Time Series Analysis by Soft Computing Methods

N. Yarushkina
Information Systems Department
Ulyanovsk State Technical University, UlSTU
Ulyanovsk, Russia
jng@ulstu.ru

Abstract—**Qualitative evaluation and comparison of changes of indications of objects having different nature is used by designers, managers, people making decisions (PMD) and experts to make the decisions more reasonable. For suport of such activity on the analysis of changes of data connected with certain dates and time intervals, models of fuzzy time series are applied. In this article a model of fuzzy tendency the carrier of which is a fuzzy time series and its variety — elementary tendency model — is offered. The offered models are aplied for solution of the problem of summarization of fuzzy time series in terms of tendencies.**

Keywords Fuzzy Time Series, Fuzzy Tendency, Elementary Tendency, Summarization.

I. INTRODUCTION

In connection with increasing volume and speed of storage of data connected with certain dates and time intervals in data bases, the new technology of data analysis Data Mining for Time Series Data Base (TSDM) is actively developed. Forming these time sequences on the basis of domains permits to consider them in the form of time series (TS). A number of distinctions of such time sequences from traditional (classic) TS, considered in statistic theory of analysis and forecast of TS, can be noted: these time sequences are short TS; such time sequences can be represented not only by numerical but also by linguistic values; it is difficult or impossible to determine suppositions about random nature of values for such time sequences. Distinctive properties of Time Series Data Base generate changes in the aggregate of aims and tasks which form the direction Data Mining for Time Series Data Base[1]: segmentation, clusterisation, classification, indexing, summarization, disclosing anomalies, frequency analysis, forecasting, extraction of associative rules. At present the methodology of solving the indicated problems is formed on the basis of methods and models of intelligent analysis, representation and processing of data [2, 3, 4, 5]. The basis of the offered methods of intelligent data analysis is the concept of the fuzzy time series (FTS) constructed on the base of levels of TS [5, 6] or their first differences [3]. In this article an original model of FTS on the basis of fuzzy tendencies is offered. The offered model describes the behaviour of FTS in the form of the sequence of fuzzy tendencies represented by information granules. Application of information granules for problems of Data Mining for Time Series Data Base is presented in the work [7]. An indisputable advantage of data granulation is the possibility of representation of models of TS at different levels of abstraction in the linguistic form, which permits to widen the accessibility of using models of TS by application users and to improve their interpretability.

In the second part of the article models of a fuzzy time series and a fuzzy tendency are considered. In the third part a special scale for generating an elementary fuzzy tendency is given. Procedures of granulation of FTS in the basis of FT are presented in the fourth part, there is also described the solution of the problem of summarization of TS in terms of FT and presented the experiment on the solution of the problem of summarization of a synthetic TS as the problem of determination of a fuzzy tendency.

II. FUZZY TENDENCY MODEL

A. *Concept of linguistic evaluation*

One of the problems of the analysis of TS is the analysis of FTS behaviour, that is change of values of TS levels. The solution of the problem of the analysis of TS behaviour expressed in linguistic form can serve as the linguistic evaluation of the behaviour. Linguistic evaluations (LE) are the means of qualitative evaluation and comparison of characteristics and indications of objects having different nature used by designers, managers, people making decisions (PMD), experts. An important property of linguistic evaluations is wide application in practice in making decisions for expressing knowledge about the degree of correspondence of the object being observed or its characteristics to some objective or subjective criterion. The stated property determines the class of absolute LE, which reflects the static aspect of evaluation. The following expressions can be examples of such evaluations: "Satisfactory", "Good", "Bad", "Big", "Small", "Medium", etc. The semantics of absolute linguistic evaluations depends on the context of the environment in which they are used and modeled by fuzzy sets.

Another important property of linguistic evaluations is conditioned on the possibility of ranking them, it permits to present the aggregate of LE in the form of some system with relations. Binary relations formed on the set of absolute LE generate comparative linguistic evaluations by different criteria such as "More", "Less", "Approximately Equal", "Earlier", "Later", "Rather", "Better", etc. Comparative evaluations made on the basis of absolute LE can represent changes in different universes: in the universe of objects, in the time universe, in the universe of problems and they express dynamic aspect of evaluation. The semantics of comparative evaluations is also context-dependent and can be modeled on the basis of fuzzy sets.

It is noted in the article [4] that linguistic evaluation has indications expressing the degree of intensity of this evaluation. These indications can be represented in the linguistic form, usually used by people: "Very", "Insignificantly", "Approximately", etc.

Context-dependent linguistic evaluations considered above are given in expert way as a rule, and they are called expert evaluations. In case of impossibility of receiving expert evaluations of indications of objects, abstract linguistic evaluations are used, let us consider such evaluations among the class of context-free linguistic evaluations.

Let the aggregate of all linguistic evaluations forms the finite set $\widetilde{X}=\{\widetilde{x}_j\}$, where $j\in[1,nj]$. Let us call the linguistic evaluation of the indication x from the universe B a fuzzy label, if the fuzzy set is determined for it such that $\widetilde{x}_j=\{\langle w_m, \mu_{\widetilde{x}_j}(w_m)\rangle \mid w_m\in w, w\subseteq B\}$, where w_m – is the carrier of the fuzzy set, $\mu_{\widetilde{x}_j}(w_m)$ – is the membership function.

B. Model and kinds of a fuzzy tendencys

Let us introduce definitions.

Let some time series (TS) $Xt=\{t_i,x_i\}$ is given, $i\in[1,n]$, n – is the quantity of members of the series, $x_i\in B$, $t_i\in Bt$.

Let us call the ordered sequence of observations over some phenomenon the states of which change in time if the value of the state at the instant t_i is expressed with the help of the fuzzy label $\widetilde{x}_i\in\widetilde{X}$, $i\in[1,n]$, n – is the quantity of members of series, a *fuzzy time series (FTS)*. That is we represent a fuzzy time series in the form $\widetilde{X}_t=\{t_i,\widetilde{x}_i\}$, where $\widetilde{x}_i$ – i-th fuzzy set (fuzzy label), t_i – i-th value of the instant of time, $t_1\le t_i\le t_n$, n – is the quantity of members of FTS. Any TS can be represented in the form of the sequence of fuzzy labels $\widetilde{x}_j=\{\langle w_m, \mu_{\widetilde{x}_j}(w_m)\rangle \mid w_m\in w, w\subseteq B\}$ on the basis of linguistic (context-dependent or context-free) evaluation of levels of TS $x_i\in B$.

Let us call the fuzzy label $\tau_k\in\widetilde{X}$ expressing the character of change (systematic motion) of the sequence of fuzzy values of FTS $\widetilde{X}_t$ in the given interval of time the *fuzzy tendency (FT)* of a fuzzy time series. A fuzzy tendency determines the nature of FTS not in analytic, but in the linguistic form.

Each fuzzy tendency τ_k of the fuzzy time series $\widetilde{X}$ can be represented by the fuzzy set $\tau_k=\{<\tau i,\mu_{\widetilde{X}_t}(\tau i)>, i\in[1,n]\}$ with the function of membership in the fuzzy time series $\mu_{\widetilde{X}_t}(\tau i)$, where τi is the model of the following form:

$\tau i=<\nu_i,\alpha_i,\Delta t_i>$, where

ν_i – is the type of the tendency. Let us compare fuzzy labels τ_k and base types of tendencies "Increase", "Decrease", "Stability". On the basis of base types derivative types of tendencies, such as "Fluctuations", "Chaos", "Load", "Idle time", etc., can be formed.

α_i– intensity of the tendency.

Δt_i – duration of the tendency.

In case if $\Delta t_i=1$, let us consider the fuzzy tendency among the class of elementary fuzzy tendencies, if Δt_i = n-1, we consider the fuzzy tendency among the class of general fuzzy tendency of FTS, if $1<\Delta t_i<$ n-1, let us consider the fuzzy tendency among the class of local (derivative) fuzzy tendencies of FTS.

In the aspect of content an elementary fuzzy tendency models changes between two neighbouring values of the fuzzy time series $\widetilde{x}_i,\widetilde{x}_{i+1}$, and it can be compared with the instant of time t_i of FTS. A local fuzzy tendency is determined between two chosen values of the fuzzy time series $\widetilde{x}_i$ and $\widetilde{x}_j$, when $i<j$ can be compared with the instant of time t_i. Any local FT can be expressed by the sequence of elementary ET. The general fuzzy tendency characterizes the behaviour of all FTS and it is representable in the form of the sequence of local and, therefore, also elementary FT. Thus, the time series of elementary FT of the form $\{t_i, \tau_k(t_i,\widetilde{x}_i)\}$, $i\in[1,n-1]$ can be made for any FTS.

The analysis of fuzzy labels used when evaluating levels and behaviour of a time series permits to make the following conclusions:

1. Models of elementary, local, general FT of a fuzzy time series have common structure.
2. Local and general fuzzy tendencies of FTS can be expressed through the time series of elementary FT.
3. A time series of elementary FT is an invariant method of linguistic representation of behaviour of any FTS.
4. Representation of time series in the form of fuzzy time series and time series of elementary fuzzy tendencies permits to take into account additional knowledge in the form of semantics of application area during their analysis owing to use of context-dependent fuzzy labels.

In the next part we will introduce a special scale for generation of the model of the elementary fuzzy tendency.

III. ACL-SCALE FOR GENERATION OF THE MODEL OF THE ELEMENTARY TENDENCY

In this part a special linguistic scale is offered as a tool of both absolute and comparative linguistic evaluation – *ACL-scale (Absolute & Comparative Linguistic)*. This scale will be applied for construction of the model of the elementary fuzzy tendency.

The model of ACL-scale S_x for determination of absolute and comparative linguistic evaluations is representable in the form of linguistic variable with relations

$$S_x = \langle Name\_S_x, \widetilde{X}, B, G, P, TTend, RTend \rangle,$$

where $Name\_S_x$ – is the name of ACL-scale, $\widetilde{X}$ – is the base term-set of absolute LE (linguistic name of gradations), for example, $\widetilde{X}$ ={Bad, Satisfactory, Good, Excellent, ..}, $\widetilde{x}_i \in \widetilde{X}$, B – is the universal set on which the scale is determined, $x \in B$. G – are syntactic rules of deduction (generation) of chains of evaluation propositions (derivative of terms which do not enter into the base term-set), P - are semantic rules which determine membership functions for each term (they are usually given in an expert way), $TTend(\widetilde{x}_i, \widetilde{x}_j)$ – is the linguistic relation fixing the type of change between two evaluations $\widetilde{x}_i$, $\widetilde{x}_j$ of the scale, $RTend(\widetilde{x}_i, \widetilde{x}_j)$ – is the linguistic relation fixing the intensity of difference between two evaluations $\widetilde{x}_i$, $\widetilde{x}_j$ of the scale.

The relation $TTend(\widetilde{x}_i, \widetilde{x}_j)$ is the fuzzy linguistic relation which is applied for determination of comparative linguistic evaluation $v_{ij} = TTend(\widetilde{x}_i, \widetilde{x}_j)$ which characterizes the direction of change (increase or decrease) of the value of the absolute LE $\widetilde{x}_i$ with respect to $\widetilde{x}_j$ which can be represented by linguistic expressions, for example, by values from the set {INCREASE, DECREASE, STABILITY, ZERO}. Let us note that each evaluation $v_{ij} = TTend(\widetilde{x}_i, \widetilde{x}_j)$ is representable by its fuzzy set. The relation $TTend$ is antireflexive, antisymmetric and transitive:

$$\forall \widetilde{x} \in \widetilde{X} \quad TTend\ (\widetilde{x}, \widetilde{x}) = 0,$$

$$\forall \widetilde{x}, \widetilde{y} \in \widetilde{X}\ (\widetilde{x} \neq \widetilde{y}) \quad TTend\ (\widetilde{x}, \widetilde{y}) \wedge TTend\ (\widetilde{y}, \widetilde{x}) = 0$$

$$\forall \widetilde{x}, \widetilde{y}, \widetilde{z} \in \widetilde{X} \quad TTend\ (\widetilde{x}, \widetilde{z}) > TTend\ (\widetilde{x}, \widetilde{y}) \wedge TTend\ (\widetilde{y}, \widetilde{z})$$

The stated properties of the relation $TTend$ permit to classify it as an ordering relation. Then the aggregate of all possible evaluations $V=\{v_{ij}\}$ forms the fuzzy ordinal scale $S_v = \langle Name\_TTend, V, \widetilde{X}, G_v, P_v \rangle$.

The relation $RTend\ (\widetilde{x}_i, \widetilde{x}_j)$ is also the fuzzy linguistic relation applied for determination of comparative linguistic evaluation $\alpha_{ij} = RTend\ (\widetilde{x}_i, \widetilde{x}_j)$ which characterizes the degree of difference, "non-metric" distance between $\widetilde{x}_i$, $\widetilde{x}_j$ which can be expressed linguistically, for example, by values from the set {BIG, MEDIUM, SMALL, ZERO}. This evaluation α_{ij} is also representable by its fuzzy set. The relation $RTend$ is antireflexive and symmetric:

$$\forall \widetilde{x} \in \widetilde{X} \quad RTend\ (\widetilde{x}, \widetilde{x}) = 0,$$

$$\forall \widetilde{x}, \widetilde{y} \in \widetilde{X} \quad RTend\ (\widetilde{x}, \widetilde{y}) = RTend\ (\widetilde{y}, \widetilde{x})$$

The indicated properties of the relation $RTend$ permit to classify it as a relation of difference, with it the aggregate of all possible evaluations $A=\{\alpha_{ij}\}$ forms the fuzzy scale $S_a = \langle Name\_RTend, A, \widetilde{X}, G_a, P_a \rangle$.

Thus, the ACL-scale S_x for determination of linguistic evaluations is the two-level scale. At the first level of hierarchy from its universal set the ACL-scale S_x permits to determine linguistic evaluations $\widetilde{x}_i$ for values $x \in X$ which characterize their qualitative aspects. Such linguistic evaluations relate to the class of absolute LE. At the second level of hierarchy for values $\widetilde{x}_i$ and $\widetilde{x}_j$ – linguistic evaluations of their changes (v_{ij}, α_{ij}) which characterize qualitative aspects of differences or "difference of the first order" by scales S_v, S_a. Such linguistic evaluations are related to comparative LE.

Let us consider peculiarities of ACL-scales. The offered linguistic ACL-scale S_x is related to the class of fuzzy evaluation scales which enter into the class of ordinal scales, difference and the degree of difference can be additionally evaluated in it. This property permits to consider the linguistic evaluation ACL-scale S_x as quasi-interval and to determine "evaluation" and "computing" operations for it.

Let us introduce the following "evaluation" operations of the ACL-scale S_x generating linguistic evaluations:

1. The operation of determination of the absolute linguistic evaluation $\widetilde{x}_i$ by the value of characteristic of the object x_j being evaluated
$$\widetilde{x}_i = Fuzzy(x_j),\ x_j \in B,\ \widetilde{x}_i \in \widetilde{X}.$$
2. The operation of determination of the value of characteristic of the object x_j being evaluated by the absolute linguistic evaluation $\widetilde{x}_i$
$$x_j = DeFuzzy(\widetilde{x}_i),\ x_j \in B,\ \widetilde{x}_i \in \widetilde{X}.$$
3. The operation of determination of the type of difference (comparative linguistic evaluation)
$$v_{ij} = TTend(\widetilde{x}_i, \widetilde{x}_j),\ \widetilde{x}_i \in \widetilde{X}, \widetilde{x}_j \in \widetilde{X}.$$
The operation $TTend$ is non-commutative.
4. The operation of determination of intensity of difference (comparative linguistic evaluation)
$$\alpha_{ij} = RTend\ (\widetilde{x}_i, \widetilde{x}_j),\ \widetilde{x}_i \in \widetilde{X}, \widetilde{x}_j \in \widetilde{X}.$$
The operation $RTend$ is commutative.

Let us determine the aggregate of "computing" operations of the ACL-scale for generated linguistic evaluations:

1. Computing the absolute linguistic evaluation
$$\widetilde{x}_j = Comp(\widetilde{x}_i, v_{ij}, \alpha_{ij}).$$
2. The difference of intensities of differences
$$\alpha_{ij} = Diff(\alpha_i, \alpha_j)$$

3. The union of intensities of the difference

$$\alpha_{ij} = Union(\alpha_i, \alpha_j).$$

4. The intersection of intensities of differences

$$\alpha_{ij} = Inter(\alpha_i, \alpha_j).$$

Operations *Diff*, *Union*, *Inter* are commutative, associative, bounded.

Linguistic evaluations received by the indicated linguistic ACL-scale will be used in the next part as semantic-dependent for solution of the problem of summarization of FTS in terms of fuzzy tendencies.

IV. Application of elementary tendency model for summarization of fuzzy time series

Let us consider the application of the offered ACL-scale in solution of the problem of summarization of FTS as the problem of identification of its general fuzzy tendency.

For this purpose, let us design the hierarchical granular model for the initial time series $X=\{t_i,x_i\}$, $i \in [1,n]$, n – is the quantity of members of the series. Let us introduce four levels of granulation forming of which corresponds to the bottom-up approach. The zero level of granules will be represented by fuzzy labels of the initial TS. For forming the granules of the zero level, let us use the "evaluation" operation of the ACL-scale: $\widetilde{x}_i = Fuzzy(x_i)$.

Let us compare fuzzy tendencies of FTS and information granules which have structural commonness. Let us define the operation of granulation of the first level in the form of the functional *ETend* forming the granules of elementary fuzzy tendencies: $\tau i = ETend(\widetilde{x}_i, \widetilde{x}_{i+1})$, $i \in [1,n-1]$, n – is the quantity of members of FTS.

The functional *ETend* generates granules of elementary fuzzy tendencies on the basis of "evaluation" operations *TTend* and *RTend* of the ACL-scale and can be realized in the subsystem of fuzzy deduction with rules of the following form:

$R_1 :=$ IF X_t is A_{11} AND X_{t+1} is A_{12} THEN v_t is B_1 *AND* a_t *is* N_1

....

$R_m :=$ IF X_t is A_{m1} AND X_{t+1} is A_{m2} THEN v_t is B_m AND a_t is N_m

The semantics of rules of realization of the functional *ETend* is represented in the following table.

TABLE I. TABLE OF RULES

No. of the rule	$\widetilde{X}_t$	$\widetilde{X}_{t+1}$	$v_t = TTend(\widetilde{X}_t, \widetilde{X}_{t+1})$	$\alpha_t = RTend(\widetilde{X}_t, \widetilde{X}_{t+1})$
1.	Sat	Sat	St	Ze
2.	Go	Go	St	Ze
3.	Ex	Ex	St	Ze
4.	Bad	Bad	St	Ze
5.	Ze	Ze	St	Ze
6.	Ze	Bad	Inc	Sm
7.	Ze	Sat	Inc	Me
8.	Ze	Go	Inc	Bi
9.	Ze	Ex	Inc	VeBi
10.	Bad	Ze	Dec	Sm
11.	Bad	Sat	Inc	Sm
12.	Bad	Go	Inc	Me
13.	Bad	Ex	Inc	Bi
14.	Sat	Ze	Dec	Me
15.	Sat	Bad	Dec	Sm
16.	Sat	Go	Inc	Sm
17.	Sat	Ex	Inc	Me
18.	Go	Ze	Dec	Bi
19.	Go	Bad	Dec	Me
20.	Go	Sat	Dec	Sm
21.	Go	Ex	Inc	Sm
22.	Ex	Ze	Dec	VeBi
23.	Ex	Bad	Dec	Bi
24.	Ex	Sat	Dec	Me
25.	Ex	Go	Dec	Sm

The following abridgements were used in this table: *Ze(Zero), Bad(Bad), Sat(Satisfactory), Go(Good), Ex(Excellent)* for fuzzy values of FTS; *Inc(INCREASE), Dec(DECREASE), St(STABILITY)* for values of types of changes; *Bi(Big), Me(Medium), Sm(Small) and modifiers, such as Ve(Very), Si(Significantly), No(No)* for values of intensity of changes.

Let us define the operation of granulation of the second level in the form of the functional *STend* forming the granules of local fuzzy tendencies: $\tau_j = STend(\tau_i, \tau_s)$, where τ_i, τ_s are granules of the first level.

The introduced functional *STend* is computed as the result of the union of one-type elementary tendencies on the base of the "computing" operation *Union* of the ACL-scale. Then the union $\tau_j = STend(\tau_i, \tau_s)$ is the such fuzzy tendency for which $v_j = v_i$, $\alpha_j = Union(\alpha_i, \alpha_s)$, $\mu_j = \mu_i \cup \mu_s$, the duration $\Delta t_j = \Delta t_i + \Delta t_s$.

The operation of union of one-type tendencies defines the granules of the second level.

The generalized form of rules of granulation of the second level on the basis of the functional *STend* has the form:

$R_{11} := IF\ v_t\ is\ Inc\ THEN$

$IF\ a_t\ is\ A_{11}\ AND\ a_{t+1}\ is\ A_{12}\ THEN\ TInc\ is\ B_1$

....

$R_{m1} := IF\ v_t\ is\ Inc\ THEN$

$IF\ a_t\ is\ A_{m1}\ AND\ a_{t+1}\ is\ A_{m2}\ THEN\ TInc\ is\ B_m$,

$R_{12} := IF\ v_t\ is\ Dec\ THEN$

$IF\ a_t\ is\ A_{11}\ AND\ a_{t+1}\ is\ A_{12}\ THEN\ TDec\ is\ B_1$

....

$R_{m2} := IF\ v_t\ is\ Dec\ THEN$

$IF\ a_t\ is\ A_{m1}\ AND\ a_{t+1}\ is\ A_{m2}\ THEN\ TDec\ is\ B_m$

The semantics of rules of granulation of the second level is given below:

TABLE 2. TABLE OF RULES LEVEL 2

	TInc	*TDec*	*Tend*
1	Bi	Me	Inc
2	VeSm	Sm	Dec
4	Me	Bi	Dec
5	Bi	Sm	Inc
6	Me	Sm	Inc
7	Sm	Me	Dec
	….	…	…

On the basis of introduced functionals we defined the procedure of summarization of TS as the procedure of identification of the fuzzy tendency *Tend*. This procedure is the sequential generation of information granules which model TS at different abstract levels. The result of the procedure of summarization of FTS is the granule of the general fuzzy tendency which is the convolution of elementary tendencies into the linguistic evaluation of behaviour of a FTS:

$$ETend(\widetilde{x}_i, \widetilde{x}_{i+1}) \rightarrow STend(\tau_i, \tau_s) \rightarrow GTend(\tau_i, \tau_s).$$

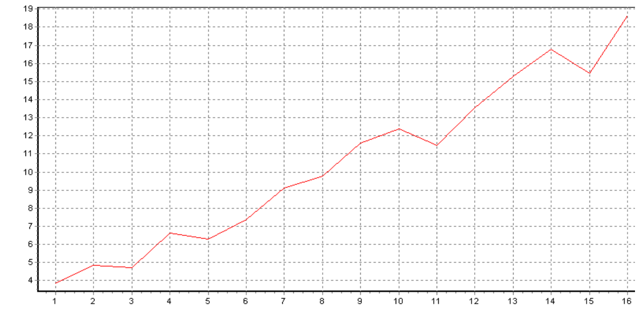

Figure 1. The time series No. 1. The result of summarization procedure: "General tendency =Increase"

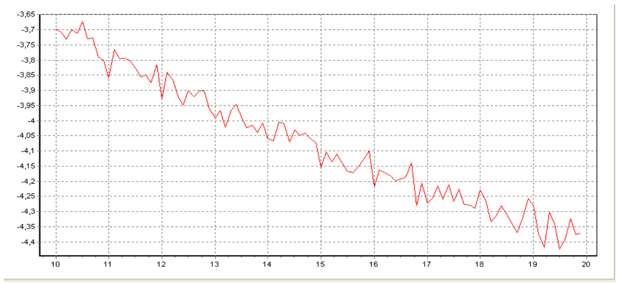

Figure 2. The time series No. 2. The result of summarization procedure "General tendency=Decrease".

The offered approach to the solution of the problem of summarization of FTS on the basis of fuzzy tendency model and granular computing was realized as software in the system generating artificial time series with noise. On figures 1, 2 examples of execution of the procedure of summarization of the artificial TS and its results are presented.

REFERENCES

[1] Batyrshinand L. Sheremetov. Perception Based Time Series Data Mining for Decision Making. /IFSA'07 Theoretical Advances and Applications of Fuzzy Logic, pp.209-219.

[2] N.G. Yarushkina Principles of the theory of fuzzy and hybrid systems: Training aid. - Moscow: Finances and statistics, 2004. - 320 p.

[3] M. Şah ,K.Y. Degtiarev. Forecasting Enrollment Model Based on First-Order Fuzzy Time Series. İnternational Conference on Computational İntelligence (İCCİ) 17-19 December 2004, İstanbul, Turkey

[4] Dvorak A.,Novak V.:Formal Theories and Linguistic Description. Fuzzy Sets and Systems, 143(2004), 169-188 .

[5] Song, Q., Chissom, B.S. Fuzzy time series and its models. Fuzzy Sets and Systems , 54 (1993) 269–277

[6] Huarng, K. Heuristic models of fuzzy time series for forecasting. Fuzzy Sets and Systems, 123: 369-386. Young, *The Technical Writers Handbook,* Mill Valley, CA: University Science, 1989.

[7] L.A. Zadeh. Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. Fuzzy Sets and Systems , 90(1997), 111-127