

Applying method of phase plane to a fuzzy trend when predicting the project metrics presented by time series

I.A. Moshkina¹, E.N. Egov¹, A.A. Romanov¹

¹Ulyanovsk State Technical University, Severny Venec street 32, Ulyanovsk, Russia, 432027

Abstract. In the article is proposed the method of applying the entropy measure to the fuzzy trend of time series of design metrics. An analysis of project performance is described. A prediction algorithm has been developed and implemented. The algorithm is based on the indices of fuzzy trends in time series. The formula for the entropy measure according to the fuzzy trend for a fuzzy time series is defined. The measure of entropy uses statistical information on the behavior of a fuzzy trend on the phase plane. The hypothesis of trend stability is used for forecasting. Experiments based on this approach are presented.

1. Introduction

Effective management of a set of projects through forecasting their development is one of the most important tasks for the successful organization of project activities. The problem of monitoring the status of project activities can be solved with the help of a tool for measuring the characteristics of project activities. The tool allows you to monitor the dynamics of a set of projects, identify problems and timely adjust regulatory actions [1]. Also, it is need to reallocate resources between projects. Therefore, the creation of a tool for project management based on forecasting the values of project indicators appears to be an urgent and important scientific and technical task. Often project metrics can be expressed in verbal form. Also, metrics are accumulated in the development process. As a result, project metrics can be represented as fuzzy time series (FTS).

The objectives of the FTS analysis include the identification of the general trend of the series, the forecast of the next fuzzy label and the evaluation of the series levels at some point in time. There are situations when it is required to assess the trend of FTS change over a certain period of time. The length of the window is set linguistically ("long", "short", etc.). In this case, the series will have an indistinctness not only on the scale of the value, but also on the time scale. These kinds of problems arise when it is necessary to apply expert knowledge about trends in the values of indicators when analyzing FTS in expert systems or to simulate the operation of the system under certain conditions by applying "if-then" rules.

In addition to the classical time series (TS) analysis methods, there is a direction for analyzing Time-Series DataMining. The main objectives of this direction are modeling and analysis of processes. These processes have a greater degree of uncertainty (including "non-stochastic" process). The methods are aimed at revealing the hidden regularities and extracting new knowledge from the TS. Based on this methodology, an expanded set of TS analysis tasks is solved.

The article is devoted to solving the problem of project monitoring by means of analyzing the state of metrics of the software project. Metrics are loaded into the version control system with subsequent forecasting and their adjustment.

The project management tool consists of the following steps:

- select metrics, by mathematical nature;
- consider the dependencies;
- construct a formal model;
- build a project management tool.

Metrics are properties of entities. The authors propose to use follow metrics extracted from Control Versions System:

- commit (developer, date);
- release (set of developers, master, date);
- branch;
- bug;
- improvement;
- new feature.

The international standards (ISO 10006, ISO 21500) and national (Russia: GOST R 54869 - 2011, GOST R 54870 - 2011, GOST R 54871 - 2011) management standards are used to make decisions in project management. A significant contribution to the analysis of project activities from the point of view of management was made by Klein G., Charette R., SeungHeonHan, Aucoin B. M., Gold M., Norenkov IP, Tarasenko F.P., Dulzon A.A. and others.

Foreign scientists who made a significant contribution to the analysis and prediction of TS are Box D., Jenkins G., Chem S.M., Granger C., Herbst G. and etc.

Foreign and Russian scientists who have made a significant contribution to the field of fuzzy modeling, analysis and prediction of TS are Novak V., Perfilieva I.G., Batyrshin I.Z., Kovalev S.M., Yarushkina N.G., Afanasyeva T.V. and etc [2], [3], [4].

2. Analysis of project activity metrics

The model of analysis and management of a set of projects in the process of project activity is developed.

$$\{C_t, R_t, B_t, I_t, F_t, R^{BI}, R^{IF}\},$$

where C_t – TS commits, R_t – TS of releases, B_t – TS of bugs, I_t – TS of improvement, F_t – TS of NewFeature, R^{BI} – dependence of the number of bugs on improvements, R^{IF} – the dependence of new functional properties on the number of improvements (New Features from improvements).

DiscreteTime Series represent project data.

Model of TS for solving this problem is constructed in two stages:

- A discrete TS $Y = \{t_i, x_i\}$, $i \in [1, n]$, where t_i – some point in time, x_i – the level of TS, is transformed into an fuzzy TS $\tilde{Y} = \{t_i, \tilde{x}_i\}$, $i \in [1, n]$, $X = \{x_i\}$, $\tilde{x}_i \in \tilde{X}$, where \tilde{x}_i – fuzzy label [5].
- Fuzzy TS is transformed into a TS of fuzzy elementary trends. It is indicated by $\tau_i = ETend(\tilde{x}_i, \tilde{x}_{i+1})$, where $ETend$ – functional, which is implemented on the basis of operations: $Ttend$ – determination of the type of trend and $Rtend$ – determination of the intensity of the trend on a special linguistic scale constructed for the initial TS [6].

Several TS, clear numerical and fuzzy linguistic ones are obtained as a result of the described transformations (for example, a TS of parameters of fuzzy trends).

A generalized model of fuzzy TS of a fuzzy trend is presented [7]:

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

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

Every fuzzy tendency $\tau \in \mathfrak{S}$ of a fuzzy TS is represented by a structural model [8]:

$$\tau = \langle \tilde{v}, \tilde{a}, \Delta t, \mu \rangle,$$

where τ – name of a fuzzy tendency from the set \mathfrak{T} , $\tau \in \mathfrak{T}$; \tilde{v} – type of fuzzy trend, $\tilde{v} \in \tilde{V}$, { Fall, Growth, Stability };

\tilde{a} – fuzzy intensity, $\tilde{a} \in \tilde{A}$, { Strong, Medium, Weak }; Δt – duration of fuzzy trend, $\Delta t \in \Delta T$; μ – the membership function of an fuzzy TS section bounded by an interval Δt , a fuzzy trend τ .

The fuzzy tendency of the fuzzy TS is classified as a function of Δt :

- elementary fuzzy tendency $T \in \mathfrak{T}$ ($\Delta t = 1$),
- local fuzzy tendency $NT \in N\mathfrak{T}$ ($1 < \Delta t < n - 1$)
- basic (general) fuzzy tendency $GT \in G\mathfrak{T}$ ($\Delta t = n - 1$).

The TS of the fuzzy elementary trends is represented in the form (TTend, RTend - procedures for obtaining fuzzy sets \tilde{v}_t and \tilde{a}_t):

$$\tilde{v}_t = \text{TTend}(\tilde{x}_t, \tilde{x}_{t+1}), \quad \tilde{a}_t = \text{RTend}(\tilde{x}_t, \tilde{x}_{t+1}), \quad \mu_t = \min(\mu(\tilde{x}_t), \mu(\tilde{x}_{t+1})).$$

The levels of the TS are described by the components of the fuzzy trend - the type and intensity: $\tau = \{t_i, \tau_i\}_{i \in [1, n]}$, where $\tau = \langle \tilde{v}, \tilde{a}, \mu \rangle$.

The clusters $S = \{s_1, s_2, \dots, s_k\}$ of similar fuzzy labels \tilde{x}_i to identify the dominant trend are defined. For each cluster $s \in S$, N is the cardinality.

The FCM-algorithm is used for fuzzy trend clustering. It is based on the minimization of the objective function [9]:

$$J = \sum_{j=1}^C \sum_{i=1}^N \mu_{ij}^m \|x_i - c_j\|,$$

where N - the number of indicators; C – number of clusters; μ_{ij} – membership function of the object i to the cluster j ; m – real number, >1 ; x_i – i -th trend of many trends TS; c_j – j -th cluster; $\|x_i - c_j\|$ – the norm characterizes the distance from the center of cluster j to the tendency i .

The fuzzy elementary trend modeling method is used to predict numerical values and fuzzy trends in the state of an organization project in a given product. The forecast uses hypothesis testing:

Hypothesis 1. The hypothesis of conservation of trend. The Forecast is constructed on base the previous period. The formula for the predicted value

$$\tau_{t+1} = \tau_t + \tau_p,$$

where τ_{t+1} – forecast for the next period of time; τ_t – the real value at time t ; τ_p – the value of the trend over the previous period of time.

Hypothesis 2. The hypothesis of stability of the trend. The moving average is used to predict

$$\tau_{t+1} = \tau_t + G\tau_p,$$

where $G\tau_p$ – the importance of a dominant fuzzy trend.

Consider the trend of the previous selected period. We select the predominant cluster of trends. The forecast for the above formula is calculated. The trend is built. Optimistic forecast for the some number of occurrences of trends used. The highest average trend is selected.

Hypothesis 3. Forecasting for a given period on the basis of fuzzy elementary trends. Stages of the prediction algorithm for the period based on trends:

The expert sets the number of considered trends for the previous period. For example, for half a year - a set of trends A . Either he sets the pattern set of trends. The presumed trend following this set is known.

$$\{\tau_{t_{n-m}}, \dots, \tau_{t_{n-1}}, \tau_{t_n}\}$$

1. Search for a set of trends A in all other previous periods.

$$\{\tau'_{t_{n-l-k}}, \dots, \tau'_{t_{n-l-(k-1)}}, \tau'_{t_{n-l}}\}$$

2. If such a set of B is found in which the C trend is located after this found set B then trend C is considered into account. The forecast equal to the trend C is constructed.

$$\tau_{t+1} = \tau_t + \tau'_{t_{n-l+1}}$$

3. If the set B , which would coincide with the set A , was not found then the search for the set is repeated, but it is already not looking for its complete coincidence. We select new pattern A is shorter into one trend. This is repeated until a suitable set of trends B [10].

4. To select the best hypothesis, we proposed to use the phase plane method of the fuzzy trend [11], [12].

3. Forecasting ts using the adaptation algorithm

To predict new values, the phase plane method of the fuzzy trend is used [13]. The method uses a change in the fuzzy trend in the phase plane. For this, each fuzzy trend is given a weight. To construct the phase plane by a fuzzy trend, each trend is given a value, weight. The rate of increment of the trend is defined as the difference between the weight of the trend in the previous point and the weight of the trend at the current point.

It is necessary to unify the information on the relationship between the weight of the trend and its speed. The encoding of possible locations of points on the coordinate plane is introduced (table 1).

Table 1. Scheme for coding points on the phase plane.

Code	Weight trends	Trend dynamics	Point in the plane	Interpretation
0	=0	=0	origin	A point at the origin of coordinates means the system is in a stable state
1	>0	>0	first quarter	In the system there is growth or stabilization after the fall
2	<0	>0	second quarter	A smaller drop after a big fall
3	<0	<0	third quarter	In the system there is a fall or stabilization after growth
4	>0	<0	fourth quarter	Smaller growth after large growth
5	=0	>0	of axes X is 0	Stabilization after falling (appears only after 3 states)
6	=0	<0	of axes X aligns to 0	Stabilization after growth (appears only after 1 state)
7	>0	=0	of axis Y from 0	Growth after growth
8	<0	=0	of axes Y lower than 0	Fall after fall

The next step is to get the predicted value. Determination of the correspondence between the displacement of points on the phase plane and trends. The trend values for each code are stored. The following rules are obtained (table 2).

Table 2. Rules for the transition to phase plane.

Stability	Growth	A fall
0 or 5 or 6 -> 0	0 or 2 or 3 or 4 or 5 or 6 or 8 -> 4	0 or 1 or 2 or 3 or 4 or 5 or 6 or 7 -> 2
1 or 4 or 7 -> 5	1 or 4 or 7 -> 7	2 or 8 -> 8
2 or 3 or 8 -> 6	4 -> 1	2 -> 3

From these rules, a number of conclusions are drawn:

1. If the point of the phase plane for the predicted point lies in the center of coordinates (code 0) or on the X axis (codes 5 and 6), the value of the fuzzy trend will correspond to the Stability.
2. If the point is in the first or fourth quarter, or lie on the Y axis above 0, then the value of the fuzzy trend will match Growth. The intensity can be taken into account. In this case, the Growth trend is broken down into 3 growth trends, taking into account the intensity: Weak, Medium or Strong.
3. If the point is in the second or third quarters, or on the Y axis below 0, then the value of the fuzzy trend will match the Fall. The intensity can be taken into account. In this case, the tendency of the Fall is divided into 3 tendencies of the Fall, taking into account the intensity: Weak, Medium or Strong.

When analyzing the TS, the schemes of the displacement of a point on the phase plane are determined. The frequency of the transition of a point from one area (labeled code) to another is determined. On the basis of frequency, a frequently encountered displacement pattern is determined. In this scheme, the code of the starting point is equal to the location code of the point on the phase plane for the last known point BP. By the code value for the end point, the scheme determines possible trends (there

may be more than one). The predicted values are then calculated on the basis of these trends. The drawbacks of the method include the acquisition of several equally probable values from which you want to choose one.

4. Experiments

The metrics of the events of the open project "FreeNAS10" were taken for research. The time series of "closed" "Bug" and "Feature" were taken [14]. The results of the similarity of these metrics for determining the dependencies are presented in Table 3. The results of the analysis and forecasting of project activity metrics are presented in Tables 3 and 4. Forecasting the appearance of errors in the FreeNAS10 project, taking into account the effect of adding new functionality to the project (the hypothesis of trend stability) is shown in Figure 1.

Table 3. Analysis of project data.

Label type	The general trend	The dominant trend	A measure of similarity	Correlation	Interpretation of correlation
Bug	Fall	Stability	0.82	0.9736	Strong
New Feature	Fall	Stability			

Table 4. The results of forecasting taking into account the influence of the predictor time series.

Dependent time series	Time series predictor	Hypothesis 1	Hypothesis 2	Hypothesis 3
Bug	New Feature	Stability	Growth average	Falling strong

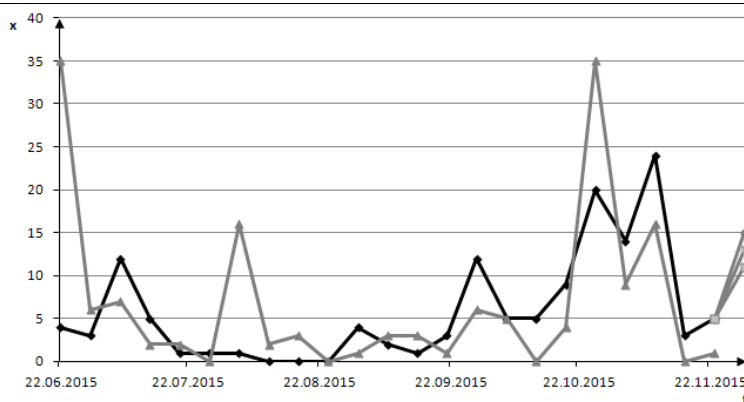


Figure 1. Forecasting in the project "FreeNAS10" (hypothesis 2).

Figure 1 shows a variant of prediction using hypotheses, where the x – axis is the TS value, the t – axis is the time; "◆◆" Stands for TS Y (BUG); "■" Forecast without adjustment; "×" Forecast after adjustment TS Y (BUG) for the main trend TS Y (BUG); "●" Forecast after adjustment TS Y (BUG) for the main trend TS predictor (New Feature).

We obtain the predicted values for the time series presented by the phase plane of the fuzzy trend method. It is presented in Table 5.

Table 5. The predicted values of the phase-plane of the fuzzy trend method.

TS	Forecast
New Feature	Fall
Bug	Stability

On the basis of the obtained data, we can draw the following conclusion. In the first, one should choose the hypothesis of preserving the trend. This is confirmed by the forecast by the phase plane of

the fuzzy trend method. In the second case, there is a discrepancy in the forecast by the method of the phase plane of the fuzzy trend with the hypothesis of maintaining the trend. The forecast by the method of the phase plane of the fuzzy trend coincides with the hypothesis of the dominant trend. This hypothesis should be chosen to predict the value in the second case.

5. Results

The research of a set of metrics series of various projects was conducted with the aim of predicting subsequent values. Based on the obtained values, the average value of MAPE for the developed method was determined. It is no more than 0.5-1.5% of the average value of the forecast metrics.

The proposed approach is a modified method for predicting fuzzy elementary trends. The obtained results show that this approach can be used for short-term forecasting of FTS. The approach uses the capabilities of the FTS. This allows you to predict the values for series with non-numeric values.

6. References

- [1] Moshkin, V.S. Intellectual Data Analysis and the Ontological Approach in Project Management / V.S. Moshkin, A.N. Pirogov, I.A. Timina, V.V. Shishkin, N.G. Yarushkina // Automation of control processes. – 2016. – Vol. 4(46). – P. 84-92.
- [2] Perfilieva, I. Time series analysis using soft computing methods / I. Perfilieva, N. Yarushkina, T. Afanasieva, A. Romanov // International Journal of General Systems. – 2013. – Vol. 42(6). – P. 687-705.
- [3] Afanasieva, T. Time series forecasting using fuzzy techniques / T. Afanasieva, N. Yarushkina // International Joint Conference IFSA-EUSFLAT, 2015. – P. 1068-1075.
- [4] Novák, V. A general methodology for managerial decision making using intelligent techniques / V. Novák, I. Perfilieva, N. Yarushkina // Chapter Recent Advances in Decision Making. Series Studies in Computational Intelligence. – 2009. – Vol. 222. – P. 103-120.
- [5] Herbst, G. Online Recognition of fuzzy time series patterns / G. Herbst, S.F. Bocklisch // International Fuzzy Systems Association World Congress and European Society for Fuzzy, 2009.
- [6] Kacprzyk, J. Using Fuzzy Linguistic summaries for the comparison of time series / J. Kacprzyk, A. Wilbik // International Fuzzy Systems Association World Congress and European Society for Fuzzy Logic, 2009.
- [7] Pedrycz, W. Time Series Analysis, Modeling and Applications: A Computational Intelligence Perspective (e-book Google) / W. Pedrycz, S.M. Chen // Intelligent Systems Reference Library. – 2013. – Vol. 47. – P. 404.
- [8] Afanasieva, T.V. Forecasting local time series trends in large data analysis problems // Automation of control processes. – 2014. – Vol. 4(38). – P. 42-47.
- [9] Krol, T.Ya. Methods for solving the problem of clustering and forecasting in an electronic archive / T.Ya. Krol, M.A. Charin // Young Scientist. – 2011. – Vol. 1(6). – P. 135-137.
- [10] Timina, I.A. Correction of forecast hypotheses to extract knowledge of time series // XIV National Conference on Artificial Intelligence with International Participation, 2014. – P. 68-76.
- [11] Egov, E.N. Fuzzy modeling and genetic optimization of time series in the intellectual system of technical diagnostics / E.N. Egov, N.G. Yarushkina, D.V. Yashin // Radiotekhnika. – 2016. – Vol. 9. – P. 64-71.
- [12] Yarushkina, N.G. Model and controls for designing an automated system based on the dynamics of program code metrics / N.G. Yarushkina, I.A. Timina // Automation of control processes. – 2015. – Vol. 3(41). – P. 73-81.
- [13] Timina, I.A. Identification of anomalies of time series of metrics of project activity based on entropic measures / I.A. Timina, E.N. Egov, Yu.P. Egorov, D.V. Yashin, S.K. Kiselev // Radio engineering. – 2017. – Vol. 6. – P. 128-135.
- [14] FreeNAS 10. Bug Tracking System [Electronic resource]. – Access mode: <https://bugs.pcbsd.org/projects/freenas-10> (30.11.2018).

Acknowledgments

The article was supported by the Russian Foundation for Basic Research (grant No. 19-07-00999).