On Tolerant Fuzzy c-Means Clustering with L_1-Regularization

HAMASUNA Yukihiro[1] ENDO Yasunori[2] MIYAMOTO Sadaaki[2]

1. Doctoral Program in Risk Engineering, University of Tsukuba
1-1-1 Tennodai, Tsukuba, Ibaraki 305-8573, Japan
2. Faculty of Systems and Information Engineering, University of Tsukuba
1-1-1 Tennodai, Tsukuba, Ibaraki 305-8573, Japan
Email: yhama@soft.risk.tsukuba.ac.jp, {endo, miyamoto}@risk.tsukuba.ac.jp

Abstract— *We have proposed tolerant fuzzy c-means clustering (TFCM) from the viewpoint of handling data more flexibly. This paper presents a new type of tolerant fuzzy c-means clustering with L_1-regularization. L_1-regularization is well-known as the most successful techniques to induce sparseness. The proposed algorithm is different from the viewpoint of the sparseness for tolerance vector. In the original concept of tolerance, a tolerance vector attributes to each data. This paper develops the concept to handle data flexibly, that is, a tolerance vector attributes not only to each data but also each cluster. First, the new concept of tolerance is introduced into optimization problems. These optimization problems are based on conventional fuzzy c-means clustering (FCM). Second, the optimization problems with tolerance are solved by using Karush-Kuhn-Tucker conditions and an optimization method for L_1-regularization. Third, new clustering algorithms are constructed based on the explicit optimal solutions. Finally, the effectiveness of the proposed algorithm is verified through some numerical examples.*

Keywords— fuzzy c-means clustering, L_1-regularization, optimization, tolerance, uncertainty

1 Introduction

The aim of data mining is to discover important knowledge from a large quantity of data. From this viewpoint, clustering techniques have been actively studied. Clustering is one of the well-known unsupervised classification methods. For example, hard c-means clustering (HCM) is the most basic method. Fuzzy c-means clustering (FCM) is one of the well-known and useful clustering methods. For example, standard fuzzy c-means clustering (sFCM) [1] and entropy regularized fuzzy c-means clustering (eFCM) [2] are representatives. The entropy regularized fuzzy c-means clustering is constructed by regularization with maximum entropy function.

By the way, there are some difficulties of handling a set of data by clustering methods. Some clustering algorithms have been proposed to solve significant problems, for example, data with uncertainty, cluster size, noise or isolated data and so on. When we handle a set of data, data contains inherent uncertainty. For example, errors, ranges or some missing value of attributes are much caused. In these cases, each data is represented by an interval or a set instead of a point. In case of handling data with uncertainty, some significant methods have been proposed [3, 4]. These methods can not only handle data with uncertainty but also obtain high quality results by considering data with uncertainty. Thus, handling data with uncertainty is a very important problem in the field of data mining.

Therefore, some of the authors have proposed the original concept of tolerance [5, 6] which handle data with the above-mentioned uncertainty by using tolerance vector, and constructed some clustering algorithms [7, 8]. In these algorithms, tolerance is defined as hypersphere [5, 6] or hyper-rectangle [7, 8]. In case of hyper-rectangle, the missing value of attributes are handled successfully.

On the other hand, it is difficult to obtain clusters with different size or shape by a single-objective function, e.g., conventional FCM. In general, multi-objective optimization is considered to solve such problems. However, there are some problems by optimization, for example, how to select objective function. We have thought to be solved the above-mentioned problems by introducing a kind of “flexibility” for a pattern space and we have proposed the method to handle such “flexibility” by tolerance vector [9].

It is considered that the constraint for tolerance vector much affects classification results. It means that unsuited parameter makes trivial solutions. Therefore, regularization technique, e.g., Tikhonov’s regularization [10] method have been used to solve these ill-posed problems. The quadratic and maximum entropy functions are typical regularization techniques. Moreover, L_1-regularization is the most efficient technique to induce sparseness. In the field of regression models or machine learning, some methods with L_1-regularization have been proposed and given sparse classifiers [12, 13, 14]. In algorithms applied to L_1-regularization, a lot of vari-

ISBN: 978-989-95079-6-8

ables are calculated zero. Non-zero variables are essential to understand classification results.

First, we consider optimization problems of tolerant fuzzy c-means clustering with L_1-regularization. Second, the optimal solutions are derived by Karush-Kuhn-Tucker (KKT) conditions and an optimization method for L_1-regularization . Third, we construct a new clustering algorithm by above-mentioned processes. Finally, the effectiveness of the proposed algorithm is verified through some numerical examples.

2 Preparation

Let data set, cluster and its cluster center be $X = \{x_k \mid x_k = (x_k^1, \ldots, x_k^p)^T \in \Re^p, k = 1 \ldots n\}$, C_i $(i = 1, \ldots, c)$ and $v_i = (v_i^1, \ldots, v_i^p)^T \in \Re^p, (i = 1, \ldots, c)$, respectively. Moreover, u_{ki} is the membership grade of x_k belonging to C_i and we denote the partition matrix $U = (u_{ki})_{k=1\sim n,\ i=1\sim c}$.

2.1 Fuzzy c-means clustering

Fuzzy c-means clustering is based on optimizing an objective function under constraint for membership grade.

We consider following two types of objective functions J_s and J_e.

$$J_s(U, V) = \sum_{k=1}^{n} \sum_{i=1}^{c} (u_{ki})^m \|x_k - v_i\|^2,$$

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

J_s is a well-known objective function of standard fuzzy c-means clustering (sFCM) proposed by Bezdek [1] and J_e is an entropy regularized fuzzy c-means clustering (eFCM) [2].

The constraint for u_{ki} is as follows :

$$\sum_{i=1}^{c} u_{ki} = 1 \ , \ u_{ki} \in [0, 1] \ , \ ^\forall k.$$

2.2 Regularization

As above-mentioned, regularization is efficient techniques in the field of data mining. Tikhonov's regularization [10] method has been used to solve ill-posed problems. Many data mining algorithms which are applied to regularization have been actively studied to determine a variety of solution [12, 13, 14].

In the field of clustering, many membership regularization techniques have been proposed to obtain a variety of membership functions.

$$\min_U J(U, V) + \lambda \Omega(U),$$

where λ is a regularization parameter and $\Omega(U)$ is a regularization term.

The choice of a regularization term is important to determine a shape of a membership function. The quadratic function and maximum entropy function are typical examples.

$$\Omega(U) = \frac{1}{2} \sum_{k=1}^{n} \sum_{i=1}^{c} (u_{ki})^2,$$

$$\Omega(U) = \sum_{k=1}^{n} \sum_{i=1}^{c} u_{ki} \log u_{ki}.$$

3 The concept of tolerance

In the new concept of tolerance, the vector which attributes not only each data but also cluster center is defined as tolerance vector ε_{ki}. The tolerance κ_{ki} is defined as upper bound of tolerance vector.

From these formulation, the proposed methods can handle data more flexible than conventional methods.

We define κ_{ki}^j as the upper bound of each attribute of tolerance vector $\kappa_{ki} = (\kappa_{11}^1, \ldots, \kappa_{nc}^p) \geq 0$ and tolerance vector $E = \{\varepsilon_{ki} \mid \varepsilon_{ki} = (\varepsilon_{ki}^1, \ldots, \varepsilon_{ki}^p)^T \in \Re^p \}$ which mean the admissible range of each data, and the vector within the range of tolerance, respectively. The constraint for ε_{ki}^j is as follows :

$$|\varepsilon_{ki}^j|^2 \leq (\kappa_{ki}^j)^2 \ (\kappa_{ki}^j \geq 0), ^\forall k, \ i, \ j.$$

Figure 1 is an illustrative example about the new concept of tolerance defined as hyper-rectangle in $\Re^2$.

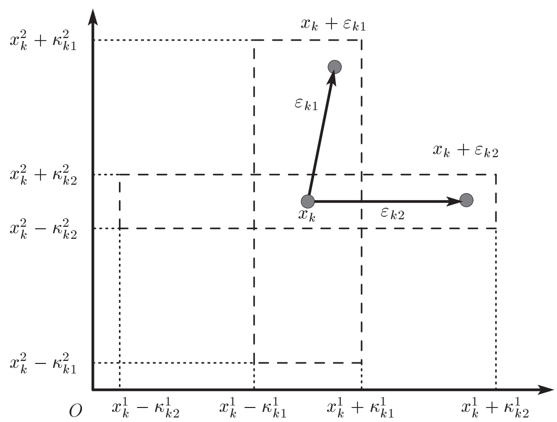

Figure 1: The new concept of tolerance defined as hyper-rectangle in $\Re^2$.

4 Tolerant fuzzy c-means clustering

In this section, we discuss optimization problems for clustering. We formulate the tolerant fuzzy c-means clustering (TFCM) by introducing the notion of tolerance into optimization problems and consider the way

ISBN: 978-989-95079-6-8

to optimize this objective function under the constraints for u_{ki} and ε^j_{ki}. The squared Euclidean-norm is used as dissimilarity, that is,

$$d_{ki} = \|x_k + \varepsilon_{ki} - v_i\|_2^2 = \sum_{j=1}^{p}(x^j_k + \varepsilon^j_{ki} - v^j_i)^2.$$

4.1 Tolerant standard fuzzy c-means clustering

The optimization problem is as follows :

$$J_{ts}(U, E, V) = \sum_{k=1}^{n}\sum_{i=1}^{c}(u_{ki})^m d_{ki}, \qquad (1)$$

under the following constraints,

$$\sum_{i=1}^{c} u_{ki} = 1 \ , \ u_{ki} \in [0, 1] \ , \ ^\forall k, \qquad (2)$$

$$|\varepsilon^j_{ki}|^2 \le (\kappa^j_{ki})^2 \ (\kappa^j_{ki} \ge 0), ^\forall k, \ i, \ j. \qquad (3)$$

The goal is to find the solutions which minimize the objective function (1) under the constraints (2) and (3).

From the convexity of (1), we introduce the following Lagrangian function to solve this optimization problem.

The Lagrangian function L_{tsr} is as follows :

$$\begin{aligned} L_{tsr} =& J_{ts}(U, E, V) + \sum_{k=1}^{n}\gamma_k(\sum_{i=1}^{c} u_{ki} - 1) \\ &+ \sum_{k=1}^{n}\sum_{i=1}^{c}\sum_{j=1}^{p}\delta^j_{ki}(|\varepsilon^j_{ki}|^2 - (\kappa^j_{ki})^2). \end{aligned}$$

Karush-Kuhn-Tucker conditions (KKT conditions) are as follows:

$$\begin{cases} \frac{\partial L_{tsr}}{\partial v^j_i} = 0, \ \frac{\partial L_{tsr}}{\partial \varepsilon^j_{ki}} = 0, \ \frac{\partial L_{tsr}}{\partial u_{ki}} = 0, \ \frac{\partial L_{tsr}}{\partial \gamma_k} = 0, \\ \frac{\partial L_{tsr}}{\partial \delta^j_{ki}} \le 0, \ \delta^j_{ki}\frac{\partial L_{tsr}}{\partial \delta^j_{ki}} = 0, \ \delta^j_{ki} \ge 0. \end{cases} \qquad (4)$$

First we consider u_{ki}, from $\frac{\partial L_{tsr}}{\partial u_{ki}} = 0$, we have,

$$u_{ki} = \left(\frac{\gamma_k}{m d_{ki}}\right)^{\frac{1}{m-1}}. \qquad (5)$$

In addition, from the constraint (2),

$$\sum_{l=1}^{c}\left(\frac{\gamma_k}{m d_{kl}}\right)^{\frac{1}{m-1}} = 1. \qquad (6)$$

From (5) and (6), we have,

$$u_{ki} = \frac{\frac{1}{(d_{ki})^{\frac{1}{m-1}}}}{\sum_{l=1}^{c}\frac{1}{(d_{kl})^{\frac{1}{m-1}}}}. \qquad (7)$$

If some $x_k + \varepsilon_{ki} = v_i$, we set $u_{ki} = 1/|C'|$. Here, $|C'|$ is number of cluster centers which satisfies $x_k + \varepsilon_{ki} = v_i$.

For v^j_i, from $\frac{\partial L_{tsr}}{\partial v^j_i} = 0$,

$$v^j_i = \frac{\sum_{k=1}^{n}(u_{ki})^m(x^j_k + \varepsilon^j_{ki})}{\sum_{k=1}^{n}(u_{ki})^m}. \qquad (8)$$

For ε^j_{ki} from $\frac{\partial L_{tsr}}{\partial \varepsilon^j_{ki}} = 0$, we can get

$$\varepsilon^j_{ki} = -\frac{(u_{ki})^m(x^j_k - v^j_i)}{(u_{ki})^m + \delta^j_{ki}}. \qquad (9)$$

From $\delta^j_{ki}\frac{\partial L_{tsr}}{\partial \delta^j_{ki}} = 0$,

$$\delta^j_{ki}(|\varepsilon^j_{ki}|^2 - (\kappa^j_{ki})^2) = 0. \qquad (10)$$

From (10), we should consider two cases, i.e., $\delta^j_{ki} = 0$ and $|\varepsilon^j_{ki}|^2 = (\kappa^j_{ki})^2$. First, we consider the case of $\delta^j_{ki} = 0$. In this case, the constraint (3) is not considered. From (9), we can get,

$$\varepsilon^j_{ki} = -(x^j_k - v^j_i).$$

On the other hand, in case that $|\varepsilon^j_{ki}|^2 = (\kappa^j_{ki})^2$,

$$|\varepsilon^j_{ki}|^2 = |-\frac{(u_{ki})^m(x^j_k - v^j_i)}{(u_{ki})^m + \delta^j_{ki}}|^2 = (\kappa^j_{ki})^2.$$

From $(u_{ki})^m + \delta^j_{ki} > 0$,

$$\frac{(u_{ki})^m}{(u_{ki})^m + \delta^j_{ki}} = \frac{\kappa^j_{ki}}{|x^j_k - v^j_i|}. \qquad (11)$$

From (9), (11),

$$\varepsilon^j_{ki} = \frac{-\kappa^j_{ki}(x^j_k - v^j_i)}{|x^j_k - v^j_i|}.$$

From the above, we can get an optimal solution for ε^j_{ki} as follows :

$$\begin{aligned} \varepsilon^j_{ki} =& -\alpha^j_{ki}(x^j_k - v^j_i), \qquad (12) \\ \alpha^j_{ki} =& \min\left\{\frac{\kappa^j_{ki}}{|x^j_k - v^j_i|}, 1\right\}. \end{aligned}$$

4.2 Tolerant entropy regularized fuzzy c-means clustering

The optimization problem is as follows :

$$\begin{aligned} J_{te}(U, E, V) =& \sum_{k=1}^{n}\sum_{i=1}^{c} u_{ki} d_{ki} \\ &+ \lambda^{-1}\sum_{k=1}^{n}\sum_{i=1}^{c} u_{ki}\log u_{ki}. \end{aligned}$$

ISBN: 978-989-95079-6-8

Constraints are same as (2) and (3).

The Lagrangian function L_{er} is as follows :

$$L_{ter} = J_{te}(U,E,V) + \sum_{k=1}^{n} \gamma_k (\sum_{i=1}^{c} u_{ki} - 1) + \sum_{k=1}^{n}\sum_{i=1}^{c}\sum_{j=1}^{p} \delta_{ki}^j (|\varepsilon_{ki}^j|^2 - (\kappa_{ki}^j)^2).$$

KKT conditions are same as (4).

The optimal solutions are derived as follows :

$$u_{ki} = \frac{\exp(-\lambda d_{ki})}{\sum_{l=1}^{c}\exp(-\lambda d_{kl})}, \quad (13)$$

$$v_i^j = \frac{\sum_{k=1}^{n} u_{ki}(x_k^j + \varepsilon_{ki}^j)}{\sum_{k=1}^{n} u_{ki}}, \quad (14)$$

$$\varepsilon_{ki}^j = -\alpha_{ki}^j (x_k^j - v_i^j),$$

$$\alpha_{ki}^j = \min\left\{ \frac{\kappa_{ki}^j}{|x_k^j - v_i^j|}, 1 \right\}.$$

5 Tolerant fuzzy c-means clustering with L_1-regularization

In this section, we will consider tolerant fuzzy c-means clustering with L_1-regularization. In this method, constraint for ε_{ki}^j is not considered.

5.1 L_1-regularization

It is well-known that L_1-regularization can induce the strong sparseness of the variables [12, 13, 14]. A lot of ε_{ki}^j become zero, by L_1-regularization described as follows :

$$\Omega(E) = \sum_{k=1}^{n}\sum_{i=1}^{c}\sum_{j=1}^{p} |\varepsilon_{ki}^j|.$$

Here, we will describe objective function with L_1-regularization as J_{tl}. The partial derivatives of J_{tl} which respect to ε_{ki}^j will be uniformly zero, as follows :

$$\left|\frac{\partial J_{tl}}{\partial \varepsilon_{ki}^j}\right| = \nu \qquad \text{if } |\varepsilon_{ki}^j| > 0,$$

$$\left|\frac{\partial J_{tl}}{\partial \varepsilon_{ki}^j}\right| < \nu \qquad \text{if } |\varepsilon_{ki}^j| = 0.$$

Here, $\nu > 0$ is a regularization parameter. This denotes that if the partial derivatives of J below ν, ε_{ki}^j will be set exactly zero.

5.2 Tolerant standard fuzzy c-means clustering with L_1-regularization

We will consider the following objective function with L_1-regularization.

$$J_{tsl}(U,E,V) = \sum_{k=1}^{n}\sum_{i=1}^{c} (u_{ki})^m d_{ki} + \nu \sum_{k=1}^{n}\sum_{i=1}^{c}\sum_{j=1}^{p} |\varepsilon_{ki}^j|.$$

In the tolerant fuzzy c-means clustering, each tolerance vector can be solved separately. So, we consider the following semi-objective function J_{tsl}^{kij} :

$$J_{tsl}^{kij}(E) = (u_{ki})^m (x_k^j + \varepsilon_{ki}^j - v_i^j)^2 + \nu|\varepsilon_{ki}^j|.$$

To obtain partial derivatives respect to ε_{ki}^j, we will decompose $\varepsilon_{ki}^j = \xi^+ - \xi^-$, where all element of ξ^+ and ξ^- are nonnegative. Thus, the semi-objective function can be written as follows :

$$J_{tsl}^{kij}(E) = (u_{ki})^m (x_k^j + \varepsilon_{ki}^j - v_i^j)^2 + \nu(\xi^+ + \xi^-).$$

The constraints are as follows :

$$\varepsilon_{ki}^j \le \xi^+, \quad \varepsilon_{ki}^j \ge -\xi^-, \quad \xi^+, \xi^- \ge 0.$$

Introducing the Lagrange multiplier β^+, β^-, δ^+ and $\delta^- \ge 0$, Lagrangian L_{tsl} is as follows :

$$L_{tsl} = (u_{ki})^m (x_k^j + \varepsilon_{ki}^j - v_i^j)^2 + \nu(\xi^+ + \xi^-) + \beta^+(\varepsilon_{ki}^j - \xi^+) + \beta^-(-\varepsilon_{ki}^j - \xi^-) - \delta^+\xi^+ - \delta^-\xi^-.$$

Here,

$$\frac{\partial L_{tsl}}{\partial \xi^+} = \nu - \beta^+ - \delta^+, \quad (15)$$

$$\frac{\partial L_{tsl}}{\partial \xi^-} = \nu - \beta^- - \delta^-. \quad (16)$$

Since $\delta^+,\ \delta^- \ge 0$, conditions $0 \le \beta^+ \le \nu$ and $0 \le \beta^- \le \nu$ are obtained from (15) and (16), respectively. By using (15) and (16), the Lagrangian L_{tsl} is simplified as follows :

$$L_{tsl} = (u_{ki})^m (x_k^j + \varepsilon_{ki}^j - v_i^j)^2 + \beta\varepsilon_{ki}^j. \quad (17)$$

Here, $\beta = \beta^+ - \beta^-$ and satisfies condition $-\nu \le \beta \le \nu$.

From $\frac{\partial L_{tsl}}{\partial \varepsilon_{ki}^j} = 0$,

$$\frac{\partial L_{tsl}}{\partial \varepsilon_{ki}^j} = 2(u_{ki})^m (x_k^j + \varepsilon_{ki}^j - v_i^j) + \beta = 0.$$

From above,

$$\varepsilon_{ki}^j = -(x_k^j - v_i^j) - \frac{\beta}{2(u_{ki})^m}. \quad (18)$$

Introducing (18) to (17), the Lagrangian dual problem is written as follows :

$$L_{tsld} = -\frac{\beta^2}{4(u_{ki})^m} - \beta(x_k^j - v_i^j).$$

ISBN: 978-989-95079-6-8

This objective function is a quadratic function respect to β. From $\frac{\partial L_{tsld}}{\partial \beta} = 0$, this dual problem is solved as,

$$\beta = -2(u_{ki})^m (x_k^j - v_i^j). \tag{19}$$

From considering (18) and (19), the optimal solution of primal problem is derived. First, if $\beta \leq -\nu$, the optimal solution is $\beta = -\nu$. Second, if $-\nu \leq \beta \leq \nu$, the optimal solution is $\beta = -2(u_{ki})^m (x_k^j - v_i^j)$. Third, if $\nu \leq \beta$, the optimal solution is $\beta = \nu$. Finally, the optimal solution for ε_{ki}^j of L_1-regularized objective function is derived as follows :

$$\varepsilon_{ki}^j = \begin{cases} -(x_k^j - v_i^j) + \frac{\nu}{2(u_{ki})^m} & (\beta \leq -\nu), \\ 0 & (-\nu \leq \beta \leq \nu), \\ -(x_k^j - v_i^j) - \frac{\nu}{2(u_{ki})^m} & (\nu \leq \beta). \end{cases} \tag{20}$$

5.3 Tolerant entropy regularized fuzzy c-means clustering with L_1-regularization

From the same procedure, the optimal solution for ε_{ki}^j of tolerant entropy regularized fuzzy c-means clustering with L_1-regularization is derived as follows :

$$\varepsilon_{ki}^j = \begin{cases} -(x_k^j - v_i^j) + \frac{\nu}{2u_{ki}} & (\beta \leq -\nu), \\ 0 & (-\nu \leq \beta \leq \nu), \\ -(x_k^j - v_i^j) - \frac{\nu}{2u_{ki}} & (\nu \leq \beta). \end{cases} \tag{21}$$

6 Algorithms

Algorithms of TFCM derived in the above section are called as follows. In case of hyper-rectangle, we call these methods TsFCM(R) and TeFCM(R). In case of L_1-regularization, we call these methods TsFCM-L_1R, and TeFCM-L_1R.

Each algorithm of TFCM is calculated according to the following procedure. Eqs. **A**, **B** and **C** used in each algorithm follow **Table 1** .

Algorithm 1

TFCM1 *Set the initial values and parameters.*

TFCM2 *Calculate $u_{ki} \in U$ by Eq.* **A**.

TFCM3 *Calculate $v_i \in V$ by Eq.* **B**.

TFCM4 *Calculate $\varepsilon_{ki} \in E$ by Eq.* **C**. *If convergence criterion is satisfied, stop. Otherwise, go back to* **TFCM2**.

In these algorithms, the convergence criterion is convergence of each variable, value of objective function or number of repetition.

Table 1: The optimal solutions of each algorithm.

Algorithm	Eq.**A**	Eq.**B**	Eq.**C**
TsFCM(R)	(7)	(8)	(12)
TeFCM(R)	(13)	(14)	(12)
TsFCM-L_1R	(7)	(8)	(20)
TeFCM-L_1R	(13)	(14)	(21)

7 Numerical examples

In this section, some numerical examples of classification are shown. In these examples, '$\circ$', '$\triangle$', '$\square$' and '$*$' mean each cluster and cluster centers, respectively. Moreover, tolerance vectors are expressed by arrowed line. The value of each data are normalized between 0 and 10. In addition, $m = 2.0$ in TsFCM and $\lambda = 1.0$ in TeFCM.

A polaris data set is mapped into two dimensional pattern space and consists of 51 points. This data set should be classified into three clusters [1]. Fig. 2 and 3 are classification results of TFCM with L_1-regularization. In Fig. 2, $\nu = 2.0$ and the number of non-zero tolerance vector is 30. In Fig. 3, $\nu = 4.0$ and the number of non-zero tolerance vector is 15. Moreover, Fig. 4 shows the relation between regularization parameter ν and the average of zero parameter ratio out of 1000 trials.

From these results, it is verified that proposed algorithm with L_1-regularization can induce the strong sparseness and its sparseness is controlled by regularization parameter ν.

8 Conclusions

In this paper, we have formulated the optimization problems based on concept of tolerance and derived the optimal solutions of tolerant fuzzy c-means clustering with L_1-regularization. From these results, we have constructed new clustering algorithm. Moreover, we have verified the effectiveness of proposed algorithm through some numerical examples.

The proposed technique is essentially different from the past one from the viewpoint of handling data more "flexible" than conventional methods.

In future works, we will calculate with real data which includes data with uncertainty, e.g., the missing value of attributes. Next, we will consider another type of regularization term or constraint for tolerance vector to induce sparseness.

Acknowledgment

This study is partly supported by Research Fellowships of the Japan Society for the Promotion of Science for Young Scientists and the Grant-in-Aid for Scientific Research (C) and (B) (Project No.21500212 and

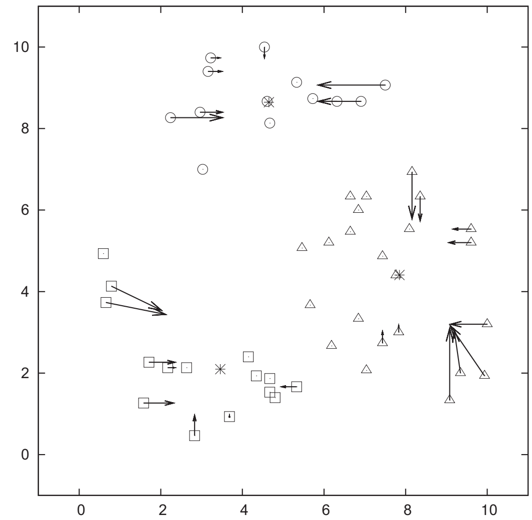

Figure 2: Result of TsFCM with L_1-regularization.

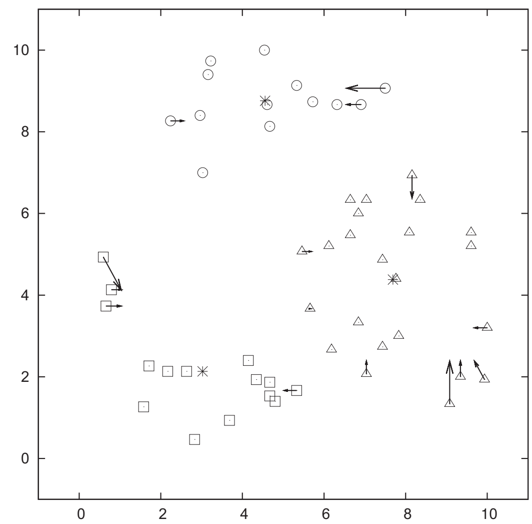

Figure 3: Result of TeFCM with L_1-regularization.

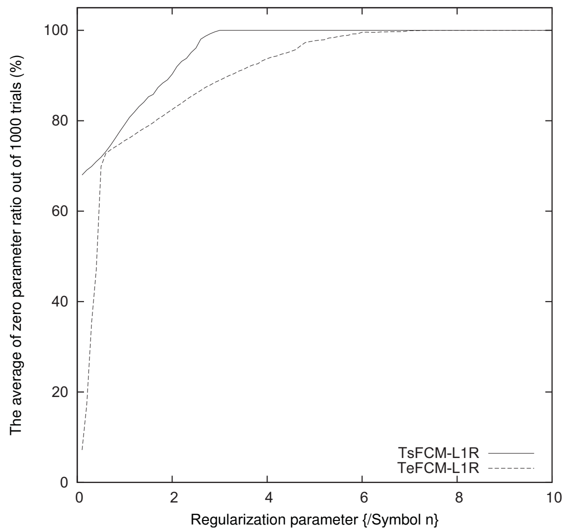

Figure 4: The relation between regularization parameter ν and the average of zero parameter ratio out of 1000 trials

No.19300074) from the Ministry of Education, Culture, Sports, Science and Technology, Japan.

References

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

[2] 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), June 25-30, 1997, Prague, Czech, Vol. 2, pp. 86–92, 1997.

[3] W. K. Ngai, B. Kao, C. K. Chui, R. Cheng, M. Chau and K. Y. Yip, "Efficient Clustering of Uncertain Data", Proc. of the Sixth International Conference on Data Mining (ICDM'06), December 18-22, 2006, Hong Kong, China, pp. 436–445, 2006.

[4] O. Takata, and S. Miyamoto, "Fuzzy clustering of Data with Interval Uncertainties", Journal of Japan Society for Fuzzy Theory and Systems, Vol.12, No.5, pp. 686–695, 2000 (in Japanese).

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

[6] R. Murata, Y. Endo, H. Haruyama and S. Miyamoto, "On Fuzzy c-Means for data with Tolerance", Journal of Advanced Computational Intelligence and Intelligent Informatics 10(5), pp. 673–681, 2006.

[7] Y. Hasegawa, Y. Endo, Y. Hamasuna and S. Miyamoto, "Fuzzy c-Means for Data with Tolerance Defined as Hyper-rectangles", Modeling Decisions for Artificial Intelligence (MDAI2007), pp. 237-248. Springer, Heidelberg 2007.

[8] Y. Hamasuna, Y. Endo, S. Miyamoto and Y. Hasegawa, "On Hard Clustering for Data with Tolerance", Journal of Japan Society for Fuzzy Theory and Intelligent Informatics, Vol.20, No.3, pp. 388–398, 2008 (in Japanese).

[9] Hamasuna Y., Endo Y. and Miyamoto S., "On Tolerant Fuzzy c-Means", Journal of Advanced Computational Intelligence and Intelligent Informatics (JACIII), (to be published).

[10] Tikhonov, A. N. and V. Y. Arsenin, 'Solutions of Ill-Posed Problems', V. H. Winston, 1977.

[11] P. M. Williams, "Bayesian Regularization and Pruning Using a Laplace Prior", Neural Computation, Vol. 7, No.1, pp. 117–143, 1995.

[12] K. Tsuda and T. Kudo, "Clustering Graphs by Weighted Substructure Mining", Proc. of the 23rd international conference on Machine learning, pp. 953–960, 2006.

[13] R. Tibshirani, "Regression shrinkage and selection via the Lasso", Journal of the Royal Statistical Society, Series B, Vol. 58, No.1, pp. 267-288, 1996.

[14] J. Kazama and J. Tsuji, "Evaluation and extension of maximum entropy models with in equality constraints", In Proc. of EMNLP, pp. 137–144, 2003.

ISBN: 978-989-95079-6-8