A NEW PRUNING APPROACH FOR BETTER AND COMPACT DECISION TREES

Ali Mirza Mahmood [1] Pavani Gudapati[2] Venu Gopal Kavuluru[3] Mrithyumjaya Rao Kuppa[4]
[1] Assistant Professor, DMSSVH College of Engineering, Machilipatnam, India.
[2] Student, PVP Sidhartha Instituite and Technology, Vijayawada, India.
[3] PVP Sidhartha Instituite and Technology, Vijayawada, India
[4]Professor, Vaagdevi College of Engineering, Warangal, India.

ABSTRACT:

The development of computer technology has enhanced the people's ability to produce and collect data. Data mining techniques can be effectively utilized for analyzing the data to discover hidden knowledge. One of the well known and efficient techniques is decision trees, due to easy understanding structural output. But they may not always be easy to understand due to very big structural output. To overcome this short coming pruning can be used as a key procedure .It removes overusing noisy, conflicting data, so as to have better generalization. However, In pruning the problem of how to make a trade-off between classification accuracy and tree size has not been well solved.

In this paper, firstly we propose a new pruning method aiming on both classification accuracy and tree size. Based upon the method, we introduce a simple decision tree pruning technique, and evaluated the hypothesis – Does our new pruning method yields Better and Compact decision trees? The experimental results are verified by using benchmark datasets from UCI machine learning repository. The results indicate that our new tree pruning method is a feasible way of pruning decision trees.

Keyword: Pre-Pruning, Post-Pruning, EBP, Laplace-Estimate.

I. INTRODUCTION

One of the research hotspot in the field of machine learning is classification. In [1], [2], [3], authors made recent improvements in decision trees. There are different types of classification models such as decision trees, SVM, neural networks, Bayesian belief networks, Genetic algorithm etc.. These above mentioned methods have provided satisfactory results, but still the most widely used classification models is the decision trees. The simple structure, the wide applicability on real time problems, the high efficiency and the high accuracy are the strengths for decision trees. Different dimensionality reduction techniques can be applied to the decision trees to improve their accuracy [4] The most common methods for creating decision trees are from data and rule, popularly known as Data-based decision tree and Rule-based decision tree respectively [5]. Decision tree is induced by Quinlan for inducing classification models [6].Decision tree induction is one of the most important branch of inductive learning, and it is one of the widely used practical method for inductive inference.

In decision tree induction the entire data in the training set is used as root node for the tree. Then the root node is split into several sub-nodes depending upon some heuristic function. Splitting of sub-node continues, till all leaf nodes are generated else if all the instances in the sub-node belong to the same class. The different variation of decision trees can be generated depending upon two main parameters, one is heuristic function used and the other is pruning method involved. The heuristic function used can be Gini index, Entropy, Information gain, Gain ratio and recently the large margin heuristic is proposed by Ning li et.al. [7].The most commonly used decision tree algorithms are ID3 [6] and C4.5 [8]. In ID3 the heuristic function used for splitting the data is Information Gain, which is the quality of information gained by partitioning the set of instances. The defect of this heuristic function is it has a strong bias in favor of the attributes with many outcomes. To solve this problem C4.5 uses another heuristic function, which penalizes the attribute that produces a wider distribution of data. This measure is commonly known as Gain Ratio.

This Paper is organized as follows. In section II, we discuss the related work. In section III, we presented the new pruning method based on Classification accuracy and tree size. In section IV, The methodology and datasets used for the experiments are introduced. In section V, We present the improvements of the new pruning criteria on classification accuracy and tree size and discuss the results. Section VI, closes the paper by presenting the conclusion and the direction for future research work.

II. RELATED WORK

Pruning is one of the most successful methods used in decision tree construction. The original work in pruning is proposed to tolerate noise in the training data [9],[10].In [11], [12], [13], the authors made through comparison of various pruning methods.

Two broad classes of methods are proposed for pruning.

Pre-pruning: Stop growing the tree earlier based on some stopping criteria, before it classifies the training set perfectly. One of the simplest method is setting a threshold for each sample when arriving the node; other method is to calculate the impact of system performance on each expansion and it is restricted if the gain is less than the threshold. In pre-pruning, the advantage is not generating full tree, disadvantage is horizon effect phenomenon [8].

Post-pruning: It has two major stages: Fitting and Simplification. First of all, it allows to over-fit the data, and then post-prunes the grown tree. In practice post-pruning methods has a better performance than pre-pruning. A lot of methods are presented based on different heuristics. In [9], the author proposed Minimal Cost Complexity Pruning (MCCP). Pessimistic Error Pruning (PEP), is proposed by J.R.Quinlan [8] which uses continuity correction for the binomial distribution to provide a more realistic error rate instead of the optimistic of error rate in training set.

In [8], the author proposed Error Based Pruning (EBP) which uses prediction of error rate (a revised version of PEP).In [10], the author proposed Reduced Error Pruning (REP), which finds the smallest version of the most accurate sub-tree but it tends to over-prune the tree. Recently in [14], the author proposed Cost and Structural complexity (CSC) pruning, which takes into account both classification accuracy and structural complexity.

Post-pruning can be further divided into two categories. One exploit the training set alone, other withhold a part of the training set for validation. Pessimistic Error Pruning (PEP), Error Based Pruning (EBP), comes under first category and Minimum Error Pruning (MEP), Critical value Pruning (CVP) comes under second.

III. Pruning Method Based On Classification Accuracy And Tree Size

In this section we present some necessary definitions and the main contribution of this work, in order to set the stage for the rest of the paper.

Let us see some definitions regarding decision tree representation.

An information system is a pair I S = (U; A; X; f) where U is a non – empty finite set of objects called universe. A denotes the set of attributes, it is usually divided into two subsets C and D, which denote the set of condition attributes and the set of decision attributes, respectively. f: U×A→X is an information function. A decision tree T = <V, E> is a directed connected acyclic graph induced by an information set I S = (U; A; X; f) Where V is the node set which encompasses both internal and external nodes. V is denoted by $V = V_{in} \cup V_{lf}$ .Where E ={<Vi, Vj>} is the set of directed edges, where vi€ V_{in} ,vj€ V_{lf} .The direction is from vi to Vj. Vi is the parent node of Vj. Vj is the child node of Vi. Except for the root node, each node has in-degree one. Each leaf node has out-degree zero.

New Pre-pruning Method:

In the tree growing phase, employing tightly stopping criteria tend to create small and under-fitted decision trees and employing a loosely stopping criteria tend to create over-fitted decision tree. To solve the trade-off between under-fitting and over-fitting of decision trees, an optimal pre-pruning and post-pruning have to be employed. As discussed in related work, the pre-pruning can be applied to the decision tree in the following ways [15],

1. To calculate the impact on each expansion of system performance, and it will not be extended if the gain is less than the threshold **s.**

2. Setting a threshold **m** for each sample when arriving the node. When the number of the samples is less than the threshold, the growth of decision tree will be stopped.

In pre-pruning the threshold s and m are used to measure only the size of the tree but not accuracy. Due to this, while pruning the accuracy of the decision trees decreases. Our main aim is by using pre-pruning we want to decrease the size of the tree, at the same time to increase the accuracy. This simple idea can be converted into a new pre-pruning technique by having optimal threshold for s and m. The formulation for the new pre-pruning is given as,

$$pre\_pruning = s \times m \qquad (1)$$

Where,
s = optimal threshold for each sample when arriving the node.
m = optimal no. of instances in each leaf node.

New Post-pruning Method:

The post-pruning is usually carried out after the decision tree is constructed. In post-pruning the efficiency of generating optimal decision tree is not implemented although it can achieve the purpose of knowledge rule simplification.

In C4.5, Error-based Pruning is implemented. Error-based pruning [8] is an evolution of the pessimistic pruning. As in pessimistic pruning the error rate is estimated using the upper bound of the statistical confidence interval for proportions

$$\varepsilon UB(T,S) = \varepsilon(T,S) + Z\alpha.\sqrt{\frac{\varepsilon(T,S)(1-\varepsilon(T,S))}{|S|}} \quad (2)$$

where $\varepsilon(T,S)$ denote the misclassification rate of the tree T on the training set S. Z is the inverse of the standard normal cumulative distribution and α is the desired significance level.

Let subtree (T,t) denote the sub tree rooted by the node t. Let maxchild(T,t) denote the most frequent child node of t .and let S_t denote all instances in S that reach the node t .The procedure performs bottom-up traversal over all nodes and compares the following values:

$\varepsilon UB(subtree(T,t), St)$
$\varepsilon UB(pruned(\ subtree(T,t),t), St)$
$\varepsilon UB(subtree(T, \max child(T,t)), S \max child(T,t))$

According to the lowest value the procedure either leaves the tree as is, prune the node t , or replaces the node t with the subtree rooted by maxchild(T,t).

To further simplify the knowledge rules – we have applied smoothing at the leaves. The Bayesian estimation is one of the popular techniques for smoothing and it can also be called as M-estimate. The M-estimate can be defined as,

$$Pi = \frac{Ni + m.p}{[\ \sum_{i \varepsilon C} Ni\ ] + m} \quad (3)$$

Where c is the number of classes. The probability p is the expected probability without any additional knowledge, and it is usually considered uniform, i.e. p=1/c. In M-estimate, one of the classical variation is Laplace-estimate. It is implemented by many author. The Laplace-estimate can be defined as,

$$Pi = \frac{Ni + 1}{[\ \sum_{i \varepsilon C} Ni\ ] + C} \quad (4)$$

Laplace-estimate, is a particular case of M-estimate where m = c. The Laplace-estimate was first introduced in machine learning by Niblett [16]. Clark and Bowell [17] implemented it into the CN2 rule learner. For decision tree learning the laplace-estimation has been used by certain researchers and practitioners [18][19].

We used this optimization technique along with the Error-Based Pruning (EBP). This effective idea can be converted into a new post-pruning technique. The formulation for the new post-pruning is given as,

$$Post\_pruning = EBP + \frac{Ni + 1}{\left[\sum_{i \varepsilon C} Ni\right] + C} \quad (5)$$

Based upon the above discussion, the new pruning technique can be applied to decision tree by implementing pre-pruning at the phase of inducing the trees and post-pruning after the decision trees have been induced.

IV. Experiments On Some Benchmark Datasets

In this section, we presented the evaluation of the proposed pruning method, we exploited C4.5 [8] to induce decision trees. We implemented the proposed pruning method in C4.5 and compared the generated decision trees with that obtained by the benchmark algorithms.

Table 1. Datasets.

Dataset	Tr-n	Te-n	Cn	An
Anneal.O	593	305	6	39
Audio	150	76	24	70
Balance	413	212	3	5
Breast	189	97	2	10
Diabetes	507	261	2	9
Glass	142	73	7	10
Heart	200	103	5	14
Hepatitis	103	52	2	20
Thyroid-h	2490	1282	4	30
Ionosphere	232	119	2	35
Iris	99	51	3	5
Labor	38	19	2	17
Lympho	98	50	4	19
Tumor	224	115	21	18
Sonar	138	70	2	61
Vehicle	559	287	4	19
Vote	288	147	2	17
Zoo	67	34	7	18

We are going to estimate the method presented by using 18 bench mark datasets that can be obtained from the UCI Machine learning repository. In the table 1, Tr-n, Te-n, Cn, An, indicate the number of training samples, test samples, classes and Conditional attributes respectively. The datasets we had included are of wide variety.

There are 5 datasets which are of large size – Anneal.O, Balance, Diabetes, Thyroid-h, and vehicle. There are 9 datasets which are of medium size - Audio, Breast, Glass, Heart, Hepatitis, Ionosphere, Tumor, Sonar, and Vote. There are 4 datasets which are of small size – iris, Labor, Lympho, and Zoo. The size of the datasets range from 67/34 to

2490/1282 each kind of data set has both training set and test set.

For example, the smallest data set 'zoo', the size of the training set is 67, the size of the test set is 34. The largest data set 'Thyroid', the size of the training set is 2490, the size of the test set is 1282.We run our new pruning technique and other benchmark algorithms with a 66/33% train/test split on each dataset and repeated 20 times. The mean values of tree size, classification accuracy and number of leaves are taken from 20 runs.

V. Results And Discussion

We present some of the results of our empirical studies of the New Pruning on the datasets from UCI Machine Learning Database repository [20].

We are interested in verifying the following two hypotheses:

1. Does our new pruning technique yields compact decision trees with small size than other benchmark algorithms?

2. Does our new pruning technique will generate better decision trees with improved accuracy than other benchmark algorithms?

We have conducted several experiments to verify that the threshold s and m are used not only to decreases the size but also to increase the accuracy. i.e. we want to have optimal s and m values. According to our investigation we have concluded that the optimal pre-pruning values are all concentrated around a small area where s and m are 0.41 and 5 respectively. In post-pruning we set up a laplace estimate along with Error-based pruning. The detail experimental results for hypothesis of the optimal values for s and m will be given some where else.

In the tables, the column C4.5/REP/CART, New Algor. and Concl. Indicates the Size of the tree /Accuracy of C4.5/REP/CART, Size of the tree/Accuracy with New Pruning and the column entitled "Concl." indicate the Conclusion of the evaluation; while the symbols are "√" or " ≈ " are used to denote whether our algorithm is better(small size/ more accuracy) or as best (same size/ same accuracy); while the symbol " × " implies whether our algorithm is worse (large size/ less accuracy) than the compared algorithm.

A. Performance Comparison in Tree size and Classification accuracy between C4.5 and New algorithm.

Table 2.Tree size for C4.5 and New Algor.

Dataset	C4.5	New Algor.	Concl.
Anneal.O	55.4	49.75	√
Audio	41.9	23.1	√
Balance	62.2	40.4	√
Breast	16.4	13.2	√
Diabetes	35.7	31.1	√
Glass	35.2	20.8	√
Heart	29.5	20.3	√
Hepatitis	11.9	7.9	√
Thyroid-h	21.8	15.25	√
Ionosphere	19.3	14.10	√
Iris	6.7	5.5	√
Labor	5.3	3.7	√
Lympho	19.6	10.40	√
Tumor	67.9	36.05	√
Sonar	22.2	16.5	√
Vehicle	92.4	62.6	√
Vote	9.7	7.1	√
Zoo	13.5	10.9	√

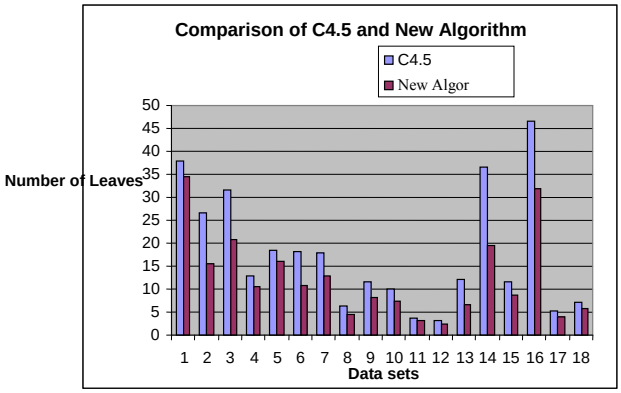
Table 3. Accuracy for C4.5 and New Algor.

Dataset	C4.5	New Algor.	Concl.
Anneal.O	90.30	88.82	×
Audiology	77.90	70.56	×
Balance	78.19	78.78	≈
Breast	71.23	70.77	≈
Diabetes	73.46	73.50	≈
Glass	67.32	66.84	≈
Heart	78.29	77.44	≈
Hepatitis	78.14	78.23	≈
Thyroid-h	99.49	99.39	≈
Ionosphere	88.80	88.50	≈
Iris	94.12	93.82	≈
Labor	79.38	78.71	≈
Lympho	76.68	77.79	√
Tumor	39.10	38.55	≈
Sonar	71.12	70.48	≈
Vehicle	70.31	69.56	≈
Vote	95.54	94.89	≈
Zoo	93.26	89.49	×

In our empirical studies, we have conducted experiments to evaluate Tree size, Classification accuracy, Number of leaves, Training errors on datasets, Test errors on datasets for new pruning technique and benchmark algorithms.

However due to the shortage of the space we are not able to present all the results. We tried to accommodate as many results as possible. Our main aim is to evaluate the hypothesis about Better and Compact decision trees so; we have presented the results of tree size and accuracy.

Figure 1. Number of Leaves in C4.5 and New Algor.



From Table 2 we can see that for all datasets, the new pruning technique does yield compact decision trees. Therefore our strategy of optimal threshold's' and optimal number of instances in each leaf m have worked in decreases the size of the tree.

In Figure 1, the results of the number of leaves of C4.5 and New Algorithm is shown, we can see that the generalization accuracy of new algorithm is outperformed over C4.5

Table 3 shows the mean classification accuracy scores for both C4.5 and new pruning method on each of the 18 test problems. Each entry is an average of 20 trails with 66% train test split. As expected, we can see that the classification accuracy for new pruning algorithm is almost equal to original C4.5.

The classification accuracy of 15 out of 18 datasets – Balance, Breast, Diabetes, Glass, Heart, Hepatitis, Thyroid-h, Ionosphere, Iris, Labor, Lympho, Tumor, Sonar, Vehicle and Vote are near or superior to C4.5, and however 1% upgrade or degrade in classification accuracy in general is not considered as a change, and where in many literatures it is neglected therefore we considered 1% upgrade or degrade as similar.

B.Performance Comparision in Tree size and Classification accuracy between REP and New algorithm.

Table 4.Tree size for REP and New Algor.

Dataset	REP	New Algor.	Concl.
Anneal.O	53.45	49.75	√
Audio	28.50	23.1	√
Balance	34.80	40.4	×
Breast	23.65	13.2	√
Diabetes	35.7	31.1	√
Glass	22.50	20.8	√
Heart	14.90	20.3	×
Hepatitis	5.40	7.9	×
Thyroid-h	16.95	15.25	√
Ionosphere	7.20	14.10	×
Iris	5.50	5.5	√
Labor	4.90	3.7	√
Lympho	10.40	10.40	≈
Tumor	27.30	36.05	×
Sonar	7.0	16.5	×
Vehicle	45.20	62.6	×
Vote	6.90	7.1	≈
Zoo	1.0	10.9	×

Table 5. Accuracy for REP and New Algor.

Dataset	REP	New Algor.	Concl.
Anneal.O	90.61	88.82	×
Audiology	71.86	70.56	≈
Balance	77.63	78.78	√
Breast	66.96	70.77	√
Diabetes	74.82	73.50	≈
Glass	63.57	66.84	√
Heart	74.12	77.44	√
Hepatitis	79.76	78.23	≈
Thyroid-h	99.33	99.39	≈
Ionosphere	89.22	88.50	≈
Iris	94.90	93.82	×
Labor	77.81	78.71	√
Lympho	72.82	77.79	√
Tumor	37.80	38.55	≈
Sonar	68.77	70.48	√
Vehicle	68.91	69.56	≈
Vote	95.10	94.89	≈
Zoo	40.54	89.49	√

The Classification accuracy of remaining 3 out of 18 datasets - Anneal.O, Audiology and Zoo has degraded. The overall experiments with C4.5 verify explicitly that our new pruning technique is practically effective and scale well on both tree size and accuracy.

Quinlan [10] has suggested a simple procedure for pruning decision trees known as Reduced-Error-Pruning. It has been shown that this procedure ends with the smallest accurate sub tree with respect to a given pruning set. To realize the goodness of our new pruning technique, we designed experiments with Reduced-Error-Pruning [REP].

The basic tree building process is same as discussed in section IV.

From table 2 and 4 we can see that REP has reduced the size of the trees further more than C4.5 and we can notice that the new pruning method have given a good competition to REP. In table 4, there exits Ten cases out of eighteen where the size of the tree is best or better than REP. In the remaining eight datasets the tree size is increased. For the dataset Glass the degradation is around 2% only.

In table 5, we may firstly notice that most of the datasets classification accuracy is increased. There are sixteen out of eighteen datasets where the accuracy obtained is similar or significant. However, only two datasets – Anneal.O and Iris where the accuracy is degraded.

From the above results we hypotheses that our algorithms performance is remarkable not only on C4.5 but also on REP which is one of the most compact tree generator

Breiman et al [9], implemented Minimal Cost Complexity Pruning (also known as weakest link pruning or error complexity pruning) in CART. This algorithm is also one of the benchmark algorithm in decision trees.

C. Performance Comparison in Tree size and Classification accuracy between CART and New algorithm.

Table 6.Tree size for CART and New Algor.

Dataset	CART	New Algor.	Concl.
Anneal.O	76.10	49.75	√
Audio	26.30	23.1	√
Balance	44.60	40.4	√
Breast	5.0	13.2	×
Diabetes	16.40	31.1	×
Glass	21.60	20.8	≈
Heart	12.50	20.3	×
Hepatitis	7.60	7.9	≈
Thyroid-h	16.80	15.25	√
Ionosphere	8.30	14.10	×
Iris	5.80	5.5	≈
Labor	7.20	3.7	√
Lympho	10.70	10.40	≈
Tumor	28.20	36.05	×
Sonar	11.10	16.5	×
Vehicle	50.60	62.6	×
Vote	6.60	7.1	≈
Zoo	1.0	10.9	×

Table 7. Accuracy for CART and New Algor.

Dataset	CART	New Algor.	Concl.
Anneal.O	91.82	88.82	×
Audiology	72.96	70.56	×
Balance	78.40	78.78	≈
Breast	69.07	70.77	√
Diabetes	74.88	73.50	≈
Glass	68.91	66.84	×
Heart	78.02	77.44	≈
Hepatitis	78.99	78.23	≈
Thyroid-h	99.54	99.39	≈
Ionosphere	88.63	88.50	≈
Iris	94.61	93.82	×
Labor	81.24	78.71	×
Lympho	77.19	77.79	≈
Tumor	41.13	38.55	×
Sonar	71.26	70.48	≈
Vehicle	69.60	69.56	≈
Vote	95.06	94.89	≈
Zoo	40.54	89.49	√

We have evaluated our new algorithm with CART and the results are given in table 6 and 7. Let us have a look at table 6, the results indicate that our new algorithm have performed equally well on ten out of eighteen datasets. In these ten datasets there are four datasets where decrease in the tree size is well above 4% indicating effectiveness of our algorithm, at the same time there are eight datasets where our algorithm is degraded

Let us show in table 7, the results of the classification accuracy for CART and new algorithm. The last column "Concl." Indicates that for our new algorithm there are well above 66% datasets where the accuracy is near or superior to CART. These evaluation results from table 6 and 7 show that our new algorithm is practically applicable to obtain better results than CART.

To find the minute differences between our new algorithm and other methods we would like to run more trails on some more different datasets, but fortunately with only 20 trails and 18 datasets we are able to discern interesting differences between the methods.

Despite the work that remains to be done we believe that our initial studies have reveled interesting insights into the relative abilities of our new algorithm. Finally, according to the nature and number of the datasets, and the quantity and quality of work developed for improving decision trees, we think that this is a significant achievement.

VI. CONCLUSION

Even though decision trees are one of the most common data mining and machine learning methodologies, they may not always be easy to understand due to very big structural output.

The implementation of optimal trade-off between classification accuracy and tree size can solve the above problem.

In this paper, we introduces a pruning decision tree method which takes into account both classification accuracy and tree size. In pre-pruning we take advantage of threshold s, and minimum no. of instances in each leaf m, and In post-pruning we embedded Laplace-estimate along with EBP. The experimental results on benchmark datasets from UCI machine learning repository shown that our new algorithm have produced better and compact decision trees when compared to benchmark algorithms.

In the future work we will investigate on the same pruning method about how to provide users more flexibility in incorporating domain specific knowledge in decision trees to obtain optimal results in prediction.

ACKNOWLEDGMENT:

The authors would like to thank Joan Albert López-Vallverdú, David Riaño and Antoni Collado for their suggestions and help during the project. The authors would also like to thank UCI repository of machine learning databases. Special thanks also to Liangxiao Jiang, Chaoqun Li, Zhihua Cai for being in touch in mail through out the implementation.

REFERENCES:

[1] Ali Mirza Mahmood, K.Mrutunjaya Rao, Kiran Kumar Reddi(2010) A Novel Algorithm for Scaling up the Accuracy of Decision Trees, (IJCSE) International Journal on Computer Science and Engineering Vol. 02, No. 02, 2010, Page no:126-131. http://www.enggjournals.com/ijcse/issue.html?issue=20100202.

[2] Chen Jin, Luo De-lin, Mu Fen-xiang,(2009) An Improved ID3 Decision Tree Algorithm, Proceedings of 2009 4th International Conference on Computer Science & Education

[3] Smith Tsang, Ben Kao, Kevin Y. Yip, Wai-Shing Ho, Sau Dan Lee(2009) Decision Trees for certain Data, IEEE International Conference on Data Engineering.

[4] Kiran Kumar Reddi, Ali Mirza Mahmood, K.Mrithyumjaya Rao (2010), Generating Optimized Decision Tree Based on Discrete Wavelet Transform, (IJEST) International Journal of Engineering Science and Technology Vol.2(3), 2010,Page no:157-164

[5] Amany Abdelhalim, Issa Traore(2009) A New Method for Learning Decision Trees from Rules, Proceedings of International Conference on Machine Learning and Applications, 2009.

[6] Quinlan J R.(1986) "Induction of decision tree", Machine Learning, 1986, 1: 81~106.

[7] Ning Li, Li Zhao, Ai-Xia Chen, Qing-Wu Meng, Guo-Fang Zhang(2009) A New Heuristic Of The Decision Tree Induction , Proceedings of the Eighth International Conference on Machine Learning and Cybernetics, Baoding, 12-15 July 2009

[8] Quinlan J R.(1993) C4.5: Programs for machine learning [M]. California: Morgan Kaufmann Publishers, Inc,1993.

[9] L. Breiman, J. Friedman, R. olshan, and C.Stone (1984), Classification and Regression trees ,California,Wadsworth international 1984.

[10] Quinlan J R.(1987) , Simplifying decision trees, International journal of Man-Machine studies, Vol 27,pp 221-234.

[11] Floriana Esposito, Donato Malerba, (1997) A comparative analysis of methods for Pruning decision trees, IEEE Transcations on pattern analysis and machine intelligence, Vol 19, no 5,

[12] L.A. Breslow , D.W.Aha "Simplifying decision trees: A Survey", Knowledge engineering review, vol 12, no.1, pp 1-40, 1997.

[13] Wang Xizhao , You Ziying, "A Brief survey of methods for Decision tree Simplification." Computer Engineering and Applications. Vol 40, No.27, pp.66-69 , 2004

[14] Jin-Mao Wei,Shu Qin Wang, Gang Yu,Li Gu, Guo-Ying Wang, Xiao-Jie Yuan (2009), A Novel method for pruning decision tree, Proceedings of the Eighth International Conference on Machine Learning and Cybernetics, Baoding, 12-15 July 2009.

[15] Juanli Hu, Jiabin Deng, Mingxiang Sui(2009) A New Approach for Decision Tree Based on Principal Component Analysis. IEEE Transcation.

[16] Niblett,T.(1987).Constructing decision trees in noisy domains. Proceedings of the Second European Working Session on Learning (pp.67–78).Wilmslow,England:SigmaPress.

[17] Clark,P.,&Boswell,R.(1991).Rule induction with CN2:Some recent improvements. Proceedings of the Sixth European Working Sessionon Learning (pp.151– 163). Berlin:Springer.

[18] Pazzani,M.,Merz,C.,Murphy,P.,Ali,K.,Hume, T.,&Brunk,C.(1994).Reducing misclassification costs Proceedings of the Eleventh International Conference on Machine Learning (pp.217–225).San Francisco: Morgan Kaufmann.

[19] Bradford,J.P.,Kunz,C.,Kohavi,R.,Brunk,C.,& Brodley,C.E.(1998). Pruning decision trees with misclassification costs. Proceedings of the Tenth European Conferenceon Machine Learning (pp.131–136).Berlin:SpringerVerlag. [20] Blake, C., & Merz, C. J. (2000). UCI Repository of machine learning databases Machine-readable data repository, Department of Information and Computer Science, University of California at Irvine, Irvine, CA. Available at http://www.ics.uci.edu/mlearn/MLRepository.html

Ali Mirza Mahmood is Pursuing his PhD in Computer Science and Engineering at Acharya Nagarjuna University under the guidance of Dr. K. Mrithyumjaya Rao. He received his Masters degree in Computer Science in 2003. Now, he is an Assistant Professor in department of Computer Science in DMSSVH College of engineering, Machilipatnam (India). His current research interest includes Data Mining and Knowledge Discovery, Machine Learning, and Artificial Intelligence.

Pavani Gudapati is Pursuing his M.Tech in Computer Science and Engineering at Jawaharlal Nehru Technological

University, Kakinada. She received her B.Tech in Computer Science in 2005. Her current research interest includes Data Mining and Knowledge Discovery, Machine Learning, and Artificial Intelligence.

Venu Gopal Kavuluru,(phd) M.Tech (cse), is Pursuing his PhD in Computer Science and Engineering at Andhra University, VisakhaPatnam , India . He is working as Assistant Professor in department of Computer Science in P.V.P.Siddhartha

Institute of Technology His current research interest includes Algorithms, image processing, web services

K. Mrithyumjaya Rao received a Ph.D. degree from Kakatiya University in 1979. Now, he is a professor in Faculty of Computer Science and Engineering in Vaagdevi College of engineering, Warangal (India). His current research interest includes data mining techniques with applied to real world problems.