Machine Learning, 43, 5–6, 2001

Guest Editorial

The papers in this issue were selected from among those presented at the 8th International Conference on Inductive Logic Programming (ILP98) held in Madison, Wisconsin, July 22 to 24, 1998. This is the first time that there is a special issue of the Machine Learning Journal originating from an Inductive Logic Programming Conference.

To create this special issue, all authors of ILP98 were given the possibility of submitting an expanded version of their conference papers which then went through the standard refereeing process of the Machine Learning journal. The five papers contained in this issue were selected out of a total of 11 submissions, and the final versions printed here were thoroughly revised based on reviewer comments.

As can be seen from the topics of the selected papers, this special issue exemplifies the intimate relationship among the fields of Machine Learning and Inductive Logic Programming in terms of techniques, methodology and applications.

Two papers in the issue are dealing with the upgrading of well-known propositional learning techniques and tasks to first-order representations. The paper by Sašo Džeroski, Luc De Raedt and Kurt Driessens, "Relational Reinforcement Learning" addresses the generalization problem in reinforcement learning. The novelty of relational reinforcement learning is that an inductive logic programming algorithm (the TILDE-RT system) is employed for generalization. More precisely, TILDE-RT is combined with Q-learning in order to obtain Q-trees, which are first order regression trees that encode the Q-function. Due to the use of a first order representation, relational reinforcement learning can cope with complex and structured state spaces. The authors demonstrate that relational reinforcement learning offers advantages over more traditional approaches in such domains using experiments in the block's world.

The paper by Tamás Horváth, Stefan Wrobel and Uta Bohnebeck, "Relational Instance Based Learning with Lists and Terms" deals with the issue of upgrading propositional instance-based approaches. They show how a first-order similarity measure capable of handling lists and terms can naturally be defined based on the algorithmically well-studied concept of edit distances, and investigate the measure's theoretical and computational properties. The distance measure is at the heart of RIBL2, a first-order instance-based learner which is empirically evaluated on the problem of recognizing signal structure in mRNA molecules. Experiments are carried out demonstrating that in this domain, the capability of using lists and term-based representations allows remarkable improvements in predictive accuracy compared to earlier approaches.

The paper by Marcel Turcotte, Stephen Muggleton and Michael Sternberg, "The Effect of Relational Background Knowledge On Learning of Protein Three-Dimensional Fold Signatures" also deals with a problem from molecular biology. Their paper provides a detailed and in-depth discussion of the application area, experimentally investigates the questions of whether ILP really provides significant advantages in such domains, and provides insightful discussion about why this turns out to be case. Three experiments are carried out with propositional and first-order background knowledge, and the above question is posed in a scientific way with precise null hypotheses that are to be rejected by the experiments. Since a direct comparison with a state-of-the-art propositional learner is made (C5.0), this paper certainly will be of interest to many readers.

The remaining two papers show interesting algorithmical advances in an area that is relatively new for ILP: dealing with text documents and learning of language.

The paper by Mark Craven and Seán Slattery, "Relational Learning with Statistical Predicate Invention: Better Models for Hypertext" deals with an important application domain of machine learning: learning about web hypertext documents. Their paper exemplifies that predicate invention techniques from ILP provide an elegant and general method to combine the power of first-order learners with propositional techniques: during the search for clauses of the first-order learner, the current clause defines an example set for the propositional learner (here Naive Bayes with feature selection) which is then used to learn a new feature that is then available to the first-order learner's top-level search. As shown in the paper, this combined method is better than either of its constituents alone on three tasks: classifying hypertext pages, learning relations between pages and finding pieces of information in the internal structure of page.

The paper by Dimitar Kazakov and Suresh Manandhar, "Unsupervised Learning of Word Segmentation Rules with Genetic Algorithms and ILP", falls within the Logic, Language and Learning Paradigm, as it studies word segmentation within the ILP framework. The technique proceeds in two steps. In the first step, a genetic algorithm (guided by bias) performs word segmentation on an unannotated corpus. In the second step, the results of the first step are input into a novel inductive logic programming system, CLOG, and used to induce word segmentation rules. The resulting hypothesis is a first order decision list. The authors of this paper show that it is possible to induce meaningful word segmentation rules from a limited corpus of unannotated words.

We finally would like to thank all authors for their contributions, and all the reviewers of the papers for their help in bringing this issue to its current form. Last but not least, our special thanks goes to the executive editor Robert Holte for his advice and to Lorenza Saitta for editing the papers co-authored by one of us. We hope you will enjoy reading this issue as much as we enjoyed preparing it.

Luc De Raedt
C. David Page
Stefan Wrobel