Proc. Third International Machine Learning Workshop, Skytop, PA, June 1985.

ODYSSEUS: A LEARNING APPRENTICE

David C. Wilkins, William J. Clancey and Bruce G. Buchanan

Computer Science Department
Stanford University
Stanford, CA 94305

Abstract

Human specialists employ powerful learning methods during their apprenticeship training period, to augment their fledgling expertise. We describe a system under construction to allow an expert system to use some of these same methods.

The Odysseus learning apprenticeship program watches the observable actions of a specialist. Justifications are created for each action via a process of differential modeling between the specialist and an expert system. A learning opportunity occurs when no action justification is judged sufficiently plausible. This paper describes the three phases that Odysseus uses to learn via differential modeling: generation and justification of an expanded rule base, generation and ranking of action justifications for each observable action, and rationalization of any discrepancies to effect repair.

1. Introduction

An apprenticeship learning period is an important phase on the path to master expert status for human specialists. During this phase, an apprentice specialist *learns by watching master specialists* and *learns by doing problem solving* under the supervision of master specialists. Our research investigates how to give an expert system the benefits of an apprenticeship period.

This paper describes a computer program, Odysseus, that learns by watching specialists in the domain of medical diagnosis. The central task of Odysseus is to *rationalize* each observable action of a specialist during problem solving sessions. In medical diagnosis, these actions consist of all data requests made by a physician and the final diagnosis. Actions are rationalized by a process of *differential modeling* between the expert system and the specialist. Failure to find an adequate rationalization signals a possible deficiency in the expert system's domain or strategy knowledge. Using a taxonomy of deficiencies in conjunction with theoretical and experiential knowledge of the application domain, Odysseus automatically generates and tests conjectures to explain its inability to justify a specialist's action.

Beginning with an existing expert system, the apprenticeship learning process goes through three phases, to be described in the next three sections. These phases are (1) generation and justification of an expanded rule base, (2) generation and ranking of action justifications for each observable step of a problem-solving session, and (3) rationalization of discrepancies to effect repair.

2. Generating Rule Justifications

The Heracles expert system shell, a domain-independent version of Neomycin, is used by Odysseus for differential modeling. Neomycin (Heracles and a particular medical knowledge base) is a reorganization of the Mycin expert system that simulates the diagnostic process of experts [1]. It differs from Mycin in its factorization of the different types of knowledge originally contained in the Mycin rules, and its method of hypothesis-directed reasoning, which employs a body of domain-independent strategy knowledge and a strategy script for diagnosis. A taxonomy of abstract strategy goals permits comparison between the actions of Heracles and the specialist.

There are two ways in which an expert system must be augmented before differential modeling of a human specialist can commence. First, the set of heuristic rules must be replaced by a larger set obtained by induction over past problem solving cases. The original set of rules is adequate for problem solving but is too impoverished to explain the problem solving behavior of a community of specialists. Having Odysseus responsible for generating all initial and subsequent rules guarantees uniformity in the method of assigning strengths to rules. Second, all domain rules must be justified by theory or experience. Definitional, subsumption, and causal rules are justified by references to

determines the likelihood that $J(A_i)$ contains $j_{i,s}$, the action justification of the specialist. This involves, first, ranking $j_{i,1},...,j_{i,n}$ in order of likelihood of being equal to the unknown $j_{i,s}$. An example of ranking rule is: given two elements of a $J(A_i)$, where A_i occurs early in the problem solving session, the $j_{i,k}$ that relates to the more general hypothesis is more likely to be $j_{i,s}$. Second, Odysseus must decide if the top ranked justification(s) is likely to be equal to $j_{i,s}$ — a very difficult problem, although the ability to "know that you don't know" why an action is taken is easy for human specialists who are watching other specialists.

On Neomycin test cases, the elements of $J(A_i)$ generated using the Neomycin rule set as a basis were correctly ranked. The specialist's (Neomycin's) justification was always the top-ranked justification. However, generating and ranking justifications became more difficult when protocols of actual physicians where used. Indeed, $j_{i,s}$ was not in the $J(A_i)$ generated from the Neomycin rule set 75% of the time. The solution to the incompleteness of $J(A_i)$ is the use of an initial induction phase to expand the rule base. The solution to inaccurate ranking is more complex ranking heuristics. A method being implemented to produce these is as follows.

Automatic generation of new ranking heuristics begins with the collection of protocols annotated with $j_{i,s}$ for each A_i. These protocols allow identification of the features of the problem solving session relevant to ranking elements of $J(A_i)$. These features fall into a number of broad categories: patterns of interpretation in the problem solving session, the distance between the expert system's preferred strategic action and the strategic action associated with each $j_{i,k}$, a user model that records individual diagnostic styles, and a history of past problem solving sessions. Each of these features may relate to features of $j_{i,k}$, such as its focus, rules, and strategic goal.

Given the set of problem features that contribute evidence to determining $j_{i,s}$ and the features