2005 International Symposium on Nonlinear
Theory and its Applications (NOLTA2005)
Bruges, Belgium, October 18-21, 2005

Fuzzy c-Means for Data with Tolerance

Yasunori Endo†, Ryuichi Murata†, Hideyuki Haruyama†, Sadaaki Miyamoto†

†Department of Risk Engineering,
Faculty of System and Information Engineering,
University of Tsukuba
Email: endo@risk.tsukuba.ac.jp

Abstract—In this paper, two new clustering algorithms are proposed, which are based on the entropy regularized fuzzy c-means and can treat data with some errors. The first, the tolerance which means the permissible range of the error is introduced into optimization problems which relate with clustering, and formulated. The next, the problems are solved using Kuhn-Tucker conditions. The last, the algorithms are constructed based on the results of solving the problems.

1. Summary

Fuzzy c-means (FCM) is very famous and representative method in clustering algorithms [1]. The FCM is based on hard c-means (HCM) and has been constructed by fuzzification of HCM. Some FCMs is used in the field of clustering. Each FCM corresponds with the way to fuzzify the HCM. Particularly, the entropy regularized FCM [2] is known as effective in FCMs.

By the way, there are many cases that data has some errors in clustering. Until now, the errors have been represented by interval values [3, 4] in these case. But the way is not adequate because only the boundary of interval values are considered and calculated frequently in these algorithms for the data with the errors.

Therefore, we try to formulate these error problems into the optimization problems with inequality constraints and construct new clustering algorithms through solving the problems in this paper. The first, we will define the tolerance ε_k which means the permissible range of the error, introduce the tolerance into the optimization problems and formulate the problems. The next, we will solve the problems by using Kuhn-Tucker conditions. The last, we will construct new algorithms based on the solutions.

2. Theory

In this section, we have the mathematical discussions about two optimization problems.

The objective functions of the two problems are same and the constraints are different.

2.1. Optimization Problems

The data set $X = \{x_k \mid x_k \in R^p,\ k = 1, \dots, n\}$ is given and let C_i $(i = 1, \dots, c)$ be the cluster. $w_i \in W$ means the cluster center of C_i. We assume that μ_{ik} and $\varepsilon_k \in E$ mean the similarity degree between x_k and C_i and the tolerance of x_k, respectively. We call $U = [\mu_{ik}]$ partition matrix.

We consider the following objective function:

$$J(U, E, W) = \sum_{k=1}^{n}\sum_{i=1}^{c} \mu_{ik}\|x_k + \varepsilon_k - w_i\|^2 + \lambda^{-1}\sum_{k=1}^{n}\sum_{i=1}^{c} \mu_{ik}\log\mu_{ik} \tag{1}$$

The last term is used for entropy regularization proposed by Miyamoto *et al.* [2].

Here, we consider the following different constraints (A) and (B).

Common Constraint

$$\sum_{i=1}^{c} \mu_{ik} = 1 \tag{2}$$

Constraint (A)

$$\|\varepsilon_k\|^2 \le \kappa_k^2 \quad (\kappa_k > 0) \tag{3}$$

Constraint (B)

$$\sum_{k=1}^{n} \|\varepsilon_k\|^2 \le K^2 \quad (K > 0) \tag{4}$$

The purpose of the problem is to find the solutions μ_{ik}, ε_k and w_i $(k = 1, \dots, n,\ i = 1, \dots, c)$ which minimize the objective function (1) under the constraints (2) and (3) (called "problem (A)"), and (2) and (4) (called "problem (B)"), respectively.

2.2. Problem (A)

To find the solution, we use the Lagrange multiplier method. The first, we introduce a function L_1 as follows:

$$\begin{aligned}
&L_1(U,E,W)\\
&=J(U,E,W)+\sum_{k=1}^{n}\gamma_k(\sum_{i=1}^{c}\mu_{ik}-1)\\
&\quad+\sum_{k=1}^{n}\delta_k(\|\varepsilon_k\|^2-\kappa_k^2)\\
&=\sum_{k=1}^{n}\sum_{i=1}^{c}\mu_{ik}\|x_k+\varepsilon_k-w_i\|^2\\
&\quad+\lambda^{-1}\sum_{k=1}^{n}\sum_{i=1}^{c}\mu_{ik}\log\mu_{ik}\\
&\quad+\sum_{k=1}^{n}\gamma_k(\sum_{i=1}^{c}\mu_{ik}-1)\\
&\quad+\sum_{k=1}^{n}\delta_k(\|\varepsilon_k\|^2-\kappa_k^2)
\end{aligned}$$

From Kuhn-Tucker conditions, The necessary conditions that the solutions should satisfy are as follows:

$$\begin{cases}\dfrac{\partial L_1}{\partial w_i}=0,\quad \dfrac{\partial L_1}{\partial \mu_{ik}}=0,\quad \dfrac{\partial L_1}{\partial \varepsilon_k}=0\\ \dfrac{\partial L_1}{\partial \gamma_k}=0,\quad \dfrac{\partial L_1}{\partial \delta_k}\le 0\\ \delta_k\dfrac{\partial L_1}{\partial \delta_k}=0\\ \delta_k\le 0\end{cases} \tag{5}$$

For w_i,

$$\frac{\partial L_1}{\partial w_i}=-\sum_{k=1}^{n}2\mu_{ik}(x_k+\varepsilon_k-w_i)=0$$

Hence,

$$w_i=\frac{\sum_{k=1}^{n}\mu_{ik}(x_k+\varepsilon_k)}{\sum_{k=1}^{n}\mu_{ik}}$$

For μ_{ik},

$$\frac{\partial L_1}{\partial \mu_{ik}}=\|x_k+\varepsilon_k-w_i\|^2+\lambda^{-1}(\log\mu_{ik}+1)+\gamma_k=0$$

$$\mu_{ik}=e^{\lambda(-\gamma_k-\|x_k+\varepsilon_k-w_i\|^2)-1} \tag{6}$$

On the other hand, from the constraint (2),

$$\sum_{i=1}^{c}\mu_{ik}=\sum_{i=1}^{c}e^{\lambda(-\gamma_k-\|x_k+\varepsilon_k-w_i\|^2)-1}=1$$

$$e^{-\lambda\gamma_k}=\frac{1}{\sum_{i=1}^{c}e^{-\lambda\|x_k+\varepsilon_k-w_i\|^2-1}} \tag{7}$$

From (6) and (7), we get

$$\mu_{ik}=\frac{e^{-\lambda\|x_k+\varepsilon_k-w_i\|^2}}{\sum_{i=1}^{c}e^{-\lambda\|x_k+\varepsilon_k-w_i\|^2}}$$

We should note that the objective function J is convex for μ_{ik} and $0<\mu_{ik}<1$.

For ε_k,

$$\frac{\partial L_1}{\partial \varepsilon_k}=\sum_{i=1}^{c}2\mu_{ik}(x_k+\varepsilon_k-w_i)+2\delta_k\varepsilon_k=0$$

Therefore

$$\varepsilon_k=-\frac{\sum_{i=1}^{c}\mu_{ik}(x_k-w_i)}{\delta_k+1}$$

From (5)

$$\delta_k(\|\varepsilon_k\|^2-\kappa_k^2)=0$$

Hence we get $\|\varepsilon_k\|^2=\kappa_k^2$ or $\delta_k=0$.

The first, we consider the case of $\delta_k=0$. In this case, the problem becomes the minimization without the constraint (3) because $\frac{\partial L_1}{\partial\varepsilon_k}=\frac{\partial J}{\partial\varepsilon_k}$. From

$$\frac{\partial J}{\partial \varepsilon_k}=\sum_{i=1}^{c}2\mu_{ii}(x_k+\varepsilon_k-w_i)=0$$

We get

$$\begin{aligned}\varepsilon_k&=-\frac{\sum_{i=1}^{c}\mu_{ik}(x_k-w_i)}{\sum_{i=1}^{c}\mu_{ik}}\\&=-(x_k-\sum_{i=1}^{c}\mu_{ik}w_i)\end{aligned} \tag{8}$$

In the second place, we consider the case of $\|\varepsilon_k\|^2=\kappa_k^2$. From $\|\varepsilon_k\|^2=\kappa_k^2$,

$$\|\varepsilon_k\|^2=\|-\frac{\sum_{i=1}^{c}\mu_{ik}(x_k-w_i)}{\delta_k+1}\|^2=\kappa^2$$

Therefore

$$\delta_k+1=\pm\frac{\|\sum_{i=1}^{c}\mu_{ik}(x_k-w_i)\|}{\kappa_k}$$

Hence,

$$\begin{aligned}\varepsilon_i&=\pm\frac{\kappa_k\sum_{i=1}^{c}\mu_{ik}(x_k-w_i)}{\|\sum_{i=1}^{c}\mu_{ik}(x_k-w_i)\|}\\&=\pm\frac{\kappa_k(x_k-\sum_{i=1}^{c}\mu_{ik}w_i)}{\|x_k-\sum_{i=1}^{c}\mu_{ik}w_i\|}\end{aligned} \tag{9}$$

From the fact that (8) corresponds to (9) when $\|\varepsilon_k\|^2=\kappa_k^2$ and that κ_k and the denominator of the above equation are positive, we find that the sign of the desirable solution of ε_k is minus, that is,

$$\varepsilon_k=\frac{-\kappa_k(x_k-\sum_{i=1}^{c}\mu_{ik}w_i)}{\|x_k-\sum_{i=1}^{c}\mu_{ik}w_i\|}$$

From the above discussion, the optimization solutions of the problem are as follows:

$$w_i = \frac{\sum_{k=1}^{n} \mu_{ik}(x_k + \varepsilon_k)}{\sum_{k=1}^{n} \mu_{ik}}$$

$$\mu_{ik} = \frac{e^{-\lambda\|x_k+\varepsilon_k-w_i\|^2}}{\sum_{i=1}^{c} e^{-\lambda\|x_k+\varepsilon_k-w_i\|^2}}$$

$$\varepsilon_k = -\alpha_k(x_k - \sum_{i=1}^{c} \mu_{ik} w_i)$$

Here

$$\alpha_k = \min\left\{\frac{\kappa_k}{\|x_k - \sum_{i=1}^{n} \mu_{ik} w_i\|}, 1\right\} \tag{10}$$

2.3. Problem (B)

To find the solution, we also use the Lagrange multiplier method. Same as the problem (A), we introduce a function L_2 as follows:

$$\begin{aligned} L_2(U,E,W) &= J(U,E,W) + \sum_{k=1}^{n} \gamma_k(\sum_{i=1}^{c} \mu_{ik} - 1) \\ &\quad + \delta(\sum_{k=1}^{n} \|\varepsilon_k\|^2 - K^2) \\ &= \sum_{k=1}^{n}\sum_{i=1}^{c} \mu_{ik}\|x_k + \varepsilon_k - w_i\|^2 \\ &\quad + \lambda^{-1}\sum_{k=1}^{n}\sum_{i=1}^{c} \mu_{ik}\log\mu_{ik} \\ &\quad + \sum_{k=1}^{n} \gamma_k(\sum_{i=1}^{c} \mu_{ik} - 1) \\ &\quad + \delta(\sum_{k=1}^{n} \|\varepsilon_k\|^2 - K^2) \end{aligned}$$

The procedure to solve the problem is as same as the problem (A).

For w_i,

$$\frac{\partial L_2}{\partial w_i} = -\sum_{k=1}^{n} 2\mu_{ik}(x_k + \varepsilon_k - w_i) = 0$$

Hence,

$$w_i = \frac{\sum_{k=1}^{n} \mu_{ik}(x_k + \varepsilon_k)}{\sum_{k=1}^{n} \mu_{ik}}$$

For μ_{ik},

$$\frac{\partial L_2}{\partial \mu_{ik}} = \|x_k + \varepsilon_k - w_i\|^2 + \lambda^{-1}(\log\mu_{ik} + 1) + \gamma_k = 0$$

$$\mu_{ik} = e^{\lambda(-\gamma_k - \|x_k+\varepsilon_k-w_i\|^2)-1}$$

On the other hand, from the constraint (2),

$$\sum_{i=1}^{c} \mu_{ik} = \sum_{i=1}^{c} e^{\lambda(-\gamma_k - \|x_k+\varepsilon_k-w_i\|^2)-1} = 1$$

$$e^{-\lambda\gamma_k} = \frac{1}{\sum_{i=1}^{c} e^{-\lambda\|x_k+\varepsilon_k-w_i\|^2-1}}$$

From (2.3) and (2.3), we get

$$\mu_{ik} = \frac{e^{-\lambda\|x_k+\varepsilon_k-w_i\|^2}}{\sum_{i=1}^{c} e^{-\lambda\|x_k+\varepsilon_k-w_i\|^2}}$$

We should note that the objective function J is convex for μ_{ik} and $0 < \mu_{ik} < 1$.

For ε_k,

$$\frac{\partial f}{\partial \varepsilon_k} = \sum_{i=1}^{c} 2\mu_{ik}(x_k + \varepsilon_k - w_i) + 2\delta\varepsilon_k = 0$$

Therefore,

$$\varepsilon_k = -\frac{\sum_{i=1}^{c} \mu_{ik}(x_k - w_i)}{\delta + 1} \tag{11}$$

On the other hand, the constraint (4) can be changed to the following equation:

$$\sum_{k=1}^{c} \|\varepsilon_k\|^2 = K^2 \tag{12}$$

This reason is as follows. The first, we suppose that J is minimized under the constraint:

$$\sum_{k=1}^{n} \|\varepsilon_k\|^2 = \kappa^2 \quad (0 < \kappa < K) \tag{13}$$

Here, we choose a vector ξ which satisfies $\sum_{k=1}^{n} \|\varepsilon_k + \xi\|^2 = K^2$. For example, from

$$\begin{aligned} \sum_{k=1}^{n} \|\varepsilon_k + \xi\|^2 &= \sum_{k=1}^{n} \|\varepsilon_k\|^2 + 2\langle\sum_{k=1}^{n} \varepsilon_k, \xi\rangle + n\|\xi\|^2 \\ &= \kappa^2 + 2\langle\sum_{k=1}^{n} \varepsilon_k, \xi\rangle + n\|\xi\|^2 \end{aligned}$$

we can get the vector ξ which is orthogonal to the vector $\sum_{k=1}^{n} \varepsilon_k$ and satisfies that $\|\xi\| = \sqrt{\frac{K^2-\kappa^2}{n}}$. If the ξ is added to each x_k, all x_k are parallel translated and the $w_i + \xi$ which is given from adding the solutions w_i under the constraint (13) to ξ obviously minimize J under the constraint (12). Therefore, it is sufficient that we consider the constraint (12) instead of (4).

From (11) and (12),

$$\sum_{k=1}^{n} \|\varepsilon_k\|^2 = \sum_{k=1}^{n} \| -\frac{\sum_{i=1}^{c} \mu_{ik}(x_k - w_i)}{\delta + 1}\|^2 = K^2$$

Therefore,

$$\delta + 1 = \pm\frac{\sqrt{\sum_{k=1}^{n} \|\sum_{i=1}^{c} \mu_{ik}(x_k - w_i)\|^2}}{K}$$

Hence,

$$\varepsilon_k = \pm \frac{K\sum_{i=1}^{c}\mu_{ik}(x_k - w_i)}{\sqrt{\sum_{k=1}^{n}\|\sum_{i=1}^{c}\mu_{ik}(x_k - w_i)\|^2}} = \pm \frac{K(x_k - \sum_{i=1}^{c}\mu_{ik}w_i)}{\sqrt{\sum_{k=1}^{n}\|x_k - \sum_{i=1}^{c}\mu_{ik}w_i\|^2}}$$

K, and the denominator of the above equation are positive so that we find that the sign of the desirable solution of ε_k is minus from the correlation with the problem (A), that is,

$$\varepsilon_k = -\frac{K(x_k - \sum_{i=1}^{c}\mu_{ik}w_i)}{\sqrt{\sum_{k=1}^{n}\|x_k - \sum_{i=1}^{c}\mu_{ik}w_i\|^2}}$$

From the above discussion, the optimization solutions of the problem are as follows:

$$w_i = \frac{\sum_{k=1}^{n}\mu_{ik}(x_k + \varepsilon_k)}{\sum_{k=1}^{n}\mu_{ik}}$$

$$\mu_{ik} = \frac{e^{-\lambda\|x_k+\varepsilon_k-w_i\|^2}}{\sum_{i=1}^{c}e^{-\lambda\|x_k+\varepsilon_k-w_i\|^2}}$$

$$\varepsilon_k = \frac{-K(x_k - \sum_{i=1}^{c}\mu_{ik}w_i)}{\sqrt{\sum_{k=1}^{n}\|x_k - \sum_{i=1}^{c}\mu_{ik}w_i\|^2}}$$

3. Algorithms

In this section, we propose two new algorithms, FCM-T(A) and FCM-T(B), which correspond with the above discussions of the problems (A) and (B), respectively. Both two algorithms are constructed using by alternate optimization process.

Algorithm 1 (FCM-T(A))

Step 1 *Give the value λ and κ_i. Set the initial values of $\varepsilon_k \in E$ $(k = 1, \ldots, n)$ and $w_i \in W$ $(i = 1, \ldots, c)$.*

Step 2 *Calculate $\mu_{ik} \in U$ such that*

$$\mu_{ik} = \frac{e^{-\lambda\|x_k+\varepsilon_k-w_i\|^2}}{\sum_{i=1}^{c}e^{-\lambda\|x_k+\varepsilon_k-w_i\|^2}}$$

Step 3 *Calculate ε_k such that*

$$\varepsilon_k = -\alpha_k(x_k - \sum_{i=1}^{c}\mu_{ik}w_i)$$

Here

$$\alpha_k = \min\left\{\frac{\kappa_k}{\|x_k - \sum_{i=1}^{n}\mu_{ik}w_i\|}, 1\right\}$$

Step 4 *Calculate w_i such that*

$$w_i = \frac{\sum_{k=1}^{n}\mu_{ik}(x_k + \varepsilon_k)}{\sum_{k=1}^{n}\mu_{ik}}$$

Step 5 *Check the stopping criterion for (U, E, W). If the criterion is not satisfied, go back to* Step 2.

Algorithm 2 (FCM-T(B))

Step 1 *Give the value λ and K. Set the initial values of $\varepsilon_k \in E$ $(k = 1, \ldots, n)$ and $w_i \in W$ $(i = 1, \ldots, c)$.*

Step 2 *Calculate $\mu_{ik} \in U$ such that*

$$\mu_{ik} = \frac{e^{-\lambda\|x_k+\varepsilon_k-w_i\|^2}}{\sum_{i=1}^{c}e^{-\lambda\|x_k+\varepsilon_k-w_i\|^2}}$$

Step 3 *Calculate ε_k such that*

$$\varepsilon_k = \frac{-K\sum_{i=1}^{c}\mu_{ik}(x_k - w_i)}{\sqrt{\sum_{k=1}^{n}\|\sum_{i=1}^{c}\mu_{ik}(x_k - w_i)\|^2}}$$

Step 4 *Calculate w_i such that*

$$w_i = \frac{\sum_{k=1}^{n}\mu_{ik}(x_k + \varepsilon_k)}{\sum_{k=1}^{n}\mu_{ik}}$$

Step 5 *Check the stopping criterion for (U, E, W). If the criterion is not satisfied, go back to* Step 2.

4. Conclusion

In this paper, we proposed two new clustering algorithms. These algorithms are base on the entropy regularized FCM and can treat the data with some errors which is represented by ε_k. In the forthcoming paper, we will verify the effectiveness of these algorithms through numerical examples.

Acknowledgment

We thank Associate Professor Masahide Kashiwagi of Waseda University for his helpful advice. This research was partially supported by the Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Scientific Research (B), 16300065, 2005.

References

[1] Bezdek, J.P., Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum, New York, 1981.

[2] Miyamoto, S. and Umayahara, K., Methods in hard and fuzzy clustering, in: Liu, Z.-Q. and Miyamoto, S. (eds), Soft computing and human-centered machines, Springer-Verlag Tokyo, 2000.

[3] Endo, Y. and Horiuchi, K., On Clustering Algorithm for Fuzzy Data, Proc. 1997 International Symposium on Nonlinear Theory and Its Applications, pp.381–384, 1997.

[4] Endo, Y., Clustering Algorithm Using Covariance for Fuzzy Data, Proc. 1998 International Symposium on Nonlinear Theory and Its Applications, pp.511–514, 1998.