Clustering Interval-valued Proximity Data using Belief Functions

Marie-Hélène Masson[1*†] and Thierry Denœux[†]

[*] Université de Picardie Jules Verne
[†] Université de Technologie de Compiègne
UMR CNRS 6599 Heudiasyc
BP 20529 - F-60205 Compiègne cedex - France

July 10, 2016

[1]Corresponding author. Email: mmasson@hds.utc.fr. Fax: (33) 3 44 23 44 77.

Abstract

The problem of clustering objects based on interval-valued dissimilarities is tackled in the framework of the Dempster-Shafer theory of belief functions. The proposed method assigns to each object a basic belief assignment (or mass function) defined on the set of clusters, in such a way that the belief and the plausibility that any two objects belong to the same cluster reflect, respectively, the observed lower and upper dissimilarity values. Experiments with synthetic and real data sets demonstrate the ability of the method to detect meaningful clusters, even in the presence of imprecise data and outliers.

Keywords: Relational Data, Clustering, Unsupervised Learning, Dempster-Shafer theory, Evidence theory, Belief Functions, Multi-dimensional Scaling.

1 Introduction

Cluster analysis refers to a wide range of numerical methods for discovering groups in data. The groups are defined as subsets of more or less "similar" objects [16]. Many clustering methods start from an $n \times n$ matrix Δ, the elements of which are quantitative measurements of closeness or dissimilarity between objects. The dissimilarity is either directly available, or derived from a $n \times p$ multivariate matrix of attributes. Reviews of classical methods for clustering such relational or proximity data, including hard or fuzzy methods, may be found in [2, chapter 3]. All these methods assume that the dissimilarity δ_{ij} between objects i and j is expressed as a real number reflecting with certainty the closeness of the two objects. However, in certain situations, the dissimilarity between i and j is only known to lie within a certain interval $[\delta_{ij}^-; \delta_{ij}^+]$. Here are some examples of such situations:

- The initial data may consist in interval-valued feature vectors whose components are intervals. Such a data type may be useful to describe a set of entities or the range of a variable observed during a certain period. It may be also a good way to summarize a large amount of data in data mining applications (see, for instance, [13]).
- The dissimilarities may be directly elicited from human evaluators who have some difficulty in precisely quantifying the proximity of two objects. An interval rather than a real value may be more suitable to account for the vagueness of the evaluation.
- The dissimilarities may be measured independently by several sensors, so that the available information concerning the dissimilarity between any two objects takes the form of an empirical distribution. One way to analyze such data is to describe the distribution by an interval such as the interquartile range or the mean plus or minus the standard deviation.

In recent years, a great deal of attention has been paid to the analysis of imprecise or fuzzy data. Several references may be found in the literature focusing on inferential statistics, regression or classification [14] [12] [15][7] [4]. For representing interval-valued dissimilarities [9] or fuzzy dissimilarities [17] in a low dimensional space, we have proposed several extensions of classical multidimensional scaling methods (MDS). Regarding cluster analysis, there have been only few studies. In the framework of Symbolic Data Analysis [13], a hierarchical divisive clustering method has been proposed based on a generalized within-cluster inertia criterion [5]. Chavent and Lechevallier have also proposed a c-means-like algorithm for interval data [6]. In [1], a "linguistic" version of the popular fuzzy c-means algorithm (FCM) is presented. A FCM-like procedure for symbolic data has also been described by El Sonbaty and Ismail [23]. These methods do not start from a dissimilarity matrix, but from interval-valued or fuzzy attributes describing a set of objects; a generalized similarity measure between objects or groups of objects si then defined, and a classical clustering procedure is applied.

The method presented in this paper uses a different approach. It generalizes a clustering algorithm for crisp relational data using belief function theory introduced in [10][11]. In this approach, the membership of an object to a cluster or a subset of clusters is represented by a "mass of belief". Belief masses are determined in such a way that the belief and the plausibility that any two objects belong to the same class reflect, respectively, their observed lower and upper dissimilarities. The possibility to assign masses not only to single classes but also to subsets of classes or to the empty set makes this partitioning model more general than the classical hard or fuzzy ones. The method allows to gain a deeper insight into the structure of the data, as will be shown by experimental results.

The rest of the paper is organized as follows. First, the necessary background on belief functions theory is recalled in Section 2. Section 3 shows how to apply these general concepts to the problem of clustering interval-valued relational data. Finally, experimental results, with synthetic and real datasets, are described in Section 4. Section 5 concludes this paper.

2 Evidence theory

Let Ω denote a finite and unordered set, called the frame of discernment, composed of c singletons or atomic propositions ω_k, $k = 1, c$. A *basic belief assignmen*t (bba) on Ω [19, 22] is a function m from the power set 2^Ω to [0,1] verifying:

$$\sum_{A \subseteq \Omega} m(A) = 1. \tag{1}$$

The term $m(A)$ is called the basic belief mass (bbm) given to A. It represents the part of belief that supports A witout supporting any more specific subset. The subsets A from Ω such that $m(A) > 0$ are *the focal sets* of m. One of the main attractive features of belief functions is that they provide a nice and flexible way to represent uncertainty, from complete ignorance ($m(\Omega) = 1$) to complete knowledge ($m(\{\omega_k\}) = 1$ for some k). A bba such that $m(\emptyset) = 0$ is said to be normal. This condition was originally imposed by Shafer [19] but it may be relaxed if one accepts the *open-world assumption* stating that the set Ω might not be complete [20] and that the true proposition might not lie in Ω. Given a bba m, one can define a *belief function* $\text{bel} : 2^\Omega \mapsto [0, 1]$ as:

$$\text{bel}(A) \triangleq \sum_{\emptyset \neq B \subseteq A} m(B) \quad \forall A \subseteq \Omega \ , \tag{2}$$

and a dual *plausibility function* $\text{pl} : 2^\Omega \mapsto [0, 1]$ as

$$\text{pl}(A) \triangleq \sum_{B : A \cap B \neq \emptyset} m(B) = \text{bel}(\Omega) - \text{bel}(\overline{A}) \quad \forall A \subseteq \Omega \ , \tag{3}$$

where $\overline{A}$ denotes the complement of A. The degree of belief $\text{bel}(A)$ (also called the *credibility* of A) quantifies the amount of *specific* support given to A. It is obtained by adding the masses allocated to subsets of A. The degree of plausibility $\text{pl}(A)$ quantifies the amount of *potential* support given to A, obtained by adding the masses given to propositions B compatible with A ($B \cap A \neq \emptyset$). Eq. (3) reflects the complementarity of these two functions: the more A is plausible, the less $\overline{A}$ is credible; conversely, the more A is credible, the less $\overline{A}$ is plausible.

Let us now assume that we have two bba's m_1 and m_2 representing distinct items of evidence. The standard way of combining them is through the conjunctive sum operation $\bigcirc\!\!\!\!\cap$ (also called unnormalized Dempster's rule of combination), defined as:

$$(m_1 \bigcirc\!\!\!\!\cap\, m_2)(A) \triangleq \sum_{B \cap C = A} m_1(B) m_2(C) \ , \tag{4}$$

for all $A \subseteq \Omega$. The quantity

$$K \triangleq (m_1 \textcircled{\cap} m_2)(\emptyset) = \sum_{B \cap C = \emptyset} m_1(B) m_2(C) \tag{5}$$

is called the *degree of conflict* between m_1 and m_2. It may be seen as a degree of disagreement between the two information sources.

Let us now consider two bba's m^{Ω_1} and m^{Ω_2} defined on two different frames of discernement Ω_1 and Ω_2 (from now on, the frame of a bba will be indicated as a superscript when necessary). To combine these two bba's, it is necessary to extend their respective frame of discernment to the Cartesian product $\Omega = \Omega_1 \times \Omega_2$. This extension, referred to as the *vacuous extension* [19, 21], can be seen as the reverse operation of marginalization. The vacuous extension of m^{Ω_1} on Ω is defined for all $B \subseteq \Omega$ as:

$$m^{\Omega_1 \uparrow \Omega}(B) \triangleq \begin{cases} m^{\Omega_1}(A) & \text{if } B = A \times \Omega_2 \text{ for some } A \subseteq \Omega_1, \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

This definition of the vacuous extension results from the Principle of Minimal Commitment [21], which formalizes the idea that one should never give more support than justified to any proposition. A similar expression can be easily obtained for the vacuous extension of m^{Ω_2} on Ω. The conjunctive sum of m^{Ω_1} and m^{Ω_2} may be then obtained by combining their vacuous extensions on Ω, using (4). We thus obtain:

$$(m^{\Omega_1} \textcircled{\cap} m^{\Omega_2})(A \times B) = m^{\Omega_1}(A) m^{\Omega_2}(B) \ , \tag{7}$$

for all non empty subsets $A \subseteq \Omega_1$ and $B \subseteq \Omega_2$, the remaining mass being assigned to $\emptyset$.

3 Proposed method

3.1 Preliminaries

In this section it is proposed to use the concepts of belief functions theory to derive a partition of a collection $O = \{o_1, o_2, ..., o_n\}$ of n objects into c classes from interval-valued dissimilarities. The frame of discernment $\Omega = \{\omega_1, ..., \omega_c\}$ is composed of the c classes.

Let us assume that we have partial information concerning the class membership of each object o_i, and this information is represented by a bba m_i with focal sets in $\mathcal{F} = \{\{\omega_1\}, ..., \{\omega_c\}, \emptyset, \Omega\}$ (this restriction is imposed for the sake of simplicity, but it could easily be relaxed). Let m_{ik}, $m_{i\Omega}$ and $m_{i\emptyset}$ denote, respectively, the masses allocated to the singleton $\{\omega_k\}$, Ω and the empty set. The interpretation of the different masses will be given below. The problem is to find a way to infer the m_i on the basis on the input dissimilarities.

Consider two object o_i and o_j and two bba's m_i and m_j quantifying one's belief regarding their respective class. Let S denote the proposition "o_i and o_j belong to the same class". This proposition is not defined on Ω but on the product space $\Omega^2 = \Omega \times \Omega$ by:

$$S = \{(\omega_1, \omega_1), (\omega_2, \omega_2), \ldots, (\omega_c, \omega_c)\}.$$

The plausibility and the credibility of S may be determined from m_i and m_j. For that purpose, it is necessary to compute the vacuous extensions of m_i and m_j using (6), and combine them in Ω^2 using (7). Let $m_{i \times j}$ denote the result of this combination. We have:

$$m_{i \times j}(A \times B) = m_i(A) \cdot m_j(B), \qquad \forall A, B \subseteq \Omega, A \neq \emptyset, B \neq \emptyset \ . \tag{8}$$

Let $\mathrm{pl}_{i\times j}$ and $\mathrm{bel}_{i\times j}$ be, respectively, the plausibility and the belief function associated to $m_{i\times j}$. We have:

$$\begin{aligned}\mathrm{bel}_{i\times j}(S) &= \sum_{\{A\times B\subseteq\Omega^2 \mid \emptyset\neq(A\times B)\subseteq S\}} m_{i\times j}(A\times B) \\ &= \sum_{k=1}^{c} m_{ik}m_{jk} \ , \end{aligned} \tag{9}$$

and

$$\begin{aligned}\mathrm{pl}_{i\times j}(S) &= \sum_{\{A\times B\subseteq\Omega^2 \mid (A\times B)\cap S\neq\emptyset\}} m_{i\times j}(A\times B) \\ &= \sum_{A\cap B\neq\emptyset} m_i(A)\cdot m_j(B) \\ &= \mathrm{bel}_{i\times j}(S) + m_{i\Omega}\sum_{k=1}^{c} m_{jk} + m_{j\Omega}\sum_{k=1}^{c} m_{ik} + m_{i\Omega}m_{j\Omega} \ . \end{aligned} \tag{10}$$

Remembering that all the masses sum up to 1, $\mathrm{pl}_{i\times j}(S)$ can more conveniently be written in the following way:

$$\begin{aligned}\mathrm{pl}_{i\times j}(S) &= 1 - \sum_{A\cap B=\emptyset} m_i(A)\cdot m_j(B) \\ &= 1 - K_{ij}^{-}, \end{aligned} \tag{11}$$

where K_{ij}^{-} is the degree of conflict between m_i and m_j (see Eq. 5). By analogy, we can write, from Eq. (10) and Eq. (11):

$$\mathrm{bel}_{i\times j}(S) = 1 - K_{ij}^{+}, \tag{12}$$

with K_{ij}^{+} defined as:

$$\begin{aligned}K_{ij}^{+} &= K_{ij}^{-} + m_{i\Omega}\sum_{k=1}^{c} m_{jk} + m_{j\Omega}\sum_{k=1}^{c} m_{ik} + m_{i\Omega}m_{j\Omega}, \\ &= 1 - \sum_{k=1}^{c} m_{ik}m_{jk}. \end{aligned} \tag{13}$$

Remark 1 K_{ij}^{-} represents the degree of conflict between m_i and m_j as defined classically in evidence theory. K_{ij}^{+} reveals, not only the part of conflict that is certain (K_{ij}^{-}), but also a part of "potential" conflict due to the uncertainty represented by the mass allocated to Ω. In this sense, K_{ij}^{-} and K_{ij}^{+} may be considered, respectively, as the minimum and the maximum bounds of the conflict between m_i and m_j.

3.2 Fundamental principle

Let $\delta_{i,j}^{-}$ and $\delta_{i,j}^{+}$ denote the lower and upper bounds for the unknown true dissimilarity between any two objects o_i and o_j. As already remarked in Section 1, objects within a group are more likely to be close to each other than objects of different groups. Two pieces of information about the proximity between the objects are available, namely δ^-_{ij} and δ^+_{ij}, which have a completely different meaning: if δ^-_{ij} is near 0, it is possible for the two objects to be very close from each other, which makes the membership to the same cluster plausible. But, at the same time, if δ^+_{ij} is large, it is also plausible that they do not belong to the same cluster, or equivalently, because of the duality bel-pl (see Eq (3)), it is not very credible that they belong to the same cluster. This fundamental and rather intuitive principle may be summarized as follows:

- the smaller δ^-_{ij}, the more plausible is the proposition "object o_i and o_j belong to the same cluster";
- the greater δ^+_{ij}, the less credible is the proposition "object o_i and o_j belong to the same cluster".

Given any two pairs of objects (o_i, o_j) and $(o_{i'}, o_{j'})$, it is then natural to impose the following conditions:

$$\begin{cases} \delta^-_{ij} > \delta^-_{i'j'} \Rightarrow \mathrm{pl}_{i\times j}(S) \leq \mathrm{pl}_{i'\times j'}(S) \\ \delta^+_{ij} > \delta^+_{i'j'} \Rightarrow \mathrm{bel}_{i\times j}(S) \leq \mathrm{bel}_{i'\times j'}(S) \end{cases} \tag{14}$$

or, equivalently:

$$\begin{cases} \delta^-_{ij} > \delta^-_{i'j'} \Rightarrow K^-_{ij} \geq K^-_{i'j'} \\ \delta^+_{ij} > \delta^+_{i'j'} \Rightarrow K^+_{ij} \geq K^+_{i'j'} \end{cases} \tag{15}$$

Remark 2 In the former version of the algorithm where dissimilarities were assumed to be crisp numbers [11], only one condition was required:

$$\delta_{ij} > \delta_{i'j'} \Rightarrow \mathrm{pl}_{i\times j}(S) \leq \mathrm{pl}_{i'\times j'}(S).$$

3.3 Inferring bba's from dissimilarities

Let $M = (m_1, m_2, ..., m_n)$ denote the n-tuple of bba's related to the n objects. Such a structure is called a *credal partition* in [11]. The fundamental principle introduced in the previous section provides a basis for inferring M from the input data $\Delta = ([\delta^-_{ij}; \delta^+_{ij}])$. Given observed interval-valued dissimilarities Δ, we want to find a credal partition M such that constraints (15) are at least approximately satisfied. As explained in [11], a similar problem is addressed by multidimensional scaling methods. Such methods attempt to find a d-dimensional configuration of points (with $d \ll n$) such that the interpoint distances match as well as possible the input dissimilarities. For that purpose, a suitable criterion (also called stress function) is introduced to measure the discrepancy between the distances and the dissimilarities. This criterion is then minimized with respect to the configuration of points. Among various existing methods, one can distinguish between the *non metric* and the *metric* approaches. In the first one, only the rank order of the dissimilarities are considered, and one identifies the relationship between dissimilarities and distances using isotonic regression, an ordinal regression method. In the *metric* ones, a linear, parameterized relationship is assumed. The first method is more general but may be quite heavy computationally. The interested reader may refer to [3] for a complete description of various existing MDS methods.

In our problem, the masses of the different objects can be seen as the coordinates of the objects in a $(c+2)$-dimensional space, and the quantities K_{ij}^- and K_{ij}^+ may be seen as minimum and maximum "pseudo-distances" between objects. Condition (15) defines an *ordinal* relation between dissimilarities and pseudo-distances, a relation that can be determined through isotonic regression. However, for computational reasons, the simplest and more restrictive metric approach was chosen in this paper. We assume an affine relationship and we propose to minimize the following criterion:

$$J_1(M,a,b) \triangleq \frac{1}{C_1}\sum_{i<j}(aK_{ij}^- + b - \delta_{ij}^-)^2 \; + \frac{1}{C_2}\sum_{i<j}(aK_{ij}^+ + b - \delta_{ij}^+)^2 \tag{16}$$

where C_1 and C_2 are two normalizing constants defined as:

$$C_1 = \sum_{i<j} {\delta_{ij}^-}^2, \tag{17}$$

and

$$C_2 = \sum_{i<j} {\delta_{ij}^+}^2. \tag{18}$$

Alternatively, if one wishes to give a smaller weight to imprecise input data (i.e., large $\delta_{ij}^+ - \delta_{ij}^-$), one may minimize the following criterion:

$$J_2(M,a,b) \triangleq \frac{1}{C}\sum_{i<j}\frac{(aK_{ij}^- + b - \delta_{ij}^-)^2}{(\delta_{ij}^+ - \delta_{ij}^- + \Delta\delta)} \; + \frac{1}{C}\sum_{i<j}\frac{(aK_{ij}^+ + b - \delta_{ij}^+)^2}{(d_{ij}^+ - \delta_{ij}^- + \Delta\delta)}, \tag{19}$$

C is a normalizing constant defined as:

$$C \triangleq \sum_{i<j}(\delta_{ij}^+ - \delta_{ij}^-) \; . \tag{20}$$

The constant $\Delta\delta$, fixed arbitrarily to a small value, is used to avoid a division by zero in case of crisp input data.

The minimization of J_1 or J_2 with respect to M, a and b can be performed using any standard unconstrained non linear programming package.

Remark 3 Each bba m_i must take values in $[0,1]$ and satisfy Eq.(1). Hence, the optimization of J_1 or J_2 with respect to M is a constrained optimization problem. However, the constraints vanish if one uses the following parameterization:

$$m_{ik} = \frac{\exp(\alpha_{ik})}{\sum_l \exp(\alpha_{il})}, \tag{21}$$

where the α_{ik} for $i \in \{1,\ldots,n\}$ and $k \in \{1,\ldots,c,\Omega,\emptyset\}$ are $n(c+2)$ real parameters.

4 Example

4.1 Synthetic dataset

In order to illustrate the proposed approach, a two-dimensional artificial data set was generated. It is composed of 24 vectors $\tilde{\mathbf{x}}_{\mathbf{i}}$, each component of which is an interval

$[x_{i\ell}^-, x_{i\ell}^+]$ $\ell = 1, 2$. The data is given in Table 1. Each vector can be represented as a rectangle in the plane as shown in Figure 5. An interval-valued dissimilarity matrix $D = ([\delta_{ij}^-, \delta_{ij}^+])$ was computed in the following way: δ_{ij}^- and δ_{ij}^+ are defined, respectively, as the minimum and maximum of:

$$d(\mathbf{x}_i, \mathbf{x}_j) = \sqrt{\sum_{\ell=1}^{p} |x_{i\ell} - x_{j\ell}|^2},$$

under the constraints

$$\begin{aligned} x_{i\ell}^- \leq x_{i\ell} \leq x_{i\ell}^+ \qquad & \ell = 1, 2 \\ x_{j\ell}^- \leq x_{j\ell} \leq x_{j\ell}^+ \qquad & \ell = 1, 2. \end{aligned}$$

The range of $x_{i\ell} - x_{j\ell}$ under the above constraints is the interval

$$[x_{i\ell}^- - x_{j\ell}^+, x_{i\ell}^+ - x_{j\ell}^-].$$

Using the following properties:

$$\begin{aligned} \min_{a \leq x \leq b} |x| &= \max(0, a, -b) \\ \max_{a \leq x \leq b} |x| &= \max(b, -a), \end{aligned}$$

we get:

$$\delta_{ij}^- = \sqrt{\sum_{\ell=1}^{2} \max(0, x_{i\ell}^- - x_{j\ell}^+, x_{j\ell}^- - x_{i\ell}^+)^2}, \tag{22}$$

$$\delta_{ij}^+ = \sqrt{\sum_{\ell=1}^{2} \max(x_{i\ell}^+ - x_{j\ell}^-, x_{j\ell}^+ - x_{i\ell}^-)^2}. \tag{23}$$

We applied the algorithm with $c = 2$ classes with the two criteria J_1 and J_2. The quality of the resulting partitions can be judged from scatterplots, usually referred to as the *Shepard diagrams*, given in Figure 2. These diagrams provide, for each criterion, a representation of the dissimilarities vs. the degrees of conflict between bba's. It can be seen that a best linear fit is obtained with J_2. This criterion was thus retained to derive a partition. Figure 3 shows this partition, in which each point is assigned either to a class or to the empty set according to the highest mass. The masses are shown in Table 1. The partition is consistent with what was expected. The two main classes are well recovered. The method is thus able to provide meaningful partitions of data expressed as interval values. A second interest of the method is its ability to detect atypical observations, such as points 23 and 24 for which the greatest mass is allocated to the empty set. The empty set plays the role of a "noise" cluster (as proposed in [8]) allowing the detection of outliers in the data set. Finally, the algorithm provides additional information via the analysis of the mass allocated to Ω. It is represented in Figure 4 by shaded discs, the diameters of which reflect the masses allocating to Ω for each point. It can be seen that there is a close relationship between the mass of Ω and the imprecision of the data. For instance, points 22 and 23, whose components are crisp numbers, are given a mass equal to zero, whereas point 21, which is the most imprecise object of the data set, is given the highest mass to Ω. By offering an additional degree of freedom, the mass of Ω allows to manage the imprecision in the data in a very natural and efficient way. It can be used to assess the validity of hard assignments to the clusters.

4.2 Sensory data set

This real data set comes from a sensory tasting experiment reported in [18]. Ten colas were presented to 10 subjects, who were asked to rate the perceived dissimilarity between each pair of beverages on a scale ranging from 0 to 100 (the values of the 45 dissimilarities for the ten subjects may be found in [18]). The dissimilarity between each pair of objects is thus characterized by an empirical distribution gathering the answers of the ten subjects. To test our approach, we chose to define the interval-valued dissimilarity between two colas as the interquartile range of the empirical distribution. The colas used in the experiment are known to fall into two categories: the normal colas (Pepsi, Coca, Pepper, Shasta, RC-cola, Yukon) and the diet colas (D-pepper, D-pepsi, D-rite, Tab). It was thus of interest to see if this partition could be recovered from the dissimilarities. We applied our algorithm with $c = 2$ classes. The Shepard diagrams are shown in Figure 6. First, it can be seen that the fit is not as good as in the case of the synthetic data set. This fact can be explained by inconsistencies in the input data, due to erroneous evaluations by the subjects on the one hand, and to the construction of the dissimilarity matrix on the other hand. Secondly, the best fit seems to be obtained using the criterion J_1 and is thus retained to compute the partition. A multidimensional scaling method for interval-valued dissimilarities [9] was used to find a two-dimensional representation of the colas. It is given in Figure 5, together with the partition obtained with the clustering method. The hard assignment of the colas is done according to the maximum allocated mass (see Table 2). The size of the symbols used in the representation are proportional to the mass of the class of the different colas. One can see that the dichotomy diet/non diet beverages is well-recovered. Additionally, two colas (Yukon, Pepper) are detected as atypical beverages.

5 Conclusion

The problem of clustering interval-valued dissimilarities has been addressed. In the proposed approach, based on evidence theory, each object is assigned a bba over a given set of classes. The set of bba's allows to derive a partition of the objects. The allocation of a mass to the empty set can be viewed as a means to improve the robustness of the algorithm with respect to atypical data (outliers). On the other hand, the mass allocated to the whole set of hypothesis gives to the method the flexibility needed to manage the imprecision and the inconsistencies in the data. The interpretability of the resulting belief masses has been demonstrated experimentally.

References

[1] S. Auephanwiriyakul and J.M. Keller. Analysis and efficient implementation of a linguistic fuzzy c-means. *IEEE Transactions on Fuzzy Systems*, 10(5):563–582, 2002.

[2] J.C. Bezdek, J. Keller, R. Krishnapuram, and N.R. Pal. *Fuzzy models and algorithms for pattern recognition and image processing.* Kluwer Academic Publishers, Boston, 1999.

[3] I. Borg and P. Groenen. *Modern multidimensional scaling.* Springer, New-York, 1997.

[4] G. Bortolan and W. Pedrycz. Linguistic neurocomputing: the design of neural networks in the framework of fuzzy sets. *Fuzzy sets and systems*, 128:389–412, 2002.

[5] M. Chavent. Criterion-based divisive clustering for symbolic objects. In H.H. Bock and E. Diday, editors, *Analysis of symbolic data*. Springer, 2000.

[6] M. Chavent and Y. Lechevallier. Dynamical clustering of interval data. optimization of an adequation criterion based on hausdorff distance. In K. Jaguga et al., editor, *Clustering and Data Analysis*, pages 53–60. Springer, 2002.

[7] K.H. Chen, H.L. Chen, and H.M. Lee. A multiclass neural network classifier with fuzzy teaching inputs. *Fuzzy sets and systems*, 91:15–35, 1997.

[8] R.N. Davé. Clustering relational data containing noise and outliers. In *FUZZ'IEEE 98*, pages 1411–1416, Anchorage, 1998.

[9] T. Denoeux and M.H. Masson. Multidimensional scaling of interval-valued dissimilarity data. *Pattern Recognition Letters*, 21:83–92, 2000.

[10] T. Denoeux and M.H. Masson. Clustering of proximity data using belief functions. In *Proceedings of IPMU'2002*, volume I, pages 609–616, Annecy, France, July 2002.

[11] T. Denoeux and M.H. Masson. Evclus: evidential clustering of proximity data. *IEEE Transactions on Systems, Man and Cybernetics, Part B (to appear)*, 2003.

[12] P. Diamond and H. Tanaka. Fuzzy regression analysis. In R. Slowinski, editor, *Fuzzy sets in decision analysis, operations research and statistics*, pages 349–387. Kluwer Academic Publishers, Boston, 1998.

[13] E. Diday and H. Bock. *Analysis of symbolic data. Exploratory methods for extracting statistical information from complex data.* Springer Verlag, Heidelberg, 2000.

[14] J. Gebhardt, M. A. Gil, and R. Kruse. Fuzzy set-theoretic methods in statistics. In R. Slowinski, editor, *Fuzzy sets in decision analysis, operations research and statistics*, pages 311–347. Kluwer Academic Publishers, Boston, 1998.

[15] H. Ishibuchi, K. Morioka, and I.B. Turksen. Learning by fuzzified neural networks. *International Journal of Approximate Reasoning*, 13:327–358, 1995.

[16] A. K. Jain and R. C. Dubes. *Algorithms for clustering data.* Prentice-Hall, Englewood Cliffs, NJ., 1988.

[17] M.H. Masson and T. Denoeux. Multidimensional scaling of fuzzy dissimilarity data. *Fuzzy sets and systems*, 128(33):55–68, 2002.

[18] S. Schiffman, M. Reynolds, and F. Young. *Introduction to Multidimensional Scaling: Theory, Methods and Applications.* Academic Press, New-York, 1981.

[19] G. Shafer. *A mathematical theory of evidence.* Princeton University Press, Princeton, N.J., 1976.

[20] P. Smets. The combination of evidence in the Transferable Belief Model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(5):447–458, 1990.

[21] P. Smets. Belief functions: the disjunctive rule of combination and the generalized Bayesian theorem. *International Journal of Approximate Reasoning*, 9:1–35, 1993.

[22] P. Smets and R. Kennes. The Transferable Belief Model. *Artificial Intelligence*, 66:191–243, 1994.

[23] Y. El Sonbaty and M. Ismail. Fuzzy clustering for symbolic data. *IEEE Transactions on Fuzzy Systems*, 6:195–204, 1998.

Table 1: Artificial data set and masses allocated by the clustering algorithm

i	$[x_{i1}^-,x_{i1}^+]$	$[x_{i2}^-,x_{i2}^+]$	$m_i(\{\omega_1\})$	$m_i(\{\omega_2\})$	$m_i(\emptyset)$	$m_i(\Omega)$
1	[6.62,8.41]	[0.24,0.79]	0.70	0.14	0.07	0.09
2	[6.10,6.35]	[-1.69,-0.44]	0.66	0.28	0.02	0.04
3	[7.52,8.81]	[-0.10,1.54]	0.72	0.08	0.11	0.09
4	[7.80,8.67]	[1.61,2.27]	0.68	0.08	0.18	0.06
5	[5.96,7.69]	[-0.83,0.13]	0.67	0.19	0.03	0.11
6	[8.43,9.69]	[0.66,1.52]	0.71	0.00	0.20	0.09
7	[8.73,9.90]	[1.06,1.60]	0.69	0.00	0.24	0.07
8	[7.94,8.39]	[-2.20,-1.30]	0.69	0.11	0.16	0.04
9	[7.51,8.63]	[-1.71,-1.10]	0.69	0.11	0.14	0.06
10	[7.73,8.36]	[0.37,1.11]	0.75	0.11	0.10	0.03
11	[-0.80,0.01]	[-0.42,0.33]	0.07	0.75	0.13	0.05
12	[-0.07,1.41]	[-0.06,1.00]	0.14	0.71	0.06	0.09
13	[-1.51,-0.29]	[0.37,1.65]	0.00	0.70	0.20	0.10
14	[1.44,2.72]	[-0.22,1.28]	0.28	0.63	0.00	0.09
15	[-0.31,0.01]	[0.82,1.66]	0.10	0.74	0.14	0.03
16	[-0.29,0.81]	[0.25,1.37]	0.11	0.73	0.09	0.07
17	[0.13,1.45]	[0.34,1.74]	0.15	0.69	0.07	0.09
18	[-0.86,0.92]	[-1.73,-0.76]	0.09	0.66	0.14	0.11
19	[-0.51,0.76]	[-0.22,0.67]	0.10	0.74	0.09	0.07
20	[-1.73,-0.24]	[-0.83,0.46]	0.00	0.70	0.19	0.11
21	[3.00,6.00]	[-2.00,2.00]	0.42	0.34	0.00	0.25
22	[4.00,4.00]	[0.00,0.00]	0.50	0.50	0.00	0.00
23	[2.00,2.00]	[-6.00,-6.00]	0.19	0.37	0.44	0.00
24	[4.00,6.00]	[-6.00,-5.00]	0.30	0.21	0.35	0.14

Table 2: Sensory data set. Masses allocated by the clustering algorithm

Drinks	$m_i(\{\omega_1\})$	$m_i(\{\omega_2\})$	$m_i(\emptyset)$	$m_i(\Omega)$
D-pepsi	0.22	0.60	0.11	0.07
RC-cola	0.67	0.33	0.00	0.00
Yukon	0.47	0.00	0.53	0.00
Pepper	0.00	0.08	0.63	0.29
Shasta	0.78	0.06	0.16	0.00
Coca	0.72	0.00	0.28	0.00
D-pepper	0.00	0.65	0.32	0.03
Tab	0.00	0.42	0.36	0.22
Pepsi	0.70	0.20	0.08	0.02
D-rite	0.23	0.68	0.00	0.09

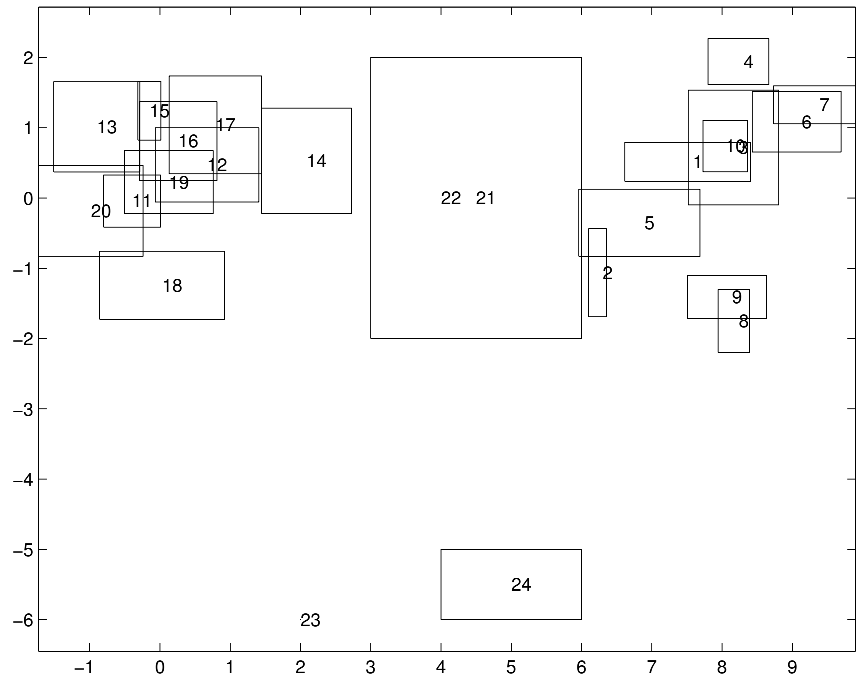

Figure 1: Two-dimensional data set

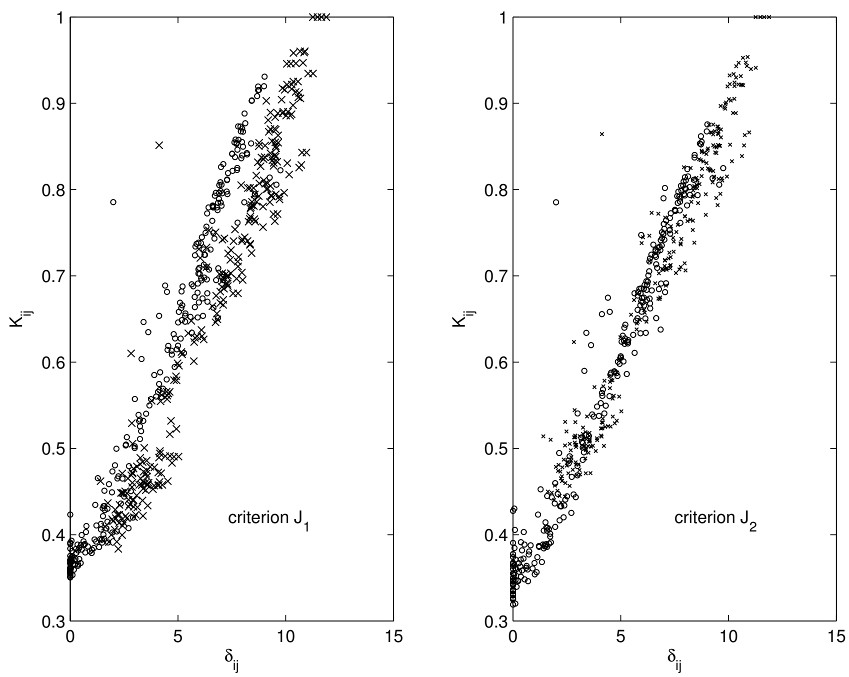

Figure 2: Dissimilarities vs degrees of conflict for the synthetic data set (o: lower dissimilarities; x: upper dissimilarities).

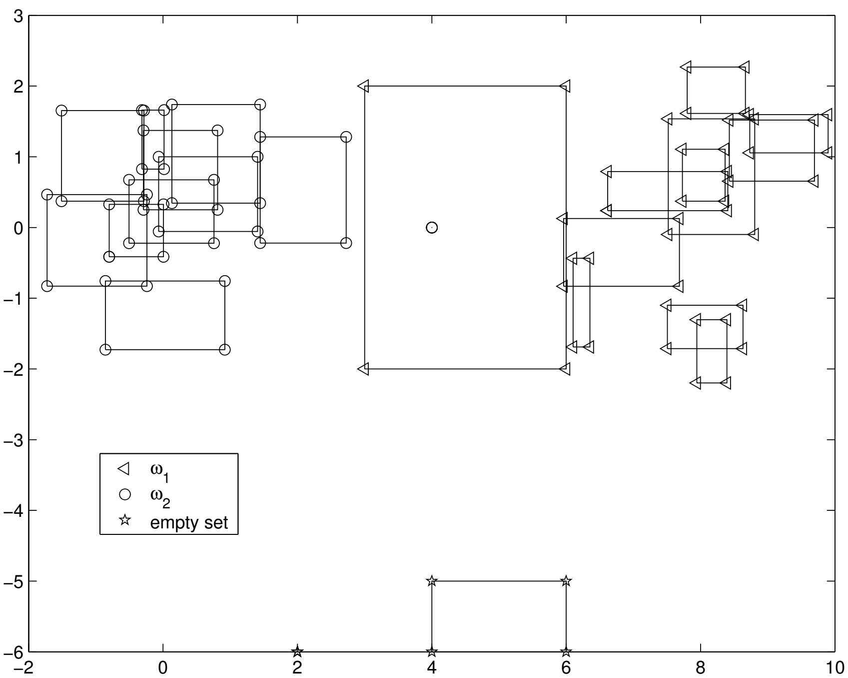

Figure 3: Synthetic data set. Partition into two classes with outlier rejection. Each point is either assigned to a class or to the empty set according to the maximum mass allocated.

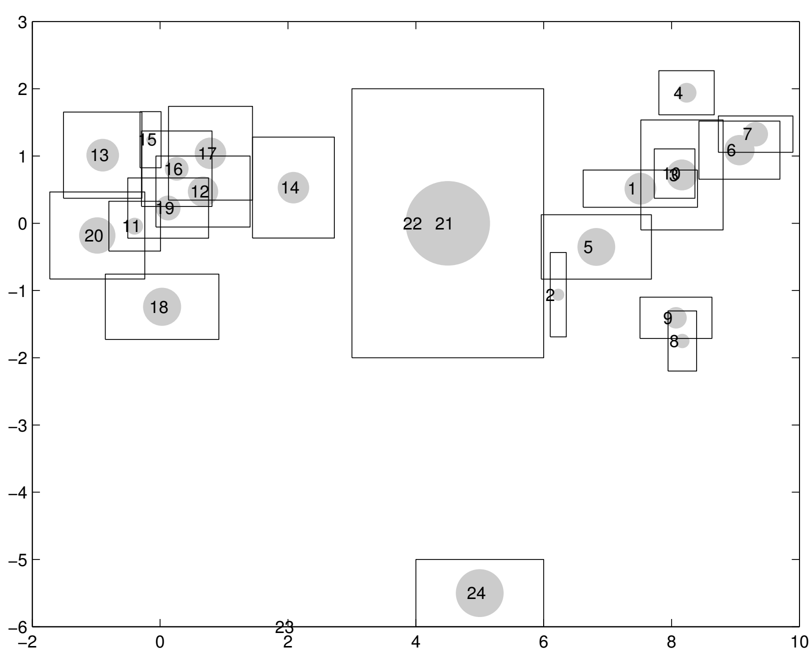

Figure 4: Synthetic data set. Mass allocated to Ω. This mass is proportional to the radius of the grey disc centered on each point of the data set.

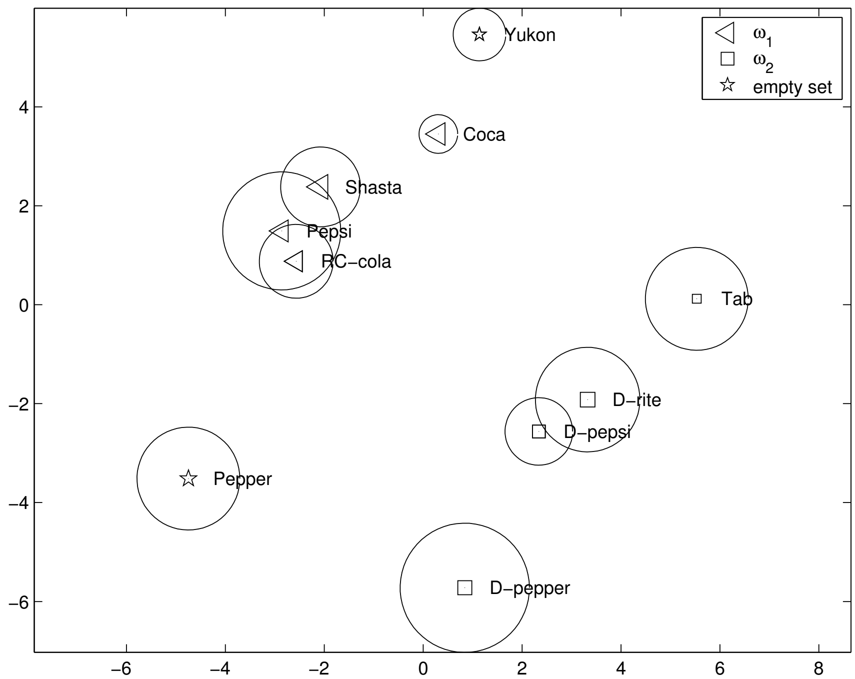

Figure 5: Sensory data set. Partition into two classes with outlier rejection. Each point is either assigned to a class or to the empty set according to the maximum mass allocated. The size of the symbols is proportional to the mass of the class.

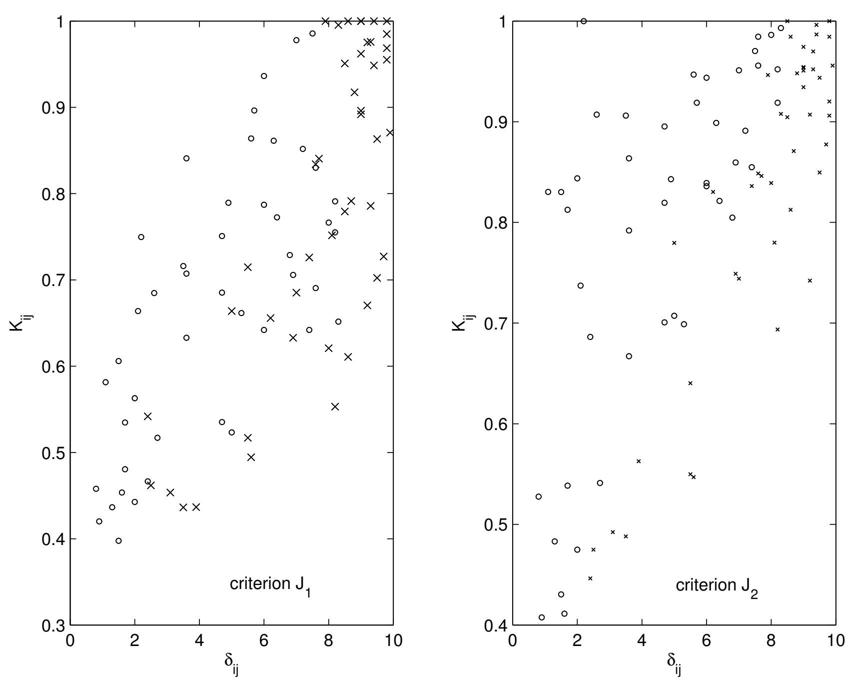

Figure 6: Dissimilarities vs degrees of conflict for the sensory data set (o: lower dissimilarities; x: upper dissimilarities).