Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 103 (2017) 214 – 221

Procedia
Computer Science

XIIth International Symposium «Intelligent Systems», INTELS'16, 5-7 October 2016, Moscow, Russia

Attractors in the development problem of the forecasting model on the base of the strictly binary trees

Liliya Demidova[a,b]*, Zhdankina Elena[a], Krylova Olga[a], Reshetova Vera[a], Sofyin Evgeny[a]

[a]*Ryazan State Radio Engineering University, Gagarina Str., 59/1, Ryazan, 390005, Russia*
[b]*Moscow Technological Institute, Leninskiy pr., 38A, Moscow, 119334, Russia*

Abstract

In this paper the questions of the choice of the training data sequence for the forecasting model on the base of the strictly binary trees which is usually applied to time series with a short length of actual part (about 15 – 30 elements) have been considered. It is offered to use the principles of the attractors' forming in the presence of the long time series that will allow creating the training data sequence more reasonably. The reviewed examples confirm the efficiency of the attractors' use in sense of minimization of the affinity indicator of the forecasting model, and also the forecasting errors on 1 – 5 steps forward. Besides, the minimization of time expenditures on development of the forecasting model is provided.

Keywords: time series; forecasting model; strictly binary tree; modified clonal selection algorithm; attractor.

---

1. Introduction

Usually the forecasting model on the base of the strictly binary trees is developed with use of the short training data sequence which length constitutes 15-30 elements. In the presence of the long time series often it is impossible to allocate reasonably at once that part of the data sequence which shall be used as training. At the same time application of the long training data sequences will be followed by increase in time expenditures at receipt of the

---

* Corresponding author. Tel.: +7-491-246-0303; fax: +7-491-292-2215.
E-mail address: demidova.liliya@gmail.com

doi:10.1016/j.procs.2017.01.068

required forecasting model, and also can lead to receive of not quite adequate forecasting model which will yield acceptable results of training, but bad results of forecasting.

In this regard it is possible to draw a conclusion on relevance of the choice problem of the training data sequence length.

In this article for the solution of this problem it is offered to use the approach on the base of attractors, an assessment of their length and decision making about a possibility of use of the attractor (if it is single) or the last of attractors (if there are several attractors) as the training data sequence.

The rest of this paper is structured as follows. Section II describes the main ideas of the development of the forecasting models on the base of the strictly binary trees (SBT) and the modified clonal selection algorithm (MCSA). Section III details the approach to the choice of the training data sequence for the forecasting model on the base of attractors. The experimental results which clearly highlight the efficiency of the offered approach to the choice of the training data sequence for the forecasting model follow in Section IV. Finally, conclusions are drawn in Section V.

2. Forecasting model on the base of the strictly binary trees

The MCSA simulates the natural laws of the immune system functioning[1,2] and provides the formation of quite complex analytical functions. The principles of development for the forecasting models with the use of the MCSA were investigated in[2]. The MCSA allows forming an analytical dependence on the base of the SBT at an acceptable time expenses, that describes certain TS values and provides a minimum value of the average forecasting error rate (*AFER*):

$$AFER = \frac{\sum_{j=k+1}^{n} \left| \left(f^{j} - d^{j} \right) / d^{j} \right|}{n-k} 100\% \tag{1}$$

where d^j and f^j are respectively the actual (fact) and forecasted values for the j-th element of the TS; n is the number of TS elements.

Also the *AFER* indicator (1) can be named as the affinity indicator.

Possible options for analytical dependences are presented in the form of antibodies Ab which recognize antigens Ag (certain TS values). An antibody Ab is selected as «the best one». It provides the minimum value of the affinity indicator Aff [1,2]. Coding of an antibody Ab is carried out by recording signs in a line. The signs are selected from three alphabets[1, 2]:

- the alphabet of arithmetic operations (addition, subtraction, multiplication and division) – $Operation = \{'+','-','\cdot','/'\}$;
- the functional alphabet, where letters $\{'S','C','Q','L','E'\}$ define mathematical functions «sine», «cosine», «square root», «natural logarithm», «exhibitor», and the sign $'\_'$ means the absence of any mathematical function, – $Functional = \{'S','C','Q','L','E','\_'\}$;
- the alphabet of terminals, where letters $'a','b',...,'z'$ define the arguments required analytical dependence and the sign $'?'$ defines a constant, $Terminal = \{'a','b',...,'z','?'\}$.

The use of these three signs alphabets provides a correct conversion of randomly generated antibodies into the analytical dependence. The structure of such antibodies can be described with the help of SBT[1,2].

The number of signs in the alphabet of terminals $Terminal$ in the antibody Ab determines maximal possible order K of the forecasting models with $K \geq k$, where k is the real model order), i.e. having the value of the element d^j in the forecasting TS at the j-th moment of time, K values of TS elements can be used as: d^{j-K},…, d^{j-2}, d^{j-1} [1,2,3,4,5,6].

The use of the SBT type, illustrated in Fig. 1, allows building the complex analytical dependence and provides high accuracy of the forecasting TS[2].

Such SBT can be generated as a composition result of one «left» subtree of the maximum possible order $K = 3$ and