

Predicative analytics for developing software

Nadejda Yarushkina
Dept.of Information systems
Ulyanovsk state technical University
Ulyanovsk, Russia
jng@ulstu.ru

Tatiana Afanasieva
Dept.of Information systems
Ulyanovsk state technical University
Ulyanovsk, Russia
tv.afanasjeva@gmail.com

Irina Timina
Dept.of Information systems
Ulyanovsk state technical University
Ulyanovsk, Russia
i.timina@ulstu.ru

Abstract—The article is devoted to the problem of applying the formal data mining tool – forecasting – for the developing of new software and for reengineering the present software. We propose the algorithm adjustments of the time series forecasting. This algorithm takes into account the dependence of the current state of time series from the previous one, the influence of basic fuzzy projected trends in the time series. The proposed algorithm expands the opportunities of time series short-term forecasting on the base of fuzzy trends, as the historical software time series are of small length. The proposed algorithm was examined experimentally and showed the efficiency

Keywords—data mining, development, software, fuzzy time series, forecasting, fuzzy tendency

I. INTRODUCTION

Predicative analytics is the major formal tool for developing of new software and for reengineering the present software. To develop not only new, but competitive software, it is necessary to research of appearing trends by using formal methods at an analysis stage. This research should be directed on the perspective functions and software technologies, scientific achievements and user requirement. So, to fulfilled such research for development of the competitive software the data mining algorithms, extracting new trends and predicative analytics have to be used. The predicative analytics is the effective tool in the analysis of trends, if we want to create new competitive software. The core of the predicative analytics undoubtedly is the time series analysis of the important software parameters.

Forecasting is one of the problems of Time series analysis. The results of Time series forecasting and its trends are useful for business and management. Particularly the forecasting of

economic indicators and its trends is a part of planning process in the enterprise. Unfortunately, the trend forecasting, as a data mining formal tool, for developing of new software and for reengineering the present software practically isn't used. The regular analysis of the trend and dependencies in technologies, scientific achievements and user requirement, expressed in time series, allows to uncover ways to create a new useful ideas for software development.

There are two approaches to Time series forecasting and analysis. The first one is based on forecasting only one Time series. This approach is widely spread, and many methods and models have been proposed: statistical [1-4], fuzzy [5-9], and their combination [10-13]. In this approach regression models on time and autoregression models are used to predict values and global trends.

In The second approach to Time series forecasting the predictive model includes another Time series values in addition. In this approach it is supposed that variation of one time series causes the variation of another time series. To estimate this dependency the regression model has to be identified and to build the adequate regression model the cointegration between TS has to be studied.

If time series are not cointegrated, then the estimation of their dependency makes no sense. Fundamental results in this problem were generated by Granger, Engle, Johansen, Phillips in [14-16]. There are well-known methods for cointegration testing: Engle–Granger test, the Johansen test, Phillips–Ouliaris cointegration test. Cointegration is an important property of many economic indicators. To forecast them the ECM- time series modeling with correction errors was proposed [15]. The main idea of ECM is to correct model for short-term dynamics in accordance with long-term dependence between time series.

However, most of the methods are required long-term time series, and the time series might be the same length. The mentioned fact makes a serious problem for an applied researcher when forecasting time series of software indicators, particularly, in case time series are not long enough, for example, 20 values (short-term time series).

For the analysis of a dependencies between short-term time series fuzzy models could be used computational Intelligence techniques [7-9], [17], [19].

According to fuzzy modeling the components of time series model are considered as fuzzy sets, and there were proposed a lot of techniques. In [17] three groups of fuzzy time-series data models are considered: (1) a regression model-based analysis [6], [11] by using a fuzzy regression coefficient, (2) a Box-Jenkins model-based analysis by using a fuzzy autocorrelation coefficient [18] (3) a fuzzy reasoning (IF-THEN rule)-based analysis by using fuzzy time series model [7-8].

In [9],[12] the modification of fuzzy time series model is proposed for modeling fuzzy short-term tendencies.

However, the problem of modeling fuzzy tendencies (local trends) based on dependences between fuzzy time series, has received far less attention and is still an open problem. The importance of this problem lies in the fact that many time series of software indicators are interdependent and the situation that fuzzy trends of TS Z might be a predictor of TS Y is really exist.

In this paper we proposed a new time series forecasting algorithm, using fuzzy trends time series model and hypothesis about dependences between two time series. We embed explicit hypothesis from an expert of applied field, that fuzzy trends of time series Z is a predictor of fuzzy trends of time series Y . The term predictor is used for determination of significant convincing characteristics providing the most precise forecast of any phenomenon. In other words the predictor is a pre-requisite of a certain important event evaluated and being a part of a respective equation for the forecast.

The structure of the article is the following: Part 2 briefly describes basic provisions of the proposed time series forecasting algorithm; Part 3 shows theoretic basics of fuzzy trends time series model; Part 4 demonstrates the new algorithm of time series forecasting and computing experiment results.

II. BASIC PROVISIONS OF PROPOSED ALGORITHM

In the development of new software applications on the analysis of the IT market is one of the important tasks. Indicators of the IT market for some time period can be represented by time series. It should be noted the following features of the time series: short time series (less than 20 values), nonstationary time series, the existence of the relationship between time series, time series with missing values. The main purpose of market analysis software is forecasting the trends in sales volumes in different segments of the software applications.

Short length, non-stationary behavior, inaccurate data and problem of selection of a proper model are the factors that complicate the use of classic statistical models and methods [1-4].

The analysis and forecast of such time series is held usually by software experts, forming results in the form of linguistic terms of tendencies: Small growth, Rapid Fall, Stability.

It is known that to design linguistic values the fuzzy-set theory is used [5], which is the base for fuzzy TS forecasting models. Models for fuzzy values forecast are called fuzzy time series [8].

But the problem of fuzzy trends modeling is under-investigated.

A. Analysis of fuzzy trends definitions

In statistical approach the analysis of behavior characterizing long-term dynamics of TS is connected with the trend concept describing long-term dependence of time series values from time. So in accordance with [3] mathematic model of time series is represented in the form:

$$y_t = f(t) + u_t \quad (1)$$

Herewith it is assumed the presence of deterministic trend (or the aggregate of trends) upon the whole of time series.

Diversity of possible variants of TS trend component behavior approximate quantitative estimation and absence of opportunities to identify qualitative assessments caused the idea to use fuzzy trend concept.

In 1982 H. Tanaka [11] proposed the model of linear regression with fuzzy coefficient and applied methods of linear programming. However the use of fuzzy coefficients did not allow to solve the problem of qualitative TS trend identification.

Further research of FTS brought out a new task of FTS object description, modeling and forecasting – fuzzy trend as a representative of qualitative changes expressing the changes not in numerical but in fuzzy values of TS [7].

The sequence of FTS fuzzy trends in time dimension brings out fuzzy time series with fuzzy trend [9].

Our research showed that the elimination of the above mentioned restrictions by introducing the forecast correction procedure considering main trend allows to get a more accurate forecast of a future TS trend of a software market indicators.

III. THEORETICAL PROPOSITION OF PROPOSED TIME SERIES FORECASTING ALGORITHM

Suppose there are given a discrete time series $Y = \{t_i, x_i\}, i = 1, 2, \dots, n$. According to the basic provisions of FTS theory, developed by Zadeh [5], Song and Chissom [8], any finite discrete time series – numeric, nonnumeric, mixed – might be transformed into FTS $\tilde{Y} = \{t_i, \tilde{x}_i\}, i = 1, 2, \dots, n$, given that its value set $X = \{x_i\}$ will be covered by specific functions (fuzzy sets) $\tilde{x}_i \in \tilde{X}, j = 1, 2, \dots, m, m < n$.

Definition 1. Fuzzy trend τ , assigned on the segment $[t_i, t_j], t_j \geq t_i$ with \tilde{x}_i, \tilde{x}_j of fuzzy time series \tilde{Y} is a fuzzy term assigning fuzzy increment $\tau = \tau((t_i, \tilde{x}_i), (t_j, \tilde{x}_j))$.

Generalized model of FTS FT will be:

$$\tau_i = f(\tau_{i-1}, \dots, \tau_{i-d}) \quad (2)$$

where d – a fixed number, model parameter;
 $\tau_i, \tau_{i-1}, \dots, \tau_{i-d}$ – a sequence of fuzzy trends;
 f – some fuzzy dependence.

Substantial analysis allows to conclude that the term *Trend* defines qualitative changes upon the time domain and is used in sentences along with general linguistic assessments connected with content function, type and intensity, for example, Long Growth Trend, Strong Fall Trend, High Quality Stability Trend, etc.

Therefore it is worthwhile to mark out the following characteristics for FT:

- **Fuzziness.** Fuzziness is a fact that FT is built on the base of fuzzy values of FTS and inherits the fuzziness of these values, time series might correspond with various fuzzy trends with different grade of membership.
- **Duration.** Duration is a characteristic of various duration of fuzzy trend.
- **Typicality.** FT typicality property allows to discern classes, FT types, which have fuzzy trends considered as homogeneous within.
- **Significance.** For various fuzzy trends of one type and equal duration application of level of FT significance of intensity characteristics is appropriate.
- **Time awareness.** This property shows that fuzzy trends are determined between two values of time interval.
- **Linguistic interpretability.** This fuzzy trend property follows the definition of fuzzy trend as a quality changes characteristic. Fuzzy trend is defined as a fuzzy mark matching linguistic term.

We suggest a more detailed description of fuzzy trend which has fuzzy time series. For this purpose let the following statements and definitions be introduced.

Assume that the linguistic variables Fuzzy Time Series, Fuzzy Trend, Type_Trend, Intensity_Trend, Duration_Trend are assigned with basic finite term-sets $\tilde{X}, \tilde{\mathfrak{T}}, \tilde{V}, \tilde{A}, \Delta T$ respectively.

Definition 2. Each fuzzy trend $\tau \in \mathfrak{T}$ of fuzzy time series $\tilde{Y} = \tilde{x}_t, t = 1, 2, \dots$ might be shown a structure model in the form of relation tuple built on Cartesian product of fuzzy trend properties $\tilde{V} \times \tilde{A} \times \Delta T \rightarrow \mathfrak{T}$:

$$\tau = \langle \tilde{v}, \tilde{a}, \Delta t, \mu \rangle \quad (3)$$

where τ – a name of fuzzy trend from the set $\mathfrak{T}, \tau \in \mathfrak{T}$;
 \tilde{v} – a type of fuzzy trend (change type) $\tilde{v} \in \tilde{V}$ shows basic quality dependences of time series {Fall, Growth, Stability}.
 \tilde{a} – intensity of fuzzy trend, $\tilde{a} \in \tilde{A}$, might be introduced linguistically, e.g. values from the set {Intense, Average, Weak};
 Δt – duration of fuzzy trend, $\Delta t \in \Delta T$;
 μ – a membership function of a FTS segment bounded by interval Δt of fuzzy trend τ

Classify fuzzy trends of fuzzy time series in accordance with duration into elementary $T \in \mathfrak{T}(\Delta t = 1)$ local $N_\tau \in N_\mathfrak{T}(1 < \Delta t < n - 1)$ and basic (general) $G_\tau \in G_\mathfrak{T}(\Delta t = n - 1)$.

Definition 3. Elementary fuzzy trend (EFT) of fuzzy time series $\tilde{Y} = \tilde{x}_t, \tilde{x} \in \tilde{X}, t = 1, 2, \dots, n$ is a fuzzy trend $\tau_t = \langle \tilde{v}_t, \tilde{a}_t, \mu_t \rangle$ showing the character of change of FTS segment between two neighboring fuzzy FTS marks $\tilde{x}_{t-1}, \tilde{x}_t$ with membership degree $\mu_t = \min(\tilde{x}_{t-1}(x_{t-1}), \tilde{x}_t(x_t))$.

Types of elementary trends are basic types of fuzzy trends of FTS from the set $\tilde{V}_1 = \{\tilde{v}_1, \tilde{v}_2, \tilde{v}_3\}$, \tilde{v}_1 =Stability, \tilde{v}_2 =Fall, \tilde{v}_3 =Growth.

Definition 4. Finite elementary trend $\tau_s = \langle \tilde{v}_s, \tilde{a}_s, \Delta t_s, \mu_s \rangle$ is an elementary trend built on the last pair of neighboring FTS values.

Definition 5. Elementary fuzzy trend (EFT) time series is introduced in the form

$$\begin{aligned} \tilde{v}_t &= TTend(\tilde{x}_t, \tilde{x}_{t+1}), \tilde{a}_t = RTend(\tilde{x}_t, \tilde{x}_{t+1}), \\ \mu_t &= \min(\mu(\tilde{x}_t), \mu(\tilde{x}_{t+1})) \end{aligned} \quad (4)$$

Statement 1. Any finite discrete time series might be transformed into time series of EFT.

Specify the generalized model of EFT time series and define main components EFT changes admitting that this model might behave differently.

Definition 6.

Let $X_t, (t = 1, 2, \dots) \subset R^1, R^1$ is a universe discourse, where fuzzy sets $\tilde{x}_t^i, (i = 1, 2, \dots), \tilde{v}_t^j, (j = 1, 2, \dots), \tilde{a}_t^s, (s = 1, 2, \dots)$ are defined and \tilde{X}_t is a collection of $\tilde{x}_t^i, (i = 1, 2, \dots), \tilde{V}_t$ is a collection of $\tilde{v}_t^j, (j = 1, 2, \dots), \tilde{A}_t$ is a collection of $\tilde{a}_t^s, (s = 1, 2, \dots)$. Let relations $R_V: \tilde{X} \times \tilde{X} \rightarrow \tilde{V}, R_A: \tilde{X} \times \tilde{X} \rightarrow \tilde{A}$ exist, then the model of *fuzzy dynamic process with fuzzy differences* is

$$\tilde{X}_t = (\tilde{X}_{t-1} \times \tilde{V}_t \times \tilde{A}_t) \circ R(t, t - 1) \quad (5)$$

where

$$\begin{aligned} \tilde{V}_t &= \tilde{V}_{t-1} \times \tilde{V}_{t-2} \times \dots \times \tilde{V}_{t-p} \circ R_{\tilde{v}}(t, t - p), \\ \tilde{A}_t &= \tilde{A}_{t-1} \times \tilde{A}_{t-2} \times \dots \times \tilde{A}_{t-p} \circ R_{\tilde{a}}(t, t - q) \end{aligned} \quad (6)$$

there $\tilde{X}_t, \tilde{X}_{t-1}$ is a state of fuzzy process, coded by fuzzy sets (linguistic terms);

$R(t, t - 1)$ – fuzzy relation defining the first-order model in terms of fuzzy values \tilde{X}_t , which can be represented by sets of fuzzy “equations” in the form of *IF-THEN*;

$\tilde{V}_t = \tilde{V}_{t-1} \times \tilde{V}_{t-2} \times \dots \times \tilde{V}_{t-p} \circ R_{\tilde{v}}(t, t - p)$ is p -th order fuzzy time series model of FT-type changes (changes type) shows basic quality dependences of time series {Fall, Growth, Stability},

$\tilde{A}_t = \tilde{A}_{t-1} \times \tilde{A}_{t-2} \times \dots \times \tilde{A}_{t-p} \circ R_{\tilde{a}}(t, t - q)$ is q -th order fuzzy time series model of FT intensity changes (changes intensity),

\circ – composition sing in fuzzy theory; $p > 0; q > 0$.

The model of the numeric time series $Y = \{t_i, x_i\}, (i = 1, 2, \dots, n)$ is represented in the form of:

$$x_t = x_{t-1} + v_t \cdot \alpha_t + \varepsilon_t \quad (7)$$

where x_t, x_{t-1} – numeric values of time series, generated by defuzzification of FTS fuzzy values $\tilde{Y} = \tilde{x}_t, \tilde{x} \in \tilde{X}, t = 1, 2, \dots, n$:

$x_t = deFuzzy(\tilde{x}_t)$, $x_{t-1} = deFuzzy(\tilde{x}_{t-1})$, $t = 1, 2, \dots, n$
 here v_t – numeric values, defining EFT type, obtained as a result of defuzzification of fuzzy values $v_t = deFuzzy(\tilde{v}_t)$; α_t – numeric value defining EFT intensity, obtained as a result of defuzzification $\alpha_t = deFuzzy(\tilde{\alpha}_t)$; ε_t – errors.

To defuzzificate fuzzy trend type the following formula is used

$$DeFuzzy(\tilde{v}_i) = \begin{cases} 0, & \text{if } \tilde{v}_i = \text{"Stability"} \\ -1, & \text{if } \tilde{v}_i = \text{"Fall"} \\ 1, & \text{if } \tilde{v}_i = \text{"Growth"} \end{cases} \quad (8)$$

To defuzzificate intensity the centroid method ($nmin$, $nmax$ – determined by min and max differences of TS values) is used:

$$DeFuzzy(\tilde{\alpha}_t) = \frac{\int_{nmin}^{nmax} x \cdot \tilde{\alpha}(x) dx}{\int_{nmin}^{nmax} \tilde{\alpha}(x) dx} \quad (9)$$

According to the approach suggested in this work the results of EFT forecasting should be corrected considering the main trend $G\tau$. To identify the main fuzzy trend $G\tau$ and to define its TS components we suggest heuristic algorithm where the applicable components are determined experimentally. To make algorithm function the source TS was transformed in FTS.

Algorithm 1.

Step 1. Deriving of fuzzy elementary trend time series $\tau_t = \langle \tilde{v}_t, \tilde{\alpha}_t, \mu_t \rangle$, $t = 2, 3, \dots, n$ and defuzzification of EFT intensities is according formula (9): $\alpha_t = deFuzzy(\tilde{\alpha}_t)$.

Step 2. Calculate the cumulative intensity of sane-type EFT upon the whole of time series $\tau_t = \langle \tilde{v}_t, \tilde{\alpha}_t, \mu_t \rangle$ ($t = 2, 3, \dots, n$):

IF ($v_t = \text{"Growth"}$), THEN $ST_{growth} = ST_{growth} + \alpha_t, \mu_{growth} = \max(\mu_{growth}, \mu_t)$;
 IF ($v_t = \text{"Fall"}$), THEN $ST_{fall} = ST_{fall} + \alpha_t, \mu_{fall} = \max(\mu_{fall}, \mu_t)$.

Step 3. If ($ST_{growth} = 0$ and $ST_{fall} = 0$) or ($ST_{growth} = ST_{fall}$), then the type of the main FT $\tilde{v}_{G\tau} = \text{"Stability"}$ and after defuzzification $v_{G\tau} = 0$, dynamics of time series is stationary, otherwise Step 4.

Step 4. On the base of comparative analysis of values ST_{growth} and ST_{fall} determine the type of main fuzzy trend. If $ST_{growth} \geq 2 \cdot ST_{fall}$, then $\tilde{v}_{G\tau} = \text{"Growth"}$ and after defuzzification $v_{G\tau} = 1$, otherwise the type of main fuzzy trend $\tilde{v}_{G\tau} = \text{"Fall"}$ and after defuzzification $v_{G\tau} = -1$. Time series dynamics is non-stationary.

Step 5. Then the main trend intensity is:

$$\alpha_{G\tau} = |ST_{growth} - ST_{fall}|.$$

For the model FTS (sample of 50 TS of short length), the accuracy of basic trend identification $G\tau$ suggested by the algorithm was 99 %.

IV. ALGORITHM OF SHORT-TERM TIME SERIES FUZZY TRENDS FORECASTING

Consider the algorithm of TS for casting $Y = \{t_i, x_i\}$, ($i = 1, 2, \dots, n$), on the assumption that expert's hypothesis that TS fuzzy trend $Z = \{t_i, z_i\}$, ($i = 1, 2, \dots, k$) is a predictor of TS Y is reasonable. Algorithm consists of 3

phases. During the first phase forecast EFT of TS Y , according to (3):

$$\tau_{t+1}^Y = f(\tau_t^Y)$$

therein τ_{t+1}^Y – prognostic fuzzy elementary trend of time series Y ,

τ_t^Y – current fuzzy elementary trend of time series Y ,

f – dependence in fuzzy elementary trends of time series Y .

The second phase involves correction of prognostic fuzzy elementary trend of time series Y in accordance with the components of main trends of the analyzed time series $G\tau_Y$ and TS predictor $G\tau_Z$ respectively:

$$\hat{\tau}_{t+1}^Y = r(\tau_{t+1}^Y, G\tau_Y, G\tau_Z),$$

therein τ_{t+1}^Y – is a prognostic fuzzy elementary trend of time series Y , $\hat{\tau}_{t+1}^Y$ – prognostic fuzzy elementary trend of time series Y after correction, $G\tau_Y$ main fuzzy trend of time series Y , $G\tau_Z$ – main fuzzy trend of time series Z , r – correction rules.

The third phase serves for estimation of prognostic value of numeric time series Y , according to (7).

On this base we suggest the following algorithm for TS.

Algorithm 2.

Step 1. Transformation of numeric TS $Y = \{t_i, x_i\}$, ($i = 1, 2, \dots, n$), into fuzzy TS $\tilde{Y} = \tilde{x}_t, \tilde{x} \in \tilde{X}, t = 1, 2, \dots, n$:

$$\tilde{x}_i = Fuzzy(x_j), x_j \in X, \tilde{x}_i \in \tilde{X},$$

Here at the intervals where fuzzy sets defined, its form and name are set up by user from object domain characteristics.

Step 2. Transformation of fuzzy TS $\tilde{Y} = \tilde{x}_t, \tilde{x} \in \tilde{X}, t = 1, 2, \dots, n$ into fuzzy TS of fuzzy elementary trends, is according to (3,4):

$$\tau_t^Y = \langle \tilde{v}_t, \tilde{\alpha}_t, \mu_t \rangle, \\ \tilde{v}_t = TTend(\tilde{x}_t, \tilde{x}_{t+1}), \tilde{\alpha}_t = RTend(\tilde{x}_t, \tilde{x}_{t+1}), \\ \mu_t = \min(\mu(\tilde{x}_t), \mu(\tilde{x}_{t+1})).$$

Beforehand determine a set of FT type names $\tilde{V} = \{\text{Fall, Growth, Stability}\}$, and a set of FT intensity names $\tilde{A} = \{\text{Intense, Average, Weak}\}$.

Step 3. Generation of EFT components change models of TS Y and its forecasting for one period according to (6):

$$\tilde{v}_{t+1} = \tilde{v}_t \times \tilde{v}_{t-1} \times \dots \times \tilde{v}_{t-p} \circ R_{\tilde{v}}(t, t-p), \\ \tilde{\alpha}_{t+1} = \tilde{\alpha}_t \times \tilde{\alpha}_{t-1} \times \dots \times \tilde{\alpha}_{t-p} \circ R_{\tilde{\alpha}}(t, t-q)$$

Step 4. Forecast of numeric time series Y with preliminary defuzzification according to formula (7) of FT components

$$\tau_{t+1}^Y = \langle \tilde{v}_{t+1}, \tilde{\alpha}_{t+1}, \mu_{t+1} \rangle \\ x_{t+1} = x_t + v_{t+1} \cdot \alpha_{t+1}.$$

Step 5. Application of main trend identification algorithm (see Part 3. Algorithm 1) for TS Y and determination of its components $G\tau_Y = \langle \tilde{v}_{G\tau}^Y, \tilde{\alpha}_{G\tau}^Y, \tilde{\mu}_{G\tau}^Y \rangle$.

Defuzzification of TS Y basic fuzzy trend components is according to (8) and (9).

Step 6. Application of basic trend identification algorithm (see Part 3. Algorithm 1) for time series Z and determination of its components $G\tau_Z = \langle \tilde{v}_{G\tau}^Z, \tilde{\alpha}_{G\tau}^Z, \tilde{\mu}_{G\tau}^Z \rangle$.

Defuzzification of TS Z basic fuzzy trend components is according to (8) and (9).

Step 7. Correction of TS prognostic fuzzy elementary trend Y

$$\hat{\tau}_{t+1}^Y = r(\tau_{t+1}^Y, G\tau_Y, G\tau_Z): \\ \hat{\tau}_{t+1}^Y = v_{t+1} \cdot a_{t+1} + v_{G\tau}^Y \cdot a_{G\tau}^Y + v_{G\tau}^Z \cdot a_{G\tau}^Z$$

Step 8. Calculation of corrected prognostic value of numeric TS Y for a one period

$$x'_{t+1} = x_t + \hat{t}_{t+1}^Y.$$

The proposed approach to forecast was tested for short-term forecast of IT companies sales of Ulyanovsk region in Russia (time series Y consists of 12 values), the predictor was expertly selected time series Z as an average number of employees of IT companies engaged in production of new software.

Table I shows the results obtained estimates predict.

Table I. Evaluation of forecasting

Evaluation	Song model[8]	Fuzzy Tend[9]	Proposed model
MSE	0,025	0,202	0.0123

To check the obtained forecast values we used the criterion MSE:

$$MSE = \frac{1}{n} \sum_{t=1}^n (x_t - x'_t)^2$$

The results of the experiment demonstrates that the suggested approach implementing modified method of EFT forecasting might be used for short-term forecasting of time series when there is an expert assumption about existence of predictor time series.

V. CONCLUSION

This article proposes the approach to the analysis of the software market trends in technologies, scientific achievements and user requirement, expressed in time series. The proposed approach allows to uncover ways to create a new useful ideas for software development. We propose the algorithm expanding the opportunities of TS forecasting on the base of fuzzy trends, as the historical software TS are of small length. The experiments carried out demonstrate the functionality and increase of forecast accuracy when applying the suggested algorithm.

The advantages of the proposed approach:

- The results of forest consider the results of the analyzed time series main trend.
- It does not demand for highly qualified users.
- The analyzed time series might be of short length.
- The analyzed time series might be of various length.

The future research will be connected with involving of time series similarity coefficient into forecasting algorithm.

ACKNOWLEDGMENT

This work has been partially funded by the projects no. 13-01-00324 and no. 14-07-00247 of the Russian Foundation for Basic Research.

REFERENCES

[1]. Holt C.C. Forecasting trends and seasonals by exponentially weighted moving averages // O.N.R. Memorandum, Carnegie Inst. of Technology. 1957. № 2.

[2]. Kendall, M. Time series. Translated from English, Yu. P. Lukashina. – Moscow: Finansy I statistika, 1981. 199 p. (rus)

[3]. Anderson T. W. Statistical Analysis of Time Series.. New York: John Wiley and Sons, Inc., 1971.

[4]. Box, J. Time series analysis: Forecasting and control: Translated from English; ed. by F. Pisarenko – Moscow: Mir, 1974. – 406 p (rus).

[5]. Zadeh, A. Lotfi. Fuzzy Sets / Lotfi A. Zadeh // Information and Control. – 1965.

[6]. Sabic, D.A. Evaluation on fuzzy linear regression models / D. A. Sabic, W. Pedrycz // Fuzzy Sets and Systems. 1991. №23. P. 51-63.

[7]. Chen, S. M. Forecasting enrollments based on high-order fuzzy time series/ S.M. Chen // Cybernetics and Systems: An International Journal. 2002. – № 33. P. 1–16.

[8]. Song, Q. Fuzzy time series and its models / Q. Song, B. Chissom // Fuzzy Sets and Systems. 1993. № 54.- P. 269–277.

[9]. Afanaseva T. V., Yarushkina N. G. Fuzzy Time series with fuzzy tendency. In Vestnik Rostovskogo Gosudarstvennogo Universiteta Puty Soobzhenia' (Vestnik RGUPS) Rostov-on-Don, Russia, 2011.- P. 7-16.(rus)

[10]. Perfilieva, I. et al. Relaxed Discrete F-Transform and its Application to the Time Series Analysis / I Perfilieva, N Yarushkina, T Afanaseva // Da Ruanetal (Eds.): Computational Intelligence. Foundations and Applications (Proc.of the 9th Int. FLINS Conf.), pp. 249 --255, World Scientific, Emei, Chengdu, China, 2-4 August, 2010.

[11]. Tanaka, H. Linear regression analysis with fuzzy model / H. Tanaka, S. Uejima, K. Asai // IEEE Transactions on Systems, Man and Cybernetics. 1982. № 12(6). P. 903–907.

[12]. Yarushkina N., et al. Time Series Processing and Forecasting using Soft Computing Tools. - Lecture Notes in Computer Science, Vol. 6743, Proceedings of 13th International Conf. RSFDGrC-2011. Springer-Verlag, Berlin Heidelberg, 2011, XIII.-p. 155-163.

[13]. Perfilieva, I. et al. Soft computing tools for time series analysis and forecast/ I. Perfilieva, N. Yarushkina, T. Afanasieva, A. Igonin, A. Romanov, V. Shishkina Proceedings of the 9th Int. Conf. on Application of Fuzzy Systems and Soft Computing (ICAFS 2010) Eds. R. A. Aliev, K. W. Bonfig, M. Jamshidi, W. Pedrycz, I.B. Turksen, Prague, August 26-27, 2010, VERLAG b- Quadrat Verlag, pp. 50—60.

[14]. Gregory, Allan W.; Hansen, Bruce E. (1996). Residual-based tests for cointegration in models with regime shifts. Journal of Econometrics 70 (1): 99–126.

[15]. Engle, Robert F.; Granger, Clive W. J. (1987). Co-integration and error correction: Representation, estimation and testing. Econometrica 55 (2): 251–276. .

[16]. Granger, Clive. (1981) Some Properties of Time Series Data and Their Use in Econometric Model Specification. Journal of Econometrics 16 (1): 121–130.

[17]. Yarushkina N.G. Osnovy teorii nechetkikh I gibridnykh sistem [The fundamentals of fuzzy and hybrid systems theory]: tutorial / N.G. Yarushkina. – Moscow: Finansy i statistika, 2004. 320 p. (rus)

[18]. Tsenga, F. M. Fuzzy ARIMA model for forecasting the foreign exchange market / F. M. Tsenga, G. H. Tzengb, H. C. Hsiao-Cheng Yua // Fuzzy Sets and Systems. 2001. №118.

[19]. Pedrycz W., Chen S.M. (Eds). Time Series Analysis, Modeling and Applications: A Computational Intelligence Perspective (e-book Google). – Springer-Verlag, Berlin Heidelberg, 2013.- (Intelligent Systems Reference Library, Vol. 47). 404 pp.