Fuzzy Decision Trees

Malcolm J. Beynon
Cardiff University, UK

INTRODUCTION

The inductive learning methodology known as decision trees, concerns the ability to classify objects based on their attributes values, using a tree like structure from which decision rules can be accrued. In this article, a description of decision trees is given, with the main emphasis on their operation in a fuzzy environment.

A first reference to decision trees is made in Hunt *et al.* (1966), who proposed the Concept learning system to construct a decision tree that attempts to minimize the score of classifying chess endgames. The example problem concerning chess offers early evidence supporting the view that decision trees are closely associated with artificial intelligence (AI). It is over ten years later that Quinlan (1979) developed the early work on decision trees, to introduced the Interactive Dichotomizer 3 (ID3). The important feature with their development was the use of an entropy measure to aid the decision tree construction process (using again the chess game as the considered problem).

It is ID3, and techniques like it, that defines the hierarchical structure commonly associated with decision trees, see for example the recent theoretical and application studies of Pal and Chakraborty (2001), Bhatt and Gopal (2005) and Armand *et al*. (2007). Moreover, starting from an identified root node, paths are constructed down to leaf nodes, where the attributes associated with the intermediate nodes are identified through the use of an entropy measure to preferentially gauge the classification certainty down that path. Each path down to a leaf node forms an 'if .. then ..' decision rule used to classify the objects.

The introduction of fuzzy set theory in Zadeh (1965), offered a general methodology that allows notions of vagueness and imprecision to be considered. Moreover, Zadeh's work allowed the possibility for previously defined techniques to be considered with a fuzzy environment. It was over ten years later that the area of decision trees benefited from this fuzzy environment opportunity (see Chang and Pavlidis, 1977). Since then there has been a steady stream of research studies that have developed or applied fuzzy decision trees (FDTs) (see recently for example Li *et al*., 2006 and Wang *et al*., 2007).

The expectations that come with the utilisation of FDTs are succinctly stated by Li *et al*. (2006, p. 655);

"Decision trees based on fuzzy set theory combines the advantages of good comprehensibility of decision trees and the ability of fuzzy representation to deal with inexact and uncertain information."

Chiang and Hsu (2002) highlight that decision trees has been successfully applied to problems in artificial intelligence, pattern recognition and statistics. They go onto outline a positive development the FDTs offer, namely that it is better placed to have an estimate of the degree that an object is associated with each class, often desirable in areas like medical diagnosis (see Quinlan (1987) for the alternative view with respect to crisp decision trees).

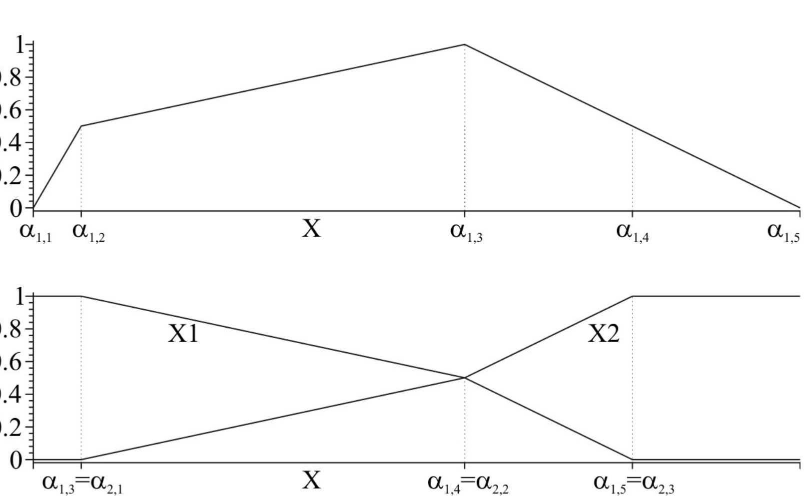
The remains of this article look in more details at FDTs, including a tutorial example showing the rudiments of how an FDT can be constructed.

BACKGROUND

The background section of this article concentrates on a brief description of fuzzy set theory pertinent to FDTs, followed by a presentation of one FDT technique.

In fuzzy set theory (Zadeh, 1965), the grade of membership of a value x to a set S is defined through a membership function $\mu_S(x)$ that can take a value in the range [0, 1]. The accompanying numerical attribute domain can be described by a finite series of MFs that each offers a grade of membership to describe x, which collectively form its concomitant fuzzy number. In this article, MFs are used to formulate linguistic variables for the considered attributes. These linguistic variables are made up of sets of linguistic terms which are defined by the MFs (see later).

Figure 1. Example membership function and their use in a linguistic variable



Surrounding the notion of MFs is the issue of their structure (Dombi and Gera, 2005). Here, piecewise linear MFs are used to define the linguistic terms presented, see Figure 1.

In Figure 1(top), a single piecewise linear MF is shown along with the defining values that define it, namely, $\alpha_{1,1}$, $\alpha_{1,2}$, $\alpha_{1,3}$, $\alpha_{1,4}$ and $\alpha_{1,5}$. The associated mathematical structure of this specific form of MF is given below;

$$\mu(x)\begin{cases} 0 & \text{if } x \leq \alpha_{j,1} \\ 0.5\dfrac{x-\alpha_{j,1}}{\alpha_{j,2}-\alpha_{j,1}} & \text{if } \alpha_{j,1} < x \leq \alpha_{j,2} \\ 0.5+0.5\dfrac{x-\alpha_{j,2}}{\alpha_{j,3}-\alpha_{j,2}} & \text{if } \alpha_{j,2} < x \leq \alpha_{j,3} \\ 1 & \text{if } x = \alpha_{j,3} \\ 1-0.5\dfrac{x-\alpha_{j,3}}{\alpha_{j,4}-\alpha_{j,3}} & \text{if } \alpha_{j,3} < x \leq \alpha_{j,4} \\ 0.5-0.5\dfrac{x-\alpha_{j,4}}{\alpha_{j,5}-\alpha_{j,4}} & \text{if } \alpha_{j,4} < x \leq \alpha_{j,5} \\ 0 & \text{if } \alpha_{j,5} < x \end{cases}$$

As mentioned earlier, MFs of this type are used to define the linguistic terms which make up linguistic variables. An example of a linguistic variable X based on two linguistic terms, X1 and X2, is shown in Figure 2(bottom), where the overlap of the defining values for each linguistic term is evident. Moreover, using left and right limits of the X domain as $-\infty$ and ∞, respectively, the sets of defining values are (in list form); X1 - $[-\infty, -\infty, \alpha_{1,3}, \alpha_{1,4}, \alpha_{1,5}]$ and X2 - $[\alpha_{2,1}, \alpha_{2,2}, \alpha_{2,3}, \infty, \infty]$, where $\alpha_{1,3} = \alpha_{2,1}$, $\alpha_{1,4} = \alpha_{2,2}$ and $\alpha_{1,5} = \alpha_{2,3}$.

This section now goes on to outline the technical details of the fuzzy decision tree approach introduced in Yuan and Shaw (1995). With an inductive fuzzy decision tree, the underlying knowledge related to a decision outcome can be represented as a set of fuzzy '*if* .. *then* ..' decision rules, each of the form;

If (A_1 is $T^1_{i_1}$) and (A_2 is $T^2_{i_2}$) … and (A_k is $T^k_{i_k}$) then C is C_j, where A_1, A_2, .., A_k and C are linguistic variables for the multiple antecedents (A_i's) and consequent (C) statements used to describe the considered objects, and $T(A_k) = \{T^k_1, T^k_2, .. T^k_{S_i}\}$ and $\{C_1, C_2, \ldots, C_L\}$ are their respective linguistic terms, defined by the MFs $\mu_{T^k_j}(x)$ etc. The MFs, $\mu_{T^k_j}(x)$ and $\mu_{C_j}(y)$, represent the grade of membership of an object's antecedent A_j being T^k_j and consequent C being C_j, respectively.

A MF $\mu(x)$ from the set describing a fuzzy linguistic variable Y defined on X, can be viewed as a possibility distribution of Y on X, that is $\pi(x) = \mu(x)$, for all $x \in X$ the values taken by the objects in U (also normalized so $\max_{x\in X} \pi(x) = 1$). The possibility measure $E_\alpha(Y)$ of ambiguity is defined by

$$E_\alpha(Y) = g(\pi) = \sum_{i=1}^{n}(\pi^*_i - \pi^*_{i+1})\ln[i],$$

Related Content