Machine Learning, 25, 115–116 (1996)

Guest Editor's Introduction

THOMAS HANCOCK

This special issue of *Machine Learning* contains papers that describe work presented at the Seventh Annual ACM Conference on Computational Learning Theory (COLT '94), held on July 12-15, 1994, in New Brunswick, New Jersey. The papers were selected with the aim of highlighting results that are both exciting theoretically and broadly interesting to the entire machine learning research community. The latter concern is especially appropriate for COLT '94, which was co-located with the Machine Learning Workshop with the intent of bringing the experimental and theoretical research communities closer together. The papers were invited at the choice of the program committee and went through the standard refereeing process for *Machine Learning*.

Broadly, much of the research in Computational Learning Theory falls into one of three major categories. One is algorithm development, in which a technique is presented to solve some particular learning problem. Another is sample complexity, in which statistical analysis is performed to determine how much data is required to solve some particular learning problem. The third is lower bounds in which some learning problem is shown to require a minimal amount of resources (e.g. computation time or examples) that exceeds some threshold. All three categories are represented in this journal issue.

Ron, Singer, and Tishby present a new learning algorithm directed at sequence understanding problems arising from areas such as language processing and computational biology. They define a new subclass of probabilistic automata, which they name probabilistic suffix automata (PSAs). These automata are limited in memory, like Markov models, but may have differing amounts of memory in different contexts. The author's algorithm takes as input a set of sample strings (or a single long string) and finds a PSA close to optimal (in a sense similar to Valiant's PAC model). The authors present impressive applications of this scheme to correcting white noise in text and modeling DNA strands.

Frazier and Pitt present an algorithm that solves a learning problem for the CLASSIC description language. CLASSIC describes first order logics of the kind commonly used as knowledge representations in AI systems. Using membership and equivalence queries (the model pioneered by Angluin), the authors give a polynomial time algorithm that learns descriptions in this language. Applying the techniques of learning theory to learn the representations that system builders actually use is an important research direction.

Haussler, Kearns, Seung, and Tishby apply techniques from statistical physics to study the learning curves that describe how the accuracy of a learned function increases with the quantity of sample data available for training. Earlier work in learning theory has emphasized the Vapnik-Chervonenkis (VC) dimension of a class of functions as a parameter characterizing the sample complexity of learning within that class. But VC analysis gives worst case sample complexity and fails to explain many learning curve phenomena observed in practice, such as phase transitions. The authors apply techniques from statistical mechanics to obtain more precise bounds on learning curves, including a variety of interesting functional forms, in the case where the distribution of samples is known. Rigorous sample complexity results that are able to explain many of the phenomena that practitioners observe are an exciting new development.

Pillaipakamnatt and Raghavan prove new lower bounds on the computation time required to learn certain classes of disjunctive normal form boolean formulas (DNF). They show that a learner who gets random examples from an unknown formula within these classes and runs in polynomial time will be unable to produce the target formula (or a close approximation) from the class. As with other similar work, the results hold based on certain assumptions from complexity theory such as P$\neq$NP. The authors develop several elegant new techniques that allow them to strengthen previous results from the literature. This is a significant step in the long standing attempts to characterize the learnability of such basic classes of boolean formulas.

I thank the authors and reviewers for their excellent work that made this issue possible. I also thank program chair Manfred Warmuth and the rest of the program committee for their valuable help in selecting papers from a strong field.