

# Initial Training Data Selection for Active Learning

Weiwei Yuan  
Department of Computer Science  
Harbin Engineering University  
Harbin, China  
weiwei@oslab.khu.ac.kr

Yongkoo Han  
Department of Computer  
Engineering  
Kyung Hee University  
Yongin-Si, Korea  
ykhan@khu.ac.kr

Donghai Guan  
College of Automation  
Harbin Engineering University  
Harbin, China  
donghai@oslab.khu.ac.kr

Sungyoung Lee  
Department of Computer Engineering  
Kyung Hee University  
Yongin-Si, Korea  
sylee@oslab.khu.ac.kr

Young-Koo Lee  
Department of Computer Engineering  
Kyung Hee University  
Yongin-Si, Korea  
yklee@khu.ac.kr

## ABSTRACT

The crucial issue in many classification applications is how to achieve the best possible classifier with a limited number of labeled training data. Active learning is one method which addresses this issue by selecting the most informative data for training. In this work, we argue that the performance of active learning could be improved through carefully selecting the initial training samples. To confirm our argument, we propose three initial training data selection mechanisms based on fuzzy clustering method: center-based selection, border-based selection and hybrid selection. Center-based selection selects the samples with high degree of membership in each cluster as initial training data. Border-based selection selects the samples around the border between clusters. Hybrid selection is the combination of center-based selection and border-based selection. The effects of them are empirically studied on a set of UCI data sets. Experimental result indicates that, compared with randomly selecting initial training samples, hybrid selection can effectively enhance the performance of active learning.

## Categories and Subject Descriptors

I.5.2 [Pattern Recognition]: Design Methodology

## General Terms

Algorithms

## Keywords

Active Learning, Classification, Feature Selection, Labeled Training data

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ICUIMC '11, February 21-23, 2011, Seoul, Korea.  
Copyright 2011 ACM 978-1-4503-0571-6...\$10.00.

## 1. INTRODUCTION

Supervised learning is one primary sub-field of classical machine learning. In supervised learning, we need a collection of labeled data for training purpose; the problem is to label a newly encountered, yet unlabeled, sample [1].

In many applications, labeling the training examples is difficult, expensive, or time consuming [2]. For instance, if one is building a speech recognizer, it is easy enough to get raw speech samples, but labeling even one of these samples is a tedious process in which a human must examine the speech signal and carefully segment it into phonemes. Another example is Web page classification in which unlabeled samples are readily available, but labeled ones are fairly expensive to obtain. In these applications, the crucial issue is how to achieve the best possible classifier with a small number of labeled data.

An important topic addressing above issue is selecting the valuable data to label, considering that labeling data is a costly job. This topic is known as active learning [3][4]. In active learning, the learning process iteratively queries unlabeled samples to select the most valuable samples to annotate and update its learned models. Therefore, the unnecessary and redundant annotation is avoided.

Most existing works on active learning focus on designing the refined data selection mechanisms to improve the learning performance. Two major data selection mechanisms are: certainty-based selection and committee-based selection. Different with them, in this work, we propose another way to improve the performance of active learning. We argue that initial training data can influence the following data selection result. If initial training data can reflect the distribution of data to be classified more accurately, then more valuable data could be selected.

We propose three methods to select initial training samples. All of them are based on fuzzy clustering method. In this work, fuzzy c-means is used for data clustering. Our methods first partition all the given unlabeled samples into clusters and then selects the most representative ones from each cluster to label. These three methods are center-based selection (CS), border-based selection

(BS) and hybrid selection (HS). In CS, the data with high degree of membership in each cluster will be selected. Center-based selection is named because usually these data samples are close to the cluster centers. BS will select training samples around the borders between clusters and HS is a hybrid selection method through combining CS and BS.

Experimental results show that through using the initial training samples got from hybrid selection, the performance of active learning is effectively improved as compared to randomly selecting initial training samples.

The rest of this paper is organized as follows: In Section 2, related work in active learning is presented. Section 3 presents the basic knowledge of fuzzy c-means. Then three data selection mechanisms (center-based selection, border-based selection and hybrid selection) are presented. Section 4 reports on the empirical study and discusses some observations. Section 5 discloses conclusions and future work.

## 2. Related Work

In typical active learning, for a data set  $D = \{x_1, \dots, x_n\} \subset R^d$ , firstly, a small number of labeled samples will be randomly selected. Let  $D_l$  denotes this labeled set and  $D_u = D \setminus D_l$ . The active learning system comprises two parts: a learning engine and a selection engine. At each iteration the learning engine uses a supervised learning algorithm to train a classifier on  $D_l$ . The selection engine then selects a sample from  $D_u$  and requests a human expert to label the sample before passing it to the learning engine. The major goal is to achieve a good classifier as best as possible within a reasonable number of calls for labeling by human help. Table 1 gives the general active learning process.

**Table 1. Active learning process**

1.	Randomly select a small number of samples from unlabeled sample pool, assign a class label to each of them, and add them into the training set						
2.	Train the classifier						
3.	Perform the following loop until the stopping criteria are satisfied <ol style="list-style-type: none"> <tr> <td>a.</td> <td>Query a set of samples from unlabeled sample pool and assign a class label to each of them</td> </tr> <tr> <td>b.</td> <td>Add the new queried and annotated samples into training set</td> </tr> <tr> <td>c.</td> <td>Retrain the classifier</td> </tr> </ol>	a.	Query a set of samples from unlabeled sample pool and assign a class label to each of them	b.	Add the new queried and annotated samples into training set	c.	Retrain the classifier
a.	Query a set of samples from unlabeled sample pool and assign a class label to each of them						
b.	Add the new queried and annotated samples into training set						
c.	Retrain the classifier						

In Table 1, step 3-a, samples selection is central to active learning. And if data selection process is refined, then the performance of active learning could be improved.

Most existing work in active learning has concentrated on two data selection strategies: certainty-based [5][6][7][8] and committee-based selection [9][10][11][12]. In the certainty-based strategy, an initial system is trained using  $D_l$ . Then the system labels the samples in  $D_u$ , and determines the certainties of its

predictions of them. The example with the lowest certainty is then selected and presented to the experts for annotation. In the committee-based methods, a distinct set of classifiers is created using  $D_l$ . The sample in  $D_u$ , whose label differs most when presented to different classifiers are presented to the experts for annotation. In both paradigms, a new system is trained using the new set of annotated examples, and this process is repeated until it reaches the predefined rounds or some stopping criteria are satisfied.

We agree that step 3-a is important for active learning. However, we argue that step 1, initial training data selection, is also important. Hence, it is interesting to see whether the performance of active learning could be improved by refining the selection of initial training data.

In this work, we propose three methods to select initial training samples. They are center-based selection, border-based selection and hybrid selection. These methods are based on fuzzy clustering method. Our proposed active learning process is given in Table 2.

**Table 2. Our proposed active learning process**

1.	Select a small number of samples from unlabeled sample pool based on fuzzy clustering, assign a class label to each of them, and add them into the training set						
2.	Train the classifier						
3.	Perform the following loop until the stopping criteria are satisfied <ol style="list-style-type: none"> <tr> <td>a.</td> <td>Query a set of samples from unlabeled sample pool and assign a class label to each of them</td> </tr> <tr> <td>b.</td> <td>Add the new queried and annotated samples into training set</td> </tr> <tr> <td>c.</td> <td>Retrain the classifier</td> </tr> </ol>	a.	Query a set of samples from unlabeled sample pool and assign a class label to each of them	b.	Add the new queried and annotated samples into training set	c.	Retrain the classifier
a.	Query a set of samples from unlabeled sample pool and assign a class label to each of them						
b.	Add the new queried and annotated samples into training set						
c.	Retrain the classifier						

## 3. Initial Training Data Selection

### 3.1 Fuzzy C-mean

Fuzzy C-means clustering (FCM) [13] is a popular data clustering algorithm and it combines K-means clustering with fuzzy logic. As with fuzzy sets [14], using FCM, each data point can be a member of more than one cluster with different degrees of membership function between 0 and 1. FCM is an objective function based clustering method. Here objective function measures the overall dissimilarity within clusters. By minimizing the objective function we can obtain the optimal partition. Let  $X = \{x_1, x_2, \dots, x_n\}$  denote the measured data set. The FCM objective function  $J$  is defined as:

$$J = \sum_{i=1}^n \sum_{j=1}^c (u_{ij})^m \|x_i - v_j\|^2 \quad (1)$$

Clustering of FCM is carried out through an iterative minimization of  $J$  according to the following steps. Here  $u_{ij}$  is the degree of membership of  $x_i$  in cluster  $j$  and  $m$  is the fuzzy factor that determines the degree of fuzziness ( $m > 1$ ). As  $m$  approaches one, fuzziness degrades and the FCM algorithm

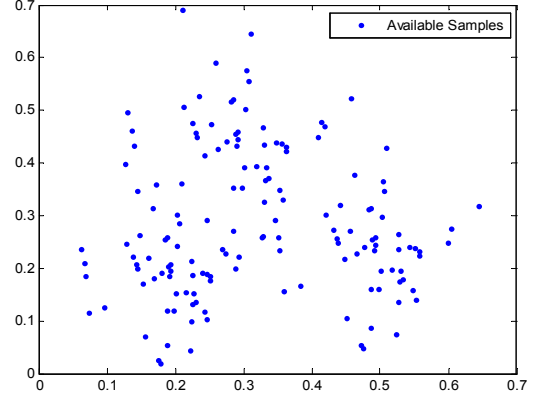
approaches to the standard K-means algorithm.  $V = \{v_1, v_2, \dots, v_c\}$  is the vector of cluster centers.  $\|x_i - v_j\|^2$  is any norm expressing the similarity between the measured data  $x_i$  and the center  $v_j$ .

**Table 3. Clustering of FCM**

<p><b>S1:</b> Choose fuzzy factor (<math>m</math>), number of clusters (<math>c</math>) and <math>c</math> initial cluster centers <math>v_j</math>.</p> <p><b>REPEAT</b></p> <p><b>S2:</b> At iteration <math>t</math>, compute <math>u_{ij}</math> with <math>v_j</math> by</p> $u_{ij} = \left( \frac{\ x_i - v_j\ ^{2/(m-1)}}{\sum_{k=1}^c \ x_i - v_k\ ^{2/(m-1)}} \right)^{-1} \quad (2)$ <p><b>S3:</b> Update <math>v_j</math> with <math>u_{ij}</math>, by</p> $v_j = \frac{\sum_{i=1}^n u_{ij}^m \cdot x_i}{\sum_{i=1}^n u_{ij}^m} \quad (3)$ <p><b>UNTIL</b> (<math>\ V_t - V_{t-1}\  \leq \varepsilon</math>, <math>V_t</math> and <math>V_{t-1}</math> denote the vector of clusters centers at iteration <math>t</math> and <math>t-1</math> respectively, <math>\varepsilon</math> is convergence criterion with <math>0 &lt; \varepsilon &lt; 1</math>)</p>
--

### 3.2 Our Proposed Methods

Fuzzy c-means computes the cluster centers and generates the class membership matrix. In this work, center-based selection, border-based selection and hybrid selection mechanisms are implemented through analyzing membership matrix. To visually see the difference between them, an artificial data set is used as shown in Fig.1. This data set includes 150 2-D samples and 3 classes. 21 samples will be selected as initial training data.



**Figure 1. Artificial data set**

Our proposed methods are:

**1. Center-based selection.** This selection method selects the samples with high degree of membership in each cluster. We extract these samples through analyzing membership matrix  $U_{(n \times m)}$ . Here  $n$  is the number of samples partitioned and  $m$  is the number of clusters.  $u_{ij}$  is the element at the  $i^{th}$  row and  $j^{th}$  column in  $U$ , which denotes the degree of membership of sample  $x_i$  in cluster  $j$ . In each cluster  $j$  ( $j = 1:m$ ), if  $i^* = \arg \max_{i=1:n} u_{ij}$ , then sample  $x_{i^*}$  is regarded as the most representative sample in this cluster to be selected. The next selected sample is  $x_{i^{**}}$  with  $i^{**} = \arg \max_{i=1:n, i \neq i^*} u_{ij}$ . In turn, other samples in cluster  $j$  will be selected, until the number of data equals to  $k_j$  (the number of training data allocated to cluster  $j$ ). Usually in active learning, we are given the total number of training data  $K$  ( $K = \sum_{j=1}^m k_j$ ) instead of  $k_j$ , so how to determine  $k_j$  with the knowledge of  $K$  is another issue in center-based selection. A simple way we used is to select same or similar number of samples from each cluster ( $k_j \cong \frac{K}{m}$ ). Based on C-S, if we select 21 training samples from above artificial data set (7 samples each cluster), the result of selection is shown in Fig. 3.

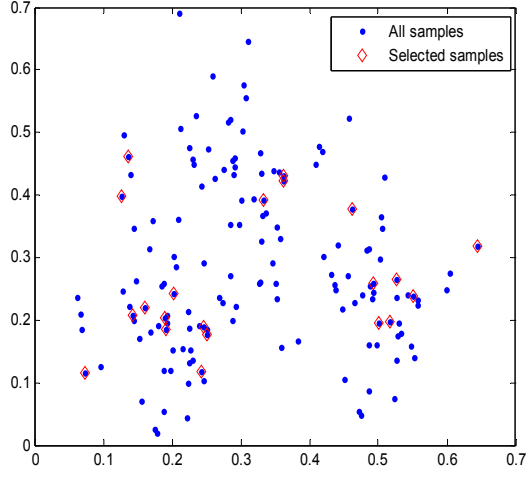


Figure 2. Training samples from random selection

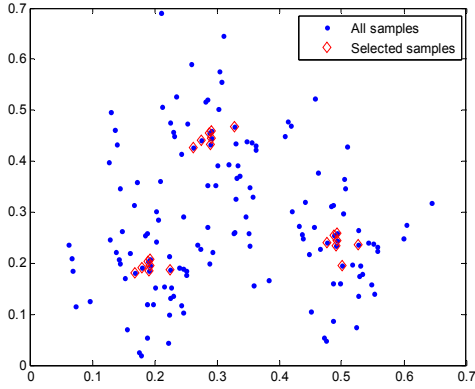


Figure 3. Training samples from center-based selection

**2. Border-based selection.** This selection strategy selects the samples around borders of clusters. Here we say a sample is located at the border between clusters when its two high degrees of membership are very similar. For example, a data set comprises three clusters. For a sample of it, when its degrees of membership in each cluster is  $[0.5, 0.49, 0.01]$ , its two high membership degree (0.5 & 0.49) are very similar. In this case, we can say that this sample is located at the border between cluster 1 and 2. Membership matrix  $U_{(n \times m)}$  is also used in this part. Here  $n$  is the number of samples partitioned and  $m$  is the number of clusters. For each sample  $x_i$  ( $i = 1:n$ ), if  $j^* = \arg \max_{j=1:m} u_{ij}$  and  $j^{**} = \arg \max_{j=1:m, j \neq j^*} u_{ij}$ , then  $T_i$  ( $T_i = u_{ij^*} - u_{ij^{**}}$ ) is calculated. Finally sample  $x_{i^*}$  with  $i^* = \arg \min_{i=1:n} T_i$  is regarded as the most

representative sample. In turn  $x_{i^{**}}$  with  $i^{**} = \arg \min_{i=1:n, i \neq i^*} T_i$  is the next valuable data to be selected. Other samples will be selected using this way until the number of selected reaches the limitation. The result of border-based selection on the artificial data set is given in Fig. 4.

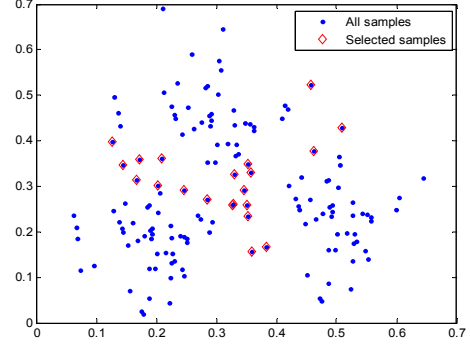


Figure 4. Initial training samples from border-based selection

**3. Hybrid selection.** This strategy is a hybrid selection method combining above two methods. It assumes that both the samples from CS and BS are representative. Combining them might provide better result than either alone. For a data pool  $D$ , let  $K$  denote the number of data to be selected as initial training data. A simple combination scheme in this work is to select about  $K/2$  samples from center-based selection and border-based selection respectively. Of course, it is not required to exactly follow this combination scheme in the real applications. For example, if it is obvious that samples got from center-based selection are redundant, then more samples could be extracted from border-based selection. On above artificial data set, if 9 samples are selected by CS (3 sample each cluster) and 12 samples are selected by BS. Then 21 training samples determined by HS are shown in Fig. 5.

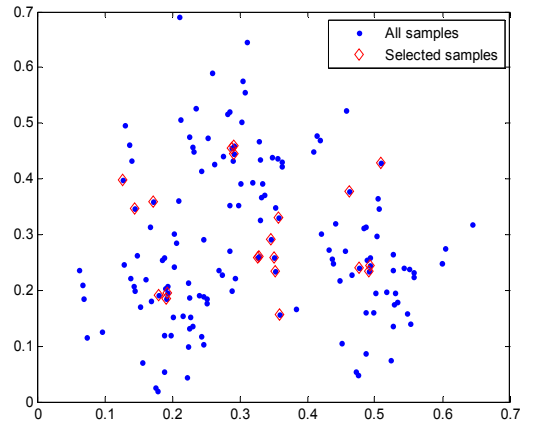


Figure 5. Initial training samples from hybrid selection

## 4. Empirical Study

### 4.1 Configuration

In this work, training samples are selected by analyzing membership matrix computed by fuzzy c-means. Fuzzy c-means is configured as follows: Fuzzy factor( $m$ , in Section 3.A) is set to 2.

Convergence criterion ( $\varepsilon$ , in Section 3.A) is set to 0.00001. Maximum iteration is set to 100. Euclidean distance is used as the similarity measure.

To test the effect of our proposed methods, a classifier is needed. In this empirical study, Multilayer Perceptron neural network with back propagation (BP) training algorithm is used. In all the experiments, the network with one hidden layer is adopted. TANSIG, LOGSIG activation functions are used in the hidden layer and output layer respectively. Let  $n_1, n_2, n_3$  denote the number of input nodes, hidden nodes and output nodes respectively.

In our experiments,  $n_1$  is the number of attributes in each sample,  $n_2 = 2 \times n_1 + 1$ ,  $n_3$  is the number of classes. Consider the example of the iris data set. It contains four attributes and classifies them into three classes. In this case, a 4-9-3 network is used. Each network is trained to 100 epochs. Note that, since the relative instead of absolute performance of the proposed methods are concerned, the architecture and training process of the neural networks have not been finely tuned.

Four well-known data sets are used in our study. The first data set used is the well-known Iris dataset [15]. It contains of four characteristics of iris plants and classifies them into three classes of iris with 50 exemplars in each class. One class is linearly separable from the other two which are not linearly separable from each other.

The second data set is soybean [15]. It includes four classes with the number of samples 10,10,10 and 17 respectively.

The third data set is Dr. William W. Wolberg’s Wisconsin Breast Cancer Dataset [15]. Originally this dataset contains 699 samples with 458 samples in the class Benign and 241 samples in the class Malignant. Each sample has 9 input features. There are 16 samples with incomplete features. After filtering out those samples, 683 samples are used in our experiment.

The last data set used is the Wine dataset [15]. It contains 178 samples with 59, 71 and 48 in three classes respectively. Each sample has 13 input features.

To evaluate the effects of our proposed methods on active learning, the experiment process is given as follows. Let  $D$  represent a dataset. Firstly of all,  $D$  is randomly partitioned into two sets:  $D_{test}$  and  $D_{non-test}$ .  $D_{test}$  represents test set. Then a small number of samples are selected from  $D_{non-test}$  as initial training data by using random selection, center-based selection, border-based selection and hybrid selection respectively. They are represented by  $D_{i-rs}, D_{i-cs}, D_{i-bs}$  and  $D_{i-hs}$ . After the initial training data is selected, the initial classifier could be obtained

through training. Then more training data are selected based on this initial classifier and a certain data selection mechanism. The data selection mechanism used in our work is a kind of certainty-based method [5]. At each iteration, it selects the sample with smallest difference between the two highest probability classes.

Let  $D_{rs}, D_{cs}, D_{bs}$  and  $D_{hs}$  represent the training data after a set of iterations ( $D_{rs}, D_{cs}, D_{bs}$  and  $D_{hs}$  includes the initial training data  $D_{i-rs}, D_{i-cs}, D_{i-bs}$  and  $D_{i-hs}$  respectively). Finally  $D_{rs}, D_{cs}, D_{bs}$  and  $D_{hs}$  will be used as training data for Multilayer Perceptron. We conducted 100 trails on each data set and average the result. In each trail, the partition of  $D_{test}$  and  $D_{non-test}$  is different.

In our experiment, we set the number of training data and initial training data to some small values, because data selection mechanisms aim to improve learning performance in the case when training data is insufficient. Also, to objectively compare the performance of our proposed initial data selection methods, experiments are conducted with different numbers of training data. The important parameters for the experiment are given in Table 4.

**Table 4. Size of test data, number of initial training data and number of training data for each data set**

	Data Set			
	iris	soybean	breast-w	wine
Test data num	75	25	100	50
Initial training data num	6	8	4	9
Training data num	9, 12, 15, 18	10,12, 14, 16	6, 8, 10,12	12,15, 18,21

The measure to evaluate the performance of our proposed methods on each data set is the average classification accuracy on different numbers of training data. For example, experiment on iris data includes four different numbers of training data (9, 12, 15 and 18). Then the performance evaluation on iris is based on the average classification accuracy on them.

### 4.2 Experiment Result

The experimental results of our proposed initial data selection methods on different data sets are shown in the following tables. The value following “ $\pm$ ” gives the standard deviation and the best result on each training data number is shown in bold face. “T” refers to the number of training data.

**Table 5. Performance comparison on iris**

	Dataset: iris			
	RS	CS	BS	HS
T=9	0.852 $\pm$	0.826 $\pm$	0.649 $\pm$	<b>0.877 <math>\pm</math></b>
	0.108	0.082	0.143	<b>0.080</b>

T=12	0.869 ± 0.086	0.858 ± 0.099	0.723 ± 0.173	<b>0.902 ±</b> <b>0.070</b>
T=15	0.860 ± 0.119	0.881 ± 0.070	0.802 ± 0.152	<b>0.916 ±</b> <b>0.042</b>
T=18	0.870 ± 0.128	0.912 ± 0.040	0.810 ± 0.172	<b>0.928 ±</b> <b>0.038</b>
Ave.	0.863	0.869	0.746	<b>0.906</b>

**Table 6. Performance comparison on soybean**

	Dataset: soybean			
	RS	CS	BS	HS
T=10	0.909 ± 0.109	<b>0.942 ±</b> <b>0.059</b>	0.749 ± 0.158	0.918 ± 0.083
T=12	0.930 ± 0.087	<b>0.959 ±</b> <b>0.051</b>	0.802 ± 0.134	0.938 ± 0.080
T=14	0.955 ± 0.057	<b>0.972 ±</b> <b>0.041</b>	0.877 ± 0.126	0.958 ± 0.060
T=16	0.958 ± 0.085	<b>0.963 ±</b> <b>0.050</b>	0.840 ± 0.135	0.960 ± 0.081
Ave.	0.938	<b>0.959</b>	0.817	0.944

**Table 7. Performance comparison on breast-w**

	Dataset: breast-w			
	RS	CS	BS	HS
T=6	0.876 ± 0.103	0.915 ± 0.056	0.692 ± 0.271	<b>0.921 ±</b> <b>0.049</b>
T=8	0.907 ± 0.072	0.920 ± 0.050	0.788 ± 0.220	<b>0.923 ±</b> <b>0.043</b>
T=10	0.911 ± 0.077	0.923 ± 0.061	0.907 ± 0.086	<b>0.930 ±</b> <b>0.046</b>
T=12	0.930 ± 0.047	0.931 ± 0.038	0.928 ± 0.050	<b>0.946 ±</b> <b>0.034</b>
Ave.	0.906	0.922	0.829	<b>0.930</b>

**Table 8. Performance comparison on wine**

	Dataset: wine			
	RS	CS	BS	HS
T=12	0.830 ± 0.079	0.850 ± 0.072	0.862 ± 0.081	<b>0.880 ±</b> 0.062
T=15	0.860 ± 0.056	0.856 ± 0.061	0.859 ± 0.065	<b>0.887 ±</b> 0.052
T=18	0.865 ± 0.060	0.860 ± 0.071	0.871 ± 0.062	<b>0.880 ±</b> 0.055

T=21	0.872 ± 0.055	0.871 ± 0.062	0.875 ± 0.052	<b>0.885 ±</b> 0.059
Ave.	0.857	0.859	0.867	<b>0.883</b>

Table 5-8 exhibits that compared with randomly selecting initial training data, hybrid selection and center-based selection are better on all the experimental data sets. While, border-based selection is better than random selection only on the last data set. If we further compare the performance of hybrid selection and center-based selection, we can see hybrid selection outperforms center-based selection in most cases. Actually this result is not difficult to understand. For center-based selection, it selects the samples around cluster centers as initial training data. Usually these samples are representative and informative. However, there will be redundant samples selected if we use center-based selection alone. For border-based selection, it selects the samples around borders between clusters. For these samples, if they are used alone for training, training would easily go to the wrong way. However, if they are used together with center-based selection, they can refine the result of center-based selection. This is the reason why hybrid selection could provide good performance.

## 5. Conclusion and Future Work

To achieve the best possible classifier with a small number of labeled data, in this paper, we propose three initial training data selection methods for active learning. They are center-based selection, border-based selection and hybrid selection. Center-based selection chooses the samples with high degree of membership in each cluster. In border-based selection, the samples located at the borders between clusters are selected. Hybrid selection is the combination of them. Experimental results indicate that both center-based selection and hybrid selection could effectively improve the performance of active learning. And hybrid selection is better than center-based selection.

In current work, the samples around centers and borders are simply combined without considering their distributions and densities. Therefore, it is interesting to see whether the information of distribution and density could further improve the performance of hybrid-selection mechanism.

## 6. ACKNOWLEDGMENTS

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2010-0018941).

## 7. REFERENCES

- [1] Jain, A. K., Murty, M.N., and Flynn, P. J., "Data Clustering: a review," ACM Computing Surveys, 1999, 264-323.
- [2] Zhu, X. J., "Semi-supervised learning literature survey," Technical report 1530, Computer Science, University of Wisconsin-Madison, 2006.
- [3] Lewis, D. D. and Gale, W. A., "A sequential algorithm for training text classifiers," In Proceedings of 17th ACM International Conference on Research and Development in Information Retrieval, 1994, 3-12.
- [4] Mackay, D. "Information-based objective functions for active data selection," Neural Computation, 4, 4 (1992) 305-318.

- [5] Lewis, D. D. and Catlett, J., "Heterogeneous uncertainty sampling for supervised learning," In Proceedings of 11th International Conference on Machine Learning, 1994, 148-156.
- [6] Thompson, C., Califf M. E. and Mooney, R. J., "Active learning for natural language parsing and information extraction.," In Proceeding of 16th International Conference on Machine Learning, 1999.
- [7] Tang, M., Luo, X. and Roukos, S., "Active learning for statistical natural language parsing," In Proceeding of 40th Anniversary Meeting of Association for Computational Linguistics, 2002.
- [8] Hwa, R., "On minimizing training corpus for parser acquisition," In Proceeding of 5th Computational Natural Language Learning Workshop, 2001, 84-89.
- [9] Cohn, D., Atlas, L. and Ladner, R., "Improving generalization with active learning," Machine Learning Journal, 15, 1994, 210-221.
- [10] Dagan, I. and Engelson, S.P., "Committee-based sampling for training probabilistic classifiers," In Proceeding of 12th International Conference on Machine Learning, 1995, 150-157.
- [11] Muslea, I. A., "Active learning with multiple views," Ph.D. dissertation, Univ. Southern California, 2000.
- [12] Liere, R., "Active learning with committees: an approach to efficient learning in text categorization using linear threshold algorithms," Ph.D. dissertation, Oregon State Univ. 2000.
- [13] Bezdek, J. C., "Pattern recognition with fuzzy objective function algorithms," Plenum Press, New York, 1981.
- [14] Zadeh, L. A., "Fuzzy sets as a basis for a theory of possibility.," Fuzzy Sets and Systems, 1, 1 (1978), 3-28.
- [15] UCI Machine Learning Repository, <http://www.ics.uci.edu/~mllearn/MLRepository.html>.