Int. J. Appl. Math. Comput. Sci., 2002, Vol.12, No.3, 347–357

RELATIONS OF GRANULAR WORLDS

WITOLD PEDRYCZ[*], GEORGE VUKOVICH[**]

[*] Department of Electrical and Computer Engineering
University of Alberta, Edmonton, Canada
e-mail: pedrycz@ee.ualberta.ca

Systems Research Institute, Polish Academy of Sciences
01–447 Warsaw, Poland

[**] Canadian Space Agency, Spacecraft Engineering
6767 Route de l'Aeroport
Saint-Hubert, Quebec J3Y 8Y9, Canada

In this study, we are concerned with a two-objective development of information granules completed on a basis of numeric data. The first goal of this design concerns revealing and representing a structure in a data set. As such it is very much oriented towards coping with the underlying *relational* aspects of the experimental data. The second goal deals with a formation of a mapping between information granules constructed in two spaces (thus it concentrates on the *directional* aspect of information granulation). The quality of the mapping is directly affected by the information granules over which it operates, so in essence we are interested in the granules that not only reflect the data but also contribute to the performance of such a mapping. The optimization of information granules is realized through a collaboration occurring at the level of the data and the mapping between the data sets. The operational facet of the problem is cast in the realm of fuzzy clustering. As the standard techniques of fuzzy clustering (including a well-known approach of FCM) are aimed exclusively at the first objective identified above, we augment them in order to accomplish sound mapping properties between the granules. This leads to a generalized version of the FCM (and any other clustering technique for this matter). We propose a generalized version of the objective function that includes an additional collaboration component to make the formed information granules in rapport with the mapping requirements (that comes with a *directional* component captured by the information granules). The additive form of the objective function with a modifiable component of collaborative activities makes it possible to express a suitable level of collaboration and to avoid a phenomenon of potential competition in the case of incompatible structures and the associated mapping. The logic-based type of the mapping (that invokes the use of fuzzy relational equations) comes as a consequence of the logic framework of information granules. A complete optimization method is provided and illustrated with several numeral studies.

Keywords: granular computing, clustering, models of collaborative computing, fuzzy models, collaboration and competition, FCM, granular modeling

1. Introduction

Clustering and fuzzy clustering (Anderberg, 1973; Bezdek, 1981; Duda *et al.*, 2001; Dunn, 1973; Hoppner *et al.*, 1999; Pedrycz, 1995; 1998; Setnes, 2000) give rise to information granules (either sets, fuzzy sets or relations and fuzzy relations). These are formed through a minimization of a certain criterion (objective function) which operates within a given data set. As a result, the obtained clusters reflect the nature (structure) of these specific data. Consider a situation when we are provided with two separate data sets. For each of them, the clustering reveals the corresponding structure. In addition to that, we are interested in finding a mapping between the information granules developed in these data. Ideally, we would like to construct the granules in such a way that the mapping itself is also optimized, i.e., it transforms information granules defined in one space (domain) into some granules in another space (co-domain) without any significant distortion (mapping error). Evidently, this problem statement goes far beyond the standard clustering optimization as encountered in the literature. The essence of the problem is portrayed in Fig. 1.

Owing to the nature of the optimization task, we will be referring to it as a *collaborative* information granulation. The collaboration effect is crucial here in order to meet the two fundamental *relational* and *directional* requirements of the information granules. The satisfaction of these helps reflect the structure in the data as well as optimize the mapping itself (in the sense outlined above).

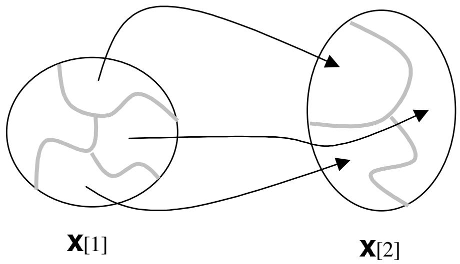


Fig. 1. The essence of collaborative information granulation; information granules capture the structure of data and simultaneously optimize the mapping between the granules in $\mathbf{X}[1]$ and $\mathbf{X}[2]$.

It is pertinent to cast the problem in the general setting of fuzzy modeling that is primarily aimed at the modeling activities carried out at the linguistic (granular) level. This paradigm of modeling entails two fundamental pursuits, that is, a development of information granules themselves (that are regarded as building blocks of the model) and a construction of a web of links between the information granules. It is needless to say that different families of information granules imply a different web of connections between them. And conversely, if we start with an initial structure of the connections, they may lead to further modifications of the information granules. The granularity of information being used implies the character of the model: a higher granularity (viz. more specific, that is, smaller granules) gives rise to more detailed models. By recognizing the two fundamental components of the fuzzy (granular) model, we become aware that these two phases are interrelated as far as their optimization is concerned. At the same time, the phases look at a different character of processing. Building information granules is inherently direction-free (namely, relational). The links (representing the mapping) are evidently directional constructs and, as such, they capture the function-like form of the model (obviously we mean the function between information granules, and not individual numeric values as traditionally encountered in numeric modeling and system identification). Building information granules can be realized through data clustering and the methods commonly available today are direction-free (say, direction-neutral). As these two phases of building granular models are strongly linked, it becomes evident that the clustering techniques need to be revisited to allow for coping with the directional character of the constructs and help optimize these features. This augmentation of the clustering techniques is an ultimate goal of this study.

In this paper, we propose a new scheme of building information granules in the collaborative fashion to capture the relational and directional aspects of information granulation. We formulate the problem first and then cast it in an operational framework of fuzzy clustering, especially the commonly used FCM environment. It is discussed how the objective function has to be modified and what type of constraint optimization problem arises there. The type of the granular mapping to be optimized is in the form of the logic-based expression using OR and AND logic connectives.

The material is arranged into seven sections. In Section 2, we start with a problem formulation where its main components are outlined and put together. The derivation of the complete optimization algorithm is discussed in Section 3 that is followed by a discussion of the flow of the optimization pursuits. Experimental studies are covered in Section 5 while conclusions are included in Section 6.

2. Problem Formulation

The collaborative way of information granulation leads to the following optimization problem. To formulate it, let us proceed with all necessary notation. To a high extent, we will be alluding to the standard terminology used in fuzzy clustering. Given are two data sets $\mathbf{X}[1]$ and $\mathbf{X}[2]$, where $\mathbf{X}[1] = \{\mathbf{x}_1[1], \mathbf{x}_2[1], \cdots, \mathbf{x}_N[1]\}$ and $\mathbf{X}[2] = \{\mathbf{x}_1[2], \mathbf{x}_2[2], \ldots, \mathbf{x}_N[2]\}$ for $\mathbf{x}_k[1] \in \mathbb{R}^{n_1}$ and $\mathbf{x}_k[2] \in \mathbb{R}^{n_2}$. The clustering of $\mathbf{X}[1]$ involves $c[1]$ clusters. For the second data set, the clustering gives rise to $c[2]$ clusters. The clusters constructed in $\mathbf{X}[1]$ are denoted by $A_1, A_2, \ldots, A_{c[1]}$. For $\mathbf{X}[2]$ the resulting clusters are $B_1, B_2, \ldots, B_{c[2]}$. In what follows, the terms 'cluster' and 'information granule' are used interchangeably. Furthermore, as we are dealing with two data sets, all constructs pertaining to the data are indexed by square brackets, namely (Anderberg, 1973; Bezdek, 1981).

The mapping between the granules in $\mathbf{X}[1]$ and $\mathbf{X}[2]$ exhibits an evident logic flavor meaning that we assume a logic form of the relationship between the information granules, namely, we express an information granule B_i as a logic aggregation (ϕ) involving the information granules developed in $\mathbf{X}[1]$, i.e.,

$$B_i = \phi(A_1, A_2, \ldots, A_{c[1]}, \mathbf{w}_i)$$

for $i = 1, 2, \ldots, c[2]$. The above expression also includes a weight vector (parameters) $\mathbf{w}_i$ that is used to calibrate the links between A_j and B_i. We discuss the details of the logic expression in the next section. More descriptively, we are interested in the development of information granules $\{A_i\}$ and $\{B_j\}$ so that they satisfy the requirements of relational and directional natures.

2.1. The Objective Function and Its Generalization

The entire optimization starts from the objective function defined in such a way that it encapsulates the way of the collaborative formation of the information granules. For illustrative purposes, we start with a specific form of the clustering method. This form can be generalized further on.

The FCM algorithm is commonly used. This motivates us to view it as a standard vehicle of forming information granules. The generic objective function that is minimized reads as a weighted sum of distances of the patterns from the prototypes. Using the notation already introduced in the previous section, the performance index (objective function) assumes the following form:

- For $\mathbf{X}[1]$ the clusters minimize the expression

$$Q[1] = \sum_{i=1}^{c[1]} \sum_{k=1}^{N} u_{ik}^2[1] d_{ik}^2[1], \tag{1}$$

where

$$d_{ik}^2[1] = \left\| \mathbf{x}_k[1] - \mathbf{v}_i[1] \right\|^2$$

is a distance function between the pattern (i.e., $\mathbf{x}_k[1]$) and the prototype (denoted here by $\mathbf{v}_i[1]$) with both of them being located in $\mathbf{X}[1]$. While the above objective function is well known in the literature, we rewrite it slightly to highlight the origin and the role of the fuzzy sets. Note that each row of the partition matrix is just an information granule (more specifically a fuzzy relation) so we can view $U[1]$ in the following form:

$$U[1] = \left[A1\ A2\ \ldots\ A_{c[1]} \right]^T = \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_{c[1]} \end{bmatrix}.$$

Note that each A_i is defined in the finite space $\mathbf{X}[1]$ (which implies that A_i has a discrete membership function).

In other words, the performance index can be sought as a weighted sum of the distances between the prototypes and the data points with the weights being the membership grades of the fuzzy relations,

$$Q[1] = \sum_{i=1}^{c[1]} A_i^2 \bullet D_i, \tag{2}$$

where

$$A_i^2 \bullet D_i = \sum_{k=1}^{N} u_{ik}^2 d_{ik}^2. \tag{3}$$

In more detail, we have

$$A_i^2 = \left[u_{i1}^2\ u_{i2}^2\ \ldots\ u_{ic[1]}^2 \right].$$

The minimization of (1), or equivalently (2), is completed with respect to the partition matrix and the prototypes.

- For $\mathbf{X}[2]$ the clusters built there are affected by the collaboration coming from $\mathbf{X}[1]$ as we are concerned about the mapping (logic transformation between $\mathbf{X}[1]$ and $\mathbf{X}[2]$). This is taken into consideration by expanding the objective function in the following additive form:

$$Q[2] = \sum_{i=1}^{c[2]} \sum_{k=1}^{N} u_{ik}^2[2] d_{ik}^2[2] + \beta \sum_{i=1}^{c[2]} \sum_{k=1}^{N} \big(u_{ik}[2] - \phi_i(U[1]) \big)^2 d_{ik}^2[2]. \tag{4}$$

The first term is just a standard component encountered in the FCM that looks like the structure in $\mathbf{X}[2]$. The second one captures differences between $U[2]$ and the mapping of the structure detected in $\mathbf{X}[1]$ (viz. the fuzzy partition $U[1]$) to $\mathbf{X}[2]$, that is $\phi_i(U[1])$. It characterizes the performance of the mapping between the information granules. The weight coefficient (β) is used to quantify a balance between the structure in $\mathbf{X}[2]$ and the impact from the mapping requirement (the second term in the above objective function). Considering the two goals of this process of information granulation, we say that β strikes a compromise between the relational and directional aspects of such an optimization.

The optimization of (1) is standard. The optimization of (4) requires detailed investigation. The minimization of the objective function $Q[2]$ is completed with respect to the partition matrix $U[2]$ (structure), prototypes and the parameters of the logic transformation (ϕ_i).

2.2. The Logic Transformation

The granular mapping from $\mathbf{X}[1]$ to $\mathbf{X}[2]$ is realized as a logic transformation between the information granules. It is worth stressing that there is a panoply of possible types of mappings and our choice is implied by the transparency of the logic mapping that comes hand in hand with the logic type of the spaces between which the mapping takes place. Two classes of mappings are discussed. The first of them is *OR-based*. As the name stipulates, we consider the information granule B_i to be an OR aggregation of the granules in the input space, i.e.,

$$B_i = A_1 \text{ or } A_2 \text{ or } \ldots \text{ or } A_{c[1]}. \tag{5}$$

Neither all fuzzy relations in the input space contribute to the formation of B_i, nor may each of them exhibit an equal impact on the membership of B_i. To gain this flexibility, we allow for a weight vector (connections) $\mathbf{w}_i$

whose role is to articulate (quantify) the contribution coming from A_j's. The following modification is made to (5):

$$B_i = (A_1 \text{ and } w_{i1}) \text{ or } (A_2 \text{ and } w_{i2}) \text{ or } \dots \text{ or } (A_{c[1]} \text{ and } w_{ic[1]}), \quad (6)$$

where $\mathbf{w}_i = [w_{i1}\ w_{i2}\ \dots\ w_{ic[1]}]$ are weights with the values confined to the unit interval. The logic operations are realized by t- and s- norms (t-conorms) and this leads to the equivalent expression of (6) (s-t realization of the logic expression):

$$B_i = (A_1\ t\ w_{i1})\ s\ (A_2\ t\ w_{i2})\ s\ \dots\ s\ (A_{c[1]}\ t\ w_{ic[1]})$$

and

$$B_i = \mathop{S}_{j=1}^{c[1]} (A_j\ t\ w_{ij}). \quad (7)$$

This augmented expression requires more attention. To provide more insight into the functioning of (6), let us consider several cases: (a) if all weights are equal to 1, then, owing to the boundary condition of any t-norm, the expression $A_k\ t\ w_{ik}$ returns A_i, (b) if the weights are all equal to zero, then B_i is an empty fuzzy set, (c) if all but one connection are equal to zero (the remaining nonzero connection w_k is set to zero), then B_i is equal to A_k. We note that the weight vector is used to select fuzzy sets A_i and calibrate their impact on the formation of the fuzzy set in the output space. We can regard it as a collection of the parameters that could be adjusted and used in the optimization of the logic mapping.

Note that the above logic mapping concerns a single fuzzy relation in $\mathbf{X}[2]$. In a similar fashion, we can realize the mapping for the remaining information granules. After a careful examination of the mappings viewed together, we come up with the following concise notation. Arrange all weights into the matrix form

$$R = [r_{ij}] = \begin{bmatrix} w_{11} & \dots & & w_{1c[2]} \\ \dots & & & \\ & & w_{lj} & \\ w_{c[1]1} & \dots & & w_{c[1]c[2]} \end{bmatrix}. \quad (8)$$

Then the mapping from information granules in $\mathbf{X}[1]$ to information granules in $\mathbf{X}[2]$ is nothing but a fuzzy relational equation with a standard s-t composition (Di Nola *et al.*, 1989; Pedrycz, 1991; Pedrycz and Rocha, 1993; Pedrycz *et al.*, 1995) (denoted here by a small dot)

$$\mathbf{B} = A \circ R, \quad (9)$$

where we set $A = [A_1\ \ A2\ \ \dots\ \ A_{c[1]}]$ and $B = [B_1\ B_2\ \dots\ B_{c[2]}]$.

Similarly, we can introduce an AND type of aggregation of the information granules meaning that we consider B_i to be a combination of A_j's aggregated AND-wise, i.e.,

$$B_i = A_1 \text{ and } A_2 \text{ and } \dots \text{ and } A_{c[1]}. \quad (10)$$

A direct generalization of the above aggregation includes a weight vector and, subsequently, the combination of the t-s type

$$B_i = (A_1 \text{ or } w_{i1}) \text{ and } (A_2 \text{ or } w_{i2}) \text{ and } \dots \text{ and } (A_{c[1]} \text{ or } w_{ic[1]}), \quad (11)$$

$$B_i = (A_1\ s\ w_{i1})\ t\ (A_2\ s\ w_{i2})\ t\ \dots\ t\ (A_{c[1]}\ s\ w_{ic[1]}) = \mathop{T}_{j=1}^{c[1]} (A_j s w_{ij}). \quad (12)$$

The motivation behind the use of the connections ($\mathbf{w}$) is the same as in the case of the previous type of the logic aggregation. The intent is to equip the mapping with some parametric flexibility. Essentially, the connections help calibrate an impact that a given input fuzzy set (or relation) A_i has on the output. Owing to the nature of the s-norm, we note that higher values of w_i reduce the impact that A_i comes with. In two limit cases we have: (a) $w_i = 0$ returns an original input fuzzy set in the sense that $A_i\ s\ 0 = A_i$, (b) $w_i = 1$ eliminates (masks) the associated fuzzy set, namely $A_i\ s\ 1 = 1$.

These two aggregation mechanisms are dual in the sense of their functionality. Owing to the character of the AND and OR operations, we intuitively use them depending on the number of the information granules existing in the respective spaces. If $c[1] > c[2]$, we consider the OR type of aggregation (in anticipation that the element in the output space is constructed as a union of several information granules in the input space). Similarly, for $c[1] < c[2]$, the AND-type of aggregation is more appealing (as we project that B_i is made more specific in relation to the information granules existing in $\mathbf{X}[1]$).

3. The Algorithm

Now we are ready to proceed with the computational details that lead us to the complete algorithm. The objective function implies the following optimization task:

$$\min_{U[2], v_1[1], v_2[2], \dots, v_{c[2]}]} Q[2]$$

subject to

$$U[2] \in \mathbf{U} \text{ and } R \in \mathbf{R}, \quad (13)$$

where the family of partition matrices $\mathbf{U}$ is defined in a usual manner (namely, we require that the elements in each column of $U[2]$ sum up to 1 and the sum of elements in each row of R is nonzero and less than N). R is an element of the family of the fuzzy relations $\mathbf{R}$ (viz. matrices with elements confined to the unit interval). The above optimization problem concerns the way of forming a structure in $\mathbf{X}[2]$ with an inclusion of the mapping properties. The clustering mechanisms in $\mathbf{X}[1]$ follow the standard FCM and will not be discussed here.

The optimization of the partition matrix $U[2]$ in the objective function uses a technique of Lagrange multipliers (because of the constraint existing in the development of the partition matrix). For a given data point (k), we form an augmented objective function

$$V = \sum_{i=1}^{c[2]} u_{ik}^2[2] d_{ik}^2[2] + \beta \sum_{i=1}^{c[2]} \big(u_{ik}[2] - \varphi_i(U[1])\big)^2 d_{ik}^2[2] + \lambda\Big(\sum_{i=1}^{c[2]} u_{ik}[2] - 1\Big), \tag{14}$$

where λ is a Lagrange multiplier. Proceeding with the necessary conditions for the minimum of V:

$$\frac{\partial V}{\partial u_{st}[2]} = 0, \quad \frac{\partial V}{\partial \lambda} = 0,$$

we calculate

$$2u_{st}[2]d_{st}^2[2] + 2\beta\big(u_{st}[2] - y_{st}\big)d_{st}^2[2] + \lambda = 0. \tag{15}$$

Here y_{st} stands for a logic-based mapping between the information granules

$$y_{st} = \mathop{S}_{j=1}^{c[1]} \big(u_{st}[1]\ t\ r_{sj}\big).$$

Computing $u_{st}[2]$ from (15), we obtain

$$u_{st}[2] = \frac{\beta y_{st} d_{st}^2[2] - \lambda}{2d_{st}^2[2](1+\beta)}.$$

As the membership grades sum up to 1, this leads us to the expression

$$\sum_{j=1}^{c[2]} \frac{\beta y_{jt} d_{jt}^2[2] - \lambda}{2d_{jt}^2[2](1+\beta)} = 1,$$

and in the sequel,

$$\frac{\lambda}{1+\beta} = \frac{-1 + \dfrac{\beta}{1+\beta}\displaystyle\sum_{j=1}^{c[2]} y_{jt}}{\displaystyle\sum_{j=1}^{c[2]} \frac{1}{d_{jt}^2[2]}}.$$

Introduce the notation

$$\tilde{u}_{st}[2] = \frac{1}{\displaystyle\sum_{j=1}^{c[2]} \frac{d_{st}^2[2]}{d_{jt}^2[2]}}.$$

Finally, we get

$$u_{st}[2] = \tilde{u}_{st}[2] + \frac{\beta}{1+\beta}\Big(y_{st} - \tilde{u}_{st}[2]\sum_{j=1}^{c[2]} y_{jt}\Big), \tag{16}$$

for $s = 1, 2, \dots, c[2]$ and $t = 1, 2, \dots, N$.

The above formula has an interesting interpretation: if β is equal to zero, then it reduces to the well-known formula for the partition matrix encountered in the FCM. When β increases, then $u_{st}[2]$ is affected by the second term in (16).

The calculations of the prototypes do not come with any constraints, so we follow the necessary condition for the minimum of $Q[2]$, namely $\partial Q[2]/\partial v_{st}[2] = 0$, $s = 1, 2, \dots, c[2]$ and $t = 1, 2, \dots, n_2$.

In light of the weighted Euclidean distance governed by the expression

$$d_{ik}^2[2] = \sum_{j=1}^{n_2} \frac{\big(x_k[j] - v_{ij}[2]\big)^2}{\sigma_j^2[2]} \tag{17}$$

(where $\sigma_j^2[2]$ denotes the variance of the j-th variable), the above derivative is equal to

$$\frac{\partial Q[2]}{\partial v_{st}[2]} = 2\beta \sum_{k=1}^{N} u_{sk}^2[2] \frac{(x_k[t] - v_{st}[2])}{\sigma_t^2[2]} - 2\beta \sum_{k=1}^{N} \psi_{sk} \frac{(x_k[t] - v_{st}[2])}{\sigma_t^2[2]} \tag{18}$$

with the following notation:

$$\psi_{sk} = \big(u_{sk}[2] - y_{sk}\big)^2.$$

Bearing in mind the necessary condition for the minimum of $Q[2]$ with respect to the prototypes, they are equal to

$$v_{st}[2] = \frac{\displaystyle\sum_{k=1}^{N} x_k[t]\big(u_{sk}^2 + \beta\psi_{sk}\big)}{\displaystyle\sum_{k=1}^{N} \big(u_{sk}^2 + \beta\psi_{sk}\big)}. \tag{19}$$

Noticeably, when $\beta = 0$, we arrive at the standard expression for the prototypes that is identical to the one encountered in the FCM method.

Finally, we optimize the fuzzy relation R describing the logic mapping between the spaces. In general, the solution is not expressed analytically and we have to proceed with some iterative optimization. The underlying expression governing this optimization reads as

$$R(\text{iter}+1) = R(\text{iter}) - \beta\nabla_{R(\text{iter})} Q, \tag{20}$$

where the fuzzy relation is transformed on the basis of the gradient of the performance index Q. The learning rate shown above ($\beta > 0$) controls the pace of changes of the updates of the fuzzy relation. The gradient itself is computed for specific triangular norms. In what follows (and all experiments shown in Section 5 will exploit these assumptions) we consider two common models of the logic connectives such as a product (t-norms) and probabilistic sum (s-norms). On this basis, the gradient reads as follows:

$$\frac{\partial y_{sk}}{\partial r_{st}} = (1 - A_{st}) u_{tk}[1], \qquad (21)$$

where A_{st} denotes an s-t composition that excludes the currently optimized element of the fuzzy relation

$$A_{st} = \mathop{S}_{\substack{j=1\\ j\neq t}}^{c[1]} \big(u_{jk}[1] \; t \; r_{sj}\big). \qquad (22)$$

4. The Overall Development Framework: A Flow of Optimization Activities

The way in which the information granules are built stipulates a certain flow of optimization activities. These can be grouped into two main phases, see Fig. 2. The initial phase concentrates on the clustering completed independently for the two data sets $\mathbf{X}[1]$ and $\mathbf{X}[2]$. The intent here is to establish some preliminary structure in the data so that we could have a reasonable starting point to proceed with the collaboration and further refine the initial relationships. During the second phase, the clustering processes start to collaborate through the mapping. At the same time the fuzzy relation is subject to the gradient-based optimization (as illustrated in Fig. 2, this as an integral portion of the collaboration process and negotiation of the granular structures). Because of the direction of the mapping, the clustering in $\mathbf{X}[1]$ is not affected per se while the relational and directional facets of the clusters emerge at the side of $\mathbf{X}[2]$.

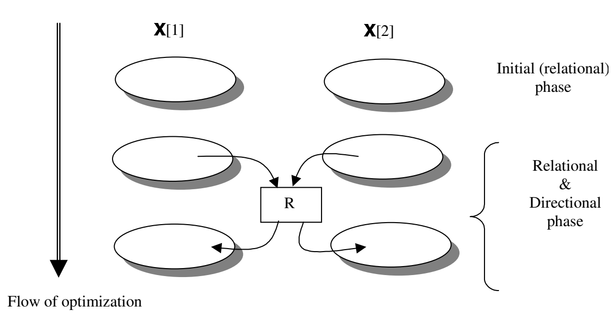


Fig. 2. Relational and directional optimization of information granules—an overall development scheme.

5. Experiments

The proposed algorithmic framework is illustrated by means of numeric experiments. They include some synthetic as well as real-world data available on the WWW.

Synthetic data. The experiment concerns three-dimensional data. The first data set includes input variables (x_1 and x_2) while the second involves one-dimensional data (y), see Table 1.

Table 1. Synthetic data used in the experiment.

x_1	1.2	0.8	0.2	0.9	3.5	4.2	4.3	4.8	6.1	6.5	6.9	6.6	6.4	6.1
x_2	1.8	1.5	1.6	1.2	3.9	3.4	4.7	4.1	7.0	6.2	6.4	5.7	5.8	5.7
y	1.5	1.1	1.4	0.9	3.5	3.2	2.9	3.4	3.6	2.7	2.5	2.8	3.7	3.9

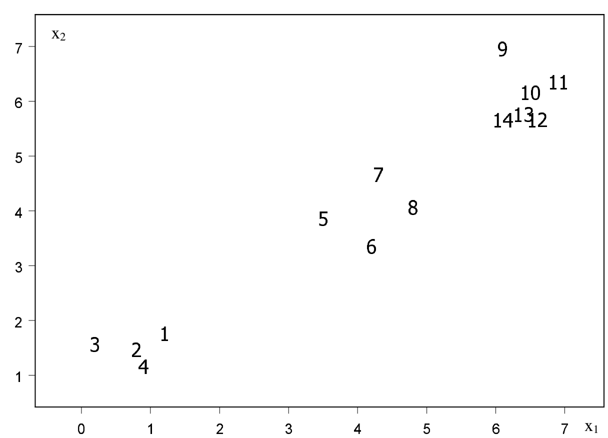


Fig. 3. Plots of the synthetic data in the input space (x_1, x_2) and the output data (y).

The same data points are shown in a 3D space in Fig. 4. This helps reveal the structure. The output variable comes with two clearly visible clusters. Moreover, the three clusters in the input space relate to the two clusters in the output space. In more detail, we note that the two clusters in the input space $\mathbf{X}[1]$ map on a single cluster in $\mathbf{X}[2]$.

Following these observations (which are easy to arrive at as we are dealing with low-dimensional synthetic data), we set up $c[1] = 3$ and $c[2] = 2$. In the experiment, the learning rate of the gradient-based learning is equal to 0.1. This relatively low value of the learning rate helps avoid oscillations (and this is more crucial to us than an eventual slowdown of the learning process itself). Without any collaboration ($\beta = 0.0$) the obtained clusters are described in terms of the partition matrices:

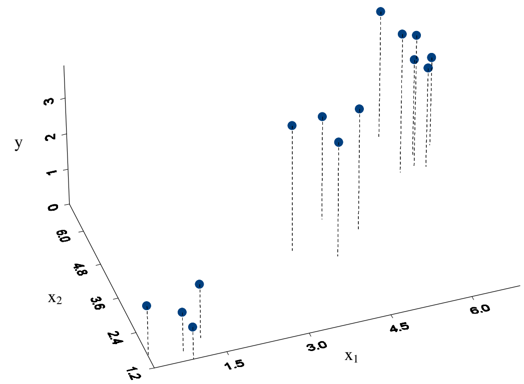


Fig. 4. 3D plot of the synthetic data.

Partition—space of input variables:

$$U[1] = \begin{bmatrix} 0.017434 & 0.005164 & 0.977402 \\ 0.000068 & 0.000022 & 0.999910 \\ 0.013630 & 0.004868 & 0.981502 \\ 0.007460 & 0.002543 & 0.989997 \\ 0.938615 & 0.030536 & 0.030849 \\ 0.936182 & 0.032897 & 0.030921 \\ 0.894294 & 0.082808 & 0.022897 \\ 0.946818 & 0.039856 & 0.013326 \\ 0.069128 & 0.914942 & 0.015930 \\ 0.001119 & 0.998669 & 0.000212 \\ 0.020588 & 0.974971 & 0.004441 \\ 0.027533 & 0.967885 & 0.004582 \\ 0.015457 & 0.982032 & 0.002511 \\ 0.045319 & 0.948161 & 0.006520 \end{bmatrix}.$$

Partition—output variable:

$$U[2] = \begin{bmatrix} 0.982731 & 0.017269 \\ 0.994259 & 0.005740 \\ 0.994833 & 0.005167 \\ 0.976867 & 0.023133 \\ 0.010273 & 0.989727 \\ 0.001394 & 0.998606 \\ 0.049299 & 0.950701 \\ 0.003565 & 0.996435 \\ 0.019316 & 0.980684 \\ 0.137239 & 0.862761 \\ 0.281143 & 0.718857 \\ 0.086491 & 0.913509 \\ 0.029929 & 0.970070 \\ 0.053702 & 0.946298 \end{bmatrix}.$$

Subsequently, the prototypes are equal to

- input space: $v_1[1] = [4.204584\ \ 4.015867]$, $v_2[1] = [6.439176\ \ 6.117275]$, $v3[1] = [0.777533\ \ 1.524844]$,
- output space $v_1[2] = 1.265061,\ v2[2] = 3.272303$.

The clusters emerging in both the spaces are very well delineated with a very limited overlap. The collaborative clustering is carried out for several levels of collaboration (β). Following the general scheme (Section 4), we implement the collaboration after the initial clustering of the individual data. Here we go for five iterations. The performance index achieved throughout the clustering and learning of the relations is shown in Fig. 5. (Note that the term “cycle” used there concerns the performance index recorded for a single clustering iteration and 20 learning epochs of the gradient-based learning.) The optimization is efficient as the values of the performance index are reduced from cycle to cycle.

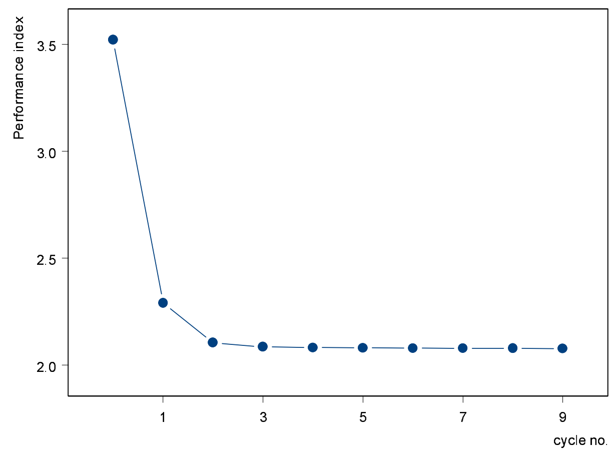


Fig. 5. Performance index in successive development cycles ($\beta = 0.1,\ \alpha = 0.1$).

Once the optimization has been completed, the fuzzy partition in the output space is as presented below:

$$U[2] = \begin{bmatrix} 0.980664 & 0.019336 \\ 0.994456 & 0.005544 \\ 0.993309 & 0.006691 \\ 0.976525 & 0.023475 \\ 0.010258 & 0.989742 \\ 0.001312 & 0.998688 \\ 0.045444 & 0.954556 \\ 0.003853 & 0.996147 \\ 0.023591 & 0.976409 \\ 0.128656 & 0.871345 \\ 0.259470 & 0.740530 \\ 0.082898 & 0.917102 \\ 0.033662 & 0.966339 \\ 0.055437 & 0.944563 \end{bmatrix}.$$

We do not report the results of clustering in the input space as in this model these fuzzy sets have not been affected. When comparing this partition matrix with the one obtained for the clustering without any collaboration, we conclude that there are no substantial differences. Obviously, the collaboration effect is quite limited and this may be a reason behind an evident coincidence in the information granules (conveyed in the respective partition matrices). The prototypes do not change when the collaboration effect comes into play at this level (namely, for $\beta = 0.1$). At this level of collaboration, there are no changes in the values of the resulting partition matrix.

What becomes of interest is a fuzzy relation revealing main relationships between the information granules (fuzzy sets) in the input and output spaces,

$$R = \begin{bmatrix} 0.000000 & 0.063547 & 0.738940 \\ 0.710587 & 0.889359 & 0.005974 \end{bmatrix}.$$

There is a strong dependence (relationship) between the granules quantified by high membership grades. Denoting the fuzzy sets by A_1, A_2 and A_3 (input space), and B_1 and B_2 (output space), we translate the above fuzzy relation into two logic expressions

$$B_1 = A_3(0.73),$$
$$B_2 = A_1(0.71) \text{ or } A_2(0.89).$$

(Note that we have included only the terms with high levels of association; the associations themselves are simply the corresponding entries of the fuzzy relation.)

Now we increase the collaboration level to 0.4. This results in the partition matrix whose entries start to diverge from those without any collaboration. Figure 6 illustrates these new membership grades of the partition matrices.

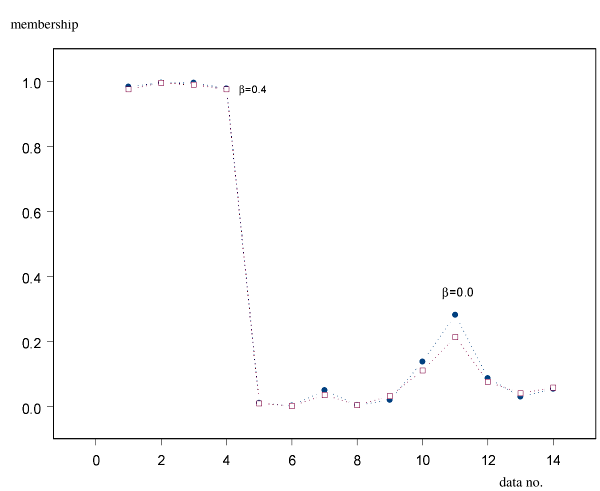


Fig. 6. Changes in membership grades as a result of collaboration.

The differences between the prototypes are still negligible as they are equal to $v_1[2] = 1.25503$ and $v_2[2] = 3.255817$, respectively. The fuzzy relation of the associations is now equal to

$$R = \begin{bmatrix} 0.000000 & 0.060559 & 0.947464 \\ 1.000000 & 0.908921 & 0.005793 \end{bmatrix}.$$

Subsequently, the list of logical expressions is similar to the one obtained before

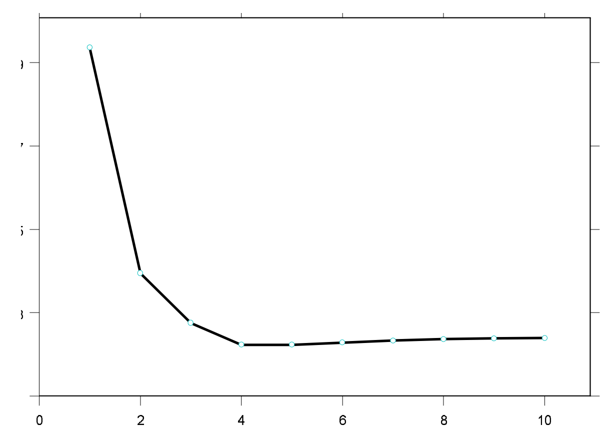
$$B_1 = A_3(0.95),$$
$$B_2 = A_1(1.00) \text{ or } A_2(0.91).$$

However, now the strength of the associations between the information granules has been elevated.

Noticeably, higher values of β may lead to an instability as the mechanisms used in the method tend to "compete." This is visible in Fig. 7: for higher β, the performance index exhibits some tendency to oscillate. These tend to become more visible once the structures begin to rely on each other more significantly (β increases). The lack of stability points out that we are now faced with a sort of competition between the structures as they do not collaborate any longer but tend to compete.

The *auto-mpg dataset* comes from the UCI repository of machine learning (`http://www.ics.uci.edu/~mlearn/MLSummary.html`) and concerns a collection of vehicles described in terms of their displacement, weight, country of origin, etc. We consider all features but the fuel efficiency (expressed in miles per gallon) as inputs. The fuel efficiency is treated as the output variable.

The clustering is carried out for different numbers of the clusters in the input and output space. The level of collaboration (β) is maximized as much as the stability is retained. The results are summarized in the form of the



$\beta = 0.5$

Fig. 7. Performance index $Q[2]$ in successive cycles of optimization for two selected values of β.

fuzzy relations (with the most essential links being highlighted) and the prototypes in the input and output spaces. Table 2 contains a sample of the findings. Noticeably, there are no significant changes to the information granules. The granules in the input space start to become more specific once their number increases.

Table 2. Results of collaboration between clusters in the input and output spaces for a selected number of the clusters in the input space, $c[1] = 4, 5$, and 7 and $c[2] = 2$.

$\beta = 0.1$

$$R = 0.1 \begin{bmatrix} 0.000000 & \mathbf{1.000000} & \mathbf{0.623695} & 0.000000 \\ \mathbf{0.853638} & 0.000000 & 0.000000 & 0.359311 \end{bmatrix}$$

prototypes

input space

$$\mathbf{v}_1[1] = [4.079673 \;\; 103.923645 \;\; 76.339424 \;\; 2219.613770 \;\; 16.554533 \;\; 77.514397 \;\; 2.661081]$$

$$\mathbf{v}_2[1] = [7.937488 \;\; 347.367065 \;\; 160.888031 \;\; 4164.818848 \;\; 12.664634 \;\; 73.340729 \;\; 1.011035]$$

$$\mathbf{v}_3[1] = [5.876760 \;\; 218.628098 \;\; 99.783264 \;\; 3196.265625 \;\; 16.621513 \;\; 75.807190 \;\; 1.114213]$$

$$\mathbf{v}_4[1] = [4.225328 \;\; 127.185654 \;\; 83.869286 \;\; 2475.715088 \;\; 16.138067 \;\; 77.310410 \;\; 1.423118]$$

output space

$$v_1[2] = 17.535294, \quad v_2[2] = 31.191629$$

$\beta = 0.15$

$$R = \begin{bmatrix} \mathbf{1.000000} & 0.000000 & \mathbf{0.509261} & 0.000000 & 0.020741 \\ \mathbf{0.628279} & \mathbf{1.000000} & 0.305735 & \mathbf{0.521905} & 0.353167 \end{bmatrix}$$

prototypes

input space

$$\mathbf{v}_1[1] = [4.076893 \;\; 102.475128 \;\; 73.396721 \;\; 2188.470947 \;\; 16.731644 \;\; 78.909592 \;\; 2.786299]$$

$$\mathbf{v}_2[1] = [7.960478 \;\; 350.072906 \;\; 162.005402 \;\; 4189.602539 \;\; 12.605497 \;\; 73.222733 \;\; 1.006485]$$

$$\mathbf{v}_3[1] = [4.274063 \;\; 137.074478 \;\; 85.290085 \;\; 2578.334717 \;\; 16.201082 \;\; 79.018456 \;\; 1.178627]$$

$$\mathbf{v}_4[1] = [6.068239 \;\; 230.507248 \;\; 102.261139 \;\; 3290.023682 \;\; 16.560471 \;\; 75.631432 \;\; 1.064134]$$

$$\mathbf{v}_5[1] = [4.156690 \;\; 113.290359 \;\; 84.537315 \;\; 2359.725342 \;\; 15.943533 \;\; 74.026688 \;\; 2.031541]$$

output space

$$v_1[2] = 31.105927, \quad v_2[2] = 17.784975$$

$\beta = 0.2$

$$R = \begin{bmatrix} 0.209803 & 0.261901 & 0.032446 & 0.000000 & 0.010196 & \mathbf{0.751416} & \mathbf{1.000000} \\ 0.224672 & \mathbf{1.000000} & \mathbf{1.000000} & \mathbf{0.623729} & 0.181158 & 0.135858 & 0.318719 \end{bmatrix}$$

prototypes

input space

$$\mathbf{v}_1[1] = [4.091851 \;\; 108.241699 \;\; 83.893974 \;\; 2309.296143 \;\; 15.864615 \;\; 73.814117 \;\; 2.213534]$$

$$\mathbf{v}_2[1] = [7.968541 \;\; 362.461090 \;\; 168.896362 \;\; 4264.135742 \;\; 12.252587 \;\; 72.357430 \;\; 1.005682]$$

$$\mathbf{v}_3[1] = [7.791941 \;\; 309.141937 \;\; 138.760895 \;\; 3885.156006 \;\; 13.943682 \;\; 76.404419 \;\; 1.025403]$$

$$\mathbf{v}_4[1] = [5.951785 \;\; 226.265488 \;\; 99.462135 \;\; 3241.927002 \;\; 16.694593 \;\; 75.186249 \;\; 1.057520]$$

$$\mathbf{v}_5[1] = [4.338152 \;\; 130.680313 \;\; 84.464478 \;\; 2507.565674 \;\; 16.512613 \;\; 76.256737 \;\; 1.500398]$$

$$\mathbf{v}_6[1] = [4.217593 \;\; 135.321823 \;\; 84.600761 \;\; 2569.331543 \;\; 16.173941 \;\; 79.529884 \;\; 1.176432]$$

$$\mathbf{v}_7[1] = [4.043318 \;\; 99.765396 \;\; 71.445312 \;\; 2154.744873 \;\; 16.819048 \;\; 79.410835 \;\; 2.868308]$$

output space

$$v_1[2] = 31.248861, \quad v_2[2] = 17.654987$$

It is interesting to note that the collaboration can be made more vigorous without sacrificing stability when the number of the clusters in the input space increases. This could have been expected, as the resulting information granules tend to be smaller (of a higher granularity) and therefore could be moved around more freely while not causing too much distortion (and hence instabilities) during the collaboration process. The dependencies between the information granules as expressed by the fuzzy relations discriminate quite well between strong and weak links. In other words, the fuzzy relations start to contain values either close to 0 or 1. This points out that some information granules relate to one another very strongly.

Now let us consider the same number of the clusters in both the spaces, cf. Table 3. This arrangement helps us reveal how the granules relate in the two spaces. Because of the same number of the fuzzy sets, the logic formula may be of the form of a one-to-one mapping, namely a mapping of a single information granule in the input space to some other information granule in the output space. Obviously, this happens at the level of information granules rather than numeric quantities. By considering the entries of the fuzzy relations, this observation about one-to-one is fully legitimate. In each row of the fuzzy relation, we have only one dominant membership grade (indicated in boldface in Table 3). Noticeably, these are not necessarily high membership values. This is, however, justified as the partition matrices start having lower entries once the number of the clusters goes up (recall that these membership grades have to add up to 1).

Table 3. Fuzzy relations of connections for $c[1] = c[2] = 2, 3$ and 6 with $\beta = 0.05$.

$$R = \begin{bmatrix} 0.000000 & \mathbf{0.949054} \\ \mathbf{0.420400} & 0.000000 \end{bmatrix}$$

$$R = \begin{bmatrix} 0.110320 & 0.025040 & \mathbf{0.247469} \\ 0.000000 & \mathbf{0.877086} & 0.000000 \\ \mathbf{0.210435} & 0.000000 & 0.000000 \end{bmatrix}$$

$$R = \begin{bmatrix} \mathbf{0.090541} & 0.003166 & 0.000000 & 0.000788 & 0.000000 & 0.003596 \\ 0.004518 & 0.012588 & 0.076320 & 0.017965 & \mathbf{0.106748} & 0.006710 \\ 0.000000 & 0.051439 & \mathbf{0.156852} & 0.109466 & 0.000000 & 0.000000 \\ \mathbf{0.160533} & 0.000564 & 0.000000 & 0.000000 & 0.000000 & 0.037163 \\ 0.031722 & 0.004176 & 0.000000 & 0.003358 & \mathbf{0.117081} & 0.075881 \\ 0.000000 & \mathbf{0.706356} & 0.000000 & 0.000000 & 0.000000 & 0.000000 \end{bmatrix}$$

The graph of links between the information granules for $c = 2$ and 3 is included in Fig. 8 (we show only the most dominant connections).

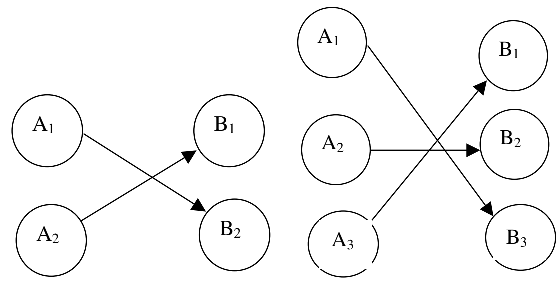


Fig. 8. A graphical illustration of linkages between information granules in $\mathbf{X}[1]$ and $\mathbf{X}[2]$ (the most significant connections included).

6. Conclusions

In this study, we raised an issue of designing information granules (fuzzy sets or fuzzy relations) that takes into consideration a structure in a data set and addresses the mapping aspects occurring at the level of such information granules.

The essential novelty of this approach resides with this multifaceted aspect of information granulation. Fuzzy clustering itself occurs after the structure of the data and does not look into the nature of possible mappings. It is evident that no matter which technique we study, fuzzy clustering tackles a *relational* nature of data (so no direction when searching for a structure is taken into consideration). The augmented objective function includes an additional collaboration component to make the information granules in rapport with the mapping requirements (that comes with a *directional* component). The additive form of the objective function with a modifiable component of collaborative activities makes it possible to model the level of collaboration and avoid the phenomenon of potential competition in the case of incompatible structures and the associated mapping. It is worth stressing that the granular mapping developed in this study can be viewed as a general development phase of fuzzy modeling and in this sense it applies to a broad class of identification problems in granular systems.

The logic-based type of the mapping (that invokes fuzzy relational equations) comes as a consequence of the logic framework of information granules. One can, however, apply other types of mappings including those implemented via neural networks. This generalizes the approach and promotes it as a general model of collaborative granular computing.

The collaborative scheme of information granulation is also in line with a broad range of techniques of fuzzy modeling, cf. (Delgado *et al.*, 1997; 1998; Jang, 1993; Kandel, 1986; Kowalczyk and Bui, ; Ma *et al.*, 2000; Pedrycz, 1995; Pedrycz and Vasilakos, 1999; Sugeno and Yasukawa, 1993; Zhang and Kandel, 1998). The agenda of fuzzy modeling is to build transparent models operating at the level of information granules viewed as fuzzy sets or fuzzy relations. At the operational end, the construction of the tangible and semantically sound information granules becomes crucial to the integrity of any fuzzy model. Furthermore, the agenda of fuzzy modeling includes an important aspect of the accuracy of the model. This calls for the adjustment of the fuzzy sets quite often to the point where their modifications tend to be difficult to justify from the standpoint of the interpretability of the model. The point is in the construction of the information granules so that they maintain their semantics and contribute to the accuracy of the model. In essence, there are two requirements to meet. The collaborative form of the fuzzy clustering is a suitable framework of fuzzy modeling in which these two needs are fulfilled in unison.

Acknowledgements

The support from the Natural Sciences and Engineering Research Council of Canada (NSERC) and ASERC (Alberta Software Engineering Research Consortium) is gratefully acknowledged.

References

Anderberg M.R. (1973): *Cluster Analysis for Applications*. — New York: Academic Press.

Bezdek J.C. (1981): *Pattern Recognition with Fuzzy Objective Function Algorithms*. — New York: Plenum Press.

Di Nola A., Sessa S., Pedrycz W. and Sanchez E. (1989): *Fuzzy Relational Equations and Their Applications in Knowledge Engineering*. — Dordrecht: Kluwer.

Delgado M., Gomez-Skarmeta F. and Martin F. (1997): *A fuzzy clustering-based prototyping for fuzzy rule-based modeling*. — IEEE Trans. Fuzzy Syst., Vol. 5, No. 2, pp. 223–233.

Delgado M., Gomez-Skarmeta A.F. and Martin F. (1998): *A methodology to model fuzzy systems using fuzzy clustering in a rapid-prototyping approach*. — Fuzzy Sets Syst., Vol. 97, No. 3, pp. 287–302.

Duda R.O., Hart P.E. and Stork D.G. (2001): *Pattern Classification, 2-nd Ed.* — New York: Wiley.

Dunn J.C. (1973): *A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters*. — J. Cybern., Vol. 3, No. 3, pp. 32–57.

Hoppner F., Klawonn F., Kruse R. and Runkle T. (1999): *Fuzzy Cluster Analysis*. — Chichester: Wiley.

Jang J.-S.R. (1993): *ANFIS: Adaptive-network-based fuzzy inference system*. — IEEE Trans. Syst. Man Cybern., Vol. 23, No. 3, pp. 665–685.

Kandel A. (1986): *Fuzzy Mathematical Techniques with Applications*. — Reading: Addison-Wesley.

Kowalczyk R. and Bui V. (2000): *On constraint-based reasoning in e-negotiation agents*, In: Agent Mediated Electronic Commerce (F. Didnum and U. Cortes, Eds.). — Berlin: Springer-Verlag, pp. 31–46.

Ma M., Zhang Y.-Q., Langholz G. and Kandel A. (2000): *On direct construction of fuzzy systems*. — Fuzzy Sets Syst., Vol. 112, No. 1, pp. 165–171.

Pedrycz W. (1991): *Neurocomputations in relational systems*. — IEEE Trans. Pattern Anal. Mach. Intell., Vol. 13, No. 3, pp. 289–296.

Pedrycz W. and Rocha A. (1993): *Knowledge-based neural networks*. — IEEE Trans. Fuzzy Syst., Vol. 1, No. 4, pp. 254–266.

Pedrycz W., Lam P. and Rocha A.F. (1995): *Distributed fuzzy modelling*. — IEEE Trans. Syst. Man Cybern., Vol. 1, No. 4, pp. 769–780.

Pedrycz W. (1995): *Fuzzy Sets Engineering*. — Boca Raton, FL: CRC Press.

Pedrycz W. (1998): *Conditional fuzzy clustering in the design of radial basis function neural networks*. — IEEE Trans. Neural Netw., Vol. 9, No. 4, pp. 601–612.

Pedrycz W. and Vasilakos A.V. (1999): *Linguistic models and linguistic modeling*. — IEEE Trans. Syst. Man Cybern., Vol. 29, No. 6, pp. 745–757.

Setnes M. (2000): *Supervised fuzzy clustering for rule extraction*. — IEEE Trans. Fuzzy Syst., Vol. 8, No. 4, pp. 416–424.

Sugeno M. and Yasukawa T. (1993): *A fuzzy-logic-based approach to qualitative modeling*. — IEEE Trans. Fuzzy Syst., Vol. 1, No. 1, pp. 7–31.

Zhang Y.-Q. and Kandel A. (1998): *Compensatory Genetic Fuzzy Neural Networks and Their Applications*. — Singapore: World Scientific.