Special Issue on New Theoretical Challenges in Machine Learning

Avrim Blum · Philip M. Long

Received: 12 September 2014 / Accepted: 22 September 2014 / Published online: 26 September 2014
© Springer Science+Business Media New York 2014

The papers in this special issue represent some exciting current directions in the theoretical study of machine learning. The papers were selected by special invitation, and further reviewed using the usual rigorous *Algorithmica* process. We aimed to select papers that brought mathematical analysis to bear to guide the design of machine learning algorithms in new ways. As you will see, significant progress is being made in unsupervised learning of models of using latent variables to capture structure in the data, in new forms of bounds that describe in more detail how algorithms and data interact, and in active algorithms to reduce human labeling effort. Also, theoretically-inspired tools for algorithm design are being brought to bear on some unexpected new problems that substantially expand the range of scenarios where they can be applied.

In “A Spectral Algorithm for Latent Dirichlet Allocation”, Hsu et al describe a provably efficient algorithm for learning topic models in which an observation can be an instance of multiple topics; their analysis applies in the practically popular case of Latent Dirichlet Allocation. “Provable ICA with Unknown Gaussian Noise, and Implications for Gaussian Mixtures and Autoencoders”, by Arora et al, presents new algorithmic tools for Independent Component Analysis, a popular model for blind signal separation; these tools lead to algorithms with strong theoretical guarantees. In “Randomized partition trees for nearest neighbor search”, Dasgupta and Sinha provide new tools for analyzing recently proposed practical variants of the $k - d$ tree, that capture the amenability of inputs to the new algorithms, enabling analysis that departs from the worst case. In “Randomized Algorithms for Low-Rank Matrix Factoriza-

A. Blum (✉)

Department of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213, USA
e-mail: avrim@cs.cmu.edu

P. M. Long

Microsoft, 1020 Enterprise Way, Sunnyvale, CA 94089, USA
e-mail: plong@microsoft.com

tions: Sharp Performance Bounds”, Candès and Witten perform a new analysis of a randomized algorithm for low-rank approximation of matrices previously proposed by Martinsson, Rokhlin and Tygert, along with lower bounds that match up to the leading constants. Balcan and Feldman, in “Statistical Active Learning Algorithms for Noise Tolerance and Differential Privacy,” develop a modification to the Statistical Query model of Kearns that allows design of label-efficient noise-tolerant active learning algorithms that also have strong privacy guarantees. Daskalakis, Diakonikolas and Servedio, in “Learning Poisson Binomial Distributions,” provide an algorithm for the fundamental problem of learning sums of independent Boolean variables, called Poisson Binomial Distributions, with excellent bounds on the amount of data and computation required.

These papers represent exciting new directions on algorithms for machine learning guided by theory. With the growing impact of machine learning technology on society, and so much interesting development in this rich and fascinating subject, it is a great time to be in the field.