The New Approach on Fuzzy Decision Forest

Jooyeol Yun, Jun Won Seo, and Taeseon Yoon

Abstract—The introduction of decision trees in the 1980's lead on to the thriving research until the 90's. However, handling data that are not in a clear form was a problem to be solved since computers were designed to only calculate crisp data. This paper presents a method that allows decision trees to handle fuzzy data. Also, for better accuracy, this paper introduces decision forests, which is a bundle of decision trees that decide the case together. Additionally we did an experiment to prove the performance of new algorithm by comparing it with support vector machine, which is known as the best algorithm in data mining field.

Index Terms—Fuzzy decision tree, decision forests, bagging, entropy, classification, data mining, iris data set.

I. INTRODUCTION

The decision tree is a type of algorithm that enables us to predict the upcoming cases using the past case's attributes. The decision tree takes the value of them to the attributes of the case that the tree needs to decide. Many algorithms such as ID3 and C5.0 have been developed for the decision tree construction. Those algorithms are frequently used in data mining and artificial intelligence field. These decision tree algorithms basically determine a case by asking questions that correspond to the determination of an attribute. When there are many attributes, entropy decides the order of questions to ask. The tree is able to decide the case quicker if there is more information gain. All these existing algorithms just deal with crisp data. However, due to observation errors, uncertainty, and so on, data collecting in real world contains a lot of fuzzy and uncertain data. Most of data collecting in real world happens in fuzzy forms, which reveal degrees of membership. Thus, some algorithms that can be applied to the real world situation and have a high accuracy are in need. It would be available to get output data in real world situation using fuzzy set theory in decision tree. This paper is concerned with combining the fuzzy set theory and entropy for constructing fuzzy decision tree. The method of decision forest is going to be used in order to obtain higher accuracy. At training time, it will construct multiple decision trees and make a single prediction by multiple individual models. It introduces a procedure of constructing fuzzy decision forest and the experiment that compares its performances with previous researches' algorithm.

II. RELATIVE RESEARCH

A. The Decision Tree

Decision tree is an algorithm which is widely used in data mining and machine learning. It simulates decisions with a computer programming by classifying the data [1]. Decision trees make some attributes end up into a leaf node from specific branches. In other words, output data is obtained from input data through a standard. Each attribute can be various, depending on individual and groups. So it determines the order of questions to be asked in order to eliminate the maximum amount of possibility. The order of questions acts as a standard of a decision tree. Entropy, which represents uncertainty, is used in ID3 algorithms (Iterative dichotomiser 3), which was first proposed by Ross Quinlan [2], [3], and Information gain is used in C4.5, and C5.0 algorithm. Entropy was quantified by Shannon [1] in 1948. Entropy stands for the degree of complexity of the data set. In other words, if the data set contains various classes of data, entropy increases. Otherwise, if it contains similar classes of data, entropy decreases. Entropy is valued from 0 to 1. And entropy can be defined as $p(x)$ stands for the possibilities of each answer. If all the $p(x)$ values are equal, the entropy gets the highest value, which leads to the most uncertain data. In ID3 algorithm, the decision tree classifies data in the tree-like frame from the data that has the highest entropy to the data which has the lowest entropy among the data. Information gain means classifying the data in better way by choosing an attribute. For example, if the mathematics score has a lot higher discrimination than the science-test. Then the attribute of mathematics score's information gain is higher than that of the science-test. Information gain (IG) for an attribute is expressed as S_v stands for a subset of S when attribute A contains " v " [4]. The formal function means the entropy of " S " and latter function means sum of entropy of subset S_v . Information gain(S, A) stands for the expected result of the decline of entropy when attribute A is used in distribution. Decision trees use these conceptions to classify the data and constructing the decision trees, and it makes inapplicable in which needs a numerical decision.

B. Fuzzy Set Theory

Fuzzy sets are the ones whose elements have degrees of membership. Fuzzy sets were first introduced by Zadeh in 1965 [5]. In classical sets, a set is just assessed in binary system that classifies an element into only two sets; "belonging to the set" that is encoded to 1) and "not belonging to the set(0)". However there is a limit in using binary system in actual situation. On the contrary, a fuzzy set classifies an element using the membership function that is values in the real unit interval $[0, 1]$. It is applied to uncertain data, which

Manuscript received September 12, 2014; revised February 12, 2015.

Jooyeol Yoon and Jun Won Seo are with the Department of Natural Science, Hankuk Academy of Foreign Studies, Yong-in, Korea (e-mail: blizzard072@naver.com, jw987@naver.com).

Taeseon Yoon was with Korea University, Seoul, Korea. He is now with Hankuk Academy of Foreign Studies, Yong-In, Korea (e-mail: tsyoon@hafs.hs.kr).

usually happens in real world situation. The degree of membership for set A can be denoted as. And for a finite set $U=\{x_1, \dots, x_n\}$, the fuzzy set (U, m) is often denoted by $\{m(x_1)/x_1, \dots, m(x_n)/x_n\}$.

C. The Fuzzy Decision Tree

Because data collecting in real world contains a fuzzy and uncertain data, a decision tree needs to contain fuzzy data [6]. In previous attempt to construct the fuzzy decision tree, a fuzzy decision tree induction method for iris flower data set was suggested [7]. As an iris data is composed of 4 features and 3 species of flower, each flower has 4 features and it expresses a dot in a four dimensional space. Many iris data can be obtained in the number of the data. If each feature is divided into 3 parts as a fuzzy data, one flower would have 12 fuzzy membership values. It obtains the entropy from the distance between an average valued of the data and a particular value. Uncertainty grows higher when the distance gets longer. On the contrary, the uncertainty gets smaller value as the distance between an average and the certain data gets smaller. If α is the distance between average and a certain data, the entropy can be defined as a following function. $E=1/(1+\alpha^2)$ And this form is appropriate to apply to the ID3 algorithm. In the experiment, it showed a similar performance compared to the previous ID3 algorithm.

D. The Decision Forest

Decision forest, also known as Random Forest, is an ensemble learning method that constructs a multitude of decision trees at training time. It makes a single prediction with the aid of multiple individual models. The process tends to ignore peculiar cases and is influenced more by the majority. This is believed to increase the accuracy when given only a small number of train sets. This random forest algorithm is first proposed by Tin Kam Ho of Bell Labs in 1995 [8]. The full training set is named S_0 . And the randomly sample subset of training data that made available for the tree t is $S_0^t \subset S_0$. In confining many decision trees that are made of different training set, the value can average the prediction or it can take the majority value. For example there is an example forest to predict continuous variables, $p_{t=1}(y|v)$, $p_{t=2}(y|v)$, $p_{t=3}(y|v)$, $p_{t=4}(y|v)$, as in Fig. 1.

Fig. 1. An example forest to predict continuous variables.

$$p(y|v) = \frac{1}{T} \sum_t p_t(y|v)$$

Fig. 2. Example forests using the average of all trees.

Fig. 3. Example forest using the most predicted case.

It can be expressed by averaging the prediction as a Fig. 2. And it also can be expressed as taking majority value as Fig. 3.

III. METHOD

In order to construct the newly suggested decision forest, several decision trees need to be constructed. The train set can be reused if they have the slightest differences. For example, if there is every other 100 data sets, different train sets can be made by alternating the cases which the set does not include. Also, this numbers increase when you choose the cases to train close to half of the number of total data. The fuzzy trees are going to have different mean values with different combinations of cases. That can result in a different decision. If the combination between two classes is unclear, it means that the slight change has influenced the value of fuzzy decision tree. When a small change affects the value of the tree, a big tree should be divided into smaller individual trees. This process is called bagging. The bagging process will divide the existing train data into data sets, which usually have similar mean values compared to others. This increases the influence of test cases which are closer to the mean of the whole data. This process will lead to constructing the tree. In addition, since all trees have one result, the results will be voted for the final decision forest's result. This method is expected to decrease the errors in the process of decision making and the chance of a peculiar case leading the tree to make errors.

In order to construct individual decision trees, the method of previous research is going to be used [7]. Iris flower data set consists of 4 features and 3 species of flower [9]. $f_i[a_{i1}, a_{i2}, a_{i3}, a_{i4}]$. 'a' represents the features. It expresses a dot in a four dimensional space. If one feature is divided into 3 parts as a fuzzy set, the membership values of 3 parts can be obtained in 4 features. As one flower has 3 features, 12 fuzzy membership values are obtained in one flower. It can be written as $F=[\mu_{n(a1)}, \mu_{m(a1)}, \mu_{w(a1)}, \dots, \mu_{w(ai)}]$. This set 'F' can be appeared as a vector in a twelve dimensional space. The average value can be obtained and the distance between average and particular value can be applied to the entropy. If " Z_{ik} " refers to the distance, entropy obtained by the distance can be defined as $\frac{1}{1+Z^2}$. The bagging process starts with sorting the cases for an attribute, following numeric order. A more complex and well-constructed Fuzzy Decision Forest method can alternate the distance function into a more proper one that fits the situation. Then, the first part and the last part of the data set are combined in order to form an individual tree's training set. Also the "second part" and the "second part from the last" of the data set are combined in order to train the second

individual tree's training set; as its mean value is similar -not the same- compared to the first one. The trees can all trained by the chosen data set along this process. The trees will test the individual cases and will output the result. And all the output decision is going to be collected. Then the forest follows the decisions of majority.

IV. RESULT

In the process of experiment, the data is divided into 2 parts for the experimental convenience. As the iris data has 4 features for each flower, it is appeared as a fuzzy eight-dimensional vector. The experiment shows the performances of each former ID3, fuzzy decision tree, and fuzzy decision forest. The five fold cross-validation was used in the process of the experiment. In the five fold cross validation, a single model was used as the validation model, and the other four models were used as training data. 150 pieces of data were divided into 3 groups. Each group has 10 data; data number 1 to 10, data number 11 to 20, and so on. The test data was changed regularly. This experimenting technique is useful in situations when there is only a small number of train data given. In this case, 150 flowers is considered a small number of train data since it requires thousands of data to show high accuracy in other algorithms such as the Support Vector machine. If the group 1 and the group 2 are compared, the result is obtained in this form.

A: the number of cases, when the test determined group 1 from the real group 1

B: the number of cases, when the test determined group 1 from the real group 2

C: the number of cases, when the test determined group 2 from the real group 1

D: the number of cases, when the test determined group 2 from the real group 2

When the real group and the determined group are the same, it means the test result is accurate. In other words, "A and D" are the accurate results, and "B and C" are the inaccurate results. In both cases of ID3 and fuzzy decision tree, the experiment result between "group 1 and group 2" and "group 1 and group 3" showed 100 percent accuracy. However, the result obtained between "group 2 and group 3" did not show 100 percent accuracy. The table behind is the result from the comparison between group 2 and 3.

TABLE I: THE RESULT OF ID3

Exp. Count	A	B	C	D
1	9	0	1	10
2	9	3	1	7
3	10	1	0	9
4	9	0	1	0

TABLE II: THE RESULT OF FUZZY DECISION TREE

Exp. Count	A	B	C	D
1	9	0	1	10
2	10	0	0	10
3	9	2	1	8
4	9	1	1	9

The result of ID3 algorithm, Table I, showed 80 percent to 95 percent in accuracy. And the result of the fuzzy decision tree, Table II, showed 85 percent to 95 percent in accuracy,

which in overall shows a 90 percent accuracy.

Each five-fold result showed different accuracy but it showed similar accuracy in total. Fuzzy decision forest showed 100 percent accuracy in all cases. So the additional experiment that dealt the data groups that consist of 5 data was conducted. In general, it might have lower accuracy compared to the previous experiment whose group was composed of 10 data. However, it also showed 100 percent accuracy in the additional experiment.

In order to prove performance of the new algorithm, we compared the new algorithm with support vector machine, which is known to be highly efficient in two class classification. Support vector machine used in the experiment only apply for 2-dimensional data, which have two features in it. In order to use support vector machine with iris data, which has 4 features, we made 6 support vector machines so that 2 features are chosen among entire 4 features. (For example, Sepal length and Sepal width, Sepal length and Petal length, Sepal length and Petal width, Sepal width and Petal width, Petal length and Petal width, and Sepal width and Petal length) Then the final decision of each data was decided by majority of every support vector machine which is using the same method as the random forest algorithm. We used SVC with linear kernel.

TABLE III: THE RESULT OF SUPPORT VECTOR MACHINE SEPAL LENGTH AND SEPAL WIDTH

Exp. Count	A	B	C	D
1	10	1	0	9
2	8	0	2	10
3	7	1	3	9
4	6	2	4	8
5	7	1	3	9

TABLE IV: THE RESULT OF SUPPORT VECTOR MACHINE SEPAL LENGTH AND PETAL LENGTH

Exp. Count	A	B	C	D
1	10	3	0	7
2	8	0	2	10
3	6	3	4	7
4	7	1	3	9
5	8	1	2	9

TABLE V: THE RESULT OF SUPPORT VECTOR MACHINE SEPAL LENGTH AND PETAL WIDTH

Exp. Count	A	B	C	D
1	9	2	1	8
2	10	1	0	9
3	4	3	6	7
4	10	0	0	10
5	10	0	0	10

TABLE VI: THE RESULT OF SUPPORT VECTOR MACHINE SEPAL WIDTH AND PETAL WIDTH

Exp. Count	A	B	C	D
1	10	2	0	8
2	9	0	1	10
3	8	3	2	7
4	8	3	2	7
5	10	2	0	8

All data classified correctly, but one case. We suppose that is because the number of train data was so small, thus the convex hull made a little error.

Formal ID3 algorithm and the previous fuzzy decision tree algorithm showed similar performance. However the newly presented method, the fuzzy decision forest algorithm,

showed much better performance compared with them. Hence enhancing the performance from the previous version. We also did an experiment using support vector machine in order to prove the performance of the new algorithm. Even the support vector machine has shown some errors in its experiment since the experiment uses only a limited number of train data. However the fuzzy decision forest didn't make any mistake, which means it can show high performance even with only a small number of train data. It could have much higher accuracy because it classified the data of very color flower and virginica flower well, which is known to be harder to classify than the other pairs.

TABLE VII: THE RESULT OF SUPPORT VECTOR MACHINE PETAL LENGTH AND PETAL WIDTH

Exp. Count	A	B	C	D
1	10	0	0	10
2	9	2	1	8
3	6	1	4	9
4	10	0	0	10
5	7	0	3	10

TABLE VIII: THE RESULT OF SUPPORT VECTOR MACHINE SEPAL WIDTH AND PETAL LENGTH

Exp. Count	A	B	C	D
1	10	2	0	8
2	9	0	1	10
3	8	3	2	7
4	8	3	2	7
5	10	2	0	8

TABLE VIII: THE RESULT OF SUPPORT VECTOR MACHINE

Exp. Count	A	B	C	D
1	10	0	0	10
2	10	0	0	10
3	10	0	0	10
4	10	1	0	9
5	10	0	0	10

V. CONCLUSION

This paper is concerned with fuzzy sets, decision tree, and the decision forest. We have presented a method of fuzzy decision forest which improved the performance of previous research, fuzzy decision tree, using the conception of the decision forest. An experiment is used in order to prove the validity of the method, and compared it with former ID3 and fuzzy decision tree. First, we applied fuzzy set theory to the decision tree. Second we used iris flower data set in order to construct a decision tree. Third we applied decision forest in order to improve performance of decision. According to the result of the experiment, it showed much better accuracy compared to ID3 and fuzzy decision tree. Also, in comparison with support vector machine, the new algorithm showed better result than support vector machine-support vector machine made 1 mistake in classifying iris data but the new algorithm made no mistake at all. The Support Vector Machine algorithm showed poor results when only given a

small amount of test data because the algorithm can make different sets of hyper planes that have considerable aberration. However, the final algorithm suggested in the paper is able to make precise decision with even a limited amount of test data. This algorithm would be useful in a situation when only a few precedent cases are given. Also, this algorithm has a high flexibility when adjusted into specific environments since the fuzzy decision forest uses a simple method of using means and bagging.

REFERENCES

- [1] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, pp. 379–423, 623–656, July, October, 1948.
- [2] K.-M. Leet, K.-M. Lee, J.-H. Lees, and L.-K. Hyung, "A fuzzy decision tree induction method for fuzzy data," in *Proc. 1999 IEEE International Fuzzy Systems Conference*, Seoul, Korea, August 22-25, 1999.
- [3] R. Kohavi and R. Quinlan, *Decision Tree Discovery*, October 10, 1999.
- [4] T. M. Mitchell, *Machine Learning*, New York: McGraw-Hill, 1997.
- [5] L. A. Zadeh, *Fuzzy Sets*, 1965.
- [6] T.-C. Wang and H.-D. LEE, "Constructing a fuzzy decision tree by integrating fuzzy," in *Proc. ACOS'06 the 5th WSEAS International Conference on Applied Computer Science*, 2006, pp. 306-311.
- [7] J. Yun, J. W. Seo, and T. Yoon, "The new approach on fuzzy decision trees," *International Journal of Fuzzy Logic Systems (IJFLS)*, vol. 4, no. 3, July 2014.
- [8] T. K. Ho, "The random subspace method for constructing decision forest," *Pattern Analysis and Machine Intelligence*, pp. 832-844, 1998.
- [9] P. S. Hoey, *Statistical Analysis of the Iris Flower Dataset*, University of Massachusetts At Lowell, 2002.

Jooyeol Yun was born in BunDang, Korean, in 1998. Since 2014, he has been studying in the Department of Natural Science, Hankuk Academy of Foreign Studies. His research interest is in artificial intelligence.

Jun Won Seo was born in Seoul, Korea, in 1998. Since 2014, he has been studying in the Department of Natural Science, Hankuk Academy of Foreign Studies. His research interest is in Artificial Intelligence.

Taeseon Yoon was born in Seoul, Korea, in 1972. He received the Ph.D. candidate degree in computer education from the Korea University, Seoul, Korea, in 2003. From 1998 to 2003, he was with EJB analyst and SCJP. From 2003 to 2004, he joined the Department of Computer Education, University of Korea, as a lecturer and Ansan University, as an adjunct professor. Since December 2004, he has been with the Hankuk Academy of Foreign Studies, where he was a computer science and statistics teacher. He was the recipient of the Best Teacher Award of the Science Conference, Gyeonggi-Do, Korea, 2013.