Introduction

Welcome to this special issue of Machine Learning devoted to the Third Workshop on Computational Learning Theory held in Rochester, New York on August 6–8, 1990. As one can see from the complete papers in this issue, and even more readily from the extended abstracts presented in the Proceedings (Fulk & Case, 1990), the field of Computational Learning Theory continues to expand and to address a wider variety of issues of interest to the machine learning community in general. In selecting papers for this issue we have tried to choose papers for their interest to this larger community, so this should not be thought of as a compilation of best papers, though we believe these papers to be among the best. We would, therefore, like to use part of this space to draw attention to some other papers presented at this conference that can be found in the Proceedings and elsewhere.

Rūsiņš Freivalds delivered the invited lecture surveying old to recent results (mostly due to Freivalds or Efim Kinber) on the inductive inference or learning in the limit of minimal size or nearly minimal size programs (Fulk & Case, 1990, pp. 3–20). One draws the programs to be learned from some programming system. In most cases the size restrictions limit what can be learned in some way. The *acceptable programming systems* are essentially those into which the programs from any programming system can be translated. All naturally occurring programming systems provide acceptable systems. Non-acceptable systems are known only from mathematical construction. Consider first the cases where the programs to be inferred or learned are merely required to be within a computable function of minimal size. In these cases what can be learned does not depend on the choice of acceptable programming system from which the programs to be learned are drawn, and a surprising technical characterization is provided in terms of special cases of limiting solutions to the program equivalence problem. In the cases in which the learned programs are required to be more strictly minimal size, there is a dependence on the choice of acceptable programming system. For some acceptable systems only finite classes of functions can be so learned; for others, rich infinite classes are so learnable. The comparison of this kind of learning across acceptable systems is shown to provide the algebraic structure of a lower semi-lattice. The *Kolmogorov systems* are those acceptable systems in which translations exist from other systems with linearly bounded growth. For Kolmogorov systems it is shown that one can always learn minimal size programs for some infinite class of functions.

Mark Fulk's paper on robust separation in inductive inference provides a surprising refutation of Barzdin's Conjecture (Fulk, 1992). Self-referential examples are used to witness the truth of many existence theorems in inductive inference, yet many researchers consider these self-referential examples to be unrealistic. J.M. Barzdin had provided a technical way of ruling out such examples and essentially conjectured that many of those existence theorems would no longer hold if his restrictions were added. Fulk, in effect, strengthens Barzdin's

restrictions and shows that the existence results Barzdin conjectured to fail actually hold, even under Fulk's stronger restrictions. There is some indication that Fulk's results may give rise to new and useful inference strategies.

In their paper in this issue, Maass and Turán explore the relationships among various models of learning in which the learner can choose its queries and is guaranteed correct answers. They also give relationships between these models and several combinatorial parameters that measure the complexity of the concept class from which the concept to be learned is chosen. The types of queries allowed are membership queries and equivalence queries, including partial equivalence queries. If membership queries are allowed, the learner can ask the teacher whether any point in the instance space is or is not included in the concept to be learned. (Here concepts are viewed as subsets of the instance space—for a generalization see the description of Yamanishi's paper below.) If equivalence queries are allowed, the learner can propose a concept, and the teacher will respond with "Correct" or give a counterexample. In the most restrictive model, the query hypothesis must be a concept from the concept class being learned. One generalization studied in this paper is how much power is gained from allowing arbitrary hypotheses. Another generalization is to allow partial hypotheses, where some point in the instance space can be labeled as "don't-care." The teacher will either respond with "Correct outside of the don't cares" or give a counterexample that is one of the instances where the learner made a definite choice. This model includes membership queries, since the learner can propose a hypothesis with all points except the one being queried set to don't-care.

In the second paper in this issue, Angluin, Frazier, and Pitt make a significant advance in applications of this area by showing that Horn sentences are learnable in polynomial time if membership and equivalence queries are allowed, even if the equivalence queries must be Horn sentences. A Horn clause is a disjunction of literals in which either all the literals are negated or just one is in its positive form. A Horn sentence is a conjunction of Horn clauses. If this result can be extended to handle universal quantifiers, it will apply to the learning of Prolog programs. This result also gives an example of a separation result of the kind proved in Maass and Turán's paper, since Angluin had previously shown that Horn sentences are not learnable in polynomial time if queries are restricted to only one of the two types.

Another result along these lines can be found in Hancock's paper in the Proceedings (Fulk & Case, 1990, pp. 23–37). He shows that Boolean decision trees in which each variable appears only once are learnable in polynomial time if both membership and equivalence queries are allowed, but not if only one type of query is allowed.

In Valiant's PAC (Probably Approximately Correct) learning model the learner has even less control over the instances observed than in the equivalence query model, since these instances are generated independently from an unknown fixed probability distribution on the instance space. Since some instances or regions of instances may not be seen due to low probability, success is *approximate*, measured with respect to this probability distribution. In a probabilistic setting there is no absolute guarantee that even a high-probability instance will be seen, thus the learner is only *probably* approximately correct. The learning of concepts that are subsets of the instance space has been studied extensively in this model. In the third paper in this issue, Yamanishi further extends the applicability of PAC learning, and brings in new tools from statistics and information theory, by proposing a

generalization to stochastic concepts. He also gives criteria by which stochastic concepts can be said to be PAC-learned, and relations among these criteria. In his model, instances are not merely members or nonmembers of a concept, but each instance has a probability distribution giving the probabilities of its classification into one of a finite number of classes. He gives a learning algorithm based on Rissanen's Minimum Description Length principle. One application given in his paper is to the learning of stochastic decision lists, a generalization of Rivest's decision lists.

In a related paper presented at the conference, Kearns and Schapire explore the learning of these probabilistic concepts where the classifications are limited to positive and negative (Kearns & Schapire, 1992). They generalize the Vapnik-Chervonenkis dimension to classes of these concepts, and show how this characterizes their learnability in case the loss criterion is assumed to be quadratic.

One particular case of learning probabilistic concepts is treated by Abe and Warmuth in the fourth paper in this issue. In their paper the class of distributions is that generated by hidden Markov models and more general classes of probabilistic automata. They give polynomial sample size bounds for learning these classes, but show that in most cases the problem of finding an automaton that is close enough to the one generating the examples is intractable, even if the class of distributions is restricted to those generated by automata with only two states. On the other hand, if the automata are restricted to a class in which the current state can be uniquely determined from the start state and the sequence of output symbols, the problem is solvable in polynomial time.

A new model for learning is proposed by Ben-David, Itai, and Kushilevitz in their paper in the Proceedings (Fulk & Case, 1990, pp. 232-245). They assume that whenever the learner proposes a hypothesis, the teacher responds by giving an approximation of the distance between that hypothesis and the correct concept. They show that learnability of a concept class is equivalent to its Vapnik-Chervonenkis dimension being finite, as in the usual PAC learning model, but that their model eliminates the measurability requirements.

On the more philosophical side, the last paper in this issue considers the relationship of machine learning to the scientific method. Osherson, Stob, and Weinstein explore the formalization of the scientific method within the context of first-order logic. They formalize a scientist who proposes new hypotheses that are not just logical consequences of the current theory and data, but that are believable and interesting (assuming the existence of an oracle or recursively enumerable function to determine this latter quality) in this context. They show that their method is universal in the sense that it is successful in every situation where success is possible.

We would like to thank all those who contributed to making this a very enjoyable and productive workshop, and especially the authors of the papers in this issue for taking so much time expanding and revising their papers. We would also like to thank the referees for reading the papers carefully and providing many helpful comments. Special thanks are also due to David Haussler and Ron Rivest for special help preparing this issue.

John Case
University of Delaware

Anselm Blumer
Tufts University

References

- Fulk, M.A. (1992). Robust separations in inductive inference. *SIAM J. Computing*, to appear.
- Fulk, M.A. & Case, J. (eds). (1990). *Proceedings of the Third Annual Workshop on Computational Learning Theory*. San Mateo, CA: Morgan Kaufmann Publishers, Inc.
- Kearns, M.J. & Schapire, R.E. (1992). Efficient distribution-free learning of probabilistic concepts, *JCSS*, to appear.