Robust Extension of FCMdd-based Linear Clustering for Relational Data using Alternative c-Means Criterion

Takeshi Yamamoto†, Katsuhiro Honda† Akira Notsu† and Hidetomo Ichihashi†

† Osaka Prefecture University, 1-1 Gakuen-cho, Nakaku, Sakai, Osaka, JAPAN

Summary
Relational clustering is an extension of clustering for relational data. Fuzzy *c*-Medoids (FCMdd) based linear fuzzy clustering extracts intrinsic local linear substructures from relational data. However this linear clustering was affected by noise or outliers because of using Euclidean distance. Alternative Fuzzy *c*-Means (AFCM) is an extension of Fuzzy *c*-means, in which a modified distance measure based on the robust *M*-estimation concept can decrease the influence of noise or outliers more than the conventional Euclidean distance. In this paper, robust FCMdd-based linear clustering model is proposed in order to extract linear substructure from relational data including outliers, using a pseudo-*M*-estimation procedure with a weight function for the modified distance measure in AFCM.

Key words:
Fuzzy clustering, Relational data, Robust clustering.

1. Introduction

Clustering is an unsupervised classification method for extracting intrinsic structure from objects. Fuzzy *c*-Means (FCM) [1] was proposed by Bezdek, which is a fuzzy extension of *k*-means [2]. Relational Fuzzy *c*-Means (RFCM) [3] is an extension of FCM for relational data. RFCM can be identified with FCM when mutual relations are given by Euclidean distances. In Fuzzy *c*-Medoids (FCMdd) [4], cluster prototypes are chosen from objects instead of cluster mean vectors. The representative objects are called ‘medoids’. Because its clustering criteria are calculated only using mutual distances among objects, the clustering algorithm can be seen as a kind of RFCM.

In linear fuzzy clustering, Fuzzy *c*-Lines (FCL) [5] and Fuzzy *c*-Varieties [6] were proposed using linear-shape prototypes such as lines, planes and linear varieties. Because the subspace learning model in each cluster can be identified with fuzzy principal component analysis (fuzzy PCA) [7], they are often regarded as a kind of local principal component analysis (local PCA) [8-10].

Haga *et al*. [11] proposed an extended model of FCMdd for linear fuzzy clustering of relational data spanning linear prototypes, and demonstrated that local features of intrinsic linear dependencies in relational data can be found in a scheme of multi-cluster-type multi-dimensional scaling [12], [13].

In using Euclidean distance, clustering results are affected by noise or outliers. Alternative Fuzzy *c*-Means (AFCM) using a robust distance measure was proposed by Wu and Yang [14], and is not easily influenced by noise. In this paper, the FCMdd-based linear clustering model is extended to robust linear clustering by applying the modified distance measure in AFCM. Experimental results are shown to demonstrate the applicability of the proposed method in relational data including noise or outliers.

The remaining parts of this paper are organized as follows: Section 2 gives a brief review on various fuzzy clustering methods. In Section 3, FCMdd-based linear clustering is introduced and applied alternative distance measure in AFCM. Several experimental results are shown in Section 4. Section 5 includes summary conclusions.

2. Brief Review of Fuzzy Clustering Methods

2.1 Fuzzy *c*-Means Clustering

Let $\mathbf{x}_i = (x_{i1},\ldots,x_{im})^{\mathrm{T}},\ i=1,\ldots,n$ be m-dimensional observations of n samples. In FCM [1], the goal is to partition the n samples into C clusters through an iterative procedure of prototype and fuzzy membership estimation. The objective function of FCM is defined as:

$$L_{fcm} = \sum_{c=1}^{C}\sum_{i=1}^{n} u_{ci}^{\theta} D_{ci}, \tag{1}$$

where u_{ci} is the fuzzy membership degree of sample i to cluster c, and θ is the fuzzification parameter. The larger the θ, the fuzzier the membership assignment. This parameter is recommended to use from 1.5 to 2.0 [1]. D_{ci} is the clustering criterion given as the squared Euclidean distance between sample $\mathbf{x}_i$ and prototype of cluster $\mathbf{b}_c$:

$$D_{ci} = \| \mathbf{x}_i - \mathbf{b}_c \|^2 . \tag{2}$$

The updating rules for membership u_{ci} and the cluster center $\mathbf{b}_c$ are derived as:

Manuscript received November 15, 2010
Manuscript revised November 20, 2010

$$u_{ci} = \left[\sum_{l=1}^{C} \left(\frac{D_{ci}}{D_{li}} \right)^{\frac{1}{\theta-1}} \right]^{-1}, \tag{3}$$

$$\mathbf{b}_c = \frac{\sum_{i=1}^{n} u_{ci}^{\theta} \mathbf{x}_i}{\sum_{i=1}^{n} u_{ci}^{\theta}}. \tag{4}$$

2.2 Relational Fuzzy *c*-Means and Fuzzy *c*-Medoids Clustering for Relational Data

Assuming that we have relational data composed of mutual relations among samples $D = \{d_{ij}^2\}$, FCM cannot be applied to relational data directly. Therefore Relational Fuzzy *c*-Means (RFCM) was proposed and the FCM-type objective function is redefined as [3]:

$$L_{rfcm} = \sum_{c=1}^{C} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{u_{ci}^{\theta} u_{cj}^{\theta} d_{ij}^2}{2\sum_{t=1}^{n} u_{ct}^{\theta}}, \tag{5}$$

where d_{ij} can be any type of dissimilarity between samples *i* and *j*, but is assumed to be Euclidean-like one in RFCM. The relational clustering model is equivalent to FCM with object type data when the mutual relation d_{ij} is given by the Euclidean distance.

Fuzzy *c*-Medoids (FCMdd) [4] is a fuzzy extension of *k*-Medoids clustering [15], in which prototype of cluster $\mathbf{b}_c$ is not calculated but selected from objects $\mathbf{x}_i$ considering fuzzy memberships. The representative objects are called 'medoids' and D_{ci} is given by d_{ci}^2 where $\mathbf{b}_c = \mathbf{x}_c$. Because of medoids, FCMdd can be identified with a kind of RFCM and applied to relational data directly.

2.3 Alternative *c*-Means Clustering

Least square models using Euclidean distance are easily influenced by noise or outliers. In order to decrease the noise sensitivity in *c*-Means-type clustering, Wu and Yang proposed a robust measure [14]:

$$d_{ci} = 1 - \exp(-\beta \|\mathbf{x_i} - \mathbf{b_c}\|^2), \tag{6}$$

where β is a weight parameter for tuning the noise sensitivity and Wu and Yang recommended to use the sample variance. This measure can be regarded as a type of robust *M*-estimator used in maximum likelihood estimation. In Alternative Fuzzy *c*-Means (AFCM), the cluster center is updated according to an additional weight parameter w_{ci} as follows:

$$\mathbf{b}_c = \frac{\sum_{i=1}^{n} u_{ci}^{\theta} w_{ci} \mathbf{x}_i}{\sum_{i=1}^{n} u_{ci}^{\theta}}, \tag{7}$$

where the typicality weight w_{ci} is given as:

$$w_{ci} = \exp(-\beta \|\mathbf{x_i} - \mathbf{b_c}\|^2). \tag{8}$$

$w_{ci} = 1$ for $\|\mathbf{x_i} - \mathbf{b_c}\|^2 = 0$ and w_{ci} becomes smaller as $\|\mathbf{x_i} - \mathbf{b_c}\|^2$ is larger. The minimum value is $w_{ci} \to 0$. By decreasing the typicality weight of outliers, robust prototypes are estimated.

2.4 Fuzzy *c*-Lines Clustering

Besides point-type prototypes in FCM, Fuzzy *c*-Lines (FCL) [5] extended to linear prototypes defined as:

$$Line_c(\mathbf{b}_c, \mathbf{a}_c) = \{\mathbf{x} \mid \mathbf{x} = \mathbf{b}_c + t\mathbf{a}_c;\, t \in R\}, \tag{9}$$

where $\mathbf{a}_c$ is the basis vector of the principal sub-space and $\mathbf{b}_c$ is the centroid, which the linear prototypes pass through. The clustering criterion can be calculated as:

$$D_{ci} = \| \mathbf{x}_i - \mathbf{b}_c \|^2 - | \mathbf{a}_c^{\mathrm{T}} (\mathbf{x}_i - \mathbf{b}_c) |^2. \tag{10}$$

The updating rules for membership u_{ci} and the cluster center $\mathbf{b}_c$ are also derived according to Eq. (3), (4). The basis vectors $\mathbf{a}_c$ are the principal eigenvectors of the generalized fuzzy scatter matrix:

$$S_{fc} = \sum_{i=1}^{n} u_{ci}^{\theta} (\mathbf{x}_i - \mathbf{b}_c)(\mathbf{x}_i - \mathbf{b}_c)^{\mathrm{T}}. \tag{11}$$

In Fuzzy *c*-Varieties (FCV) [6], prototypical linear varieties spanned by multiple are also derived by solving the above eigenvalue problem.

Honda *et al.* [16] extended the AFCM clustering model to linear fuzzy clustering, which is called Alternative Fuzzy *c*-Lines (AFCL). In AFCL, the robust measure of AFCM is applied to linear prototypes and robust linear substructures are extracted by minimizing least square criterion in an *M*-estimation-like iterative procedure.

3. An Alternative Model of Robust FCMdd-based Linear Fuzzy Clustering for Relational data

3.1 FCMdd-based Linear Clustering

Let d_{ij} be the mutual Euclidean distance such that

$$d_{ii} = 0, \quad d_{ij} \geq 0, \quad d_{ij} = d_{ji}, \quad i, j = 0, \ldots, n. \tag{12}$$

Haga *et al*. [12] applied FCMdd to linear fuzzy clustering, in which two representative medoids are selected in each cluster and the prototypical line is defined as:

$$Line_c(\mathbf{x}_{c_1}, \mathbf{x}_{c_2}) = \{\mathbf{x} \mid \mathbf{x} = \mathbf{x}_{c_1} + t(\mathbf{x}_{c_2} - \mathbf{x}_{c_1}); t \in R\}. \tag{13}$$

The clustering criterion is given as the squared Euclidean distance between object *i* and the prototypical line $Line_c$:

$$D_{ci} = d_{ic_1}^2 - \frac{\left(d_{ic_1}^2 - d_{ic_2}^2 + d_{c_1c_2}^2\right)^2}{4d_{c_1c_2}^2}. \tag{14}$$

With fixed fuzzy memberships u_{ci}, the optimal medoids are derived by the following combinatorial optimization problem:

$$(c_1, c_2) = \mathop{\arg\min}_{\substack{(k_1,k_2) \\ 1 \leq k_1,k_2 \leq n \\ k_1 \neq k_2}} \sum_{i=1}^{n} u_{ci}^{\theta} D_{ci}. \tag{15}$$

To reduce the computational cost of the combinatorial optimization, the optimal medoid set of (c_1, c_2) can be selected from a subset X_c of objects having large memberships:

$$(c_1, c_2) = \mathop{\arg\min}_{\substack{(k_1,k_2) \\ \mathbf{x}_{k_1},\mathbf{x}_{k_2} \in X_c \\ k_1 \neq k_2}} \sum_{i=1}^{n} u_{ci}^{\theta} D_{ci}, \tag{16}$$

where $X_c = \{\mathbf{x}_i : u_{ci} > M_{\min}\}$. This linear fuzzy clustering model was also extended to Plane-like 2-D prototypes defined by using three medoids [12].

3.2 Application of AFCM Measure to FCMdd-based Linear Fuzzy Clustering

In this paper robust linear fuzzy clustering for relational data is proposed by applying the robust measure in AFCM to FCMdd-based linear fuzzy clustering. The goal is to extract local intrinsic linear dependencies rejecting noise. By extending the least square measure of Eq.(13) into a robust one, the clustering criterion for sample *i* in cluster *c* is rewritten as follows:

$$D_{ci} = 1 - \exp\left(-\beta\left(d_{ic_1}^2 - \frac{\left(d_{ic_1}^2 - d_{ic_2}^2 + d_{c_1c_2}^2\right)^2}{4d_{c_1c_2}^2}\right)\right). \tag{17}$$

Medoids of each cluster and fuzzy memberships are estimated considering the robust measure. This robust measure emphasizes the deviations in local area near prototypes but weaken the influences of noise. Then, the sample near prototypes can be assigned to a single cluster while noise samples have similar distances to all prototypes, i.e., noise have relatively small and equal memberships to all clusters and medoids are selected rejecting the influences of noise.

The responsibility of each sample in estimating the cluster prototype $Line_c$ is estimated by the typicality weight w_{ci} in a similar manner to AFCM as follows:

$$w_{ci} = \exp\left(-\beta\left(d_{ic_1}^2 - \frac{\left(d_{ic_1}^2 - d_{ic_2}^2 + d_{c_1c_2}^2\right)^2}{4d_{c_1c_2}^2}\right)\right), \tag{18}$$

which is only used in regarding points as outliers because prototypes are not calculated but selected in FCMdd-type clustering.

A sample procedure is summarized as follows:

Step 1 Set β. Randomly initialize the prototypical medoids (two representative objects) of each cluster.

Step 2 Calculate clustering criteria D_{ci} according to Eq.(17).

Step 3 Update fuzzy memberships by Eq.(3) using robust measure of Eq.(17).

Step 4 Search medoids in each cluster based on Eq.(15) or Eq.(16) using robust measure of Eq.(17).

Step 5 Repeat Steps 2-5 until a certain stop criterion is satisfied.

4. Numerical Experiments

In this section, three experimental results are shown in order to demonstrate the characteristic features of the proposed approach. First experiment compares the proposed approach with the conventional one using artificial data. Second experiment inspects the influence of noise for the various value of β in proposed method. The last is applied to relational data of real world data to demonstrate the applicability of alternative distance measure.

4.1 Comparison of Proposed and Conventional Approaches using Artificial Data Set

The first numerical experiment was performed using an artificially generated relational data set. The relational data set composed of 80 objects including outliers was generated from a 2-D data set, in which objects form two line-shape clusters shown in Fig. 1. A 80×80 relational matrix was constructed by calculating mutual Euclidean distances among objects and the FCMdd-based linear clustering models were applied to this relational matrix. The iterative algorithm was performed until the medoids become unchanged and the model parameters were set as $(C,\theta) = (2,2)$. In order to emphasize the characteristics of the algorithm, the initial memberships were given in a supervised manner, i.e., $(u_{c_1}, u_{c_2}) = (0.9, 0.1)$ for the first visual cluster and $(u_{c_1}, u_{c_2}) = (0.1, 0.9)$ for the second one. The noise sensitivity weight β was given as $\beta = 0.7$ that is almost equal to the sample variance.

Figure.2 compares the clustering results of the proposed approach with that of the conventional FCMdd-based linear clustering model, where clustering results are depicted in the original 2-D data space. Objects were partitioned into two clusters of circles and times, and bigger black circles indicate cluster medoids. Smaller black circles are outliers whose typicality weights w_{ci} were smaller than 0.8 for both clusters.

The conventional method using Euclidean distance measure failed to extract linear substructure from noisy relational data, in which noise or outliers affected clustering results. Proposed method in $\beta = 0.7$, which is almost equal to the sample variance, succeeded in extracting linear substructure from relational data by rejecting the influence of noise.

4.2 The Influence of β in Proposed Method

In next numerical experiment, the influence of β on clustering results was verified. In first experiment, β was set as the sample variance of relational data, but this experiment was performed by various value of β. The model parameters were also set as $(C,\theta) = (2,2)$ and the initial memberships were given in a supervised manner.

Clustering results are depicted in Fig.3. Smaller value of β such as $\beta = 0.05$, resulted in Fig.3(a), which was similar partition to the conventional method. Therefore proposed method in smaller β could be regarded as conventional one. Fig.3(b), (c) indicated that bigger value of β decreased the influence of noise or outliers and more points are regarded as noise correctly. However much bigger β such resulted in Fig.3(d), in which some points belonging to linear substructure were regarded as noise. That's why it was necessary to select β appropriately and setting β as the sample variance recommended in [15] worked well.

4.3 Rothkopf Morse Code Confusion Data

Next, the proposed approach is applied to real world data, the classical Rothkopf Morse code confusion data [14]. The 36×36 relational data is composed of confusion rates among 36 Morse code signals (26 letters and 10 digits). Before experiment, the asymmetric relational data matrix, whose element s_{ij} represents the percentage of trials that *i* was identified as *j*, was transformed into a symmetric relational data matrix as follows:

$$d_{ij} = s_{ii} + s_{jj} - s_{ij} - s_{ji} \tag{19}$$

Figure.4 shows the 2-D plots of the relational data constructed by MDS [13]. In the 2-D plots, the majority of objects form a rough curve in the upper area while other few objects are assigned to lower area. In this experiment, the goal is to reveal the curve structure using piecewise linear prototypes by rejecting noise. The model parameters were set as $C = 2$ and $\theta = 2.0$.

Figure.5 compares the clustering results of the proposed method in $\beta = 0.001$ and $\beta = 0.002$, where clustering results are depicted in the 2-D feature space given by MDS. In $\beta = 0.001$, noise or outliers could not be regarded as noise, but in $\beta = 0.002$, rejecting the effect of noise was succeeded. $\beta = 0.002$ is much smaller than the average of relational elements d_{ij} and is expected to be useful for emphasizing noise in real world data. Bigger β resulted in regarding many points as noise.

5. Conclusions

This paper considered the applicability of the alternative distance measure in FCMdd-type linear clustering for relational data including outliers. In numerical experiments, conventional and proposed methods were compared using an artificially generated data set and the influence of β was verified using an artificial and a real one. The experimental results indicated that alternative distance measure worked well for both relational data sets by decreasing the effect of outliers. The proposed model using the robust measure in AFCM is useful for extracting intrinsic linear substructure from noisy relational data.

In the experiment with Morse code data, a smaller value rather than the sample variance recommended in [15] worked well. The robust model with linear prototype may be more sensitive to noise than the point-prototype model such as FCM. How to automatically tune the noise sensitivity weight β was remained for a potential future work.

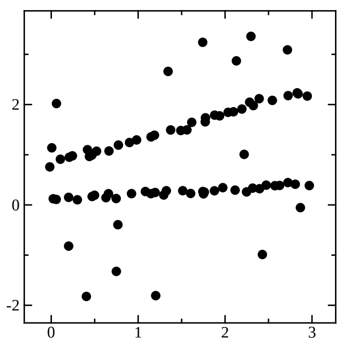
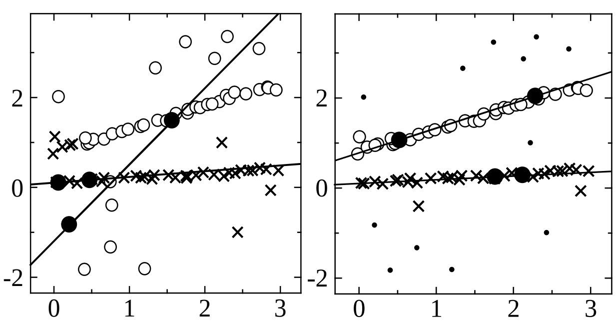
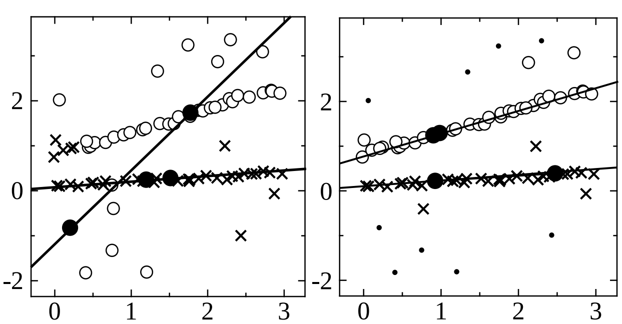


Fig. 1 2-D plots of artificial data set

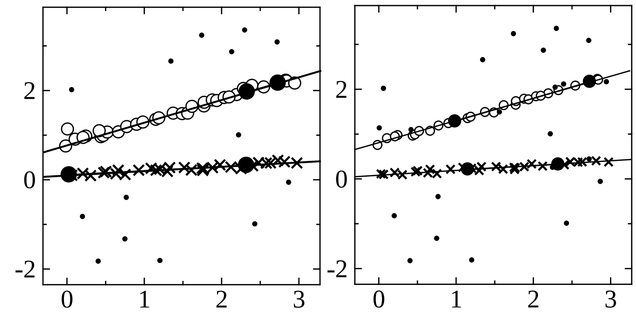


(a) Conventional method (b) Proposed method($\beta = 0.7$)

Fig. 2 Comparison of cluster partitions of artificial relational data



(a) $\beta = 0.05$ (b) $\beta = 0.3$



(c) $\beta = 2.0$ (d) $\beta = 50.0$

Fig. 3 Comparison of influence of β in artificial relational data

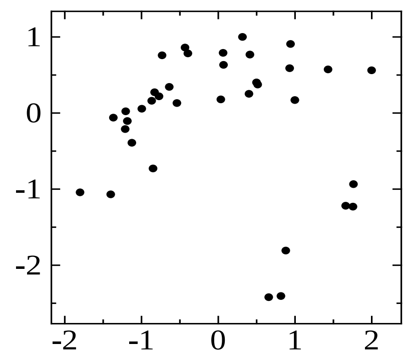
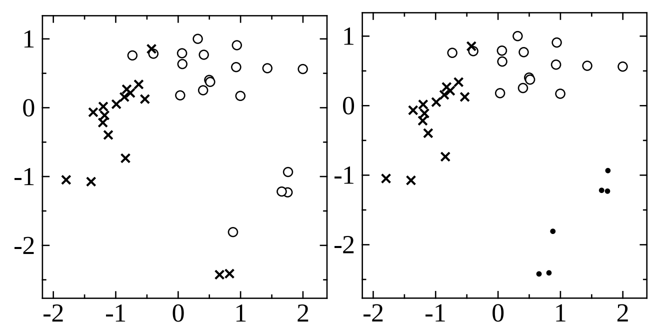


Fig. 4 2-D plots of Morse code by MDS



(a) $\beta = 0.001$ (b) $\beta = 0.002$

Fig. 5 Comparison of cluster partitions of Morse code data

Acknowledgments

This work was supported in part by the Ministry of Education, Culture, Sports, Science and Technology, Japan, through a Grant-in-Aid for Scientific Research (#23500283).

References

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

[2] J. B. MacQueen, “Some methods of classification and analysis of multivariate observations,” Proc. of 5th Berkeley Symposium on Math. Stat. and Prob., pp. 281-297, 1967.

[3] R. J. Hathaway, J. W. Davenport and J. C. Bezdek : Relational duals of the c-means clustering algorithms, PatternRecognition, vol. 22, no. 2 , pp. 205-212, 1989.

[4] R. Krishnapuram, A. Joshi, O. Nasraoui and L. Yi: “Low-complexity fuzzy relational clustering algorithms for web mining,” IEEE Trans. Fuzzy Systems, vol.9, no.4, pp.595-607, 2001.

[5] J. C. Bezdek, C. Coray, R. Gunderson and J. Watson, “Detection and characterization of cluster substructure I. Linear structure fuzzy c-lines,” SIAM Jour. of Appl. Math., vol.40, no.2, pp.339-357, 1981.

[6] J. C. Bezdek, C. Coray, R. Gunderson, and J. Watson, “Detection and characterization of cluster substructure 2. Fuzzy c-Varieties and convex combinations thereof,” SIAM J. Appl. Math., vol.40, no.2, pp.358-372, 1981.

[7] Y. Yabuuchi and J. Watada, Fuzzy principal component analysis and its application, Biomedical Fuzzy and Human Sciences, vol. 3, pp. 83-92, 1997.

[8] N. Kambhatla, and T. K. Leen, “Dimension reduction by local principal component analysis,” Neural Computation, vol. 9, no. 7, pp. 1493-1516, 1997.

[9] G. E. Hinton, P. Dayan, and M. Revow, “Modeling the manifolds of images of handwritten digits,” IEEE Trans. on Neural Networks, vol. 8 no. 1, pp. 65-74, 1997.

[10] K. Honda and H. Ichihashi, “Regularized linear fuzzy clustering and probabilistic PCA mixture models,” IEEE Trans. Fuzzy Systems, vol. 13, no. 4, pp. 508-516, 2005.

[11] N. Haga, K. Honda, A. Notsu and H. Ichihashi, “Local sub-space learning by extended fuzzy c-medoids clustering,” International Journal of Knowledge Engineering and Soft Data Paradigms, vol. 2, no. 2, pp. 169-181, 2010.

[12] W. S. Torgerson, Theory & Methods of Scaling, Wiley, 1958.

[13] E. Z. Rothkopf, “A measure of stimulus similarity and errors in some paired-associate learning tasks,” Jour. Experimental Psychology, vol. 53, pp.94-101, 1957.

[14] Wu. K-L, and Yang. M-S, “Alternative c-means clustering algorithms,” PatternRecognition, vol. 35, 2002, pp.2267-2278.

[15] L. Kaufman and P. J. Rousseeuw, Finding Groups In Data: An Introduction To Cluster Analysis, Wiley-Interscience, 1990.

[16] K. Honda, S. Nakao, A. Notsu and H. Ichihashi, “Alternative fuzzy c-lines and local principal component extraction,” International Journal of Knowledge Engineering and Soft Data Paradigms, vol. 3, no. 2, 2011 (in print).

Takeshi Yamamoto received the B.E. degree in industrial engineering from Osaka Prefecture University, Osaka, Japan in 2010.
He is currently the student of Graduate School of Engineering in the Department of Computer Sciences and Intelligent Systems. His research interests include relational clustering, fuzzy clustering and data mining.

Katsuhiro Honda received the B.E., M.E., and D.Eng. degrees in industrial engineering from Osaka Prefecture University, Osaka, Japan in 1997, 1999, and 2004, respectively.
From 1999 to 2009, he was a Research Associate and Assistant Professor at Osaka Prefecture University, where he is currently an Associate Professor in the Department of Computer Sciences and Intelligent Systems. His research interests include hybrid techniques of fuzzy clustering and multivariate analysis, data mining with fuzzy data analysis, and neural networks. He received a paper award and a young investigator award from the Japan Society for Fuzzy Theory and Intelligent Informatics (SOFT) in 2002 and 2005, respectively, and gave a tutorial on “Introduction to Clustering Techniques” at 2004 IEEE Int. Conf. Fuzzy Systems (FUZZ-IEEE 2004).

Akira Notsu received the B.E., M.I., and D. Informatics degrees from Kyoto University in 2000, 2002, and 2005, respectively.
He is currently an Assistant Professor, Department of Computer Sciences and Intelligent Systems, Osaka Prefecture University. His research interests include agent-based social simulation, communication networks, game theory, human-machine interface, and cognitive engineering.

Hidetomo Ichihashi received the B.E. and D.Eng. degrees in industrial engineering from Osaka Prefecture University, Osaka, Japan in 1971 and 1986, respectively.
From 1971 to 1981, he was with the Information System Center of Matsushita Electric Industrial Co., Ltd.. From 1981 to 1993, he was a Research Associate, Assistant Professor, and Associate Professor at Osaka Prefecture University, where he is currently a Professor in the Department of Computer Sciences and Intelligent Systems. His fields of interest are adaptive modeling of GMDH-type neural networks, fuzzy c -means clustering and classifier, data mining with fuzzy data analysis, human-machine interface, and cognitive engineering.