

УДК 004.724.4

МОДЕЛЬ «ЖИЗНЕННОГО ЦИКЛА» СООБЩЕНИЯ В СОЦИАЛЬНЫХ СЕТЯХ

*Д.т.н Ландэ Д. В.¹, Грайворонская А.Н.², Березин Б.А.¹,
¹ИПРИ НАН Украины, dwlande@gmail.com,
²НТУУ «Киевский политехнический институт», г. Киев,*

В работе приведены правила функционирования мультиагентной сети, определены статистические закономерности, связанные с «жизненным циклом» агентов – информационных сообщений. Выявлены статистические закономерности, относящиеся к количеству лайков и репостов, распределение которых, как видно по результатам моделирования, соответствуют распределению Вейбулла. Данные моделирования проверены путем исследования реальной сети микроблогов Twitter.

Введение

Моделирование распространения информации в социальных сетях позволяет выявлять закономерности, которые могут использоваться как в изучении механизмов передачи информации в таких сетях, так и уровня ее воздействия, что является актуальной задачей [1].

В работе рассматривается мультиагентная модель, ключевым элементом (агентом) в которой является сообщение [2]. Каждое сообщение может вызывать разные типы реакции, такие как положительные/отрицательные комментарии, выражение поддержки или отторжения (like/dislike, лайк/дизлайк), текст сообщения может быть скопирован, размещен на другой странице (repost, репост), также одним сообщением может быть размещена ссылка на другое сообщение (link, линк). Эволюция агента будет связана с событиями, которые с ним происходят. В качестве основной характеристики агента введем «энергию» (E), которая отображает актуальность сообщения и степень интереса к нему. При этом старение информации или отрицательная реакция уменьшает энергию сообщения, а положительная реакция или появление ссылки на сообщение увеличивает его энергию.

Эволюция агента в модели

Агент появляется с начальной энергией E_0 и с каждым дискретным отсчетом времени его энергия уменьшается на 1. Будем рассматривать события типичные для социальных сетей: like, dislike, repost, link (предоставления ссылки на одного агента другим). Эти события влияют на энергию агента следующим

образом: like повышает энергию на 1, dislike уменьшает на 1, repost повышает на 2, link повышает на 1. С другой стороны, вероятность того, что какое-то из этих событий произойдет, зависит от актуальности сообщения, интереса к информации в нем, что в терминах модели выражается энергией. В связи с этим определим вероятности того, что с сообщением с энергией E произошло определенное событие, следующим образом:

$$p_{like}^{(E)} = p_{l_0} \varphi(E); \quad p_{dislike}^{(E)} = p_{d_0} \varphi(E); \quad p_{repost}^{(E)} = p_{r_0} \varphi(E),$$

где p_{l_0} , p_{d_0} , p_{r_0} – параметры модели, а φ – это некоторая монотонно неубывающая функция от текущей энергии агента со значениями в $[0, 1]$. При падении энергии до 0, агент «умирает» и больше не рассматривается.

Моделирование динамики информационного потока

Рассмотрим жизненный путь одного агента. Агент появляется с начальным значением энергии E_0 и далее его энергия изменяется в зависимости от событий, которые с ним происходят. Будем считать, что возможны два события: лайк и репост. За единицу времени может произойти одно из этих событий, оба одновременно или не произойти ни одного.

Обозначим ε_t как значение энергии агента в момент времени t . Тогда значение энергии в следующий момент времени можно записать следующим образом

$$\varepsilon_{t+1} = \varepsilon_t + \delta_t,$$

где δ_t является случайной величиной со значениями в $\{-1, 0, 1, 2\}$. Согласно с правилами изменения энергии, введенными выше, увеличение энергии на 2 соответствует тому, что произошли одновременно лайк и репост; увеличение на 1 – произошел только репост; энергия не меняется, если был лайк; и уменьшается на 1, если не произошло ни одно из событий. Следовательно, можно указать условное распределение δ_t при известной энергии ε_t :

$$P(\delta_t = 2 | \varepsilon_t = E) = p_{like}^{(E)} p_{repost}^{(E)};$$

$$P(\delta_t = 1 | \varepsilon_t = E) = (1 - p_{like}^{(E)}) p_{repost}^{(E)};$$

$$P(\delta_t = 0 | \varepsilon_t = E) = p_{like}^{(E)} (1 - p_{repost}^{(E)});$$

$$P(\delta_t = -1 | \varepsilon_t = E) = (1 - p_{like}^{(E)}) (1 - p_{repost}^{(E)}).$$

Данные формулы справедливы при $E > 0$. Далее будем использовать обозначение $P_{\Delta}^{(E)} = P(\delta = \Delta | \varepsilon = E)$.

Процесс изменения энергии агента можно рассматривать как целочисленное случайное блуждание с переходными вероятностями

$$p_{ij} = \begin{cases} P_{j-i}^{(i)}, & (j-i) \in \{-1, 0, 1, 2\}, \quad i > 0; \\ 1, & i = j = 0; \\ 0, & \text{иначе.} \end{cases}$$

Так как значение энергии в следующий момент времени зависит только от значения энергии в предыдущий момент времени, то стохастическая последовательность $(\varepsilon_0, \varepsilon_1, \dots, \varepsilon_t, \dots)$ является марковской цепью с переходными вероятностями p_{ij} . Граф, соответствующий такой марковской цепи показан на рис. 1.

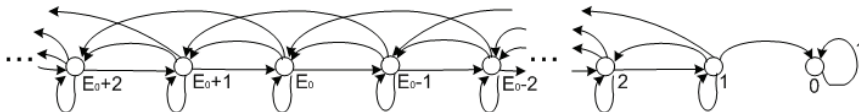


Рис. 1 – Граф, соответствующий случайному блужданию энергии агента

Можно явно указать вероятностное распределение последовательности $(\varepsilon_0, \varepsilon_1, \dots, \varepsilon_t, \dots)$ через переходные вероятности:

$$P((\varepsilon_0, \varepsilon_1, \dots, \varepsilon_t, \dots) = (E_0, E_1, \dots, E_t, \dots)) = \prod_{i=1}^{\infty} P_{E_{i-1}E_i}.$$

Или аналогичное распределение для $(\delta_0, \delta_1, \dots, \delta_t, \dots)$:

$$P((\delta_1, \delta_2, \dots, \delta_t, \dots) = (\Delta_1, \Delta_2, \dots, \Delta_t, \dots)) = \prod_{i=1}^{\infty} P_{\Delta_j}^{(E_0 + \sum_{j=1}^{i-1} \Delta_j)},$$

при условии, что $\Delta_i \in \{-1, 0, 1, 2\}$ и если для некоторого k : $\sum_{i=1}^k \Delta_i = -E_0$ (энергия агента упала до 0 на k -ом шаге), то $\Delta_i = 0, \forall i > k$. Для последовательностей $(\Delta_1, \Delta_2, \dots, \Delta_t, \dots)$, которые не удовлетворяют этим условиям, вероятность равна 0. Далее будем рассматривать только те последовательности, которые удовлетворяют указанным условиям.

Теоретически существуют последовательности $(\Delta_1, \Delta_2, \dots, \Delta_t, \dots)$, которые реализуются с положительной вероятностью, и при этом k : $\sum_{i=1}^k \Delta_i = -E_0$

может быть как угодно большим. Заметим, что k соответствует длине жизни агента и имеет смысл времени, на протяжении которого информационное сообщение остается актуальным. Следовательно, длина жизни агента должна быть конечной с большой вероятностью.

Обозначим τ_{E_0} длину жизни агента с начальным значением энергии E_0 или, что то же самое, время, за которое из E_0 попали в 0. В реалистичной мо-

дели хотелось бы иметь оценку $P(\tau_{E_0} > T_{\max}) < \varepsilon$ для малого ε и не очень большого значения T_{\max} , для того, чтобы можно было вместо бесконечных последовательностей $(\Delta_1, \Delta_2, \dots, \Delta_i, \dots)$ рассматривать конечные.

Рассмотрим функцию $\rho_T(E) = P(\tau_E > T)$. Справедливо рекуррентное соотношение:

$$\rho_T(E) = P_2^{(E)} \rho_{T-1}(E+2) + P_1^{(E)} \rho_{T-1}(E+1) + P_0^{(E)} \rho_{T-1}(E) + P_{-1}^{(E)} \rho_{T-1}(E-1).$$

Систему таких рекуррентных соотношений можно решить, используя начальные условия:

$$\rho_0(E) = \begin{cases} 0, & E = 0; \\ 1, & E \neq 0. \end{cases}$$

При начальных параметрах $p_{i_0} = 0.4$, $p_{r_0} = 0.1$ из решения рекуррентного уравнения можно получить оценку $P(\tau_{E_0} > 1.5E_0) < 10^{-3}$, т.е. время жизни агента ограничено $1.5E_0$ с большой вероятностью и, следовательно, для получения достаточно точных оценок распределения для $(\delta_0, \delta_1, \dots, \delta_i, \dots)$ можно рассматривать вектора конечной длины $T_{\max} = 1.5E_0$.

Представляет интерес распределение количества событий, которые происходят с агентами. Рассмотрим распределение количества лайков. Заметим, что если в момент времени t произошел лайк, то $\delta_t \in \{0, 2\}$, если же нет, то $\delta_t \in \{-1, 1\}$. Обозначим $(\Delta_1, \dots, \Delta_{T_{\max}})$ вектор, который удовлетворяет условию, что $\Delta_t \in \{0, 2\}$, при $t = t_1, \dots, t_n$ и $\Delta_t \in \{-1, 1\}$ иначе, где $0 < t_1 < \dots < t_n < T_{\max}$. Тогда для агента, получившего лайк справедлива формула:

$$P\{like\} = \sum_{t_1 < \dots < t_n} \sum_{(\Delta_1, \dots, \Delta_{T_{\max}})} \prod_{i=1}^{T_{\max}} P_{\Delta_i}^{(E_0 + \sum_{j=1}^{i-1} \Delta_j)}.$$

Численные результаты

На рис. 2 показана плотность полученного распределения, при начальных параметрах $p_{i_0} = 0.4$, $p_{r_0} = 0.1$. Точками, соединенными отрезками прямых линий, обозначены полученные значения $P\{\text{агент получил } n \text{ лайков}\}$.

Красная кривая соответствует плотности распределения Вейбулла [3]:

$$f(x) = \begin{cases} \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-\left(\frac{x}{\lambda}\right)^k}, & x \geq 0; \\ 0, & x < 0. \end{cases}$$

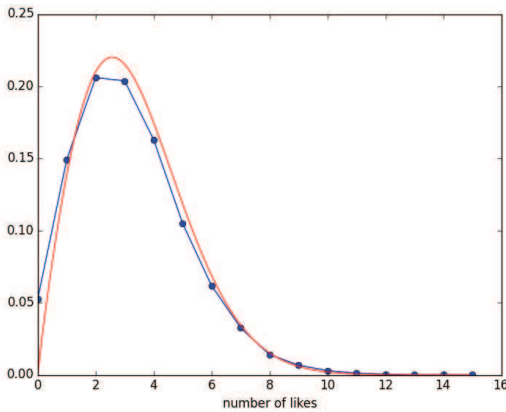


Рис. 2 – Плотность распределения количества лайков, полученных агентом при начальных параметрах $p_{l_0} = 0.4$, $p_{r_0} = 0.1$

Параметры распределения Вейбулла k и λ были получены методом максимального правдоподобия. При указанных начальных параметрах, полученные значения $k = 1.9$, $\lambda = 3.8$.

Исследование информационных потоков стети Twitter

Полученные результаты моделирования сравнивались с проведенными авторами результатами исследования жизненного цикла новостных сообщений в сети микроблогов Twitter, где, в частности, анализировались характеристики роста количества специальных репостов (ретвиттов) выбранных сообщений [4]. Распределение лайков и ретвиттов в этом случае, как и в модели, соответствовало стандартному распределению Вейбулла, причем параметр k с высокой точностью совпал с модельным (рис. 3).

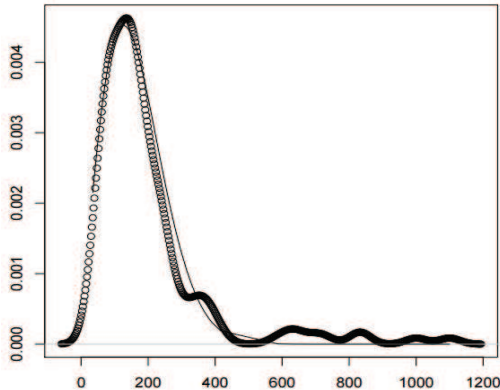


Рис. 3 – Плотность распределения количества ретвитов из реальной сети (и аппроксимация распределением Вейбулла при $k = 1.9$, $\lambda = 180$)

Выводы

Таким образом, построена мультиагентная модель «жизненного цикла» новостей в информационных сетях. Выявлены статистические закономерности, относящиеся к количеству лайков и репостов отдельных сообщений, распределение которых, как видно по результатам моделирования, соответствуют распределению Вейбулла.

Данные моделирования проверены путем исследования реальной сети микроблогов Twitter. Совпадение результатов моделирования и параметров распределения реальной сети позволяют говорить о закономерности, присущей реальным сетям, а также об адекватности модели.

Найденные закономерности могут использоваться при определении аномалий в жизненном цикле отдельных сообщений, искусственно поддерживаемых информационных кампаний [5].

Литература

1. Додонов А.Г., Ландэ Д.В., Прищепя В.В., Путятин В.Г. Конкурентная разведка в компьютерных сетях. – К.: ИПРИ НАН Украины, 2013. – 248 с.
2. Грайворонська А.М., Ланде Д.В. Дослідження інформаційних потоків, як динамічних мультиагентних систем // Системний аналіз і інформаційні технології: матеріали 17-й Міжнародної науково-технічної конференції SAIT 2015, Київ, 22-25 червня 2015 г. / УНК "ІПСА" НТУУ "КПІ". – К.: УНК "ІПСА" НТУУ "КПІ", 2015. – С. 62-63.

3. *Нешиной В.В.* Математико-статистические методы анализа в библиотечно-информационной деятельности – Минск: Белорус. гос. ун-т культуры и искусств, 2009. – 203 с.
4. *Li R., Lei K. H., Khadiwala R., Chang K. C.* TEDAS: A Twitter-based Event Detection and Analysis System // Data Engineering (ICDE), 2012 IEEE 28th International Conference, 2012. – P. 1273-1276.
5. *Додонов А.Г., Ландэ Д.В., Додонов В.А.* Распознавание информационных операций: мультиагентный подход // Открытые семантические технологии проектирования интеллектуальных систем (OSTIS-2016): материалы VI междунар. науч.-техн. конф. (Минск 18-20 февраля 2016 года) / – Минск: БГУИР, 2016. – С. 253-256.

Министерство образования и науки Украины
Российская ассоциация искусственного интеллекта
Белорусский государственный университет
информатики и радиоэлектроники
Национальный технический университет Украины «КПИ»
Факультет прикладной математики
УНК «Институт прикладного системного анализа»
Издательство «Просвіта»

***Международная научная конференция
имени Т. А. Таран***

Интеллектуальный анализ информации

ИАИ-2016

Киев, 18 – 20 мая 2016 г.

Сборник трудов

Рекомендовано Ученым советом
факультета прикладной математики

**Київ
«Просвіта»
2016**

УДК 004.8/.9+001.102](06)

ББК 32.973я43+73я43

I-73

Редакционная коллегия:

д.т.н., проф. Валькман Ю.Р., д.т.н., проф. Голенков В.В., д.т.н., проф. Дичка И.А., д.ф.-м.н., проф. Жиликова Л.Ю., д.т.н., проф., академик НАНУ Згуровский М.З., д.ф.-м.н., проф. Ивохин Е.В., д.т.н., проф. Кондраченко Ю.П., к.ф.-м.н., проф. Крейнович В., д.т.н., проф. Кузнецов О.П., д.т.н., проф. Кулаков Ю.А., д.т.н., проф. Литвинов В.В., д.т.н., проф. Смородин В.С., д.т.н., проф. Стефанюк В.Л., д.т.н., проф. Чертов О.Р.

Главный редактор к.т.н., доц. Сирота С.В.

Ответственный редактор Темникова Е.Л.

Ответственный за выпуск Копычко С.Н.

Рекомендовано Ученым советом факультета прикладной математики НТУУ «КПИ»,
протокол №12 от 31 мая 2016 г.

«Интеллектуальный анализ информации», междунар. науч. конф.
им. Т.А.Таран (16 ; 2016 ; Киев).

I-73 Международная научная конференция имени Т.А. Таран
«Интеллектуальный анализ информации» ИАИ-2016, Киев, 18–20 мая
2016 г. : сб. тр. – К. : Просвіта, 2016. – 296 с. : ил.

ISBN 978–617–7010–11–0

В сборнике опубликованы доклады, представленные на конференции по следующим направлениям: сетевые и многоагентные модели, знания и рассуждения, онтологический инжиниринг, анализ данных, мягкие вычисления, обработка естественного языка и речи, социальные проблемы и образование.

УДК 004.8/.9+001.102](06)

ББК 32.973я43+73я43

Використання матеріалів збірки можливе
за умови обов'язкового посилання.

Использование материалов сборника
возможно при условии обязательной
ссылки.

ISBN 978–617–7010–11–0

© ФПМ НТУУ «КПИ», 2016

Івохін Є. В., Махно М. Ф. Про одну гібридну модель задачі теорії розкладів	90
Калюжний М.С. Використання інтелектуального аналізу тексту для структурування знань з різних текстових джерел.....	97
Клименко В. П., Лялецкий А.В. О символьных и дедуктивных построениях в среде формализованных текстов	100
Кузнецов О.П. Пороговые модели ансамблей в нейробиологических и социальных науках	107
Кулаков Ю. А., Коган А. В., Диброва М. А. Способ формирования множества непересекающихся путей от одной к нескольким вершинам в компьютерной сети большой размерности ...	115
Кулаков Ю.А., Копычко С.Н. Способ планирования задач в распределенных системах с учетом их гранулярности	119
Кургаєв О.П., Григор'єв С.М. Відновлення граматики мови	124
Ландэ Д.В., Березин Б.А., Грайворонская А.Н. Модель «жизненного цикла» сообщения в социальных сетях	130
Лахно В.А., Двірний К.І. Технології Data Mining для аналізу стохастичних систем	137
Литвинов В. А., Майстренко С. Я., Хурцилава К. В. Целостность информационных ресурсов ГИС – типовые ограничения и методы контроля	140
Машкова А.Л. Анализ сетей информационного воздействия на основе методов агентного и когнитивного моделирования.....	147
Науменко Ю. О. Моделивання процесу розповсюдження реклами на основі гібридного рівняння дифузії.....	154
Petrakieva S. K. New Add-in for Excel for automatic solution of transport problems	157
Піднебесна Г.А., Степашко В.С. Структуризація знань у пред- метній галузі індуктивного моделювання для побудови онтологій	167

НТУУ «КПІ» Факультет прикладної математики

**Збірка праць
XVI міжнародної наукової конференції
імені Тетяни Архипівни Таран
ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ІНФОРМАЦІЇ
IAI-2016**

наукове видання

РЕДАКЦІЙНА КОЛЕГІЯ

д.т.н., проф. Валькман Ю.Р.,
д.т.н., проф. Голенков В.В.,
д.т.н., проф. Дичка І.А.,
д.ф.-м.н., проф. Жилякова Л.Ю.,
Д.т.н., проф., академік НАН України Згуровський М.З.,
д.ф.-м.н., проф. Івохін Є. В.
д.т.н., проф. Кондратенко Ю.П.,
к.ф.-м.н., проф. Крейнович В.,
д.т.н., проф. Кузнєцов О.П.,
д.т.н., проф. Кулаков Ю. О.,
д.т.н., проф. Литвинов В.В.,
д.т.н., проф. Смородин В.С.,
д.т.н., проф. Стефанюк В.Л.
д.т.н., доц. Чертов О.Р.,

Головний редактор к.т.н., доц. Сирота С.В.

Відповідальний редактор Темнікова О.Л.

Відповідальний за випуск Копичко С.М.



Промислово-торговельна фірма «Просвіта»
у формі товариства з обмеженою відповідальністю.
01032, Київ, бульвар Т. Шевченка, 46,
тел. (044) 234-15-86, 234-95-23 (факс).
Свідоцтво ДК № 221 від 16.10.2000 р.

Підписано до друку з оригінал-макета 25.05.2016. Формат 60x84 1/16. Гарнітура
таймс. Спосіб друку – ризогр. Ум. друк. арк. 16,51 Обл.-вид. арк.12,8 Наклад 250 пр.
(перший завод 80 пр.)

Віддруковано НТУУ «КПІ» ВПІ ВПК «Політехніка»

Свідоцтво ДК 1665 від 28.01.2004 р.

03056, Київ вул. Політехнічна, 14 корп. 15 тел (044) 406-81-78

Зам. №