Available online at www.sciencedirect.com
ScienceDirect
International Journal of Approximate Reasoning
48 (2008) 36–46
INTERNATIONAL JOURNAL OF APPROXIMATE REASONING
www.elsevier.com/locate/ijar

Fuzzy transform in the analysis of data ☆

Irina Perfilieva *, Vilém Novák, Antonín Dvořák

University of Ostrava, Institute for Research and Applications of Fuzzy Modeling, 30. dubna 22, 701 03 Ostrava 1, Czech Republic

Accepted 14 June 2007
Available online 1 July 2007

Abstract

Fuzzy transform is a novel, mathematically well founded soft computing method with many applications. In this paper, we present this technique with applications to data analysis. First, we show how it can be used for detection and characterization of dependencies among attributes. Second, we apply it to mining associations that have a *functional* character. Moreover, the mined associations are characterized linguistically which means that their antecedent consists of fuzzy numbers and the consequent is characterized using pure evaluative linguistic expressions (i.e. expressions such as *small, very big, more or less medium*, etc).

Keywords: Fuzzy transform; Evaluative linguistic expressions; Linguistic associations; Data-mining; Association rules

1. Introduction

The method of *fuzzy transform* (F-transform, for short) has been developed by Perfilieva in [10–12] and applied to many practical problems such as the construction of approximate models, filtering, solution of differential equations, and data compression. The principal constituent of F-transform, reflected in its name, is a correspondence (transformation) between two universes: one is composed of fuzzy objects (fuzzy sets) and the other one consists of real numbers. The form of fuzzy objects in the first universe enables us to consider them as an interpretation of certain linguistic expressions where the universe represents a specific context of their use. The fuzzy transform puts the fuzzy objects into a correspondence with real numbers characterizing typical values of the related monitoring attribute. The way of assigning the corresponding value is predetermined by the method and has a number of great properties that have been investigated in the related literature [10–12,15]. Among them, we highlight the optimality of the corresponding value with respect to the weighted square root deviation.

The goal of this paper is to demonstrate that applications of F-transform are wide and can also be used in data analysis. *First*, we deal with the detection and characterization of dependencies among given attributes. In this case, one attribute is fixed (we call it monitoring) and is considered to be dependent (function) on other attributes regarded as independent. The number and choice of independent attributes is a priori not known.

☆ This paper has been supported by the project 1M0572 and partially also MSM 6198898701 of the MŠMT ČR.
* Corresponding author. Tel.: +420 596 120 478; fax: +420 738 511 408.
E-mail addresses: Irina.Perfilieva@osu.cz (I. Perfilieva), Vilem.Novak@osu.cz (V. Novák), Antonin.Dvorak@osu.cz (A. Dvořák).

0888-613X/$ - see front matter
doi:10.1016/j.ijar.2007.06.003

By constructing inverse F-transforms of all possible hypothetic functions, we verify existence of a functional dependence and propose its formal expression.

Second, we demonstrate that F-transform can also be used as a specific data-mining method, namely as a method for mining associations that exist in numerical data. There are many publications (see [1,3,2,5,4,13,14] and elsewhere) on data-mining theory. In most cases, the methods presented there work with boolean-valued data. The data-mining method presented in this paper mines associations directly from the real-valued data, even in linguistic form. Our main claim is that associations should have a *functional* character. Moreover, if we can establish an association between two events, we often can establish a graded association between them. For example, the association *age* $\sim$ *health* assumes that "the younger a person is, the less problems he/she has with his/her health" which leads to an isotonic function between grades of being "young" and grades of being "healthy". Similarly, *skillfulness* $\sim$ *salary* assumes that "a skillful person has a higher salary" and leads to an isotonic function between grades of "skillfulness" and grades of having a "high salary", etc. Based on this claim, we conclude that to mine associations from (especially numerical) data means to extract a functional dependence among related attributes. Of course, as in all data-mining methods, our method provides a hypothesis about functional dependence that may exist in the given data and supports it by parameters of degrees of support and confidence.

The structure of this contribution is the following: Section 2 contains a brief introduction to the theory of fuzzy transforms. The main part is Section 3, where we demonstrate the manner in which F-transform can be applied to the analysis of functional dependencies and how it can be used for mining associations that linguistically characterize such dependencies.

2. The method of fuzzy transform

The fuzzy transform is a method that can be applied to a continuous function on a bounded domain or to a function known on a finite domain (the latter is called a *discrete* function). The domain is supposed to be partitioned by fuzzy sets. The F-transform takes a function and produces a set-to-point correspondence between fuzzy sets from the partition and certain average values of that function.

In this section, we will consider functions with one variable because of their simplicity. However, the F-transform of a function of two or more variables can easily be obtained as a straightforward extension of the definitions provided below (see [12,15] for more details).

Definition 1. ([12]) Let $[a,b]$ be an interval of real numbers and $x_1 < \cdots < x_n$ be fixed nodes within $[a,b]$ such that $x_1 = a$, $x_n = b$ and $n \geqslant 2$. We say that fuzzy sets $A_1, \ldots, A_n$ identified with their membership functions $A_1(x), \ldots, A_n(x)$ and defined on $[a,b]$[1] form a fuzzy partition of $[a,b]$ if they fulfil the following conditions for $k = 1, \ldots, n$:

(1) $A_k : [a,b] \to [0,1], A_k(x_k) = 1$;
(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 $x \in [a,b]$

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

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

We say that a fuzzy partition $A_1(x), \ldots, A_n(x), n > 2$, is *uniform* if the nodes $x_1, \ldots, x_n$ are equidistant, i.e. $x_k = a + h(k-1), k = 1, \ldots, n$, where $h = (b-a)/(n-1)$, and two more properties are fulfilled for $k = 2, \ldots, n-1$:

[1] The membership function of a fuzzy set on $[a,b]$ is a mapping from $[a,b]$ to $[0,1]$.

(6) $A_k(x_k - x) = A_k(x_k + x)$, for all $x \in [0, h]$,
(7) $A_k(x) = A_{k-1}(x - h)$, for all $x \in [x_k, x_{k+1}]$ and $A_{k+1}(x) = A_k(x - h)$, for all $x \in [x_k, x_{k+1}]$.

Let $C[a,b]$ be the set of continuous functions on interval $[a,b]$. The following definition (see also [10,12]) introduces the fuzzy transform of a function $f \in C[a,b]$.

Definition 2. Let $A_1, \ldots, A_n$ be basic functions which form a fuzzy partition of $[a,b]$ and f be any function from $C[a,b]$. We say that the n-tuple of real numbers $[F_1, \ldots, F_n]$ given by

$$F_k = \frac{\int_a^b f(x)A_k(x)\mathrm{d}x}{\int_a^b A_k(x)\mathrm{d}x}, \quad k = 1, \ldots, n, \tag{2}$$

is the (integral) F-transform of f with respect to $A_1, \ldots, A_n$.

The F-transform of a function $f \in C[a,b]$ with respect to $A_1, \ldots, A_n$ is denoted by $\mathbf{F}_n[f]$. Then, according to Definition 2, we can write $\mathbf{F}_n[f] = [F_1, \ldots, F_n]$. The elements $F_1, \ldots, F_n$ are called components of the F-transform (see Fig. 1 for illustration).

The F-transform with respect to $A_1, \ldots, A_n$ establishes a linear mapping from $C[a,b]$ to $\mathbb{R}^n$ so that

$$\mathbf{F}_n[\alpha f + \beta g] = \alpha \mathbf{F}_n[f] + \beta \mathbf{F}_n[g]$$

holds for $\alpha, \beta \in \mathbb{R}$ and arbitrary functions $f, g \in C[a,b]$. This linear mapping is denoted by $\mathbf{F}_n$, where n is a dimension of the image space.

At this point, we will refer to [12] for useful properties of the components of F-transform. The most important property concerns the following problem: how accurately is the original function f represented by its F-transform? We can show that under certain assumptions of the original function, the components of its F-transform are *weighted mean values* of the given function where the weights are given by the basic functions.

Theorem 1. ([12]) *Let f be a continuous function on $[a,b]$ and $A_1, \ldots, A_n$ be basic functions which form a fuzzy partition of $[a,b]$. Then the kth component of the integral* F*-transform minimizes the function*

$$\Phi(y) = \int_a^b (f(x) - y)^2 A_k(x)\mathrm{d}x \tag{3}$$

defined on $[f(a), f(b)]$.

Let us now consider a discrete case where the original function f is known (can be computed) only at some nodes $p_1, \ldots, p_s \in [a,b]$. We assume that the set $P = \{p_1, \ldots, p_s\}$ is *sufficiently dense with respect to the fixed partition*, i.e.

$$(\forall k)(\exists j) A_k(p_j) > 0. \tag{4}$$

Then the (discrete) F-transform of f is introduced as follows.

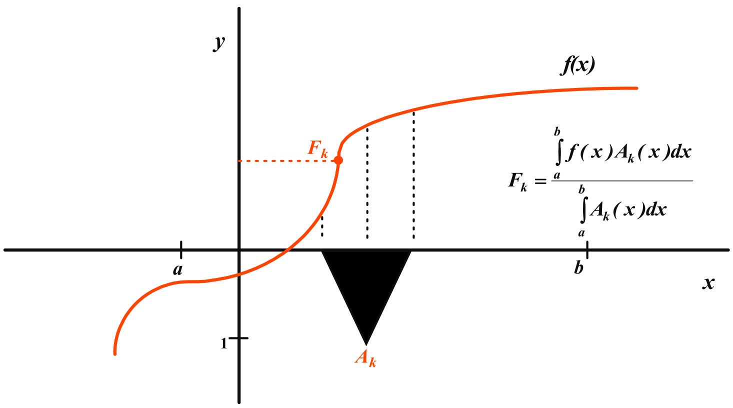

Fig. 1. F_kth component of the F-transform of a function f computed over basic function A_k of a triangular shape.

Definition 3. ([12]) Let a function f be given at points $p_1, \ldots, p_s \in [a,b]$ and $A_1, \ldots, A_n, n < s$, be basic functions which form a fuzzy partition of $[a,b]$. We say that the n-tuple of real numbers $[F_1, \ldots, F_n]$ is the discrete F-transform of f with respect to $A_1, \ldots, A_n$ if

$$F_k = \frac{\sum_{j=1}^{s} f(p_j)A_k(p_j)}{\sum_{j=1}^{s} A_k(p_j)}. \tag{5}$$

Similar to the integral F-transform, we may show that components of the discrete F-transform are the *weighted mean values* of the given function where the weights are given by the basic functions.

Theorem 2. ([12]) *Let function f be given at points $p_1, \ldots, p_s \in [a,b]$ and $A_1, \ldots, A_n$ be basic functions which form a fuzzy partition of $[a,b]$. Then the kth component of the discrete* F-*transform minimizes the function*

$$\Phi(y) = \sum_{j=1}^{s} (f(p_j) - y)^2 A_k(p_j) \tag{6}$$

defined on $[f(a), f(b)]$.

Recall that the F-transform can also be extended to k-dimensional function f for $k > 1$; see [12,15] for the details.

2.1. Inverse F-transform

The inverse F-transform (with respect to $A_1, \ldots, A_n$) takes an n-dimensional vector of the F-transform components $\mathbf{F}_n[f] = [F_1, \ldots, F_n]$ of a function f and produces a linear combination of $A_1, \ldots, A_n$ with coefficients given by those components. Thus, the obtained formula is called an *inversion formula*.

Definition 4. ([12]) Let $A_1, \ldots, A_n$ be basic functions which form a fuzzy partition of $[a,b]$ and f be a function from $C[a,b]$. Let $\mathbf{F}_n[f] = [F_1, \ldots, F_n]$ be the integral F-transform of f with respect to $A_1, \ldots, A_n$. Then the function

$$f_{F,n}(x) = \sum_{k=1}^{n} F_k A_k(x) \tag{7}$$

is called *the inverse* F-*transform* (of f with respect to the fuzzy partition $A_1, \ldots, A_n$).

The theorem below shows that the inverse F-transform $f_{F,n}$ of f can approximate that function f with an arbitrary precision.

Theorem 3. ([12]) *Let f be a continuous function on $[a,b]$. Then for any $\varepsilon > 0$ there exists n_ε and a fuzzy partition $A_1, \ldots, A_{n_\varepsilon}$ of $[a,b]$ such that for all $x \in [a,b]$*

$$|f(x) - f_{F,n_\varepsilon}(x)| \leqslant \varepsilon, \tag{8}$$

where f_{F,n_ε} is the inverse F-*transform of f with respect to the fuzzy partition $A_1, \ldots, A_{n_\varepsilon}$.*

In the discrete case, we define the inverse F-transform only at points where the original function is given

$$f_{F,n}(p_j) = \sum_{k=1}^{n} F_k A_k(p_j). \tag{9}$$

Analogous to Theorem 3, we may show that the inverse discrete F-transform $f_{F,n}$ can approximate the original function f at common points with arbitrary precision (see [12]).

3. Applications of F-transform to data analysis

In this section, we present two applications of F-transform for data analysis. The first application discovers dependencies among attributes and the second one mines associations. Both methods are based on the data of the following general form:

$$\begin{array}{c|ccccc|} & X_1 & \cdots & X_i & \cdots & X_n \\ \hline o_1 & f_{11} & \cdots & f_{1i} & \cdots & f_{1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ o_j & f_{j1} & \cdots & f_{ji} & \cdots & f_{jn} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ o_m & f_{m1} & \cdots & f_{mi} & \cdots & f_{mn} \end{array}, \tag{10}$$

where $o_1, \ldots, o_m$ are some objects (processes, transactions, etc.), $X_1, \ldots, X_n$ are their attributes. The value $f_{ji} \in \mathbb{R}, j = 1, \ldots, m, i = 1, \ldots, n$ is a value of ith attribute measured on jth object. For each attribute X_i, we specify its context $w_i = [a_i, b_i] \subset \mathbb{R}$ (a universe of discourse). This is usually done by setting a_i to be the smallest and b_i the greatest value of $f_{ji}, j = 1, \ldots, m$, in the data. However, quite often the context is clear from the meaning of the given attribute, e.g., if X_i is *age* then $a_i = 0$ and $b_i = 100$, etc.

We are looking for dependencies among some attributes which have a general (simplified) form

$$X_z = H(X_1, \ldots, X_k), \tag{11}$$

where $X_z, z \in \{1, \ldots, n\}$, is chosen as a dependent attribute, $X_1, \ldots, X_k, k \leqslant n$ and $1, \ldots, k \neq z$, are chosen as independent attributes[2] and $H : w_1 \times \cdots \times w_k \to w_z$ is a function whose existence is expected and should be discovered by the F-transform.

In particular, we are looking for dependencies, which characterize a chosen attribute X_z, and therefore, serve as models of behavior of X_z. It is worth mentioning that a priori, there is no evidence that such models exist. We propose the following procedure which verifies the existence of a model which is formally expressed by (11):

- Assume that H is a continuous function of arguments $X_1, \ldots, X_k$.
- Choose certain fuzzy partitions of universes $w_1, \ldots, w_k$ and construct the discrete F-transform of H (as a function of arguments $X_1, \ldots, X_k$) using the values of the respective attributes $X_1, \ldots, X_k$ given in the data.
- Construct the inverse F-transform H_F of H (according to (7)) and estimate the difference between it and H (or X_z) at common points (below we use statistical index of determinacy[3] for this purpose).
- If the difference is appropriate then take the inverse F-transform H_F of H as a model of X_z; otherwise we can be sure that a continuous model of X_z as a function of $X_1, \ldots, X_k$ does not exist (Theorem 3).

If the proposed procedure ends up with a model of X_z, then we may wish to find the best model of X_z with respect to the chosen accuracy (difference between the model and X_z at common points). In this case, we shall apply our procedure to other combinations of independent attributes and choose the combination that leads to a model with the least value of accuracy. If the proposed procedure does not end up with a model of X_z, then we may repeat it with other combinations of independent attributes until all have been exhausted. This approach is demonstrated in the next subsection.

[2] The simplification consists of avoiding double subscripts. Clearly, $X_1, \ldots, X_k$ some attributes are chosen from the set of all attributes $\{X_1, \ldots, X_n\}$.

[3] The index of determinacy is $r_c^2 = \frac{s_{\hat{y}}^2}{s_y^2}$, where s_y and $s_{\hat{y}}$ are the standard deviations of the data points and the estimates given by the regression curve.

3.1. Analysis of dependencies

The following application demonstrates that F-transform can be applied to the detection and characterization of dependencies among attributes in the form (11). The context (universe of discourse) w_i of each attribute $X_i, i = 1, \ldots, k$, z is partitioned into n_i basic functions. The expected dependence is then obtained by the respective inverse F-transform of the discrete function (11) from the given data. Thus, F-transform, in a certain sense, can replace regression analysis with the verification of existence of a dependence.

We will demonstrate our method on a concrete application.[4] The test data contain values of gross domestic product (GDP) and six other economical factors of the Czech Republic measured in quarters starting from 1997:

Year/ quarter	Final consumption	Gross capital	Gross product	Unemployment	Import rate	Inflation	GDP
1997/1	298778	146496	402031	221500	224315	81.93	414740
1997/2	339007	138911	419728	231300	257425	83.00	450409
1997/3	334682	142208	418809	259800	263575	87.20	458630
1997/4	362366	117778	411755	280700	291061	88.47	461352
1998/1	323331	138134	432098	306700	273625	92.80	455489
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
2004/1	463529	179746	636171	443800	425937	109.10	646761
2004/2	500132	211808	689695	419100	526972	109.60	709589
2004/3	505042	207514	686198	420400	496122	110.00	707744
2004/4	542491	170541	670315	420200	536025	110.00	703623

The purpose of this application was to create an optimal mathematical model of the GDP, i.e. to find a minimal set of attributes (independent variables) that determine its dynamics. We considered all possible combinations of attributes placed in the first six columns of the given database. To realize our goal, we verified whether the GDP can be modeled as a function of these chosen attributes.

We started with one attribute and increased their number. For each combination of attributes, we computed the index of determinacy and considered those combinations with the highest values within that index. For each chosen combination of attributes, we constructed the respective inverse (discrete) F-transform (of the GDP as a function of those attributes) with nine basic functions of cosine type and considered it a possible (hypothetic) model of the GDP. Finally, the model which gives the least value of the square root error has been taken as the resulting optimal model. The sequence of Figs. 2–5 shows two best and two worst models in each chosen combination with one and two attributes. The resulting optimal model presented in Fig. 6 contains three attributes.

3.2. Mining associations

The method presented in the previous subsection demonstrates the power of F-transform for detection and verification of a dependence. In this subsection, we will describe the use of F-transform for mining associations in the data of the form (10), i.e. finding hypotheses of possible dependencies among attributes.

As stated in the introduction, the mined associations have a functional character (11). Each of them characterizes the dependence locally, and the whole dependence is characterized by all of the observed associations together. Moreover, unlike the previous subsection, associations discovered here are not cases of ordinary function, but cases of a fuzzy function which establishes a correspondence between universes of fuzzy sets. We suppose that the universes $w_1, \ldots, w_k$ are partitioned by fuzzy sets (according to Definition 1) and the associations functionally join some fuzzy sets from partitions of $w_1, \ldots, w_k$ with fuzzy sets over respective F-trans-

[4] This application has been prepared together with Lorenziniová as a part of her diploma work.

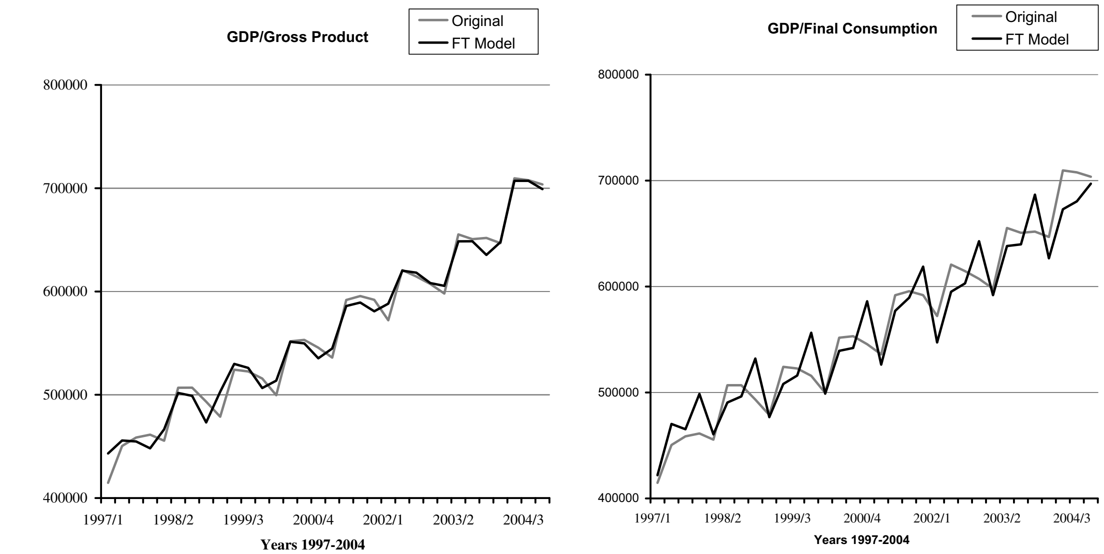

Fig. 2. Two best models of GDP as a function of one variable: GDP as a function of gross product (left) and GDP as a function of final consumption (right).

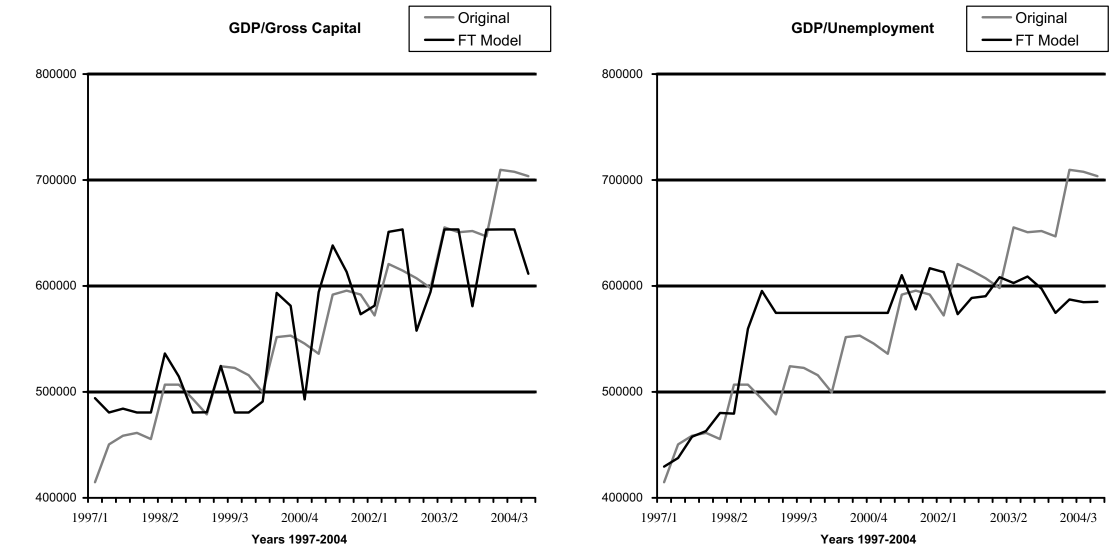

Fig. 3. Two worst models of GDP as a function of one variable: GDP as a function of gross capital (left) and GDP as a function of unemployment (right).

form components. Each association is supported by two parameters, namely the degrees of support r and confidence γ defined below.

Our data-mining method discovers associations using F-transform and expresses them linguistically using the theory developed in [6]. The mined associations using F-transform have the following form:

$$(X_1 \text{ is } Fn[l_1]) \ \mathsf{AND} \ \cdots \ \mathsf{AND} \ (X_k \text{ is } Fn[l_k]) \overset{F}{\sim}_{r,\gamma} (\mathtt{mean}\ X_z \text{ is } \mathscr{C}). \tag{12}$$

The value of the antecedent attribute X_i is a fuzzy set $Fn[l_i]$, which is an element of the fuzzy partition of w_i over the node l_i and which models (differently in different contexts) the meaning of the linguistic expression

$$\mathtt{approximately}\ l_i, \quad i = 1, \ldots, k.$$

Hence, the expression of the form 'X_i is $Fn[l_i]$', occurring in (12), should be read as "X_i is approximately l_i".

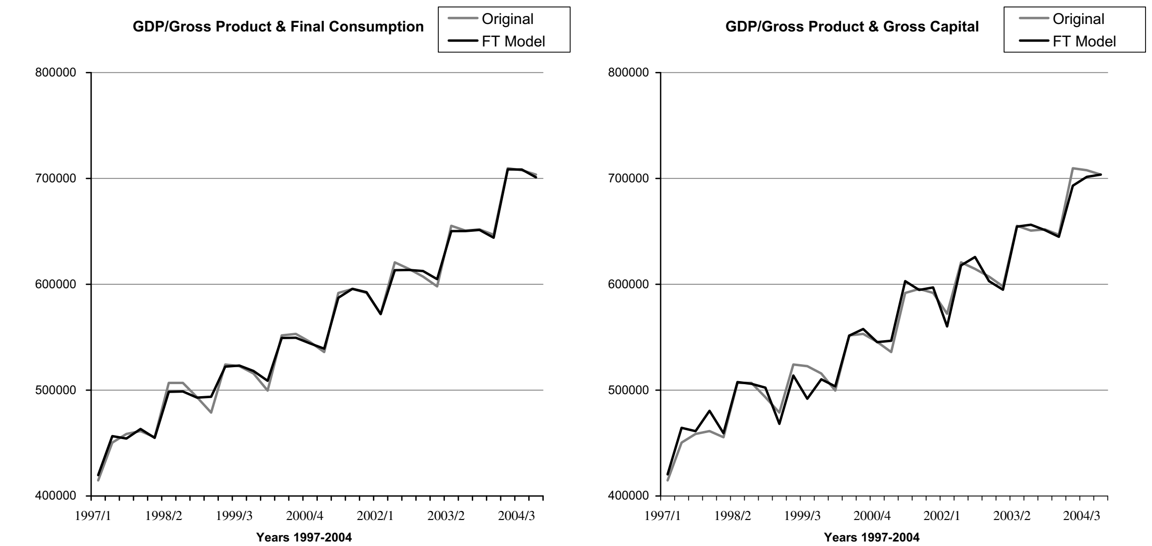

Fig. 4. Two best models of GDP as a function of two variables: GDP as a function of gross product and final consumption (left) and GDP as a function of gross product and gross capital (right).

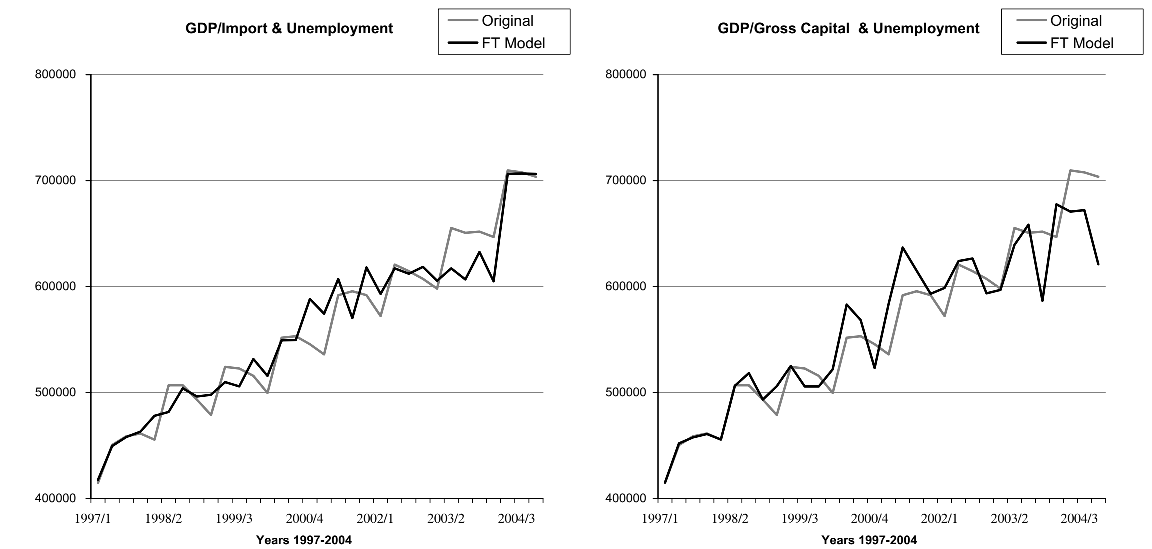

Fig. 5. Two worst models of GDP as a function of two variables: GDP as a function of import and unemployment (left) and GDP as a function of gross capital and unemployment (right).

The $\mathscr{C}$ is a *pure evaluative linguistic expression* (see [6,8]) which linguistically characterizes a component $F_{l_1 \ldots l_k}$ of F-transform corresponding to the k-tuple of fuzzy sets $\langle Fn[l_1], \ldots, Fn[l_k] \rangle$ (analogously to (5)). $\mathscr{C}$ is one of the expressions Sm (small), Me (medium), Bi (big), that has been possibly combined with the following linguistic hedges: Ex (extremely), Si (significantly), Ve (very), *empty hedge*, ML (more or less), Ro (roughly), QR (quite roughly), VR (very roughly). The reason why we put the adjective "mean" in (12) is a result of the fact that $\mathscr{C}$ evaluates the component $F_{l_1 \ldots l_k}$ which is a certain weighted mean of X_z over $Fn[l_1] \times \cdots \times Fn[l_k]$.

The symbol $\overset{F}{\sim}_{r,\gamma}$ in (12) expresses an association between the independent and dependent attributes found using F-transform. The parameters r, γ express degrees of *support* and *confidence* defined as follows. First, for

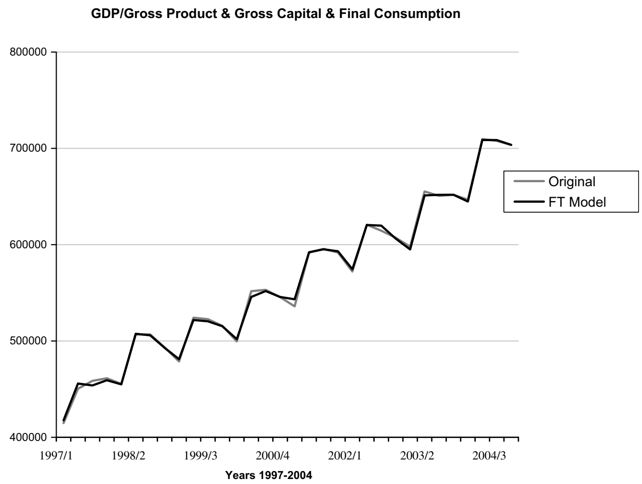

Fig. 6. The best model of GDP as a function of three variables: gross product and gross capital and final consumption.

a vector of nodes $\langle l_1, \ldots, l_k \rangle$, we define a membership function of an induced fuzzy set on the set of objects $\{o_1, \ldots, o_m\}$ as follows:

$$Fn[l_1 \ldots l_k](o_j) = Fn[l_1](f_{j1}) \ldots Fn[l_k](f_{jk}), \tag{13}$$

where $Fn[l_i](f_{ji})$ is a membership degree of f_{ji} (the value of the attribute X_i measured on the object o_j) in the fuzzy set $Fn[l_i]$.

The degree of support is defined by

$$r = \frac{|\{o_j | Fn[l_1 \ldots l_k](o_j) > 0\}|}{m}, \tag{14}$$

where $|\cdot|$ denotes a number of elements of the given set. It can be seen that r characterizes a relative number of objects that belong with non-zero membership degree to the induced fuzzy set $Fn[l_1 \ldots l_k]$.

The degree of confidence γ of each association (12) is computed as follows. First, we create an auxiliary function on the set of objects $\{o_1, \ldots, o_m\}$ induced by the inverse F-transform with components $F_{l_1 \ldots l_k}$:

$$f_F(o_j) = \sum_{l_1 \ldots l_k} F_{l_1 \ldots l_k} \cdot Fn[l_1 \ldots l_k](o_j), \tag{15}$$

where the sum is taken over all vectors $\langle l_1, \ldots, l_k \rangle$ of nodes used for the partitions of respective universes. Then we define

$$\gamma = \sqrt{\frac{\sum_{j=1}^{m} (f_F(o_j) - F_{l_1 \ldots l_k})^2 \cdot Fn[l_1 \ldots l_k](o_j)}{\sum_{j=1}^{m} (f_{jz} - F_{l_1 \ldots l_k})^2 \cdot Fn[l_1 \ldots l_k](o_j)}}. \tag{16}$$

We developed a special experimental program based on the software system LFLC 2000 (developed at the University of Ostrava), which is configured to work with the above linguistic expressions and applied it to several sets of data.

Below, we present some results of testing using data taken from a public WEB page.[5] We chose a subsample of 500 observations from a data set from a study where air pollution on a specific road is related to traffic

[5] http://lib.stat.cmu.edu/.

volume and meteorological variables, collected by the Norwegian Public Roads Administration. Few of the measured attributes are:

X_1 – logarithm of the number of cars per hour,
X_2 – temperature 2 m above ground (°C),
X_3 – wind speed (m/s),
X_z – hourly values of the logarithm of concentration of NO_2 (particles), measured at Alnabru in Oslo, Norway, between October 2001 and August 2003.

The corresponding contexts are $w_1 = [4.127, 8.349], w_2 = [-18.6, 21.1], w_3 = [0.3, 9.9], w_z = [1.2, 6.0]$. We obtained them from maximal and minimal values of the corresponding attribute in the data using histograms. It turned out that the data are not too rich and, as a result, we had to set the parameters in (12) to $r = 0.1$ and $\gamma = 0.15$.

Below are some examples of observed associations using F-transform. Choosing X_1 as the independent attribute and $n_1 = 5$ nodes, we obtained five associations, for example

A1 $(X_1 \text{ is } Fn[6]) \overset{F}{\sim}_{0.37,0.19} (\texttt{mean}\ X_z \text{ is ML Me})$.
A2 $(X_1 \text{ is } Fn[8]) \overset{F}{\sim}_{0.53,0.17} (\texttt{mean}\ X_z \text{ is QR Bi})$.

The linguistic form of these associations is clear. A1 should be read as

> If logarithm of the number of cars per hour is approximately six then mean hourly values of the logarithm of concentration of NO_2 are more or less medium with the degree of support $r = 0.37$ and confidence $\gamma = 0.19$.

When comparing both associations, we may conclude that they suggest a hypothesis with the degree of support at least 0.37 and confidence at least 0.17 about dependence (11) of the form $X_z = H(X_1)$ stating that the "greater number of cars results in a higher concentration of NO_2".

Let us stress that the fact that'the 'if-then" linguistic form presented above has been chosen for better readability. Recall, however, that this does not mean that the above association indeed characterizes a *real dependence* but only suggests a *hypothesis* about its existence, which is supported by the above parameters. To verify, other methods (and other sets of data) must be used, e.g. various statistical methods and possibly the method described in the previous subsection.

An example of more complex associations is depicted in the following. The number of nodes was $n_1 = 5, n_2 = 8, n_3 = 6$. Altogether 18 associations have been found, for example

B1 $(X_1 \text{ is } Fn[7])\ \mathsf{AND}\ (X_2 \text{ is } Fn[-1.6])\ \mathsf{AND}\ (X_3 \text{ is } Fn[2.2]) \overset{F}{\sim}_{0.11,0.3} (\texttt{mean}\ X_z \text{ is ML Me})$
B2 $(X_1 \text{ is } Fn[7])\ \mathsf{AND}\ (X_2 \text{ is } Fn[-1.6])\ \mathsf{AND}\ (X_3 \text{ is } Fn[0.3]) \overset{F}{\sim}_{0.18,0.19} (\texttt{mean}\ X_z \text{ is ML Bi})$
B3 $(X_1 \text{ is } Fn[7])\ \mathsf{AND}\ (X_2 \text{ is } Fn[-1.6])\ \mathsf{AND}\ (X_3 \text{ is } Fn[2.2]) \overset{F}{\sim}_{0.34,0.32} (\texttt{mean}\ X_z \text{ is Ro Bi})$
B4 $(X_1 \text{ is } Fn[7])\ \mathsf{AND}\ (X_2 \text{ is } Fn[-1.6])\ \mathsf{AND}\ (X_3 \text{ is } Fn[4.1]) \overset{F}{\sim}_{0.23,0.39} (\texttt{mean}\ X_z \text{ is QR Bi})$

These associations suggest a hypothesis with a degree of support of at least 0.11 and confidence of at least 0.19 about dependence (11) of the form $X_z = H(X_1, X_2, X_3)$ stating that "if logarithm of number of cars is approximately 7 and temperature 2 m above ground is approximately -1.6 then the stronger wind slightly decreases the concentration of NO_2".

The associations observed above are more or less obvious and could be stated on the basis of our common sense knowledge without analyzing the data. But, our method was still able to identify them. Moreover, these associations are also in accordance with experiments conducted with similar data but different method (presented in [7,9]). This leads us to believe that our method is indeed capable of finding associations, which correspond to expert knowledge. We argue that the linguistic form of the observed associations makes them easily understood to the expert. If it turns out that they characterize real dependencies, then they can be directly taken as fuzzy IF-THEN rules and used to characterize expert knowledge about the problem. Further research will focus on the development of other measures using the mined associations, improving search algorithms, characterization of their complexity and refining the linguistic form of associations.

4. Conclusion

In this paper, we demonstrated that the F-transform is a universal method that can be applied for data analysis. First, we focused on detection and characterization of dependencies among attributes. Second, we presented a new method for mining associations characterizing dependencies among attributes that are obtained directly from the numerical data and are expressed as specific natural language expressions. Our method is quick and the observed associations have a linguistic form and are easily understood by many. The method is illustrated in the practical example using the software LFLC 2000.

References

[1] G.Q. Chen, Q. Wei, E.E. Kerre, Fuzzy logic in discovering association rules: An overview, in: E. Triantaphyllou, G. Felici (Eds.), Data Mining and Knowledge Discovery Approaches Based on Rule Induction Techniques, Springer, NY, USA, 2006, pp. 459–493.
[2] P. Hájek, The question of a general concept of the GUHA method, Kybernetika 4 (1968) 505–515 (in Czech).
[3] P. Hájek, T. Havránek, Mechanizing Hypothesis Formation (Mathematical Foundations for a General Theory), Springer-Verlag, Berlin, Heidelberg, New York, 1978.
[4] S. Mitra, S.K. Pal, P. Mitra, Data mining in soft computing framework: A survey, IEEE Transactions on Neural Networks 13 (1) (2002) 3–14.
[5] R.J. Miller, Y. Yang, Association rules over interval data, ACM SIGMOD 26 (1997) 452–461.
[6] V. Novák, A comprehensive theory of trichotomous evaluative linguistic expressions. in: Proceedings of 26th Linz Seminar on Fuzzy Set Theory (Research Report No.71), in press, <http://irafm.osu.cz/>.
[7] V. Novák, I. Perfilieva, Dvořák: Mining pure linguistic associations on the basis of perceptions in numerical data. in: Proceedings of International Conference NAFIPS 2006, Montreal, Canada, pp. 237–240.
[8] V. Novák, I. Perfilieva, A. Dvořák, G.Q. Chen, Q. Wei, P. Yan, Mining linguistic information from numerical databases, in: Fuzzy Logic, Soft Computing and Computational Intelligence I, Eleventh IFSA World Congress, Tsinghua University Press and Springer, Beijing, 2005, pp. 546–551.
[9] V. Novák, I. Perfilieva, A. Dvořák, G.Q. Chen, Q. Wei, P. Yan, Mining pure linguistic associations from numerical data. International Journal of Approximate Reasoning (this issue).
[10] I. Perfilieva, Fuzzy transform: application to reef growth problem, in: R. Demicco, G.J. Klir (Eds.), Fuzzy Logic in Geology, Academic Press, Amsterdam, 2003, pp. 275–300.
[11] I. Perfilieva, Fuzzy transforms, in: D. Dubois et al. (Eds.), Rough and Fuzzy Reasoning: Rough versus Fuzzy and Rough and Fuzzy, Springer-Verlag, Heidelberg, 2004, pp. 63–81.
[12] I. Perfilieva, Fuzzy transforms: theory and applications, Fuzzy Sets and Systems 157 (2006) 993–1023.
[13] G. Piatecky-Shapiro, W.J. Frawley, Knowledge Discovery in Databases, AAAI Press/MIT Press, Cambridge, MA, 1991.
[14] R. Srikant, R. Agrawal, Mining quantitative association rules in large relational tables. in: Proceedings of the ACM SIGMOD International Conference on Management of Data, Montreal, Canada, 1996.
[15] M. Štepnička, R. Valášek, Fuzzy transforms for functions with two variables, in: J. Ramí k, V. Novák (Eds.), Methods for Decision Support in Environment with Uncertainty – Applications in Economics, Business and Engineering, University of Ostrava, IRAFM, 2003, pp. 96–102.