBLEM OF Q-ANALYSIS OF COMPLEX SYSTEMS STRUCTURE

Polutsyganova V. I., Smirnov S. A.
National Technical University of Ukraine
"Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine
medvika@gmail.com, sergsmr@gmail.com

It is considered the inverse problem of Q-analysis. In the course of the research, an algorithm for the recovery of simplicial complexes from elementary simplex using local maps and a structural tree was developed. This algorithm will reduce the amount of data stored and improve the management process if the simplicial complex describes a real big complex system.

Introduction

Q-analysis was first described by Atkin [1]. The approach allows to analyze the structure of a complex system, taking into account the variety of connections between its components. In order to use Q-analysis, it is necessary to formalize the description of the system in terms of the simplicial complexes.

The simplicial complex is called the finite set of simplexes, which satisfies the following conditions [5]:

1. Together with any simplex, its facets of all dimensions belong to this set;

2. Two simplexes can intersect (have common points) only along the entire facet of any dimension and thus only on one facet.

The simplicial complex is a multidimensional topological structure, so it better describes the relationship between parts of the system and its elements than graphs. Also the concept of q-connectivity is used in this context, that is, the level of connectivity between simplexes in a complex. At each level of such connectivity, a simplex complex can be described as a set of chains, that is, a graph whose nodes are simplexes and edges are connections which dimension is not less than a given level of connectivity (in current level). Such graphs are also called local maps [3] because they reflect the internal structure of the chains.

Definition 1. The local map of a simplicial complex is called the graph of the binary relation q-adjacent whose nodes are the simplexes of dimension k> q, and the edges correspond to the q-relation between them.

Definition 2. Q-adjacency is called the binary relation between simplexes in a simplicial complex, which occurs at the cross section of two simplexes and whose dimension is greater than or equal to "q" [7].

Definition 3. A Q-link is a binary relation that occurs when a q-junction is transitively closed in a simplicial complex [7].

Basing on the imitation of circuits of adjacent levels of connectivity, it is possible to build a corresponding tree [2]. The nodes of a Q-tree are chains of

simplexes that are connected at a certain level of connectivity, and the depth levels of the trees are the same q-levels of connectivity. That is, a structural analysis of a complex system (a direct problem) results in a structural tree and local maps of the corresponding complex. The Atkin structural vector and the eccentricities of the simplexes are only a small part of the complete information about the structure of the system set up in this way. This method allows to quantify the set of connections between simplexes that form chains in a complex, which is important for the study of the structure of the system as a whole [4].

The question naturally arises about the inverse task. Suppose, that a system has a large number of elements (not simplexes), they make up certain subsystems that can be considered as simplexes or circuits of simplexes. Then the simplicial complex represents the structure of connections in the system, that knowledge gives the opportunity to visit the whole system. Working with large systems, we would like to reduce the amount of information on their structure for efficient storage and processing. Merging into a simplex of a vertex can be considered as a single vertex on some scale of representation of a simplicial complex and it creates opportunities for more efficient (scaled) processing of system related information.

In addition, management systems may encounter tasks related to individual subsystems, when it is advisable to work only with certain simplexes, without affecting the complex as a whole, that is, maintaining their external connections. Therefore, you need not only to decompose the system (which allows you to do Q-analysis), but also to synthesize it to its original state as needed.

Problem discussion

In the studied sources, the analysis is considered for the application and use of the results of its work [8-10]. This provides some (incomplete) information about all nodes, simplexes, chains in the complex. As stated earlier, this is a very large amount of data. Sometimes it is enough to know only the set of simplexes and the levels of connectivity between them. This study proposes a new approach to the analysis, synthesis, and control of structurally complex systems based on the use of their structural tree and local maps. We will work with subsystems or simplexes and circuits, store information about connections, that is, trees and local maps, and, if necessary, restore the overall structure of the system. So far, the formulation of tasks in this form is unknown to the authors. It is possible only with the exception of the general meta-rule of “act locally, think globally”. The main purpose of this study is to prove the possibility of correct restoration of a structurally complex system (simplicial complex) by information about its structural tree and local maps.

In addition, this methodology can be used to build complex systems with a defined structure of connections, but unknown nodes. Such problems are typical in linear systems, when the analysis of the system itself is not difficult, and the construction of a system with given characteristics is not a trivial task.

In this article, we will analyze and synthesize not a large system, but with non-trivial binary relationships. This will allow you to use this methodology not

only for compact data storage, which will reduce data load, but also allow you to synthesize new systems only on known attributes.

Methodology

To solve this problem, an appropriate synthesis method has been developed, that is, a certain algorithm for the recovery of the simplicial complex has been constructed. The basic idea behind a recovery is that local maps give us the topology of connections at each level of connectivity, and the structural tree defines the rules of inheritance between chains when the level of connectivity changes. More details about the algorithms for constructing local maps and structural tree can be found in the article [8]. Therefore, it is well known how and what simplexes are related. But there is a problem: how or which nodes / edges / facets are connected by simplexes. To do this, you need to enter a certain equivalence ratio [5]. Therefore, it was hypothesized that there is a certain isomorphism that, if necessary, will return the simplex in the way it was built into the complex.

We define such an isomorphism as a reflection of the set (denote such a set C) of nodes of the simplex itself $f: C \to C$. The essence of such a mapping is that it renames the nodes in the simplicial complex obtained after restoration according to the names that were in the original complex.

Hypothesis. For any two simplicial complexes in which the structure (number of nodes, simplexes, local maps, and structural tree) coincides, there is an isomorphism $f: C \to C$ that rearranges the nodes in the simplex so that the complexes become identical.

At this stage of the study there is no doubt about the existence of such an isomorphism, so we will use this assumption to describe the algorithm of complex recovery.

Therefore, to reconstruct a complex, it is necessary to have a structural vector in the concept of Q-analysis, that is, the number of symmetry chains at each q-connectivity level, local maps or incidence matrix between the simplexes at these levels, and the structural tree. In general, the recovery algorithm looks like this:

Incoming data:

- Structural tree;
- Local maps;
- Sets of simplexes according to local maps;

The output of the algorithm:

- Simplicial complex of complex system.

The definition of isomorphism will be specified in the course of the inverse algorithm. In the first step, a subset of nodes is formed by simplexes of maximum dimension. In the next step, the simplexes are glued in accordance with the scheme predetermined by local map, while the sub-complexes (facets) of the different simplexes are identified - each such sub-complexes corresponds to the real subsystem, which correspond as a common part of the higher-level subsystems. That is, the nodes that form it are identified with precision by permutation according to the subsystem connection information. Subsequent

bonding may clarify the correspondence between the nodes of the complex (points) and the real elementary (unstructured) subsystems. Some nodes uniquely correlate with such subsystems, and some up to permutations that can characterize the internal stuctural symmetry of the system.

This algorithm makes it possible to recover a simplicial complex from a structural tree and local maps, but an important step is a well-defined isomorphism. Conflict may arise if several simplexes are connected through the same simplex but smaller dimension. But the local map at the appropriate level is always shown which simplexes are interconnected, and the isomorphism "returns" the simplexes so that the input complex of Q-analysis and the output of the recovery algorithm are identical.

When using the recovery algorithm, it is needs to remember that the simplicial complex presents some real system. In the process of joining a simplex to a complex, we considered that all nodes / edges / facets are equivalent, that is, by choosing a facet or node to attach another simplex, one can choose any existing one (except when one node / edges / facets with several simplexes).

Therefore, after recovery, we use an isomorphism that will determine each vertex according to the connections that were in the real system. If it is not defined before the algorithm starts, then expert evaluation can be used to accurately determine the transformation. That is, specialists can provide information on how subsystems (in our terminology - simplexes) fit into the system (complex). Therefore, using isomorphism, the nodes will be determined according to the original simplicial complex.

In the report presents a real example of using the algorithm for the recovery of simplicial complex structure, but it size is too large for presentation here.

Conclusion

Thus, the developed recovery algorithm makes it possible to reduce the amount of data on the structure of the system and improve management capabilities. The amount of data is reduced due to the fact that instead of storing all nodes and relations of the simplicial complex, it is sufficient to know only the simplexes and their dimensions, and, in addition, it is necessary to have a structural graph and local maps at each level of q-connectedness in order to be able to fully recover a structure of the system. The most important part of the synthesis of a complex is the determination of an isomorphism that will ensure that the complex being analyzed corresponds to the complex formed at the output of the recovery algorithm. Although this part is formed, it may not be sufficiently formalized, but there is no doubt about its feasibility (at least through busting).

Therefore, the inverse problem of Q-analysis can have both theoretical and practical value for structuring and managing complex systems.

References

[1] Atkin R. H. (1973) “Mathematical structure in human affairs”, *Heinemann Educational Books*, 143.

[2] Pierre Mazzega, Claire Lajaunie and Etienne Fieux (2018) “Governance Modeling: Dimensionality and Conjugacy”, *Graph Theory - Advanced Algorithms and Applications.*

[3] Beaumont J.R., Gatrell A.C. (1982) “An introduction to Q-analysis” *Catmog 34*.

[4] Duckstein, L. and Nobe, S. A. (1997) “Q-analysis for modeling and decision making”, *European Journal of Operational Research*, 103/3, 411–425.

[5] Dobrovin B. A., Novikov S. P., Fomenko A. T. (1984) “Modern geometry. Method of the theory of homology ”, *The main edition of the physics and mathematics literature*, 344.

[6] Doktorova M. (2012) “Constructing simplicial complexes over topological spaces”, *A thesis submitted dartmouth college*.

[7] Jeffrey H. J. (1981) “Some structures and notation of Q-analysis”, *Environment and Planning B Planning and Design*.

[8] Macgill S. M. (1986) “Q-analysis and transport: a false start?” *Pergamon Journals Lid*, 4, 271–281.

[9] Kachinsky A. B., Agarkova N. V. (2015) “The nature of the interconnectedness of the elements of the safety system for hydraulic structures”, *System technology and information technology*, 3, 72–83.

[10] Johnson J. H. (1978) “A q-analysis of television programmes”, *International Journal of Man-Machine Studies*, 10, 461–479.

НАУКОМЕТРИЧНИЙ АНАЛІЗ ТЕМИ «ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ ТА БЕЗПЕКА»

Балагура І.В.
Інститут проблем реєстрації інформації НАН України, м.Київ
balaguraira@gmail.com

Мета роботи – наукометричний аналіз теми «Інформаційні технології та безпека» на основі матеріалів XIX, XX Міжнародної науково-практичної конференції «Інформаційні технології та безпека» для визначення основних напрямів досліджень в публікаціях з допомогою побудови хмари термінів. Джерелом даних було обрано публікації конференції в CEUR workshop proceeding. Визначено основні терміни конференції, побудовано відповідні хмари. Наукометричний аналіз теми може застосовуватись для визначення пріоритетних напрямів розвитку галузі, обрання авторами конференцій, визначення основних напрямів розвитку журналу чи конференції.

Вступ

«Наукова конференція» – інтерактивна медіаосвітня технологія, мета якої – підвищення професійної компетентності фахівців, обмін досягненнями, окреслення шляхів подальшого розвитку проблеми, що обговорювалась [1]. Наукові конференції є осередком багатосторонньої комункації та можливістю заглиблення в обрану тематику.

Конференція «Інформаційні технології та безпека» щорічно проходить і організовується Інститутом проблем реєстрації інформації Національної академії наук України [2]. Також у різні роки організаторами конференції виступали Інститут системних досліджень й інформаційних технологій УАН, Науково-дослідний інститут інформатики і права НАПрН України, Навчально-науковий центр інформаційного права та правових питань інформаційних технологій НТУУ "КПІ ім. Ігоря Сікорського". Інформаційну підтримку надавали науково-технічних журналів "Реєстрація, зберігання і обробка даних", "Інформаційні технології та спеціальна безпека", Видавничий дім "СофтПрес" [2]. Обрані тези доповідей публікуються [3]. Тематика конференції «Інформаційні технології та безпека»: безпека та живучість критичних інфраструктур; комп'ютерне моделювання складних систем;технології аналітики великих обсягів даних (Big Data); аналітичні системи на основі відкритих джерел інформації (OSINT);моделювання, аналіз та прогнозування процесів мережевої взаємодії; методи та засоби підтримки прийняття рішень [2]. Обрані матеріали публікуються на сервісі відкритих публікацій CEUR Workshop Proceedings (CEUR-WS.org).

Мета роботи – наукометричний аналіз теми «Інформаційні технології та безпека» на основі матеріалів XIX, XX Міжнародної науково-практичної

конференції «Інформаційні технології та безпека» для визначення основних напрямів досліджень в публікаціях з допомогою побудови хмари термінів.

Основні результати досліджень

Джерелом даних було обрано публікації конференції в CEUR workshop proceeding 2019, 2020 років [4,5]. Всього за 2 роки було опубліковано 44 наукові праці авторами Інституту проблем реєстрації інформації НАН України, Національного технічного університету України «КПІ ім. І. Сікорського», Інституту проблем моделювання в енергетиці ім. Пухова НАН України, Київського національного університету ім. Тараса Шевченка, Національного авіаційного університету, Інституту комп'ютерних систем НАН України, Міжнародного науково-навчального центру інформаційних технологій та систем НАН України та МОН України, Державного університету телекомунікацій, Харківського національного університету радіоелектроніки, Джедзянського університету технологій (Китай).

Для наукометричного аналізу використано частотний аналіз термінів та побудовано хмару термінів (рис.1).

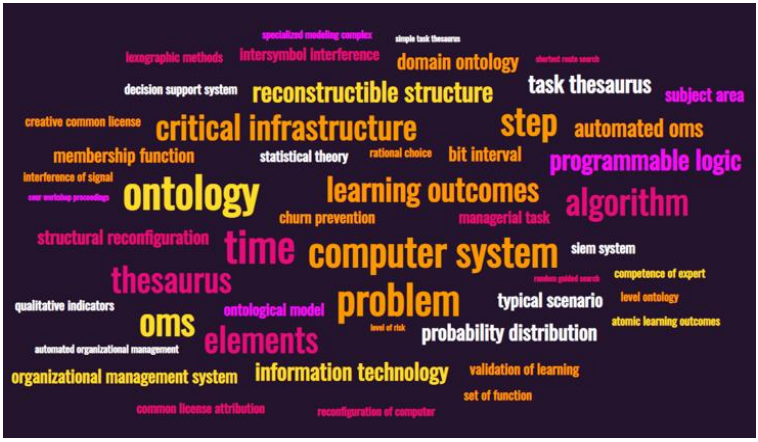


Рисунок 1 – хмара термінів на основі матеріалів доповідей XIX, XX міжнародних наукових конференцій «Інформаційні технології та безпека»

Хмара термінів або хмара тегів — це візуальне представлення ключових слів у тексті, що відображає слова та фрази відповідно до частоти їх вживання у тексті. Для досліджень використано програмне забезпечення для автоматизованого процесу побудови хмар термінів на основі алгоритмів штучного інтелекту MonkeyLearn WordCloud [6]. В основі алгоритму виділення термінів лежить алгоритм TFIDF. Term Frequency-Invert Document

Frequency, «дискримінантна сила» – статистична міра, що використовується для оцінки важливості слів у контексті документа, що є частиною колекції документів або корпуса) [10]. За даним методом значну вагу отримують слова з великою частотою у межах конкретного документу і низькою частотою вживання в інших документах. TFIDF дорівнює добутку частоти слова у фрагменті тексту та двійкового логарифму величини, оберненої до кількості фрагментів тексту, в яких це слово присутнє [7].

Серед найбільш вживаних термінів у матеріалах доповідей міжнародної наукової конференції «Інформаційні технології та безпека» 2019, 2020 років: computer system, ontology, critical infrastructure, reconstructible structure, oms, programmable logic, management, infrastructure, functioning, security, functions, survivability, tasks, functionality, implementation. Терміни відповідають темам конференції: кібернетична безпека критичних інфраструктур, моделювання та протидія інформаційним операціям, проблеми технологічного та правового забезпечення інформаційної та кібернетичної безпеки, можливості онтологічного підходу, семантичних мереж, сценарного аналізу при забезпеченні інформаційної підтримки прийняття рішень, комп'ютерному моделюванні процесів і систем.

Наукометричний аналіз тематики конференції дає можливість організаторам наукових конференцій порівняти основні терміни із пріоритетними напрямами у галузі. Наприклад, компанія Kaspersky визначила наступні тренди у кібербезпеці: кібербезпека для забезпечення віддаленої роботи, розвиток інтернету речей, зростання кількості програм-вимагачів, загрози хмарної безпеки, атаки в соціальних мережах, конфіденційність даних, покращення багатофакторної аутентифікації, розвиток штучного інтелекту, мобільна кібербезпека [8].

Висновки

Проведено наукометричний аналіз матеріалів міжнародної наукової конференції «Інформаційні технології та безпека». Визначено основні терміни і побудовано відповідну хмару термінів.

Серед найбільш вживаних термінів у матеріалах доповідей міжнародної наукової конференції «Інформаційні технології та безпека» 2019, 2020 років: computer system, ontology, critical infrastructure, reconstructible structure, oms, programmable logic, management, infrastructure, functioning, security,f unctions, survivability, tasks, functionality, implementation.

Наукометричний аналіз теми може застосовуватись для визначення пріоритетних напрямів розвитку галузі, обрання авторами конференцій, визначення основних напрямів розвитку журналу чи конференції, порівняння тематики наукових конференцій із світовими тенденціями.

Перелік посилань

1. Онкович Г. Наукова конференція як інтерактивна медіаосвітня технологі. Вища освіта України. – 2014. - №4 - С.85-93

2. Міжнародна науково-практична конференція "Інформаційні технології та безпека" (ІТБ) (електронний ресурс): http://its.ipri.kiev.ua/
3. CEUR Workshop Proceedings (CEUR-WS.org)(електронний ресурс): http://ceur-ws.org/
4. Selected Papers of the XIX International Scientific and Practical Conference "Information Technologies and Security" (ITS 2019), Kyiv, Ukraine, November 28, 2019 (електронний ресурс): http://its.ipri.kiev.ua/

СПЕЦІАЛІЗОВАНИЙ АЛГОРИТМ ФОРМУВАННЯ ІНФОРМАЦІЙНОГО ПРОСТОРУ

В.Є. Мухін, В.В. Завгородній, Я.І. Корнага, Л.В. Барановська
Національний технічний університет України "Київський політехнічний інститут ім. Ігоря Сікорського"

Створення інформаційного простору покликане забезпечити доступ до загальної інформації без обмеження місця та часу. Основою інформаційного простору є сукупність комп'ютерних систем, локальних мереж, відкритих мереж (Інтернет), програмного забезпечення (операційна система, прикладні програми, бази даних, поштові служби тощо). Також при створенні інформаційного простору формуються засоби поєднання різних інформаційних систем одна з одною [1-3].

З урахуванням вищесказаного наведемо алгоритм формування інформаційного простору на основі інформаційної системи (рис. 1):

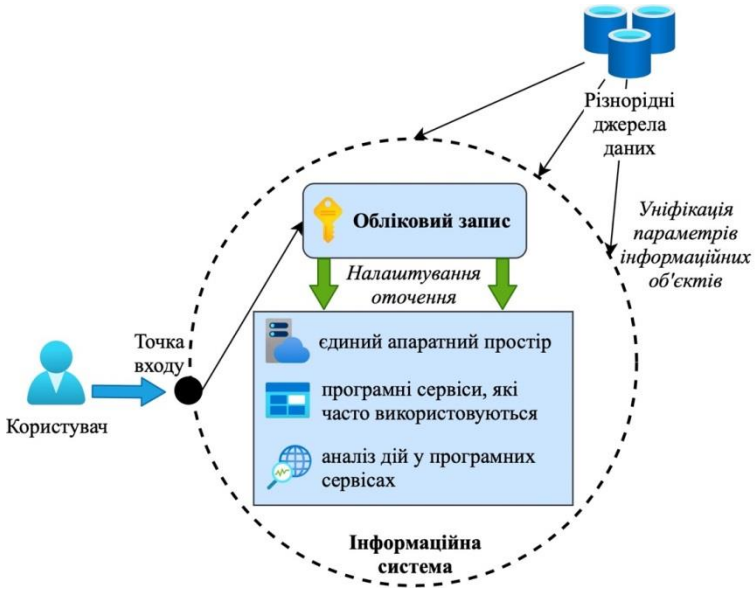


Рис. 1. Алгоритм формування інформаційного простору

1. Інформаційна система є розподіленою системою для виконання перетворення інформації, що надходить від різних джерел і, як правило, в різних форматах, в єдиний набір параметрів інформаційних об'єктів, за яким

користувачі інформаційного простору зможуть однозначно ідентифікувати вхідний об'єкт.

2. Виділяється точка входу користувача в інформаційну систему. Таких точок входу може бути безліч.

3. Для користувача формується єдиний апаратний простір незалежно від точки входу. Тобто для обслуговування інформаційного запиту будь-якого користувача, використовуються однакові апаратні засоби.

4. Набір програмних сервісів доступний для користувача з будь-якої точки входу.

5. Кожен користувач має обліковий запис, в якому зафіксовані програмні сервіси, які він часто використовував, а також його дії, в тому числі відносно конкретних сервісів.

6. Облікові записи користувачів зберігаються віддалено і доступ до них можливий з різних точок входу.

7. При підключенні користувача через точку входу виконується налаштування його оточення, тобто вибираються сервіси, які він часто використовував, а також аналізуються його дії відносно конкретних сервісів.

8. В результаті інформаційна система подається користувачу у вигляді єдиного простору, незалежно від поточної точки входу.

Основним завданням інформаційного простору є однозначне зіставлення кожного вхідного об'єкту, з інформаційним об'єктом, що знаходиться в ньому.

Ідентифікація вхідного об'єкта в інформаційному просторі дозволяє однозначно ідентифікувати його за відповідними ознаками. Для цього може використовуватися метод ідентифікації, який заснований на покроковому аналізі характеристик вхідного об'єкта.

Ідентифікація інформаційного об'єкта проводиться за певними зовнішніми або внутрішніми його ознаками з урахуванням взаємодії інформаційного об'єкта в інформаційному просторі. Для цього кожен інформаційний об'єкт забезпечується набором параметрів які, в певній мірі, характеризують об'єкт, тобто образом інформаційного об'єкта. Аналогічно, у вигляді короткої характеристики формується інформаційний запит до вхідного об'єкта.

Завдяки цьому процедура ідентифікації вхідного об'єкта зводиться до простого порівняння його ознак із заданим образом запита. Якщо параметри вхідного об'єкта в необхідній та достатній мірі збігаються із образом запита, то вважається, що вхідний об'єкт успішно ідентифіковано.

Розглянемо множину інформаційних просторів (*Information Space*), таких як IS_1, IS_2, ..., IS_m, де m – це кількість інформаційних просторів. На рисунку 2 представлена загальна схема існування множини інформаційних просторів.

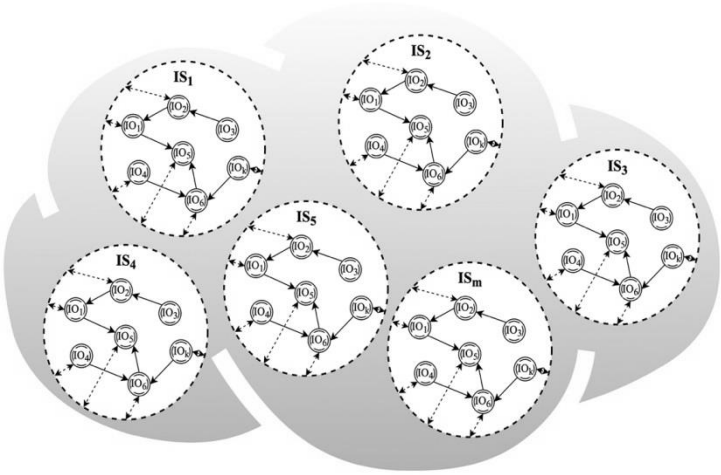


Рис. 2. Загальна схема існування множини інформаційних просторів

Кожен такий інформаційний простір має свій набір інформаційних об'єктів (*Information Object*): $IO_1, IO_2, ..., IO_k$, кожен з яких має набір параметрів $P_1, P_2, ... P_n$, де k – кількість інформаційних об'єктів, а n – кількість параметрів інформаційного об'єкта. Тобто, $IO_k (P_1, P_2, ..., P_n)$. Кожен інформаційний об'єкт має однакову кількість параметрів. Якщо деякі параметри відсутні, то їх значення дорівнює *NONE*.

Параметри інформаційних об'єктів повинні бути в єдиному форматі та їх кількість повинна бути однаковою. В інформаційних просторах можуть бути однакові інформаційні об'єкти, наприклад $IS_1(IO_1) = IS_m(IO_k)$.

Зовнішні джерела інформації для інформаційного простору по-різному представляють вхідні об'єкти $O_1, O_2, ..., O_i$, де i – це кількість вхідних об'єктів. Інформація про ці об'єкти у вигляді набору значень ознак, отримується шляхом зчитування їх за допомогою сенсорів.

В інформаційному просторі не існує двох абсолютно однакових інформаційних об'єктів: $IO_1 \neq IO_2 \neq IO_3 \neq ... \neq IO_k$. У зв'язку з цим, інформаційний простір може бути представлений як довідкова система, яка буде постійно оновлюватися та наповнюватися. Для цього потрібно виконати наступні етапи:

1. Сформувати інформаційний простір з набором унікальних інформаційних об'єктів, тобто розпізнати вхідні об'єкти без прив'язки до конкретного інформаційного об'єкта.

2. Зчитати характеристики вхідного об'єкта у вигляді ознак за допомогою сенсорів, які є свого роду вимірювачами. Якщо з якої-небудь

причини сенсору не вдалося цього зробити, то інформація про цей параметр буде відсутня, тобто відповідає значенню *NONE*.

3. Система отримує вхідний об'єкт O_i з деяким набором характеристик. Сенсори повинні зчитувати їх значення, а інформаційний простір повинен однозначно визначити, чи є в ньому інформаційний об'єкт з точно такими ж значеннями характеристик, тобто провести ідентифікацію вхідного об'єкта.

На рисунку 3 представлена загальна схема формування інформаційного простору [4].

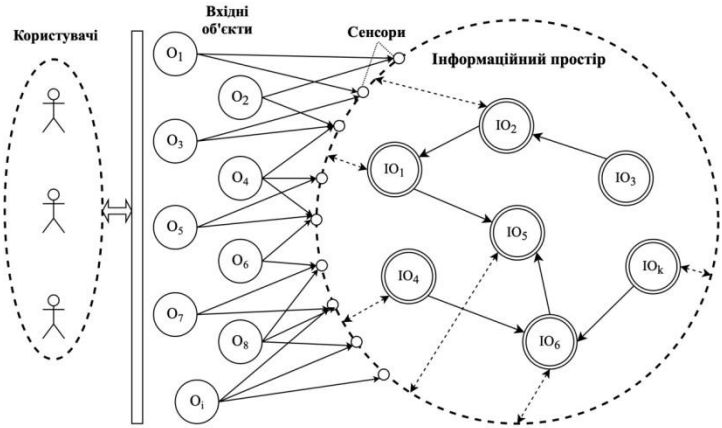


Рис. 3. Загальна схема інформаційного простору

Збір даних в інформаційному просторі та його формування представляється у вигляді графа, який відображає цілісну модель інформаційного простору, в якому вузли графа відповідають відповідним інформаційним об'єктам, а дуги – відношенням між цими об'єктами.

Формально таку модель можна описати у вигляді:

$$IS = < IO, P, R, C >, \quad (1)$$

де IO – множина інформаційних об'єктів;

P – множина параметрів інформаційних об'єктів;

R – множина типів зв'язків;

C – відображення, що задає конкретне відношення з множини типів зв'язків R між інформаційними об'єктами.

Множину типів зв'язків R знаходимо як об'єднання множини зв'язків між інформаційними об'єктами та множини зв'язків між параметрами інформаційних об'єктів:

$$R = R_1 \bigcup R_2, \quad (2)$$

де R_1 – множина зв'язків між інформаційними об'єктами;

R_2 – множина зв'язків між параметрами інформаційних об'єктів.

В свою чергу єдиний інформаційний простір представляє собою об'єднання інформаційних просторів між собою:

$$UIS = \bigcup_{i=1}^{m} IS_i , \tag{3}$$

де m – кількість інформаційних просторів.

При чому, кожен єдиний інформаційний простір можна в свою чергу представити у вигляді графа:

$$UIS = < IS(IO, P, R, C), \tilde{R}, \tilde{C} > , \tag{4}$$

де IS – множина інформаційних просторів;

$\tilde{R}$ – множина типів зв'язків, яка знаходиться аналогічно виразу (2) як об'єднання множини зв'язків між інформаційними об'єктами інформаційних просторів $\tilde{R}_1$ та множини зв'язків між параметрами інформаційних об'єктів інформаційних просторів $\tilde{R}_2$, тобто

$$\tilde{R} = \tilde{R}_1 \bigcup \tilde{R}_2 . \tag{5}$$

$\tilde{C}$ – відображення, що задає конкретне відношення з множини типів $\tilde{R}$ між множиною інформаційних просторів IS.

Для кожного вхідного об'єкта O_i $(i = \overline{1, m})$, який подається на вхід інформаційного простору та інформаційного об'єкта IO_j $(j = \overline{1, k})$, який вже існує в інформаційному просторі визначається множина параметрів $P_{i,j}$, при чому:

– якщо $P_{i,j} = P_i \bigcap P_j$ не співпадає повністю з набором параметрів O_i та IO_j, то такий вхідний об'єкт O_i є новим та додається до інформаційного простору IS;

– якщо $P_{i,j} = P_i \bigcap P_j$ співпадає з набором параметрів O_i та IO_j, то такий вхідний об'єкт O_i вважається ідентифікованим та не додається до інформаційного простору IS .

При чому додаються зв'язки C_i між новим інформаційним об'єктом та вже існуючими інформаційними об'єктами інформаційного простору IS .

Важливим завданням інформаційного простору є перетворення вхідної інформації таким чином, щоб кожен інформаційний об'єкт в інформаційному просторі представлявся однозначно.

В інформаційному просторі параметри інформаційних об'єктів повинні бути визначені в єдиному форматі та їх кількість повинна бути однаковою.

Користувачами інформаційного простору інформаційні об'єкти також повинні сприйматися однозначно. Фактично вхідна інформація в інформаційному просторі є гетерогенною за її поданням. Завдання внутрішніх механізмів інформаційної системи полягає в тому, щоб виконати перетворення різнорідної інформації, що надходить в різних форматах і від різних джерел в єдиний набір параметрів інформаційних об'єктів, за яким користувачі інформаційного простору зможуть однозначно ідентифікувати вхідний об'єкт.

Висновки. Розроблено алгоритм формування інформаційного простору з використанням інформаційної системи, яка фактично є програмним базисом для підтримки інформаційного простору. Ідентифікація вхідного об'єкта в інформаційному просторі дозволяє однозначно ідентифікувати його за відповідними ознаками. Для цього використовується метод ідентифікації заснований на покроковому аналізі характеристик об'єкта.

Література

1. Ожерельева, Т. А. (2014). Об отношении понятий информационное пространство, информационное поле, информационная среда и семантическое окружение. Международный журнал прикладных и фундаментальных исследований, (10-2), 21-24. Режим доступа: https://applied-research.ru/ru/article/view?id=5989
2. Карин, С. А. (2012). Интеграция в едином информационном пространстве разнородных геопространственных данных. Информационно-управляющие системы, (2 (57)). Режим доступа: http://www.i-us.ru/index.php/ius/article/view/13797
3. Герасименко А.Ю. Формирование единого информационного пространства научной библиотеки. Библиосфера. 2019. № 4. С. 78-84.
4. Dodonov A., Mukhin V., Zavgorodnii V., Kornaga Ya., Zavgorodnya A. Method of searching for information objects in unified information space. System research and information technologies, 2021. Vol. 1. P. 34-46. doi: https://doi.org/10.20535/SRIT.2308-8893.2021.1.03

ВИЗНАЧЕННЯ ІНТЕГРАЛЬНОЇ ЯКОСТІ НАУКОВОЇ ПОДІЇ У КОНТЕКСТІ ІНТЕРЕСІВ ОРГАНІЗАЦІЇ

Григорій Гнатієнко [0000-0002-0465-5018],
Віталій Снитюк Євгенович[0000-0002-9954-8767],
Наталія Тмєнова[0000-0003-1088-9547]
Київський національний університет імені Тараса Шевченка, вул. Володимирська, 64/13, 01601 Київ, Україна
G.Gna5@ukr.net, snytyuk@gmail.com, tmyenovox@gmail.com

Розглянуто нові підходи до оцінювання якості організації та проведення наукової події. Запропоновано математичну модель кількісного визначення індексу результативності наукової події у контексті інтересів організації. Досліджено підходи до агрегування атрибутів наукових заходів та види згорток часткових критеріїв. Введено евристики для доозначення математичної моделі інформацією, якої не вистачає при наповненні бази даних наукових заходів. Запропоновано розділити проблему визначення результативності наукової події на три етапи. Розглянуто атрибути кожного з трьох етапів проведення та оцінювання результативності наукового заходу. Наведено перспективи подальших досліджень проблеми оцінювання результативності наукових заходів, які проводяться конкретною організацією.

Вступ

Визначення наукового рівня наукової події у сучасних масштабних потоках інформації є важливою науковою проблемою. Основним чинником при наукометричному підході до обчислення рівня наукової події мають бути її дієвість та результативність. Проблема визначення якості та результативності наукових досліджень є слабкоструктурованою, тому при її вирішенні необхідно на усіх етапах враховувати суб'єктивні фактори. При цьому слід вибрати критерії, які відображують мету і всебічно характеризують якість проведення наукової події.

Висока якість проведення наукових заходів, які забезпечують високий рівень достовірності наукової інформації є необхідною умовою сталого розвитку науки та технологій, адже наука спирається на принципи презумпції доведеного [1].

Зростання загальної кількості наукових заходів, збільшення обсягу робіт, поданих на рецензування і представлених на окремих наукових заходах, поширення наукового контенту наукових подій як результат розвитку технологій, еволюція університетів та наукових досліджень призвели до того, що зріс ризик поширення неякісної наукової інформації [2]. Наявність публікацій низької якості у збірниках матеріалів наукових подій не лише

знижує рейтинг та загрожує репутації окремих авторів, самого наукового заходу, видавців, будь-якого репозиторію, в якому вони розміщені, а й послаблює прагнення всіх учасників цього процесу до забезпечення високої якості матеріалів наукових подій [2].

Класифікація наукових заходів

На сьогодні існує велике різноманіття наукових заходів: конгрес, з'їзд, симпозіум, форум, конференція, круглий стіл, школа, семінар, виставка тощо. Усі ці заходи після їх реалізації будемо позначати терміном наукова подія (НП). НП залежно від напрямку прийнято диференціювати на науково-теоретичні, науково-технічні, науково-практичні. Крім того, залежно від охопленої території, НП може бути міжнародною, всеукраїнською, міжрегіональною, регіональною тощо.

НП – це форма організації наукової діяльності, при якій дослідники представляють і обговорюють свої роботи. Чітких правил проведення НП не існує – їх зазвичай визначає оргкомітет наукового заходу. Існують лише деякі орієнтири, які визначаються критеріями включення у План проведення наукових заходів Міністерства освіти та науки України. Зокрема, у міжнародній конференції, як правило, має бути не менше п'яти країн-учасниць, а кількість учасників – понад сто осіб.

НП є важливим каналом як обміну усною інформацією між вченими, а також публікації письмового звіту про проведення дослідження, представленого на НП [2].

Постановка задачі

Нехай кількість наукових заходів, які слід порівняти та/або визначити їхню інтегральну результативність (ефективність, якість, значимість) в контексті інтересів організації, дорівнює n, а множину їхніх індексів позначимо через $J = \{1,...,n\}$.

Слід розділити проблему визначення результативності НП на кілька етапів, які надають науковому заходу статус НП і впливають на рівень інтегральної якості НП:

1. Визначення політики організації та проведення наукового заходу.
2. Прийняття рішення щодо підготовки та опублікування матеріалів проведення наукового заходу.
3. Визначення кількісних показників результативності НП з точки зору інтересів організації.

Для підготовки математичної моделі визначення ефективності НП розглянемо множину k атрибутів наукового заходу $z_i^i, i \in I = \{1,...,k\}, j \in J$, які можуть бути використані для визначення

кількісних показників результативності НП та застосовані для порівняння результативності різних НП.

Підходи до визначення якості наукової події

Підходи до визначення якості наукової події можуть бути подібними до визначення імпакт-фактора наукових журналів. Але, на відміну від журналів, серед наукових заходів існує суттєво більше різноманіття. Крім того, вони не є регулярними та періодичними – як правило, одноразові або щорічні.

Забезпечення адекватного визначення інтегрального показника якості проведення НП вимагає побудови адекватної моделі для релевантного та стійкого оцінювання ефективності НП.

І. Етап. Здійсненення експертних оцінок деяких параметрів та границь зміни атрибутів НП.

ІІ. Етап. Впровадження динамічного підходу до визначення якості НП, уточнення значень параметрів та границь зміни атрибутів.

ІІІ. Етап. Застосування статичного підходу до визначення якості НП на основі аналізу бази даних НП.

Висновки

У цій роботі досліджено підходи до дослідження результативності НП. Результатом наукового дослідження є публікація, яка підтверджує факт наукового звершення, з яким вчений зможе ознайомити не тільки своїх колег, а й світову спільноту. Наукова робота є не завершеною до тих пір, поки вона не буде опублікована та не проіндексована у наукометричній базі даних. У сучасному науковому світі все більшої значущості набирає публікаційна активність для кожного вченого та наукової організації чи університету в цілому, незалежно від напряму його наукових досліджень.

Одержано такі основні наукові результати:

– здійснено теоретичний аналіз проблеми створення системи оцінки та контролю (результативності (ефективності, якості, значущості) НП;

– розглянуто та досліджено роль суб’єктивної складової у системі підтримки прийняття рішень щодо якості наукових досліджень;

– розроблено підходи до вимірювання кваліметричних показників результативності наукових досліджень;

– визначено критерії оцінки якості та результативності наукової діяльності;

– запропоновано підходи до визначення рейтингів НП з урахуванням необхідності мотивації працівників організацій;

– запропоновано та обґрунтовано інтерпретацію інтегральної результативності НП.

В перспективі проведення досліджень у цьому напрямку доцільно автоматизувати та реалізувати у вигляді СППР інструмент автоматизованого

здобування інформації з наукометричної бази даних та створити програмне забезпечення аналізу якості проведення НП.

Джерела

[1] Симонов П.В. Избранные труды: В 2 т. Т. 1. Мозг, эмоции, потребности, поведение. – М.: Наука, 2004. – С. 396.

[2] Kulamer B., Meester W., Salk J., Blair-DeLeon N., MacPherson G., Moses W., Philippidis A., Chapman C., Smith D., Stoneham I., Vukmirovic K. Recommended Practices to Ensure Technical Conference Content Quality. Впервые представлены Gordon MacPherson на 4th World Conference on Research Integrity Rio De Janeiro, Brazil, June 2, 2015. URL: https://www.ieee.org/conferences_events/conferences/publishing/paper_acceptance_criteria.pdf

[3] Гнатієнко Г.М., Снитюк В.Є. Експертні технології прийняття рішень: Монографія. – К.: ТОВ «Маклаут», 2008. – 444 с.

IMPROVEMENT OF PAIR-WISE COMPARISON METHODS BASED ON GRAPH THEORY CONCEPTS

Sergii Kadenko[a *] https://orcid.org/0000-0001-7191-5636,
Vitaliy Tsyganok[a,b] https://orcid.org/0000-0002-0821-4877,
Zsombor Szádoczki[c,d] https://orcid.org/0000-0003-2586-5660,
Sándor Bozóki[c,d] https://orcid.org/0000-0003-4170-4613,
Patrik Juhász [d], Oleh Andriichuk[a,b] https://orcid.org/0000-0003-2569-2026

[a]Institute for Information Recording of the National Academy of Sciences of Ukraine, Kyiv, Ukraine
[b]Taras Shevchenko National University of Kyiv, Kyiv, Ukraine
[c]Hungarian Academy of Sciences Institute for Computer Science and Control, Budapest, Hungary
[d]Corvinus University of Budapest, Budapest, Hungary
*seriga2009@gmail.com

The paper briefly summarizes the currently available results of recent research on improvement of decision support methods, based on pair-wise comparisons of alternatives. Pair-wise comparison structures (especially, incomplete ones) can be easily represented by non-directed incidence graphs. It turns out that smaller-diameter quasi-regular graphs ensure higher credibility and consistency of expert session results, and maintain stability of preference structures. Particular pair-wise comparison patterns, both taking and not taking the rough ranking of alternatives into account, allow expert session organizers to significantly reduce the minimum required number of pair-wise comparisons, especially on large numbers of alternatives, without compromising the quality of expert data and credibility of expert session results. Following these patterns also allows us to reduce the computational complexity of pair-wise comparison-based decision support methods. Improved pair-wise-comparison-based decision support methods are widely applicable to decision-making problems in weakly-structured subject domains, calling for expert estimation.

Introduction: challenges of pair-wise comparison methods in decision-making

Decision-making in weakly-structured subject domains is characterized by high levels of uncertainty. In such domains it is problematic to measure or numerically describe (rate) objects or decision options. Moreover, even the criteria according to which the objects are compared cannot be easily formalized. Measuring of objects according to these “intangible” [1] criteria is also a challenge.

At the same time, examples of weakly-structured domains span almost all the spheres of human activity. From selection of a car or house [2] to space industry [3], from supply chain management [4] to strategic planning [5]. That is why decision-making in these areas calls for formal description and numeric representation of options and criteria according to which these options are evaluated.

Technically, any measurement of an object is the result of its comparison to some unit value. However, one of the challenges of weakly-structured subject domains is absence of such measurement units. So, as many researchers (from Condorcet to Saaty [1], from Kendall to Hwang & Yoon [6]) show, the best to measure the objects and significance of criteria, by which these objects are evaluated, is to compare them with each other. This assumption provides the basis for a whole family of expert pair-wise comparison (PWC) methods.

The methods prove to be highly efficient, especially in weakly-structured subject domains. However, they have some common problems, which researchers around the globe are trying to solve. The key problems are as follows.

1) Subjectivity of experts. Reasons: cognitive biases, mindsets, background, experience etc.
2) Often, low credibility of expert session results. Reasons: expert estimation errors, inconsistency, incompatibility, incompleteness, insufficiency, low degree of detail of expert data.
3) Large numbers of pair-wise comparisons required to obtain the resulting priorities, and high computational complexity of priority calculation methods.

While selection of competent and unbiased experts (problem 1) is, mostly, the responsibility of the decision-maker (DM), high quality of expert data representation (problem 2), sufficient numbers of comparisons, and manageable computational complexity of priority calculation methods (problem 3) can be ensured through certain mathematical procedures.

So, in our paper we are going to outline several approaches, based on graph theory, targeted at reduction of labor-intensity (problem 3) and improvement of credibility (problem 2) of PWC-based decision support methods.

Problem statement

Let us start by formulating the common problem of priority calculation based on a set of PWC (as posed in AHP and other related methods [1,2]).

We have a set of objects (or alternatives) $\{A_i; i = 1..n\}$, which are compared among themselves by one or several experts. A multiplicative pair-wise comparison matrix (PCM) looks as follows: $A = \{a_{ij}; i,j = 1..n; \forall i,j: a_{ij} = 1/a_{ji}\}$.

We need to find the set of normalized relative alternative weights $W = \{w_i; i = 1..n; \sum_{i=1}^{n} w_i = 1\}$, which would allow us to rate the alternatives.

Methods of priority elicitation from a PCM are quite numerous. They include eigenvector method [1,2], best/worst method [7,8], geometric mean (GM) [9], combinatorial method of spanning tree enumeration [10], logarithmic least squares [11]. All these methods have a lot in common, but produce different results. In this paper we are going to address some modifications of available priority calculation methods, based on graph theory.

Representation of PWC structures through graphs

In the context of this paper, it is important to note, that alongside PCM, graphs represent a handy instrument for representation of PWC. That is, any object can be represented by a graph node (vertex), while a PWC between any two objects can be represented by the respective edge of a graph, and a complete PCM of dimensionality n can be represented by a complete graph with n nodes.

When it comes to PWC methods, credibility improvement and computational complexity reduction can be achieved through several conceptual ways or approaches. The approaches are as follows.

1) Comparisons of alternatives can be performed and ordered according to specific patterns (represented by respective graphs) [12,13].
2) Comparisons can be performed and ordered according to the ranking of alternatives [7,8,14].
3) Comparisons can be performed according to both specific patterns (incidence graphs) and given alternative ranking [15].

In subsequent sections we will address these conceptual approaches and patterns in greater detail.

Modified combinatorial method of spanning tree enumeration

In order to be able to rate a set of n objects, it is enough to perform at least $n-1$ independent PWC (build a connected PWC graph, spanning all the objects, called a spanning tree). For example, we can compare all objects with the 1st one, the last one, the best one, the worst one, the neighboring one in the ranking etc. At the same time, in order to obtain a complete set of PWC, an expert has to compare all objects with each other, and, thus, perform $n(n-1)/2$ PWC. Combinatorial method is based on enumeration of all basic sets of $n-1$ independent PWC. Each of these basic sets can be represented by the respective spanning tree graph (Fig. 1).

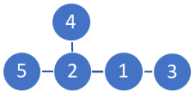


Fig. 1. Spanning tree example for PWC of 5 alternatives

At the same time, every basic PWC set allows us to reconstruct an ideally consistent PCM (ICPCM). Any row or column of this ICPCM (or any other set of $(n-1)$ independent PWC) is, in fact, a set of alternative weights. Priorities which we need to find in the above problem statement, are calculated as GM across all ICPCM (spanning trees) (1).

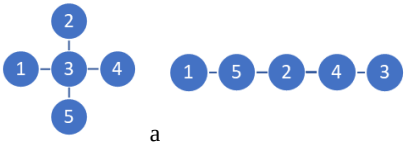
$$w_j^{aggregate} = (\prod_{q=1}^{T} w_j^q)^{1/T}; j = 1..n (1)$$

According to Cayley's theorem on trees [16], the total number of trees, which can be built on n objects, amounts to $T = n^{n-2}$. So, for example, the maximum number of trees we need to analyze in order to calculate the relative weights of 6 objects is $6^{6-2} = 1296$; 7 objects – $7^{7-2} = 16807$; 8 objects – $8^{8-2} = 262144$ etc. These numbers illustrate considerable computational complexity of the method. The ordinary combinatorial method, based on formula (1) turns out to be mathematically equivalent to row GM [9] and LLSM [11]. However, we are using a modified method. Instead of ordinary GM formula (1) it uses weighted GM (2).

$$w_j^{aggregate} = \prod_{k,l=1}^{m} (\prod_{q^k=1}^{T_k} (w_j^{(kq_kl)})^{\frac{R_{kq_kl}}{\sum_{u,p,v} R_{upv}}}); j = 1..n (2)$$

In formula (2) m is the number of experts, and R is the rating of the respective ICPCM, reflecting completeness, detail, consistency, and compatibility of expert data (as explained in [10, 12]). So, beside "transitive" weights $w_j^{(kq_kl)}$, we need to calculate the respective ICPCM ratings. As a result, for larger numbers of objects, the computational complexity of the modified method (2) becomes really tremendous.

So, how can we reduce the computational complexity of the method without compromising the quality (credibility) of the result? We suggest sorting the spanning trees according to their diameter. A diameter of a graph is the longest shortest distance between two nodes. The smallest possible spanning tree diameter value is 2. It is the diameter of a star-type spanning tree, where one of the alternatives is compared to all other alternatives from the set (Fig. 2a). The respective PWC form one row or column of a PCM. The largest possible spanning tree diameter value is $n-1$. It is the diameter of a path-type spanning tree, where each alternative is compared to the neighboring ones (Fig. 2b). The respective PWC are located above the principal diagonal of the PCM.



a b

Fig. 2. Examples of a star-type and path-type trees for 5 alternatives

If we assume that the maximum error made by an expert during PWC $\max_{i,j} abs(a_{ij}^{true} - a_{ij}^{estimated})/a_{ij}^{true} = \delta$, then the maximum error

accumulated on a path-type tree equals approximately $(n-1)\delta$, on a star-type tree 2δ [12]. If we have a spanning tree of diameter k, then

$$a_{ij}^{estimated} = a_{ij}^{true}\frac{(1\pm\delta)^{k_1}}{(1\mp\delta)^{k_2}} = a_{ij}^{true}\left(1 \pm \frac{k\delta}{(1\mp\delta)^{k_2}}\right) \quad (3)$$

In (3) $k_1 + k_2 = k$. Under small δ the accumulated error equals approximately $k\delta$. So, based on these considerations, it makes sense to start with enumerating smaller-diameter spanning trees, because they accumulate smaller expert errors. Experiments from [12] show that weighted GM across smaller-diameter spanning trees yields approximately the same results as weighted GM across all trees.

Enumeration of smaller-diameter trees only significantly reduces the computational complexity of combinatorial method. For example, of 1296 trees, built on 6 alternatives, there are only 6 trees of diameter 2 and 210 trees of diameter 3. This example is illustrated by Fig. 3.

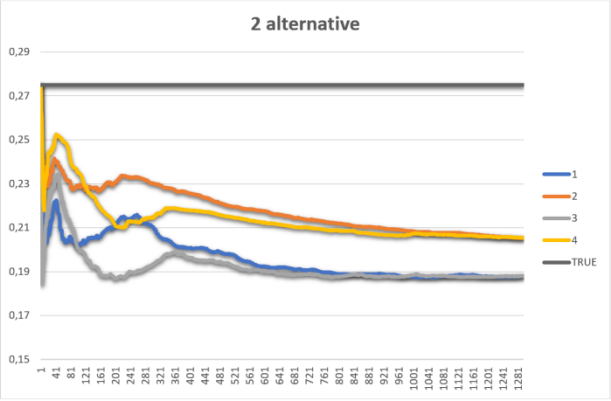


Fig. 3. Alternative weight calculated as GM across all spanning trees. 1 – ordinary GM; unsorted trees; 2 – weighted GM; unsorted trees; 3 – ordinary GM; sorted trees; 4 – weighted GM; sorted trees. Straight line – true non-perturbed value.

Conclusions

Graph representation of PWC structures in decision-making problems is an efficient tool. It allows to reduce computational complexity of decision-making problems and the number of PWC experts need to perform in order to obtain credible results. Graphs with smaller diameter tend to accumulate smaller expert estimation errors, while regular graphs help an expert session organizer to maintain stability of preference structures and patterns on a set of objects.

Future research on the subject will be dedicated to finding the best PWC patterns (in terms of different cardinal and ordinal indicators), both taking and not taking the initial ranking of alternatives into account.

References

1. Saaty, T. (1980). The Analytic Hierarchy Process. McGraw-Hill, New York.
2. Saaty T. (1996). Decision Making with Dependence and Feedback: The analytic Network Process. RWS Publicaitons, Pittsburgh.
3. Tsyganok, V., Kadenko, S., & Andiichuk, O. (2015) Using different pair-wise comparison scales for developing industrial strategies. International Journal of Management and Decision Making, 14(3), 224-250. DOI:10.1504/IJMDM.2015.070760.
4. Oliveira Ramos, M., da Silva, E. M., & Rodrigues Lima-Júnior, F. (2020). A fuzzy AHP approach to select suppliers in the Brazilian food supply chain. Production, vol.30, e20200013. DOI: 10.1590/0103-6513.20200013.
5. De Felice, F. (Ed.) (2016) Applications and Theory of Analytic Hierarchy Process. Decision Making for Strategic Decisions. IntechOpen.
6. Hwang, C.L., & Yoon, K. (1981) Multiple attribute decision making: methods and applications: a state-of-the-art survey. Berlin, New York: Springer-Verlag.
7. Rezaei, J. (2015). Best-worst multi-criteria decision-making method. *Omega. 53*: 49–57.
8. Rezaei, J. (2016). Best-worst multi-criteria decision-making method: Some properties and a linear model. *Omega. 64*: 126–130.
9. Lundy, M., Siraj, S., & Greco, S. (2017). The Mathematical Equivalence of the “Spanning Tree” and Row Geometric Mean Preference Vectors and its Implications for Preference Analysis. Eu-ropean Journal of Operational Research 257(1), 197-208. DOI: 10.1016/j.ejor.2016.07.042.
10. Tsyganok V., Kadenko S., Andriichuk O., & Roik P. (2018). Combinatorial Method for Aggregation of Incomplete Group Judgments. Proceedings of 2018 IEEE 1st International Conference on System Analysis & Intelligent Computing (SAIC), 25–30. DOI:10.1109/SAIC.2018.8516768.
11. Bozóki, S., & Tsyganok, V. (2019). The (logarithmic) least squares optimality of the arithmetic (geo-metric) mean of weight vectors calculated from all spanning trees for incomplete additive (multiplicative) pairwise comparison matrices. International Journal of General Systems 48(4), 362-381. DOI: 10.1080/03081079.2019.1585432
12. Kadenko, S., Tsyganok, V., Szádoczki, Z. & Bozóki, S. (2021). An update on combinatorial method for aggregation of expert judgments in AHP. Production, 31, https://doi.org/10.1590/0103-6513.20210045.

13. Szádoczki, Z., Bozóki, S., & Tekile, H.A. (2022). Filling in pattern designs for incomplete pairwise comparison matrices:(quasi-) regular graphs with minimal diameter. Omega, 107. https://doi.org/10.1016/j.omega.2021.102557.
14. Andriichuk, O., Tsyganok, V., Kadenko, S., & Porplenko, Y. (2020) Experimental Research of Impact of Order of Pairwise Alternative Comparisons upon Credibility of Expert Session Results. 2020 IEEE 2nd International Conference on System Analysis & Intelligent Computing (SAIC), pp. 1-5, doi: 10.1109/SAIC51296.2020.9239126.
15. Juhász, P., Bozóki, S., Szádoczki, Zs., Kadenko, S., & Tsyganok, V. (2021) Nem teljesen kitöltött páros összehasonlítás mátrixok kitöltési mintázatainak vizsgálata. XXXIV. Magyar Operációkutatási Konferencia 2021.08.31.-2021.09.02. Available at: https://www.gazdasagmodellezes.hu/images/stories/konferenciak/GMT2021/MOK_absztraktok.pdf. Accessed December 2, 2021.
16. Cayley, A. (1889). A Theorem on Trees. Quarterly Journal of Mathematics, 23, 376-378.
17. Stevens, S.S., Galanter, E.H. (1957). Ratio Scales and Category Scales for a Dozen Perceptual Continua. Journal of Experimental Psychology, Vol. 54, No 6, 377-411.

МЕТОДЫ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ ДЛЯ ОПРЕДЕЛЕНИЯ ПРИОРИТЕТНОСТИ МЕРОПРИЯТИЙ КОРПОРАТИВНОЙ БЕЗОПАСНОСТИ В РАСПРЕДЕЛЕННЫХ ОРГАНИЗАЦИОННЫХ СИСТЕМАХ

Григорий Гнатиенко [1] [0000-0002-0465-5018] ,
Николай Киктев [2] [0000-0001-7682-280X] ,
Татьяна Бабенко [1] [0000-0003-1184-9483] ,
Алёна Десятко [3] [0000-0003-2860-2188]

[1] Киевский национальный университет имени Тараса Шевченко, ул. Владимирская, 64/13, 01601 Киев, Украина
[2] Национальный университет биоресурсов и природопользования Украины, ул. Героев обороны, 15, 03041 Киев, Украина
[3] Киевский национальный торгово-экономический университет, ул. Киото, 90, 02156 Киев, Украина
g.gna5@ukr.net; nkiktev@ukr.net

Успешная корпоративная защита любой организационной системы состоит в создании и реализации комплексной многоуровневой системы мер, охватывающих основные аспекты функционирования организации: организационную составляющую, управление персоналом, компьютерную технику, локальные сети, программное обеспечение, базы данных и т.д. Как правило, такие системы являются распределенными. Для постоянной и действенной защиты от угроз все указанные аспекты корпоративной защиты должны комплексно взаимодействовать и взаимодополнять друг друга. Распространенной практической задачей является ранжирование альтернатив разной природы. Это производится экспертами высокой компетентности в пределах зон ответственности. С целью сравнения различных способов достижения результирующего ранжирования альтернатив рассматривается формализация задачи в классах однокритериальных и многокритериальных моделей для метрик Кука, Хеминга, Евклида и Литвака. Вводится понятие модифицированной медианы Литвака и компромиссной медианы Литвака, находящейся с использованием минимаксного критерия.

Введение

Под распределенными организационными системами (РИС) будем понимать системы, которые не располагаются на одной контролируемой территории или на одном объекте.

Защита информации в РИС имеет ряд особенностей. С точки зрения защиты информации в РИС необходимо разделить компьютерные сети на корпоративные и общедоступные. В корпоративных сетях все элементы

принадлежат одному ведомству, за исключением каналов связи. В общедоступных коммерческих сетях главным является распространение информации, а вопросы защиты собственных информационных ресурсов решаются, как правило, на уровне пользователей.

Корпоративные сети в свою очередь могут быть связаны с общедоступными сетями. В таком случае администрации (владельцам) корпоративных сетей необходимо предпринимать дополнительные меры безопасности для блокирования возможных угроз со стороны общедоступных сетей.

При построении системы защиты информации в любой РИС следует учитывать: сложность системы, которая определяется количеством подсистем и разнообразием их типов и выполняемых функций; невозможность обеспечения эффективного контроля за доступом к ресурсам, распределенным на больших расстояниях и за пределами страны; возможность принадлежности ресурсов сети различным владельцам.

Особенностью защиты информации от непреднамеренных угроз в РИС по сравнению с сосредоточенными сетями является обеспечение гарантированной передачи информации по коммуникационной подсети. Для этого в РИС необходимо предусмотреть дублирующие маршруты доставки сообщений, предпринять меры против искажения и потери информации в каналах связи.

В РИС все потенциальные преднамеренные угрозы безопасности информации делят на две группы: пассивные и активные. К *пассивным* относятся угрозы, целью которых является получение информации о системе путем прослушивания каналов связи (злоумышленник может получить информацию путем перехвата незашифрованных сообщений или путем анализа трафика. *Активные* угрозы предусматривают воздействие на передаваемые сообщения в сети и несанкционированную передачу фальсифицированных сообщений для воздействия на информационные ресурсы объектов РИС и её дестабилизацию. В РИС все угрозы связаны с территориальной разобщенностью объектов системы. Поэтому, в РИС, наряду с мерами по безопасности информации сосредоточенных ИС, реализуются механизмы для защиты информации при передаче по каналам связи и от несанкционированного воздействия на неё. Все методы и средства, обеспечивающие безопасность информации в защищенной сет можно распределить по группам: обеспечение безопасности информации в пользовательской подсистеме; защита информации на уровне управления сетью; защита информации в каналах связи; обеспечение контроля подлинности взаимодействующих процессов. Неполнота экспертной информации - естественное явление, атрибут многих ситуаций принятия решений, нередко возникает на практике и является одним из видов неопределенности - вместе с нечеткостью, неточностью, ненадежностью, неопределенностью, некорректностью, неадекватностью, недостаточностью и т.д. Неполнота данных и невозможность их дополнить закономерно сопровождает экспертов и принимающих решения в их деятельности [1].

Литературный обзор

Разработку последовательности выполнения этих мероприятий логично формализовать в классе задач группового выбора, ориентированного на определение подмножества равноценных решений или составление нескольких выделенных альтернатив [2]. Такой гибкий подход к упорядочению объектов, при котором каждый эксперт может предлагать свое частичное упорядочение выбранного им подмножества альтернатив, и поиск полного результирующего ранжирования объектов алгебраическими методами является актуальным и перспективным. По избранным подходам, характеру постановки проблем и форме обратной связи при проведении экспертных оценок выделяют следующие основные направления [3, 4]: абсолютные оценки, бальные оценки, метризированные оценки относительной тяжести альтернатив, их параметров, критериев или относительной компетентности экспертов, бинарные отношения, отражающие результаты попарных сравнений альтернатив, – ранжирование альтернатив. Эта работа касается последних двух подходов, и в ней будут описаны задачи, связанные именно с этими направлениями.

Постановка задачи

Идею исследований можно сформулировать следующим образом: подразделения организации задают приоритетность мер по корпоративной безопасности, на выходе – согласованное упорядоченное ранжирование; полное ранжирование, ранее не исследовавшееся; поиск результирующего ранжирования с шагом длиной 1 от молифицированной; провести тестирование алгоритмов распределенной информационной среде.

Материалы и методы

Основными методами, используемыми в данной работе, являются методы теории принятия решений [5], экспертные технологии [3], генетический алгоритм [3] и эвристический алгоритм [4]. Методы теории принятия решений [5] берут начало от теории исследования операций. Экспертные технологии широко применяются в различных сферах человеческой жизнедеятельности и активно развиваются в последние десятилетия. Одним из ключевых понятий экспертных технологий является эвристика [3], которыми могут выступать аксиомы, постулаты, предположения, презумпции, парадигмы, гипотезы, дополнения, предложения и т.д. Эвристики являются эмпирическими методологическими правилами, которые могут помочь найти решение и способствовать определенности некорректно поставленных задач. С помощью эвристических алгоритмов генерируются варианты решений, близкие к оптимальным, но не гарантирующим нахождение оптимального решения задачи. Генетические алгоритмы являются эволюционными алгоритмами, которые используются, в частности, для решения задач оптимизации и

теоретически могут обеспечить нахождение глобального оптимума задачи [2].

Результаты

Результаты вычислительного эксперимента с применением генетического алгоритма

В целях исследования описанного алгоритма авторами были проведены вычислительные эксперименты с разным количеством экспертов и альтернатив. Многочисленные вычислительные эксперименты, проведенные с использованием генетического алгоритма, показывают перспективность применения такого подхода [4, 5]. Для случайно сгенерированных ранжировок 100-200 альтернатив программа вычисляет в пространстве всех возможных ранжировок медианы, которые примерно на 10% ближе к медианам, заданным экспертами, чем модифицированные медианы. Для улучшения решения применяется генетический алгоритм, который улучшает опорные решения, но является субоптимальным.

Результаты вычислительного эксперимента с применением эвристического метода

Авторами был проведен вычислительный эксперимент с применением эвристического алгоритма для матриц размерностей от 6x6 до 10x10, то есть, когда задаются несколько десятков случайно упорядоченных 6-10 альтернатив. Эксперименты с применением указанного метода, проверенные точными методами, дали следующие результаты:

– при слабой согласованности заданных экспертами матриц множество эффективных решений может быть очень большим – количество медиан Кемени-Снелла для некоторых профилей предпочтений составляет до 15% процентов от общего количества ранжировок на множестве альтернатив, то есть – до 0,15* n!;

– среди найденных с помощью указанных алгоритмов решения при применении метода строчных сумм до 50% ранжировок альтернатив являются медианами Кэмэни-Снелла, а при применении метода столбцовых сумм – до 83%.

Таким образом, указанные алгоритмы являются удобным эвристическим алгоритмом для определения множеств медиан Кемени-Снелла и ГВ-медиан [6, 7]. Проведенное исследование подтверждает взаимосвязь между ординальными и кардинальными моделями задания экспертных преимуществ, а также перспективность применения эвристических алгоритмов в задачах экспертного оценивания [8].

Выводы

В данной работе исследованы подходы к определению результирующего ранжирования альтернатив на основе неполных экспертных ранжировок и получены следующие основные результаты:

- описаны особенности принятия решений при определении приоритетности мероприятий по обеспечению функционирования распределенных информационных систем:
- предложены постановки задач определения группового ранжирования альтернатив на основе неполных экспертных ранжировок;
- разработаны алгоритмы для решения задач вычисления коллективного ранжирования большого размера;
- проведены вычислительные эксперименты по исследованию описанных алгоритмов и особенностей задач ранжирования.

Литература

[1] Волошин О.Ф., Гнатієнко Г.М., Дробот О.В. Метод косвенного определения интервалов весовых коэффициентов параметров для метризованных отношений между объектами//Проблемы управления и информатики. 2003. №2, С.34-41.

[2] Додонов А.Г., Ландэ Д.В., Цыганок В.В., Андрейчук О.В., Каденко С.В., Гайворонская А.Н., Распознавание информационных операций, Киев: ООО «Инжиниринг», 2017, 282 с.

[3] Гнатієнко Г.М., Снитюк В.Є. Експертні технології прийняття рішень: Монографія. – К.: ТОВ «Маклаут», 2008. – 444с.

[4] Dodonov, A., Lande, D., Tsyganok, V., Andriichuk, O., Kadenko, S., Graivoronskaya, A.: Information Operations Recognition. From Nonlinear Analysis to Decision-Making. Lambert Academic Publishing, (2019).

[5] Каденко С.В., Циганок В.В., Андрійчук О.В., Карабчук О. В. Аналіз інструментарію підтримки прийняття рішень у контексті вирішення задач стратегічного планування // Реєстрація, зберігання і обробка даних, 2020, Т. 22, № 2. С.77-91.

[6] Гнатієнко Г.М., Тмєнова Н.П. Визначення пріоритетності заходів кібербезпеки при неповних експертних ранжуваннях / Безпека інформаційних систем і технологій. – 2020. – №1(2), с. 9-15.

[7] Hnatiienko H., Tmienova N., Kruglov A. (2021) Methods for Determining the Group Ranking of Alternatives for Incomplete Expert Rankings. In: Shkarlet S., Morozov A., Palagin A. (eds) Mathematical Modeling and Simulation of Systems (MODS'2020). MODS 2020. Advances in Intelligent Systems and Computing, vol 1265. Springer, Cham. https://doi.org/10.1007/978-3-030-58124-4_21. Pp. 217-226.

[8] Гнатієнко Г.М, Круглов О.І., Тмєнова Н.П. Обчислення результуючого ранжування альтернатив на основі використання неповних експертних ранжувань // Безпека інформаційних систем і технологій. – 2020. №3-4, с.27-37.

ПІДХІД ДО ПРОГНОЗУВАННЯ БЕЗПЕКОВОГО СЕРЕДОВИЩА НА ОСНОВІ СТРУКТУРИЗАЦІЇ ПРОЦЕСУ ПЕРЕДБАЧЕННЯ ТА МЕТОДУ ЦІЛЬОВОГО ДИНАМІЧНОГО ОЦІНЮВАННЯ АЛЬТЕРНАТИВ

Віталій Циганок[1], Олександр Григоренко[2], Віктор Голота[2]
[1]Інститут проблем реєстрації інформації НАН України, м.Київ (Україна)
[2]Воєнно-дипломатична академія імені Євгенія Березняка, м.Київ (Україна)
vitaliy.tsyganok@gmail.com, gregorenko@ukr.net, v.holota@ukr.net

У 2021 році в Україні було вперше розроблено опис майбутнього безпекового середовища (БС) до 2030 року. Разом з тим методологія стратегічного прогнозування в інтересах сектору безпеки і оборони держави потребує подальшого розвитку та удосконалення. Пропонується використати американський та вітчизняний досвід для вирішення цього питання за рахунок структуризації процесу прогнозування та використання методу цільового динамічного оцінювання альтернатив.

Вступ

Прогнозування розвитку безпекового середовища (далі – БС) довкола України є одним із пріоритетних напрямів інформаційно-аналітичної діяльності органів державної влади, які беруть участь у проведенні оборонного огляду в нашій державі. В основу процесу прогнозування покладено розроблення імовірних сценаріїв виникнення та розвитку ситуацій воєнного характеру на середньо- та довгострокову перспективу [1]. Вирішення цього завдання, серед іншого, є запорукою побудови ефективної системи національної стійкості як здатності держави швидко адаптуватися до змін БС й підтримувати стале функціонування шляхом мінімізації зовнішніх і внутрішніх загроз [2].

Основний виклад

У 2021 році в Україні було вперше розроблено опис майбутнього безпекового середовища до 2030 року. Українські фахівці сектору безпеки і оборони під БС розуміють сукупність суб'єктів, які впливають або можуть впливати на національні інтереси, становлять або можуть становити для них небезпеку, для нейтралізації якої необхідне залучення (застосування) сил оборони. До БС входять окремі держави та їх національні інтереси (економічні, політичні, військові та інші), навколишнє природне середовище, окремі фізичні та юридичні особи, політичні партії, інформаційний простір держави тощо [3].

Водночас серед завдань розвитку спроможностей сил оборони нашої держави, визначено запровадження мережецентричного підходу до застосування

сучасних систем управління, розвідки та обміну інформацією та створення єдиного розвідувально-інформаційного середовища шляхом автоматизації процесів добування, обробки і доведення розвідувальної інформації [3].

У рамках розгляду питання прогнозування безпекового середовища доцільно врахувати американський досвід. США як одна з провідних держав світу мають досить розгалужену мережу урядових та неурядових "мозкових центрів", які розробляють ті чи інші прогнози. У цьому контексті одним із найпотужніших сегментів американського сектору безпеки і оборони, який займається стратегічним прогнозуванням є розвідувальне співтовариство США.

В американських експертних колах неодноразово піднімалося питання осмислення існуючих проблем стратегічного прогнозування. Так, у 2008 р. провідний американський "мозковий центр" RAND Corporation опублікував доповідь "Оцінка професійних компетентностей аналітиків розвідки". У документі було наведено перелік невдалих спроб представників розвідувального співтовариства США передбачити ті чи інші історичні події, що мали значний вплив на подальший розвиток безпекового середовища [3]. Серед них:

- неточний прогноз щодо набуття Радянським Союзом спроможності створення ядерної зброї (1950-ті рр.);
- нездатність американських аналітиків передбачити швидкий розпад СРСР (1980-ті рр.);
- хибна (перебільшена) оцінка "проблеми 2000 року" у сфері розвитку комп'ютерного програмного забезпечення (1990-ті рр.);
- викривлені прогнози щодо набуття іракським режимом С.Хусейна спроможностей
  з виробництва зброї масового ураження (2000-ні рр.).

З огляду на сьогодення до цього переліку можливо додати відсутність прогнозів щодо: початку збройної агресії РФ проти України (2014 р.); виникнення "Ісламської держави в Сирії та Іраку" (2014 р.); рішення Великої Британії про вихід з ЄС (2016 р.) та виникнення пандемії кронавірусу COVID-2019 (2019 р.).

Саме тому, в розвідувальному співтоваристві США регулярно постає питання щодо удосконалення підходів до організації діяльності аналітиків-прогнозистів. У цьому контексті, значний інтерес, становить дослідження, яке було проведене у 2011-2013 рр. за участю професора Пенсільванського університету (м.Філадельфія, США) Ф.Тетлока та ряду інших науковців цього освітнього закладу на замовлення американського Агентства передових досліджень у сфері розвідки (*Intelligence Advanced Research Projects Activity, IARPA*). Окремі результати цього дослідження були опубліковані у статті Ф.Тетлока "Психологія розвідувального аналізу: чинники точності прогнозування ходу світової політики" [4] та його книзі "Суперпрогнозування. Мистецтво та наука передбачення" (2016 р.) [5].

Метою дослідження було шляхом експерименту встановити чинники, які можуть сприяти підвищенню точності (достовірності) прогнозів, що розробляються аналітиками розвідки.

У дослідженні в якості учасників експерименту взяли участь близько 750 американських аналітиків урядових і неурядових установ, діяльність яких була пов'язана з розробкою геополітичних прогнозів, тобто у значній мірі відповідала завданням з прогнозування безпекового середовища. Упродовж двох років вони давали прогностичні відповіді на тестові запитання щодо близько 200 геополітичних подій різної важливості.

За результатами обробки отриманих у ході експерименту даних Ф.Тетлок зі своїми колегами дійшов висновку, що точність прогнозів зростає, коли аналітики володіють рядом таких компетентностей:

диспозиційні – високий особистий інтелект; відкритий для дискурсу стиль мислення; вміння розробляти різноманітні моделі у контексті прогнозованої ситуації (явища); вміння зробити вибір на користь тієї чи іншої інтерпретації подій; готовність на докорінну зміну своєї думки (парадигми) в разі отримання нової інформації; спроможність коректно оцінювати ступені невизначеності своїх суджень та/або імовірності майбутніх подій;

ситуаційні – дослідження показало, що значний вплив на точність прогнозування здійснює експертно-аналітичне середовище, в якому працюють прогнозисти. Найбільш сприятливим є середовище, в якому аналітики мають можливість розвивати свої здібності та отримувати зворотній зв'язок у контексті об'єкту прогнозування. Іншою запорукою створення сприятливого середовища є об'єднання прогнозистів у об'єкто-орієнтовані експертно-аналітичні групи, що займаються прогнозуванням тієї чи іншої ситуації (явища). Зазначене дає змогу активізувати обмін досвідом та знанням всередині групи та підвищити мотивацію аналітиків, оскільки у такий спосіб вони відчувають увагу зі сторони своїх колег, а подекуди й конкуренцію, що їх також мотивує. Поряд з цим робота в команді сприяє виникненню явища, яке американські дослідники назвали "когнітивним альтруїзмом" – аналітики всередині однієї групи починають ініціативно ділитися один з іншим своїми думками, судженнями та новою цікавою інформацією, наприклад науковими та/або аналітичними статтями. Водночас, працюючи в команді, аналітики більш адекватно реагують на критику та є більш самокритичними, що підвищує ступінь об'єктивності їхньої аналітичної роботи. У ході дослідження було встановлено, що після об'єднання декількох аналітиків в об'єкто-орієнтовані робочі групи точність їхніх прогнозів значно зростала на відміну від аналітиків, які продовжували працювати самостійно.

Слід відмітити, що такої ж думки дотримуються американські дослідники Д.Канеман та Дж.Клейн. Вони вважають, що для розвитку професійних компетентностей, притаманних прогнозистам, необхідно створити експертно-аналітичне середовище, в якому аналітики зможуть постійно розвиватися за рахунок обміну досвідом, знаннями та судженнями, а також регулярно практикуватися в прогнозуванні [6];

біхевіористські – здатність аналітиків-прогнозистів до постійного саморозвитку, а також спроможність ставити під сумнів свою аналітичну парадигму щодо прогнозованої ситуації (явища). У цьому сенсі аналітики повинні намагатися регулярно перевіряти свої ключові припущення, постійно розмірковуючи, шукаючи відповіді на нові запитання та дискутуючи зі своїми колегами. Загалом, це культивування необхідних аналітичних компетентностей шляхом постійного самоаналізу.

Водночас Ф.Тетлок відмічає, що характерною особливістю роботи аналітиків-прогнозистів
у розвідці є той факт, що вони здебільшого працюють з вербальною (нечисловою) інформацією, що суттєво ускладнює процес її автоматизованої обробки (на відміну від числових даних) [4].

Характерною особливістю американських публікацій у сфері розвідувальної діяльності є те, що в них здебільшого піднімаються проблемні питання без конкретних шляхів їх вирішення. Проте, як відомо, у правильно поставленому запитанні закладено 50% відповіді на нього. Тому для вирішення цих та інших проблем прогнозування безпекового середовища пропонується використати вітчизняний досвід інформаційно-аналітичної діяльності, а саме:

(I) практичні результати курсової підготовки аналітиків сектору безпеки і оборони. У 2016-2021 рр. було проведено десятки профільних семінарів (тренінгів), під час яких встановлено, що найбільш точні та обґрунтовані результати прогнозування БС отримуються внаслідок застосування комбінації методів структурованого аналізу. Ключові з них: структурований мозковий штурм; аналіз зацікавлених сторін; когнітивне моделювання; аналіз силових полів; DIMEFIL/PMESII-аналіз; SWOT-аналіз; метод "Альтернативи майбутнього" та інші.

При цьому процес прогнозування БС доцільно розділити на такі етапи:

1) передпрогнозна орієнтація;

2) підготовка масиву формалізованих вхідних даних;

3) формулювання основної проблеми (запитання);

4) декомпозиція основної проблеми (запитання);

5) формулювання основних тенденцій розвитку БС;

6) створення моделі БС;

7) генерування сценаріїв розвитку БС;

8) оцінювання сценаріїв розвитку БС з точки зору ступеню ймовірності їх реалізації та рівня небезпеки для національних інтересів держави;

(II) застосування методу цільового динамічного оцінювання альтернатив для автоматизації запропонованого алгоритму дій. Цей метод розроблений для оцінювання альтернатив рішень [8, 9] був удосконалений для можливості оцінювання альтернатив заходів при побудові стратегічних планів [10] та для оцінювання важливості та ймовірностей альтернативних сценаріїв у ході їх генерування [11]. Для цього доцільно використати напрацювання лабораторії систем підтримки прийняття рішень Інституту проблем реєстрації інформації НАН України. Йдеться про застосування системи підтримки прийняття рішень "Солон-3" [12] та системи проведення експертиз

розподіленими групами експертів "Консенсус-2" [**Ошибка! Источник ссылки не найден.**]. Саме остання програмна система дозволяє структуризувати безпекове середовище, шляхом побудови його ціле-орієнтованої моделі із залученням до процесу побудови груп експертів-аналітиків.

Застосування цих програмних засобів дає змогу:

- створити мережне середовище для дистанційної роботи прогнозистів у складі об'єктно-орієнтованих аналітичних груп;
- забезпечити адміністрування та протоколювання всіх дій аналітиків, залучених до процесу прогнозування;
- автоматизовано проводити попарні порівняння щодо важливості та впливовості різних альтернатив (гіпотез, сценаріїв, тенденцій), що є вкрай ефективним при роботі з вербальною (нечисловою) інформацією;
- створювати когнітивні моделі розвитку безпекового середовища або конкретної ситуації;
- визначати шляхом розподіленого експертного оцінювання ступінь імовірності реалізації сценаріїв розвитку безпекового середовища та рівень небезпеки для національних інтересів держави.

Висновки

Поєднання американського та вітчизняного досвіду інформаційно-аналітичної діяльності дає можливість виробити підходи до вирішення окремих проблем прогнозування безпекового середовища за рахунок структуризації самого процесу прогнозування на вісім етапів та використання методу цільового динамічного оцінювання альтернатив.

Перелік посилань

1. Про затвердження Порядку проведення оборонного огляду Міністерством оборони України (Постанова Кабінету Міністрів України від 31 жовтня 2018 року № 941). URL: https://zakon.rada.gov.ua/laws/show/941-2018-%D0%BF#n10.
2. Стратегія національної безпеки України (Указ Президента України від 14 вересня 2020 року № 392/2020). URL: https://zakon.rada.gov.ua/laws/ show/392/2020#Text.
3. Стратегічний оборонний бюлетень України (Рішення Ради національної безпеки і оборони України від 20 серпня 2021 року, введе в дію Указом Президента України від 17 вересня 2021 року №473/2021). URL: https://www.president.gov.ua/documents/4732021- 40121.
4. P. Tetlock, B. Mellers, , E. Stone, P. Atanasov, N. Rohrbaugh, S. E. Metz, L. Ungar, M. M. Bishop, M. Horowitz, E. Merkle. The psychology of intelligence analysis: Drivers of prediction accuracy in world politics. Journal of Experimental Psychology. 2015. Applied 21(1), p. 1–14. URL: https://doi.org/10.1037/xap0000040.
5. Philip E. Tetlock, Dan Gardner. Superforecasting. The Art and Science of Prediction. Crown. 2016. 352 pp.

6. D. Kahneman, G. Klein. Conditions for intuitive expertise: A failure to disagree. American Psychologist, 2009. №64, p. 515–526. URL: http://dx.doi.org/10.1037/a0016755.

7. Gregory F. Treverton, C. Bryan Gabbard. Assessing the Tradecraft of Intelligence Analysis. RAND, 2008. 76 pp.

8. Тоценко В.Г. Об одном подходе к поддержке принятия решений при планировании исследований и развития. Часть 2. Метод целевого динамического оценивания альтернатив / В.Г.Тоценко // Проблемы управления и информатики. – 2001. – №2. – С. 127-139.

9. Totsenko V.G. One Approach to the Decision Making Support in R&D Planning. Part 2. The Method of Goal Dynamic Estimating of Alternatives / V.G.Totsenko // Journal of Automation and Information Sciences. – 2001. – vol. 33, #4. – P. 82-90.

10. Циганок В.В. Удосконалення методу цільового динамічного оцінювання альтернатив та особливості його застосування / В.В.Циганок // Реєстрація, зберігання і обробка даних. 2013, т.15, №1 – с. 90-99.

11.Циганок В.В., Роїк П.Д. Технологія генерування сценаріїв в умовах невизначеності / Реєстрація, зберігання і обробка даних: зб. наук. праць за матеріалами Щорічної підсумкової наукової конференції 28 вересня 2020 року / НАН України. Інститут проблем реєстрації інформації. – К: ІПРІ НАН України, 2020. – С. 119-121.

12. Свідоцтво про державну реєстрацію авторського права на твір №8669. Міністерство освіти і науки України державний департамент інтелектуальної власності. Комп'ютерна програма "Система підтримки прийняття рішень СОЛОН-3" (СППР СОЛОН-3) / Тоценко В.Г., Качанов П.Т., Циганок В.В.// зареєстровано 31.10.2003.

13. Свідоцтво про реєстрацію авторського права на твір №75023. Комп'ютерна програма „Система розподіленого збору та обробки експертної інформації для систем підтримки прийняття рішень – «Консенсус-2»" / Циганок В.В., Роїк П.Д., Андрійчук О.В., Каденко С.В. // від 27/11/2017.

USING THE INTELLIGENT EXPERT SYSTEMS FOR A STRUCTURING OF EDUCATIONAL AND METHODICAL MATERIALS IN EDUCATIONAL INSTITUTIONS

Sergiy Zagorodnyuk, Bohdan Sus, Oleksandr Bauzha, Valentyna Maliarenko, and Tetiana Zahorodniuk
Taras Shevchenko National University of Kyiv, Ukraine

Application of expert intelectual systems for structuring and automation of laboratory activities in the educational process has been analyzed. It is shown that combination of teaching materials with cross-references and transitions by the means of expert systems allow teachers to to make optimal decisions on time, set priorities and prepare more effective laboratory classes. The implementation of expert systems for classification sections of academic disciplines in both engineering and technical sciences and humanities has been discussed in detail. An approach that allows the development of a new expert system to prepare teachers to perform remote laboratory work, and also to use of virtual simulators or effective practical training has been suggested. Ii is show that expert systems can be used for classifying thematic sections of a wide range of disciplines, including natural sciences, engineering, and humanities. It is demonstrated that a group of teachers of an educational institution can use a common expert system with a branched and structured classifier, thus they can prepare a new course and train a new teacher in different areas of humanities, in particular, philosophy, philology, history and law. As example, it is shown that classifier of history can be built by tree criterions: chronological criterion, criterion for wars and armed conflicts, and criterion for the name and form of government of the state. Thus expert systems are described as high-scale software units without restrictions on the depth of the question tree and the number of logical branches of the classifier.

Introduction

Representatives of educational institutions are forced to widely use the means of structuring and automation in the educational process [1] to ensure its competitiveness and compliance with modern standards of development [2]. Such tools include automated learning systems, electronic reference books and textbooks [3], tools for assessment and control of knowledge [4]. High-tech support of the educational process allows the opportunity for teachers to prepare and conduct classes at a high level. In particular, students of one specialization may be taught related disciplines, which often contain common topics, elements, results, and conclusions. On the contrary, for students of different specialties, the same discipline may contain different material, or the same material, the topics and issues of which are presented in different sequences. As a result, it has

different hierarchical nesting and structure. That formulates the need to organize methodological material in the form of a nested hierarchical structure. The article demonstrates that this technical problem can be solved using intelligent expert systems [8].

The expert system can record teaching materials prepared by one teacher and, through cross-references and transitions, combine them with the teaching materials of another teacher, as well as provide access to such mutually integrated materials to other fellow teachers. Expert systems can be applied to solve a variety of multi-purpose problems [10], which relate to different fields: science [9,14], technology [5,21], business [6], medicine [1, 7, 11-13]. In educational institutions, they allow you to conveniently formalize the experience and traditions of the educational process. The training material can be saved in the form of a hierarchically structured database. Different levels of access can be arranged for professionals with different experiences. If a new teacher appears among the professors of the educational institution, he can be assigned the mode "Read-only", which allows him to quickly get acquainted with the educational material. Later, such a teacher can become a supervisor and author of new educational material.

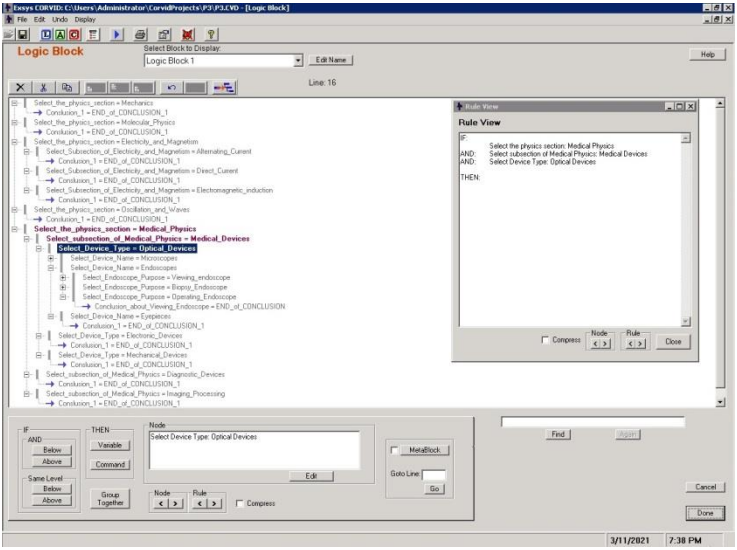


Figure 1: Exsys CORVID administrative application interface.

Thanks to the recommendations and conclusions of the expert system, teachers can prepare and conduct more effective practical, seminar, or laboratory classes. The article demonstrates that expert systems can be used to classify

sections of academic disciplines in both engineering and technical sciences and humanities.

Formation and Structure of the Expert System

Teachers can form and fill expert systems in different ways. The created expert system can be used for students to perform laboratory work, manage virtual simulators, conduct practical classes in history or other humanities. Of course, planning and setting up a virtual laboratory requires a lot of time from the teacher. Moreover, laboratory work in natural, technical, and IT disciplines can differ significantly in the methodological description. However, at the stage of planning laboratory work, the expert system will allow the teacher to make optimal decisions on time, set priorities, and place emphasis.

It is in the process of designing and constructing an expert system that the teacher can introduce the desired interface elements of communication with other disciplines, replace inefficient and outdated issues,and classify practical assignments. Most expert systems are in the form of web applications that can be used on any modern operating system, including GNU Linux, Microsoft Windows, Apple macOS, Google Android. For example, to work with the popular Exsys Corvid test program (see Figure 1), you need to deploy the Oracle Java java runtime environment and the Apache Tomcat web server. That will allow for remote connection and long-term operation of other users of the expert system. Both the Apache Tomcat and Oracle Java software platforms are cross-platform and run on the operating systems noted above.

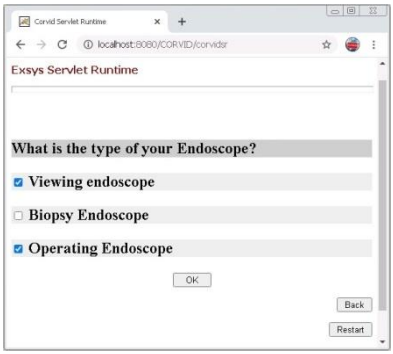


Figure 2: Questions of the expert system with several options answer.

The user of the configured expert system uses a specialized web program in which he can work with distributed multi-level menus configured by the author of the expert system. For example, Figure 1 shows that the endoscope type selection menu item is expanded in the administrative program, and the corresponding web interface in which the user must select the types of endoscopes he uses is shown in

Figure 2. From this figure, it is obvious that the questions can have only one answer or several options, as in this case.

The expert system has an extensive classifier question tree. The user must answer the question of the current level of the classifier menu and then he goes to the next level, where the expert system asks a more specific and specific question. If the system receives answers to all questions to the user - it generates and displays the conclusion of the expert system. At each level of the classifier, the expert system editor is obliged to assign a new interactive question to the user or to formulate an expert decision for him.

Conclusions

1. Expert systems can organize and systematize the knowledge and experience of teachers of educational institutions to ensure a quality level of assignment preparation and conduct of the educational process. Such regulation and structuring can be performed according to different criteria. Therefore the topics and sections of one discipline can be designed in the form of several parallel classifiers. On the other hand, passing through the question tree of different classifiers can lead the user to one conclusion. Expert systems are effective for classifying laboratory and practical activities. Depending on the chosen topic or section of the discipline, the expert system may advise the teacher to prepare remote laboratory work with real equipment or a virtual simulation.

2. Execution of laboratory works in virtual simulators takes place according to a certain structured algorithm. The result of the algorithm is to achieve the object of learning. A large number of laboratory works on one topic and a significant number of options can be combined at the design stage into one common model of data processing and decision-making, described by a common classifier of the expert system. The conclusion of the expert system should contain a sample of the correct execution of virtual laboratory work, as well as control questions with links to related virtual simulations.

3. Expert systems can be used for classifying thematic sections of a wide range of disciplines, including natural sciences, engineering, and humanities. These systems are high-scale software units without restrictions on the depth of the question tree and the number of logical branches of the classifier.

References

[1] S. Hossain, D. Sarma, R. J. Chakma, W. Alam, M. M. Hoque, and I. H. Sarker, “A Rule-Based Expert System to Assess Coronary Artery Disease Under Uncertainty,” in Computing Science, Communication and Security, vol. 1235, N. Chaubey, S. Parikh, and K. Amin, Eds. Singapore: Springer Singapore, 2020, pp. 143–159.

[2] Atomic scanning microscopy URL: https://hub.knu.ua/course/view.php? id=165

[3] Fundamentals of modern AFM microscopy, URL: https://hub.knu.ua/course/view.php?id=170

[4] Genemo, Hussein and Miah, Shah Jahan, "Assessment Issues in Mathematics: Design Science Approach for developing an Expert-System Based Solution" (2015). ACIS 2015 Proceedings. 63. URL: https://aisel.aisnet.org/acis2015/63

[5] I. Merino, J. Azpiazu, A. Remazeilles, and B. Sierra, "2d Image Features Detector and Descriptor Selection Expert System," in 8th International Conference on Natural Language Processing (NLP 2019), Sep. 2019, pp. 51–61, doi: 10.5121/csit.2019.91206.

[6]M. Bohlouli, N. Mittas, G. Kakarontzas, T. Theodosiou, L. Angelis, and M. Fathi, "Competence assessment as an expert system for human resource management: A mathematical approach," Expert Systems with Applications, vol. 70, pp. 83–102, Mar. 2017, doi: 10.1016/j.eswa.2016.10.046.

[7] I. Almarashdeh et al., "Real-Time Elderly Healthcare Monitoring Expert System Using Wireless Sensor Network," SSRN Journal, 2018, doi: 10.2139/ssrn.3415732.

[8] Faculty of Automatic Control and Computers, University "Politehnica" of Bucharest, Romania, C.-O. Truică, A. Barnoschi, and Faculty of Automatic Control and Computers, University "Politehnica" of Bucharest, Romania, "Innovating HR Using an Expert System for Recruiting IT Specialists – ESRIT," JSSD, pp. 1–11, Jun. 2015, doi: 10.5171/2015.762987.

[9]A. K. Akanbi and M. Masinde, "Towards the Development of a Rule-Based Drought Early Warning Expert Systems Using Indigenous Knowledge," in 2018 International Conference on Advances in Big Data, Computing and Data Communication Systems (icABCD), Durban, South Africa, Aug. 2018, pp. 1–8, doi: 10.1109/ICABCD.2018.8465465.

[10] A. J. Paul, "Randomised fast no-loss expert system to play tic-tac-toe like a human," Cognitive Computation and Systems, vol. 2, no. 4, pp. 231–241, Dec. 2020, doi: 10.1049/ccs.2020.0018.

[11] H. Yared Agizew, "Adaptive Learning Expert System for Diagnosis and Management of Viral Hepatitis," IJAIA, vol. 10, no. 02, pp. 33–46, Mar. 2019, doi: 10.5121/ijaia.2019.10204.

[12] S. Mojrian et al., "Hybrid Machine Learning Model of Extreme Learning Machine Radial basis function for Breast Cancer Detection and Diagnosis; a Multilayer Fuzzy Expert System," in 2020 RIVF International Conference on Computing and Communication Technologies (RIVF), Ho Chi Minh, Vietnam, Oct. 2020, pp. 1–7, doi: 10.1109/RIVF48685.2020.9140744.

[13] D. Santra, J. K. Mandal, S. K. Basu, and S. Goswami, "Medical expert system for low back pain management: design issues and conflict resolution with Bayesian network," Med Biol Eng Comput, vol. 58, no. 11, pp. 2737–2756, Nov. 2020, doi: 10.1007/s11517-020-02222-9.

[14] Chaikivskyi T, Sus' B, Bauzha O and Zagorodnyuk S "Multicomponent analyzer of volatile compounds characterization based on

artificial neural networks," CMIS-2020 2608 pp 819-831. URL: http://ceur-ws.org/Vol-2608/paper61.pdf

[15] B. B. Sus, S. P. Zagorodnyuk, O. S. Bauzha, and A. Kozinetz, "Development and Modeling of Remote Laboratory Works for Engineering Education," IOP Conf. Ser.: Mater. Sci. Eng., vol. 1016, p. 012006, Jan. 2021, doi: 10.1088/1757-899X/1016/1/012006.

[16] B. Sus, I. Revenchuk, N. Tmienova, O. Bauzha, and T. Chaikivskyi, "Software System for Virtual Laboratory Works," in 2020 IEEE 15th International Conference on Computer Sciences and Information Technologies (CSIT), Zbarazh, Ukraine, Sep. 2020, pp. 396–399, doi: 10.1109/CSIT49958.2020.9322046.

[17] Sus, B., Tmienova, N., Revenchuk, I., Bauzha, O., Stirenko, S. "Gamification approach to the creation of virtual laboratory works and educational courses," CEUR Workshop Proceedings, 2020, 2711, pp. 68–78. URL: http://ceur-ws.org/Vol-2711/paper6.pdf.

[18] Chaikivskyi, T., Bauzha, O., Sus, B. B., Tmienova, N., Zagorodnyuk, S.: 3D simulation of virtual laboratory on electron microscopy. In: CEUR Workshop Proceedings 2533, pp. 282-291. (2019). URL: http://ceur-ws.org/Vol-2533/paper26.pdf.

[19] Interactive Simulations for Science and Math: 806 million simulations delivered. (2021). URL: https://phet.colorado.edu.

[20] Open Educational Resources: Simulations And Virtual Labs URL: https://libguides.mines.edu/oer/simulationslabs

[21] Gang Li a, Zhenguo Bab i Hongzhi Zhangc DOI: 10.4028/ www.scientific.net/AMR.1055.371

[22] Crocodile ICT. User interface to control the on-screen characters and animation. URL: http://www.aertia.com/en/productos.asp?pid=333&pg=ds

[23] Віртуальні лабораторні роботи з фізичних методів досліджень будови хімічних сполук. URL: http://iht.univ.kiev.ua/virtual-lab/chem/phys-met-chem-comp/story_html5.html

[24] Turnbull J. The docker book containerization is the new virtualization / James Turnbull., 2014. – 321 c.

[25] Yadav A.K., Garg M.L., Ritika M.N. Docker containers versus virtual machine-based virtualization // Advances in Intelligent Systems and Computing,- V. 814. – 2019. – P. 141-150 DOI: 10.1007/978-981-13-1501-5_12

SOME ASPECTS OF OPTIMIZATION OF ELECTRICAL NETWORK MODELING

Amir Sanhinov and Viktor Gurieiev
G.E. Pukhov Institute for Modelling Problem in Energy Engineering of NAS of Ukraine, Kyiv, Ukraine
amir.sanhinov@gmail.com, viktor.gurieiev@ipme.com.ua

Modeling of electrical networks is a computationally expensive task. Besides improving the abstracted versions of the algorithms and theoretical framework, optimizing implementation is essential to ensure short response times in applications. There exists a wide variety of optimization techniques, however, they depend on various factors such as available memory, programming language, compiler, and processor architecture. In this article, we have taken a look at three of the following optimization techniques: branchless programming, multithreading and preprocessing macros, and how they affect mode calculation of the United Electrical Power System of Ukraine in practice. As a reference program, a modified version of the PORT system by Infotec Ltd. translated to C language is used.

Introduction

In order to ensure standardized input and relevance of the optimization techniques, PORT program is modified by removing an interface and database interactions, leaving only mode calculation. The parameters of a calculation such as number of iterations, graph mapping and pivot nodes are then recorded to create an additional version of the program that contains an equivalent number of operations in a single for loop. In this article, the base version of the program is referred to as non-linearized, and the modified version is referred to as the linearize. GNU C Compiler (GCC), version 8.5.0 is used as a compiler.

Branchless programming

Conditional statements are a sizable part of the mode calculation, with the common conditions being whether a node is enabled, whether a branch was viewed and whether a convergence condition is met. The disadvantage of branching code is that in case of branch mispredictions, unnecessary computations are performed. The solution to this is to replace the branching code with a branchless version, which is a common optimization technique. As an example, the following line:

```
if (c == 0) { b = a; } else { b = 0; }
```

can be replaced with:

```
b = (1 >> c) * a;
```

It is important to note that the behavior of a right shift performed this way is unspecified in C language standards and can produce unexpected results on different processor architectures or with different compilers. Additionally, the performance of binary operators can vary depending on the architecture.

GCC compiler is capable of recognizing some patterns and optimizing them with branchless code, however, manual adjustments are required to guarantee the best result. Care should be taken when substituting the branching code, as sometimes it is possible to degrade performance.

Additional advantage of branchless programming is that it multiplicatively increases the multithreading optimizations, since a code is executed per thread. A comparison table of branching and branchless version of the base PORT program is shown in table 1.

Version	Branching	Branchless
Average time with compiler optimizations, ms	6396	6198
Average time without compiler optimizations, ms	15369	6371

Multithreading

Since there are practical limitations on the clock rate of a processor core, there is a limit on the number of instructions per second that can be executed by a program, and by extension on the performance increase. Therefore, a common optimization technique is to divide the work among several cores. While this approach has its limit too due to Amdahl's law, it can provide a performance increase several times larger compared to a non-parallelized code. To achieve this on Von Neumann architecturem, however, certain criteria must be met: For the purposes of this article, we divide parallelization approaches on modern computer systems into three categories:

- Central Processing Unit (CPU) approach;
- Graphical Processing Unit (GPU) approach;
- Cloud computing approach.

GPU parallelization introduces a problem of passing data between Random Access Memory (RAM) and Video RAM (VRAM), and therefore is beyond the scope of this article. In CPU, parallelization is achieved via multithreading. There are two ways to add multithreading: operating system calls, or using a library. Since PORT has a system-independent codebase, OpenMP library was selected. OpenMP is a compiler-side solution for creating code that runs on multiple cores/threads. Because OpenMP is built into a compiler, it does not require using external libraries[7].

While OpenMP gives us significant results when input is linearized, such as a single for-loop, or on embarrassingly parallel code, inside a complicated codebase, such as several nested for- and while loops, the performance decreases

drastically, and the overhead overshadows the benefits gained from parallelization. Below is a comparison table of the equivalent amount of computation (2525783628 elements) in linearized and non-linearized form:

Form	Time on 8 threads, ms	Single-threaded time, ms
Linearized program	2346	8394
Non-linearized program	213787	10409

Conclusions

In this paper, 3 optimization approaches to computer modelling of electrical networks were analyzed. The preprocessing approach showed stable, but small performance increase. In the particular test case that we enacted, branchless code also showed small and stable performance increase, however, it is dependent on the CPU architecture, and therefore must be optionally enabled (based on performance tests) before deployment for reliable gains. The multithreading approach showed approximately a four time performance increase on a linear system of the same size, however, when using nested loops it showed a twenty one times decrease in performance due to overhead.

References

GNU C Compiler profile driven optimizations
https://gcc.gnu.org/news/profiledriven.html
GNU C compiler manual
https://www.gnu.org/software/gnu-c-manual/gnu-c-manual.html
Sorting algorithm optimizations with branchless approach
https://link.springer.com/chapter/10.1007/978-3-642-30347-0_14

An experimental study of sorting and branch prediction

https://dl.acm.org/doi/abs/10.1145/1227161.1370599
Compilers: Principles, Techniques and Tools, 2nd edition
Aho, A., et al. "Compilers: Principles, Techniques and Tools, 2nd Editio." (2007).
Function Call Stack Demonstration
https://cs.gmu.edu/~kauffman/cs222/stack-demo.html
John Carmack preprocessing code
http://www.number-none.com/blow/john_carmack_on_inlined_code.html
OpenMP specification
https://www.openmp.org/specifications/
OpenMP compiler tools
https://www.openmp.org/resources/openmp-compilers-tools/

COMPARATIVE ANALYSIS OF MACHINE LEARNING MODELS WITH DIFFERENT TYPES OF DATA PRESENTATION FOR DETECTING MALICIOUS FILES

Alan Nafiiev[1], Hlib Kholodulkin[2], Andrii Rodionov[1]
[1]National Technical University of Ukraine «Igor Sykorsky Kiev Polytechnic Institute», Kyiv, Ukraine
[2]Taras Shevchenko National University of Kyiv, Faculty of Computer Science and Cybernetics, Kyiv, Ukraine
kholodulkin.hlib@gmail.com

Nowadays, one of the most critical cyber security problems is the fight against malicious software, precisely, the problem of detecting it. Every year, new modern computer viruses are created that are capable of mutation and changing while running. Machine learning methods are increasingly being used to detect such programs. However, these solutions can consume a lot of computing resources to perform their operations. Therefore, creating an optimal machine learning model arises regarding malware detection accuracy and the learning rate. This article presents the recommended methods for selecting features and generating input data for a machine learning model. These methods make it possible to find the best option for preparing features that describe a malicious file that will be used in training and determining the model with the best parameters.

Introduction

More and more progressive, sophisticated malware programs are being created every day. And more and more often, various methods of detecting malicious programs that are not based on the signature approach are proposed [1, 2]. These approaches mainly use heuristic analysis, behavior analysis, or a combination of both. Such methods are actively researched because of their ability to detect zero-day malware without any prior knowledge of it. Such approaches often involve the use of different machine learning models. However, their use always entails several essential tasks that require careful research. First, there is the problem of identifying and extracting a suitable feature set that best represents the input dataset. Secondly, these features need to be correctly formed for a machine learning model so that when performing operations, as little time and computational resources as possible are used. Therefore, it becomes necessary to create an optimal machine learning model that does a good job of detecting malicious programs and, at the same time, consumes a minimum of computing resources. This work aims to research several machine learning models based on support vector machines. Compare them with each other and identify the best model for two different types of data presentation.

Types of data presentation

The dataset used was collected from various internet sources. It consists of 12824 executable .exe files, 11844 of which are malicious and 980 are benign. Figure 1 shows the names of the malicious files.

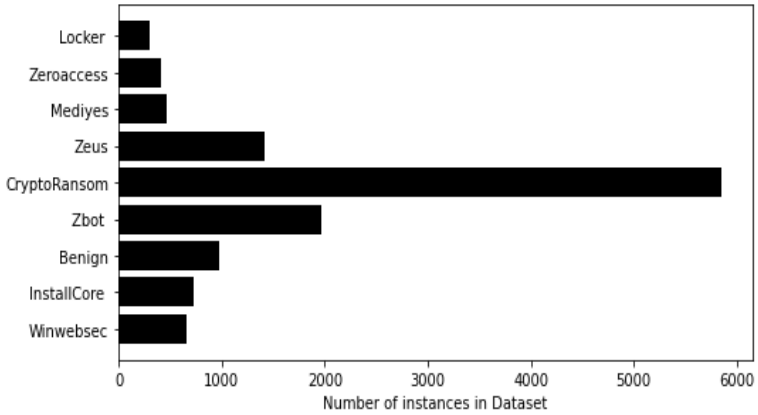


Figure 3. Distribution of files in the dataset

Executable files of the PE format contain a lot of information of various types that can be used to detect malicious programs [3]. We can think of each file as a series of bytes, a trace of instructions or function calls [4]. Different file representation types require different feature extraction, selection, and preprocessing approaches. In this work, two types of data presentation are used. The first is based on the bit information of the file, and the second is based on trace instructions.

For the binary representation of the file, you need to pull out the useful bit information. Each file is processed in turn using the pefile library from python. The functions of this module allow you to get the bits of specific fields from various sections of the PE file. All received bits are alternately formed into one large list. In this paper, two models of binary data representation are considered. One model includes a file data section (Resource), while the other does not.

Further, 2-grams are used to process the bit sequence [5]. For each file, an adjacency matrix is built as follows: all possible 256 bytes (00, 01, ..., ff) are used to create a two-dimensional quadratic adjacency matrix 256 * 256, in which for each pair of bytes in the matrix it is counted how many times in the sequence after the first byte immediately followed by the second byte. After that, the final matrix with dimensions (12824 * 256 ^ 2) is formed, in which each line corresponds to one file and includes its adjacency matrix converted into vector form. Each such vector contains 65536 features that will be used in training the model.

For the second type of data representation, tracing of instructions of the executable file is used. The IDA Pro program generates the disassembled code.

The program saves the disassembled instructions of each file in text format. Then, for each file, instructions (mov, push, call, jz, etc.) are parsed in turn, collected into one large array. From this array, instructions are deleted that occur in total in all files less than a certain threshold (N = 50, 400, 2500, 9500). Increasing this hyperparameter can significantly reduce the dimension of the final matrix and, therefore, reduce the training time for the machine learning model. Figure 2 shows the dependence of the hyperparameter *N* on the number of instructions used.

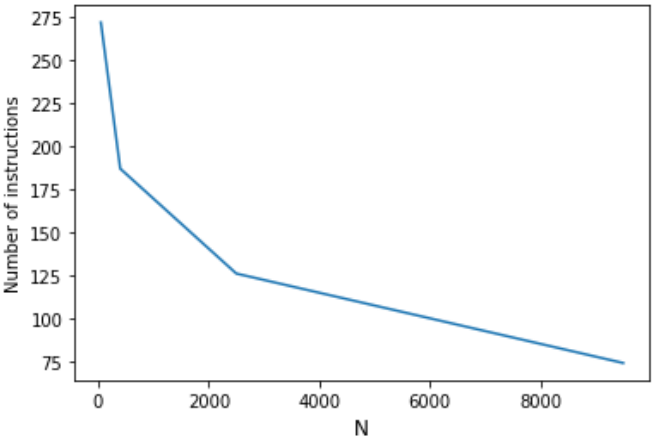


Figure 4. The dependence of the hyperparameter **N** on the number of instructions used

Further, for each dataset element, an adjacency matrix is constructed in the same way as in the binary representation. The number of unique instructions (74, 126, 187, 272) sets the dimension of the two-dimensional quadratic adjacency matrix, in which for each pair of instructions in the matrix it is counted how many times after the first instruction the second one immediately followed. Adjacency matrices are added into vectors to form the final matrix.

Experiment and results

We chose a support vector machine to train the models on the final matrices, which uses a square exponential kernel. The tests were carried out for two types of data presentation with a 40% test to training sample ratio. A wide range of metrics are used to compare model accuracy results, such as F-score, precision, recall, ROC_ AUC, and precision-recall AUC. Figure 3 shows the graph of Receiver operating characteristic AUC.

The results show that the number of used features in the tracing form of data presentation has little effect on the classifier's result. However, you can still select the best option relying on the ROC-AUC metric. The highest score was obtained for the variant with 74 instructions, with a roc_auc result of 0.9946. Also, the

training time for this model is significantly less than that of options with 126, 187, and 272 features, which is a significant plus.

Among the binary type of data presentation, a slight advantage in all metrics was shown by the model, which includes the data section of the file. However, the training time also increased as more data was used. It is also worth noting that the binary representation of the data showed a comparable accuracy result with the best model trained on instruction traces.

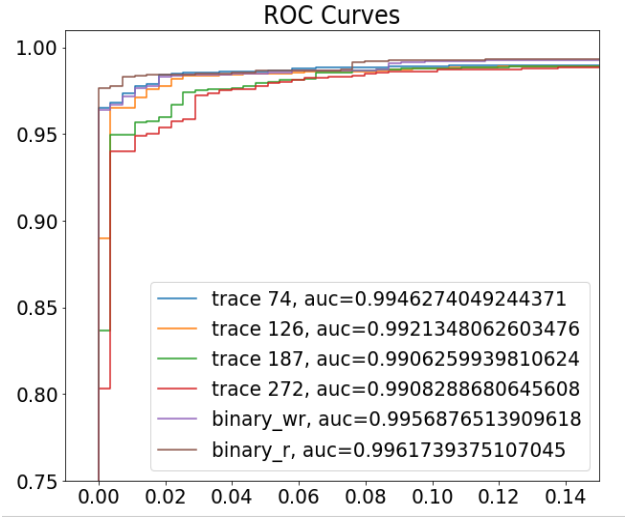


Figure 5. Receiver operating characteristic curves

Conclusions

We trained machine learning models with two different types of data representations in this work. All models demonstrate perfect results in detecting malicious files. The best accuracy indicators were demonstrated by models trained on a binary data representation. The model containing the data section scored the highest across all metrics. In particular, the metrics F0-score - 0.9090, F1-score - 0.9928 and roc_auc - 0.9961. We also managed to identify the optimal variant of the number of used tracing instructions for training the model using the support vector machine. The best was the model trained on the smallest number of instructions - 74. In addition, this model requires the least time for its training, which is a significant plus for using the model in real classification systems.

Reference

1. Faranak Abri, Sima Siami-Namini, Mahdi Adl Khanghah, Fahimeh Mirza Soltani, Akbar Siami Namin « The Performance of Machine and

Deep Learning Classifiers in Detecting Zero-Day Vulnerabilities», 2019.
2. Prachi Chaudhary, «PE File-Based Malware Detection Using Machine Learning», January 2021.
3. Imad Abdessadki, Saiida Lazaar, «A New Classification Based Model for Malicious PE Files Detection», June 2019, International Journal of Computer Network and Information Security.
4. M. Zubair Shafiq, S. Momina Tabish, Fauzan Mirza, Muddassar Farooq, «PE-Miner: Mining Structural Information to Detect Malicious Executables in Realtime», September 2009.
5. Edward Raff, Richard Zak, Russell Cox and other, «An investigation of byte n-gram features for malware classification», February 2018, Journal of Computer Virology and Hacking Techniques.

О ТОЧНОСТИ ИЗМЕРЕНИЙ ПРИ МОНИТОРИНГЕ ЛОКАЛЬНЫХ МАГНИТНЫХ ПОЛЕЙ С ПОМОЩЬЮ СУПЕРПРОВОДЯЩИХ КВАНТОВЫХ ИНТЕРФЕРОМЕТРОВ

Белобородов О.А. [1], Довгополый А.С.1, Токалин О.А.[2]
[1]Central Scientific Research Institute of Armament and Military Equipment of the Armed Forces of Ukraine, MOD of Ukraine, Kyiv, Ukraine,
[2]Institute for Information Recording, NAS of Ukraine, Kyiv, Ukraine

Based on the principle of equivalence the general theory of relativity and the definition of the non-Euclidean metric of space in a non-inertial frame of reference associated with rotation, a geometric (topological) phase is presented that occurs when traversing any closed loop. This approach made it possible to establish a deep physical analogy between various wave effects (both classical and quantum) in closed waveguides that arise under the conditions of their rotation. Due to the coherence of the wave function of spinless charge carriers (Cooper pairs of conduction electrons with oppositely directed spins) in super-conductors in the ground quantum state (superconducting state), the appearance of a geometric phase in closed loops un-der rotation conditions can lead to interference effects in the presence of weak bonds in the loop. To register interference effects, it was proposed to use superconducting quantum interferometers placed in the electric field of a cylindrical or spherical capacitor. In accordance with the general theory of the geometric phase of rotation, the study obtained the basic relations between the geometric phase of rotation and the phase of the wave function induced by an external magnetic field. Taking into account the dependence of the magnitude of the geometric phase of rotation on the voltage across the capacitor and the size of the super-conducting quantum interferometers, and also taking into ac-count the specific range of angular velocities of rotation, the sensitivity of superconducting quantum interferometers to the angular velocity can be regulated by a rational choice of these parameters. As a result of the analysis of the influence of non-uniform motions on the accuracy of measurements of local magnetic fields by SQUIDs during their monitoring on moving carriers, a method is proposed for their minimization by double measurements of the magnetic induction in different modes or simultaneous measurements by two SQUIDs. Based on the results obtained, a new method for measuring magnetic fields was proposed using two superconducting quantum interferometers under conditions of rotational movements, which makes it possible to compensate for the disturbing influence of rotations on magnetic measurements, as well as to simultaneously determine the magnetic induction and angular velocity of rotation.

Введение

Мониторинг локальных магнитных полей Земли имеет большое значение в геофизике, геологии и других областях человеческой деятельности. При дистанционном, например аэромагнитном мониторинге точность измерений считается допустимой, если ошибка измерений не превышает величины 10^{-8} Тл [1]. Наиболее точными современными инструментами для измерения магнитных полей являются сверхпроводящие квантовые интерферометры (СКВИДы), в которых благодаря наличию слабых связей реализуется интерференция волновых функций носителей заряда - куперовских пар электронов проводимости с противоположно направленными спинами, которые в основном состоянии представляют собой бозе-конденсат и имеют общую волновую функцию (ВФ) [2]. Интерференция возможна при наличии разности фаз ВФ с двух сторон слабой связи, которая пропорциональна потоку магнитной индукции через поверхность, охватываемую контуром СКВИДа. В то же время, разность фаз в контуре кольцевого интерферометра при любых неравномерных перемещениях его носителя изменяется вследствие изменения метрического тензора пространства, связанного с перемещающимся объектом [3,4], что не может не оказывать влияния на точность измерений. Такая дополнительная фаза не связана со свойствами объекта и определяется только геометрией связанного с объектом движущегося пространства и называется геометрической фазой [5].

Теоретический анализ

В качестве примера движения неинерциальных систем отсчета рассмотрено вращение с произвольной угловой скоростью Ω вокруг мгновенной оси. В неинерциальной системе отсчета метрический тензор пространства отличается от единичного тензора $\delta_{\alpha\beta}$:

$$\gamma_{\alpha\beta} = \delta_{\alpha\beta} + \sqrt{1-\beta^2}\,G_\alpha G_\beta\,,\ G_\alpha = \beta_\alpha \Big/ \sqrt{1-\beta^2}\,,\ \vec{\beta} = c^{-1}\vec{\Omega}\times\vec{r}\,, \qquad (1)$$

где используются общепринятые обозначения и c - скорость света, $\vec{r}$ - мгновенный радиус-вектор в неинерциальной системе отсчета. Геометрическая фаза с точностью до размерного множителя определяется величиной контурного интеграла:

$$\Delta\theta_C = \oint_C \vec{G}\cdot d\vec{r} = \oint_C G_\alpha dx^\alpha\ . \qquad (2)$$

Сверхпроводимость является квантовым эффектом. Для квантово-механического описания, используя стандартный метод [3], находим интеграл действия для заряженной частицы с нулевым спином в электромагнитном поле с 4-потенциалом $A_i = \left(\phi, -\vec{A}\right)$:

$$S = \int\left(-mcds - \frac{q}{c}A_i dx^i\right) = \int\left[-mc^2\sqrt{1-\left(\frac{v}{c}\right)^2} + \frac{q}{c}\left(A_\alpha - \phi G_\alpha\right)v^\alpha - \frac{q\phi}{\sqrt{1-\beta^2}}\right]dt \text{ , (3)}$$

где q – заряд и m – масса частицы. Затем определяется функция Гамильтона:

$$H\left(\vec{P}\right) = \vec{v}\cdot\frac{\partial L}{\partial \vec{v}} - L = \frac{q\phi}{\sqrt{1-\beta^2}} + \sqrt{\left(mc^2\right)^2 + \left(\vec{P} - \frac{q}{c}\vec{A} + \frac{q}{c}\phi\vec{G}\right)^2 c^2} \text{ ,}$$

(4)

которая используется для построения оператора Гамильтона $\hat{H}\left(\hat{\vec{P}}\right)$ (Гамильтониан) при помощи замены обобщенного импульса $\vec{P}$ на оператор импульса $\hat{\vec{P}} = -i\hbar\vec{\nabla}$ (здесь i - мнимая единица и $\hbar = 1{,}055\cdot10^{-27}\,эрг\cdot с$ или = $1{,}055\cdot10^{-34}\,Дж\cdot с$ - приведенная постоянная Планка, $\vec{\nabla}$ - оператор градиента). В формулах (3) и (4) L – функция Лагранжа, $\vec{v}$ - скорость частицы с компонентами $v^\alpha = \partial x^\alpha / \partial t$, причем $v^2 = \gamma_{\alpha\beta}v^\alpha v^\beta$, по совпадающим верхним и нижним индексам подразумевается суммирование.

Таким образом, в пространстве с метрикой $\gamma_{\alpha\beta}$ (1) на заряженную частицу действует эффективный калибровочный векторный потенциал $\vec{A}_{eff} = \vec{A} - \phi\vec{G}$, который не влияет на ее состояние, но приводит к сдвигу фазы ВФ при интегрировании по замкнутому контуру:

$$\Delta\theta_C = \frac{q}{\hbar c}\oint_C\left(A_\alpha - \phi G_\alpha\right)dx^\alpha = \frac{q}{\hbar c}\left[\Phi\left(\vec{A}\right) - \iint_{\Sigma_C} d\vec{\sigma}\cdot\vec{\nabla}\times\frac{\phi\vec{\beta}}{\sqrt{1-\beta^2}}\right] = \Delta\theta_A - \Delta\theta_R \text{ , (5)}$$

где Σ_C - ограниченная контуром С поверхность. Этот сдвиг фазы распадается на 2 компонента: сдвиг из-за магнитного поля $\Delta\theta_A$, пропорциональный величине магнитного потока через эту поверхность $\Phi\left(\vec{A}\right)$ и сдвиг фазы из-за вращения $\Delta\theta_R$. Как было показано в [6] для СКВИДа, помещенного в сферический или цилиндрический конденсатор под напряжением U, в нерелятивистском случае $\left(\beta \ll 1\right)$ с учетом того, что $\left(q/\hbar c\right) = 2\pi\Phi_0^{-1}$, где $\Phi_0 = 2{,}068\cdot10^{-7}\,Гс\cdot см^2$ или $\Phi_0 = 2{,}068\cdot10^{-15}\,Вб$ – квант магнитного потока, для центрального или осесимметричного скалярного потенциала ϕ выражение (5) упрощается [6]:

$$\Delta\theta_C = 2\pi\left[\frac{\Phi\left(\vec{A}\right)}{\Phi_0} - 2\frac{\Omega_\perp}{c\Phi_0}\iint_{\Sigma_C}\phi d\sigma\right] = 2\pi\left[\frac{\bar{B}_\perp S_{SQID}}{\Phi_0} - 2K_{S,Z}\frac{\Omega_\perp S_{SQID}U}{c\Phi_0}\right], \quad (6)$$

где $\bar{B}_\perp$ - усредненное по площади СКВИДа значение нормальной компоненты вектора магнитной индукции, $\Omega_\perp$ - проекция аксиального вектора угловой скорости вращения на нормаль к плоскости СКВИДа, S_{SQID} - площадь СКВИДа, U - напряжение на контактах цилиндрического или сферического конденсатора, $K_{S,Z}$ - коэффициент, зависящий от формы и размеров конденсатора, используемого для создания электрического поля. Если рассматривать влияние внешнего электрического поля, например атмосферного электричества при магнитной аэросъёмке, то в силу малости размеров СКВИДа поле можно считать постоянным и вынести потенциал в (5 и 6) за интеграл, т.е. считать $K_{S,Z} = 1$.

Измеряемой величиной является бездиссипационный ток через слабую связь (например, переход Джозефсона [2]), который однозначно связан с градиентом фазы ВФ и ограничен критическим током i_C, что определяет рабочий диапазон СКВИДа. Теоретическая зависимость, определяющая квантовую интерференцию ВФ при туннелировании носителей заряда через переход, имеет простой вид:

$$i(\Delta\theta) = i_C \sin\Delta\theta. \quad (7)$$

В результате сдвиг фазы ВФ за счет вращения приводит к ошибкам измерений. Так даже при скорости вращения $\sim 1 град/с$ площади $S_{SQID} \sim 1 см^2$ и напряжении $U \sim 10 В$ поправка достигает $\sim 2{,}5\Phi_0$, что многократно превышает порог чувствительности СКВИДа. Для минимизации ошибок необходимо производить по крайней мере двукратные измерения с разными значениями напряжения на конденсаторе или одновременные измерения двумя СКВИДами. Тогда в соответствии с (6):

$$\Omega_\perp = c\Phi_0\frac{\Delta\theta_1 - \Delta\theta_2}{2\pi S_{SQID}K_{S,Z}\left(U_2 - U_1\right)}, \bar{B}_\perp = \Phi_0\frac{U_2\Delta\theta_1 - U_1\Delta\theta_2}{2\pi S_{SQID}\left(U_2 - U_1\right)}. \quad (8)$$

Связь между измеряемыми величинами и сдвигами фаз ВФ определяется из конкретных рабочих характеристик приборов.

Выводы

Таким образом, в результате анализа влияния неравномерных движений на точность измерений локальных магнитных полей СКВИДами при их мониторинге на подвижных носителях предложена методика их минимизации путем двукратных измерений магнитной индукции при разных режимах или одновременных измерений двумя СКВИДами.

Литература

1. Хмелевский В.К., Геофизические методы исследования земной коры. Международный университет природы, общества и человека «Дубна». -1997.
2. Шмидт В.В., Введение в физику сверхпроводников. Изд. 2, испр. и доп. М.: МЦНМО. 2000.-XIV, 402 с.
3. Ландау Л.Д., Лившиц Е.М., Теоретическая физика. Учебное пособие: для ВУЗов в 10 тт. Т. 2. Теория поля. 8-е изд. М.: ФИЗМАТГИЗ. 2003, 536 с. –ISBN 5-9221-0056-4.
4. Довгополый А.С., Токалин О.А., Геометрическая фаза вращения и её наблюдение в условиях квантовой интерференции//Письма в ЖТФ. -1991. Т.17. вып. 5. –С.44-48.
5. Berry M.V.//Proc. Roy. Soc.:A. -1984. V. 160. Iss. 6. Pp. 1-49.
6. Довгополий А.С., Токалін О.О., Білобородов О.О., Дослідження чутливості надпровідних квантових інтерферометрів до обертальних рухів. ВІСНИК СХІДНОУКРАЇНСЬКОГО НАЦІОНАЛЬНОГО УНІВЕРСИТЕТУ ім. В. Даля, вип. 8(264). Січень 2021, С. 27-33. Doi: 10.33216/1998-7927-2020-264-8-27-33.

МАТЕМАТИЧНА ФОРМАЛІЗАЦІЯ СИСТЕМИ ГЛОБАЛЬНОЇ МАРШРУТИЗАЦІЇ МЕРЕЖІ ІНТЕРНЕТ У ВИГЛЯДІ ТОПОЛОГІЧНОГО ПРОСТОРУ

Віталій Зубок, Андрій Давидюк
Інститут проблем моделювання в енергетиці ім. Г.Є. Пухова НАН України,
м. Київ, Україна
vitaly.zubok@gmail.com, andrey19941904@gmail.com

Стрімкий розвиток топології глобальної комп'ютерної мережі Інтернет вимагає її постійного вдосконалення. При цьому важливим аспектом функціонування мережі Інтернет є захист її системи глобальної маршрутизації від кібератак. Однією з актуальних проблем забезпечення кіберзахисту системи глобальної маршрутизації є відсутність конкретного визначення топології Інтернет як поняття.

У даному дослідженні запропоновано варіант формулювання поняття «топологія Інтернет» засобами математичного визначення топологічного простору, що утворений системою глобальної маршрутизації на множині з'єднань між вузлами – автономними системами. У результаті застосування такого підходу продемонстровано, що маршрути, якими здійснюється трансфер інформації про доступність мережевих префіксів, є елементами топології. Таке представлення створює можливості для використання методів топології з метою дослідження топологічного простору Інтернет.

Вступ

Словосполучення «топологія мережі Інтернет» найчастіше використовується у двох значеннях. Перше - коли йдеться глобальний інформаційний простір, що утворює всесвітню павутину (WWW) з використанням гіперпосилань (URL) між інформаційними ресурсами. Друге - термін «топологія» згадується підчас опису структури комп'ютерної мережі Інтернет. В обох випадках завдяки аналізу окремих груп зв'язків проводяться успішні дослідження властивостей цієї «топології».

У рамках цього дослідження розглянуто аспекти визначення термінології, пов'язаної з топологією Інтернет як комп'ютерної мережі, опису топології та принципів її формування. Це необхідно, для усунення невизначеності, що виникає в наслідок відсутності чіткого визначення топології Інтернет як поняття, та під час проведення дослідження цієї топології. Узагальнене та інтегральне визначення, яке здатне пояснити, з яких елементів складається саме глобальна комп'ютерна мережа Інтернет і як саме вони взаємодіють, допоможе прибрати перепону на шляху пошуку

методів вдосконалення мережі та захисту від кібератак, зокрема, її системи глобальної маршрутизації [1].

Огляд існуючих досліджень топології Інтернет

Інтернет є об'єднанням комп'ютерних мереж, і сутність його саме в об'єднанні за принципами загальної логічної адресації та маршрутизації.

У [2] топологією Інтернет називають неформалізовану конструкцію, у якій кінцеві пристрої, маршрутизатори, автономні системи з'єднані одне з одним. Таке визначення містить очевидні протиріччя. Перше протиріччя полягає в тому, що з'єднання кінцевих пристроїв та маршрутизаторів властиве будь-якій складовій комп'ютерної мережі і, таким чином, таке визначення не є специфічним для мережі Інтернет. Друге протиріччя полягає в тому, що автономна система за визначенням є результатом адміністративного об'єднання маршрутизаторів, а тому припущення про кінцевий пристрій, з'єднаний з автономною системою, є некоректним.

Робота [3], на яку посилаються автори [2], пропонує інше визначення топології Інтернет, а саме – сукупність взаємопоєднаних доменів маршрутизації (routing domains). Ці домени являють собою «групу вузлів (маршрутизатори, комутатори та хости) під єдиним технічним адмініструванням, в яких наявна спільна маршрутизаційна інформація та правила».

Отже, тут топологія являє собою з'єднання груп неоднорідних вузлів, при томі визначення цих груп співпадає з наведеними раніше визначенням автономної системи. Тобто, синтезуючи це визначення з більш загальними поняттями топології комп'ютерних мереж, можна запропонувати два наступні формулювання:

1) топологією комп'ютерної мережі Інтернет є визначений системою глобальної маршрутизації спосіб з'єднання її автономних систем, до яких входять окремі мережеві префікси зі спільною політикою маршрутизації.

2) топологією комп'ютерної мережі Інтернет є структура у вигляді графа, яка може представити взаємодію її автономних систем, до яких входять окремі мережеві префікси зі спільною політикою маршрутизації.

Такі визначення не містять заздалегідь відомих протиріч, і, відповідно до них, топологія Інтернет базується на з'єднаних одна з одною автономних системах (далі - AS). Вона є найбільш часто досліджуваною [3 - 6]. При цьому дослідники акцентують увагу на наступному:

- топологія Інтернет на рівні AS є найглибшою деталізацією мережі Інтернет, інші рівні топології Інтернет частково залежать від топології рівня AS;
- отримати топологію рівня AS відносно просто, а інші рівні топології іноді розглядаються як приватна інформація, і отримати їх значно важче;
- топологія рівня AS не розробляється безпосередньо людьми; натомість це спільний результат технологічних та економічних

потужностей, а отже, його походження та еволюція викликають значний інтерес з боку дослідників.

Уявлення про «відносну простоту» отримання інформації про зв'язки між AS, яке було наведене в [2] та інших роботах, дуже спрощене. Функціонування системи глобальної маршрутизації не дає можливості в окремій AS отримати інформацію про усі зв'язки в мережі безпосередньо за допомогою протоколу BGP. Відсутність єдиної точки огляду (vantage point) для всіх зв'язків та маршрутів підтверджує також Джоф Хостон, видатний вчений і розробник Інтернет, який в [7] акцентує увагу на тому, що «не існує абсолютної правди про топологію; є лише набір відносних маршрутних карт, зібраних окремими BGP-системами».

Вище було дане визначення всіх елементів глобальної комп'ютерної мережі Інтернет, які складають її архітектуру, і яка відрізняє її від будь-якої іншої комп'ютерної мережі. Необхідно визначити, що є топологією Інтернету, який складається з цих елементів. Це дасть змогу усунути наведене протиріччя.

Визначення топологічного простору мережі Інтернет

Існує чітке математичне визначення поняття топології. Воно походить з визначення поняття топологічного простору, яке використовується у загальній топології. Визначення топологічного простору спирається лише на теорію множин, і є найбільш загальним поняттям математичного простору, що дозволяє визначити наступні концепції, такі як безперервність, зв'язність та конвергентність [8].

Нехай існує множина елементів X. Система T відкритих підмножин її елементів є топологією на X, що відповідає вимогам, які звуться аксіомами топології:

- об'єднання довільного сімейства підмножин L з елементів X належить T:
  $$\forall L', L'' \in T : \left(L' \cap L'' \in T\right)$$
- перетин довільного скінченного сімейства L з підмножин елементів X належить T:
  $$\forall L', L'' \in T : \left(L' \cup L'' \in T\right)$$
- сама множина X та порожня множина належать T:
  $$\varnothing \in T,\ X \in T$$

Окремим випадком топологічного простору є простір з дискретною топологією. У дискретному топологічному просторі множина точок не є безперервною, всі точки простору в певному сенсі ізольовані одна від одної. Топологією дискретного топологічного простору (дискретною топологією) є сімейство всіх його підмножин, що відповідають аксіомам топології. Особливістю дискретної топології є те, що її базою послуговують всі підмножини множини X, що складаються з одного елемента.

Мережеві структури утворюють топологічний простір. Метричні структури часто моделюють у вигляді графа. У роботі [9] представлено

топологічний простір логістичного графа на множині його ребер, з чого слідує, що граф є прикладом моделі.

Ми також вбачаємо топологію Інтернет у вигляді «структури як графа, що може представити взаємодію її AS, до яких входять окремі мережеві префікси зі спільною політикою маршрутизації», оскільки представлення мережі у вигляді графа є типовим і загально прийнятим. Якщо граф Інтернет теж є прикладом дискретного топологічного простору, необхідно визначити на множині яких елементів задано цю топологію.

Оскільки топологія – це система підмножин з елементів множини, на якій вона задається, необхідно визначити, на якій саме множині задається топологія.

Основною функцією системи глобальної маршрутизації є обмін інформацією про доступність мережі. Ця інформація про доступність мережі включає список AS, крізь які проходить інформація про доступність мережі. Цієї інформації достатньо для побудови графа зв'язків AS.

Маршрут – це одиниця інформації, яка поєднує набір пунктів призначення з атрибутами шляху до цих пунктів призначення. Набір пунктів призначення – це системи, IP-адреси яких містяться в одному IP-префіксі, розташованому в повідомленні про доступність мережі (network layer reachability information, NLRI). Шлях – це дані у вигляді списку AS, через які проходить інформація про доступність мережі. Дані містяться в полі атрибутів шляху того самого повідомлення [10].

Початкові дані, що доступні нам з системи глобальної маршрутизації, виглядають так.

- множина вузлів AS;
- множина мережевих префіксів (ідентифікатори IP-мереж, до яких прокладаються маршрути), причому кожен префікс прив'язаний до певного вузла, званого «джерела»;
- з'єднання – спрямований зв'язок двох суміжних AS, по якому передається анонс конкретного префіксу; з'єднання можуть розглядатись виключно в контексті передачі по них інформації про доступність мережевих префіксів;
- маршрут до певного префіксу – комбінація з послідовних зв'язків, що закінчується джерелом префіксу;
- множина маршрутів до певного префіксу, яка охоплює всі можливі комбінації послідовних зв'язків, за якими анонсується певний префікс, з кінцевою точкою в джерелі префіксу;
- множина всіх маршрутів до всіх префіксів.

Необхідним є встановлення відповідності між структурами, що утворюються в результаті процесів глобальної маршрутизації та математичним визначенням топології.

Твердження 1. Окремий маршрут до префіксу є топологією на множині з'єднань.

Твердження 2. Сукупність усіх маршрутів до префіксу є топологією на множині з'єднань.

Твердження 3. Сукупність усіх маршрутів до всіх префіксів є розширеною топологією на множині з'єднань.

Наведемо обґрунтування тверджень 1,2,3, попередньо сформулювавши засади та представивши визначення, на яких базується обґрунтування.

За визначенням, маршрут в BGP-4 містить мережевий префікс та шлях до нього. Оскільки нами прийнято, що з'єднання між суміжними AS ми розглядаємо виключно в контексті передачі по ньому маршруту до певного префікса, то маршрути відрізняються виключно шляхами, тому далі замість «шлях» ми використовуватимемо термін «маршрут».

Мережа Інтернет є системою сполучених комп'ютерних мереж, об'єднаних принципами маршрутизації [11]. Отже, мережевий префікс є ідентифікатором окремої комп'ютерної мережі, яку можна вважати сполученою з Інтернет тоді і лише тоді, коли до неї існує маршрут з кожної іншої сполученої мережі. Тому вважатимемо, що в кожній AS наявний принаймні один маршрут до кожного мережевого префіксу.

AS за визначенням обмінюється інформацією про маршрути до префіксів з іншими AS [10,12]. Тому вважаємо, що будь-яка AS має з'єднання з принаймні однією іншою AS, і кожне з'єднання задіяне принаймні в одному маршруті, тобто через нього може бути прокладено шлях від хоча б однієї AS до джерела хоча б одного префіксу.

Обґрунтування Твердження 1.

Маршрут m(p) до префікса p за визначенням є множиною послідовних з'єднань. Дискретна топологія включає в себе всі можливі комбінації елементів множини, на якій вона визначається. Дискретна топологія на множині з'єднань, які належать префіксу, включає в себе всі комбінації з'єднань. Впорядкована безперервна послідовність з'єднань є однією з можливих комбінацій з'єднань. Отже, окремий маршрут до префіксу належить топології на множині з'єднань.

Обґрунтування Твердження 2.

Нехай M є сімейством всіх комбінацій з послідовних з'єднань, по яких передається інформація про доступність префіксу p:

$$M_p = \cup m(p)$$

$$m(p) = \{l_1, l_2, l_3, ..., l_i, ...l(p)\}.$$

Очевидно, що тоді кожне окреме з'єднання l належить якомусь елементові m сімейства M, і не існує жодного з'єднання, яке не належало б до жодного елемента сімейства M. Отже, сукупність усіх маршрутів до префіксу належить топології на множині з'єднань.

Обґрунтування Твердження 3.

Нехай в кортежі (V,E) до V належать всі AS, а до E належать всі з'єднання між AS. Нехай T є топологією, заданою на множині всіх з'єднань. З визначення AS слідує, кожен зв'язок використовується при побудові принаймні одного маршруту до певного префіксу. Тоді всі маршрути до всіх префіксів «задіюють» всі зв'язки у тій чи іншій комбінації, а отже –

сукупність всіх маршрутів до всіх префіксів є розширеною топологією, заданою на множині всіх зв'язків між AS в мережі Інтернет.

Крім того, до топології має належати порожня множина елементів, на яких вона задана. Для топології Інтернет порожньою множиною може послуговувати комбінація зв'язків, по якій AS анонсує самій собі власний мережевий префікс.

Розглянемо приклад на рис. 1. Нехай AS1, AS2,..., AS6 – ідентифікатори автономних систем (AS), при чому AS1 є джерелом префіксу p. Нехай L={1,2,3,4,5,6} – направлені зв'язки суміжних AS, якими передається інформація про доступність мережевого префіксу p.

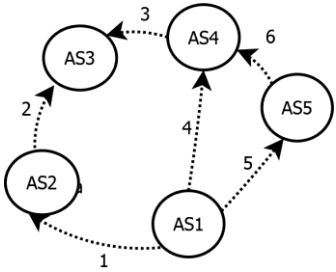


Рисунок 1 – Приклад утворення маршрутів з елементів топології

Розглянемо сімейство підмножин з елементів L, які могли б стати маршрутами до префіксів, що анонсуються AS1. Обов'язкова умова за визначенням маршруту – послідовність з'єднань має бути безперервною, а закінчуватись джерелом префіксу (AS1):

{$\varnothing$}: (AS1,AS1);
{1}: (AS2,AS1);
{1,2},{4,3},{5,6,3}: (AS3, AS1);
{4},{5,6}: (AS4,AS1);
{6}: (AS5,AS1).

Резюмуємо міркування. Нехай існує множина L всіх з'єднань l між вузлами. Нехай існує система елементів множини цих з'єднань T, до якої належать порожня множина, сама множина всіх з'єднань, та будь-які комбінації (об'єднання та перетини) з цієї множини:

$$\exists T \ : \ \varnothing, L \in T; \ \forall L', L'' \in T : \left(L' \cap L'' \in T; \ L' \cup L'' \in T\right)$$

У такому випадку система T відповідає визначенню топології на множині з'єднань L, а пара $G(L,T)$ відповідає визначенню топологічного простору, де множина з'єднань є «носієм» топології. Формалізуємо визначення маршруту до певного префіксу як безперервної послідовності унікальних з'єднань, що закінчується джерелом префіксу. Відповідно до цього визначення всі маршрути належать топології, бо безперервна

послідовність елементів може бути утворена об'єднанням елементів і за визначенням є елементом топології:

$$\left(\begin{array}{l} m(p)=\{l_1,l_2,l_3,...,l_i,...l(p)\}; \\ \\ l_i=\{v_i,v_j\}; \quad v_i,v_j\in V \end{array}\right) \Rightarrow \ \forall p, m(p)\in T$$

$$\forall m(p'), m(p'')\in T: \left(\begin{array}{l} m(p')\cup m(p'')\in L\subset T; \\ m(p')\cap m(p'')\in(\varnothing, L)\subset T \end{array}\right)$$

Сутність «порожня множина» у даній топології позначає з'єднання, по якому вузол-джерело передає анонс власного префікса сам собі.

Таким чином, алгоритм протоколу BGP обирає кращі шляхи для кожного префікса p з елементів топології, що належать топологічному простору цього префіксу та складаються виключно із з'єднань, по яких передається анонс цього префіксу:

$$G_p=\left(L_p, T_p\right).$$

Отже G - сукупність топологічних просторів всіх мережевих префіксів охоплює всі існуючі з'єднання між вузлами Інтернету:

$$G=\bigcup_p G_p$$

Таким чином, даний вираз є математичною формалізацією топологічного простору Інтернет, який утворено системою глобальної маршрутизації.

Висновки

Отже, основні проаналізовані дослідження не дають визначення топології Інтернет, натомість вважають топологією Інтернет щось саме собою зрозуміле – з'єднання на рівні AS. Відсутність конкретного визначення топології Інтернет як поняття та, водночас, проведення дослідження топології Інтернет, є протиріччям, що є перепоною на шляху пошуку нових методів захисту від кібератак на систему глобальної маршрутизації.

Тому запропоновано варіант формулювання поняття «топологія Інтернет» через математичне визначення топологічного простору глобальної комп'ютерної мережі Інтернет, що утворений системою глобальної маршрутизації на множині з'єднань між вузлами – автономними системами. Також показано, що маршрути, які несуть інформацію про доступність мережевих префіксів, є елементами топології. Перехід від нечіткого розуміння поняття топології глобальної комп'ютерної мережі до математичної формалізації топологічного простору відкриває шлях застосування нових методів дослідження властивостей Інтернет.

Перелік посилань

1. Зубок В.Ю. Дослідження зв'язку між топологією та ризиком в наслідок кібератак на глобальну маршрутизацію / В.Ю. Зубок, В.В. Мохор //

Моделювання та інформаційні технології. – ISSN: 2309-7647. – 2018. – №85. – С.23-26.

2. He Y., Siganos G., Faloutsos M., “Internet Topology” In: Meyers R. (eds) Encyclopedia of Complexity and Systems Science. Springer. – 2009. – DOI:10.1007/978-0-387-30440-3_293

3. K. L. Calvert, M. B. Doarand E. W. Zegura, "Modeling Internet topology," in IEEE Communications Magazine, vol. 35, no. 6, pp. 160-163, June 1997. – DOI:10.1109/35.587723

4. Faloutsos M. On Power Law Relationships of the InternetTopology / Faloutsos M., Faloutsos P., Faloutsos C .// Comput. Commun. Rev. – 1999. – №.29. – С.251-263

5. Krioukov D. The Internet AS-Level Topology: Three Data Sources and One Definitive Metric / D. Krioukov // SigComm Computer Communication Review. – Vol.36,issue 1. – С.17-26.

6. Barabasi A.-L. Scale-FreeNetworks / A.-L. Barabasi, E. Bonebau // ScientificAmerican. – 2003.- Vol.5. – С.50-59.

7. G. Huston. “Why Securing BGP is So Damn Hard”. [Електронний ресурс] Режим доступу: https://blog.apnic.net/2019/09/19/why-is-securing-bgp-just-so-damn-hard/. Дата звернення: Sep.15, 2019.

8. Виро О.Я., Иванов О.А., Нецветаев Н.Ю., Харламов В.М. Элементарная топология. – М.: МЦНМО. – 2010. – 352 с.

9. Живицкая Е. Топологические свойства сложных логистических систем / Е.Н. Живицкая // Доклады БГИУР. – №8. – 2012.

10. Y.Rekhter, T.Li and S.Hares. A Border Gateway Protocol 4 (BGP-4). [Електронний ресурс] Режим доступу: https://tools.ietf.org/html/rfc4271. Дата доступу: Вер.29, 2020.

11. В. Г. Олифер, Н. А. Олифер. Компьютерные сети. Принципы, технологии, протоколы: Учебник для вузов. 3-е изд, 2006.

12. Y.Rekhter, D. Estrin, and S.Hots. A Unified Approach to Inter-Domain Routing. [Електронний ресурс] Режим доступу: https://tools.ietf.org/html/rfc1322. Дата доступу: Вер.29, 2020.

ТЕХНОЛОГІЇ ТА ПОБУДОВА СУКУПНОСТІ АНАЛІТИЧНИХ МОДЕЛЕЙ ТУМАННИХ ОБЧИСЛЕНЬ

О.Я. Матов
Інститут проблем реєстрації інформації НАН України, Київ, Україна

Туманні обчислення доповнюють хмарні обчислення, теріторіально наближують вузли обробки і зберігання даних до користувачів. За рахунок скорочення трафіку це дозволяє уникнути безлічі проблем в традиційних хмарних інфраструктурах, які можуть виникнути при необхідності переміщення зеттабайтних обсягів даних. Одночасно скорочується час доставки рішень задач користувачів. Розглянуті архітектура та технології туманних обчислень.
Запропоновані аналітичні моделі туманних обчислень для вирахування характеристик з використанням багатьох потоків та багатьох приоритетів заявок на рішення задач, різних дисципліни обслуговування та їх комбінацій, з врахуванням відмов і різних дисципліни дообслуговування та накопичення в чергах на час відновлення.

Вступ

Розробник цієї відносно нової технології туманних обчислень міжнародний консорціум OpenFog Consortium. Еталонна архітектура OpenFog Reference Architecture (OpenFog RA) туманних обчислень націлена на певний клас бізнес-задач, для яких хмарні структури або інтелектуальні граничні пристрої самі по собі недостатньо ефективні. Вона доповнює традиційну модель хмарних обчислень, забезпечуючи виконання властивих їм функцій на різних рівнях мережевої топології Коротко розглянуті основні технологічні принципи (Pillars) архітектури OpenFog RA, що характеризують приналежність систем до класу OpenFog: безпека, масштабованість, відкритість, автономність, програмованість, працездатність і ремонтопридатність, адаптивність, ієрархічність.

Архітектура туманних обчислень

Еталонна архітектура туманних обчислень доповнює традиційну модель хмарних обчислень, забезпечуючи виконання властивих їм функцій на різних рівнях мережевої топології зі збереженням таких технологічних переваг, як віртуалізація, контейнеризація, оркестровка, керованість, ефективність. Така архітектура усуває обмеження централізованих хмарних рішень, надаючи необхідні для конкретних завдань ресурси і канали зв'язку [1].

Інфраструктура туманних обчислень, за визначенням OpenFog Consortium, є горизонтальною системної архітектурою, яка розподіляє обчислювальні потужності, засоби зберігання, мережеві функції, засоби

управління в усьому континуумі - від «речей» до хмар - з метою наблизити ці ресурси до кінцевих користувачів і прискорити процеси прийняття рішень.

Еталонна архітектура OpenFog Reference Architecture (OpenFog RA) туманних обчислень націлена на певний клас бізнес-задач, для яких хмарні структури або інтелектуальні граничні пристрої самі по собі недостатньо ефективні.

Вузли (Fog Nodes) - це фізичні і логічні компоненти, які виконують обчислювальні функції в туманних мережах. Певною мірою вузли служать аналогами серверів хмарних структур. Вузли, розміщені на кордоні мережі, здійснюють управління доступом і шифрування даних. Вони повинні забезпечувати їх контекстуальну цілісність і ізоляцію, а також агрегувати конфіденційні дані, перш ніж вони будуть спрямовані на наступні рівні.

Хоча вузли і є уніфікованими компонентами туманних мереж, їх архітектурні елементи (включаючи центральні та графічні процесори, а також програмовані логічні матриці і мережеві модулі) варіюються в залежності від місця вузлів в туманною ієрархії і виконуваних ними функцій.

Вузли здатні формувати пористі структури, які здійснюють балансування навантаження, підвищують відмовостійкість і знижують обсяг хмарного трафіку.

У більш складних структурах, що підтримують сервісні моделі FaaS (Fog as a Service), ланцюжки довіри формуються від вузла до вузла і поширюються на хмарні компоненти. Так як вузли можуть динамічно створюватися і розформовуватися, використовувані для цього програмні і апаратні ресурси повинні бути довіреними і атестованими.

Технологічні принципи туманних обчислень

Архітектура OpenFog RA базується на восьми основних технологічних принципах (Pillars), що характеризують приналежність систем до класу OpenFog: безпека, масштабованість, відкритість, автономність, програмованість, працездатність і ремонтопридатність, адаптивність, ієрархічність [1,2].

Архітектура OpenFog RA дозволяє створювати обчислювальні середовища, оснащені широким спектром засобів безпеки, які використовуються в усіх їх компонентах - від «речей» і пристроїв (IoT) до туманних структур і хмарних середовищ. До захисту інфраструктури пред'являється певний набір вимог, яких необхідно дотримуватися при розробці рішень на основі еталонної архітектури. Перш за все це конфіденційність, анонімність, цілісність, довіру, атестація, перевірка і вимірювання. Відповідність перерахованим вимогам гарантує, що рішення OpenFog будуть розгортатися в безпечної обчислювальної середовищі, що забезпечує захист вузлів, мережевих компонентів, процесів управління та оркестровки.

Масштабованість забезпечується динамічним відповідністю технологічних можливостей туманних середовищ бізнес-вимогам з урахуванням таких факторів, як робочі навантаження, продуктивність, вартісні показники. Масштабованість передбачає внесення змін в окремі вузли мережі шляхом додавання апаратного і програмного забезпечення, збільшення числа вузлів на особливо завантажених рівнях або на сусідніх з ними або зменшення їх кількості в міру необхідності, а також додавання систем зберігання даних і засобів аналітики. Це дозволяє нарощувати продуктивність в туманних структурах, змінювати розміри мереж при збільшенні числа додатків, «речей» або кінцевих користувачів, розширювати функціональність засобів забезпечення надійності та безпеки. Крім того, для адекватної підтримки додатками, інфраструктурою управління і оркестровкі величезного числа підключених до мережі «речей» і об'єктів здійснюється модифікація апаратних конфігурацій компонентів вузлів, а також програмного забезпечення.

Відкритість є фундаментальним принципом, який сприяє формуванню масштабної екосистеми туманних обчислень, що забезпечує розгортання і підтримку платформ і додатків Інтернету речей. Вона перешкоджає обмеженню числа постачальників і переважання пропрієтарних «фірмових» рішень, яке може привести до подорожчання продуктів, зниження їх якості і уповільнення інноваційного розвитку систем. Завдяки відкритості вузли можна розміщувати в будь-яких сегментах туманних мереж, розширювати мережі шляхом додавання вузлів, використовувати програмно-конфігуровані вузли, які динамічно формуються.

Відкритість забезпечує інтероперабельність, підтримує побудову компонованої інфраструктури, надання завантаженим додаткам можливість використовувати вільні ресурси, дозволяє реалізувати принцип прозорості місця розташування (Location Transparency). Прозорість гарантує вузлам їх розміщення на будь-якому ієрархічному рівні туманною мережі, а «речам» IoT - оптимізацію мережевих підключень і вибір найбільш зручних маршрутів доступу до обчислювальних та інших ресурсів.

Автономність визначається можливістю виконання вузлами туманних мереж запропонованих їм функцій в умовах відмов або відсутності підтримуючих їх зовнішніх сервісів. В архітектурі OpenFog RA цей принцип поширюється на всі рівні мережевий ієрархії. Наприклад, в разі операційної автономності централізоване прийняття рішень в хмарі не є єдино можливою опцією. Завдяки автономної реалізації вузлів подібні функції виконуються локальними вузлами на кордоні мережі на основі оброблюваних ними даних.

Серед інших областей, де автономність необхідна, - визначення та реєстрація підключень до мережі об'єктів, оркестровка, управління, забезпечення безпеки.

Програмованість, яка повинна підтримуватися і апаратними компонентами, забезпечує високий рівень адаптованості програм, розгорнутих в туманних середовищах. Це дозволяє повністю автоматизувати

повторну постановку завдань, які повинні виконуватися вузлами туманною мережі або кластерами, що складаються з декількох вузлів.

Працездатність і ремонтопридатність туманних інфраструктур є невід'ємним атрибутом сучасної системної архітектури, сприяючи зменшенню часу простоя, зниження ризиків порушення безперервності бізнес-процесів і спрощення технічної підтримки.

Технології, необхідні для їх реалізації і представляють собою найважливіші компоненти архітектури OpenFog RA, охоплюють апаратні системи, програмне забезпечення та операції.

За допомогою архітектури OpenFog RA можливо наблизити обробку даних до джерел їх генерації, щоб приймати оперативні рішення відразу після того, як дані будуть перетворені в осмислений контекст. Цю властивість архітектури її розробники назвали адаптивностью (Agility). Адаптивність дозволяє приймати стратегічні рішення на різних рівнях ієрархії туманних структур, швидко впроваджувати інновації та здійснювати масштабування в рамках загальної інфраструктури.

Моніторинг та управління здійснюються микроконтроллерами: вони відповідають за перевірку стану процесів, що відбуваються, формування сигналів тривоги, запуск додатків для залучення уваги персоналу або автоматичне коректування ситуації, коли спостерігаються істотні відхилення параметрів від заданих значень.

Прикладом становлення одного з сегментів ринку можуть служити безпілотні автономні автомобілі, за управління якими відповідає безліч взаємопов'язаних вузлів. Ці вузли повинні взаємодіяти з вузлами інших машин, вузлами дорожньої інфраструктури, системами управління рухом і хмарними додатками вищого рівня. Таким чином, створюються в результаті розподілені туманні системи будуть охоплювати досить значні за площею території [2].

Архітектура OpenFog RA відкриває можливість побудови інфраструктури, що підтримує послуги FaaS, які допомагають спростити і прискорити розгортання туманних рішень. Як стверджують розробники, до складу FaaS увійдуть не тільки такі добре відомі сервіси, як Infrastructure as a Service (IaaS), Platform as a Service (PaaS), Software as a Service (SaaS), але і багато інших послуг, створені з урахуванням специфічних вимог хмарних структур [1].

Загальна характеристика моделей туманних обчислень

Розробка математичних моделей туманних обчислень є важливим напрямком для виявлення та покращення їхніх характеристик. Стохастичний характер головних чинників і необхідність кількісної оцінки масових процесів на основі теорії імовірності обумовлює використання теорії масового обслуговування. Пропонуються аналітичні моделі туманних обчислень для вирахування характеристик з використанням багатьох потоків та багатьох приоритетів заявок на рішення задач, Розглянуто моделі функціонування вузлів (Fog Nodes) з адаптацією до відмов

характеризуються наступними параметрами: вхідним потоком заявок і потоком відмов вузлів; дисципліною обслуговування заявок; дисципліною поновлення обслуговування заявок після відновлення вузла, що відмовив; дисципліною прийому заявок до черги під час цього відновлення. Сполучення однієї з дисциплін обслуговування з однією з дисциплін поновлення обслуговування після відновлення вузла, що відмовив, і поводженням заявки, обслуговування якої було перервано відмовленням, задає умови самостійних моделей. В моделях використан довільний закон розподілу випадкових величин обслуговування та відновлення вузла, що надає додаткові можливості при дослідженні конкретних туманних обчислень. Розглянуті моделі придатні для аналізу не тільки туманих але і гранічних (перефірійних, крайніх) вузлів, а також логічно ізольованих виконавчих середовищ різних користувачів туманих і гранічних вузлів. Моделі базуються на роботах [3-5].

Всі обчислювальні, регулюючі, управляючі функції в туманних мережах виконуються в вузлах, які є аналогами серверів хмарних структур. Більш того, завдяки реалізованого в ТО принципа (Pillar) автономності централізоване прийняття рішень в хмарі може здійснюватися вузлами на основі даних, що обробляються в них. Тому вузел ТО (Fog Node) в моделюючій системі масового обслуговання (СМО) розглядається, як обслуговуючий прибор.

Математична постановка моделей туманних обчислень

На вхід вузла, як одноканальної СМО з очікуванням, надходять N пуассоновських потоків різнотипних заявок на вирішення задач з інтенсивністю λ_i, $i=\overline{1,N}$. Потоки занумеровані в порядку убування важливості заявок, тобто заявки *i-го* потоку володіють *i-м* пріоритетом в обслуговуванні. Час обслуговування заявок є випадковою величиною з функцією розподілу $B_i(t)$ і двома кінцевими моментами b_i і $b_i^{(2)}$, $t=1,N$.

Природний процес обслуговування заявок у СМО порушується відмовленнями обслуговуючого приладу. Обслуговуючий прилад ненадійний і може виходити з ладу по пуассоновському закону з параметром λ_0. Час відновлення приладу - випадкова величина з функцією розподілу $B_0(t)$ і двома кінцевими моментами b_0 і $b_0^{(2)}$. Прилад може вийти з ладу як при обслуговуванні заявок (при цьому можливі два випадки: заявки повертаються в чергу; заявки губляться), так і у вільному стані. Відмовлення обслуговуючого приладу приводять до росту черги заявок і додаткових затримок у їхньому обслуговуванні. Один з можливих способів адаптації СМО до непродуктивних відмов обчислювальних ресурсів - пріоритетний прийом заявок від різних джерел у чергу до цих ресурсів на час їхнього відволікання. Таке регулювання надходженням потоків може бути досягнуте за рахунок зворотного зв'язку вузлів з джерелами заявок.

У період відновлення обслуговуючого приладу заявки одних потоків у чергу приймаються, а інші – не приймаються. Ця умова задається матрицею-рядком коефіцієнтів $n_i, i = \overline{1,N}$, причому $n_i = 1$ в тому випадку, якщо заявки *i-го* потоку в чергу приймаються, і $n_i = 0$, якщо заявки одержують відмовлення.

Після відновлення обслуговуючого приладу можливі дві дисципліни поновлення обслуговування: із заявок старшого пріоритету і з заявок, обслуговування яких було перервано відмовленням приладу (за умови, що вони не губляться під час відмовлення).

Обслуговування заявок у системі може бути організоване за правилами відносних, абсолютних, змішаних і комбінованих пріоритетів.

При відносних пріоритетах можливість заявок враховується тільки в момент їхнього призначення на обслуговування. На прилад, що звільнився, призначається заявка з найбільшим пріоритетом і її подальше обслуговуванням не переривається іншими заявками.

Абсолютні пріоритети припускають переривання обслуговування заявок низького пріоритету заявками, що надходять на вхід системи, більш високого пріоритету. Перервані заявки в цьому випадку повертаються в чергу й очікують свого дообслуговування.

Проміжної стосовно розглянутого вище дисциплінам є дисципліна обслуговування з комбінованими пріоритетами. При комбінованих пріоритетах час обслуговування всіх заявок, крім заявок старшого пріоритету, розбито на два часових відрізки: на першому відрізку діє абсолютний пріоритет, на другому - відносний.

Змішані пріоритети являють собою сполучення абсолютних і відносних пріоритетів, причому для окремих заявок може бути використане безпріоритетне обслуговування.

Потрібно визначити наступні характеристики обслуговування заявок в вузлах ТО:

w_i - середній час чекання початку обслуговування заявок *i-го* потоку в *i-й* черзі;

v_i - середній час перебування заявок *i-го* потоку в системі (час відгуку системи);

q_i - середнє число заявок *i-го* потоку в *i-й* черзі;

l_i - середнє число заявок *i-го* потоку в системі.

Сполучення однієї з дисциплін обслуговування з однією з дисциплін поновлення обслуговування після відновлення приладу, що відмовив, і поводженням заявки, обслуговування якої було перервано відмовленням, задає умови самостійних задач. Наведені рішення всіх задач.

Висновки

1. Запропоновані аналітичні моделі придатні для аналізу не тільки туманих але і гранічних (перефірійних, крайніх) вузлів, а також *логічно ізольованих виконавчих середовищ різних користувачів* туманих і гранічних вузлів. Останню можливість надає програмованість реалізації в туманних та гранічних структурах сервісних моделей.

2. Ефективність функціонування туманих вузлів може визначатися якістю організації в них обчислювальних процесів, які кількісно оцінюється показниками ефективності [3,4,6]. що базуються на оцінці середнього чи максимального часу перебування задач у системі, часу відгуку обчислювальної системи, затримки щодо припустимих часових термінів і т.п. Отримані формульні вирази характеристик різних моделей придатні для оцінки ефективності туманих вузлів.

Література

1. OpenFog Reference Architecture for Fog Computing. OpenFog_Reference_Architecture_2_09_17-FINAL-1.pdf. https://site.ieee.org/denver-com/files/2017/06/

2. Алексей Чернобровцев. Стандартизация туманной инфраструктуры. Журнал сетевых решений/LAN. 2018. – № 04. https://www.osp.ru/lan/2018/04/13054572

3. Матов А.Я., Шпилев В.Н., Комов А.Д. и др. Организация вычислительных процессов в АСУ. Под ред. А.Я.Матова. Киев, 1989. – 200с.

4. Матов О.Я., Храмова І.О. Проблеми користування і математичне моделювання хмарних обчислень для інтегрованої інформаційно-аналітичної системи державного управління. Реєстрація, зберігання і обробка даних. 2010. Т.12, №2. С. 113-127.

5. Aleksandr Matov. Mathematical models of cloud computing with absolute-relative priorities of providing of computer resources to users in conditions of functioning features and failures // CEUR Workshop Proceedings (ceur-ws.org). Vol-2318 urn:nbn:de:0074-2318-4. Selected Papers of the XVIII International Scientific and Practical Conference on Information Technologies and Security (ITS 2018) PP. 150-159 [http://ceur-ws.org/Vol-2318/paper13.pdf] .

6. Aleksandr Matov Adaptation of cloud computing as optimization of the process of rendering services to users in the conditions of limited computing resources // Selected Papers of the XIX International Scientific and Practical Conference "Information Technologies and Security" (ITS 2019). CEUR Workshop Proceedings (ceur-ws.org). - Vol-2577. - pp. 210-221. ISSN 1613-0073 [http://ceur-ws.org/Vol-2577/paper17.pdf].

ПІДХІД ДО ФОРМУВАННЯ ІНФОРМАЦІЙНОГО РЕСУРСУ ЄДИНОГО ІНФОРМАЦІЙНОГО ПРОСТОРУ СИСТЕМИ ОРГАНІЗАЦІЙНОГО УПРАВЛІННЯ

Володимир Юзефович[0000-0002-6336-9548], Євгенія Цибульська
Інститут проблем реєстрації інформації НАН України, Київ, Україна
uzefv71@gmail.com, evts68@gmail.com

Вступ

На сьогодні питанню створення єдиного інформаційного простору (ЄІП) в різних предметних областях присвячено багато робіт, зокрема [1-3]. Під ЄІП загалом розуміється сукупність баз і банків даних, технологій їх ведення та використання, інформаційно-телекомунікаційних систем і мереж, які функціонують на основі єдиних принципів і за загальними правилами і забезпечують інформаційну взаємодію та задоволення інформаційних потреб користувачів [3]. В роботах зазначається, що ЄІП включає: інформаційні ресурси, що містять дані (знання); організаційні структури, що забезпечують функціонування та розвиток ЄІП; засоби інформаційної взаємодії для забезпечення доступу до інформаційних ресурсів (ІР). Останні дві складові утворюють інформаційну інфраструктуру ЄІП. Більшість публікацій присвячено саме питанням створення та вдосконалення інформаційної інфраструктури ЄІП. Така інфраструктура включає програмні та апаратні засоби, які в сукупності мають забезпечувати надійне зберігання інформації, швидкий доступ до неї (у відповідності до повноважень користувачів) та обмін даними. Разом з тим, очевидно, що ЄІП створюється саме з метою забезпечення інформаційних потреб користувачів, яке неможливо здійснити без формування якісного, з точки зору користувача, інформаційного ресурсу (ІР), який складається з множини пов'язаних інформаційних об'єктів (ІО). Інформаційний об'єкт - цілісна сутність («атомарна», або складана) інформаційного ресурсу, що має певний стан, властивості та функціональне навантаження.

Наприклад, в роботі [1] розглядається реалізація ЄІП у мілітарних і безпекових системах як розподілена база даних з жорстким розподілом прав доступу користувачів до інформації та механізмом фільтрації на основі контексту. Роботу [2] присвячено формулюванню та обґрунтуванню принципів проектування архітектури ЄІП для реалізації спільного управління складним виробничим процесом з розподіленими джерелами інформації. В роботі [3] обговорюються проблеми інформаційної взаємодії користувачів ЄІП окремої предметної області та пропонуються засоби організації швидкого пошуку і візуалізації результатів виконання запитів користувачів. Проблеми інтелектуальної обробки інформації при побудові ЄІП загалом залишаються поза увагою дослідників.

В даній роботі розглядається питання забезпечення інформаційних потреб користувачів системи організаційного управління (СОУ), під якою

розуміється система, об'єкт та суб'єкт управління якої включає соціальну (людську) складову [4].

Мета роботи

Вироблення комплексного (цілісного) підходу до формування інформаційного ресурсу єдиного інформаційного простору СОУ, орієнтованого на якісне забезпечення інформаційних потреб користувачів.

Основний зміст

Очевидно, що будь-яка з відомих СОУ є ієрархічною (багаторівневою) системою. Розгляд інформаційних потреб користувачів різних ієрархічних рівнів СОУ свідчить про суттєву їх відмінність, яка пояснюється відмінністю покладених на них функціональних завдань та задач. На вищих ієрархічних рівнях вирішуються переважно задачі стратегічного (в рамках системи, що розглядається) рівня. Відповідно, користувачі ІР такого рівня оперують здебільшого комплексними, узагальненими інформаційними об'єктами. Що стосується користувачів нижніх ієрархічних рівнів, які вирішують задачі (умовно) тактичного рівня, то їх інформаційні потреби переважно зосереджені на даних про об'єкти, які прямо характеризують поточні процеси та явища. Такими є об'єкти об'єктивної реальності, що переважно (однак не завжди) належать до матеріального світу. Принциповою відмінністю таких об'єктів від ІО користувачів стратегічного рівня є доступність для безпосереднього спостереження (вимірювання, оцінки) їх основних властивостей, для чого в СОУ організаційно, або принаймні функціонально, виділяють підсистему моніторингу. З іншого боку, навіть на одному ієрархічному рівні системи управління, окрім персоналу, що вирішує функціональні задачі за прямим призначенням СОУ, існує персонал, що вирішує кадрові, фінансові та інші логістичні задачі. Очевидно, що такі користувачі інформаційного ресурсу ЄІП оперують здебільшого спільними інформаційними об'єктами, однак розглядають їх із суттєво різних сторін. Отже, їх інформаційні потреби задовольняються даними про спільні об'єкти, однак ці дані характеризують зовсім різні властивості ІО. Таким чином, можна зробити висновок, що *дані щодо одних і тих самих ІО не здатні ефективно забезпечити інформаційні потреби користувачів ієрархічної СОУ*.

Разом з тим, очевидно, що комплексні та узагальнені ІО на самому початку формуються на основі інформації, що надходить від підсистеми моніторингу, шляхом цілеспрямованої багатоетапної її обробки в деякій інформаційно-аналітичній підсистемі (ІАП) СОУ. При цьому, узагальнені об'єкти вищого рівня переважно формуються з ІО попередніх рівнів. Таким чином, поетапна обробка інформації із утворенням об'єктів більш високого ієрархічного рівня забезпечує єдність інформаційного простору з точки зору процедури формування інформаційного ресурсу. Така єдність полягає в тому, що усі інформаційні об'єкти з самого початку формуються на основі

спільних даних від підсистеми моніторингу, а отже, і всі рішення, що приймаються СОУ, мають спільну інформаційну основу.

На рисунку 1 показано узагальнену схему формування інформаційного ресурсу ЄІП СОУ.

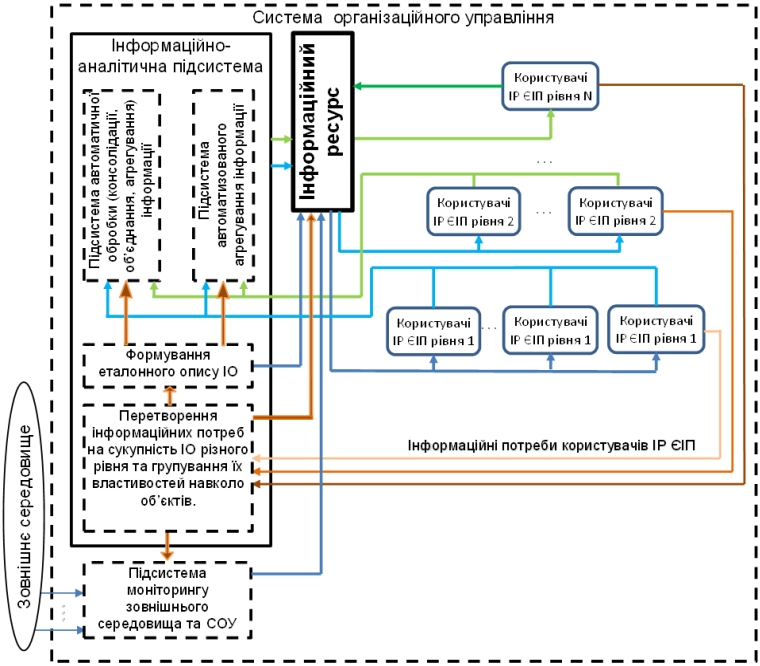


Рисунок 1 – Схема формування інформаційного ресурсу

Схема передбачає виділення в ІАП СОУ складових, де здійснюється перетворення інформаційних потреб на сукупність ІО, формується еталонний опис таких об'єктів, здійснюється автоматична або, якщо це неможливо, автоматизована обробка інформації.

На рисунку 2 показано порядок визначення інформаційних потреб користувачів СОУ.

Першим етапом визначення інформаційних потреб є формування переліку усіх функціональних завдань, що вирішуються її персоналом. Важливим є також врахування регламенту вирішення функціональних завдань на різних ієрархічних рівнях системи. Очевидно, що завдання стратегічного рівня управління вирішуються із більш тривалим інтервалом часу, ніж задачі тактичного рівня. Це пояснюється меншою динамікою стратегічної обстановки та більшими часовими витратами, необхідними для реалізації стратегічних рішень. Отже, оновлення даних про ІО вищих рівнів управління СОУ може здійснюватися у меншому темпі, що, відповідно, може зменшити навантаження на підсистему моніторингу. Далі, на основі

переліку функціональних завдань, визначається множина усіх ІО, які використовуються користувачами під час їх вирішення. Отриманий перелік ІО підлягає багатоетапній декомпозиції із утворенням вичерпного переліку «атомарних» об'єктів, які, у різних їх комбінаціях, дозволяють сформувати будь-який ІО, що складає інформаційні потреби. Під «атомарними» ІО розуміються об'єкти, властивості яких піддаються вимірюванням або оцінюванню засобами підсистеми моніторингу. Тобто, перелік «атомарних» ІО складає інформаційні потреби користувачів СОУ, які мають бути забезпечені її підсистемою моніторингу. Інші ІО мають формуватися ІАП СОУ в процесі багатоетапної обробки даних «атомарних» ІО.

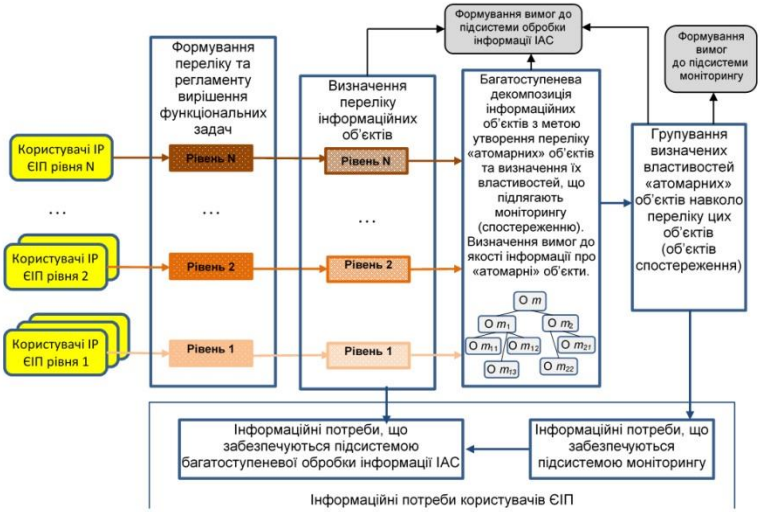


Рисунок 2 – Порядок визначення інформаційних потреб користувачів СОУ

Враховуючи те, що (як вже зазначалося) різних користувачів цікавлять різні властивості одних і тих самих об'єктів, останнім кроком визначення інформаційних потреб є групування різних наборів властивостей кожного «атомарного» об'єкта для надання їх різним користувачам. Таке групування властивостей спростить постановку задачі підсистемі моніторингу СОУ та організацію її виконання. Окремою важливою задачею є визначення вимог до якості інформації про «атомарні» ІО, яка по суті визначатиме (обмежуватиме) якість забезпечення інформаційних потреб користувачів в цілому.

До таких показників якості інформації відносяться:

актуальність - відповідність даних поточному моменту часу (період актуальності інформації є пропорційним швидкості плину процесів, які така інформація відображає (характеризує));

точність - наближеність значень інформаційних одиниць даних до реального значення (похибка визначення);

достовірність - ступінь відображення об'єктивної реальності;

своєчасність (на противагу оперативності) - формування усього необхідного користувачам обсягу інформації безпосередньо до моменту її затребуваності;

повнота - відповідність складу та обсягу наявної інформації (з урахуванням її точності та достовірності) інформаційним потребам користувачів.

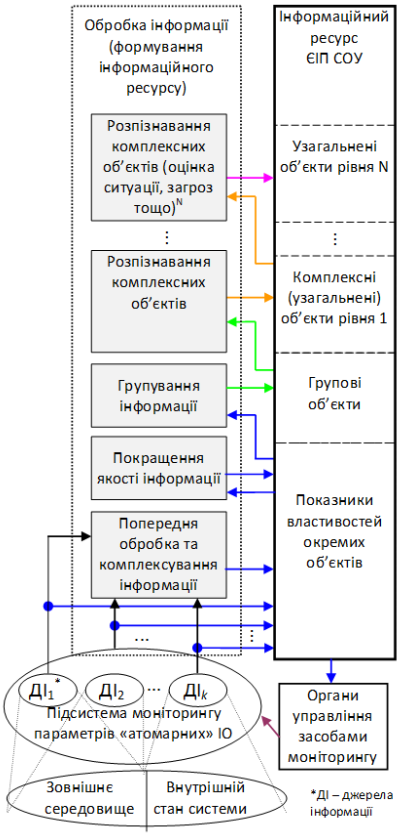


Рисунок 3 - Етапи перетворення (обробки)

Оцінка показника точності інформації, що отримується за допомогою технічних засобів, як правило, не складає проблем. При цьому таку інформацію можна вважати достовірною. Показники точності та

достовірності інформації, отриманої від інших джерел, встановлюються дослідним шляхом.

Врахування регламенту вирішення завдань персоналу СОУ забезпечує можливість використовувати показник своєчасності інформації замість показника оперативності. Показник повноти може бути оцінено через співставлення переліку та обсягу даних про ІО, які визначають інформаційні потреби користувачів СОУ, та обсягів наявної у інформаційному ресурсі інформації із урахуванням її якості.

На рисунку 3 показано процедуру багатоетапної обробки інформації в інформаційно-аналітичній підсистемі СОУ з метою формування інформаційного ресурсу ЄІП. Забезпечення такої процедури можливе за рахунок комплексного застосування:

- методів та способів попередньої обробки інформації (оцінка «вхідної» якості; формалізація різнорідних даних, що надходять від різнотипних джерел);
- методів та способів ідентифікації об'єктів спостереження зовнішнього середовища та консолідації показників їх властивостей навколо ідентифікованих інформаційних об’єктів;
- методів об'єднання різнорідної (числової та лінгвістичної) нерівноточної та різної достовірності інформації для підвищення її якості;
- методів згладжування параметрів ІО за рядом отриманих значень у часі із похибками від одного або декількох джерел;
- методів прогнозування значень параметрів ІО та стану ІО в цілому (зокрема – використання моделей прогнозування часових рядів);
- методів групування ІО для подальшого розпізнання комплексних (узагальнених) об'єктів;
- методів агрегування даних про ІО із утворенням комплексних (узагальнених) об'єктів;
- методів розпізнавання ІО різного рівня.

Необхідно зазначити, що способи попередньої обробки інформації тісно пов’язані із методами її подальшої обробки. Наприклад, при використанні нечіткої логіки чи нечіткого інтегрування для обробки інформації, вхідні дані мають подаватись та зберігатись у вигляді нечітких множин.

Повертаючись до рисунку 1, зазначимо, що при виборі конкретних методів вирішення задач обробки інформації, перевагу необхідно надавати підходам, які забезпечують автоматичне вирішення завдань. Наприклад, у [5] розглядається метод автоматичної кластеризації, що дозволяє вирішувати задачу групування рухомих об’єктів без участі оператора.

В цілому, запропонований підхід до формування інформаційного ресурсу ЄІП СОУ, як ієрархії взаємопов’язаних інформаційних об’єктів різного рівня узагальнення для різних рівнів користувачів СОУ, має вказані нижче переваги.

1. Цільова спрямованість підходу на визначення та задоволення інформаційних потреб користувачів різних рівнів, що відповідає головній меті формування ЄІП.

2. Високий рівень автоматизації процесу формування інформаційного ресурсу ЄІП.

3. Можливість накопичення інформації «навколо» інформаційних об'єктів, що складають інформаційні потреби користувачів для забезпечення швидкого доступу до даних щодо поточного стану таких об'єктів. Можливість використовувати для пошуку власну (притаманну даній групі користувачів) термінологію.

4. Відсутність (мінімальна кількість) зайвої інформації, що надходить на запит користувача до ІР ЄІП.

5. Можливість швидкого цілеспрямованого детального аналізу стану інформаційних об'єктів за рахунок прямого (безпошукового) доступу до складових об'єктів, що його визначають.

6. Більш низька динаміка інформаційних об'єктів верхніх ієрархічних рівнів дозволяє продовжувати вирішувати завдання користувачам ЄІП протягом певного часу, в умовах відсутності (часткової втрати) інформації про «атомарні» об'єкти чи об'єкти нижчого ієрархічного рівня.

7. Можливість відновлення інформації про інформаційні об'єкти нижчих ієрархічних рівнів у випадку зворотності процедури отримання даних про стан інформаційних об'єктів верхніх рівнів.

До недоліків підходу слід віднести:

1. Помітне збільшення обсягів інформації, що підлягає накопиченню та збереженню в ІР ЄІП.

2. Висока аналітична складність та неоднозначність процедури багатоетапного формування ІР ЄІП.

3. Необхідність передбачення в організаційно-штатній структурі СОУ додаткових фахівців щодо формування та ведення інформаційного ресурсу із високими кваліфікаційними вимогами до них.

Висновки

У роботі запропоновано комплексний підхід до формування інформаційного ресурсу єдиного інформаційного простору системи організаційного управління, як ієрархії взаємопов’язаних інформаційних об’єктів, що визначаються інформаційними потребами користувачів СОУ та у сукупності здатні максимально їх забезпечити. Запропоновано перелік методів обробки інформації для автоматизації процесу створення ІР ЄІП СОУ.

Оскільки будь-яку систему, що характеризується функціональною спеціалізацією елементів, можна розглядати як ієрархічну, запропонований підхід може бути поширено і на інші соціотехнічні системи.

Перелік посилань

1. Fabian Angelstorf; Stefan Apelt; Nico Bau; Norman Jansen; Sylvia Käthner: Shared Information Space // 2017 International Conference on

Military Communications and Information Systems (ICMCIS) 15-16 May 2017;
2. Pavel V. Senchenko, Yury B. Gritsenko, Oleg I. Zhukovsky, R.V. Mesheryakov Architectural Principles of Common Information Space Development for Control of Complex Production Processes. Online: https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8687007;
3. B. Baliś, T. Bartyński, M. Bubak, G. Dyk, T. Gubała, M. Kasztelnik: The Common Information Space: A Framework for Early Warning Systems, // Cracow Grid Workshop 2012, Kraków, Poland, 22 October 2012. Online: http://dice.cyfronet.pl/products/cis;
4. Dodonov O.G., Gorbachyk O.S., Kuznietsova M.G. Dynamic Reconfiguration in Automated Organizational Management Systems // Selected Papers of the XX International Scientific and Practical Conference "Information Technologies and Security" (ITS 2020), Kyiv, Ukraine, December 10, 2020. Online: http://ceur-ws.org/Vol-2859/paper11.pdf. Pp.129-141.
5. Юзефович В.В. Вирішення задачі групування рухомих об’єктів у системах моніторингу з використанням методів кластеризації. Реєстрація, зберігання і обробка даних. – 2015. – Т. 17, № 4, С. 31-37.

ВІЗУАЛЬНИЙ АНАЛІЗ ДАНИХ — КІБЕРЗАГРОЗИ ПЗ

Катерина Кунєва
Київський національний економічний університет імені Вадима Гетьмана,
Київ, Україна
kunieva.kateryna@gmail.com

Зловмисне програмне забезпечення є одним з найпоширеніших кіберзагроз. Питання кібербезпеки постає особливо гостро. Візуалізація у кібербезпеці допомагає організувати великі набори даних для легшого розпізнавання закономірностей, контексту та інформації, це може заощадити значну кількість часу на аналіз даних або відсутність ключових точок даних. Візуалізацію можна використовувати як дослідний інструмент для пошуку та зниження ризиків. Дозволяє це реалізувати софт Maltego — інструмент для побудови та аналізу зв'язків між різними суб'єктами та об'єктами. Maltego пропонує можливість підключення даних і функцій із різних джерел за допомогою Transforms. Дана програма використовується аудиторією від професіоналів у сфері безпеки до судових слідчих, журналістів-розслідувачів та дослідників ринку, завдяки діапазону можливих випадків використання, починаючи від розвідки загроз до розслідування шахрайства. Maltego пропонує можливість підключення даних і функцій із різних джерел за допомогою Transforms — фрагменти коду, які беруть біт інформації (у формі Entity) як вхідні дані, а потім повертають пов’язану інформацію у вигляді додаткових сутностей як вихід, або це механізм, який дає змогу досліджувати посилання. Також в ньому присутній ряд трансформацій, особливо корисних для фахівців з кібербезпеки, які прагнуть виявити приховані загрози в мережі організації та простежити походження цих загроз. MISP — платформа розвідки загроз з відкритим вихідним кодом для обміну, зберігання та співвіднесення показників компрометації цільових атак, розвідки загроз та інформації про фінансове шахрайство, вразливості чи навіть боротьбу з тероризмом. Virus Total Public API — служба, яка аналізує файли та URL-адреси на наявність вірусів, хробаків, троянів та іншого шкідливого вмісту. ZETAlytics Massive Passive надає дані, включаючи мільярди записів про історичні домени, адреси електронної пошти, IP-адреси та сервери імен.

Вступ

Зловмисне програмне забезпечення є одним з найпоширеніших кіберзагроз. Пандемія показала, що наше життя можливо перенести повністю в он-лайн, саме тому питання кібербезпеки, невід’ємним елементом якої є аналіз кіберінцидентів, постає особливо гостро.

Кіберзлочинці зазвичай намагаються зламати комп’ютер або мережеві системи, встановлюючи шкідливе ПЗ на цільовий комп’ютер через фішинговий домен або вкладення в email. При цьому, мета може варіюватися від паралізування внутрішніх систем до викрадення облікових даних користувачів, конфіденційної інформації і так далі.

Для зручнішого оперування даними пов’язаними з кіберзагрозою, використовується візуалізація. В цій роботі буде розглянута візуалізація даних після кіберінциденту, зокрема на прикладі софту Maltego — інструмент для побудови та аналізу зв'язків між різними суб'єктами та об'єктами. Він дозволяє зібрати воєдино інформацію, отриману з відкритих і закритих джерел.

Огляд Maltego

Maltego — це інструмент для аналізу графічних зв'язків. Його особливостями є візуалізація отриманих даних, розвідка на основі відкритих джерел, комбінування для глибокого аналізу даних, отриманих із закритих та відкритих джерел, автоматичний аналіз відкритих джерел та автоматична побудова взаємозв'язків між виявленими об'єктами. Даний софт:

1) Надає потужний пошук, що дає кращі результати. Це дозволяє заощадити час і працювати точніше.

2) Допомагає у процесі мислення, візуально демонструючи взаємопов’язані зв’язки між інформацію.

3) Зображує результати в широкому діапазоні графічних макетів, які дозволяють групувати інформацію, що робить перегляд відносин миттєвим і точним – це дає можливість знаходити приховані зв’язки, навіть якщо вони знаходяться на відстані трьох або чотирьох ступенів поділу.

Maltego пропонує можливість підключення даних і функцій із різних джерел за допомогою Transforms.

Transforms

Transforms або трансформації — це фрагменти коду, які беруть біт інформації (у формі Entity) як вхідні дані, а потім повертають пов’язану інформацію у вигляді додаткових сутностей як вихід, або це механізм, який дає змогу досліджувати посилання. Вони також можуть бути написані користувачами Maltego, забезпечуючи гнучкість підключення до власних даних.

Використання

Maltego використовується для збору інформації (загальнодоступної в Інтернеті), пов’язаної з безпекою. Одним із найпоширеніших варіантів використання є дослідження та отримання інформації про сайти. Дана програма використовується аудиторією від професіоналів у сфері безпеки до судових слідчих, журналістів-розслідувачів та дослідників ринку, завдяки

діапазону можливих випадків використання, починаючи від розвідки загроз до розслідування шахрайства.

Для використання софту потрібно спочатку завантажити Maltego з офційного сайту — https://www.paterva.com/. Далі створити на цьому сайті акаунт. У самій програмі обрати версію, після чого увійти в акаунт. Провести налаштування.

Можна переходити безпосередньо до використання Maltego. Для цього в лівому верхньому куті тиснемо кнопку New або можна просто натиснути Ctrl+T.

Відкриється новий робочий простір, в якому надалі працюватимемо.

У верхній частині екрана знаходиться панель керування. Тут можна змінювати налаштування, спосіб відображення інформації, зберігати або завантажувати проекти, керувати трансформаціями та зовнішнім виглядом. Ліворуч знаходиться палітра об'єктів. Звідти ми будемо брати різні об'єкти, яким надаватимемо певні значення, тобто ту інформацію, яку ми знаємо. Саме до цих об'єктів ми застосовуватимемо трансформації для отримання потрібної нам інформації. У центрі знаходиться вікно графіків — основна робоча область. Тут буде проводитися основна маса маніпуляцій і тут відображатиметься результат проведених трансформацій та відображатимуться зв'язки між різними об'єктами. Під ним знаходиться консоль виводу, в якій буде відображатися лог виконання кожної операції та інформація про помилки, якщо такі виникатимуть. Праворуч розташовано три вікна:

1. Overview – це мініатюра вікна графіків. Дозволять бачити загальну картину, а якщо графік досить великий, то швидко ним переміщатися. Там є вкладка Machines — в ній відображається хід виконання трансформацій.
2. Detail View — відображається докладна інформація про поточний (виділений) об'єкт.
3. Property View — це атрибути поточного (виділеного) об'єкта, тут їх можна редагувати.

Щоб застосувати будь-яку трансформацію, потрібно натиснути правою кнопкою миші на мету на графіку, відкриється меню трансформацій, що розподілені за розділами.

Визначення загроз ПЗ

Maltego має ряд трансформацій, особливо корисних для фахівців з кібербезпеки, які прагнуть виявити приховані загрози в мережі організації та простежити походження цих загроз.

Зокрема, це:

4. ATT&CK – MISP
5. Virus Total Public API
6. ZETAlytics Massive Passive

MISP — платформа розвідки загроз з відкритим вихідним кодом для обміну, зберігання та співвіднесення показників компрометації цільових

атак, розвідки загроз та інформації про фінансове шахрайство, вразливості чи навіть боротьбу з тероризмом. Використовуючи MISP Transforms, дослідники можуть запитувати дані з екземпляра обміну загрозами MISP і переглядати інші події, атрибути, об'єкти, теги та галактики MISP.

Virus Total Public API — служба, яка аналізує файли та URL-адреси на наявність вірусів, хробаків, троянів та іншого шкідливого вмісту. Запитуючи дані VirusTotal Public API за допомогою Maltego Transforms, дослідники можуть повернути інформацію про пов'язування IP-адрес, хешів, доменів та URL-адрес.

ZETAlytics Massive Passive надає дані, включаючи мільярди записів про історичні домени, адреси електронної пошти, IP-адреси та сервери імен. Використовуючи ZETAlytics Transforms, мисливці за загрозами можуть ідентифікувати зв'язки між історичними IP-адресами та іменами хостів, з'єднання хешів шкідливих програм з доменами, зв'язки між серверами імен і доменами тощо.

Висновки

В цій роботі було розглянуто програмне забезпечення Maltego, що виявилося потужним та в той же час простим засобом у кібербезпеці. Діапазон його використання надзвичайно широкий: від професіоналів у сфері безпеки до судових слідчих, журналістів-розслідувачів та дослідників ринку, завдяки своїй універсальності, доступності та багатофункціональності.

Робота в Maltego побудована на основі використання трансформацій, що пропонують підключення даних і функцій із різних джерел. Було розглянуто деякі трансформації, які застосовуються фахівцями з кібербезпеки: ATT&CK – MISP, Virus Total Public API, ZETAlytics Massive Passive.

Візуалізація у кібербезпеці допомагає організувати великі набори даних для легшого розпізнавання закономірностей, контексту та інформації, це може заощадити значну кількість часу на аналіз даних або відсутність ключових точок даних. Візуалізацію можна використовувати як дослідний інструмент для пошуку та зниження ризиків.

Ресурси

1. https://www.maltego.com/
2. https://hacker-basement.ru/2020/04/21/instrukcia-po-maltego-sbor-informacii/

ANALYSIS OF SOFTWARE FOR PROJECT MANAGEMENT

Lyudmyla Kaminskaya
Zhytomyr Polytechnic State University, Zhytomyr, Ukraine
kaminskayalyuda@gmail.com

In today's world, the success of IT companies depends on project management software. Realizing this, companies are investing in software. Highly qualified specialists are involved in system administration. Experience in working with a project management system and deep knowledge of their capabilities is the advantage of the candidate during hired to work.

Project management software is comprehensive software for task management, work and resource planning, reporting, price and budget management, documentation and administration, collaboration and communication.

The leaders of the project management software market are such solutions as: Jira, Wrike, Celoxis, Procore.

Each of these software has features, benefits, and nuances.

Let's consider them.

Jira - developed by Atlassian, is used by development teams that use an Agile approach in their work. Jira is great for both Scrum sprint project management and Kanban project management methodology.

Applying this software will allow you to organize effective work planning, track the results and progress of projects on a non-stop basis, organize the release of software versions and support.

In addition, Jira allows you to schedule sprints, track software issues and bugs, and generate reports that help improve teamwork. You can follow the default business process or create your own, as the system is flexible. It is possible to integrate with a large number of third-party solutions. It is designed for each member of the software development team and is able to meet the needs of each. Whether it is logging the tester's working time, or generating reports at the program level for the needs of the CEO.

Wrike is software used by more than 2.3 million users worldwide. Since 2006, it has received a significant number of awards. Wrike was first presented at Le Web 3 in Paris, where he won the B2B nomination.

This software is designed for workflow management, project planning and tracking, effective teamwork and real-time collaboration. It allows you to create automated reports, use custom query forms, and personalized dashboards. Process automation can increase efficiency by up to 50%, and 360-degree visibility increases trust through time and budget control. Deep analysis capabilities allow you to comfortably manage project portfolios. And a reliable level of security allows you not to worry about the leakage of confidential information.

The system can be integrated with more than 400 applications from Microsoft, Google and Salesforce.

Celoxis is a software development company of the same name Celoxis that was founded in 2001 and located in India. Celoxis combines project management, finance, integration, resources, technical issues, documents, schedules, project plans, costs and staff to create a single integrated project management program. Celoxis is available as a SaaS and as a local version.

Celoxis runs on Android or iOS mobile devices.

This comprehensive platform that allows you to manage projects within different organizations has recently updated its functionality, and added the ability to report manager and multi-level approval of work schedules.

The system has flexible installation capabilities and allows project management to more than 2,800 users worldwide.

Procore - this project management application schedules, quickly closes requests, tracks project emails, archives documents and photos, manages documents, manages daily work logs, project change requests, cost estimates and agreed work lists, integrates with MS Project and Sage Timberline Office. Launched on the market by the American company Procore Technologies in 2002.

Such project management services will continue to evolve in the future, increase functionality and cover more and more areas of influence. Their application is appropriate both in the first stages of the project, such as the process of selling the idea, concluding a contract, agreeing on budgets and deadlines, and in the final stages of implementing the system, tracking problems in the work and project support.

List of sources and literature

1. Agile-манифест разработки программного обеспечения [Електронний ресурс] – Режим доступу до ресурсу: https://agilemanifesto.org/iso/ru/manifesto.html
2. Лучшее программное обеспечение для управления проектами на 2021 год [Електронний ресурс] – Режим доступу до ресурсу: https://habr.com/ru/company/otus/blog/577090/
3. Power the Modern, Agile Enterprise [Електронний ресурс] – Режим доступу до ресурсу: https://www.wrike.com/
4. Обзор Wrike [Електронний ресурс] – Режим доступу до ресурсу: https://www.onlineprojects.ru/tool/205/
5. The All-in-One Project Management Software [Електронний ресурс] – Режим доступу до ресурсу: https://www.celoxis.com/
6. Connect everyone on your project with one platform. [Електронний ресурс] – Режим доступу до ресурсу: https://www.procore.com/

ІМІТАЦІЙНА МОДЕЛЬ НЕЧІТКОЇ СИСТЕМИ ВИЯВЛЕННЯ КІБЕРАТАК

Ігор Субач[1,2 [0000-0002-9344-713X],] Віталій Фесьоха[2[0000-0001-6612-1970],] Артем Микитюк[1[0000-0002-8307-9978],]
Володимир Кубрак[1[0000-0001-8877-5289],] Станіслав Коротаєв[1[0000-0003-3823-8375].]
[1] Інститут спеціального зв'язку та захисту інформації Національного технічного Університету України "Київський політехнічний інститут імені Ігоря Сікорського"
[2] Військовий інститут телекомунікацій та інформатизації імені Героїв Крут
igor_subach@ukr.net

Розглянуто методику застосування імітаційної моделі нечіткої системи виявлення кібератак. Наведено функціональну схему імітаційної моделі. Розглянуто структурну схеми імітаційної моделі та описано призначення її елементів. Описано основні кроки застосування імітаційної моделі для проведення експериментального дослідження щодо оцінки ефективності моделей та методів виявлення кібератак, що ґрунтуються на теорії нечітких множин та нечіткого логічного виводу. Наведено порядок формування початкових даних, визначено класи кібератак, що підлягають виявленню, виділено вектори ознак кібератак, наведено опис параметрів досліджуваного трафіку, визначено типи функцій належності для формалізації експертних знань і представлення їх у базі знань у вигляді нечітких продукційних правил. Розглянуто питання параметричної адаптації функцій належності для уточнення суб'єктивних суджень експертів. Для реалізації можливості виявлення поліморфних кібератак, описано порядок визначення необхідної кількості найбільш важливих ознак для кожного відомого класу кібератак, представлених нечіткими множинами та лінгвістичними змінними, які достатньо повно їх характеризують. Проведено порівняльний аналіз результатів моделювання процесу виявлення кібератак на основі запропонованого підходу з існуючими методами виявлення кібератак: на основі теорії нечітких множин та нечіткої логіки, штучних імунних систем та нейронних мереж за показником точності.

Вступ

У [1, 2] було запропоновано архітектуру нечіткої інтелектуальної системи виявлення вторгнень, а у [3, 4, 5] – моделі та методи виявлення вторгнень відомих та поліморфних кібератак, що ґрунтуються на нечітких правилах. Для оцінки ефективності запропонованих рішень була розроблена

імітаційна модель системи (FIDM – Fuzzy Intrusion Detection Model) з використанням пакету Fuzzy Logic Toolbox™, який забезпечує функції MATLAB® та блоку Simulink® [6, 7, 8] для проектування та моделювання систем на основі нечіткої логіки.

Функціональна схема імітаційної моделі

Функціональна схема імітаційної моделі представлена на рисунку 1, де $x_1..x_n$ – вхідні параметри, а y – вихідна змінна.

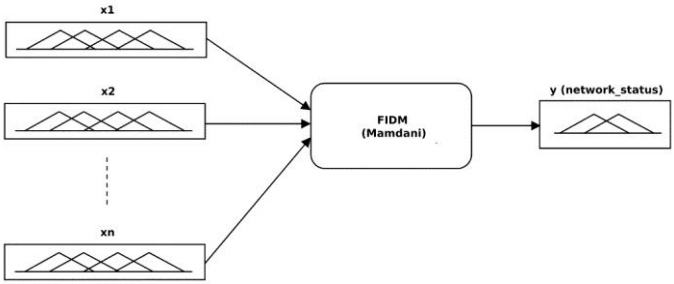


Рис. 1. Функціональна схема імітаційної моделі FIDM

Основу запропонованої імітаційної моделі становить модульна схема організації послідовної ітераційної взаємодії між її компонентами: модулем введення даних для аналізу, модулем фазифікації, базою знань, модулем нечіткого логічного виводу та дефазифікації [6, 7, 8]. Структурну схему розробленої імітаційної моделі представлено на рисунку 2.

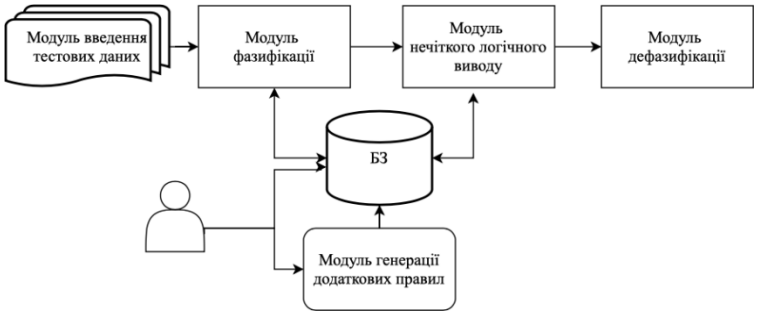


Рис. 2. Структурна схема розробленої імітаційної моделі

Для роботи модулю введення тестових даних [9] застосовувався набір статистичних даних про кібератаки KDD Cup 1999 Data (xlsx-файли).

Призначенням модулю фазифікації є представлення кількісних та якісних значень досліджуваних параметрів за допомогою терм-множин та лінгвістичних змінних. Урахування неповних та невизначених даних про

мережеву активність здійснювалося шляхом застосування розробленої моделі виявлення кібератак.

Модуль генерації додаткових правил застосовувався для створення нових нечітких продукційних правил у базі знань шляхом перетину нечітких множин лінгвістичних змінних раніше існуючих правил та попередньо визначених експертом найбільш значущих лінгвістичних змінних для кожного класу кібератак.

База знань представляє собою набір нечітких продукційних правил, які побудовані експертом.

Призначенням модулю нечіткого логічного виводу є генерування рішення про стан інформаційно-телекомунікаційної мережі на основі визначення залежності між вхідними даними та експертними висновками засобами нечіткої логіки.

Модуль дефазифікації застосовувався для перетворення отриманих значень нечіткого логічного виводу у чіткі.

Методика застосування імітаційної моделі

Методика застосування імітаційної моделі полягала у наступному.

Крок 1. Визначення набору статистичних даних про кібератаки: KDD Cup 1999 Data [9].

Крок 2. Визначенні класів кібератак, що підлягають подальшому виявленню [9]: Denial of Service, Remote to Local, User to Root, Probe та нормальні стани інформаційної системи.

Крок 3. На вхід імітаційної моделі подавалися вектори ознак кібератак набору KDD Cup 1999 Data у кількості: відомих – Denial of Service – 4264, Remote to Local – 1020, User to Root – 52, Probe – 3231; нормальних станів інформаційної системи – 1000; поліморфних кібератак, побудованих на основі відомих – 100. Таким чином, загальна кількість векторів ознак склала – 9667.

Крок 5. Визначення типу функцій належності для опису діапазонів значень досліджуваних параметрів та потужності терм-множин для вхідних та вихідної лінгвістичних змінних: трикутні функції належності, завдяки властивості піддаватися параметричній адаптації (уточненню) зі збереженням прийнятного рівня обчислювальної складності, потужність терм-множини – 7 (кількість: “ДМ” – дуже мала, “М” – мала, “НС” – нижче середньої, “С” – середня, “ВС” – вище середньої, “В” – велика, “ДВ” – дуже велика).

Крок 6. Визначення вхідних та вихідних лінгвістичних змінних: 38 вхідних лінгвістичних змінних, які відповідають кількості досліджуваних параметрів мережевого трафіку та одна вихідна – показник стану інформаційно-телекомунікаційної системи. Кожному вхідному значенню (x_i) відповідає параметр мережевого трафіку відповідно до KDD Cup 1999 Data, а налаштовані функції належності експертом (автором) представлені терм-множинами.

Крок 7. Підготовка формату тестових даних KDD Cup 1999 Data для програмного забезпечення Fuzzy Logic Toolbox™.

Крок 9. Створення нечітких продукційних правил для БЗ на основі набору даних KDD Cup 1999 Data із використанням алгоритмів пошуку асоціативних правил.

Крок 10. Отримання експертних висновків щодо стану ІС відповідно до класифікації кібератак, яка представлена у KDD Cup 1999 Data: Denial of Service, Remote to Local, User to Root, Probe та нормальний стан: загальна кількість правил: 335, з них Denial of Service – 68; Remote to Local – 55; User to Root – 18; Probe – 107; Normal – 87.

Крок 11. Параметрична адаптація побудованих функцій належності [10, 11] з метою уточнення суб'єктивної точки зору експерта засобами пакету Optimization Tool програмного забезпечення MATLAB® [6, 7, 8]: на основі знайдених частих наборів даних на попередньому етапі для вищевказаних термів кожної досліджуваної змінної здійснювалась параметрична оптимізація функцій належності (пошук оптимуму вектору параметрів системи рівнянь аналітичної моделі функції належності трикутного типу).

Крок 12. Визначення необхідної кількості найбільш важливих (інформативних) параметрів (ознак) для кожного відомого класу кібератак, представлених у вигляді нечітких множин лінгвістичних змінних, які достатньо повно їх характеризують з метою отримання можливості виявлення поліморфних модифікацій відомих кібератак.

Крок 13. Отримання додаткових нечітких продукційних правил для БЗ на основі проведених дій на попередньому кроці та видаленні правил, які дублюються. У результаті чого було отримано нові правила у наступній кількості: Denial of Service – 48; Remote to Local – 14; User to Root – 13; Probe – 16. Загальна кількість правил у БЗ – 426.

Крок 14. Застосування розробленого програмного коду для коректного введення даних з набору даних про кібератаки з метою здійснення подальшого їх аналізу бібліотекою Fuzzy Logic Toolbox™.

Крок 15. Проведення експериментальних досліджень виявлення кібератак розробленою імітаційною моделлю, функціонування якої ґрунтується на застосуванні моделей та методів та порівняльний аналіз результатів моделювання процесу виявлення кібератак на основі запропонованого підходу з існуючими методами виявлення кібератак: на основі теорії нечітких множин та нечіткої логіки, штучних імунних систем та нейронних мереж за показником точності (Accuracy).

Висновки

Застосування на практиці розробленої імітаційної моделі нечіткої системи виявлення кібератак, показало доцільність її використання для оцінки моделей та методів виявлення кібератак, що ґрунтуються на теорії нечітких множин та нечіткого логічного виводу. Так, оцінка розробленого науково-методичного апарату у порівнянні з відомими, показала, що його застосування дозволяє підвищити ефективність кіберзахисту ІС за

показником точності виявлення відомих кібератак в середньому на 10%, а також забезпечити виявлення поліморфних кібератак за показником точності не менше ніж 80%.

Література

1. І. Субач, В. Кубрак, та А. Микитюк, "Архітектура та функціональна модель перспективної проактивної інтелектуальної системи SIEM-системи для кберзахисту об'єктів критичної інфраструктури", Information Technology and Security, Vol 7., Iss. 2., 2019, pp. 208–215, DOI: https://doi.org/10.20535/2411-1031.2019.7.2.190570.

2. І. Субач, В. Фесьоха, та Н. Фесьоха. "Аналіз існуючих рішень запобігання вторгненням в інформаційно-телекомунікаційні мережі, відкритих на основі загальнодоступних ліцензій". Information Technology and Security, Vol 5, Iss 1, 2017 pp. pp. 29–41, DOI: https://doi.org/10.20535/2411-1031.2017.5.1.120554.

3. V. Fesokha, I. Subach, V. Kubrak, A. Mykytiuk, and S. Korotaiev. "Zero-day polymorphic cyberattacks detection using fuzzy inference system" Austrian Journal of Technical and Natural Sciences, № 5-6, 2020, pp. 8-13, DOI: https://doi.org/10.29013/AJT-20-5.6-8-13.

4. І. Субач, Ю. Здоренко, та В. Фесьоха, "Методика виявлення кібератак типу JS(HTML)/Scrinject на основі застосування математичного апарату теорії нечітких множин", Збірник наукових праць Військового інституту телекомунікацій та інформатизації імені Героїв Крут, № 4, 2018, С. 125–131.

5. І. Субач, та В. Фесьоха, "Модель виявлення кібернетичних атак на інформаційно-телекомунікаційні системи на основі описання аномалій їх роботи зваженими нечіткими правилами", Information Technology and Security. 2017, Vol, 5, Iss 2., pp. 145–152, DOI: https://doi.org/10.20535/2411-1031.2017.5.2.136984.

6. Документація Matlab. Експонента: веб-сайт. URL: https://docs.exponenta.ru/matlab/index.html (дата звернення: 22.05.2020).

7. Sivanandam S.N., Sumathi S., Deepa S.N. Introduction to Fuzzy Logic using MATLAB. Springer-Verlag Berlin Heidelberg, 2007. 430 p.

Леоненков А.В. Нечеткое моделирование в среде MATLAB и fuzzyTECH. Санкт-Петербург: БХВ-Петербург, 2003, 736 с.

8. UCI Knowledge Discovery in Databases Archive. University of California, Irvine, CA 92697–3425. KDD Archive : website. URL: http://kdd.ics.uci.edu/databases/kddcup99/task.html (дата звернення: 22.05.2020).

9. A.Rothstein. Intelligent identification technologies: fuzzy sets, genetic algorithms, neural networks. – Vinnytsia: UNIVERSUM, 1999. – 320 p.

10 Zaichenko Y.P. Operations Research: Fuzzy Optimization. – Kiev: High school, 1991. – 191 p.

A UNIVERSAL MODEL OF DECISION-MAKING SUPPORT TOOLS IN AUTONOMOUS CYBERPHYSICAL SYSTEMS

Bohdan Zimchenko
Lviv Polytechnic National University, Lviv, Ukraine
Bohdan.V.Zimchenko@lpnu.ua

This article discusses the issues of decision-making in autonomous cyberphysical systems which is relevant today and considers the tools that can be implemented in the architecture of cyberphysical systems to achieve their bigger autonomy. It also discusses the key elements in the structure of any cyberphysical system, the relationship between them and their functional features, and reviews existing tools that can help the decision-making process in cyberphysical systems that have different architectures and solve problems of different nature and complexity. A universal model of decision-making support tools in autonomous cyberphysical systems is proposed. This model can be implemented in most existing cyberphysical systems, as well as implemented at the stage of planning the architecture and direct development of the cyberphysical system. The review of the model is carried out within the architecture of the cyberphysical system and considers the functional features of the nodes of the cyberphysical system and the decision support system that is a part of it. The main technologies that can be used in the construction of an autonomous cyberphysical system, the architecture of which includes a decision support system, based on the proposed universal model, were also considered. The article gives a generalized idea of how a cyber-physical system can look from an architectural point of view, containing a decision support system, which is illustrated by the corresponding structural diagrams. The role of the human expert in the process of design and development of an autonomous cyberphysical system and the methods of interaction of the human expert with the decision support system are reviewed.

Keywords: cyberphysical systems, decision support systems, expert systems, artificial intelligence, human expert

Introduction

Autonomous CPS are systems that are able to make decisions and perform actions independently, without human intervention. In such systems, the problem of decision-making is very acute, and to achieve their autonomy, different technological approaches are used, which are based on the architecture of the CPS and the tasks it should solve.

The concept of decision-making support emerged mainly through theoretical studies of organizational decision-making in the late 1950s and was first implemented in the 1960s [1]. Since then, decisions support systems(DSS) have

become widely used in different areas of human activity. In order to achieve autonomy in the work of the CPS, it is becoming more common to use such DSSs that contain an expert system(ES). This approach helps the CPS to use the analytical skills of one or more experts in a particular field of application, to draw logical conclusions, thus ensuring the solution of specific problems. Also, given the large amount of data coming from the physical environment to the CPS, and data, generated directly in CPS, during its usage, artificial intelligence systems are often used as a part of the DSS for CPS.

Expert system as a decision-making tool

An ES is a computer system that can emulate a human expert behavior in the case of decision-making. It is designed to solve problems with the help of knowledge, which is represented mainly as IF-THEN rules rather than through conventional procedural code. Knowledge is a theoretical or practical understanding of a subject or domain and is possessed by experts. Expert is someone, who has deep knowledge and strong practical experience in a particular domain. Knowledge can be represented by rules, frames, or semantic networks.

Rules are the most common way to represent knowledge. They are defined as an IF-THEN structure, where the “IF” (antecedent) part contains facts and the “THEN” part(consequent) contains some actions, that can be released. One rule can be described with the usage of multiple antecedents, joined by the keywords: AND(conjunction), OR(disjunction).

A frame is a data structure with typical knowledge about a particular object or concept [2]. In a frame-based representation of knowledge, each frame has its own name and a set of attributes associated with it, which, in turn, have an attached value. Frames are an application of object-oriented programming for ES.

Semantic networks are used for propositional information and are also called propositional networks. They consist of nodes, links, and link labels and can be defined as a labeled directed graph in math.

The following specialists are usually involved in the creation of the ES: the domain expert, the knowledge engineer, the programmer, the project manager. Success in creating an ES directly depends on the coordination of actions of all specialists.

Artificial intelligence as a decision-making tool

Decision-making is a complex process that ordinary computer programs usually perform structurally, which limits the decision-making process. In addition, the machine itself is not able to think and formulate rules based on its own experience. Not surprisingly, researchers have sought to improve the quality of solutions by developing computer technology to increase and empower people.

Artificial Intelligence (AI) can be defined as a space for the design and development of systems that can provide effective solutions to real problems and imitate the way of thinking, the behavior of humans and other living organisms in nature [3-4]. Advances in Artificial Intelligence (AI) have made the goal of a

machine's ability to think by itself, a reality in many applications. CPSs, integrated with artificial intelligence, or with intelligent decision support systems (IDSS), are increasingly used to help make decisions in various areas of human activity. The term intelligent is used to describe systems that mimic human cognitive capabilities in some way. These systems use AI tools to reason, learn, remember, plan and analyze. AI tools can be used to empower people, for example, by researching and selecting relevant information from extremely large and distributed data sources, applying analytical tools to unstructured data, creating generalized solutions from sets of rules and probabilities, and finding associations in information from multiple sources that may affect decisions. Tools such as Artificial Neural Networks, Fuzzy Logic, Intelligent Agents, Agent Teams, Case-Based Reasoning, Evolutionary Computing, and probabilistic reasoning, when combined with other decision support systems, can help a decision-maker evaluate and choose alternatives.

A universal model of decision-making support tools in autonomous cyberphysical systems

Almost any CPS can be divided into physical and cybernetic parts. The physical part consists of sensory and actuation systems, and the cybernetic part consists of various computing devices combined with a communication environment. The basic component of the CPS is the computing and measuring node, which contains the sensor, executive, and computing systems. The number of such nodes in the CPS can be unlimited. Each individual node can be both independent in decision-making and the formation of the node's response to input data from the sensor system, and dependent on the core of the CPS - the center for collecting and processing information.

Communication and exchange of information within the cyberphysical system between its structural parts occurs through the communication environment. The basic communication environment for information transfer includes wires as well as optical and wireless media. The connection between electronic devices is usually performed through wires, but it has a number of disadvantages, namely the high cost and extensive efforts to deploy and maintain them. In the case of cyberphysical systems, it is reinforced by their very nature, as they cover large areas including hard-to-reach places, and often work in harsh environments. The use of digital systems almost everywhere led to the development of a wide range of protocols communication using wires. Ethernet technology as part TCP / IP (Transmission Control Protocol / Internet Protocol) and the UDP (User Datagram Protocol) stack is the core standard used in home and office applications, while Fieldbus protocols are found in industrial environments. Some communication protocols, implemented between digital controllers and software in industrial environments include: OPC UA, PROFIBUS, CAN, CANopen, PROFINET, INTERBUS, Foundation Fieldbus, HART, Modbus, and others [5]. But anyway, the current trend is the widespread usage of wireless technologies in monitoring

and management applications, which is motivated by the undeniable advantages they have. Lower prices due to cable replacement, simplified network topology, scalability, and low maintenance costs are just some of them. However, you have to pay for it - wireless communication poses a wide range of problems and tasks to engineers. In most cases, the accompanying sensors and actuators in such applications are systems that are limited in terms of energy consumption, communications, and energy resources, and issues such as data transmission, reliability, real-time data delivery, and energy efficiency should be considered from the design stage to the continuing the development and finishing with the implementation phase.

Communication between the computing and measuring node with other nodes and the kernel occurs through the communication environment. It is proposed to expand the structural model of such a node, including in its composition a decision-making subsystem, and to expand the structural model of the CPS, including a DSS in its composition as is shown in fig.1.

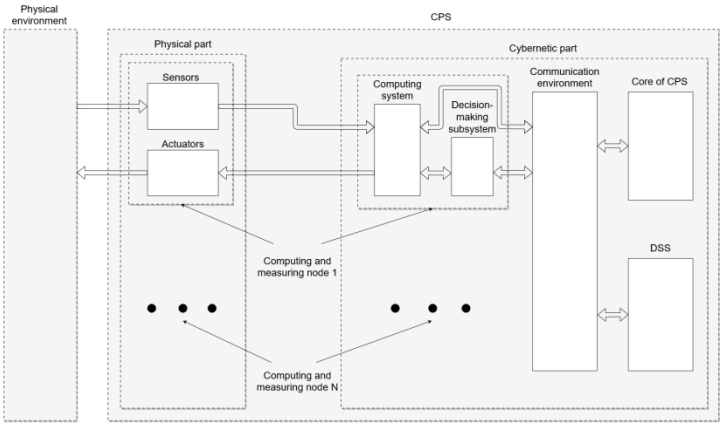


Fig.1. Structural model of CPS containing DSS

The computing system collects and pre-processes the information obtained using the sensor system, transmits information in processed or unaltered form to the communication environment, as well as to the decision-making subsystem. It also receives information from the communication environment and decision-making subsystem and is able to activate the actuation system, send it executive commands. Such a computing system can be represented as a microcontroller or a microprocessor system. When designing such a system, it is necessary to pay attention to the needs of the cyberphysical system itself, and take into account the complexity and type of calculations that it must perform.

The decision-making subsystem is able to extract from a set of data (transferred from a computing system), such data that may influence a process of decision-making by node, group it according to certain characteristics, and transmit to the communication environment and also receive a response from the DSS and send it back to the computing system. Such a subsystem can be a tool that helps to make decisions together with the DSS, performing the functions of communication of the computing system with the DSS, standardization, and verification of data that can participate in the decision-making process. Such a subsystem can be designed as a software solution and can also include additional hardware if needed.

It is proposed the CPS include DSS, which has: information collection and storage node, active and cooperative DSSs, ES, user interface, and the DSS core. The structural model of the proposed DSS is presented in fig.2.

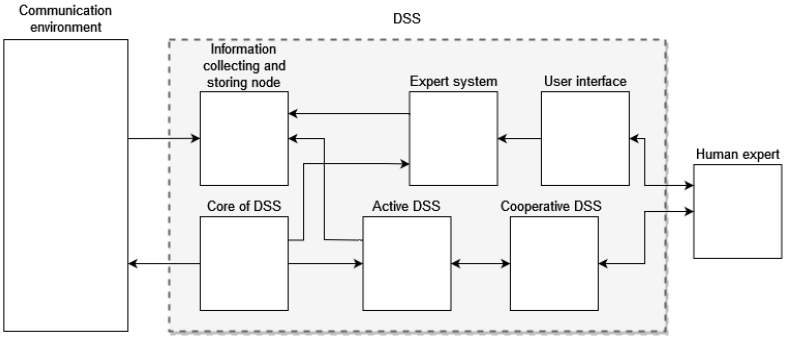


Fig.2. Structural model of DSS in CPS

The information collection and storage node performs the function of a buffer by receiving information through the communication environment from the subsystems of decision-making and storing it for further transmission to other nodes of the DSS.

Active DSS is a system that assists the decision-making process, receives data from the information collection and storage node, and is able to submit clearly formulated decisions or proposals. Active DSS includes AI tools that are capable of self-learning, such as neural networks that work with large amounts of data and make decisions or assumptions based on experience. Large and efficient neural networks require significant computing resources. In addition, the developer often needs to transmit signals through many of these connections and associated neurons, which require enormous CPU power and time. Using accelerators, such as FPGAs and GPUs can reduce training time from months to days.

Cooperative DSS can be useful at the stage of teaching neural networks that are part of active DSS, by a human expert. Also, a human expert can use the user

interface to communicate with the ES, providing the system with the necessary knowledge that can help the decision-making process. The user interface is a method of interaction of the expert system with the user. It can be performed through a dialog box, command line, form, or other input methods. Some expert systems interact with other computer programs and do not interact directly with humans.

The main element for the decision-making process in the DSS is its core. It receives data from active DSS and ES and if this data does not make a conflict, the assumption or decision made by active DSS is not considered incorrect by ES or coincides with the assumption or decision proposed by the ES, the DSS core determines it as correct and transmits it to the communication environment.

Conclusions

The issue of autonomy in cyberphysical systems today remains relevant and scientists face many problems with decision-making in cyberphysical systems. The proposed universal model of decision-making support tools can be considered as a generalized platform for creating and designing autonomous cyberphysical systems of any complexity, but it should be noted, that it may be necessary to be expanded and changed, depending on the needs and characteristics of the cyberphysical system that is being designed.

References

1. Keen, P. G. W., & Scott, M. M. S. Decision support systems: An organizational perspective. Reading, Mass: Addison-Wesley Pub. Co. (1978).
2. *M.Minsky, A Framework for Representing Knowledge, MIT-AI Laboratory Memo 306, DOI: doi.org/10.1016/B978-1-4832-1446-7.50018-2 (1974).*
3. L. Rabelo, S. Bhide, E. Gutierrez, Artificial Intelligence: Advances in Research and Applications, Nova Science Publishers, Inc. (2018).
4. J. Romportl, E. Zackova, J. Kelemen, Beyond Artificial Intelligence, Springer International (2016).
5. G. Mois, S. Folea, T.Sanislav, L. Miclea, Communication in Cyber-Physical Systems, Conference: 19th International Conference on System Theory, Control and Computing (ICSTCC) (2015).

ОГЛЯД ПРОГРАМНИХ ЗАСОБІВ ДЛЯ РОЗРОБКИ ЛОКАЛЬНИХ МЕРЕЖ

Андрій Савчук, Андрій Єфіменко, Тетяна Вакалюк
Державний університет «Житомирська політехніка»

При розробці будь-якої мережі необхідно вибрати правильне обладнання та програмне забезпечення, адже від кожного вибору буде залежати працездатність всієї мережі, її надійність та захист.

Cisco – це одна із небагатьох компаній, яка є лідером серед розробки мережевих приладів [2]. Дана компанія розробляє прилади для всього світу і відповідає за якість своєї продукції; прилади компанії розробляються так, щоб вони мали змогу якнайдовше працювати. Це саме те, що потрібно для великих підприємств та компаній; налаштування мережевих приладів від компанії Cisco максимально просте і зрозуміле. Прилади підтримують можливість графічного інтерфейсу користувача. Також можна керувати мережею і одночасно налаштувати прилади; комутатори Cisco мають вбудовану систему безпеки, за допомогою якої вони бачать, які прилади підключаються до мережі задля запобігання атак. Компанія має велику кількість обладнання, яке підходить як для великих підприємств та компаній, так і для малих, що дозволяє підібрати такі прилади, що будуть оптимальні, як і для роботи, так і для бюджету цих самих компаній та підприємств [2].

GNS3 (Graphical Network Simulator) – це безкоштовне програмне забезпечення, яке дає можливість створювати топології від маленьких, які складаються з декількох пристроїв, так і до великих, що складаються з багатьох пристроїв, віртуальних машин тощо [3]. Для цього не потрібне реальне обладнання, адже GNS використовує емулятори того самого обладнання, що дає змогу максимально наблизитися до реального. Основна перевага GNS заключається в тому, що він дозволяє використовувати віртуальні машини операційних систем. Тобто це дозволяє створювати майже реальну мережу певного підприємства чи компанії. Все ж таки будь-яке програмне забезпечення має свої недоліки, і GNS не виключення. Для того, щоб зручно і безперешкодно працювати в GNS, потрібно мати достатньо потужний ПК, адже розробка мереж це ресурсозатратний процес. Тому не на всіх ПК є можливість працювати в даному симуляторі [3].

Virtual Box – це програмне забезпечення, яке дозволяє додавати, використовувати, видаляти та працювати в різних операційних системах, при цьому не змінюючи своєї операційної системи [4]. До основних переваг можна віднести:

- зручний та простий інтерфейс;
- тестування різних операційних систем, не змінюючи своєї основної;
- налаштування ресурсо-затратності різних операційних систем [4].

Wireshark – це програма для аналізу трафіку в мережі [5]. Вона допомагає відстежувати атаки на мережу та розробляти план захисту від

них. За допомогою даної програми можна фільтрувати трафік, слідкувати за ним в своїй мережі а також перехоплювати пакети даних в інших мережах до яких маєш доступ [5].

Kali Linux - дана операційна система розроблялась для того, щоб тестувати мережі та знаходити слабкі місця [1]. Тобто за допомогою даної системи можна проводити атаки на мережу. Kali Linux має такі переваги:

- зручний інтерфейс.
- безкоштовність;
- має багато утиліт, за допомогою яких можна тестувати мережі на захищеність [1].

Список використаних джерел та літератури

1. Kali Linux [Електронний ресурс] URL: https://www.kali.org/
2. 10 Reasons to Choose Cisco Switching [Електронний ресурс] URL: https://www.cisco.com/c/dam/m/en_sg/cisco-start/assets/pdfs/singapore_cust_aag_flyer_v1.pdf
3. GNS3 – Графический Сетевой Симулятор [Електронний ресурс] URL: http://www.ciscolab.ru/labs/40-gns3-graficheskiy-setevoy-simulyator.html
4. Oracle VM VirtualBox что это за программа и нужна ли она? [Електронний ресурс] URL: http://virtmachine.ru/oracle-vm-virtualbox-chto-eto-za-programma-i-nuzhna-li-ona.html
5. Что такое WireShark? Для чего нужна эта программа? [Електронний ресурс] URL: https://nordd.ru/chto-takoe-wireshark-dlya-chego-nuzhna-eta-programma/

АНАЛІЗ ТА ПОРІВНЯННЯ ПОПУЛЯРНИХ PHP FRAMEWORKS

Юрій Клименко
Житомирська політехніка

Перед PHP розробниками початкового рівня часто постає питання, який PHP фреймворк обрати. Основна мета фреймворку – це економія часу та ресурсів при розробці php-додатків. Фреймворк включає в себе набір базових функцій таорозширень, які за звичайної розробки потрібно писати з нуля. Завдяки володінню фреймворками розробка php-додатків займає менше часу, завдяки чому підвищується ефективність праці розробника.

На порівняння буде винесено 3 фреймворки, а саме: Laravel, CodeIgniter та Yii, які на даний момент користуються найбільшою популярністю у розробників php-додатків. Всі з представлених сьогодні фреймворків є безкоштовними та мають ряд відмінностей.

Laravel – є одним з самих відомих фреймворків, що має зручну структуру,в якій можна розібратись навіть не використовуючи його раніше. Основними особливостями Laravel є велика бібліотека сторонніх розширень та зручна інтеграція зі сторонніми сервісами, а також можливість суттєво заощаджувати час при розробці php-додатків.

Переваги Laravel:

- швидкий розвиток;
- MVC (Model View Controller);
- зручна debug консоль, котра працює одразу “з коробки”;
- достатня кількість документації різними мовами;
- керування доступом на основі ролей (RBAC);
- велика кількість розширень;
- швидка інтеграція зі сторонніми сервісами.

Недоліки Laravel:

- складність вивчення порівняно з деякими іншими фреймворками;
- нижча швидкість роботи в порівнянні з іншими фреймворками.

CodeIgniter – фреймворк, який серед початківців відомий менше за попередній, але він також користується популярністю серед розробників php- додатків. CodeIgniter використовує архітектуру MVC. Головними перевагами CodeIgniter для молодого розробника є простота та наявність великої кількості документації, що значно полегшує його вивчення. Але є і недолік – непостійність його оновлення.

Переваги CodeIgniter:

- MVC;
- наявність великої кількості документації;
- низький вхідний рівень;
- швидкість роботи.

Недоліки CodeIgniter:

- проблеми з безпекою;

- неактивна спільнота;
- мала кількість бібліотек.

У випадку розробки проектів з високим рівнем захисту, варто віддати перевагу Yii (Yes, it is) який представлений на ринку як швидкий, ефективний та захищений фреймворк. Головні його особливості включають швидкість налаштування та роботи, яка може зрівнятися з одним з найшвидшихфреймворків Phalcon. Крім того варто відмітити високий рівень захисту та велику кількість розширень, що значно пришвидшує розробку проекту.

Переваги Yii:

- досить висока швидкість роботи;
- наявність великої кількості розширень;
- велика кількість документації;
- високий рівень захисту;
- швидкість налаштування.

Таким чином, сказати, який фреймворк кращий неможливо. Все залежить від потреб користувача. Однозначно можливо сказати лише те, що використання PHP фреймворків дозволяє суттєво заощаджувати час як розробнику-початківцю, так і досвідченому розробнику.

FORMATION OF SHIFT INDEXES VECTORS OF RING CODES FOR INFORMATION TRANSMISSION SECURITY

Serhii Toliupa [1][0000-0002-1919-9174], Liubov Berkman[2][0000-1111-2222-3333], Serhiy Otrokh [1][0000-0003-3990-5205], Bohdan Zhurakovskyi [3][0000-0003-3990-5205], Valeriy Kuzminykh[3][0000-0002-8258-0816], and Hanna Dudarieva[2][0000-0002-9887-021X]

[1]Taras Shevchenko National University of Kyiv, Ukraine
[2]State University of Telecommunications, Kyiv, Ukraine
[3] National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine

tolupa@i.ua, berkmanlubov@gmail.com, 2411197@ukr.net, zhurakovskiybyu@tk.kpi.ua, vakuz0202@gmail.com, Annett.13.86@gmail.com

The paper considers a method of transformation a ring codes into a shift indexes vectors, which are intended to compress information and protect it unauthorized access. The structure of shift index vectors and regularities of change of decimal values of shift index vectors are analyzed. The properties of shift indexes vectors, created by conversion a ring codes using of binary transformations of the *XOR*, *AND*, *OR* elements of the initial sequence (first line) of the ring code and successively on each subsequent line, are investigated. It is established that the limits of change of decimal values of elements of the shift index vectors depend both on length and number of ones in code combinations of ring codes and on the ratio of ones and zeros in the code combination. Analysis of the structure of index vectors of shift indexes of ring codes shows that for a family of ring codes of a certain type there is a mutually-ambiguous dependence of decimal values of elements of shift index vectors and the limits of their location in the index vector on the number of elements and units in the code combination. The set of decimal values of the index vector consists of three sequences. The set of decimal values of the index vector consists of three sets. The example of the family of ring codes of type 000111 shows that the limits of each of the three sets are uniquely described by the dependence on the number of elements and units in the code combination. This method can be used to build an effective channel for the transmission of the future network.

Keywords: ring codes, shift indexes vector, decimal values, binary transformations

Introduction

The ring code is built on the principle of forming a cyclic code by shifting a certain number of elements to the right or left in the code combination and differs from the cyclic code in that the elements of the high bits are shifted to the beginning of the low bits, as if forming a ring. The ring code matrix is a square of size $N \times N$, each line containing m units and N - m zeros [1]. At the same time each line of the matrix repeats the previous line with a simultaneous ring shift of characters by a certain number of digits to the right or left. Based on the above, the ring code, which is formed by shifting one character of the code sequence from right to left, can be represented as the following matrix G :

$$G = \begin{bmatrix} k_{N-1}2^{N-1} + k_{N-2}2^{N-2} \dots + k_1 2^1 + k_0 2^0 \\ k_{N-2}2^{N-1} + k_{N-3}2^{N-2} \dots + k_0 2^1 + k_{N-1}2^0 \\ \cdot \\ \cdot \\ \cdot \\ k_0 2^{N-1} + k_{N-1}2^{N-2} \dots + k_2 2^1 + k_1 2^0 \end{bmatrix} \quad (1)$$

where k_i is a binary element of the code combination, which takes on the value 0 or 1 depending on its structure. According to [1] the ring code is characterized by a of shift indexes vector (SIV), which is formed as follows:

- the binary logical transformation *XOR, AND* or *OR* is performed alternately over the elements of the first row and subsequent rows of the ring code matrix placed at the same positions of the code sequences. In this case the matrix of the ring code is converted into a of shift indexes matrix;
- in each row of the resulting matrix of shift indicators the number of elements corresponding to one is counted. At that, the matrix of shift indices is transformed into a shift indexes vector.

Then the formula of transformation of matrix G into the shift indexes vector using the binary logical transformation *XOR* takes on the following form:

$$SIV_{XOR} = \begin{bmatrix} k_{N-1}2^{N-1} XOR\ k_{N-2}2^{N-1} + k_{N-2}2^{N-2} XOR\ k_{N-3}2^{N-2} + \cdots + k_0 2^0 XOR\ k_{N-1}2^0 \\ k_{N-1}2^{N-1} XOR\ k_{N-3}2^{N-1} + k_{N-2}2^{N-2} XOR\ k_{N-4}2^{N-2} + \cdots + k_0 2^0 XOR\ k_{N-2}2^0 \\ \cdot \\ \cdot \\ \cdot \\ k_{N-1}2^{N-1} XOR\ k_0 2^{N-1} + k_{N-2}2^{N-2} XOR\ k_{N-1}2^{N-2} + \cdots + k_0 2^0 XOR\ k_1 2^0 \end{bmatrix} \quad (2)$$

The formula of transformation of matrix G into the shift indexes vector using the binary logical transformation *AND* takes **on** the following form:

$$SIV_{AND} = \begin{bmatrix} k_{N-1}2^{N-1} AND\ k_{N-2}2^{N-1} + k_{N-2}2^{N-2} AND\ k_{N-3}2^{N-2} + \cdots + k_0 2^0 AND\ k_{N-1}2^0 \\ k_{N-1}2^{N-1} AND\ k_{N-3}2^{N-1} + k_{N-2}2^{N-2} AND\ k_{N-4}2^{N-2} + \cdots + k_0 2^0 AND\ k_{N-2}2^0 \\ \cdot \\ \cdot \\ \cdot \\ k_{N-1}2^{N-1} AND\ k_0 2^{N-1} + k_{N-2} AND\ k_{N-1}2^{N-2} + \cdots + k_0 2^0 AND\ k_1 2^0 \end{bmatrix} \quad (3)$$

The formula of transformation of matrix *G* into the shift indexes vector using the binary logical transformation *OR* takes on the following form:

$$SIV_{OR} = \begin{bmatrix} k_{N-1}2^{N-1} OR\ k_{N-2}2^{N-1} + k_{N-2}2^{N-2} OR\ k_{N-3}2^{N-2} + \cdots + k_0 2^0 OR\ k_{N-1}2^0 \\ k_{N-1}2^{N-1} OR\ k_{N-3}2^{N-1} + k_{N-2}2^{N-2} OR\ k_{N-4}2^{N-2} + \cdots + k_0 2^0 OR\ k_{N-2}2^0 \\ \cdot \\ \cdot \\ \cdot \\ k_{N-1}2^{N-1} OR\ k_0 2^{N-1} + k_{N-2}2^{N-2} OR\ k_{N-1}2^{N-2} + \cdots + k_0 2^0 OR\ k_1 2^0 \end{bmatrix} \quad (4)$$

Analysis of the dependence of the shift indexes vectors structure on the type of the ring codes family

Each line (code sequence) of the ring code is characterized by a delta factor - the distribution of zeroes and units between two extreme units, separated by the largest number of zero symbols for a given initial vector. Ring codes having a delta factor of a particular type form a family of ring codes. Each family of ring codes is characterized by the length of *N* code sequences and the number of *m* units [2].

In [4, 5], the properties of families of ring codes are analyzed based on the delta factor of type 0011100 (units in the code sequence are placed without interruption), type 010101 (units and zeroes alternate) and type 001011. For these types of families of ring codes, mathematical models are built the formation of families based on the analysis of the values of code combinations in the decimal number system.

Table 1 shows the results of converting the families of ring codes of types 0011100, 010101, and 001011 to shift indexes vectors using the logical transformations *XOR*, *AND* and *OR*

Analysis of the structure of the shift indexes vectors allows us to conclude that, within the family of ring codes, the sequence of changes in the decimal values of the shift indexes vector is identical. The shift indexes vectors within the family differ in the number of decimal values and their values, which depend on the length of the codeword and the number of ones in each codeword.

Let us consider in more detail the patterns of change in the values of the elements of the shift indexes vectors for family of the type 0011100.

Table1. Results of indexes ring codes to shear indexes vectors using logical transformations *XOR*, *AND* and *OR*

N	*m*	Structure of the ring code	XOR transformation		AND transformation		OR transformation	
			SIV structure	Sum	SIV structure	Sum	SIV structure	Sum
6	3	000111	24642	18	21012	6	45654	24
		001011	44244	18	11211	6	55455	24
		010101	60606	18	03030	6	63636	24
8	4	00001111	2468642	32	3210123	12	5678765	44
		00010111	4464644	32	2212122	12	6676766	44
		01010101	8080808	32	0404040	12	8484848	44
10	5	0000011111	2468108642	50	432101234	20	6789109876	70
		0000101111	446868644	50	332121233	20	778989877	70
		0101010101	10 0 10 0 10 0 10 0 10	50	050505050	20	10 5 10 5 10 5 10 5 10	70

Common factors of change in the elements values of the shift indexes vectors for ring codes family of the type 001110

Table 2 shows the results transforming ring codes of the 0011100 family into shift indexes vectors using the logical transformations XOR, AND and OR.

Fig. 1 shows a graph of the dependence of the decimal values of the elements of the shift indexes vectors, formed using the logical transformations XOR, AND and OR from the position number of their placement in the shift indexes vector V from right left. The graph is presented for the ring code of the 000111 family of length N = 7 with the number of units m = 4.

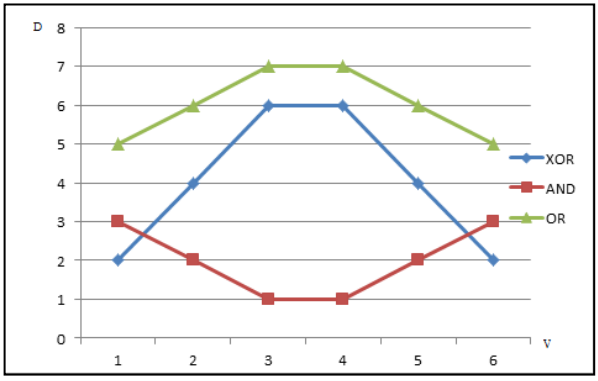


Fig. 1. Graph of the dependence of the decimal values *D* of the elements of the shift indexes vectors from the position number of their placement in the shift indexes vector *V*

Table 2. Results of transforming ring codes of the 0011100 family into vectors of shift exponents using logical transformations XOR, AND and OR

N	*m*	Structure of the ring code	XOR transformation		AND transformation		OR transformation	
			SIV structure	Sum	SIV structure	Sum	SIV structure	Sum
7	1	0000001	222222	12	000000	0	222222	12
	2	0000011	244442	20	100001	2	344443	22
	3	0000111	246642	24	210012	6	456654	30
	4	0001111	246642	24	321123	12	567765	36
	5	0011111	244442	20	433334	20	677776	40
	6	0111111	222222	12	555555	30	777777	42
8	1	00000001	2222222	14	0000000	0	2222222	14
	2	00000011	2444442	24	1000001	2	3444443	26
	3	00000111	2466642	30	2100012	6	4566654	36
	4	00001111	2468642	32	3210123	12	5678765	44
	5	00011111	2466642	30	4322234	20	6788876	50
	6	00111111	2444442	24	5444445	30	7888887	54
	7	01111111	2222222	14	6666666	42	8888888	56
9	1	000000001	22222222	16	00000000	0	22222222	16
	2	000000011	24444442	28	10000001	2	34444443	30
	3	000000111	24666642	36	21000012	6	45666654	36
	4	000001111	24688642	40	32100123	12	56788765	36
	5	000011111	24688642	40	43211234	20	67899876	60
	6	000111111	24666642	36	54333345	30	78999987	66
	7	001111111	24444442	28	65555556	42	89999998	70
	8	011111111	22222222	16	77777777	56	99999999	72

An analysis of the dependence of the decimal values *D* of the elements of the shift indexes vectors from the position number of their placement in the shift indexes vector *V* suggesting that the set of decimal values of the shift indexes vector P_V consists of 3 sets:

$$P_V = P_1 \cup P_2 \cup P_3, \tag{5}$$

where P_1 is a sequence of decimal values that are within the position of their placement in the shift indexes vector from left to right from 1 to *m* at $m \leq N\text{-}m$ and from 1 to *N-m* at $m > N\text{-}m$. In this case, *N* is the number of elements in the code combination, and *m* is the number of units;

P_2 is a sequence of decimal values, located within the position of their placement in the shift indexes vector from left to right from $m + 1$ to $N\text{-}m\text{-}1$ when $m < N\text{-}m$ and from $N\text{-}m + 1$ to $m\text{-}1$ when $|m\text{-}N| > 1$. For $|m\text{-}N| \leq 1$, the sequence is zero;

P_3 is a sequence of decimal values that are within the position of their placement in the shift indexes vector from left to right from $N\text{-}m$ to $N\text{-}1$ at $m < N\text{-}m$ and from *m* to $N - 1$ at $m > N\text{-}m$. Provided that $m = N\text{-}m$, the sequence P_3 is in the range of positions from $N\text{-}m + 1$ to $N\text{-}1$.

In general, the mathematical model for determining the limits of the sequences P_1, P_2 and P_3 is as follows:

$$\lim_{n\to\infty} P_{1n} \in \begin{cases} [1, m], m \leq N - m; \\ [1, N - m], m > N - m. \end{cases} \tag{6}$$

$$\lim_{n\to\infty} P_{2n} \in \begin{cases} [m+1, N-m-1], m < N-m; \\ [N-m+1, m-1], \mathrm{m} - \mathrm{N} > 1; \\ [\emptyset], |m-N| \le 1. \end{cases} \quad (7)$$

$$\lim_{n\to\infty} P_{3n} \in \begin{cases} [N-m, N-1], m < N-m; \\ [m, N-1], m > N-m; \\ [N-m+1, N-1], \ m = N-m. \end{cases} \quad (8)$$

The above specified sequence limits are valid for the decimal values of the elements of the shift indexes vectors formed using the logical transformations *XOR*, *AND* and *OR*.

Fig. 2 shows a graph of the dependence of the decimal values of the elements of the shift indexes vectors on the position number of their placement in the shift indexes vector *V*, formed using the *XOR* logical operation of of the 011100 ring codes family with different length of the code combination N and the number of units in the code combination $m = 4$.

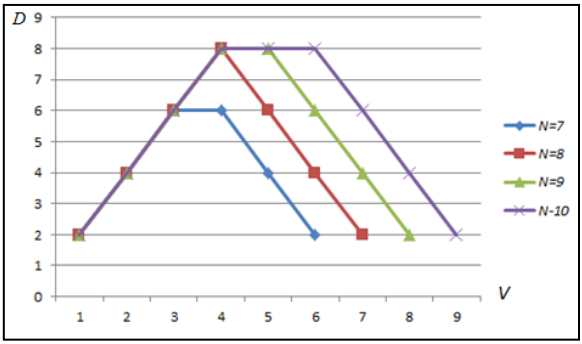


Fig.2. The dependence of the decimal values of the elements of the shift indexes vectors from the position number of their placement in the shift indexes vector, formed using the logical *XOR* operation

Conclusions

The method of transformation ring codes into vectors of shift indicators may be used for compressing information transmitted over communication channels and increasing of information transmission security. In this case, at the receiving end of the communication channel, it is necessary to solve the problem of decoding the vector of the shift indices into a ring code in order to obtain reliable information. The analysis of the structure of the vectors of the shift indicators, presented in this work, allows you to see the patterns of change in the decimal values of the elements of the vectors of the shift indicators and their dependence on the length of the code combination and the number of units in the code

combination. These regularities will allow in the future to develop an algorithm for converting shift indexes vectors into ring codes.

References

1. S.I. Otrokh, L.M. Hryshchenko, V.V. Dubrovsky, U.V. Melnik Peculiarities of the Formation of Commemoration of Kilts Kodiv Type 001011: Mathematical Model – Kyiv: Communication –2018 – No 1 – p.33-40 (in Russian).

2. V.B. Tolubko, S.I. Otrokh, L.N. Berkman, V.I. Kravchenko Manipulation coding of signal n-dimensional multi-position constellations based on the optimal noise immunity of regular structures– Kyiv: Telecomunication and information texnology –2017 – No 3(56) – p.5-11 (in Russian).

3. S.I. Otrokh, V.A. Kuzminykh, I.O. Sosnovsky Future network in action, online life– Kyiv: Communication –2018 – No 6 – p.42-45 (in Russian).

4. L.M. Hryshchenko Patterns of formation of ring codes. Mathematical model – Kyiv: Communication – 2016 – No 5(123). – p.27-31 (in Ukrainian).

5. L.M. Hryshchenko Mathematical model for creating the 010101 type family ring code– Kyiv: Communication – 2017– No 1(125). – p.58-61 (in Ukrainian).

6. E.V. Havrylko, S.I. Otrokh, V.I. Yarosh, L.M. Hryshchenko Improving the quality 8. of the future network by using the ring– Minsk: Communication Herald –2018 –No2 –p.60-64(in Russian).

7. Security of information systems (in Russian) [Electronic Resource]. Mode of access: http://intuit.valrkl.ru/course-1312/index.html.

8. Otrokh S., Kuzminykh V., Hryshchenko O. Method of forming the ring codes// CEUR WorkshopProceedings - Vol-2318, - 2018, pp. 188-198.

9. Berkman, L., Otrokh, S., Kuzminykh, V., Hryshchenko O. Method of formation shift indexes vector by minimization of polynomials// CEUR Workshop Proceedings -Vol-2577, -2019, pp. pp. 259-269.

10. Kravchenko, V., Hryshchenko, O., Skrypnik, V., & Dudarieva, H.(2021). Investigation of the dependence of the structure of shift indexes vectors on the properties of ring codes in the mobile networks of the internet of things. Informatyka, Automatyka, Pomiary W Gospodarce I Ochronie Środowiska, 11(1), 62-64. https://doi.org/10.35784/iapgos.2404.

11.Zhurakovskyi B., Toliupa S., Otrokh, S., Dudarieva H., Zhurakovskyi V. Coding for information systems security and viability //CEUR Workshop Proceedings, 2021,Vol -2859, P. 71–84.

SOCIAL ENGINEERING INFLUENCE ACTORS SET FORMATION

Volodymyr Mokhor[1[0000-0001-5419-9332]], Rostyslav Herasymov[1[0000-0002-4115-8344]], Olha Kruk[1[0000-0003-2994-6804]], Oksana Tsurkan[1[0000-0002-5524-8834]] and Vadym Yashenkov[2[0000-0002-9015-1181]]

[1]Pukhov institute for modeling in energy engineering of National academy of sciences of Ukraine, Kyiv-164, Ukraine

[2]E. O. Paton electric welding institute of the National academy of sciences of Ukraine, Kyiv-150, Ukraine

v.mokhor@gmail.com, gerasimov.rostislav@gmail.com, o.n.kruk@gmail.com, o.tsurkan24@gmail.com, vadym.yashenkov@gmail.com

Abstract. The process of ensuring safety of social and technical systems was investigated. Due to that their technical and social subsystems were singled out. Their interaction via environment, in particular organization, was demonstrated. Within the composition of the social subsystem the users are singled out. They are interpreted as workers of the organization. The forms of consciousness manipulation (e.g. weaknesses, needs, manias (addictions) and interests) were taken as the process base. Thus, relations between the social engineer and the user are reflected in the social network. Their use allowed initiate orientation of the influence. And the social engineer, user and forms of consciousness manipulation (e.g. weaknesses, needs, manias (addictions) and interests) are defined as the elements of actors set. The interaction between them is set up using relations. Thus, by forming of the actors set the peaks of the unclear social graph are initiated. The graph reflects influence of the social engineer on the user of social and technical system.

Keywords: socio-technical system, social engineering, social engineering influence, actors set, actors set formation.

Introduction

One of the main elements of the social and technical system are users, in particular workers of the organization. They are singled out within social subsystem, which interacts with the technical one via environment, in particular organization. On one hand, it allows take users as an essential part of data processing in the organization. On the other hand - to be an object of influence of social engineering. Thus, ensuring of safety of social and technical systems are reduced to opposition of using the weak points of the users, in particular weaknesses, needs, manias (addictions) and interests) [1], [2].

Social engineering influence actors set formation

Vulnerabilities of the social and technical systems users are analyzed in view of the forms of consciousness manipulation. These forms are, e.g. [2], fraud, cheating, affair, intrigue and hoax. It is possible due to setting up of relations between the social engineer and the user. This structure is interpreted as social network [3]. Its use allowed initiate orientation of influence of the social engineer on the user of the social and technical system.

Demonstration of influence of the social engineer on the social network user stipulated interpreting them as actors. In view that they may be able or unable to act, it is complemented by the manipulation forms. Thus, the actors set is created by the elements like social engineer, user and manipulation form.

First of all, it is important when the unclear social graph that reflects relations between social engineer and user of the social and technical system is identified [3], [4].

Conclusion

Thus, the actors set is created within analyzing of the social and technical systems' users vulnerability to influence of social engineering. It is stipulated by the fact that the influence is reflected by a social network. Its composition includes actors (social engineer, user and manipulation form) and relations between them. And to demonstrate them an unclear social graph is used, which peaks are identified by a created actors set.

References

1. Volobuev, S. V.: Security of socio-technical systems. Obninsk, Russia: Viking (2012).
2. Mokhor, V. V., Tsurkan, O. V, Herasymov, R. P., Tsurkan, V. V.: Information Security Assessment of Computer Systems by Socio-engineering Approach, Selected Papers of the XVII International Scientific and Practical Conference Information Technologies and Security. Kyiv, 2017, pp. 92-98, http://ceur-ws.org/Vol-2067/paper13.pdf, last accessed 2021/10/22.
3. Wasserman, S., Faust, K.: Social Network Analysis: Methods and Applications. Cambridge, England: Cambridge University Press (2012), https://doi.org/10.1017/ CBO9780511815478, last accessed 2021/10/22.
4. Moderson, J. N., Nair, P. S.: Fuzzy Graphs and Fuzzy Hypergraphs. Heidelberg, Germany: Physica-Verlag Heidelberg (2000). https://doi.org/10.1007/978-3-7908-1854-3, last accessed 2021/10/22.

SYSTEM OF INFORMATION SECURITY SYSTEMS

Volodymyr Mokhor[1][0000-0001-5419-9332], Oleksandr Bakalynskyi[1][0000-0001-9712-2036], Yaroslav Dorohyi[2][0000-0003-3848-9852] and Vasyl Tsurkan[2][0000-0003-1352-042X]
[1] Pukhov institute for modeling in energy engineering of National academy of sciences of Ukraine, Kyiv, Ukraine
[2] National technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine

v.mokhor@gmail.com, baov@meta.ua., argusyk@gmail.com, v.v.tsurkan@gmail.com

The process of ensuring information security was investigated. It has been established that preserving, first, the confidentiality, integrity, and accessibility of information is a strategic decision for the organizations. Its implementation achieved the use of the framework ISO/IEC 27k series of international standards. This framework is complemented by the risk management guidelines. This is due to the need to provide internal and/or external stakeholders with security for the organizations' activities. Because a proactive measure is used to develop information security management systems. At the same time, it is proposed to interpret each of their elements as a separate independent system. Their combination of relationships will allow the development of information security management system, and private information management. While, in general, they will determine the uniqueness of the organization's information security management activities.

Introduction

One of the strategic decisions of the organizations is to preserve above all the properties of confidentiality, integrity, and availability of information. For this purpose, appropriate systems are being implemented. One of the most common frameworks for their development is the ISO/IEC 27k series of international standards. These are complemented by the ISO/IEC 31k Risk Management Guide. This is because organizations need to guarantee the security of their activities to internal and/or external stakeholders. Guarantee is achieved through the proper handling of information security risks. Therefore, the development of information security management systems by the guidelines of the ISO / IEC 27k series of international standards is used as a proactive measure in organizations. [1], [2].

Information security system of systems

Information security management systems are developed according to the needs, expectations, and constraints of internal and external stakeholders. Based on these, requirements are formulated and, as a consequence, interactions with the environment, in particular the organization, is expressed. This establishes functional boundaries for

the development of information security management systems. For each of the functions, the start and end conditions for satisfying stakeholder requirements are set. As a result, prerequisites are created for defining the architecture of information security management systems. It is expressed by the architecture their basic concepts, properties through elements, and the relationships between them [1].

Each of the expressed elements can be interpreted as a separate independent system. Their combination of relationships allows the development of an information security management system as a system of systems. On the one hand, elements of such a system are characterized by individual architecture and behavior, for example, risk management, cybersecurity management, and private information management [1], [3], [4]. On the other hand, the uniqueness of an organization's information security management is generally determined.

Conclusion

Therefore, the interpretation of information security management systems as a system allows firstly to develop the architecture and behavior of individual independent systems. While, secondly, in general, based on their combination, determine the uniqueness of the organization's information security management activities.

References

1. ISO/IEC 27001:2013. Information technology. Security techniques. Information security management systems. Requirements. [Valid from 2013-09-25; revised 2018-12-04]. URL: https://www.iso.org/standard/54534.html, last accessed 2021/10/30.
2. Mokhor V., Bakalynskyi O., Tsurkan V.: Probabilistic criterion of information security management system development. Information Technologies and Security: selected papers of the XIX international scientific and practical conference (Kyiv, 28 November 2019). Vol. 2577. Aachen, Germany: CEUR Workshop Proceedings, 2019. P. 159-168. URL: http://ceur-ws.org/Vol-2577/paper13.pdf, last accessed 2021/10/30.
3. ISO/IEC 27032:2012. Information technology. Security techniques. Guidelines for cybersecurity. [Valid from 2012-07-16; revised 2018-12-13]. URL: https://www.iso.org/ standard/44375.html, last accessed 2021/10/30.
4. ISO/IEC 27701:2019. Security techniques. Extension to ISO/IEC 27001 and ISO/IEC 27002 for privacy information management. Requirements and guidelines. [Valid from 2019-08-05]. URL: https://www.iso.org/standard/71670.html, last accessed 2021/10/30.

ПІДХОДИ ДО УРАХУВАННЯ КОГНІТИВНИХ УПЕРЕДЖЕНЬ ЕКСПЕРТІВ В СИСТЕМАХ ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ

Андрійчук О.В.[1,2][0000-0003-2569-2026], Циганок В.В.[1,2][0000-0002-0821-4877], Каденко С.В.[1][0000-0001-7191-5636], Порпленко Я.В.[1][0000-0001-5914-8438], Мінглей Фу[3], Власенко О.О.[2]

[1] Інститут проблем реєстрації інформації Національної академії наук України, Київ, Україна

[2] Київський національний університет імені Тараса Шевченка, Київ, Україна

[3] Академія Шаосін Чжецзянського технологічного університету, Шаосін, Китай

andriichuk@ipri.kiev.ua, tsyganok@ipri.kiev.ua, seriga2009@gmail.com, daliss@ukr.net, fuml@zjut.edu.cn, vo12062007@gmail.com

Вступ

Дослідження когнітивних упереджень при підтримці прийняття рішень наразі широко досліджується науковою спільнотою [1, 2]. Системи підтримки прийняття рішень (СППР) використовуються для управління об’єктами в слабко структурованих предметних областях [3]. Результатом роботи СППР є рекомендації для осіб, що приймають рішення. [4]. Властивість неповноти опису в слабко структурованих предметних областях пов'язана з неточністю, неповнотою, невизначеністю та недостовірністю даних, що описують об’єкт [3]. Це вимагає використовувати знання експертів наряду з об'єктивною інформацією та відкритими джерелами. Таким чином актуальним питанням є дослідження і розробка підходів до урахування когнітивних упереджень експертів в СППР.

Основні етапи застосування СППР в слабко-структурованих предметних областях

Розглянемо основні етапи застосування СППР в слабко-структурованих предметних областях, на яких можуть виникнути експертні когнітивні упередження та відповідні підходи до їх урахування.

1) Після того, як особа, що приймає рішення (замовник експертизи) формулює головну ціль, інженер знань (організатор експертизи) має попередньо знайомиться з відповідною предметною областю. Для цього слід використовувати відкриті джерела (Інтернет, статі, лекції), а також і спеціалізовані засоби, наприклад системи контент-моніторингу (СКМ).

Ефект ілюзії правди – це когнітивне упередження, що виявляється у схильності вірити у достовірність інформації після її багаторазового сприйняття [5]. Це особливо актуально при дослідженні інформаційних операцій [6]. Але в інших предметних областях не можна брати експертні

оцінки базуючись безпосередньо на статистиці текстових джерел (наприклад: книг, статей, Інтернету).

2) Відповідно до головної цілі організатор експертизи проводить підбір групи експертів, що мають достатній рівень компетентності в предметній області. Це особливо актуально для малих експертних груп [7]. В подальшій роботі враховується рівень компетентності експерта у кожному з питань експертизи [4]. При цьому використовуються такі показники, як об'єктивна компонента, самооцінка, взаємооцінка [4].

Ефект Даннінга-Крюгера – це когнітивне, яке полягає в тому, що "люди, які мають низький рівень кваліфікації, роблять помилкові висновки і приймають невдалі рішення, але не здатні усвідомлювати свої помилки через свій низький рівень кваліфікації" [8]. Таким чином слід особливо ретельно підходити до підбору експертів базуючись на пошуку в спеціалізованих соціальних мережах та відкритих джерелах. Бажано використовувати метод "сніжного кому", коли наявні експерти рекомендують нових [4].

3) Експертна декомпозиція полягає в розкритті кожної цілі на підцілі/критерії, які безпосередньо впливають на неї [4]. При роботі експертної групи використовують системи розподіленого збору та обробки експертної інформації (СРЗОЕІ).

Ефект Рінгельмана, або "соціальна лінь", полягає в тому, що ефективність роботи кожного члена групи знижується на 7% при додаванні кожного наступного учасника [9]. Таким чином експерт не повинен знати, що він в групі не один. Це забезпечується засобами СРЗОЕІ.

Ефект фокусування виникає, коли люди приділяють надто багато уваги якомусь одному аспекту явища; викликає помилки у правильному передбаченні корисності майбутнього результату [10]. Це може призвести до незбалансованості гілок ієрархії при декомпозиції. За цим має слідкувати організатор експертизи за допомогою засобів СРЗОЕІ.

Помилка того, хто вижив - різновид систематичної помилки відбору, коли по одній групі («тим хто вижив») є багато даних, а по іншій («загиблим») — практично немає, в результаті чого дослідники намагаються шукати спільні риси серед тих, що «вижили» і упускають із виду, що не менш важлива інформація ховається серед «загиблих» [11]. Отже, при декомпозиції може бути однобокість розкриття цілей експертами. Тому організатору експертизи бажано подивитися відкриті джерела, застосувати СКМ, щоб переконатися, що нічого не пропустили.

Ефект велосипедного сараю - тенденція людей витрачати багато часу на несуттєві питання, часто при цьому нехтуючи найважливішими [12]. Може свідчити про низьку компетентність або про занадто вузьку спеціалізацію. Може виявитися при декомпозиції, коли на досить високому рівні ієрархії порушують стратифікацію і відразу вказують підцілі надто низького та детального рівня або навіть конкретні проекти. Це може призвести до незбалансованості гілок ієрархії при декомпозиції. За цим має слідкувати організатор експертизи за допомогою засобів СРЗОЕІ.

4) Експертне оцінювання має за мету достовірно визначити ступені переваги між альтернативами/критеріями. Можливе безпосереднє оцінювання в балах та парні порівняння.

Помилка розрізнення є тенденція сприймати два варіанти, як більш різні, коли вони реалізуються одночасно, ніж коли вони реалізуються окремо [13]. Таким чином більш пріоритетним шляхом (але більш затратним за часом) є використання парних порівнянь за допомогою засобів СРЗОЕІ.

Ефект прив'язки є особливістю оцінки невідомих чисельних значень людини, завдяки чому ця оцінка зміщується на сторону попередньо сприйнятих чисел, навіть якщо ці числа не мають відношення до значення, що оцінюється [14]. Тому слід по можливості не застосовувати безпосереднього оцінювання. Краще використовувати парні порівняння.

Ефект контрасту - це когнітивні спотворення, коли ці емоції, які відчувають особу під час першої дії, впливають на його сприйняття після наступного [15]. Тобто, відчуття значною мірою визначаються шляхом порівняння. Таким чином при експертних парних порівняннях слід враховувати порядок альтернатив [16].

Дослідження Джорджа Міллера, показали, що короткочасна пам'ять людини, як правило, не дозволяє одночасно маніпулювати більше ніж з 7 ± 2 елементами [17]. Таким чином не слід давати одному експерту на оцінювання більше ніж 7 альтернатив. Якщо альтернатив більше, то слід розподілити їх між різними експертами групи.

Висновки

Дослідження когнітивних упереджень експертів при застосуванні СППР є актуальним питанням. Розглянуто основні етапи застосування СППР в слабко структурованих предметних областях. Запропоновано підходи до урахування когнітивних упереджень експертів в СППР.

Перелік посилань

1. Phillips-Wren, G., Power, D. J., & Mora, M. (2019). Cognitive bias, decision styles, and risk attitudes in decision making and DSS. Journal of Decision Systems, 28(2), 63–66. doi:10.1080/12460125.2019.1646509.
2. Ермошина Г. П., Андреева А. А., Добрынина М. В. Снижение негативных последствий субъективной иррациональности экономических субъектов с помощью системы поддержки принятия решений //Московский экономический журнал. – 2020. – №. 1.
3. Таран Т.А., Зубов Д.А. Штучний інтелект. Теорія і застосування. – Луганськ: Вид-во СНУ ім. В.Даля, 2006. – 242 с.
4. Тоценко В.Г. Методы и системы поддержки принятия решений. Алгоритмический аспект / Киев. : Наукова думка, 2002. – 382 с.

5. Zubko, JD; Fugelsang, J (January 2011). "Remembering makes evidence compelling: retrieval from memory can give rise to the illusion of truth". Journal of Experimental Psychology: Learning, Memory, and Cognition. 37 (1): 270–6. doi:10.1037/a0021323
6. Ланде Д.В., Андрійчук О.В., Дмитренко О.О., Циганок В.В., Порпленко Я.В. Побудова баз знань систем підтримки прийняття рішень з використанням направлених мереж термінів при дослідженні інформаційних операцій // Інформаційні технології та безпека – Т. 7 – №1(12). – С. 153-163. DOI: 10.20535/2411-1031.2019.7.1.184221
7. Цыганок В.В., Каденко С.В., Андрейчук О.В. Имитационное моделирование экспертных оценок для тестирования методов обработки информации в системах поддержки принятия решений Международный научно-технический журнал «Проблемы управления и информатики». – 2011. – №6. – С. 84-94.
8. David Dunning (2011). "The Dunning–Kruger Effect: On Being Ignorant of One's Own Ignorance". Advances in Experimental Social Psychology. 44: 247–296. doi:10.1016/B978-0-12-385522-0.00005-6
9. Czyż, SH; Szmajke, A; Kruger, A; Kübler, M (December 2016). "Participation in Team Sports Can Eliminate the Effect of Social Loafing". Perceptual and Motor Skills. 123 (3): 754–768. doi:10.1177/0031512516664938
10. Ni, Feng; Arnott, David; Gao, Shijia (2019-04-03). "The anchoring effect in business intelligence supported decision-making". Journal of Decision Systems. 28 (2): 67–81. doi:10.1080/12460125.2019.1620573
11. Redelmeier, Donald A.; Singh, Sheldon M. (2001-05-15). "Survival in Academy Award–Winning Actors and Actresses". Annals of Internal Medicine. 134 (10): 955–62. doi:10.7326/0003-4819-134-10-200105150-00009.
12. Parkinson, C. Northcote (1958). Parkinson's Law, or the Pursuit of Progress. John Murray.
13. Brooks, Margaret; Guidroz, Ashley M.; Chakrabarti, Madhura (12 November 2009). "Distinction Bias in Applicant Reactions to Using Diversity Information in Selection". International Journal of Selection and Assessment. 17 (4): 377–390. doi:10.1111/j.1468-2389.2009.00480.x.
14. Ni, Feng; Arnott, David; Gao, Shijia (2019-04-03). "The anchoring effect in business intelligence supported decision-making". Journal of Decision Systems. 28 (2): 67–81. doi:10.1080/12460125.2019.1620573.
15. Гудкова М. В. Специфика протекания процессуальных и операциональных компонентов критического социального

мышления субъекта под действием эффекта контраста // Вестник Тамбовского университета. Серия: Гуманитарные науки. – 2010. – № 1(81). – С. 210-214.
16. Oleh Andriichuk, Vitaly Tsyganok, Sergii Kadenko, Yaroslava Porplenko Experimental Research of Impact of Order of Pairwise Alternative Comparisons upon Credibility of Expert Session Results Proceedings of the 2nd IEEE International Conference on System Analysis & Intelligent Computing, (SAIC), 05-09 October, 2020, Kyiv, Ukraine, pp. 1-5. DOI: 10.1109/SAIC51296.2020.9239126
17. Miller, G. A. (1956). "The magical number seven, plus or minus two: Some limits on our capacity for processing information". Psychological Review. 63 (2): 81–97. CiteSeerX 10.1.1.308.8071. doi:10.1037/h0043158

FUNCTIONAL APPROACH TO REVERSE ENGINEERING MALWARE

Dmitry Voloshin[0000-0002-4600-2816]
Institute of special communication and information protection National technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute",
Kyiv, Ukraine
cost.filatoff@gmail.com

The distribution of malware is analyzed. The existence of various methods of counteracting its implementation is shown. Among them, the reverse engineering of malicious software is highlighted. The orientation of this method towards determining the patterns of its behavior has been established. At the same time, attention is focused on the presentation of the reverse engineering of malicious software by a variety of functions. The existence of the need for their coordination is shown, first, to the input and output data. Therefore, to overcome this limitation, it is proposed to use a functional approach to reverse-engineering malicious software. This achieved the consistency of the reflection of this activity by the multitude of functions of the upper and lower levels. They are presented in terms of inputs, outputs, constraints, and implementation devices. Each of the levels reflects activities, processes, operations, and actions. Whereas the action shows an elementary function within the malware reverse engineering activity. For its implementation in practice, the use of a separate component is considered. The unification of this representation is achieved with the graphical notation IDEF0 by the established goals and points of view.
Keywords: malware, reverse engineering malware, functional approach, functions set, IDEF0.

Introduction

Currently, there is a trend towards the proliferation of malicious software. This can be prevented by developing appropriate methods. Reverse engineering stands out among them. Its use in practice is focused on setting patterns of behavior of malicious software [1], [2].

At the same time, the use of reverse engineering of malicious software as an activity is characterized by a variety of presentations of functions. In this case, it is necessary to determine the input, output data, restrictions, and mechanisms for their implementation. This establishes the relevance of using a functional approach to reverse-engineering malicious software.

Functional approach to reverse engineering malware

The use of a functional approach to reverse-engineering malicious software is focused on representing it as an activity. Such activity is defined as a top-level function displayed according to a set starting point, for example, by cybersecurity professional. With this in mind, the goal of reverse engineering is formulated, first of all, the installation of malicious software templates. For this activity, input, initial data, restrictions, and mechanisms for its implementation are set [3], [4].

Detailing the reverse engineering of malicious software is achieved by representing it using lower-level functions. These levels are processes, operations, and activities. In this case, the action is interpreted as an elementary level and can be implemented, for example, by a separate component. By analogy with mirroring the reverse engineering of malware, input, output, constraints, and mechanisms are established for each lower-level function. To implement the described functional approach, the graphic notation IDEF0 [4] was chosen.

Conclusion

Therefore, the use of a functional approach to reverse-engineering malicious software is focused on the display of this activity by a variety of functions. This display is carried out according to the established goals and point of view, in particular, of the cybersecurity specialist. In addition, for each function, input, initial data, constraints, and mechanisms for their implementation are established. Their presentation is unified using the graphical notation IDEF0.

References

1. Alrammal, M., Naveed, M., Sallam, S., Tsaramirsis, G.: Malware analysis: Reverse engineering tools using santuko linux, Materials Today: Proceedings, https://doi.org/10.1016/j.matpr.2021.10.243.
2. Xiang, Z., Miller, D. J., Kesidis, G.: Reverse engineering imperceptible backdoor attacks on deep neural networks for detection and training set cleansing. Computers & Security, vol. 106 (2021), https://doi.org/10.1016/j.cose.2021.102280, last accessed 2021/11/12.
3. International Organization for Standardization. (2016, May 27). ISO/IEC/IEEE 24748-4:2016. Systems and software engineering. Life cycle management. Part 4: Systems engineering planning. Geneva, 2016, 62 p.
4. International Organization for Standardization. (2012, Sept. 15). ISO/IEC/IEEE 31320-1:2012. Information technology. Modeling Languages. Part 1: Syntax and Semantics for IDEF0. Geneva, 2012, 120 p.

ПРЕДСТАВЛЕННЯ ІНФОРМАЦІЇ В РЕФЕРАТИВНІЙ БД «УКРАЇНІКА НАУКОВА» ТА УРЖ «ДЖЕРЕЛО»

Світлана Добровська
Інститут проблем реєстрації інформації НАН України, Київ, Україна
e-mail: djereloipri@ukr.net

Розглянуто Національну систему реферування: реферативну базу даних «Україніка наукова» та УРЖ «Джерело». Наведено технологію представлення інформації в реферативній БД «Україніка наукова». Визначено пріоритетні напрями розвитку національних реферативних ресурсів. Показано, що створення та вдосконалення національної системи реферування не має аналогів в Україні за інформаційною та технічною складовою.

Постановка проблеми

Світовий досвід свідчить, що реферативні бази даних та реферативні журнали здійснюють не тільки оперативне інформування різних категорій користувачів про видання в науковій літературі, а й забезпечують ретроспективний пошук публікацій, зменшують негативний вплив пов'язаного з диференціацією наук розсіянням публікацій, інформують про досягнення в суміжних областях наук, сприяють інтеграції наукових напрямків і дисциплін. Відсутність цілісних систем реферування наукових джерел призводить до втрат інформації, перешкоджає якісному обслуговуванню вчених і фахівців, участі в міждержавному обміні науково-технічною інформацією. [1]

Бібліометричні дослідження з використанням баз даних реферативної інформації спрямовані на проведення кількісних досліджень, орієнтованих не на отримання конкретної інформації про проблеми розвитку окремих галузей науки, а на виявлення тенденцій, причому головним чином довгострокових тенденцій, що пов'язано зі стратегічним відстеженням (моніторингом) еволюції наукової діяльності [2].

Мета роботи

Сьогодні актуальною стає проблема координації зусиль провідних бібліотечних, інформаційних, наукових та видавничих установ України з метою вдосконалення Національної системи реферування, наповнення її дійсно вичерпною інформацією щодо результатів наукової діяльності українських учених і фахівців, а також створення на її основі системи оперативного пошуку необхідних документів. Ефективному використанню реферативної бази даних «Україніка наукова» сприятиме вдосконалення пошукових механізмів бази даних та стандартів наукових метаданих, узгоджених з міжнародними нормативами обміну науковою інформацією і науковими цифровими ідентифікаторами (ORCID, DOI). Зокрема, для

інтеграції з міжнародними науково-інформаційними базами даних необхідно забезпечити обов'язкове наведення критичних елементів метаданих, таких як автор(и), назва, анотація, ключові слова англійською мовою.

Результати дослідження

Загальнодержавна реферативна база даних «Україніка наукова» – національний інформаційний ресурс, єдина в Україні універсальна база даних, що містить бібліографічні описи та реферативну інформацію про вітчизняні наукові публікації з усіх галузей знань (природничі, технічні, суспільні, економічні, гуманітарні, медичні науки). Це інтеграційна основа наукової інформаційної галузі України. Її використання має багатоаспектний характер (підготовка та випуск УРЖ, підтримка он-лайнового доступу до реферативної інформації засобами глобальних комп'ютерних мереж, створення в перспективі української служби електронної доставки документів, формування наукової електронної бібліотеки шляхом повнотекстового розширення реферативних записів, проведення широкого спектру бібліометричних, інформетричних і наукометричних досліджень, організація внутрішньо- та міждержавного обміну інформацією). РЖ вже зіграв і продовжує грати важливу роль в розвитку вітчизняної і світової науки. [3]. Вже 25 років УРЖ «Джерело» є національним надбанням України, де накопичено великий інформаційний масив даних про проведення фундаментальних і прикладних досліджень вченими, фахівцями різних галузей, аспірантами та студентами. Видання реферативного журналу дозволяє проводити пошук, відбір та систематизацію джерел інформації, компенсує наслідки розсіювання інформації, а також встановлює єдину науково-технічну термінологію та рубрикатор. Вимоги до подання інформації в реферативному журналі зростають і його якість залежить від таких факторів як оперативність та повнота представлення інформації, якість рефератів. Для електронних видань повинен бути ще й зручний пошуковий апарат. Реферативні видання призначені для науковців і спеціалістів, яким потрібна конкретна, а не загальна інформація, у тому числі для наукометричного аналізу.

Концепція побудови реферативної БД передбачає поєднання принципів:

- розподіленого аналітико-синтетичного опрацювання потоку наукових видань, які вийшли друком в Україні;
- централізованої кумуляції кооперативно створеної реферативної інформації з формуванням загальнодержавної РБД «Україніка наукова» та використанням її інформаційних ресурсів

Головним напрямком розвитку реферативної бази даних «Україніка наукова» є розширення бібліометричних сервісів, які надає вона для проведення наукометричних досліджень. Слід відзначити, що РБД «Україніка наукова» дозволяє бібліометричні дослідження в першу чергу з використанням показників публікаційної активності, що дає змогу здійснювати визначення пріоритетних напрямів, робити висновки та

прогнози на розвиток різних галузей, зокрема, енергетики, інформаційних технологій, економіки, педагогіки, політики тощо. Для окремих галузей спостерігається досить швидке зростання кількості публікацій, збільшується кількість журналів, в яких вони представлені. За результатами моніторингу публікаційної активності вітчизняних учених можна виявити найперспективніші напрями досліджень у різних галузях, а також з'ясувати, котрі з них насамперед потребують державного фінансування.

Для проведення більш глибоких наукометричних досліджень з виявлення тенденцій розвитку наукових досліджень доцільно використовувати аналіз складних мереж термінів у наукових публікаціях та мереж співавторів, новим методом досліджень може стати застосування так званих кореляційних мереж. Проведення таких досліджень дозволить розширити можливості наукометричного аналізу з виявлення пріоритетних напрямків розвитку наукових досліжень та визначення груп експертів. Для ефективного використання реферативної бази даних необхідна розробка наукометричного апарату для дослідження тенденцій розвитку української науки в тому числі з використанням технології складних мереж [3].

Понад 25 років проєкт РБД «Україніка наукова» існує автономно на сайті НБУВ на власній технологічній платформі і оновлюється приблизно раз на півроку з технологічної бази FULL. Найближчим часом у ході оптимізації архітектури РБД «Україніка наукова» планується її інтеграція до масиву електронних ресурсів сайту НБУВ. Відтоді співробітники служби реферування наповнюватимуть РБД «Україніка наукова» безпосередньо на сайті НБУВ, що значно підвищить оперативність подання інформації. Послуги інших служб НБУВ зі створення РБД «Україніка наукова» стануть доступні в комп'ютерному режимі, що підвищить ефективність нинішнього ручного режиму процесу систематизації.

В РБД «Україніка наукова» існує такий розподіл публікацій за видами: статті з журналів і збірників наукових праць – 66%, монографії – 17%, автореферати дисертацій – 17%. Значну частину наукових публікацій складають статті з журналів і збірників наукових праць.

Наведемо порівняльну таблицю наповнення інформації в реферативній базі даних «Україніка наукова» та випуску журналу «Джерело» (існуюча та вдосконалена системи представлення інформації).

Пошук за контентом РБД «Україніка наукова» здійснюється шляхом уведення таких пошукових елементів як прізвища авторів, редакторів та укладачів публікації; назва публікації; ключові слова (пошук за будь-яким словом із бібліографічного опису або тексту реферату); галузь знання; назва періодичного видання; індекс Рубрикатора НБУВ; рік видання; вид видання.

Таблиця 1

Технологія представлення інформації в реферативній БД «Україніка наукова»

	Існуюча система представлення інформації в реферативній БД «Україніка наукова»	**Вдосконалена система представлення інформації в реферативній БД «Україніка наукова»**
1	Наповнення РБД «Україніка наукова» здійснюється шляхом включення рефератів з електронних ресурсів НБУВ, з сайтів журналів, а також сканування анотацій рефератів з дисертацій, книжок та періодичних видань	Наповнення РБД «Україніка наукова» здійснюється шляхом включення рефератів з електронних ресурсів НБУВ, з сайтів журналів, а також сканування анотацій рефератів з дисертацій, книжок та періодичних видань
2	РБД «Україніка наукова» існує автономно на сайті НБУВ на власній технологічній платформі	У ході оптимізації архітектури РБД «Україніка наукова» планується її інтеграція до масиву електронних ресурсів сайту НБУВ
3	РБД «Україніка наукова» оновлюється приблизно раз на півроку з технологічної бази FULL	РБД оновлюватиметься на сайті кожен робочий день, що забезпечуватиме регулярне оновлення веб-сторінок сайту
4	Співробітники служби реферування наповнюють РБД «Україніка наукова» на робочих компютерах не підключених до сайту НБУВ та передають інформацію в електронному вигляді	Співробітники служби реферування наповнюватимуть РБД «Україніка наукова» безпосередньо на сайті НБУВ, тобто буде можливість дистанційного доступу до РБД.
5	Фахівці із систематизації вносять індекси Рубрикатора, зробленого на основі бібліотечно-бібліографічної класифікації (ББК), до рефератів з наукової періодики в ручному режимі	Співробітники відділу комплексного опрацювання документів зможуть проводити індексацію рефератів із наукової періодики безпосередньо в електронних полях РБД зі своїх робочих комп'ютерів.
6	Інформація зосереджується в систематичному порядку згідно з розділами тематичного пошуку	Інформація зосереджуватиметься в систематичному порядку згідно з розділами тематичного пошуку
7	Формування серій проводиться шляхом відбору записів з РБД «Україніка наукова» за датами та наповненням по серіям.	Формування серій проводиться шляхом відбору записів з РБД «Україніка наукова» за датами.
8	Для підготовки друкованого видання УРЖ «Джерело» проводилось редагування журналу редакторами НБУВ.	Для підготовки друкованого видання УРЖ «Джерело» проводиться редагування журналу верстальщиками ІПРІ НАНУ.

Періодичність випуску УРЖ «Джерело» складає 6 разів на рік та видається чотири серії за тематичними розділами наукових знань. Кожна серія є окремим друкованим виданням, має свій код ISSN.

Таблиця 2

Кількість записів рефератів в УРЖ «Джерело» за 2018–2020 рр.

Назва серії	2018 р.	2019 р.	2020 р.
Сер.1 Природничі науки	3 954	4 151	3 172
Сер.2 Техніка. Промисловість. Сільське господарство	8 297	7 328	6 092
Сер.3 Суспільні та гуманітарні науки. Мистецтво	9 749	10 049	7 277
Сер.4 Медицина	6 172	4 913	4 988
Всього	28 172	26 441	21 529

Поточне наповнення РБД «Україніка наукова». За весь період з 1999 до 2021 р. РБД «Україніка наукова» відображено 697 863 тис. наукових документів. Найбільша кількість публікацій, відображених в РБД «Україніка наукова» та УРЖ «Джерело» стосується суспільних і гуманітарних наук, на другому місці – технічні науки та сільське господарство, далі – природничі науки .

Збільшення надходжень реферативної інформації досягається завдяки поступовому залученню дедалі більшої кількості редакцій періодичних видань і видавничих організацій до подання в РБД «Україніка наукова» відомостей про їх друковану продукцію в електронній формі. Розвиток реферативного журналу та РБД здійснюється у напрямах підвищення вимог до першоджерел та якості рефератів і анотацій, вдосконалення програмного та апаратного забезпечення, збільшення обсягів інформації.

У цьому контексті можна визначити пріоритетні напрями розвитку національних реферативних ресурсів (РБД «Україніка наукова» та УРЖ «Джерело») на найближчу перспективу:

- інтеграція загальнонаціональних наукових інформаційних ресурсів: РБД «Україніка наукова» та повнотекстової електронної бібліотеки «Наукова періодика України»; створення на їх основі наукової електронної бібліотеки з розвиненим пошуковим інтерфейсом;

- поліпшення користувацького інтерфейсу електронної системи пошуку наукової інформації України завдяки впровадженню системи авторитетних файлів: назв періодичних видань, імен науковців, назв наукових установ, наукового рубрикатора, предметних рубрик тощо;
- забезпечення наявних служб реферування розподіленою платформою реферування та систематизації наукових публікацій; створення єдиної постійно поповнюваної реферативної бази даних;
- підвищення зацікавленості вітчизняних учених, наукових інституцій та видавництв у постачанні необхідних метаданих публікацій та повних текстів до загальнонаціональних наукових інформаційних ресурсів, які дають можливість здійснювати наукометричний аналіз публікаційної активності науковців, що в підсумку дозволяє визначити ефективність результатів наукової діяльності в різних галузях;
- створення професійних проблемно-орієнтованих баз даних на основі реферативної БД «Україніка наукова»
- створення українського індексу цитування [4].

Висновки

Реферативна БД «Україніка наукова» є розгалуженим, диференційованим за галузями знань та інтегрованим в масштабах країни інформаційним продуктом, який не має аналогів в Україні. РБД «Україніка наукова» та УРЖ «Джерело» слугують проведенню бібліо- та наукометричних досліджень на теренах України з метою виявлення пріоритетних напрямів розвитку окремих галузей науки. На підставі бібліометричного аналізу наукових публікацій з наукових напрямів, які є провідними для НАН України (фізичних, математичних, біологічних і хімічних наук), можна виявити певні закономірності та проблеми їх розвитку в Україні за останні роки.

Література

1. Костенко Л. Наукометрія: методологія та інструментарій / Л. Костенко, О. Жабін, О. Кузнєцов, Є. Кухарчук, Т. Симоненко // Вісник Книжкової палати. - 2015. - № 9. - С. 25-29. - Режим доступу: http://nbuv.gov.ua/UJRN/vkp_2015_9_8.

2. Гонашвили А.С. Наукометрические базы данных и работа с ними: научно-методическое пособие / А.С. Гонашвили. - СПб.: Университет при МПА ЕврАзЭС, 2020. - 57 с.

3. Петров В., Крючин А., Лобузіна К., Гарагуля С., Балагура І., Мініна Н. Технологія формування реферативної бази даних «Україніка наукова»: наукометричний потенціал. Бібліотечний вісник. - 2020. - № 6. - С.7-14, **DOI:** http://doi.org/10.15407/bv2020.06.007

4. Добровська С. В. Розвиток національної системи реферування / С. В. Добровська // Наука та наукознавство. – 2015. – № 3. – С. 45-51.

INCREASING THE MULTI-POSITION SIGNALS NOISE IMMUNITY OF MOBILE COMMUNICATION SYSTEMS BASED ON HIGH-ORDER PHASE MODULATION

Mykhailo Klymash[1][0000-0002-1166-4182], Liubov Berkman[2][0000-0002-6772-1596], Serhiy Otrokh [3][0000-0001-9008-0902], Volodymyr Pilinsky[3][0000-0002-2569-9503], Oleksandr Chumak[4][0000-0003-3876-8149], and Olena Hryshchenko [2][0000-0001-8198-5056]

[1]Lviv Polytechnic National University, Lviv, Ukraine
[2]State University of Telecommunications, Kyiv, Ukraine
[3] National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine
[4]Military Diplomatic Academy named after Yevheniy Bereznyak, Kyiv, Ukraine

klymash@journal.kh.ua, berkmanlubov@gmail.com, 2411197@ukr.net, pww@ukr.net, a_ch_i@ukr.net, lgrischenko88@ukr,net

The article presents the results of a study of situations typical for 4G (Long Term Evolution - LTE) mobile communication systems, in which there are conditions for the use of the autocorrelation method for processing multi-position signals, and the noise immunity of demodulators that implement this method is determined. It has been proved using the theory of catastrophes that for modern mobile communication systems it is advisable to use energy methods with high-order phase-difference modulation (PDM), which will provide the system with the property of invariance to a certain class of electromagnetic interference (EMI). A comparative analysis of the application of systems with PDM of the first and second order has been carried out and it has been determined that in the case of autocorrelation reception of signals from PDM-1, it is necessary to ensure sufficiently stringent requirements for the stability of the frequency of the carrier wave. It was found that in the case of using PDM-1, the autocorrelation technique is much simpler, but the PDM-2 has a unique property of invariance to the frequency of the carrier wave, which neither coherent nor optimal incoherent methods have. An energy (autocorrelation) demodulator of signals with double PDM-1 is proposed, which has the property of relative or absolute invariance to changes in the frequency of the carrier wave, which can occur in digital information transmission systems during communication with rapidly moving objects. satellite communication systems, as well as fiber-optic communication systems, mobile broadband access with support for the fourth generation technology.

Keywords: phase-difference modulation, autocorrelation method, multi-position signals

Introduction

At present, in the case of using information transmission rates close to the capacity of communication channels, when calculating the noise immunity of a system, it is necessary to take into account not only white Gaussian noise, but also multiplicative noise distributed in the spectrum and time.

In this case, the use of correlation reception methods is not effective, since the possible variants of the sent signal are not known, but only their energies. Therefore, in this case, it is advisable to use energy methods of receiving signals.

A large number of optimal algorithms have been created in the theory of signal reception and processing. The best algorithm (according to the fundamental position of the theory of systems) under certain reception conditions can be considered an algorithm that meets the criteria of some optimality depending on the established restrictions due to the characteristics of the channel and signal and the peculiarities of the communication system as a whole according to certain specified criteria [1].

The most adequate classification of conditions under which one or another optimal algorithm can be synthesized is based on what information about the signal is known before it arrives, that is, according to the degree of a priori uncertainty of the situation. There is a wide range of intermediate situations between a signal with a known initial phase and a signal with a completely unknown initial phase.

The hierarchy of optimal algorithms is opened by coherent reception, which is optimal under conditions when the possible realizations of the transmitted signal are fully known. Optimal incoherent reception is optimal for signals with an unknown but uniformly distributed initial phase. If we go further by reducing the a priori information about the sent signal, we can to synthesize other reception methods that can be applied under appropriate conditions. In this case, the less we have a priori information about the parameters of the signal during its processing, the less noise immunity. Completing the hierarchy of good methods for receiving signals are methods using methods for processing multi-position signals [2] of unknown shape. These include algorithms for autocorrelation processing of phase-modulated signals. The essence and conditions of application of autocorrelation signal processing are given below.

Algorithms for autocorrelation processing of phase-modulated signals

Let there are two versions of the transmitted signal $S_1(t)$ or $S_2(t)$ (signals of unknown shape) and the known interval of their existence $(0, \tau)$. These signals can be presented as a sum of orthogonal transformation functions with bases $\{\varphi_i\}$ and $\{\psi_i\}$:

$$S_1(t) = \sum_{i=1}^{2\beta} a_i \varphi_i(t), \qquad S_2(t) = \sum_{i=1}^{2\beta} \beta_i \psi_i(t), \tag{1}$$

where 2β is the base of expected signals;

$$a_i = \int_0^\tau S_1(t)\,\varphi_i(t)dt;\ \beta_i = \int_0^\tau S_2(t)\psi_i(t)dt. \quad (2)$$

The signals $S_1(t)$ and $S_2(t)$ are called signals of unknown shape if there is no information about the coefficients a_i and β_i. Provided that the basis functions φ_i and ψ_i are known, the system of functions $\{\varphi_i\}$ defines the space of possible signal shapes $S_1(t)$, and the system of functions $\{\psi_i\}$ defines the space of possible signal shapes $S_2(t)$. We can assume that in this case the spaces of the expected signals are known, determined by the corresponding set of basis functions and the time interval of their existence.

If the distributions of the coefficients a_i and β_i are known, then a simple maximum likelihood rule can be used to synthesize the optimal algorithm for receiving signals $S_1(t)$ and $S_2(t)$; if the distributions a_i and β_i are unknown, then the general maximum likelihood rule should be applied, according to which, out of two possible hypotheses $S_1(t)$ or $S_2(t)$, one should choose the one for which the maximum of the likelihood function will acquire a greater value.

For signals with the same energy, when it is known that

$$\sum_{i=1}^{2\beta} a_i^2 = \sum_{i=1}^{2\beta} \beta_i^2 \quad (3)$$

Signal $S_1(t)$ should be considered as transmitted if

$$\sum_{i=1}^{2\beta}\left[\int_0^\tau x(t)\varphi_i(t)dt\right]^2 > \sum_{i=1}^{2\beta}\left[\int_0^\tau x(t)\psi_i(t)dt\right]^2 \quad (4)$$

where i is a signal $S_2(t)$ when the inequality is opposite.

The mathematical expressions in square brackets of inequality (4) are the coefficients of the orthogonal transformation of the received signal according to the basis functions φ_i and ψ_i, and the sums of the squares of these coefficients (the left and right sides of (4) are equal to the energy of the received signal in the spaces functions $\{\varphi_i\}$ and $\{\psi_i\}$ respectively. Based on the remarks made, inequality (4) can be written as:

$$\int_0^\tau x_1^2(t)\,dt > \int_0^\tau x_2^2(t)\,dt, \quad (5)$$

where $x_1(t)$ and $x_2(t)$ is the estimate of the membership of the received signal in the $\{\varphi_i\}$ } and in the space $\{\psi_i\}$. Based on inequality (5), the following algorithm for estimating the signal arriving at the receiver is proposed. First, you need to calculate the energy of the received signal, in the spaces of the first and second possible signal realization, and assume that a signal has arrived, in the space of which the calculated energy is greater. In this regard, acceptance algorithms based on the application of criteria (4) and (5) are called energy or autocorrelation. The last name of the algorithm follows from the fact that in the case of its application in the receiver, it is necessary to calculate the convolution of the received signal with its time-delayed copy with a different delay time or, otherwise, with a

different time offset (in (5) this offset is equal to zero). These algorithms can be applied to signals with different types of signal modulation. To obtain the corresponding separate algorithm, one should apply the basis functions corresponding to the applied type of modulation [3] to calculate inequality (4).

Autocorrelation demodulators of signals with PDM-1 attract attention with their extreme simplicity, since their implementation does not require either coherent reference waves extractor devices as in coherent demodulators, or correlators with quadrature reference waves, as in optimal incoherent demodulators [4]. But it should be borne in mind that potentially autocorrelation demodulators are inferior in noise immunity to coherent and optimal incoherent. In contrast, the noise immunity of autocorrelation (energy) demodulators depends not only on the ratio h^2 of the signal energy to the spectral power density of the noise, but also on the base of the received signal-to-noise mixture, which, in turn, depends on the bandwidth F of the receiver input filter. The larger the base $2FT$, the lower the noise immunity at the same value h^2. The more complex the signal, the greater its dimension and the more other identical conditions are the probability of error. These are the features of the autocorrelation method of receiving. At the same time, in the case of receiving narrow-band signals with a base of $B \approx 2$, autocorrelation demodulators are slightly inferior to the optimal incoherent ones with respect to noise immunity. When comparatively assessing the noise immunity of coherent and maximum incoherent demodulators, on the one hand, and autocorrelation ones, on the other, it should be taken into account that the former have an advantage only when the conditions for their optimality and performance are met. If these conditions are not met and it is necessary to provide a priori reception of a signal of unknown shape, then the autocorrelation demodulator can provide a lower probability of error. Therefore, we can conclude that the autocorrelation technique (as a method arising from the generalized maximum likelihood rule) provides a minimum error probability under the condition of uniformly distributed unknown parameters (conversion coefficients of a signal of an unknown shape).

Algorithms for autocorrelation processing of signals with PDM-1 directly follow from the general maximum likelihood rule and, in this sense, are optimal for signals of unknown shape. The use of an optimal algorithm for receiving signals of an unknown shape, in which the general maximum likelihood rule is implemented, when using PRM-1 determines the structure of the autocorrelation signal processing circuit shown in Fig. 1.

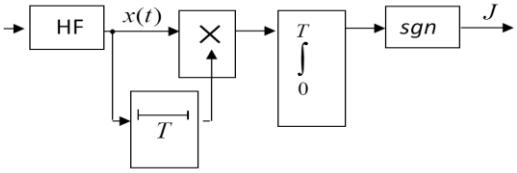


Fig.1. Functional diagram of an optimal autocorrelation demodulator of signals with a single PDM

The input signal goes to a bandpass filter (HF), which is designed in the autocorrelation demodulator to perform two functions:

- firstly, for the usual frequency selection function of the useful signal. Such selection is necessary in cases where the demodulator input arrives at the baseband signal of the transmission system with FDC (frequency distribution of channels) and in other cases. To implement the same function, a similar input bandpass filter is used at the input of coherent and good incoherent demodulators;
- secondly, the input bandpass filter provides a limitation of the spectrum (and, consequently, the power) of the fluctuation noise entering the demodulator input, since in the case of using the autocorrelation method of signal reception, in contrast to the correlation methods - coherent and optimal incoherent, the noise immunity depends on the width frequency bands (and, as a consequence, - power) of noise, and not only from its spectral density.

On the basis of a similar optimal algorithm for processing signals from PDM-2, a functional diagram of an energy demodulator can be proposed. As in the case of optimal incoherent reception, the energy demodulator of signals from the PDM-2 is superior in noise immunity to the equivalent autocorrelation demodulator of signals from the PDM-1. At the same time, the noise immunity of these types of demodulators depends on the stability of the frequency of the carrier wave. At the same time, such dependence in the PRM-2 autocorrelation demodulator is not.

If in the case of PDM-1 the autocorrelation acceptance is attractive mainly for its simplicity, then in the case of PRM-2 the autocorrelation acceptance, which in this case is one of the submaximal modifications of the algorithm for receiving the corresponding signals of an unknown shape, has a unique property of invariance to the frequency of the carrier wave. This property is absent in the coherent and maximum incoherent acceptance method. Therefore, autocorrelation algorithms for receiving signals from the PDM-2 are important in mobile networks. It is advisable to use them in channels with an undefined signal frequency [5].

Conclusions

The considered algorithms can be effective not only in the case of an unknown waveform, but also in the case of variable waveform signals. However, algorithms can only be used for signals whose changes occur within a certain space of signal implementations. With phase difference modulation (PDM-1), this means that the signals must be repeated with sign accuracy over an interval of two bursts. This, in particular, implies rather stringent requirements for the stability of the frequency of the carrier wave when using the autocorrelation method for receiving PDM-1 signals.

Autocorrelation modems with PDM-2 are characterized by the property of relative or absolute invariance to changes in the frequency of the carrier vibration, they are used in digital information transmission systems in communication systems with fast moving objects, in satellite communication systems, as well as

in fiber-optic communication systems, mobile broadband access with support for LTE technology [6].

Thus, when transmitting digital information by different communication channels, conditions arise under which the receiver must process a signal with an unknown or inaccurately known frequency of the carrier wave - channels with an undefined signal frequency. For such channels and conditions, when using signals from the second-order PRM, it is better to use the autocorrelation method of reception.

References

1. V. B. Tolubko, L. N. Berkman,, S. I. Otrokh, V. I. Kravchenko. Manipulation coding of signal n-dimensional multi-position constellations on the basis of regular structures that are optimal in terms of noise immunity / Kyiv: Telecommunication and information technologies – 2017. – No 3 (56). – p. 5-11 (in Ukrainian).

2. Vladimir Tolubko, Lyubov Berkman, Sergei Otrokh, Oleksandr Pliushch and Vladislav Kravchenko. Noise Immunity Calculation Methodology for Multi-Positional Signal Constellations / 14th IEEE International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET'2018): Conference Proceedings. – Lviv, 20-24 of February, 2018. – p. 436.

3. Okunev Yu.B.//Digital transmission of information with phase-modulated signals/ M .: Radio and communication, 1991 -- 295 p. (in Russian).

4. Steklov V.K., Berkman L.N., Starodub N.M.//The choice of the generalized criterion of optimality of information network management systems / Zvyazok.-2000. - No 5. - p.12-16. (in Russian).

5. Levin L.S., Plotkin M.A.// Digital information transmission systems/ M.: Radio and communication. 1982. – 216 p. (in Russian).

6. Tikhvinsky V.O., Terentyev S.V.//Mobile networks LTE: technologies and architecture / M .: Eco-Trends. – 2010 p. 192-196. (in Russian).

ЗМІСТ

Національна академія наук України
Інститут проблем реєстрації інформації НАН України

ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ ТА БЕЗПЕКА

Матеріали XXI Міжнародної науково-практичної конференції

Випуск 21

Підп. до друку 28.12.2021. Формат 60x84 1/16. Папір офс. Гарнітура Times.
Спосіб друку – ризографія. Ум. друк. арк. 10,5. Обл.-вид. арк. 24,36. Наклад 100 пр.
Зам. № 15-200.

ТОВ "Інжиніринг" 978-966-2344-84-4