A Brief Overview of Prototype Based Clustering Techniques

Olfa Nasraoui
Department of Computer Engineering & Computer Science
University of Louisville,
olfa.nasraoui_AT_louisville.edu

Prototype-Based Clustering Techniques

Clustering aims at classifying the unlabeled points in a data set into different groups or clusters, such that members of the same cluster are as similar as possible, while members of different clusters are as dissimilar as possible. Because there is no *a priori* knowledge about the class labels, clustering is also called unsupervised classification. Several approaches to clustering exist, and probably differ because they have originated in different domains of artificial intelligence and statistics. For example, graph-theoretic and tree-based techniques are popular in the machine learning community; while objective function-driven or prototype-based clustering methods such as the $K-$Means and Gaussian Mixture modeling have long been used in statistical pattern recongnition. We concentrate on prototype based clustering methods because they lend themselves more easily to robustifying efforts based on robust statistics. Most prototype based clustering methods are based on the $K-$Means and its fuzzy counterpart, the Fuzzy $C-$Means (FCM) [Bez81] algorithms.

Let $\mathcal{X} = \{\mathbf{x}_j | j = 1, \ldots, N\}$ be a set of feature vectors in an $n-$dimensional feature space with coordinate axis labels $[x_1, x_2, \ldots, x_n]$, where $\mathbf{x}_j = [x_{j1}, x_{j2}, \ldots, x_{jn}]$. Let $\mathbf{B} = (\beta_1, \ldots, \beta_c)$ represent a C-tuple of prototypes each of which characterizes one of the C clusters. Each β_i consists of a set of parameters. The $K-$Means has the following objective

$$\min_{\mathbf{B}} \quad \left\{ J = \sum_{i=1}^{C} \sum_{\mathbf{x}_j \in \mathcal{X}_i} d^2(\mathbf{x}_j, \beta_i) \right\}, \tag{1}$$

where $d^2(\mathbf{x}_j, \beta_i) = d_{ij}^2$ represents the distance from a feature point $\mathbf{x}_j$ to the prototype β_i, and $\mathcal{X}_i$, the i^{th} cluster, is given by

$$\mathcal{X}_i = \{\mathbf{x}_j \in \mathcal{X} | \, d_{ij}^2 = \min_{k=1}^{C} d_{kj}^2\} \tag{2}$$

The optimal prototype parameters β_i of the i^{th} cluster are derived by setting $\frac{\partial J}{\partial \beta_i} = \mathbf{0}$. For instance if d_{ij}^2 is the squared Euclidean distance $d_{ij}^2 = \|\mathbf{x}_j - \mathbf{c}_i\|^2$, then the center $\mathbf{c}_i$ is given by

$$\mathbf{c}_i = \frac{\sum_{\mathbf{x}_j \in \mathcal{X}_i} \mathbf{x}_j}{N} \tag{3}$$

The $K-$Means algorithm consists of alternating updates of the centers using (3) and the partition using (2), until convergence or for a maximum number of iterations.

It is known that for complex data sets containing overlapping clusters, fuzzy partitions model the data better than their crisp counterparts. In particular, fuzzy memberships are richer than crisp memberships in describing the degrees of belongingness of data points lying in the areas of overlap. Moreover, fuzzy partitions generally make the optimization process less prone to local or sub-optimal solutions. With a fuzzy partition, a data point $\mathbf{x}_j$ belongs to each cluster, $\mathcal{X}_i$, to a varying degree called fuzzy membership u_{ij}. A fuzzy partiton, usually represented by the $C \times N$ matrix $\mathbf{U}$ is called a constrained fuzzy $C-$partition of $\mathcal{X}$ if the entries of $\mathbf{U}$ satisfy the following constraints [Bez81],

$$0 \leq u_{ij} \leq 1$$

$$\sum_{i=1}^{C} u_{ij} = 1 \; \forall j = 1 \cdots N$$

$$0 < \sum_{j=1}^{N} u_{ij} < N \; \forall i = 1 \cdots C$$

The Fuzzy C Means (FCM) [Rus69,Dun74,Bez81] algorithm uses the following criterion:

$$\min_{\mathbf{B},\mathbf{U}} \quad \sum_{i=1}^{C} \sum_{j=1}^{N} u_{ij}^m d_{ij}^2. \tag{4}$$

The optimal prototype parameters β_i of the i^{th} cluster are derived by setting $\frac{\partial J}{\partial \beta_i} = \mathbf{0}$. For instance if d_{ij}^2 is the squared Euclidean distance, then the center $\mathbf{c_i}$ is given by

$$\mathbf{c_i} = \frac{\sum_{j=1}^{N} u_{ij}^m \mathbf{x}_j}{\sum_{j=1}^{N} u_{ij}^m} \tag{5}$$

The optimal FCM memberships for (4) can be shown to be [Bez81]

$$u_{ij} = \frac{\left(\frac{1}{d_{ij}^2}\right)^{\frac{1}{m-1}}}{\sum_{k=1}^{C} \left(\frac{1}{d_{kj}^2}\right)^{\frac{1}{m-1}}}. \tag{6}$$

Therefore, the optimization process consists of alternating updates of the memberships, as given by (6), and the cluster centers as given by (5). The fuzzy memberships allow each data point to belong to all clusters to a varying degree of membership. Hence, fuzzy partitions involve less commitment than their hard or crisp counterparts in every step of the optimization process. This results in a lower sensitivity to initialization. By changing the distance measure in the objective function of the FCM, it can be generalized to seek clusters of various shapes such as lines, curves, planar and quadric surfaces [GK79,KFN95]

Acknowledgment

This work is supported by the National Science Foundation (CAREER Award IIS-0133948 to O. Nasraoui). Partial support of earlier stages of this work by the Office of Naval Research grant (N000014-96-1-0439) and by the National Science Foundation Grant IIS 9800899 is also gratefully acknowledged.

Bibliography

Bez81

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

Dun74

J. C. Dunn.
A fuzzy relative of the isodata process and its use in detecting compact, well separated clusters.
J. Cybernetics, 3:32-57, 1974.

GK79

E. E. Gustafson and W. C. Kessel.
Fuzzy clustering with a fuzzy covariance matrix.
In *IEEE CDC*, pages 761-766, San Diego, California, 1979.

KFN95
R. Krishnapuram, H. Frigui, and O. Nasraoui.
Fuzzy and possibilistic shell clustering algorithms and their application to boundary detection and surface approximation-parts i and ii.
IEEE Trans. Fuzzy Systems, 3(1):29-60, 1995.

Rus69
E. Ruspini.
A new approach to clustering.
Information Control, 15:22-32, 1969.

About this document ...

A Brief Overview of Prototype Based Clustering Techniques

This document was generated using the **LaTeX2HTML** translator Version 2002 (1.62)

The command line arguments were:
latex2html `-split 0 -image_type gif PrototypeBasedClustering.tex`