DOCUMENTATION FOR SOFTWARE SYSTEM

TIME SERIES PREDICTION TOOL

Theoretical background

Time series analysis and prediction is an important task that can be used in many areas of practice. The task of getting the best prediction to given series may bring interesting engineering applications in wide number of areas like economics, geography or industry. Solution to the problem of obtaining best results in prediction of time series can be based on well-known and simple methods like Winters or Linear method. In this paper we present a tool based on two methods originally developed by members of Institute for Research and Applications of Fuzzy Modeling. The aim of the paper is not to present the details of the methods already published, but to present a tool implementing them. The first method is based on the notion of F-transform devised by the group of Prof. Perfilieva. The second approach use the linguistic rules utilizing fuzzy logic and deduction that is a well-known formalism with very good results in variety of practical applications like industrial ones.

F-transform

The core idea of the F-transform technique is a fuzzy partition of the universe. It can be simply presented like set of intervals fulfilling some criteria. It is described in the following definition.

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

(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) $A_k(x)$, $k = 2, \ldots, n$, monotonically increases on $[x_{k-1}, x_k]$ and $A_k(x)$, $k = 1, \ldots, n-1$, monotonically decreases on $[x_k, x_{k+1}]$;
(5) for all $x \in [a, b]$

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

The membership functions $A_1(x), \ldots, A_n(x)$ are called *basic functions*.

These partitions form a base for F-transform which lead to the tuple of numbers representing original transformated function. The n-tuple can be obtained using the following notion.

Definition 2

Let $f \in V_l$ be given and $A_1, \ldots, A_n$, $n < l$, be basic functions which constitute a fuzzy partition of $[a, b]$. 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

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

To forecast a time series we will use its F-transform representation and separately forecast the next component Y_{n+1} of the F-transform(of y_t) and a respective residuum. We will consider three methods for the forecasting a component of the F-transform: the F-transform of the second order an extrapolation of the inverse fuzzy transform and a logical deduction.

Linguistic rules

The theory of linguistic term and variables is well-known approach in the fuzzy logic community. It enables to work with rules containing terms of natural language like small or big and modifiers like very, roughly etc. The rule interpretation is then done by logical deduction based on which is based on fuzzy set theory and fuzzy logic to enable to deduce conclusions on the basis of imprecise description of the given situation using the linguistically formulated fuzzy IF-THEN rules [1].

The usage of this theory within a frame of time series prediction lies in the learning of these rules from the serie and then application to future (predicted) members of the serie. These learning algorithms are already prepared within the LFLC software [1], which is intended to perform logical deduction on linguistic rules. The core of the system serve also for the presented application.

Example

An example fuzzy IF-THEN rule is

IF *the obstacle is near* AND *the speed of the car is high*
THEN *the breaking force is very strong*.

The "obstacle", "speed" and "breaking force" are variables while "near", "high" and "very strong" are expressions characterizing vaguely the magnitude of the variable.

The part before THEN is called the *antecedent* and the part after it the *succedent*. The variables $X_1, \ldots, X_n$ are called *input*, or *independent* variables. The variable Y is called *output*, or *dependent* variable.

The fuzzy IF-THEN rules are usually put together to form the *linguistic description*

$$\mathcal{R}_1 := \text{IF } X_1 \text{ is } \mathcal{A}_{11} \text{ AND } \cdots \text{ AND } X_n \text{ is } \mathcal{A}_{1n} \text{ THEN } Y \text{ is } \mathcal{B}_1$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$
$$\mathcal{R}_m := \text{IF } X_1 \text{ is } \mathcal{A}_{m1} \text{ AND } \cdots \text{ AND } X_n \text{ is } \mathcal{A}_{mn} \text{ THEN } Y \text{ is } \mathcal{B}_m$$

Time series tool – the software application

The software for prediction of time series based on the previously presented formalisms is currently in the development, but first alpha version is now complete. It was implemented on MS-Windows platform in C++ using free GUI library WxWidgets. The predecessor of the tool has been an console application without GUI, but shares the same core like the tool.

The main tasks of the tool are the following:

- Loading and presentation of a prepared file with a time serie in a graph.
- Setting up the methods and parameters for prediction.
- Computation of prediction according to the methods.
- Presentation and selection of predictions in a graph with reports generated during prediction (difference to the original serie etc.)
- Export of selected results.

The application is an SDI (single document interface) application divided into two main windows – Plot and Methods (Fig. 1). Plot window allows opening of user chosen serie through standard file open dialog (File menu). Then it performs desired predictions on loaded serie and presents it in a graph. The graph window contains two basic panels – graphical information panel and text information panel. Graphical information panel enables to present standard graph of the serie and predictions, it can be zoomed (important method is to fit the curve into graph which is connected to ‘s’ key). Some basic graph operations can be obtained by Plot menu. Last menu item - Help - shows the application info and basics of application control (Fig.2).
Text information panel describes performed predictions projected in graphical information panel. It presents information concerning particular prediction in one line (color refers to the same color in graph). The information given could be presented on the following simple example:

No.4 Predictor, name: AvgTrend[k=5] {IntRule[v=9]}, error:0.0105458

It divides into predictor number, prediction description (methods used) – Trend (k = seasons number included for prediction), {method [specification]}, error of prediction. For our example it means:
4-th best prediction, trend is computed from average,[number of seasons included = 5]{logical deduction was used with [variable number = 9]}.

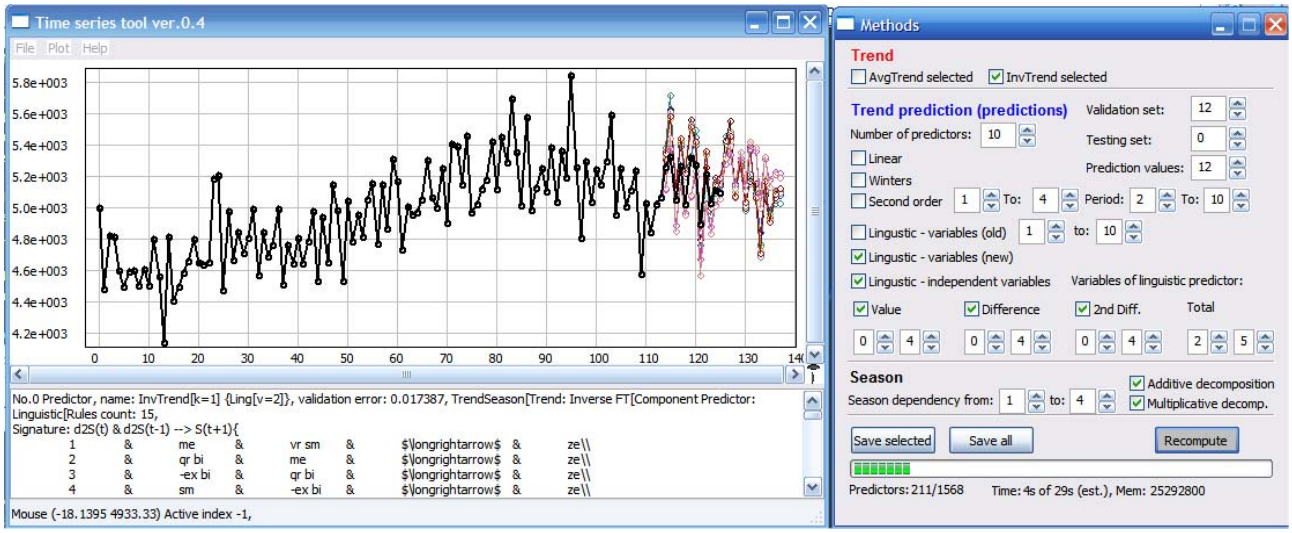

Fig. 1:Time series tool layout

Methods window (Fig. 3) enables to set up the details for prediction process. It consists of four basic parts:

1. Trend computation selection.
2. Method computation selection.
3. Season part computation selection.
4. Operations with application – computation of prediction, export of the selected curve, export of all curves into file representation.

Trend could be selected either to be computed via standard average method or via inverse F-transform. There are basically four possibilities how to predict serie – standard Linear or standard Winters method or new method of logical deduction or second order F-transform. Season computation can be additionally set up for season's dependency.

The input file with a serie should conform to a simple format, which could be observed from the following example:

```
NN3_101     4998  4480  4824  4814  4602  4499  4594  4600  4507  4606  4503
        4801  4564  4142  4818  4408  4496  4587  4656  4799  4652  4638  4650
```

First item should be name of the serie followed by unlimited number of real numbers delimited by a blank space (TAB, space, etc.).

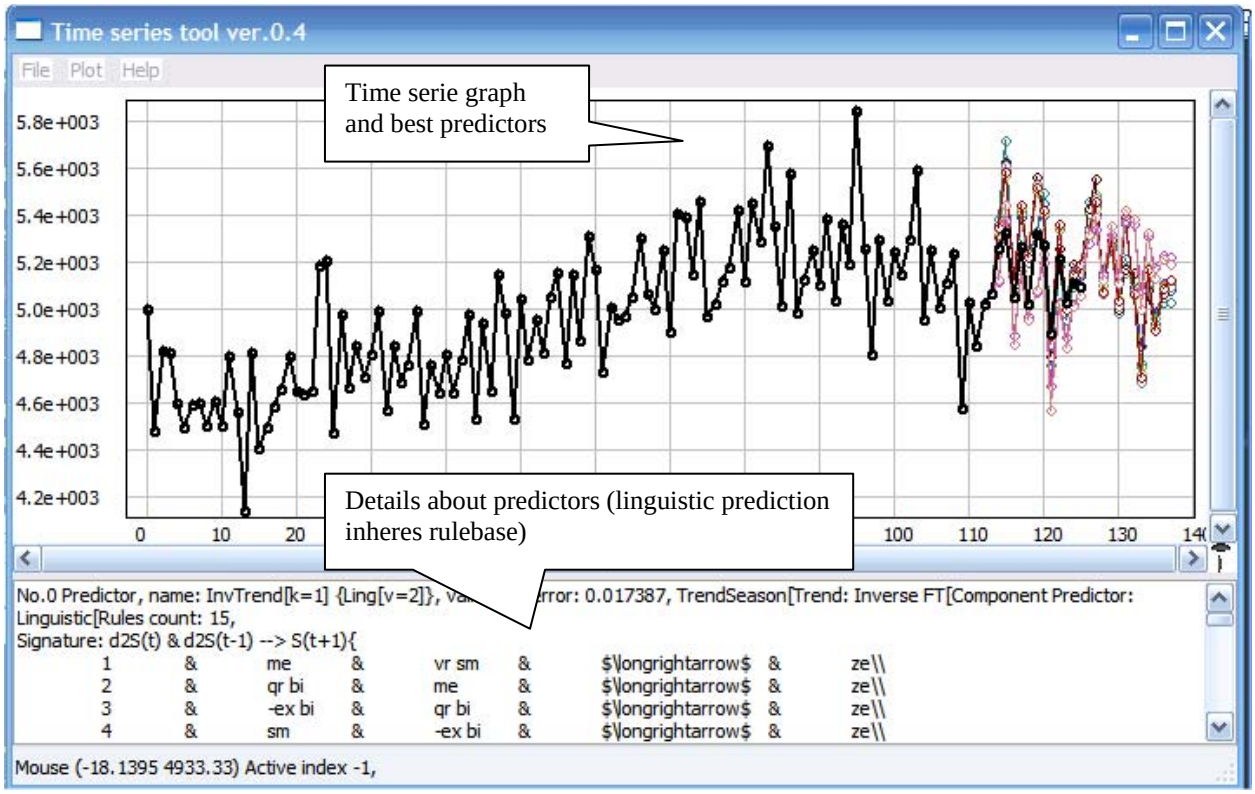

Fig. 2: Plot window

Plot window enables functions as follows:

- Move origin: Scrollbars, left-mouse drag on axis labels, arrow keys, shift+ctrl+alt and left mouse drag.
- Zoom plot: Select area with mouse or [qwe|asd|zxc] keys, where 's' fits plot.
- Select curve: shift key cycles through them.
- Delete curve: delete key.
- Select points: shift + left mouse drag to select area.
- Unselect points: ctrl + left mouse drag to select area.
- Move cursor: Select point w/ left mouse button, move forward and, backwards using ',<' or '.>' keys.

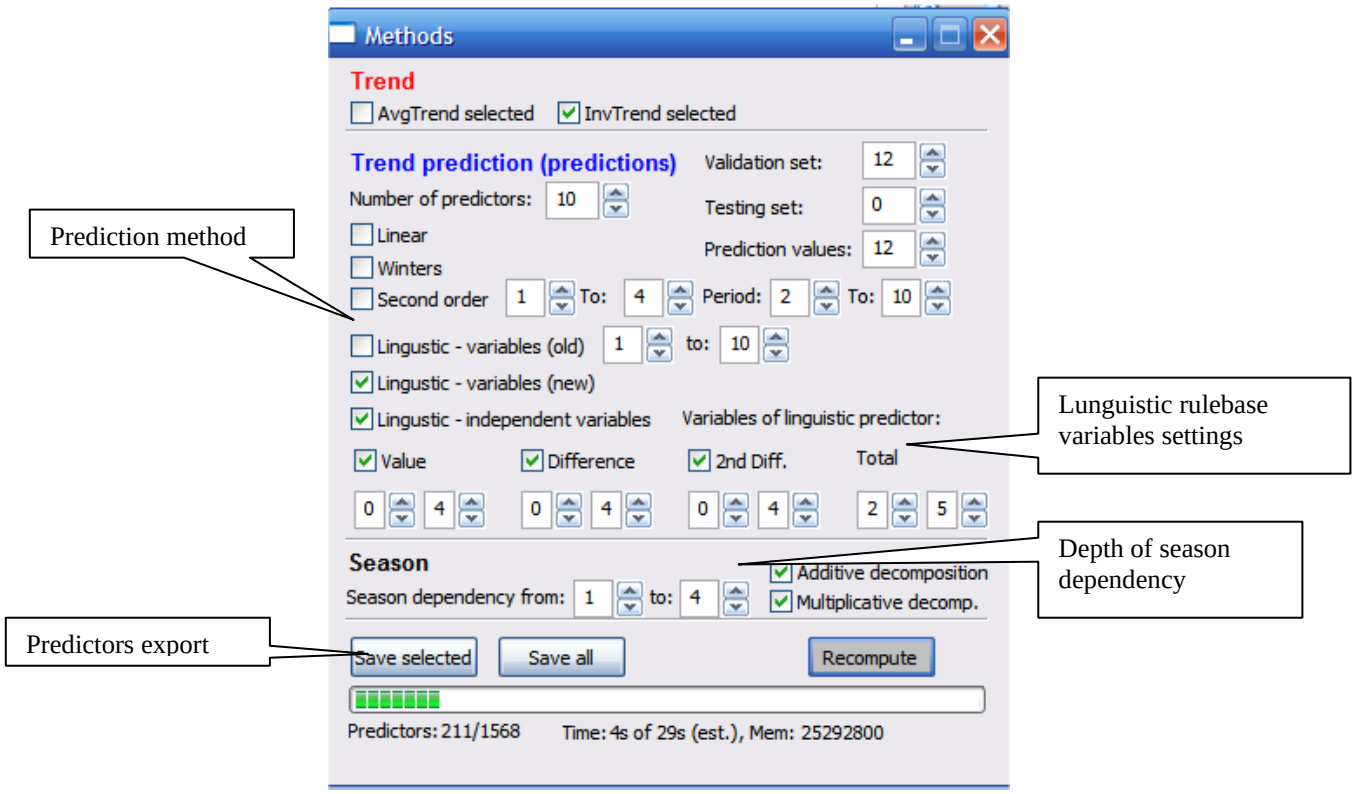

Fig. 3: Methods window