Exploiting the Ordering of Observed Problem-solving Steps for Knowledge Base Refinement: an Apprenticeship Approach

Steven K. Donoho and **David C. Wilkins**
Department of Computer Science, University of Illinois
Urbana, IL 61801
donoho@cs.uiuc.edu, wilkins@cs.uiuc.edu

Apprenticeship is a powerful method of learning among humans in which a student refines his knowledge by observing and analyzing the problem-solving steps of an expert. In this paper we focus on knowledge base (KB) refinement for classification problems and examine how the *ordering* of the intermediate steps of an observed expert can be used to yield leverage in KB refinement. In the classical classification problem, the problem-solver is given an example consisting of a set of attributes and their corresponding values, and it must put the example in one of a pre-enumerated set of classes.

Consider a slightly different situation, though, in which the problem-solver is not given *all* the attribute/value pairs from the outset but rather must request attributes one at a time and make his classification decision once sufficient evidence is gathered. This situation would arise when it is too costly or otherwise unreasonable to simply be given all the attribute values. When a mechanic is troubleshooting a malfunctioning car, he does not run every test possible and then stop to examine his data and make his decision. Rather he checks one thing, and based on the result of that, he decides what to check next. Thus the order in which attributes are requested reflects the internal problem-solving process going on in the mind of the observed expert. By watching the order in which a superior problem-solver requests attributes, we should be able to refine the KB of a weaker problem-solver.

The ordering of attribute requests can be used to detect KB shortcomings because it allows the analysis of an attribute request with respect to what was and what was not known at the time of the request. As an example from the audiology domain, if the expert requests the attribute *history_noise* after *age_gt_60* is known to be *true*, it can be assumed that knowing *history_noise* is important to know even when *age_gt_60* is *true*. If in the KB we are refining, *history_noise* is not worth knowing given that *age_gt_60* = *true*, then our KB contradicts the actions of the expert indicating a KB shortcoming. Using our KB, we cannot explain why the expert would ask *history_noise* given what he already knew; thus, the expert must have some knowledge which our KB lacks.

Once a KB shortcoming has been detected, an attempt is made to repair it by adding a rule of the form: $condition_1 \wedge \ldots \wedge condition_N \longrightarrow class_i$ where each condition is an attribute/value pair such as *history_dizziness* = *true*. The repair is built by starting with an initial single-condition rule and greedily adding conditions. The condition in the initial rule consists of the **unexplained attribute** (*history_noise* in the above example) and one of its possible values. The class of the initial rule is any class that has not been ruled out by the attribute requests preceding the unexplained attribute (with respect to a set of training examples). Since there may be multiple feasible classes and multiple values for the unexplained attribute, multiple initial rules may have to be explored. The conditions which are greedily added each consist of an attribute requested before the unexplained attribute and its known value — this is because knowledge of these attributes gave rise to the request of the unexplained attribute; therefore, they may be related to the unexplained attribute and to the shortcoming. Conditions are added until the purity of the set of training examples covered by the rule no longer improves.

The principles explored have been implemented, and experiments were run using the audiology dataset. In a test with a training set of size 100, an initial KB was created using C4.5 which achieved an accuracy of 67.9%. This initial KB was refined, and the final KB achieved an accuracy of 82.1%, a net improvement of 14.2%. The ordered sequences of requested attributes were generated by C4.5 using all 226 audiology examples. The power of the attribute ordering method of apprenticeship is that it does not rely solely on empirical calculations to discover attribute/class relationships. Rather, these attribute/class relationships are suggested by attribute ordering and are only *verified* empirically requiring less empirical evidence.