

Utilizing Fuzzy Decision Trees in Decision Making

Malcolm J. Beynon,
Cardiff University, UK

INTRODUCTION

The seminal work of Zadeh (1965), namely 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 (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, where this restriction is mitigated when operating in a fuzzy environment (Crockett, Bandar, Mclean, & O'Shea, 2006).

This chapter investigates the use of fuzzy decision trees as an effective tool for data mining. Pertinent to data mining and decision making, Mitra, Konwar and Pal (2002) succinctly describe a most important feature of decision trees, crisp and fuzzy, which is their capability to break down a complex decision-making process into a collection of simpler decisions and thereby, providing an easily interpretable solution.

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-Serpulveda, Herrera, Pavez-Lazo, Knight, & Coonick, 2003) and the antecedents of company audit fees (Beynon, Peel, & Yang, 2004). Even in application based studies, the linguistic formulation to decision making is continually investigated (Chakraborty, 2001; Herrera, Herrera-Viedma, & Martinez, 2000).

Appropriate for a wide range of problems, the fuzzy decision trees approach (with linguistic variables) allows a representation of information in a direct and adequate form. A linguistic variable is described in Herrera, Herrera-Viedma, & Martinez (2000), highlighting it differs from a numerical one, with it instead 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 characterisation. 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 the structure of, the MFs that is fundamental to the utilization 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.

Within 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 $(A_1 \text{ is } T_1^1)$ and $(A_2 \text{ is } T_2^2) \dots$ and $(A_k \text{ is } T_k^k)$ then C is C_j ,

where A_1, A_2, \dots, A_k and C are linguistic variables in the multiple antecedents (A_i s) and consequent (C) statements, respectively, and $T(A_k) = \{T_1^k, T_2^k, \dots, T_{S_i}^k\}$ and $\{C_1, C_2, \dots, C_L\}$ are their concomitant linguistic terms. Each linguistic term T_j^k is defined by the MF $\mu_{T_j^k}(x)$, which transforms a value in its associated domain to a grade of membership value to between 0 and 1. The MFs, $\mu_{T_j^k}(x)$ and $\mu_{C_j}(y)$, represent the grade of membership of an object's attribute value for A_j being T_j^k and C being C_j , respectively (Wang, Chen, Qiang, & 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:

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

with the respective *defining values* in list form, $[\alpha_{j,1}, \alpha_{j,2}, \alpha_{j,3}, \alpha_{j,4}, \alpha_{j,5}]$. A Visual representation of this MF form is presented in Figure 1, which elucidates its general structure along with the role played by the respective sets of defining values.

The MF form presented in Figure 1 shows how the value of a MF is constrained within 0 and 1. The implication of the specific defining values is also illustrated, including the idea of associated support, the domain $[\alpha_{j,1}, \alpha_{j,5}]$ in Figure 1. Further, the notion of dominant support can also be considered, where a MF is most closely associated with an attribute value,

the domain $[\alpha_{j,2}, \alpha_{j,4}]$ in Figure 1 (see Kovalerchuk & Vityaev, 2000).

Also included in Figure 1, using dotted lines are neighbouring MFs (linguistic terms), which collectively would define a linguistic variable, describing a continuous attribute. 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 decreasing relevance.

MAIN THRUST

Formulation of Fuzzy Decision Tree

The first fuzzy decision tree reference is attributed to Chang and Pavlidis (1997). A detailed description on the concurrent work of fuzzy decision trees is presented in Olaru & Wehenkel (2003). It highlights how methodologies include the fuzzification of a crisp decision tree post its construction (Pal & Chakraborty, 2001), 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.

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 $E_\alpha(Y) = g(\pi) = \sum_{i=1}^n (\pi_i^* - \pi_{i+1}^*) \ln[i]$, where $\pi^* = \{\pi_1^*, \pi_2^*, \dots, \pi_n^*\}$ is the permutation of the normalized possibility distribution $\pi = \{\pi(x_1), \pi(x_2), \dots, \pi(x_n)\}$, sorted so that $\pi_i^* \geq \pi_{i+1}^*$ for $i = 1, \dots, n$, and $\pi_{n+1}^* = 0$. In the limit, if $\pi_2^* = 0$, then $E_\alpha(Y) = 0$, indicates no ambiguity, whereas if $\pi_n^* = 1$, then $E_\alpha(Y) = \ln[n]$, which indicates all values are fully possible for Y , representing the greatest ambiguity.

The ambiguity of attribute A (over the objects u_1, \dots, u_m) is given as:

$$E_\alpha(A) = \frac{1}{m} \sum_{i=1}^m E_\alpha(A(u_i)),$$

where

5 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/utilizing-fuzzy-decision-trees-decision/11097

Related Content

Computation of OLAP Data Cubes

Amin A. Abdulghani (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition* (pp. 286-292). www.irma-international.org/chapter/computation-olap-data-cubes/10834

Cost-Sensitive Learning

Victor S. Sheng and Charles X. Ling (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition* (pp. 339-345). www.irma-international.org/chapter/cost-sensitive-learning/10842

Data Mining for Fraud Detection System

Roberto Marmo (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition* (pp. 411-416). www.irma-international.org/chapter/data-mining-fraud-detection-system/10853

XML Warehousing and OLAP

Hadj Mahboubi (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition* (pp. 2109-2116). www.irma-international.org/chapter/xml-warehousing-olap/11111

Tabu Search for Variable Selection in Classification

Silvia Casado Yusta and Joaquín Pacheco Bonrostro (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition* (pp. 1909-1915). www.irma-international.org/chapter/tabu-search-variable-selection-classification/11080