

A Feature Selection-based Ensemble Method for Arrhythmia Classification

Erdenetuya Namsrai*, Tsendsuren Munkhdalai*, Meijing Li*,
Jung-Hoon Shin**, Oyun-Erdene Namsrai*** and Keun Ho Ryu*

Abstract—In this paper, a novel method is proposed to build an ensemble of classifiers by using a feature selection schema. The feature selection schema identifies the best feature sets that affect the arrhythmia classification. Firstly, a number of feature subsets are extracted by applying the feature selection schema to the original dataset. Then classification models are built by using the each feature subset. Finally, we combine the classification models by adopting a voting approach to form a classification ensemble. The voting approach in our method involves both classification error rate and feature selection rate to calculate the score of the each classifier in the ensemble. In our method, the feature selection rate depends on the extracting order of the feature subsets. In the experiment, we applied our method to arrhythmia dataset and generated three top disjointed feature sets. We then built three classifiers based on the top-three feature subsets and formed the classifier ensemble by using the voting approach. Our method can improve the classification accuracy in high dimensional dataset. The performance of each classifier and the performance of their ensemble were higher than the performance of the classifier that was based on whole feature space of the dataset. The classification performance was improved and a more stable classification model could be constructed with the proposed approach.

Keywords—Data Mining, Ensemble Method, Feature Selection, Arrhythmia Classification

1. INTRODUCTION

Changes in the normal rhythm of a human heart may result in different cardiac arrhythmias. This may cause death or result in irreparable damage to the heart over a long period of time. Early diagnosis of cardiac arrhythmia makes it possible to choose appropriate anti-arrhythmic drugs, and is thus very important for improving arrhythmia therapy. Various machine learning and data mining methods have been applied to improve the accuracy of arrhythmia detection. The arrhythmia dataset, in which there are many features given, is challenging and is also where

※ The research was supported by the International Science and Business Belt Program through the Ministry of Education, Science and Technology(2012K001552) and by the National Research Foundation of Korea grant (NRF), which is funded by the Korean government (MEST) (No. 2012-0000478)

Manuscript received February 17, 2012; first revision October 22, 2012; accepted November 22, 2012.

Corresponding Author: Keun Ho Ryu

* Database/Bioinformatics Laboratory, Chungbuk National University, Cheongju, South Korea ({nerka, mjlee, khryu}@dblab.chungbuk.ac.kr)

** Dept. of Software Engineering, Chonbuk National University, Jeonju, South Korea (shinjh@jbnu.ac.kr)

*** Dept. of Information Technology, Mongolian National University, Ulaanbaatar, Mongolia (oyunerdene@num.edu.mn)

the addition of a feature selection step is necessary. Feature selection is a process of choosing a subset of features from the original features. This method is frequently used as a preprocessing technique in data mining. It has proven to be effective in reducing dimensionality of dataset.

However, the use of a subset of a feature set may eliminate important information contained in other subsets, and thus classification performance is reduced. Therefore, this paper proposes an ensemble method that uses a novel feature selection schema. Firstly, a number of feature subsets are extracted by applying the feature selection schema to the original dataset. Then classification models are built by using the each feature subset. Finally, we combine the classification models by adopting a voting approach to form a classification ensemble. The classification models are then weighted according to their feature set order and training errors to an ensemble classifier. We used the number of feature selection methods on the data set of the UCI cardiac arrhythmia dataset to get the three best dataset classifiers. In the next step we combined classifiers as an ensemble where each classifier is assigned a weight according to the order its selected feature set and training error.

The ensemble method performs classification by giving the weight to the predictions made by each best classifier. In the experiment, we applied our method to arrhythmia data and generated the top three disjoint feature sets. We then built three classifiers based on the top three feature sets and formed the classifier ensemble via the voting approach. The experiment results showed that not only was the performance of the three classifiers but also the performance of the ensemble classifier was even higher than the performance of the classifier that was based on the whole features in the data.

This paper is organized as follows: in Section 2, related research is discussed. In Section 3, we review the learning algorithms and research methods. The proposed method is presented in Section 4. In Section 5, the experimental results are given. Finally, we give our conclusions in Section 6.

2. RELATED WORK

There has been much work done with feature selection methods for creating an ensemble of classifiers. The ensemble feature selection method is where a set of the classifiers, each of which solve the same original task, are combined in order to obtain a better composite global classifier, with more accurate and reliable estimates or decisions than can be obtained from using a single classifier. The aim of designing and using the ensemble method is to achieve a more accurate classification by combining many weak learners.

Previous studies pointed out that methods like bagging improve generalization by decreasing variance, while methods similar to boosting achieve this by decreasing the bias [1]. [2] presented a technique for building ensembles from simple Bayes classifiers in random feature subsets. [3] described tree based ensembles for feature selection. It uses the approximately optimal feature selection method and classifiers constructed with all variables from the TIED dataset.

David W. Opitz et al. presented the genetic ensemble feature selection strategy, which uses a genetic search for an ensemble feature selection method. It begins with creating an initial population of classifiers where each classifier is generated by randomly selecting a different subset of features. The final ensemble is composed of the most fitted classifiers. [5] proposed a nested ensemble technique for real time arrhythmia classification. A classifier model was built for each

training sets with enhanced majority voting technique. The nested ensembles can relieve the problem of the unlikelihood of a classifier being generated when learning the classifier by an old dataset and limited input features. In our previous work, we introduced an ensemble method with a voting principle. Since the voting is performed according to only the feature set order, the performance improvement was not good as we expected.

One of the reasons that make the ensemble method popular is that ensemble methods tend to solve dataset problems. The authors [6] proposed an ensemble learning method to solve the data imbalance problem. They built classifiers by learning on a subset of training data. The subset is selected by an example selection criterion. The classification models are then composed into an ensemble classifier with a weighting principle using a training error. Also T.Soman and P.O.Bobbie [7] showed machine learning schemes, OneR, J48 and Naïve Bayes to classify an arrhythmia dataset. This work aims to automatically classify cardiac arrhythmias and to compare the accuracy and learning time dataset between classification algorithms.

The main goal of these approaches is not to select good subsets of features, but to create diverse classifiers. The previous [7, 8] works examined the accuracy of the arrhythmia dataset. But results of these works were lower than our work result.

3. RESEARCH METHODS

3.1 Feature Selection and Ensemble Methods

Feature selection is a process that selects a subset of original features while examining the importance of the regarding class label. The number of features (attributes) and number of instances in the raw dataset can be enormously large. Feature selection must be conducted to identify and remove irrelevant features [9]. Feature selection aims to maximize classification accuracy. There are three major categories of feature selection methods. Filter methods select features that are based on discriminating criteria, which are relatively independent of classification. Wrapper methods form a second group of feature selection methods. The prediction accuracy of a model directly measures the value of a feature set. Embedded methods form a third group of feature selections. Feature selection occurs naturally as part of the data-mining algorithm. These methods use all the features to generate a classifier and then they analyze the classifier to infer the importance of the features. We used a number of feature selection methods on the dataset of the UCI cardiac arrhythmia dataset to reduce the dimensions and to get the three best dataset classifiers.

Ensemble methods are learning algorithms that construct a set of base classifiers to combine and then classify new data points by taking a vote on their predictions [10, 11]. The learning procedure for ensemble algorithms is divided into two sections. The first is a construction of the base classifiers, and the second is for a voting task. The vote is used to combine classifiers. There are various kinds of voting systems. There are two main voting systems, which are namely weighted voting and un-weighted voting. In the weighted voting system, each base classifier holds different voting power. In the un-weighted voting, the individual base classifier has equal weight, and the winner is the one with the most of votes [11]. We used an ensemble method to combine classifiers as an ensemble where each classifier is given a weight according to the order of feature sets selected and their training errors. The ensemble method performs classification by giving the weight of the predictions made by each best classifier.

3.2 Machine Learning Algorithms

We used several classification algorithms and the accuracy of the classifiers were obtained by using a percentage split of 70%. The study comparatively evaluated the performance of Naïve Bayes, SVM, the Decision Tree, and the Bayes Network as described below.

The Naïve Bayes classification algorithm is based on the Bayes' theorem of poster probability. A given instance is where the algorithm computes the conditional probability of the classes and picks the class with the highest posterior. Naïve Bayes classification assumes that attributes are independent. The probabilities for nominal attributes are estimated by assuming all normal distribution for each attribute and class. Unknown attributes are simply skipped [7].

Support Vector Machine (SVM) learning supervised tools are used to analyze complex patterns by classification or regression data. Support Vector Machines are built by mapping the training patterns into a higher dimensional feature space where points can be separated by using a hyperplane. In WEKA, SVMs are implemented as the Sequential Minimal Optimization (SMO) algorithm [7].

The J48 algorithm is an implementation of the C4.5 Decision Tree learner. The algorithm uses the greedy technique to induce decision trees for classification. A decision tree model is built by analyzing training data and the model is used to classify unseen data. An information-theoretic measure is used to select the attribute tested for each non-leaf node of the tree. Decision tree induces a learner that usually is able to learn a set of rules with high accuracy [7].

Bayesian networks are a powerful probabilistic representation, and their use for classification has received considerable attention. This classifier learns data where the conditional probability of each attribute A_i is given the class label C . Classification is then done by applying the Bayes' rule to compute the probability of C given the particular instances of $A_1 \dots A_n$ and then predicting the class with the highest posterior probability. The goal of classification is to correctly predict the value of a designated discrete class variable given a vector of predictors or attributes [4]. In particular, the Naive Bayes classifier is a Bayesian network where the class has no parents and each attribute has the class as its sole parent [8].

4. A FEATURE SELECTION-BASED ENSEMBLE METHOD

This section describes an ensemble method based on feature selection for arrhythmia classification. The process of Ensemble based on Feature selection algorithm involves the following: (1) the partitioning of original dataset into different subsets by applying the feature selection technique. The result would be more than three disjoint subsets of features, (2) estimation of classification accuracy of the subsets, (3) forming an ensemble with voting, and (4) implementing an evaluation step for the combined classifier using four machine learning algorithms. Using this methodology, we have performed experiments on the cardiac arrhythmia dataset. The working of the proposed ensemble method with feature selection is shown in Fig. 1 and Table 1.

The weighting is done as shown in the following equation:

$$w_j = (e(C_j) - e(C_{j-1})) \cdot (n - j) \quad (1)$$

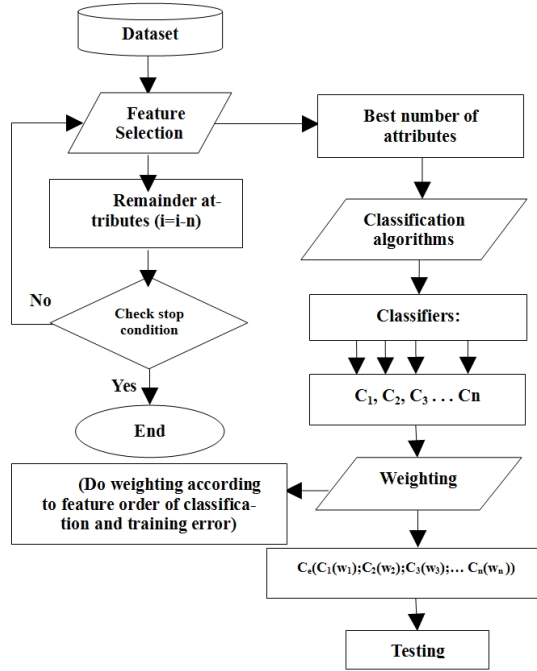


Fig. 1. The proposed ensemble method with feature selection

Table 1. The proposed ensemble learning method

	Input: Arrhythmia dataset
	Output: Improved classification accuracy
	Given: Training data, D;
	Feature selection schema, $R(F) \rightarrow f$;
	Threshold for selected feature subset, t
1	while (the number of selected features $> t$)
2	$f \leftarrow R(D)$;
3	train_classifier(C_i, f, D);
4	$i++$;
5	end while
6	for $j = 1$ to i do
7	$w_j = (e(C_j) - e(C_{j+1})) \times (i - j)$; // equation(1)
8	end for
9	classification: equation (2)

where $e(C_j)$ is the training error of the C_j classification that was obtained during the learning phase, and n is the number of classifiers. Once the ensemble classifiers have been built, class distribution for an instance is calculated as follows:

$$f(x) = \arg \max_{c_j \in C} \sum_{j=1}^n w_j c_j(x) \quad (2)$$

5. EXPERIMENT AND RESULTS

5.1 Experimental Dataset

The arrhythmia database from the UCI Machine Learning Repository was used [12]. This dataset contains 452 instances each with 279 attributes in which 206 of these are linear valued and the rest are nominal. The input dataset is in the WEKA ARFF file format [13]. The experiments were conducted in the WEKA 3.7.4 environment. The aim of our task was to reduce dimensionality and to improve the classification accuracy using the ensemble based feature selection method. Four machine-learning algorithms were used with the accuracy of three different classifiers and a combined classifier. The accuracy of classifiers was obtained by using a percentage split of 70%. The study comparatively evaluated the performance of Naïve Bayes, SVM, the Decision Tree, and the Bayes Network.

5.2 Performance Evaluation

To perform the feature selection method on the above experimental set we first converted the entire dataset into arff format, which WEKA can read. After applying the CFS feature selection filter approach of the feature selection in WEKA, the selected attributes were 24 variables: field5-QRS duration, field7-Q-T interval, field8-T interval, field11-T, field15-Heart rate, field40-of channel DIII, field76-of channel AVF, field90-of channel V1, field93-of channel V2, field100-of channel V2, field103-of channel V2, field112-of channel V3, field114-of channel V4, field190-of channel AVR, field197-of channel AVR, field211-of channel AVF, field217-of channel AVF, field224-of channel V2, field228-of channel V1, field247-of channel V4, field248-of channel V3, field267-of channel V6, field277-of channel V6, and field279-of channel V6.

In the next step, we removed those selected 24 attributes from the arrhythmia dataset using WEKA's filter function. After repeating the previous step of the feature selection we obtained the best 27 attributes, which contained: field1-Age, field28-of channel DII, field33-of channel DII, field57-of channel AVR, field69-of channel AVL, field91-of channel V1, field95-of channel V1, field102-of channel V2, field126-of channel V4, field167-T wave, field171-of channel DII, field177-of channel DII, field179-of channel DII, field181-of channel DIII, field199-of channel AVR, field207-of channel AVR, field231-of channel V2, field234-of channel V2, field238-of channel V2, field240-of channel V3, field241-of channel V3, field243-of channel V3, field249-of channel V3, field257-of channel V4, field260-of channel V5, field269-of channel V5, field270-of channel V6 and 12 attributes field65-of channel AVL, field101-of channel V2, field113-of channel V3, field117-of channel V3, field141-of channel V5, field149-of channel V6, field160-JJ wave, field230-of channel V2, field233-of channel V2, field251-of channel V4, field258-of channel V4, and field259-of channel V.

We created new top three disjoint classifiers datasets, which had 24, 27, and 12 attributes respectively. These datasets are shown in Table 2. After obtaining the best three datasets, we evaluated classifiers like Naïve Bayes, the Decision Tree, SVM, and the Bayes Network, with 70% of the dataset being used as training and the remaining 30% was used as testing data. Classification was performed using WEKA. Table 3 shows the percentage of the classification arrhythmia dataset of the three best disjoint feature sets. Also, this classification process showed different accuracy results.

We gave a decreased score to each classifier by considering the subset order and training error.

Table 2. Created feature set list table

Classifiers	The number of attributes	Notation
First subset	24	C ₁
Second subset	27	C ₂
Third subset	12	C ₃

Table 3. Accuracy of the different dataset

F-score /Percentage split 70%/	Classifiers			
	Naïve Bayes	SVM	Decision Tree	Bayes Network
C _{all}	70.5	64.4	77.4	78
C ₁	74.38	63.07	81.75	77.84
C ₂	73.53	63.37	78.19	66.89
C ₃	57.85	43.52	67.16	48.73

For example, we can give the weight of 1 to classifier 3, 2 to classifier 2, and 3 to classifier 1 based on their training errors and subset orders. After applying a weighted vote, we obtained three different classifiers and combined one ensemble. The voting result and the comparison of the accuracy of each classifier are presented in Table 4 and Fig. 2.

We also compared our method with well-known state of the art ensemble methods. The comparison result is shown in Table 5. It is clear that our method outperforms the methods by 37-28% in terms of F-measures.

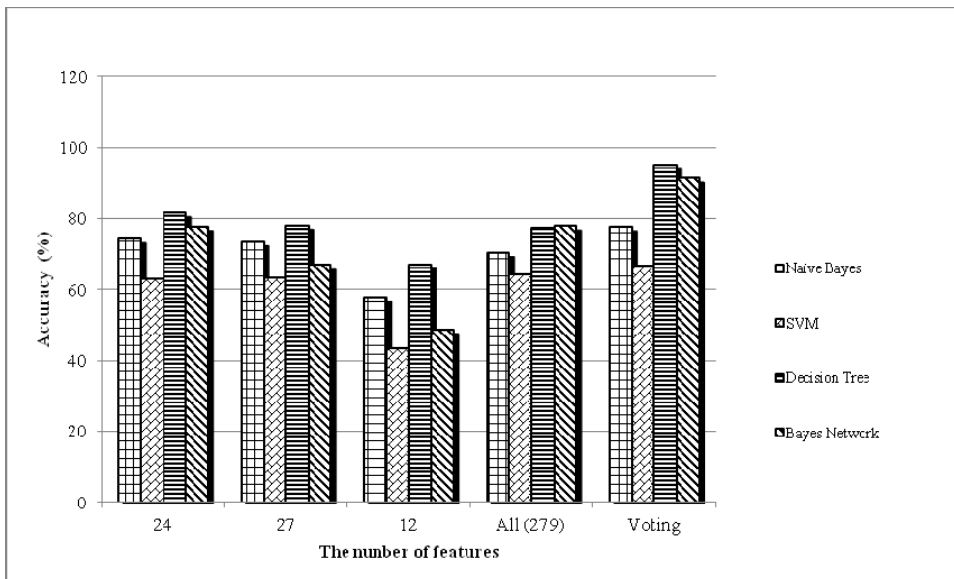


Fig. 2. Accuracy of each Classifiers by the number of selected features

Table 4. Classification accuracy with the voting result

Classification Accuracy /Percentage split 70%/	Classifiers			
	Naïve Bayes	SVM	Decision Tree	Bayes Network
C _{all}	70.5	64.4	77.4	78
C ₁	74.38	63.07	81.75	77.84
C ₂	73.53	63.37	78.19	66.89
C ₃	57.85	43.52	67.16	48.73
C _e	77.78	66.67	95.24	91.43

Table 5. Classification performance as compared with state of the art ensemble methods

Classification Accuracy /Percentage split 70%/	Classifiers			
	Bagging with Bayes Network	Bagging with Naïve Bayes	Random Forest	Proposed Ensemble Method
F-Measure	67.4	63.3	58.6	95.24

6. CONCLUSION

We proposed a novel method to build an ensemble of classifiers. The method uses a feature selection schema to derive a number of feature subsets from original dataset. The feature subsets are then used to train classification models, each presenting different levels of classification performance. In order to combine the models efficiently, a voting approach that takes both classification error rate and importance of the each feature subset into account is adopted in our method.

In the experiment, we built three different best classification models by using different feature subsets. The each feature subsets consists of 24, 27, and 12 attributes from the original dataset, respectively. We then combined the models in the voting principle to form the ensemble classifier. It was reported that the ensemble classifier outperforms the classifiers that were learned on the original dataset by 7.28% for Naïve Bayes, 2.27% for SVM, 17.84% for the Decision Tree, and 13.43% for the Bayes Network of F-score. The result showed that our proposed method performs better on the high dimensional dataset.

REFERENCES

- [1] Schapire RE, Freund Y, Bartlett P, Lee WS, “Boosting the margin: A new explanation for the effectiveness of voting methods”, *Statistics*, 1998, pp.1651-1686.
- [2] Alexey Tsybal, Seppo Puuronen, David Patterson, “Feature selection for Simple Bayesian Classifiers”, *ISSU*, 2002, pp.592-600.
- [3] Eugene Tuv, Alexander Borisov, George Runger, “Feature selection with Ensembles, Artificial variables, and Redundancy elimination”, *Journal of Machine Learning Research*, Vol.10, 2009, pp.1341-1366.
- [4] Ching Wei Bang, “New Ensemble Machine Learning Method for Classification and Prediction on Gene Expression Data”, *IEEE*, 2006, pp.3478-3481.
- [5] Mohamed Ezzeldin A.Bashir, Dong Gyu Lee, Keun Ho Ryu, “Nested Ensemble Technique for Excellence Real time cardiac Health Monitoring”, *BIOCOMP*, 2010.

- [6] S.Oh, M.S Lee, B. Zhang, "Ensemble learning with active example selection for imbalanced biomedical data classification", *IEEE/ACM transactions on computational biology and bioinformatics*, Vol.8, No.2, 2011.
- [7] Thara Soman, Patrick O.Bobbie, "Classification of Arrhythmia Using Machine Learning Techniques", *WSEAS Transactions on computers*, Vol.4, June, 2005, pp.548-552.
- [8] Mohd Fauzi bin Othman, Thomas Moh Shan Yau, "Comparison of Different Classification Techniques Using WEKA for Breast cancer", *IFMBE Proceedings Vol.15*, 2007, pp.520-523.
- [9] Asha Gowda Karegowda, M.A.Jayaram, A.S. Manjunath, "Feature Subset Selection Problem using Wrapper Approach in Supervised learning", *International journal of Computer applications*, Vol.1, No.7, 2010, pp.13-17.
- [10] Pengy Yang, Yee Hwa Yang, Bing B.Zhou, "A Review of Ensemble Methods in Bioinformatics", *Current Bioinformatics*, Vol.5, No.4, 2010, pp.296-308.
- [11] Ching Wei Bang, "New Ensemble Machine Learning Method for Classification and Prediction on Gene Expression Data", *IEEE*, 2006, pp.3478-3481.
- [12] UCI Machine Learning Repository, <http://archive.ics.edu/ml/datasets/Arrhythmia>
- [13] Weka web site, <http://www.cs.waikato.ac.nz/ml/WEKA/>
- [14] F.Yaghouby, A.Ayatollahi, R.Soleimani, "Classification of Cardiac Abnormalities Using Reduced Features of Heart Rate Variability Signal", *World Applied Sciences Journal*, Vol.6, 2009, pp.1547-1554.
- [15] M.R.Homaeinezhad, E.Tavakkoli, M.Habibi, "Combination of Different Classifiers for cardiac Arrhythmia Recognition", *World Academy of Science, Engineering and technology*, 2011, pp.1189-1200.
- [16] Zhi Hua Zhou, "Ensemble Learning", *Encyclopedia of Bioinformatics*, 2009, pp.270-273.
- [17] Ho Sun Shon, Kyung-Sook Yang, Keun Ho Ryu, "Feature Selection Method using WF-LASSO for Gene Expression Data analysis", *ACM-BCM*, 2011, pp.522-525.
- [18] Jon Atli Benediktsson, Javier Ceamanos Garcia, Bjorn Waske, Jocelyn Chanussor, "Ensemble Methods for Classification of Hyperspectral Data", *IEEE*, 2008, pp.62-65.
- [19] Dymitr Ruta, Bogdan Gabrys, "Classifier selection for majority voting", *Information fusion*, Vol.6, 2005, pp.63-81.
- [20] Guangrong Li, Xiaohua Hu, Xiajiong Shen, Xin Chen, "A Novel Unsupervised feature Selection Method for Bioinformatics Data Sets through Feature Clustering", *IEEE*, 2008, pp.41-47.
- [21] Lior Rokach, Barak Chizi, "A Methodology for Improving the Performance of Non-ranker Feature Selection Filters", *International Journal Pattern Recognition and Artificial Intelligence*, 2007, pp.1-20.



Erdenetuya Namsrai

She received the BS degree in Computer Science from Mongolian National University, Ulaanbaatar, Mongolia in 2006. In 2012, she received a MS degree at Database and Bioinformatics Laboratory of the Department of Computer Science, Chungbuk National University, Cheongju, South Korea. Her research interests include data mining and bioinformatics.



Tsendsuren Munkhdalai

He received the BS in Computer Science from Mongolian National University, Ulaanbaatar, Mongolia in 2010. In 2012, He received a MS degree at Database and Bioinformatics Laboratory of the Department of Computer Science, Chungbuk National University, Cheongju, South Korea and has joined Ph.D. program at the same laboratory. His research interests include database, bioinformatics and data mining, specially bio data, information extraction and text mining.



Meijing Li

She received a MS degree at Database and Bioinformatics Laboratory, Chungbuk National University, Cheongju, South Korea in 2010. She received BS degree in the School of Information and Computing Science from Dalian University, China, in 2007. Currently, she is a Ph.D. candidate at the same laboratory of the Department of Computer Science, Chungbuk National Univ., Rep. of Korea since 2010. Her major research interests include database, bioinformatics and data mining.



Jung-Hoon Shin

Jung-Hoon Shin received the B.S. degree in computer science from the Soongsil University in 1982, the M.S., and Ph.D. degrees in computer science from the Chungbuk National University, Republic of Korea, in 1991 and 1999, respectively. He joined the faculty of the Chonbuk National University in 1992. His research interests include multimedia DBMS and embedded system.



Oyun-Erdene Namsrai

She received a Ph.D. degree in Computer Science from Chungbuk National University, Cheongju, South Korea in 2008. She has been an assistant professor at Mongolian National University since 2010. Her research interests are in the area of data mining, Bioinformatics, and advanced database systems.



Keun Ho Ryu

He received a Ph.D. degree from Yonsei University, Seoul, Korea. Currently he is a professor with Chungbuk National University and a Leader of Database and bioinformatics laboratory, Cheongju, Korea. He served the Korean Army as ROTC. He was not only a Postdoctoral Researcher and Research Scientist at the University of Arizona, Tucson, but also the Electronic and Telecommunications Research Institute, Daejeon, Korea. His research interests include temporal databases, the spatiotemporal database, stream data processing, knowledge-

base information retrieval, database security, data mining, bioinformatics, and biomedical.

Dr. Ryu has served on numerous program committees, including Demonstration Co-Chair of the International Conference on Very large Databases, the PC committee member of APWeb, and the Advanced Information Networking and Applications. He has been a member of the ACM and IEEE since 1983.