Artificial Neural Network Hybrid Algorithm Combimed with Decision Tree and Table

Taehwan Kim and Taeseon Yoon

Abstract—**Nowadays, the method of data mining, widely known as KDD (Knowledge-discovery in databases), is researched actively due to increasing amount of data. Decision tree, which is human-readable and understandable, is widely used in the field of rule extraction, data mining, and pattern recognition. Although decision tree is easy for quickly understanding the data, it has many faults in algorithm itself. In this paper, to compensate for these faults, we combined decision table (rough set theory) and ANN (Artificial Neural Network). With this combined algorithm, we could easily obtain more precise rules from iris data. Our experiment was in this order: 1) Find weights by using back propagation (bp) algorithm. (ANN). 2) Use dimensionality reduction method to delete huge amount of irrelevant data. (Rough set theory and decision table). 3) Make decision tree with pruned weights and rules. With ANN algorithm and decision table (dimensionality reduction method), we could easily notice the most influential (high entropy) character of iris data. With this algorithm, we could extract more precise and pruned rule from iris data sets.**

Index Terms—**Decision tree, rough set, dimensionality reduction method, decision table, ANN (Artificial Neural Networks), Iris data.**

I. INTRODUCTION

As of late, the need for automated processing became clear because of the appearance of massive databases. To mine these massive data sets, many methods exist such as Support Vector Machine (SVM), Artificial Neural Networks (ANNs), Decision Tree (DT), and rough set theory. Studies for mining more precise classification patterns from large databases have been actively researched nowadays. Among such methods, neural networks were found apt in dealing with non-linear data sets. In this paper, decision table was used as pruning method, and notably, dimensionality reduction was used in decision table to find enthalpy of each node (which is connected by weights in neural networks). More specifically, to remove uninfluential data relations, we recreated decision table where the data with the same decision value was expressed as ø The fundamental theorem and explanation for the reduction method in decision table is well explained in [1]. Then, as the final process, decision tree was used to find more precise and accurate rule from databases. ID3, C4.5 and CART decision tree algorithm are widely used today and among them, ID3 was used in this paper. Although decision tree is well known as human-readable and understandable, there are many systematic limitations in using only the decision tree algorithm. Those problems are as following:

1) Because unlike algorithms such as the Artificial Neural Networks, decision tree algorithm cannot consider multiple variables simultaneously, when certain characteristics of a data does not fit into specific classification perfectly horizontally or vertically, it causes the algorithm to be extremely complicated and impossible to divide into nodes.
2) Decision tree utilizes Hill Climbing and Greedy method and just like the methods utilized in the algorithm, the algorithm doesn't guarantee optimal solution value.
3) Decision tree can be dramatically changed even from slight changes between two variables with similar enthalpies.

To complement for these difficulties, we combined decision tree with decision table and neural networks. Decision table was used for irrelevant data and dealing with vague, imprecise data from data sets. With decision table, high volume of data was reduced and only relevant data remained. With this process, it was easier to make decision table with large data sets. Upper class and lower class are defined by enthalpy which is obtained from reduction of decision table in decision tree. The node with small enthalpy was defined as lower class, and the node with high enthalpy was defined as upper class. With this classification, we could easily classify the most influential input and also find more specific rules from big data sets. Iris data, which is in [2], was used in this paper as big data sets. With this new algorithm, by complementing for the downfalls of exclusive use of decision tree algorithm, we could obtain more precise and pruned rule from massive data sets.

A. Artificial Neural Networks (ANNs)

Artificial Neural Networks were originally developed to imitate the human brain. Like the human brains, ANNs consist of processing units (artificial neurons) and connections (weights) between the units. The greatest feature of ANNs is that they have the capacity to learn, memorize, and create relationships amongst data. This is done by a process called back propagation (bp). This is the process of adjusting the weight values that are connecting the units in the ANNs to match correctly the input and output data. With this process, new set of data can be applied to the networks but still be able to produce the correct output values from novel input values. We call this "learning by example". Through this method, ANNs can be used to interpret a set of data which we lack knowledge to understand fully but have ample amount of training data.

B. Rough Set (RS) Theory

Rough Set theory is used when dealing with data that contains vague or imprecise concepts. It is based on the usage of upper approximation and lower approximation. The values

contained in the lower approximation are certain to belong to the subset of interest and the values in the upper approximation have the possibility of belong to the subset. By using these two approximations, a decision table can be made. Then discernibility matrix can be used from the results of the table and find the values with the highest enthalpy values. This ultimately helps o find out the most influential values that decide the output and therefore is a crucial step in rule mining.

C. Decision Tree (DT)

Decision tree is widely known as human-readable and easy for realization, but there are also many problems in this algorithm, which is introduced in introduction. To compensate these problems, we combined decision table (rough set) with decision tree algorithm. There are many decision tree algorithms, such as ID3, C4.5, and C.5.0. C4.5 and C5.0 are advanced version of ID3, and ID3 is most commonly used algorithm. ID3 algorithm, which we used in this paper, classifies upper and lower class regarding entropy of each data. More specifically, ID3 algorithm classifies data with information gain value. Data with high information gain value is being upper class, and data with lower information gain value is being lower class. The equations about calculating entropy and information gain are well explained in Method and Results section (Heading3). With decision tree, we could easily classify rules of iris data in this paper.

D. Iris Data Sets

The Iris flower data set is the data that was used for the proposed algorithm in this paper. Iris flower data set is a multivariable data set introduced by Sir Ronald Fisher. The data set consists of 50 samples for each of the three species of Iris flower: Iris setosa, Iris virginica, and Iris versicolor. Four features of each flower were measured: the length and the width of both sepals and petals, in centimeters. The ultimate goal of the proposed algorithm was to extract a reliable, highly accurate rule that can discriminate the species of the flower by only looking at the features of it.

II. RELATIVE RESEARCH

Reference [1] presents the basic premise for the rough sets philosophy, gives its fundamental concepts and discusses briefly some areas of applications including ones in decision support. The data set that was used in this paper, Iris flower data set, is in [2]. Reference [3] explains how the use of rough set theory in an organism can be successfully used in helping to overcome the downfalls of using the decision tree by reducing the complexity of tree and in addition increases the accuracy. Reference [4] explains the synergy between Rough set theory and decision tree can improve efficiency, simplicity and generalization capability. Reference [5] explains in detail the purpose of data mining in today's society as well as some of the most popular algorithms for data mining including, Artificial Neural Networks, Regression, and Decision Trees. It also explains in detail the basic concept of how the Artificial Neural Network works. Reference [6] discusses the basic idea of how Artificial Neural Networks functions and explains many different types of neural networks. Reference [7] explains how Artificial

Neural Networks is used for Data Mining. It discusses the process of training the Artificial Neural Networks for accuracy, the commonly applied pruning process used to rid of redundant connections, and the effectiveness of the algorithm itself. Reference [8] introduces the fundamental concepts of rough set theory. It also explains how the data is processed and presented through attribute-value blocks, information tables reducts, indiscernibility relation and decision table. Also the basic concept of upper and lower approximation is explained. It also includes description of some examples of application of rough set theory in data mining system. Reference [9] introduces an algorithm that incorporates both Rough Set Theory and Artificial Neural Networks. The resulting algorithm shows how Rough Set theory can effectively reduce the number of input vector to Artificial Neural Network without reducing the basic knowledge of the system itself, increasing the prediction accuracy of the system. Reference [10] proves through the experiment with the HIV data set that the use of Rough Set Theory improved the Decision Tree Theory by allowing it to find patterns that weren't recognized by the Artificial Neural Networks.

III. METHOD AND RESULTS

To obtain more precise rules from iris data sets, we integrated the decision tree algorithm with decision table and neural networks. We reduced huge amount of data with dimensionality reduction method of decision table and obtained entropy of each data, and with this obtained entropy, made a decision tree. There are many decision tree algorithms such as ID3, C4.5 and C5.0, and in this paper, ID3 algorithm was used. ID3 algorithm classifies class according to each data's entropy, a measure of disorder. The range of entropy is from 0 to 1, and we calculated entropy with (1) and (2).

$$\text{Entropy(s)} = -\sum_{i=1}^{m} p_i \log_2(p_i) \quad (1)$$

$$p_i = \frac{freq(C_i, S)}{|s|} \quad (2)$$

The meaning of each variable is well explained in Table I.

TABLE I: THE MEANING OF EACH VARIABLE

S	set of given data
C	$\{C_1 C_2 C_3 \dots C_k\}$: set of class value
$freq(C_i, S)$	The number of record which is classified as C_i in S
$|s|$	The number of data in given data sets

For example, let's assume there are 10-members in math class. If 10-members are all American, entropy is 0, and if five are American and five are Asian, entropy is 1. This is well explained in (3) and (4). (The entropy is higher when the data is unorganized)

$$\text{Entropy(5,5)} = -\frac{5}{10}\log_2\frac{5}{10} - \frac{5}{10}\log_2\frac{5}{10} = 1.0 \quad (3)$$

$$\text{Entropy(10,0)} = -\frac{10}{10}\log_2\frac{10}{10} - \frac{0}{10}\log_2\frac{0}{10} = 0.0 \quad (4)$$

After calculating each node's entropy, information gain

should also be calculated. Equation (5) is equation about calculating information gain, which indicates importance degree of data set.

$$\text{Gain(A)} = I(s_1 s_2 \ldots s_m) - E(charcter\ A) \qquad (5)$$

$I(s_1 s_2 \ldots s_m)$ indicates the entropy of the upper class, and E(A) indicates average entropy of lower class when there are m nodes in character A. In other words, (5) is the upper class's entropy subtracted by lower class's entropy. If information gain value of A is huge, it means set A is influential data set.

For example, the equation to obtain information gain in example of Fig. 1 is (6).

$$\text{Infogain(Age)} = E(node) - [\tfrac{5}{14} * E(2,3) + \tfrac{4}{14} * E(4,0) + \tfrac{5}{14} * E(3,2)]$$
$$= 0.940 - [\tfrac{5}{14} * 0.971 + \tfrac{4}{14} * 0 + \tfrac{5}{14} * 0.971] \qquad (6)$$
$$= 0.940 - [0.694] = 0.246$$

When decision tree is made, data with high information gain value is the upper class, and data with lower information gain value is the lower class in decision tree. As decision tree is known for being human-readable and understandable and graphical way of representing the steps in making a decision, we could easily make more sophisticated rules from iris data sets. As rule pruning process, decision table was used, and especially, dimensionality reduction method was used in decision table to reduce huge amount sets of irrelevant data sets. Iris data sets were used in this table, which is from [2],

and we categorized three parts regarding the type of iris flower.
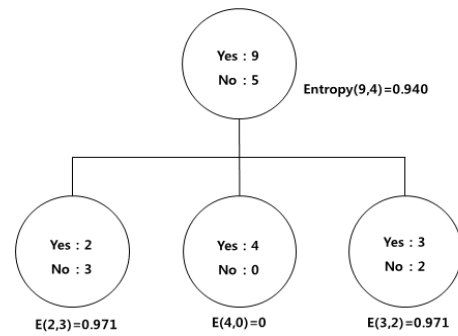

Fig. 1. Diagram used to obtain information gain.

Table II clearly shows decision table of iris data. In this paper, we did experiment with 10-data sets, which are the average of 150-data sets. Experiments were done with three setosa data, three versicolor data, and two verginica data (eight flower sets). Weights of sepal length are expressed from 1 to 4 (which connects sepal length input factor and four hidden factors), and weights of sepal width are expressed from 5 to 8 (which connects sepal width input factor and four hidden factors). Weights of petal length are expressed from 9 to 12 (which connects petal length and four hidden factors), and weights of petal width are expressed from 13 to 16 (which connects petal width and four hidden factors). Finally, weights of hidden data are expressed from 17 to 20 (which connect one bias hidden node with four output values).

TABLE II: DECISION TABLE

	Weight position									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	0.251	0.253	0.248	0.248	0.172	0.173	0.171	0.17	0.073	0.082
1	0.245	0.247	0.243	0.243	0.168	0.169	0.167	0.166	0.081	0.042
1	0.252	0.253	0.248	0.248	0.15	0.159	0.157	0.156	0.041	0.061
2	0.204	0.206	0.204	0.198	0.093	0.094	0.094	0.091	0.138	0.138
2	0.24	0.243	0.238	0.227	0.096	0.097	0.096	0.091	0.168	0.17
2	0.194	0.195	0.187	0.105	0.106	0.105	0.102	0.102	0.158	0.158
3	0.171	0.165	0.172	0.164	0.089	0.086	0.091	0.086	0.163	0.156
3	0.192	0.184	0.194	0.184	0.097	0.094	0.099	0.094	0.166	0.159
	11	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
1	0.068	0.038	0.04	0.028	0.021	0.05	0.05	0.05	0.049	0.049
1	0.068	0.053	0.068	0.02	0.02	0.024	0.046	0.047	0.046	0.045
1	0.098	0.092	0.021	0.028	0.03	0.03	0.047	0.048	0.047	0.046
2	0.137	0.133	0.041	0.045	0.042	0.081	0.03	0.03	0.03	0.029
2	0.166	0.159	0.048	0.089	0.048	0.045	0.049	0.049	0.048	0.046
2	0.157	0.152	0.08	0.06	0.076	0.077	0.034	0.034	0.033	0.032
3	0.164	0.157	0.068	0.065	0.169	0.065	0.028	0.027	0.028	0.026
3	0.167	0.159	0.167	0.159	0.06	0.155	0.033	0.032	0.033	0.031

[a.] 1=SETOSA, 2= VERSICOLOR, 3= VIRGINICA

New decision table (Table III) is made in respect of information reduction. Different from Table II, weights in Table III are complete three parts of twenty weights value with one output value (weights which connect four inputs with four hidden and four hidden with one output). From weights, which are in Table II, we calculated the difference of weights and these are the component of Table III. Data sets

with same weight result are expressed as Φ, and we expressed sets with little to no differences as blanks. (pw is abbreviation of petal width and pl is abbreviation of petal length). If the difference of weight values is large, we set the component of flowers (among sepal length, sepal width, petal length, petal width) as the component of table. As you can see in Table II, the difference between 13~16 weights, which connect petal

width and four hidden values, are largest. This means petal length is most influential factor which decides the outcome of iris data. Table III shows this result.

$$C_{ij} = \{ a \in A \mid a(x_i) \neq a(x_{ij}) \} \; for \; i,j = 1, \dots, n \quad (7)$$

TABLE III: NEW DECISION TABLE

	Weight position									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	Φ	Φ	Φ	Φ					PL	PL
2					Φ	Φ	Φ	Φ		
3									Φ	Φ
	11	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
1	PL	PL	PW	PW	PW	PW				
2			PW	PW	PW	PW				
3	Φ	Φ	PW	PW	PW	PW				

b. 1=SETOSA, 2= VERSICOLOR, 3= VIRGINICA

Each data's entropy is calculated in Table IV with weight values. With this table, we could also conclude petal width has highest information gain value because of its highest entropy. In conclusion, we concluded petal width is most influential factor with decision table and tree.

TABLE IV: ENTROPY

Iris data	Entropy			
	S_length	S_width	P_length	P_width
Setosa	0.0015	0.0079	0.0169	**0.0230**
Versicolor	0.0051	0.0102	0.0140	**0.0194**
Virginica	0.0077	0.0097	0.0160	**0.0166**

Decision tree is made to categorize iris data sets effectively. High entropy character was classified as upper class, and low entropy character was classified as lower class. As Table III and Table IV prove, petal width has highest entropy. Thus, we set petal width as upper class and sepal length as lower class. Sketchy decision tree is expressed in Fig. 2.
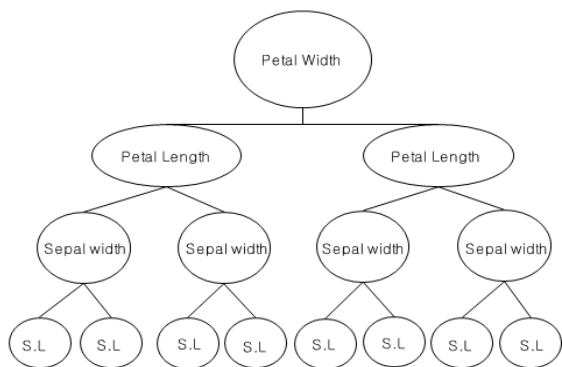

Fig. 2. Decision tree.

IV. CONCLUSIONS

Today's society relies heavily on finding patterns or rules that can be used to help us make more prudent judgments whether it is making smart decisions in stock market or making the right diagnosis on patients in hospitals. Because of this, importance of finding effective ways to fruitful data mining has been rising. Unlike the traditional method of rule extraction from large data by using rule pruning algorithm or the conventional decision tree algorithm, in the proposed algorithm, decision table and discernibility matrix were used to discern the weight values that were acquired from the use of Artificial Neural Networks whose input was the Iris flower data set. It was observed that because of the use of discernibility matrix, the process of making a decision tree was much less time-consuming because the matrix aided to reveal clearly the weight values that had significantly higher enthalpies-the most powerful weight value being the petal length for identifying iris setosa and petal width for iris versicolor and iris virginica. Thus, decision tree was made that returned three rules, which is explained in equation (8),(9), and (10). (Petal width has highest entropy, and petal length has second highest entropy; 1.9, 1.4, 2.1 are length of flowers (cm) not weight values in Table II)

If petal length <=1.9(0.0169) then Iris setosa (8)

If petal width <= 1.4(0.0194) then Iris versicolor (9)

If petal width >= 2.1(0.0166) then Iris virginica (10)

By using Artificial Neural Networks and decision table (rough set theory) in the proposed algorithm, the problems that occurred when using only the decision tree algorithm, such as multiple variables making the tree too complicated, the tree not being able to guarantee optimal solution value, and the tree being easily shifted by minor changes in variables, were solved because the weight values were already filtered to reveal which data have more enthalpy by the use of decision table and the weight values were accurately found through the use of Artificial Neural Networks. And by this, the algorithm became more efficient and accurate.

REFERENCES

[1] Z. Pawlak, "Rough set approach knowledge-based decision support ," *European Journal of Operational Research*, vol. 99, pp. 48-57, 1997.
[2] Machine Learning Repository. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Iris
[3] S. K. Shirvastava and M. Tantway, "A decision tree algorithm based on rough set theory after dimensionality redcution," *International Journal of Computer Applications*, vol. 17, no. 7, March 2011.
[4] R. Yellasiri, C. R. Rao, and V. Reddy, "Decision tree induction using rough set theory-comparative study," *Journal of Theoretical and Applied Information Technology*.
[5] Y. Singh and A. S. Chauhan, "Neural networks in data mining," *Journal of Theoretical and Applied Information Technology*.
[6] A. Kumar, *Artificial Networks for Data Mining*, I.A.S.R.I., Library Avenue, Pusa, New Delhi-110 012.
[7] H. Lu, R. Setiono, and H. Liu, "Effective data mining using neural networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 8, no. 6, December 1996.
[8] S. Thabet, "Application of rough set theory in data mining," *arXiv preprint arXiv: 1311.4121*, 2013.
[9] M. Durairaj and K. Meena, "A hybrid prediction system using rough sets and artificial neural netwoks," *International Journal of Innovative Technology & Creative Engineering*, vol. 1, no. 7, July 2011.
[10] H. Ko, J. Sung, S. M. Cho, and T. Yoon, "Comparison of the performances of the decision tree algorithm C5 using rough set and the neural network," *IJCSITAR*, vol. 2, issue 2, Feb. 2014.
[11] J. Y. Lim, J. Ryoo, J. Shin, and T. Yoon, "Modification of the Chou-Fasman method using a decision Tree," in *Proc. Hong Kong*

International Conference on Engineering and Applied Science, 2013, vol. 12, pp. 19-21.

Taehwan Kim was born in Seoul, Korea, in 1997. Since 2013, he has been studying in the International Department, Hankuk Academy of Foreign Studies. He is interested in artificial intelligence, especially in hopfield network and perceptron network. He wrote three articles about artificial intelligence and got a best speaker award in Hong Kong ICMMD (International Conference of Mechatronics and Mechanical Design) conference.

Taseon Yoon was born in Seoul, Korea, in 1972. He got a 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 in 2013.