Machine Learning, 43, 171–172, 2001

Guest Editor's Introduction

The papers in this special issue were selected from conferences on learning theory that were held from 1997 through 1999. The authors submitted expanded versions of their conference papers that included additional material and elaborated proofs that did not appear in the conference papers. The papers went through the standard review process of Machine Learning before appearing here.

The five papers represent top quality current research in computational learning theory. The papers focus on theoretical and algorithmic issues in online learning, game theory and boosting. While being abstract and formal, the papers set the ground for more applied research that will use the algorithmic results for designing improved classification learning tools. The papers also draw connections to other fields including statistics and probability theory, game theory, and information theory.

A short version of the first paper entitled "General Convergence Results for Linear Discriminant Updates" by Grove, Littlestone, and Schuurmans appeared in the Tenth Annual Conference on Computational Learning Theory, held in Vanderbilt University, Nashville, Tennessee, on July 6–9, 1997. The paper, which attracted significant attention at the conference, studies the problem of learning linear discriminants in the online mistake bound model. The authors define and analyze a general family of "quasi-additive" algorithms. This family includes, as special cases, the well studied Perceptron and Winnow algorithms. The new family also includes new algorithms that interpolate between additive-update algorithms like the Perceptron and multiplicative-update algorithms as Winnow.

The second paper entitled "Relative Loss Bounds for On-line Density Estimation with the Exponential Family of Distributions" by Azoury and Warmuth is based on a paper that appeared in the Fifteenth Conference on Uncertainty in Artificial Intelligence, held in Stockholm, Sweden, from July 30th through August 1st, 1999. There are close connections between the paper by Grove, Littlestone, and Schuurmans and this paper. Azoury and Warmuth analyze in the mistake bound model the problem of online density estimation. The main tool used in the paper is the Bregman divergence which was extensively studied by researchers in information theory. Using the Bregman divergence the authors provide a general analysis for prediction and regression algorithms in probabilistic settings.

The last three papers in this special issue are based on papers that appeared in the Twelfth Annual Conference on Computational Learning Theory, held in the University of California at Santa Cruz, on July 7–9, 1999. The first paper of the three, entitled "Worst-case Bounds for Logarithmic Loss of Predictors" by Cesa-Bianchi and Lugosi, is also concerned with online learning of sequences. The authors study the problem of prediction using experts' advice and, similar to Azoury and Warmuth, use the logarithmic loss to measure the performance of a master algorithm that combines the experts' advice. Cesa-Bianchi and Lugosi build on a result by Shtarkov to prove an upper bound on the excess loss, called the regret, of the master algorithm and apply the bound to parametric and non-parametric classes of prediction algorithms.

In "Drifting Games", Schapire studies online learning in a game theoretic setting. The paper introduces a game played between two players called the shepherd and the adversary. The author describes an algorithm for this game and proves an upper bound on its performance. This algorithm subsumes Freund's boost-by-majority, Freund and Schapire's Adaboost algorithm, and Cesa-Bianchi et al.'s Binomial Weights algorithm. Schapire also describes an efficient implementation of the shepherd algorithm that might prove to be useful in machine learning applications.

The last paper by Freund called "An Adaptive Version of the Boost by Majority Algorithm" provides a generalization of boost-by-majority and Adaboost in a different direction. Freund gives an adaptive version, motivated by Brownian motion, of boost-by-majority by looking at the limit where on each boosting iteration the booster makes an infinitesimal change to a distribution over the examples it is provided with. The result is a noise-robust algorithm that can be especially useful in boosting using hypotheses of high VC-dimension.

Thanks to the authors for their contributions, and to the reviewers of the papers their thorough work.

Yoram Singer
School of Computer Science & Engineering
Hebrew University, Jerusalem, Israel
singer@cs.huji.ac.il