Linguistic Models and Linguistic Modeling

Witold Pedrycz, *Fellow, IEEE*, and Athanasios V. Vasilakos

***Abstract*— The study is concerned with a linguistic approach to the design of a new category of fuzzy (granular) models. In contrast to numerically driven identification techniques, we concentrate on building meaningful linguistic labels (granules) in the space of experimental data and forming the ensuing model as a web of associations between such granules. As such models are designed at the level of information granules and generate results in the same granular rather than pure numeric format, we refer to them as linguistic models. Furthermore, as there are no detailed numeric estimation procedures involved in the construction of the linguistic models carried out in this way, their design mode can be viewed as that of a rapid prototyping. The underlying algorithm used in the development of the models utilizes an augmented version of the clustering technique (context-based clustering) that is centered around a notion of linguistic contexts—a collection of fuzzy sets or fuzzy relations defined in the data space (more precisely a space of input variables). The detailed design algorithm is provided and contrasted with the standard modeling approaches commonly encountered in the literature. The usefulness of the linguistic mode of system modeling is discussed and illustrated with the aid of numeric studies including both synthetic data as well as some time series dealing with modeling traffic intensity over a broadband telecommunication network.**

***Index Terms*—Context-based clustering, fuzzy sets, information granularity, linguistic model, rapid prototyping, telecommunication networks.**

I. INTRODUCTION

FUZZY modeling has emerged as an interesting, attractive, and powerful modeling environment applied to numerous system identification tasks. The key features being emphasized very often in this setting concern a way in which fuzzy sets enhance or supplement the existing identification schemes. It was Zadeh [19], [20] who has introduced the concept of fuzzy models and fuzzy modeling. The enhancements of system modeling conceived within this framework take place at the conceptual level as well as at the phase of detailed algorithms. In a nutshell, fuzzy models are concerned with the modeling pursuit that occurs at the level of linguistic granules (fuzzy sets or fuzzy relations) rather than the one that happens at a detailed and purely numeric level encountered in other modeling approaches. What fuzzy sets offer in system modeling is another more general and holistic view at the resulting model that gives rise to their augmented interpretation and better utilization. From a computational point of view, fuzzy sets are inherently nonlinear (viz. their membership functions are nonlinear mappings). As a consequence of such nonlinear character, one may anticipate that this feature augments the representation power of the fuzzy models. There have been a substantial number of various schemes of fuzzy modeling along with specific algorithmic variations that help eventually capture some specificity of the problem at hand and contribute to the efficiency of the overall identification schemes, cf., [5]-[7], [9], [11], [12], [14], [18]. Quite often, in order to take advantage of numeric experimental data, the modeling algorithms exploit neurofuzzy techniques, see e.g., [2]–[4], [13], [15].

The intent of this study is to design a generic linguistic granules-oriented modeling technique that reveals and binds linguistic granules in an explicit fashion. Our primary thrust is in the delivery of the transparency of the modeling scheme itself as well as assure a high readability of the resulting construct. The fuzzy model designed in this way is fully developer-controllable (meaning that one can easily modify some crucial components, that is information granules to affect the entire model). The outstanding features of the proposed novel approach to the linguistic modeling are as follows.

1) The fuzzy model captures relationships (associations) between linguistic granules in the data space and does not confine itself to any specific structural relationships between the variables (such as rules, nonlinear mappings, linear dependencies, etc.). Practically, we do not make any assumption about logic operations used within the fuzzy models (as this has been usually found in practice in the existing models such as inference schemes of approximate reasoning). The lack of any strong preliminary commitments as to the structure of the model is another essential feature of the linguistic approach.
2) Being structure-free, the obtained linguistic models can serve as a general modeling blueprint. If required, the linguistic model can be easily enhanced to capture and refine further details (e.g., such refinements can be realized in the form of a series of local regression models [6], [7], [9], [11], [12], [15], [18], [21]).

The material is arranged in such a way so that it unveils the construction of the fuzzy model in a systematic, step-by-step, and highly readable fashion. First, we discuss a role of information granularization in fuzzy modeling (Section II). Subsequently, we proceed with more detailed construction of these linguistic landmarks (Section III)—the process that lends itself to a so-called context-oriented fuzzy clustering. We show that the linguistic granules are fully data compliant,

Manuscript received May 12, 1998; revised April 24, 1999 and September 18, 1999. This research was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC) and EU project ACTS MISA. This paper was recommended by Associate Editor R. Rada.

W. Pedrycz is with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton T6G 2G7, Alta., Canada. He is also with the Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland.

A. V. Vasilakos is with the Telecommunications and Networks Group, ICS-FORTH, 711 10 Herakion Crete, Greece.

Publisher Item Identifier S 1083-4419(99)09701-0.

namely they reflect the topology of the experimental data. In Section IV we reveal a more detailed way in which the model is utilized. Experimental studies involving both synthetic and real-world data originating from the area of telecommunications are reported in Section V.

II. Linguistic Granularization of Data as a Basis of System Modeling: An Underlying Concept

In this section, we study the generic concept of linguistic modeling. The starting point is to look at any linguistic model as an association of information granules (linguistic terms) defined over some variables of the system. Quite descriptively, one may allude to such linguistic granules or linguistic landmarks as being focal point of all modeling activities. Linguistic granules are viewed as clumps of objects (data points, in particular) drawn together by the criteria of indistinguishability, similarity or functionality. Such collections can be modeled in several formal environments including set theory, rough sets, random sets or fuzzy sets. The last option is of particular interest. Fuzzy sets offer two interesting and useful features supporting processes of information granulation and the form of information granules resulting therein. First, fuzzy sets support modeling of concepts that exhibit continuous boundaries. The overlap between fuzzy sets (that is an inherent phenomenon to the theory of fuzzy sets) help avoid brittleness effect when moving from one concept to another. Second, fuzzy sets exhibit a well-defined semantics and emerge as fully meaningful conceptual entities—building modules defined in problem solving.

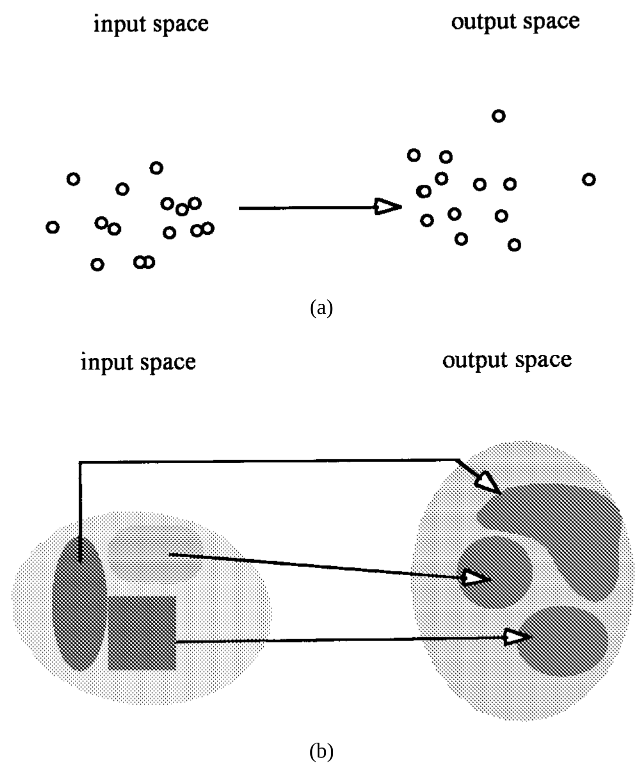

Fig. 1. Numeric models versus linguistic models. (a) Numeric models attempt to build a detailed numeric or nonlinear mapping based on available numeric data. (b) In linguistic modeling, all pursuits are concentrated at the level of linguistic granules and the resulting mappings (models) are developed at the level of information granularity implied by the already selected information granules.

It is eventually highly instructive to contrast the methodology of *granular* modeling with its counterpart that is *numeric* system modeling. It goes without saying that the pursuit of any modeling is to make sense of data and represent them in a concise and easy to interpret form of numeric dependencies. The numeric style of system modeling concerns a determination of linear or nonlinear relationships between systems variables, refer to Fig. 1(a). In particular, we may talk about linear or nonlinear regression or neural networks—all of these fall under the same category of detailed numeric models. In contrast, fuzzy models attempt to establish meaningful dependencies between linguistic granules—modeling landmarks defined in the respective spaces of data as exemplified in Fig. 1(b). This means that we start off any modeling pursuit by constructing these landmarks. More descriptively, they tend to form a conceptual skeleton of the model. One of the plausible options available to us when it comes to the formation of these design artifacts arises in the setting of fuzzy clustering. We discuss this development of the linguistic model in the ensuing section.

Prior to proceeding with any design details, it is instructive to highlight the key features of the linguistic models and its development methodology.

1) The linguistic models exhibit a significant level of conceptual flexibility. This flexibility comes as a direct consequence of the use of information granules of varying specificity. The specificity of an information granule correspond to its size that is a number of elements embraced by this granule: the higher the specificity, the lower the number of elements encapsulated by the granule. For information granules represented by sets, we can compute their cardinality to gauge the level of specificity. The corresponding instrument of expressing granularity of fuzzy sets comes in the form of σ-count (cardinality), cf., [6], [7], [11], [21]. By increasing the specificity of the information granules, the models can capture more details (it becomes more specific). The lower specificity of the information granules used in the construct contributes to the effect of hiding some details that may not be pertinent to the particular modeling pursuits. Similarly, this information hiding effect reduces an overall computational effort.
2) The way of linguistic modeling discussed there puts the designer in a more active position as far as the constructed model is concerned. This is accomplished in two ways. First, the specification of the level of granularity helps the designer to cast the problem (and the ensuing model) in a certain perspective that is the most suitable for the future use of the model. Too many details may not be beneficial to the user who could be easily confused by their amount especially if the intent is to gain a comprehensible and quite condensed description of the overall system under modeling. On the other hand, too low granularity may easily result in the model that hides too many essential aspects and therefore may easily become irrelevant to some groups

of the users. The same information granules can help the designer to concentrate modeling pursuits around some linguistic landmarks that are deemed essential to the potential users. This focus of attention can be easily accomplished by moving fuzzy sets along the corresponding variables so that the modal values of these fuzzy sets (or fuzzy relations in a many-dimensional case) are placed over the regions of the variables that are of particular interest. Interestingly, granulation of information as realized by fuzzy sets is very much in line of the main pursuits of the modeling fundamentals as expressed by S. Karlin at the 11th R.A. Fisher Memorial Lecture in 1983 "*... the purpose of models is not to fit the data but to sharpen the questions*."

3) Linguistic models are examples of rapid prototypes. Building linguistic models is an example of rapid prototyping. As the overall linguistic model relies on the information granules (both predefined by the designer and induced by the clustering method) and does not require any further optimization (estimation of the values of the parameters of the model), once information granulation has been completed, the model is quickly assembled by downloading the granules onto the models architecture. By working with the information granules, we do not commit ourselves to any detailed structure of the model as it becomes required in other modeling approaches (say rigidly structured regression models).
4) Owing to the granular form of the results produced by the linguistic models, the models become more user-friendly. In essence, the user is provided with a host of possible results (encapsulated in the form of the output fuzzy set) along with their degrees of preference (that is membership values). The visualization of the results delivers another dimension of the modeling by helping the user understand the nature of the results and become more cognizant about their relevance.

III. Context-Inclined Construction of the Linguistic Landmarks in the Data Space

To illustrate the very idea in which clustering, and fuzzy clustering, in particular plays in system modeling, let us consider a relational table (array) $\boldsymbol{X}$ comprising data elements regarded as vectors of real numbers. We are interested in revealing (discovering) a structure and eventual quantifying functional dependencies manifesting throughout this experimental log. The focal nature of fuzzy modeling is achieved by specifying linguistic term(s) prior to launching any detailed analysis and running computationally intensive algorithms. Let us consider the output variable to form a *context variable* and define therein a fuzzy set (linguistic term of focus) A

$$A\colon \boldsymbol{Y} \rightarrow [0,1].$$

In the above, $\boldsymbol{Y}$ stands for a universe of discourse of this attribute (variable). The clustering problem reads in the following way:

reveal structure in the space of input variables $\boldsymbol{X}$ in context $\boldsymbol{A}$

where the corresponding context is established through a family of fuzzy sets

$$\mathcal{A} = \{A\colon \boldsymbol{X} \rightarrow [0,1]\}.$$

It is instructive to recall a basic clustering method and establish notation before getting into the respective modifications capturing the essence of the granular modeling problem. The so-called objective function-based clustering hinges solely upon a certain performance index Q whose minimization should reveal some meaningful structures in the data set. Quite commonly, one adopts a sum of distances between the individual data points and the prototypes (centroids) arising as an integral part of the clustering procedure. The performance index (objective function) assumes the following form:

$$Q = \sum_{i=1}^{c} \sum_{k=1}^{N} u_{ik}^{m} \|\boldsymbol{x}_k - \boldsymbol{v}_i\|^2 = \sum_{i=1}^{c} \sum_{k=1}^{N} u_{ik}^{m}\, d_{ik}^2.$$

In the above "c" denotes the number of clusters. The fuzzification coefficient "m" $(m > 1)$ is used to affect the values of the partition matrix (we discuss its role in the following discussion). The partition matrix (U) is used to store all results of clustering (partitioning) the patterns (data points) into clusters. Depending whether we are oriented toward set-oriented or fuzzy set oriented partitioning (clustering), the partition matrix satisfies the following conditions:

1) for set-oriented clustering $u_{ik} \in \{0,1\}$

$$\begin{array}{l} 0 < \displaystyle\sum_{k=1}^{N} u_{ik} < N \quad \text{for } i = 1, 2, \cdots c \\ \displaystyle\sum_{i=1}^{c} u_{ik} = 1 \text{ for } k = 1, 2, \cdots, N \end{array};$$

2) for fuzzy-set-oriented clustering (fuzzy clustering) we assume that $u_{ik} \in [0,1]$ with the same two conditions as stated above.

These conditions give rise to a family of partition matrices; we will be using the pertinent notation $\mathcal{U}$ to describe the entire family of them:

$$\mathcal{U} = \left\{ U | 0 <, \sum_{k=1}^{N} < N, 0 < \sum_{k=1}^{N} u_{ik} < N \right\}.$$

Note that the above conditions assume a straightforward and intuitively legitimate interpretation:

1) each cluster is nonempty;
2) the element (data point, pattern) belongs to a single cluster (set-oriented clustering) or the sum of membership values to all clusters is equal to 1 (fuzzy clustering).

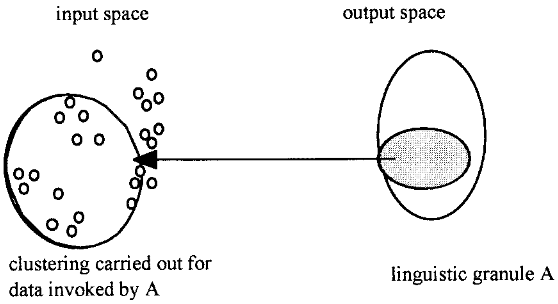

Fig. 2. Context-based clustering for fuzzy modeling.

The corresponding rows of the partition matrix form characteristic sets or fuzzy sets formed over the entire data set, namely

$$U = \begin{bmatrix} \boldsymbol{u}_1 \\ \boldsymbol{u}_2 \\ \boldsymbol{u}_c \end{bmatrix}$$

where $\boldsymbol{u}_i$ embraces the membership function of the ith cluster. The optimization is realized through an iterative procedure that can be described as a series of steps:

(initialization) Define the number of clusters (c), fix the distance function and decide upon the value of the fuzzification factor (m) in the objective function (quite commonly we assume that m is equal to 2). Initiate randomly the fuzzy partition.

(iterative computation) Compute the prototypes $(\boldsymbol{v}_i)$ and update the partition matrix (U) based upon the first-order conditions of the minimized objective function. The computations are terminated once some termination criterion is satisfied.

The conditioning aspect (context sensitivity) of the clustering mechanism is introduced into the algorithm by taking into consideration the conditioning variable (context) assuming the values $f_1, f_2, \cdots, f_N$ on the corresponding data points. More specifically, f_k describes a level of involvement of $\boldsymbol{x}_k$ and y_k in the assumed context, $f_k = A(k)$. In other words, A acts as a data mining filter (or a focal element sometimes referred to as a data window) by focusing attention on some specific subsets of experimental data.

The concept of the context-based clustering [10] attempts to reflect upon the output variable while clustering the remaining data. This means that we first agree upon some granulation of the output variable of the model (which again could be guided by some semantically sound criterion) and afterwards produce some information granules being, in fact, induced by the successive fuzzy sets already formed for the output variable. This phenomenon of focus of attention (invoking only a subset of data in the input space) is illustrated in Fig. 2.

As an illustration, one can view the linguistic granule of context as a fuzzy set of low or large values of the variable. The way in which f_k can be associated with or allocated among the computed membership values of $\boldsymbol{x}_k$, say $u_{1k}, u_{2k}, \cdots, u_{ck}$, is not unique. Two possibilities are immediately envisioned:

1) we admit f_k to be distributed additively across the entries of the kth column of the partition matrix meaning that

$$\sum_{i=1}^{c} u_{ik} = f_k \qquad k = 1, 2, \cdots, N;$$

2) we may request that the maximum of the membership values within the corresponding column equals f_k

$$\max_{i=1}^{c} u_{ik} = f_k \qquad k = 1, 2, \cdots, N.$$

In general, one can envision some other alternatives, that is regard f_k as a function of the resulting membership values.

In the sequel, we confine ourselves to the first way of distribution of the conditioning variable. Bearing this in mind, let us modify the requirements to be met by the partition matrices and define the family

$$\mathcal{U}(\boldsymbol{f}) = \left\{ u_{ik} \in [0,1] \middle| \sum_{i=1}^{c} u_{ik} = f_k \ \forall_k \text{ and } 0 < \sum_{k=1}^{N} u_{ik} < N \ \forall_i \right\}.$$

The optimization problem guiding the clustering of data is now reformulated in the following manner [10], [11]:

$$\min_{U, \boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_c} Q$$
$$\text{subject to} \quad \mathcal{U} \in U(\boldsymbol{f}).$$

Let us now proceed with a derivation of a complete solution to this optimization problem. Essentially, this task splits into two separate subproblems:

1) an optimization of the partition matrix U;
2) an optimization of the prototypes.

As these tasks can be handled independently from each other, we start with the optimization of the partition matrix. Moreover, one can notice that each column of U can be optimized independently, so let us fix the index of the data point (k) and rephrase the resulting problem as follows:

$$\min \sum_{i=1}^{c} u_{ik}^m \, d_{ik}^2$$
$$\text{subject to} \quad \sum_{i=1}^{c} u_{ik} = f_k$$

(thus with the fixed data index, we end up with solving "N" independent optimization problems). In the sequel, to make the notation more concise, we introduce the notation d_{ik} to denote the distance between the kth pattern and the ith prototype, namely $d_{ik}^2 = \|\boldsymbol{x}_k - \boldsymbol{v}_i\|^2$.

As the above is an example of optimization with constraints, we can easily convert this into unconstrained optimization by considering the technique of Lagrange multipliers. Not getting into details, the derived iterative optimization involves the calculations of the partition matrix and the prototypes completed in the following way:

$$u_{ik} = \frac{f_k}{\displaystyle\sum_{j=1}^{c} \left(\frac{\|\boldsymbol{x}_k - \boldsymbol{v}_i\|}{\|\boldsymbol{x}_k - \boldsymbol{v}_j\|} \right)^2}$$

and

$$\boldsymbol{v}_i = \frac{\sum_{k=1}^{N} u_{ik}^{m} \boldsymbol{x}_k}{\sum_{k=1}^{N} u_{ik}^{m}} \qquad i = 1, 2, \cdots, c, k = 1, 2, \cdots, N.$$

The convergence conditions for the method are the same as discussed in the case of the original FCM algorithm [1].

There are two essential design components of the clustering method:

1) the distance function $||\cdot||$;
2) the fuzzification parameter (m).

The distance addresses an important issue of defining a similarity between two elements (data). In general, a Minkowski distance is often used with the Hamming, Euclidean, and Tschebyschev distance being viewed as its special cases. The fuzzification factor (m) affects the form of the clusters being produced (or, equivalently, the form of membership function). With the increasing values of "m" there is a profound rippling effect where the membership functions tend to show up more local minima. For lower values of the membership functions tend to resemble characteristic functions of sets meaning that we are getting less elements with intermediate membership values. It is interesting to underline that $m = 2$ constitutes a reasonable compromise between set-like membership function and those with excessive oscillations in the membership values.

It is of interest to elaborate more on the remaining parameter of the clustering method that is the fuzzification factor (m). Its values become reflected in the form of the clusters being produced (or, equivalently, the form of membership function). In order to clarify this effect, let us refer to Fig. 3. With the increasing values of "m" there is a profound rippling effect where the membership functions tend to show up more local minima. For lower values of the membership functions tend to resemble characteristic functions of sets meaning that we are getting less elements with intermediate membership values.

The context A has a dominant effect on the performance of the clustering mechanism. If $\boldsymbol{f} < \boldsymbol{f}'$ then the population of the data involved in grouping (clustering) and placed under context $\boldsymbol{f}'$ is lower. Similarly, the number of eventual clusters could be lowered as well. The above inclusion relation between the contexts holds if the context fuzzy sets are made more specific or if the contexts consist of more constraints (focal points).

IV. Granular Fuzzy Model: A Linguistic Blueprint of Fuzzy Modeling

The context-based clustering leaves us with the number of contexts and induced clusters. The links (associations) between these entities are assumed by the method but not quantified at all. What we are provided with once the contextual clustering has been completed is a structure one can portray in Fig. 4.

This figure summarizes the associations as being formed by the context-based clustering. What has been constructed

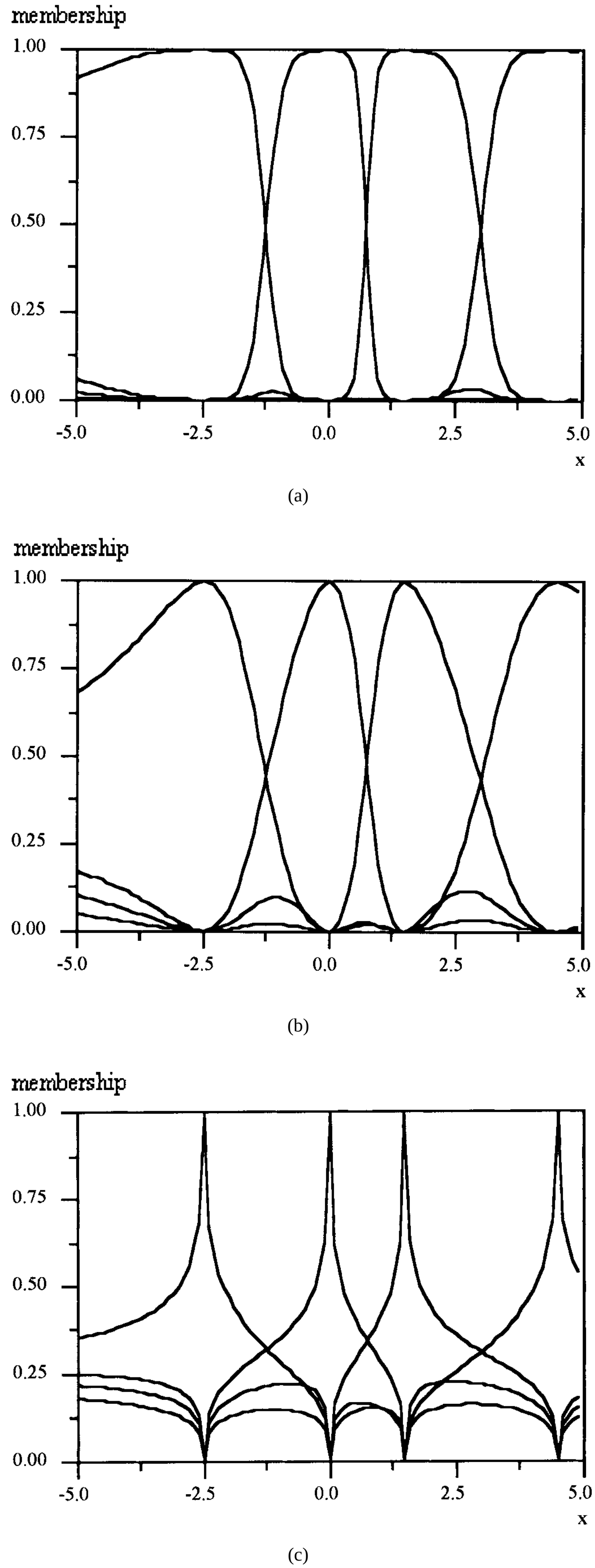

Fig. 3. Examples of membership functions generated for selected values of "m" for the same prototypes (a) = 1.5, (b) $m = 2$, and (c) $m = 5$.

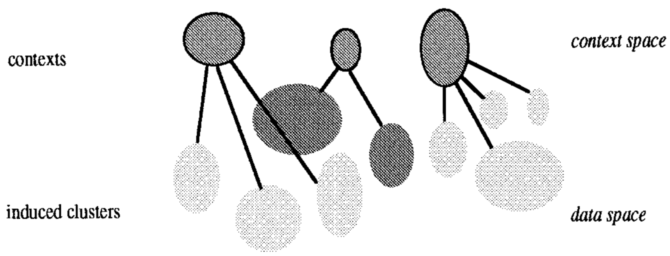

Fig. 4. Linguistic contexts and induced clusters-formation of a web of generic associations.

in this manner, becomes eventually the most *descriptive* and least *restrictive* (but yet operational) realization of the fuzzy model. Still a lot of details are missing whose calibration (adjustments) could help design more precise fuzzy model. This is, however, beyond our current interest.

But the clusters, contexts and the associations are enough to construct an operational fuzzy model. These components can be redrawn in an alternative yet equivalent form to this in Fig. 5. By doing this we arrive at an interesting fuzzy neural network—the network that produce genuine fuzzy sets (more precisely fuzzy numbers) as a result of processing going therein. Let us note that the architecture in Fig. 5 comprises a single hidden layer formed by the induced clusters; the activation levels are denoted by z_{ii}. These two indexes help distinguish between successive clusters (j) implied by a specific context (i). Each unit is connected to the output layer composed of a single linear unit. While the activation levels are numerical, the connections between the hidden and output layer are linguistic and they are just the contexts used for the previous clustering purposes. As a matter of fact, the summation completed at the output layer can be read as

$$\begin{aligned} Y = &(z_{11} \otimes A_1 \oplus z_{12} \otimes A_1 \oplus \cdots \oplus z_{1n_1} \otimes A_1) \\ &\oplus (z_{21} \otimes A_2 \oplus z_{22} \otimes A_2 \\ &\oplus \cdots \oplus z_{2n_2} \otimes A_1) \\ &\oplus (z_{c1} \otimes A_c \oplus z_{c2} \otimes A_c \oplus \cdots \oplus z_{cn_c} \otimes A_c) \end{aligned}$$

where the addition $\oplus$ and multiplication $\otimes$ are completed with the use of the standard techniques of fuzzy arithmetic.

If the contexts have been specified as triangular fuzzy numbers, say

$$A_i = (a_{i-}, a_i, a_{i+})$$

where these three numbers denote the lower, modal and upper bound of A_i, then the output of the network is again a triangular fuzzy number $Y = (y_-, y_m, y_+)$ with the bounds computed based upon the bounds of the individual contexts. The calculations follow the fundamentals of fuzzy arithmetic. For the sake of completeness, the pertinent results are summarized in the Appendix. Using them in the multivariable case, we obtain the parameters of the fuzzy number of the output

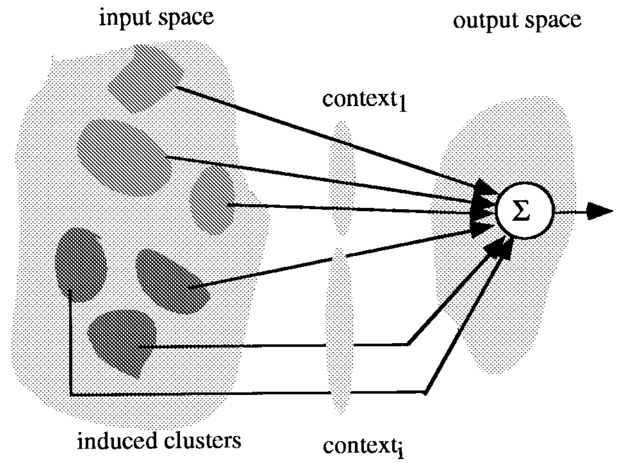

Fig. 5. A fuzzy granular model: a general topology.

1) lower bound

$$\begin{aligned} y_- = &(z_{11}a_{1-} + z_{12}a_{1-} + \cdots + z_{1n1}a_{1-}) \\ &+ \cdots + (z_{c1}a_{c-} + z_{c2}a_{c-} + \cdots + z_{cnc}a_{c-}); \end{aligned}$$

2) modal value

$$\begin{aligned} y_m = &(z_{11}a_1 + z_{12}a_1 + \cdots + z_{1n1}a_1) \\ &+ \cdots + (z_{c1}a_c + z_{c2}a_c + \cdots + z_{cnc}a_c); \end{aligned}$$

3) upper bound

$$\begin{aligned} y_+ = &(z_{11}a_{1+} + z_{12}a_{1+} + \cdots + z_{1n1}a_{1+}) \\ &+ \cdots + (z_{c1}a_{c+} + z_{c2}a_{c+} + \cdots + z_{cnc}a_{c+}). \end{aligned}$$

V. Numerical Experiments

This section includes two detailed numeric examples illustrating the design of the linguistic model in detail. The first example concerns synthetic two-dimensional data that are useful from the standpoint of visualization of the method. The second one comes from the telecommunication domain and deals with a time series of load reported in a telecommunication network.

Example 1: We consider a nonlinear function of two variables

$$y = f(x_1, x_2) = 0.6 + 2x_1 + 4x_2 + 0.5x_1x_2 + 25\sin(0.5x_1x_2)$$

defined in the Cartesian product of two intervals, $\boldsymbol{X} = [-4, 6] \times [-2, 4]$. The three-dimensional plot of the function as well as its contour map are included in Figs. 6 and 7.

The training data set is composed of 100 input–output data and was generated randomly from the already defined universe of discourse $\boldsymbol{X}$ (see Fig. 8).

We use nine contexts. Each of them is described in the form of the triangular membership function whose parameters are listed below (note that each triangular fuzzy set is described

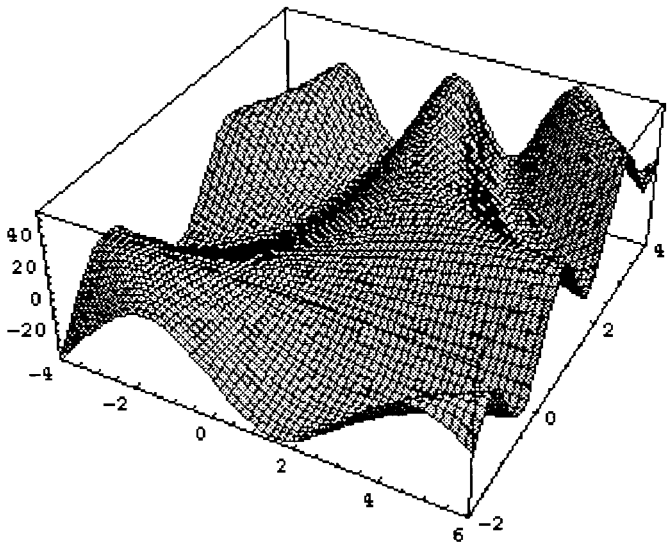

Fig. 6. A 3-D plot of the two-variable nonlinear function.

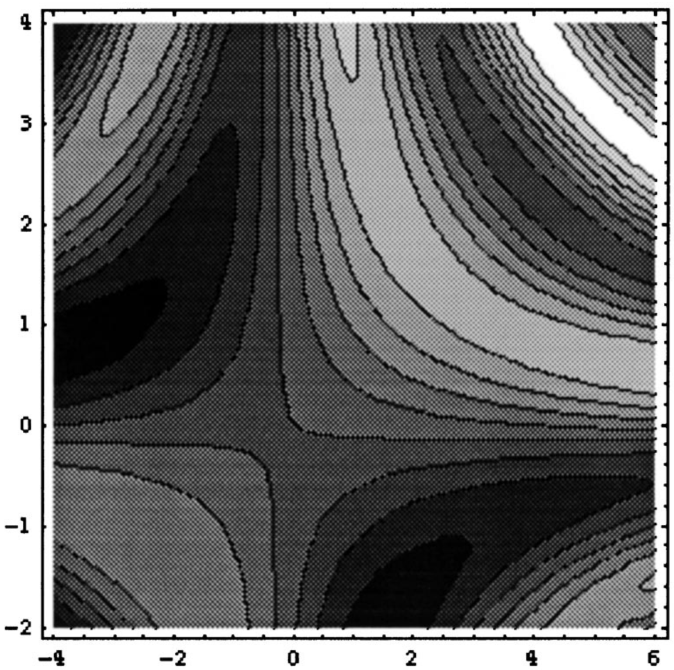

Fig. 7. Contour plot of the function used in the experiment.

by its lower bound, modal value, and upper bound)

context no.	Triangular fuzzy set of context
1	$(-50\ -25\ -5)$
2	$(-10\ -5\ 0)$
3	$(-5\ 0\ 5)$
4	$(0\ 5\ 10)$
5	$(5\ 10\ 20)$
6	$(10\ 20\ 30)$
7	$(20\ 30\ 40)$
8	$(30\ 40\ 50)$
9	$(40\ 80\ 100)$.

The context-based clustering is completed for $c = 4$ clusters for each context. In total, this generates 36 prototypes—centroids forming the backbone (blueprint) of the linguistic model. The prototypes for selected contexts are illustrated in Fig. 9; these are shown in a superposition with the original training data.

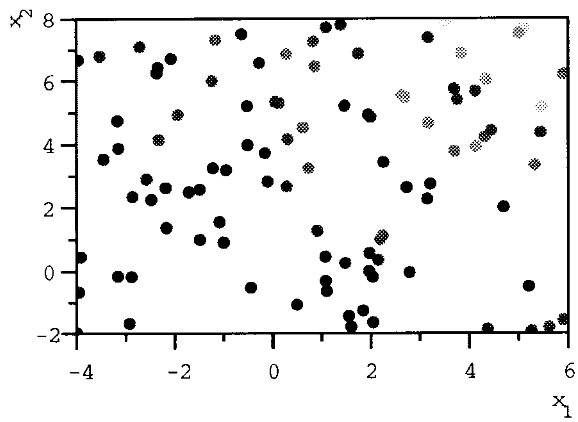

Fig. 8. A set of training data used for the construction of the linguistic model.

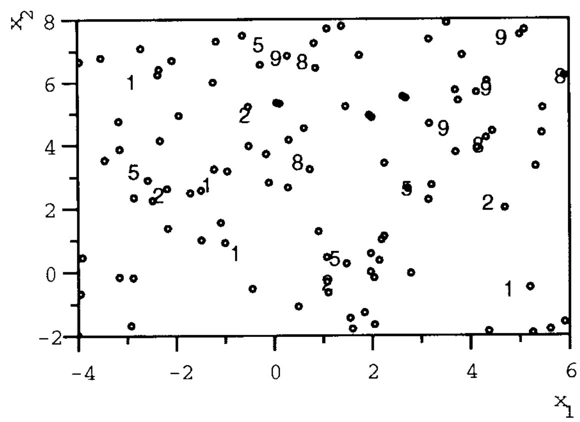

Fig. 9. Prototypes of the clusters (they are indicated by showing the number of the corresponding context).

The results produced by the linguistic model are visualized in Fig. 10; we include the lower and upper bound of the triangular fuzzy sets generated by the model. Observe that the size of the bounds is not fixed and fluctuates quite substantially when moving from one data point to another. Moreover in most cases the original training data fall within the bounds of the fuzzy set generated by the linguistic model.

It is instructive to contrast the linguistic model with the linear regression model regarded as a general yardstick commonly utilized in system modeling. The regression model minimizing the sum of squared errors assumes the form

$$y = 3.652 + 3.86x_1 + 3.489x_2.$$

To make this comparison possible, one has to use numeric representation of the linguistic output of the model. The modal value of the resulting fuzzy set arises as a legitimate choice. The sum of squared errors (differences between the data and model's output) Q is taken as a performance index expressing the quality of the model. The linguistic model surpasses the linear regression model; the values of the performance index

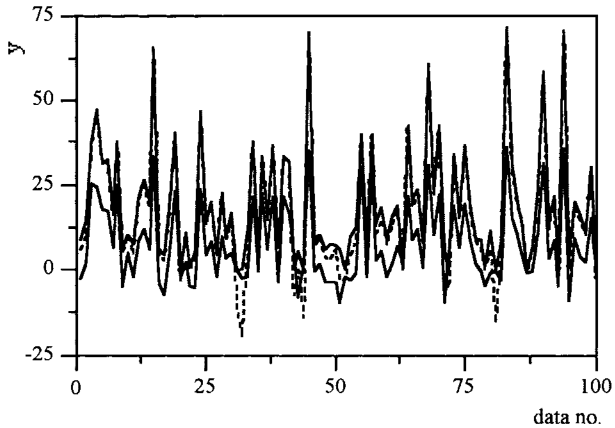

Fig. 10. The output of the linguistic model; visualized are bounds of the triangular numbers (outputs) of the model (solid lines) while the original training data are shown using a dashed line.

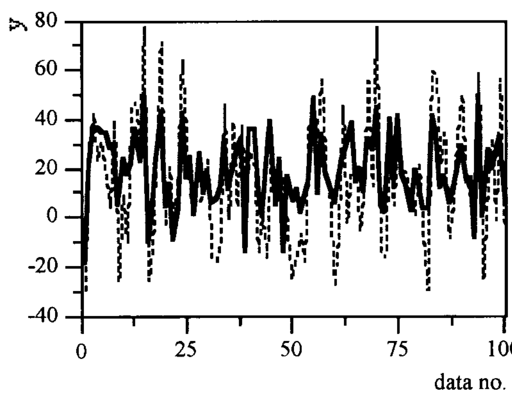

Fig. 11. Training data (dotted line) versus the linear regression model (solid line).

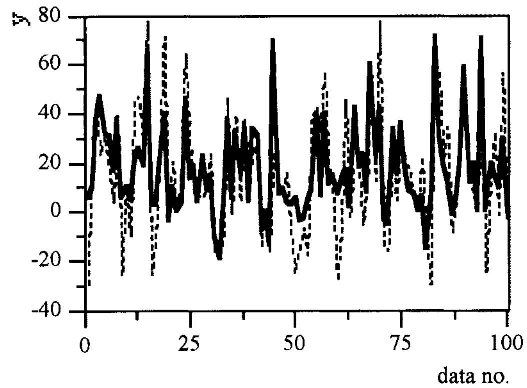

Fig. 12. Training data (dotted line) versus the fuzzy model (solid line).

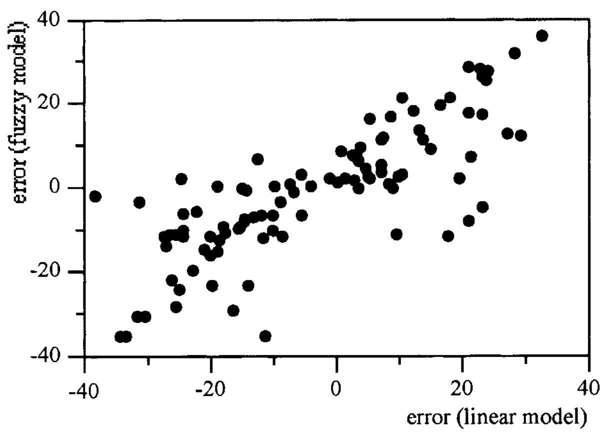

Fig. 13. Error of the linguistic model versus error of the linear regression model.

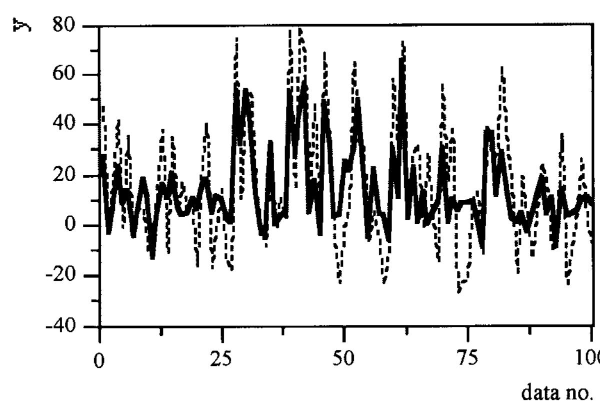

Fig. 14. Testing data (dotted line) versus the results of the linguistic model (solid line).

are equal to

$$\text{regression model:} \quad Q = 33558.2$$
$$\text{linguistic model:} \quad Q = 26\,311.6.$$

It is also worth stressing that even though we have not minimized this performance index for the linguistic model, it performs better than the linear regression model (which was constructed through the minimization of the same performance index). The results produced across the entire training data set are summarized in Figs. 11 and 12.

The individual error values contrasted for the two models are summarized in Fig. 13; the linear relationship between them can be characterized by the relationship

$$\text{error (linguistic model)} \\ = 0.6991 \text{ error (linear regression model)} + 0.4992.$$

Then the performance of the linguistic model is examined for the testing set. This set consists of 100 data points randomly drawn from the universe of discourse $\boldsymbol{X}$. Again in this case the linguistic model performs better than the regression one yielding the value of Q equal to 29 573.7. The linear model is characterized by the value of the performance index equal to 35 337.4. The performance of the two models is included in Figs. 14 and 15.

Example 2: The experimental data set discussed in this example represents a network load recorded in a certain broadband telecommunication network. Data were collected in 5 min intervals over a period of more than 1 year, using

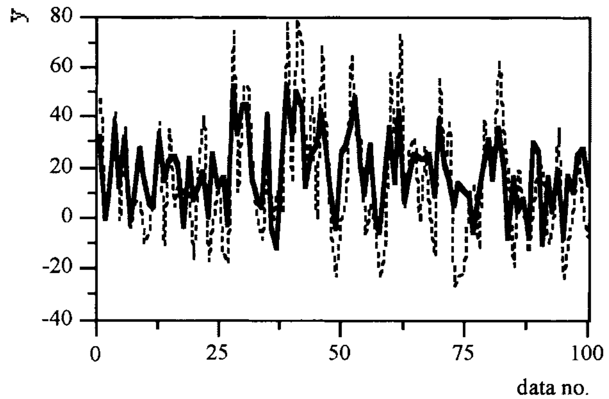

Fig. 15. Testing data (dotted line) versus results of the linear regression model (shown by solid line).

typical SNMP get requests from a WAM routers interface on a long-distance WAN link. The load is expressed in terms of total bandwidth utilization. These traffic statistics exhibit some of the typical characteristics, mainly regular traffic patterns (hourly, daily, and weekly cycles), and self similarity occurring on several different time scales and long-term chaotic behavior [8], [17]. This phenomenon becomes apparent from the sample data plots, as shown in Fig. 16.

As described in Vasilakos *et al.* [16], the predictor will be used in an integrated routing-scheduling architecture to route incoming connections with specific Quality of Service (QoS) requirements through a specific path, which will have the least impact on global and local network performance, as measured by a specific metric. Assuming that the connection characteristics do not change during the connections lifetime, to calculate these metrics, one has to predict the future trends of the utilization level of the link. An evident advantage of this scheme is that the routing process can be made aware and automatically adapt to the temporal patterns that are present in the network utilization (for example, it could route a connection through paths which are likely to be underutilized due to different time zones). We anticipate that this scheme will be able to exploit a wide range of situations in a way that seems to be both natural and exhibit some intelligent behavior.

The experimental data deal with an intensity of traffic over a certain part of a telecommunication network. The training section of the data set (consisting of around 900 data points) is shown in Fig. 17.

The fuzzy model is regarded as a time series of the fifth order meaning that the input space consists of five consecutive samples occurring in the previous time instances. Generally, this produces the relationship

$$x(k+1) = \text{association}\ (x(k), x(k-1), \cdots, x(k-5))$$

(note that we have decided to indicate a sort of association as the model itself is built using linguistic granules rather than plain numbers). Hence the above notation needs to be interpreted accordingly. The granularization of the data space and revealing the associations follows the way already

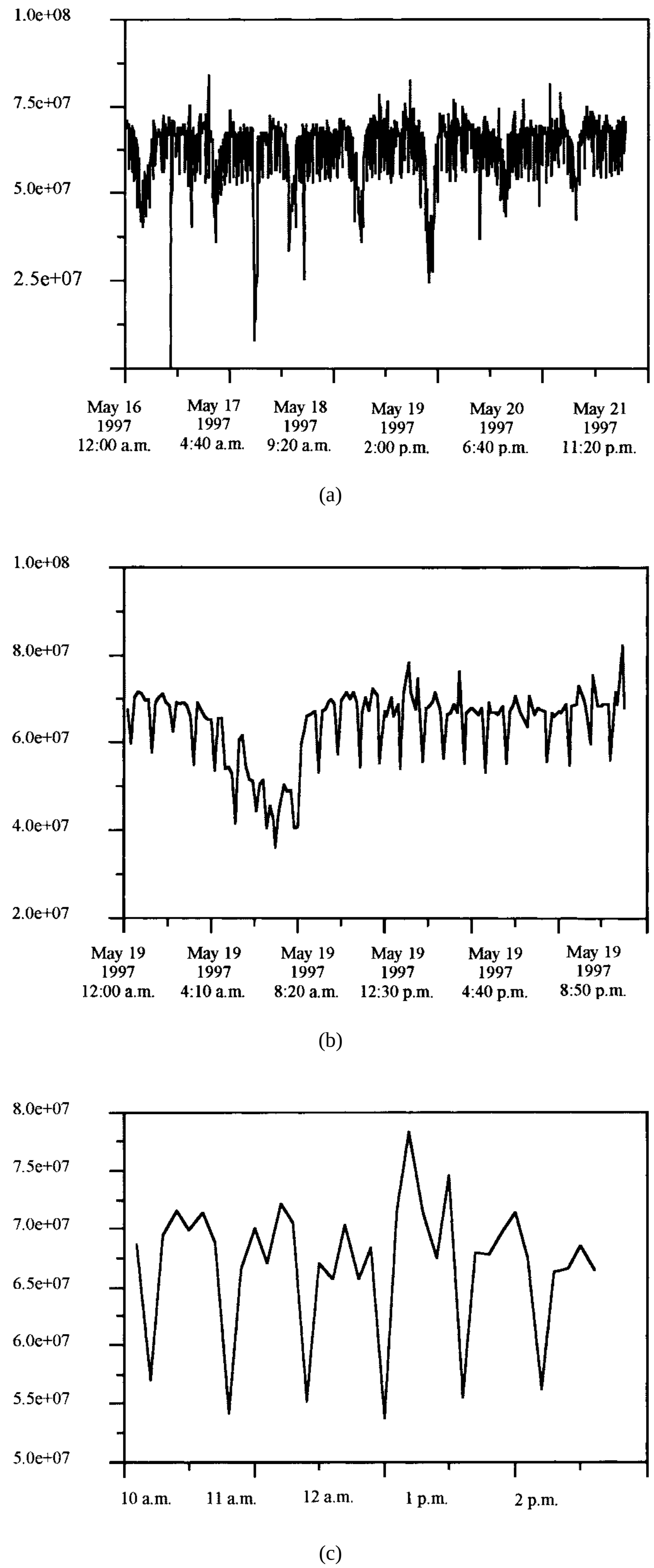

Fig. 16. Typical network utilization patterns (a) weekly, (b) daily, and (c) 5-h period ($x(t)$ denotes network load utilization in bits per second).

described in Sections III and IV. We start with the linguistic contexts defined in the output space. Here we restrict ourselves to seven triangular fuzzy numbers with an overlap halfway between two adjacent linguistic terms. The selection of these

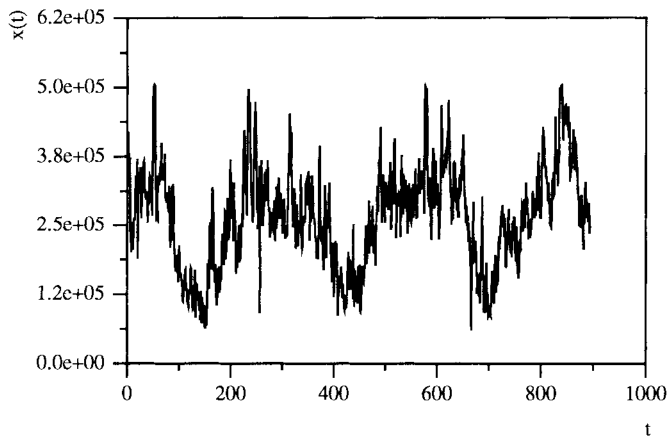

Fig. 17. Traffic data—a training set. (x(t) denotes network load utilization in bits per second).

contexts is rather subjective and relates to the use of the model. The parameters of the fuzzy sets of context are the following:

45000	50000	150000
50000	150 000	260 000
150 000	260 000	350 000
260 000	350 000	450 000
350 000	450 000	500 000
450 000	490 000	500 000
490 000	500 000	560 000

(as the membership functions are triangular, there is enough to use three numbers to fully describe their location (lower bound, modal value, and upper bound). For each context we performed clustering with $c = 5$ clusters. In total this has yielded $7 * 5 = 35$ induced clusters in the input space. The prototypes of these clusters are listed in Table I.

It is illuminating to visualize a distribution of the prototypes vis a vis the original data set. To make this visualization possible, this is done for two coordinates of the input data, that is $x(t-1)$ and $x(t)$, refer to Fig. 18.

The prototypes along with the contexts are enough to compute the output of the fuzzy neural network, refer to Fig. 5. The results are shown for the modal values contrasted with the experimental data (Fig. 19). The envelope of the model formed by the upper and lower bound of the triangular fuzzy number. Subsequently the output of the fuzzy model is shown in Fig. 20.

The series of plots in Fig. 21 summarizes a distribution of the data versus the results produced by the fuzzy model and represented in terms of their bounds and modal values. It is worth indicating that in most cases the bounds of the model embrace the original data—which is another nice feature of the fuzzy model.

Subsequently, the same model was used on a testing set composed of around 1000 data points (Fig. 22). Note also that the statistical characteristics of the testing set and training set are quite different as indicated in the respective histograms of the variable to be predicted (Fig. 23).

It is quite apparent that the fuzzy model performs well (Fig. 22). To gain a better insight into the performance of the fuzzy model, we have used a standard MSE index (note however, that this index has never been attempted to minimize) which, nevertheless, serves as a useful and highly informative quantitative descriptor of the model. For the training set we

TABLE I
PROTOTYPES OF THE CLUSTERS INDUCED BY THE LINGUISTIC CONTEXTS FOR THE TRAFFIC DATA

151717.9459	161898.9965	161259.1247	151957.8822	137246.0613
107030.7803	101524.5683	93672.0943	97694.4527	93809.3609
124394.0149	120695.7551	112958.0098	104627.9970	108917.3536
context 2:				
209402.8761	210490.5839	212237.0709	214956.7038	198803.8440
143882.7150	149172.0179	152120.3984	148061.6433	144293.5075
280711.7848	283760.2761	275224.5637	250317.8410	220103.9642
179350.3398	176315.8080	168003.7702	164357.5214	164667.3364
120894.2568	115092.5864	115847.6771	123296.3618	134085.5367
context 3:				
245927.4666	243954.3066	244481.5787	248824.3063	251342.0511
293057.0284	289639.9356	282441.7040	274804.9254	269688.6503
187321.4961	187130.7364	189944.1260	197797.0628	210329.1190
299974.3025	303311.7871	305646.0620	301977.9017	289296.2070
403497.5525	384986.6080	358574.1199	324443.9656	297007.7856
context 4:				
260787.3885	254832.8448	253496.8329	265678.9177	296181.2192
411983.9166	427630.1153	418029.5715	401366.0627	381413.1777
323385.2467	323798.1339	323304.3867	323549.5541	326201.8151
351921.1227	357850.3367	365316.1900	369932.6168	360995.9468
293887.5600	287863.1848	289286.3597	298779.5852	305724.3401
context 5:				
470910.5123	471860.8175	460911.4491	445446.4584	440490.1030
267827.5147	268545.9485	273432.4148	293635.7417	352964.9896
265970.1094	289065.4519	329082.9589	365644.2598	416123.0506
342844.4697	354176.8298	360826.0508	369075.8750	386865.0747
427563.2518	437324.4621	444738.2448	453079.5719	443370.5159
context 6:				
418076.1330	428235.6246	449453.1665	445277.1714	449688.8095
330691.8832	323636.2772	313905.2041	304340.6458	283497.0650
330673.1371	339028.5474	376465.6680	433262.3614	421945.3130
327539.7955	311461.8275	280158.7008	278102.7527	467687.0902
487514.8656	493676.4525	493992.7542	496372.0431	497335.1297
context 7:				
296716.6601	276525.6921	370310.2885	493601.7996	497240.1053
325646.5787	319622.9490	316182.7184	279090.4914	422976.3464
320734.7809	308994.9106	279385.0622	426139.9158	493639.7401
309088.8868	363426.1896	493062.7076	496955.1786	499252.6038
363951.1307	490333.5292	496001.9184	494630.5481	501849.1075

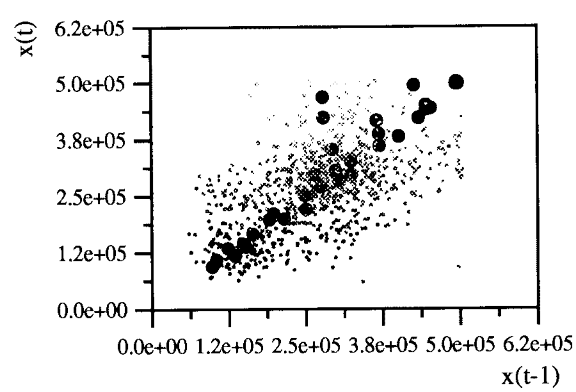

Fig. 18. Prototypes of the clusters versus data in $x(t-1)-x(t)$ coordinates for the traffic data.

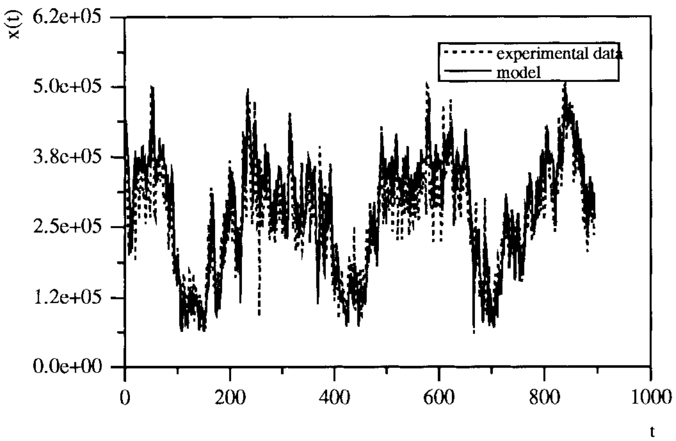

Fig. 19. Fuzzy model versus experimental data.

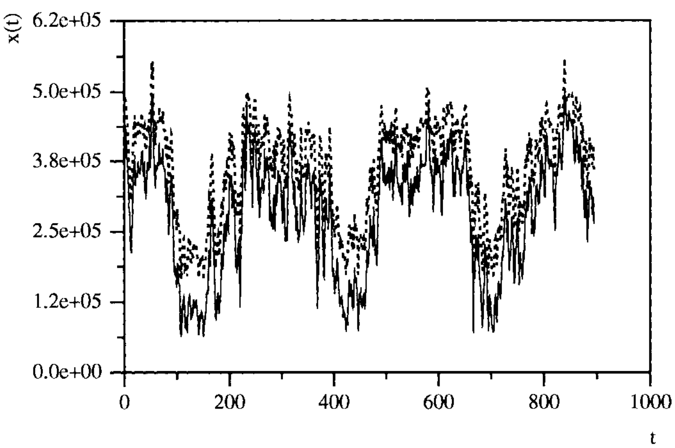

Fig. 20. Upper and lower bound of prediction realized by the fuzzy model.

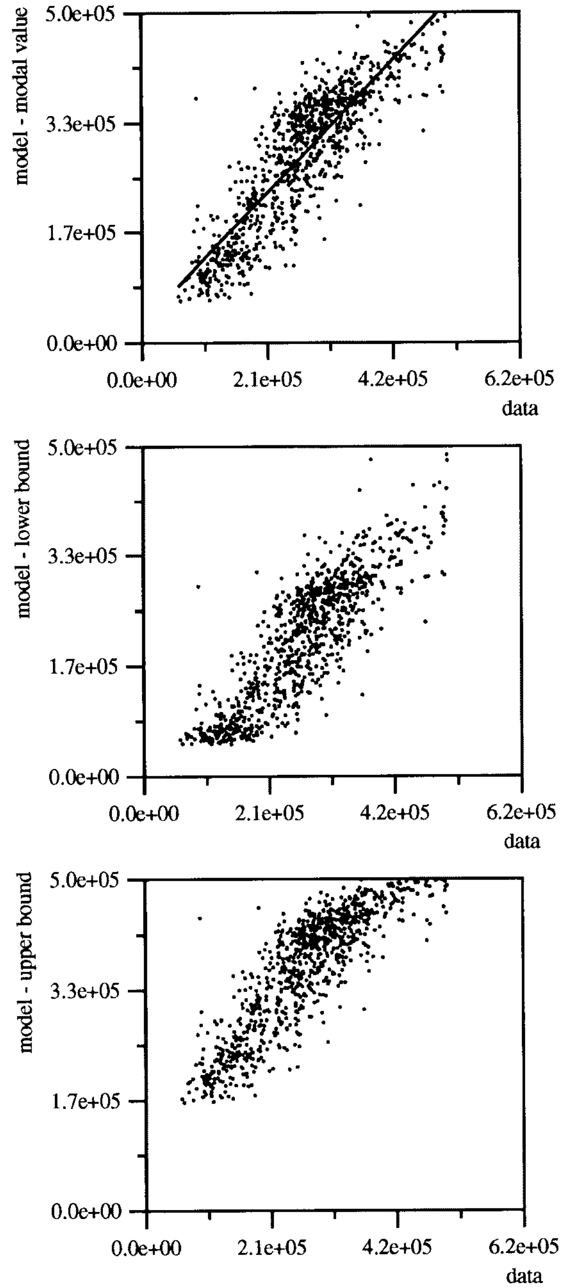

Fig. 21. Fuzzy model (bounds and modal values) versus experimental data.

obtained 54 771 while on the testing set this value was lower and equal 46 529.

VI. Conclusions

We have introduced a notion of linguistic (granular) modeling whose intent is to carry out system modeling at the level of information granules—fuzzy sets rather than numeric quantities. While still dwelling on the principles of fuzzy sets , the discussed methodology departures quite substantially from fuzzy modeling in several ways, namely

1) it makes modeling more designer-oriented;
2) it emphasizes the role of information granules as the generic modeling entities.

The assumed level of information granularity at which the modeling pursuits take place becomes crucial from two points of view:

1) Information granularity helps assume a proactive position with respect to the way in which modeling is going to take place. Especially, by modifying fuzzy sets and distributing them along the variables of interest one can easily affect the depth of modeling and emphasize its focal points.
2) The linguistic granules help develop some regularization effect that becomes of interest in presence of a noisy environment.

The concept of granular modeling comes fully supported by the detailed algorithm that hinges on the context-based fuzzy clustering.

From an application point of view, the presented experiments identify the proposed method as a highly promising technique for predicting both short term and long term performance of the network.

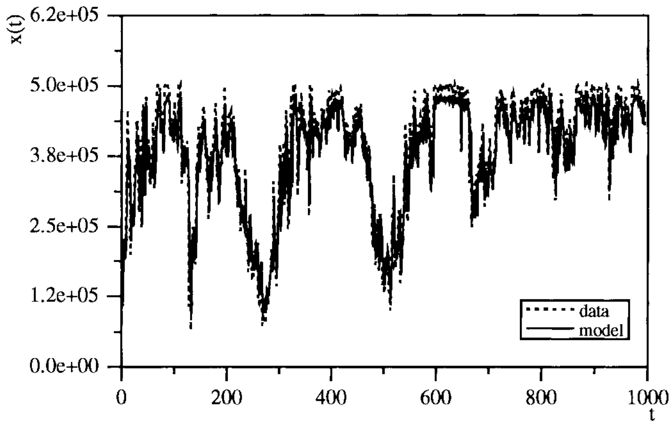

Fig. 22. Fuzzy model versus testing data.

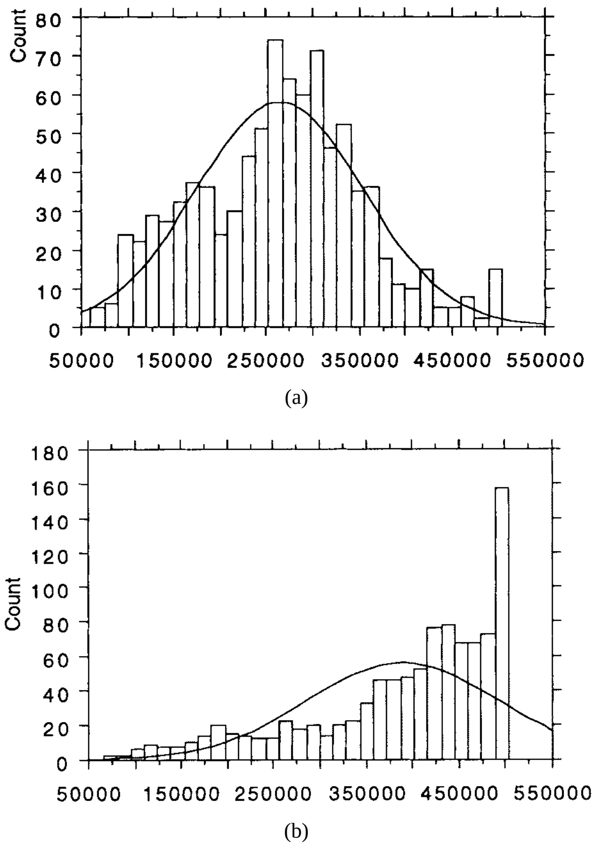

Fig. 23. Histograms of the predicted variable for the (a) training and (b) testing set.

The further optimization of the algorithm will allow to address the needs of the most interesting and promising applications, such as strategic routing [16] and help incorporate it into high speed switching and routing equipment.

APPENDIX

We are interested in two operations on triangular fuzzy numbers that is a multiplication of the number by a positive real number and an addition of two fuzzy numbers. A triangular fuzzy number A is defined by the following membership function

$$A(x;a,m,b) = \begin{cases} \dfrac{x-a}{m-a}, & \text{if } x \in [a,m] \\ \dfrac{b-x}{b-m}, & \text{if } x \in [m,b] \\ 0, & \text{otherwise.} \end{cases}$$

We denote a triangular fuzzy number A by $A = T(a,m,b)$ with $a, b,$ and m being its parameters (that is lower bound, modal value, and upper bound, respectively).

Multiplication of A by a Positive Constant z_0: Consider the increasing part of A and fix the membership value (ω). Then we get

$$\omega = \frac{x-a}{m-a}$$

or equivalently

$$x = \omega(m-a) + a.$$

Multiplying x by z_0 we get the corresponding argument of the membership function of the fuzzy set $A \otimes \{z_0\}$. Denote it by y

$$u = xz_0 = z_0\omega(m-a) + z_0 a.$$

Treat the above relationship as a function of $\omega, F(\omega)$. The membership function of $A \otimes \{z_0\}$ is equal to its inverse F^{-1} (more precisely, this concerns the increasing section of the resulting fuzzy number)

$$(A \otimes \{z_0\})(y) = F^{-1}(\omega)$$

that is

$$(A \otimes \{z_0\})(y) = \frac{y - z_0 a}{z_0 m - z_0 a}.$$

Similarly, we complete the computations for the decreasing part of the membership function. We start from the relationship

$$\omega = \frac{b-x}{b-m}.$$

Repeating the same computations as we did for the increasing part of A, the decreasing part of the membership function of the result is equal to

$$(A \otimes \{z_0\})(y) = \frac{bz_0 - y}{bz_0 - mz_0}.$$

Summarizing, the result of this multiplication (scaling) is a fuzzy number with the parameters of A multiplied by z_0

$$(A \otimes \{z_0\})(y) = T(z_0 a, z_0 m, z_0 b).$$

Addition of Two Triangular Fuzzy Numbers: For the addition of $A = T(a, m, b)$ and $B = T(c, n, d)$ we start with the extension principle [21]

$$C(z) = \sup_{x,u:z=x+y} [\min(A(x), B(y))]$$

and discuss the increasing and decreasing sections of the two membership functions separately. Consider first that $z < m+n$. There exists values x and y such $x < m$ and $y < n$ such that

$$A(x) = B(y) = \omega \ \omega \in [0, 1].$$

The linearly increasing sections of the membership functions of A and B give rise to the relationships

$$\frac{x-a}{m-a} = \omega$$

where

$$\frac{y-c}{n-c} = \omega.$$

Owing to the addition operation used for the arguments of $A \oplus B$, we get

$$z = x + y = a + c + (m + n - a - c)\omega.$$

Expressing the above as a function of ω one has

$$\omega = \frac{z - (a+c)}{(m+n) - (a+c)}$$

that holds for $z < m + n$. Similarly, when dealing with the decreasing sections of the membership functions of A and B we obtain

$$\omega = 1 - \frac{z - (m+n)}{(b+d) - (m+n)}.$$

Apparently, the result of addition $A \oplus B$ arises as a triangular fuzzy number with the following parameters

$$A \oplus B = T(a + c, m + n, b + d).$$

In other words, the parameters of the sum of two fuzzy numbers are produced as sums of the corresponding bounds of the arguments. The obtained results easily generalize to the same operations involving many arguments.

ACKNOWLEDGMENT

The authors would like to thank M. Petsagourakis (mp@forthnet.gr) and FORTHnet (http://www.forthnet.gr), for kindly providing us with real data to experiment with.

REFERENCES

[1] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York: Plenum, 1981.
[2] J. Buckley and Y. Hayashi, "Fuzzy neural networks: A survey," *Fuzzy Sets Syst.*, vol. 66, pp. 1–14, 1994.
[3] C. J. Harris, C. G. Moore, and M. Brown, *Intelligent Control—Aspects of Fuzzy Logic and Neural Nets*. Singapore: World Scientific, 1993.
[4] J. S. R. Jang, C. T. Sun, and E. Mizutani, *Neuro-Fuzzy and Soft Computing*. Englewood Cliffs, NJ: Prentice-Hall, 1997.
[5] A. Kandel, *Fuzzy Mathematical Techniques with Applications*. Reading, MA: Addison-Wesley, 1986.
[6] G. J. Klir and T. A. Folger, *Fuzzy Sets, Uncertainty, and Information*. Englewood Cliffs, NJ: Prentice Hall, 1988.
[7] R. Kruse, J. Gebhardt, and F. Klawonn, *Foundations of Fuzzy Systems*. New York: Wiley, 1994.
[8] V. Paxson, "Fast, approximate synthesis of fractional Gaussian noise for generating self similar network traffic," *ACM Comput. Commun. Rev.*, vol. 27, p. 5, Oct. 1997.
[9] W. Pedrycz, *Computational Intelligence: An Introduction*. Boca Raton, FL: CRC, 1997.
[10] ——, "Conditional fuzzy C-means," *Pattern Recognit. Lett.*, vol. 17, 1996, pp. 625–632.
[11] W. Pedrycz and F. Gomide, *An Introduction to Fuzzy Sets*. Cambridge, MA: MIT Press, 1998.
[12] M. Sugeno and T. Yasukawa, "A fuzzy logic based approach to qualitative modeling," *IEEE Trans. Fuzzy Syst.*, vol. 1, pp. 7–31, 1993.
[13] H. Takagi and I. Hayashi, "NN-driven fuzzy reasoning," *Int. J. Approx. Reas.*, vol. 5, pp. 191–212, 1991.
[14] *Fuzzy Systems Theory and Its Applications*, T. Terano, K. Asai, and M. Sugeno, Eds. New York: Academic, 1992.
[15] L. H. Tsoukalas and R. E. Uhrig, *Fuzzy and Neural Approaches in Engineering*. New York: Wiley, 1997.
[16] A. Vasilakos, C. Ricudis, K. Agnagnostakis, W. Pedrycz, and A. Pitsillides, "Evolutionary fuzzy prediction for strategic QoS routing in telecommunication networks," in *Proc. IEEE-FUZZ'98*, May 4–9, 1998, pp. 1488–1493.
[17] W. Willinger, M. Taqqu, W. Leland, and D. Wilson," Self-similarity in high-speed packet traffic analysis and modeling of Ethernet traffic measurements," *Statist. Sci.*, vol. 10, pp. 67–85, 1995.
[18] R. R. Yager and D. Filev, *Essentials of Fuzzy Modeling and Control*. New York: Wiley, 1994.
[19] L. A. Zadeh, "Toward a theory of fuzzy systems," in *Aspects of Network and System Theory*, R. E. Kalman and N. De Claris, Eds. New York: Holt, Rinehart & Winston, 1971.
[20] ——, "Fuzzy sets and information granularity," in *Advances in Fuzzy Set Theory and Applications*, M. M. Gupta, R. K. Ragade, and R. R. Yager, Eds. Amsterdam, The Netherlands: North-Holland, 1979, pp. 3–18.
[21] H. J. Zimmermann, *Fuzzy Set Theory and Its Applications*, 2nd ed. Norwell, MA: Kluwer, 1991.

Witold Pedrycz (M'88–SM'94–F'99), for a photograph and biography, see p. 415 of the June 1999 issue of this TRANSACTIONS.

Athanasios V. Vasilakos was born in Greece in 1959. He received the B.Sc. degree in electrical engineering in 1983 from the University of Thrace, Greece, the M.Sc. degree in electrical and computer engineering in 1986 from the University of Massachusetts, Amherst, and the Ph.D. degree in computer engineering in 1988 from the University of Patras, Patras, Greece.

From 1988 to 1991, he was with the Computer Engineering Department, University of Patras, and the Computer Technology Institute, Patras. From 1991 to 1995, he was Professor of informatics at the Hellenic Air Force Academy. Since 1995, he has been with the Institute of Computer Science, Foundation for Research and Technology Hellas, FORTH, Greece, and a Consultant to the Greek government. His main interests are communications networks, B-ISDN, ATM networks, mobile nets, learning theory, and computational intelligence. He is Chairman of the Technical Committee of Telecommunications, EURIDIT Network of Excellence for Fuzzy Logic in Europe. He is a member of the program committee of a number of conferences and an editor of the journal *Computer Communications*. He is a co-editor (with W. Pedrycz) of the book *Computational Intelligence in Telecommunications Networks* (to be published by CRC Press). He has published more than 70 papers, most of them on the applications of computational intelligence to computer networks.

Dr. Vasilakos is a member of the ACM and ACM SIGCOMM Societies.