ELSEVIER

Pattern Recognition Letters 17 (1996) 625–631

Pattern Recognition Letters

Conditional Fuzzy C-Means

Witold Pedrycz *

Department of Electrical & Computer Engineering, University of Manitoba, Winnipeg, Manitoba, Canada R3T 5V6

Received 3 September 1995; revised 15 November 1995

Abstract

A Fuzzy C-Means-based clustering method guided by an auxiliary (conditional) variable is introduced. The method reveals a structure within a family of patterns by considering their vicinity in a feature space along with the similarity of the values assumed by a certain conditional variable. The usefulness of the algorithm is exemplified in the problems of data mining.

Keywords: Fuzzy clustering; Fuzzy C-Means; Conditional variable; Data mining; Radial basis functions

1. Introduction

Fuzzy clustering (Zadeh, 1977; Bezdek, 1981) arises as a commonly used conceptual and algorithmic framework for data analysis and unsupervised pattern recognition. Evident are applications of fuzzy clustering to fuzzy modelling (primarily rule-based systems) and radial basis function neural networks. Quite often fuzzy clustering emerges as a synonym of unsupervised learning. The variants of objective-based (Fuzzy C-Means) clustering are numerous, cf. (Bezdek, 1981; Bezdek et al., 1987; Dave, 1992; Hathaway and Bezdek, 1994; Krishnapuram and Keller, 1993; Ruspini, 1970). These generalizations deal with various shapes of clusters (hyperellipsoidal, spherical, linear, etc.) being promoted (favored) by the specific form of the a priori assumed objective function. What is common to all these alternatives is their unsupervised form of learning (optimization of the objective function). The new idea to be pursued in this study can be exemplified by a simple example. It is obvious that the patterns in Fig. 1(a) form three evident clusters (ω_1, ω_2, ω_3).

Now consider that associated with each pattern $x = (x_1, x_2)$ is an auxiliary variable (z) whose values assumed over the individual patterns are around 1 and 2, respectively. Then the results of the clustering that takes this variable into account look very distinct, Fig. 1(b). Some patterns even very close in the feature space (x_1–x_2 plane) become very different when considered in light of this extra variable. The clustering involving both the features and the auxiliary variable can be called *conditional* as the patterns are structured into several categories not only based upon their vicinity in the feature space but also conditioned by the values of the variable used in this data. The conditional clustering emerges as one among various techniques of data mining that can be regarded as a linguistically oriented (focused) search for dependencies in multivariable data sets.

* E-mail: pedrycz@ee.umanitoba.ca.

PII S0167-8655(96)00027-X

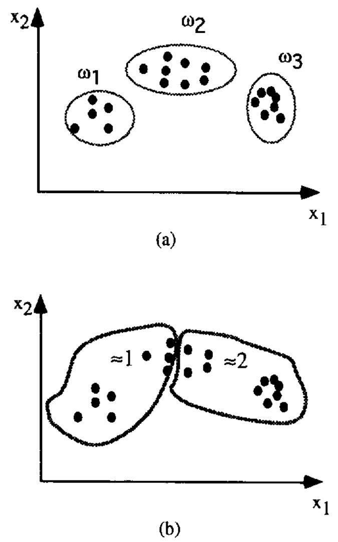

Fig. 1. Examples of (a) fuzzy clustering and (b) conditional fuzzy clustering.

Instead of the standard (condition-free) data analysis studied through clustering methods

reveal structure (groups) in data X,

we now consider

reveal groups in data X in context A

(in brief, the context A is established by considering some auxiliary explanatory variables).

The aim of this study is to propose a relevant extension of Fuzzy C-Means that is capable of addressing the idea of conditional clustering.

2. The algorithm

Assume that $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N$ are n-dimensional patterns defined in $\mathbb{R}^n$. The method comes as a certain modification of the generic version of Fuzzy C-Means. Let us briefly recall that the objective function as encountered in (Bezdek, 1981) reads as

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

with $U = [u_{ik}]$ being a partition matrix, $U \in \mathscr{U}$, and c standing for the number of the clusters; $p > 1$, and $\| \cdot \|$ forming a distance function computed between $\boldsymbol{x}_k$ and the prototype ($\boldsymbol{v}_i$) of the ith cluster (group). $\mathscr{U}$ denotes a family of $c \times N$ partition matrices, namely

$$\mathscr{U} = \left\{ u_{ik} \in [0, 1] \,\middle|\, \sum_{i=1}^{c} u_{ik} = 1 \ \forall k \text{ and } \middle| \ 0 < \sum_{k=1}^{N} u_{ik} < N \ \forall i \right\}. \tag{1'}$$

Formally, the optimization problem comes in the form

$\min_{U, \boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_c} Q$

subject to $U \in \mathscr{U}$.

The iterative scheme leading either to a local minimum or a saddle point of Q is well known (Hathaway and Bezdek, 1994) and involves a series of updates of the partition matrix written as

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

$i = 1, 2, \ldots, c, \ k = 1, 2, \ldots, N.$

The prototypes of the clusters are obtained in the form of weighted averages of the patterns

$$\boldsymbol{v}_i = \sum_{k=1}^{N} u_{ik}^{p} \boldsymbol{x}_k \Bigg/ \sum_{k=1}^{N} u_{ik}^{p}.$$

The conditioning aspect of the clustering mechanism is introduced into the algorithm by taking into consideration the conditioning variable assuming the values $f_1, f_2, \ldots, f_N$ for the corresponding patterns. More specifically, f_k describes a level of involvement of $\boldsymbol{x}_k$ in the constructed clusters. The way in which f_k can be associated with or allocated among the computed membership values of $\boldsymbol{x}_k$, say $u_{1k}, u_{2k}, \ldots, u_{ck}$, is not unique. Two possibilities are worth exploring:

– We admit f_k to be distributed additively across the entries of the kth column of the partition matrix meaning that

$$\sum_{i=1}^{c} u_{ik} = f_k, \quad k = 1, 2, \ldots, N. \tag{2}$$

– We can also request that the maximum of membership values within the corresponding column equals f_k,

$$\max_{i=1,\ldots,c} u_{ik} = f_k, \quad k = 1, 2, \ldots, N. \tag{3}$$

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

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

Note that the standard normalization condition in (1′) is replaced by the involvement (conditioning) constraint. The optimization problem is now reformulated accordingly

$$\min_{U, \boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_c} Q$$

$$\text{subject to } U \in \mathscr{U}(f). \tag{4}$$

Again the minimization of the objective function is carried out iteratively where the partition matrix is updated accordingly,

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

We arrive at the above formula by transforming (4) to a standard unconstrained optimization problem by making use of Lagrange multipliers and determining a critical point of the resulting function. The prototypes are modified as in the original method.

One should emphasize that the membership values do not sum up to 1; a similar phenomenon can be witnessed in possibilistic clustering (Krishnapuram and Keller, 1993) and clustering with noise cluster (Dave, 1992), although the origin of these two departures from the constraint is completely different.

The method can be readily extended to several conditional variables. With some linguistic terms of interest defined for each variable we develop a composite logic formula (e.g., (f and g) or h) used to guide the clustering algorithm. For instance,

reveal groups in data $\boldsymbol{X}$ considering that

(y_1 is *small* and y_2 is *large*) or y_3 is *medium*

More formally, the previous optimization task involves more constraints, say $\boldsymbol{f}$, $\boldsymbol{g}$, $\boldsymbol{h}$,

$$\min_{U, \boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_c} Q$$

subject to $U \in \mathscr{U}(\boldsymbol{f}, \boldsymbol{g}, \boldsymbol{h})$,

U belongs to the family of the partition matrices defined as

$$\mathscr{U}(\boldsymbol{f}, \boldsymbol{g}, \boldsymbol{h}) = \left\{ u_{ik} \in [0,1] \,\middle|\, \sum_{i=1}^{c} u_{ik} = L(f_k, g_k, h_k) \ \forall k \,\middle|\, \text{and } 0 < \sum_{k=1}^{N} u_{ik} < N \ \forall i \right\},$$

where L is a logic expression of $\boldsymbol{f}$, $\boldsymbol{g}$, $\boldsymbol{h}$ bin fuzzy sets (linguistic labels).

4. Numerical studies

Two experiments are reported in this section. The first one shows the details of the algorithm, whereas the second example alludes to linguistic data mining of multidimensional data.

Example 1. The two-dimensional data set under discussion is shown in Fig. 2 and Table 1.

The patterns are conditioned by a certain variable whose values are reported in the third column of Table 1. Not including f_k's, two clusters show up quite distinctly, see Fig. 3. Dark shading indicates higher membership values.

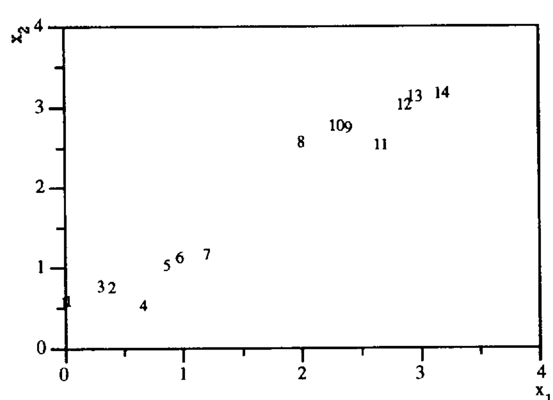

Fig. 2. Two-dimensional patterns.

Table 1
A collection of patterns along with their conditional variable (f)

Pattern no.	x_1	x_2	f
1	0.00	0.50	0.96
2	0.40	0.67	0.84
3	0.30	0.70	0.70
4	0.67	0.45	0.82
5	0.87	0.95	0.21
6	0.97	1.05	0.12
7	1.20	1.10	0.18
8	2.00	2.50	0.96
9	2.40	2.67	0.84
10	2.30	2.70	0.70
11	2.67	2.45	0.82
12	2.87	2.95	0.21
13	2.97	3.05	0.12
14	3.20	3.10	0.18

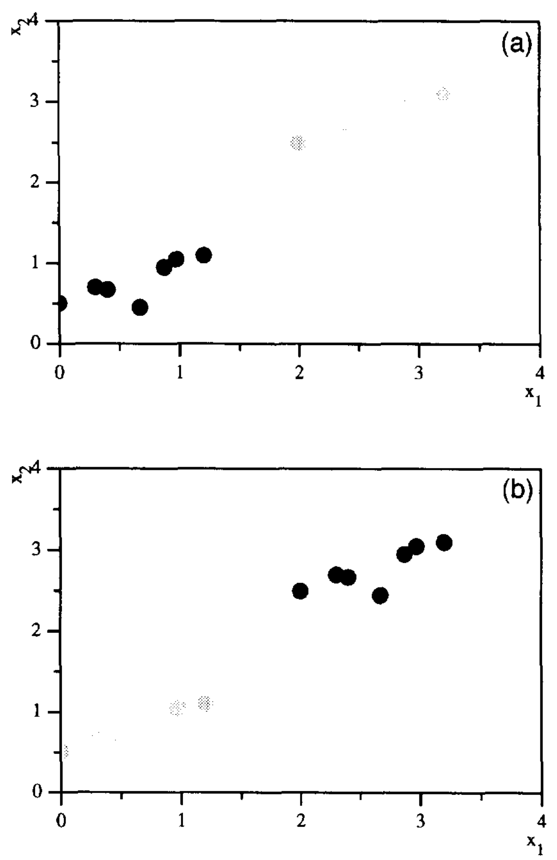

Fig. 3. Results of fuzzy clustering into two clusters (a) and (b) – no conditioning effect.

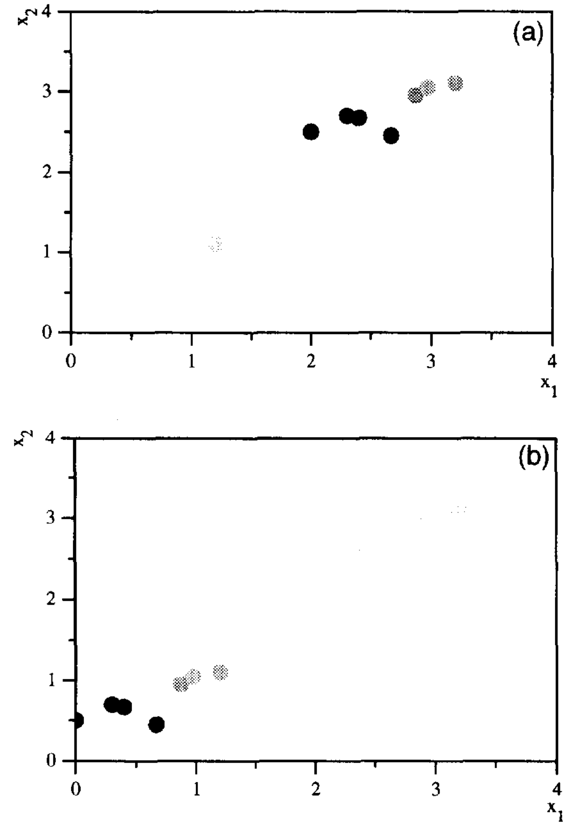

Fig. 4. Results of fuzzy clustering into two clusters (a) and (b) – conditioning effect included.

The conditional variable does make a difference. Hence the clustering reveals a very different organization of the patterns producing two clusters indicated in Fig. 4.

It is interesting to observe that some of the patterns (in particular 5, 6, 7, and 12, 13, 14) even though conditioned quite low are still assigned to the clusters primarily because of their very high resemblance to the patterns forming the core of the clusters. Table 2 summarizes the prototypes of the clusters.

Example 2. In general, the problem of data mining is aimed at revealing relationships in multivariable data sets. The problem becomes of a genuine concern when dealing with rapidly growing databases. Moreover the mechanism of this search should be very much user-oriented meaning that one should be

Table 2
Prototypes of the clusters

Conditional clustering		
prototype 1	2.3355	2.5793
prototype 2	0.3312	0.5785
Clustering		
prototype 1	2.6381	2.7758
prototype 2	0.6225	0.7703

able to customize and focus the search of these dependencies according to some specific requirements formulated by the user. Being very much context-sensitive, conditional clustering can be a useful addition to a toolbox of data analysis of this type. To focus our attention, let us consider a data set provided in a tabular format. The dependencies between the variables (features) are to be revealed given a context specified linguistically and pertaining to a certain variable, see Fig. 5.

An instance of data mining can be articulated as

reveal dependencies between $x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n$ given x_i is *small*

where *small* is a fuzzy set specified for the ith variable.

From the above formulation it becomes obvious that the problem calls for the mechanisms of conditional clustering whose conditional variable comes in the form of the corresponding membership function.

The considered data set comes from (Chambers et al., Table 2, p. 347), and deals with air quality in the New York metropolitan area. Daily readings concern air quality characterized by the four-dimensional feature vector:

x_1 – mean ozone level (parts per billion),
x_2 – solar radiation,
x_3 – average wind speed,
x_4 – maximal daily temperature.

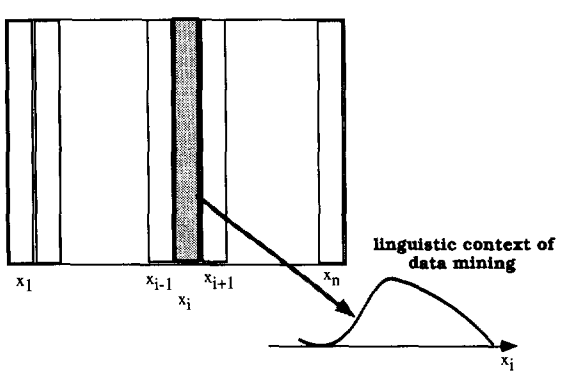

Fig. 5. Data mining completed in the context of linguistic labels.

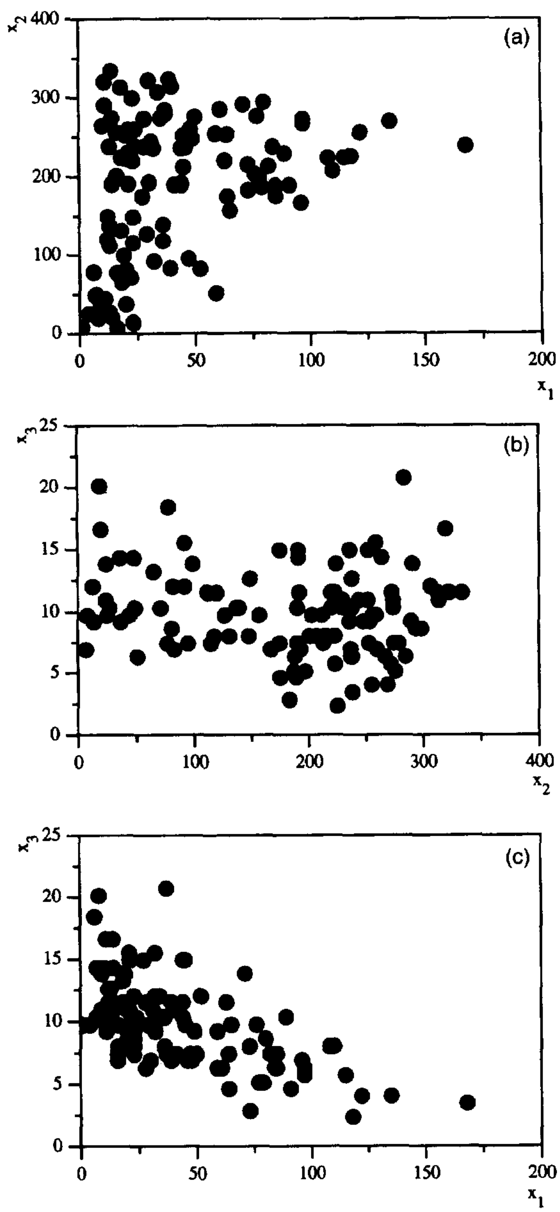

Fig. 6. Two-dimensional plots of data.

Table 3
Prototypes of the clusters – conditional clustering

prototype 1	71.2914	249.8229	8.3880
prototype 2	43.3526	99.7794	12.2129
prototype 1	64.5342	280.5529	7.7655
prototype 2	41.9109	90.1966	12.6245
prototype 3	73.6518	217.3061	9.1198
prototype 1	61.8981	214.0438	9.8788
prototype 2	59.3803	283.4610	7.8542
prototype 3	41.7349	88.9482	12.6864
prototype 4	112.1120	235.0014	7.0208
prototype 1	60.4605	285.9387	7.7377
prototype 2	48.9608	191.6477	10.1027
prototype 3	41.8340	88.2855	12.7346
prototype 4	64.0542	224.3242	10.0292
prototype 5	112.4965	233.6293	7.1390
prototype 1	111.1535	229.1800	7.4612
prototype 2	65.4006	289.8749	7.6478
prototype 3	51.1749	252.4707	8.8493
prototype 4	41.8718	88.1751	12.7448
prototype 5	67.6474	218.0532	9.9440
prototype 6	47.3570	190.4026	10.4619

Only complete vectors (no missing data) are analyzed. This yields 111 four-dimensional patterns, see Fig. 6.

Table 4
Prototypes of the clusters – Fuzzy C-Means

prototype 1	50.7674	245.3164	9.5784
prototype 2	22.1233	71.7794	10.9758
prototype 1	38.6772	270.6956	10.5910
prototype 2	19.3078	57.2232	11.2832
prototype 3	64.7932	197.4344	8.1412
prototype 1	35.1091	272.0611	10.8517
prototype 2	77.0283	210.9928	7.5520
prototype 3	14.2086	33.6573	11.6601
prototype 4	26.5386	122.3671	10.2254
prototype 1	36.7870	293.1449	10.9922
prototype 2	26.3074	112.6052	10.3256
prototype 3	13.4287	30.5839	11.7253
prototype 4	87.3194	207.9694	6.8342
prototype 5	31.8851	234.5482	10.5984
prototype 1	95.3058	218.4979	6.7660
prototype 2	33.9147	301.5518	11.2822
prototype 3	32.3318	246.1810	10.7107
prototype 4	12.6250	27.1755	11.8100
prototype 5	38.4703	183.0297	10.0110
prototype 6	27.1667	96.8553	10.6623

Data mining can be realized in many different ways. Let us consider the context formed by the fourth variable and the linguistic term *about* 85 F specified therein where this concept is modelled through the membership function defined as

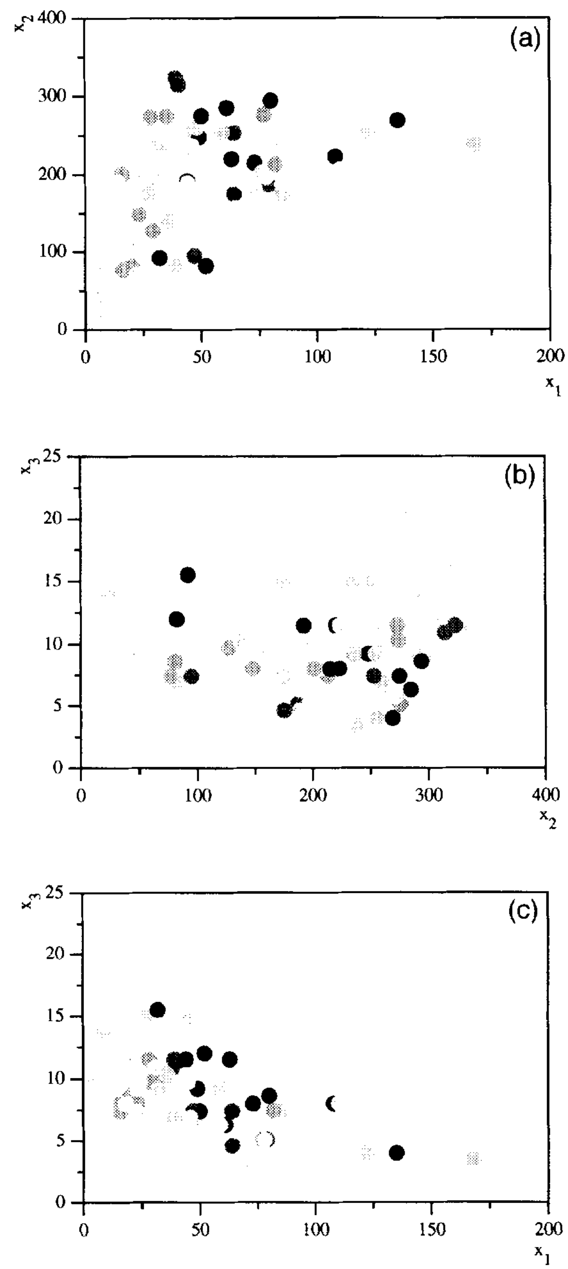

Fig. 7. Two-dimensional plots of data with superimposed membership grades of the linguistic term *about* 85 F.

$\exp(-(x_4-85)^2/5)$. The conditional clustering is carried out for the feature space composed of x_1, x_2, x_3. The results of this clustering reported in terms of the prototypes are outlined in Table 3.

To support comparative analysis, Table 4 summarizes the prototypes derived through Fuzzy C-Means without conditioning (this could also be viewed as a special form of context set up by the linguistic term *unknown*).

Fig. 7 shows three two-dimensional plots of data described by the values of the assumed membership function.

The conditional variable exhibits a visible impact on the location of the prototypes. Consider two classes, $c = 2$. Without conditioning, one of the prototypes ([22.12 71.78 10.98]) comes as an effect of averaging of the patterns and therefore gets moved down to the lower values of the features.

4. Conclusions

Starting off with the standard version of Fuzzy Isodata, we have proposed its generalization involving the conditioning effect imposed by some auxiliary variable(s) (features). The clustering algorithm is given and illustrated with the use of numerical examples showing its usefulness in data mining. Some other areas in which the effect of conditioning becomes essential include:
- the design of radial basis function neural networks;
- the construction of fuzzy rule-based classifiers whose classification outcomes are continuous rather than binary.

In both of them we are concerned with the determination of regions in the input or feature space of highest possible homogeneity.

It is worth underlining that other approaches such as fuzzy c-lines, relational Fuzzy C-Means, etc. can be "conditionalized" in the same way. This emphasizes the generality of the proposed approach.

Acknowledgements

Support from the Natural Sciences and Engineering Research Council of Canada, MICRONET and University Research Grants Program (URGP) of the University of Manitoba is gratefully acknowledged.

References

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

Bezdek, J.C., R.J. Hathaway, M.J. Sabin and W.T. Tucker (1987). Convergence theory for fuzzy c-means: counterexamples and repairs. *IEEE Trans. Syst. Man Cybernet.* 17, 873–877.

Chambers, J.M., W.S. Cleveland, B. Kleiner and P.A. Tukey (1983). *Graphical Methods for Data Analysis*. Wadsworth Int. Group, Belmont, CA.

Dave, R. (1992). Characterization and detection of noise in clustering. *Pattern Recognition Lett.* 12, 657–664.

Hathaway, R.J. and J.C. Bezdek (1994). NERF c-means non-Euclidean relational fuzzy clustering. *Pattern Recognition* 27, 429–437.

Krishnapuram, R. and J.M. Keller, (1993). A possibilistic approach to clustering. *IEEE Trans. Fuzzy Systems* 1, 98–110.

Ruspini, E. (1970). Numerical methods for fuzzy clustering. *Information Sci.* 2, 319–350.

Zadeh, L.A. (1977). Fuzzy sets and their applications to classification and clustering. In: J. van Ryzin, Ed., *Classification and Clustering*. Academic Press, New York, 251–299.