Proceedings of Machine Learning Research vol 65:1–4, 2017

Open Problem: Meeting Times for Learning Random Automata

Benjamin Fish BFISH3@UIC.EDU
Lev Reyzin LREYZIN@MATH.UIC.EDU
Department of Mathematics, University of Illinois at Chicago

Abstract

Learning automata is a foundational problem in computational learning theory. However, even efficiently learning random DFAs is hard. A natural restriction of this problem is to consider learning random DFAs under the uniform distribution. To date, this problem has no non-trivial lower bounds nor algorithms faster than brute force. In this note, we propose a method to find faster algorithms for this problem. We reduce the learning problem to a conjecture about meeting times of random walks over random DFAs, which may be of independent interest to prove.

Keywords: automata theory, PAC learning, uniform distribution, social choice theory

1. Introduction

We are motivated by understanding the complexity of learning deterministic finite automata. The question of learning automata, or regular languages, has been studied since the advent of computational learning theory. While automata are known to be learnable in various query models (e.g. MQ+EQ (Angluin, 1987) and label queries (Angluin et al., 2009)), cryptographic lower bounds give strong evidence against their efficient learnability in Valiant's PAC model (Pitt and Warmuth, 1990).

One natural easier task is to design algorithms for learning automata on average. In particular, one can ask whether automata with random transitions and accept/reject behavior can be efficiently PAC-learned. This restriction, however, still presents barriers – namely, even uniformly random automata have super-polynomial statistical query dimension (Angluin et al., 2010). Even when in addition the target distribution over input strings is uniform, no algorithms better than brute-force are known, except when additional information is also given (Angluin and Chen, 2015).

Towards the goal of faster algorithms for learning random DFAs under the uniform distribution, we reduce the learning problem via the powerful tools of boolean Fourier analysis to a conjecture about random walks over random DFAs. We propose proving (or disproving) this conjecture as an avenue towards progress on the learning problem.

Our conjecture may also be interesting in its own right and posits a statement about the meeting time of random walks on DFAs, which is also related to the concept of synchronizing strings over DFAs.

2. Definitions

Here, we consider the uniform distribution over n-state DFAs, so the transitions between states are chosen uniformly at random, as are whether each state is an accept state or a reject state, and we will call this distribution D. The goal is to PAC-learn uniformly random DFAs under the uniform distribution over inputs in time poly(n). That is, we wish to return a hypothesis that has at most

ϵ error with respect to the unknown DFA except with probability at most δ, where the probability is taken both over the sample the learning algorithm receives and D. For m the length of the input strings, even if we are allowed to choose m as a function of n where $m = \text{poly}(n)$, nothing faster than brute force is known when each input string is chosen uniformly at random (which we will denote as U_m). Note that there are $\tilde{O}(2^{n \log n})$ DFAs on n states, so merely improving to an algorithm that takes $O(2^n)$ time remains an open problem.

To achieve this goal, we suggest using the tools of boolean Fourier analysis. Namely, for any function $f : \{-1,1\}^m \to \mathbb{R}$, we may write it as $f(x) = \sum_{S \subseteq [m]} \hat{f}(S)\chi_S(x)$, where $\chi_S(x) = \Pi_{i \in S} x_i$ and $\hat{f}(S)$ is a unique real number called the Fourier coefficient of f on S. The set of all such coefficients is called the Fourier spectrum of f.

Definition 1 (ϵ-concentration) *Let $\mathscr{F}$ be a collection of subsets $S \subseteq [m]$. The Fourier spectrum of $f : \{-1,1\}^m \to \mathbb{R}$ is ϵ-concentrated in $\mathscr{F}$ if*

$$\sum_{S \notin \mathscr{F}} \hat{f}(S)^2 \leq \epsilon.$$

Theorem 2 *For unknown f, suppose an algorithm A can identify a collection $\mathscr{F}$ of subsets on which f's Fourier spectrum is $\epsilon/2$-concentrated. Then using poly$(|\mathscr{F}|, m, 1/\epsilon, 1/\delta)$ additional time (and samples), A can learn f under the uniform distribution with error at most ϵ.*

This approach was pioneered by Linial et al. (1993). We will show that a random DFA has a concentrated Fourier spectrum by showing that most influences of the random DFA are small, as defined below. For any string $x \in \{-1,1\}^m$, call $x^{\oplus i} = (x_1, \ldots, x_{i-1}, -x_i, x_{i+1}, \ldots, x_m)$.

Definition 3 (ith influence) *The ith influence of a function $f : \{-1,1\}^m \to \{-1,1\}$ is*

$$\operatorname*{Inf}_i[f] = P_{x \sim U_m}[f(x) \neq f(x^{\oplus i})] = \sum_{S \ni i} \hat{f}(S)^2.$$

Influence was first introduced in the theory of social choice by Ben-Or and Linial (1985). See O'Donnell (2014) for a proof that these two definitions of the ith influence are equivalent.

3. Partial results and previous work

Note that showing that the ith influence of a DFA is small is both a statement about the concentration of f and a statement about a random walk on the graph of a DFA. That is, the ith influence will be small if the following holds: given a random walk on the DFA and the same walk where the ith bit is flipped, these two walks are likely to end in both accept states or both reject states.

In particular, if the walks coincide at any time after i, then they will land in the same state. Related to this is the concept of a 'synchronizing' string, which is a string that brings all of the states of a DFA to the same state. If a DFA has such a string, it is said to be synchronizing. A long-standing open problem is the maximum length of a synchronizing string, raised by Černỳ and others (Černỳ, 1964). Černỳ's conjecture states that all synchronizing DFAs have a synchronizing string of length no more than $(n-1)^2$. For recent results on random DFAs see for example the work of Skvortsov and Zaks (2010). However, in our case, we don't need the string to be the same for every set of states, but instead just need a random string to bring together two states.

Also closely related is the concept of a meeting time of a random walk. In the simplest version, the meeting time of random walks starting from states u and v is the number of steps it takes for the two walks to first meet at any vertex. Bounds on expected meeting time on random walks in various settings have appeared in the literature (Aldous, 1991; Cooper et al., 2013; Kanade et al., 2016). Closest to our conjecture is the work of Cooper et al. (2009), who describes expected meeting times on random r-regular, undirected graphs for $r \geq 3$. However, we need high-probability results on the meeting time on random 2-regular *directed* graphs (so that the out-degree but not necessarily in-degree is always two, i.e. it is a DFA).

Namely, we conjecture that with high probability the meeting time is $O(n)$. To state this precisely, for any DFA f, let $\delta(q, x)$ be the function that applies the string x on that DFA starting at state q and returns the state that it ends in.

Conjecture 4 *There is some constant $c > 0$ such that for any two states u and v, and any $\lambda > 0$,*

$$P_{f\sim D}\left[P_{x\sim U_{cn}}[\delta(u,x) \neq \delta(v,x)] \leq O\left(\frac{1}{n^\lambda}\right)\right] \geq 1 - o_n(1).$$

We propose this conjecture as a concrete open problem; proving it would lead to progress on learning DFAs in this setting.

Specifically, using this conjecture, we can show that a random DFA with high probability has many small influences, implying that its Fourier spectrum is sufficiently concentrated:

Proposition 5 *Assume that Conjecture 4 holds with constant c. Consider $x \in \{-1,1\}^m$ to be a uniformly random string and $f : \{-1,1\}^m \to \{1,1\}$ a uniformly random DFA on n states. There is a $2^{O(n)}$-time learning algorithm for random DFAs under the uniform distribution over examples of length $m = \Omega(n)$ that needs at most $2^{O(n)}$ samples for any $\epsilon = 1/poly(n)$.*

Proof Using Theorem 2, it suffices to give a collection of subsets of size $2^{O(n)}$ that is $\epsilon/2$-concentrated with high probability. Let $\mathscr{F} = \{S : S \subseteq \{\lceil m - cn\rceil, \ldots, m\}\}$.

$$\sum_{S\notin\mathscr{F}} \hat{f}(S)^2 \leq \sum_{i<m-cn}\sum_{S\ni i}\hat{f}(S)^2 = \sum_{i<m-cn}\operatorname*{Inf}_i[f].$$

Then it suffices to show that for $i < m - cn$, $\text{Inf}_i[f] \leq \frac{\epsilon}{2(m-cn)}$.Consider $i < m - cn$, setting $m > cn$. Calling q_0 the start state and $x|_k$ the first k bits of x, if $\delta(q_0, x|_{i+cn}) = \delta(q_0, x^{\oplus i}|_{i+cn})$, then certainly $f(x) = f(x^{\oplus i})$: Then the flip of the bit at the ith position is 'corrected' to the same state, at which point the rest of the strings x and $x^{\oplus i}$ will end in the same state, at which point $f(x) = f(x^{\oplus i})$. Let $u = \delta(q_0, x|_i)$ and $v = \delta(q_0, x^{\oplus i}|_i)$. Conjecture 4 implies that with high probability,

$$P_{y\sim U_{cn}}[\delta(u,y) \neq \delta(v,y)] \leq 1/n^\lambda$$

for any $\lambda > 0$, which immediately implies that

$$P_{x\sim U_m}[f(x) \neq f(x^{\oplus i})] \leq P_{x\sim U_m}[\delta(q_0, x|_{i+cn}) \neq \delta(q_0, x^{\oplus i}|_{i+cn})] \leq \frac{\epsilon}{2m} \leq \frac{\epsilon}{2(m-cn)}$$

for $\epsilon \geq \Omega\left(\frac{1}{\text{poly}(n)}\right)$.

■

References

David J Aldous. Meeting times for independent markov chains. *Stochastic Processes and their applications*, 38(2):185–193, 1991.

Dana Angluin. Learning regular sets from queries and counterexamples. *Inf. Comput.*, 75(2):87–106, 1987.

Dana Angluin and Dongqu Chen. Learning a random DFA from uniform strings and state information. In *Algorithmic Learning Theory - 26th International Conference, ALT 2015, Banff, AB, Canada, October 4-6, 2015, Proceedings*, pages 119–133, 2015.

Dana Angluin, Leonor Becerra-Bonache, Adrian-Horia Dediu, and Lev Reyzin. Learning finite automata using label queries. In *Algorithmic Learning Theory, 20th International Conference, ALT 2009, Porto, Portugal, October 3-5, 2009. Proceedings*, pages 171–185, 2009.

Dana Angluin, David Eisenstat, Leonid Kontorovich, and Lev Reyzin. Lower bounds on learning random structures with statistical queries. In *Algorithmic Learning Theory, 21st International Conference, ALT 2010, Canberra, Australia, October 6-8, 2010. Proceedings*, pages 194–208, 2010.

Michael Ben-Or and Nathan Linial. Collective coin flipping, robust voting schemes and minima of banzhaf values. In *26th Annual Symposium on Foundations of Computer Science, Portland, Oregon, USA, 21-23 October 1985*, pages 408–416, 1985.

Ján Černỳ. Poznámka k homogénnym experimentom s konečnỳmi automatmi. *Matematicko-fyzikálny časopis*, 14(3):208–216, 1964.

Colin Cooper, Alan M. Frieze, and Tomasz Radzik. Multiple random walks in random regular graphs. *SIAM J. Discrete Math.*, 23(4):1738–1761, 2009.

Colin Cooper, Robert Elsässer, Hirotaka Ono, and Tomasz Radzik. Coalescing random walks and voting on connected graphs. *SIAM J. Discrete Math.*, 27(4):1748–1758, 2013.

Varun Kanade, Frederik Mallmann-Trenn, and Thomas Sauerwald. On coalescence time in graphs-when is coalescing as fast as meeting? *CoRR*, abs/1611.02460, 2016.

Nathan Linial, Yishay Mansour, and Noam Nisan. Constant depth circuits, fourier transform, and learnability. *J. ACM*, 40(3):607–620, 1993.

Ryan O'Donnell. *Analysis of boolean functions*. Cambridge University Press, 2014.

Leonard Pitt and Manfred K. Warmuth. Prediction-preserving reducibility. *J. Comput. Syst. Sci.*, 41 (3):430–467, 1990.

Evgeny S. Skvortsov and Yulia Zaks. Synchronizing random automata. *Discrete Mathematics & Theoretical Computer Science*, 12(4):95–108, 2010.