

A Framework for a Decision Tree Learning Algorithm with Rough Set Theory

Masaki Kurematsu^(✉), Jun Hakura, and Hamido Fujita

Faculty of Software and Information,
Iwate Prefectural University, Takizawa, Japan
{kure, hakura, issam}@iwate-pu.ac.jp

Abstract. In this paper, we improve the conventional decision tree learning algorithm using rough set theory. First, our approach gets the upper approximate for each class. Next, it generates the decision tree from each upper approximate. Each decision tree shows whether the data item is in this class or not. Our approach classifies the unlabeled data item using every decision trees and integrates the outputs of these decision trees to decide the class of unlabeled data item. We evaluated our method using mechanically-prepared datasets whose the proportion of overlap of classes in datasets differs. Experimental result shows our approach is better than the conventional approach when the dataset has the high proportion of overlap of classes and few data items which have the same set of attributes. We guess it is possible to get better classification rules from uncertain and dispersed datasets using our approach. However, we don't use enough datasets to show this advantage in this experiment. In order to evaluate and enhance our approach, we analyze various and big datasets by our approach.

Keywords: Decision tree learning algorithm · Rough set theory · Future reduction

1 Introduction

A decision tree learning algorithm [1] is one of well-known supervised machine learning algorithms. We use this algorithm to make a decision tree from the transaction dataset and predict a class label of the unlabeled data item using the decision tree. A decision tree shows the set of classification rules which show how to decide a class label of each data item based on attribute values. Though there are some statistical methods or machine learning algorithms to make classifiers from dataset, for example, SVM (support vector machine [2]), ANN (Artificial Neural networks [3]), analysts and researchers use this algorithm to make rules from dataset and predict a class of a unlabeled data item. One of reasons for using this algorithm is that it is easy to understand a decision tree made by this algorithm. The other reason is that there are some tools based on this algorithm, for example, Weka [4], R [5] and SPSS [6]. So this algorithm is used in many research filed, such as marketing, psychology and medical [7,8].

Though there are many researches using decision tree type algorithms, this algorithm has some weak points. One of well-known weak points is that this algorithm has the impact of the training dataset. The quality of a decision tree made by this algorithm

depends on the quality and the quantity of the training dataset. If the training dataset demonstrates the area most clearly, we can make the good decision tree. It is not easy to make a good decision tree from small training dataset. If the training dataset has ambiguity, for example, there is a data item in one class which has same attribute values as other data item in the other class. We have to prepare good a training dataset in advance, however, it is difficult to do. So we have to manage the impact of the training dataset.

In order to solve this issue, we make decision trees which predict whether a data item is in the target class or not and predict the class label by voting every decision trees [9]. Before making a decision tree, we make the dataset for each class label. There are two class labels in each dataset. One class label shows the target class label and the other class label shows remaining class labels. We make a decision tree from each dataset. We predict a class label of the unlabeled data item by these decision trees and select the class label which has the maximum predict score. Experimental results said this approach was better than a conventional decision tree, however, this approach has some issues, for example, this approach does not make a good decision tree when the training dataset is small or ambiguity. We have to enhance this approach.

On the other hand, the research about Rough set theory is actively pursued. Rough set theory is a mathematical approach to vague and uncertain data analysis and imperfect knowledge proposed by Zdzisław I. Pawlak [10]. This approach has the some advantages, for example, it provides efficient algorithms for finding hidden patterns in data, it finds reduced sets of data and it identifies relationships that would not be found using statistical methods. So, this theory has possibilities to improve our approach.

In this paper, we propose a new approach to improve a decision tree learning algorithm with rough set theory. In this approach, we make some training datasets form an original training dataset by rough set theory and make some decision trees from each datasets. We predict unlabeled data item's class by voting these decision trees.

Next section gives an outline of related works. We describes our approach in Sect. 3 and the experiment for evaluation our research in Sect. 4. We also describe discussion about our approach according to the experimental result and future works in Sect. 4. Finally, we conclude this paper in Sect. 5.

2 Related Works

2.1 Overview of Enhancing a Decision Tree Learning Algorithm

As we mentioned in Sect. 1, a decision tree learning algorithm is considered to be appropriate if the tree can classify the unlabeled data items accurately and the size of the tree is small. We need good training dataset in order to make an appropriate decision tree because the performance of a decision tree depends on training dataset. However, it is not easy to prepare good training dataset in advance. So we need to improve a decision tree learning algorithm to manage the impact of the training dataset. There are some researches to improve a decision tree learning algorithm. There are two big streams, one is to make a multitude of trees and the other is to preprocess dataset.

First approach makes some difference decision trees and predicts using them. Random Forest [11] is one well-known method in first approach. Random forests are an ensemble learning method for classification that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes output by individual trees. The training algorithm for random forests repeatedly selects a random sample with replacement of the training dataset and makes a decision tree for each sample. After training, predictions for unseen samples can be made by averaging the predictions from all the individual regression trees.

Second approach adds extra information to dataset before learning. For example, Treabe et al. proposed a novel approach that the knowledge on attributes relevant to the class is extracted as association rules from the training data [12]. The new attributes and the values are generated from the association rules among the originally given attributes. They elaborate on the method and investigate its feature. The effectiveness of their approach is demonstrated through some experiments. Liang focused on the current automotive maintenance industry and combining K-means method and decision tree theory to analyze customer value and thus promote customer value [13]. This investigation first applies the K-means method to establish a customer value analysis model for analyzing customer value. By the results of the K-means method, the customers are divided into high, middle and low value groups. Moreover, further analysis is conducted for clustering variables using the LSD and Turkey HSD tests. Subsequently, decision tree theory is utilized to mine the characteristics of each customer segment. Gaddam et al. presented “*K-Means + ID3*,” a method to cascade k-Means clustering and the *ID3* decision tree learning methods [14] for classifying anomalous and normal activities in a computer network, an active electronic circuit, and a mechanical mass-beam system [15]. The *K-Means* clustering method partitions the training instances into k clusters using Euclidean distance similarity. On each cluster, representing a density region of normal or anomaly instances, they build an *ID3* decision tree. The decision tree on each cluster refines the decision boundaries by learning the subgroups within the cluster. To obtain a final decision on classification, the decisions of the *K-Means* and *ID3* methods are combined using two rules: the Nearest-neighbor rule and the Nearest-consensus rule.

Though there are some approaches to try to improve a decision tree learning algorithm, they have advantage and disadvantage. One of big disadvantage is that these algorithms make the comprehensibility of a decision tree worse. Some algorithms use additional attributes made by them. Users don't know these additional attributes. So it becomes difficult for users to understand these decision trees. Other algorithms make multiple decision trees. That is, users have to understand more than one decision tree. Understanding multiple decision trees need more time than understanding a decision tree. So we should continue to improve a decision tree learning algorithm.

2.2 Overview of Rough Set Theory

Rough set theory is a mathematical approach to vague and uncertain data analysis and imperfect knowledge proposed by Zdzisław I. Pawlak. This approach has the some advantages, for example, it provides efficient algorithms for finding hidden patterns

in data, it finds reduced sets of data and it identifies relationships that would not be found using statistical methods. We can use rough set theory for many real-life applications in various domains.

The basic concept of rough set theory is reduct. A reduct is a set of attributes that preserves partition. It means that a reduct is the minimal subset of attributes that enables the same classification of elements of the original data as the whole set of attributes. In other words, attributes that do not belong to a reduct are superfluous with regard to classification of elements of the original data. We can reduce superfluous attributes for classification using rough set theory.

Other basic concepts of rough set theory are the lower and the upper approximation. They are used to draw conclusions from data. Informal definition of the lower approximation of a set X with respect to data D is the set of all facts that can be for certain classified as X in view of the data D and informal definition of the upper approximation of a set X with respect to data D is the set of all facts that can be possibly classified as X in view of the data D. We use the lower and the under approximations to understand and manipulate uncertainty. Table 1 shows the example of rough set theory. In Table 1, A_i means i-th attribute and C means class label. Now, we consider the set of items with the class label Y. The correct set is {#1, #4, #6}. Item #2 and item #6 have same attributes but belong to difference class. It is important to manage such items. The lower approximate for class label Y is {#1, #4}, i.e., it removes {#6} from the correct set. All items in the lower approximate should belong to class Y. On the other hand, the upper approximate for the class label Y is {#1, #2, #4, #6}, i.e., it includes {#2} which doesn't belong to class Y.

Table 1. Example of rough set theory

ID	A_0	A_1	A_2	A_3	C
#1	V_1	V_3	V_5	V_7	Y
#2	V_1	V_4	V_5	V_8	N
#3	V_2	V_4	V_6	V_9	N
#4	V_2	V_3	V_6	V_7	Y
#5	V_2	V_4	V_6	V_7	N
#6	V_1	V_4	V_5	V_8	Y

2.3 Overview of a Decision Tree Learning Algorithm with Rough Set Theory

There are some researches about a decision tree learning algorithm with rough set theory, because it is possible to manage the uncertainty of dataset using rough set theory. We show some existing works in as the following;

Longjun Huang et al. [16] proposed the degree of dependency of decision attribute on condition attribute is used as a heuristic for selecting the attribute based on rough set theory. One of the keys to constructing decision tree model is to choose standard for

testing attribute, for the criteria of selecting test attributes influences the classification accuracy of the tree. There exists diversity choosing standards for testing attribute based on entropy, Bayesian, and so on. They proposed the degree of dependency of decision attribute on condition attribute, based on rough set theory, is used as a heuristic for selecting the attribute that will best separate the samples into individual classes. The results of example and experiments showed that compared with the entropy-based approach, their approach were a better way to select nodes for constructing decision tree.

Chang-Sik Son et al. [17] proposed the decision-making model for early diagnosis of congestive heart failure using rough set and decision tree approaches. The accurate diagnosis of heart failure in emergency room patients is quite important, but can also be quite difficult due to their insufficient understanding of the characteristics of heart failure. The purpose of their study was to design a decision-making model that provides critical factors and knowledge associated with congestive heart failure (CHF) using an approach that makes use of rough sets (RSs) and decision trees. Among 72 laboratory findings, it was determined that two subsets (RBC, EOS, Protein, O2SAT, Pro BNP) in an RS-based model, and one subset (Gender, MCHC, Direct bilirubin, and Pro BNP) in a logistic regression (LR)-based model were indispensable factors for differentiating CHF patients from those with dyspnea, and the risk factor Pro BNP was particularly so. To demonstrate the usefulness of the proposed model, they compared the discriminatory power of decision-making models that utilize RS- and LR-based decision models by conducting 10-fold cross-validation. The experimental results showed that the RS-based decision-making model (accuracy: 97.5 %, sensitivity: 97.2 %, specificity: 97.7 %, positive predictive value: 97.2 %, negative predictive value: 97.7 %, and area under ROC curve: 97.5 %) consistently outperformed the LR-based decision-making model (accuracy: 88.7 %, sensitivity: 90.1 %, specificity: 87.5 %, positive predictive value: 85.3 %, negative predictive value: 91.7 %, and area under ROC curve: 88.8 %). In addition, a pairwise comparison of the ROC curves of the two models showed a statistically significant difference ($p < 0.01$; 95 % CI: 2.63–14.6).

Jin-Mao Wei et al. [18] proposed rough set based approach for inducing decision trees. In their research, they proposed a new approach for inducing decision trees based on Variable Precision Rough Set Model. The presented approach was aimed at handling uncertain information during the process of inducing decision trees and generalizes the rough set based approach to decision tree construction by allowing some extent misclassification when classifying objects. In their research, two concepts, i.e. variable precision explicit region, variable precision implicit region, and the process for inducing decision trees are introduced. They discussed the differences between the rough set based approaches and the fundamental entropy based method. The comparison between the presented approach and the rough set based approach and the fundamental entropy based method on some data sets from the UCI Machine Learning Repository [19] were also reported.

There are some researches about improving a decision tree learning algorithm with rough set theory. Most approaches reduce attributes by using reduction of rough set theory to get a good decision tree.

3 A Framework for a Decision Tree Learning Algorithm with Rough Set Theory

3.1 Overview of Our Approach

There are some datasets which some data items have same attributes but belong to the different class. There are some reasons that we make such a dataset. One is that we take some mistake at making dataset and the other is that we miss some necessary attributes. We can say such a dataset is uncertainty dataset. It is difficult to divide data items into each class, so the decision tree learning algorithm cannot make a good decision tree. So we have to manage the uncertainty.

There are some proposed methods to manage the uncertain data. We focus on the rough set theory, because, we can modify class information of datasets by using the upper and lower approximates. We show it is possible that modified class information is useful to enhance the decision tree in previous research. We decide to use the rough set theory. Our approach makes the decision tree from the upper approximate of each class and it classifies new data items by integrating the prediction results by using these trees. Basically, our approach uses ID3 algorithm to generate the decision tree, however, our approach predict an unlabeled data item using the different method.

In next subsection, we will explain our approach in detail.

3.2 The Algorithm of Our Approach

First, we explain how to generate a decision tree and extract classification rules from this tree.

Our approach collects class labels from the given dataset and does the following steps for each class label.

- (1) It makes the upper approximate for a class label C_i from the given dataset, where C_i is one of class labels in the given dataset.
- (2) It redefines a class label of all data items in the given dataset. If a data item in the given dataset is the member of the upper approximate for the selected class label C_i , our approach modifies the class label of this data item to “the class label $C_i = Y$ ”. Otherwise, it modifies the class label of this data item to “the class label $C_i = N$ ”. After that, it makes a new dataset based on the new dataset. So, there are only two class labels in the new dataset.
- (3) Our approach generates a decision tree from the new dataset using the conventional decision tree learning algorithm. It extracts every paths of the decision tree as classification rules. These rules consist of conditions and the class information. Leaf node shows the class information and other nodes and brunches show conditions. There are one or two class labels with proportion in each leaf node. The proportion equals the number of data items in one class divided by the number of nodes in the leaf node. When there is one class label, it extracts a path as a classification rule for the class. Otherwise, our approach extracts a path as a classification rule for the class whose proportion is higher than other’s one. If proportion of each class label is

same, it extracts classification rules for each class label. Moreover, it uses the proportion as the weight of a classification rule.

Next, we explain how to predict an unlabeled data item’s class label using classification rules.

Our approach compares an unlabeled data item’s attributes and each classification rule’s conditions. If the unlabeled data item matches all conditions of the classification rule, our approach multiplies the weight of this rule by the number of the classification rule’s conditions and adds this value to the point of the class label in the classification rule.

After comparing an unlabeled data item and all classification rules, our approach lets the point of “the class label $C_i = Y$ ” as Zero, if the point of “the class label $C_i = Y$ ” is smaller than the point of “the class label $C_i = N$ ”.

Next, our approach calculates the sum of the point of “the class label $C_i = Y$ ” and divides the point of “the class label $C_i = Y$ ” by it. This value is the final point of “the class label $C_i = Y$ ”.

Finally, our approach selects the class label which has the maximum value of the final point as an answer and it returns the class label with its point.

4 Experiments

4.1 Overview of Our Experiments

In order to evaluate our approach, we do experiments using some datasets. According to the algorithm, our method may not be effective in all cases. So, we use some datasets to find the good condition for our approach. We will explain the overview of our experiments next.

First, we explain about dataset. Each data item in the dataset has four attributes and one class label. These attributes are categorical data. We define four class labels and make the subset for each class label. We control the overlapping region among some subsets. We make datasets to show in Table 2 using a program. In Table 2, # of data items shows the number of data items in each dataset and the size of each class is same. Overlapping ratio means the percentage of data items whose attributes equals other data item belongs to other class. If overlapping rate is more than zero, the data set is uncertain dataset. We use these datasets except #A4, #B4 and #C4 to generate classification rules and to predict #A4, #B4 and #C4 using these classification rules.

In order to generate classification rules, we use our approach and the conventional decision tree learning algorithm. When we use the conventional decision tree learning algorithm, we select the path of the decision tree whose proportion is more than 0.5 as the classification rule. We can run our approach using the lower approximate or the

Table 2. The outline of datasets

Dataset ID	#A1	#A2	#A3	#A4	#B1	#B2	#B3	#B4	#C1	#C2	#C3	#C4
# of data items	40	200	400	800	40	200	400	800	40	200	400	800
Overlapping Ratio	0.0	0.0	0.0	0.0	0.3	0.3	0.3	0.3	0.7	0.7	0.7	0.7

original class, instead of the upper approximate. In order to indicate the upper approximate is the best training sets for our approach, we generate classification rules with the lower approximate or the original class.

There are two points for evaluation the decision tree, one is the simplicity of the decision tree and the other is the accuracy of them. Our approach is weaker than the conventional approach about the simplicity of the tree, because our approach makes more than one decision trees. So we evaluate about the accuracy of them and focus on the number of data items which has the extracted class label is same as the given class label, the number of data items which has the extracted class label is not same as the given class label, the number of data items which doesn't have the extracted class label.

4.2 The Result of Experiments

Table 3 shows the outline of experiments. In Table 3, RS shows the set of classification rules by each method. DT means the set of rules generated by the conventional decision tree learning algorithm. UA means the set of rules generated by the proposed approach with the upper approximate. LA means the set of rules generated by the proposed approach with the lower approximate. OS means the set of rules generated by our approach with original class and this rule set is same as the set of rules generated by our previous approach. Every columns show the result of predication the data items in a test dataset. When we use #A1, #A2 and #A3 to generate classification rules, #A4 is a test dataset. When #B1, #B2 and #B3 are training datasets, #B4 is a test dataset. We use #C4 as the test dataset when we use #C1, #C2 and #C3 as training dataset. The column whose label is OK shows the number of data items predicted correctly. NG means the number of data items predicted in wrong class and MISS means the number of data

Table 3. The outline of experiments

RS	ID	OK	NG	MISS	ID	OK	NG	MISS	ID	OK	NG	MISS
DT	#A1	800	0	0	#B1	420	180	200	#C1	0	600	200
UA	#A1	800	0	0	#B1	210	340	250	#C1	60	690	50
OS	#A1	800	0	0	#B1	420	180	200	#C1	0	600	200
LA	#A1	800	0	0	#B1	140	60	600	#C1	0	200	600
DT	#A2	800	0	0	#B2	560	240	0	#C2	0	800	0
UA	#A2	800	0	0	#B2	140	660	0	#C2	60	740	0
OS	#A2	800	0	0	#B2	560	240	0	#C2	0	800	0
LA	#A2	800	0	0	#B2	0	0	800	#C2	0	0	800
DT	#A3	800	0	0	#B3	560	240	0	#C3	560	240	0
UA	#A3	800	0	0	#B3	400	400	0	#C3	400	400	0
OS	#A3	800	0	0	#B3	560	240	0	#C3	560	240	0
LA	#A3	800	0	0	#B3	0	0	800	#C3	0	0	800

items not predicted using classification rules, i.e., the system cannot classify data items. We judge a approach is better than others, if this approach has the biggest number of OK.

When we use #A1, #A2 and #A3 as training dataset and #A4 as a test dataset, there are no different among each method. When we use #B1, #B2 and #B3 as a training dataset and #B4 as a test dataset, the conventional algorithm is better than our approach with the upper approximate. In #C1 and #C2, our approach is better than other. However, the conventional approach is better than our approach in #C3.

5 Discussion

According to the experimental results, our approach is better than others when the overlapping region between classes is big and most data items differ each other, where the overlapping region between classes means data items which has same attribute patterns but belongs to the different class. On the other words, the dataset is uncertain and few data items in the dataset has same attributes set as others one. It is difficult for the decision tree learning algorithm to divide data items which have same attributes set but belong to the different class, because this algorithm tries to divide data items based on the attribute sets. Our approach changes the class of data items using the upper approximate from the rough set theory before classification, so, it can divide these data items. The decision tree learning algorithm, however, can divide such data items when the number of data items is high, because, it can use the proportion of data items in one class as a means of supplementing for processing these data items. This method is same as a means of supplementing for noise. We don't use some classification rule with low weight, i.e., the proportion of data items in a class is low, so, it is possible to enhance the accuracy of classification using the decision tree by using all classification rules.

In order to identify the feature of our approach, we compare classification rules generated by four approaches. Table 4 shows the result of comparing. In Table 4, #P means the number of positive classification rules and #N means the number of negative classification rules. Positive classification rules for a class judge an unlabeled data item as a member of the class. On the other hand, negative classification rules for a class judge an unlabeled data item as a not member of the class. The conventional decision tree learning algorithm doesn't generate negative classification rules and generating these rules is one of main points for our approach. According to result, the number of classifications generated by the conventional algorithm is smaller than our approach with the upper approximate and the number of positive rules is smaller than the number of negatives, because, the number of data items belonging negative class is bigger than the number of data items belonging the class. When the number of data items is low, the number of rules seems to be proportionate to the ratio of overlapping. However, the number of rules doesn't increase when the number of data items increases. When the number of data items, these algorithm can integrate some rules based on the proportion of classes in a leaf nodes, i.e., cut off rules with low probability. So, the number of rules doesn't increase when the number of data is high. Moreover, \$P shows the mean of the weight of positive classification rules and \$N shows the mean of the weight of negative

Table 4. The result of comparing of rules generated by approaches

		Number				Weight						Number				Weight				
RS	ID	#P	#N	\$P	\$N	ID	#P	#N	\$P	\$N	ID	#P	#N	\$P	\$N	ID	#P	#N	\$P	\$N
DT	#A1	9	0	1		#B1	20	0	0.67		#C1	20	0	0.67						
UA	#A1	9	15	1	1	#B1	14	18	1	1	#C1	18	22	1	1					
EA	#A1	9	15	1	1	#B1	20	34	0.67	0.8	#C1	20	34	0.67	0.8					
LA	#A1	9	15	1	1	#B1	4	16	1	1	#C1	4	16	1	1					
DT	#A2	9	0	1		#B2	16	0	0.7		#C2	16	0	0.7						
UA	#A2	9	15	1	1	#B2	19	19	1	1	#C2	19	19	1	1					
EA	#A2	9	15	1	1	#B2	15	32	0.7	0.85	#C2	16	30	0.7	0.84					
LA	#A2	9	15	1	1	#B2	0	0			#C2	0	0							
DT	#A3	9	0	1		#B3	16	0	0.7		#C3	16	0	0.7						
UA	#A3	9	15	1	1	#B3	17	17	0.93	0.93	#C3	17	17	0.93	0.93					
EA	#A3	9	15	1	1	#B3	10	30	0.7	0.84	#C3	16	28	0.7	0.85					
LA	#A3	9	15	1	1	#B3	0	0			#C3	0	0							

classification rules in Table 4. \$N of our approach with the upper approximation is same as \$P of one.

In addition to, we compare rules generated by each approach and show the result in Table 5. The denominator shows the total number of classification rules and the numerator shows the number of common classification rules. When there is not the overlapping region between each class, the classification rules generated by each approach are same. Fewer than half of classification rules generated by our approach are same as the classification rules generated by the conventional decision tree algorithm. Table 5 shows the accuracy of classification rules generated by the conventional approach is better than the accuracy of our approach. The classification rules generated by our approach are worse than rules generated by the conventional rules.

According to the experimental results, it is difficult to say our approach is better than the conventional approach. Our approach, however, is better than others in a few cases. So, it is worth using the rough set theory to generate the decision tree. Selecting our approach and the conventional approach is possible to enhance the accuracy of the

Table 5. The result of comparing between DT and UA

RS#1:RS#2	#A1	#A2	#A3	#B1	#B2	#B3	#C1	#C2	#C3
DT:UA (P)	9/9	9/9	9/9	1/20	2/16	2/16	9/20	10/16	6/16
DT:OS (P)	9/9	9/9	9/9	5/20	8/16	2/16	7/20	6/16	8/16
DT:LA (P)				1/20			1/20		
UA:OS (P)	9/9	9/9	9/9	0/14	1/19	2/17	2/18	4/19	4/17
UA:OS (N)	15/15	15/15	15/15	11/18	7/19	3/17	2/22	3/19	4/17

decision tree. Our future works are as follow. We continue experiments with other datasets which have more attributes and patterns in order to identify the feature of datasets for using our approach. After that, we select the conventional approach or our approach based on the feature of given datasets. We also consider to use our idea when we generate child nodes, because the decision tree learning algorithm is implemented as the recurrence function and the dataset changes on recurrence process. Moreover, we can cut out conflicted data items which have same attributes set but belong to different class using the lower approximate, so, we reconsider the lower approximate, too. Alongside of modification our approach, we evaluate our modified approach with some open datasets and real-world datasets.

6 Conclusion

In this paper, we proposed a modified decision tree learning algorithm. We have improved the conventional decision tree learning algorithm using rough set theory. Our approach generates the decision tree as follows. First, our approach gets the upper approximate for each class. Next, it generates the decision tree from each upper approximate. Each decision tree shows whether the data item is in this class or not. Our approach classifies the unlabeled data item using every decision trees and integrates the outputs of these decision trees to decide the class of unlabeled data item. We evaluated our method using mechanically-prepared datasets whose the proportion of overlap of classes in datasets differs. Experimental result shows our approach is better than the conventional approach when the dataset has the high proportion of overlap of classes and few data items which have the same set of attributes, i.e., the dataset is uncertain and small. So, we guess it is possible to get better classification rules from uncertain and dispersed datasets using our approach, such as medical datasets. However, we don't use enough datasets to show this advantage in this experiment. In order to evaluate and enhance our approach, we prepare various and big datasets and analyze them by our approach. After improving our approach based on experimental results, we analyze some open and real-world datasets using our approach and the other approaches and evaluate our approach based on the results.

Acknowledgment. We would like to thank Mr. Yasutomo FUKADA who graduated Iwate Prefectural University in March, 2015 and Ms. Saori AMANUMA who has completed a master's course of the graduate school of Iwate Prefectural University.

References

1. Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J.: Classification and Regression Trees (Wadsworth Statistics/ Probability). CRC-Press, Boca Raton (1984)
2. Corinna, C., Vladimir, N.V.: Support-vector networks. *Mach. Learn.* **20**(3), 237–297 (1995). Kluwer Academic Publishers, Hingham
3. Rosenblatt, F.: Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms. Spartan Books, Michigan (1962)

4. Weka 3 - Data Mining with Open Source Machine Learning Software in Java. <http://www.cs.waikato.ac.nz/ml/weka/>. Accessed 12 March 2015
5. The R Project for Statistical Computing. <http://www.r-project.org/>. Accessed 12 March 2015
6. IBM SPSS software. <http://www-01.ibm.com/software/analytics/spss/>. Accessed 12 March 2015
7. Bach, M.P., Čosić, D.: Data mining usage in health care management: literature survey and decision tree application. *Medicinski Glasnik* **5**(1), 57–64 (2008). Bosnia and Herzegovina
8. MacQueen, J.B.: Some methods for classification and analysis of multivariate observations. In: *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*, pp. 281–297. University of California Press, Berkeley (1967)
9. Kurematu, M., Hakura, J., Fujita, H.: An extraction of emotion in human speech using speech synthesizer and each classifier for each emotion. *WSEAS Trans. Inf. Sci. Appl.* **3**(5), 246–251 (2008). World Scientific and Engineering Academy and Society, Greece
10. Pawlak, Z.: Rough sets. *Int. J. Parallel Program.* **11**(5), 341–356 (1982). Springer, Heidelberg
11. Breiman, L.: Random forests. *Mach. Learn.* **45**(1), 5–32 (2001). Academic Publishers, Hingham
12. Terabe, M., Katai, O., Sawaragi, T., Washio, T., Motoda, H.: Attribute generation based on association rules. *Knowl. Inf. Syst.* **4**(3), 329–349 (2002). Springer, Heidelberg
13. Liang, Y.-H.: Combining the K-means and decision tree methods to promote customer value for the automotive maintenance industry. In: *IEEE International Conference on Industrial Engineering and Engineering Management 2009*, pp. 1337–1341. IEEE, Hong-Kong (2009)
14. Quinlan, J.R.: Induction of decision trees. *Mach. Learn.* **1**(1), 81–106 (1986). Springer, Heidelberg
15. Gaddam, S.R., Phoha, V.V., Balagani, K.S.: K-Means + ID3: a novel method for supervised anomaly detection by cascading k-means clustering and ID3 decision tree learning methods. *IEEE Trans. Knowl. Data Eng.* **9**(3), 345–354 (2007). IEEE, Washington
16. Huang, L., Huang, M., Guo, B., Zhang, Z.: A new method for constructing decision tree based on rough set theory. In: *Granular Computing, GRC 2007*, p. 241. IEEE, Washington (2007)
17. Son, C.-S., Kim, Y.-N., Kim, H.-S., Park, H.-S., Kim, M.-S.: Decision-making model for early diagnosis of congestive heart failure using rough set and decision tree approaches. *J. Biomed. Inf.* **45**(5), 999–1008 (2012). Elsevier, New York
18. Wei, J.-M., Wang, S.-Q., Wang, M.-Y., You, J.-P., Liu, D.-Y.: Rough set based approach for inducing decision trees. *Knowl.-Based Syst.* **20**(8), 695–702 (2007). Elsevier, New York
19. UCI Machine Learning Repository. <http://archive.ics.uci.edu/ml/datasets/>. Accessed 29 March 2015