Trans-Pacific machine learning research: The Calgary/Waikato axis

Ian H. Witten

Department of Computer Science, University of Calgary, Calgary T2N 1N4 Canada
Department of Computer Science, University of Waikato, Hamilton, New Zealand
Email: ian@cpsc.ucalgary.ca

The following five contributions summarize selected research projects on machine learning that are being undertaken in the Computer Science Departments at the Universities of Calgary and Waikato:

- The WEKA machine learning workbench (Bob McQueen *et al.*);
- Knowledge-rich induction of classification rules (Brent Martin);
- LINUS: a transformation-based system for function discovery (Thong Phan);
- Modeling sequences (Craig Nevill-Manning *et al.*);
- Instructible agents (Dave Maulsby).

Most of the authors are graduate students: the exceptions are Bob McQueen, who is a faculty member at Waikato, and his co-authors who are research assistants. Many have been fortunate enough to visit the other group: Dave Maulsby and Thong Phan visited Waikato from Calgary for several months last year; Brent Martin and Craig Nevill-Manning are visiting Calgary from Waikato this year.

The twin themes that unite these projects are *application* of machine learning techniques to real-world problems, and the benefits of *interaction* in expediting learning.

The WEKA workbench forms part of a project on the development, integration, and application of machine learning techniques, which places special emphasis on application to economically important problems in the New Zealand agricultural industries. The project has stimulated some work on the development of machine learning techniques. It has demonstrated the necessity for extending similarity-based learning to incorporate some background knowledge into an interactive system for inducing classification rules, which has prompted Brent Martin to extend the MARVIN framework for interactive learning. Moreover, it has shown the desirability of inducing numeric functions as a component of real-world learning systems. Along these lines, Thong Phan's LINUS has proved capable of discovering rather complex single-valued real functions from examples. Interaction again plays a crucial role: the key problem—that numerical operations such as differentiation and root-finding inevitably involve approximation and introduce noise—has been tackled by allowing the system to request examples interactively when appropriate.

Dave Maulsby's work on instructible agents that are programmable by example, by demonstration, and by natural-language user hints, is another real-world application of machine learning techniques, this time in the service of effective human-computer interaction. Dave has experimented extensively with an advanced Wizard-of-Oz prototype of an instructible agent called TURVY, with numerous prototypes and mock-ups of various interface aspects collectively called MOCTEC, and with an implementation of an actual learning program called CIMA. Related to programming by example, Craig Nevill-Manning's work began as a mechanism for inferring the structure of sequences used in demonstrating tasks, but the techniques he developed turned out to have interesting applications in their own right, ranging from text compression to the inference of graphic instance structures for ray-tracing.