ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 2, Issue 2, March 2013

Facilitating Decision Support through Decision Tree

Pankaj Pathak, Parashuram Pal
PhD Scholar (Pacific University Udaipur, Raj), Professor (LNCT Bhopal, MP)

***Abstract*— A decision tree is a simple representation for classifying examples. Decision tree learning is the most successful technique for classification learning. Each element of the domain of the classification is called a class. In the context of classification problems, decision tree applies decision tests based on a combination of attributes. In this paper we presented the overview of decision tree and how the classification is performed in the tree with the help of example. The important technique for the extraction of information from a table is splitting. So further we study the binary and multiway splitting. When we got our target node we have to stop the growth of decision tree, for which well specified criteria for stopping the tree should be present. The reasons for popularity of decision tree are also explained. The big issue for representation of a decision tree is to determine which attribute is going to be the first chosen. The representation and enhancement of decision tree induction is also described in the next section.**

***Index Terms*—Attribute, Classification, Decision Tree, Induction.**

I. INTRODUCTION

The use of data for data mining techniques is very important for data mining practitioners. The data which is used for particular applications responsible for taking the decision. Data mining is used to distinguish or classify the decisions based on available information like who can get a bank loan, who can get a special offer, or who is the most eligible candidates for the certain designations. The point is that data mining is just a tool in the whole decision making process which is used by people for obtaining the results as well as applying the further knowledge and decide what action to be taken. The data is only the facts and it becomes the information which can be the set of patterns after applying certain operations. Knowledge can be the aggregation of set of patterns or expectations on the basis of which the decisions are made. There are different styles of learning are used in data mining applications[1]. The first step in learning is the classification in which the previously classified examples are presented to draw the conclusions. And it is expected to classify the unseen examples based on the drawn conclusions. The whole process of classification, drawing conclusions and predicting the decision for the new situations can be implemented in the form of a system which can be named as a decision support system. The outcome of classification is getting a class on which decisions are based. The success of the outcome of the classification is to be judged by the test data and then to be applied on real world situation. The success rate of outcome of the classification process determines the accuracy of the rules which are used in classification process.

II. DECISION TREE- HIERARCHICAL MULTI-ATTRIBUTE STRUCTURE

Decision trees are one of the popularly used technique in data mining scenarios for the classifying data or building decision models to extract useful information, and it is widely applied in a variety of contexts such as market analysis, fraud detection in banking and financial sector, medical diagnosis, and geographic studies in transportation planning[1], [2]. A decision tree is a classified form of nodes which makes a rooted tree. In a decision tree the internal nodes are successively splits into two or more nodes according to a certain discrete function which is based on the values of input attributes. The splitting process continues until leaf node is not reached. Each leaf node represents one class of appropriate target value. **Classification** assumes that a set of objects—characterized by some attributes or features—belong to different classes [3]. The class label is a discrete qualitative identifier; for example, large, medium, or small. The objective is to build classification models that assign the correct class to previously unseen and unlabeled objects. Classification models are mostly used for predictive modeling. Discriminant analysis, decision tree, rule induction methods To construct a decision tree for each outcome class, the original instances in the training data set are categorized into two revised classes: yes (*Y)* and no (*N*). If an instance has a decision outcome that belongs to the outcome class for which the tree has been constructed, it is assigned to the revised class *Y*. Otherwise, and it is assigned to the revised class N**.**

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 2, Issue 2, March 2013

a) One of the examples of a decision tree is as under mentioned.

Table 1: Raw data for decision tree

Age	Gender	Credit_Rating	buys_car
<=30	Male	good	no
<=30	Male	excellent	yes
30…40	Male	good	yes
>40	Male	good	yes
>40	Female	good	no
>40	Female	excellent	no
31…40	Female	excellent	yes
<=30	Male	good	no
<=30	Female	good	no
>40	Female	good	yes
<=30	Female	excellent	yes
31…40	Male	excellent	yes
31…40	Female	good	yes
>40	Male	excellent	no

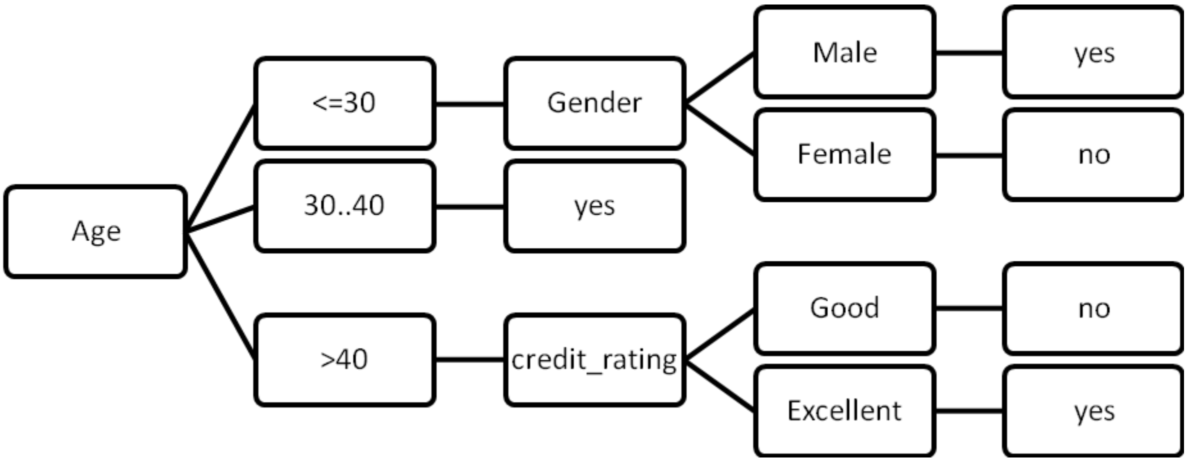

Fig 1: Hierarchical structure to predict the potential customer for buying car

Figure 1 describes a decision tree that infers whether or not a potential customer will respond to a direct mailing for buying a car. The mentioned decision tree having both nominal and numeric attributes. With the help of given classifier, the analyst can predict the potential customer (by sorting it down the tree), and understand the characteristics of the entire potential customers population regarding prospects of buying a car. Each node is labeled with the attribute it tests, and its branches are labeled with its corresponding values[3]. Decision-makers prefer less complex decision trees, since they may be considered more comprehensible. Furthermore, according to Breiman *et al*. (1984) the tree complexity has a crucial effect on its accuracy. The tree complexity is explicitly controlled by the stopping criteria used and the pruning method employed. Usually the tree complexity is measured by one of the following metrics: the total number of nodes, total number of leaves, tree depth and number of attributes used. Decision tree induction is closely related to rule induction. Each path from the root of a decision tree to one of its leaves can be transformed into a rule simply by conjoining the tests along the path to form the antecedent part, and taking the leaf's class prediction as the class value. For example, one of the paths in Figure 1 can be transformed into the rule: "If customer age is less than or equal to or equal to 30, and the gender of the customer is "Male" – then the customer is the prospect of buying a car and will respond to the mail".

III. SPLITTING CRITERIA

The important technique for the extraction of information from a table is splitting. A splitting gives mutually exclusive, collectively exhaustive set of subsets (called "components"). One possible basis for the partitioning of a table is the value of one or more attributes. In database systems such a partition is often realized as an index, i.e. a table that maps from the attribute value to the tuples of the partition component. A common reason to implement an index is to provide a fast access path to components of the partition [3].The binary criteria are used for creating binary decision trees. These measures are based on division of the input attribute domain into two parts. The obtained value for the optimal division of the attribute domain into two mutually exclusive and exhaustive sub-domains is used for comparing attributes [4],[5]. In multivariate splitting criteria, several attributes may participate in a single node split test. Obviously, finding the best multivariate criteria is more complicated than finding the best univariate split**.** Furthermore, although this type of criteria may dramatically improve the tree's performance, these criteria are much less popular than the univariate criteria.

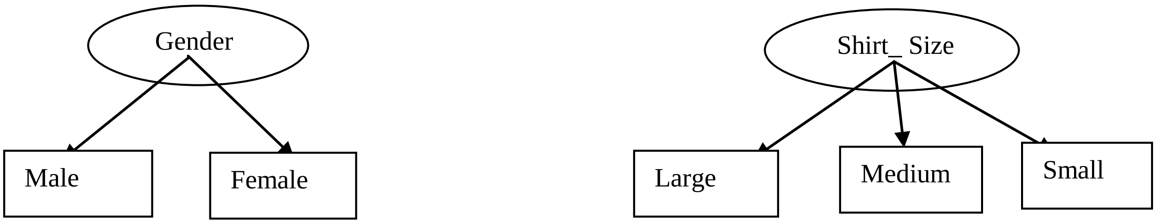

Fig 2: Multiway verses Binary Split

IV. STOPPING CRITERION FOR HIERARCHICAL PREDICTIVE MODEL

For the hierarchical multi-attribute predictive model the stopping rules usually terminate the hierarchical structure. This process is continued along each branch of the tree until a stopping condition is satisfied. The following conditions are common stopping rules:

1. All instances in the training set belong to a single value of *y*.
2. The maximum tree depth has been reached.
3. The number of cases in the terminal node is less than the minimum number of cases for parent nodes.
4. If the node were split, the number of cases in one or more child nodes would be less than the minimum number of cases for child nodes.
5. The best splitting criteria is not greater than a certain threshold.

V. WHY DECISION TREE IS POPULAR

Several advantages of the decision tree as a classification tool have been identified so that it becomes a popular technique:

1. Decision tree is the important technique of data mining which very easy to understand and self-explanatory [6]. If the decision tree has a less number of leaves, it can be comprehended by non–professional users. Decision trees can be converted to a set of rules. So, this representation is considered as most comprehensible.
2. Decision trees can handle both nominal and numeric input attributes.
3. Decision tree representation is rich enough to represent any discrete–value classifier.
4. Decision trees are capable of handling large datasets that may have errors.
5. Decision trees are capable of handling large datasets that may have missing values.
6. Relatively faster learning speed (than other classification methods).
7. Decision trees can use SQL queries for accessing databases.
8. Comparable classification accuracy with other methods.

VI. REPRESENTATION OF DECISION TREE

The representation of a decision tree can be made as follows:

1. Select an attribute.
2. Extend the tree by adding a branch for each value of this attribute.
3. Show the examples in the leaves.
4. For each leaf, if all examples are from the same class, add this class to the leaf. If not, repeat steps 1 to 4.

The big issue for representation of a decision tree is to determine which attribute is going to be the first chosen. We must choose first the ones that have the best information [7], [9]. Decision trees use the concept of entropy to test how much information has an attribute. From the information theory field, we can define entropy (Claude Shannon) as a measure of randomness of a variable. With the help of this concept it is possible to measure whether an attribute is explicitly or not a good one. Information, expressed as a mathematical quantity, reflects this. For example, consider a very simple classification problem that requires creating a decision tree to decide yes or no based on some data. This is exactly the scenario visualized in the decision tree. Each attributes values will have a

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 2, Issue 2, March 2013
certain number of yes or no classifications. If there are equal numbers of yeses and no's, then there is a great deal of entropy in that value. In this situation, information reaches a maximum [8]. Conversely, if there are only yeses or only no's the information is also zero. The entropy is low, and the attribute value is very useful for making a decision. The formula for calculating intermediate values is as follows:

$$Info = -\sum_{i=1}^{m} p_1 log_2 p_1 \ldots$$

Let's understand this with the example. Considering calculating the information gain for three variables for choosing one splitting attribute. The attributes as a whole has the total of nine yeses and five no's. The first variable has two yeses and three no's. The second has four yeses and zero no's. The third variable has three yeses and two no's. Our first step is to calculate the information for each of the variables.

Starting with the first variable, the formula says us that

info ([2,3]) being equal to -2/5 x log 2/5 - 3/5 x log 3/5.

This comes to 0.971

The second variable is easy to calculate. It only has yeses, so it has a value of 0

The third variable is just the reverse of the first -- the value is also 0.971.

Having the information for all the variables, we need to calculate the information for the attribute as a whole: 9 yeses and 5 no's. The calculation is-

info ([9,5])= -9/14 x log 9/14 - 5/14 x log 5/14.

This comes to 0.940

In decision tree induction, our objective is to find the overall information gain. This is found by averaging the information value of the attribute values. In our case, this is equivalent to finding the information of all the attributes together. We would use the formula-

info([2,3],[4,0],[3,2]) = (5/14)x0.971+(4/14)x0+(5/14)x0.971.

This comes to 0.6931

The last step is to calculate the overall information gain. Information gain is obtained by subtracting the information value average by the raw total information of the attribute. We calculate-

gain = info([9,5]) - info([2,3],[4,0],[3,2]) = 0.940 - 0.693 = 0.247.

The decision tree induction algorithm will compute this sum for every attribute, and select the one with the highest information gain as the root node, and continue the calculation recursively until the data is completely classified. This approach is one of the fundamental techniques used for decision tree forming. It has a number of possible lacunas. One common problem arises when an attribute has a large number of unique values. An example of this could be employee_id, or other types of identification numbers. In that case, there is an artificially high decision-value to the information -- the ID classifies each employee, and distorts the tree induction process by over fitting the data. One solution is to use an information gain ratio that biases attributes with large numbers of unique values. For example, an attribute can be windy and other attributes are sunny, rainy and cloudy. After the calculation of the windy attribute, the tree should also perform the calculation for the other attributes. The attribute which the GAIN info value shows to better information will be used first.

VII. ENHANCEMENT TO BASIC DECISION TREE

We can move to build a complex decision tree from the basic decision tree by adopting following concepts:

- We can allow for continuous-valued attributes and dynamically define new discrete-valued attributes that partition the continuous attribute value into a discrete set of intervals [6], [10].
- In large databases several attributes values can be missing so we can handle missing attribute values by assigning the most common value of the attribute and by assigning probability to each of the possible values of the attributes.
- By constructing new attributes based on existing ones that are sparsely represented, so it will reduces fragmentation, repetition, and replication.

VIII. CONCLUSION

This paper has introduced a most popular data mining technique i.e. decision tree. Further, the technique enables the classification which is used to from predictive models. Decision tree induction is also explained with the help of an example. The paper shows the splitting criteria which are based on attribute values. The rules also have to be identifying for stopping the hierarchical multi-attribute structure. The reasons for popularity of decision tree are also explained. The big issue while using decision tree for decision making is to determine which attribute is

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 2, Issue 2, March 2013

going to be the first chosen. For selecting the best splitting attribute one measure is used which is called information gain. It is explained by an example.

REFERENCES

[1] Data Mining Practical Machine Learning Tools and Techniques, Third Edition by Ian H. Written.

[2] Agrawal, R., Imielinski, T., & Swami, A. (1993a). Database mining: A performance perspective. IEEE Transactions on Knowledge and Data Engineering, 5(6), 914–925.

[3] Breiman L., Friedman J., Olshen R., and Stone C. Classification and Regression Trees. Wadsworth Int. Group, 1984.

[4] Brodley C. E. and Utgoff. P. E., Multivariate decision trees. Machine Learning, 19:45-77, 1995.

[5] Buntine W., Niblett T., A Further Comparison of Splitting Rules for Decision- Tree Induction. Machine Learning, 8: 75-85, 1992.

[6] Han, J., & Kamber, M. (2006). Data mining: Concepts and techniques (2nd Ed.). San Francisco: Morgan Kaufmann.

[7] Decision-Tree-Based Knowledge Discovery: Single- vs. Multi-Decision-Tree Induction by Namsik Chang et. El, INFORMS Journal on Computing Vol. 20, No. 1, winter 2008, pp. 46–54.

[8] http://www.cse.unsw.edu.au/~billw/cs9414/notes/ml/06prop/id3/id3.html.

[9] Witten IH, Frank E. Data Mining: Practical Machine Learning Tools and Techniques. Second edition, 2005. Morgan Kaufmann.

[10] V. Chercassky and F. Mulier, Learning from Data: Concepts, theory and methods. Wiley Interscience 1998.

AUTHOR BIOGRAPHY

Mr. Pankaj Pathak received Bachelor of Science (Computers) in 2003 from Vikram University. He received his Master's Degree in Information Technology. He joined ICFAI National College in 2006 as a faculty of IT. Now currently doing his PhD in Comp. Sc. From Pacific University Udaipur. He is doing his research in the field of Data Mining.

Dr. Parashuram Pal has received his Masters in Computer Application and PhD in Computers. He has worked with Shri Vaishnav Institute of Management, Indore and Ajay Kumar Garg Engineering College, Ghaziabad as a Professor. Now he is working with LNCT, Bhopal. He has done his research in the field of Data Mining.