TYPE 2 FUZZY CLUSTERING ALGORITHM FOR FUZZY DATA

HIDEYUKI IMAI, SHIN-ICHI OHNISHI, AND YOSHIHARU SATO

1. Introduction

In many situations, precise measurements are vary costly, unnecessary, and even impossible to obtain in some cases. Thus, it often occurs that fuzzy data such as real intervals or real fuzzy numbers are available. In recent literatures, some method to handle fuzzy data have been developed [2, 4]. These are based on a parameterization representing centers and radiuses of fuzzy data. It is pointed out that different parameterizations of fuzzy sets can lead to different computational result even in very simple cases. In this paper, we propose other method to classify fuzzy data. This is based on a interval computations of α-cut sets of fuzzy data and a extension principle of fuzzy sets. This method enables us to obtain both type 2 fuzzy prototypes and type 2 fuzzy partitions. It requires high computing cost to determine exact intervals for a lot of α-cut sets. Therefore, we also propose a method to obtain approximate intervals which is based on a sensitivity analysis.

2. FCM Algorithm

Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N \in \mathbf{R}^k$ be data points in k-dimensional real Euclidean space $\mathbf{R}^k$. Thus, the FCM algorithm is to solve the following problem:

$$\text{minimize } J_m(U, V) = \sum_{i=1}^{N} \sum_{\beta=1}^{c} (u_{\beta i})^m \|\boldsymbol{x}_i - \boldsymbol{v}_\beta\|^2 \tag{1}$$

where $\|\cdot\|$ denotes the Euclidean distance in $\mathbf{R}^k$, c is the fixed and known number of clusters, m is the arbitrary chosen scalar greater than 1, $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_c \in \mathbf{R}^k$ are the unknown prototypes (cluster centers), $V \equiv [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_c]$ is a $k \times c$ matrix, and $U \equiv (u_{\beta i})$ is a $c \times N$ matrix whose elements satisfy the following condition (a), (b), and (c) :

$$\text{(a) } u_{\beta i} \in [0, 1], \forall \beta \in J, \forall i \in I, \qquad \text{(b) } \sum_{\beta=1}^{c} u_{\beta i} = 1, \forall i \in I,$$

$$\text{(c) } 0 < \sum_{i=1}^{N} u_{\beta i} < N, \forall \beta \in J,$$

where $I = \{1, \ldots, N\}$ and $J = \{1, \ldots, c\}$. The scalar m, called the weighting exponent of FCM, determines the fuzziness of the clustering.

When m is small, the result is the same as hard clustering, that is, $u_{\beta i}$'s tend to $\{0, 1\}, \forall i \in I, \forall \beta \in J$, and when m is large, the result is quite fuzzy, that is, $u_{\beta i}$'s tend to $\frac{1}{c}, \forall i \in I, \forall \beta \in J$. For more details about FCM, refer [1].

3. Fuzzy Data

In ordinary cluster analysis including FCM, each (crisp) data is represented by a k-tuple of real numbers, such that $\mathbf{x}_i = (x_{i1}, \ldots, x_{ik}), i \in I$. In this paper, we assume that fuzzy data points are available, and that they are real numbers, real intervals or real fuzzy numbers[2]. Thus, each data point is represented by k-tuple of them. Since they are a kind of fuzzy sets, their components are represented by membership functions. The α-cut sets ($\alpha \in [0, 1]$) of membership functions are denoted by $\mu^{(\alpha)}$, which are real intervals, that is, $\mu^{(\alpha)} = \{s \in \mathbf{R} \mid \mu(s) \geq \alpha\}$. In what follows, fuzzy data points in $\mathbf{R}^k$ is denoted by $\boldsymbol{\mu}_i = (\mu_{i1}, \ldots, \mu_{ik}), i \in I$, and the k-tuple of their α-cut sets is denoted by $\boldsymbol{\mu}_i^{(\alpha)} = (\mu_{i1}^{(\alpha)}, \ldots, \mu_{ik}^{(\alpha)}), i \in I$. Moreover, the center of each interval $\mu_{ij}^{(\alpha)}$ is denoted by $x_{ij}^{(\alpha)}$.

4. Fuzzy clustering method for fuzzy data

Since α-cut sets of fuzzy data points are intervals, the result of clustering becomes also intervals. Thus, to classify fuzzy data points, we use interval analysis. To obtain approximate membership functions of partitions and prototypes, we combine the results of α_i-cut sets with $0 = \alpha_0 < \alpha_1 \cdots < \alpha_M = 1$. The proposed method is based on sensitivity analysis for FCM [3].

Let a $\{c \times N + c \times k\}$ vector $\boldsymbol{\theta}$ and a $\{(c-1) \times N + c \times k\}$ vector $\bar{\boldsymbol{\theta}}$ be

$$\boldsymbol{\theta} = (u_{11}, \ldots, u_{c1}, \ldots, u_{1N}, \ldots, u_{cN}, v_{11}, \ldots, v_{1k}, \ldots, v_{ck})',$$
$$\bar{\boldsymbol{\theta}} = (u_{11}, \ldots, u_{c-1,1}, \ldots, u_{1N}, \ldots, u_{c-1,N}, v_{11}, \ldots, v_{1k}, \ldots, v_{ck})',$$

respectively, where these components are the result of ordinary FCM for crisp data. Note that the value $u_{ci}, i \in I$ can be calculated from the condition (b), and that the objective function J_m is regarded as a function of $\bar{\boldsymbol{\theta}}$, which is denoted by $\bar{J}_m$. By the Theorem 1 of [3], we obtain the following

Theorem 1. Suppose the data points $\mathbf{y}_1^{(\alpha)}, \ldots, \mathbf{y}_N^{(\alpha)}$ is denoted by

$$\mathbf{y}_i^{(\alpha)} = \mathbf{x}_i^{(\alpha)} + \mathbf{d}_i^{(\alpha)} \epsilon, \ i \in I,$$

where $\mathbf{x}_i^{(\alpha)} = (x_{i1}^{(\alpha)}, \ldots, x_{ik}^{(\alpha)})$, and $\mathbf{d}_i^{(\alpha)} = (d_{i1}^{(\alpha)}, \ldots, d_{ik}^{(\alpha)})$, $d_{ij}^{(\alpha)}$ is a half of the length of the interval $\mu_{ij}^{(\alpha)}$. Moreover, $\boldsymbol{\theta}_0$ denotes the solution of Eq.(1) for crisp data $\mathbf{x}_i^{(\alpha)}, i \in I$. When either the condition $m > 2$ or

$u_{\beta i} \neq 0, 1, \forall i \in I, \forall j \in J$, is satisfied, and the Hessian matrix of $\bar{J}_m(\bar{\theta})$, which is denoted by $H(\bar{\theta})$, is nonsingular at $\bar{\theta} = \bar{\theta}_0$, then

$$\frac{\partial \bar{J}_m}{\partial \bar{\theta}}(\bar{\theta}, \epsilon) = 0$$

can be analytically solved as $\bar{\theta} = \bar{\theta}(\epsilon)$ in the neighborhood of $(\bar{\theta}, 0)$ and $\bar{\theta}(0) = \bar{\theta}_0$. Moreover,

$$\bar{\theta}(\epsilon) - \bar{\theta}_0 = H^{-1}(\bar{\theta}_0) L(\bar{\theta}_0) \text{vec}(D)\epsilon + O(\epsilon^2),$$

where vec is the vec-operator.

For calculation of $H(\bar{\theta})$ and $L(\bar{\theta})$, refer [3]. From Theorem 1 and the condition (b), we get

$$\theta(\epsilon) - \theta_0 = A(\theta_0) \text{vec}(D)\epsilon + O(\epsilon^2), \tag{2}$$

where A is some $p \times (c \times N + k)$ matrix . For calculation of A, also refer [3]. Using Eq.(2), we obtain approximate intervals of α-cut sets of fuzzy data points.

References

[1] J. C. Bezdek. *Pattern Recognition with Fuzzy Objective Function Algorithm*. Plenum, 1981.

[2] R. J. Hathaway, J. C. Bezdek, and W. Pedrycz. A parametric model for fusing heterogeneous fuzzy data. *IEEE Trans. Fuzzy Systems*, 4(3):270–281, Aug. 1996.

[3] H. Imai, A. Tanaka, and M. Miyakoshi. A method of identifying influential data in fuzzy clustering. *IEEE Transactions on Fuzzy Systems*, 6(1):90–101, Feb. 1998.

[4] W. Pedrycz, J. C. Bezdek, R. J. Hathaway, and G. W. Rogers. Two nonparametric models for fusing heterogeneous fuzzy data. *IEEE Trans. Fuzzy Systems*, 6(3):411–425, Aug. 1998.

(H. Imai) Division of Systems and Information Engineering, Hokkaido University, Sapporo, 060–8628 Japan

E-mail address: `imai@main.eng.hokudai.ac.jp`

(S. Ohnishi) Department of Electronics and Information Engineering, Hokkai-Gakuen University, Sapporo 064–0926, Japan

E-mail address: `ohnishi@eli.hokkai-s-u.ac.jp`

(Y. Sato) Division of Systems and Information Engineering, Hokkaido University, Sapporo, 060–8628 Japan

E-mail address: `ysato@main.eng.hokudai.ac.jp`