Machine Learning, 18, 127–129 (1995)

Guest Editor's Introduction

This special issue of Machine Learning is devoted to the Sixth Annual ACM Conference on Computational Learning Theory (COLT '93) which was held in Santa Cruz, California on July 26–28, 1993. These papers were selected from those presented at the 1993 COLT conference (ACM Press, 1993) based on both the significance of their results from a theoretical viewpoint and the relevance of the results to the machine learning community. While the authors were invited to submit their papers to this special issue, all papers went through the standard refereeing procedures of this journal.

I would like to thank the authors and the referees for their help in making timely publication of this special issue possible. This issue could not have been produced without them. In addition to this special issue, other selected papers from the 1993 COLT conference will be appearing in a special issue of the *Journal of Computer and System Sciences.*

As the field of computational learning theory has matured, the breadth of topics addressed has continued to grow, and the papers in this special issue give a sampling of such breadth. For readers who would like to obtain some general background about computational theory there are three textbooks (Kearns & Vazirani, 1994; Anthony & Biggs, 1992; Natarajan, 1991) as well as two survey articles written by Dana Angluin (Angluin, 1992, 1993). There are also special issues of *Machine Learning* devoted to previous COLT conferences (Hellerstein, 1994; Li & Valiant, 1994; Blumer & Case, 1992; Pitt, 1990).

The field of computational learning theory began with the seminal paper of Valiant (1984) in which he introduced the PAC (probably approximately correct) model of learning. There has been a tremendous amount of research related to this model. In the PAC model, a major contribution to the understanding of sample complexity was made by Blumer et al. (Blumer, Ehrenfeucht, Haussler & Warmuth, 1989). Building on the work of Vapnik and Chervonenkis (Vapnik & Chervonenkis, 1971), they showed that the combinatorial parameter of the VC-dimension of a concept class essentially characterizes the needed sample complexity. In particular, if you can show the VC-dimension of a concept class is polynomial in the relevant parameters, then any hypothesis from the class that is consistent with a polynomial size sample will satisfy the PAC criteria. While polynomial upper bounds on the VC-dimension are easily obtained for discrete concept classes, little was known about what general conditions guarantee polynomial bounds on the VC-dimension for non-discrete classes. The first paper of the special issue, by P.W. Goldberg and M.R. Jerrum, shows that for non-discrete classes the VC-dimension is polynomially bounded if the containment of an instance in a concept is testable in polynomial time or if the criterion for membership of an instance in a concept can be expressed as a formula with fixed quantification depth and exponentially-bounded length.

Another central issue with regard to the PAC model is whether or not learning algorithms can be made robust against noise. While there has been tremendous progress in this area, there are still many interesting and important open questions to answer. Recently, Kearns (Kearns, 1993) has introduced the statistical query model and shown that any algorithm that can be formulated in this model is also PAC learnable with random classification noise in the random examples. However, much less is known about the variant of the PAC model in which there are membership queries that are incorrectly answered. The second paper in this issue, by D. Ron and R. Rubinfeld, presents a polynomial time algorithm that uses corrupted random examples and membership queries to learn an arbitrarily good approximation to an unknown deterministic finite state automaton where the random examples are drawn from the uniform distribution on the inputs. In particular, they consider the model of noise in which both the answers to the queries and random examples are corrupted with independently distributed errors with an error rate bounded away from $1/2$, and these errors are persistent.

While there is a fairly well developed theory for concept learning, where each example is classified as either a positive or negative example of the target concept, very little is known about the theory of learning real-valued functions. In particular, there have been few results about general properties of function learning and only a few non-trivial function classes for which positive results had been exhibited. The third paper of this issue, by P. Auer, P.M. Long, W. Maass, and G.J. Woeginger, presents new results on the complexity of function learning in the most common non-probabilistic models of on-line learning. They provide a combinatorial max-min definition that characterizes the optimal learning cost for learning a class of real-valued functions. They also relate the cost for learning a union of function classes to the learning costs for the individual function classes. Furthermore, they give an efficient algorithm for learning convex piecewise linear functions from $\Re^d$ into $\Re$. They also give a sufficient condition for learnability that they can apply to obtain efficient learning algorithms for a number of additional non-trivial classes of functions from $\Re$ to $\Re$.

In addition to exploring the learnability of various concept classes under the existing learning models, a central problem in computational learning theory is to further develop and investigate new learning models and to formalize learning tasks into new learning problems so that they can be studied in a formal setting. The final two papers in this special issue fall into these categories. The fourth paper of this issue, by M. Betke, R.L. Rivest, and M. Singh, introduces the problem of learning a graph by a piecemeal search, in which the learner must occasionally return to its starting point (say, to refuel). They present two linear-time piecemeal-search algorithms for learning grid graphs with rectangular obstacles. Recently, Baruch, et al. (Baruch, Betke, Rivest, & Singh, 1994) have given an algorithm that will explore, in a piecemeal manner, every vertex and edge in an arbitrary graph $G = (V, E)$ in $O(|E| + |V|^{1+o(1)})$ time. There has also been some recent work on learning graphs (Bender & Slonim, 1994).

The fifth paper of this issue, by M. Kearns and H.S. Seung, considers the question of how one can make optimal use of multiple independent runs of a mediocre learning algorithm. They introduce a new formal model in which the learning algorithm must combine a collection of potentially poor but statistically independent hypotheses in order to approximate an unknown target function arbitrarily well. A second motivation for this study is the setting in which many hypotheses are obtained by a distributed population of identical learning agents and one wants to combine them to obtain an agent that is significantly better than any individual in the population.

Once again, let me thank all of the authors and reviewers for their help. I hope you enjoy reading the papers contained in this special issue, and I look forward to hearing about the future research that builds upon them.

Sally A. Goldman
Department of Computer Science
Washington University
St. Louis, MO 63130

References

ACM Press (1993). *Proceedings of the Sixth Annual ACM Conference on Computational Learning Theory.* Santa Cruz, CA: ACM Press.

Angluin, D. (1992). Computational learning theory: Survey and selected bibliography. In *Proceedings of the Twenty-Fourth Annual ACM Symposium on Theory of Computing,* 351–369. Victoria, BC: ACM Press.

Angluin D. (1993). Learning with Queries. In E.B. Baum, (Ed.), *Computational Learning & Cognition: Proceedings of 3rd NEC Research Symposium.* Proceedings in Applied Mathematics 64. Philadelphia, PA: Society for Industrial and Applied Mathematics.

Anthony, M. & Biggs, N. (1992). *Computational Learning Theory: An Introduction.* Cambridge, England: Cambridge University Press.

Awerbuch, B., Betke, M., Rivest, R.L. & Singh, M. (1994). How to do BFS without teleportation. Unpublished manuscript.

Bender, M.A. & Slonim, D.K. (1994). The power of team exploration: Two robots can learn unlabeled directed graphs. *Proceedings of the 35th Annual Symposium on Foundations of Computer Science,* 75–85. Santa Fe, NM: IEEE Computer Society Press.

Blumer, A., & Case. J. (Eds.). (1992). [Special Issue on Computational Learning Theory] *Machine Learning, 9* (2/3).

Blumer, A., Ehrenfeucht, A., Haussler, D., & Warmuth, M.K. (1989). Learnability and the Vapnik-Chervonenkis Dimension. *Journal of the Association for Computing Machinery, 36,* 929–965.

Hellerstein, L. (Ed.). (1994). [Special Issue on Computational Learning Theory] *Machine Learning, 17* (2/3).

Kearns, M. J. (1993). Efficient noise-tolerant learning from statistical queries. *Proceedings of the 25th Annual ACM Symposium on Theory of Computing,* 392–401. San Diego, CA: ACM Press.

Kearns, M.J. & Vazirani, U.V. (1994). *An Introduction to Computational Learning Theory.* Cambridge, MA: The MIT Press.

Li, M., & Valiant, L. G. (Eds.). (1994). [Special Issue on Computational Learning Theory] *Machine Learning, 14* (1).

Natarajan, B.K. (1991). *Machine Learning: A Theoretical Approach.* San Mateo, CA: Morgan Kaufmann.

Pitt, L. (Ed). (1990). [Special Issue on Computational Learning Theory] *Machine Learning, 5* (2).

Valiant, L.G. (1984). A Theory of the Learnable. *Communications of the ACM, 27,* 1134–1142.

Vapnik, V.N., & Chervonenkis, A. Ya. (1971). On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability and its Applications, 16,* 264–280.