A Decision Tree Algorithm based on Rough Set Theory after Dimensionality Reduction

Shailendra K. Shrivastava
RGPV, Bhopal, India

Manisha Tantuway
RGPV,Bhopal, India

ABSTRACT

Decision tree technology has proven to be a valuable way of capturing human decision making within a computer. As ID3 select those attribute as splitting attributes which have different values whether it classify dataset properly or not. There is another drawback of ID3 which repeat sub tree many times and select same attribute many times. These drawbacks can be removed by the proposed algorithm. By the proposed algorithm firstly large volume of dataset which contain redundant instance are reduced. These redundant instance doesn't make any contribution to take decision hence can be deleted from the dataset. After reducing the volume of the dataset decision tree is constructed through rough set. The main concept of rough set theory is degree of dependency which is used in the proposed algorithm to select splitting attribute on the compressed data. Thus the proposed algorithm reduces the complexity of tree and in addition increases the accuracy. We have used some UCI machine learning repository. By the experimental result it is shown that proposed algorithm gives better accuracy and diminishes the complexity of tree.

General Terms

Data mining, Classification, Machine Learning

Keywords

Classification, Rough Set, Decision Tree, Dimensionality Reduction, ID3.

1. INTRODUCTION

Now-a–days the data stored in a database and which is used for applications is huge. This explosive growth in data and databases has generated an urgent need for new techniques and tools that can intelligently and automatically transform the processed data into useful information and knowledge. Hence data mining has become a research area with increasing importance. Classification in data mining has gained a lot of importance in literature and it has a great deal of application areas from medicine to astronomy, from banking to text classification. The aim of the classification is to find the similar data items which belong to the same class. For example, as the tail length, ear size, number of teeth etc are the variables which may vary from one specie to another, the variables 'cat' and 'dog' will be determined according to the values of the other variables. Classification is a predictive model of data mining that predicts the class of a dataset item using the values of some other variables. Many algorithms have been introduced with different models. Classification algorithms may be categorized as follows:

1) Distance based algorithms
2) Statistical algorithms
3) Neural networks
4) Genetic algorithms
5) Decision tree algorithms

Decision tree is widely used classification algorithm which can be useful for finding structures in high dimensional spaces and in problems with mixed data, continuous and categorical. ID3, C4.5 and CART algorithms are best known decision tree algorithms. The most commonly used ID3 algorithm uses information gain to measure impurity of the data. It selects condition attribute as splitting attribute which have highest information gain. The ID3 algorithm is simpler algorithm but it have strong disadvantage that it selects those condition attribute as a splitting attribute which have different attribute values whether this attribute contain noise or irrelevant information. ID3 algorithm also generates tree which contain repeated sub tree [1], [2], [3], [4] and [5].

In the proposed algorithm, rough set theory is used to select features (splitting attribute) and to reduce irrelevant data. In the proposed algorithm, initially duplicate rows or instance are eliminated which reduces the size of dataset significantly. After eliminating duplicate instance, degree of dependency of decision attribute on condition attribute is selected from the compressed data as splitting attribute that will best separate dataset into a particular class. By deleting duplicate instance processing of decision tree becomes faster and it also decreases the storing size of dataset. By the deletion of duplicate instance the degree of dependency of decision attribute on all condition attribute is calculated. If dependency of decision attribute on condition attribute is zero then this condition attribute is irrelevant and doesn't make any contribution to take decision. After calculating degree of dependency of decision attribute on condition attribute, the condition attribute on which degree of dependency of decision attribute has maximum is selected as splitting attribute. This process is repeated in every level until all samples are classified.

2. DECISION TREE AND ROUGH SET

2.1 Decision Tree Classification

The idea of decision tree algorithms is to analyze a set of so-called training samples each assigned a class label. The decision tree system splits the training samples into subsets so that the data in each of the descendant subsets is purer than the data in the parent super set. As a classifier, the decision tree can then be used to classify an unknown sample, i.e. a sample with no class information associated according to the previously analyzed training set. From more formal point of view, a decision tree method is a supervised learning technique that builds a decision tree from a set of training samples during the machine learning process.

For inductive learning, decision tree learning is attractive for 3 reasons:
1) Decision tree is a good generalization for unobserved instance, only if the instances are described in terms of features that are correlated with the target concept.
2) The methods are efficient in computation that is proportional to the number of observed training instances.
3) The resulting decision tree provides a representation of the concept those appeals to human because it renders the classification process self-evident [6] and [7].

2.2 Rough Set

In 1982, Pawlak was introduced rough set theory. Rough set theory is a mathematical concept which is used to deal vague, imprecise concept. For example painting can be categorized into two category i.e. beautiful painting and not beautiful painting. But there is some painting which cannot be decided whether it is beautiful or not beautiful. This incomplete information which cannot be classified into particular class is called rough set. To deal incomplete information approximation is used. In the approximation, two crisp set are used i.e. lower approximation and upper approximation. Lower approximation consist all object that can be classified certainly. And upper approximation consist all possible information that can be certainly classified. Rough set theory has very large area of application like in medicine, finance, telecommunication, image analysis, pattern recognition, marketing etc.

2.3 Basic Concepts of Rough Set Theory

In 1985, Pawlak derived rough dependency of attributes in information systems. Some of the concepts of RS are given below:

2.3.1 Knowledge Base

In rough set theory, a decision table is denoted by T = (U, A, C, D), where U is universe of discourse, A is a set of primitive features, and C, D $\subset$ A are the two subsets of features that are called condition and decision features respectively.

Let a $\in$ A, P $\in$ A, A binary relation IND (P) called the Indiscernibility relation, is defined as follows:

IND (P) = {(x, y) $\in$ U x U: for all a $\in$ P, a (x) = a (y)}
Let U/ IND (P) denote the family of all equivalence classes of the relation IND (P). For simplicity of notation U/P will be written instead of U/ IND (P). Equivalence classes U/IND(C) and U/IND (D) will be called condition and decision classes respectively.

Let R$\in$ C and X $\in$U, $\underline{R\ X}$ = U {Y $\in$ U/R: Y $\in$ X} and $\overline{RX}$ = U {Y $\in$ U/R: Y∩X ≠ Φ}

Here $\underline{RX}$ and $\overline{RX}$ are said to be R-lower and upper approximations of X and ($\underline{RX}$, $\overline{RX}$) is called *R-rough set.* If X is R-definable then $\underline{RX}$= RX otherwise X is R-Rough. The boundary BNR(X) is defined as BNR(X)= $\underline{RX}$ - $\overline{RX}$. Hence if X is R-definable, then BNR(X) = Φ.

2.3.2 Dispensable and Indispensable Features

Let c $\in$ C. A feature c is dispensable in T, if POS ($_{C\text{-}(c)}$) (D) = POS $_{C}$ (D); otherwise feature c is indispensable in T. c is an independent if all c $\in$ C are indispensable.

2.3.3 Reduct and CORE

Reduct- A set of features R $\in$ C is called a reduct of C, if T' = {U, A, R, D} is independent and POSr(D). In other words, a reduct is the minimal feature subset preserving the above condition. CORE (C) denotes the set of all features indispensable in C. We have CORE (C) = ∩ RED(C) where RED(C) is the set of all reducts of C.

2.3.4 Discernibility Matrix

Matrix entries for a pair with different decision value are list of attributes in which the pair differs [1], [2], [8], [9] and [10].

3. RELATED WORK

Many researches has performed on the decision tree based on the rough set theory. some of them are summarizes in the following:

In 2006, Zhang et al. presents [11], which stepwise investigates condition attributes and outputs the classification rules induced by them, which is just like the strategy of on the fly. The theoretical analysis and empirical study shows that on the fly method is effective and efficient. They compared a proposed method with the traditional method like naïve bays, ID3, C4.5 and k-nearest neighbor. By the experimental result, a novel rough set approach gave better accuracy then traditional method. But accuracy determined by 10-fold cross validation the proposed method doesn't give best performance.

In 2007, Longjun et al. in [12] proposed a method to construct decision tree that used the degree of dependency of decision attribute on condition attribute for selecting the attribute that separate the samples into individual classes. First of all degree of dependency of decision attribute on all condition attribute calculated. The condition attribute having maximum dependency on decision attribute is selected for splitting attribute. Then this process is repeated until all samples are classified in individual class. They used four dataset labor, Monk1, Monk2 and Vote dataset are used to calculate accuracy of algorithm. A new method proposed in [12] gives higher average accuracy

78.2% than C4.5. This algorithm also produces limited node tree than C4.5.

In 2008, Cuiru et al. in [2] proposed An Algorithm for Decision Tree Construction Based on Rough Set Theory. They proposed a novel and effective algorithm in which knowledge reduction of rough set theory is used to reduce irrelevant information from the decision table. In [2], first of all degree of dependency of all condition attribute on decision attribute is determined. The condition attribute which have highest degree of dependency is selected as splitting attribute. In case if there is more than two attribute which have same degree of dependency then β-dependability is used to select splitting attribute. They used weather dataset for experimental result and compared this result to the ID3 decision tree algorithm. The decision tree generated in [2], consist limited node and produce simple and efficient decision tree.

In 2009, Baoshi et al. in [1] developed FID3 (Fixed information gain) algorithm. In the FID3, a new parameter fixed information gain is used to select splitting attribute. In FID3, attribute is reduced by calculating degree of dependency and then fixed information gain of each attribute is selected. The attribute which have highest information gain is selected as splitting attribute. FID3 algorithm removes the drawback of ID3 in which attribute is selected as splitting attribute which have different attribute values. There is one more drawback of ID3 is the instability of the decision tree built by information gain is removed by using fixed information gain.

In 2010, Baowei et al. in [13] proposed a new algorithm to construct decision tree. They stress on reducing the size of dataset and to eradicate irrelevant attributes from the dataset to reduce dimensionality. Firstly they reduced irrelevant attribute by the rough set theory then condensed the sample by removing duplicate instance. Subsequently they used the condensed dataset to construct decision tree by ID3 algorithm. From the experiments it is shown that the algorithm proposed in [13] improved greatly the number of attributes, the volume of the training samples, also and the running time. The results illustrate that the improved algorithm based on rough set theory is efficiency and robust, not only waste the store space of the data but improve the implementation efficiency.

4. PROPOSED TECHNIQUE

4.1 Steps of the proposed algorithm

In the proposed algorithm mainly three steps are performed:

(1) Deletion of redundant instance
(2) Selecting best splitting attribute to categorize data
(3) Construction of Decision Tree

(1) Deletion of redundant instance

In rough set theory, information is represented as decision table which includes condition and decision attributes. Decision table may consist attributes that have significant information but it may also contain irrelevant information which are not relevant or which are not important to classify data. Similarly decision table may contain duplicate instance which increase the size of the dataset significantly. Thus firstly the size of the dataset is decreased by removing redundant instance.

(2) Selecting best splitting attribute to categorize data

In the proposed algorithm degree of dependency is used as standard to select best splitting attribute. The degree of dependency means how much decision is effected by condition attribute. If any condition attribute is deleted from the dataset then how much accurately data is classified, is concerned in the degree of dependency. To calculate degree of dependency, first of all dependency of decision attribute on all condition attribute is calculated i.e. $r_c(d)$. If $r_c(d) = 1$ then decision table is harmonious and don't need to simplify decision table. After then degree of dependency of particular condition attribute is calculated i.e. $r_c(d)$- $r_{c-ci}(D)$. If $k_i=r_c(d)$- $r_{c-ci}(D)=0$ that means this condition attribute is not relevant and can be deleted from the decision table. After calculating degree of dependency k_i the condition attribute having highest k_i is selected as a splitting node. This process is repeated successively in each level. If more then one condition attribute having same degree of dependency then β – dependability is used. If more than one attribute having same β – dependability then attribute is selected according to the index.

(3) Construction of Decision Tree

After selecting splitting attribute, branches are grown according to different values. These branches are grown until all data are classified.

4.2 The Proposed Algorithm

Input: The training sets;
Output: A decision tree;
Step 1: Build a decision table T, C is the condition attribute set, D is the decision attribute set;
Step 2: Delete the redundant instance from decision table.
Step 3: Calculate $r_c(D) = |(POS_C(D)| / |(U)|$
if $r_c(D)=1$ then it shows that the decision table is completely harmonious; Otherwise it is inharmonious and should be simplified [1].
Step 4: For each attribute in c, calculate $r_{\{c-ci\}}(D)$, then the result of $k_i=r_c(D)$- $r_{\{c-ci\}}(D)$ can be obtained. If $k_i=0$ then C_i is irrelevant and can't be used for splitting attribute.
Step 5: Choose splitting attribute with the highest ki as the root node. If there are several attribute having the highest value then β -dependability is used.
Step 6: Grow branches according to different values c_i, and the samples are partitioned accordingly.
Step 7: If samples in a certain value are all of the same class, then generate a leaf node and is labeled with that class.
Step 8: Otherwise use the same process recursively to form a decision tree for the samples at each partition;

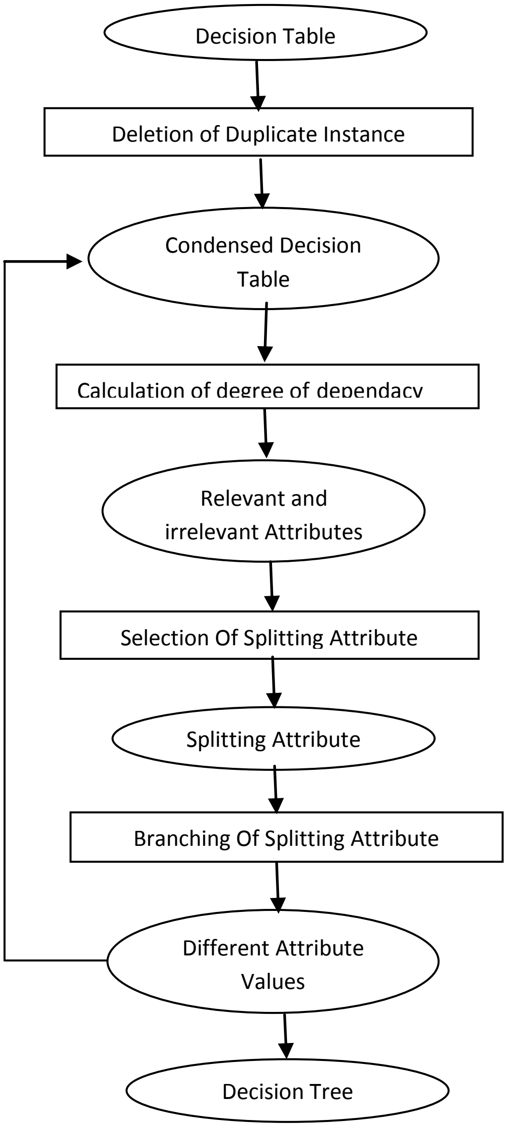
Decision Table
Deletion of Duplicate Instance
Condensed Decision Table
Calculation of degree of dependacy
Relevant and irrelevant Attributes
Selection Of Splitting Attribute
Splitting Attribute
Branching Of Splitting Attribute
Different Attribute Values
Decision Tree

Fig 1: The Architecture Of The Proposed Algorithm.

Example: The proposed algorithm is implemented by three dataset, one of the dataset fruit is shown in table 1 and decision tree constructed by the proposed algorithm is also depicted in the figure 3.

Table 1. Fruit Dataset

Size	Color	Surface	Class
Small	Yellow	Smooth	A
Medium	Red	Smooth	A
Medium	Red	Smooth	A
Big	Red	Rough	A
Medium	Yellow	Smooth	B
Medium	Yellow	Smooth	B
Big	Red	Smooth	A
Big	Yellow	Smooth	A

By the proposed algorithm, at first duplicate instance is eliminated. So we get condensed size of dataset as shown in table 2.

Table 2. Fruit Dataset after Dimensionality Reduction

Size	Color	Surface	Class
Small	Yellow	Smooth	A
Medium	Red	Smooth	A
Big	Red	Rough	A
Medium	Yellow	Smooth	B
Big	Red	Smooth	A
Big	Yellow	Smooth	A

Then degree of dependency of all condition attribute on decision attribute is calculated. After implementation following results are calculated $r_c\text{-}_{\{}r_{c\text{-size}\}}$ (class) =1, $r_c\text{-}_{\{}r_{c\text{-color}\}}$ (class) =1 and $r_c\text{-}_{\{}r_{c\text{-surface}\}}$ (class) =0. Since two attributes having same degree of dependency on decision attribute hence β- dependability is used. Following β-dependability of condition attribute are calculated beta_dep (size) = 0.6250, beta_dep (color) = 0.6250 and beta_dep (surface) = 0.3750. Hence color attribute is selected as root node of decision tree. This process is repeated until all data are classified.

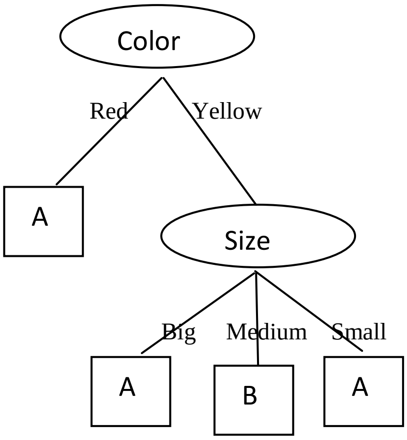

Fig 2: The Decision Tree Is Constructed By The Proposed Algorithm.

5. RESULTS AND DISCUSSION

We have used weather, sunburn and fruit dataset from UCI machine learning repository [14] for the classification. The accuracy of the decision tree is increased significantly by the proposed algorithm. We have used five-fold cross validation technique to determine accuracy. After testing the datasets the proposed algorithm gives 100% accuracy for weather and sunburn dataset whereas for fruit dataset 67% accuracy is encountered.

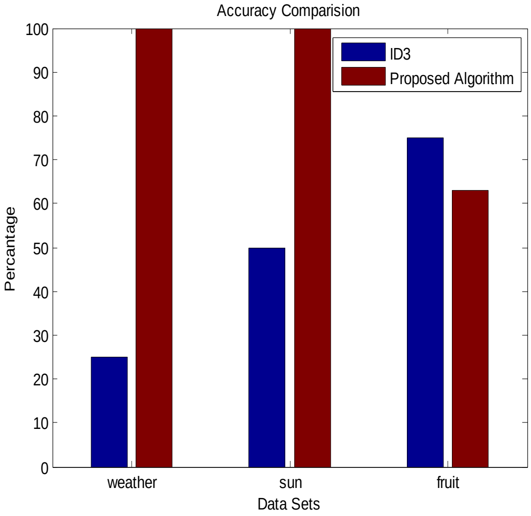

Fig 3: The Accuracy Result Of ID3 And Proposed Algorithms.

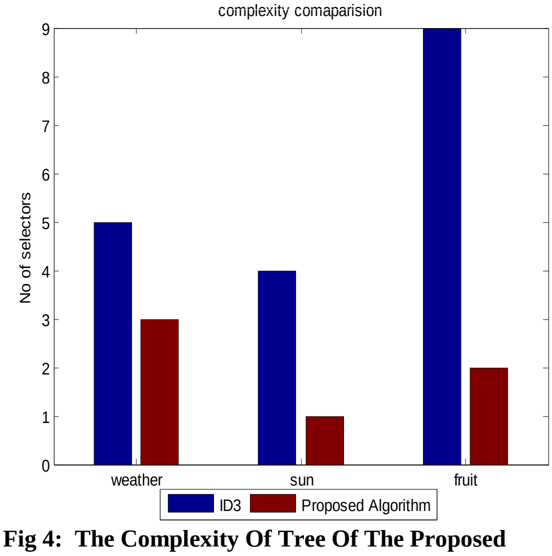

Fig 4: The Complexity Of Tree Of The Proposed Algorithm And ID3 Algorithms

The complexity of tree means total no of splitting attribute is greatly reduced by the proposed method. Since the proposed algorithm selects only relevant attributes as splitting attributes hence tree contains only limited node that is important to make decision. Thus tree contains limited no of rules as well as limited no of nodes. From the graph, it is clearly shown that the proposed algorithm has decreased complexity of tree drastically than ID3. ID3 produces large no. of nodes when size of dataset is decreased whereas the proposed algorithm gives less no. of splitting attribute while decreasing the size of dataset.

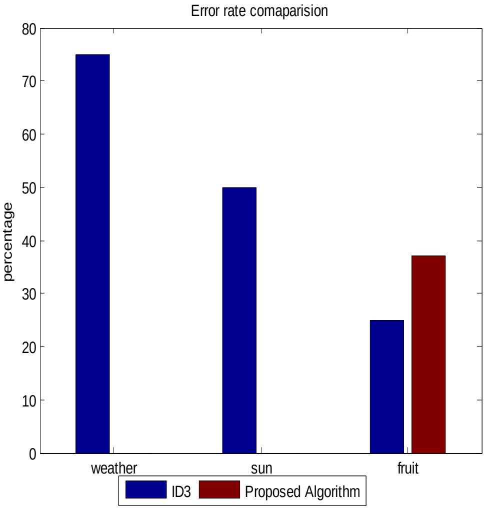

Fig 5: The Misclassification Of Data Or Error Rate Of The Different Algorithms.

If an instance is assigned to the wrong class, we say that it is misclassified. The predictive performance of a classifier is measured by its error rate. From the fig. 6, it is clearly shown that the proposed algorithm classifies data with great accuracy than ID3 algorithm.

6. CONCLUSION

In the proposed algorithm, dimensionality of the dataset is condensed significantly by rough set theory which reduces the storage capacity and makes faster processing of the algorithm. It is observed from the experimental result that the proposed algorithm not only decreased dimensionality of the dataset but also provide best result than ID3 algorithm. The proposed algorithm has also contributed to reduce noise and irrelevant information from the tree. From the experimental result on the dataset shows that the proposed algorithm serves a good classifier which generates less no. of nodes and removes irrelevant information from the tree and produces efficient tree.

7. ACKNOWLEDGEMENT

First and foremost I want to thank Prof. Shailendra Shrivastava for his excellent guidance throughout my Dissertation. I am deeply grateful to the Computer Lab for providing me with an excellent work environment and simply for consisting of so many great people. It was a real pleasure to work with my colleagues. I extend a special word to my family member, particularly my father Shri Laxmi Prasad Tantuway & my brothers & sister for the unceasing moral & financial support & the enthusiasm showered on me from time to time & a very very special person without her blessing I can’t progress at all, my mother Late. SMT. Kashi Bai Tantuway.

8. REFERENCES

[1] Baoshi Ding, Yongqing Zheng, Shaoyu Zang, A New Decision Tree Algorithm Based on Rough Set Theory, Asia-Pacific Conference on Information Processing, IEEE 2009, pp. 326-329.

[2] Cuiru Wang and Fangfang OU, An Algorithm for Decision Tree Construction Based on Rough Set Theory, International Conference on Computer Science and Information Technology, IEEE 2008,pp. 295-299.

[3] Subrata Pramanik, Md. Rashedul Islam, Md. Jamal Uddin, Pattern Extraction, Classification and Comparison Between Attribute Selection Measures, International Journal of Computer Science and Information Technologies, Vol. 1 (5), 2010,pp. 371-375.

[4] Gökhan Silahtaroğlu , An Attribute-Centre Based Decision Tree Classification Algorithm, World Academy of Science, Engineering and Technology 2009, pp:302-306.

[5] Rosiline Jeetha B. and Punithavalli M. An Integrated Study on Decision Tree Induction Algorithms in Data Mining, 2009.

[6] Tom Mitchell, Machine Learning, McGraw Hill, Computer Science Series. 1997, pp: 334.

[7] E. B. Hunt, J. Marin, and P. J. Stone. Experiments in Induction. Academic Press, 1966.

[8] Ramadevi Yellasiri, C.R.Rao,,Vivekchan Reddy, Decision Tree Induction Using Rough Set Theory – Comparative Study, Journal of Theoretical and Applied Information Technology, 2007,pp. 230-241.

[9] Jinmao Wei and Dao Huang, Rough Set Based Decision Tree, Proceedings of the 4 World Congress on Intelligent Control and Automation, IEEE 2002.

[10] Pawlak, Rough Sets, International Journal of Information and Computer Science, vol lI, 1982, pp. 341 -356.

[11] Zhang Li-Juan and Li Zhou-Jun,A Novel Rough Set Approach for Classification, 1999.

[12] Longjun Huang, Minghe Huang, Bin Guo, Zhiming Zhang, A New Method for Constructing Decision Tree Based on Rough Set Theory ,IEEE 2007, pp: 241-245.

[13] Baowei Song Chunxue Wei, Algorithm of Constructing Decision Tree Based on Rough Set Theory, International Conference on Computer and Communication Technologies in Agriculture Engineering, 2010.

[14] UCI Repository of Machine Learning Databases. http://www.ics.uci.edu/~mlearn.