Multiobjective Optimization for the Forecasting Models on the Base of the Strictly Binary Trees

Nadezhda Astakhova

Ryazan State Radio Engineering
University
Ryazan, Russia

Liliya Demidova

Moscow Technological Institute
Ryazan State Radio Engineering
University
Moscow, Russia

Evgeny Nikulchev

Moscow Technological Institute
Moscow,
Russia

Abstract—The optimization problem dealing with the development of the forecasting models on the base of strictly binary trees has been considered. The aim of paper is the comparative analysis of two optimization variants which are applied for the development of the forecasting models. Herewith the first optimization variant assumes the application of one quality indicator of the forecasting model named as the affinity indicator and the second variant realizes the application of two quality indicators of the forecasting model named as the affinity indicator and the tendencies discrepancy indicator. In both optimization variants the search of the best forecasting models is carried out by means of application of the modified clonal selection algorithm. To obtain the high variety of population of the forecasting models it is offered to consider values of the crowding-distance at the realization of the second optimization variant. The results of experimental studies confirming the use efficiency of the modified clonal selection algorithm on the base of the second optimization variant are given.

Keywords—forecasting model; strictly binary tree; modified clonal selection algorithm; multiobjective optimization; affinity indicator; tendencies discrepancy indicator

I. INTRODUCTION

The main problem dealing with the development of the forecasting models is the problem of the right choice of the best forecasting model. The forecasting model based on the strict binary trees (SBT) and the modified clonal selection algorithm (MCSA) [1, 2] is presented in the form of antibody, which is coded by a line of symbols randomly selected from the corresponding alphabets. This antibody can be transformed to the analytical dependence, which is used for forecasting of a time series (TS). Obviously, the correct selection of antibodies is very important for the effective use of the MCSA [1 – 6].

The traditional approach in the short-term forecasting models choice consists in the quality estimation of the forecasting models by means of the average forecasting error rate (*AFER*), calculated for the training data sequence. Herewith the *AFER* should be minimized [1 – 6]. However, the use of the *AFER* as the unique quality indicator of the forecasting model is not always sufficient to determine the best forecasting model. Often it is required to consider the additional quality indicators of the forecasting model, such as the compliance to the seasonal tendencies of TS, the compliance to the trend of TS, lack of emissions, complexity of the forecasting model, etc. [6]. It is expedient to use the additional quality indicator, which will allow estimating the

general tendency of values' change of the known elements of TS (for example, the tendencies discrepancy indicator) along with the *AFER* [6]. It is possible to increase the efficiency of the forecasting models on the base of the SBT, using the multiobjective MCSA at the solution of the problem of the medium-term forecasting. Herewith the affinity indicator based on the AFER and the tendencies discrepancy indicator can be used in the role of the objective functions.

The rest of this paper is structured as follows. Section 2 presents the main ideas of the original MCSA. Section 3 details the multiobjective optimization variant for the MCSA. Experimental results comparing two optimization variants (with the original MCSA and with the multiobjective MCSA) follow in Section 4. Finally, conclusions are drawn in Section 5.

II. THE MAIN IDEAS OF THE MODIFIED CLONAL SELECTION ALGORITHM

The MCSA simulates the natural laws of the immune system functioning and provides the formation of quite complex analytical functions [1], [2]. The principles of developing forecasting models of k-order 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 affinity indicator *Aff* based on the *AFER*:

$$AFER = (100\%/(n-k)) \cdot \sum_{j=k+1}^{n} |(f^{\,j} - d^{\,j})/d^{\,j}| \qquad (1)$$

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.

The possible variants for analytical dependences are presented in the form of antibodies Ab, which recognize antigens Ag (the TS values). The antibody Ab is selected as "the best one". It provides the minimum value of the affinity indicator *Aff*. The antibody coding is carried out by recording signs in a line. These signs are selected from three alphabets: the alphabet of arithmetic operations (addition, subtraction, multiplication and division) $Operation = \{'+','-','\cdot','/'\}$; the functional alphabet $Functional = \{'S','C','Q','L','E','\_'\}$,

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, the alphabet of terminals $Terminal = \{'a', 'b', ..., 'z', '?'\}$, where letters $'a', 'b', ..., 'z'$ define the arguments required analytical dependence and the sign $'?'$ defines a constant. The use of these alphabets provides a correct conversion of randomly generated antibodies into the analytical dependence. The structure of these antibodies can be described with the help of the SBT. The number of signs in the alphabet of terminals $Terminal$ in the antibody Ab determines the maximal possible order K of the 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 the TS elements can be used as: $d^{j-K}, ..., d^{j-2}, d^{j-1}$ [1], [2].

The use of the SBT type, illustrated in Fig. 1, allows building the complex analytical dependence and provides high accuracy of the forecasting TS [1], [2].
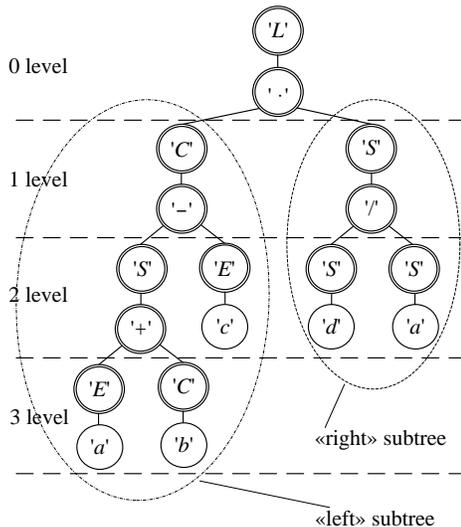
Fig. 1.   An example of a strict binary tree, which is used to form antibodies

This SBT can be created as the composition of one "left" subtree of the maximum possible order $K = 3$ and some "right" subtrees of the maximum possible order $K = 2$. The term "left" ("right") subtree is used for the branch (left or right) of the SBT level where a new subtree should be included. It is rational to form antibodies by subdividing SBT into subtrees, then execute the subtree-walk of each vertex forming the ordered symbol lists on its vertices and then combining these lists consistently. Forming the symbol ordered list on the base of a subtree the consecutive double subtree-walk is carried out: at first moving the subtree bottom-up left to right it is necessary to bypass the vertices containing the alphabetic terminal signs $Terminal$ in pairs and correspondingly above placed vertices containing the alphabetic functional symbols $Functional$ and then moving in the same direction it is necessary to go around in pairs the vertices containing the alphabetic arithmetic operation signs $Operation$ and

correspondingly above placed vertices containing the alphabetic functional signs $Functional$. The first two signs of such antibody contain the pair of zero level SBT from the functional alphabet $Functional$ and arithmetic operation alphabet $Operation$. Then there are the lists of the signs corresponding to the "right" maximum possible ordered subtrees $K = 2$ (moving the SBT bottom-up) and finally the symbol list of the «left» maximum possible ordered subtree $K = 3$.

For example, the antibody formed on the base of the SBT as shown in Figure 1 is coded by the line of signs: $L \cdot S/SeSdC - S + EaCbEa$, which can be transformed into the analytical dependence for the forecasting model with $k = 4$:

$$f(d^{j-1}, d^{j-2}, d^{j-3}, d^{j-4}) = ln(cos(sin(exp(d^{j-1}) + \\ + cos(d^{j-2})) - exp(d^{j-3})) \cdot sin(sin(d^{j-4})/sin(d^{j-1})).$$

Interpreting the antibodies into the analytical dependences it is rational to use the recursive procedure of interpretation [2]. The MCSA applied to the searching for "the best" antibody defining "the best" analytic dependence includes the preparatory part (realizes the formation of the initial antibody population) and iterative part (presupposes the ascending antibodies ordering of affinity Aff the selection and cloning the part of "the best" antibodies, that are characterized by the least affine value Aff the hypermutation of the antibodies clones; self-destruction of the antibodies clones "similar" to the other clones and antibodies of the current population; calculating the affinity of the antibodies clones and forming the new antibodies population; suppression of the population received; generation of the new antibodies and adding them to the current population until the ingoing size; the conditional test of the MCSA completion).

III.   MULTIOBJECTIVE OPTIMIZATION

The average forecasting error rate $AFER$, which is also called the affinity indicator Aff (in the context of working with the MSCA) can be used as the first quality indicator for the forecasting models. The rate of discrepancy between the tendencies of two time series (the tendencies discrepancy indicator $Tendency$) can be used as the second quality indicator for the forecasting models [6].

The tendencies discrepancy indicator $Tendency$ can be calculated as:

$$Tendency = h/(n - r - 1), \qquad (2)$$

where h is the number of negative multiplications $(f^{j-1} - f^j) \cdot (d^{j-1} - d^j)$; $j = \overline{r+2, n}$; d^j and f^j are respectively the actual (fact) and forecasted values for the j-th element of TS; n is the number of TS elements; r is the model order; $n - r - 1$ is the total number of multiplications $(f^{j-1} - f^j) \cdot (d^{j-1} - d^j)$.

This indicator allows adapting the forecasting models on the base of the SBT and MCSA for the medium-term forecasting.

The affinity indicator (1) and the tendencies discrepancy indicator (2) must be used simultaneously at the quality assessment of the forecasting models on the base of the SBT and MCSA to solve the problem of the medium-term forecasting.

Various well proved approaches can be applied to the solution of the problem of the simultaneous accounting of two quality indicators for the development of the forecasting models. Herewith it is necessary especially to allocate approach, based on the several multiobjective optimization algorithms, including, evolutionary algorithms. In recent years a number of multiobjective evolutionary algorithms (MOEA) have been suggested [7] – [15]. The main reason for this is their ability to find the multiple Pareto-optimal solutions in one single simulation run. These algorithms work with a population of solutions. Therefore, the primary attention has to be paid to maintaining the diversity and spread of solutions. Such MOEAs provide a solution of the account's problem of the several objective functions (quality indicators) at the analisys of various applied problems. The multiobjective genetic algorithms (MOGA) [7] – [11] are the most known algorithms of the multiobjective optimization. It is necessary to say about the multiobjective clonal selection algorithms (MOCSA) [12] – [15]. However, these algorithms are less designed and, in the majority, borrow the principles of multiobjective optimization underlain in the genetic algorithms. The possibility of this loan can be explained with many similar mechanisms of the evolutionary process realization in the MOGA and MOCSA. The analisys of merits and demerits of the MOEAs shows that such the MOGAs as the NSGA-II and the NSGA-III are significantly better than othes because they can successfully solve more difficult problems of the multiobiective optimization [6].

In this regard the decision on expediency of the adaptation of the ideas put in the NSGA-II at the realization of the multiobjective MCSA which is applied for the selection of the forecasting models on the base of the SBT had been made. Herewith, it is necessary to understand the forecasting model (and the antibody corresponding to it) as the decision, and the quality indicator of the forecasting model as the objective function at the realization of the multiobjective optimization algorithm. All forecasting models with use of the notion "Pareto-dominance" can be divided to dominated and nondominated models [6].

Let $Q_{s,v}$ be a value of the v-th quality indicator for the s-th forecasting model ($v = \overline{1,V}$; $s = \overline{1,S}$); let V be a quantity of the quality indicators of the forecasting model; let S be a quantity of the forecasting models. The s-th forecasting model is dominated by the z-th forecasting model ($s = \overline{1,S}$; $z = \overline{1,S}$), if the following conditions are satisfied: the s-th forecasting model is dominated by the z-th forecasting model, if the following conditions are satisfied for all quality indicators: $Q_{s,v} \geq Q_{z,v}$ ($v = \overline{1,V}$), and also there is at least one the v^*-th ($1 \leq v^* \leq V$) indicator for which the condition $Q_{s,v^*} > Q_{z,v^*}$ is satisfied. A herewith all quality indicators must be minimized. The rank R_s must be calculated for every s-th forecasting

model ($s = \overline{1,S}$). The rank R_s is equal to the quantity of the forecasting models which dominate over the s-th forecasting model. The rank R_s of the s-th nondominated forecasting model is equal to zero [6]. Let $V = 2$, $Q_{s,1} = Aff_s$, $Q_{s,2} = Tendency_s$, where Aff_s and $Tendency_s$ are the values of the affinity indicator (1) and the tendencies discrepancy indicator (2) for the s-th forecasting model ($s = \overline{1,S}$) accordingly. The s-th forecasting model is dominated by the z-th forecasting model ($s = \overline{1,S}$; $z = \overline{1,S}$), if the following conditions are satisfied: ($Q_{s,1} \geq Q_{z,1}$ and $Q_{s,2} > Q_{z,2}$) or ($Q_{s,1} > Q_{z,1}$ and $Q_{s,2} \geq Q_{z,2}$), that is ($Aff_s \geq Aff_z$ and $Tendency_s > Tendency_z$) or ($Aff_s > Aff_z$ and $Tendency_s \geq Tendency_z$) [6].

The crowing distances τ_s ($s = \overline{1,S}$) can be calculated using the following algorithm [10, 11].

Step 1. To calculate ranks for all forecasting models in the population. To unite the models with identical values of the rank into one group.

Step 2. For every group of the forecasting models:

- to sort the forecasting models according to each quality indicator value in ascending order of magnitude;

- to assign the infinite distance to boundary values of the forecasting models in the group, i.e. $\tau_1 = \infty$ and $\tau_{G_w} = \infty$, where G_w is the quantity of the forecasting models in the w-th group ($w = \overline{1,W}$); W is the groups' quantity; to assign $\tau_s = 0$ for $s = \overline{2, G_w - 1}$;

- to calculate the the crowding distance τ_s as:

$$\tau_s = \sum_{v=1}^{V} (Q_{s-1,v} - Q_{s+1,v})/(Q_v^{max} - Q_v^{min}), \qquad (3)$$

where $Q_{s-1,v}$ and $Q_{s+1,v}$ are the values of the v-th quality indicator ($v = \overline{1,V}$) for the forecasting models with the numbers ($s-1$) and ($s+1$), which are the nearest "neighbors" for the s-th model; Q_v^{min} and Q_v^{max} are the minimum and maximum values of the v-th quality indicator ($v = \overline{1,V}$) accordingly.

Fig. 2 shows how we can calculate the crowding distance on the base of two quality indicators. The points, marked with solid circles, correspond to the models with the minimum (zero) value of the rank (i.e. these points correspond to the Pareto front with the zero rank). To calculate the crowing distance for the s-th forecasting model it is required to define values of both quality indicators for the ($s-1$)-th and the ($s+1$)-th models, which are the nearest "neighbors" for the s-th model and have the same rank. Also, it is necessary to define the best and worst values of each quality indicator.
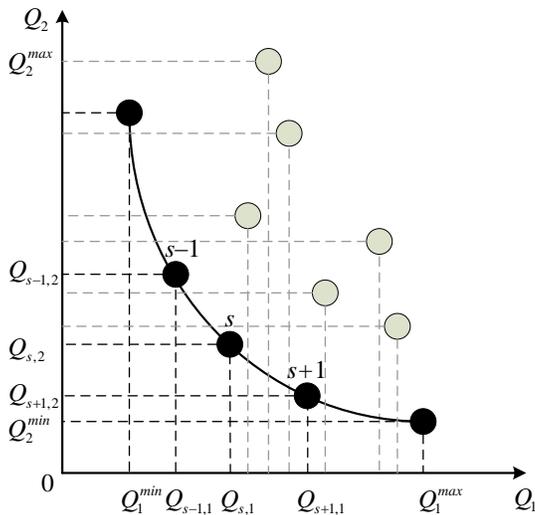
Fig. 2.   The points used for the crowding distance calculation An example of a strict binary tree, which is used to form antibodies

The crowing distances τ_s ($s = \overline{1, S}$) for the s -th forecasting model on the base of two quality indicators can be calculated as [6]:

$$\tau_s = (Q_{s-1,1} - Q_{s+1,1})/(Q_1^{max} - Q_1^{min}) + \qquad (4)$$
$$+ (Q_{s-1,2} - Q_{s+1,2})/(Q_2^{max} - Q_2^{min})$$

At the realization of the multiobjective MCSA the s -th forecasting model will be better than the z -th forecasting model, if:

$$R_s < R_z \text{ or } (R_s = R_z \text{ and } \tau_s > \tau_z).$$

If the s -th forecasting model is better than the z -th forecasting model, the s -th forecasting model is the candidate for transfer into the new generation.

For confirmation of prospects of the offered transformation of the MCSA it is offered to realize the following multiobjective optimization algorithm [6].

Step 1. To generate initial population of antibodies. Each antibody is coded on the base of the SBT and represents some forecasting model.

Step 2. To perform the nondominated sorting to population of antibodies on the base of two indicators of quality for the forecasting model (the affinity indicator (1) and the tendencies discrepancy indicator (2)).

Step 3. To choose the parents-antibodies for the next generation of the clones-antibodies based on the values of the rank and crowding distance.

Step 4. To pass to step 5 if desirable values of the quality indicators are reached or the quantity of generations in the MCSA is settled. Otherwise to pass to step 2.

Step 5. To accept the antibody with the minimum value of the affinity indicator (1) in the last population as the optimum decision. To use the forecasting model corresponding to this antibody for the forecasting.

As a result of application of the offered multiobjective clonal selection algorithm the Pareto set of the nondominated forecasting models will be received. These models provide the best combinations of values of the used quality indicators for the forecasting models.

IV.   EXPERIMENTAL STUDIES

Both variants of optimization have been applied for the development of the forecasting models intended for forecasting of the names' references' quantity of the E-Commerce systems in the requirements to vacancies posted on the websites of 2 famous recruiting network services – HeadHunter.ru (Russia) and Indeed.com (USA). The obtained forecasting results can be used for the analysis of tendencies of the labour market. Each of the analyzed time series contains information on the number of vacancies which include a specific keyword (Magento, OpenCart, PrestoShop, Hybris, Demandware). This keyword defines the name of E-Commerce system for development of online stores. Herewith 7 TSs have been considered: 4 TSs with information on vacancies in Russia and 3 TSs with information on vacancies in USA marked (in brackets after the name of a keyword) respectively as RF and USA:

Hybris (RF) (monitoring period 03.02.2016 – 06.04.2016, unit of measure is the number of references)=[55; 55; 57; 56; 56; 55; 56; 62; 62; 61; 59; 57; 57; 57; 57; 58; 58; 58; 58; 58; 58; 56; 56; 56; 57; 57; 57; 57; 57; 55; 54; 54; 53; 53; 54; 52; 51; 53; 53; 51; 51; 51; 54; 56; 58; 58; 58; 58; 58; 58; 60; 63; 63; 62; 63; 65; 67; 67; 67; 67; 66; 66; 69; 70];

Magento (RF) (monitoring period 03.02.2016 – 06.04.2016, unit of measure is the number of references)=[116; 128; 130; 132; 131; 126; 125; 120; 128; 130; 133; 130; 128; 124; 121; 120; 123; 123; 127; 129; 123; 127; 129; 131; 133; 133; 123; 123; 123; 123; 123; 120; 120; 118; 116; 117; 118; 129; 124; 120; 120; 127; 130; 131; 129; 129; 124; 124; 129; 133; 134; 132; 133; 133; 132; 135; 136; 136; 136; 136; 136; 142; 140; 147];

OpenCart (RF) (monitoring period 03.02.2016 – 06.04.2016, unit of measure is the number of references)=[84; 84; 82; 81; 80; 81; 89; 88; 89; 85; 85; 85; 88; 91; 93; 90; 88; 85; 90; 90; 91; 85; 83; 83; 77; 77; 77; 79; 79; 76; 72; 75; 73; 74; 76; 80; 75; 80; 80; 79; 86; 92; 93; 89; 90; 90; 90; 92; 94; 96; 98; 99; 99; 101; 115; 116; 116; 119; 120; 120; 112; 114; 114; 118];

PrestoShop (RF) (monitoring period 03.02.2016 – 06.04.2016, unit of measure is the number of references)=[32; 31; 31; 34; 32; 33; 33; 32; 34; 34; 32; 31; 33; 34; 34; 33; 33; 32; 32; 32; 32; 30; 30; 30; 29; 29; 29; 26; 26; 24; 23; 23; 22; 22; 22; 25; 24; 25; 25; 25; 28; 32; 30; 30; 29; 29; 29; 31; 31; 32; 31; 31; 31; 31; 31; 31; 29; 28; 28; 28; 28; 27; 27; 28];

Hybris (USA) (monitoring period 03.02.2016 – 06.04.2016, unit of measure is the number of references)=[674; 688; 677; 672; 664; 680; 690; 693; 710; 706; 697; 692; 688; 689; 676; 668; 663; 648; 643; 637; 635; 642; 641; 641; 629; 629; 629; 631; 631; 666; 671; 663; 667; 672; 711; 703; 662; 715; 715; 709; 690; 665; 668; 657; 662; 659; 659; 660; 657; 658; 656; 653; 653; 646; 643; 630; 631; 649; 647; 647; 652; 654; 662; 650];

Magento (USA) (monitoring period 03.02.2016 – 06.04.2016, unit of measure is the number of references)=[1093; 1102;

1082; 1076; 1076; 1077; 1087; 1095; 1080; 1073; 1072; 1070; 1073; 1086; 1103; 1110; 1118; 1110; 1098; 1107; 1114; 1133; 1126; 1126; 1124; 1124; 1124; 1134; 1134; 1131; 1135; 1125; 1111; 1114; 1138; 1137; 1125; 1155; 1155; 1145; 1103; 1006; 1023; 1013; 1015; 1008; 1008; 1009; 1012; 1021; 1024; 1022; 1022; 1013; 1022; 1023; 1045; 1049; 1040; 1040; 1022; 1018; 1038; 1042];

Demandware (USA) (monitoring period 03.02.2016 – 06.04.2016, unit of measure is the number of references)=[335; 334; 334; 332; 332; 335; 339; 331; 332; 329; 331; 330; 326; 338; 341; 342; 342; 341; 339; 343; 344; 340; 344; 344; 357; 357; 357; 367; 367; 385; 390; 389; 389; 400; 403; 404; 307; 397; 397; 395; 388; 383; 389; 381; 373; 374; 374; 381; 376; 377; 372; 373; 373; 370; 369; 364; 372; 370; 371; 371; 374; 377; 378; 377].

The first 59 values and the last 5 values of elements of each TS were used as the training data sequence and the test data sequence correspondingly. The forecasting models had been developed for each TS with the use of the MCSA on the base of two variants of optimization (Table 1). The forecasting results with use of these models received for the training and test sequences of data are shown in Fig. 3 and 4. The averaged values of the relative forecasting errors at the 5 steps, the averaged values of the affinity indicator and the averaged values of the tendencies discrepancy indicator received by the results of 10 runs of MCSA for each TS are presented in Table 2.

TABLE I.   THE FORECASTING MODELS

Time series	The forecasting model on the base of one quality indicator	The forecasting model on the base of two quality indicators
Hybris (RF)	$cos(exp(cos(cos(cos(d(t-4))+cos(d(t-1))) - sin(d(t-2))))) - exp(ln(d(t-3)))$	$exp(sin(sin(587.749/cos(d(t-2))) - 0.670) + sin(d(t-4)+sin(d(t-3))) - sin(0.667 - ln(d(t-1)))) - d(t-1)+cos(d(t-2))$
Magento (RF)	$ln(cos(ln(sin(sin(ln(d(t-1)) - sin(d(t-3)))\cdot d(t-4)) + cos(d(t-5)) - exp(d(t-2)))/exp(exp(d(t-7))\cdot d(t-7)))+exp(d(t-1)+sin(d(t-1)))$	$cos(sin(exp(d(t-3)+d(t-2)) - ln(d(t-5)) + exp(d(t-2)) - sin(d(t-5))+cos(ln(d(t-6))\cdot exp(d(t-7))))/ln(sin(d(t-4))+exp(d(t-2)))$
OpenCart (RF)	$sin(sin((ln(d(t-2))\cdot cos(d(t-1)))\cdot sin(d(t-5))) - cos(sin(d(t-6))\cdot d(t-4))\cdot sin(cos(d(t-3))/d(t-3)) - ln(d(t-3)) - d(t-2))$	$exp(sin(cos(sin(sin(d(t-4))+ln(d(t-3))+ln(d(t-2))) + cos(exp(d(t-6))\cdot ln(d(t-1))) - exp(cos(d(t-3))\cdot sin(d(t-2)))) - ln(exp(d(t-1)) - sin(d(t-5))$
PrestoShop (RF)	$ln(sin(exp(sin(ln(d(t-3))\cdot sin(d(t-2)))/ln(d(t-4))) + exp(sin(d(t-2))\cdot cos(d(t-5))))\cdot exp(d(t-1)) - cos(d(t-5))\cdot cos(ln(d(t-1))\cdot sin(d(t-1))$	$sin(ln(sin(sin(cos(d(t-1))\cdot sin(d(t-3)) - sin(d(t-6))) + sin(cos(d(t-2))\cdot d(t-4)) - cos(ln(d(t-3)) - ln(d(t-5)))) - ln(exp(d(t-1))\cdot 2.3))$
Demandware (USA)	$ln(exp(sin(cos(d(t-2)) - cos(d(t-3)))\cdot 0.002 + cos(cos(d(t-2)) - d(t-4))\cdot exp(d(t-1) - sin(d(t-3))))+exp(sin(d(t-2)) - ln(d(t-2))$	$ln(exp(sin(cos(ln(d(t-2))/sin(d(t-4))) - sin(d(t-3))) - ln(exp(d(t-1)) - ln(d(t-3)))+sin(d(t-1)/d(t-3)))+exp(sin(d(t-3)) - ln(d(t-1))$
Magento (USA)	$exp(exp(sin(ln(sin(d(t-6))+sin(d(t-1)))+cos(d(t-5))\cdot cos(sin(d(t-2))\cdot ln(d(t-1))))/exp(exp(d(t-3)) - cos(d(t-4)))\cdot ln(d(t-1)+cos(d(t-3))$	$exp(sin(sin(sin(sin(ln(d(t-6))\cdot d(t-1))\cdot ln(d(t-3))) + cos(sin(d(t-4))/sin(d(t-2))))\cdot sin(sin(d(t-5))\cdot 0.013))\cdot ln(ln(d(t-4))+d(t))$
Hybris (USA)	$ln(exp(sin(sin(sin(d(t-1))/exp(d(t-3)))\cdot(-0.499)) + exp(cos(d(t-4)) - cos(d(t-2)))) - ln(d(t-4)) - d(t-4))+sin(sin(d(t-1)) - ln(d(t-4))$	$sin(sin(exp((cos(d(t-2)) - 3.530) - cos(d(t-3))) + sin(0.968\cdot ln(d(t-4)))\cdot exp(sin(d(t-1))/cos(d(t-5)))) - d(t-1) - sin(d(t-2)))$

TABLE II.   THE AVERAGED VALUES OF THE FORECASTING ERRORS AT THE 5 STEPS AND THE AVERAGED VALUES OF THE QUALITY INDICATORS (AFF (AFER) AND TENDENCY)

№	The name of the TS	Aff (AFER), %	The value of the forecasting error					Tendency		
			1-st step	2-nd step	3-rd step	4-th step	5-th step	Average error of 5 steps, %	for the training sequence	for the test sequence
one quality indicator										
1.	Hybris (RF)	2,56	1,12	0,67	0,30	1,55	3,54	1,43	0.45	0.6
2.	Magento (RF)	0,44	0,45	0,45	0,45	3,79	2,28	1,48	0.38	0.4
3.	OpenCart (RF)	6,86	3,96	3,98	4,27	4,20	4,14	4,11	0.2	0
4.	PrestoShop (RF)	1,52	1,32	0,83	0,74	0,55	1,21	0,93	0.09	0.4
5.	Demandware(USA)	0,22	0,19	0,20	0,58	0,75	0,44	0,43	0.09	0.2
6.	Magento (USA)	2,46	1,30	1,41	0,45	0,20	1,37	0,94	0.43	0.4
7.	Hybris (USA)	2,40	3,45	3,45	1,41	1,70	3,07	2,61	0.4	0.4
	Average value	2.35	1.68	1.57	1.17	1.82	2.29	1.71	0.27	0.34
two quality indicators										
1.	Hybris (RF)	1,06	0,52	0,02	0,02	0,98	0,44	0,39	0.11	0
2.	Magento (RF)	0,32	0,27	0,28	0,35	0,48	0,45	0,37	0.06	0
3.	OpenCart (RF)	0,32	0,07	0,06	0,12	0,20	0,17	0,12	0	0
4.	PrestoShop (RF)	1,04	0,52	0,55	0,44	0,71	0,67	0,58	0.09	0.2
5.	Demandware(USA)	0,17	0,15	0,15	0,20	0,23	0,15	0,17	0	0
6.	Magento (USA)	0,06	0,05	0,04	0,04	0,09	0,10	0,06	0	0
7.	Hybris (USA)	0,08	0,05	0,06	0,05	0,03	0,17	0,07	0	0
	Average value	0.43	0.23	0.16	0.17	0.39	0.31	0.25	0.04	0.03

Forecasting Hybris (RF) based on a single indicator   Forecasting Hybris (RF) based on the two indicators

Forecasting Magento (RF) based on a single indicator   Forecasting Magento (RF) based on the two indicators

Forecasting OpenCart (RF) based on a single indicator   Forecasting OpenCart (RF) based on the two indicators

Forecasting PrestoShop (RF) based on a single indicator   Forecasting PrestoShop (RF) based on the two indicators

— Fact Value   — Forecast value (for the training data)   — Forecast value (for the test data)

Fig. 3. The Forecasting of TSs, determining the number of references of E-Commerce systems for HeadHunter.ru (Russia)
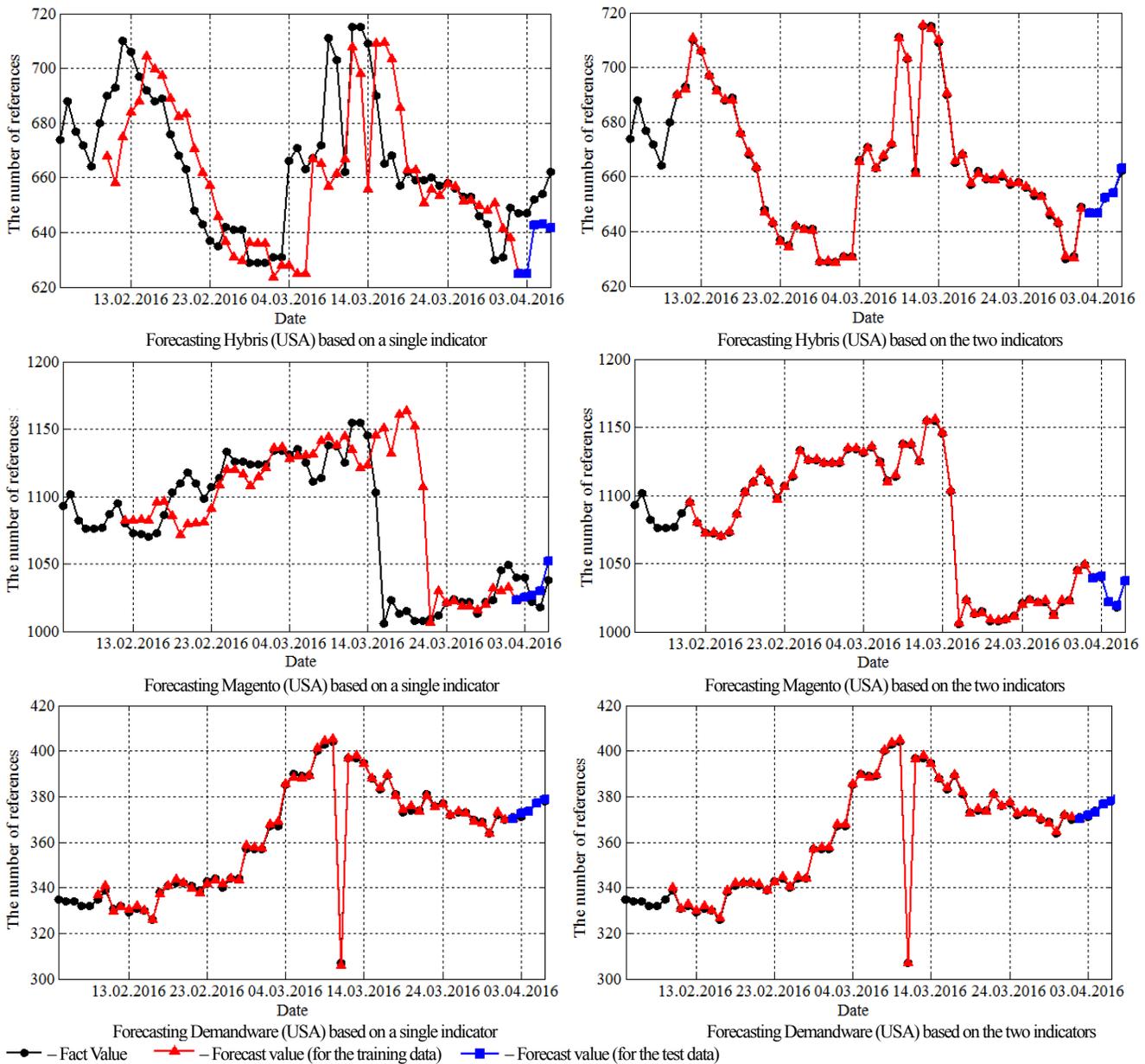
Fig. 4.   The Forecasting of TSs, determining the number of references of E-Commerce systems for Indeed.com (USA)

The averaged values of the relative forecasting errors at the 5 steps and the averaged values of the affinity indicator in the context of all TSs are presented graphically in Fig. 5, a. It is clear, that the second optimization variant is more effective as for the solution of problems of short-term forecasting (for $1 - 3$ step forward), as for the solution of problems of medium-term forecasting (for 4 and 5 steps forward). Herewith the second optimization variant allows not only receiving the smaller value of the tendencies discrepancy indicator *Tendency* in

comparison with the first optimization variant (Table 3 and Fig. 6), but also in many cases reducing the value of the affinity indicator *Aff* (Fig. 5, b) thanks to the corresponding correction of the search direction of the forecasting model.

The most essential influence on the development time of the forecasting model on the base of the SBT and MCSA is rendered by such parameters as the number of iterations, size of antibodies' population, coefficient of antibodies' cloning and coefficient of clones' reproduction.
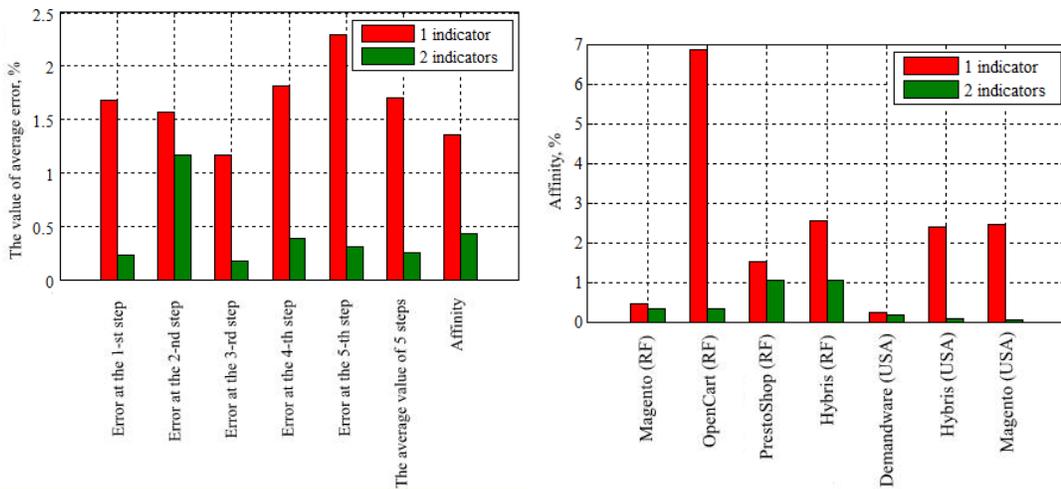
a – The averaged values of the forecasting errors at the 5 steps and the averaged value of the affinity indicator for all TSs

b – The averaged values of the affinity indicator for each TS

Fig. 5.   The comparison of the averaged values of the forecast errors at the 5 steps and the averaged values of the affinity indicator for two variants of optimization
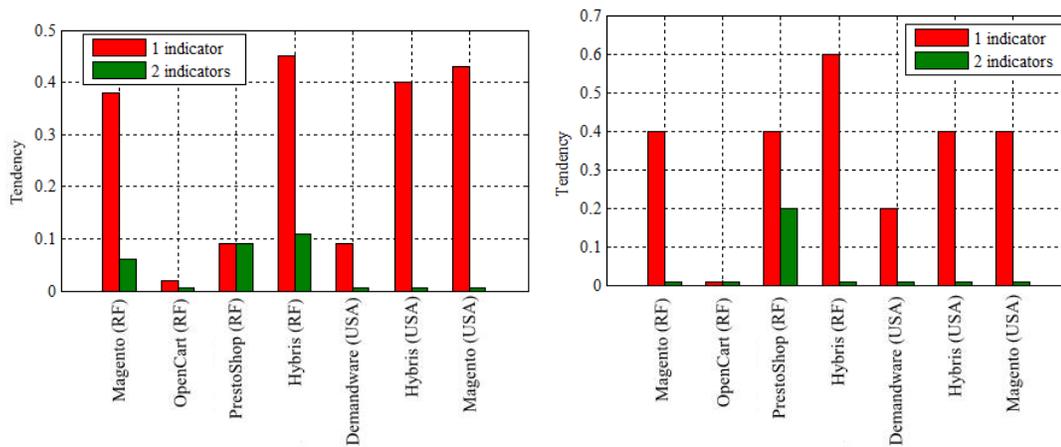
a – The averaged values of the tendencies discrepancy indicator for the training data sequence

b – The averaged values of the tendencies discrepancy indicator for the test data sequence

Fig. 6.   The comparison of the averaged values of the tendencies discrepancy indicator for two variants of optimization

In the reviewed example 400 iterations of MCSA for population of 20 antibodies were executed. Coefficient of antibodies' cloning was equal to 0.3. Coefficient of clones' reproduction was equal to 0.8. Computer working under the 64-bit Windows 7 version with RAM of 2 Gb and the two-nuclear Pentium 4 processor with a clock frequency of 3.4 GHz was used for experiment. 108.5 seconds on average were spent for creation of one forecasting model on the base of one quality indicator (for the first optimization variant). To build the forecasting model on the base of two quality indicators (for the second optimization variant) it is necessary to spent 120.9 seconds on average that on 12.4 seconds (on 10.3%) more, than for the first optimization variant.

V.   CONCLUSION

The comparative analysis of two optimization variants in the context of the development problem of the forecasting models on the base of the SBT shows the expediency and perspective of use of the second optimization variant which realizes the accounting of two quality indicators of the forecasting model – the affinity indicator and the tendencies discrepancy indicator.

Use of the principles of Pareto-dominance during the MCSA realization at the development of the forecasting models on the base of the SBT allows receiving the effective solution of the accounting problem of two quality indicators of the forecasting model at the acceptable time expenditures. A herewith it is possible to expand the scope of the forecasting models on the base of the SBT and MCSA.

Thus, the expediency of researches on further improvement of the multiobjective optimization algorithms for the purpose of their application to the search problem of the adequate forecasting model of TS is obvious.

REFERENCES

[1] L.A. Demidova, "Time Series Forecasting Models on the Base of Modified Clonal Selection Algorithm", 2014 International Conference on Computer Technologies in Physical and Engineering Applications (ICCTPEA), pp. 33 – 34, 2014.

[2] L.A. Demidova, "Assessment of the quality of the forecasting models based of the strict on binary trees and the modified clonal selection algorithm", Cloud of Science, 1 (2014), pp. 202-222 [in Russian] (http://cloudofscience.ru/sites/default/files/pdf/CoS_2_202.pdf).

[3] N.N. Astakhova, L.A. Demidova, E.V. Nikulchev, "Forecasting Method For Grouped Time Series With The Use Of K-Means Algorithm", Applied Mathematical Sciences, 9, no. 97, pp. 4813–4830, 2015.

[4] N.N. Astakhova, L.A. Demidova, E.V. Nikulchev, "Forecasting Of Time Series' Groups With Application Of Fuzzy C-Mean Algorithm," Contemporary Engineering Sciences, 8, no 35, pp. 1659–1677, 2015.

[5] N. Astakhova, L. Demidova, V. Konev, "The Description Problem Of The Clusters' Centroids", 2015 International Conference "Stability and Control Processes" in Memory of V.I. Zubov (SCP), pp. 448–451, 2015.

[6] N. Astakhova, L. Demidova, "Using of the notion "Pareto set" for development of the forecasting models based on the modified clonal selection algorithm," 6th Seminar on Industrial Control Systems: analysis, modeling and computation, art. 02001, 2016.

[7] C.M. Fonseca, P.J. Fleming, "Multiobjective optimization and multiple constraint handling with evolutionary algorithms – Part I: A unified formulation," Technical report 564, University of Sheffield, Sheffield, UK, pp. 1–16, 1995.

[8] J. Horn, N. Nafpliotis, D.E. Goldberg, "A niched Pareto genetic algorithm for multiobjective optimization," Proceedings of the First IEEE Conference on Evolutionary Computation, 1, Piscataway, pp. 82–87, 1994.

[9] E. Zitzler, L. Thiele, "Multiobjective optimization using evolutionary algorithms – A comparative case study," Parallel Problem Solving From Nature, V, A. E. Eiben, T. Back, M. Schoenauer, and H.-P. Schwefel, Eds. Berlin, Germany: Springer-Verlag, pp. 292–301, 1998.

[10] K. Deb, A. Pratap, S. Agarwal, T. Meyarivan, "A Fast and Elitist Multiobjective Genetic Algorithm: NSGA II, " KanGAL Report No. 200001, Indian Institute of Technology, Kanpur, India, pp. 182–197, 2000.

[11] H. Seada, K. Deb, "U-NSGA-III: A Unified Evolutionary Optimization Procedure for Single, Multiple, and Many Objectives: Proof-of-Principle Results," Evolutionary Multi-Criterion Optimization, 9019, pp. 34–49, 2015.

[12] P. Coello Coello C.A., Cruz Cortés N. "An approach to solve multiobjective optimization problems based on an artificial immune system," Proceedings of the First International Conference on Artificial Immune Systems, University of Kent at Canterbury, UK, pp. 212–221, 2002.

[13] G.-C. Luh, C.-H. Chueh, W.-W. Liu, "MOIA: Multi-Objective Immune Alghorithm," Computers and Structures, 82, pp. 829–844, 2004.

[14] X.L. Wang, M. Mahfouf, "ACSAMO: An Adaptive Multiobjective Optimization Algorithm using the Clonal Selection Principle," 2nd European Symposium on Nature-Inspired Smart Information Systems, pp. 1–12, 2006.

[15] L. Jiao, M. Gong, H. Du, L. Bo, "Multiobjective immune algorithm with nondominated neighbor-based selection," Evolutionary Computation, 16, issue 2, pp. 225–255, 2008.