2010 International Symposium on Nonlinear Theory and its Applications
NOLTA2010, Krakow, Poland, September 5-8, 2010

Kernelized Fuzzy *c*-Means Clustering for Uncertain Data with L_1-Regularization Term of Penalty Vectors Using Explicit Mapping

ENDO Yasunori†, TAKAYAMA Isao‡, HAMASUNA Yukihiro§† and MIYAMOTO Sadaaki†

†Department of Risk Engineering, University of Tsukuba,
Tennodai 1-1-1, Tsukuba, Ibaraki, 305-8573, Japan
Email: {endo@,yhama@soft,miyamoto@}.tsukuba.ac.jp
‡Graduate School of Systems and Information Engineering, University of Tsukuba,
Tennodai 1-1-1, Tsukuba, Ibaraki, 305-8573, Japan
Email: takayama@soft.risk.tsukuba.ac.jp
§Research Fellow of the Japan Society for the Promotion of Science
Email: yhama@soft.risk.tsukuba.ac.jp

Abstract—Recently, fuzzy *c*-means clustering with kernel functions is remarkable in the reason that these algorithms can handle datasets which consist of some clusters with nonlinear boundary. However the algorithms have the following problems: (1) the cluster centers can not be calculated explicitly, (2) it takes long time to calculate clustering results. By the way, we have proposed the clustering algorithms with regularization terms of penalty vectors to handle uncertain data. In this paper, we propose new clustering algorithms with L_1-regularization term by introducing explicit mapping of kernel functions to solve the following problems.

1. Introduction

In recent years, according to large-scale and complicated data of computer systems by progress of the hardware technology, the importance of data analysis techniques has been increasing. Clustering, one of the data classification technique, is the method without any external criterion and it classifies data automatically. One of most typical and useful clustering is fuzzy *c*-means clustering (FCM) [1] and we consider FCM in this paper.

Now, studies of clustering methods for detecting nonlinear boundary have been widely discussed and kernel trick [2, 3] is focused as a very useful tool to handle datasets which consist of some clusters with nonlinear boundary. The kernel trick uses some function called "kernel function".

Here, we consider a mapping $\phi : \Re^p \to \Re^s$ $(p \ll s)$. The kernel function is an operator $K : \Re^p \times \Re^p \to \Re$ which satisfies $K(x, y) = \langle \phi(x), \phi(y) \rangle$. We can calculate the value of the inner product in the high-dimensional feature space with with low-cost calculation by the kernel functions.

However, special problems frequently arise when we introduce kernel functions into FCM. First, the cluster centers can not be calculated directly. The reason is that the mapping ϕ can be not represented explicitly in general. The second problem is caused by the first one, that is, it takes long time to calculate clustering results. The procedure to calculate the belongingness or membership grades of each data to clusters becomes complicated. To solve the above problems, Miyamoto et al. introduced an explicit mapping into kernel trick and succeeded to calculate the cluster centers directly [4].

By the way, the data we handle have uncertainties in many cases. Until now, we have two methodologies to analyze such data, deterministic and parametric ones.

In deterministic methodology, the uncertainty introduced data is represented as an interval. Distance or dissimilarities are the most essential measures for data analysis and therefore, the measures between intervals must be defined when we handle uncertainty of data as interval. There are many kinds of such measures, e.g., Hausdorff distance. However, we have two problems when we try to classify the dataset of such data. First, we don't have clear indication to select which measures to use. Second, we can't discuss clustering in the framework of optimization with such measures.

On the other hand, the uncertainty is represented by some probabilistic density function (PDF) in parametric methodology. In this methodology, the clustering is regarded as a method to determine some parameters of the PDF. However, the methodology have also at least one problem, that is, selection which type of PDFs.

Therefore, we have proposed and developed the concept of tolerance in Ref. [5, 6, 7] to solve the above problems. In our concept, tolerance vectors [5] and penalty ones [6, 7] play main role. Uncertain data is allowed to allocate any position by those vectors as far as the constraints for those vectors are satisfied and the position is derived as an optimal solution of a given objective function. Hence, We can say that this concept is in the framework of methodology of soft computing. Penalty vectors are similar to tolerance ones and the methods using penalty vectors become more flexible than tolerance vectors because no constraint for the vectors is needed. Moreover, the concept has been developed by using kernel trick in Ref. [8]. The method can classify datasets which consist of clusters of uncertain data with nonlinear boundary.

In this paper, we try to construct new FCMs for uncertain data with L_1-regularization term of penalty vectors using explicit kernel mapping (EK-FCMP$_{L_1}$). The proposed methods have the following advantages:

1. The method can classify datasets which consist of clusters of uncertain data with nonlinear boundary because of kernel trick.

2. It takes shorter time to calculate clustering result than the conventional method because of explicit mapping.

2. Kernelized FCM for Uncertain Data with L_1-Regularization Term of Penalty Vectors Using Explicit Mapping

In this section, we try to construct FCM for uncertain data with L_1-regularization term of penalty vectors using explicit mapping of kernel functions (EK-FCMP$_{L_1}$). First, we describe kernelized FCM for uncertain data with L_1-regularization term of penalty vectors (K-FCMP$_{L_1}$). Next, we introduce an explicit mapping of kernel functions into K-FCMP$_{L_1}$ to construct EK-FCMP$_{L_1}$. Last, we show an algorithm of EK-FCMP$_{L_1}$.

2.1. Kernel Functions

First, we define some symbols to introduce kernel functions. A mapping from the pattern space to the feature space is expressed as $\phi : \Re^p \to \Re^s$ $(p \ll s)$. Each data in the feature space is denoted $\phi(x_k) = x_k^\phi = (x_{k1}^\phi, \ldots, x_{ks}^\phi)^T \in \Re^s$ and the dataset $X^\phi = \left\{x_1^\phi, \ldots, x_n^\phi\right\}$ is given. Each cluster $C_i (i = 1, \ldots, c)$ has a cluster center $v_i^\phi = (v_{i1}^\phi, \ldots, v_{is}^\phi)^T \in \Re^s$. V^ϕ means a set of cluster centers $\left\{v_1^\phi, \ldots, v_c^\phi\right\}$. A membership grade for x_k^ϕ to C_i, which means belongingness of x_k^ϕ to C_i, is denoted by u_{ki}. U means a partition matrix $(u_{ki})_{1\le k\le n, 1\le i\le c}$. Moreover, $\delta_k^\phi = (\delta_{k1}^\phi, \ldots, \delta_{ks}^\phi)^T \in \Re^s$ and $w_k \in [0, +\infty)$ mean a penalty vector and a weight of the penalty vector δ_k^ϕ, respectively. Δ and W mean a set of penalty vectors $\{\delta_1^\phi, \ldots, \delta_n^\phi\}$ and a set of weights $\{w_1, \ldots, w_n\}$, respectively.

Now, a kernel function $K : \Re^p \times \Re^p \to \Re$ satisfies the following relation:

$$K(x, y) = \langle \phi(x), \phi(y) \rangle \tag{1}$$

From Mercer's theorem [10], K is a continuous symmetric non-negative definite kernel iff a mapping ϕ which satisfies the above relation exists. We should notice that ϕ is not explicit.

2.2. Objective Functions

Kernelized fuzzy c-means clustering (K-FCM) is FCM in which a kernel function is introduced into dissimilarity. An extended K-FCM by introducing penalty vectors of the quadratic term have been proposed in Ref. [8]. Here, we consider the objective function $J_{ksp_{L_1}}$ of the standard EK-FCMP$_{L_1}$ (sEK-FCMP$_{L_1}$) and another function $J_{kep_{L_1}}$ of the entropy based EK-FCMP$_{L_1}$ (eEK-FCMP$_{L_1}$) as follows:

$$J_{ksp_{L_1}}(U, \Delta^\phi, V^\phi) = \sum_{k=1}^{n}\sum_{i=1}^{c}(u_{ki})^m d_{ki} + 2\sum_{k=1}^{n} w_k \|\delta_k^\phi\|_1, \tag{2}$$

$$J_{kep_{L_1}}(U, \Delta^\phi, V^\phi) = \sum_{k=1}^{n}\sum_{i=1}^{c} u_{ki} d_{ki} + \lambda^{-1}\sum_{k=1}^{n}\sum_{i=1}^{c} u_{ki}\log u_{ki} + 2\sum_{k=1}^{n} w_k \|\delta_k^\phi\|_1. \tag{3}$$

The dissimilarity d_{ki} is the square of the norm as follows.

$$d_{ki} = d(x_k + \delta_k, v_i) = \|x_k + \delta_k - v_i\|^2 = \sum_{j=1}^{p}(x_{kj} + \delta_{kj} - v_{ij})^2. \tag{4}$$

$\| \bullet \|_1$ means the L_1-norm.

2.3. Optimal Solutions of sEK-FCMP$_{L_1}$

We can derive the optimal solutions of the membership grades u_{ki} and the cluster centers v_i^ϕ by using Lagrange multiplier. The optimal solutions of standard K-FCMP$_{L_1}$ (sK-FCMP$_{L_1}$) are as follows:

$$\begin{cases} u_{ki} = \dfrac{(1/d_{ki})^{\frac{1}{m-1}}}{\sum_{l=1}^{c}(1/d_{kl})^{\frac{1}{m-1}}}, & (5) \\[2ex] v_i^\phi = \dfrac{\sum_{k=1}^{n}(u_{ki})^m (x_k^\phi + \delta_k^\phi)}{\sum_{k=1}^{n}(u_{ki})^m}. & (6) \end{cases}$$

Next, we derive the optimal solution of penalty vectors δ_k^ϕ. Now, we define a semi-objective function $J_{kj}(\delta_{kj}^\phi)$.

$$J_{kj}(\delta_{kj}^\phi) = \sum_{i=1}^{c}(u_{ki})^m (x_{kj}^\phi + \delta_{kj}^\phi - v_{ij}^\phi)^2 + 2w_k|\delta_{kj}^\phi|. \tag{7}$$

The δ_{kj}^ϕ which minimizes J_{kj} is the optimal solution of $J_{ksp_{L_1}}$ because $J_{kep_{L_1}}(U, \Delta^\phi, V^\phi) = \sum_{j=1}^{s} J_{kj}(\delta_{kj}^\phi)$. However, the term $|\delta_{kj}^\phi|$ is piecewise linear and J_{kj} cannot be differentiated at $\delta_{kj}^\phi = 0$. Therefore, we have to consider another way to find the solutions.

Here, we consider the value of $\frac{\partial J_{ki}}{\partial \delta_{kj}}$. We can find the optimal solution is δ_{kj} as the point at which the value of $\frac{\partial J_{ki}}{\partial \delta_{kj}}$ turns into positive from negative. Thus, We consider two cases, $\delta_{kj}^\phi > 0$ and $\delta_{kj}^\phi < 0$.

We denote the semi-objective functions when $\delta_{kj}^\phi > 0$ and $\delta_{kj}^\phi < 0$ as J_{kj}^+ and J_{kj}^-, respectively and show the functions as follows:

$$J_{kj}^+(\delta_{kj}^\phi) = \sum_{i=1}^{c}(u_{ki})^m (x_{kj}^\phi + \delta_{kj}^\phi - v_{ij}^\phi)^2 + 2w_k\delta_{kj}^\phi, \ (\delta_{kj}^\phi > 0) \tag{8}$$

$$J_{kj}^-(\delta_{kj}^\phi) = \sum_{i=1}^{c}(u_{ki})^m (x_{kj}^\phi + \delta_{kj}^\phi - v_{ij}^\phi)^2 - 2w_k\delta_{kj}^\phi. \ (\delta_{kj}^\phi < 0) \tag{9}$$

Therefore, we can get $\delta_{kj}^{\phi+}$ and $\delta_{kj}^{\phi-}$ from $\frac{\partial J_{kj}^+}{\partial \delta_{kj}^\phi} = 0$ and $\frac{\partial J_{kj}^-}{\partial \delta_{kj}^\phi} =$

0 as follows:

$$\begin{cases} \delta_{kj}^{\phi+} = \dfrac{-\left(\sum_{i=1}^{c}(u_{ki})^m(x_{kj}^{\phi} - v_{ij}^{\phi}) + w_k\right)}{\sum_{i=1}^{c}(u_{ki})^m}, & (10) \\ \delta_{kj}^{\phi-} = \dfrac{-\left(\sum_{i=1}^{c}(u_{ki})^m(x_{kj}^{\phi} - v_{ij}^{\phi}) - w_k\right)}{\sum_{i=1}^{c}(u_{ki})^m}. & (11) \end{cases}$$

Therefore, we obtain the optimal solution of δ_{kj}^{ϕ} is as follows:

$$\delta_{kj}^{\phi} = \begin{cases} \delta_{kj}^{\phi+}, & (\delta_{kj}^{\phi+} > 0) \\ \delta_{kj}^{\phi-}, & (\delta_{kj}^{\phi-} < 0) \\ 0. & (\text{otherwise}) \end{cases} \quad (12)$$

2.4. Optimal Solutions of eEK-FCMP$_{L_1}$

We can derive the optimal solutions by using Lagrange multiplier. The optimal solutions of entropy based K-FCMP$_{L_1}$ are as follows:

$$\begin{cases} u_{ki} = \dfrac{\exp(-\lambda d_{ki})}{\sum_{l=1}^{c}\exp(-\lambda d_{kl})}, & (13) \\ v_i^{\phi} = \dfrac{\sum_{k=1}^{n} u_{ki}(x_k^{\phi} + \delta_k^{\phi})}{\sum_{k=1}^{n} u_{ki}}, & (14) \\ \delta_{kj}^{\phi} = \begin{cases} \delta_{kj}^{\phi+}, & (\delta_{kj}^{\phi+} > 0,\ m = 1) \\ \delta_{kj}^{\phi-}, & (\delta_{kj}^{\phi-} < 0,\ m = 1) \\ 0. & (\text{otherwise}) \end{cases} & (15) \end{cases}$$

We can formally obtain the optimal solutions as above. However, we cannot calculate v_i^{ϕ} and δ_k^{ϕ} directly by this method. The reason is that the map ϕ is not explicit and we have to use x_k, not x_k^{ϕ}. Hence, We introduce an explicit mapping into K-FCMP$_{L_1}$ to calculate v_i^{ϕ} directly.

2.5. Optimal Solutions with Explicit Mapping of Kernel Functions

In this paragraph, we introduce an explicit mapping of kernel functions [4]. As above, the kernel function K satisfies $K(x, y) = \langle \phi(x), \phi(y) \rangle = \langle x^{\phi}, y^{\phi} \rangle$. Now, we consider a kernel function with the kernel matrix $K = (K_{kl}) = (K(x_k, x_l))$. As K is a positive symmetric matrix, we can introduce a inner product $\langle x, y \rangle_K = xKy$ into $\Re^s$. We assume that $\phi : X \to \Re^n$ and $\phi(x_k) = x_k^{\phi} = e_k$. Here $e_k = (e_{k1}, \ldots, e_{ks})$ is a unit vector and e_{kj} is Kronecker delta. Because ϕ can be represented explicitly, we call ϕ using the kernel matrix K explicit mapping. Therefore, the inner product is as follows:

$$\langle e_g, e_h \rangle_k = K_{gh}. \quad (16)$$

Thus, the dissimilarity d_{ki} can be calculated as follows:

$$\begin{aligned} d_{ki} &= d(x_k^{\phi} + \delta_k^{\phi}, v_i^{\phi}) = d(e_k + \delta_k^{\phi}, v_i^{\phi}) \\ &= \langle e_k + \delta_k^{\phi} - v_i^{\phi}, e_k + \delta_k^{\phi} - v_i^{\phi} \rangle_K \\ &= (e_k + \delta_k^{\phi} - v_i^{\phi})^T K (e_k + \delta_k^{\phi} - v_i^{\phi}). \end{aligned} \quad (17)$$

We can calculate the optimal solutions by using the above inner product and dissimilarity and replacing x_k and x_{kj} by e_k and e_{kj}, respectively. We had to omit the details for want of space.

2.6. Algorithm

Here, we construct an algorithm for EK-FCMP$_{L_1}$ using the above optimal solutions with the explicit mapping as follows.

Algorithm 1 (sEK-FCMP$_{L_1}$ and eEK-FCMP$_{L_1}$)

Step 1 *Set the initial values U and Δ^{ϕ}. Generate the kernel matrix K.*

Step 2 *Calculate V^{ϕ} by the optimal solutions with the explicit mapping on fixing U and Δ^{ϕ}.*

Step 3 *Calculate U by the optimal solutions with the explicit mapping on fixing Δ^{ϕ} and V^{ϕ}.*

Step 4 *Calculate Δ^{ϕ} by the optimal solutions with the explicit mapping on fixing V^{ϕ} and U.*

Step 5 *Finish if the solutions are convergent, else go back to* Step 2.

3. Numerical Examples

We show numerical examples in Fig. 1 ~ Fig. 8. The data size is 133 and the pattern space is $[0, 1] \times [0, 1]$. We use Gaussian kernel:

$$K(x, y) = \exp(-\beta \|x - y\|^2).$$

We use the same initial values at all case and give the same weight of penalty vectors to all data.

Fig. 1 ~ Fig. 4, and Fig. 5 ~ Fig. 8 are the results by sEK-FCMP$_{L_1}$ and eEK-FCMP$_{L_1}$, respectively. We also show the cluster centers by using kernel principal component analysis (K-PCA) [12].

In Table 1 we show the range of w_k when desirable clustering results are obtained. In the range, the larger w_k gets, the faster the algorithm converges and the clearer the shape of the fuzzy classification functions becomes. In addition, we get a desirable clustering results by K-sFCMP$_{L_1}$ even when w_k is very small, and the range of w_k hardly changes even if a value of β changes. Thus, we can say that K-sFCMP$_{L_1}$ is robust for the parameters.

4. Conclusion

In this paper, we developed FCM for uncertain data with quadratic regularization term of penalty vectors by introducing explicit mapping of kernel functions. The proposed algorithm has the following features:

1. The algorithm can handle datasets which consist of some clusters with nonlinear boundary.

Table 1: The range of w_k when obtaining desirable clustering results

sEK-FCMP$_{L_1}$ ($\beta = 70, m = 1.2$)	$10^{-6} \le w_k \le 0.91$
sEK-FCMP$_{L_1}$ ($\beta = 40, m = 1.2$)	$10^{-6} \le w_k \le 0.94$
eEK-FCMP$_{L_1}$ ($\beta = 70, \lambda = 10$)	None
eEK-FCMP$_{L_1}$ ($\beta = 40, \lambda = 5.0$)	$0.94 \le w_k$

2. Because of explicit mapping of kernel functions, we reduce the calculation time.

3. The algorithm becomes more simple than the conventional ones.

In the forthcoming paper, we will discuss how to find appropriate parameters and apply to real data.

Acknowledgment

This study is partly supported by the Grant-in-Aid for Scientific Research (C) (Project No.21500212) from the Ministry of Education, Culture, Sports, Science and Technology, Japan.

References

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

[2] V. N. Vapnik: *Statistical Learning Theory*, Wiley, New York, 1998.

[3] V. N. Vapnik: *The nature of Statistical Learning Theory*, 2nd edn., Springer, New York, 2000.

[4] S. Miyamoto, Y. Kawasaki, K. Sawazaki: *An Explicit Mapping for Fuzzy c-Means Using Kernel Function and Application to Text Analysis*, IFSA/EUSFLAT 2009, 2009.

[5] Y. Endo, R. Murata, H. Haruyama, S. Miyamoto: *Fuzzy c-Means for Data with Tolerance*, Proc. 2005 International Symposium on Nonlinear Theory and Its Applications, pp.345-348, 2005.

[6] Y. Hasegawa, Y. Endo, Y. Hamasuna: *On Fuzzy c-Means for Data with Uncertainty using Spring Modulus*, SCIS&ISIS 2008, 2008.

[7] T. Miyamoto, Y. Endo, Y. Hamasuna, S. Miyamoto: *On L_1-Norm Based Fuzzy c-Means Clustering with Penalty Term*, The 6th International Conference on Modeling Decisions for Artificial Intelligence (MDAI2009), 2009.

[8] Y. Kanzawa, Y. Endo, S. Miyamoto: *Entropy Regularized Fuzzy c-Means for Data with Tolerance introducing Penalty Term in Feature Space*, SCIS&ISIS 2008, 2008.

[9] S. Miyamoto and M. Mukaidono: *Fuzzy c-means as a regularization and maximum entropy approach*, Proc. of the 7th International Fuzzy Systems Association World Congress (IFSA'97), Vol.2, pp.86–92, 1997.

[10] J. Mercer: *Functions of Positive and Negative Type and Their Connection with the Theory of Integral Equations*, Philosophical Transactions of the Royal Society A, 209, pp.415–446, 1909.

[11] S. Miyamoto, K. Umayahara, M. Mukaidono: *Fuzzy Classification Functions in the Methods of Fuzzy c-Means and Regularization by Entropy*, Journal of Japan Society for Fuzzy Theory and Systems Vol.10, No.3, pp.548–557, 1998.

[12] B. Schölkopf, A. Smola, and K.-R. Müller: *Nonlinear Component Analysis as a Kernel Eigenvalue Problem*, Neural Computation, Vol.10, pp.1299–1319, 1998.

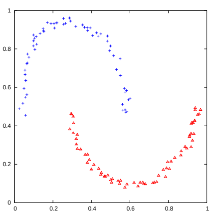

Figure 1: sEK-FCMP$_{L_1}$ ($\beta = 70, m = 1.2, w_k = 0.1$)

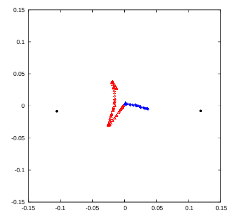

Figure 2: K-PCA for sEK-FCMP$_{L_1}$

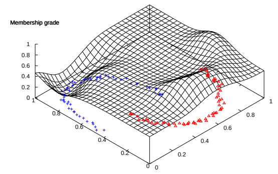

Figure 3: The fuzzy classification function for sEK-FCMP$_{L_1}$ (the 1st cluster)

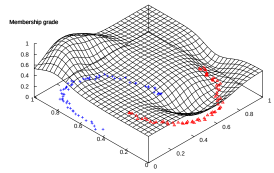

Figure 4: The fuzzy classification function for sEK-FCMP$_{L_1}$ (the 2nd cluster)

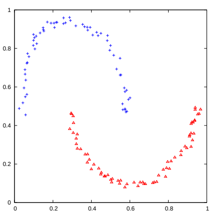

Figure 5: eEK-FCMP$_{L_1}$ ($\beta = 40, \lambda = 5.0, w_k = 5.0$)

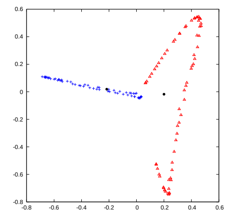

Figure 6: K-PCA for eEK-FCMP$_{L_1}$

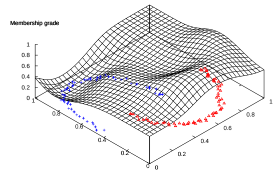

Figure 7: The fuzzy classification function for eEK-FCMP$_{L_1}$ (the 1st cluster)

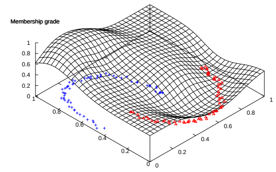

Figure 8: The fuzzy classification function for eEK-FCMP$_{L_1}$ (the 2nd cluster)