ACKNOWLEDGMENT

The authors wish to thank Y.-T. Yang for her help in programming our experiments and anonymous reviewers for their valuable comments.

REFERENCES

[1] G. H. John, R. Kohavi, and K. Pfleger, "Irrelevant features and the subset selection problem," in *Machine Learning: Proceedings of the Eleventh International Conference (ICML '94)*. San Mateo, CA: Morgan Kaufmann, 1994.

[2] R. Kohavi and D. Sommerfield, "Feature subset selection using the wrapper method: Overfitting and dynamic search space topology," in *Proceedings of the First International Conference on Knowledge Discovery and Data Mining (KDD '95)*. Menlo Park, CA: AAAI Press, 1995.

[3] R. Carunan and D. Freitag, "Greedy attribute selection," in *Machine Learning: Proceedings of the Eleventh International Conference*. San Mateo, CA: Morgan Kaufmann, 1994.

[4] H. Liu and R. Setiono, "Feature selection and classification—A probabilistic wrapper approach," in *Proc. 9th Int. Conf. Industrial and Engineering Applications of AI and ES*, 1996.

[5] D. Schuschel, "Automation of attribute selection for neural nets," M. S. Thesis, Dept. Computer Science and Engineering, Arizona State Univ., Tempe, 1998.

[6] C. J. Merz and P. M. Murphy. UCI repository of machine learning databases. Dept. Computer Science, Univ. California, Irvine. [Online]. Available: http://www.ics.uci.edu/~mlearn/MLRepository.html.

[7] J. T. McClave and F. H. Dietrich, *Statistics*. San Francisco, CA: Dellen, 1991.

[8] M. Richeldi and P. L. Lanzi, "Performing effective feature selection by investigating the deep structure of the data," in *Proceedings of the Second International Conference on Knowlwdge Discovery and Data Mining (KDD '96)*, E. Simoudis, J. Han, and U. M. Fayyad, Eds. Menlo Park, CA: AAAI Press, 1996.

[9] R. Setiono and H. Liu, "Neural network feature selector," *IEEE Trans. Neural Networks*, vol. 8, pp. 654–662, May 1997.

[10] J. Yang and V. Honavar, "Feature subset selection using a genetic algorithm," in *Feature Extraction, Construction, and Subset Selection: A Data Mining Perspective*, H. Motoda and H. Liu, Eds. Norwell, MA: Kluwer, 1998, ch. 8.

[11] J. Yang, R. Parekh, and V. Honavar, "DistAL: An inter-pattern distance-based constructive learning algorithm," *Intell. Data Anal.*, vol. 3, pp. 55–73, 1999.

[12] M. Dash and H. Liu, "Hybrid search of feature subsets," in *Proc. PRICAI*, Singapore, Nov. 1998.

[13] H. Liu and R. Setiono, "A probabilistic approach to feature selection—A filter solution," in *Proceedings of the 13th International Conference on Machine Learning (ICML'96)* Bari, Italy, 1996, pp. 319–327.

[14] D. Koller and M. Sahami, "Toward optimal feature selection," in *Machine Learning: Proc. of the Thirteenth International Conference (ICML '96)*, L. Saitta, Ed. San Mateo, CA: Morgan Kaufmann, 1996.

[15] S. Downing and D. Socie, "Simple rainflow counting algorithms," *Int. J. Fatigue*, Jan. 1982.

[16] *Standard Practices for Cycle Counting in Fatigue Analysis*, Americam Society for Testing and Materials, ASTM E ASTM 1049-85, 1985.

Granular Clustering: A Granular Signature of Data

Witold Pedrycz and Andrzej Bargiela

***Abstract*—The study is devoted to a granular analysis of data. We develop a new clustering algorithm that organizes findings about data in the form of a collection of information granules—hyperboxes. The clustering carried out here is an example of a granulation mechanism. We discuss a compatibility measure guiding a construction (growth) of the clusters and explain a rationale behind their development. The clustering promotes a data mining way of problem solving by emphasizing the transparency of the results (hyperboxes). We discuss a number of indexes describing hyperboxes and expressing relationships between such information granules. It is also shown how the resulting family of the information granules is a concise descriptor of the structure of the data—a granular *signature* of the data. We examine the properties of features (variables) occurring of the problem as they manifest in the setting of the information granules. Numerical experiments are carried out based on two-dimensional (2-D) synthetic data as well as multivariable Boston data available on the WWW.**

***Index Terms*—Complex systems, confidence limits analysis, data mining, feature analysis, granular time series, hyperboxes, information abstraction, information granules and granulation, interval analysis, principle of balanced information granularity.**

I. INTRODUCTORY COMMENTS

Making sense of data has been a motto of data mining. Any in-depth analysis of data that leads to comprehensive and interpretable results has to address an issue of transparency of final findings. In one way or another, arises a need for casting the results in the language of information granules—conceptual entities that capture the essence of the overall data set in a compact manner. It is worth stressing that information granules are a vehicle of abstraction that supports a conversion of clouds of numeric data into more tangible information granules [2], [3], [5], [12], [13], [16]–[18].

The area of clustering with its long history has been an important endeavor of finding structures in data and representing the essence of such finding in terms of prototypes, dendrograms, self-organizing maps [8], [9] and alike [1], [4]. Commonly, if not exclusively, the direct aspect of granulation has not been tackled. The intent of this study is to address this important problem by introducing an idea of granular clustering. Being more descriptive, the simplest scenario looks like this: we start from collection of numeric data (points in $\mathbf{R}^n$) and form information granules whose distribution and size reflects the essence of the data. Forming the clusters (information granules) may be treated as a process of *growing* information granules—as the clustering progresses, we expand the clusters, enhance the descriptive facet of the granules while gradually reduce the amount of details being available to us. The information granules we are interested in this study are represented as hyperboxes positioned in a highly dimensional data space. The mathematical formalism of the interval analysis provides a robust framework for the analysis of information density of the granular structures that

Manuscript received February 3, 2001; revised November 6, 2001. This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC), Alberta Consortium of Software Engineering (ASERC) and by the Engineering and Physical Sciences Research Council (U.K.). This paper was recommended by Associate Editor P. K. Bhattacharya.

W. Pedrycz is with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, Canada and also with the Systems Research Institute, Polish Academy of Sciences 01-447 Warsaw, Poland (e-mail: pedrycz@ee.ualberta.ca).

A. Bargiela is with the Department of Computing, The Nottingham Trent University, Nottingham NG1 4BU, U.K. (e-mail: andre@doc.ntu.ac.uk).

Publisher Item Identifier S 1083-4419(02)00703-3.

emerge in the process of clustering. The study reflects the intuitive objective of matching the granularity of data items used to describe the physical systems to the structure of these systems. In this sense the granulation process is attempting to achieve the highest possible generalization while maintaining the specificity of data structures.

This paper is organized into seven sections. In Section II, we start off with an introduction to information granules and spell out a rationale behind information granulation. As we are concerned with information granulation carried out in terms of sets (hyperboxes), this formalism is supported with all pertinent notation. The principle of granular clustering is covered in Section III that is followed by a complete algorithm (Section IV). Data analysis completed in the framework of information granules is studied in Section V. Finally, experimental studies are included in Section VI.

II. Information Granules and Information Granulation

Most experimental data available in a raw form are numeric. Granulation of information happens through a process of data organization and data comprehension. Interestingly; humans granulate information almost in a subconscious manner. This eventually makes the ensuing cognitive processes so effective and far superior over processes occurring under the auspices of machine intelligence. Two representative categories of problems in which information granulation emerges in a profound way involve processing of one-dimensional (1-D) and two-dimensional (2-D) signals. The first case, we are concerned primarily with temporal signals. The latter case pertains to image processing and image analysis. In signal processing, analysis and interpretation the granules arise as a result of temporal sampling and aggregation. Several samples in the same time window can be represented as an information granule. In the simplest case, such interval can be constructed by taking a minimal and maximal value of the signal occurring in this window of granulation; refer to Fig. 1. Some other ways of forming information granules may rely on statistical analysis: one determines a mean or median as a representative of the numeric data points and then build a confidence interval around it (obviously, the use of this mechanism requires assumptions about the statistical properties of the population contained in the window as well as the numeric representative under discussion). Similarly, in image processing one combines pixels exhibiting some spatial neighborhood. Again, various features of an image can be granulated, say brightness, texture, RGB, etc.

Information granulation has been studied in [2], [3], [10], [12] both in terms of the concept itself, computational aspects of it as well as resulting structures.

A. Set-Based Framework of Information Granules: The Language of Hyperboxes

In the overall presentation we adhere to a standard notation. A hyperbox defined in $\mathbf{R}^n$ is denoted by B and is fully described by its lower ($\mathbf{l}$) and upper corner ($\mathbf{u}$). To use explicit notation, we use $B(\mathbf{l}, \mathbf{u})$ where $\mathbf{l}, \mathbf{u} \in \mathbf{R}^n$ and obviously a strict inclusion relationship holds true $\mathbf{l} \leq \mathbf{u}$. If $\mathbf{l} = \mathbf{u}$ the box reduces to a single point (numeric datum) $B(\mathbf{l}, \mathbf{l}) = \{\mathbf{l}\}$. Hyperboxes are elements of a family of relations defined in $\mathbf{R}^n$ More specifically we state that $B \in \mathcal{P}(\mathbf{R}^n)$ with $\mathcal{P}(.)$ being a class of sets. The volume of B, denoted by $V(B)$, is viewed as a measure of specificity of the information granule. The point, $B(\mathbf{l}, \mathbf{l})$ comes with the highest specificity that becomes reduced once the volume increases. Computationally, it is advantageous to consider the expression $\exp(-V)$ which captures the same aspect of granularity yet this measure is normalized as it attains 1 for the numeric datum and reduces to zero once the hyperbox starts growing.

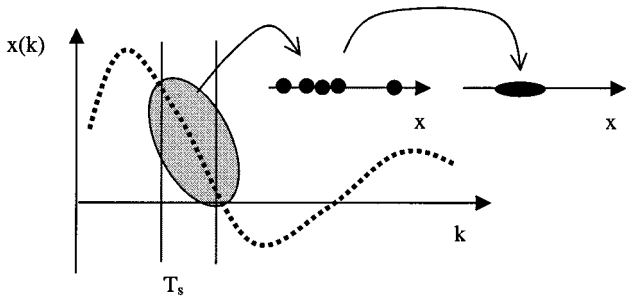

Fig. 1. Fragment of a time series and its granulation trough sampling (T_s denotes a sampling interval).

III. The Principle of Granular Clustering

Before we proceed with the details of the clustering technique for granular data, it is instructive to discuss the underlying principle, learn how the process proceeds and concentrate on the interpretation of some results generated by the proposed clustering mechanism.

As emphasized in the literature [1], [4], the essence of clustering (unsupervised learning) is to discover a structure in data. In essence, almost all existing clustering techniques operate on numeric objects (vectors in $\mathbf{R}^n$) and produce representatives (say, prototypes) that are again entirely numeric. In this sense, their form does not reflect how much data points they represent and how the distribution of these data points looks like (obviously, the nature of data is captured by a pertinent allocation of the prototypes). In the design of the clustering method, we add an extra dimension of granularity that helps sense the structure in the data as it becomes unveiled during the formation of the clusters.

A. The Design

This approach introduced here is very much different in many ways from the others. The leitmotiv is the following

— abstraction (no matter whether dealing with numeric or granular elements) is achieved through *condensation* of original data elements into granules whose location and granularity reflects the essence of the structure of data. The more condensation, the larger the sizes of the information granules that realize this aggregation.

At the qualitative end, the granular clustering is carried out as the following iterative process:

- find the two closest information granules (where the idea of compatibility guiding this search of information will be quantified later on) and on this basis build a new granule embracing them. In this way, one condenses the data while reducing the size of the data set
- repeat the first step until enough data condensation has been accomplished (here one has to come up with a certain termination criterion or introduce a sound validation mechanism).

Fig. 2 illustrates how the clustering works. We start from a collection of small information granules (these are original data) and start growing larger information granules. Noticeably, through their growth they tend reflecting the essential characteristics of the original data. The size of the granules reflects quite evidently how much they incorporated the original data and convey an extra message about their dispersion (distribution).

Considering a way in which the data points are merged together this approach resembles techniques of aggregative hierarchical clustering. There is a striking difference though: in hierarchical clustering we deal with numeric objects and the clusters are sets of the same objects. While a structure is revealed, no conceptually new entities are formed. Here,

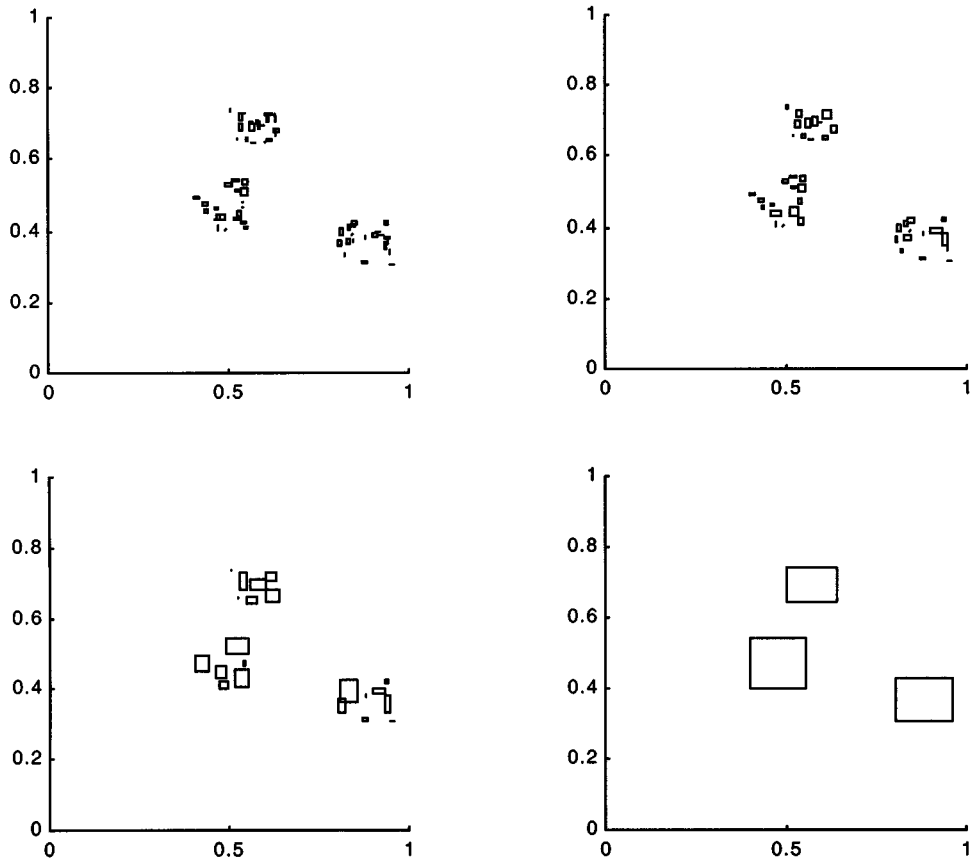

Fig. 2. Several snapshots of cluster growing over the clustering process; observe that small information granules forming at the initial stage (first iteration) that are grouped in some well-confined regions and give rise to three apparent large information granules at the later stage of clustering.

we "grow" the clusters: from iteration to iteration they tend to form larger hyperboxes. Moreover the nature of these hyperboxes help monitor the clustering process more thoroughly and raise awareness about terminating the clustering. Essentially, once we have found that the evolved boxes become distant in the state space, the process of clustering (forming combined boxes) is terminated.

By the same token, this concept should be contrasted with the idea of min–max clustering discussed by Simpson [14], [15] as this technique seems to bear some resemblance with the method studied here. The similarity is superficial though. First, the Simpson's method deals with point-size data while we consider data that is represented by either points or hyperboxes in pattern space. Second, the fuzzy membership functions of the information granules (clusters) as proposed by Simpson promote formation of clusters that are having largely varying sizes in various dimensions which is exactly the opposite to what we are trying to promote through the "compatibility measure" (discussed in Section IV). To emphasize the latter point we present in Fig. 3 a representative of a class of membership functions proposed by Simpson and refer the reader to Fig. 9 for comparison with the functions that have been utilized in our clustering algorithm.

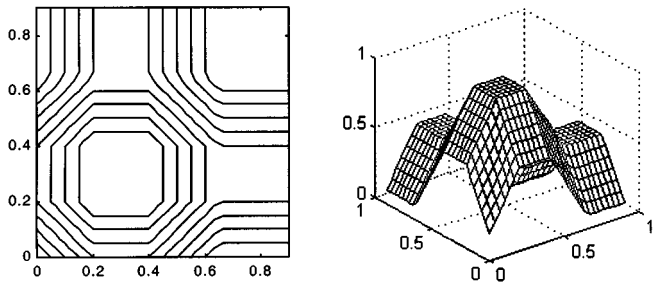

Fig. 3. Simpson's membership function (as presented in [15]) for the hyperbox defined by the min point $V = [0.2\ \ 0.2]$ and max point $W = [0.4\ \ 0.4]$. Sensitivity parameter $\gamma = 4$.

B. Interpretation and Validation of Granular Clustering

Clustering comes with a significant number of cluster validity indexes whose role is to decide upon a certain number of clusters. They help navigate the clustering process by stating what number of clusters should be. Commonly, their behavior does not lead to clear conclusions. What could be even worse, they may generate conflicting suggestions as to the termination condition (that is the number of clusters).

In granular clustering, we take another position. As the clusters capture the core of the data (and obviously, this is regarded as an important benefit of the method), our conjecture is that such core should help establish a sound platform of assessment of the structure (granular clusters).

When progressing with an expansion of the information granules, a certain criterion worth investigating deals with measuring a volume of the smallest granule ($V_{\min}$) that is constructed at this particular step (more specifically, we determine $e^{-V_{\min}}$; the details will be covered in Section IV-A). The main point is that if such minimal volume grows quickly to cluster two granules, then it can be deduced that the compatibility of the component granules is low and the clustering process can be completed.

Again, it is worth emphasizing that the granularity of data adds an extra important dimension to any processing. Not only a location of the information granule is essential but also its size plays a crucial role

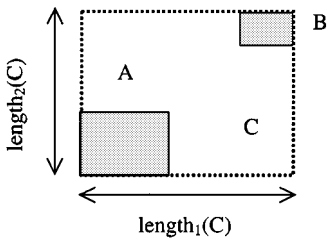

Fig. 4. Information granule C as a result of combining A and B.

in the process of clustering and afterwards during the validation of the clusters.

IV. Computational Aspects of Granular Computing

There are two essential functional elements of granular clustering that need to be constructed prior to moving to the detailed computing. These concern a way in which a distance between two information granules is determined and how we compute an inclusion relation between them. While the definitions generalize to a multidimensional case, we focus here on a 2-D case. Note also that these two concepts work for heterogeneous data that is granules and numeric entities.

A. Defining Compatibility Between Information Granules

In this section, we discuss details in which a compatibility and inclusion between two information granules are computed. The issue is more complicated than in a numeric case as these notions are granular and therefore the definitions of compatibility and inclusion should reflect this aspect as well.

Consider two information granules (hyperboxes) A and B. More explicitly, we follow a full notation $A(\mathbf{l}_a, \mathbf{u}_a)$ and $B(\mathbf{l}_b, \mathbf{u}_b)$ to point at their location in the space. The expression of compatibility, $compat(A, B)$ involves two components that is distance between A and B, $d(A, B)$, and a size of a newly formed information granule that comes when merging A and B. The distance is defined in the form

$$d(A, B) = (\|\mathbf{l}_b - \mathbf{l}_a\| + \|\mathbf{u}_b - \mathbf{u}_a\|)/2 \tag{1}$$

that is an average of the two distances. Obviously $\|.\|$ is a distance defined between the two numeric vectors. To make the framework general enough, we treat $\|.\|$ as an L_p distance , $p > 1$. By changing the value of "p" we sweep across a spectrum of well known distances that depend upon a particular value of "p." For instance, $p = 1$ yields a Hamming distance, L_1. The value $p = 2$ produces the well-known Euclidean distance, L_2. For $p = \infty$ we refer to a Tchebyschev distance, L_∞.

Once A and B have been combined giving rise to a new information granule C, its granularity can be captured by a volume, $V(C)$ computed in a standard way

$$V(C) = \prod_{i=1}^{n} length_i(C) \tag{2}$$

where

$$length_i(C) = \max(u_{\mathbf{b}}(i), u_{\mathbf{a}}(i)) - \min(l_{\mathbf{b}}(i), l_{\mathbf{a}}(i)) \tag{3}$$

$i = 1, 2, \ldots, n$. For details, refer to Fig. 4.

Expressions (1) and (2) are the contributing factors to the compatibility measure $compat(A, B)$ to be defined now in the form

$$compat(A, B) = 1 - d(A, B)e^{-\alpha V(C)}. \tag{4}$$

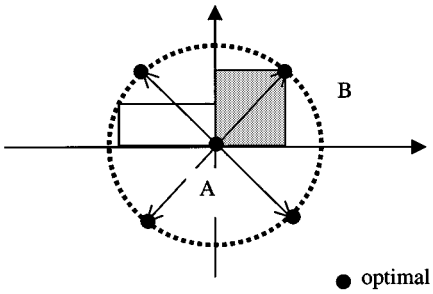

Fig. 5. Calculations of the compatibility measure; note that there are four possible candidates (Bs) on the circle that maximize this measure.

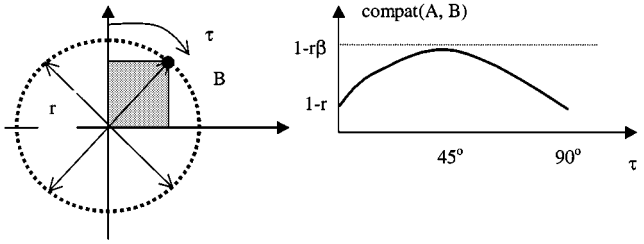

Fig. 6. Compatibility measure expressed as a function of τ (the plot here plot is restricted to the first 90° degrees); $\beta = e^{-\alpha r.r/2}$.

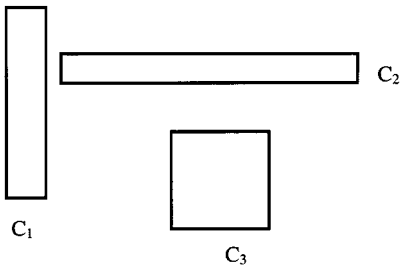

Fig. 7. Examples of information granules characterized by various degrees of balance of information granularity; note that C_1 and C_2 are highly unbalanced as exhibiting different levels of information specificity along one of the variables (C_1 and C_2 with high specificity along $x1$ and $x2$, respectively) while C_3 is well-balanced.

The rationale behind the above form of the compatibility measure is as follows. In clustering we aggregate (cluster) two information granules that are the closest *viz.* their compatibility measure is the highest, $compat(A, B) = 1$. In light of the above criterion, the candidate granules to be clustered should not only be "close" enough (which is reflected by the distance component) but the resulting granule should be "compact" (meaning that the size of the granule in every dimension is approximately equal). The second requirement favors such A and B that give rise to a maximum volume for a given $d(A, B)$, in other words it stipulates formation of hyperboxes that are as similar to hypercubes as possible. The particular exponential form of this expression has to do with the normalization criterion so that all values are kept in the unit interval. In particular, the volume of a point produces $e^{-0} = 1$ While the volume increases, its exponential function goes down to zero. The parameter α balances the two concerns in the compatibility measure and is chosen so as to control an extent to which the volume impacts the compatibility measure.

The compactness factor ($e^{-\alpha V(C)}$) introduced in the compatibility measure is critical to the overall processing (*viz.* clustering) of the information granules. By contrast, it is not essential and does not play any role when we proceed in a standard way and do not attempt to develop granules but retain a cluster of numeric data.

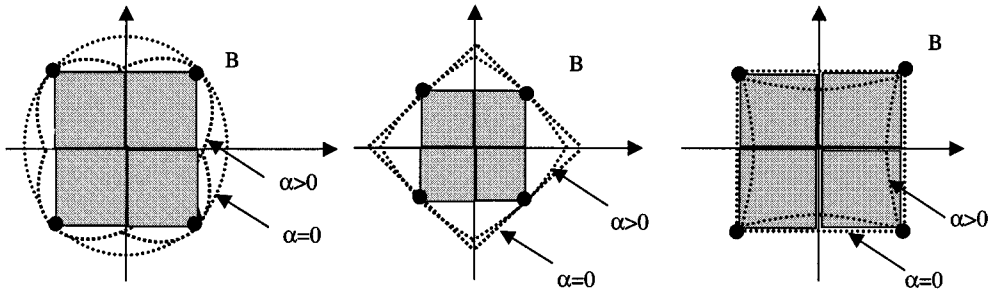

Fig. 8. Identification of Bs leading to the highest value of the compatibility measure.

To retain the values of the compatibility measure to the unit interval, we consider the data to lie in the unit hypercube $[0,\ 1]^n \subset \mathbf{R}^n$ [in other words we normalize the data before computing the value of (4)].

To gain a better insight of what really is accomplished when using the above compatibility measure, let us study two points (numeric values) A and B situated in $\mathbf{R}^2$. Furthermore let A be fixed and located at the origin of the coordinates while we allow B with some flexibility. $d(A,\ B)$ is just a standard Euclidean distance. It becomes obvious that all elements (Bs) located on a circle of a fixed radius exhibit the same distance value. Restrict now a choice of Bs from this pool. If we connect A and any of such Bs, the resulting volume changes its value depending upon the location of B. Interestingly, out of all Bs, there are four of them (located on this circle) for which the volume of the resulting attains its maximum. This happens if such box (*viz.* the information granule formed by clustering A and B) is a square, refer again to Fig. 5. In other words, the compatibility measure attains a maximal value there.

If we plot the compatibility measure as a function of τ (where τ is an angular position of B), we can easily see that the values of the compatibility measure is modulated by the angle (or equivalently the shape of the resulting information granule C); see Fig. 6.

More importantly, the above graphical considerations shed light on the geometry of the information granules that are preferred by the introduced compatibility measure. Such preference reflects a principle that may be coined as a *principle of balanced information granularity*. In a nutshell, in building new information granules, we prefer to have entities whose granularity is balanced along all dimensions (variables) rather than constructing granules that are highly unbalanced. A number of selected examples of varying granularity are portrayed in Fig. 7.

When changing the distance function to the Hamming ($p = 1$) and Tchebyschev distance ($p = \infty$), and carrying out the calculations of the compatibility measure, see Fig. 5, now we have a number of Bs to choose from yet this selection can be made from different geometrical figures (that is a diamond and a square), Fig. 8.

Moving on to the case where both A and B are two information granules, the resulting plots visualizing the compatibility measure are collected in Fig. 9.

As the clustering proceeds (refer to Fig. 2) the process of merging the progressively less closely associated patterns finds its reflection in the gradual reduction of the compatibility measure (4). A typical plot of the evolution of the compatibility measure over the complete clustering cycle is shown in Fig. 10.

It is self evident that the proximity of patterns that are being merged into granules at the early stages of the clustering process, is reflected in the relatively small gradient of the compatibility measure curve. By contrast, a large gradient of the curve, at the final stages of the clustering, indicates merging of incompatible clusters. The compatibility measure curve provides therefore a convenient reference for identifying which number of clusters captures the essential characteristics of the input data while providing the best generalization of them. The intersection of the two gradient lines (as indicated in Fig. 10) can be used as an approximation to the optimal number of clusters. This number provides a good starting point in the subsequent optimization of the overlap of the identified clusters as discussed below.

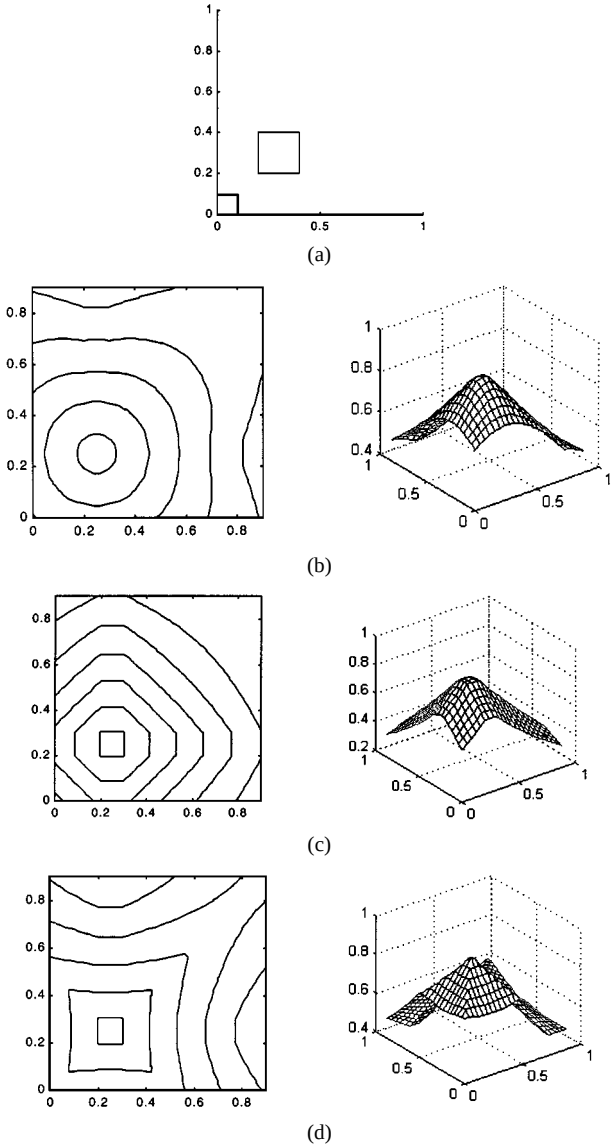

Fig. 9. Comparison of compatibility measures obtained with various distance measures. Note the preference that the compatibility measure gives to hyperboxes that are well balanced in all dimensions. This contrasts with the membership function proposed in [15] and illustrated in Fig. 3. (a) Two hyperboxes representing information granules in a unit box in $\mathbf{R}^2$. (b) Compatibility measure with L_2 distance measure. (c) Compatibility measure with L_1 distance measure. (d) Compatibility measure with L_∞ distance measure.

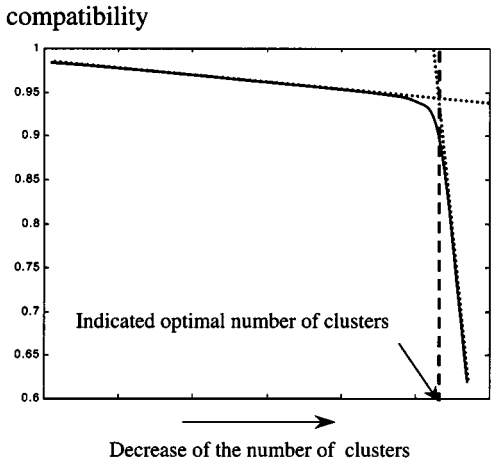

Fig. 10. Example of the evolution of the compatibility measure over the full cycle of the clustering process.

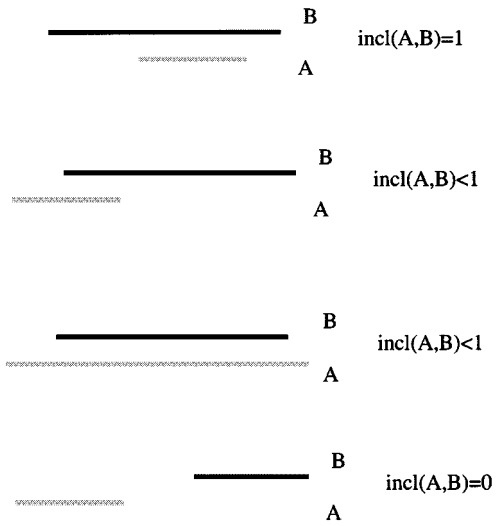

Fig. 11. Computing the inclusion for two information granules A and B.

B. Expressing Inclusion of Information Granules

The inclusion relation expressing an extent to which A is included in B is defined as a ratio of two volumes

$$incl(A,\, B) = \frac{V(A \cap B)}{V(A)}. \tag{5}$$

It is clear from the above that the inclusion measure is monotonic, noncommutative and satisfies the following boundary conditions: $Incl(A,\, X) = 1$ and $Incl(A,\, \emptyset) = 0$ where X and $\emptyset$ are the unit hyperbox and the empty set in $\mathbf{R}^n$, respectively. The calculations are straightforward; Fig. 11 enumerates all cases for 1-D granules along with the pertinent values of this measure.

It is worth mentioning that the value of the inclusion measure drop down quite substantially (at a rate of a^{-n} where $a \in (0,\, 1\}$ with the increasing dimension of the space in which the information granules are distributed. For example if there is an 1/2 overlap ($a = 2$) in each variable in an n-dimensional space, the inclusion level expresses as 2^{-n}.

Clearly, the objective of effective information abstraction through clustering of information granules translates into identifying for which number of clusters there is a minimum overlap between the clusters. To encourage merging of clusters that have significant overlap we calculate an average of the maximum inclusion rates of each cluster in every other cluster: (6)

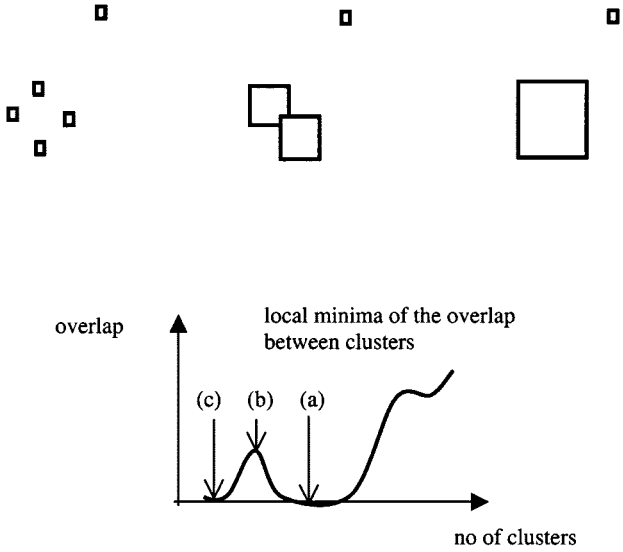

Fig. 12. Progression from five to two clusters involves stage (b) during which clusters overlap. This is reflected in $overlap(3) > 0$ while $overlap(5) = 0$ and $overlap(2) = 0$.

$$overlap(c) = \frac{1}{c-1} \sum_{i=1}^{c} \max_{\substack{j=1,\,\ldots c \\ j \neq i}} (incl(A_i,\, A_j)) \tag{6}$$

where c is the current number of clusters and A_i and A_j are ith and jth cluster respectively. It has to be pointed out however that, while the measure (5) is monotonic for any two pairs of clusters i.e., if $A \subset B$ and $C \subset D$, then $incl(A,\, C) \leq incl(B,\, D)$, the change of the number and the size of clusters during the clustering process results in the collective measure, (6), having various local optima. We illustrate this effect in Fig. 12.

Because of the local minima of the $overlap(.)$ function it is important to have a good initial estimate of the number of clusters as a starting point for the local minimization of the function. Such an estimate is provided by our earlier analysis of the compatibility measure as discussed in the previous section.

Having accomplished the clustering process the quality of data abstraction afforded by the given set of data clusters is measured using an independent validation data set. The generality of each of the identified clusters is well quantified by the sum of the inclusion rates of the validation data items in the respective cluster

$$INCL(i) = \sum_{j=1}^{M} incl(V_j,\, A_i) \qquad i = 1,\, \ldots c \tag{7}$$

where c is the number of clusters and M is the cardinality of the validation data set. As well as indicating whether a given cluster is representative for a large proportion of data the $INCL(.)$ measure can be used to assess how representative are the training and the validation data sets. If the sets are representative, then $INCL(.)$ should correlate closely with the cardinality of the individual clusters.

V. The Granular Analysis

The hyperboxes constructed during the design phase are helpful in a thorough analysis. They shed light on the nature of data as they are perceived from the standpoint of information granularity established during the design of the hyperboxes. Two main aspects are distinguished. First, we characterize the hyperboxes themselves. Second,

we analyze the properties of the variables (features) forming the data space. We should emphasize that the granular analysis follows the synthesis phase and does not impact it in any way. To maintain conciseness of the presentation, we consider that each out of "c" hyperboxes located in the n-dimensional space is fully described by vectors of its lower and upper corners (coordinates), that is $\mathbf{B}(k) = \{\mathbf{l}(k), \mathbf{u}(k)\}$, $k = 1, 2, \ldots, c$ where $\mathbf{l}(k)$ and $\mathbf{u}(k)$ are vectors of the corresponding coordinates, that is $\mathbf{l}^T(k) = [l_1(k)\ l_2(k) \cdots l_n(k)]$ and $\mathbf{u}^T(k) = [u_1(k)\ u_2(k) \cdots u_n(k)]$.

A. Characterization of Hyperboxes

The most evident characterization of the hyperboxes can be provided in their volumes, $V(B(k))$. The computations are obvious. First, we determine a ratio (normalized length)

$$norm\_length_i(B(k)) = \frac{u_i(k) - l_i(k)}{range_i(B(k))}$$

where $range_i(B(k))$ is a range of the ith feature (variable). Since the data is normalized to a unit hypercube the $range_I(B(k)) = 1$ for all i. Second, the volume is taken as a product

$$V(B(k)) = \prod_{i=1}^{n} norm\_length_i(B(k)).$$

The volume quantifies the essence of granularity of the hyperboxes. Intuitively, it states how "large" (detailed) the hyperboxes are and how much details each of them captures. One can take an average of the volumes of the hyperboxes that gives a general summary of the hyperboxes

$$\overline{v} = \frac{1}{c} \sum_{k=1}^{c} V(B(k)).$$

If one side of the hyperbox is zero then the volume measure returns a zero value. This occurs because of the multiplicative nature of volume. To alleviate such problem, we may also introduce a measure of an additive character. A plausible descriptor of a hyperbox could reflect a "circumference" of the hyperbox and read as follows

$$\sum_{i=1}^{n} norm\_length_i(B(k)).$$

B. Granular Feature Analysis

The granulation of the data space (and each feature) provides an interesting insight into the nature of the variables occurring in the problem. In what follows, we provide their description in terms of sparsity and discrimination abilities. These two descriptors are exclusively implied by the granular nature of the hyperboxes.

Sparsity: When looking at a certain variable of the hyperboxes, we can visualize how much of the entire range of the variable is occupied by the hyperboxes (i.e., how *sparse* the boxes are in the given space). Take the ith feature and calculate the sum of length of the corresponding sides of the hyperboxes that is

$$tot\_length_i = \sum_{k=1}^{c} length_i(B(k))$$

where $length_i(B(k)) = u_i(k) - l_i(k)$. The sparsity defined in the form

$$sparsity_i = \frac{tot\_length_i}{range_i} \frac{1}{c}$$

assumes values in the unit interval. If $sparsity_i$ is less than 1 then this represents a situation when hyperboxes (more precisely its ith coordinate) occupy a portion of the entire range of the feature. We may state that the variable is "underutilized." In other words, we witness a highly localized usage of this feature. The sparsity around 1 means a complete utilization of the variable. The effect of overutilization happens when sparsity achieves values higher than 1 (in this case we have some hyperboxes overlapping along this variable).

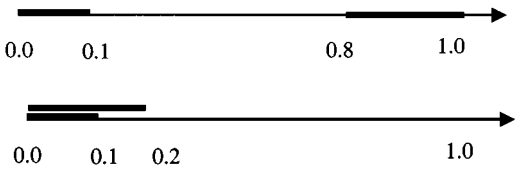

Fig. 13. Two different distributions of hyperboxes (ith feature) producing the same value of the sparsity index; in both cases the sparsity is equal to 0.3.

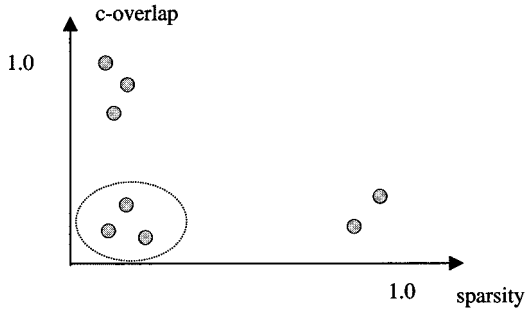

Fig. 14. Sparsity-overlap space and feature arrangement; note a collection of highly discriminative features of low sparsity.

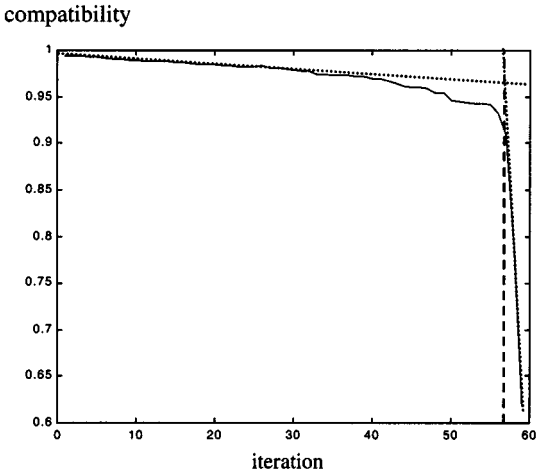

Fig. 15. Compatibility measure for a single clustering process.

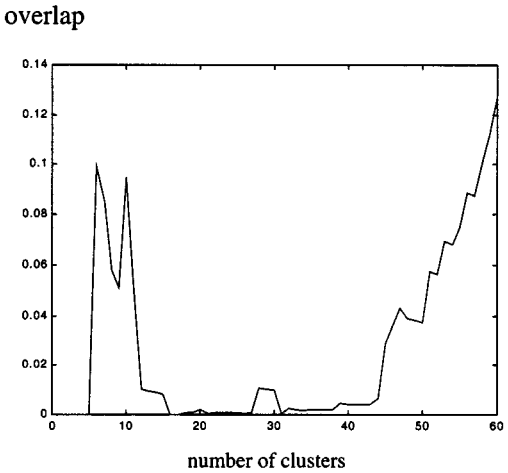

Fig. 16. Average degree of overlap of clusters.

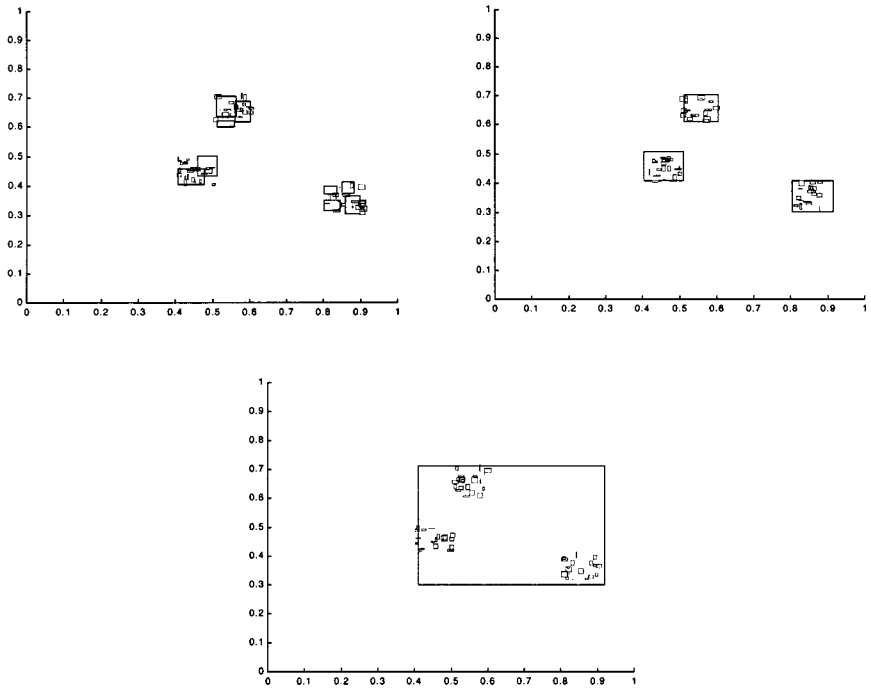

Fig. 17. Inclusion of the validation data in ten, three, and one cluster respectively.

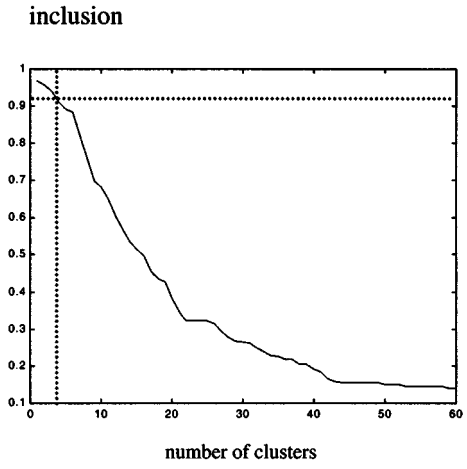

Fig. 18. Average inclusion rate for the validation data set.

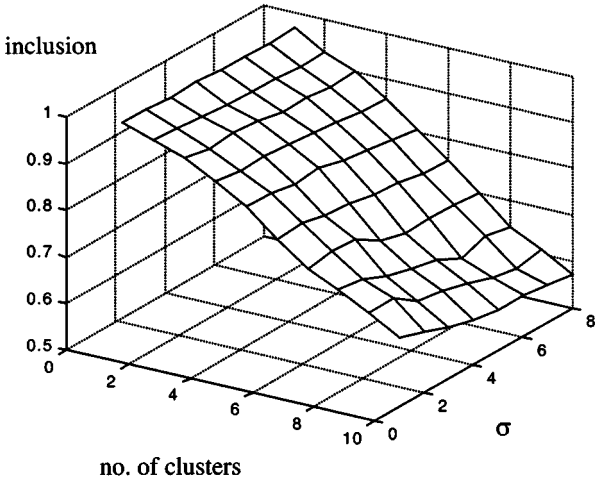

Fig. 19. Average inclusion measure evaluated for 8000 training and validation sets.

The sparsity does not capture the entire picture. A situation illustrated in Fig. 13 shows two cases where the distribution of the hyperbox along the given feature is very different yet we end up having the same value of the sparsity. This leads us to another index (descriptor) that describes an overlap between the hyperboxes.

Overlap Index: We define the following index called coordinate overlap

$$c-overlap_i = \frac{2}{c(c-1)} \sum_{k=1}^{c-1} \sum_{l>k}^{c} \frac{length_i(I(k) \cap I(l))}{length_i(I(k) \cup I(l))}$$

$i = 1, 2, \ldots, n$. In this definition, $I(k)$ and $I(l)$ are intervals (sides) of the hyperboxes for the ith variable. The higher the value of this index, the more overlap between the hyperboxes expressed along the given variable. When $I(k)$ and $I(l)$ are pairwise disjoint then the overlap is equal to zero. This means that the feature is highly discriminative as it separated the hyperboxes. The higher the overlap measure, the lower the discriminatory aspects of the feature.

Each of the measures leads to a linear ordering of the variables. We can easily state which of the variables is highly "utilized" and which of them comes with the most significant discriminatory properties. To form a comprehensive picture, one can localize each feature in the sparsity-overlap space (see Fig. 14). By doing this, one can distinguish between the variables that are essential to the problem. More specifically, we prefer features that exhibit low overlap (as those come with strong discriminative properties) along with low values of sparsity that points at the issue of the localized usage of the variable.

It should be stressed that the above descriptors (sparsity and overlap) of the features emerge as important quantifiers because of the existence of information granules forming the hyperboxes.

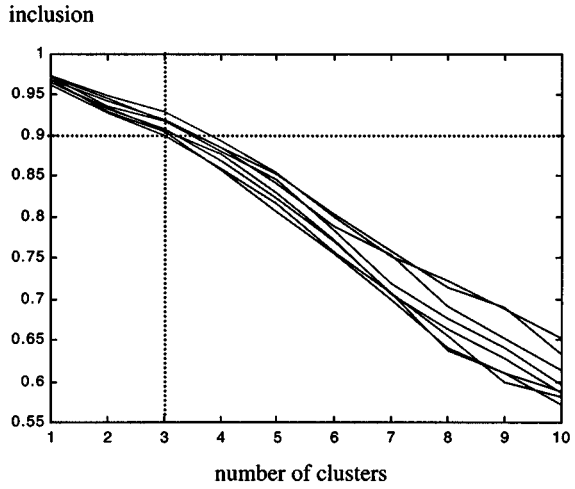

Fig. 20. Two-dimensional projection of the surface from Fig. 15 resulting in a family of curves illustrating average inclusion rates of the validation data in clusters for the various values of σ.

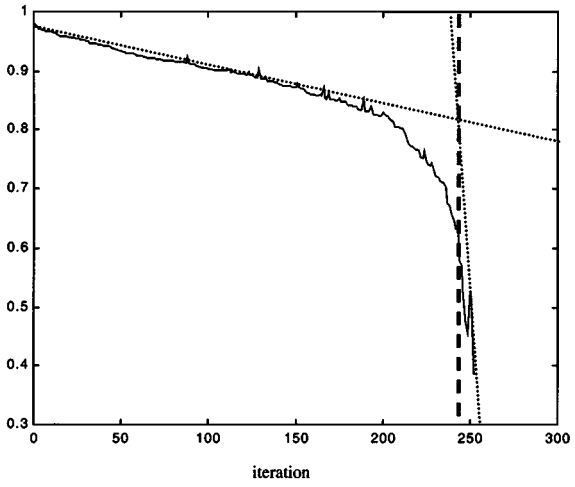

Fig. 21. Compatibility measure of clusters formed from the odd-numbered records in the Boston housing data set. Iteration step 245 corresponds to seven clusters.

VI. Experimental Studies

The series of experiments is aimed at visualizing the most essential features of granular clustering. We consider both synthetic data set and the one available on the WWW (Boston housing data).

A. Synthetic Data

The synthetic data sets consist of three groups of information granules (hyperboxes), $A_i \in [0,\ 1] \times [0,\ 1]$, generated by a random number generator with a uniform distribution. Each group comprises 20 granules dispersed around pre-defined points: $c_1 = [0.4,\ 0.4]$; $c_2 = [0.5,\ 0.6]$; and $c_3 = [0.8,\ 0.3]$. The dispersion factor σ is varied between 0.08 and 0.15 to establish the sensitivity of the clustering process to the dispersion of the data. The clustering process is governed by the compatibility measure, (4), with the distance defined according to L_2 norm and the "compactness" factor $\alpha = 0.5$.

An example of the evolution of the compatibility measure throughout the clustering process is shown in Fig. 15. The intersection of the two asymptotes to the compatibility measure traced at the beginning and at the end of the clustering process indicates that three clusters (iteration 57) mark a natural "change over" point in the behavior of the system. So, the clustering process should terminate with three clusters

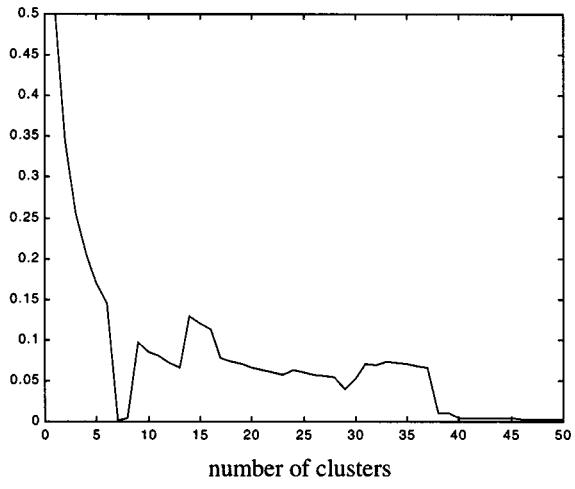

Fig. 22. Degree of average overlap of clusters in the last 50 out of 252 iterations.

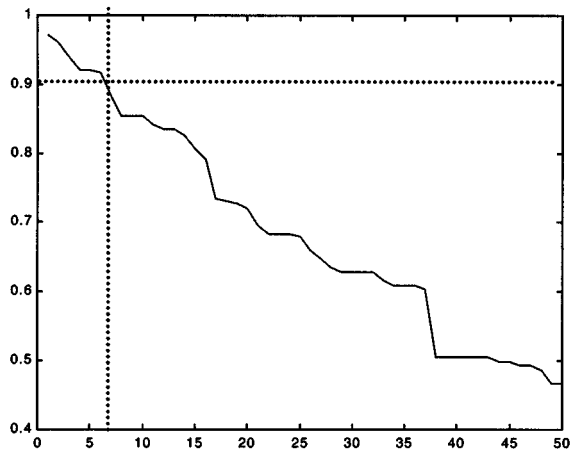

Fig. 23. Average inclusion measure evaluated for one to 50 clusters.

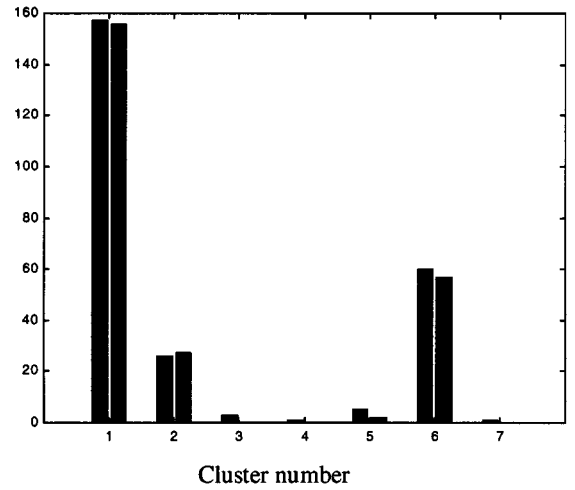

Fig. 24. Cardinality (first bar) and the aggregate inclusion rate (second bar) for each of the seven clusters.

providing that the degree of overlap of clusters is also minimized for this number of clusters.

The degree of overlap of clusters was evaluated at each of the 59 iterative steps of the clustering process, according to (6), and is illustrated in Fig. 16.

As expected, the results of the cluster overlap analysis clearly confirm that the test data naturally falls into three clusters.

TABLE I
DESCRIPTION OF THE SEVEN CLUSTERS. (L_i REPRESENTS MINIMUM COORDINATES OF THE iTH HYPERBOX AND U_i REPRESENTS MAXIMUM COORDINATES)

Variables 1 through 7

L1	0.0063	0	0.7399	0	0.3850	4.9730	6.0004
U1	2.6354	95.0000	19.5800	1.0000	0.6470	8.3980	100.0000
L2	0.0686	0	8.1399	0	0.5200	4.9030	69.6999
U2	2.7795	0	27.7400	0	0.8710	6.4580	100.0000
L3	1.1265	0	19.5800	1.0000	0.8710	5.0120	88.0004
U3	3.3213	0	19.5800	1.0000	0.8710	6.1290	100.0000
L4	2.0099	0	19.5800	0	0.6050	7.9290	96.2005
U4	2.0099	0	19.5800	0	0.6050	7.9290	96.2005
L5	3.4744	0	18.1001	1.0000	0.6310	5.8750	82.8997
U5	8.9834	0	18.1001	1.0000	0.7700	8.7800	97.4997
L6	2.3783	0	18.1001	0	0.5320	4.1380	41.9002
U6	73.5337	0	18.1001	0	0.7700	7.0610	100.0000
L7	88.9762	0	18.1001	0	0.6710	6.9680	91.8999
U7	88.9762	0	18.1001	0	0.6710	6.9680	91.8999

Variables 8 through 14

L1	1.7984	1.0000	192.9998	12.6000	288.9906	1.9199	12.7000
U1	10.7103	8.0000	469.0011	22.0000	396.9000	30.8101	50.0000
L2	1.3459	2.0000	188.0008	14.7000	70.8002	6.4300	8.1000
U2	3.9900	4.9999	711.0000	21.2000	396.9000	29.6801	24.3000
L3	1.3216	4.9999	402.9980	14.7000	321.0184	12.1200	13.4002
U3	1.7494	4.9999	402.9980	14.7000	396.9000	26.8200	17.0002
L4	2.0459	4.9999	402.9980	14.7000	369.2980	3.7000	50.0000
U4	2.0459	4.9999	402.9980	14.7000	369.2980	3.7000	50.0000
L5	1.1296	24.0000	665.9989	20.2000	347.8787	2.9600	17.7998
U5	2.7227	24.0000	665.9989	20.2000	395.4287	17.5999	50.0000
L6	1.1370	24.0000	665.9989	20.2000	0.3200	3.2601	5.0000
U6	3.7240	24.0000	665.9989	20.2000	396.9000	37.9700	50.0000

Variable no.	sparsity	c-overlap
1	0.135	0.1826
2	0.136	0.7143
3	0.201	0.2194
4	0.143	0.3333
5	0.291	0.1933
6	0.326	0.3255
7	0.307	0.3759
8	0.210	0.3397
9	0.062	0.2109
10	0.218	0.2381
11	0.241	0.2234
12	0.344	0.4357
13	0.458	0.4399
14	0.426	0.3759

The quality of data abstraction achieved through clustering is assessed by evaluating the inclusion rate, (7), of the independently generated data set (with the same statistical properties) in the constructed clusters. An example of the output of the validation process for ten, three, and one cluster is illustrated in Fig. 17.

The change of the overall inclusion rate of the validation data throughout the clustering process is illustrated in Fig. 18. It is not surprising to see that the high value of the average inclusion rate for three or fewer clusters confirms that three clusters capture the essential features of the data while the high value of the compatibility measure confirms that the clusters retain high specificity. Should the number of clusters be reduced to two or one, the inclusion rate of the validation data set would only be improved marginally while there would be a very significant reduction of specificity of the cluster(s).

In order to achieve a degree of independence from the statistical characteristics of the random number generator the evaluation of the inclusion of the validation data sets in the clusters was repeated 100 times for each value of $\sigma \in \{0.08, 0.09, 0.10, 0.11, 0.12, 0.13, 0.14, 0.15\}$ and the number of clusters varying from one to ten. A total of 8000 training sets and 8000 validation sets were processed.

Fig. 19 illustrates how the inclusion measure, (5), depends on the dispersion parameter σ and the number of clusters. It is interesting to note that σ has little influence on the value of the inclusion measure. This is a very desirable characteristic of the clustering process since it suggests that the precise statistical properties of data sets do not need to be known for the clustering to be effective.

It is easy to note, from Figs. 19 and 20, that the inclusion rate of 0.9 or higher is attained consistently with three or fewer clusters.

B. Boston Housing Data

Although for 2-D data sets $B \in \mathcal{P}(\mathbf{R}^2)$ the number of clusters can be easily established by visual inspection, the higher dimensional data presents a significant challenge. We have applied therefore the algorithm to a realistic 14-dimensional (14-D) data set representing factors affecting house prices in the Boston, MA area. The data set has been originally compiled by Harrison and Rubinfeld [6], and is available from the Machine Learning Database at University of California at Irvine (http://www.ics.uci.edu/~mlearn/MLSummary.html). The data

TABLE II
DESCRIPTION OF SEVEN CLUSTERS FOR THE TRAINING AND VALIDATION SETS SWITCHED ROUND (L_i REPRESENTS MINIMUM AND U_i REPRESENTS MAXIMUM COORDINATES OF THE HYPERBOX)

Variables 1 through 7

L1	0.0108	0	0.4600	0	0.3890	5.0190	2.9000
U1	2.9236	100.0000	25.6501	0	0.6470	8.7250	100.0000
L2	0.7617	0	3.9701	0	0.6470	4.9260	62.8000
U2	4.0972	20.0000	19.5800	0	0.8710	6.5100	100.0000
L3	3.5349	0	19.5800	1.0000	0.8710	6.1520	82.5997
U3	3.5349	0	19.5800	1.0000	0.8710	6.1520	82.5997
L4	0.0615	0	6.2000	1.0000	0.4470	5.3440	27.6003
U4	1.5188	40.0000	19.5800	1.0000	0.6050	8.3750	100.0000
L5	0.0152	90.0000	1.2099	1.0000	0.4010	7.9230	24.7999
U5	0.0152	90.0000	1.2099	1.0000	0.4010	7.9230	24.7999
L6	2.8187	0	18.1001	0	0.5320	3.5610	40.3000
U6	67.9206	0	18.1001	1.0000	0.7700	7.3930	100.0000
L7	0.1060	0	27.7400	0	0.6090	5.4140	98.2998
U7	0.1834	0	27.7400	0	0.6090	5.9830	98.7998

Variables 8 through 14

L1	1.4394	1.0000	187.0000	12.6000	227.6119	1.7300	11.8999
U1	12.1265	8.0000	437.0004	22.0000	396.9000	34.4101	50.0000
L2	1.4118	4.9999	264.0018	13.0000	172.9116	7.3900	13.8002
U2	1.9865	4.9999	402.9980	14.7000	396.9000	29.5301	23.3001
L3	1.7455	4.9999	402.9980	14.7000	88.0118	15.0199	15.6002
U3	1.7455	4.9999	402.9980	14.7000	88.0118	15.0199	15.6002
L4	2.1620	3.0001	222.9988	14.7000	388.4489	3.3198	19.3001
U4	4.8628	8.0000	402.9980	18.6000	396.9000	23.9799	50.0000
L5	5.8850	1.0000	197.9988	13.6000	395.5199	3.1600	50.0000
U5	5.8850	1.0000	197.9988	13.6000	395.5199	3.1600	50.0000
L6	1.1691	24.0000	665.9989	20.2000	2.5210	3.7301	5.0000
U6	4.0983	24.0000	665.9989	20.2000	396.9000	34.7700	50.0000
L7	1.7554	3.9999	711.0000	20.1000	344.0517	18.0699	6.9998
U7	1.8682	3.9999	711.0000	20.1000	390.1106	23.9701	13.6000

Variable no.	sparsity	c-overlap
1	0.117	0.1414
2	0.229	0.2667
3	0.284	0.1432
4	0.143	0.5238
5	0.258	0.0985
6	0.348	0.3391
7	0.393	0.3144
8	0.221	0.1674
9	0.075	0.1769
10	0.155	0.1560
11	0.228	0.0760
12	0.303	0.3276
13	0.443	0.3762
14	0.412	0.3175

set comprises of 506 records where each of them describes real estate through 14 attributes such as crime rate, average number of rooms, index of accessibility to radial highways, etc.

Study A: We divided the original set into two sets. The training set, comprising 253 odd-numbered records and the validation set comprising 253 even-numbered records. It should be noted that, as a preprocessing step, all data has been mapped into a 14-D unit hyperbox. The compatibility measure provided direction for the clustering process and the evolution of this measure throughout the whole process is presented in Fig. 21. The gradients of the compatibility measure at the beginning and the end of the process indicate that seven clusters represent a good abstraction of the training data.

In the vicinity of seven clusters the cluster overlap indicator is minimized for seven and eight clusters, as shown in Fig. 22. Of these two possible numbers of clusters we select the smaller number so as to achieve greater granulation of the original data.

The generality of the identified clusters was tested by evaluating average inclusion of the validation data set (even-numbered records from the original data set) in the sets of clusters identified in the last 50 steps of the clustering process. This is illustrated in Fig. 23. The value of over 90%, achieved for seven clusters, indicates a good abstraction of the detailed data that is achieved with this number of clusters.

To gain a more detailed insight into the makeup of the seven clusters we evaluated an aggregate inclusion measure (7), using the validation set, and compared the results with the cardinality of each cluster. It is clear, from Fig. 24, that out of seven clusters, three have a significant support in the two data sets while the other four clusters represent data that could be described as significant exceptions. It is interesting to note however that the zero inclusion rates of the validation data in clusters three, four, and seven indicate that the small data sample makes it difficult to do a proper evaluation of the clusters.

The full description of the identified clusters is given in Table I.

The results of feature analysis is summarized in terms of their sparsity and overlap values. This analysis provides with an interesting insight into the discriminatory properties of the variables in the problem. The most dominant ones are: crime rate (1), nitric oxide concentra-

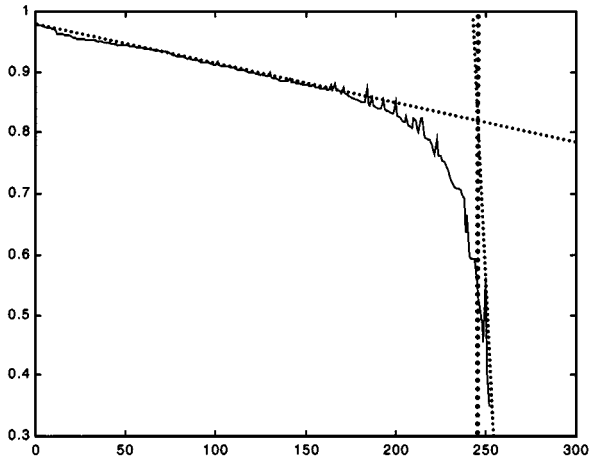

Fig. 25. Compatibility measure of clusters formed from the even-numbered records in the Boston housing data set. Iteration step 246 corresponds to six clusters.

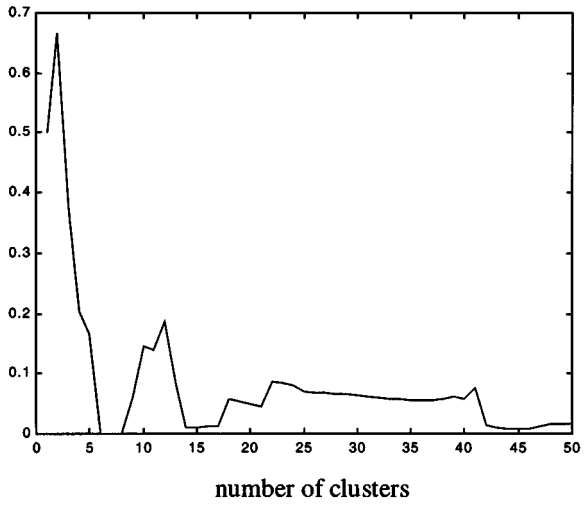

Fig. 26. Degree of average overlap of clusters in the last 50 iterations.

tion (5), index of accessibility to radial highways (9), and proportion of nonretail business acres (3). In other words, these are the variables that discriminate between hyperboxes (we stress that the discriminatory aspects have been raised in the setting of the information granules):

Variable no.	sparsity	c-overlap
1	0.135	0.1826
2	0.136	0.7143
3	0.201	0.2194
4	0.143	0.3333
5	0.291	0.1933
6	0.326	0.3255
7	0.307	0.3759
8	0.210	0.3397
9	0.062	0.2109
10	0.218	0.2381
11	0.241	0.2234
12	0.344	0.4357
13	0.458	0.4399
14	0.426	0.3759

Study B: In order to ascertain whether the selection of records for the training and the validation data sets had influenced significantly conclusions regarding the number of clusters abstracting the original data set, we repeated the clustering process with the training and validation sets switched round (see Table II). Again the compatibility measure directed the clustering process and the asymptotic evolution of the measure, at the initial and final stages of the process, indicated that six data clusters mark a "change-over" point in the clustering process (Fig. 25).

The curve showing the average degree of overlap between the clusters, illustrated in Fig. 26, indicates that a minimum overlap is achieved with six, seven, and eight clusters. For the ease of comparison with the Study A case we select seven clusters for the validation stage. The average inclusion rate of the validation data set (odd-numbered records from the original data set) in the seven clusters is approximately 30%

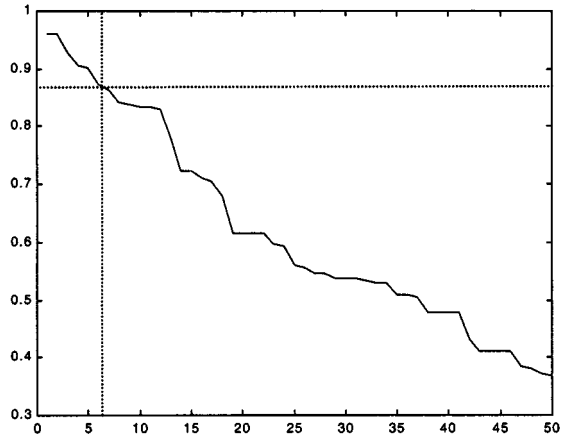

Fig. 27. Inclusion measure evaluated for one to 50 clusters.

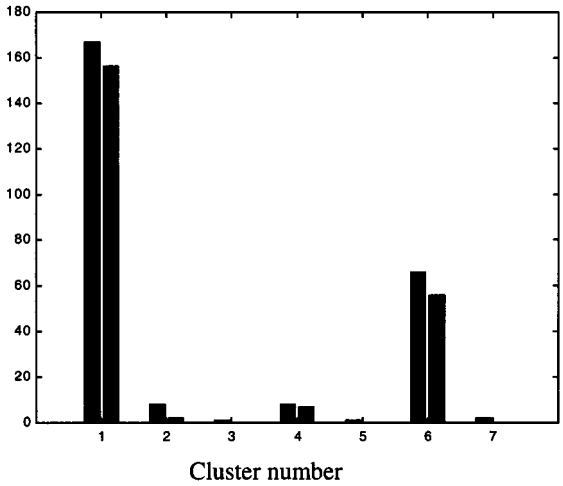

Fig. 28. Cardinality (first bar) and the aggregate inclusion rate (second bar) for each of the seven clusters.

worse than in the previous case, averaging at 86%. This is illustrated in Fig. 27. The reduction of the average inclusion rate in this case suggests that the training and validation sets contain a small number of unique patterns that do not have counterparts in the other set. The result is that although the distinctiveness of these patterns warrants their inclusion in separate clusters, the cross-comparison of these "minority clusters" is very limited. This is further verified by the inspection of Fig. 28, which

shows that the clusters thre, five, and seven are representing by one, one, and two patterns, respectively, with no corresponding patterns in the validation set. It is also interesting to note that, compared to the Study A, there is a greater discrepancy between the cardinality of the clusters and the inclusion rate. We conclude therefore that the size of the data supports only firm conclusions about two clusters and the characterization of further clusters requires an order of magnitude greater data sample.

The sparsity and c-overlap of the features (variables) are very similar as in Study A meaning that some global properties discovered in the data set have been retained:

Variable no.	sparsity	c-overlap
1	0.117	0.1414
2	0.229	0.2667
3	0.284	0.1432
4	0.143	0.5238
5	0.258	0.0985
6	0.348	0.3391
7	0.393	0.3144
8	0.221	0.1674
9	0.075	0.1769
10	0.155	0.1560
11	0.228	0.0760
12	0.303	0.3276
13	0.443	0.3762
14	0.412	0.3175

VII. Conclusions

The study has articulated another look at data analysis by providing a constructive way of forming information granules that capture the essence of the large collections of numeric data. In this sense, the original data are compressed down to a few information granules whose location in the data space and granularity reflect the structure in the data. The approach promotes a data mining way of problem solving by emphasizing the transparency of the results (hyperboxes). The way in which information granules is guided by two aspects that is distance between information granules and a size (granularity) of the potential information granule formed through merging two other granules. These two aspects are encapsulated in the form of the compatibility measure. Moreover we discussed a number of indexes describing the hyperboxes and expressing relationships between such information granules. It has been shown how to validate the granular structure. The resulting family of the information granules is a concise descriptor of the structure of the data—we may call them a granular *signature* of the data.

Some further extensions of the hyperbox approach may deal with more detailed instruments of information granulation such as fuzzy sets [7], [11].

It should be stressed that the proposed approach to data analysis is *noninvasive* meaning that we have not attempted to formulate specific assumptions about the distribution of the data but rather allow the data to "speak" freely. This is accomplished in two main ways:

- first, the hyperboxes are easily understood by a user as each dimension (variable) comes as a part of the construct
- second, the approach finds relationships that are direction-free meaning that we do not distinguish between input and output variables (which could be quite restrictive as we may not know in advance what implies what). Obviously, this feature is quite common to all clustering methods.

Furthermore the granulation mechanism puts the variables (features) existing in the problem in a new perspective. The two indexes such as sparsity and overlap are useful in understanding the relevance of the variables, in particular their discriminatory abilities.

While the study was concerned with the development of information granules (hyperboxes), there are interesting inquires into their use in granular modeling. In particular, we are concerned with the fundamental inference problem

— given an input datum (information granule and numeric datum, in particular) X defined in a certain subspace of dimension n' of the original space $\mathbf{R}^{n'} \subset \mathbf{R}^n$ and a collection of information granules $\mathbf{B} = \{B(1)\ B(2), \ldots,\ B(c)\}$ determine the corresponding information granule Y.

The current paper provides a basis for this investigation.

References

[1] M. R. Anderberg, *Cluster Analysis for Applications*. New York: Academic, 1973.

[2] A. Bargiela, "Interval and ellipsoidal uncertainty models," in *Granular Comput.*, W. Pedrycz, Ed: Springer-Verlag, 2001, pp. 23–57.

[3] A. Bargiela and W. Pedrycz, "Information granules: Aggregation and interpretation issues," *IEEE Trans. Syst. Man, Cybern. B*, submitted for publication.

[4] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York: Plenum, 1981.

[5] B. Gabrys and A. Bargiela, "General fuzzy Min–Max neural network for clustering and classification," *IEEE Trans. Neural Networks*, vol. 11, pp. 769–783, May 2000.

[6] D. Harrison and D. L. Rubinfeld, "Hedonic prices and the demand for clean air," *J. Environ. Econ. Manage.*, vol. 5, pp. 81–102, 1978.

[7] A. Kandel, *Fuzzy Mathematical Techniques with Applications*. Reading, MA: Addison-Wesley, 1986.

[8] T. Kohonen, "Self-organized formation of topologically correct feature maps," *Biol. Cybern.*, vol. 43, pp. 59–69, 1982.

[9] ——, *Self-Organizing Maps*. Berlin: Springer-Verlag, 1995.

[10] W. Pedrycz, *Computational Intelligence: An Introduction*. Boca Raton, FL: CRC, 1997.

[11] W. Pedrycz and F. Gomide, *An Introduction to Fuzzy Sets*. Cambridge, MA: MIT Press, 1998.

[12] W. Pedrycz, "Fuzzy equalization in the construction of fuzzy sets," *Fuzzy Sets Syst.*, vol. 119, no. 2, pp. 329–335, 2001.

[13] W. Pedrycz, M. H. Smith, and A. Bargiela, "A granular signature of data," in *Proc. 19th Int. (IEEE) Conf. NAFIPS'2000*, Atlanta, GA, July 2000, pp. 69–73.

[14] P. K. Simpson, "Fuzzy Min–Max neural networks—Part 1: Classification," *IEEE Trans. Neural Networks*, vol. 3, pp. 776–786, Sept. 1992.

[15] ——, "Fuzzy Min–Max neural networks—Part 2: Clustering," *IEEE Trans. Neural Networks*, vol. 4, pp. 32–45, Feb. 1993.

[16] L. A. Zadeh, "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.

[17] ——, "Fuzzy logic = Computing with words," *IEEE Trans. Fuzzy Syst.*, vol. 4, no. 2, pp. 103–111, 1996.

[18] ——, "Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic," *Fuzzy Sets Syst.*, vol. 90, pp. 111–117, 1997.