

Chapter XX

The Exposition of Fuzzy Decision Trees and Their Application in Biology

Malcolm J. Beynon

Cardiff University, UK

Kirsty Park

University of Stirling, UK

ABSTRACT

This chapter employs the fuzzy decision tree classification technique in a series of biological based application problems. With its employment in a fuzzy environment, the results, in the form of fuzzy 'if.. then ..' decision rules, bring with them readability and subsequent interpretability. The two contrasting applications considered concern, the age of abalones and the lengths of torpor bouts of hibernating Greater Horseshoe bats. Emphasis is on the visual results presented, including the series of membership functions used to construct the linguistic variables representing the considered attributes and the final fuzzy decision trees constructed. Technical details presented further offer the opportunity to readers to future employ the technique in other biological applications.

INTRODUCTION

Fuzzy set theory (FST) stands out as a general methodology that has contributed to the development of already established techniques used throughout many areas of science, including biology and medicine (recent examples include, Morato *et al.*, 2006; Mastorocostas and Theocharis, 2007). Since its introduction in Zadeh (1965),

the utilisation of FST in such development has been with the inclusion of the acknowledgement of the presence of vagueness and ambiguity during its operation. Further, it has also invoked the ability to interpret the structuring and subsequent results from data analysis in a linguistic orientated language (Grzegorzewski and Mrówka, 2005). Indeed, artificial intelligence, with respect to FST, is closely associated with the mimicry of human

cognition and linguistic language (Trillas and Guadarrama, 2005).

The issue of interpretability is particularly relevant in classification based problems, but overlooked, since so often the concomitant analysis is more oriented to the resultant classification accuracy, rather than interpretability. Indeed, Breiman (2001), in an informative discussion on the cultures of statistical modelling, comes down heavily on the need to accommodate the ability to interpret results in analysis undertaken. Their discussion offers a pertinent illustrative argument, describing the example of a medical doctor, with experimental data, in a choice between accuracy and interpretability they would choose interpretability. This issue of interpretability over accuracy is also pertinent in general biology.

This chapter considers fuzzy decision trees (FDTs), an example of an already established technique that has been developed using FST. The fundamentals of the decision tree technique, within a crisp or fuzzy environment, is concerned with the classification of objects described by a number of attributes, with concomitant decision rules derived in the constructed decision tree. The inherent structure a consequence of the partitioning algorithm used to discern the classification impact of the attributes. An early FDT reference is attributed to Chang and Pavlidis (1977). In the area of medicine, for example, Podgorelec *et al.* (2002) offer a good review of decision trees, with a most recent employment of FDTs presented in Armand *et al.* (2007), which looked into gait deviations. Relative to this, there is a comparative dearth of their employment in a biological setting, one exception being Beynon *et al.* (2004a).

An important feature of FDTs is the concomitant sets of fuzzy ‘*if... then ...*’ decision rules constructed, whose condition and decision parts, using concomitant attributes, can be described in linguistics terms (such as low, medium or high). The suggested FDT approach employed here was presented in Yuan and Shaw (1995) and Wang *et*

al. (2000), and attempts to include the cognitive uncertainties evident in the data values. This FDT approach has been used in Beynon *et al.* (2004b) and Beynon *et al.* (2004a), the latter investigating the songflight of the Sedge Warbler, expositing the relationship between the birds’ characteristics like, repertoire size and territory size, against their song flight duration.

Central to the utilisation of FDTs is the fuzzification of the considered data set, through the employment of FST related membership functions (MFs), which further enable the linguistic representation of the attributes considered (Kecman, 2001), present also in the subsequent decision rules constructed. Alongside the exposition of FDTs in this chapter, the results from two contrasting biology based applications are considered; the first using a well known data set from the UCI data repository and relates to the prediction of the age of abalones (Waugh, 1995), the second is a more contemporary application looking at the torpor bouts of hibernating Greater Horseshoe bats (Park *et al.*, 2000).

The contribution of this book chapter is the clear understanding of the potential advantages of the utilization of FDTs in biological processes. The small hypothetical example allows the reader to clearly follow the included analytical rudiments, and the larger applications demonstrate the interpretability allowed through the use of this approach. It expected a reader will gain effective insights into the type of technical details necessary for a FDT analysis, and its analysis within biological problems.

BACKGROUND

The contents of this section are partitioned into three subsections; the first introduces fuzzy set theory including the fuzzification of a small example data set, the second outlines the FDT technique used to exposit the concomitant general

18 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/exposition-fuzzy-decision-trees-their/4988

Related Content

Development of Highly Sensitive Compact Chemical Sensor System Employing a Microcantilever Array and a Thermal Preconcentrator

Takashi Mihara, Tsuyoshi Ikehara, Mitsuo Konno, Sunao Murakami, Ryutaro Maeda, Tadashi Fukawa and Mutsumi Kimura (2013). *Human Olfactory Displays and Interfaces: Odor Sensing and Presentation* (pp. 199-228).

www.irma-international.org/chapter/development-highly-sensitive-compact-chemical/71925

Particle Swarms: Optimization Based on Sociocognition

James Kennedy (2005). *Recent Developments in Biologically Inspired Computing* (pp. 235-269).

www.irma-international.org/chapter/particle-swarms-optimization-based-sociocognition/28330

Identifying Subtypes of Cancer Using Genomic Data by Applying Data Mining Techniques

Tejal Upadhyay and Samir Patel (2019). *International Journal of Natural Computing Research* (pp. 55-64).

www.irma-international.org/article/identifying-subtypes-of-cancer-using-genomic-data-by-applying-data-mining-techniques/231573

AGE-P: An Evolutionary Platform for the Self-Organization of Smart-Appliance Ensembles

Ralf Salomon and Stefan Goldmann (2010). *Nature-Inspired Informatics for Intelligent Applications and Knowledge Discovery: Implications in Business, Science, and Engineering* (pp. 182-203).

www.irma-international.org/chapter/age-evolutionary-platform-self-organization/36316

Modeling an Artificial Stock Market

S. Lavigne and S. Sanchez (2007). *Handbook of Research on Nature-Inspired Computing for Economics and Management* (pp. 909-925).

www.irma-international.org/chapter/modeling-artificial-stock-market/21173