2009 International Symposium on Nonlinear Theory and its Applications
NOLTA'09, Sapporo, Japan, October 18-21, 2009

p-Norm Based Fuzzy c-Means

Takatomi Miyata† and Yuchi Kanzawa‡

†Department of Information and Computer Science, Kanazawa Institute of Technology
7-1 Ohgigaoka, Nonoichi, Ishikawa, 920-8501 Japan

‡Faculty of Engineering, Shibaura Institute of Technology
3-7-5, Toyosu, Koto, Tokyo, 135-8548 Japan

Email: takatomi-miyata@neptune.kanazawa-it.ac.jp, kanzawa@sic.shibaura-it.ac.jp

Abstract—In this paper, we propose two variants of Fuzzy c-Means (FCM) clustering algorithms. Both models are extension of conventional FCM in which not only squared Euclidian norm, L_1 norm but also p-norm to the qth power are used as the dissimilarity. One is constructed by using Newton method, and the other, only for the case of $p = 2$, is constructed by using alternative optimization.

1. Introduction

Clustering is a technique of data classification without external criterion. There are two methods in clustering. One is hierarchical and the other is non-hierarchical. Fuzzy c-means (FCM) is one of the most typical method of non-hierarchical clustering. FCM classifies the data into c clusters by optimizing an objective function, where c, the number of clusters, is given in advance. Standard fuzzy c-means (sFCM)[1][2] and entropy regularized fuzzy c-means (eFCM)[3] are well-known methods among FCM algorithms.

As the dissimilarity in FCM, squared Euclidian norm is most commonly used, thanks to its easily implementable alternative optimization solution. On the other hand, L_1 norm based FCM also has been studied actively. About the reason, Jajuga[4] pointed out that the L_1 norm based FCM is little affected on outliers, and Bobrowski et al.[5] concluded that L_1 norm based FCM matches “boxy” data well, and better accuracy than FCM can using squared Eucilidian norm.

Bobrowski et al. treated not only L_1 norm but also L_∞ norm as the dissimilarity. Miyamoto et al.[6] has studied not only L_1 norm based sFCM but also L_1 norm based eFCM and L_1 norm based mixture models.

But the most of previous works use squared Euclidian norm, L_1 norm and L_∞ norm as the dissimilarity of FCM. It has been hardly examined the algorithm, the numerical result, and the effectiveness of FCM using p-norms to the q-th power as the dissimilarity. It is impossible to obtain the cluster center V analytically in that case, that is one of the reason why the p-norm based FCM (nFCM) has not been studied.

In this paper, we propose algorithms for nFCM which is the generalization of the conventional FCM. In our new algorithm, p-norms to the q-th power is used as the dissimilarity, and V is obtained not analytically but numerically.

In the section 2, we prescribe our notation and explain the conventional FCM. In the section 3, we propose an algorithm for nFCM using Newton method. In the section 4, we propose another algorithm. It is available if we use the Euclidian norm to the q-th power as the dissimilarity of FCM. In the section 5, we show some numerical examples. In the section 6, we summarize our conclusions.

2. Preliminaries

In this section, we define some notations which are data for clustering, the membership by which the each data belongs to the each cluster, and the cluster prototypes of FCM. Next, we show two types of objective function. At the end of this section, we show the algorithm of conventional FCM.

2.1. Data, Membership and Cluster Prototypes

Let the data set $X = \{x_k \in \boldsymbol{R}^s \mid k = 1, \ldots, n\}$ be given in advance. We classify X into c clusters $G_1, \ldots, G_c$ without the external criterion ($1 < c < n$). x_k^j denotes the j-th component of x_k, where $k = 1, \ldots, n$ and $j = 1, \ldots, s$. Namely,

$$x_k = (x_k^1, \cdots, x_k^s)^T . \tag{1}$$

The membership $U \in \boldsymbol{R}^{n\times c}$ by which x_k belongs to the i-th cluster G_i is denoted by $u_{ki} \in [0, 1]$. The prototype of the cluster G_i is denoted by $v_i \in \boldsymbol{R}^s$, and the cluster prototype set is denoted by $V \in \boldsymbol{R}^{cs}$. Namely,

$$V = (v_1^1, v_1^2, \cdots, v_c^s)^T . \tag{2}$$

2.2. Dissimilarity

We define the dissimilarity as follows:

Definition 1 (Dissimilarity) *Let X be an arbitrary set. A function $d : X \times X \to \boldsymbol{R}$ is dissimilarity if d satisfies the following for any elements x_1, x_2 of X:*
(a) $d(x_1, x_2) \geq d(x_1, x_1) = 0$,
(b) $d(x_1, x_2) = d(x_2, x_1)$. ∎

Any norms in $\boldsymbol{R}^s$ always satisfy the property (a), (b) of the definition 1, hence we can use each norm as the dissimilarity of FCM.

2.3. Objective Function

FCM is an optimized clustering to obtain the membership U and the cluster center V, which minimize the objective function $J(U, V)$ under the constraint M_f

$$M_f = \left\{(u_{ki}) : u_{ki} \in [0, 1], \sum_{i=1}^{c} u_{ki} = 1 \text{ for all } k\right\}. \tag{3}$$

The objective function of the standard regularized fuzzy c-means (sFCM) is represented as follows:

$$J_{\mathrm{sFCM}}(U, V) = \sum_{k=1}^{n} \sum_{i=1}^{c} (u_{ki})^m d_{ki} \tag{4}$$

where $m \in [1, \infty)$. The objective function of the entropy regularized fuzzy c-means (eFCM) is represented as follows:

$$J_{\mathrm{eFCM}}(U, V) = \sum_{k=1}^{n}\sum_{i=1}^{c} u_{ki} d_{ki} + \lambda^{-1} \sum_{k=1}^{n}\sum_{i=1}^{c} u_{ki} \log u_{ki} \tag{5}$$

where $\lambda > 0$.

d_{ki} denotes the dissimilarity between x_k and v_i. We usually use the squared Euclidian norm as the dissimilarity. Namely,

$$d_{ki} = \|x_k - v_i\|_2^2 = \sum_{j=1}^{s} (x_k^j - v_i^j)^2. \tag{6}$$

In the L_1-norm based FCM([4]-[6]), the following L_1 norm is used as the dissimilarity:

$$d_{ki} = \|x_k - v_i\|_1 = \sum_{j=1}^{s} |x_k^j - v_i^j| . \tag{7}$$

2.4. FCM alternative optimization algorithm

In this section, we show the conventional alternative optimization algorithm of FCM:

Algorithm 1 (FCM)

FCM1. *Set the value of the cluster prototype* $V^{\langle 0 \rangle}$ *and Let* $K = 0$.

FCM2. *Calculate* $U^{\langle K+1 \rangle}$ *by*

$$U^{\langle K+1 \rangle} = \arg\left(\min_{U \in M_f} J(U, V^{\langle K \rangle})\right).$$

FCM3. *Calculate* $V^{\langle K+1 \rangle}$ *by*

$$V^{\langle K+1 \rangle} = \arg\left(\min_{V} J(U^{\langle K+1 \rangle}, V)\right).$$

FCM4. *Check the stopping criterion. If the criterion is not satisfied, let* K *be* $K + 1$ *and go back to FCM2.*

End of FCM. ∎

We use J_{sFCM} of (4) or J_{eFCM} of (5) as the objective function J in the above algorithm.

2.4.1. Update of the membership U

J_{sFCM} is a convex function according to U if $m > 1$. And J_{eFCM} is a convex function according to U if $\lambda > 0$.

Both J_{sFCM} and J_{eFCM} are convex functions with regard to V if we use the squared Euclidian norm as the dissimilarity. Hence $U^{\langle K+1 \rangle}$ in FCM2. can be expressed as the formula of $V^{\langle K \rangle}$ explicitly via the Lagrange multiplier method.

Actually, in the case of $x_k \neq v_i^{\langle K \rangle} (i = 1, \cdots, c)$ in J_{sFCM},

$$u_{ki}^{\langle K+1 \rangle} = \left[\frac{\left(\frac{1}{d_{ki}^{\langle K \rangle}}\right)^{\frac{1}{m-1}}}{\sum_{j=1}^{c}\left(\frac{1}{d_{kj}^{\langle K \rangle}}\right)^{\frac{1}{m-1}}} \right]^{-1}, \tag{8}$$

in the case of $x_k = v_i^{\langle K \rangle}$ for some i in J_{sFCM},

$$u_{ki}^{\langle K+1 \rangle} = 1, \qquad u_{kj}^{\langle K+1 \rangle} = 0 \ (j \neq i) . \tag{9}$$

With regard to J_{eFCM}

$$u_{ki}^{\langle K+1 \rangle} = \frac{e^{-\lambda d_{ki}^{\langle K \rangle}}}{\sum_{j=1}^{c} e^{-\lambda d_{kj}^{\langle K \rangle}}} . \tag{10}$$

2.4.2. Update of the cluster center V

The objective function $J(U, V)$ is a convex function according to V if we use the squared Euclid norm as the dissimilarity. Hence in order to obtain $V^{\langle K+1 \rangle}$ in FCM3., solve the following equation with regard to V:

$$\nabla_V J(U, V) = 0. \tag{11}$$

In sFCM, we obtain $V^{\langle K+1 \rangle}$ in FCM3. as follows:

$$v_i^{\langle K+1 \rangle} = \frac{\sum_{k=1}^{n} (u_{ki}^{\langle K \rangle})^m x_k}{\sum_{k=1}^{n} (u_{ki}^{\langle K \rangle})^m} . \tag{12}$$

And in sFCM, we obtain $V^{\langle K+1 \rangle}$ in FCM3. as follows:

$$v_i^{\langle K+1 \rangle} = \frac{\sum_{k=1}^{n} u_{ki}^{\langle K \rangle} x_k}{\sum_{k=1}^{n} u_{ki}^{\langle K \rangle}} . \tag{13}$$

However, it is difficult to solve the equation (11) with regard to V for any p, q. Equation (12), (13) can be solved just because the dissimilarity is denoted by equation (6).

2.5. Fuzzy Classification Function

Fuzzy classification functions[7] are available in FCMs which show how prototypical an arbitrary point in the data space is to a cluster by extending the membership U to the whole space. Fuzzy classification function for sFCM with respect to a brand-new datum $\tilde{x} \in \boldsymbol{R}^s$ is defined as

$$U_i^s(\tilde{x}) = \left[\frac{\left(\frac{1}{\tilde{d}_i}\right)^{\frac{1}{m-1}}}{\sum_{j=1}^{c}\left(\frac{1}{\tilde{d}_k}\right)^{\frac{1}{m-1}}} \right]^{-1}, \tag{14}$$

and fuzzy classification function for eFCM with respect to a brand-new datum $\tilde{x} \in \boldsymbol{R}^s$ is defined as

$$U_i^e(\tilde{x}) = \frac{e^{-\lambda \tilde{d}_i}}{\sum_{j=1}^{c} e^{-\lambda \tilde{d}_j}}. \tag{15}$$

In these equations,

$$\tilde{d}_i = \|\tilde{x} - v_i\|_2^2 \tag{16}$$

in the case of squared Euclidian norm based sFCM, and

$$\tilde{d}_i = \|\tilde{x} - v_i\|_1 \tag{17}$$

in the case of L_1norm based sFCM. Fuzzy classification function is valid to investigate the features of FCM since it clarify the classifying situation in whole space than only memberships for finite number of data.

3. Proposed Method I

In this section, we propose a p-norm based FCM (nFCM) algorithm using Newton iteration. First we explain how to update U, V if we use the p-norm to the q-th power as the dissimilarity d_{ki}. Namely,

$$d_{ki} = \|x_k - v_i\|_p^q = \left(\sum_{j=1}^{s} |x_k^j - v_i^j|^p\right)^{q/p} \tag{18}$$

Next, we show the algorithm in which V is obtained not analytically but numerically.

3.1. Update of the membership U

In nFCM, we can obtain optimized U as the formula of $V^{\langle K\rangle}$ explicitly as well as the discussion in the section 2.4.1.

In the case of $x_k \neq v_i^{\langle K\rangle} (i = 1, \cdots, c)$ in J_{sFCM},

$$u_{ki}{}^{\langle K+1\rangle} = \left[\frac{\left(\frac{1}{\|x_k - v_i^{\langle K\rangle}\|_p^q}\right)^{\frac{1}{m-1}}}{\sum_{j=1}^{c}\left(\frac{1}{\|x_k - v_j^{\langle K\rangle}\|_p^q}\right)^{\frac{1}{m-1}}} \right]^{-1}, \tag{19}$$

in the case of $x_k = v_i{}^{\langle K\rangle}$ for some i in J_{sFCM}, $U^{\langle K+1\rangle}$ can be expressed like (9).

Furthermore, with regard to J_{eFCM}, $U^{\langle K+1\rangle}$ can be expressed from equation (10) and (18) as follows:

$$u_{ki}{}^{\langle K+1\rangle} = \frac{\exp\left(-\lambda \|x_k - v_i^{\langle K\rangle}\|_p^q\right)}{\sum_{j=1}^{c} \exp\left(-\lambda \|x_k - v_j^{\langle K\rangle}\|_p^q\right)} . \tag{20}$$

3.2. Update of the cluster center V

In this section, we explain how to obtain the optimized V in nFCM for U given in advance. It is enough to solve the equation (11) with regard to V in order to obtain the optimized V, because $J(U, V)$ becomes convex function according to V if $p \geq 1, q \geq 1$ holds.

In sFCM, we obtain the optimized V as follows:

$$\begin{aligned} &\frac{\partial}{\partial v_i^j} J_{\text{snFCM}}(U, V) \\ &= \sum_{k=1}^{n} \left[\text{sgn}(v_i^j - x_k^j) q (u_{ki})^m \|x_k - v_i\|_p^{q-p} |x_k^j - v_i^j|^{p-1}\right] \\ &= 0 . \end{aligned} \tag{21}$$

Equation (21) is equivalent to equation (12) in the case of $p = q = 2$. In eFCM, we obtain the optimized V as follows:

$$\begin{aligned} &\frac{\partial}{\partial v_i^j} J_{\text{enFCM}}(U, V) \\ &= \sum_{k=1}^{n} \left[\text{sgn}(v_i^j - x_k^j)\; q\; u_{ki}\; \|x_k - v_i\|_p^{q-p}\; |x_k^j - v_i^j|^{p-1}\right] \\ &= 0 . \end{aligned} \tag{22}$$

Equation (22) is equivalent to equation (13) in the case of $p = q = 2$.

Neither (21) nor (22) can be solved with regard to $V^{\langle K+1\rangle}$ analytically for any p, q.

3.3. Algorithm

In this section, we propose our new algorithm, nFCM using Newton iteration. Because we can't obtain V in equation (11) analytically for any p, q, we calculate V by solving $f(V) = 0$ numerically, where $f : \boldsymbol{R}^{cs} \to \boldsymbol{R}^{cs}$ is expressed as follows:

$$f(V) = \nabla_V J(U, V)\Big|_{U = \arg\left(\min\limits_{V} J(U,V)\right)} \tag{23}$$

Now, we show the algorithm:

Algorithm 2 (nFCM by using Newton iteration)
nFCM1. *Set the value of the cluster prototype $V^{\langle 0\rangle}$ and Let* $K = 0$.
nFCM2. *Solve the following linear equations*

$$f'(V^{\langle K\rangle})\left(V^{\langle K\rangle} - V^{\langle K+1\rangle}\right) = f(V^{\langle K\rangle})$$

to obtain $V^{\langle K+1\rangle}$. This is equivalent to Newton *iteration.*
nFCM3. *Check the stopping criterion. If the criterion is not satisfied, let K be $K + 1$ and go back to nFCM2.* ∎
End of nFCM.

Stopping criterion is defined by determining an acceptable difference between iterations.

4. Proposed Method II

In this section, we propose another nFCM algorithm which is using not Newton iteration but alternative optimization. It is available only in the case of $p = 2$. That is, the Euclid norm to the q-th power is used as the dissimilarity in this algorithm.

Since $p = 2$,

$$\text{sgn}(v_i^j - x_k^j)|x_k^j - v_i^j|^{p-1} = v_i^j - x_i^j. \tag{24}$$

Hence we can construct the algorithm using alternative optimization by introducing the following $g : \boldsymbol{R}^{n\times c} \times \boldsymbol{R}^{cs} \to \boldsymbol{R}^{cs}$.

In sFCM, if we define $g(U, V)$ by

$$g_{(i-1)s+j}(U, V) = \frac{\sum_{k=1}^{n} (u_{ki})^m \|x_k - v_i\|_2^{q-2} x_k^j}{\sum_{k=1}^{n} (u_{ki})^m \|x_k - v_i\|_2^{q-2}}, \tag{25}$$

the equation (21) becomes equivalent to the following:

$$V = g(U, V). \tag{26}$$

In eFCM, if we define $g(U, V)$ by

$$g_{(i-1)s+j}(U, V) = \frac{\sum_{k=1}^{n} u_{ki} \|x_k - v_i\|_2^{q-2} x_k^j}{\sum_{k=1}^{n} u_{ki} \|x_k - v_i\|_2^{q-2}} \tag{27}$$

the equation (22) becomes equivalent to the equation (26).

Now we show the nFCM algorithm using alternative optimization.

Algorithm 3 (nFCM by alternative optimization)
nFCM'1. *Set the value of the cluster prototype* $V^{\langle 0 \rangle}$ *and Let* $K = 0$.
nFCM'2. *Calculate* $U^{\langle K+1 \rangle}$ *by the equation (19), (9), and the value of* $V^{\langle K \rangle}$. **nFCM'3.** *Let* $V^{\langle K+1 \rangle} = g(U^{\langle K+1 \rangle}, V^{\langle K \rangle})$.
nFCM'4. *Check the stopping criterion. If the criterion is not satisfied, let K be K + 1 and go back to nFCM'2.*
End of nFCM'
∎

5. Numerical Examples

In this section, we show some examples of fuzzy classification functions of nFCM. In each example, 100 trials for each nFCM algorithm (Algorithm2 and Algorithm3) with different initial cluster centers are tested and the solution with the minimal objective function value is selected as the final classification result. Using the obtained cluster centers v_i, fuzzy classification function values $U_i^s(x)$ and $U_i^e(x)$ $(i = 1, \ldots, c)$ are calculated for points $\tilde{x}$ around the initially given data x_k $(k = 1, \ldots, n)$ by Algorithm2 and Algorithm3.

The data set[5] shown in Fig.1 is constructed by 11 elements in the two dimensional Euclidian space. We classify the data set into two clusters.

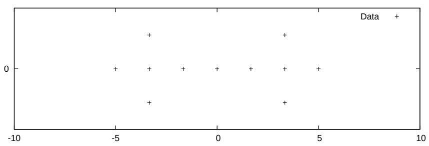

Figure 1: Data

We show the result of snFCM. Fig.2 represents the contour map of the fuzzy classification function $U_1^s(x)$ for various value of p, q. Fig.2 consists of nine graphs. In each one, m equals to 2.0. In the three graphs of upper row, $p = 10.0$. Similarly, in the three graphs of middle row, lower row, p equals to 2.0, 1.5, respectively. In the three graphs of left column, center column, right column, q equals to 1.5, 2.0, 10.0, respectively.

In every case, the contour line of $U_1^s(x) = 0.5$ is line of $x_1 = 0$, because data set consists of 11 points is symmetrically arranged with regard to the line of $x_1 = 0$. For each value of p, the larger the value of q is, the wider the area of $U_1^s(x) < 0.1$ or $0.9 < U_1^s(x)$ becomes.

6. Conclusion

In this paper, we proposed two types of new algorithms of FCM, in which dissimilarity is neither squared Euclidian norm nor L_1 norm.

The first algorithm uses p-norm to the q-th power as the dissimilarity of FCM. It updates the cluster center V by Newton iteration in order to obtain optimized (U, V).

And the second algorithm uses the Euclidian norm to the q-th power as the dissimilarity of FCM. It is the algorithm in order to obtain optimized (U, V) by using alternative optimization. It becomes equivalent to the conventional FCM in the case of $q = 2$, hence it can be regarded as the generalization of the conventional FCM.

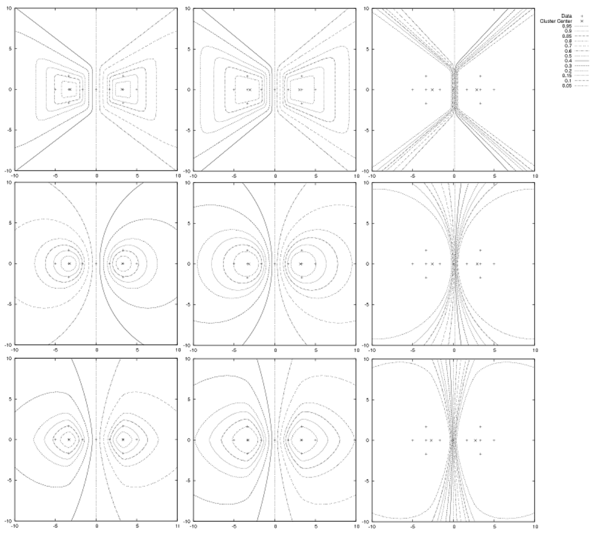

Figure 2: contour lines of $U_1^s(x)$

References

[1] J. C. Dunn, "Well-separated Clusters and Optimal Fuzzy Partitions", J. of Cybern., Vol.4, pp. 95-104, 1974.

[2] J. C. Bedzek, "Pattern Recognition with Fuzzy Objective Function Algorithms", Plenum, New York, 1981.

[3] S. Miyamoto and M. Mukaidono, "Fuzzy c-means as a regularization and maximum entropy approach", Proc. of IFSA'97, June 25-30, 1997, Prague, Chech, Vol.II, pp.86-92, 1997.

[4] K. Jajuga, "L_1-norm based fuzzy clustering", Fuzzy Sets and Systems, Vol. 39, pp. 43-50, 1991.

[5] L. Bobrowski and J. C. Bezdek, "c-means clustering with the l_1 and l_∞ norms", IEEE Trans. on Syst., Man. and Cybern., Vol. 21, No. 3, pp. 545-554, 1991.

[6] S. Miyamoto and Y. Agusta, "An efficient algorithm for l_1 fuzzy c-means and its termination", Control and Cybern., Vol. 24, No.4, pp. 421-436, 1995.

[7] S. Miyamoto, and K. Umayahara, "Methods in Hard and Fuzzy Clustering", in: Z.-Q. Liu, and S. Miyamoto, (eds), Soft computing and humancentered machines, Springer-Verlag Tokyo, (2000).