Forecasting Time Series Using Fuzzy Transform

Irina Perfilieva, Vilém Novák, Viktor Pavliska, Antonín Dvořák, Martin Štěpnička

***Abstract*— A new methodology for forecasting time series which is based on combination of two techniques: fuzzy transform and perception-based logical deduction is proposed.**

I. Introduction

We propose a new methodology for forecasting time series which is based on combination of two techniques: *fuzzy transform* and *perception-based logical deduction*. The technique of fuzzy transforms has been introduced in [8] and then elaborated in e.g., [6], [7]. The computation technique based on logical deductions is presented in [2], [5].

The proposed methodology consists of two phases: analysis of a time series and its forecast. In the first phase, a time series is decomposed into two components, namely its trend and residua. The trend is represented either by a vector of fuzzy transform components, or by the inverse fuzzy transform (see below). By the residuum we understand the difference between the original and the corresponding trend value of the time series.

In the second phase, both trend as well as residua are forecast and then put together. We use one of three possibilities: second order fuzzy transform, extrapolation of the inverse fuzzy transform, or perception-based logical deduction. Forecast of the residua is obtained by a linear combination of previous residua using optimization. A number of parameters are involved in this methodology. They are obtained by training. The best combination of parameters is taken for the final forecast. We outline the technique of fuzzy transform in Section II and the technique of perception-based logical deduction in Section III. The time-series forecasting method is presented in Section III.

II. Preliminaries. Fuzzy (F)-Transform

In this section, we will briefly describe the main theoretical tools that have been used in the proposing time series analysis and forecast.

The *fuzzy transform* (F-transform) has been introduced in [8] for continuous functions and later extended to functions defined at finite set of points [6], [7]. In that case, we called it the *discrete fuzzy transform*. For the time-series analysis we need the discrete fuzzy transform only and so, we will omit the adjective “discrete” in the sequel.

We will use the ordinary algebra of reals and fix an interval $[a, b]$ and a linear space V_l of functions defined on $[a, b]$ at fixed points $p_1, \dots, p_l \in [a, b]$, $l \geq 3$. For these functions we will define the direct and inverse fuzzy transform with respect to a fuzzy partition of $[a, b]$. The direct fuzzy transform of a function from V_l is its homomorphic image in the space of n-dimensional vectors, $n < l$. The inverse F-transform (given by the inversion formula), converts an n-dimensional vector into a function from V_l which approximates the original one.

A. *Fuzzy partition of* $[a, b]$

The fuzzy transform of a function from V_l is defined with respect to a *fuzzy partition* of its domain, i.e. the interval $[a, b]$. The details are presented below.

Fuzzy sets on $[a, b]$ are identified with their membership functions, i.e., they are mappings from $[a, b]$ into $[0, 1]$. If A is a fuzzy set on $[a, b]$ then we write $A \underset{\sim}{\subset} [a, b]$.

Definition 1
Let $x_1 < \dots < x_n$ be fixed nodes within $[a, b]$, such that $x_1 = a$, $x_n = b$ and $n \geq 2$. We say that fuzzy sets $A_1, \dots, A_n$, identified with their membership functions $A_1(x), \dots, A_n(x)$ defined on $[a, b]$, constitute a fuzzy partition *of $[a, b]$ if they fulfill the following conditions for $k = 1, \dots, n$:*

1) *$A_k : [a, b] \longrightarrow [0, 1]$, $A_k(x_k) = 1$;*
2) *$A_k(x) = 0$ if $x \notin (x_{k-1}, x_{k+1})$ where for the uniformity of denotation, we put $x_0 = a$ and $x_{n+1} = b$;*
3) *$A_k(x)$ is continuous;*
4) *$A_k(x)$, $k = 2, \dots, n$, strictly increases on $[x_{k-1}, x_k]$ and $A_k(x)$, $k = 1, \dots, n-1$, strictly decreases on $[x_k, x_{k+1}]$;*
5) *for all $x \in [a, b]$*

$$\sum_{k=1}^{n} A_k(x) = 1. \tag{1}$$

The membership functions $A_1, \dots, A_n$ are called basic functions.

Let us remark that basic functions are specified by a set of nodes $x_1 < \dots < x_n$ and the properties 1)–5). The shape of basic functions is not predetermined and therefore, it can be chosen additionally according to further requirements (e.g. smoothness).

The following formulas represent a fuzzy partition of $[x_1, x_n]$ given by n triangular membership functions:

$$A_1(x) = \begin{cases} 1 - \frac{(x - x_1)}{h_1}, & x \in [x_1, x_2], \\ 0, & \text{otherwise}, \end{cases}$$

$$A_k(x) = \begin{cases} \frac{(x - x_{k-1})}{h_{k-1}}, & x \in [x_{k-1}, x_k], \\ 1 - \frac{(x - x_k)}{h_k}, & x \in [x_k, x_{k+1}], \\ 0, & \text{otherwise}, \end{cases}$$

$$A_n(x) = \begin{cases} \frac{(x - x_{n-1})}{h_{n-1}}, & x \in [x_{n-1}, x_n], \\ 0, & \text{otherwise}. \end{cases}$$

where $k = 1, \dots n - 1$, and $h_k = x_{k+1} - x_k$.

Address: University of Ostrava, Institute for Research and Applications of Fuzzy Modelling, 30. dubna 22, 70103 Ostrava, Czech republic (email: {Irina.Perfilieva, Vilem.Novak, Viktor.Pavliska, Antonin.Dvorak, Martin.Stepnicka}@osu.cz).

B. Fuzzy Transform

Let $[a, b]$ be the fixed universe, $x_1 < \ldots < x_n$ fixed nodes and $p_1, \ldots, p_l$ fixed points from $[a, b]$, such that $n \geq 2$, $l > n$ and $x_1 = a$, $x_n = b$. Let $A_1, \ldots, A_n$ be fixed basic functions which constitute a fuzzy partition of $[a, b]$. We assume that the set $P_l = \{p_1, \ldots, p_l\}$ is *sufficiently dense with respect to the partition*, i.e.

$$(\forall k)(\exists j) \quad A_k(p_j) > 0. \tag{2}$$

Let us consider the space V_l of real valued functions defined on the set P_l, i.e.

$$V_l = \{f : P_l \longrightarrow \mathbb{R}\}.$$

If for each $f \in V_l$ we denote $f_j = f(p_j)$, $j = 1, \ldots, l$, then V_l can be identified with the set of all l-dimensional vectors with real components. The (discrete) F-transform of $f \in V_l$ is introduced as follows.

Definition 2
Let $f \in V_l$ be given and $A_1, \ldots, A_n$, $n < l$, be fixed basic functions. We say that the n-tuple of real numbers $[F_1, \ldots, F_n]$ is the F-transform of f with respect to $A_1, \ldots, A_n$ if

$$F_k = \frac{\sum_{j=1}^{l} f(p_j) A_k(p_j)}{\sum_{j=1}^{l} A_k(p_j)}. \tag{3}$$

The F-transform of f with respect to $A_1, \ldots, A_n$ will be denoted by $\mathbf{F}_n[f] = [F_1, \ldots, F_n]$. It has been proved in [7] that the components of the F-transform are the *weighted mean values* of an original function where the weights are given by the basic functions.

The original function f can be approximately reconstructed (with the help of the inversion formula) from its fuzzy transform $\mathbf{F}_n[f]$. The function represented by the inversion formula is called the *inverse F-transform*. We consider the inverse F-transform at the same points where the original function is given.

Definition 3
Let function $f \in V_l$ be given and $\mathbf{F}_n[f] = [F_1, \ldots, F_n]$ be the F-transform of f with respect to $A_1, \ldots, A_n$. Then the function

$$f_{F,n}(p_j) = \sum_{k=1}^{n} F_k A_k(p_j), j = 1, \ldots, l, \tag{4}$$

defined on the same set P_l, is the inverse discrete F-transform.

III. Main Results

A. Analysis of Time Series with the Help of the Fuzzy Transform

Assume that y_t, $t = 1, \ldots, T$, $T \geq 3$, is a time series. We can consider it as a function which is defined on the set $P_T = \{1, \ldots, T\}$ and which belongs to the space V_T. Let $A_1, \ldots, A_n$, $n < T$, be triangular basic functions which constitute a fuzzy partition of the interval $[1, T]$ such that the set P_T is sufficiently dense with respect to $A_1, \ldots, A_n$. Denote P_i, $i = 1, \ldots, n$, a subset of P_T consisting of points "covered" by A_i, i.e.$t \in P_i \quad \text{iff} \quad A_i(t) > 0$. Note that each P_i is not empty.

Denote $\mathbf{F}_n[y] = [Y_1, \ldots, Y_n]$ the F-transform of y_t with respect to $A_1, \ldots, A_n$. Then $\{y_t - Y_i \mid t \in P_i\}$ is the i-th residuum of y_t with respect to A_i, $i = 1, \ldots, n$. For $t = 1, \ldots, T$, $i = 1, \ldots, n$ we denote

$$r_{ti} = \begin{cases} y_t - Y_i, & \text{if } t \in P_i, \\ -\infty, & \text{otherwise} \end{cases}$$

so that $R_{T \times n} = (r_{ti})$ is the matrix of residua. It is easy to see that y_t can be reconstructed from its F-transform $\mathbf{F}_n[y]$ and the matrix of residua R:

$$y_t = \bigvee_{i=1}^{n} (Y_i + r_{ti}). \tag{5}$$

B. Forecasting a Time Series

To forecast a time series we will use its representation (5) and separately forecast the next component Y_{n+1} of the F-transform (of y_t) and a respective residuum. From now on we will explicitly refer to the time series which have been chosen for the competition. We assume that a fuzzy partition $A_1, \ldots, A_n$ of the time interval $\{1, \ldots, T\}$ is uniform and moreover, n is such that each A_i, $i = 2, \ldots, n-1$, covers 12 points corresponding to one year and $h = 6$. These points constitute the set P_i. Let us extend the set P_T by new points $T+1, \ldots, T+12$ so that they constitute the new set P_{n+1}. Let us also extend the fuzzy partition $A_1, \ldots, A_n$ by the new basic function A_{n+1} which covers P_{n+1}.

We will consider two methods for the forecast of the Y_{n+1}-th component: the F-transform of the second order and a logical deduction. Then we will consider the method for the forecasting a new residua with respect to the new Y_{n+1}-th component. We will create possible combinations and train them on the last (one-year long) part of y_t. Then we choose the best combination and use it for the forecast.

C. Forecast an F-Transform Component by the F-transform of the second order

In this subsection we explain the first method of the forecasting the Y_{n+1}-th component of the F-transform (of y_t).

The F-transform will be applied to the vector $\mathbf{F}_n[y] = [Y_1, \ldots, Y_n]$ which we consider as a function from the space V_n and which is defined on the set $P_n = \{1, \ldots, n\}$. We will choose another basic functions $B_1, \ldots, B_s$, $s < n$, which constitute a fuzzy partition of the interval $[1, n]$ such that the set P_n is sufficiently dense with respect to $B_1, \ldots, B_s$. The F-transform of the second order of y_t is the vector $\mathbf{F}_s^2[y] = [Y_1^2, \ldots, Y_s^2]$ of the F-transform components of $\mathbf{F}_n[y]$ with respect to $B_1, \ldots, B_s$. Analogous to the above, we define the matrix of residua so that the $\mathbf{F}_n[y]$ can be reconstructed similar to the reconstruction of y_t expressed in (5). Thus, to forecast the next component Y_{n+1} we will forecast the next component Y_{s+1}^2 of the F-transform of the second order and a respective residuum.

To forecast Y^2_{s+1} we will extrapolate the linear model using the least square method based on the preceding components $Y^2_1, \dots, Y^2_s$.

The forecast of the respective residua will be discussed in Subsection III-E.

D. Forecasting F-transform components from linguistic description

In this subsection, we will briefly describe the second method of the forecasting the Y_{n+1}-th component of the F-transform (of y_t).

Let $\mathbf{F}_n[y] = [Y_1, \dots, Y_n]$ be the F-transform of y_t as described in Subsection III-A. We can view $\mathbf{F}_n[y]$ as a new time series and forecast it using a *perception-based logical deduction*. The procedure is based on a linguistic description which characterizes its behavior.

The considered linguistic description is a set of fuzzy IF-THEN rules of the form

$$\begin{aligned}
&\text{IF } Y_1 \text{ is } \mathcal{A}_{11} \text{ AND } \cdots \text{ AND } Y_q \text{ is } \mathcal{A}_{1q} \\
&\qquad\qquad\qquad \text{THEN } Y_{q+1} \text{ is } \mathcal{B}_1 \\
&\text{IF } Y_2 \text{ is } \mathcal{A}_{21} \text{ AND } \cdots \text{ AND } Y_{q+1} \text{ is } \mathcal{A}_{2q} \\
&\qquad\qquad\qquad \text{THEN } Y_{q+2} \text{ is } \mathcal{B}_2 \\
&\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \qquad (6)\\
&\text{IF } Y_{n-q} \text{ is } \mathcal{A}_{m1} \text{ AND } \cdots \text{ AND } Y_{n-1} \text{ is } \mathcal{A}_{mq} \\
&\qquad\qquad\qquad \text{THEN } Y_n \text{ is } \mathcal{B}_m
\end{aligned}$$

where $m = n - q$ and $\mathcal{A}_{11}, \dots, \mathcal{A}_{mq}$ are specific linguistic expressions of the form

$$\langle\text{linguistic hedge}\rangle\{\textit{small, medium, big}\}.$$

The linguistic hedge is, e.g., *very, extremely, roughly, more or less*, etc. These expressions have a specific semantics that has been in detail described in [3]. Let us stress that the rules in (6) are taken as genuine conditional expressions of natural language that are interpreted accordingly — see [4]. Each rule in (6) thus linguistically expresses that the value of the component Y_{q+j}, $j = 1, \dots, n-q$ is determined by certain values of the previous q components $Y_j, \dots, Y_{j+q-1}$.

We have implemented a learning method (see [1]) that can learn of linguistic expressions $\mathcal{A}_{11}, \dots, \mathcal{A}_{mq}$ from values of the components $\mathbf{F}_n[y]$. The number q is a free parameter that has to be determined during leaning.

After the learning, the forecasting of the new value Y_{n+1} is done from the previous q components $Y_{n-q+1}, \dots, Y_n$ using the method called *perception-based logical deduction* that has been described in details in [2], [5]. Roughly speaking, this method chooses and fires the rule whose antecedent (IF part) gives the best linguistic characterization of the values $Y_{n-q+1}, \dots, Y_n$. The final result of this deduction is the best linguistic characterization of Y_{n+1} which can be represented by a fuzzy set having a specific shape. The latter is defuzzified by the method called DEE (Defuzzification of Evaluating Expressions). The result of the defuzzification is the desired value of Y_{n+1}.

E. Forecast a Residua

We will explain how the matrix of residua $R_{T\times n} = (r_{ti})$, introduced in Subsection III-A, can be extended by the $(n+1)$-th column. The latter is actually the desired forecast of residua.

By the agreement on a fuzzy partition, each set P_i, $i = 2, \dots, n-1$, contains 12 points which will be denoted by $t_{1,i}, \dots, t_{12,i}$. Then, each vector of residua $\mathbf{r}_i = (r_{1i}, \dots, r_{Ti})$, $i = 2, \dots, n-1$, contains 12 components which are different from $-\infty$ and are represented by $y_t - Y_i$, if $t \in P_i$, . Let they constitute the subvector $\hat{\mathbf{r}}_i = (\hat{r}_{1,i}, \dots, \hat{r}_{12,i})$. It is easy to see that

$$\frac{\sum_{j=1}^{12}(\hat{r}_{j,i} A_i(t_{j,i}))}{\sum_{j=1}^{12} A_i(t_{j,i})} = 0$$

so that $\hat{\mathbf{r}}_i$, $i = 2, \dots, n-1$, is the sequence of random variables with the distribution functions $\frac{A_i(t_{j,i})}{\sum_{j=1}^{12} A_i(t_{j,i})}$ and the zero expected values.

A new vector of residua $\hat{\mathbf{r}}_{n+1}$ should contain a subvector with 12 components such that they constitute a random variable with a distribution function determined by A_{n+1} and the zero expected value. Let us explain how new 12 components $(\hat{r}_{1,n+1}, \dots, \hat{r}_{12,n+1})$ of $\hat{\mathbf{r}}_{n+1}$ can be forecasted.

We assume that the sequence $\hat{\mathbf{r}}_i$, $i = 2, \dots, n+1$ is stationary so that each $k+1$-th vector is a linear combination of the k preceding ones where $k < n$ in an unknown parameter. Coefficients of the linear combination can be found by solving the respective system of linear equations. Having coefficients, the new 12 components $(\hat{r}_{1,n+1}, \dots, \hat{r}_{12,n+1})$ of $\hat{\mathbf{r}}_{n+1}$ can be easily computed. The value of k can be found by comparison a forecasted series with the original one on the last (one-year long) part of y_t. The value of k which gives the best possible absolute difference is chosen.

F. Forecasting Procedure

Let y_t, $t = 1, \dots, T$, $T \geq 3$, be a given time series. We divide it into two parts y^f_t, $t = 1, \dots, T-12$ and y^l_t, $t = T-11, \dots, T$ and use y^l_t to train parameters. Let us choose initial values of parameters s, k_1, k_2. Then we proceed as follows.

- The first part y^f_t is analyzed with the help of the F-transform (Subsection III-A) so that the components $[Y_1, \dots, Y_{n-1}]$ of y^f_t with respect to $A_1, \dots, A_{n-1}$ and their respective residua $\mathbf{r}_i = (r_{1i}, \dots, r_{Ti})$, $i = 1, \dots, n-1$, are to be computed.
- The components $[Y_1, \dots, Y_{n-1}]$ are analyzed in the same way as above with the help of the F-transform of the second order. The latter results in the vector of components $\mathbf{F}^2_s[y] = [Y^2_1, \dots, Y^2_s]$ and their residua which are obtained with respect to basic functions $B_1, \dots, B_s$.
- We forecast the next component Y^2_{s+1} either by an extrapolation of the linear model (Subsection III-C) or by a logical deduction (Subsection III-D).

- We forecast the $s+1$-th vector of residua (with respect to the basic function B_{s+1}) by a linear combination of k_1 preceding residua vectors (Subsection III-E).
- By (5), we compute the component Y_n of the F-transform y_t^f with respect to A_n.
- We compute the n-th vector of residua $\mathbf{r}_n$ (with respect to the basic function A_n) by a linear combination of k_2 preceding residua vectors.
- By (5), we compute the components of the second part y_t^l of the original time series y_t.

For each possible combination of parameters s, k_1, k_2 we apply the described above procedure and obtain (by the computation) the respective components of the second part y_t^l. Then the triple of parameters which gives the best possible absolute difference between the computed and the actual series y_t^l is chosen. The chosen parameters are then used for the forecast of the original time series y_t. The forecast is obtained using the procedure described above where y_t^f is replaced by y_t.

IV. Conclusions

The proposed methodology for forecasting time series is based on combination of two techniques: fuzzy transform and perception-based logical deduction. It consists of two phases: analysis of a time series and its forecast. In the first phase, a time series is decomposed into two components: a trend and residua. The trend is represented either by a vector of fuzzy transform components, or by the inverse fuzzy transform. The residuum is the difference between the original and the corresponding trend value of the time series.

In the second phase, both trend as well as residua are forecast and then put together. We use one of three possibilities: second order fuzzy transform, extrapolation of the inverse fuzzy transform, or perception-based logical deduction. Forecast of the residua is obtained by a linear combination of previous residua using optimization. A number of parameters are involved in this methodology. They are obtained by training. The best combination of parameters is taken for the final forecast.

Acknowledgment

The research was supported by projects MSM 6198898701 and 1M0572 of the MŠMT ČR.

References

[1] R. Bělohlávek and V. Novák, "Learning rule base of the linguistic expert systems," *Soft Computing*, vol. 7, pp. 79–88, 2002.

[2] V. Novák, "Perception-based logical deduction," in *Computational Intelligence, Theory and Applications*, B. Reusch, Ed. Berlin: Springer, 2005, pp. 237–250.

[3] ——, "A comprehensive theory of evaluating linguistic expressions," in *Proc. of 26th Linz Seminar on Fuzzy Set Theory*, S. Gottwald and E. P. Klement, Eds. Berlin: Springer, (to appear, (*Research Report* No. 71, http://irafm.osu.cz/irafm).

[4] V. Novák and S. Lehmke, "Logical structure of fuzzy IF-THEN rules," *Fuzzy Sets and Systems*, vol. 157, pp. 2003–2029, 2006.

[5] V. Novák and I. Perfilieva, "On the semantics of perception-based fuzzy logic deduction," *International Journal of Intelligent Systems*, vol. 19, pp. 1007–1031, 2004.

[6] I. Perfilieva, "Fuzzy transform: Application to reef growth problem," in *Fuzzy Logic in Geology*, R. B. Demicco and G. J. Klir, Eds. Amsterdam: Academic Press, 2003, pp. 275–300.

[7] ——, "Fuzzy transforms: Theory and applications," *Fuzzy Sets and Systems*, vol. 157, pp. 993–1023, 2006.

[8] I. Perfilieva and E. Haldeeva, "Fuzzy transformation," in *Proc. of the Joint 9th IFSA World Congress and 20th NAFIPS Int'l Conf. CD-ROM*, Vancouver, Canada, July, 25.–28. 2001, pp. 127–130.