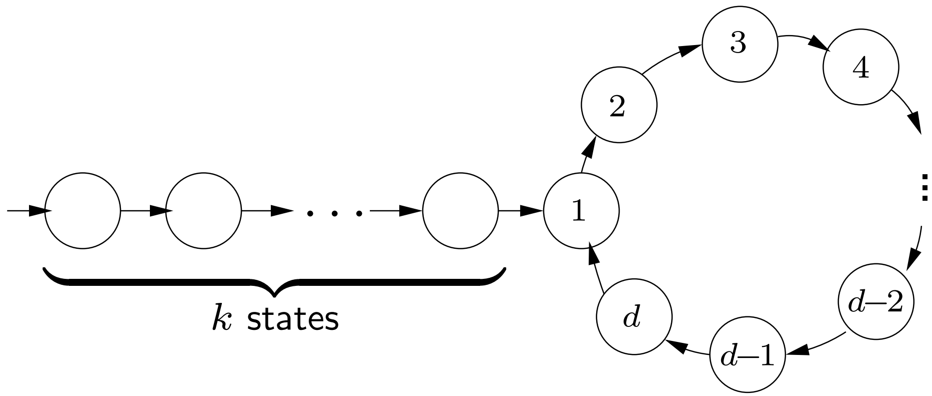
Probabilistic and Nondeterministic Unary Automata

Gregor Gramlich
Institute of Computer Science
Johann Wolfgang Goethe-Universität Frankfurt

August 2003

Unary Regular Languages

- Unary Language $L \subseteq \{a\}^*$.
- Unary DFA



- $\mathrm{period}(L)$ = number of states in the cycle of the minimal DFA.
- Cyclic language for $k = 0$, empty path.
- L cyclic $\Rightarrow$ $\mathrm{period}(L)$ = number of states of the minimal DFA.

Topics

- Determinism versus probabilism:
  - Comparing the number of states.

- Approximating minimal NFA's:
  - Given a unary NFA, how complex is it to determine an equivalent (almost) minimal NFA?
  - Given a unary cyclic DFA, does the possibly exponentially larger input size allow efficient approximation?

Unary PFA's

- Unary PFA $M = (Q, A, \pi, \eta)$.
  - $Q =$ set of states.
  - $A =$ stochastic transition matrix – describes a Markov chain.
  - $\pi =$ initial distribution (stochastic row vector).
  - $\eta =$ vector indicating final states.
- Acceptance probability for input a^j: $\pi A^j \eta$.
- Cutpoint λ specifies the language $L(M, \lambda) = \{a^j \mid \pi A^j \eta > \lambda\}$.
- Cutpoint λ is ϵ-isolated if $\forall j \in \mathbb{N}_0 : |\pi A^j \eta - \lambda| \geq \epsilon$.

Previous Results

Given a PFA, what is the size of the equivalent minimal DFA?

- Fixed isolation for arbitrary alphabets:

  Exponential blow up.

- Fixed isolation for unary alphabet:

  Exponential blow up for the initial path. (Freivalds, 1982)

- Arbitrarily small isolation for unary alphabet:

  Blow up $\Theta(e^{\sqrt{n \ln n}})$ for the cycle. (Milani and Pighizzini, 2000)

A tight polynomial bound for the period

a) For any unary PFA M with n states and ϵ-isolated cutpoint λ

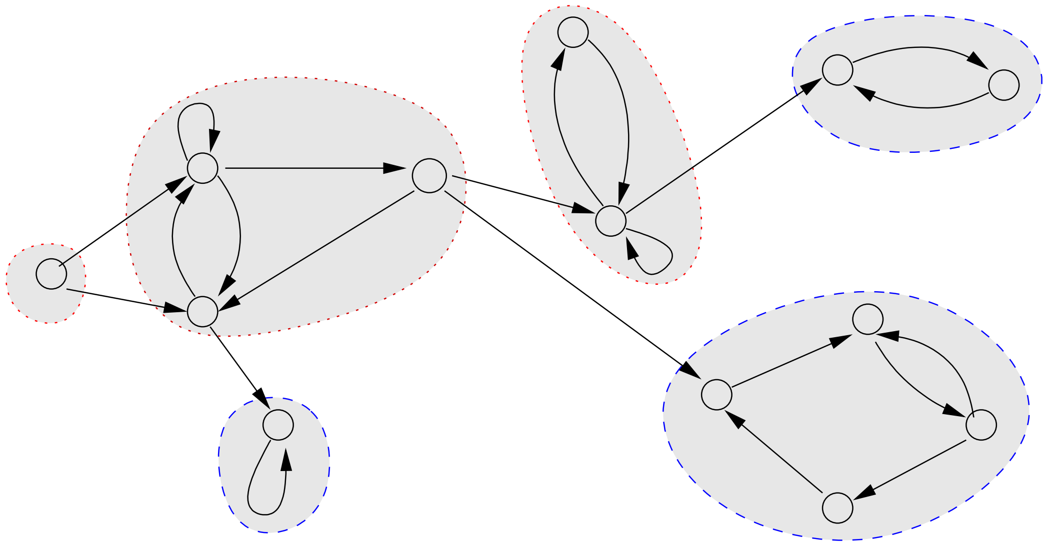
$$\text{period}(L(M,\lambda)) \leq n^{\frac{1}{2\epsilon}}.$$

Polynomial relationship for fixed ϵ.

b) Result is almost tight:
For any $\alpha < 1$ and any ϵ there is a PFA M with n states and ϵ-isolated cutpoint λ, such that

$$\text{period}(L(M,\lambda)) > n^{\alpha\frac{1}{2\epsilon}}.$$

Finite Markov Chain



- Strong components with no outgoing arc are ergodic components.

- Ergodic component B_i:

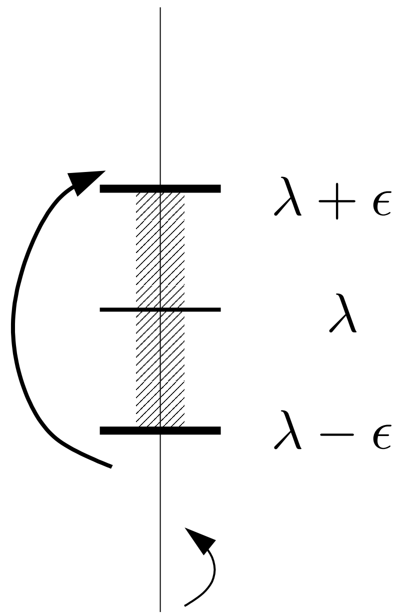
  d_i = period of B_i.
  r_i = absorption probability = prob[B_i is eventually reached].

Behaviour of a Markov Chain in the Long Run

- For a PFA with individual ergodic periods $d_1, d_2, \ldots, d_k$ let $D := \operatorname{lcm}\{d_1, d_2, \ldots, d_k\}$.
  - Size of the PFA $\geq \sum d_i$.
  - For large m : $\quad \pi A^m \eta \approx \pi A^{m+D} \eta$.
  - $\mathrm{period}(L)$ divides D.

Do we need ALL the prime powers dividing D?

How to leap over the gap?



- An ergodic component can provide at most its absorption probability.

- Ergodic components can add up their absorption probabilities, if they accept and reject in a "synchronized manner": periods have common divisors.

Leaping Strength of a Prime Power

- Ergodic component B_i:

  d_i = period of B_i.
  r_i = absorption probability of B_i.

- For a prime power $q = p^\alpha$

$$\text{leap}(q) = \sum_{i:q \text{ divides } d_i} r_i$$

is the leaping strength of q.

Proof Sketch:
Only Prime Powers with High Leaping Strength Count

- Call a prime power q weak if $\operatorname{leap}(q) < 2\epsilon$.

- Crucial Step 1:
  - A weak prime power cannot divide $\operatorname{period}(L)$.
  - Hence $\operatorname{period}(L)$ divides $D := \operatorname{lcm}\{q \mid \operatorname{leap}(q) \geq 2\epsilon\}$.

Proof Sketch:
$D = \mathrm{lcm}\{q \mid \mathrm{leap}(q) \geq 2\epsilon\}$ and the Number of States

- Crucial Step 2:
  - If q is not weak: $q^{2\epsilon} \leq q^{\mathrm{leap}(q)} = q^{\sum_{i:q \text{ divides } d_i} r_i}$.
  - Consequence: $D^{2\epsilon} \leq \prod d_i^{r_i}$.
  - Conclusion:

$$n \geq \sum d_i \geq \sum r_i d_i \geq \prod d_i^{r_i} \geq D^{2\epsilon} \geq \mathrm{period}(L)^{2\epsilon}.$$

$$\Rightarrow \ \mathrm{period}(L) \leq n^{\frac{1}{2\epsilon}}$$

Computing Minimal NFA's, Previous Results

For a regular language L let $\mathrm{nsize}(L)$ be the size of a minimal NFA accepting L.

- L is specified by a DFA (or NFA):

  Determining $\mathrm{nsize}(L)$ is $PSPACE$-hard.

- L is specified by a **unary** NFA:

  Determining $\mathrm{nsize}(L)$ is NP-hard.

- L is specified by a **unary cyclic** DFA:

  Determining $\mathrm{nsize}(L)$ efficiently implies $NP \subseteq DTIME(n^{O(\log n)})$.

How hard is approximation?

Approximating Minimal NFA's

Given a unary cyclic DFA accepting L with n states, an NFA for L with at most $\mathrm{nsize}(L) \cdot (1 + \ln n)$ can be efficiently constructed.
This result implies an approximation with factor $O(\ln n)$ or $O\left(\sqrt{\mathrm{nsize}(L) \cdot \ln \mathrm{nsize}(L)}\right)$.

Shown by reduction to a set cover problem, easy to approximate within ratio $O(\ln n)$.

Given a unary NFA N with s states, it is impossible to efficiently approximate $\mathrm{nsize}(L(N))$ within a factor of $\frac{\sqrt{s}}{\ln s}$ unless $P = NP$.

Moreover, every approximation algorithm with approximation factor bounded by a function with $\mathrm{nsize}(L)$ as its only argument solves an NP-hard problem.

NP Hardness of the Universe Problem for NFA's

Result of Stockmeyer and Meyer (1973):

> For a given unary NFA N, it is NP-hard to decide, if $L(N) \neq a^*$.

- Given an instance Φ of the 3SAT problem, construct a unary NFA N_Φ that accepts a^*, iff Φ is not satisfiable.
- We put this into an approximation framework:
- **Reduction is gap producing!**
  - $\Phi \notin$ 3SAT $\Rightarrow L := L(N_\Phi) = a^*$ and thus $\mathrm{nsize}(L) = 1$.
  - $\Phi \in$ 3SAT $\Rightarrow \mathrm{nsize}(L) = \Omega(n^2 \ln n)$ for Φ defined over n variables.

Conclusions

- Approximating minimal NFA's:
  - NP-hard to approximate within $\frac{\sqrt{s}}{\ln s}$ if the language is represented by an s-state unary NFA.
  - Factor $(1 + \ln n)$ is possible if the language is represented by an n-state unary cyclic DFA.

- Determinism versus probabilism with fixed isolation:
  - Short-term behaviour (length of the initial path) with exponential blow up, but
  - long-term behaviour (length of the cycle) with only polynomial blow up.