Machine Learning, 30, 5–6 (1998)

Guest Editor's Introduction

DANA RON
Laboratory of Computer Science, MIT, Cambridge, MA 02139

This special issue of Machine Learning is devoted to the Ninth Annual Conference on Computational Learning Theory (COLT '96), which was held in Desenzano del Garda, Italy, on June 28th–July 1st, 1996. All but one of the papers in this issue are full versions of extended abstracts that were presented at the conference and appear in the conference's proceedings. All papers went through the standard reviewing process of Machine Learning.

In the first paper, Long and Tan describe and analyze an algorithm for learning axis-aligned rectangles from multiple-instance examples when the examples are distributed according to a product distribution. In the multiple-instance learning model, each example consists of several elements of the domain, and the label of the example is positive if at least one of these elements belongs to the target concept, and negative otherwise. This model, which is motivated by a problem in drug discovery, was introduced by Dietterich, Lathrop and Lozano-Perez. The central technique employed by Long and Tan is running in parallel several algorithms for learning certain probabilistic concepts, and combining the hypotheses these algorithms construct into a single (deterministic) hypothesis.

Blum and Kalai describe a general reduction from the problem of learning from multiple-instance examples to PAC-learning with one-sided classification noise. Therefore, every concept class that is efficiently learnable from noisy examples is efficiently learnable from multiple-instance examples. They also describe a more efficient reduction to learning in the Statistical-Query model. By using the latter reduction, they get an algorithm for learning axis-aligned rectangles from multiple-instance examples with fairly low sample and time complexities. This work was not presented in COLT '96 but rather was triggered by Long and Tan's work described above. Thus it provides an example of the development of a research direction in the computational-learning-theory community since COLT '96.

Antos and Lugosi prove lower bounds for learning several natural and well-studied concept classes such as linear halfspaces, and polyhedra with a bounded number of faces. The novelty in their results is that the lower bounds they present are of a stronger type than those previously known. Namely, previous results were of the following form: For every sample size m and learning algorithm (which uses m examples), there exists a distribution on the examples and a target concept, such that the error of the learning algorithm (with respect to the distribution and target) is $\Omega(d/m)$, where d is the VC-dimension of the concept class. Antos and Lugosi show, for several k-parameter concept classes (where in most cases $k = \Theta(d)$), that for every family of learning algorithms (one for every sample size), there exists a *fixed* distribution and a *fixed* target concept, such that for infinitely many sample sizes m, the error of the algorithm is $\Omega(k/m)$. Their lower bounds, as opposed to the former ones, provide us with information concerning the error decrease as a function of m for a worst-case but fixed choice of distribution and target concept.

Cohen proves hardness results for a "dual" DFA learning problem. In this problem the examples are deterministic finite automata, and the concepts are strings, where each string corresponds to the set of automata that accept it. This result, which is representation-independent and based on cryptographic assumptions, implies the hardness of several more natural learning problems such as learning the description logic CLASSIC from subconcepts.

Birkendorf, Dichterman, Jackson, Klasner, and Simon, further the study of learning with restricted focus of attention (RFA). In the k-RFA learning model, the learner is allowed to view only k of the n attributes of each example, where this set of attributes is determined by the learner. A motivating example for this model is that of medical diagnosis of a disease. In this example, the attributes chosen correspond to (possibly costly) medical tests that can be performed on given patients. This paper continues to explore the relationship between the PAC and RFA learning models and presents several results in the latter model. In particular the authors develop an information theoretic characterization of the RFA model and use this characterization to prove hardness results. By combining some of their results they show that as opposed to the PAC model, in the RFA model, weak learning does not imply strong learning.

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