Fuzzy Technique-based Identification of Close and Distant Clusters in Clustering

Kyung Mi Lee and Keon Myung Lee*

Department of Computer Science, Chungbuk National University, PT-ERC
Cheongju, 361-763, Korea

Abstract

Due to advances in hardware performance, user-friendly interfaces are becoming one of the major concerns in information systems. Linguistic conversation is a very natural way of human communications. Fuzzy techniques have been employed to liaison the discrepancy between the qualitative linguistic terms and quantitative computerized data. This paper deals with linguistic queries using clustering results on data sets, which are intended to retrieve the close clusters or distant clusters from the clustering results. In order to support such queries, a fuzzy technique-based method is proposed. The method introduces distance membership functions, namely, close and distant membership functions which transform the metric distance between two objects into the degree of closeness or farness, respectively. In order to measure the degree of closeness or farness between two clusters, both cluster closeness measure and cluster farness measure which incorporate distance membership function and cluster memberships are considered. For the flexibility of clustering, fuzzy clusters are assumed to be formed. This allows us to linguistically query close or distant clusters by constructing fuzzy relation based on the measures.

Keywords : cluster identification, clustering, fuzzy clustering, fuzzy set theory, linguistic interface

1. Introduction

With the remarkable growth of hardware performance, user-friendly interfaces become one of major concerns in information systems. Linguistic conversation is a very natural way of human communications. Fuzzy techniques have been employed to liaison the discrepancy between the qualitative linguistic terms and quantitative computerized data. Many data analysis techniques have been developed to extract the embedded structures in data of interest. Meanwhile, it is becoming more and more important to provide analysis results in a convenient and understandable way.

Clustering is one of crucial data analysis techniques that groups a data set into possibly overlapped subsets such that subsets consist in similar entities and one subset are quite different from the others in their members' characteristics. Clustering has been widely exploited such various purposes as identifying natural structures embedded in data set, helping choose an interesting subset of data for further analysis.[1] It has been widely studied in various fields such as pattern recognition, statistical data analysis, data mining.[1,3,8] Clustering studies have been mainly focused on cluster formation, determination of the number of clusters, quality evaluation of clustering results.[3-8] Here we are concerned with how to provide clustering results to the analysts. To this end, various visualization techniques have been proposed such as 2D or 3D visualization based on PCA, MDS, and so on[1]. Here we propose an effective way to answer the questions like which clusters are close to or far from the given cluster. This question answering could take place in the situations as follows: In microarray data analysis, distant clusters with respect to a given interesting cluster could be a good target for further analysis in the pursuit of meaningful markers. In recommender services, closer groups for a group could be useful information for collaborative filtering.

In cluster analysis, the distance metrics between entities are supposed to be defined to measure their similarity. Such distance metrics usually gives unbound distance values. It is therefore not natural to assert that a pair of entities is close or far away with respect to specific threshold. In some situations, it would be more useful to give a degree ranged on the domain $[0,1]$ rather than to give a value on the unbounded range like metric distance, when comparing the clusters. Due to noise or inherent characteristics in data, clusters would be partially overlapped. Hence, it is desirable to take into account the extent of overlapping of clusters under consideration.

This paper proposes a method to provide for clusters the degree of closeness or farness which considers both distance and overlapping between clusters using fuzzy set techniques. The method introduces fuzzy membership functions called distance membership functions which transform metric distance between two objects into the degree of closeness or farness. In order to measure the degree of closeness or farness between two clusters, it defines both cluster closeness measure and cluster farness measure which incorporate distance membership function and cluster memberships. For the flexibility of clustering, fuzzy clusters are assumed to be

formed. It allows us to linguistically query close or distant clusters by constructing fuzzy relation based on the measures.

The remainder of this paper is organized as follows: Section 2 presents some works about distance measures, and Section 3 specifies in detail the close or distant cluster identification problems. Section 4 proposes the definitions of distance membership functions, cluster closeness measures and cluster farness measures, and the induced fuzzy relation for distance, and how to choose close or distant clusters. Section 5 shows an example about how to apply the proposed method and in final Section 6 draws the conclusions.

2. Related Works

In order to determine close or far clusters from a given cluster, there is the need to have some measures to give a degree value on [0,1] which tells how they are close or distant each other. To this notion, the distance and dissimilarity measures are related in some sense. Although distance and dissimilarity are interchangeably used in some situations, here we distinguish them each other. A distance measure gives a unbounded distance value, yet a dissimilarity measure tells a value bounded on [0,1]. The dissimilarity between clusters may consider either their degree of overlapping or their distance, or both. When all attributes of entities take values from continuous numeric domains, the distance between two entities could be evaluated by various measures such as Euclidian distance, Manhattan distance, Minkowski distance, Mahalanobis distance, and cosine distance as follows[4]: Here $\mathbf{d_i} = \left(d_{i1}, d_{i2}, \cdots, d_{ip}\right)$ denotes the i-th data with p attribute values, $\bar{\mathbf{d}}$ the mean value of the data set, and $\mathbf{S}$ the covariance of the data set.

· Euclidian distance

$$D_E(\mathbf{d_i}, \mathbf{d_j}) = \sqrt{\sum_k (d_{ik} - d_{jk})^2}$$

· Manhattan distance

$$D_{Mh}(\mathbf{d_i}, \mathbf{d_j}) = \sum_k |d_{ik} - d_{jk}|$$

· Minkowski distance

$$D_{Mk}(\mathbf{d_i}, \mathbf{d_j}) = \left(\sum_k |d_{ik} - d_{jk}|^p\right)^{1/p}$$

· Mahalanobis distance

$$D_{Mn}(\mathbf{d_i}, \mathbf{d_j}) = \sqrt{(\mathbf{d_i} - \bar{\mathbf{d}})\mathbf{S}^{-1}(\mathbf{d_i} - \bar{\mathbf{d}})}$$

· Cosine distance

$$D_{cos}(\mathbf{d_i}, \mathbf{d_j}) = \frac{\mathbf{d_i} \cdot \mathbf{d_j}}{|\mathbf{d_i}||\mathbf{d_j}|}$$

If attributes are categorical or discrete, specialized measures are employed like Jaccard's coefficient, Hamming distance.[9]

· Jaccard's coefficient

$$D_{Jaccard}(\mathbf{d_i}, \mathbf{d_j}) = \frac{M_{diff}}{M_{diff} + M_{same}}$$

· Hamming distance

$$D_{Hamming}(\mathbf{d_i}, \mathbf{d_j}) = \frac{\sum_k \delta(d_{ik}, d_{ij})}{p}$$

M_{diff} denotes the number of attributes with different values in $\mathbf{d_i}$ and $\mathbf{d_j}$, M_{same} that with the same values.

The distance measures between clusters have been studied in agglomerative hierarchical clustering algorithms. The algorithms take an approach to merge clusters based on the distance between them. The agglomerative hierarchical clustering algorithms are usually classified into average linkage clustering, complete linkage clustering, and single linkage clustering.[1] In average linkage algorithms, the distance between clusters is determined by calculating the average distance between all pairs of objects in the two different clusters. In complete linkage clustering, the distance between two clusters is defined as the greatest distance between a member of one cluster and a member of the other cluster. In single linkage clustering, the distance of two clusters is measured as the minimum distance between members of the two clusters. All these distance measures between clusters give an unbounded value.

In this study we are concerned with identifying close or far clusters for a given clusters. The distance measures giving unbounded value are not convenient for such tasks. In order to facilitate linguistic query for close or far clusters with respect to clustering results, we propose a fuzzy technique-based approach. It is assumed that clustering results are in a form of fuzzy clusters which allow clusters to overlap with each other in part. The approach is to adopt membership functions to map unbounded distance domain into the range [0,1], and to define the so-called closeness and farness measures which evaluate how much close or far away two cluster are respectively. Based on these measures, the intended clusters are identified.

3. The Close and Distant Cluster Identification Problems

For the convenience of description, the following notations are used:

$D = \left\{d_4, d_2, \cdots, d_n\right\}$: a data set of size n on which clusters is performed

$d_i = \left(d_{i1}, d_{i2}, \cdots, d_{ip}\right)$: the i th data with p attribute values

$C = \left\{C_1, C_2, \cdots, C_m\right\}$: a set of fuzzy clusters constructed over D

$C_i = \left\{(c_{i1}, \mu_{i1}), (c_{i2}, \mu_{i2}), \cdots, (c_{iL_i}, \mu_{iL_i})\right\}$: the i th cluster of C where μ_{ij} is the membership degree of data c_{ij} to cluster C_i

$O_i = \left(o_{i1}, o_{i2}, \cdots, o_{ip}\right)$: the center of cluster C_i

$\|d_i - d_j\|$: distance between data d_i ad d_j evaluated by a distance measure such as Euclidean distance, Manhattan distance

$LS = \{\alpha_i \mid i = 1,\ldots,q \quad 0 \le \alpha_1 < \alpha_2 < \cdots < \alpha_q \le 1\}$: a level set with which α-cut sets are defined

$IS_k(C_i) = \{c_{ij} \mid \alpha_k \le \mu_{ij} < \alpha_{K+1}\}$: the subset of C_i whose data membership degree is in the interval $[\alpha_k, \alpha_{k+1})$

θ_c : a threshold for cluster membership

γ_C : cluster radius which is the maximum distance from cluster center in order for an object to be considered as a member of the cluster

θ_d : distance threshold which is the minimum distance in order for an object to be regarded as a distant one from a cluster center

θ_k : distance threshold which is the maximum distance in order for an object to be regarded as a close one from a cluster center

C_i^{α} : the distinct cluster member set of cluster C_i which consists of object whose membership degree is not less than threshold θ_{α}

$$C_i^{\alpha} = \{c_{ir} \mid \mu_{ir} \ge \theta_{\alpha}\}$$

The close and distant cluster identification problems take place when there are queries to find close or distant clusters with respect to a specific cluster, or to find close or distant cluster pairs. In the setting, it is assumed that there is a fuzzy cluster C for a given data set D, already formed by some clustering method.

4. Close or Far Cluster Identification

In order to determine close or far clusters in a clustering result, we introduce two measures to evaluate how close a cluster is to a given cluster or how far it is from the cluster: cluster closeness measure and cluster farness measure. The measures are one to give for a pair of clusters a value on the range [0,1]. The higher value a cluster closeness measure gives, the closer the corresponding clusters are and vice versa. Similarly, the higher value a cluster farness measure gives for a pair of cluster, the farther they are. Compared to distance measures which could produce unbounded values, the ranges of the measures are bounded on [0,1]. The proposed method makes use of this property in deciding which clusters are closer or far away from a given cluster by comparing the measure values with a specified threshold.

4.1 Cluster Closeness Measure
The cluster closeness measures are intended to evaluate the degree to which two clusters are considered as being close together. In order to map the unbounded distance value into the range [0,1], we exploit a fuzzy membership function called a closeness membership function which is monotonously decreasing illustrated in Figure 1.
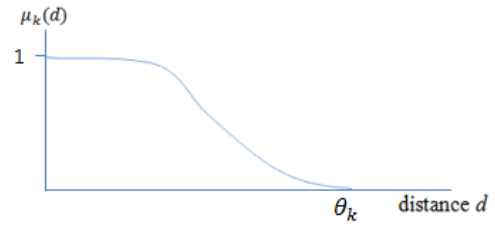

Fig. 1. A closeness membership function $\mu_k(d)$

The membership function $\mu_k(d)$ should preserve the following constraints:

$$0 \le \mu_k(d) \le 1$$
$$\mu_k(d) = 1 \text{ if } d < \theta_k$$
$$\mu_k(d_i) \ge \mu_k(d_j) \text{ if } d_i \le d_j$$

The constraint (a) is required for it to be a fuzzy membership function. The constraint (b) is required due to the definition of the threshold θ_k which is the maximum distance to be considered as somewhat close cluster. The constraint (c) enforces the monotonicity on the membership function.

The proposed measures assume that the clusters are formed using a fuzzy clustering algorithm and thus an object may have a membership degree to each cluster. Here we propose three cluster closeness measures $KC(C_i, C_j)$ as follows:

$$KC_1(C_i, C_j) = \frac{\sum\limits_{c_{ir} \in C_i, c_{js} \in C_j} \min\{\mu_{ir}, \mu_{js}\} \cdot \mu_k\left(\|c_{ir} - c_{js}\|\right)}{\sum\limits_{c_{ir} \in C_i, c_{js} \in C_j} \min\{\mu_{ir}, \mu_{js}\}}$$

$$KC_2(C_i, C_j) = \mu_k\left(\|O_i - O_j\|\right)$$

$$KC_3(C_i, C_j) = \frac{\sum\limits_{k=1}^{\alpha} \sum\limits_{c_{ir} \in IS_k(C_i), c_{js} \in IS_k(C_j)} \min\{\mu_{ir}, \mu_{js}\} \cdot \mu_k\left(\|c_{ir} - c_{js}\|\right)}{\sum\limits_{k=1}^{\alpha} \sum\limits_{c_{ir} \in IS_k(C_i), c_{js} \in IS_k(C_j)} \min\{\mu_{ir}, \mu_{js}\}}$$

In $KC_1(C_i, C_j)$, when an object c_{ir} has for a cluster C_i a membership degree not less than a specified threshold, i.e., $\mu_{ir} \ge \theta_C$, it is regarded as a member of the cluster, i.e., $c_{ir} \in C_i$. In order to define the degree of closeness, $KC_1(C_i, C_j)$ takes into account both the membership degrees of compared objects to their corresponding cluster and the membership degree of their distance to the close membership function. For all pairs of objects (c_{ir}, c_{js}) to be considered, $\min\{\mu_{ir}, \mu_{js}\}$ plays the role of a weighting factor.

$KC_2(C_i, C_j)$ just pays attention to the distance between the centers of clusters. It allows fast computation for the degree of cluster closeness, but could not reflect the distribution of objects in clusters.

$KC_3(C_i, C_j)$ is an approximation to $KC_1(C_i, C_j)$ developed to reduce the computation burden for $KC_1(C_i, C_j)$.

It first stratifies the objects of clusters into the subsets according to the membership degrees to cluster as follows:

$$IS_k(C_i) = \{c_{ij} \mid \alpha_k \le \mu_{ij} < \alpha_{K+1}\} \text{ over a level set}$$
$$LS = \{\alpha_i \mid i = 1, \dots, q \quad 0 \le \alpha_1 < \alpha_2 < \cdots < \alpha_q \le 1\}$$

Then the objects at the same level set are evaluated to determine the degree of cluster closeness. Due to this characteristics, $KC_3(C_i, C_j)$ could provide similar values to $KC_1(C_i, C_j)$ in a faster way.

4.2 Cluster Farness Measure

In order to measure cluster farness, the proposed method makes use of a fuzzy membership function so-called distant membership function $\mu_\delta(d)$ which maps distance to a degree on [0,1]. Figure 2 shows a distance membership function.
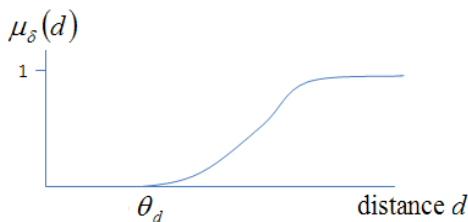

Fig. 2. A distant membership function $\mu_\delta(d)$

The membership function $\mu_\delta(d)$ has the following constraints:

$$0 \le \mu_\delta(d) \le 1$$
$$\mu_\delta(d) = 0 \text{ if } d < \theta_d$$
$$\theta_d > \gamma_C$$
$$\mu_\delta(d_i) \le \mu_\delta(d_j) \quad \text{if } d_i \le d_j$$

The constraint (a) is required for it to be a membership function. The constraint (b) is required due to the definition of the threshold θ_d which is the minimum distance to be considered as somewhat distant cluster. The constraint (c) is imposed to avoid the situations that an object belonging to a cluster is treated as a distant one. The constraint (d) enforces the monotonicity on the membership function.

Here by using a distant membership function we propose three cluster farness measures $KF(C_i, C_j)$ as follows:

$$KF_1(C_i, C_j) = \frac{\sum\limits_{c_{ir} \in C_i, c_{js} \in C_j} \min\{\mu_{ir}, \mu_{js}\} \cdot \mu_\delta(\|c_{ir} - c_{js}\|)}{\sum\limits_{c_{ir} \in C_i, c_{js} \in C_j} \min\{\mu_{ir}, \mu_{js}\}}$$

$$KF_2(C_i, C_j) = \mu_\delta(\|O_i - O_j\|)$$

$$KF_3(C_i, C_j) = \frac{\sum\limits_{k=1}^{\alpha} \sum\limits_{c_{ir} \in IS_k(C_i), c_{js} \in IS_k(C_j)} \min\{\mu_{ir}, \mu_{js}\} \cdot \mu_\delta(\|c_{ir} - c_{js}\|)}{\sum\limits_{k=1}^{\alpha} \sum\limits_{c_{ir} \in IS_k(C_i), c_{js} \in IS_k(C_j)} \min\{\mu_{ir}, \mu_{js}\}}$$

These measures are defined in the similar conceptualization

mentioned in the cluster farness measures. Hence, the formula has the same with cluster closeness measures except that the distant membership function is used in place of close membership function.

4.2 Characteristics of the Cluster Measures

The cluster closeness measures satisfy the following properties:

(a) $0 \le KC(C_i, C_j) \le 1$

(b) $KC(C_i, C_j) = KC(C_j, C_i)$

(c) $KC(C_i, C_i) = 1$ when the distance threshold θ_k is set to be large enough to cover the farthest pair of objects in the same cluster.

The property (a) is met for $KC_1(C_i, C_j)$ because it is a weighted sum of $\mu_k(\|c_{ir} - c_{js}\|)$, and $\mu_k(\|c_{ir} - c_{js}\|)$ is in the interval [0,1]. $KC_2(C_i, C_j)$ holds the property (a) due to the property of close membership function. $KC_3(C_i, C_j)$ satisfies the property (a) in the same context of $KC_1(C_i, C_j)$. The reflexivity property of (b) holds because the minimum operation and the distance operation are reflexive. The property (c) is satisfied because an object is regarded as a (fuzzy) member of cluster C_i only when its membership degree not less than a specified threshold, i.e., $\mu_{ir} \ge \theta_C$, and the threshold θ_k is greater than two times of the cluster radius γ_C which is set to cover the cluster members.

The cluster farness measures also satisfy the following properties :

(a) $0 \le KF(C_i, C_j) \le 1$

(b) $KF(C_i, C_j) = KF(C_j, C_i)$

(c) $KF(C_i, C_i) = 0$ when the distance threshold θ_d is set to be large enough to cover the farthest pair of objects in the same cluster.

The property (a) is met for $KF_1(C_i, C_j)$ because it is a weighted sum of $\mu_\delta(\|c_{ir} - c_{js}\|)$, and $\mu_\delta(\|c_{ir} - c_{js}\|)$ is in the interval [0,1]. $KF_2(C_i, C_j)$ holds the property (a) due to the property of distant membership function. $KF_3(C_i, C_j)$ satisfies the property (a) in the same context of $KF_1(C_i, C_j)$. The reflexivity property of (b) holds because the minimum operation and the distance operation are reflexive. The property (c) is satisfied because an object is regarded as a (fuzzy) member of cluster C_i only when its membership degree not less than a specified threshold, i.e., $\mu_{ir} \ge \theta_C$, and the threshold θ_d is greater than two times of the cluster radius γ_C which is set to cover the cluster members.

4.3 Fuzzy Relation for Identifying Close or Distant Clusters

The goal of this work is to enable to answer the linguistic questions about which clusters are close or distant. To this end, we build fuzzy relation which tells the degree of closeness or

farness between cluster by using cluster closeness or farness measures.

For the close cluster identification, the fuzzy close relation R_C is defined as follows using the cluster closeness measure.

$$R_C = \left\{ \left((C_i, C_j), \mu_{ij} \right) \mid C_i \in C, C_j \in C, \mu_{ij} = K(C_i, C_j) \right\}$$

The fuzzy close relation satisfies the following properties:

R_C is symmetric.

R_C is reflexive.

The properties (a) and (b) follow from the properties of cluster closeness measures shown in Section 4.2. Once

For the distant cluster identification, the fuzzy distant relation R_F is defined as follows:

$$R_F = \left\{ \left((C_i, C_j), \mu_{ij} \right) \mid C_i \in C, C_j \in C, \mu_{ij} = K(C_i, C_j) \right\}$$

The relation R_F has the following properties:

R_F is symmetric.

R_F is irreflexive.

The properties (a) and (b) follow from the properties of cluster farness measures shown in Section 4.2.

Once a fuzzy close relation R_C is constructed, it is possible to identify distant clusters by taking α-cut operation for the relation. The higher α-cut operations are applied, the more distant clusters are determined. In a similar way, when we get a fuzzy distant relation, we can determine distant clusters according to the α-value.

5. An Application Example

In order to show the applicability of the proposed method, a data set of Figure 3 was generated using the Gaussian function centered at (0.5,0.5,-0.5), (1.5, -2.0, -1.5), (0.5, -1.0, -1.5), (-1,1,-1.5), (1.5, -0.5, 1.5), for each center 10 data was generated with standard deviation 0.3. For the data set, the fuzzy k-means clustering algorithm was applied to get fuzzy clusters. The colors in the figure indicate the cluster labels for which membership degree is the highest.
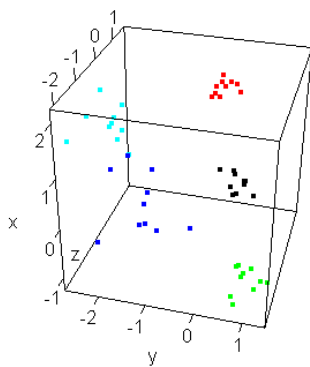

Fig. 3. Data used in the experiment

Table 1 shows the fuzzy close relation constructed for the data by the KC_1 measure. We can see that the relation is

symmetric and reflexive. From the relation R_C, we can get the close cluster pairs by taking α-cut operations. For example, if α is set to 0.5, the following pairs are chosen as the close clusters: $\{(C_2, C_4), (C_3, C_4), (C_3, C_5)\}$ When α is set to a higher value like 0.7, the more close cluster pairs are identified as follows: $\{(C_2, C_4), (C_3, C_5)\}$

Table 1. The fuzzy close relation R_C for data of Fig. 3

	C_1	C_2	C_3	C_4	C_5
C_1	1.0	0.01	0.01	0.35	0.01
C_2	0.01	1.0	0.29	0.72	0.02
C_3	0.01	0.29	1.0	0.54	0.76
C_4	0.34	0.72	0.54	1.00	0.20
C_5	0.01	0.02	0.77	0.20	1.0

Table 2 shows the fuzzy far relation constructed for the data by the KF_1 measure which is symmetric and irreflexive. From the relation, we can get the distant cluster pairs by taking α-cut operations. For example, if α is set to 0.7, the following pairs are chosen as the distant clusters: $\{(C_1, C_2), (C_1, C_3), (C_1, C_4), (C_1, C_5), (C_2, C_5), (C_4, C_5)\}$ When α is set to a higher value like 0.9, the more distant cluster pairs are identified as follows: $\{(C_1, C_2), (C_1, C_3), (C_1, C_5), (C_2, C_5)\}$

Table 2. The fuzzy distant relation R_F for data of Fig. 3

	C_1	C_2	C_3	C_4	C_5
C_1	0.0	1.0	0.99	0.82	0.99
C_2	1.00	0.0	0.64	0.24	0.98
C_3	0.97	0.64	0.0	0.23	0.17
C_4	0.82	0.24	0.23	0.00	0.79
C_5	0.99	0.98	0.17	0.79	0.0

6. Conclusion

It is becoming more and more important to present the information to users in intuitive and understandable way. We proposed a method to support linguistic queries for clustering results such as which ones are close or which ones are far away each other. The method adopted fuzzy membership functions which transform metric distance between two objects into the degree of closeness or farness. In order to measure the degree of closeness or farness between two clusters, it defined both cluster closeness measures and cluster farness measures which incorporate distance membership function and cluster memberships. Some interesting properties for the measures have been investigated. In addition, it showed that the measures could be used to define a fuzzy relation to indicate the degree of farness between clusters, and that the pairs of distant clusters could be found out from such a fuzzy relation. It allows us to linguistically query close or distant clusters by constructing

fuzzy relation based on the measures. We may find in phylogenic analysis of biology some practical applications of the proposed method. From the phylogenic analysis based on genetic sequence comparison, the distances among the entities like viruses are estimated and clustered, and sometimes formed in phylogenic tree. When the proposed method is applied, the analyst could simply pose questions like which viruses groups are closely related or evolutionally distant away.

References

[1] B. Clarke, E. Fokoue, H. H. Zhang, *Principles and Theory for Data Mining and Machine Learning*, Springer, 2009.

[2] H.-J. Zimmermann, *Fuzzy Set Theory and Its Application*, Kluwer-Nijhoff Publishing, 1985.

[3] A. K. Jain, M.N. Murty, P.J. Flynn, "Data clustering: a review, *ACM Computing Survey*," vol. 31, no. 3, 1999.

[4] R. O. Duda, P. E. Hart, D. G. Stork, *Pattern Classification*, Wiley, 2001.

[5] M. K. Pakhira, S. Bandyopadhyay, U. Maulik, "Validity index for crisp and fuzzy clusters," *Pattern Recognition*, vol. 37, no. 3, pp. 487-501, 2004.

[6] Y. Zhang, W. Wang, X. Zhang, Y. Li, "A cluster validity index for fuzzy clustering," *Information Sciences*, vol. 178, no. 4, pp. 1205-1218, 2008.

[7] M.S. Yang, A survey of fuzzy clustering, *Math. Comput. Model.* vol. 18, pp. 1–16, 1993.

[8] P Berkhin, "A survey of clustering data mining techniques," *Grouping Multidimensional Data*, Springer, 2006.

[9] G. J. Torres, R. B. Basnet, A. H. Sung, S. Mukkamala, and B. M. Ribeiro, "A Similarity Measure for Clustering and its Applications," *Int. J. of Electrical, Computer, and Systems Engineering*, vol. 3, no. 3, 2009.

[10] K. M. Lee, Sun A Lee, "Fuzzy Set-based Distant Cluster Identification," *Proc. of the 14th Int. Conf. on Computers* (ICCOMP'10), vol. II, pp. 506-509.

[11] J. Yen, R. Langari*, Fuzzy Logic: Intelligence, Control, And Information*, Prentice-Hall, 1999.

[12] M. R. Rezaee, B.P.F. Lelieveldt, J.H.C. Reiber, "A new cluster validity index for the fuzzy c-means," *Pattern Recognition Letters*, vo.31, no. 3-4, pp. 237-246, 1998.

Kyung Mi Lee

PhD course student, Department of Computer Science of Chungbuk National University

Research Area: soft computing, artificial intelligence application

E-mail : kmlee07@cbnu.ac.kr

Keon Myung Lee

Professor, Department of Computer Science of Chungbuk National University

Research Area : machine learning, data mining, bioinformatics, cloud computing

E-mail: kmlee@cbnu.ac.kr