Context Sensitive Fuzzy Clustering

Annette Keller[1] Frank Klawonn[2]

[1]Institute for Flight Guidance
German Aerospace Center
Lilienthalplatz 7
D-38108 Braunschweig, Germany
e-mail annette.keller@dlr.de

[2]Department of Electrical Engineering and Computer Science
Ostfriesland University of Applied Sciences
Constantiaplatz 4
D-26723 Emden, Germany
e-mail klawonn@et-inf.fho-emden.de

Abstract

We introduce an objective function-based fuzzy clustering technique that incorporates linear combinations of attributes in the distance function. The main application field of our method is image processing where a comparison pixel by pixel is usually not adequate, but the environmnet of a pixel or groups of pixels characterize important properties of an image or parts of it. In addition, our approach can be seen as generalization of other fuzzy clustering techniques like the axes-parellel version of the Gustafson-Kessel algorithm.

1 Introduction

The usual objective function in fuzzy clustering is of the form

$$J(X,U,v) \;=\; \sum_{i=1}^{c}\sum_{k=1}^{n}(u_{ik})^m d^2(v_i,x_k) \tag{1}$$

where c is the number of fuzzy clusters, $u_{ik} \in [0,1]$ is the membership degree of datum x_k to cluster i and $d(v_i,x_k)$ is the distance between cluster prototype v_i and datum x_k. In order to avoid the trivial solution $u_{ik}=0$, additional assumptions have to be made leading to probabilistic [1], possibilistic [7] or noise [2] clustering, for an overview see e.g. [5].

The prototypes can be simple vectors like the data as in the fuzzy c-means algorithm (FCM) or more complex structures like in the Gustafson-Kessel algorithm [4], in linear or shell clustering. In these cases the distance function d is not simply the Euclidean distance but some other measure depending on the type or form of the clusters.

In this paper we introduce a new distance measure that generalizes FCM as well as the Gustafson-Kessel algorithm restricted to diagonal fuzzy covariance matrices.

2 Context Sensitive Fuzzy Clustering

The distance between a datum x and a cluster (vector) v is defined by

$$d^2(v,x) = \sum_{I\in\mathcal{I}} \alpha_I \left(\sum_{s\in I} x^{(s)} \ - \ \sum_{s\in I} v^{(s)} \right)^2 . \tag{2}$$

$x^{(s)}$ and $v^{(s)}$ indicate the sth coordinates of the vectors x and v, respectively. $\mathcal{I}$ is a set of sets of indices (coordinates), i.e. $\mathcal{I} \subseteq 2^{\{1,\ldots,p\}}$ when we have to deal with p-dimensional data vectors. The parameters α_I can be considered as fixed or adapted during clustering individually for each cluster subject to the constraint

$$\prod_{I\in\mathcal{I}} \alpha_I = 1. \tag{3}$$

The idea of this context sensitive clustering is that certain subsets of the variables of data vectors yield similar values when we sum them up instead of comparing them one by one. A typical application of this approach is image recognition, where two similar images or regions might not correspond to each other pixel by pixel, but for instance the sum of the grey values in smaller windows might almost coincide.

Based on this approach we can derive an alternating optimization scheme for fuzzy clustering using this distance measure.

With condition (3) we obtain the Lagrange function

$$J_\lambda(X,U,v) = \sum_{i=1}^{c}\sum_{k=1}^{n}(u_{ik})^m \cdot \sum_{I\in\mathcal{I}} \alpha_I \left(\sum_{s\in I} x^{(s)} - \sum_{s\in I} v^{(s)} \right)^2 - \lambda \cdot \left(\prod_{L\in\mathcal{I}} \alpha_L - 1 \right). \tag{4}$$

If the parameter α_I should be adapted during the iteration procedure, differentiating (4) gives us a calculation instruction (5) for the parameter α_I as a necessary condition for the objective function to adopt a minimal value.

$$\alpha_I = \frac{\sqrt[card(\mathcal{I})]{\prod_{J\in\mathcal{I}} \sum_{k=1}^{n}\sum_{i=1}^{c} u_{ik}^m \cdot \left(\sum_{s\in J} \left(x_k^{(s)} - v_i^{(s)} \right) \right)^2}}{\sum_{k=1}^{n}\sum_{i=1}^{c} u_{ik}^m \cdot \left(\sum_{s\in I} \left(x_k^{(s)} - v_i^{(s)} \right) \right)^2}. \tag{5}$$

In a similar way we obtain a necessary condition for the cluster centers (6).

$$\sum_{I\in\mathcal{I}|r\in I} \left(\alpha_I \cdot \sum_{s\in I} v_i^{(s)} \right) = \frac{\sum_{k=1}^{n} \left(u_{ik}^m \cdot \sum_{I\in\mathcal{I}|r\in I} \left(\alpha_I \cdot \sum_{s\in I} x_k^{(s)} \right) \right)}{\sum_{k=1}^{n} u_{ik}^m} \tag{6}$$

Equation (6) is a system of linear equations, but variable and highly dependend on the choice of the sets in $\mathcal{I}$. So we decided to use the following heuristics (7) to estimate the parameters $v_i^{(s)}$ in (6).

$$\begin{aligned} v_i^{(q)} \quad = \quad & \frac{\sum_{k=1}^{n} \left(u_{ik}^m \cdot \sum_{I\in\mathcal{I}|q\in I} \left(\alpha_I \cdot \sum_{s\in I} x_k^{(s)} \right) \right)}{\left(\sum_{k=1}^{n} u_{ik}^m\right) \cdot \left(\sum_{I\in\mathcal{I}|q\in I} \alpha_I \right)} \\ & - \frac{\left(\sum_{k=1}^{n} (u_{ik}^m)\right) \cdot \left(\sum_{I\in\mathcal{I}|q\in I} \alpha_I \cdot \sum_{s\in I\setminus\{q\}} v_i^{(s)} \right)}{\left(\sum_{k=1}^{n} u_{ik}^m\right) \cdot \left(\sum_{I\in\mathcal{I}|q\in I} \alpha_I \right)} \end{aligned} \tag{7}$$

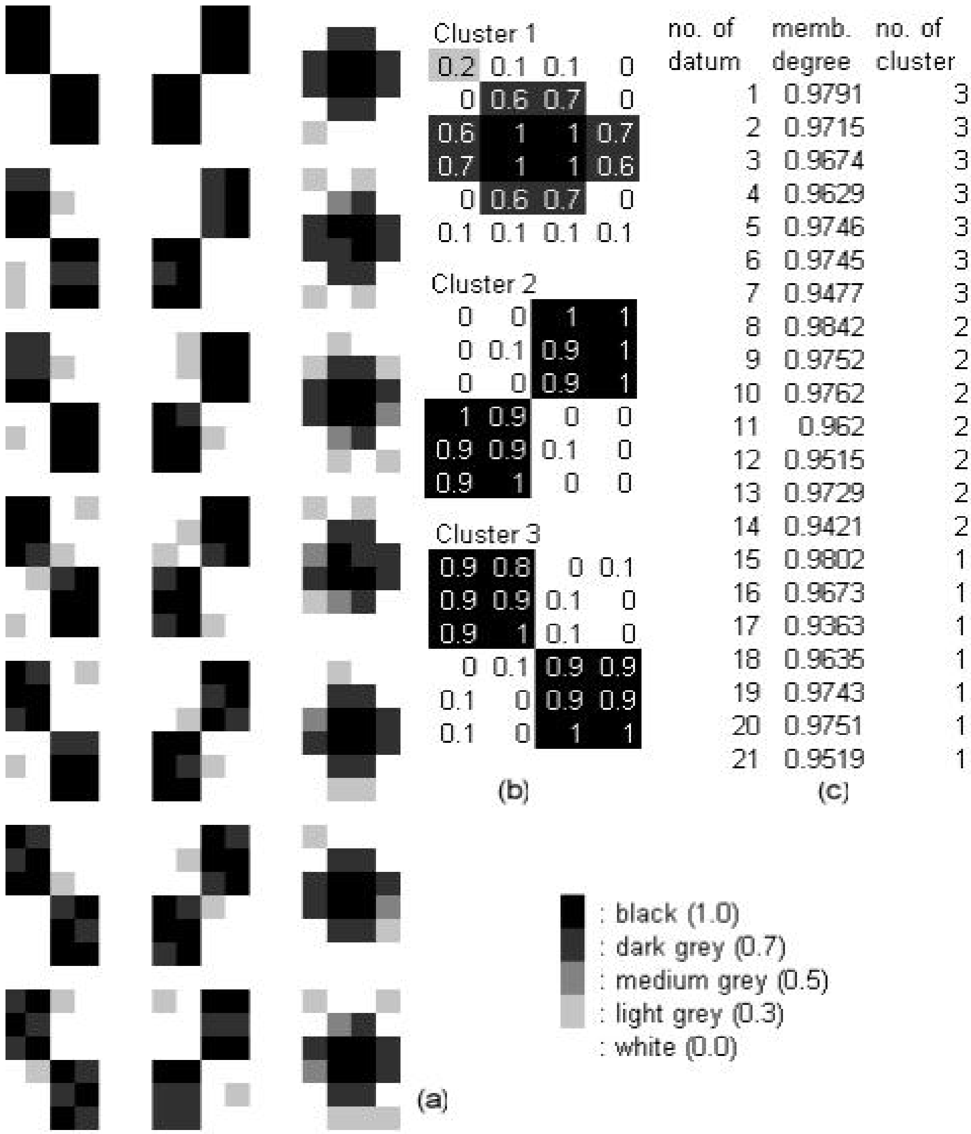

Figure 1: Box Images

In (7) we have to make sure that none of the parameters $v_i^{(q)}$ is allowed to be placed outside the domain of the corresponding data set's attribute. If values outside the domain would be allowed, the error of the former placed cluster centers would be neglected in placing the next coordinate $v_i^{(q)}$ in great distance of all observed data coordinates. In calculating $v_i^{(q)}$ the former calculated parameters $v_i^{(r)}$ $(r < q)$ are used in equation (7). Otherwise each prototype coordinate would adapt the whole error between the sum of corresponding data coordinates and the coordinates of the former prototype. A similar heuristic approach to determine the prototypes is also used for the fuzzy c-ellipses algorithm [3] and fuzzy c-rings algorithm [8] clustering techniques. The parameter α_I determines the influence of one particular subset of attributes. If e.g. the class determining areas of an image are known in advance, it is not necessary to adapt the α_I belonging to those areas (assuming that each I contains the variables of one significant area) during the clustering procedure. In the case that no supplementary information about the data set is given, it is possible to define more subsets I than are expected to be necessary for the task of pattern recognition and adapt the α_I in order to adapt the influence of certain subsets.

Our approach can be seen as a generalization of the axes-parallel version of the Gustafson-Kessel algorithm [6] that based on the distance function $d^2(x,(v,D)) = (x-v)^\top D(x-v)$ where D is a diagonal matrix with determinant 1. Therefore, if $d_1,\ldots,d_p$ are the diagonal elements of D, the distance function can be written as

$$d^2(x,(v,D)) = \sum_{s=1}^{p} d_s(x_s-v_s)^2 \tag{8}$$

with the constraint

$$\prod_{s=1}^{p} d_s = 1. \tag{9}$$

When we choose $\mathcal{I} = \{\{1\},\ldots,\{p\}\}$, then our initial equations (2) and (3) correspond to equations (8) and (9), respectively.

3 Examples

Figure 1 shows an example with three different kinds of grey-scale images. On the left side – (a) – the data set is shown. The grey scales are equivalent to real numbers as denoted at the bottom of the picture. During the clustering procedure we adapted the values for the α_I and set the number of clusters to three. The set of sets $\mathcal{I}$ is constructed of all sets containing one element – one particular point of the picture's area – and four sets with six points each (the top-left, top-right, bottom-left and bottom-right regions of the images). The resulting prototypes for the three clusters are shown in figure 1 (b). The highest membership degrees of the data points to the clusters are presented together with the corresponding clusters in (c). In figure 2 results for the data set from figure 1 (a) are shown. Here the set of sets $\mathcal{I}$ contains only the subsets with six elements each – as described above. Figure 2 (a) shows the resulting prototypes and (b) presents the highest membership degrees of the data points to the clusters together with the corresponding clusters. The corresponding clusters of the data points are determined correctly, also the prototypes for cluster one and three. Since all grey values are present in each particular subset I of $\mathcal{I}$ for the third group of data points (figure 1 (a)), the corresponding prototype for cluster two in figure 2 (a) is not able to reproduce the data images correctly – nevertheless the data points are correctly assigned.

4 Conclusions

Our clustering technique seems to be well suited to determine groups of similar images. Problems may arrise if no diverging areas for the groups or classes of images (describing a special class) can be found.

References

[1] J. Bezdek. *Pattern Recognition with Fuzzy Objective Function Algorithms.* Plenum Press, New York, 1981.

[2] R. Davé. Characterization and detection of noise in clustering. *Pattern Recognition Letters*, 12:657–664, 1991.

Cluster 1

0	0	1	1
0	0	1	1
0	0.3	0.7	0.9
1	1	0	0
1	1	0	0.1
1	0.5	0.1	0

Cluster 2

0.2	0.4	0.5	0.6
0.6	0.6	0.2	0.7
0.5	0.4	0.1	0.6
0.2	0.1	0	0
0.9	0.7	1	0.9
0.5	0.1	0.5	0.2

Cluster 3

1	1	0	0
1	1	0	0
1	0.4	0.3	0
0	0	1	1
0	0	1	1
0	0.3	0.8	0.7

(a)

no. of datum	memb. degree	no. of cluster
1	0.9759	3
2	0.9943	3
3	0.9777	3
4	0.9869	3
5	0.9985	3
6	0.9947	3
7	0.9832	3
8	0.9817	1
9	0.9828	1
10	0.9874	1
11	0.9949	1
12	0.997	1
13	0.9973	1
14	0.9737	1
15	0.9918	2
16	0.9952	2
17	0.9868	2
18	0.9951	2
19	0.9955	2
20	0.996	2
21	0.9845	2

(b)

Figure 2: Result with four subsets

[3] I. Gath and D. Hoory. Fuzzy clustering of elliptic ring-shaped clusters. *Pattern Recognition Letters*, 16:727–741, 1995.

[4] D. Gustafson and W. Kessel. Fuzzy clustering with a fuzzy covariance matrix. In *IEEE CDC*, pages 761–766, San Diego, 1979.

[5] F. Höppner, F. Klawonn, R. Kruse, and T. Runkler. *Fuzzy Cluster Analysis*. Wiley, Chichester, 1999.

[6] F. Klawonn and R. Kruse. Constructing a fuzzy controller from data. *Fuzzy Sets and Systems*, 85:177–193, 1997.

[7] R. Krishnapuram and J. Keller. A possibilistic approach to clustering. *IEEE Trans. on Fuzzy Systems*, 1:98–110, 1993.

[8] Y. Man and I. Gath. Detection and separation of ring-shaped clusters using fuzzy clustering. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 16:855–861, 1994.