Event prediction technique based on correlation mapping and Fourier decomposition

Lande, D.[1]; Shchutsky, V.[2]; Shnurko-Tabakova[2], E.; Strashnoy, L.[3]

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

[2] Index Systems ltd., Kyiv, Ukraine

[3] UCLA, Institute Infectious disease department, USA

Annotation. The method of predicting events based on correlation mapping and Fourier decomposition of relevant time series, which is presented in this paper, is intended for the predictive calculation of signs that an event will occur. As input data for calculating the forecast, in addition to the series that needs to be extended with forecast values, relevant series that are close in the content are used. The calculation is based on a model that calculates the correlations of the total values of the Fourier decomposition for all relevant time series with the historical series. In this case, a certain number of constant parameters of the Fourier decomposition is calculated as the result of solving the optimization problem. Based on the generated model and the values of constant parameters, further values of the initial series are calculated, that is, the predicted event is determined.

Keywords. Event forecasting, Relevant time series, Fourier decomposition, Correlation of time series, Correlation mapping

Introduction

The problem of event forecasting (a partial case of time series forecasting) consists in finding the values of a series corresponding to an event in the future, based on known values of a series from the past, that is, a historical series.

While the prediction of stationary time series is a practically solved problem, the prediction of non-stationary time series is currently a current problem [1-3].

This paper presents a method for predicting events based on correlation reflection and Fourier decomposition [2], designed for the predictive calculation of signs of an event that occurred, binary values of 1 ("yes") or 0 ("no"). The proposed method solves the problem of forecasting events that are a special case of time series forecasting, and this is a scientific and practical novelty [4]. In addition, the series under consideration are not stationary.

As input data for the calculation, in addition to the original historical series, which must be extended with forecast values, additional so-called relevant series of events that are close in the content are used (the number of such events and their relevance level is determined by expert analysis).

The calculation is based on a model that calculates the correlations of the total values of the Fourier decomposition for all relevant time series with the historical series. In this case, a certain number of constant parameters of the Fourier decomposition is calculated as the result of solving the optimization problem. Here, the target function is the correlation coefficient between the known values of the historical series and the series of total values of the Fourier decomposition. After that, based on the generated model and the values of constant parameters, further values of the historical series are calculated, that is, the predicted event is determined at the time points determined by the constructed model (the minimum value of the model).

Below are the steps to implement the model.

Implementation of the methodology

The preliminary stage for implementing the methodology is to prepare data for further forecasting. To do this, prepare the following rows:

Y – a set of historical data that requires forecasting;

X_i – relevant series to the original series, whose behavior should be predicted (their number and relevance are determined by the analyst);

t – time (a uniform increasing sequence of digits that can represent minutes or hours, days, weeks, etc.).

As an example of relevant time series for such events (a series of Y dynamic time series of publications obtained by the Attack Index content monitoring system [5]

for the following queries can be considered as massive Russian missile strikes on Ukraine:

1. "country, Putin, war "(corresponds to the theme of the war associated with Putin, Figure 1);
2. "war, SWO" (corresponds to the theme of war as a so-called special military operation of Russia, Figure 2);
3. "(SWO) & (humanitarian, electrical, energy, heat)" (corresponds to the theme of a special military operation and attacks on the energy sector, Figure 3);
4. "(SWO)&(Ukraine, regime)" (corresponds to the theme of the special military operation and the political regime in Ukraine, Figure 4).

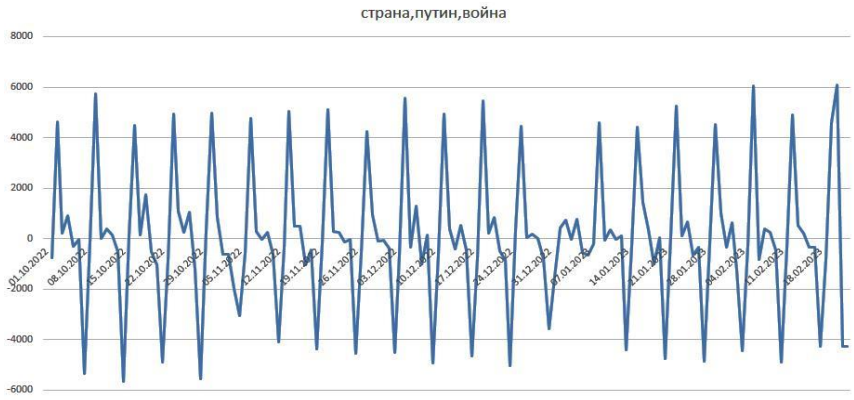

Figure 1. Time series of publications on websites on request 1

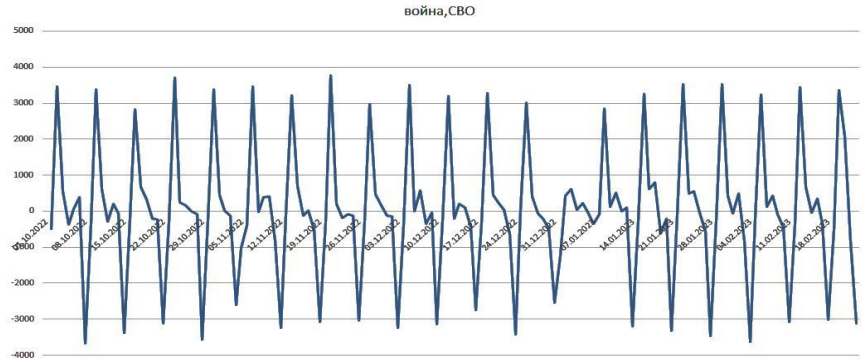

Figure 2. Time series of publications on websites on request 2

Figure 3. Time series of publications on websites on request 3

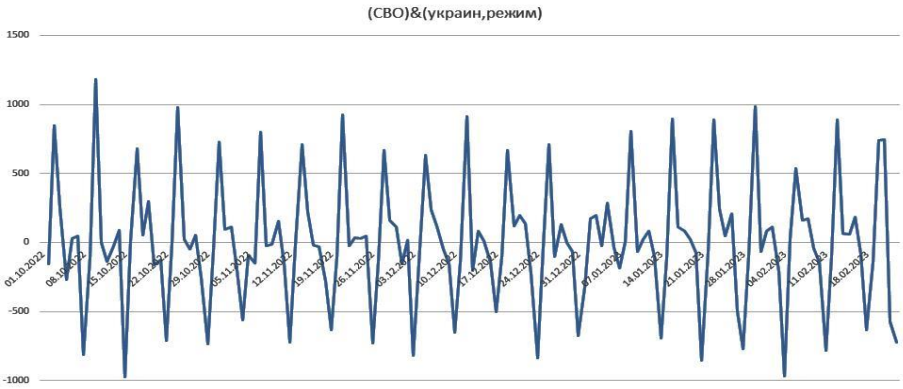

Figure 4. Time series of publications on websites on request 4

Step 1

Pairwise element differences are determined for all relevant series X_i (i – row number), like:

$$\Delta X_{ij} = X_{ij} - X_{i(j-1)} ,$$

where j – time value.

Step 2

Correlograms between Y and rows of ΔX_i, by calculating the correlation coefficients when the series is shifted Y with respect to the series ΔX_i to the set time value (Figure 5).

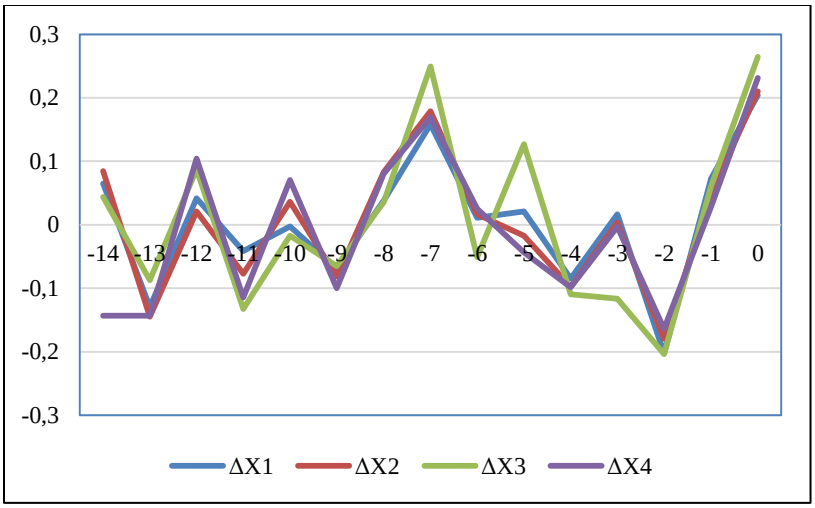

Figure 5. Correlograms

Step 3

Based on the required forecast horizon, the lag shifts that correspond to the largest absolute (without taking into account the sign) correlation values of the corresponding series are determined by expert analysis from the obtained correlograms.

Step 4

Before each lag shift, the formula is applied:

$$S_{ij} = a_1 \times \Delta X_{i(j-l))} \times \sin\left(a_2 \times t - a_3\right) + a_4 + a_5 \times \Delta X_{i(j-l))} \times \cos\left(a_6 \times t - a_7\right) + a_8 \ ,$$

where

a_1, a_2… – constant parameters, the value of which is calculated later (the initial value depends on the values of the series, for this example = 0.5);

l – lag shift.

Step 5

The intermediate result is defined as the sum of

$$S_j = \sum_i S_{ij} \ .$$

Step 6

Constant parameters are calculated using one of the methods of local or global optimization of the objective function chosen by the analyst (gradient methods, evolutionary methods, etc.). The target function is the correlation coefficient between known values of the series *Y* and *S*.

For the example under consideration, we obtain a correlation coefficient of 0.93. An example of the form of an optimized model is shown in Figure 6:

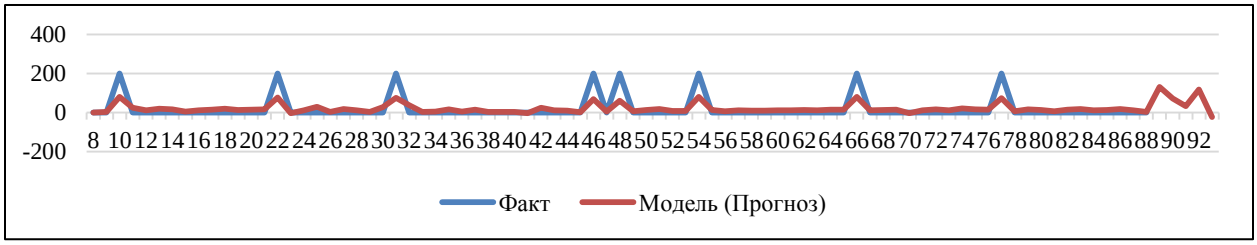

Figure 6. The shape of the optimized model

Step 7

The final solution to the problem is selected using the decision–making method-minimin (minimum probability of omission with minimum probability of an erroneous event).

To do this, determine the minimum value of the series *S (model range)*, at the moment when the tracked events appear in a number of *Y (the series that is being predicted)*.

For the given example, the minimum value of the row is *S*, at the moment when the tracked event appears in row *Y* equal to 60.1628.

Thus, the decision threshold is defined as 60.

Step 8

Based on the determined model, further values of the *Y* series are calculated, with the prediction of the event determined at the moments when the *S* series values will correspond to the minimax criterion (Figure 7).

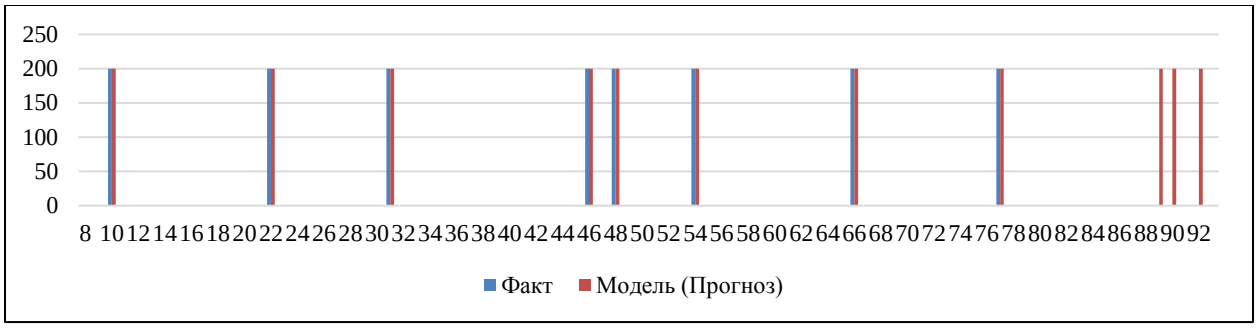

Figure 7. A series of events – massive Russian missile strikes on Ukraine

Conclusions

The proposed method solves the problem of event forecasting, which is a special case of time series forecasting, and this is a scientific and practical novelty.

The presented method is based on correlation reflection and Fourier decomposition of relevant time series and is intended for predictive calculation of signs that an event has occurred, binary values of 1 ("yes") or 0 ("no").

In addition to the historical time series that needs to be extended with forecasted values, additional relevant time series with similar content of events are used as input data for the calculation. The calculation is based on the model constructed using the proposed methodology, in which correlations of the total values of Fourier decomposition for all relevant time series with the historical series are calculated.

After that, further values of the historical series are calculated based on a certain model and values of constant parameters, that is, the predicted event is determined at the time points determined by the constructed model.

Bibliography

1. G. F. Zebende and E. F. Guedes. Detrended Correlogram Method for Non-Stationary Time-Series Analysis. Fluctuation and Noise Letters. – Vol. 21, No. 02, 2250012 (2022). DOI: 10.1142/S0219477522500122.
2. Wei Zhou, Zhongren Feng, Y.F. Xu, Xiongjiang Wang, Hao Lu. Empirical Fourier decomposition: An accurate signal decomposition method for nonlinear and non-stationary time series analysis. Mechanical Systems and

Signal Processing (2022). –Vol. 163, 108155. DOI: 10.1016/j.ymssp.2021.108155.

3. Dmytro Lande, Volodymyr Yuzefovych, Yevheniia Tsybulska. Linguistic Approach to Time Series Forecasting. E-preprint ArXiv arxiv.org/abs/2207.00985. DOI: 10.48550/arXiv.2207.00985.
4. Gurina, E., Klyuchnikov, N., Antipova, K. et al. Forecasting the abnormal events at well drilling with machine learning. Appl Intell 52, 9980–9995 (2022). DOI: 10.1007/s10489-021-03013-x
5. Dmytro Lande, Ellina Shnurko-Tabakova. OSINT as a part of the cyber defense system. Theoretical and Applied Cybersecurity, 2019. - N. 1. - pp. 103-108. DOI: 10.20535/tacs.2664-29132019.1.169091.