D

Decision Trees

John Wang
Montclair State University, USA

Dajin Wang
Montclair State University, USA

INTRODUCTION

Decision trees are part of the decision theory and are excellent tools in the decision-making process. Majority of decision tree learning methods were developed within the last 30 years by scholars like Quinlan, Mitchell, and Breiman, just to name a few (Ozgulbas & Koyuncugil, 2006). There are a number of methods and sophisticated software used to graphically present decision trees. Decision trees have a great number of benefits and are widely used in many business functions as well as different industries. However there are also disagreements and various concerns as to how useful decision trees really are. As technology evolves so do decision trees. Therefore not only do many controversies arise but also solutions and new proposals to these arguments.

BACKGROUND

Decision trees date back to the early 1960's, and were originated by C. I. Hovland and E. B. Hunt (Fu, 2000). Their book which was published in 1961 entitled, *Programming a Model of Human Concept Formation*, was the earliest published discussion of a concept learning program (CLP). The topic of CLP was further expanded in the book titled, *Experiment in Induction*, of 1966 which was written by E. B. Hunt, J. Marin, and P. J. Stone. This publication is considered a starting point to Ross Quinlan's work, whose contribution to the decision tree theory is mainly credited to his ID3 and C4.5/5.0 algorithms in the tree-based methods (Fu, 2000).

Decision trees are a useful technique for classification. Its core concept is based on the graph of decisions that presents possible consequences, with corresponding resource costs and risks, leading to the final conclusions. The main purpose of the decision tree is to make the decision-making process clearer and more understandable. They are also constructed to ease prediction about possible outcomes and alternatives of a specific situation. A set of "if-then" conditions allows the establishing of a final outcome.

Decision trees deal only with predictive values and consist of "square decision nodes, circle probability nodes, and branches representing decision alternatives" (Taylor, 2004, p. 490). The process of constructing a tree involves computation of an expected value of each outcome and makes a decision based on these expected values. Such decision trees can be drawn simply on a piece of paper or go as far as to use sophisticated software in order to present more complex trees.

MAIN FOCUS

Advantages, Problems with DTs, and Proposed Solutions

No matter which method is used, decision trees share many advantages. They provide an illustration of the decision-making process, and therefore are simple to understand and interpret. They require little data preparation and apply to both nominal and categorical values. Thanks to statistical test and formulas, decision trees are reliable. They also perform well with large data in a short time. Decision tree learning can be applied in many different business functions like finance, marketing, or management and many others. It is also very useful across various industries. It has ascribed a particular importance in data mining. Algorithms like Chi-Square Automatic Interaction Detector, CHAID Decision Tree Algorithms, or CART are considered a specifically useful tool in data mining. *Cambridge Business Review* states that "from a business perspective decision trees can be viewed as creating a segmentation of the original dataset: each segment would be one of the leaves of the tree (Ozgulbas & Koyuncugil, 2006, p. 316)

While all of the advantages may make decision trees seem perfect, they are not. Currently, a few problems exist with the decision trees modern formation. One current problem exists in some large decisions that require many calculations which lead to inefficiencies. The calculation problem can easily be shown by a simple example. Zebda (2006) observed that a problem with two decisions, ten states of nature, and one-level signal would require 699 operations using the traditional decision trees approach. The same problem with two-level signal would require 7,399 operations to solve the problem using the traditional decision tree approach. This problem has led to researchers suggesting alternatives instead of decision trees to solve problems. Some researchers do not feel that the alternatives would work better than decision trees. These researchers feel that decision trees perform better in decision making and suggest a new modified decision tree method.

The newly modified decision tree will alleviate the current calculation problem. The modified decision trees can decrease calculations needed by the traditional decision tree method by more than 75% in some examples. The modified decision trees can be used in the same situations as traditional decision trees. It can be used for single level and multi-level trees and symmetrical and non-symmetrical decision-making problems. Even with fewer calculations, the modified decision tree will still maintain the advantages of the traditional trees and make them more efficient than the current state of traditional decision trees. The modified decision tree helps fix the problem with excessive calculations and keeps the advantages of the traditional decision tree. It becomes more efficient, requires less work, and counters the claim of some researchers' problems with traditional decision trees (Zebda, 2006).

Also, there is the issue of "overfitting" rules with few data, limiting the predictive power of decision trees for previously unseen data (Quinlan, 1993). Another problem encountered in most trees is that they are axis-parallel, making them convenient to analyze but they may result in intricate and inaccurate trees if the "data can be partitioned by hyper-planes that are not axis-parallel" (Cantu-Paz & Kamath, 2003, p. 56). Cantu-Paz and Kamath (2003) evoked a revolutionary concept of oblique trees designed to reduce the flaws of the ordinary trees. Then, again, the data became more complex to analyze, requiring a pundit in the field for interpretation.

RECENT DEVELOPMENT IN DECISION TREE THEORY

Genetic Algorithm Trees (GA Trees)

The originality of a *genetic algorithm*, compared to other algorithms used in decision-tree building, is its reliance in addition to the variance of the tree's accuracy, on the expected value of the classification, and a probabilistic method of measuring the performance of the tree (Fu, Golden, Lele, Raghavan, & Wasil, 2003). Also, genetic algorithms fundamentally emulate the natural Darwinian theory of "survival of the fittest", based on a roulette wheel selection, involving a series of mutations and crossovers with the goal to yield the best trees (Michalewicz, 1996). An experiment conducted on 40 test subsets, repeated 10 times, by Fu (2006) and his colleagues revealed an average accuracy of 78.71% for a GA tree. The main advantage of using GA trees is that large, complex analytical problems can be reformulated in a manner that is computationally more efficient than the original problem. Further, their coding aspect inherently takes care of most of the constraints associated with the scheduling problem. Results from a number of test problems demonstrate that genetic algorithms are able to find optimal schedules with a reasonable computational resource (Fu, Golden, Lele, Raghavan, & Wasil, 2003).

Orthogonal Decision Trees (ODTs)

Orthogonal decision trees, as the name implies, are a set of functionally orthogonal decision trees corresponding to the principal components of the underlying function space (Kargupta & Dutta, 2004). Their major advantage is that they offer an effective way to construct redundancy-free ensembles that are easier to understand and apply. Also, they allow the monitoring of data streams from resource-constrained platforms such as PDAs, pocket PCs, and cell-phones where CPU computing power is limited. A special method called "Fourier Spectra" is used to remove redundancies, yielding an