Soft Comput (2005) 9: 155–163
DOI 10.1007/s00500-003-0339-2

ORIGINAL PAPER

A. Bargiela · W · Pedrycz

A model of granular data: a design problem with the Tchebyschev FCM

Published online: 19 December 2003

Abstract In this study, we propose a model of granular data emerging through a summarization and processing of numeric data. This model supports data analysis and contributes to further interpretation activities. The structure of data is revealed through the FCM equipped with the Tchebyschev (l_∞) metric. The paper offers a novel contribution of a gradient-based learning of the prototypes developed in the l_∞-based FCM. The l_∞ metric promotes development of easily interpretable information granules, namely hyperboxes. A detailed discussion of their geometry is provided. In particular, we discuss a deformation effect of the hyperbox-shape of granules due to an interaction between the granules. It is shown how the deformation effect can be quantified. Subsequently, we show how the clustering gives rise to a two-level topology of information granules: the core part of the topology comes in the form of hyperbox information granules. A residual structure is expressed through detailed, yet difficult to interpret, membership grades. Illustrative examples including synthetic data are studied.

Keywords Information granulation through clustering · FCM · l_∞ metric (distance) · Hyperboxes · Deformation effect in clustering · Geometry

A. Bargiela (✉)
Department of Computing,
The Nottingham Trent University,
Nottingham NG1 4BU, UK
E-mail: andre@doc.ntu.ac.uk

W. Pedrycz
Department of Electrical & Computer Engineering,
University of Alberta, Edmonton, Canada,
e-mail: pedrycz@ee.ualberta.ca

Present address: W. Pedrycz
Systems Research Institute,
Polish Academy of Sciences 01-447 Warsaw, Poland

1 Introduction

Clustering has been widely recognized as one of the dominant techniques of data analysis. The broad spectrum of the detailed algorithms and underlying technologies (fuzzy sets, neural networks, heuristic approaches) is impressive. In spite of this diversity, the key objective remains the same which is to understand the data. In this sense, clustering becomes an integral part of data mining [4, 15]. Data mining is aimed at making the findings that are inherently *transparent* to the end user. The transparency is accomplished through suitable knowledge representation mechanisms, namely a way in which generic data elements are formed, processed and presented to the user. The notion of information granularity becomes a cornerstone concept that needs to be discussed in this context, cf [15, 10].

The underlying idea is that in any data set we can distinguish between a core part of a structure of the data that is easily describable and interpretable in a straightforward manner and a residual part, which does not carry any evident pattern of regularity. The core part can be described in a compact manner through several information granules while the residual part does not exhibit any visible geometry and requires some formal descriptors such as membership formulas. The approach proposed in this study dwells on the standard FCM method equipped with a Tchebyschev distance that promotes hyperbox geometry of the information granules (hyperboxes). Starting from the results of clustering, our objective is to develop information granules forming a core structure in the data set, provide their characterization and discuss an interaction between the granules leading to their deformation.

The material is arranged into 7 sections. First, we formulate the problem and then move on to the modified clustering algorithm (Sect. 3). Section 4 is concerned with the generation of granular prototypes. In Sect. 5 we analyse the geometry of information granules and quantify the deformation of regular hyperboxes. Following that we propose a general model of granular data description (Sect. 6) and present conclusions in Sect. 7.

The concept and numerical studies are presented in parallel so that new ideas are made more tangible by some illustrative material.

2 Problem formulation

In what follows, we set up all necessary notations. The set of data (patterns) is denoted by $\mathbf{X}$, $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ while each pattern is an element in the n-dimensional unit hypercube, that is $[0, 1]^n$. The objective is to cluster X into "c" clusters and the problem is cast as an optimization task (objective function based optimization)

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

where $U = [u_{ik}]$, $i = 1, 2, \ldots, c$, $k = 1, 2, \ldots, N$ is a partition matrix describing clusters in data. The distance function (metric) between the k-th pattern and i-th prototype is denoted by d_{ik}. $d_{ik} = \mathrm{dist}(\mathbf{x}_k, \mathbf{v}_i)$ while $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_c$ are the prototypes characterizing the clusters. The type of the distance implies a certain geometry of the clusters one is interested in exploiting when analyzing the data. For instance, it is well known that a commonly used Euclidean distance promotes an ellipsoidal shape of the clusters.

Concentrating on the parameters to be optimized, the above objective function reads now as

$$\text{Min Q with respect to } \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_c \text{ and } U \tag{2}$$

with its minimization carried out for the partition matrix as well as the prototypes. With regard to the prototypes (centroids), we end up with a constraint-free optimization while the other one calls for the constrained optimization. The constraints assure that U is a partition matrix meaning that the following well-known conditions are met

$$\sum_{i=1}^{c} u_{ik} = 1 \quad \text{for all } k = 1, 2, \ldots, N \tag{3}$$

$$0 < \sum_{k=1}^{N} u_{ik} < N \quad \text{for all } i = 1, 2, \ldots, c \tag{4}$$

The choice of the distance function is critical to our primary objective of achieving the transparency of the findings. We are interested in such distances whose equidistant contours are "boxes" with the sides parallel to the coordinates. The Tchebyschev distance (l_∞ distance) is a distance satisfying this property. The boxes are decomposable that is the region within a given equidistant contour of the distance can be treated as a decomposable relation R in the feature space, viz.

$$R = A \times B \tag{5}$$

where A and B are sets (or more generally information granules) in the corresponding feature spaces. It is worth noting that the Euclidean distance does not lead to the decomposable relations in the above sense (as the equidistant regions in such construct are spheres or ellipsoides). The illustration of the decomposability property is illustrated in Fig. 1.

The above clustering problem known in the literature as an l_∞ FCM was introduced and discussed by Bobrowski and Bezdek [3] more than 10 years ago. Some recent generalizations can be found in [7]. The motivation behind the introduction of this type of distance was the one about handling data structures with "sharp" boundaries (clearly the Tchebyschev distance is more suitable with ths regard than the Euclidean distance). The solution proposed in [3] was obtained by applying a basis exchange algorithm.

In this study, as already highlighted, the motivation behind the use of the Tchebyschev distance is different. We are after the description of data structure and the related interpretability of the results of clustering so that the clusters can be viewed as basic models of associations existing in the data. Here, we derive a gradient-based FCM technique enhanced with some additional convergence mechanism.

3 The clustering algorithm – detailed considerations

The FCM optimization procedure is standard to a high extent [2] and consists of two steps: a determination of the partition matrix and calculations of the prototypes. The use of the Lagrange multipliers converts the constrained problem into its constraint-free version. The original objective function (1) is transformed to the form

$$V = \sum_{i=1}^{c} \sum_{k=1}^{N} u_{ik}^2 d_{ik} + \lambda\left(\sum_{i=1}^{c} (u_{ik} - 1)\right) \tag{6}$$

with λ being a Lagrange multiplier. The problem is then solved with respect to each pattern separately that is we consider the relationship below for each data point $(t = 1, 2, \ldots, N)$

$$\frac{\partial V}{\partial u_{st}} = 0 \tag{7}$$

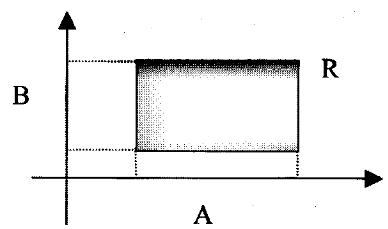

Fig. 1 Decomposability property provided by the Tchebyschev distance; the region of equidistant points is represented as a Cartesian product of two sets in the corresponding feature spaces

$s = 1, 2, \ldots, c$, $t = 1, 2, \ldots, N$. Straightforward calculations lead to the expression

$$u_{st} = \frac{1}{\sum_{j=1}^{c} \frac{d_{st}}{d_{jt}}} \tag{8}$$

The determination of the prototypes is more complicated as the Tchebyschev distance does not lead to a closed-type expression (unlike the standard FCM with the Euclidean distance). Let us start with the objective in which the distance function is spelled out in an explicit manner

$$Q = \sum_{i=1}^{c} \sum_{k=1}^{N} u_{ik}^2 \max_{j=1,2,\ldots,n} |x_{kj} - v_{ij}| \tag{9}$$

The minimization of Q carried out with respect to the prototype (more specifically its t-th coordinate) follows a gradient-based scheme

$$v_{st}(\text{iter} + 1) = v_{st}(\text{iter}) - \alpha \frac{\partial Q}{\partial v_{st}} \tag{10}$$

where α is an adjustment rate (learning rate) assuming positive values. This update expression is iterative; we start from some initial values of the prototypes and keep modifying them following the gradient of the objective function. The detailed calculations of the gradient lead to the expression

$$\frac{\partial Q}{\partial v_{st}} = \sum_{k=1}^{N} u_{sk}^2 \frac{\partial}{\partial v_{st}} \left\{ \max_{j=1,2,\ldots,n} |x_{kj} - v_{sj}| \right\} \tag{11}$$

Let us introduce the following shorthand notation

$$A_{kst} = \max_{\substack{j=1,2,\ldots,n \\ j \neq t}} |x_{kj} - v_{sj}| \tag{12}$$

Evidently, A_{kst} does not depend on v_{st}. This allows us to concentrate on the term that affects the gradient. We rewrite the above expression for the gradient as follows

$$\frac{\partial Q}{\partial v_{st}} = \sum_{k=1}^{N} u_{sk}^2 \frac{\partial}{\partial v_{st}} \{\max(A_{kst}, |x_{kt} - v_{st}|)\} \tag{13}$$

The derivative is nonzero if A_{kst} is less or equal to the second term standing in the expression,

$$A_{kst} \leq |x_{kt} - v_{st}| \tag{14}$$

Next, if this condition holds we infer that the derivative is equal to either 1 or –1 depending on the relationship between x_{kt} and v_{st}, that is –1 if $x_{kt} > v_{st}$ and 1 otherwise.

Putting these conditions together, we get

$$\frac{\partial Q}{\partial v_{st}} = \sum_{k=1}^{N} u_{sk}^2 \begin{cases} -1 & \text{if } A_{kst} \leq |x_{kt} - v_{st}| \text{ and } x_{kt} > v_{st} \\ +1 & \text{if } A_{kst} \leq |x_{kt} - v_{st}| \text{ and } x_{kt} \leq v_{st} \\ 0 & \text{otherwise} \end{cases} \tag{15}$$

The primary concern that arises about this learning scheme is not the one about a piecewise character of the function (absolute value) that is not (a concern that could be easily raised from the formal standpoint) but a fact that the derivative zeroes for a significant number of situations. This may result in a poor performance of the optimization method so it could be easily trapped in case the overall gradient becomes equal to zero. To enhance the method, we relax the binary character of the predicates (less or greater than) standing in (15). These predicates are Boolean (two-valued) as they return values equal to 0 or 1 (which translates into an expression "predicate is satisfied or it does not hold). The modification comes in the form of a degree of satisfaction of this predicate, meaning that we compute a multivalued predicate

$$\text{Degree}(a \text{ is included in } b) \tag{16}$$

that returns 1 if a is less or equal to b. Lower values of the degree arise when this predicate is not fully satisfied. This form of augmentation of the basic concept was introduced in [5, 6, 13, 14] in conjunction to studies in fuzzy neural networks and relational structures (fuzzy relational equations).

The degree of satisfaction of the inclusion relation is equal to

$$\text{Degree}(a \text{ is included in } b) = a \to b \tag{17}$$

where a and b are in the unit interval. The implication $\to$ operation is a residuation operation, cf. [13, 14]. Here we consider a certain implementation of such operation where the implication is implied by the product t-norm, namely

$$a \to b = \begin{cases} 1 & \text{if } a \leq b \\ b/a & \text{otherwise} \end{cases} \tag{18}$$

Using this construct, we rewrite (15) as follows

$$\frac{\partial Q}{\partial v_{st}} = \sum_{k=1}^{N} u_{sk}^2 \begin{cases} -(A_{kst} \to |x_{kt} - v_{st}|) & \text{if } x_{kt} > v_{st} \\ (A_{kst} \to |x_{kt} - v_{st}|) & \text{if } x_{kt} \leq v_{st} \end{cases} \tag{19}$$

In the overall scheme, this expression will be used to update the prototypes of the clusters (10).

Summarizing, the clustering algorithm arises as a sequence of the following steps

repeat

- compute partition matrix using (8);
- compute prototypes using the partition matrix obtained in the first phase. (It should be noted that the partition matrix does not change at this stage and all updates of the prototypes work with this matrix. This phase is more time consuming in comparison with the FCM method equipped with the Euclidean distance)

until a termination criterion satisfied

Both the termination criterion and the initialization of the method are standard. The termination takes into account changes in the partition matrices at two successive iterations that should not exceed a certain threshold level. The initialization of the partition matrix is random.

As an illustrative example, we consider a synthetic data involving 4 clusters, see Fig. 2. The two larger data groupings consist of 100 data-points and the two smaller ones have 20 and 10 data-points respectively.

Table 1 gives a representative set of clustering results for 2 to 8 clusters. As expected, the two larger data groupings exercise dominant influence on the outcome of the FCM algorithms. Both Euclidean and Tchebyschev distance based FCM exhibit robust performance in that they find approximately the same clusters in their successive runs (within the limits of the optimization convergence criterion). While most of the identified prototypes fall within the large data groupings, the Tchebyschev distance based FCM consistently manages to associate a prototype with one of the smaller data grouping (underlined in the table). This is clearly a very advantageous feature of our modified FCM algorithm and confirms our assertion that the objective of enhancing the interpretability of data through the identification of decomposable relations is enhanced with Tchebyschev distance based FCM.

The above results are better understood if we examine the cluster membership function over the entire pattern space. The visualization of the membership function for one of the two clusters, positioned in the vicinity of (0.2, 0.2), ($c = 2$) is given in Fig. 3.

It is easily noticed that that for higher values of the membership grades (e.g. 0.9), the shape of contours is rectangular. This changes for lower values of the membership grades when we witness a gradual departure from this geometry of the clusters. This is an effect of interaction between the clusters that manifests itself in a deformation of the original rectangles. The deformation depends on the distribution of the clusters, their number and a specific threshold β being selected. The lower is the value of this threshold, the more profound departure from the rectangular shape. For higher values of β such deformation is quite limited. This suggests that when using high values of the threshold level the rectangular (or hyperbox) form of the core part of the clusters is fully legitimate.

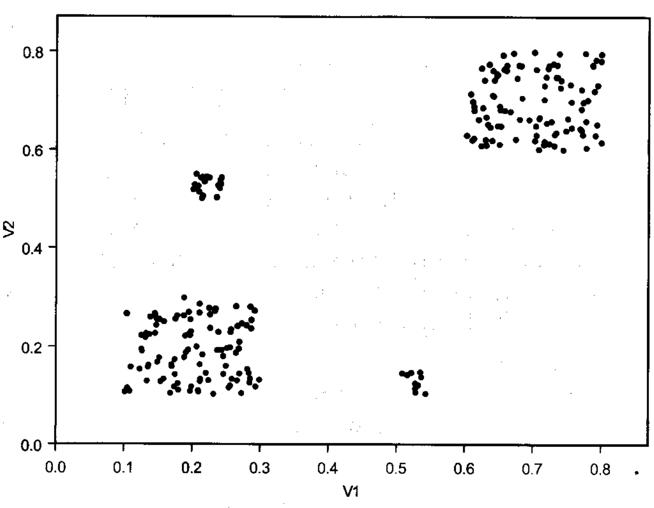

Fig. 2 Two-dimensional synthetic data with four visible clusters of unequal size

Let us contrast these results with the geometry of the clusters constructed when using a Euclidean distance. Again, we consider two prototypes, as identified by the Euclidean distance based FCM, see Fig. 4. The results are significantly different: the clusters are close to the Gaussian-like form and do not approximate well by rectangular shapes.

The above effect is even more pronounced when there are more clusters interacting with each other. We consider 8 prototypes identified by the two FCM algorithms, see Fig. 5. In the case of Chebyschev FCM, it is clear that despite strong interactions between the clusters, the rectangular shape of the cluster membership function is preserved for a range of values of this function. These undistorted rectangles cover a good pro-

Table 1 Prototypes identified by two FCM algorithms, with Euclidean and Tchebyschev distance measure respectively, for the varying number of clusters (the underlined prototypes correspond to the smaller data groupings)

Number of clusters	Prototypes for FCM with	
	Euclidean distance	Tchebyschev distance
2	0.6707 0.6706	0.2088 0.1998
	0.2240 0.2236	0.6924 0.6831
3	0.2700 0.3011	0.7000 0.6847
	0.6875 0.6841	0.24400.4914
	0.2302 0.2127	0.2124 0.1852
4	0.2255 0.2035	0.7261 0.7377
	0.2323 0.2479	0.22780.5178
	0.6872 0.6814	0.2092 0.1846
	0.6533 0.6588	0.6523 0.6498
5	0.2525 0.2784	0.2189 0.1451
	0.2282 0.2014	0.22720.5188
	0.6721 0.6757	0.1960 0.2258
	0.2343 0.2389	0.6568 0.6868
	0.6919 0.6841	0.7268 0.6593
6	0.2329 0.2562	0.7469 0.6650
	0.6809 0.6777	0.2151 0.1364
	0.6857 0.6830	0.6570 0.6840
	0.2272 0.2206	0.6570 0.6840
	0.2261 0.2008	0.2619 0.2648
	0.6447 0.6500	0.1945 0.2239
7	0.6646 0.6697	0.1967 0.2255
	0.7036 0.6619	0.2200 0.1450
	0.6993 0.7100	0.7278 0.6594
	0.2395 0.5019	0.22770.5183
	0.2382 0.1935	0.3976 0.4051
	0.2164 0.1955	0.6099 0.6117
	0.2271 0.2018	0.6588 0.6923
8	0.6962 0.6892	0.6607 0.7615
	0.23980.5088	0.2122 0.1327
	0.2360 0.1980	0.3209 0.3097
	0.2441 0.2203	0.6565 0.6830
	0.6962 0.6882	0.7267 0.6590
	0.6850 0.6756	0.6460 0.6492
	0.2385 0.1942	0.22770.5191
	0.2166 0.1965	0.2108 0.2249

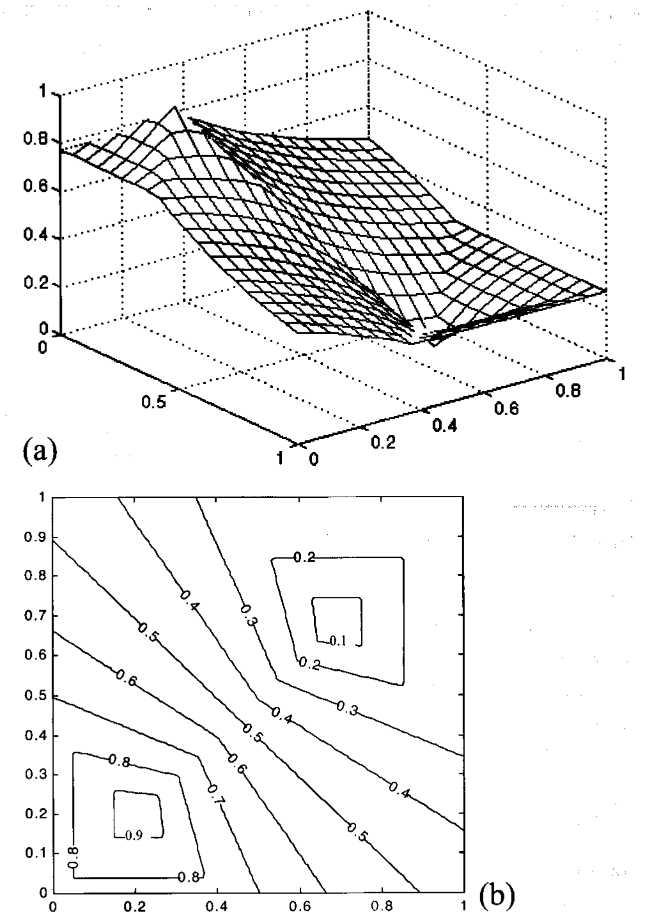

Fig. 3 Visualization of the first cluster (membership function) centered around (0.2088 0.1998): **a** 3D space and **b** contour plots

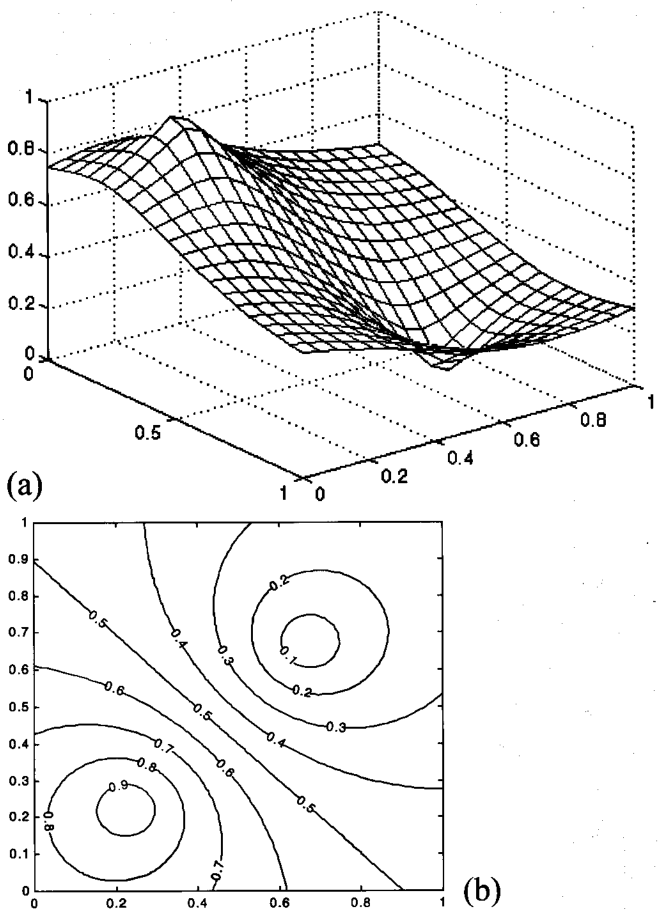

Fig. 4 Visualization of the first cluster (membership function) centered around (0.2240 0.2236): **a** 3D space and **b** contour plots. The Euclidean distance function was used in the clustering algorithm

portion of the original data, which is represented by the selected prototype. On the other hand, the Euclidean FCM results in contours of the membership function that are undistorted circles only in the very close proximity of the prototype itself. Thus the task of linking the original data with the prototype representing an association existing in the data is quite difficult for most of the data points.

4 Generation of granular prototypes

As we are dealing with the Tchebyschev metric, its underlying geometry promotes rectangular shape of the information granules. We formalize here how an information granule can be seen as a union of appropriately constructed hyperboxes. The crux of this construction is to move around the prototype by changing only a single feature. The moves are made separately toward higher and lower values (with the reference to the prototype) of the feature, see Fig. 6.

If we fix a certain threshold value (β), the resulting rectangle captures the corresponding part of the data set. The threshold values are in the unit interval. The development of the box observes the following scheme: move gradually from the prototype along one of the directions (as indicated in Fig. 6), compute the membership grade $u(\mathbf{x},\mathbf{v})$ (see (8)) until it becomes equal to β. Note that as we are moving away from the prototype, this membership grade continuously decreases. Once we reach the threshold, the move stops and the corresponding $\mathbf{x}$ is treated as the face of the box (rectangle). The construction is carried out for all remaining directions. The final result is denoted by $\mathbf{B}_\beta$.

Obviously, higher values of β confine the core part of the cluster and produce smaller information granules of the core of data structure. The resulting hyperbox is tied to the threshold level through an evident relationship

$$\mathbf{B}_\beta = \{\mathbf{x} \in [0,1]^n | u(\mathbf{x},\mathbf{v}) \geq \beta\} \tag{20}$$

where $\mathbf{B}_\beta$ is a hyperbox induced with the threshold equal to β.

Second, for each β, the corresponding hyperbox is a relation in the feature space. When looking at these relations globally (considering varying values of β), we can represent them as a fuzzy relation, meaning that $\mathbf{B}_\beta$

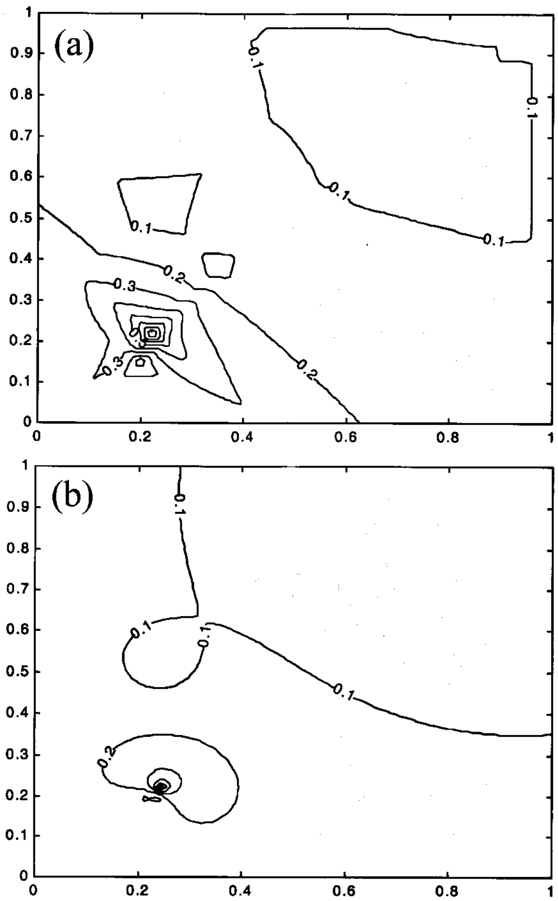

Fig. 5 Contour plots for one of the 8 clusters (membership function) centered around (0.2108 0.2248) for the Thebyschev distance **a**; and (0.2441 0.2203) for the Euclidean distance **b**

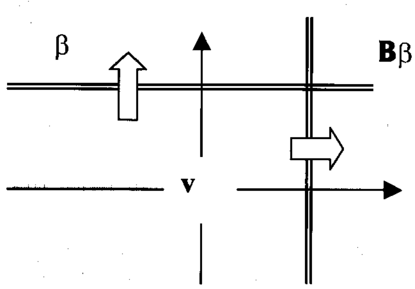

Fig. 6 Generation of information granules (2-dimensional case); the region around the prototype is constructed by analyzing the membership grades of the clusters

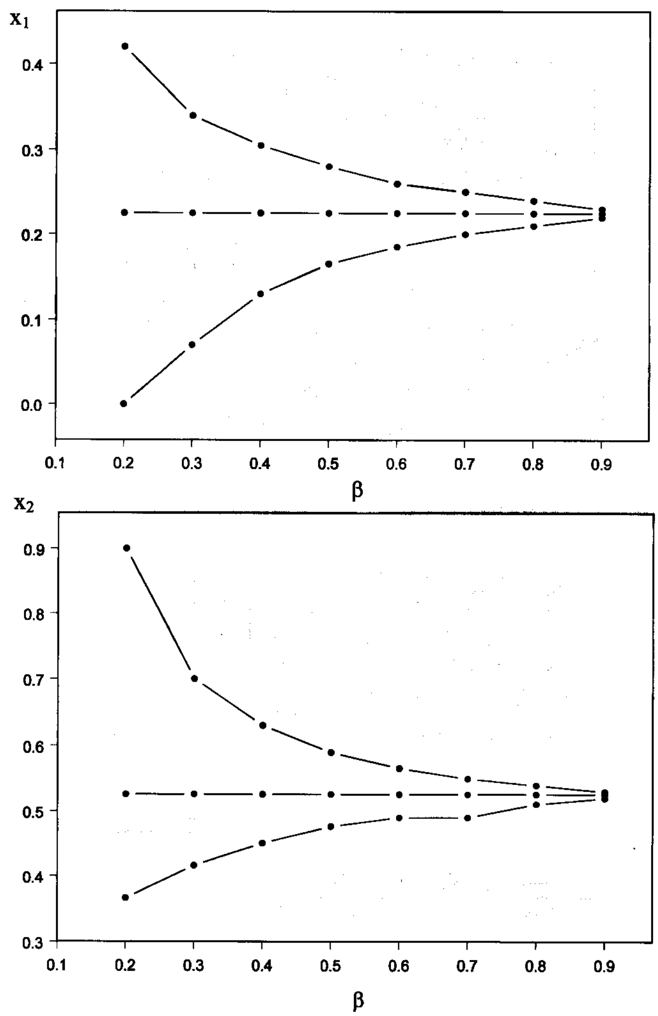

Fig. 7 Coordinates of the information granules (x_1 and x_2) of a selected cluster (0.2277, 0.5191) for clustering with $c = 8$ clusters

is a fixed β-cut of it. We then have the following relationship

$$\mathbf{B} = \bigcup_{\beta} \mathbf{B}_{\beta} \tag{21}$$

where **B** is a fuzzy relation of the core of the data structure. To make the core meaningful, the threshold value should be high enough so that the only essential part of the data set becomes qualified as the core. There is also another aspect to this design, that is a deformation of the hyperboxes in the feature space caused by an interaction between the clusters. Again, with lower values of β, the interaction tends to be more critical leading to more profound deformations of the hyperbox.

Following the discussion of the numeric example, a representative information granule (box), taken from the set of prototypes calculated for $c = 8$ clusters, is shown in Fig. 7. Evidently, the box shrinks with the increasing values of the threshold; at some point the changes to the size of the granule become very small.

5 The geometry of information granules

As the contour plots of the clusters reveal (Fig. 5), interaction between the clusters become responsible for the deformation of the hyperbox shape of the cores. This poses an interesting question as to the size of the core structure of the data. The choice of the threshold level (β) needs to be controlled by an acceptable level of deformation of the equidistant lines of the Tchebyschev distance. We require a quantification of such deformation effect. This can be done by finding differences between the "theoretical" values of the membership (dictated by the Tchebyschev metric) and those resulting from the calculations of the membership grades based on the prototypes. The details follow the notation in Fig. 8.

In the two-dimensional case, we identify four corner points of the box implied by the fixed threshold β. This means that all four of them belong to the information granule at the membership grade equal to β. Once using (8) with $\mathbf{x} = \mathbf{x}(1), \mathbf{x}(2), \ldots, \mathbf{x}(4)$ and the prototype $\mathbf{v}$, the calculated membership grades could be different from this threshold, say $u(1)$, $u(2)$, $u(3)$, $u(4)$. Let us consider the sum of differences

$$D = |\beta - u(1)| + |\beta - u(2)| + |\beta - u(3)| + |\beta - u(4)| \quad (22)$$

Since $u = \beta$ was calculated only in 'white' points (by virtue of construction (20)) and it ideally should also be satisfied by 'gray' points, which are the vertices spanned by the original vectors $\mathbf{x}(1)$, $\mathbf{x}(2)$, $\mathbf{x}(3)$ and $\mathbf{x}(4)$)

Therefore (22) serves as a useful measure of deformation of the rectangular shapes of the granules. The above construct easily expands to any dimension of the feature space; evidently for "n" features, a search is completed for all corners of the hypercube, that is 2^n.

Continuing the numeric example, we quantify the deformation of the boxes by means of (22). The approximation of the resulting dependency between the deformation measure D, viewed as a function of β, is done through a polynomial fit. A representative set of results is given in Fig. 9. The assessment of information granules was carried out for every granule identified with 2 to 8 clusters. It is evident that the deformation of the hyperboxes can be approximated by a low-order polynomial.

6 Granular data description: a general model

With the development of the granular prototypes guided by the clustering algorithm, we can concisely describe the data in the form

$$D = \mathbf{B}_1 \cup \mathbf{B}_2 \cup \ldots \mathbf{B}_c \cup R \quad (23)$$

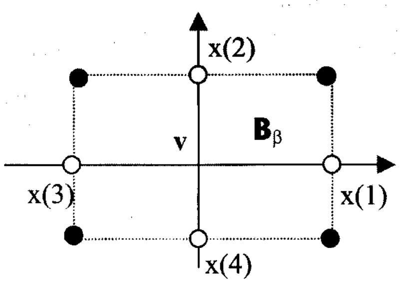

Fig. 8 A concept of a deformation index – expressing departure from the hyperbox nature of cores implied by the Tchebyschev distance

where D is a data set under discussion, $\mathbf{B}_i$ are granular prototypes and R is a residual structure of the data set. The brief characteristics of the this structure are summarized in the tabular form

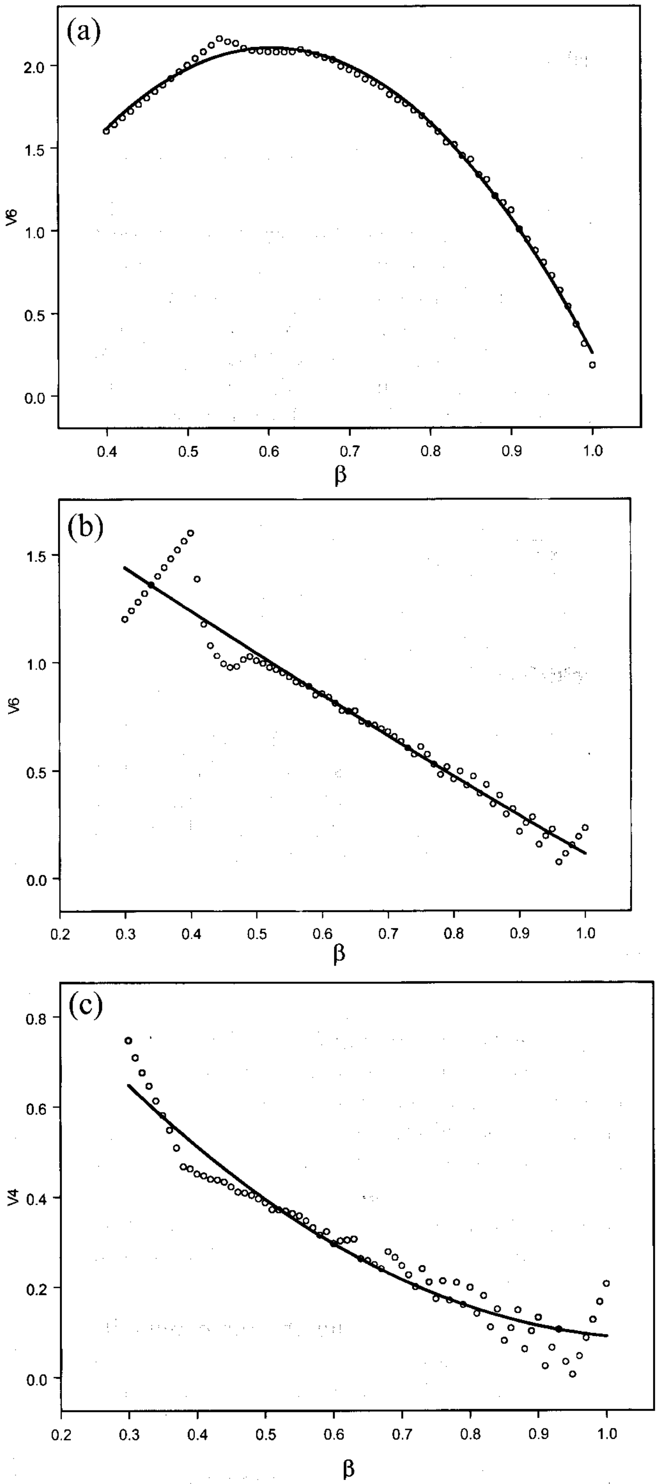

Fig. 9 Quantification of the deformation effect D for: **a** prototype v6 obtained with $c = 8$; **b** prototype v6 obtained with $c = 6$; and **c** prototype v4 obtained with $c = 4$

Data structure- descriptor	Description	Main features
hyperboxes	capture the essence of the data. Serve as focal "points" of the structure. Can be easily interpreted as relations and	Hyperboxes constructed on the basis of the Tchebyschev distance for a given threshold level. The choice of this level depends on an acceptable level of deformation of the hyperboxes allowed in the structure. The hyperboxes can be decomposed into intervals for individual features
Residual	Deals with the remaining data points and concentrates on patterns not clearly identified as focal points of the data structure	A lack of explicit description, more complex geometry of the regions. The degree of membership to the hyperbox is expressed through a formula

Any element **x** in the data space of interest can be characterized as belonging to the core or being identified with the residual portion of the data. The membership to the core is binary – we identify the hyperbox to which this membership happens. If **x** is identified to be a part of R, the degree of membership to the hyperbox is determined through the standard membership expression in which we now consider a distance between a data point (**x**) and a set (relation) $\mathbf{V}_i$

$$u_i(\mathbf{x}) = \frac{1}{\sum_{j=1}^{c} \frac{d(\mathbf{x},\mathbf{V}_i)}{d(\mathbf{x},\mathbf{V}_j)}} \tag{23}$$

7 Conclusions

In the description of data, we have developed two main components, namely cores of the data that are well-structured in the form of hyperboxes in the feature space and a far less regular structure that is described analytically through an expression for membership grades but does not carry any clear geometric interpretation. The computing backbone of this approach is based on the well-known FCM technique equipped with the Tchebyschev distance. We introduced a new way of optimizing the prototypes in this method that uses a gradient-based technique augmented by a logic-oriented mechanisms of gradient determination. The geometry and design of the hyperbox information granules have been discussed along with an important aspect of deformation of such granules. Furthermore a quantification of this effect is discussed.

The proposed approach to data analysis can be exploited in many different ways. A few options worth pursuing are as follows

- Data mining. Considering the main pursuit of data mining articulated in the language of well-defined, semantically sound and easily interpretable constructs, the information granules envisioned in this way are legitimate entities one dwell data mining activities. They are easy to interpret and thus cope with the underlying structure of data while leaving out the residual portion of data not exhibiting strong patterns of dependencies
- In any modeling pursuit, the above data description helps concentrate on the design of local models assigned to the core parts. The residual part of data can be handled separately with an anticipation that these data points may not lead to a model with a strongly manifested character
- In classification problems, the core part of the data implies a collection of simple classifiers while the residual part invokes more demanding and conceptually advanced classifiers such as neural network

Acknowledgements Support from the Engineering and Physical Sciences Research Council (UK), the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Alberta Consortium of Software Engineering (ASERC) is gratefully acknowledged.

References

1. Bargiela A (2001) Interval and Ellipsoidal Uncertainty Models. In: Pedrycz W. (ed.) Granular Computing, Springer Verlag, Heidelberg, Germany
2. Bezdek JC (1981) Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press, NY
3. Bobrowski L, Bezdek JC (1991) C-Means clustering with the l_1 and l_∞ norms. IEEE Trans Sys Man and Cybernetics 21:545–554.
4. Cios K, Pedrycz W, Swiniarski R (1998) Data Mining Techniques. Kluwer Academic Publishers, Boston
5. Dubois D, Prade H (1995) Fuzzy relation equations and causal reasoning. Fuzzy Sets and Systems 75:119–134
6. Gottwald S (1995) Approximate solutions of fuzzy relational equations and a characterization of t-norms that define metrics for fuzzy sets. Fuzzy Sets and Systems 75:189–201
7. Groenen PJF, Jajuga K (2001) Fuzzy clustering with squared Minkowski distances. Fuzzy Sets and Systems 120:227–237
8. Höppner F, Klawonn F, Kruse R, Runkler T (1999) Fuzzy Cluster Analysis. J. Wiley, Chicester
9. Kandel A (1986) Fuzzy Mathematical Techniques with Applications. Addison-Wesley, Reading, MA
10. Maimon O, Kandel A, Last M (2001) Information-theoretic fuzzy approach to data reliability and data mining. Fuzzy Sets and Systems. 117:183–194
11. Pawlak Z (1991) Rough Sets: Theoretical Aspects of Reasoning about Data. Kluwer Academic, Dordrecht
12. Pedrycz W (1997) Computational Intelligence: An Introduction. CRC Press, Boca Raton, FL
13. Pedrycz W, Gomide F (1998) An Introduction to Fuzzy Sets. Cambridge, MIT Press, Cambridge, MA
14. Pedrycz W (1989) Fuzzy Control and Fuzzy Systems. RSP/ Wiley Letchworth, England

15. Zadeh LA (1997) Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. Fuzzy Sets and Systems 90: 111–117

16. Zadeh LA (1999) From computing with numbers to computing with words-from manipulation of measurements to manipulation of perceptions. IEEE Trans Circuits and Systems 45: 105–119