Utilizing Fuzzy Decision Trees in Decision Making

Malcolm J. Beynon
Cardiff University, UK

INTRODUCTION

The seminal work of Zadeh (1965), fuzzy set theory (FST) has developed into a methodology fundamental to analysis that incorporates vagueness and ambiguity. With respect to the area of data mining, it endeavours to find potentially meaningful patterns from data (Hu & Tzeng, 2003). This includes the construction of if-then decision rule systems, which attempt a level of inherent interpretability to the antecedents and consequents identified for object classification (and prediction), (see Breiman 2001).

Within a fuzzy environment this is extended to allow a linguistic facet to the possible interpretation, examples including mining time series data (Chiang, Chow, & Wang, 2000) and multi-objective optimisation (Ishibuchi & Yamamoto, 2004).

One approach to if-then rule construction has been through the use of decision trees (Quinlan, 1986), where the path down a branch of a decision tree (through a series of nodes), is associated with a single if-then rule. A key characteristic of the traditional decision tree analysis is that the antecedents described in the nodes are crisp. This chapter investigates the use of fuzzy decision trees as an effective tool for data mining.

BACKGROUND

The development of fuzzy decision trees brings a linguistic form to the if-then rules constructed, offering a concise readability in their findings (see Olaru & Wehenkel, 2003). Examples of their successful application include in the areas of optimising economic dispatch (Roa-Sepulveda, Herrera, Pavez-Lazo, Knight, & Coonick, 2003) and the antecedents of company audit fees (Beynon, Peel, & Tang, 2004). Even in application based studies, the linguistic formulisation to decision making is continually investigated (Chakraborty, 2001; Herrera, Herrera-Viedma, & Martinez, 2000).

Appropriate for a wide range of problems, fuzzy decision trees (with linguistic variables) allows a representation of information in a more direct and adequate form. A linguistic variable is described in Herrera, Herrera-Viedma, and Martinez (2000), highlighting it differs from a numerical one, instead of using words or sentences in a natural or artificial language. It further serves the purpose of providing a means of approximate characterization of phenomena, which are too complex, or too ill-defined to be amenable to their description in conventional quantitative terms.

The number of elements (words) in a linguistic term set which define a linguistic variable determines the granularity of the characterization. The semantic of these elements is given by fuzzy numbers defined in the [0, 1] interval, which are described by their membership functions (MFs). Indeed it is the role played by, and structure of, the MFs that is fundamental to the utilisation of FST related methodologies (Medaglia, Fang, Nuttle, & Wilson, 2002; Reventos, 1999). In this context, DeOliveria (1999) noted that fuzzy systems have the important advantage of providing an insight on the linguistic relationship between the variables of a system.

With an inductive fuzzy decision tree, underlying knowledge related to a decision outcome can be represented as a set of fuzzy if-then decision rules, each of the form:

\[
\text{If } (A_i \text{ is } T_i^1) \text{ and } (A_j \text{ is } T_j^1) \ldots \text{ and } (A_k \text{ is } T_k^1) \text{ then } C \text{ is } C_j,
\]

where \( A = \{ A_1, A_2, \ldots, A_i \} \) and \( C \) are linguistic variables in the multiple antecedents \((A_i)'s\) and consequent \((C)\) statements, respectively, and \( T(A_i) = \{ T_i^1, T_i^2, \ldots, T_i^k \} \) and \( \{ C_1, C_2, \ldots, C_j \} \) are their linguistic terms. Each linguistic term \( T_j^k \) is defined by \( p_{ij}(x) : A_i \rightarrow [0, 1] \) (similar for a \( C \)). The MFs represent the grade of membership of an object’s antecedent \( A_j \) being and consequent \( C \) being \( C_j \), respectively (Wang, Chen, Qian, & Ye, 2000; Yuan & Shaw, 1995).

Different types of MFs have been proposed to describe fuzzy numbers, including triangular and trapezoidal functions (Lin & Chen, 2002; Medaglia, Fang, Nuttle, & Wilson, 2002). Yu and Li (2001) highlight that MFs may be (advantageously) constructed from mixed shapes, supporting the use of piecewise linear MFs. A general functional form of a piecewise linear MF (in the context of a linguistic term), is given by:

Copyright © 2006, Idea Group Inc., distributing in print or electronic forms without written permission of IGI is prohibited.
where \([a_j,1, a_j,2, a_j,3, a_j,4, a_j,5]\) are the defining values for the MF associated with the \(j^{th}\) linguistic term of an antecedent \(A_j\). A visual representation of this MF is presented in Figure 1, which elucidates its general structure along with the role played by the defining values.

Also included in Figure 1, using dotted lines are neighbouring MFs (linguistic terms), which collectively would define a linguistic variable. To circumvent the influence of expert opinion in analysis, the construction of the MFs should be automated. On this matter DeOliveria (1999) considers the implication of Zadeh’s principle of incompatibility - that is, as the number of MFs increase, so the precision of the system increases, but at the expense of relevance decreasing.

**MAIN THRUST**

Formulation of Fuzzy Decision Tree

The first fuzzy decision tree reference is attributed to Chang and Pavlides (1977). A detailed description on the concurrent work of fuzzy decision trees is presented in Olaru and Wehenkel (2003). It highlights how methodologies include the fuzzification of a crisp decision tree post its construction, or approaches that directly integrate fuzzy techniques during the tree-growing phase. The latter formulation is described here, with the inductive method proposed by Yuan and Shaw (1995) considered, based on measures of cognitive uncertainties.

This method focuses on the minimisation of classification ambiguity in the presence of fuzzy evidence. A membership function \(\mu(x)\) from the set describing a fuzzy 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 possibility measure \(E_{j}(Y)\) of ambiguity is defined by \(E_{j}(Y)\)

\[
g(\pi) = \sum_{i=1}^{n} (\pi_{i}^*-\pi_{i}^*) \log(\pi_i^*)
\]

where \(\pi_{i}^*\) is the permutation of the possibility distribution \(\pi = \{\pi(x_1), \pi(x_2), \ldots, \pi(x_n)\}\) so that \(\pi_{i}^* \geq \pi_{i+1}^*\) for \(i = 1, \ldots, n\), and \(\pi_{n+1}^* = 0\).

The ambiguity of attribute \(A\) (with objects \(u_1, \ldots, u_m\)) is given as:

\[
E_j(A) = \frac{1}{m} \sum_{i=1}^{m} E_i(A(u_i))
\]

where \(E_i(A(u_i)) = g(\mu_{j}(u_i)/\max(\mu_{j}(u_i)))\), with \(j\) the linguistic terms of an attribute (antecedent) with \(m\) objects. When there is overlapping between linguistic terms (MFs) of an attribute or between consequents, then ambiguity exists.

The fuzzy subethood \(S(A, B)\) measures the degree to which \(A\) is a subset of \(B\) and is given by:

\[
S(A, B) = \sum_{j=1}^{m} \min(\mu_{j}(u), \mu_{x}(u))/\sum_{j=1}^{m} \mu_{j}(u)
\]

Given fuzzy evidence \(E\), the possibility of classifying an object to consequent \(C\) can be defined as:

\[
\pi_{C}(E) = S(E, C)/\max(\pi(E, C)), \text{ where } S(E, C)
\]

represents the degree of truth for the classification rule (that is ‘if \(E\) then \(C\)’). With a single piece of evidence (a fuzzy number for an attribute), then the classification ambiguity based on this fuzzy evidence is defined as:

\[
G(E) = g(\pi(E, C))
\]

The classification ambiguity with fuzzy partitioning \(P = \{E_1, \ldots, E_n\}\) on the fuzzy evidence \(F\), denoted as \(G(P, F)\), is the weighted average of classification ambiguity with each subset of partition:

\[
G(P, F) = \frac{1}{|P|} \sum_{E \in P} w(E, F) G(E, F)
\]

where \(w(E, F)\) is the classification ambiguity with fuzzy evidence \(E \cap F\), and where \(E \cap F\) is the relative size of subset \(E\) in \(F\): \(w(E, F)\)

\[
= \frac{\sum_{j=1}^{m} \min(\mu_{j}(u), \mu_{x}(u))/\sum_{j=1}^{m} \sum_{i=1}^{m} \min(\mu_{j}(u), \mu_{x}(u))\}
\]

In summary, attributes are assigned to nodes based on the lowest level of ambiguity. A node becomes a leaf node if the level of subethood (based on the intersection of the nodes from the root) is higher than some truth value \(\beta\).

**Illustrative Application of Fuzzy Decision Tree Analysis to Audit Fees Problem**

The data mining of an audit fees model may be useful to companies in assessing whether the audit fee they are paying is reasonable. In this analysis, a sample of 120 UK companies is used for training (growing) a fuzzy decision tree (Beynon, Peel, & Tang, 2004).

The variables used in the study, are the decision attribute, AFEE: Audit Fee (£000’s) and condition attributes; SIZE: Sales (£000s), SUBS: No. of subsidiaries, FORS: Ratio of foreign to total subsidiaries, GEAR: Ratio of debt to total assets CFEE: Consultancy fees (£000s), TOTSH: Proportion of shares held by directors and sub-
Related Content

Semantic Multimedia Content Retrieval and Filtering
www.igi-global.com/chapter/semantic-multimedia-content-retrieval-filtering/11058?camid=4v1a

Financial Ratio Selection for Distress Classification
www.igi-global.com/chapter/financial-ratio-selection-distress-classification/10649?camid=4v1a

Predicting Resource Usage for Capital Efficient Marketing
www.igi-global.com/chapter/predicting-resource-usage-capital-efficient/10726?camid=4v1a

Control-Based Database Tuning Under Dynamic Workloads
www.igi-global.com/chapter/control-based-database-tuning-under/10841?camid=4v1a