Application of Empirical Discovery in Knowledge Acquisition

Jan M. Żytkow    Jieming Zhu
Computer Science Department, Wichita State University
Wichita, KS 67208    U.S.A.
e-mail: zytkw@wsuiar.wsu.ukans.edu

ABSTRACT

There is no doubt that the most fundamental method of knowledge acquisition is discovery, but the AI subfield of Knowledge Acquisition neither studies nor uses discovery methods. We argue that machine discovery is approaching the stage at which it can be useful to knowledge acquisition in two ways: as a source of useful techniques, and as a model of unified knowledge representation and application. We present the discovery system FAHRENHEIT and we discuss its various real-world applications: automated experimentation and discovery in a chemistry laboratory, mining databases for useful knowledge, and others, demonstrating FAHRENHEIT's potential as a knowledge acquisition aid. Finally, we discuss the new developments in the area of discovering basic laws and hidden structure, and we note that automation of modeling would close the cycle of automated knowledge acquisition and application.

1 Introduction

Despite progress, the state of the art in knowledge acquisition is characterized by an amalgamation of incompatible techniques, lack of standards, and narrow applications (Boose & Gaines 1989). In our search for a solution to these problems, we should consider the best mechanism for knowledge acquisition developed by humanity: modern science and engineering. Science and engineering offer us a unifying knowledge organization spanning many levels: data, empirical regularities, models, basic laws, and principles. Science and engineering offer proven methods of acquiring and verifying knowledge. Although the scientific metaphor is becoming popular in the field of knowledge acquisition, and although
researchers acknowledge that building a knowledge-based system resembles the construction of a scientific theory, the scientific method has been used only in a very limited way for knowledge acquisition and representation. There are several reasons. First, although separate elements of the scientific method have been reconstructed, many important elements are still missing. Second, although many attempts have been made at integration of various discovery capabilities (IDS: Nordhausen & Langley 1990; BLAGDEN: Slee-\text{man, Stacey, Edwards, & Gray 1989; FAHRENHEIT: Zyt-\text{kow 1987; Langley & Zyt-\text{kow 1989}), each integration is limited to only a few elements of the scientific method. Third, the applications of machine discovery are believed to be still in the research phase.}

We concentrate on the third argument. Although no-one objects to the claim that science and the scientific method are worth understanding and automating, the feasibility of the task raises doubts, in both the short and long run. To show that practical results are possible we summarize the FAHRENHEIT project, which is focused on multi-dimensional experimentation and data analysis. We demonstrate the breadth and importance of real-world applications of FAHRENHEIT. Then, in response to the first argument, we summarize progress that has been made in the last few years on automation of other elements of scientific method. We concentrating especially on discovery of basic laws and discovery of hidden structure. Finally, we consider automation of knowledge application. Although automated model construction is the key to the acquisition of practical knowledge (Morik 1989; Zyt-\text{kow & Lewenstam 1990), so far little has been done in that area and significant progress is needed.

In a short run, FAHRENHEIT can be viewed as another incompatible technique added to the repertoire of knowledge acquisition methods. But from a larger perspective, the results in the area of machine discovery fall into a master plan, as gradually reconstructed components of a unified method of experimental science. This approach is guaranteed to succeed because scientific method and scientific knowledge representation form a solid and proven theoretical platform for knowledge acquisition.

We concentrate on automated discovery rather than on user support, exemplified by systems such as BLIP (Morik 1989; Wrobel 1989), QuMAS (Mozetic 1987) and DISCI-\text{PLE (Kodratoff & Tecuci 1989).}

2 Discovery of patterns in empirical data

In this section we review the research program in machine discovery based on FAHRENHEIT. Because we summarize a large research program and many results, the review will