Fuzzy Clustering with Weighting of Data Variables *

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

Frank Klawonn
Dept. 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

Summary

We introduce an objective function-based fuzzy clustering technique that assigns one influence parameter to each single data variable for each cluster. Our method is not only suited to detect structures or groups in unevenly over the structure's single domains distributed data, but gives also information about the influence of individual variables on the detected groups. In addition, our approach can be seen as a generalization of the well-known fuzzy c-means clustering algorithm.

Keywords: Fuzzy clustering, variable selection, generalized fuzzy c-means

1 Introduction

The common objective function to be minimized in fuzzy clustering is of the form

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

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 [8] or noise [2] clustering. Parameter $m \in \mathbb{R}_{\geq 1}$ is called fuzzifier. For $m \to 1$, we have for the membership degrees $u_{ik} \to 0/1$, so the classification tends to be crisp. If $m \to \infty$, then $u_{ik} \to \frac{1}{c}$, where c is the number of clusters.

* This work was supported by the European Union under grant EFRE 98.053

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 [7]. 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. A thorough overview on objective function-based fuzzy clustering techniques can be found in [5].

In this paper we introduce a new distance measure that generalizes the FCM in adding a parameter that determines the influence of certain data attributes for some cluster.

2 Attribute Weighting Fuzzy Clustering

Especially in data where few variables determine particular clusters other variables may disguise the structure and should therefore not be considered to find these clusters. This can be done by weighting single attributes for each cluster as we have done with our new distance measure.

The distance between a datum x_k and a cluster (vector) v_i is defined by

$$d^2(v_i, x_k) = \sum_{s=1}^{p} \alpha_{is}^t \cdot \left(x_k^{(s)} - v_i^{(s)} \right)^2. \qquad (2)$$

$x_k^{(s)}$ and $v_i^{(s)}$ indicate the sth coordinates of the vectors x_k and v_i, respectively. The number of variables or attributes is denoted by p. α_{is}^t is a parameter determining the influence of attribute (coordinate) s for cluster i. The parameters α_{is} can be considered as fixed or adapted individually for each cluster during clustering due to the constraint

$$\forall i \in \{1, \cdots, c\} : \qquad \sum_{s=1}^{p} \alpha_{is} = a, \qquad (3)$$

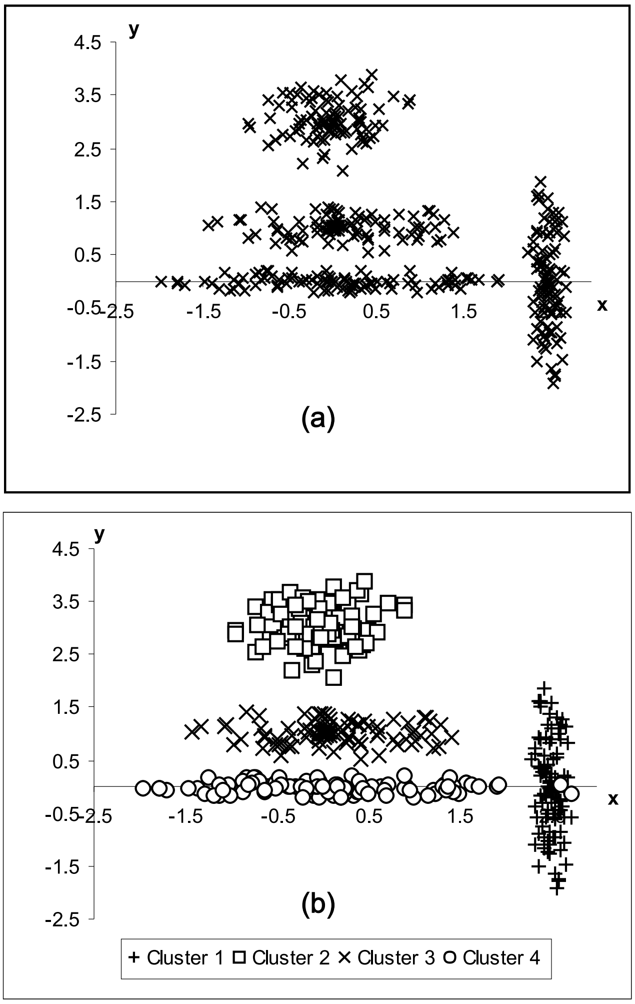

Figure 1: Ellipsoidal clusters

where $a \in \mathbb{R}$ is a constant parameter, e.g. $a = 1$ or $a = c$. If we would neglect this constraint, we would obtain the trivial solution $\alpha_{is} = 0$ for all i and s.

We will see that the exponent $t \in \mathbb{R}_{\geq 1}$ in equation (2) has a similar influence on the parameters α_{is} as the fuzzifier m on the membership degrees u_{ik}. For $t \to 1$ the α_{is} tend to be 1 or 0 – either one attribute has unrestricted influence or no influence at all. On the other hand, if $t \to \infty$, all attributes get the same influence on the cluster structure, i.e. $\alpha_{is} \to \frac{1}{a}$ for all i and s.

Based on this approach we can derive an alternating optimization scheme for fuzzy clustering using distance measure (2).

To adapt the influence parameters α_{is} we have to determine a necessary condition for the values α_{is} so that the objective function achieves an optimum value. With condition (3) we obtain the Lagrange function

$$J_\lambda(X,U,v) = \sum_{i=1}^{c}\sum_{k=1}^{n}(u_{ik})^m \cdot \sum_{s=1}^{p} \alpha_{is}^t \left(x_k^{(s)} - v_i^{(s)}\right)^2 - \sum_{i=1}^{c}\lambda_i\left(\sum_{s=1}^{p}\alpha_{is} - 1\right). \tag{4}$$

Differentiating (4) leads to equation (5) for the parameter α_{is} as a necessary condition for the objective function to have a minimum. This equation can be used for updating α_{is} during the alternating clustering procedure.

$$\alpha_{is} = \frac{1}{\sum_{r=1}^{p}\left(\frac{\sum_{k=1}^{n} u_{ik}^m \cdot \left(x_k^{(s)} - v_i^{(s)}\right)^2}{\sum_{k=1}^{n} u_{ik}^m \cdot \left(x_k^{(r)} - v_i^{(r)}\right)^2}\right)^{\frac{1}{t-1}}}. \tag{5}$$

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

$$v_i^{(s)} = \frac{\sum_{k=1}^{n} u_{ik}^m \cdot x_k^{(s)}}{\sum_{k=1}^{n} u_{ik}^m} \tag{6}$$

Equation (6) is the same as in FCM.

3 Examples

In Figure 1 a data set consisting of four ellipsoidal groups is shown. Part (a) presents the original data set and part (b) represents the clustering result obtained by our attribute weighting clustering technique where a datum is assigned to the cluster to which it has the highest membership degree (maximum defuzzification). In this case we have set the fuzzifier m and the exponent t to 2.0. For parameter a we have chosen the value 1. However, the clustering result depends more on a suitable initialization of cluster centres than the choice of parameters m, t and a. Table 1 lists the minimum and maximum attribute values for all clusters.

Table 1: Minimum/maximum data values for each cluster

	attributes			
	x		y	
	min	max	min	max
cluster 1	2.28	2.71	−1.93	1.85
cluster 2	−0.96	0.90	2.07	3.88
cluster 3	−1.42	1.40	0.54	1.40
cluster 4	−1.97	1.93	−0.21	0.21

In table 2 *cluster* 1 represents the ellipsoidal group with greatest x-values in the right part of figure 1.

From top to bottom in the left part of figure 1 are the clusters *cluster* 2, *cluster* 3 and *cluster* 4. The scale values α_{is} were adapted during the clustering procedure. It is obvious, that for each cluster the more the data coordinates are scattered around the corresponding prototype's coordinate, the less is the influence of the corresponding attribute for that cluster. In our example in figure 1 the two attribute influence parameters α_{is} for *cluster* 2 have nearly the same value. The data coordinates are approximately uniformly distributed for the two domains of this cluster. For clusters 3 and 4, the data values for attribute x are scattered widely whereas the values for attribute y have a small range – so the influence parameters α_{ix} are small in comparison to α_{iy} for clusters 3 and 4. In case of *cluster* 1 the data values for attribute y are scattered widely, resulting in a high value for influence parameter α_{1x}.

Table 2: Attribute weights for ellipsoidal data set

	attributes	
cluster i	α_{ix}	α_{iy}
cluster 1	0.99	0.01
cluster 2	0.49	0.51
cluster 3	0.08	0.92
cluster 4	0.01	0.99

Figure 2 presents the clustering result for the example data set generated by the FCM clustering technique with fuzzifier $m = 2$ as in our approach. Using the Euclidean distance measure, the FCM is not well suited to detect ellipsoidal structures in data. One indication for the suitability of a clustering result is following value. Of the c membership degrees associated with each datum, we only consider the highest membership degree (i.e. the membership degree to the cluster to which we would assign the datum by maximum defuzzification) and compute the mean value of these membership degrees. Here, the mean value for FCM is 0.81 in comparison to 0.96 for our attribute weighting clustering technique. Nevertheless are the methods from Gustafson and Kessel [4] or Gath and Geva [3] well suited to detect the structures of our example data.

4 Conclusions

The presented clustering technique gives us information about the influence of particular variables or attributes of the data set on special clusters. This knowledge can be used e.g. in classification tasks to determine or detect class defining attributes. Without ignoring one data attribute for the whole classification it is possible to reduce the influence of that at-

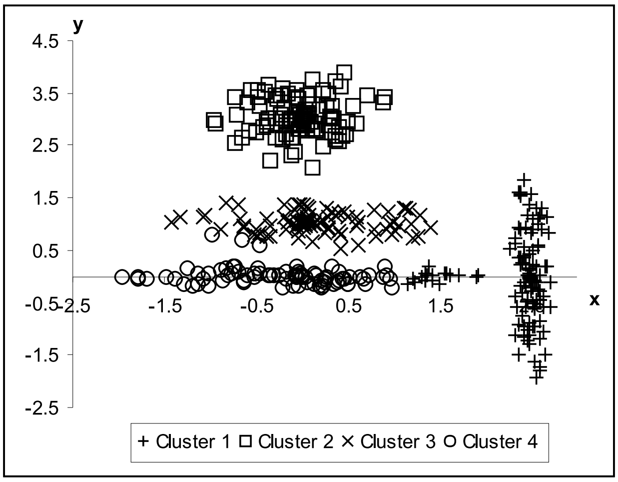

Figure 2: Results for ellipsoidal clusters with Fuzzy-c-means

tribute on only some clusters. In that way, attribute weights could help to partition the whole data set into smaller data parts depending on the same attributes. Analysing the smaller parts with a reduced number of attributes would reduce the computation effort. Real data sets soon get immense large as e.g. in the mentioned EU project where we analyse sound measured on tyres with different pressures (2520 data sets with 200 sound attributes and 12 different pressures as classification attribute). Here attribute weighting could not only be helpful in reducing computation time but also to reduce the future expense of measuring.

Our fuzzy clustering approach is also well suited for deriving rules from the clusters. Since the weighting of the attributes for each cluster provides information about the importance of the variables, we can neglect variables with very small weighting factors in the rules. In the example in the previous section this would mean that we would derive a fuzzy rule from cluster 2 invoking only the variable y.

It should be noted that our approach is also related to the simplified version of the Gustafson-Kessel algorithm described in [6] that introduces a diagonal matrix for each cluster. The diagonal elements are weights for the attributes in the same way as we use them here, except for our exponent t. However, the constraint is that the determinant is constant, i.e. the sum in equation (3) is replaced by a product. The advantage of our new approach is that we can control how strong the influence of single variables can be by the parameter t.

Note that our approach differs from the idea to carry out a cluster analysis first and then apply something like a principal component analysis to each cluster.

This would mean that the clustering must take all attributes into account, whereas in our approach the selection of relevant variables is already carried out during the clustering.

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.

[3] I. Gath and A. Geva. Unsupervised optimal fuzzy clustering. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 11:773–781, 1989.

[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, H. Frigui, and O. Nasraoui. Fuzzy and possibilistic shell clustering algorithms and their application to boundary detection and surface approximation - part 1 and 2. *IEEE Trans. on Fuzzy Systems*, 3:29–60, 1995.

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