Paper:

Algorithms for Sequential Extraction of Clusters by Possibilistic Method and Comparison with Mountain Clustering

Sadaaki Miyamoto*, **Youhei Kuroda, and Kenta Arai**

*Department of Risk Engineering, School of Systems and Information Engineering, University of Tsukuba
Ibaraki 305-8573, Japan
E-mail: miyamoto@risk.tsukuba.ac.jp

[Received October 10, 2007; accepted February 15, 2008]

In addition to fuzzy c-means, possibilistic clustering is useful because it is robust against noise in data. The generated clusters are, however, strongly dependent on an initial value. We propose a family of algorithms for sequentially generating clusters "one cluster at a time," which includes possibilistic medoid clustering. These algorithms automatically determine the number of clusters. Due to possibilistic clustering's similarity to the mountain clustering by Yager and Filev, we compare their formulation and performance in numerical examples.

Keywords: possibilistic clustering, sequential extraction of clusters, mountain clustering

1. Introduction

Fuzzy c-means clustering currently plays the central role in clustering algorithms [1–3], but there is another useful tool of possibilistic clustering [4], which both resembles and differs from fuzzy c-means. Possibilistic clustering is robust against noise in data [5]. Results of possibilistic clustering, however, tend to strongly dependent on initial values. Although there is a proposal of an algorithm generating "one cluster at a time" [5] to mitigate this drawback, this proposal has not yet been put into a concrete algorithm.

We propose algorithms for sequentially extracting clusters one by one and present an objective function having the form $J(V)$ of a variable of cluster centers, instead of $J(U,V)$ of the two variables of memberships and centers in fuzzy c-means and possibilistic clustering in the original formulation.

We also consider how these algorithms are related to the mountain clustering algorithm [6], and present examples showing properties of the proposed algorithms and compare performance of the proposed method and the mountain clustering.

Our proposal determines the number of clusters automatically, and more efficient than the mountain clustering, which is also sequential and similar to our proposal.

2. Possibilistic Clustering Formulations

An object of clustering is assumed to be a vector in p-dimensional Euclidean space: $x_k = (x^1, \ldots, x_k^p) \in \boldsymbol{R}^p$, $(k = 1, \ldots, n)$, and the set of objects is denoted by $X = \{x_1, \ldots, x_n\}$. Cluster centers are denoted by $v_i = (v_i^1, \ldots, v_i^p)^T$, $i = 1, \ldots, c$, where c is the number of clusters. Simplified symbol $V = (v_1, \ldots, v_c)$ is used for the collection of c cluster centers, whereas cluster membership matrix $U = (u_{ki})$, $(i = 1, \ldots, c, k = 1, \ldots, n)$ is used as usual, where u_{ki} is the degree of belongingness of object x_k to cluster i. In fuzzy c-means clustering, the constraint for a fuzzy partition is given by

$$M = \{U = (u_{ki}) : \sum_{i=1}^{c} u_{ki} = 1, \forall k; u_{kj} > 0, \forall j, k\},$$

while in possibilistic clustering we set

$$M = \{U = (u_{ki}) : u_{kj} \geq 0, \forall j, k\}, \quad \ldots\ldots \quad (1)$$

where we have omitted original constraint $0 < \sum_k u_{ki} \leq N$ proposed in [4], for simplicity.

The dissimilarity of clustering is the standard squared Euclidean distance between an individual and a cluster center:

$$D_{ki} = \|x_k - v_i\|^2.$$

We use the dissimilarity between generic elements $x, y \in \boldsymbol{R}^p$, denoted by

$$D(x,y) = \|x - y\|^2.$$

2.1. Fuzzy c-Means and Possibilistic Clustering

Fuzzy c-means [1] and possibilistic clustering are based on the optimization of an objective function. We consider the following two:

$$J_e(U,V) = \sum_{k=1}^{n} \sum_{i=1}^{c} \{u_{ki} D_{ki} + \lambda^{-1} u_{ki} (\log u_{ki} - 1)\} \quad (2)$$

$$J_2(U,V) = \sum_{k=1}^{n} \sum_{i=1}^{c} \{(u_{ki})^2 D_{ki} + \zeta^{-1} (1 - u_{ki})^2\} \quad (3)$$

J_e is an entropy-based objective function [7], and J_2 is a function for the possibilistic clustering [4] with a restriction to $m = 2$ and $\eta_i = \zeta^{-1}$ $(1 \leq i \leq c)$.

We focus on the possibilistic clustering, because the fuzzy c-means objective function by Bezdek [1] and Dunn [8] provides no help because it gives the trivial solution of $u_{ki} = 0$ in possibilistic clustering, while Eqs. (2) and (3) are useful for both possibilistic clustering and fuzzy c-means.

In the alternative optimization **PCM** of fuzzy c-means clustering, we use $J = J_e$ or J_2 and constraint (1) for possibilistic clustering.

Algorithm PCM
PCM0: Set an initial value $\bar{V}$.
PCM1: Find the optimal solution of J for U while V is fixed: put

$$\bar{U} = \arg\min_{U \in M} J(U, \bar{V}).$$

PCM2: Find the optimal solution of J for V while U is fixed: put

$$\bar{V} = \arg\min_{V} J(\bar{U}, V).$$

PCM3: If solution $(\bar{U}, \bar{V})$ is convergent, stop; else go to **PCM1**.
End PCM.

We show solutions for each step, writing u_{ki} instead of $\bar{u}_{ki}$, v_i instead of $\bar{v}_i$, without confusion, for simplicity.

$$u_{ki} = \exp(-\lambda D_{ki}) \quad \ldots\ldots\ldots\ldots \quad (4)$$

for J_e, or

$$u_{ki} = \frac{1}{1 + \zeta D_{ki}} \quad \ldots\ldots\ldots\ldots \quad (5)$$

for J_2, while

$$v_i = \frac{\sum_{k=1}^{n} (u_{ki})^m x_k}{\sum_{k=1}^{n} (u_{ki})^m}. \quad \ldots\ldots\ldots\ldots \quad (6)$$

in **PCM2** where $m = 1$ for J_e and $m = 2$ for J_2.

3. Sequential Extraction of Clusters

We define two functions related to Eqs. (4) and (5):

$$U_e(x_k, y) = \exp(-\lambda D(x_k, y)) \quad \ldots\ldots\ldots \quad (7)$$

$$U_2(x_k, y) = \frac{1}{1 + \zeta D(x_k, y)} \quad \ldots\ldots\ldots \quad (8)$$

Note $U_e(x_k, v_i) = u_{ki}$ for J_e; $U_2(x_k, v_i) = u_{ki}$ for J_2.

To see the properties of possibilistic clustering, we substitute $U(V) = (U_e(x_k, v_i))_{i=1,\ldots,c}$ into $J_e(U, V)$ where v_i is regarded as a variable, yielding

$$J_e(U(V), V) = -\lambda \sum_{i=1}^{c} \sum_{k=1}^{n} \exp(-\lambda D(x_k, v_i)).$$

If we put $J'_e(V) = J_e(U(V), V)$ and

$$j_e(y) = -\sum_{k=1}^{n} \exp(-\lambda D(x_k, y)),$$

we have

$$J'_e(V) = J_e(U(V), V) = \lambda \sum_{i=1}^{c} j_e(v_i).$$

Note that this substitution changes the original formulation of possibilistic clustering, so the optimization of $J'_e(V)$ is not identical to the alternative minimization of $J_e(U, V)$, so they are expected to have similar properties (cf. [9]). We therefore study the properties of $J'_e(V)$ as a function of V.

Note first that $J'_e(V)$ is the sum of $j_e(v_i)$. Since no constraint is imposed on v_i, every $j_e(v_i)$ can be minimized independently from other $j_e(v_j)$ $(j \neq i)$. If we assume the minimizing element is unique, minimization leads to $\bar{v} = v_1 = \cdots = v_c$ and therefore only one cluster center is obtained as the minimizing element of $J'_e(V)$. This means that if we want to have multiple clusters from this function, we must search for different minimizing solutions of a multimodal function, which is far more difficult than minimizing a unimodal function.

This leads us to the idea of extracting "one cluster at a time," discussed by Davé and Krishnapuram [5]. Note that this idea has not yet been fully developed, so we discuss their idea in greater detail in developing new algorithms.

We consider $J_2(U, V)$ in the same way. Substitute $U(V) = (U_e(x_k, v_i))_{i=1,\ldots,c}$ into $J_2(U, V)$ in which v_i is a variable. This yields

$$J'_2(V) = J_2(U(V), V) = \sum_{i=1}^{c} \sum_{k=1}^{n} \frac{D(x_k, v_i)}{1 + \zeta D(x_k, v_i)}.$$

We put

$$j_2(y) = \sum_{k=1}^{n} \frac{D(x_k, y)}{1 + \zeta D(x_k, y)},$$

and it follows that

$$J'_2(V) = \sum_{i=1}^{c} j_2(v_i).$$

Note again that $j_2(v_i)$ can be minimized independently from other $j_2(v_j)$. We thus have the unique solution $\hat{v} = v_1 = \cdots = v_c$ that minimizes $J'_2(V)$.

These observations justify the use of an algorithm to extract "one cluster at a time." generally done as follows.

SC: General Procedure for Sequential Clustering Algorithms
SC1. Let the initial set of objects be $X^{(0)} = X$ and $k = 0$. Let function $J(v; k) = j_e(v)$ (or $J(v; k) = j_2(v)$) with the set of objects $X^{(k)}$.
SC2. Search the minimizing element of $J(v; k)$:

$$v^{(k)} = \arg\min_{v} J(v; k)$$

SC3. Extract cluster $G^{(k)}$ that belongs to center $v^{(k)}$.
SC4. Let $X^{(k+1)} = X^{(k)} - G^{(k)}$. If $X^{(k+1)}$ does not have sufficient elements to extract one more cluster, stop; else $k := k + 1$ and go to step **SC2**.
End SC.