An Experimental Comparison of Genetic and Classical Concept Learning Methods

Gabriella Kókai[1]
[1]Department of Programming Languages
Friedrich-Alexander University
Martensstr. 3. D-91058 Erlangen, Germany
e-mail: kokai@informatik.uni-erlangen.de

Zoltán Tóth[1,2], **Szilvia Zvada**[1]
[2]Department of Informatics
University of Szeged
Árpád tér 2, H-6720 Szeged, Hungary
e-mail: {zntoth, zvada}@inf.u-szeged.hu

1 INTRODUCTION

In this work the classical learning methods *C4.5* [Qui93] and *FOIL* [Qui90] are compared with the genetic learning systems *GEA* (Generic Evolutionary Programming Library, [Tot01]) and *GeLog* (Genetic Logic Programming, [Kok01]). Two problems were involved in the comparison: A mushroom classification and a chess endgame problem. The experiences show that the evolutionary methods not only reach the performance of the traditional learning systems but in complex tasks even outperform them.

2 THE TESTED SYSTEMS

GEA is an evolutionary optimizer tool, with implementation of *EAs* and *ESs*. Due to the applied plug-in technology, it is easily extendible with new individual representation forms and evolutionary algorithms.

In contrast to *GEA*, the *GeLog* system alloys the robust searching technique of genetic algorithms with the learning approach of *ILP* (inductive logic programming). It evolves Prolog programs by means of background knowledge and positive and negative examples.

FOIL is a relational learning system developed in the framework of ILP. It learns Horn clauses by considering their coverage on the training data which is expressed as relations.

The last system partaken in the comparison is the decision tree learner *C4.5*. It builds decision trees by the guidance of information-based heuristics from the attribute-value representation of the training examples.

3 THE LEARNING TASKS

The first learning task in the comparison is the recognition of poisonous mushrooms from 22 attributes, sometimes with missing values. The database contains the descriptions of 8124 mushrooms, 48.2% of which are poisonous. In the second task, the goal is to classify board configurations for the chess endgame situation *white king and rook against black king* (KRK). A configuration is considered positive if white wins immediately or the result of the game is draw. 10.1% of the 28056 examples in the database are positive.

4 RESULTS

All of the tested systems achieved basically the same classification accuracy on the simpler mushroom classification problem. On the more complicated KRK problem, the genetic learners overperformed *C4.5* and produced rule sets that contain significantly less rules than the results of both classical greedy learning systems. The price of the simpler and more general hypotheses is the longer execution time of the learning processes.

References

[BM98] Blake, C. L., Merz, C. J.: *UCI Repository of Machine Learning Databases.* http://www.ics.uci.edu/~mlearn/MLRepository.htm

[Kok01] Kókai, G.: *GeLog - A System Combining Genetic Algorithm with Inductive Logic Programming.* In Proceedings of the 7th Fuzzy Days, Springer-Verlag pp. 326–344 LNCS 2206, 2001

[Qui90] Quinlan, J. R.: *Learning Logical Definitions from Relations.* Machine Learning, 5 pp. 239–266, 1993

[Qui93] Quinlan, J. R.: *C4.5: Programs for Machine Learning.* Morgan Kaufmann, San Mateo, CA, 1993

[Tot01] Tóth, Z., Kókai, G.: *An Evolutionary Optimum Searching Tool.* in Proceedings of IEA/AIE-2001, Springer-Verlag LNAI 2070 pp. 19–24, 2001