

Learning DNF from random walks

Nader H. Bshouty^{a,1}, Elchanan Mossel^{b,2}, Ryan O'Donnell^{c,3}, Rocco A. Servedio^{d,*,4}

^aDepartment of Computer Science, Technion, USA

^bDepartment of Statistics, University of California, Berkeley, USA

^cInstitute for Advanced Study, Princeton, NJ, USA

^dDepartment of Computer Science, Columbia University, 1214 Amsterdam Avenue, Mailcode 0401, NY 10027, USA

Received 5 January 2004; received in revised form 24 September 2004

Available online 8 December 2004

Abstract

We consider a model of learning Boolean functions from examples generated by a uniform random walk on $\{0, 1\}^n$. We give a polynomial time algorithm for learning decision trees and DNF formulas in this model. This is the first efficient algorithm for learning these classes in a natural passive learning model where the learner has no influence over the choice of examples used for learning.

© 2004 Elsevier Inc. All rights reserved.

Keywords: Computational learning theory; Noise sensitivity; Disjunctive Normal Form formulas; DNF; Random walks

* Corresponding author.

E-mail addresses: bshouty@cs.technion.ac.il (N.H. Bshouty), mossel@stat.berkeley.edu (E. Mossel), odonnell@theory.lcs.mit.edu (R. O'Donnell), rocco@cs.columbia.edu (R.A. Servedio).

¹ Supported by the Sydney Goldstein Research Fund 120-122 and the fund for promotion of research at the Technion, Research No. 120-025.

² Supported by a Miller Postdoctoral Fellowship.

³ Supported by NSF Grant 99-12342. This research was performed while the author was in the Department of Mathematics at MIT.

⁴ Supported by an NSF Mathematical Sciences Postdoctoral Fellowship and by NSF Grant CCR-98-77049.

1. Introduction

1.1. Motivation

One of the most notorious open questions in computational learning theory is whether it is possible to efficiently learn Boolean formulas in disjunctive normal form, or DNF, from random examples. This question was first posed by Valiant [36] in his seminal paper which formalized the Probably Approximately Correct (PAC) model of learning from independent random examples, and has remained stubbornly open ever since. DNF formulas achieve an attractive balance between expressiveness and clarity: any Boolean function can be represented by a sufficiently large DNF, yet DNF formulas are easily understood by humans and seem to be a natural form of knowledge representation.

Provably correct and efficient algorithms for learning DNF from random examples would be a powerful tool for the design of learning systems, and over the past two decades many researchers have sought such algorithms. Despite this intensive effort, the fastest algorithms to date for learning polynomial size DNF formulas in Valiant's original PAC model of learning (where the learner receives independent examples drawn from an arbitrary probability distribution over $\{0, 1\}^n$) run in time $2^{\tilde{O}(n^{1/3})}$ [26]. Even if we only consider learning under the uniform distribution, the fastest known algorithms for learning polynomial size DNF from independent uniform examples run in time $n^{O(\log n)}$ [37].

Since learning DNF formulas from random examples seems to be hard, researchers have considered alternate models which give more power to the learning algorithm. The most popular of these is the model of *learning from membership queries*; in this model the learner has access to a black-box oracle for the function to be learned and thus can determine the value of the function on any inputs of its choice. Several polynomial time algorithms have been given for learning in this enhanced model. Kushilevitz and Mansour [28] gave a polynomial time membership query algorithm which can learn any polynomial size decision tree under the uniform distribution (i.e., the error of the final hypothesis is measured with respect to the uniform distribution on $\{0, 1\}^n$). Building on the work of [28], Jackson [20] gave a polynomial time algorithm for learning polynomial size DNF formulas under the uniform distribution using membership queries.

While learning from membership queries is interesting in its own right, it represents a significant departure from traditional “passive” models of learning (such as the PAC model) in which the learning algorithm has no control over the data which it receives; the assumption that a learning algorithm can actively make queries is a strong one which may limit the usefulness of membership query learning algorithms. Thus an important goal is to design efficient algorithms for learning DNF formulas in natural “passive” learning models. Towards this end, researchers have considered several alternatives to the standard uniform distribution PAC model of learning from independent uniform random examples. Bshouty and Jackson [9] defined a model where the learner can access a uniform quantum superposition of all labelled examples, and showed that DNF formulas can be efficiently learned in this framework. More recently Bshouty and Feldman [8] showed that DNF can be efficiently learned in a model called $SQ\text{-}\mathcal{D}_\rho$, which is intermediate in power between standard uniform distribution learning and uniform distribution learning with membership queries; in this model the learner is allowed to make statistical queries about the target function under product distributions of the learner's choosing. While Bshouty and Feldman showed that this model is strictly weaker than the membership query model, it is still an “active” learning model since the learner selects the various distributions which will be used.

Download English Version:

<https://daneshyari.com/en/article/10333017>

Download Persian Version:

<https://daneshyari.com/article/10333017>

[Daneshyari.com](https://daneshyari.com)