ISSN:0975-9646
Rajkumar Sharma et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 3 (3) , 2012,3942-3947

An Optimize Decision Tree Algorithm Based on Variable Precision Rough Set Theory Using Degree of β-dependency and Significance of Attributes

Rajkumar Sharma, Pranita Jain, Shailendra K.Shrivastava
IT Deptt, SATI Vidisha, India

Abstract-In this paper, an optimize and effective algorithm is proposed for constructing decision tree based on variable precision rough set theory which can deal with inconsistent, uncertain or vague knowledge. FID3 has some drawbacks, it does not provide relaxation to the subset operator .Therefore, we improved FID3 algorithm based on variable precision rough set .This paper proposes a new attribute selection criterion, the enhanced information gain based on degree of β - dependency and significance of condition attributes on decision attribute is used as a heuristic for selecting the optimal splitting attribute to overcome the drawback of FID3 algorithm. Experiments prove that the improved VPRSFID3 algorithm reduces the complexity of tree and increases classification accuracy of the decision trees as compare to the FID3 algorithm.

Keywords-*Decision tree, ID3 algorithm, degree of β-dependency, variable precision rough set (VPRS), enhanced information gain*

I. Introduction

Machine learning, a branch of artificial intelligence, is a scientific discipline concerned with the design and development of algorithms. Machine learning is programming computers to optimize a performance criterion using example data or past experience.Classification is the prediction approach in data mining techniques. There are many algorithms based on classification that is Instance Based, Neural Networks, Bayesian Networks, Support Vector Machine and Decision tree. Decision tree learning is one of the most widely used and practical methods for inductive inference. Decision tree classification algorithm is one of the most popular techniques in the emerging field of data mining. ID3 algorithm [1], as a heuristic algorithm, was proposed in1986 by Quinlan, is famous and popular in the construction of decision trees. The main idea of ID3 algorithm is to choose attributes with the maximum information gain based on entropy as current classification attribute, and then expand the branches of decision tree recursively until the entire tree has been built completely. According to the performance analysis [1], information gain tends to favor those attributes with a large number of distinct values and a sub-tree may repeat several times in a single decision tree, which degrades the efficiency and accuracy of classification.

Rough set theory (RST), proposed by Poland mathematician Pawlak in 1982, is a new mathematic tool to deal with vagueness and uncertainty [2-3]. Its main idea is that to classify samples into similar classes containing objects that are indiscernible with respect to some attributes. RST can solve many problems occurred in data reduction, feature selection and pattern extraction so that we can get rid of redundant data even in the information system with null values or missing data. Rough-set-based decision tree algorithms have been studied within resent years. However, these proposed approaches also have their limitations. They only do well in accurate classification where objects are strictly classified according to equivalence classes; hence the induced classifiers lack the ability to tolerate possible noises in real world datasets. In order to improve the shortcomings of rough set model, the classical rough set model is extended, Ziarko proposed a variable precision rough set model [4], which introduced the $\beta(0\leq \beta <0.5)$ based on the basic rough set model, and allowed some degree of misclassification rate. Aijun also proposed a variable precision rough set model [6], which introduced $\beta(0.5< \beta \leq 1)$ as the correct rate. Reference [7] presented an algorithm to construct decision tree based on variable precision rough set. References [8, 9, 10, and 11] built the decision trees based on dependency of attributes and β – dependability.

The main concept of variable precision rough set theory is degree of dependency and significance of attributes which is used in the proposed algorithm to select splitting attribute,therefore this approach proposes a new attribute selection criterion, the enhanced information gain based on degree of β-dependency and significance of condition attributes on decision attribute is used as a heuristic for selecting the optimal splitting attribute to overcome the drawbacks of FID3 algorithm and also extends variable precision rough set theory. Experiments prove that the improved VPRSFID3 algorithm reduces the complexity of tree and increases classification accuracy of the decision tree as compare to the FID3 algorithm. We have used some dataset to implement the proposed algorithm and on these dataset the proposed algorithm gives better accuracy than the FID3 algorithm.

II. Basic Concepts

We introduce some basic concepts of decision tree and variable precision rough set theory [4].

1. Decision Tree Classification

Decision tree is a very practical and popular approach in the machine learning domain for solving classification problems. In decision tree approach ID3 algorithm is the most popular and used in this area.

1.1 ID3 Algorithm

Suppose S is the set of example set, and the number of equivalence class constructed by indiscernibility relation is n then entropy is defined as:

$$Entropy(S) = -\sum_{i=1}^{n} \mathrm{p_i} \log_2 \mathrm{p_i} \qquad (1)$$

Where $pi = \frac{Si}{|S|}$, $|S|$ is the number of example set S.

Given an attribute $A \in C$, C is the set of condition attributes the domain of A is denoted as V_A , then the expected information of the entropy is given as follows:

$$Info_A(S) = \sum_{i=1}^{n} \frac{|S_i|}{|S|} Entropy(S_i) \qquad (2)$$

Hence, the information gain on $A \in C$ is defined as

$$Gain(S, A) = Entropy(S) - Info_A(S) \qquad (3)$$

We compute the information gain of each condition attribute, and the attribute with the maximum information gain is the most informative attribute.

2. Variable Precision Rough Set Theory

2.1 Information Systems

An information system [5] is a pair $S=(U, W, V, f)$ where U is a non-empty finite set of objects called universe. W denotes the set of attributes, it is usually divided into two subsets P and Q, which denote the set of condition attributes and the set of decision attribute, respectively. $f: U \times W \rightarrow V$ is an information function, where $V = \square a \in WVa$ is the domain of attribute.

2.2 Relative Misclassification Rate

Variable precision rough sets (VPRS) [4] attempts to improve upon rough set theory by relaxing the subset operator. It was proposed to analyse and identify data patterns which represent statistical trends rather than functional. The main idea of VPRS is to allow objects to be classified with an error smaller than a certain predefined level.

This approach is arguably easiest to be understood within the framework of classification. Let $P, Q \subseteq$ U, the relative classification error is defined by

$$c(P,Q) = \begin{cases} 1 - \dfrac{|P \cap Q|}{|P|} & |P| > 0 \\ 0 & |P| = 0 \end{cases} \qquad (4)$$

Where $|P|$ is the cardinality of that set.

2.3 Degree of inclusion

Let P, Q be any two sets, if $0 \le \beta < 0.5$, the majority inclusion relation can be defined as:

$P \subseteq_\beta Q$ iff $c(P,Q) \le \beta$, $0 \le \beta < 0.5$

2.4 β -lower and β- upper Approximation of Set

Let R be the indiscernible relation on the universe U. Suppose (U, R) is an approximation space. $U/R = \{P_1, P_2..., P_n\}$ where Pi is an equivalence class of R. For any subset $P \subseteq U$, lower approximation $\underline{R}^{\beta} P$ and upper approximation $\bar{R}_\beta P$ of P with precision level β respect to R is respectively defined as [4]:

$$\underline{R}^{\beta} P = \cup \left\{ Q \in U/R \mid \frac{\mathrm{P \cap Q}}{\mathrm{P}} \le \beta \right. \qquad (5)$$

$$\bar{R}_\beta P = \bigcup \left\{ Q \in U/R \mid \frac{\mathrm{P \cap Q}}{\mathrm{P}} < 1 - \beta \right. \qquad (6)$$

Where the domain of β *is* $0 \le \beta < 0.5$, $\underline{R}^{\beta} X$ is also called β-Positive region (POS (P, Q)). The β boundary of P with respect to R is defined as:

$$BND^{\beta} P = \cup \left\{ Q \in U/R \mid \beta < \frac{\mathrm{P \cap Q}}{\mathrm{P}} < 1 - \beta \right. \qquad (7)$$

When $\beta = 0$, Ziarko variable precision rough set model becomes Pawlak rough set model.

III. LITRETURE SURVEY

Many researchers have worked on decision tree based on rough set theory, some of them are summarizes in the following as:

In 2006, Zhang et al. presents [12], 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 neighbour. 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.

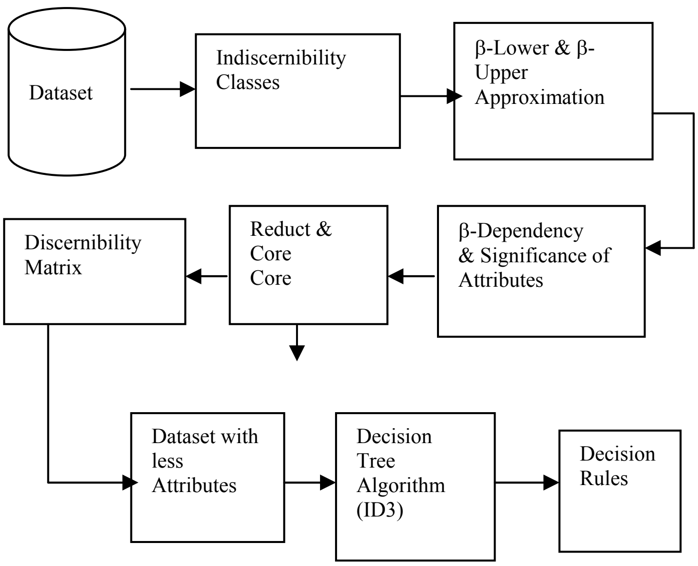
In 2007, Longjun et al. in [13] proposed a method to construct decision tree that used the degree of dependency of decision attribute on condition attribute for selecting the attribute that separate the samples into individual classes. First of all degree of dependency of decision attribute on all condition 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 labour, Monk1, Monk2 and Vote dataset are used to calculate accuracy of algorithm. A new 78.2% than C4.5. This algorithm also produces limited node tree than C4.5.method proposed in [13] gives higher average accuracy.

In 2008, Cuiru et al. in [11] 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 [11], 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 [11], consist limited node and produce simple and efficient decision tree. In 2009, Baoshi et al. in [10] 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 [14] 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 [14] 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.

IV. THE IMPROVED ALGORITHM VPRSFID3

1. *The Proposed System*



2. The *Degree of β -Dependency*

The degree of dependency of condition attributes on decision attribute $\gamma_\beta(P,Q)$ is defined as

$$k_\beta = \gamma_\beta(P,Q) = \frac{|POS_\beta\,(P,Q)|}{|U|} \tag{8}$$

Where $POS_\beta(P,Q)$is defined by (5), $\gamma_\beta(P,Q)$ implies the proportion that objects in the domain U can be correctly classified for a given value of β, and it evaluates the ability of classification to object.

3. *Significance of Attributes*

By calculating the change in dependency when an attribute is removed from the set of considered possible attributes, an estimate of the significance of that attribute can be obtained. The higher the change in dependency, the more significant the attribute is. If the significance is 0, then the attribute is dispensable without losing information .More formally, given *P, Q* and an attribute $A \in P$, the significance of attribute *A* upon *Q is* defined by:

$$\sigma P(Q,A) = \gamma P(Q) - \gamma P - \{A\}(Q) \tag{9}$$

4. The Enhanced Information Gain

The attribute chosen by information gain cannot always be the best one to split nodes in some different data sets. Another main problem is the instability of the decision tree built by information gain. Thus, we propose an enhanced information gain ($Gain_{enh}$) as the new standard for selecting splitting attributes, it is defined as:

$$Gain_{enh} = \sqrt{k_\beta * \frac{Gain(S,A)}{m}} \tag{10}$$

Where *Gain*, k_β are information gain and the degree of β-dependency of condition attributes on decision attribute defined in (3) and (8), respectively. The larger k_β is, the more determinate information is included between condition attributes and decision attribute. The condition attribute with highest degree of dependency is chosen as the test attribute. *m* means the domain of attribute *A*, $A \in C$. The greater the value of *m* is, the smaller weight factor gets. The attributes with larger number values obtain the smaller weight factor. Hence, this approach overcomes the drawback of ID3 algorithm whenever classified with information gain.

5. *Algorithm VPRSFID3*

Now we propose our algorithm to generate a decision tree in the following way:

Input: An information systems *S= (U, P ∪ Q, V, f)*, the training sets, the threshold parameter β, $0 \le \beta < 0.5$

Output: A decision tree *T*.

Step 1: Create an initial node the of tree based on maximum $Gain_{enh}$. Judge that whether the samples are all of the same class. If they are, then turns the node into a leaf and return the leaf labelled with that class.

Step 2: For each attribute in P_i, calculate $Gain_{enh}$ according to (10), choose the attribute *Ai* with the maximum value of $Gain_{enh}$ as the root node. Where P_i is the set of β- reducts.

Step3: Construct the branches according to different values of attribute P_i so that the samples are partitioned accordingly.

Step 4: If samples in a certain value are all of the same class, then generate a leaf node and is labelled with that class.

Step 5: Otherwise use the same process recursively to form a decision tree for the samples at each partition.

Step 6: Build nodes and branches repeatly until any one of the following conditions is satisfied:

- $k_\beta \le \beta$
- All samples for a given node belong to the same class, return a leaf labelled with that class.
- There are no more training samples to be classified, we can create a leaf belong to the class in majority among samples.

Step 6: Output the decision tree *T*.

V. EXAMPLE ANALYSIS

In this paper, we introduce an example using both FID3 and the proposed VPRSFID3 algorithm. According to the decision trees constructed by using two algorithms, we can clearly find the better performance of VPRSFID3.

TABLE I. A DECISION TABLE

Objects	Condition Attributes				Decision Attribute
U	**Degree**	**Experience**	**English**	**Reference**	**D**
x1	MTech	Medium	Yes	Excellent	Select
x2	MTech	Low	Yes	Neutral	Reject
x3	BSc	Low	Yes	Good	Reject
x4	MSc	High	Yes	Neutral	Select
x5	MSc	Medium	Yes	Neutral	Reject
x6	MSc	Medium	Yes	Excellent	Select
x7	MTech	High	No	Good	Select
x8	BSc	Low	No	Excellent	Reject

A decision table is shown in TABLE I .The process of constructing decision tree by VPRSFID3 can be analyzed in the following way. At first, we can create the initial node, and find that the whole samples belong to different classes, and then we begin to calculate σ_A^β according to (9). For each attribute A where A∈P.

After computation, we can get all the results:

$\sigma_{Degree}^\beta = 0$, $\sigma_{Experience}^\beta = 0.25$,

$\sigma_{English}^\beta = 0$, $\sigma_{Reference}^\beta = 0.25$

Excluding attributes Degree and English, the β -reducts attributes set P_i= {Experience, Reference}.

Calculate $Gain_{enh}$:

$Gain_{enh}$ (Experience) =0.4049,

$Gain_{enh}$ (Reference) =0.1238

Obviously, attribute Experience has the maximum $Gain_{enh}$ and should be chosen as the root node and labelled with Experience. Continue to grow branches with the different values of Experience. When Experience =low the decision attribute belong to the same class, so we create a leaf at the end of this branch and label it with Reject, Reject∈Q.And when Experience =high the decision attribute belong to the same class, so we create a leaf at the end of this branch and label it with Select, Select ∈Q. Rest branches can be built by analogue according to the computation process above.

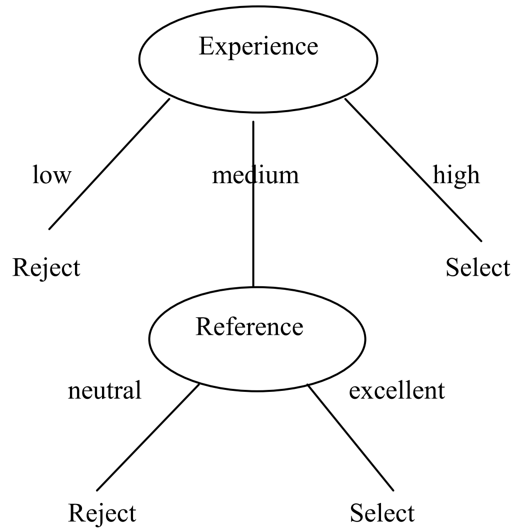


Figure 1: The decision tree constructed by VPRSFID

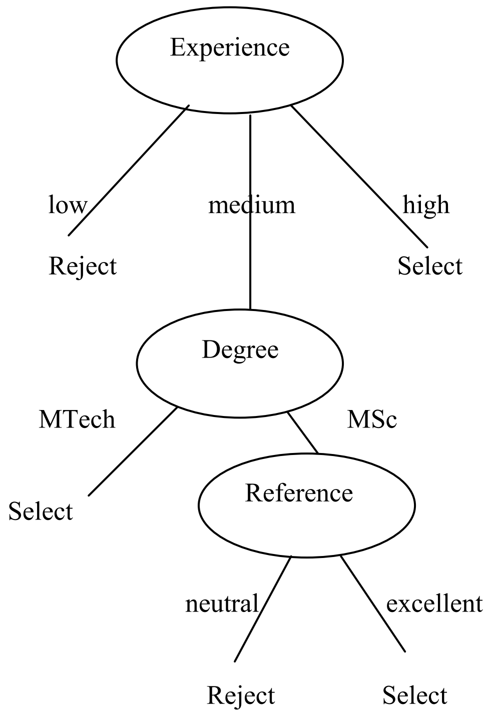


Figure 2: The decision tree constructed by FID3

Finally, we get the decision tree constructed by VPRSFID3, and it is shown in Fig.1 Meanwhile, we get the decision tree constructed by FID3 shown in Fig.2. In Fig.1, the complexity of the tree (the count of all the nodes) is 6, and it can generate 4 decision rules. Compared with the decision tree using VPRSFID3, FID3 generates more nodes and leaves. Thus, we can say that VPRSFID3 has lower complexity than FID3.In other words, the proposed algorithm can overcome the shortcomings of FID3, and generate a smaller decision tree with less space complexity.

VI. EXPERIMENTAL RESULT AND ANALYSIS

Our experiments are carried out on an Intel (R) Pentium(R) CPU B940@ 2.00 GHz, 2GB RAM, 32 bit Windows 7 Operating System. All procedures were implemented on MATLAB System. We use four groups of datasets from the UCI Machine Learning Repository [15]. In the experiments, self test validation was conducted on all data sets to calculate the classification accuracy, the average leaf nodes number and time complexity for comparing the two algorithms(FID3and VPRSFID3).TableII shows that VPRSFID3 algorithm achieves an improvement on classification accuracy and the complexity of decision tree over FID3.And the average classification accuracy and complexity of VPRSFID3 (the count of all the nodes) is 87.74%, 5, respectively, which correspondingly is 77.18%, 6.5 in FID3,respectively.

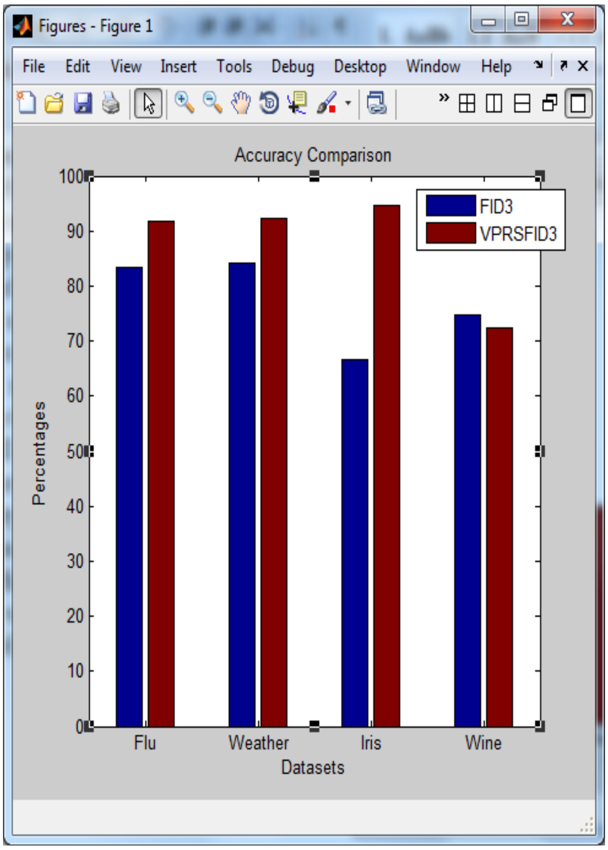


Figure3: The accuracy result of FID3 and VPRSFID3

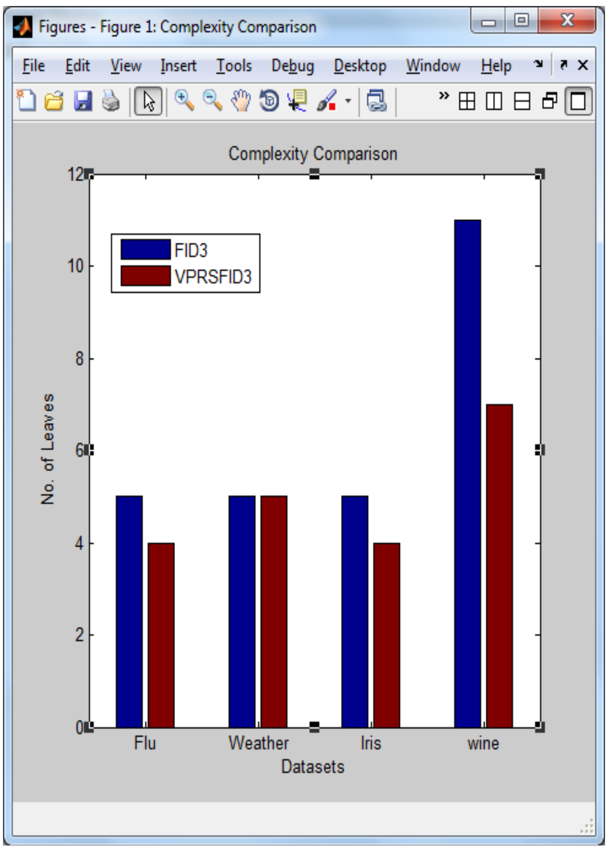


Figure4: The complexity comparison of FID3 and VPRSFID3

TABLE II. THE COMPARISON OF VPRSFID3 AND FID3

Datasets	Instances	Attributes	FID3		VPRSFID3	
			Acc (%)	Leaves	Acc (%)	Leaves
Flu	48	4	83.33%	5	91.66%	4
Weather nominal	64	5	84.20%	5	92.18%	5
Iris	150	5	66.50%	5	94.66%	4
Wine	178	14	74.70%	11	72.47%	7
Average			77.18%	6.5	87.74%	5

VI. CONCLUSIONS AND FUTURE WORKS

We proposed the concept of the enhanced information gain based on VPRS Model. This approach improved the FID3 algorithm and proposes a new attribute selection criterion. The hybrid method, however, cannot tolerate possible noises in real world datasets but it overcomes the drawback of FID3 algorithm. In this paper, we suggest an improved method (VPRSFID3) based on degree of β-dependency and significance of attributes based on variable precision rough set theory to select the splitting attributes. VPRSFID3 increases classification accuracy and reduces the size of the decision trees and thus enhance the generalization ability of the constructed decision trees. In the future, we will adapt VPRSFID3 for the application of constructing multivariate decision tree.

ACKNOWLEDGMENT

First and foremost I want to thank Prof Sandeep Raghuwanshi for his excellent guidance throughout my dissertation. I am deeply grateful to computer lab for providing me with an excellent work environment .It was a real pleasure to work with my colleagues. I extend a special word to my family member, particularly my fat her Shri Ramesh Babu Sharma for the moral and financial support and very special person without her blessing I can't progress, my mother Mrs Parvati Sharma. The enthusiasm showered on me from time to time and very special person without her help and cooperation I can't be completed my work, my wife Mrs Rinki Sharma.

REFERENCES

[1] R. Quinlan, “Induction of decision trees,” Machine Learning, vol.1, No. 1, pp. 81–106, 1986.
[2] Z. Pawlak, “Rough sets,” International Journal of Computer and Information Sciences, vol. 11, pp. 341–356, 1982.
[3] Z. Pawlak, “AI and intelligent industrial applications: the rough set perspective,” Cybernetics' and Systems, vol. 31, pp. 227–252, 2000.
[4] Ziarko, “Variable precision rough sets model,” Journal of Computer and System Sciences, vol. 46, pp. 39–59, 1993.
[5] Z. Pawlak, “Information systems theoretical foundations,” Information Systems, vol. 6, pp. 205–218, 1981
[6] Aijun An, Ning Shan, Christine Chan, et al, “Discovering rules for water demand prediction: An enhanced rough set approach,” Engineering Applications of Artificial Intelligence, vol. 9 (6) pp.645-653, 1996.
[7] X. P. Li, M. Dong, “An algorithm for constructing decision tree based on variable precision rough set model,” in Fourth International Conference on Natural Computation, pp. 280–283, 2008.
[8] Conference on Natural Computation, pp. 280–283, 2008.
[9] Conference on Computer Science and Information Technology, IEEE, pp. 295-299, 2008.
[10] Baoshi Ding, Yongqing Zheng, Shaoyu Zang, “A New Decision Tree Algorithm Based on Rough Set Theory,” Asia-Pacific Conference on Information Processing, IEEE, pp. 326-329,2009.
[11] 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, pp. 295-299, 2008.
[12] Zhang Li-Juan and Li Zhou-Jun,A Novel Rough Set Approach for Classification, 1999.
[13] 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.
[14] 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.
[15] www.ics.uci.edu/-mlearnlMLRepository.html