Forecasting of Time Series’ Groups with Application of Fuzzy c-Mean Algorithm

Nadezhda N. Astakhova
Ryazan State Radio Engineering University
Gagarin Str., 59/1, 390005 Ryazan, Russian Federation

Liliya A. Demidova
Ryazan State Radio Engineering University
Gagarin Str., 59/1, 390005 Ryazan, Russian Federation

Evgeny V. Nikulchev
Moscow Technological Institute
Leninskiy pr., 38A, 119334 Moscow, Russian Federation

Abstract
The paper is focused on the development of the forecasting method for time series’ groups with application of the clustering algorithms. Fuzzy c-means algorithm (FCM algorithm) is suggested to be a basic one for clustering. The coordinates of the clusters’ centers have been put in correspondence with summarizing time series data – the clusters’ centroids. A description of time series, the clusters’ centroids, is implemented with application of the forecasting models. They are based on the strict binary trees and the modified clonal selection algorithm. The forming possibility of analytic dependences with application of such forecasting models, is shown. It is suggested to use a common forecasting model, which is constructed for time series – the clusters’ centroids, in forecasting for the private (individual) time series in the cluster. The use advantage of FCM algorithm in comparison with k-means algorithm for the clustering of time series’ groups is shown. The promising application of the suggested forecasting method for forecasting of time series’ groups is demonstrated.
Keywords: time series, clustering, cluster centroid, k-means algorithm, fuzzy c-means algorithm

1 Introduction

In recent years development of various socio-economic processes, which are becoming increasingly interconnected, accelerates more actively. In this regard, it is necessary to use approaches, which are already well known, more extensively. It is also important to develop new approaches to solving the problem of forecasting grouped time series (TS), which describe certain process indicators [22].

The plain fact is that during the development of forecasting TS models it is necessary to consider not only the tendencies of forecasted TS, but also tendencies of other TS which can influence on it. For example, such macroeconomic indicators as gross domestic product and the level of exports described by a TS are in explicit dependence, as well as the average level of income and level of education, health care and average life expectancy.

At the same time it is also clear that many TS have similar laws of changing values of its components and they can be combined in a subgroup according to the criterion of similarity of these laws [21]. In this case, it is advisable to develop a forecasting model that is common to all TS included in the subgroup for the purpose of direct use to forecast all private TS subgroups or further refinement of forecasting models for individual private TS (in the case of lack of precise predicted values while using the general model) [20].

The process of developing the model of forecasting TS is characterized by the highest computational complexity. That means that the use of a common forecasting model for a TS subgroup can be a significant step in the development of approaches to the analysis of correlating TS. It allows to carry out forecasting of private TS in a subgroup with acceptable time-consuming.

It is obvious that the use of cluster analysis techniques, namely clustering algorithms such as k-means algorithm [18], fuzzy c-means algorithm (FCM algorithm) and its modifications [3, 4, 13], EM-algorithm [19], will bring together similar TS into subgroups (clusters) and determine TS-centroids of clusters. All of that is necessary for solving the problem of common forecasting models development for TS subgroups. In addition to actual development of forecasting TS-centroid clusters models, it is suggested to use an approach that supposes construction of forecasting models on the base of strict binary trees (SBT) and modified clonal selection algorithm (MCSA) with the formation of analytical dependencies. They describe certain values of TS and provide minimum affinity values, in other words average forecasting error rate [11].

As shows the analysis, sometimes at the solution of various applied problems the FCM algorithm in comparison with k-means algorithm allows to receive more exact clustering results. This fact is explained by that the requirement about rigid (unique) membership of objects inherent in k-means algorithm is absent in FCM algorithm: at realization of FCM algorithm objects can belong at once to several
clusters with different values of membership degree. It should be noted that property of possible fuzzy membership of objects to clusters (classes) is used at the solution of a wide range of applied problems [13 – 15]. In some cases FCM algorithm allows to receive more exact results of a clustering. Therefore, it is possible to speak about expediency of its application at the solution of TS clustering problem for the minimization purpose of average forecasting error rate for TS groups.

The aim of the paper is to develop an approach to forecasting grouped time series with application of FCM algorithm, which provides a forecast with reasonable time-consuming.

2 Theoretical part

Suppose that there is a group of TS $T: t_i \quad (i = 1,m)$. Therewith, let each TS $t_i$ consists of $n \quad (10 \leq n \leq 30)$ elements $t_{ij} \quad (j = 1,n) [2]$.

Significant interest is in the solution of the problem of developing forecasting models for all TS in a group with reasonable time-consuming.

Time series, as well as other objects of data analysis, can be joined into clusters (subgroups) taking into account values of certain features. They are expectation values, dispersion and others. They can be calculated on the base of TS elements values, maximum and minimum values of TS, etc.

The problem of selection of an analytic dependence is usually solved in the process of forecasting models development for the purpose of calculating TS future values. This analytic dependence describes the variation of TS values of the elements in time in a best way. So, it seems appropriate to use similar mathematical laws of changing TS values in time for grouping TS into clusters.

The hypothesis of possible similarities between mathematical laws of changing TS values, and consequently, correspondent kinds of analytical dependences can be explained by the fact that many socio-economic indicators are interrelated, and changing trends of one of them causes a change in the trends of the other.

Unfortunately, there is no available information about what TS form connected subgroups in a group of analyzed TS. This fact leads to a need for additional methods of analysis, such as cluster analysis techniques.

The use of such clustering algorithms as k-means algorithm [18], fuzzy c-means algorithm (FCM algorithm) and its modifications [3, 4, 13], EM-algorithm [19], during iterative calculations allows to divide groups of objects into a predetermined number of clusters according to some optimality criterion, in addition to determining the coordinates of clusters’ centroids.
During handling the problem of TS clustering, centroids’ coordinates can be used to form integrating TS-centroids, which characterize private TS included in corresponding clusters (Fig. 1). It is obvious that a certain forecasting model can be developed for TS-centroid [2]. Then it can be used either directly for forecasting private TS assigned to a particular cluster or as a basic model for the purpose of further clarification and application for forecasting private TS.

It is suggested to use FCM algorithm [3, 4] that implements the partition of objects of group \(T\) into subgroups \(T_r\) \((r = \overline{1,c})\).

It should be noted that in some cases FCM algorithm in comparison with k-means algorithm, declaring the similar principles of realization, allows to receive more exact clustering results [3, 4]. The main difference of FCM algorithm from k-means algorithm consists in use of fuzzy membership functions for objects’ membership definition to clusters. In the context of solving TS clustering problem it is supposed to understand TS as an object.

FCM algorithm implements fuzzy interpretation of uncertainty and belongs to iterative algorithms that compute values of membership functions (MF) to the clusters and the centers’ coordinates in accordance with the values of MF [3, 4].

The FCM algorithm provides minimizing of the objective function:

\[
J(U,V) = \sum_{r=1}^{c} \sum_{i=1}^{n} (u_r(x_i))^m \cdot d^2(v_r, x_i)
\]

given that

\[
\sum_{r=1}^{c} u_r(x_i) = 1 \quad (c \in N \text{ and } c > 1; \ i = \overline{1,n}),
\]

where \(U = [u_r(x_i)]\) is a fuzzy \(c\) -partition of the objects’ set \(X\) on the base of MF \(u_r(x_i)\), determining the degree of membership of the \(i\)-th object to the \(r\)-th cluster; \(V = (v_1, ..., v_c)\) is the clusters’ centers’ vector; \(d(v_r, x_i)\) is the distance between the cluster center \(v_r\) and the object \(x_i\) in accordance with (8); \(m\) is a fuzzyfier
Forecasting of time series’ groups with application

\( m \in R \; ; \; m > 1 \); \( c \) is the clusters’ number; \( n \) is the objects’ number; \( i = 1, n \); \( r = 1, c \).

The following steps should be done during the implementation of FCM-algorithm:

1. The initialization of the initial \( c \)-partition \( U = [u_r(t_i)] \), that must match the requirement (2).
2. The calculation of clusters centroids coordinates:
   \[
   v_r^j = \frac{\left( \sum_{i=1}^{n} u_r(x_i)^m \cdot x_i^j \right) / \sum_{i=1}^{n} u_r(x_i)^m}{\sum_{i=1}^{n} (d(v_r,x_i)/d(v_1,x_i))^{2/(m-1)}}.
   \] (3)
3. The calculation of the MF values:
   \[
   u_r(x_i) = \frac{1}{\sum_{i=1}^{n} (d(v_r,x_i)/d(v_1,x_i))^{2/(m-1)}}.
   \] (4)
4. Steps 2 and 3 are repeated until the desired accuracy \( \varepsilon : |J(U,V) - J'(U,V)| \leq \varepsilon \), where \( J(U,V) \), \( J'(U,V) \) are the values of the objective function (1) in two successive iterations (or until prescribed number of iterations \( H \) is done).

If it is necessary to calculate the distance \( d(v_r,t_i) \) between the center of the cluster \( v_r \) and the object \( t_i \), usually Euclidean metric is used. In this case \( d(v_r,t_i) \) is calculated as [2]

\[
\begin{align*}
\sqrt{n} \sum_{j=1}^{n} (v_r^j - t_i^j)^2 \right]^{0.5},
\end{align*}
\] (5)

where \( n \) is the features’ number of the object. So in the context of solving the problem of TS clustering on the base of similarity of mathematical laws of TS values changing, it is suggested to do modification to perform this metric so as to take into account various relevance of TS elements (most distant in time from the moment of forecasting) and larger relevance of other TS elements (closest in time to the moment of forecasting) [2]

\[
\sqrt{n} \sum_{j=1}^{n} \frac{j}{n} \cdot (v_r^j - t_i^j)^2 \right]^{0.5},
\] (6)

The utility of such a modification can be explained so that with the passage of time the dependencies between socio-economic indicators change. So, during the forming of forecasting models, larger preference should be given to the closest TS elements in time of forecasting. The use of the weighting coefficients \( j/n \) suggests the most significant differences between the values of the most relevant TS elements (for example, if \( j = n \) then the weight value is 1, and if \( j = 1 \) then it is \( 1/n \)).
The use of formula (6) will not only take into account the relevance of TS elements, but also in a view of high sensitivity to divergent TS trends, it will provide unification of clusters on the base of similarity of the trends.

Results of experimental studies confirm the use advantage of formula (6) in comparison with formula (5) both for clustering on the base of FCM algorithm and for clustering on the base of k-means algorithm. Application of formula (6) allows to build the general forecasting models for TS-centroids of clusters which are characterized by the smaller average values of average forecasting error rates for all TS in the group [2].

In the case of cluster analysis there is a problem, the solution of which is a principal. This problem is related to the different scale of the analyzed TS which characterize various socio-economic indicators. Such indicators have different unit of measurement, a variety of ranges and corresponding statistical features (statistical expectation and so on). To solve this problem it is advisable to use the algorithms of normalization, which are widely used in the statistics, mathematical economics and econometrics. Their point is to define a medium level – median, according to which all analyzed TS are aligned. A certain conditional straight can be used as a median, as well as one of TS of the analyzed TS group or TS-centroid $S$, which $S_j$ elements are defined as [2]:

$$S_j = \frac{1}{m} \sum_{i=1}^{m} t_i^j,$$

(7)

where $t_i^j$ is the $j$-th element of the $i$-th TS; $i = 1, m$; $j = 1, n$; $m$ is the number of TS; $n$ is the number of elements of TS-centroid.

Normalization algorithm of the $i$-th TS $(i = 1, m)$, which is based on TS-centroid $S$, can be represented by the following sequence of steps [2].

Step 1.

1.1. The average step $hS$ of changing TS-centroid elements is determined as

$$hS = \frac{\max(S_j) - \min(S_j)}{n},$$

(8)

where $n$ is the number of elements of TS-centroid; $S_j$ is the $j$-th element of TS-centroid $S$.

1.2. The average step $ht_i$ of changing of the $i$-th TS elements $(i = 1, m)$ is determined:

$$ht_i = \frac{\max(t_i^j) - \min(t_i^j)}{n},$$

(9)

where $n$ is the number of TS elements; $t_i^j$ is the $j$-th element of the $i$-th TS.

Step 2.

2.1. The average TS-centroid level is determined:
where \( n \) is the number of elements of TS-centroid; \( S_j \) is the \( j \)-th element of TS-centroid.

2.2. The average level \( \bar{t}_i \) of the \( i \)-th TS \((i = 1, m)\) is determined as:

\[
\bar{t}_i = \frac{\sum_{j=1}^{n} t_{ij}}{n},
\]

where \( n \) is the number of TS elements; \( t_{ij} \) is the \( j \)-th element of the \( i \)-th TS.

Step 3.

The value \( d_{it} \) is calculated. It represents a ratio of diminution between the average level \( \bar{t}_i \) of the \( i \)-th TS and the \( j \)-th element \( t_{ij} \) of the \( i \)-th TS to the average step \( h_{it} \):

\[
\Delta t_{ij} = \frac{\bar{t}_i - t_{ij}}{h_{it}},
\]

where \( t_{ij} \) is the \( j \)-th element of the \( i \)-th TS; \( \bar{t}_i \) is the average level of the \( i \)-th TS; \( h_{it} \) is the average step of changing of the \( i \)-th TS elements; \( i = 1, m; \; j = 1, n \).

Step 4.

The \( j \)-th element \( t_{ij} \) of the \( i \)-th TS is transformed into:

\[
t_{ij} = S + \Delta t_{ij} \cdot hS.
\]

Such converted TS can further be used for clusterization with the help of FCM algorithm.

Since clusters centroids express general trends for TS subgroups, which form correspondent clusters, it is appropriate to develop forecasting models for TS-centroid [2].

Nowadays, there are different approaches to building TS forecasting models. One of the most perspective approach is the implementation of evolutionary algorithms (genetic algorithms [10, 23, 24], clonal selection algorithms [5]) which based on the principles of natural selection. They provide (at acceptable time expenses) construction of TS forecasting models that describe in certain TS values a best way. They can be also characterized by acceptable indicators values of quality models.

In the context of solving the problem of forecasting models development for TS-centroid of clusters, it is appropriate to use the modified clonal selection algorithm. This algorithm simulates the natural laws of the immune system functioning [7, 26] and provides the formation of quite complex analytical functions [12].

The principles of developing forecasting models of \( k \)-order with the use of MCSA were investigated in [11, 12].
McSA allows to form an analytical dependence on the base of SBT at an acceptable time expenses, that describes certain TS values and provides a minimum affinity (affinity) – average forecasting error rate (AFER) [11, 12]:

\[
AFER = \frac{\sum_{j=k+1}^{n} |f^j - d^j|}{n-k} \times 100\% ,
\]

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

In the context of solving the problem of forecasting TS subgroups as \(d^j\), it is suggested to use, for example, the \(j\)-th element \(t_i^j\) of the \(i\)-th TS or the corresponding \(j\)-th element \(v_i^j\) of the \(r\)-th TS-centroid cluster.

Possible options for analytical dependences are presented in the form of antibodies \(Ab\), which recognize Ag-antigens (certain TS values). An antibody \(Ab\) is selected as «the best one». It provides the minimum value of affinity \(Aff\) [6]. Coding of an antibody \(Ab\) is carried out by recording signs in a line. The signs are selected from three alphabets [11, 12]:

- the alphabet of arithmetic operations (addition, subtraction, multiplication and division) – \(Operation = \{+,-,\times,\div\}\);
- 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 [11, 12]. The number of signs in the alphabet of terminals \(Terminal\) in the antibody \(Ab\) determines maximal possible order \(K\) of forecasting models (\(K \geq k\), where \(k\) is the real model order), i.e. having the value of the element \(d^j\) in forecasting TS at the \(j\)-th moment of time, \(K\) values of TS elements can be used as: \(d^{j-K}, \ldots, d^{j-2}, d^{j-1}\).

The use of SBT type, illustrated in Fig. 1, allows to build complex analytical dependence and provides high accuracy of forecasting TS [11, 12].

Such SBT can be generated as a composition result of one «left» subtree of the maximum possible order \(K = 3\) and some «right» subtrees of the maximum possible order \(K = 2\). Thus the term «left» subtree («right» subtree) is used for the branch (left or right) of SBT level in which it is necessary to include a new subtree. In this case it is rational to form antibodies by subdividing SBT into sub-
trees, then execute the subtree-walk of each vertex forming the ordered symbol lists on its vertices and then combining these lists consecutively [11, 12].

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 we walk 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 in such an 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 \( K = 3 \). Using such a way of antibody formation we ensure the visualization of the SBT structure representation in the form of the subtrees union, and the antibody is easily interpreted in the analytical dependence.

\[
L \cdot S/SeSdC - S + EaCbEa,
\]
which can be transformed into an analytical dependence:
\[
f(a,b,c,d) = \ln(\cos(\sin(\exp(a) +
+ \cos(b)) - \exp(c)) \cdot \sin(\sin(d) / \sin(a)).
\]
In the task of a forecasting model development of \( k \)-order where \( k = 4 \) and considering the order of \( a, b, c, d \) in the terminal alphabet Terminal, the analytical dependence (15) can be written as:

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

(18)

Interpreting the antibodies into the analytical dependences it is rational to use the recursive procedure of interpretation [11, 12].

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 \( \text{Aff} \); the selection and cloning the part of «the best» antibodies, that are characterized by the least affine value \( \text{Aff} \); 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.

---

**Fig. 3. Scheme of TS groups’ forecasting**

---

It is arguable that the application of the general forecasting models developed for TS-centroid of the clusters ensures the uniqueness of the forecasting result for all the particular TS labeled as a cluster since while prognostic value calculation for the particular TS based on the analytic dependences defining prognostic models for TS-centroid of the clusters private item values of TS stand for the values of variables.
On the base of the suggested approach to a TS group forecasting ensuring the forecasting result within the acceptable expenditures of time can be received according to the scheme at the Fig. 3 [2]. Thus it is expedient to execute the step 2 repeatedly at different clusters’ quantity for a choice of the optimum clustering results defined taking into account value of objective function (1) of FCM algorithm, which must be minimum. Along with objective function (1) for an assessment of clustering quality a various indicators of clustering quality [16, 17], in particular, widely applied and well proved Xie-Beni index can be used [25].

It should be noted that the TS forecasting with use of the general forecasting models doesn’t conduct to receiving the general forecasting for the TS subgroup (cluster). The forecasting model defines only the mathematical law of elements value change of TS by means of the analytical dependence formed with MCSA applied. The forecasted values for every particular TS will be unique as they will be calculated by means of substitution known private elements values of TS in the general forecasting model.

3 Experimental studies

Approbation of the suggested approach to forecasting grouped TS was executed with use of TS for 22 macroeconomic indicators of Russian Federation taken from the site World DataBank from 1999 to 2014 (http://databank.worldbank.org/data/views/reports/tableview.aspx?isshared=true#).

All indicators were divided into 4 clusters (subgroups) with application of FCM algorithm. Both formulas (5) and (6) were used for calculation of distance between TS-centroid and any TS. Information on the clusters’ contents, received with use of formula (5), is given in the first column of Table 1. Information on the clusters’ contents, received with use of formula (6), is given in the first column of Table 2.

Data from 1999 to 2011 were used for development of forecasting models. Data from 2011 to 2014 were used for forecasting of private TS on 3 steps forward.

The general forecasting models for each cluster, received with use of formula (5), were defined on the base of antibodies compared to TS-centroids of clusters:

\[ +E - C + L \cdot Q - Q \cdot E - L + EhQeCdEdEbEeQhQfEc; \]
\[ L \cdot Q/Q \cdot E - S \cdot S + _\cdot C \cdot Sb \_ EgChEdLe\_ f \_ ?Sf; \]
\[ _\cdot S - -/ -_\cdot C/C \cdot S/ -_\cdot iCfEg _\_ dQh _\_ ?ShLdSa; \]
\[ E + C - L + Q + C \cdot E - C - C \cdot CeQcQdShSfSfSgSf \]
in the form of the following analytical dependences:

\[
f(d^{j_1}, d^{j_2}, d^{j_3}, d^{j_4}, d^{j_5}, d^{j_6}) =
\exp(\cos(\exp(\ln(\exp(d^{j_5}) + \sqrt{d^{j_2}})) - \sqrt{d^{j_1}})) + \sqrt{\exp(d^{j_3})} \cdot \exp(d^{j_4}) - \sqrt{\exp(d^{j_3} - \cos(d^{j_4}))} + \ln(\sqrt{d^{j_3} \cdot \exp(d^{j_1})});
\]
\[ f(d^{-1}, d^{-2}, d^{-3}, d^{-4}, d^{-5}, d^{-6}) = \]
\[ = \ln(\sqrt{\sqrt{\cos(\sin(d^{-1}) \cdot 1) \cdot d^{-4} \cdot \sin(\ln(d^{-5}) + \exp(d^{-6})})/\sin(\cos(\sin(d^{-2}) \cdot \exp(d^{-3})) \cdot \exp(d^{-1} - \sin(d^{-7})))}; \]
\[ f(d^{-1}, d^{-2}, d^{-3}, d^{-4}, d^{-5}, d^{-6}) = \]
\[ = \sin(\sin((\sin(d^{-6}) - \ln(d^{-5}))/\sin(d^{-2})) \cdot \cos(1,547 \cdot \sqrt{\sin(d^{-2})}) - \cos(d^{-5}/\exp(d^{-1})))/\cos(d^{-4}) - d^{-1}; \]
\[ f(d^{-1}, d^{-2}, d^{-3}, d^{-4}, d^{-5}, d^{-6}) = \]
\[ = \exp(\cos(\ln(\cos(\sin(d^{-3}) \cdot \sin(d^{-2})) - \sin(d^{-3})) + \exp(\sin(d^{-3}) - \sin(d^{-3}))) - \cos(\sin(d^{-1}) \cdot \sqrt{\sin(d^{-5}))}) - \sqrt{\sin(d^{-6}) + \cos(d^{-4})}). \]

It should be noted that analytical dependence \( f(d^{-1}, d^{-2}, d^{-3}, d^{-4}, d^{-5}, d^{-6}) \) can be written down without constant «1» in shorter form as
\[ f(d^{-1}, d^{-2}, d^{-3}, d^{-4}, d^{-5}, d^{-6}) = \]
\[ = \ln(\sqrt{\sqrt{\cos(\sin(d^{-1}) \cdot d^{-4} \cdot \sin(\ln(d^{-5}) + \exp(d^{-6})})/\sin(\cos(\sin(d^{-2}) \cdot \exp(d^{-3})) \cdot \exp(d^{-1} - \sin(d^{-7})))}. \]

The general forecasting models for each cluster, received with use of formula (6), were defined on the base of antibodies compared to TS-centroids of clusters:
\[ L - C/L+_+ - S - Q + C \cdot E = \text{EhQbLbQcEe} _?EdQbQa; \]
\[ L+C \cdot E = E - C+C/C/C - Cg _iEaEgSgQeSfSdLb; \]
\[ E \cdot C/S : L + _+ + E - L \cdot E = _hLiCgEeCiCd _iQiQc; \]
\[ _-S+Q+--- _- Q/S - Q - _iCgQe _fEgQgCc _h c \]
in the form of the following analytical dependences:
\[ f(d^{-1}, d^{-2}, d^{-3}, d^{-4}, d^{-5}, d^{-6}) = \]
\[ = \ln(\cos(\ln(\cos(\exp(\sqrt{\sin(d^{-6}) - \sqrt{\sin(d^{-5}) \cdot \exp(d^{-3})}) \cdot \exp(d^{-1})) - \sqrt{0,99 + \exp(d^{-2})})/\sin(\sqrt{\sin(d^{-4}) - \ln(d^{-5}))} + \sqrt{d^{-5}) - \exp(d^{-1})}); \]
\[ f(d^{-1}, d^{-2}, d^{-3}, d^{-4}, d^{-5}, d^{-6}, d^{-7}) = \]
\[ = \ln(\cos(\exp(\cos(\ln(d^{-6}) - \sin(d^{-5}))/\sin(d^{-3})) + \cos(\sqrt{d^{-1})}) - \exp(d^{-1} - \cos(d^{-2})); \]
\[ f(d^{-1}, d^{-2}, d^{-3}, d^{-4}, d^{-5}, d^{-6}, d^{-7}) = \]
\[ = \exp(\cos(\ln(\exp(\sqrt{\sin(d^{-6}) - \sqrt{\sin(d^{-1})} \cdot d^{-1}) \cdot \exp(\cos(d^{-5}) - \cos(d^{-1}))/\exp(d^{-4})+\cos(d^{-4}) \cdot \ln(\ln(d^{-1})+d^{-2})); \]
Forecasting of time series’ groups with application

\[
f(f^{j-1}, f^{j-2}, f^{j-3}, f^{j-4}, f^{j-5}, f^{j-6}) = \]

\[= \sin(\sqrt{\sin(\sqrt{f^{j-6} - f^{j-2}}) - \cos(f^{j-6})}) - \]

\[ - \sqrt{\sqrt{f^{j-3}}/\exp(f^{j-3})} + f^{j-4} - \sqrt{f^{j-5}}) + \cos(f^{j-3} - f^{j-1}).\]

Forecasting results of private TS on 3 steps forward, values of forecasting average relative errors for 3 steps forward, and also \(AFER\) (15) values (from 1999 to 2011) with use of formula (5) are given in Table 1. The average value \(AFER\) (15) for all private TS in group equals to 1,74 %. The average value of forecasting average relative errors for 3 steps forward (\(Error\)) for all private TS in group equals 3,14 %.

Forecasting results of private TS on 3 steps forward, values of forecasting average relative errors for 3 steps forward, and also \(AFER\) (15) values (from 1999 to 2011) with use of formula (6) are given in Table 2. The average value of \(AFER\) (15) for all private TS in group equals to 1,12 %. The average value of forecasting average relative errors for 3 steps forward (\(Error\)) for all private TS in group equals 2,78 %.

Comparison of the obtained results (values of errors) allows to make a conclusion on use expediency of formula (6) for TS clustering.

It should be noted that values of forecasting average relative errors \(Error\) for 3 steps forward were calculated according the formula:

\[
Error = (100\%/3) \cdot \sum_{j=13}^{n-2} \left| \frac{(f^j - d^j)}{d^j} \right|, \quad (19)
\]

where \(n = 13\) and values \(f^j\) и \(d^j\) with \(j = 14, 15\) и \(j = 16\) correspond to the forecasted (expected) and real values of TS elements for 2012, 2013 and 2014 years.

The most essential influence on the development time of forecasting model on the base of SBT and MCSA is rendered by such MCSA parameters as iterations’ number, size of antibodies’ population, coefficient of antibodies’ cloning and coefficient of clones’ reproduction.

In the reviewed example 600 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. 115,5 seconds were spent for creation of one forecasting model. Thus, 462 seconds (7 minutes 42 seconds) are necessary for creation of 4 models, and 2541 seconds (42 minutes 21 seconds) are necessary for creation of 22 models, that in 5,5 times more.
Table 1. Results of macroindicators’ forecasting with use of formula (5)

<table>
<thead>
<tr>
<th>No</th>
<th>Indicator name</th>
<th>Measurement unit</th>
<th>AFER, %</th>
<th>2012 fact</th>
<th>2013 forecast</th>
<th>2014 fact</th>
<th>2014 forecast</th>
<th>Error, %</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Procedures' start for business registration</td>
<td>Quantity</td>
<td>1.84</td>
<td>8</td>
<td>8.3</td>
<td>8</td>
<td>8.31</td>
<td>7.95</td>
</tr>
<tr>
<td>2</td>
<td>Export of high technologies</td>
<td>%</td>
<td>1.21</td>
<td>9.1</td>
<td>8.46</td>
<td>8.4</td>
<td>9.5</td>
<td>8.27</td>
</tr>
<tr>
<td>3</td>
<td>Coefficient of fertility</td>
<td></td>
<td>1.9</td>
<td>1.6</td>
<td>1.53</td>
<td>1.6</td>
<td>1.69</td>
<td>1.07</td>
</tr>
<tr>
<td>4</td>
<td>Value added in agricultural</td>
<td>% of GDP**</td>
<td>1.52</td>
<td>13.9</td>
<td>13.98</td>
<td>13.9</td>
<td>14.52</td>
<td>13.6</td>
</tr>
<tr>
<td>5</td>
<td>Mortality aged till 5 years</td>
<td>% of GDP**</td>
<td>1.22</td>
<td>3.9</td>
<td>3.77</td>
<td>4</td>
<td>3.97</td>
<td>3.06</td>
</tr>
<tr>
<td>6</td>
<td>Coefficient of teenage fertility</td>
<td>Births' quantity/100women</td>
<td>1.07</td>
<td>27.2</td>
<td>28.07</td>
<td>25.7</td>
<td>25.95</td>
<td>29.51</td>
</tr>
<tr>
<td>7</td>
<td>Import of goods and services</td>
<td>% of GDP**</td>
<td>2.17</td>
<td>21.1</td>
<td>22.17</td>
<td>22.3</td>
<td>21.68</td>
<td>21.6</td>
</tr>
<tr>
<td>8</td>
<td>Gross accumulation of capital</td>
<td>% of GDP**</td>
<td>1.04</td>
<td>22.6</td>
<td>22.74</td>
<td>24.5</td>
<td>22.44</td>
<td>24.36</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Income (except for grants)</td>
<td>% of GDP**</td>
<td>0.46</td>
<td>26.1</td>
<td>26.07</td>
<td>29.8</td>
<td>28.87</td>
<td>28.04</td>
</tr>
<tr>
<td>10</td>
<td>Immunization against measles</td>
<td>%</td>
<td>2.5</td>
<td>98</td>
<td>100.3</td>
<td>98</td>
<td>100.3</td>
<td>98.28</td>
</tr>
<tr>
<td>11</td>
<td>The population percent with primary education</td>
<td>%</td>
<td>1.43</td>
<td>92.5</td>
<td>100.15</td>
<td>97.1</td>
<td>94.77</td>
<td>98.15</td>
</tr>
<tr>
<td>12</td>
<td>Ratio of girls and boys in system of primary and secondary education</td>
<td>%</td>
<td>2.33</td>
<td>98.5</td>
<td>100.79</td>
<td>98.8</td>
<td>100.79</td>
<td>98.39</td>
</tr>
<tr>
<td>13</td>
<td>The improved water sources</td>
<td>%</td>
<td>2.19</td>
<td>97</td>
<td>99.08</td>
<td>97</td>
<td>99.29</td>
<td>98.02</td>
</tr>
<tr>
<td>14</td>
<td>The expected life expectancy at the birth</td>
<td>Years' quantity</td>
<td>0.85</td>
<td>68.9</td>
<td>70.71</td>
<td>70.5</td>
<td>71.01</td>
<td>72.94</td>
</tr>
<tr>
<td>15</td>
<td>The improved sanitation means</td>
<td>%</td>
<td>3.18</td>
<td>70.4</td>
<td>72.73</td>
<td>70.5</td>
<td>72.53</td>
<td>74.03</td>
</tr>
<tr>
<td>16</td>
<td>Value added in a services sector</td>
<td>% of GDP**</td>
<td>2.89</td>
<td>61.4</td>
<td>63.74</td>
<td>59.2</td>
<td>63.44</td>
<td>60.48</td>
</tr>
<tr>
<td>17</td>
<td>Value added in the industry</td>
<td>% of GDP**</td>
<td>2.96</td>
<td>34.7</td>
<td>35.4</td>
<td>36.8</td>
<td>36.49</td>
<td>33.37</td>
</tr>
<tr>
<td>18</td>
<td>Export of goods and services</td>
<td>% of GDP**</td>
<td>2.95</td>
<td>29.2</td>
<td>29.6</td>
<td>29.6</td>
<td>30.92</td>
<td>28.07</td>
</tr>
<tr>
<td>19</td>
<td>Gross national income per capita on the Atlas method</td>
<td>$</td>
<td>1.15</td>
<td>10010</td>
<td>10087.27</td>
<td>12740</td>
<td>12855.95</td>
<td>11082</td>
</tr>
<tr>
<td>20</td>
<td>Gross national income per capita at par purchasing power</td>
<td>$</td>
<td>0.17</td>
<td>19910</td>
<td>19853.99</td>
<td>22710</td>
<td>22656.8</td>
<td>20627.68</td>
</tr>
</tbody>
</table>

Cluster 1

Cluster 2

Cluster 3

Cluster 4

*KOE – kilogram oil equivalent; **GDP – gross domestic product
Table 2. Results of macroindicators’ forecasting with use of formula (6)

<table>
<thead>
<tr>
<th>No</th>
<th>Indicator name</th>
<th>Measurement unit</th>
<th>AFER, %</th>
<th>2012 fact</th>
<th>2012 forecast</th>
<th>2013 fact</th>
<th>2013 forecast</th>
<th>2014 fact</th>
<th>2014 forecast</th>
<th>Error, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Immunization against measles</td>
<td>%</td>
<td>0.17</td>
<td>98</td>
<td>98</td>
<td>98</td>
<td>98</td>
<td>98</td>
<td>98,28</td>
<td>0.09</td>
</tr>
<tr>
<td>2</td>
<td>The population percent with primary education</td>
<td>%</td>
<td>1.38</td>
<td>92.5</td>
<td>98.9</td>
<td>97.1</td>
<td>98.5</td>
<td>98.15</td>
<td>98.1</td>
<td>2.8</td>
</tr>
<tr>
<td>3</td>
<td>Ratio of girls and boys in system of primary and secondary education</td>
<td>%</td>
<td>0.3</td>
<td>98.5</td>
<td>98.5</td>
<td>98.8</td>
<td>98.5</td>
<td>98.39</td>
<td>98.8</td>
<td>0.24</td>
</tr>
<tr>
<td>4</td>
<td>The improved water sources</td>
<td>%</td>
<td>0.17</td>
<td>97</td>
<td>96.8</td>
<td>97</td>
<td>97</td>
<td>98.02</td>
<td>97</td>
<td>0.42</td>
</tr>
<tr>
<td>5</td>
<td>The expected life expectancy at the birth</td>
<td>Years’ quantity</td>
<td>1.22</td>
<td>68.9</td>
<td>68.6</td>
<td>70.5</td>
<td>68.9</td>
<td>72.94</td>
<td>70.5</td>
<td>2.02</td>
</tr>
<tr>
<td>6</td>
<td>The improved sanitation means</td>
<td>%</td>
<td>0.21</td>
<td>70.4</td>
<td>70.6</td>
<td>70.5</td>
<td>70.4</td>
<td>74.03</td>
<td>70.5</td>
<td>1.73</td>
</tr>
<tr>
<td>7</td>
<td>Value added in a services sector</td>
<td>% of GDP**</td>
<td>1</td>
<td>61.4</td>
<td>61.7</td>
<td>59.2</td>
<td>61.4</td>
<td>60.48</td>
<td>59.2</td>
<td>2.11</td>
</tr>
<tr>
<td>8</td>
<td>Procedures' start for business registration</td>
<td>Quantity</td>
<td>1.82</td>
<td>8</td>
<td>8.15</td>
<td>8</td>
<td>8.15</td>
<td>7.95</td>
<td>8.15</td>
<td>2.09</td>
</tr>
<tr>
<td>9</td>
<td>Export of high technologies</td>
<td>%</td>
<td>1.73</td>
<td>9.1</td>
<td>8.22</td>
<td>8.4</td>
<td>8.07</td>
<td>8.27</td>
<td>8.35</td>
<td>2.13</td>
</tr>
<tr>
<td>10</td>
<td>Coefficient of fertility</td>
<td>%</td>
<td>1.41</td>
<td>1.6</td>
<td>1.63</td>
<td>1.6</td>
<td>1.71</td>
<td>1.07</td>
<td>1.03</td>
<td>4.16</td>
</tr>
<tr>
<td>11</td>
<td>Value added in agricultural</td>
<td>% of GDP**</td>
<td>2.03</td>
<td>13.9</td>
<td>14.96</td>
<td>13.9</td>
<td>14.43</td>
<td>13.6</td>
<td>13.67</td>
<td>3.98</td>
</tr>
<tr>
<td>12</td>
<td>Mortality aged till 5 years</td>
<td>% of GDP**</td>
<td>0.67</td>
<td>3.9</td>
<td>3.97</td>
<td>4</td>
<td>3.98</td>
<td>3.06</td>
<td>3.02</td>
<td>1.2</td>
</tr>
<tr>
<td>13</td>
<td>Coefficient of teenage fertility</td>
<td>Births’ quantity/1,000 women</td>
<td>1.43</td>
<td>27.2</td>
<td>28.85</td>
<td>25.7</td>
<td>28.13</td>
<td>29.51</td>
<td>26.18</td>
<td>8.94</td>
</tr>
<tr>
<td>14</td>
<td>Value added in the industry</td>
<td>% of GDP**</td>
<td>1.16</td>
<td>34.7</td>
<td>33.63</td>
<td>36.8</td>
<td>35.28</td>
<td>33.37</td>
<td>37.79</td>
<td>6.82</td>
</tr>
<tr>
<td>15</td>
<td>Export of goods and services</td>
<td>% of GDP**</td>
<td>1.25</td>
<td>29.2</td>
<td>26.91</td>
<td>29.6</td>
<td>30.13</td>
<td>28.07</td>
<td>30.2</td>
<td>5.74</td>
</tr>
<tr>
<td>16</td>
<td>Import of goods and services</td>
<td>% of GDP**</td>
<td>1.2</td>
<td>21.1</td>
<td>21.49</td>
<td>22.3</td>
<td>21.18</td>
<td>21.6</td>
<td>21.93</td>
<td>2.8</td>
</tr>
<tr>
<td>17</td>
<td>Gross accumulation of capital</td>
<td>% of GDP**</td>
<td>0.03</td>
<td>22.6</td>
<td>22.97</td>
<td>24.5</td>
<td>24.6</td>
<td>24.36</td>
<td>25.32</td>
<td>2</td>
</tr>
<tr>
<td>18</td>
<td>Income (except for grants)</td>
<td>% of GDP**</td>
<td>0.63</td>
<td>26.1</td>
<td>26.15</td>
<td>29.8</td>
<td>28.17</td>
<td>28.04</td>
<td>28.23</td>
<td>2.11</td>
</tr>
<tr>
<td>19</td>
<td>Gross national income per capita on the Atlas method</td>
<td>$</td>
<td>1.97</td>
<td>10010</td>
<td>4558.31</td>
<td>12740</td>
<td>4931.94</td>
<td>11082</td>
<td>5114.44</td>
<td>3.97</td>
</tr>
<tr>
<td>20</td>
<td>Gross national income per capita at par purchasing power</td>
<td>$</td>
<td>2.45</td>
<td>19910</td>
<td>9649.14</td>
<td>22710</td>
<td>12299.21</td>
<td>20627.68</td>
<td>11019.45</td>
<td>2.54</td>
</tr>
<tr>
<td>21</td>
<td>Electricity consumption</td>
<td>kWh/person</td>
<td>1.57</td>
<td>6430.8</td>
<td>19719.84</td>
<td>6486</td>
<td>22859.9</td>
<td>6538.07</td>
<td>20920.03</td>
<td>1.01</td>
</tr>
<tr>
<td>22</td>
<td>Power consumption</td>
<td>KOE*</td>
<td>0.88</td>
<td>4932.3</td>
<td>6443.82</td>
<td>5113.2</td>
<td>6141.77</td>
<td>5075.21</td>
<td>6439.58</td>
<td>2.34</td>
</tr>
</tbody>
</table>

Average AFER, % | 1.12 | Average Error, % | 2.78

*KOE – kilogram oil equivalent; **GDP – gross domestic product
At realization of the offered method the additional time-consuming caused by need of TS clustering procedure takes place. However, time spent for clustering procedure in the reviewed example takes only 0.358 seconds that much less time which needs to be spent for additional creation of 18 forecasting models [this time takes 2079 seconds (34 minutes 39 seconds)].

At application of this method for new grouped TS, that is for the first time, it is necessary to perform of TS clustering procedure for different quantity of clusters for finding of the optimum splitting determined by a minimum value of objective function (1) of FCM algorithm. But also in this case essential decrease in time expenditure on forecasting for grouped TS takes place.

Comparison of forecasting results for the considered group of TS for 22 macroeconomic indicators of Russian Federation with use of FCM algorithm and k-means algorithm allows to make a conclusion on higher quality of TS clustering with application of FCM algorithm, and consequently on higher precision of forecasting [2]. In particular, sharing of FCM algorithm and formula (6) (in comparison with sharing of k-means algorithm and formula (5)) allowed to reduce the average value of $\text{AFTER}$ (15) for all private TS in group by 0.1%, and the average value of $\text{Error}$ (19) for all private TS in group by 0.32% (thanks to some structure change of the created clusters).

It should be noted that in all cases (when using both clustering algorithms (FCM algorithm and k-means algorithm) and both formulas (formula (5) and formula (6)) the average value of $\text{AFTER}$ (15) for all private TS in group and the average value of $\text{Error}$ (19) for all private TS in group, received with application of the offered forecasting method, aren't much more, than when using individual forecasting models for everyone TS (0.87% and 2.56% respectively).

It is important to note that the demanded forecast accuracy for the separate private TS can be reached by means of improving of general forecasting model with application of MCSA.

4 Conclusion

The offered approach to forecasting of TS groups realizes the combined use of FCM algorithm and forecasting models on the base of SBT and MCSA and provides individual forecast values for all TS in group with acceptable time-consuming.

The results of experimental studies received during macroeconomic indicators’ forecasting of Russian Federation confirm prospects of application and further development for the offered approach. Thus use expediency of the offered metric for calculation of distance between two TS is experimentally confirmed.

Use of FCM algorithm allows to form clusters (subgroups) of the connected TS, having similar change laws of elements’ values, and provides increase of TS forecasting speed. Often it is possible to receive more exact results of TS clustering, than in case of k-means algorithm’s use (thanks to removal of rigid restriction on membership uniqueness of object to some concrete cluster). Application of the general forecasting models (forecasting models for TS-centroids of clusters) for
private TS, entering into the relevant clusters (subgroups), doesn't lead to essential
decrease in forecasting accuracy. Forecast accuracy for private TS can be reached
in the course of general forecasting model improving with application of MCSA.
Forecast accuracy for private TS can be reached by means of improving of gen-
eral forecasting model with application of MCSA.

Further researches can be connected with the development of forecasting
models taking into account the principles of evolutionary-based multi-objective
optimization [6, 8, 9] for the purpose of forecasting accuracy increase.

References

[1] N. N. Astakhova, Time series forecasting by means of modified clonal selec-
tion algorithm and clustering mechanisms, Modern Informatization Problems
in Simulation and Social Technologies Proceedings of the XX-th International
Open Science Conference, Yelm, WA, USA, (2015), 146 – 150.

grouped time series with the use of k-means algorithm, Applied Mathemati-
http://dx.doi.org/10.12988/ams.2015.55391

[3] J.C. Bezdek, A convergence theorem for the fuzzy ISODATA clustering al-
gorithms, IEEE Transactions on Pattern Analysis & Machine Intelligence, 1

http://dx.doi.org/10.1016/0098-3004(84)90020-7

[5] L.N. de Castro, F.J. Von Zuben, Learning and Optimization Using the Clon-
al Selection Principle, IEEE Transactions on Evolutionary Computation, 6

[6] C.A. Coello Coello, A comprehensive survey of evolutionary-based multi-
objective optimization techniques, Laboratorio Nacional de Informatica

[7] D. Dasgupta, Z. Ji, F. Gonzalez, Artificial immune system (AIS) research in
the last five years, 2003 Congress on Evolutionary Computation (CEC’03),
1 (2003), 123 – 130. http://dx.doi.org/10.1109/CEC.2003.1299565

[8] K. Deb, Multi-Objective Optimization Using Evolutionary Algorithms,


**Received:** October 27, 2015; **Published:** December 12, 2015