Detection of clusters of distinct geometry:
A step towards generalised fuzzy clustering

Department of Electronics and Electrical Communication Engineering, Indian Institute of Technology, Kharagpur 721 302, India
Received 3 May 1994; revised 16 January 1995

Abstract
This paper presents a method of identification of clusters even if they are of distinctly different geometrical categories. An optimal fuzzy partition is achieved by minimisation of a modified objective function. This technique is an elegant one and works well in segmentation of 3-D images.

Keywords: Fuzzy sets, Clustering, Fuzzy partition space

1. Introduction
The cluster analysis techniques encompass a broad spectrum of methodologies where a data set $X$ is partitioned into $C$ subsets $X_1, X_2, \ldots, X_C$ which are pairwise disjoint and reproduce $X$ via union. The inherent problem in all methodologies are hidden in the underlying axioms that each point is unequivocally assigned a partition and there is no way of ascertaining any similarity measure of each sample to other members in the sample space. The solution lies in assigning a membership grade to each sample for its belongingness to each cluster, which results in a fuzzy partition space.

The problem of fuzzy clustering may be presented as follows. A finite set of $N$ samples is partitioned in $C$ pattern classes on a fuzzy partition space $M \in \mathbb{R}^{CN} = \{ \mu_{ij} \}$ where $\mu_{ij}$ is the membership of the $j$th sample to the $i$th cluster. The membership function $\mu_{ij}$ has the following properties:

$$\mu_{ij} \in [0, 1]; \quad \sum_{j=1}^{C} \mu_{ij} = 1; \quad \sum_{i=1}^{n} \mu_{ij} \in (0, N).$$

Bezdek (1981) presents a very detailed and elegant description of objective function based clustering. Most of these approaches use a least square objective function which is iteratively minimised to achieve an optimal partition. In the Fuzzy C Means (FCM) or Fuzzy ISODATA algorithm (Bezdek, 1976), the samples are clustered into hyperspherical clusters, while Bezdek et al. (1981)'s linear varieties algorithm can detect clusters of linear structure. The infinite extension of a linear structure is limited by providing the distance as a weighted sum of the distance from the mean and the linear prototype. The problem of identification of higher-order structures has generated interest in recent times (Dave, 1992; Krishnapuram et al., 1993). The approaches include the adaptive variation of a matrix norm which in turn becomes the distance
measure for clustering. The more promising strategy is fitting prototypes of proper order and to detect the partitions based on the distance from them.

But when clusters have distinct structures, their identification becomes difficult, since the error will be minimum if the order of prototype chosen is high. Thus hyperspherical clusters will be identified as a hyperplanar cluster when we attempt to fit a linear prototype and this results in misclassification. Here we have introduced a modified least square objective function which is minimised using Lagrange multipliers to get the proper classification even if the clusters are of distinctly different geometrical structures.

2. Formulation of generalised clustering strategy

Bezdek (1981), while presenting a family of algorithms, optimised a least square objective function

\[ J = \sum_{i=1}^{C} \sum_{j=1}^{N} (\mu_{ij})^T (d_{ij})^2 \]

to achieve an optimal fuzzy partition through Picard iteration. In the FCM algorithm, \( d_{ij} \) corresponds to the distance of the \( j \)th sample point from the \( i \)th cluster center, while the linear varieties algorithm considered \( d_{ij} \) as the orthogonal distance from the \( j \)th sample point to the \( i \)th linear structure. Such algorithms are extremely well suited to identify clusters of the same geometry or the same order, i.e., the clusters should have homogeneous order. In this paper, we propose a generalised algorithm for partitioning a set of points having a combination of different geometries.

The problem of fuzzy clustering is to partition \( N \) distinct sample points into \( C \) partitions, where each cluster prototype is a combination of all the ordered prototypes, i.e., from zeroeth to \( P \)th order. This can be achieved by minimising the objective function

\[ J(M, V, \Gamma) = \sum_{i=1}^{C} \sum_{j=1}^{N} (\mu_{ij})^T (d_{ij})^2 \]

where \( M \) is a fuzzy partition; \( V \) is a set of prototypes of each cluster \( i = 1, 2, \ldots, C \) of order \( k = 0, 1, \ldots, P; \Gamma \in \mathbb{R}^{(P+1)C} \) is a matrix \([\gamma_{ki}]\) where \( \gamma_{ki} \) is the weighting of the \( i \)th cluster to the \( k \)th order prototype; \( d_{ij} \) is the minimum distance of the \( j \)th sample to the \( k \)th order prototype in the \( i \)th cluster, and \( \tau \in [1, \infty) \) is a real number which controls the amount of fuzzines in clustering. When \( \tau = 1 \) the partition becomes a hard one, which is a special case of fuzzy partition.

Here we consider that each cluster will have a weighting to each of the structures of different order, i.e., spherical, linear, quadratic etc. Here \( \gamma_{ki} \) is the weighting of the \( i \)th cluster to the \( k \)th order prototype. In the objective function (1) if the order \( P = 1 \), we consider that each cluster geometry consists of a combination of spherical and linear prototypes. Similarly \( P = 2 \) implies that the cluster geometry consists of a combination of spherical, linear and quadratic prototypes.

The minimisation of \( J(M, V, \Gamma) \) is done with respect to a single variable by fixing the other two variables constant.

2.1. Identification of cluster weights

Minimisation of \( J \) by fixing \( M \) and \( V \) constant, is performed with the constraint

\[ \sum_{i=0}^{P} \gamma_{ki} = 1. \]

Let

\[ f(\Gamma) = \sum_{i=1}^{C} \sum_{j=1}^{N} (\mu_{ij})^T \sum_{k=0}^{P} (\gamma_{ki} d_{ij}^k)^2. \]

Since the columns of \( \Gamma \) are independent, minimisation can be done independently in each column. The respective Lagrangian is

\[ F(\lambda, \gamma_i) = \sum_{j=1}^{N} (\mu_{ij})^T \sum_{k=0}^{P} (\gamma_{ki} d_{ij}^k)^2 - \lambda (\sum_{k=0}^{P} \gamma_{ki} - 1) \]

where \( \gamma_i \) is the \( i \)th column of \( \Gamma \). Differentiating (2) with respect to \( \gamma_{mn} \),

\[ \frac{\partial F}{\partial \gamma_{mn}} = 2 \sum_{j=1}^{N} (\mu_{nj})^T \gamma_{mn} (d_{nj}^P)^2 - \lambda = 0, \]

leads to
Substituting (4) into the constraint \( \sum_{j=0}^{p} y_{ij} = 1 \), gives

\[
\lambda = 1 \left( \sum_{j=0}^{p} 1 \left( \sum_{j=1}^{N} (\mu_{n_j})^2 (d_{n_j}^2) \right) \right) \tag{5}
\]

Substituting (5) into (4) we get

\[
y_{mn} = 1 \left( \sum_{j=0}^{p} \sum_{j=1}^{N} (\mu_{n_j})^2 (d_{n_j}^2) \right) \left( \sum_{j=1}^{N} (\mu_{n_j})^2 (d_{n_j}^2) \right)^{-1} \tag{6}
\]

Eq. (6) gives the method of updation of cluster weights for different orders. As may be observed from Eq. (6), the updation is carried out using the squared error for each of the clusters of all orders. It is interesting to note that a cluster should be assigned more weight to a particular geometry, i.e., a particular order, in case the squared error of that cluster for that particular order is less. It may be observed that the proposed modified objective function and the weight updation strategy is a more general one leading to an optimal set of partitions. If \( y_{k_1} = 1 \) for \( k = 0 \) and is zero for other values of \( k \), the algorithm reduces to FCM. Also if \( y_{k_1} = 1 \) for \( k = 1 \) and \( y_{k_1} = 0 \) for \( k \) other than one, the algorithm reduces to the Fuzzy C Varieties algorithm.

2.2. Calculating cluster prototypes

For different clusters, the prototypes are calculated by minimising least square functions depending on the order.

The zeroth-order prototype is the cluster centre which can be calculated in a similar manner to that of (11.3.b) of (Bezdek, 1981)

\[
\mu_j = \left( \sum_{j=1}^{N} (\mu_{n_j})^2 x_j \right) \left( \sum_{j=1}^{N} (\mu_{n_j})^2 \right) \tag{7}
\]

where \( x_j \) is the \( j \)th sample. The first-order prototype is defined by the principal eigenvectors of the fuzzy
Following the procedure as in Section 2.1,
\[ \mu_{mn} = \frac{1}{1 + \sum_{l=1}^{C} \left( \frac{\sum_{k=0}^{p} (\gamma_{km} d_{ml}^k)^2}{\sum_{k=0}^{p} (\gamma_{kl} d_{ln}^k)^2} \right)^{1/\tau} - 1}. \] (9)

Eq. (9) is similar to the membership updation formula (11.3a) of (Bezdek, 1981) in structure but the simple distance measure is replaced by a composition of distances which enables us to detect clusters even if their geometry is different. The minimisation of J is carried out iteratively with the help of (6) and (9) by taking into consideration the cluster prototypes of appropriate order depending on the requirement.

3. Results

The above described clustering method has been successfully verified in a number of cases. First we present a problem of partitioning a data set of 50 synthetic two-dimensional samples having three distinct clusters, one linear and the other two spherical. This set is shown in Fig. 1(a). When we attempt to fit linear or spherical cluster structures, misclassification results as may be observed from Fig. 1(b) and Fig. 1(c) respectively, i.e., the FCM or linear varieties algorithm fails to give a proper partition. The new strategy, however, detects the clusters properly and results are as shown in Fig. 1(d). In Figs. 1(b), 1(c), and 1(d), the fuzzy partition is hardened by putting each sample in a cluster to which it has maximum affinity. The famous Gustafson’s cross (Fig. 2(a)) has been partitioned into two clusters as shown in Fig. 2(b). This shows that the proposed strategy is more general and works well with many of the existing data sets. In Fig. 3(a) sample data of a pair of crossing clusters of linear structures along with a spherical cluster is presented and the resulting partitions are shown in Fig. 3(b). In all the cases the value of P in Eq. (1) is one and the norm is simply Euclidean. Distinct clusters are represented by different marks. As pointed out in the preceding section there is always a tendency to fit a higher-order cluster, which has been controlled here by introducing a parameter \( f(k) \) and modifying \( d_{ij}^k \) by \( f(k) d_{ij}^k \) where \( f(k) \) is a monotonically increasing function. If \( f(k) \) is very high there is a tendency to group the sample points to a lower-order cluster and...
4. Conclusion

A generalised fuzzy clustering strategy is presented here for efficient clustering. This methodology can detect true cluster substructures and has been observed to work well. The strategy has been applied to surface characterisation in 3-D image analysis for robotic vision. This will be reported shortly.

Acknowledgment

The authors wish to thank the unknown referees for their detailed comments and suggestions.

References


Fig. 3. (a) A synthetic sample set. (b) Results of applying the proposed algorithm.

it has been observed that an optimal value of \( f(k) = (k - 1)^{2.5} \) yields excellent results in all the cases.