STUDY AND ANALYSIS OF DECISION TREE BASED CLASSIFICATION ALGORITHMS

K. krishnaveni[1]

[1]Fatima College, India

ABSTRACT

Machine learning is to learn machine on the basis of various training , testing data and determines the results in every condition without explicit programmed. One of the techniques of machine learning is Decision Tree. Different fields used Decision Tree algorithms , used it in their respective application. These algorithms can be used as to find data in replacement statistical procedures, to extract text, medical certified fields and also in search engines. Different Decision tree algorithms have been built according to their accuracy and cost of effectiveness. To use the best algorithm in every situations of decision making is very important for us to know. This paper includes three different algorithms of Decision Tree which are ID3, C4.5 and CART.

1. INTRODUCTION

Decision Tree

A normal tree includes root, branches and leaves. The same structure is followed in Decision Tree. It contains root node, branches, and leaf nodes. Testing an attribute is on every internal node, the outcome of the test is on branch and class label as a result is on leaf node [3, 4]. A root node is parent of all nodes and as the name suggests it is the topmost node in Tree. A decision tree is a tree where each node shows a feature , each link shows a decision (rule) and each leaf shows an outcome . As decision trees mimic the human level thinking interpretations.

so it's so simple to grab the data and make some good interpretations. The whole idea is to create a tree like this for the entire data and process a single outcome at every leaf.

so it's so simple to grab the data and make some good

2. RELATED WORK ON DECISION TREE

Decision Tree is similar to the human decision-making process and so that it is easy to understand. It can solve in both situations whether one has discrete or continuous data as input. Talking about the characteristics of Decision Tree, the ID3 algorithm is simulated only on WEKA tool and the data type of data set is only categorical. ID3 can not take continuous data set for simulation. Similarly, CART and C4.5 have same characteristics as ID3 has. The only difference is that C4.5 and CART both can take continuous data set as input for simulation purpose .The decision tree makes explicit all possible alternatives and traces each alternative to its conclusion in a single view, to make easy comparison among the various alternatives. Transparent in nature is one of the best advantages of Decision Tree. Another main advantage is the ability to selecting the most biased feature and comprehensibility nature. It is also easy to classify and Interpretable easily. Also used for both continuous and discrete data sets. Variable screening and feature section are good enough in decision tree [19]. By talking on its performance, non-linear does not affect any of the parameters of the decision tree.

at any node to determine whether splitting is "Best" i

3. EVALUATION MECHANISM

If the values are close to each other, the set can be said to be precise. If their average is close to the true value of the quantity being measured, the set can be said to be accurate. Only if given a set of data points from repeated measurements of the same quantity then one can measure

above two terms.

$$ACCURACY = \frac{TP + TN}{(TP + TN + FP + FN)}$$

$$PRECISION = \frac{TP}{TP + FP}$$

□□□□□ □
□□□ □ □

TP = True positive, TN = True Negative

FP = False Positive, FN = False Negative

4. CONCLUSION

The Decision Tree algorithms ID3 C4.5 and CART were applied on the dataset. Decision tree outperforms others in terms of accuracy, time and precision. It quite relies on the algorithm used for recommendation to find interesting resources. At last, the comprehensive study is done about decision tree algorithms and this paper concludes that CART is the algorithm for this dataset is very precise and most

accurate among the others.

5. FUTURE WORK

In the future, this will be installed in the Apache server thus published on the internet. Datasets are updated continuously and it will take online rating for the prediction. The prediction approaches can also be tried in different datasets to check the performance of the system.

6. REFRERENCES

[1] Sorower MS. A literature survey on algorithms for multi-label learning. Oregon State University, Corvallis. 2010 Dec;18.

[2] Utku A, Hacer (Uke) Karacan, Yildiz O, Akcayol MA. Implementation of a New Recommendation System Based on Decision Tree Using Implicit Relevance Feedback. JSW. 2015