Stochastic Foundations for the Case-Driven Acquisition of Classification Rules

Megan Vazey

1 Department of Computing, Division of Information and Communication Sciences, Macquarie University, Sydney NSW 2109 Australia
megan@excellan.com.au

Abstract. A predictive mathematical model is presented for the expected case-driven transfer of classification rules. Key insights are offered for Knowledge Acquisition in expert systems, machine learning, artificial intelligence, ontology, and folksomonies.

Keywords: Knowledge Acquisition, Group Decision Support Systems, Collaborative Tagging, Folksonomies, Knowledge Based Systems, Machine Learning, Knowledge Discovery in Databases, Case Based Reasoning, Ripple Down Rules, Expert Systems.

1 Introduction

In this paper, I examine the case-driven transfer of knowledge in which one or more parties transfers classification rules to another party as a result of a continuous and randomized stream of incoming cases, and I present a predictive stochastic model for the case-driven acquisition of classification rules. Importantly, my analysis does not assume any particular underlying Case-based Knowledge Acquisition (KA) technique. The resultant trajectories reflect the natural slowing of knowledge exchange in an environment where incoming repetitive and randomized cases are mapped to a bounded set of classes, and where the class mappings are defined by rules that examine the attributes of the incoming cases. The derived model provides very good insight to the rule acquisition data presented in previous machine-learnt and case-based KA simulations for Single Classification Ripple Down Rules (SCRDR) and Multiple Classification Ripple Down Rules (MCRDR) as discussed in [1] and [2], as well the tag acquisition data observed in folksomonies i.e. collaborative tagging forums as shown in [3, Fig. 3].

Note that both over-specialisation and over-generalisation errors may result in more RuleNodes being acquired than is optimal. In the former scenario, multiple RuleNodes with different rules may refer to identical classifications. In the latter scenario, RuleNodes with different rules may be required to create exceptions to an invalid parent RuleNode.
2 An Analysis of Case-Driven Knowledge Acquisition

The type of knowledge that this research concerns itself with is that which can be codified in the form of rules or RuleNodes that examine the properties (i.e. attributes) of incoming cases, and then map those cases to representative classes or classifications. Classification Knowledge can be acquired directly as top-down knowledge-based rules, or as bottom-up experience-based rules derived by examining specific cases. This paper focuses on the latter case-driven KA approach.

2.1 A Single Classification Case-Driven Equal Frequency KA Example

Say that the target knowledge domain will be comprised of $m$ RuleNodes in a decision tree that maps incoming cases to their representative classes. In order to examine the case-driven KA process I randomly generated $N = 1000$ cases each comprising one of $m = 100$ different integers. The $m$ different integers were represented with equal frequency in the example. In this experiment, a case with a novel integer was used to represent an exemplar case for a novel class or classification. For single classification case-driven KA, each novel exemplar case represents an opportunity for the KBS to acquire a new RuleNode in the decision tree.

Next, I took the sequential set of cases and cumulatively counted the number of times a case with a novel integer was seen. I then plotted the number of novel integers seen versus the number of cases seen. Fig. 1 shows the Actual trajectories for 5 independent case-driven KA scenarios, together with the Expected trajectory, and the Best Case straight-line trajectory for $m = 100$.

![Classes (K) vs Cases (n)](image)

**Fig. 1.** Expected, Actual and Best-Case Single-Class Case-Driven KA trajectories