FuzzyDT - A Fuzzy Decision Tree Algorithm Based on C4.5

Marcos E. Cintra\textsuperscript{1}, Maria C. Monard\textsuperscript{2}, and Heloisa A. Camargo\textsuperscript{3}

\textsuperscript{1} Exact and Natural Sciences Dept. - Federal University of the Semi-arid - UFERSA
Av. Francisco Mota, 572, CEP: 59.625-900 - Mossoró - RN

\textsuperscript{2} Math. and Computer Science Institute - University of São Paulo - USP
Av. Trabalhador Sáo-carlense, 400 - CEP: 13.566-590 - São Carlos - SP

\textsuperscript{3} Computer Science Department - Federal University of São Carlos - UFSCar
Rod. Washington Luis, Km 235 - PO Box 676, 13.565-905 - São Carlos - SP
{cintra@ufersa.edu.br, mcmonard@icmc.usp.br, heloisa@dc.ufscar.br}

Abstract. Decision trees have been successfully applied to many areas for tasks such as classification, regression, and feature subset selection. Decision trees are popular models in machine learning due to the fact that they produce graphical models, as well as text rules, that end users can easily understand. Moreover, their induction process is usually fast, requiring low computational resources. Fuzzy systems, on the other hand, provide mechanisms to handle imprecision and uncertainty in data, based on the fuzzy logic and fuzzy sets theory. The combination of fuzzy systems and decision trees has produced fuzzy decision tree models, which benefit from both techniques to provide simple, accurate, and highly interpretable models at low computational costs. In this paper, we expand previous experiments and present more details of the FuzzyDT algorithm, a fuzzy decision tree based on the classic C4.5 decision tree algorithm. Experiments were carried out using 16 datasets comparing FuzzyDT with C4.5. This paper also includes a comparison of some relevant issues regarding the classic and fuzzy models.

Keywords: Fuzzy classification systems, decision trees, fuzzy decision trees.

1 Introduction

Machine learning is concerned with the development of methods for the extraction of patterns from data in order to make intelligent decisions based on these patterns. A relevant concern related to machine learning methods is the issue of interpretability, which is highly desirable for end users. In this sense, Decision Trees (DT) \cite{12} are powerful and popular models for machine learning since they are easily understandable, quite intuitive, and produce graphical models that can also be expressed as rules. The induction process of decision trees is usually fast and the induced models are usually competitive with the ones generated by other interpretable machine learning methods. Another quality of decision