The structure of polynomial growth for tree automata/transducers and MSO set queries

Lê Thành Dũng (Tito) Nguyễn (Aix-Marseille Univ.)
joint work with Paul Gallot (Univ. Bremen) & Nathan Lhote (AMU)

Highlights of Logic, Games and Automata 2025 (Saarbrücken)

Overview

Basic results on finite automata (known + obvious extension)

$\Downarrow$

"Reparameterisation" of MSO set queries on trees

$\Downarrow$

A bunch of consequences for *transducers* outputting strings or trees

Overview

Basic results on finite automata (known + obvious extension)

$\Downarrow$

"Reparameterisation" of MSO set queries on trees

$\Downarrow$

A bunch of consequences for *transducers* outputting strings or trees

Asymptotic study of the growth rate of $f : (\text{strings or trees}) \to \mathbb{N}$

$$\text{growth}[f] : n \in \mathbb{N} \mapsto \max\{f(t) \mid |t| \leqslant n\}$$

Dichotomy phenomena:
- polynomial $\Theta(n^k)$, degree $k \in \mathbb{N}$
- exponential $2^{\Theta(n)}$, degree $+\infty$ by convention

Overview

Basic results on finite automata (known + obvious extension)

$\Downarrow$

"Reparameterisation" of MSO set queries on trees

$\Downarrow$

A bunch of consequences for *transducers* outputting strings or trees

Asymptotic study of the growth rate of $f \colon$ (strings or trees) $\to \mathbb{N}$

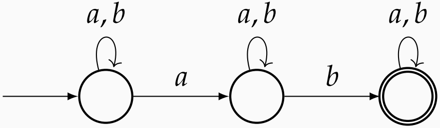
$$\mathrm{growth}[f] \colon n \in \mathbb{N} \mapsto \max\{f(t) \mid |t| \leqslant n\}$$

Dichotomy phenomena:

- polynomial $\Theta(n^k)$, degree $k \in \mathbb{N}$
- exponential $2^{\Theta(n)}$, degree $+\infty$ by convention

Ambiguity of nondeterministic finite automata

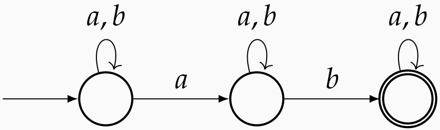
Ambiguity = number of runs of NFA on given input word; for example:



counts matches for pattern ... a ... b ...

Ambiguity of nondeterministic finite automata

Ambiguity = number of runs of NFA on given input word; for example:



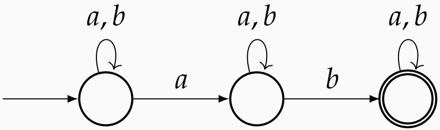
counts matches for pattern ... a ... b ...

Theorem (Weber & Seidl 1986)

deg(growth[*ambiguity of* $\mathscr{A}$]) *is well-defined in* $\mathbb{N} \cup \{\infty\}$ (*poly/exp dichotomy*)

Ambiguity of nondeterministic finite automata

Ambiguity = number of runs of NFA on given input word; for example:



counts matches for pattern ... a ... b ...

Theorem (Weber & Seidl 1986)

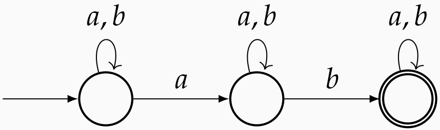
deg(growth[*ambiguity of* $\mathscr{A}$]) *is well-defined in* $\mathbb{N} \cup \{\infty\}$ (*poly/exp dichotomy*) *and*

- *it is computable in time* $O(|\mathscr{A}|^3)$
- $\deg < \infty$ *is decidable in time* $O(|\mathscr{A}|^2)$

$|\mathscr{A}| = |\mathit{states}| + |\mathit{transitions}|$

Ambiguity of nondeterministic finite automata

Ambiguity = number of runs of NFA on given input word; for example:



counts matches for pattern ... a ... b ...

Theorem (Weber & Seidl 1986)

deg(growth[*ambiguity of* $\mathscr{A}$]) *is well-defined in* $\mathbb{N} \cup \{\infty\}$ (*poly/exp dichotomy*) *and*

- *it is computable in time* $O(|\mathscr{A}|^3)$
- $\deg < \infty$ *is decidable in time* $O(|\mathscr{A}|^2)$

$|\mathscr{A}| = |states| + |transitions|$

These bounds are optimal: cf. Karolina Drabik's talk this afternoon

Ambiguity of nondeterministic finite tree automata

Ambiguity = number of runs of NFTA on given input tree

Theorem (Paul 2015)

$\deg(\text{growth}[$*ambiguity of any tree automaton*$])$ *is well-defined in* $\mathbb{N} \cup \{\infty\}$

Theorem (new)

- *it is computable in time* $O(|\mathscr{A}|^3)$ (*same complexity as for words*)
- $\deg < \infty$ *is decidable in time* $O(|\mathscr{A}|^2)$

Ambiguity of nondeterministic finite tree automata

Ambiguity = number of runs of NFTA on given input tree

Theorem (Paul 2015)

deg(growth[*ambiguity of any tree automaton*]) *is well-defined in* $\mathbb{N} \cup \{\infty\}$

Theorem (new)

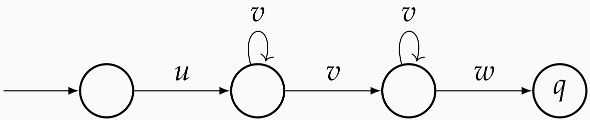
- *it is computable in time* $O(|\mathscr{A}|^3)$ (*same complexity as for words*)
- deg $< \infty$ *is decidable in time* $O(|\mathscr{A}|^2)$

Proof similar to case of strings (next slide):

- branching plays limited role
- words $\rightsquigarrow$ one-hole contexts — e.g. $a(b(\Box), c)$ — in *pumping patterns*

Ambiguity of nondeterministic finite automata: polynomial case

- Define *degrees of states* via pumping patterns, for example:

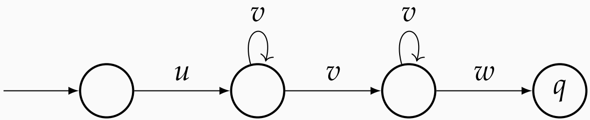


$\implies \deg(q) \geqslant 1$

- Easy: growth[ambiguity]$(n) = \Omega(n^k)$ for $k = \max(\deg(\text{co-reachable states}))$

Ambiguity of nondeterministic finite automata: polynomial case

- Define *degrees of states* via pumping patterns, for example:



$\implies \deg(q) \geqslant 1$

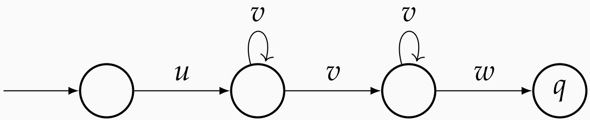
- Easy: growth[ambiguity]$(n) = \Omega(n^k)$ for $k = \max(\deg(\text{co-reachable states}))$
- Matching upper bound $O(n^k)$: at most k critical positions in a run

Definition

Critical position in a run = where the state degree increases

Ambiguity of nondeterministic finite automata: polynomial case

- Define *degrees of states* via pumping patterns, for example:



$\implies \deg(q) \geqslant 1$

- Easy: growth[ambiguity]$(n) = \Omega(n^k)$ for $k = \max(\deg(\text{co-reachable states}))$
- Matching upper bound $O(n^k)$: at most k critical positions in a run

Definition

Critical position in a run = where the state degree increases

Lemma (consequence of existing results on finite ambiguity)

There are $O(1)$ runs over a given word/tree with a given set of critical positions.

From automata to Monadic Second-Order logic

nondeterminism $\leftrightarrow$ choice of params X_i in MSO formula $\varphi(X_1, \ldots, X_m)$

ambiguity $\leftrightarrow$ number of satisfying choices = results of query φ

From automata to Monadic Second-Order logic

nondeterminism $\leftrightarrow$ choice of params X_i in MSO formula $\varphi(X_1, \ldots, X_m)$

ambiguity $\leftrightarrow$ number of satisfying choices = results of query φ

Corollary

- $k = \deg(\text{growth}[\textit{nb of results of query } \varphi]) \in \mathbb{N} \cup \{\infty\}$ *well-def & computable*

From automata to Monadic Second-Order logic

nondeterminism ↔ choice of params X_i in MSO formula $\varphi(X_1, \dots, X_m)$

ambiguity ↔ number of satisfying choices = results of query φ

Corollary

- $k = \deg(\text{growth}[\textit{nb of results of query } \varphi]) \in \mathbb{N} \cup \{\infty\}$ *well-def & computable*

f_t(query result) = list of critical positions of corresponding run over t

finite-to-one due to previous lemma: $|f_t^{-1}(\{\text{some list}\})| = O(1)$

From automata to Monadic Second-Order logic

nondeterminism $\leftrightarrow$ choice of params X_i in MSO formula $\varphi(X_1, \dots, X_m)$

ambiguity $\leftrightarrow$ number of satisfying choices = results of query φ

Corollary

- $k = \deg(\text{growth}[\textit{nb of results of query } \varphi]) \in \mathbb{N} \cup \{\infty\}$ *well-def & computable*
- *If* $k < \infty$ *then* $\exists$ "reparameterisation" $f_t \colon \{\textit{results of } \varphi \textit{ on } t\} \to \{\textit{nodes in } t\}^k$ *which is finite-to-one and* MSO-definable by some $\psi(\underbrace{X_1, \dots, X_m}_{\textit{2nd-order}}, \underbrace{z_1, \dots, z_k}_{\textit{1st-order}})$

Proof: f_t(query result) = list of critical positions of corresponding run over t

finite-to-one due to previous lemma: $|f_t^{-1}(\{\text{some list}\})| = O(1)$

Conclusion

Computing polynomial degree of growth of NFA ambiguity (over trees)

$\Downarrow$

k-variable "reparameterisation" of MSO set queries of growth $O(n^k)$

[Bojańczyk 2023]: case of $\varphi(x_1, \ldots, x_m)$ on strings, using factorisation forests

Conclusion

Computing polynomial degree of growth of NFA ambiguity (over trees)

$\Downarrow$

k-variable "reparameterisation" of MSO set queries of growth $O(n^k)$

[Bojańczyk 2023]: case of $\varphi(x_1, \ldots, x_m)$ on strings, using factorisation forests

$\Downarrow$

Optimisation of *MSO <u>set</u> interpretations* from trees (cf. next talk by Colcombet)

[Bojańczyk 2023]: case of MSO interpretations from strings a.k.a. polyregular functions

Conclusion

Computing polynomial degree of growth of NFA ambiguity (over trees)

$\Downarrow$

k-variable "reparameterisation" of MSO set queries of growth $O(n^k)$

[Bojańczyk 2023]: case of $\varphi(x_1, \dots, x_m)$ on strings, using factorisation forests

$\Downarrow$

Optimisation of *MSO <u>set</u> interpretations* from trees (cf. next talk by Colcombet)

[Bojańczyk 2023]: case of MSO interpretations from strings a.k.a. polyregular functions

$\Downarrow$

inter alia, results on asymptotics of *macro tree transducers* (cf. next^2 talk by Peyrat)

generalising [Engelfriet & Maneth 2000; Gallot, Maneth, Nakano & Peyrat 2024] w/ easier proofs

Conclusion

TheoretiCS submission successfully passed Phase 1

journal-first, never submitted to conference proceedings!

Computing polynomial degree of growth of NFA ambiguity (over trees)

$\Downarrow$

k-variable "reparameterisation" of MSO set queries of growth $O(n^k)$

[Bojańczyk 2023]: case of $\varphi(x_1, \ldots, x_m)$ on strings, using factorisation forests

$\Downarrow$

Optimisation of *MSO <u>set</u> interpretations* from trees (cf. next talk by Colcombet)

[Bojańczyk 2023]: case of MSO interpretations from strings a.k.a. polyregular functions

$\Downarrow$

inter alia, results on asymptotics of *macro tree transducers* (cf. next² talk by Peyrat)

generalising [Engelfriet & Maneth 2000; Gallot, Maneth, Nakano & Peyrat 2024] w/ easier proofs

Conclusion

Thanks for your attention!

TheoretiCS submission successfully passed Phase 1

journal-first, never submitted to conference proceedings!

Computing polynomial degree of growth of NFA ambiguity (over trees)

$\Downarrow$

k-variable "reparameterisation" of MSO set queries of growth $O(n^k)$

[Bojańczyk 2023]: case of $\varphi(x_1, \dots, x_m)$ on strings, using factorisation forests

$\Downarrow$

Optimisation of *MSO set interpretations* from trees (cf. next talk by Colcombet)

[Bojańczyk 2023]: case of MSO interpretations from strings a.k.a. polyregular functions

$\Downarrow$

inter alia, results on asymptotics of *macro tree transducers* (cf. next² talk by Peyrat)

generalising [Engelfriet & Maneth 2000; Gallot, Maneth, Nakano & Peyrat 2024] w/ easier proofs