CLASSIFICATION PROBLEM SOLVING: A TUTORIAL FROM AN AI PERSPECTIVE

Bruce Chandrasekaran and Anne Keuneke

Laboratory for Artificial Intelligence Research
Department of Computer & Information Science
The Ohio State University
Columbus, OH 43210, USA

Abstract

This paper is a tutorial discussion on classification problem solving, especially hierarchical classification. First we compare how pattern recognition and AI approaches to classification differ by pointing out how knowledge provides leverage against complexity, and thus point to the rationale behind knowledge-based systems. But we critique much of the work in knowledge-based systems, showing that important distinctions between various generic problem solving activities are often obscured by concentration on the implementation level of abstraction, such as "rules", "logic", or "frames". We then argue that a generic task approach facilitates problem analysis, system design, knowledge acquisition and explanation of problem solving. We describe MDX, a medical diagnosis system that performs knowledge-based hierarchical classification, and motivate a number of issues in classification from that perspective. We also describe a high-level language called CSRL that is specially designed for hierarchical classification problem solving and show its power and utility.

1. Introduction

In this paper, we present a tutorial describing issues involved in classification problem solving. Most of these ideas have been introduced in earlier papers from our laboratory:[6], [8], [9], [10], [3]. The purpose of this paper is to guide the reader through a progression of approaches to classification problem solving to illustrate how knowledge of the domain and of the task reduces the computational complexity and aids in system building. The flow of the argument is as follows:

1. We trace the evolution of pattern recognition techniques to complex knowledge-based operations.

2. We describe knowledge-based systems as envisioned through the rule-based approach and include a brief overview of the widely known rule-based system, MYCIN, as it performs classification.

3. We motivate a need for a higher level understanding of classification as a problem solving task, and then describe MDX, a medical diagnostic system that explicitly performs knowledge-based classification.
4. We describe CSRL, a high level language which facilitates the development of classification systems by supporting constructs which represent classification knowledge at appropriate levels of abstraction.

5. We discuss the power and utility of approaching problem solving from this perspective of matching techniques to tasks.

1.1. The Classification Task

Classification is a useful and powerful method of knowledge organization for comprehension and problem solving. In the sciences examples of its use are common. Taxonomic or hierarchical classification has long been a significant methodology in biology. Linnaeus' classification scheme is very well known, and more recently mathematical taxonomy has been used for providing better classifications in this area [25]. The periodic table of elements in chemistry is a masterful example of how man has classified patterns in nature to great advantage.

With the advent of computers, new sciences solved new problems, but the use of classification remained. In fact, in the early days of Pattern Recognition, the problem of recognition was formulated as a problem of classification, in particular one of statistical classification. Even when newer techniques, such as syntactic approaches, came into the field the problem was still often formulated as a classification problem, this time into grammatical categories.

In Expert Systems, a subarea of Artificial Intelligence, the objective is to capture in computer programs, explicitly and in symbolic form, the knowledge and problem solving methods of human experts in selected domains and tasks. If one were to examine the nature of the tasks of the current generation of expert systems, a fact emerges: most of them solve variants of problems which are intrinsically classificatory in nature.

Let us take some examples:

1. MYCIN [24], in its diagnostic phase, has the task of classifying patient data in an infectious agent hierarchy.

2. PROSPECTOR [14] classifies a geological description as corresponding to one or more mineral formation classes.

3. MDX [6,7] explicitly views a significant portion of the diagnostic task as classifying a complex description (the patient data) as an element in a disease classification hierarchy (e.g., liver disease, in particular hepatitis).

4. SACON [1] classifies structural analysis problems into classes for each of which a particular family of analysis methods will be appropriate.

This is by no means to imply that all problems are classification problems, or that all can be usefully converted into such problems. Rather it is important to note that classification seems to be a rather ubiquitous problem solving process, and a number of real world problems can be thought of as having a large classification component. Further, classification has been one of the more tractable problems for knowledge-based system technology to handle at this point in its development.