Machine Learning, 51, 215–216, 2003

Guest Editor's Introduction

This special issue collects some of the best recent papers in Learning Theory. The authors of four papers from leading theoretical conferences held in 1999 and 2001 were invited to submit expanded versions of their conference papers. These papers then went through the standard review process of Machine Learning Journal. The papers were selected according to both theoretical significance and potential interest for applications.

Altogether these papers represent a snapshot of a variety of current lines of research in Learning Theory, ranging from boosting to on-line learning, from large margin classifiers to theoretical views of recommender systems.

A short version of the first paper, "Boosting and hard-core set construction", by Klivans and Servedio appeared in the proceedings of the 40th Annual Symposium on Foundations of Computer Science (FOCS), held in New York City, on October 17–19, 1999. The paper relates hard-core set construction methods (hardness amplification methods from Computational Complexity) to boosting. In particular, the authors show that the hard-core set construction by Impagliazzo (which proves the existence of distributions under which boolean functions are "highly" inapproximable) is a boosting algorithm when the initial distribution is uniform. Boosting and hard-core set constructions share two relevant parameters: the "boundedness" of the distributions which are constructed and the number of "stages" required for the construction. As it turns out, the procedures which have been used for hard-core set construction have better "boundedness" and can be exploited to improve algorithms in Learning Theory, while boosting algorithms have smaller number of "stages" and can be exploited to improve hard-core set construction. Using alternate boosting methods Klivans and Servedio give an improved bound for hard-core set construction. Then, using techniques from Impagliazzo, the authors show how to obtain a more efficient version of Jackson's Harmonic Sieve algorithm for learning DNF formulas with membership queries under the uniform distribution.

The remaining three papers in this special issue are based on papers published in the proceedings of the Fourteenth Annual Conference on Computational Learning Theory and the Fifth European Conference on Computational Learning Theory held in Amsterdam, the Netherlands, on July 16–19, 2001.

In "Potential-based algorithms in on-line prediction and game-theory", Cesa-Bianchi and Lugosi offer a unified view of several known algorithms for sequential prediction problems (including the quasi-additive family of Grove, Littlestone and Schuurmans), for playing iterated games (including Freund and Schapire's Hedge algorithm and the Λ-strategies by Hart and Mas-Colell), and for boosting (including Freund and Schapire's AdaBoost). The authors show that all such algorithms are instances of a general sequential decision strategy centered on the notion of potential. By analyzing this strategy the authors re-derive known performance bounds and prove new bounds, as simple consequences of a single general analytic result. The authors relate potential-based analyses in Learning Theory to their counterparts independently developed in Game Theory, discuss a very general notion of regret, and prove bounds for a generalization of a method of adaptive game playing due to Hart and Mas-Colell.

In “Estimating the optimal margins of embeddings in Euclidean half spaces”, Forster, Schmitt, Simon and Suttorp use the singular value decomposition of the matrix representing a concept class to determine the optimal margin of embeddings of classes such as singletons and half intervals in homogeneous Euclidean half spaces. The singular value decomposition can not only be used to construct optimal embeddings, but also to prove the corresponding best possible upper bounds on the margin. This line of research addresses the question which concept classes can be embedded in half spaces with a large margin, and clearly connects to the literature about maximal margin classifiers and kernel methods.

Loosely speaking, a recommender system, also called a collaborative filtering system, exploits the opinions of past users (“experts”) to make recommendations to new users. The users are asked to rate a few items before any new item is recommended. Once the number of rated items is large enough, the system starts using such ratings to estimate which previous users of the system are most similar to the current user. The opinions of these similar users can then be used by the system to make recommendations to the current user. In “A theoretical analysis of query selection for collaborative filtering”, Dasgupta, Lee and Long study the problem of selecting the initial items for the users to rate, i.e., the problem of generating questions to ask the user. The authors show an algorithm requiring, in a worst-case setting, an essentially optimal number of queries. This algorithm runs in time polynomial in the relevant parameters of the problem. The analysis is complemented by proving that no polynomial-time algorithm can ask a significantly smaller number of queries in the worst-case (up to standard complexity-theoretic assumptions). Dasgupta, Lee and Long also study a more general case where the user ratings are from a finite set and a loss function (a metric) on this finite set is used to measure the distance between ratings. The authors use a variant of the membership query model similar to the one proposed by Angluin et al. and, within this model, they introduce and analyze a generalization of Seung et al.’s query-by-commitee algorithm. The proposed algorithm is polynomial in the relevant parameters. The bound on the number of queries is a seemingly tight function of both the optimal worst-case number of queries for this model and the “geometry” of the set of experts.

I am grateful to the authors for their contributions, and to the anonymous reviewers for their valuable help in bringing the papers to their current form. Thanks to Rob Holte for inviting me to edit this special issue, and to Karen Cullen, Melissa Fearon and other members of the editorial staff for their support.

Claudio Gentile
Dipartimento di Informatica e Comunicazione
Università dell’Insubria
Via Mazzini, 5, 21100 - Varese, Italy
gentile@dsi.unimi.it