Machine Learning, 37, 239–240 (1999)

Guest Editors' Introduction

JONATHAN BAXTER jon@syseng.anu.edu.au
Computer Sciences Laboratory, Australian National University, Canberra, Australia 0200

NICOLÒ CESA-BIANCHI cesabian@dsi.unimi.it
Department of Information Sciences, University of Milan, Milan, Italy

The papers in this issue were selected from among those presented at the *Eleventh Annual Conference on Computational Learning Theory,* held in Madison, Wisconsin, on July 24–26, 1998. The authors submitted expanded versions of their conference papers which went through the standard review process of Machine Learning before appearing here.

These five papers provide a sample of the variety of research themes in learning theory: from the more abstract—showing well developed connections with other fields including probability theory, statistics and logic—to the more "applied", where theoretical analysis of algorithms is complemented by the study of their empirical performance on real-world problems.

In "Learning Function-Free Horn Expressions", Khardon shows that certain classes of Horn expressions are exactly learnable using equivalence and membership queries. This result builds on a key lemma showing that the Angluin-Frazier-Pitt algorithm for learning propositional Horn expression can be lifted to first order logic. The resulting algorithm is shown to work in different learning models, such as learning from entailment, and learning to reason.

In "Large Margin Classification Using the Perceptron Algorithm", Freund and Schapire introduce and analyze an algorithm for linear classification based on Rosenblatt's Perceptron augmented with kernel functions. The analysis proves that, similarly to Vapnik's support vector machine, the algorithm performs well on data that are linearly separable with a large margin. Experimental results on a standard OCR problem indicates that the performance is slighlty inferior to Vapnik's, but with a substantial saving in programming and computation time.

Schapire and Singer give refined Boosting algorithms for weak hypotheses that produce a "confidence rated prediction", i.e. a real number whose sign indicates the classification and whose magnitude indicates the confidence of the classification. The new algorithms give improved criteria for both choosing the weak hypotheses and assigning their weights in the final linear combination. Building on a previous result of Schapire, Freund, Bartlett and Lee, the authors present a theorem showing how the generalization performance of the combined classifier can be bounded in terms of its margins on the training examples. In particular the performance is independent of the number of weak hypotheses forming the combination. The new algorithms are applied to the multi-class case in a number of different ways, with experimental comparisons between the algorithms completing the paper.

Long's paper, "The Complexity of Learning According to Two Models of a Drifting Environment", is among the more theoretically dense in this collection. The main contributions of this paper are to remove $\log(\frac{1}{\epsilon})$ factors from bounds on the rate of drift tolerable by a PAC-style learner in a drifting environment, and to provide a relatively simple proof of a similar logarithmic factor removal on the number of examples required to learn agnostically in a fixed environment. The former results close the gap (barring constants) between the upper and lower bounds in this model. This paper is particularly noteworthy in that it brings to the theoretical community the technique of "chaining" from the empirical process literature. It is the chaining argument that allows the removal of logarithmic factors in the bounds.

McAllester gives distribution-free bounds for "Bayesian" algorithms—algorithms that predict by weighting concepts according to a posterior distribution—in both the realizable and unrealizable case. The bounds hold for all possible measurable subsets of the concept class, while the key lemma upon which they depend "the quantifier reversal lemma" should be of independent interest.

Thanks to the authors for their contributions, and to the reviewers of the papers for their help in bringing this issue to its current form.