Software of Time Series Forecasting based on Combinations of Fuzzy and Statistical Models

T. Afanasieva, A. Sapunkov, A. Afanasiev
Information Systems Department
Ulyanovsk State Technical University
Ulyanovsk, Russia

Abstract—The developed software is a web application with open access and is aimed on forecasting of time series stored in database. We proposed approach of time series forecasting, combined ARIMA models with fuzzy techniques: three fuzzy time series models, fuzzy transformation (F-transform) and ACL-scale. Applications of a proposed web service have demonstrated efficiency in practical time series predictions with suitable accuracy.

Keywords—time series; fuzzy time series; software; forecasting model

I. INTRODUCTION

Time series data mining is considered as a set of techniques concerned with description, modeling and forecasting at a different level of representation. There are a lot of techniques and tools of statistical description, modeling and forecasting, i.e. ARIMA (and its modifications) [1], when time series is considered as a stochastic process with stationary or non-stationary properties [2]. In recent years in Time series data mining fuzzy models have become very popular due to the ability of human orientation and high-level interpretation. Hybrid time series forecasting methods combine fuzzy and other soft computing techniques. In the paper [3] the overview of the studies in time series forecasting is presented. The authors describe the combinations of soft computing and artificial neural networks tools (in particular fuzzy artificial neural networks). Three fuzzy time series models and fuzzy transformation by F-transform [4] are described in our work [5]. In this work the selection of the best time series forecasting model is based on the concept of fuzzy tendency. The application of different combinations of exponential and fuzzy time series models with some weights is proposed in [6] and is successfully applied in CIF-2015 competition [7].

There are some toolboxes in well-known Math Software described in review [8] that contained Fuzzy logic functions such as fuzzification, defuzzification, fuzzy inference. These functions of Free and Open-Source Tools can be successfully used in time series forecasting algorithms realized as a desktop application. However, it must be noted that the design of such application requires a professional programming.

Among the stable trends in software engineering a development of a web services is considered as a tool providing wide access for users. These users can be domain experts or managers without knowledge in programming.

Therefore the development of web service applications for time series forecasting on the basis of soft computing focused on fuzzy techniques will allow researchers and trainees to be closer to new forecasting methods.

The web service for forecasting of financial indicators of enterprise with fuzzy techniques described in [9] is an example of such web service. The linear autoregression, artificial neural network, F-transform [4] and fuzzy time series model are used as soft computing techniques in that web service for forecasting of financial indicators. In this paper we described new time series forecasting software designed as a web service. The following soft computing techniques are implemented in developed web service: three basic fuzzy time series models, considered in [10], statistical ARIMA (SARIMA) models [1], F-transform [4] for trend extraction and ACL-scale [11] as a tool for time series fuzzification. The structure of this contribution is as follows. First we briefly describe the applied soft computing techniques and represent MTSFA approach of time series forecasting. In Section III we present some aspects of the web service - its architecture and technologies that we applied. In section IV we demonstrate some results of time series forecasting which we get using developed software. Finally the conclusions will be done and future work will be considered.

I. SOFT COMPUTING TECHNIQUES IN TIME SERIES FORECASTING

A. ACL-scale as a tool for time series fuzzification

Some tool of time series fuzzification is necessary in forecasting techniques based on the fuzzy time series models. ACL-scale presented in [11] is one of such tools which transforms numerical time series into fuzzy time series.
Let \( X = \{ x_t \in W, W \subseteq \mathbb{R}, t = 1, 2, ..., n \} \) be a numerical time series. Let suppose \( r \) is a number of the partially ordered fuzzy intervals of equal length that cover a set \( W \).

Suppose these partially ordered intervals on the set \( W \) are used for building a linguistic variable \( \tilde{X} \)

\[
\tilde{X} = \{ \tilde{x}_t, \ t = 1, 2, ..., r, r < n \}. \tag{1}
\]

For given \( x_t \in W \) and for constructed linguistic variable \( \tilde{X} \) on the set \( W \) ACL-scale determines a linguistic value \( \tilde{x}_t \in \tilde{X} \). The set of these values \( \{ \tilde{x}_t, t = 1, 2, ..., n \} \) is considered as a fuzzy time series.

**B. Three fuzzy Time series models**

We apply three basic fuzzy time series models with symbolic labels, considered in [5]: S-model with fuzzified time series values [12], D-model with fuzzified first differences of time series values [13] and T-model, based on the elementary fuzzy tendencies [5]. A fuzzy logical relationship is employed in these fuzzy time series models on the assumption that the observation at time \( t \) is accumulated results of the observations at the previous times [12]. The fuzzy time series model, on fuzzified time series values, according to [12], is defined as a time-invariant model:

\[
\tilde{X}_t = (\tilde{x}_{t-1} \times \tilde{x}_{t-2} \times ... \times \tilde{x}_{t-p}) \ast R(t, ..., t - p), \tag{2}
\]

where \( \times \) is the Cartesian product, \( R(t, ..., t - p) \) is the fuzzy TS model as fuzzy relation, which can be calculated by Mamdani’s algorithm, \( p \) is the order of the model, \( \ast \) is the max-min composition. This time-invariant model is applied to other fuzzy time series determined for D-model and for T-model by an ACL-scale.

The order of the model determines the number of previous time points, which are taken in account. To obtain fuzzified values in above mentioned fuzzy time series models an ACL-scale [11] are applied. Let us consider some features of this fuzzification on the basis of ACL-scale in respect to three fuzzy time series models.

In S-model the fuzzified time series values \( \Delta \tilde{X}_t \in \tilde{X} \) are performed in respect to a given time series \( X = \{ x_t \in W, W \subseteq \mathbb{R}, t = 1, 2, ..., n \} \), to a set \( W \) and to a constructed linguistic variable \( \tilde{X} \) on the set \( W \).

D-model requires a simple transformation of a given time series \( X \) to a time series of first differences \( \Delta X = (x_t - x_{t-1}), \Delta x_t \in \Delta W, \Delta W \subseteq \mathbb{R}, t = 2, 3, ..., n \). Then a linguistic variable \( \Delta \tilde{X} \) is built on the set \( \Delta W \) and is applied to determine fuzzified first differences of time series values \( \{ \Delta \tilde{x}_t | t = 2, 3, ..., n \} \).

The elementary fuzzy tendency in T-model according to definition in [5] has two components: type and intensity. For these components two linguistic variables \( \tilde{V} \) and \( \tilde{A} \) are defined respectively. A linguistic variable \( \tilde{V} \) is built on the set \( W_V = [-\Delta x, \Delta x] \subseteq \mathbb{R} \), where \( \Delta x = \max |x_i - x_j|, x_i, x_j \in X, i, j = 1, 2, ..., n \). A set \( W_A = [0, \Delta x] \subseteq \mathbb{R} \) is used to construct a linguistic variable \( \tilde{A} \). Thus time series of elementary fuzzy tendencies is represented by two fuzzy time series

\[
\{ \tilde{V}_t | t = 2, 3, ... \}, \{ \tilde{A}_t | t = 2, 3, ... \}.
\]

**C. Fuzzy transformation of Time series**

Fuzzy transform (F-transform) was proposed in [4] as an approximate technique of the original function and then was applied in time series trend extraction. This property of F-transform is useful in time series decomposition and forecasting. We apply direct and inverse F-transform for a given time series \( \tilde{X} = \{ x_t \in W, W \subseteq \mathbb{R}, t = 1, 2, ..., n \} \) to obtain piecewise linear representation of time series trend \( f_t \) using pre-defined set of basic functions \( A = \{A_1, A_2, ..., A_q\}, q < n \). This set of basic functions can be considered as a linguistic variable that defined on the set \( T \subseteq \mathbb{R}, [1, 2, ..., n] \subseteq T \). Basic F-transform computes the value of Time series trend using following expressions (first one for direct and second one for inverse transformation) [4]:

\[
F_k(X) = \frac{\sum_{t=1}^{n} x_t A_k(t)}{\sum_{t=1}^{n} A_k(t)}, \quad f_t = \sum_{t=1}^{n} F_k(X) * A_k(t) \tag{3}
\]

**D. Time series forecasting approach**

The proposed time series forecasting approach (named MTSFA approach) implements soft computing, which combines fuzzy techniques described above with well-known statistical model ARIMA and its modification for seasonal time series (SARIMA) [1].

We assume that a time series can be defined in the additive form by following models as a combination of the time series components:

\[
x_t = f_t + \xi_t, \tag{4}
\]

\[
x_t = f_t + \psi_t + \xi_t, \tag{5}
\]

where \( x_t \) is a given time series, \( f_t \) is a time series trend, expressed as the inverse F-transform components [4] and modeled by fuzzy time series models from the set \( S_f = \{ \text{S-model [12], D-model [13], T-model [5]} \} \), \( \psi_t \) is a time series of residuals, modeled by the models from the set \( S_a = \{ \text{S-model [12], D-model [13], T-model [5]} \} \), ARIMA-model [1], SARIMA-model [1]), \( \xi_t \) is a random white noise.

For each of the above described time series component the adequate and accurate model is identified from the sets \( S_f \) or \( S_a \). The identification means the process, where the order and the type of the best forecasting model for each time series component are defined on the training part of time series.
To choose the best time series model from the sets \( S_f \) and \( S_a \) the minimum out of sample SMAPE criterion is used:

\[
SMAPE = \frac{1}{n} \sum_{t=1}^{n} \frac{|\hat{F}_t - x_t|}{|x_t| + |\hat{F}_t|}/2
\]

where \( \hat{F}_t \) is the actual predicted values, produced by the best combination of components of time series model for the given time series; \( n \) is the number of predicted points out of sample; \( x_t \) is the real values of the unknown for model part of the time series.

II. DESCRIPTION OF A PROPOSED SOFTWARE “SALX.FUZZYFMF”

MTSFA approach, presented in Section II.D defines the main functionality of the developed software. In this Section we will describe its architecture and technologies in detail. To bring soft computing techniques of time series forecasting closer to domain experts and managers we have designed the software in the form of web service with open access (http://salx.pw). The component-oriented architecture of the developed software is depicted in Fig. 1.

![System architecture of a Time series forecasting software based on soft computing techniques](image)

Fig. 1. System architecture of a Time series forecasting software based on soft computing techniques

The software includes two main components: a web service server and API (Application Programming Interface) server. The core technology of web service server is MVVM (model - view - view model). To reduce the query number to a web server there was made a part of computations on the client side. The client side of web service is based on the pattern SPA (Single Page Application) and implemented the libraries jQuery и AngularJS 1.2.9 [14]. The Web service receives data of the Time series through Web API server and is used to render the work of “Library of Time series forecasting techniques” component. Web service and Web API server were created using ASP.NET [15] technology. All time series must be stored in the MS-SQL (Microsoft SQL Server [16]) database. Up to now time series database consists of time series, used in the competitions CIF-2015 and CIF-2016 [7].

To work with the “Time series database” component, the “data source view” component is used. The interactions with time series database occur through this component. It also converts the format of data stored in a database in a model used in other components and vice versa.

A. Components of user interface

The developed web service includes two major components for interaction with the user (Fig.1):

“Forecasting model selection interface” (when web service is activated the “Select model” corresponds to this component). This component is designed for selection and for study of fuzzy time series forecasting models only. It consists of four zones. The first one is assigned to time series. In this zone user can load or select the already downloaded time series using a drop-down list and then can see its graphical representation. To work with the custom time series it is necessary to load the file with a time series previously. Representation of the values of the time series from a new line is the basic requirement for the format of the user file. The second zone is associated with the legend place of the time series graph and is used to display the chosen fuzzy time series model. The parameters of a fuzzy time series model and parameters of a prediction, which can be set by user, are located at the right zone. Below there are shown the values of accuracy criterion SMAPE.

To start forecasting of a selected time series in this interface a one of the time series models from the set \( S_f = \{S\text{-model [12], D-model[13] and T-model[5]} \} \) must be chosen and the parameters of prediction must be set by the user: horizon of prediction, an order \( p \) of fuzzy model, a number of points used as test part of time series. After that a label of a chosen fuzzy model with its order will appear at a legend place in a second zone. To obtain the graphical representation of predicted time series the chosen fuzzy model at a legend place must be activated. An average time of constructing of a fuzzy model in “Forecasting model selection interface” ranges from 200 ms (for a short time series having length less then 30 points) to 400 ms (for a large time series having length up to 1000 points).

“Forecasting models comparing interface” (when the web service is activated the “Compare model” corresponds to this component). This component applies the MTSFA approach and automatically performs varies combinations of time series models in respect to (4) and (5) with (or without) F-transform.
The following pre-defined parameters are fixed and unavailable to the user: a horizon of prediction is equal to length of test part of time series and is 10% of a time series length; an order of fuzzy models ranges from 1 to 3; the number of basic function is set first as 13 and then as 25.

This component is divided into four zones located from top to down. The first one enables user to select the already downloaded time series using a drop-down list. Then two graphical representations of a selected time series in the second zone appear: the first one consists of a selected time series only, in the second one selected series with predicted time can be displayed. Simultaneously all of the combinations with identified time series models in respect to (4) and (5) with (or without) F-transform are created and their designations appear in the third zone, which is located below the second one.

As about of 56 combinations of time series models are performed simultaneously with graphical representation of a selected time series it requires some time. An average time of constructing combinations and of visualization of a selected time series range from 30 seconds (for a short TS, having length less than 30 points) to 90 seconds (for a large TS, having length up to 1000 points).

In Fig. 2 the part of “Forecasting models comparing interface” is depicted. We can see the second zone, which consists of selected series and selected with predicted time series. Below the third zone is located with the list of combinations of time series models (the combinations used to visualize predictions are displayed more brightly).

In the third zone is located with the list of combinations of time series models (the combinations used to visualize predictions are displayed more brightly). This information is depicted. This information is represented in the interface (see Fig. 3). Here the ordered values of the criterion SMAPE shown in the training part (SMAPE_i) and on the test part of ts 49 are shown. In the first column the designations of model combinations are represented, then the values of criterion of forecasting accuracy are depicted. This information confirms the previous conclusion about the adequate combination (7) for forecasting of time series ts 49.

Let us consider a format of the designation of model combination on an example in accordance to Fig. 2:

CIF-2016-full: ts49_F(13)_D(1)+CIF-2016-full:ts49_F(13)_o_S(3) (7)

This example corresponds to the combination (5) represented in Section II.D. Here “CIF-2016-full: ts49” denotes the designation of a selected time series, F(13) denotes identified F-transform technique[4] with 13 basic functions for trend extraction, D(1) denotes identified D-model [13] with order p=1 for trend component, o_S(3) denotes that S-model [12] with order p=3 was identified for the residuals.

At the fourth zone located below the third one the values of the criterion SMAPE are shown for every received time series combination.

Among derived combinations of the identified models from the set Sa (see Section II.D) the combination with the minimum SMAPE in test part of time series can be received in two ways. The first way is the way of visual comparison of a behavior of a selected time series with the behavior of a predicted time series (one or several) on the same graph (in the second zone). This allows us to compare the results of forecasting models of different types (see Fig. 2). To perform that the user must activate the designation of combination represented in the third zone, after that the predicted time series will appear in the graph (in the second zone). As one can see in the Fig. 2 there is a good compliance in behaviors of the selected and the predicted time series. So, the combination (7) can be used in forecasting of time series ts 49.

The second way to select the best combination is to analyze the values of the criterion SMAPE shown in the fourth zone of the interface (see Fig. 3). Here the ordered values of the criterion of forecasting accuracy SMAPE (6) on the training part (SMAPE_i) and on the test part of ts 49 are shown. In the first column the designations of model combinations are represented, then the values of criterion of forecasting accuracy are depicted. This information confirms the previous conclusion about the adequate combination (7) for forecasting of time series ts 49.

Fig. 2. The second and third zone of “Forecasting models comparing interface” for time series ts 49 from dataset of CIF-2016

Let us consider a format of the designation of model combination on an example in accordance to Fig. 2:

CIF-2016-full:ts49_F(13)_D(1)+CIF-2016-full:ts49_F(13)_o_S(3) (7)

This example corresponds to the combination (5) represented in Section II.D. Here “CIF-2016-full: ts49” denotes the designation of a selected time series, F(13) denotes identified F-transform technique[4] with 13 basic functions for trend extraction, D(1) denotes identified D-model [13] with order p=1 for trend component, o_S(3) denotes that S-model [12] with order p=3 was identified for the residuals.

At the fourth zone located below the third one the values of the criterion SMAPE are shown for every received time series combination.

Among derived combinations of the identified models from the set Sa (see Section II.D) the combination with the minimum SMAPE in test part of time series can be received in two ways. The first way is the way of visual comparison of a behavior of a selected time series with the behavior of a predicted time series (one or several) on the same graph (in the second zone). This allows us to compare the results of forecasting models of different types (see Fig. 2). To perform that the user must activate the designation of combination represented in the third zone, after that the predicted time series will appear in the graph (in the second zone). As one can see in the Fig. 2 there is a good compliance in behaviors of the selected and the predicted time series. So, the combination (7) can be used in forecasting of time series ts 49.

The second way to select the best combination is to analyze the values of the criterion SMAPE shown in the fourth zone of the interface (see Fig. 3). Here the ordered values of the criterion of forecasting accuracy SMAPE (6) on the training part (SMAPE_i) and on the test part of ts 49 are shown. In the first column the designations of model combinations are represented, then the values of criterion of forecasting accuracy are depicted. This information confirms the previous conclusion about the adequate combination (7) for forecasting of time series ts 49.

Fig. 1. The fourth zone of “Forecasting models comparing interface” for time series ts 49 from dataset of CIF-2016

B. Library of time series forecasting techniques

All transformations and calculations are performed in the “Library of Time series forecasting techniques” component.
Let us consider the “Library of Time series forecasting techniques” component in details (see Fig. 4).

The “Library of Time series forecasting techniques” consists of many individual components, which are based on “Common classes” library. Thus, they readily react to each other at the code level, they can be easily integrated into third-party software and any component may be processed using other algorithms which will not affect the functionality of the entire system. “Common classes” library includes primitive objects such as a time series class, a point of time series class, a fuzzy time series class, a point of fuzzy time series class, as well as logics for basic interaction of these classes.

There are six major components in the “Library of Time series forecasting techniques”: fuzzification tools [11], fuzzy inference system, library of fuzzy models, library of statistical models (ARIMA-models), F-transform tool [4], calculation accuracy criteria component.

To convert a numerical time series into a fuzzy time series, “Fuzzification tools” component is used; it implements the logics of an ACL-scale [11], described in section II.A. Pre-defined parameters of an ACL-scale are as follows: the number of fuzzy sets of a linguistic variable \( r = 10 \); the fuzzy sets in linguistic variables are determined as the triangular fuzzy numbers performed on the sets, different for different fuzzy models, represented in section II.B.

The input module receives a time series and converts it into a fuzzy time series which is further used in “Fuzzy inference system” component.

The “Fuzzy inference system” component is intended for a construction of a fuzzy time series model from the set \( S_f \) using the algorithm of fuzzy inference, proposed by Mamdani in the paper [17]. Then this fuzzy time series model is applied in real forecasting of a time series. The parameters, which the user must set is the number of points of forecast. The defuzzification of a fuzzy value obtaining in fuzzy inference system is implemented in “Fuzzy inference system” component too and is made using centroid method [12].

The “Library of statistical models forecasting” component includes well-known statistical ARIMA and SARIMA models [1].

To extract trend and to derive time series decomposition of a MTSFA approach the “F-transform tool” component was developed. The formulas of a direct and an inverse F-transform were presented in Section II.C. The user can set the parameter \( q \), corresponding to the quantity of the basic functions.

It is necessary to mark that a time series before prediction will be divided into two parts: a training part and a test part. The training part is used to construct and to identify time series model (fuzzy or statistical). The test part of a time series is used to test the identified time series model and to calculate the accuracy of the model. Pre-defined length of a test part is determined as 10% of a time series length. On the basis of the training and of the test parts the coefficient SMAPE shows the accuracy of forecasting models. The ”Calculation accuracy criteria” component is used for calculating the criterion of accuracy by the coefficient SMAPE, presented in section II.D.

All the components of the “Library of Time series forecasting techniques” except statistical model are developed in C# using .Net Framework 4.5.1. The statistical time series models were implemented with the functions of the language R [18].
C. Examples of time series forecasts

Below there is presented the application of developed software to time series forecasting based on MTSFA approach (in the component “Forecasting models comparing interface”).

The examples of time series forecasts by the developed software are depicted in Fig. 5, 6, 7, 8.

The results of time series forecasting received by implemented soft computing techniques based on MTSFA approach in developed software may be considered as good. One can see compliances in behaviors of the selected and predicted time series as it is shown in Fig. 2, 5-8.

The developed software was applied to perform real forecasts of the dataset of the competition CIF-2016 [7]. The dataset consists of 72 time series of different length and different behaviors. After the real forecasts were received the real values of time series of that dataset became available. Then the accuracy of real forecasts on the opened time series dataset CIF-2016 [7] was calculated. An average forecasting accuracy of MTSFA approach, measured by SMAPE was 0.165 which was achieved due to combinations of soft computing techniques.

CONCLUSION

The paper describes new software in the form of web service for time series forecasting. This web service provides a wide access to soft computing tools to solve the problems of selection and application of the best forecasting time series model (simple or complex) from the set of the available ones.

The component based architecture enables to add new time series models. Two user interfaces provide time series forecasting in two different ways: by an automatic construction of a set model combinations and by setting parameters of fuzzy models by the user. The developed software may be useful in research, in education and in predicative analytics. The aim of the future work is to reduce the time of creation of the time series model combinations and to increase forecasting accuracy.

ACKNOWLEDGMENT

This work was supported by the Russian Foundation for Basic Research grant 16-47-732112 “The study and development of forecasting methods for time series based on multi-model approach”.

REFERENCES


Published online 09-Feb-2018

Copyright © Afanasieva, Sapunkov, Afanasiev 2018

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731


AngularJS. Available at: https://angularjs.org (accessed on 6 February 2017).

ASP.NET. Available at: https://www.asp.net (accessed on 6 February 2017).

