Fuzzy clustering with additional information

Witold Pedrycz
Department of Automatic Control and Computer Sci.
Silesian Technical University
44-100 Gliwice
16,Pstrowskiego
Poland

Abstract The paper deals with fuzzy clustering algorithms based on minimization of an objective function performance index in a presence of additional information conveyed by some labelled patterns.

1. Introduction

Nowadays fuzzy clustering methods(unsupervised pattern recognition) - a fundamental tool for pattern recognition, arein rapid growth [1] -[8]. They provide a data analyst an information on a structure in a data set under consideration,pointing out a grade of belonginess to a concrete class. By comparison,nonfuzzy clustering methods generate a {0,1}partition of the data set.We would like to underline here that this way of thinking may radically mask imprecision(uncertainty)existing in a real world.Remembering that clustering is an initial stage of various data analysis,which usually are concerned with complex or imprecise objects(e.g. handwritten characters,biological signals)we prefer to characterize the object by a value from the[0,1]-interval rather than giving the information about a total belonginess or nonbelonginess.

[illegible] problem we are interested in, is tied with clustering in a pre- [illegible] group of labelled patterns (elements of the data set) to which we can assign a grade of belonginess to the classes. Such a situa- [illegible] may occur while discussing a set of a few categories of handwrit- [illegible] characters with several elements carefully investigated by the da- [illegible] and labelled by him. Of course, the process of labelling is [illegible] consuming, so we can speak about a situation where only few ele- [illegible] are investigated. The classification performed by a human being [illegible] tive and therefore we can expect grades of belonginess in a [illegible] range between 0 and 1.

[illegible] the paper we shall illustrate abovestated problem by the use [illegible] c-means algorithm provided by Bezdek (cf. e.g. [2]). We shall [illegible] some new objective functions and present iterative procedures [illegible] their minimization (Section 3) A numerical example in Section [illegible] with illustration of the algorithms proposed by means of a data [illegible] as Gustafson's cross [3]. Let us start our discussion remebering [illegible] notions and notations, which will be useful in further consi- [illegible].

[illegible] Preliminaries

Consider a finite data set $X=\{\underline{x}_i\}_{i=1,2,\ldots,n}$ with elements [illegible] points in p-dimensional Euclidean space, $\underline{x}_i \in \mathbb{R}^p$. The problem [illegible] fuzzy clustering of the data set into "c" clusters (classes) [illegible] with a generation of a fuzzy partition **matrix** U, viz. the matrix with the entries from the $[0,1]$-interval,

$$U=[u_{ik}]_{\substack{i=1,2,\ldots,c \\ k=1,2,\ldots,n}} \tag{1}$$

[illegible] fixed column, say k_0, their elements reflect a strength of [illegible] of the k_0-th pattern to respective classes. Moreover the [illegible] conditions are usually imposed (e.g. [2]),

[illegible] cluster is nonempty,

$$\forall_{1 \leqslant i \leqslant c} \quad \sum_{k=1}^{n} u_{ik} > 0 \tag{2}$$

[illegible] every column is equal to 1,

$$\forall_{1 \leqslant k \leqslant n} \quad \sum_{i=1}^{c} u_{ik} = 1 \tag{3}$$

[illegible] $\mathbb{U}$ a family of fuzzy partition matrices U,

$$\mathbb{U} = \{ U \mid U \text{ satisfies } (2) \text{ and } (3) \} \tag{4}$$

[illegible] of X are clustered in order to minimize a performance index [illegible]. It may take a form (cf. [8])

$$J = \sum_{l=1}^{n} \sum_{i=1}^{c} \sum_{k=1}^{n} g(w_k, u_{ik})\, d(x_k, x_l) \tag{5}$$

subject to $U \in \mathbb{U}$

[illegible] being a dissimilarity measure of the elements of X, x_k, x_l. w_k [illegible] the a priori weights assigned to all objects of X.

[illegible] is any inner-product induced on $\mathbb{R}^p$, $g(w_k, u_{ik})$ is defined [illegible] then (5) is read as

$$J = \sum_{i=1}^{c} \sum_{k=1}^{n} u_{ik}^{m} \; d_{ik}^{2} \tag{6}$$

[illegible] is equal to

$$d_{ik}^2 = \|x_k - v_i\|^2 \qquad (7)$$

[illegible] ing a fuzzy mean of the i-th cluster,

$$v_i = \sum_{k=1}^{n} x_k u_{ik} \Big/ \sum_{k=1}^{n} u_{ik} \qquad (8)$$

$\|\cdot\|$ is [illegible] as any inner product induced on $\mathbb{R}^p$, $m \in (1,\infty)$. Very often one [illegible]

$$\|x_k - v_i\|^2 = \sum_{j=1}^{p} (x_{kj} - v_{ij})^2 \qquad (9)$$

[illegible] of the optimization task,

$$\min_{U \in \mathbb{U},\ v_1, v_2, \ldots, v_c} Q \qquad (10)$$

[illegible] is well-known as Fuzzy c-means algorithm given by Bezdek. [illegible] ization procedure which converges to a local minumum of (10), [illegible] the partition matrix and fuzzy means of respective classes. [illegible] with any initial U^0 we follow a sequence of steps, $l=0,1,2,\ldots$

1. Calculate fuzzy means v_i, $i=1,2,\ldots,c$ according to (8),
2. Update U^l as

$$u_{ik}^l = 1 \Big/ \sum_{j=1}^{c} \left(\frac{d_{ik}}{d_{jk}} \right)^{2/(m-1)} \qquad (11)$$

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

3. Compare U^l to U^{l+1}, if $\|U^l - U^{l+1}\| < \delta$ then stop, otherwise return to [illegible] 1 making use of the partition matrix calculated in the step 2.

[illegible] present possible modifications of the performance index (6) in [illegible] use information contained in the labelled points.

...ization of the objective functions with the aid of labelled patterns

... we reconsider several modifications of the objective ... (4) in order to handle labelled patterns.

... us denote by F a partition matrix representing labelled ...,

$$F = [f_{ik}]_{\substack{i=1,2,\ldots,c \\ k=1,2,\ldots,n}} \qquad (12)$$

... is assigned:

... $\in X_1$ (set of labelled elements of X) then f_{ik} are given by the data ..., with the condition $\sum_{i=1}^{c} f_{ik}$ satisfied,

... $\notin X_1$ then f_{ik} takes any value, $i=1,2,\ldots,c$.

... an n-dimensional vector with the entries,

$$g_k = \begin{cases} 1, \text{if } x_k \in X_1 \\ 0, \text{otherwise} \end{cases} \qquad (13)$$

... modifications of Q are listed below.

...

$$Q = \sum_{i=1}^{c} u_{ik}^{2} d_{ik}^{2} + \sum_{i=1}^{c} (u_{ik} - f_{ik} g_k{}^{2}) d_{ik}{}^{2} \qquad (14)$$

... modification of the performance index takes into account a sum of ... distances between u_{ik} and f_{ik} with weights being d_{ik}^{2}. Then we

$$\min_{U \in \mathbb{U}, v_1, v_2, \ldots, v_c} Q \qquad (15)$$

... find a local minimum of Q with regards to the constraint $\sum$ ..., we apply a technique of Lagrange multipliers. It leads to ... minimization of an index,

$$\ldots = Q - \lambda\left(\sum_{i=1}^{c} u_{ik} - 1\right) \tag{16}$$

$$\partial \ldots / \partial \ldots \; f_{st} g_t + 2(u_{st} - f_{st} g_t) d_{st}^2 - \lambda = 0, \tag{17}$$

$$u_{st} = (f_{st} g_t d_{st}^2 + \lambda) / 4 d_{st}^2 , \tag{18}$$

$$\sum \ldots = \sum_{j=1}^{c} (f_{jt} g_t d_{jt}^2 + \lambda) / 4 d_{jt}^2 = 1 \tag{19}$$

... range multiplier is equal to

$$\lambda = \left(4 - \sum_{j=1}^{c} f_{jt} g_t\right) / \sum_{j=1}^{c} \left(1/d_{jt}^2\right) \tag{20}$$

... leads to,

$$u_{st} = \left(4 - \sum_{j=1}^{c} f_{jt} g_t\right) / 4 \sum_{j=1}^{c} \left(\frac{d_{st}^2}{d_{jt}^2}\right) + f_{st} g_t / 4 \tag{21}$$

... iterative scheme given in the previous section is modified ... step ..., where a calculation of the partiton matrix is processed ... of (21). Note that if $g_t = 0$ for all $t = 1, 2, \ldots, n$ then (21) redu- ... formula derived before,

$$u_{st} = 1 / \sum_{j=1}^{c} \left(\frac{d_{st}^2}{d_{jt}^2}\right) \tag{22}$$

... labelled elements are an important source of information about an internal structure of the data set. It will be applied by specification of the inner product induced on $\mathbb{R}^p$. Introduce fuzzy covariance matrices generated by labelled data,

$$C_{f_i} = \frac{\sum_{k=1}^{n} f_{ik}^2 g_k (x_k - w_i)(x_k - w_i)^T}{\sum_{k=1}^{n} f_{ik}^2 g_k} \qquad (23)$$

where w_i are fuzzy means,

$$w_i = \sum_{k=1}^{n} f_{ik} g_k x_k \Big/ \sum_{k=1}^{n} f_{ik} g_k \qquad (24)$$

$i=1,2,\dots,c$. Next distances between pattern and the fuzzy means are computed with the help of C_{f_i} introduced above,

$$d'_{ik} = (x_k - v_i)^T C_{fi}^{-1} (x_k - v_i) \qquad (25)$$

The performance index we take into account is given by,

$$Q = \sum_{i,k} u_{ik}^2 \, d_{ik}'^{\,2} \qquad (26)$$

In this performance index the labelled data generate a structure of the inner product on $\mathbb{R}^p$.

C. This criterion makes use of the labelled data in two ways: firstly they are applied in order to generate the underlying structure C_{f_i} secondly, to measure a distance between partition matrices U and F,

$$= \sum_{i,k} u_{ik}^2 \; d_{ik}^2 \; + \; \sum_{i,k} (u_{ik} - f_{ik} g_k)^2 \; d_{ik}^2 \tag{27}$$

The criteria B. and C. are constructed in such a manner that the algorithms presented before can be used without significant modifications.

4. Numerical illustration

As an illustration of the clustering methods discussed above, we consider the data set known as Gustafson's crosss (see Fig. 1), which consists of two classes of points with a certain overlap. Six points indicated in Fig. 1 are labelled.

For comparison we test four algorithms: Fuzzy c-means without the use of the labelled objects, and next the method A, B, and C. In order to evaluate a character of convergence of the methods established we take into account an index, which is a norm defined in the space $\mathbb{U}$

$$e(l) = \max_{\substack{1 \le i \le c \\ 1 \le k \le n}} | u_{ik}^{l} - u_{ik}^{l-1} | \tag{28}$$

l=1,..... The results are displayed in Fig. 2. Fuzzy c-means algorithm is characterized by a slow rate of convergence; it is not suprising remembering the fact that Euclidean distance prefers hyperellipsoidal shapes of clusters, while A, B, and C indicate the same speed of convergence.

Moreover, we evaluate the methods tested calculating a sum of squared deviations between values of the membership functions of the labelled elements and values of the computed membership functions,

$$q = \sum_{i=1}^{c} \sum_{x_k \in X_l} |u_{ik} - f_{ik}| \tag{29}$$

The results are summarized in the Tab.1. It is clear that introduction of the fuzzy covariance matrices which control shapes of the generated clusters, makes it possible to diminish the values of q, while the introduction of the constraints in terms of membership functions only, has minor influence on q.

The results of the last method (C) are shown in Fig.1.

Table 1. Calculated values of the membership functions for the labelled patterns

pattern	label	Fuzzy c-means	A	B	C
x_1	1.0,0.0	0.71,0.29	0.64,0.36	0.99,0.01	0.99,0.01
x_6	0.5,0.5	0.78,0.22	0.91,0.09	0.92,0.08	0.81,0.19
x_{10}	1.0,0.0	0.16,0.84	0.66,0.34	0.99,0.01	0.99,0.01
x_{11}	0.0,1.0	0.74,0.26	0.20,0.80	0.00,1.00	0.00,1.00
x_{15}	0.5,0.5	0.89,0.11	0.01,0.99	0.00,1.00	0.13,0.87
x_{20}	0.0,1.0	0.35,0.65	0.50,0.50	0.00,1.00	0.00,1.00
q		5.78	4.60	1.88	1.40

References

1. E. Backer, Cluster analysis by optimal decomposition of induced fuzzy sets, Delft Univ. Press, Delft, 1978.

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

3. D.E.Gustafson,W.Kessel,Fuzzy clustering with a fuzzy covariance matrix, in Proc.IEEE-CDC,vol.2,F.S.Fu ed., IEEE Press,Piscataway,New Jersey, 1979,761-766.

4. ...ert,M.Roubens,Non metric fuzzy clustering algorithms,in Approximate Reasoning and Decision Analysis, M.M.Gupta,E.Sanchez eds.,North Holland,Amsterdam,1982,417-426.

5. W.Pedrycz,E.Backer,Determination of the cores of fuzzy clusters, Fuzzy Sets and Systems submitted .

6. E.Ruspini,A new approach to clustering,Information and Control,15, 1969,22-32.

7. E.Ruspini,Numerical methods for fuzzy clustering,Information Sciences, 2,1970,319-350.

8. M.Roubens,Pattern classification problems and fuzzy sets,Fuzzy Sets and Systems,1,1978,239-253.

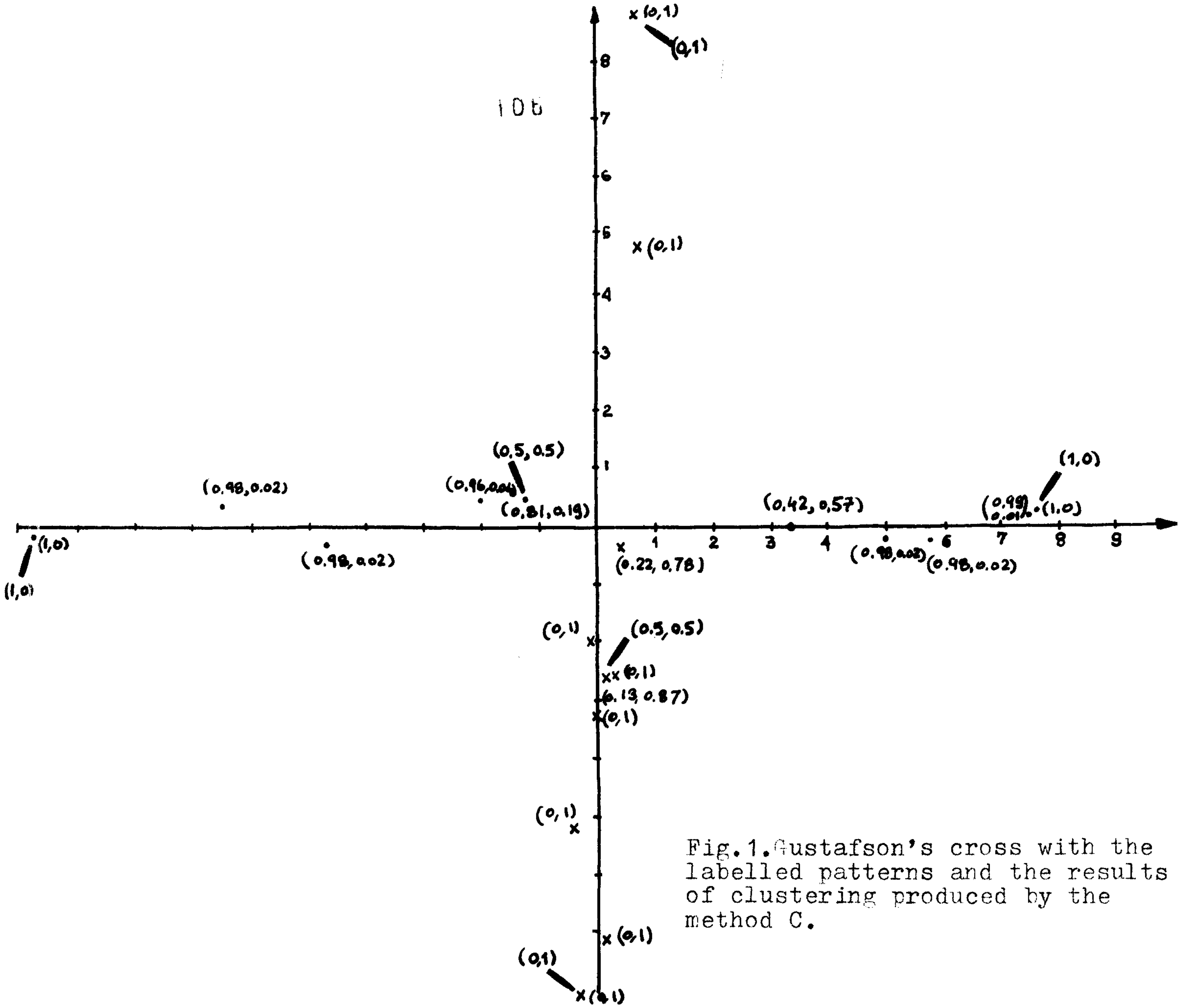

Fig.1.Gustafson's cross with the labelled patterns and the results of clustering produced by the method C.

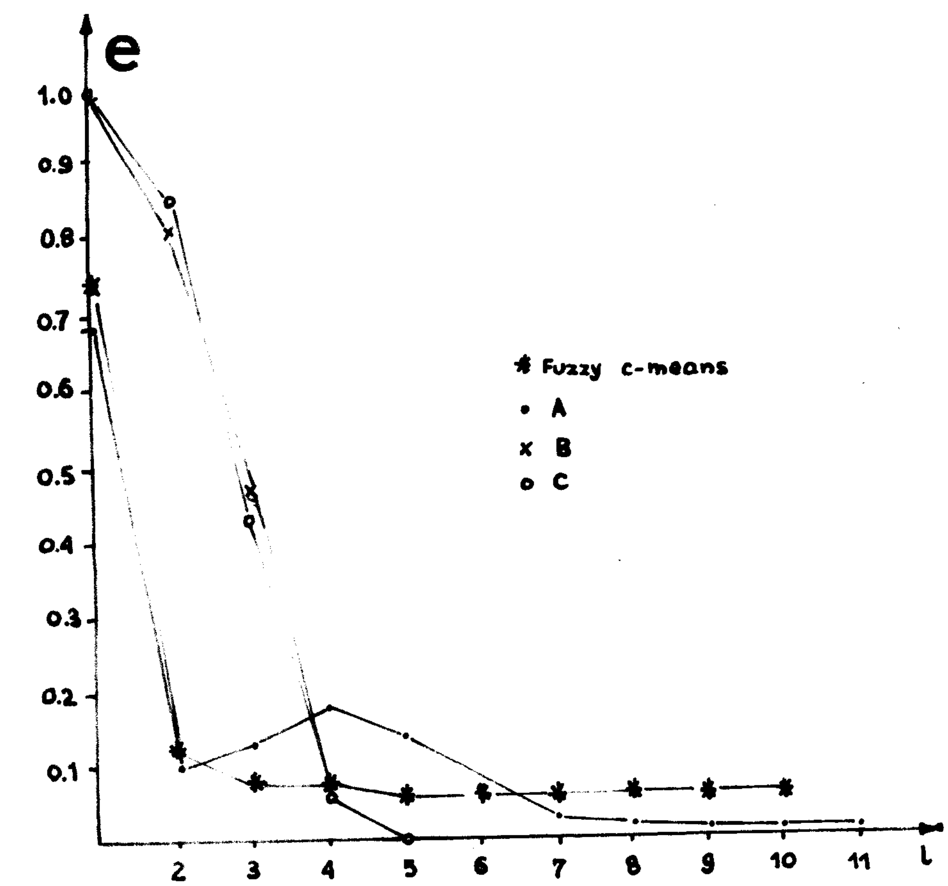

Fig.2.The index of convergence e(l) vs.number of iterations for various clustering methods.