UDC 004.89

G.S. Ivaschenko, N.M. Korablev

TIMESERIESFORECASTINGONTHEBASISOFTHECASE-BASEDREASONINGUSINGTHEMODELSOFARTIFICIALIMMUNESYSTEMS

Abstract. The article describes the immune algorithms of short-term forecasting of time series based on the clonal selection model and the immune network model using the case-based reasoning. The clonal selection model uses heterogeneous antibodies on the basis of the case-based reasoning and the simplest forecasting methods. Evaluation of the effectiveness of the models is carried out by the comparative analysis; the results of experimental studies that demonstrate the features of the proposed approaches are presented.

Keywords: forecasting, time series, case-based reasoning, artificial immune systems, antibody, antigen, affinity, cloning, mutation.

Introduction

The problem of short-term forecasting of time series plays an important role in the decision-making process in technical and economic systems. Forecasting future values of a time series allows solution of an actual problem of determining the future state based on the analysis of the already existing retrospective data.

Approaches based on the artificial intelligence methods, such as artificial neural networks and artificial immune systems (AIS) are currently actively developing [1-4]. There are various models based on the principles of immune system: clonal selection model, immune network model and others that can be used to solve the problems of short-term time series forecasting [2-4] and can be integrated with other approaches, which allows compensation for the shortcomings of some models by using the features of the others [4]. Promising is the use of the case-based reasoning (CBR), where, when considering a new problem of forecasting, a similar case is sought in history as an analogue [5].

Thepurposeofthestudyliesinthedevelopmentofimmunealgorithmsof short-term forecasting basedonthemodelsofclonalselectionandofartificialimmunenetworkusingthecase-basedreasoning.

Problem statement

This paper considers discrete time series, the values of which are obtained at times $t_1, t_2, t_3, \ldots, t_N$. Let us denote time series $Z = z(t_1), z(t_2), z(t_3), \ldots, z(t_N)$ of

length N as $Z_1^N = z_1, z_2, z_3, \ldots, z_N$. Let us call a set of consecutive values $Z_t^L = z_t, z_{t+1}, z_{t+2}, \ldots, z_{t+L-1}$, within the time series Z_1^N, a sample of this series of length L, with a reference time t, $L \in \overline{[1, N-1]}$, $t \in \overline{[1, N-L]}$. The problem of forecasting a time series is the estimation of future values by its known section.

CBR-forecasting is based on the hypothesis proposed in [6]: if the similarity measure between the samples of the values of the forecasting series Z_t^L and Z_{t-k}^L has a value close to one, then the similarity measure between the samples of length P following them, Z_{t+L}^P and Z_{t-k+L}^P, is also close to one. I.e. by determining the sample, which is the most relevant to the latest known values of a time series, it is possible to estimate its future values.

However, case-based forecasting has a number of shortcomings, the main of which include the requirement for the number of the known values of a time series and the assumption that the past patterns will be the same in the future. To compensate for these shortcomings, one needs to use other forecasting methods, applying the CBR for the segmentation of the original time series. The problem of selecting an appropriate case from those existing in the database and its adaptation to the current conditions is one of the most important in the CBR-systems [5].

To solve this problem, this paper proposes to use the models of artificial immune network and of clonal selection.

Immune network model using the case-based reasoning

Case-based reasoning is a method of analyzing data to make conclusions about the situation on the search results of analogies stored in the database. This process includes steps such as the selection of a set of cases from those available in the database based on a given similarity relation, adaptation of the selected cases in relation to current conditions, comparison of the obtained and real values of the forecasting value, and saving in the database for future use of the taken decision and the current situation as a new case or a corresponding change in the previously selected case [5]. When solving the problem of forecasting using the CBR, a case is a combination of the problem (a sample of known values of a series) and the decision taken earlier (forecast).

To construct an AIS-based forecasting model, it is necessary to compare the biological objects and processes with their analogues from the domain. Antibodies are known values of a time series. Antigens are known values of a time series immediately preceding the forecasting one. Affinity of antibodies (primary criterion for the

selection in the immune network algorithm) is a scalar value determining the measure of closeness between an antibody and an antigen. In this approach, the affinity is determined not for a separate antigen-antibody pair, but between the antigens and antibodies tuples of length L – multi-antibody $mAb = ab_1, ab_2, ab_3, \ldots ab_{L+f}$ andmulti-antigen $mAg = ag_1, ag_2, ag_3, \ldots ag_L$. Multi-antibody also includes f forecasting values – this part is not involved in the determination of affinity. The number of antibodies included in it corresponds to the forecasting horizon.

In terms of the CBR approach, multi-antibody plays the role of a case, describing the current situation (the sequence of known values of a series) and the decision taken earlier (the relevant forecast). Thus, to estimate future values of a series, one needs to find multi-antibodies with the greatest affinity – the most relevant for the current problem.

Affinity of a multi-antibody with length L is determined by the formula:

$$Aff = \frac{\sum_{i=1}^{L} (1 + d_i)^{-1}}{L} \in (0,1], \quad (1)$$

where $d_i = |ab_i - ag_i|$ –distancebetweenthepairsofthevaluesconstituting multi-antibody mAb and multi-antigen mAg.

The problem of selecting a suitable case from those existing in the database and its adaptation to the current conditions is one of the most important in the CBR-systems. These steps in the AIS-based forecasting model correspond to the selection of multi-antibodies with the highest affinity from those present in the population, and the subsequent AIS training. The feature of the immune network model is the relationship between individual antibodies – correction in the process of AIS training of antibodies belonging to one multi-antibody affects others, close to it.

Clonal selection model using the case-based reasoning

Clonal selection algorithm operates on fixed-length data lines and is often used for solving the problems of classification, identification and optimization [1]. Each data line of the algorithm called an antibody (in terms of the CBR – case) represents a set of parameters describing the set problem (a set of known values of a series) and the decision taken (proposed forecast option): $Ab_i = ab_1, ab_2, ab_3, \ldots, ab_L, \ldots, ab_{L+f}$, where i – index in the population, L – length of the sample of known values of a

series, f – value of the forecasting horizon (length of a sample of forecasting values of a series).

Antigen is the set problem – sample of known values of a series immediately preceding the forecasting values: $Ag_j = ag_1, ag_2, ag_3, \ldots, ag_L$, where j – index in the population, L – length of the sample. I.e. an antigen is a section of a series, for which it is necessary to make a forecast.

Affinity between an antibody and an antigen is determined as

$$Aff = \eta * \frac{\sum_{k=1}^{L} \left(1 + \left|ab_k - ag_k\right|\right)^{-1}}{L} \in (0,1], \quad (2)$$

where η – value of the selection coefficient for determining the priority of the various types of antibodies. Values of antibody $ab_{L+1}, \ldots, ab_{L+f}$, the affinity of which $Aff(Ab) \to 1$, are taken as the forecasting ones.

Formally, AIS model based on clonal selection can be represented as follows:

$$AIS\_ClonAlg = \left\langle Ab, n', n_c, Ab^C, Ab^M, Ag, D, \sigma_d, S, \sigma_s, \sigma_{age}, \eta \right\rangle, \quad (3)$$

where Ab – population of the antibodies of size n; n' – number of the cells selected for cloning and mutation; n_c – number of the clones created by one antibody; Ab^C – population of clones; Ab^M – population of memory cells; Ag – population of the antigens of size m; D – matrix of affinities between antigens and antibodies; σ_d – threshold coefficient of cell stimulation; S – matrix of affinities between antibodies; σ_s – threshold compression coefficient; σ_{age} – threshold age of antibodies; η – set of selection coefficients.

Clonal selection model allows combining the use of the CBR and other forecasting methods, performing segmentation of a series and selection of the most appropriate forecasting method for each section of the original series during the obtainment of a forecast. Therefore, in addition to the antibodies implementing the case-based reasoning, other types of antibodies are created based on the same fragments of the time series, forming its own version of the forecast, using the simplest forecasting methods, such as simple average forecasting, exponential average forecasting, naive models [5].

Algorithm for short-term forecasting

Algorithm to obtain a forecast includes the following steps:

1. Creation of the initial population of antibodies. To create antibodies, a part of known values of a time series is used. Unused values serve as a training and test samples.

2. Formation of an antigen (or a multi-antigen for the immune network model) based on the latest known values of the series preceding the forecasting ones.

3. Determination of Ab-Ag (mAb-mAg) affinities and selection of antibodies (or multi-antibodies for the immune network model) having an affinity above the threshold value.

4. Forecast of the antibody (multi-antibody) having the highest affinity is taken as a result.

5. If there is a real value of the forecasting value, forecast error is determined.

6. Cloning antibodies with the highest affinity (selected in step 3), during which the mutation operator is performed. In the immune network model clones replace multi-antibodies that gave rise to them.

7. Determination of affinities between the antibodies (multi-antibodies), applying the suppression operator with a view to eliminate redundancy. In the clonal selection model the remaining antibodies from the previously selected ones become memory cells.

8. Application of the aging operator and correction of the population.

The immune network model (also for antibodies in the clonal selection model implementing the case-based reasoning) uses directional proportional mutation – the most suitable case for the set problem undergoes substantial correction. Only a part of the antibody that determines its forecast and is not involved in the determination of affinity is exposed to mutation. For antibodies computing their own version of the forecast, only the first part is exposed to undirected mutation (and therefore, the proposed version of the forecast changes), which partially solves the problem of the lack of cases in the database.

After receiving the real values of the forecast AIS based on the clonal selection model performs a correction of selection coefficient η. The coefficient is incremented for that type of antibody, which showed the smallest prediction error.

The population of memory cells formed in the process of AIS training based on the model of clonal selection represents a set of patterns that are the most typical for this time series.

Changing of the population of antibodies in the current generation gen occurring as a result of applying operators of cloning, mutation and selection can be generally represented as follows:

$$Ab^{gen+1} = Edit\left(Mutate\left(Clone\left(Ab^{gen}\right)\right)\right), \quad (4)$$

where

$$Clone : Ab^{gen} \to Ab_{C}^{gen}; \quad (5)$$

$$Mutate : Ab_{C}^{gen} \to Ab_{MC}^{gen}; \quad (6)$$

$$Edit : (Ab_{MC}^{gen}, Ab^{gen}) \to Ab^{gen+1}. \quad (7)$$

The graph in fig. 1 shows the effect of the multi-antibody length value L on the accuracy of the obtained forecast.

AIS training is repeated for each antigen from the training sample a given number of times. A stop to achieve a certain number of generations, or a stop to achieve a predetermined error value is used as a criterion for stopping the immune algorithm. Fig. 2 shows the dependence of the mean absolute error (MAE) on the affinity for the untrained (series 1) and trained (series 2) AIS based on the immune network model. Thus, the purpose of AIS training is to configure the system in accordance with new values of a time series.

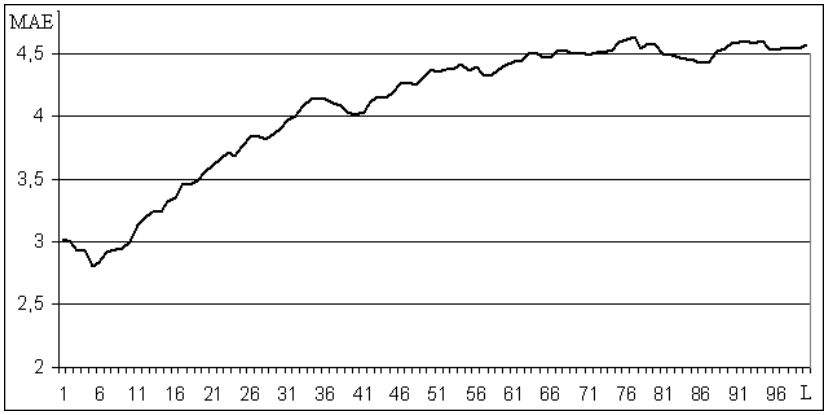

Fig.1 –Dependence of the mean forecast error on the multi-antibody length

 ISSN 1562-9945

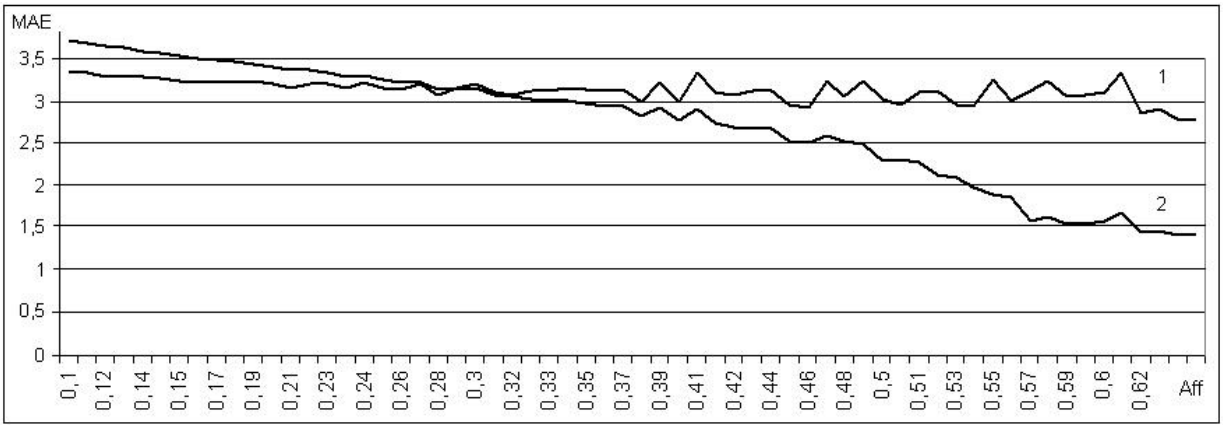

Fig.2 –Dependence of the error value on the affinity of multi-antibodies

Effectiveness of forecasting using the models considered is reduced when cases from the training sample created on the basis of similar patterns propose conflicting versions of the forecast. To solve this problem, an increase in the sample size is used, on the basis of which the first part of an antibody (multi-antibody) is built. Since the nature of the series can change over time, it is advisable to maintain sets of antibodies of various lengths in the population.

Results of the comparative analysis

Experimentalstudies included short-term forecasting for the series used in M3-Competition [7] and the series of average daily temperature readings. Forecastingresults by various methods are shown in table 1. To evaluate the prognosis,the value of the symmetrical mean absolute error (SMAPE) was used; the results of the use of the forecasting methods different from AINet ClonAlg areobtained from [7]. To estimate the Meteo time series prediction was used mean absolute error (MAE).

Comparing the results of the model based on the artificial immune network (AINet) and the model based on the clonal selection (ClonAlg), it may be noted that although the results are close for the series of temperatures (Meteo), the use of clonal selection allows better forecasting for the series with smaller background, i.e. this approach is less demanding to the number of values of the time series.

Table 1

Symmetrical mean absolute error (SMAPE, %) for different forecasting methods

Method	N704 (44)	N736 (44)	N1366 (63)	N2830 (104)	N2841 (104)	N2867 (79)	Meteo (21337)
Exp.Smooth	4,08	12,11	0,42	2,47	0,5	20,52	4,56
RobustTrend	4,76	8,88	0,41	2,18	0,52	18,83	–

HoltWinters	4,92	10,68	1,04	3,27	0,39	20,08	2,9
CombSHD	4,36	9,65	0,5	2,73	0,46	20,09	–
Box–Jenkins	3,66	7,35	0,57	2,45	0,5	26,13	2,99
ForecastPro	3,13	6,5	0,41	2,47	0,5	20,52	–
SmartFcs	5,18	8,62	0,29	2,47	0,5	23,85	–
AutoANN	3,42	8,73	0,27	1,56	0,53	26,11	–
AINet	5,36	9,71	0,46	2,37	0,68	18,93	2,45
ClonAlg	5,20	7,65	0,41	1,83	0,14	16,26	2,44

The result of forecasting a series of average daily temperature readings (Meteo) confirms the advantages of the AIS-based approaches using the CBR – sizes of the Meteo series allow creation of a population of antibodies that corresponds to the vast database of cases, and carrying out the training. However, with a small number of values of a time series, the absence of antibodies with high affinity in the database (there is no suitable case) is likely in the preparation of a forecast, and the value of a training sample will not allow full configuration (training) of the artificial immune system. As a result, in some series presented in the table the traditional methods have advantages over the considered approaches.

Conclusions

This paper proposes a solution to the problem of short-term forecasting of time series by the case-based reasoning, using the models of artificial immune systems, such as the immune network model and the clonal selection model using heterogeneous antibodies.

The main features of the models considered are as follows:

- use of a population of antibodies as a database of cases;
- the possibility to configure the size of multi-antibodies in accordance with the nature of the series;
- antibodies in the model of clonal selection implement various forecasting methods; segmentation of a series is performed in the process of the AIS training to determine the most appropriate predictor for each section of the original series;
- formation of a plurality of patterns (population of memory cells) in the process of analyzing time series in the AIS training based on the model of clonal selection;
- the possibility to determine abnormal emissions in time series using the population of memory cells formed in the process of training;

• the possibility to extend the models to account for the influence of external factors presented in the form of other time series.

The results presented in the paper confirm the efficiency of the use of the considered approaches for short-term forecasting of time series, but in a lack of baseline information it is preferable to use the approach based on the clonal selection model, as this approach allows getting a better result with a small training sample.

REFERENCES

1. Dasgupta D. Artificial Immune Systems and Their Applications / D. Dasgupta. – Springer-Verlag, 1999. – 306 p. DOI: 10.1007/978-3-642-59901-9
2. Bidjuk P. I. Algorithm Clonal Selection for Forecasting Behaviour of Non-Stationary Dynamic Systems / P.I. Bidjuk, V.I. Litvinenko, A.A. Fefelov, I.V. Baklan // Iskusstvennyiintellekt. – 2004. – No. 4. – P. 89–99 (inRussian).
3. KorablevN. M. Applicationofartificialimmunenetworksfortimeseriesforecasting / N. M. Korablev, G. S. Ivashchenko // Sistemiobrobkiinformatsii. – Kharkov, 2012. – No. 9. – P. 42–45. (inRussian).
4. KorablevN. M. Modelofclonalselectionforforecastingtimeserieswithmissingdata / N. M. Korablev, G. S. Ivashchenko // Elektrotekhnicheskieikomp'yuternyesistemy. – Odessa, 2014. – No. 13(89). – P. 170–177 (inRussian).
5. Singh S. Pattern Modelling in Time-Series Forecasting / S. Singh // Cybernetics and Systems – An International Journal. – 2000. – No. 1(31). – P. 49–65.
6. Chuchueva I. A. Extrapolation Model of Time Series by the Selection of Maximum Likeness / I. A. Chuchueva // Informatsionnye tekhnologii. – 2010. – No. 12. – P. 43–47 (in Russian).
7. Makridakis S. The M-3 Competition: Results, Conclusions and Implications / S. Makridakis, M. Hibon // International of Forecasting. – Amsterdam: Elsevier, 2000. – No. 16. – P. 451–476. DOI: 10.1016/S0169-2070(00)00057-1