Chapter 7
Fuzzy Decision Rule Construction Using Fuzzy Decision Trees: Application to E-Learning Database

Malcolm J. Beynon
Cardiff University, UK

Paul Jones
University of Glamorgan, UK

ABSTRACT

This chapter considers the soft computing approach called fuzzy decision trees (FDT), a form of classification analysis. The consideration of decision tree analysis in a fuzzy environment brings further interpretability and readability to the constructed ‘if.. then..’ decision rules. Two sets of FDT analyses are presented, the first on a small example data set, offering a tutorial on the rudiments of one FDT technique. The second FDT analysis considers the investigation of an e-learning database, and the elucidation of the relationship between weekly online activity of students and their final mark on a university course module. Emphasis throughout the chapter is on the visualization of results, including the fuzzification of weekly online activity levels of students and overall performance.

INTRODUCTION

In 1998, Bonissone (1998) perceived soft computing to be a recently coined term, describing the symbiotic use of many emerging computing disciplines, amongst these Fuzzy Logic (introduced by Zadeh, 1965). They suggested that fuzzy logic gives us a language, with syntax and local semantics, in which we can translate our qualitative knowledge about the problem to be solved, with its main characteristic the robustness of its interpolative reasoning mechanism (see Bonissone, 1998).

The association of fuzzy logic with soft computing is regularly perceived. Mitra et al. (2002) consider the role of soft computing in data mining (viewing data mining as a step in knowledge discovery in databases) (p. 6):
Fuzzy logic is capable of supporting, to a reasonable extent, human type reasoning in natural form. It is the earliest and most widely reported constituent of soft computing.

It is this characteristic of fuzzy logic, including the potential to perceive qualitative observations and calibration of commonsense rules, in an attempt to establish meaningful and useful relationships between system variables, that allows it to be a pertinent tool to be encompassed in knowledge discovery in databases, which itself is mainly concerned with identifying interesting patterns and describing them in a concise and meaningful manner (Fayyad et al., 1996).

An ability to construct ‘commonsense’ rules, which are readable and so easily interpretable, is considered here through the employment of a fuzzy decision tree approach (FDT). Amongst the FDT techniques previously constructed, include, fuzzy versions of crisp decision techniques, such as fuzzy ID3 (see Ichihashi et al., 1996; Pal and Chakraborty, 2001), and other versions (see Yuan and Shaw, 1995; Olaru and Wehenkel, 2003). Like the original crisp algorithms, such as the ID3 algorithm (Quinlan, 1979), the general approach involves the repetitive partitioning of the objects in a data set through the augmentation of attributes down a tree structure from a root node, through a series of branches ending in a number of leaf nodes, where each subset of objects is associated with the same decision class or no attribute is available for further decomposition.

With the operation of FDTs in a fuzzy environment (Zadeh, 1965), it is closely associated with the analysis of uncertain and imperfect information. Further, as previously hinted at, the fuzzy environment offers a linguistic domain, within which an analysis can be undertaken and subsequent results expressed (Herrera et al., 2000). This domain is determined through the utilisation of membership functions (MFs), which define fuzzy based analyses (Sancho-Royo and Verdegay, 1999). FDT analyses have further been praised for comprehensibility in Janikow (1998), who goes on to state Quinlan (1993), when discussing decision trees in general:

This appeals to a wide range of users who are interested in domain understanding, classification capabilities, or the symbolic rules that may be extracted from the tree.

The rudiments of FDTs are, as an inductive learning technique, for the fuzzy classification of objects that are described by a number of characteristic attributes. The results from using FDTs are a set of fuzzy ‘if.. then..’ decision rules (see Dubois and Prade, 1996). The fuzziness associated with these rules brings with it the pertinent qualities of readability and interpretability to any analysis, offering an efficient insight into the considered problem, a further facet of soft computing (Bodenhofer et al., 2007).

In this chapter, the FDT approach of Yuan and Shaw (1995) is utilised, which attempts to include the cognitive uncertainties evident in data values (for alternative FDT approaches see also Umano et al. (1994) and Pal and Chakraborty (2001)). Particularly pertinent to its applicability in areas such as business, but appropriate here, Chen and Chiou (1999, pp. 408) highlight:

In practical business, economic data are usually imprecise and fuzzy. The assessment level of a criterion often depends on judgement or approximation. A sharp classification or unnatural approximating values results in unreasonable or incorrect outcomes.

The FDT approach described in this chapter is initially demonstrated on a small data set, where explicit analytic calculations are given, for when different numbers of MFs are employed to define the linguistic terms formulating the linguistic variables of attributes. This, almost tutorial based exposition, should allow a reader of this chapter the opportunity to follow the rudiments of a
Related Content

Object-Relational Modeling in the UML
www.igi-global.com/chapter/object-relational-modeling-uml/11183?camid=4v1a

Designing Information Systems Capabilities to Create Business Value: A Theoretical Conceptualization of the Role of Flexibility and Integration
www.igi-global.com/chapter/designing-information-systems-capabilities-create/7946?camid=4v1a

A Formal Verification and Approach for Real-Time Databases
www.igi-global.com/chapter/formal-verification-approach-real-time/28557?camid=4v1a

An Investigation of the Impact of Organization Size on Data Quality Issues
www.igi-global.com/article/investigation-impact-organization-size-data/3337?camid=4v1a