An Introduction to Random DFA

Borja Balle

McGill University

October 29, 2013

Deterministic Finite Automata (DFA)

DFA is $A = \langle \Sigma, Q, \tau, q_0, F \rangle$

- Σ finite alphabet
- Q set of states
- $\tau : Q \times \Sigma \to Q$ transitions
- $q_0 \in Q$ initial state
- $F \subseteq Q$ final states

Computes $A : \Sigma^\star \to \{0, 1\}$

$$A(x) = \mathbb{I}[\tau(q_0, x) \in F]$$

$A^{-1}(1) = L_A \subseteq \Sigma^\star$ is the *language recognized* by A

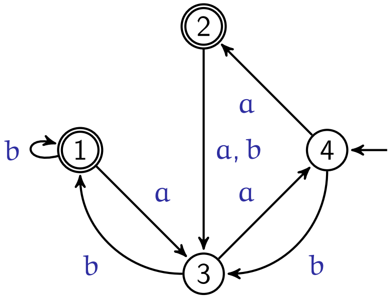
Example DFA

$\Sigma = \{a, b\}$ $\quad q_0 = 4$

$Q = \{1, 2, 3, 4\}$ $\quad \tau(1, a) = 3$

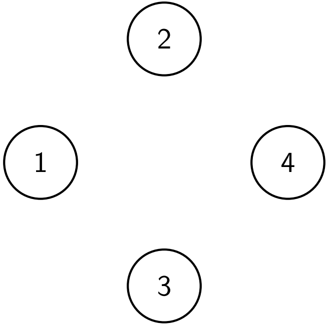
$F = \{1, 2\}$ $\quad \tau(2, b) = 3$

$$A(baa) = 1$$



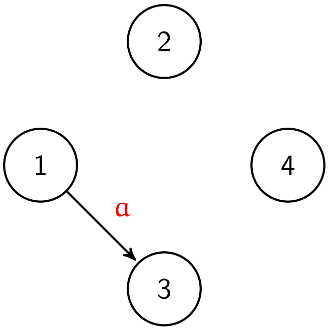
Generating a DFA at Random

Fix $\Sigma = \{a, b\}$ and $Q = \{1, 2, 3, 4\}$, and choose at *random* τ, q_0, and F



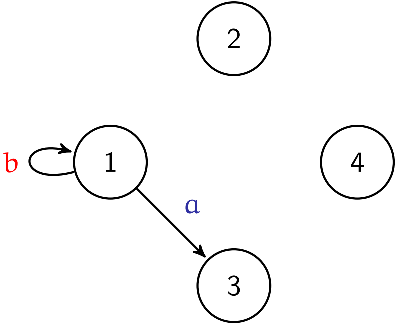
Generating a DFA at Random

Fix $\Sigma = \{a, b\}$ and $Q = \{1, 2, 3, 4\}$, and choose at *random* τ, q_0, and F



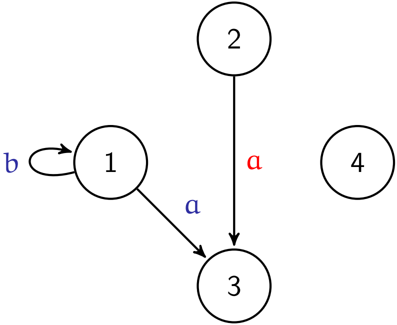
Generating a DFA at Random

Fix $\Sigma = \{a, b\}$ and $Q = \{1, 2, 3, 4\}$, and choose at *random* τ, q_0, and F



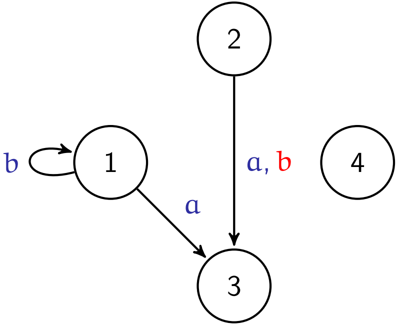
Generating a DFA at Random

Fix $\Sigma = \{a, b\}$ and $Q = \{1, 2, 3, 4\}$, and choose at *random* τ, q_0, and F



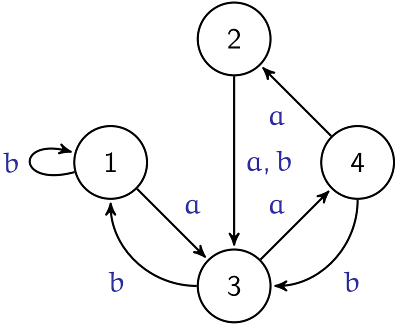
Generating a DFA at Random

Fix $\Sigma = \{a, b\}$ and $Q = \{1, 2, 3, 4\}$, and choose at *random* τ, q_0, and F



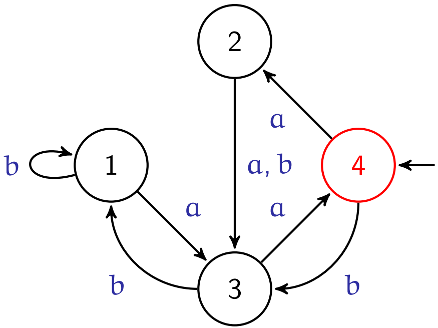
Generating a DFA at Random

Fix $\Sigma = \{a, b\}$ and $Q = \{1, 2, 3, 4\}$, and choose at *random* τ, q_0, and F



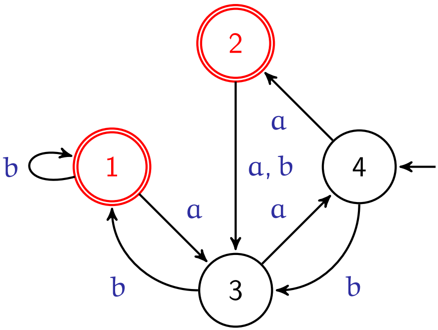
Generating a DFA at Random

Fix $\Sigma = \{a, b\}$ and $Q = \{1, 2, 3, 4\}$, and choose at *random* τ, q_0, and F



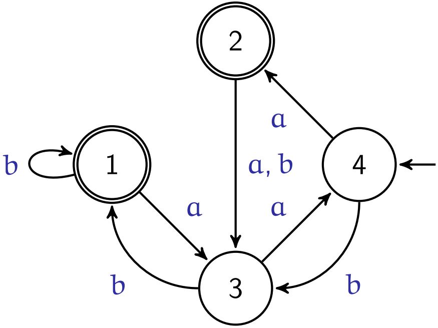
Generating a DFA at Random

Fix $\Sigma = \{a, b\}$ and $Q = \{1, 2, 3, 4\}$, and choose at *random* τ, q_0, and F



Generating a DFA at Random

Fix $\Sigma = \{a, b\}$ and $Q = \{1, 2, 3, 4\}$, and choose at *random* τ, q_0, and F



Note: All choices are independent

Outline

1. Motivation for Studying Random DFA
2. Some Properties of Generic DFA
3. Two Proofs of Grusho's Theorem

Outline

1. Motivation for Studying Random DFA
2. Some Properties of Generic DFA
3. Two Proofs of Grusho's Theorem

Why Study Random DFA?

Consider the following assertions . . .

- Almost all DFA with n states are not minimal
- The average running time of Moore's DFA minimization algorithm is $O(n \log\log n)$ for DFA with n states
- Typical DFA with n states can be learned in time $\mathrm{poly}(n)$

Why Study Random DFA?

Consider the following assertions . . .

- Almost all DFA with n states are not minimal
- The average running time of Moore's DFA minimization algorithm is $O(n \log \log n)$ for DFA with n states
- Typical DFA with n states can be learned in time $\mathrm{poly}(n)$

Why Study Random DFA?

Consider the following assertions . . .

- Almost all DFA with n states are not minimal
- The average running time of Moore's DFA minimization algorithm is $O(n \log \log n)$ for DFA with n states
- Typical DFA with n states can be learned in time $\mathrm{poly}(n)$

Why Study Random DFA?

Consider the following assertions ...

- Almost all DFA with n states are not minimal
- The average running time of Moore's DFA minimization algorithm is $O(n \log \log n)$ for DFA with n states
- Typical DFA with n states can be learned in time $\mathrm{poly}(n)$

Why Study Random DFA?

Consider the following assertions . . .

- Almost all DFA with n states are not minimal
- The average running time of Moore's DFA minimization algorithm is $O(n \log\log n)$ for DFA with n states
- Typical DFA with n states can be learned in time $\mathrm{poly}(n)$

Leit Motif

Generic/Average Case Bounds *vs* Worst Case Bounds

PAC Learning DFA — Setup

- A minimal DFA with n states over Σ
- D probability distribution over $\Sigma^\star$
- $S = ((x^1, A(x^1)), \dots, (x^m, A(x^m))$ sample with i.i.d. $x^i \sim D$

Problem

Give algorithm L such that with probability $\geqslant 1 - \delta$ the output $\hat{A} = L(S)$ is computed in time $\mathrm{poly}(m)$ and satisfies $\mathbb{P}_{x \sim D}[A(x) \neq \hat{A}(x)] \leqslant \varepsilon$ whenever $m \geqslant \mathrm{poly}(n, |\Sigma|, 1/\varepsilon, 1/\delta)$

PAC Learning DFA — Some Bounds

Negative Results

[Pitt–Warmuth '93] Assuming P $\neq$ NP

No poly-time algorithm can approximate the minimum DFA/NFA consistent with S within a polynomial factor

[Kearns–Valiant '89] Assuming RSA is secure

No poly-time algorithm can PAC learn DFA

Positive Results

[Clark–Thollard '04]

Every DFA A with n states can be learned under distributions D_A with support L_A in time $\text{poly}(n, 1/\mu_{D_A})$

PAC Learning DFA — The Random Approach

Observations

- DFA used in lower bounds are *far from random*: acyclic or with single final state
- *Adapting* distribution to target is very restrictive

Questions

- Are random DFA easier to learn than arbitrary DFA?
- Are there distributions under which most DFA can be learned?

State of The Art (Based on [Jackson–Servedio '03, Sellie '09, Angluin et al. '10])

	Random DT	Random DNF	Random DFA
PAC (dist. free)	?	?	?
PAC/SQ (uniform)	✓	✓	?[1]
SQ (dist. free)	×	×	×

[1]Positive empirical evidence: [Lang '92] and competitions Abbadingo One, Gowachin, GECCO '04, Stamina, Zulu

Learning Random DFA under the Uniform Distribution

- A *random DFA* with n states over Σ
- D *uniform distribution* over Σ^T for some $T \geqslant 1$
- $S = ((x^1, A(x^1)), \dots, (x^m, A(x^m)))$ with i.i.d. $x^i \sim D$

Conjecture

The exists an algorithm L such that with probability $1 - o(1)$ over the choice of A, on input S produces an output $\hat{A}$ in time $\mathrm{poly}(m, T)$ such that $\mathbb{P}_{x \sim D}[A(x) \neq \hat{A}(x)] \leqslant \varepsilon$ with probability $\geqslant 1 - \delta$ whenever $m \geqslant \mathrm{poly}(n, |\Sigma|, T, 1/\varepsilon, 1/\delta)$

Note: The regime $T = O(\log n)$ is trivial

Outline

Definition of Generic Property

Given DFA $A = \langle \Sigma, Q, \tau, q_0, F \rangle$

- $|\Sigma| = r \geqslant 2$, usually a fixed constant
- $|Q| = n$, the regime of interest is $n \to \infty$
- $\mathcal{U}_{r,n}$ uniform distribution over $\mathcal{DFA}(r, n)$

Definition

We say that *generic DFA over* r *symbols satisfy property* P if the following holds when $n \to \infty$:

$$\mathbb{P}_{A \sim \mathcal{U}_{r,n}}[P(A)] = 1 - o(1)$$

Diameter of Random DFA

The diameter of a DFA is the minimum d such that:

if $q' \in \tau_\star(q) = \bigcup_{x \in \Sigma^\star} \{\tau(q, x)\}$, then
there exists $x \in \Sigma^{\leqslant d}$ such that $q' = \tau(q, x)$

Theorem (Trakhtenbrot–Barzdin '70)

Generic DFA have diameter at most $(1 + C_r + o(1)) \log_r n$, where C_r is a constant depending on r such that $C_r \to 0$ as $r \to \infty$

Reachability in Random DFA

The reachability is the number of states $|\tau_\star(q_0)|$ reachable from the initial state

Theorem (Grusho '73, Carayol–Nicaud '12, Berend–Kontorovich '13)

Generic DFA have reachability $n(c_r + o(1))$, where c_r is the positive solution of $c = 1 - e^{-rc}$

Examples of c_r

r	2	3	4	5	6	7
c_r	0.796	0.940	0.980	0.993	0.997	0.999

Note: Same result proved in the form of CLT [G73,CN12] and concentration bound [BK13]

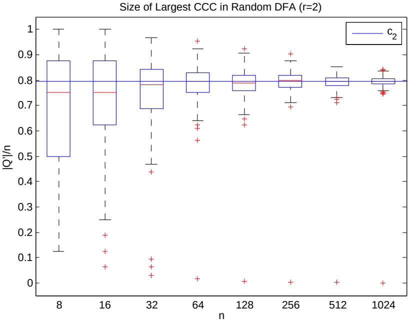
Communication Classes of Random DFA

A closed communication class (CCC) is a set of states $Q' \subseteq Q$ such that:

$$\tau_\star(Q') = \bigcup_{q \in Q'} \tau_\star(q) = Q'$$

Theorem (Grusho '73) (also follows from Berend–Kontorovich '13)

Generic DFA have a unique CCC and its size is $n(c_r + o(1))$, where c_r is the positive solution of $c = 1 - e^{-rc}$



Random Walks on Random DFA

By Grusho's theorem, the random walk starting at q_0 and ending in $\tau(q_0, x)$ with x uniform over Σ^T will end inside the CCC for large enough T

A CCC $Q' \subseteq Q$ is k-periodic if $Q' = Q_0 \sqcup \cdots \sqcup Q_{k-1}$ such that:

$$\text{for all } 0 \leqslant i \leqslant k-1 \text{ one has } \tau_\star(Q_i) = Q_{i+1 \bmod k}$$

Otherwise it is aperiodic

Theorem (Balle '13)

The unique CCC in a generic DFA is aperiodic

Note: This implies that random walks on random DFA are ergodic

Minimization of Random DFA

The minimal size of a DFA A is the size of a minimal automata accepting the same language as A

Theorem (Berend–Kontorovich '13)

Generic DFA have minimal size $n(c_r + o(1))$, where c_r is the positive solution of $c = 1 - e^{-rc}$

Running Time of DFA Minimization Algorithms

Algorithm	Worst-case	Average-case	
Hopcroft	$O(n \log n)$	$O(n \log \log n)$	[David '10]
Moore	$O(n^2)$	$O(n \log \log n)$	[David '10]
Brzozowski	exponential	super-polynomial	[De Felice–Nicaud '13]

Synchronization of Random DFA

A DFA has a reset word of length l if there exists $x \in \Sigma^l$ such that:

$$\text{for all } q' \in Q \text{ one has } \tau(q', x) = q \text{ for the same } q \in Q$$

A DFA with a reset word is called synchronizing

Conjecture (Černý '64)

Every synchronizing DFA has a reset word of length at most $(n-1)^2$

Theorem (Skvortsov–Zaks '10)

- Generic DFA on alphabets of size $r > 18 \log n$ have reset words of length less than $3n^2 \log n$.
- Generic DFA on alphabets of size $r > n^{1/2+\epsilon}$ satisfy Černý's conjecture.

Note: [SZ10] report experiments suggesting generic DFA with $r = 2$ have reset words of length $o(n)$

Outline

1. Motivation for Studying Random DFA

2. Some Properties of Generic DFA

3. Two Proofs of Grusho's Theorem

Proof 1: Kolmogorov Complexity

Whiteboard

— proof courtesy of Ricard Gavaldà —

Proof 2: Differential Equation Method

Whiteboard

Proof 2: Differential Equation Method

Find Reachable States

Input: DFA $A = \langle \Sigma, Q, \tau, q_0, F \rangle$
Output: $\tau_\star(q_0)$
$S \leftarrow \{q_0\}$
$T \leftarrow \{(q_0, \sigma_1), \ldots, (q_0, \sigma_r)\}$
while $T \neq \varnothing$ **do**
  Choose $(q, \sigma) \in T$
  Let $q' \leftarrow \tau(q, \sigma)$
  if $q' \notin S$ **then**
    Let $S \leftarrow S \cup \{q'\}$
    Let $T \leftarrow T \cup \{(q', \sigma_1), \ldots, (q', \sigma_r)\}$
  Let $T \leftarrow T \backslash \{(q, \sigma)\}$
return S

Conclusion

- Exciting research topic with many open problems
- Yields useful insights into analysis of DFA algorithms
- Almost nothing done between the 70's and 2010

An Introduction to Random DFA

Borja Balle

McGill University

October 29, 2013