

# A Study on Regularization Effects of Fuzzified Memberships in FCM Clustering

Katsuhiro Honda, Yui Matsumoto, Akira Notsu, Hidetomo Ichihashi

Department of Computer Science and Intelligent Systems

Osaka Prefecture University

1-1 Gakuen-cho, Nakaku, Sakai, Osaka 599-8531, Japan

Email: {honda, notsu, ichi}@cs.osakafu-u.ac.jp

**Abstract**—FCM clustering is a fundamental technique for capturing intrinsic cluster structures of multivariate data sets. This paper presents a comparative study on the regularization effects of Fuzzy  $c$ -Means memberships estimated by two different fuzzification approaches: standard approach and entropy regularization approach. In this paper, the characteristics of the two fuzzification approaches are also discussed in noise fuzzy clustering (NFC) and it is revealed that the noise rejection mechanism of NFC can contribute to weakening the influence of initialization problems in entropy regularization approach although the approach is generally more sensitive to initial partition than the standard approach.

**Keywords**—fuzzy clustering, regularization, noise clustering.

## I. INTRODUCTION

Fuzzy  $c$ -Means (FCM) [1] is a basic method for estimating fuzzy  $c$ -partitions and is a fuzzy variant of  $k$ -Means clustering model [2], in which bivariate cluster indicators of  $k$ -Means are fuzzified so that they represent the relative typicalities of degree of belongingness to clusters. In order to estimate “fuzzy memberships” drawn from the interval  $[0, 1]$ , the linear programming problem of  $k$ -Means is transformed into a non-linear one by introducing an additional weighting exponent.

Entropy regularization [3] is another approach to fuzzification of  $k$ -Means cluster memberships, in which a non-linear penalty term (regularization term) is added to the  $k$ -Means objective function so that the full objective function is non-linear with respect to memberships. Using an entropy-like penalty term, the FCM model has close relation to Gaussian mixture models (GMMs) with Expectation-Maximization (EM) algorithm [4], in which the weight for the penalty term corresponds to the variance parameter in GMMs [5]. Besides the entropy term, other non-linear penalty terms such as a quadratic term [6] and a K-L information term [7] were also proved to be useful in fuzzification of memberships [8].

Although the FCM models with various fuzzification approaches derive similar clustering results in many cases, an important difference between the standard fuzzification approach and the regularization approach is seen in their classification functions [9]. The classification function of the standard approach has the maximum value of “1” on cluster centers and has the “fuzziest” value in the area far distant from all cluster

centers. On the other hand, the classification function of the entropy regularization approach does not have the maximum value on cluster centers and the value becomes large as a point goes away from all cluster centers.

In this paper, the regularization effects of the standard fuzzification and entropy regularization are compared. It has been known that regularization of objective functions contributes to avoiding singular solutions. In the FCM clustering, we can expect that the fuzzified objective functions improve the possibility of avoiding local minima. Additionally, in this paper, the regularization effects are also compared in a robust clustering context. Noise fuzzy clustering [10], [11] is a robust clustering model, in which noise samples that are far distant from all cluster centers are dumped into an additional “noise cluster” in order to robustly estimate cluster centers. If we consider the noise rejection situation, the entropy regularization approach may have a similar classification function to the standard approach.

The remainder of this paper is organized as follows: Section II briefly reviews fuzzification approaches in the FCM clustering and noise fuzzy clustering. Section III considers the regularization effects of fuzzified objective function through a toy example. Section IV presents several experimental results to compare the characteristic features of the fuzzification approaches. Section V summarizes the conclusions of this paper.

## II. MEMBERSHIP FUZZIFICATION IN FCM AND NOISE REJECTION

### A. Standard FCM

Let  $\mathbf{x}_i = (x_{i1}, \dots, x_{im})^\top$ ,  $i = 1, \dots, n$  denote  $m$  dimensional observations of  $n$  samples and the goal be to partition the sample into  $C$  clusters, in which samples belonging to same clusters are mutually similar.  $k$ -Means clustering [2] uses mean vector  $\mathbf{b}_c$  as the prototype of cluster  $c$  and estimates cluster indicators (membership of each sample) by minimizing the following within-group-sum-of-errors.

$$\begin{aligned} L_{km} &= \sum_{c=1}^C \sum_{i=1}^n u_{ci} d_{ci} \\ &= \sum_{c=1}^C \sum_{i=1}^n u_{ci} (\mathbf{x}_i - \mathbf{b}_c)^\top (\mathbf{x}_i - \mathbf{b}_c), \end{aligned} \quad (1)$$

This work was funded in part by the Ministry of Education, Culture, Sports, Science and Technology, Japan, under Grant-in-Aid for Scientific Research (20700214).

where  $u_{ci}$  represents the membership of data point  $\mathbf{x}_i$  to cluster  $c$  and is given based on the nearest prototype principle as

$$u_{ci} = \begin{cases} 1; & \mathbf{x}_i \text{ belongs to cluster } c \\ 0; & \text{otherwise} \end{cases}$$

and

$$\sum_{c=1}^C u_{ci} = 1, \quad i = 1, \dots, n. \quad (2)$$

The clustering algorithm is a 2-step iterative procedure of calculation of memberships  $u_{ci}$  and mean vectors  $\mathbf{b}_c$  [2].

In fuzzy clustering, memberships are generalized to  $u_{ci} \in [0, 1]$  so that membership matrix  $U = (u_{ci})$  represents a fuzzy partition composed of relative typicalities of belongingness to clusters. Equation (1) is, however, a linear function with respect to memberships and the optimal memberships are given in the extremal points ( $u_{ci} \in \{0, 1\}$ ). Then, Standard Fuzzy  $c$ -Means (FCM<sup>s</sup>) algorithm [1] introduced an additional weighting exponent into the linear objective function so that it becomes non-linear with respect to  $u_{ci}$ .

$$L_{fcm}^s = \sum_{c=1}^C \sum_{i=1}^n u_{ci}^\theta d_{ci}, \quad (3)$$

where  $\theta (\theta > 1)$  is the additional weighting exponent of “fuzzifier”. If  $\theta = 1$ , the clustering model is reduced to the (hard)  $k$ -Means model. The larger the  $\theta$ , the fuzzier the memberships.  $\theta$  is often set as  $\theta = 2$  or sometime  $\theta = 1.5$  in real applications. The clustering algorithm has similar formula to that of  $k$ -Means. New prototypes are derived as the weighted centers of clusters (mean vectors). Memberships are calculated under the constraint of Eq.(2) called “probabilistic constraint” [12]. The constraint forces memberships to have similar meaning to posterior probabilities in the EM algorithm [4] with probabilistic mixture models so that they represent the relative typicalities in clusters.

The FCM algorithm with standard fuzzification approach is summarized as follows:

**Algorithm: Standard Fuzzy  $c$ -Means (FCM<sup>s</sup>)**

Step 1 Initialize memberships  $u_{ci}, c = 1, \dots, C, i = 1, \dots, n$  randomly, and normalize the memberships so that they satisfy the probabilistic constraint. Choose termination threshold  $\varepsilon$ .

Step 2 Update  $\mathbf{b}_c, c = 1, \dots, C$  as

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

Step 3 Update  $u_{ci}, c = 1, \dots, C, i = 1, \dots, n$  as

$$u_{ci} = \begin{cases} \left[ \sum_{l=1}^C \left( \frac{d_{ci}}{d_{li}} \right)^{\frac{1}{\theta-1}} \right]^{-1} & ; I_i = \emptyset \\ 1/|I_i| & ; I_i \neq \emptyset, c \in I_i \\ 0 & ; I_i \neq \emptyset, c \notin I_i \end{cases} \quad (5)$$

where  $I_i$  is the set of indices such that  $d_{ci}$  is zero, i.e.,  $I_i = \{c : d_{ci} = 0\}$ .

Step 4 If

$$\max_{c,i} |u_{ci}^{NEW} - u_{ci}^{OLD}| < \varepsilon,$$

then stop. Otherwise, return to Step 2.

This standard FCM algorithm has the following characteristics [9]: If we use the membership calculation rules of Eq.(5) as fuzzy classification rules  $u_c$ , the maximum value ( $u_c = 1$ ) is given on the cluster center, and  $u_c$  moves toward  $1/C$  as  $\|\mathbf{x}\| \rightarrow \infty$ .

**B. Fuzzy Clustering with Regularized Objective Function**

If the goal is to fuzzify memberships in  $k$ -Means clustering, we can add a non-linear penalty term in order to make the objective function non-linear with respect to memberships. Miyamoto and Mukaidono [3] introduced a regularization term with a positive parameter  $\lambda$  into the objective function based on the regularization concept. Adding an entropy term, the objective function of FCM with regularization by entropy (FCM<sup>e</sup>) is defined as

$$L_{fcm}^e = \sum_{c=1}^C \sum_{i=1}^n u_{ci} d_{ci} + \lambda \sum_{c=1}^C \sum_{i=1}^n u_{ci} \log u_{ci}, \quad (6)$$

where the entropy term works like the weighting exponent in the standard FCM algorithm and the parameter  $\lambda$  plays a role for tuning the degree of fuzziness of membership values. The larger the  $\lambda$ , the fuzzier the memberships.

This objective function has similar formulation to the negative log-likelihood in the EM algorithm with GMMs, in which  $\lambda$  correspond to the variance parameter of Gaussian distributions [5].

The clustering algorithm is also composed of the 2-step procedure, in which the updating rules for memberships and cluster centers are derived as follows:

$$u_{ci} = \frac{\exp(-\frac{1}{\lambda} d_{ci})}{\sum_{l=1}^C \exp(-\frac{1}{\lambda} d_{li})}, \quad (7)$$

$$\mathbf{b}_c = \frac{\sum_{i=1}^n u_{ci} \mathbf{x}_i}{\sum_{i=1}^n u_{ci}}. \quad (8)$$

From the classification viewpoint, the fuzzy classification function have different features from that of FCM<sup>s</sup> [9].  $u_c$  does NOT have the maximum value ( $u_c = 1$ ) on the cluster center and  $u_c$  moves toward 1 as  $\|\mathbf{x}\| \rightarrow \infty$ . So, a data point that is far distant from all clusters may have a very large membership value.

The regularization approach can also be performed by using other regularization terms [6], [7], [13].

**C. Noise Fuzzy Clustering**

Noise fuzzy clustering [10], [11] introduced an additional noise cluster in order to damp noise samples into the noise cluster with the goal being to estimate robust cluster centers. In the noise clustering model, it is assumed that all samples have a constant (same) distance to the noise cluster, and that all samples, whose distance to all real cluster centers are greater

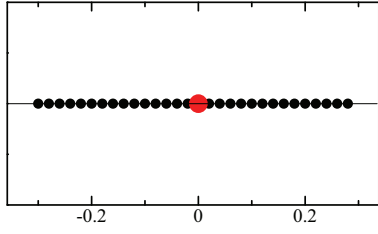


Fig. 1. an 1-D artificial data

than the constant distance, should be assigned to the noise cluster. In the FCM<sup>s</sup> formulation, the objective function is given as

$$L_{nfc}^s = \sum_{c=1}^C \sum_{i=1}^n u_{ci}^\theta d_{ci} + \sum_{i=1}^n \delta \left( 1 - \sum_{c=1}^C u_{ci} \right)^\theta, \quad (9)$$

where  $1 - \sum_{c=1}^C u_{ci}$  represent the membership degree to the noise cluster and  $\delta$  is the weight parameter that is identified with the distance to noise cluster. So, the conventional probabilistic constraint is redefined as

$$\sum_{c=1}^C u_{ci} \leq 1. \quad (10)$$

The membership updating rule of Eq.(5) (case of  $I_i = \emptyset$ ) is modified as

$$u_{ci} = \left[ \sum_{l=1}^C \left( \frac{d_{ci}}{d_{li}} \right)^{\frac{1}{\theta-1}} + \left( \frac{d_{ci}}{\delta} \right)^{\frac{1}{\theta-1}} \right]^{-1}. \quad (11)$$

In the same way, for the FCM<sup>e</sup> formulation, Eq.(7) is modified as

$$u_{ci} = \frac{\exp(-\frac{1}{\lambda} d_{ci})}{\sum_{l=1}^C \exp(-\frac{1}{\lambda} d_{li}) + \exp(-\frac{1}{\lambda} \delta)}, \quad (12)$$

### III. CONSIDERATION ON REGULARIZATION EFFECTS OF FUZZIFIED OBJECTIVE FUNCTIONS

In this section, a toy example is shown in order to demonstrate the regularization effects of fuzzified objective functions. Figure 1 shows an 1-D artificial data composed of 30 samples uniformly distributed in a 1-D space.

Assume that we partition the samples into 3 clusters and the third cluster center ( $b_3$ ) is fixed on the center (red circle). Figure 2 shows the 3D contour surface of the  $k$ -Means objective function Eq.(1) calculated by changing  $b_1$  and  $b_2$  in the interval  $[-0.35, 0.35]$ . Local minima are shown by red marks. The figure indicates that we have the optimal (global) solution of  $(b_1, b_2) = (0.21, -0.21)$  or  $(b_1, b_2) = (-0.21, 0.21)$  and that we also have several other local minima. So,  $k$ -Means is severely influenced by initial partitions and is often trapped in local minima.

Figure 3 shows the 3D contour surfaces of the FCM objective functions Eqs.(3) and (6). The fuzzification coefficients were set as  $\theta = 2$  and  $\lambda = 0.05$ , respectively. In the fuzzified objective functions, the contour surfaces were smoothened

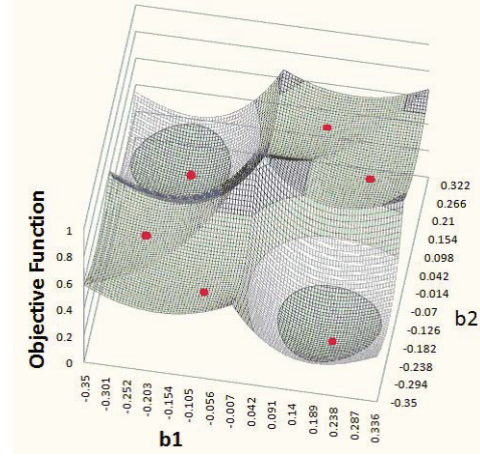
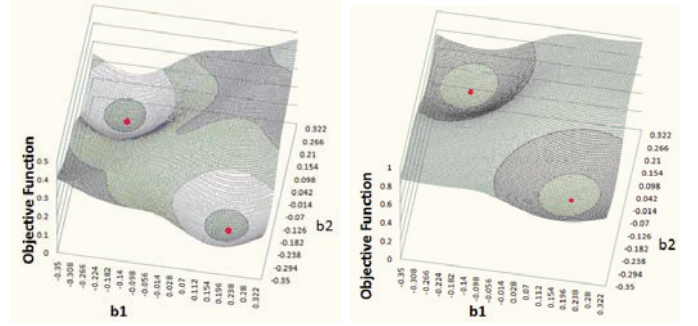


Fig. 2. Objective function of  $k$ -Means in toy example



(a) standard fuzzification (b) entropy regularization

Fig. 3. Objective function of FCM in toy example

and the possibility of being trapped into local minima was weakened. In this way, both fuzzification approaches have the regularization effects in  $k$ -Means clustering.

### IV. NUMERICAL EXPERIMENTS

In this section, several experimental results are shown in order to compare the regularization effects of fuzzification approaches.

#### A. Artificial data Set

The first experiment was performed with a 2-D ( $m = 2$ ) artificially generated data set shown in Fig. 4, in which 200 samples ( $n = 200$ ) were drawn from one of 4 Gaussian distributions (50 samples each) with unit variances whose mean vectors were  $(-1, -1)$ ,  $(-1, 1)$ ,  $(1, -1)$  and  $(1, 1)$ , respectively. In order to compare the possibilities of being trapped into local minima, the samples were partitioned into 4 clusters with various initial partitions.

Applying the  $k$ -Means algorithm, 2 different partitions shown in Fig. 5 were derived. Figure 5-(a) is the optimal  $k$ -Means partition having the smallest objective function value and was given in 62% trials. On the other hand, another local optimal solution shown in Fig. 5-(b) was also derive in the remaining trials. The result implies that the (hard)  $k$ -Means

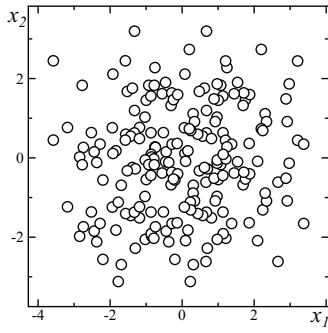


Fig. 4. 2-D artificial data set

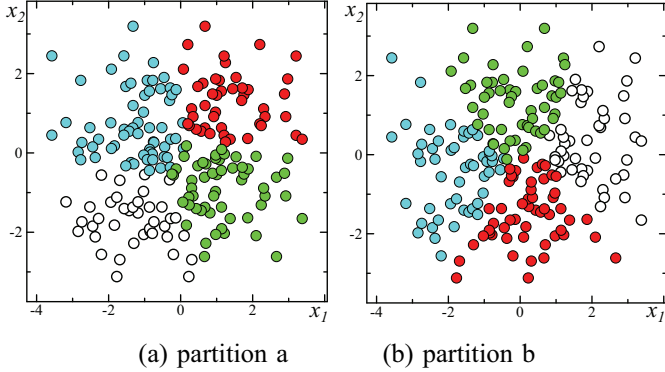


Fig. 5. Derived partitions with 2-D artificial data set

algorithm is very sensitive to initial partitions and is often trapped into local minima.

Next, fuzzy  $c$ -partitions were estimated by the standard fuzzification approach and the entropy regularization approach comparing the regularization effects. Tables I and tab: Comparison of local solution FCM-e compare the frequencies of the local solutions (LS(%)) in 1820 trials, in which the initial cluster centers were located on  $b_{cj} = \{x_{ij}^{\min}, \frac{1}{3}x_{ij}^{\min} + \frac{2}{3}x_{ij}^{\max}, \frac{2}{3}x_{ij}^{\max} + \frac{1}{3}x_{ij}^{\min}, x_{ij}^{\max}\}$  for each coordinate without duplication. The local solutions mean the case where the clustering algorithm failed to capture the  $c$ -partition of Fig. 5-(a). The tables imply that both fuzzification approach have regularization effects and contribute to avoiding being trapped into local minima as the weight parameters were increased.

Here, the fuzzy  $c$ -partitions derived by the two fuzzification approaches are compared using several clustering criteria. Partition coefficient (PC) and partition entropy (average entropy: AE) [1] are the average values of squared membership values  $u_{ci}$  and entropy of fuzzy memberships  $-u_{ki} \log u_{ki}$ . The criteria are used for evaluating the degree of fuzziness of fuzzy  $c$ -partition. A small PC and a large AE values imply a fuzzier partition with overlapping clusters. Xie-Beni index (XB) [14] is a cluster criterion considering the geometric features of FCM partitions:

$$XB = \frac{\sum_{c=1}^C \sum_{i=1}^n u_{ci} \|\mathbf{x}_i - \mathbf{b}_c\|^2}{n(\min_{c \neq j} \|\mathbf{b}_c - \mathbf{b}_j\|^2)}. \quad (13)$$

TABLE I  
COMPARISON OF FREQUENCIES OF LOCAL SOLUTIONS AND CLUSTER CRITERIA WITH 2-D ARTIFICIAL DATA SET (FCM<sup>s</sup>)

| $\theta$ | LS(%) | PC    | AE    | XB    |
|----------|-------|-------|-------|-------|
| 1.01     | 62.4  | 1.000 | 0.000 | 0.274 |
| 1.05     | 21.5  | 0.987 | 0.006 | 0.269 |
| 1.1      | 12.1  | 0.968 | 0.014 | 0.257 |
| 1.15     | 12.1  | 0.948 | 0.023 | 0.245 |
| 1.17     | 11.8  | 0.940 | 0.026 | 0.241 |
| 1.18     | 11.8  | 0.936 | 0.028 | 0.239 |
| 1.19     | 0.0   | 0.932 | 0.030 | 0.237 |

TABLE II  
COMPARISON OF FREQUENCIES OF LOCAL SOLUTIONS AND CLUSTER CRITERIA WITH 2-D ARTIFICIAL DATA SET (FCM<sup>e</sup>)

| $\lambda$ | LS(%) | PC    | AE    | XB    |
|-----------|-------|-------|-------|-------|
| 0.05      | 23.6  | 0.997 | 0.000 | 0.273 |
| 0.1       | 22.5  | 0.988 | 0.005 | 0.272 |
| 0.2       | 11.9  | 0.971 | 0.012 | 0.263 |
| 0.4       | 11.7  | 0.934 | 0.028 | 0.250 |
| 0.45      | 11.3  | 0.925 | 0.032 | 0.248 |
| 0.49      | 10.9  | 0.917 | 0.036 | 0.247 |
| 0.5       | 0.0   | 0.915 | 0.037 | 0.246 |

A small XB value implies compact-separated clusters.

Comparing the cluster criteria in the situation where the local solutions were disappeared, the entropy regularization approach needs a “fuzzier” memberships than the standard approach in order to avoid local solutions and the derived fuzzy  $c$ -partition is more compact and separated.

It may be because FCM<sup>e</sup> can assign larger memberships to samples that are far distant from all cluster centers as mentioned in Section II-2. If we gave a bad initial partition, some “noisy” samples having large memberships may illegally pull cluster centers.

### B. Iris Data Set

A similar experiment was performed with Iris data set that is downloadable from UCI Machine Learning Repository [15]. In this experiment, 150 instances ( $n = 150$ ) with 4 dimensional observation ( $m = 4$ ) drawn from 3 classes were partitioned into 3 clusters without using the class label.

Tables III and IV compare the clustering performance with various fuzzification parameters. In this experiment, the cluster partition corresponding to the 3 Iris classes were regarded as the optimal solution. The frequencies and the average values were calculated after 85320 trials with a similar initialization strategy to the previous experiment in 4-D data space. In the same way with the previous experiment, FCM<sup>s</sup> successfully derived a stable solution by using a suitable fuzzification weight. On the other hand, FCM<sup>e</sup> was still trapped into local minima using larger fuzzification weights. Especially, when  $\lambda \geq 1.0$ , Xie-Beni index significantly increased, i.e. some clusters might be severely overlapping.

In this sense, FCM<sup>s</sup> seems to work better than FCM<sup>e</sup> from the viewpoint of regularization of  $k$ -Means objective function.

TABLE III  
COMPARISON OF FREQUENCIES OF LOCAL SOLUTIONS AND CLUSTER  
CRITERIA WITH IRIS DATA SET (FCM<sup>s</sup>)

| $\theta$ | LS(%) | PC    | AE    | XB    |
|----------|-------|-------|-------|-------|
| 1.1      | 8.5   | 0.991 | 0.005 | 0.160 |
| 1.5      | 1.8   | 0.919 | 0.049 | 0.142 |
| 1.7      | 0.6   | 0.867 | 0.081 | 0.136 |
| 1.8      | 0.3   | 0.840 | 0.098 | 0.135 |
| 1.9      | 0.1   | 0.812 | 0.115 | 0.135 |
| 2.0      | 0.0   | 0.783 | 0.132 | 0.137 |

TABLE IV  
COMPARISON OF FREQUENCIES OF LOCAL SOLUTIONS AND CLUSTER  
CRITERIA WITH IRIS DATA SET (FCM<sup>e</sup>)

| $\lambda$ | LS(%) | PC    | AE    | XB    |
|-----------|-------|-------|-------|-------|
| 0.08      | 12.1  | 0.995 | 0.003 | 0.162 |
| 0.09      | 12.1  | 0.994 | 0.003 | 0.161 |
| 0.1       | 12.1  | 0.994 | 0.004 | 0.161 |
| 0.5       | 13.5  | 0.929 | 0.039 | 0.166 |
| 1.0       | 14.6  | 0.854 | 0.077 | 0.204 |
| 1.5       | 15.1  | 0.777 | 0.114 | 0.323 |

### C. Combination With Noise Fuzzy Clustering

Finally, the regularization effects of membership fuzzification are also compared in a robust clustering context using the 2-D data set shown in Fig. 4. Additional noise cluster adopted in noise fuzzy clustering [10] contributes to decreasing the membership degree of samples that are far distant from all cluster centers. We can expect that the robust feature may improve the effect of bad initial partitions in the entropy regularization approach by assigning smaller memberships to noisy samples. In noise fuzzy clustering, initial partitions are often given by FCM without noise rejection mechanism. In this experiment, however, the initial partition was given in the same initialization strategy with Section IV-A.

Table V compares the partition performance, in which the number of samples assigned to noise cluster “noise” was also shown with various noise weights  $\delta$ . The fuzzification weights were set as  $\theta = 1.19$  in order to consider the situation where the local solutions were first disappeared in Section IV-A. In this experiment, when  $\delta \leq 1.5$ , more than 30 out of 200 samples were regarded as noise and the plausible cluster centers could not be estimated. The table implies that the noise rejection mechanism increases the possibility of being trapped into local minima in FCM<sup>s</sup>. In this sense, the FCM-guided initialization in noise fuzzy clustering seems to be plausible.

Next, Table VI shows the results given by FCM<sup>e</sup>. The fuzzification parameter was set as  $\lambda = 0.49$  where local solutions were still remained in Section IV-A. The table indicates that the noise rejection mechanism with large  $\delta$  did not increase local solutions. Especially, when  $\delta = 3.0$ , the clustering performances were almost equivalent to those of  $\lambda = 0.5$  in Table II, i.e., the noise rejection mechanism contributed to avoiding being trapped into local minima. It may be because the noise cluster plays a role for decreasing the membership values of samples far distant from all clusters, which would illegally draw in some cluster centers.

TABLE V  
COMPARISON OF FREQUENCIES OF LOCAL SOLUTIONS AND CLUSTER  
CRITERIA WITH NOISE REJECTION (FCM<sup>s</sup>)

| $\delta$ | LS(%) | PC    | XB    | noise |
|----------|-------|-------|-------|-------|
| 1.0      | —     | 0.451 | 0.078 | 97    |
| 1.25     | —     | 0.570 | 0.132 | 72    |
| 1.5      | —     | 0.727 | 0.124 | 33    |
| 1.75     | 28.8  | 0.788 | 0.160 | 25    |
| 2.0      | 16.3  | 0.852 | 0.172 | 12    |
| 2.25     | 20.5  | 0.887 | 0.193 | 4     |
| 2.5      | 16.6  | 0.908 | 0.209 | 2     |
| 2.75     | 15.8  | 0.919 | 0.219 | 1     |
| 3.0      | 14.9  | 0.925 | 0.227 | 0     |

TABLE VI  
COMPARISON OF FREQUENCIES OF LOCAL SOLUTIONS AND CLUSTER  
CRITERIA WITH NOISE REJECTION (FCM<sup>e</sup>)

| $\delta$ | LS(%) | PC    | XB    | noise |
|----------|-------|-------|-------|-------|
| 1.0      | —     | 0.311 | 0.042 | 90    |
| 1.25     | —     | 0.472 | 0.070 | 66    |
| 1.5      | —     | 0.637 | 0.107 | 37    |
| 1.75     | 25.1  | 0.748 | 0.139 | 27    |
| 2.0      | 17.6  | 0.828 | 0.172 | 13    |
| 2.25     | 19.0  | 0.878 | 0.204 | 5     |
| 2.5      | 0.0   | 0.905 | 0.223 | 2     |
| 2.75     | 0.0   | 0.913 | 0.234 | 1     |
| 3.0      | 0.0   | 0.915 | 0.242 | 0     |

The result implies that the noise rejection mechanism may have another role for tackling initialization problems in the regularized objective functions of FCM.

## V. CONCLUSION

This paper discussed the effects of fuzzified memberships of  $k$ -Means clustering from the view point of the regularization of objective functions. Several experimental results revealed that the standard fuzzification approach is more useful for decreasing the effects of bad initial partitions than the entropy regularization approach, which can be identified with a sort of GMMs. In this paper, the regularization effects were also compared in the robust clustering context. An experimental result implied that the noise fuzzy clustering mechanism can also contribute to decreasing the effects of noisy samples in bad initial partitions by assigning smaller memberships in the entropy regularization approach.

Comparative study with other fuzzification approaches is remained in future work.

## 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] 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*, vol. 2, pp 86-92, 1997.
- [4] A. P. Dempster, N.M. Laird, and D.B. Rubin, “Maximum likelihood from incomplete data via the EM algorithm,” *J. of the Royal Statistical Society, Series B*, vol. 39, pp. 1–38, 1977.

- [5] K. Honda and H. Ichihashi, "Regularized linear fuzzy clustering and probabilistic PCA mixture models," *IEEE Trans. on Fuzzy Systems*, vol. 13, no. 4, pp. 508-516, 2005.
- [6] S. Miyamoto, S. Umayahara, "Fuzzy clustering by quadratic regularization," Proc. 1998 IEEE Int. Conf. Fuzzy Syst., pp.1394-1399, 1998.
- [7] H. Ichihashi, K. Miyagishi, K. Honda, "Fuzzy  $c$ -means clustering with regularization by K-L information," Proc. 2001 IEEE Int. Conf. Fuzzy Syst., pp.924-927, 2001.
- [8] S. Miyamoto, H. Ichihashi and K. Honda, *Algorithms for Fuzzy Clustering*, Springer, 2008.
- [9] Z.-Q. Liu, S. Miyamoto (eds.), *Soft Computing and Human-Centered Machines*, Springer-Verlag, 2000.
- [10] R. N. Davé, "Characterization and detection of noise in clustering," *Pattern Recognition Letters*, vol. 12, no. 11, pp. 657-664, 1991.
- [11] R. N. Davé and R. Krishnapuram, "Robust clustering methods: a unified view," *IEEE Trans. on Fuzzy Systems*, vol.5, pp.270-293, 1997.
- [12] F. Höppner, F. Klawonn, R. Kruse, and T. Runkler, *Fuzzy Cluster Analysis*, John Wiley & Sons, 1999.
- [13] K. Honda and H. Ichihashi, "A new approach to fuzzification of memberships in cluster analysis," *Modeling Decisions for Artificial Intelligence (MDAI) 2005*, Lecture Notes in Artificial Intelligence 3558, Springer, pp.172-182, 2005.
- [14] X. L. Xie and G. Beni, "A validity measure for fuzzy clustering," *IEEE Trans. on Pattern Anal. Machine Intell.*, vol. 13, no. 8, pp. 841-847, 1987.
- [15] A. Asuncion and D.J. Newman, UCI Machine Learning Repository [<http://www.ics.uci.edu/mlearn/MLRepository.html>], Irvine, CA: University of California, School of Information and Computer Science, (2007).