Nedeterminizmus v generatívnych systémoch

Študent: Bc. Ján Rosina
Školiteľ: prof. RNDr. Branislav Rovan, PhD.

9. júna 2021

Čo sú g-systémy?

- abstrakcia gramatík
- g-systém: $G = (N, T, M, \sigma)$
- M - číta vstup znak po znaku a zapisuje na výstup
- $u \Rightarrow_G v \iff v \in M(u)$
- $L(G) = \{w \in T^* | \sigma \Rightarrow_G^* w\}$
- deterministický M - deterministický g-systém (dgs)
- trieda jazykov generovaných dgs - $\mathcal{L}_{DG}$

Východisková situácia I

- dve predošlé diplomové práce:
  - M. Králik: Deterministic generative systems, 2002
  - D. Krcho: Non-determinism in Generative Systems, 2002
- sila deterministických g-systémov:
  - dokážu generovať všetky konečné jazyky
  - dokážu generovať aj niektoré jazyky z $\mathcal{L}_{CS} - \mathcal{L}_{CF}$ (napr. $\{a^n b^n c^n \mid n > 0\}$)
  - ak $\epsilon \in L$ a L je nekonečný, potom $L \notin \mathcal{L}_{DG}$
  - $\{a^i b^j \mid i > 0, j > 0\} \notin \mathcal{L}_{DG}$
  - nám sa podarilo ukázať, že ľubovoľný nekonečný jazyk obsahujúci slová u, v, v' t.ž. u je prefixom v aj v' ale tie si nie sú prefixami nepatrí do $\mathcal{L}_{DG}$

Východisková situácia II

- kľúčová slabina dgs - neschopnosť rozpoznať koniec vetnej formy
  - deterministické g-systémy s endmarkerom - \$dgs
  - \$ na konci vetnej formy, terminálna vetná forma: w\$
  - dôkaz $\mathcal{L}_{REC} \subseteq \mathcal{L}_{\$DG}$
  - platí $\mathcal{L}_{RE} = \mathcal{L}_{\$DG}$?
- koľko nedeterminizmu stačí pridať?
  - 1 nedeterministický stav
  - $|\Sigma| + 1$ nedeterministických prechodov
  - potreba výpočtovej miery nedeterminizmu

Ciele práce

- skúmať vzťah medzi triedami jazykov $\mathcal{L}_{RE}$ a $\mathcal{L}_{\$DG}$
    - platí rovnosť
- definovať a skúmať výpočtové miery nedeterminizmu v g-systémoch
    - podľa dĺžky odvodenia - vieme dosiahnuť ľubovoľne pomaly rastúcu funkciu
    - podľa dĺžky odvodeného slova - vo všeobecnosti $n|\Sigma|$, pre unárne jazyky a Σ^* 2log(n)

- vieme simulovať výpočet TS A na konkrétnom slove
- ako simulovať A na všetkých slovách?

$$\sigma\$ \Rightarrow_G \ldots \Rightarrow_G w_0\$ \Rightarrow_G \ldots \Rightarrow_G w_1\$\ldots$$

$\mathcal{L}_{\$DG} = \mathcal{L}_{RE}$ - v čom je problém

- vieme simulovať výpočet TS A na konkrétnom slove
- ako simulovať A na všetkých slovách?

$$\sigma\$ \Rightarrow_G \ldots \Rightarrow_G w_0\$ \Rightarrow_G \ldots \Rightarrow_G w_1\$\ldots$$

- 2 hlavné problémy:
    - ako pokračovať v odvodení z vygenerovanej vetnej formy $w\$$?
    - ako zamedziť simulácii nekonečného výpočtu?

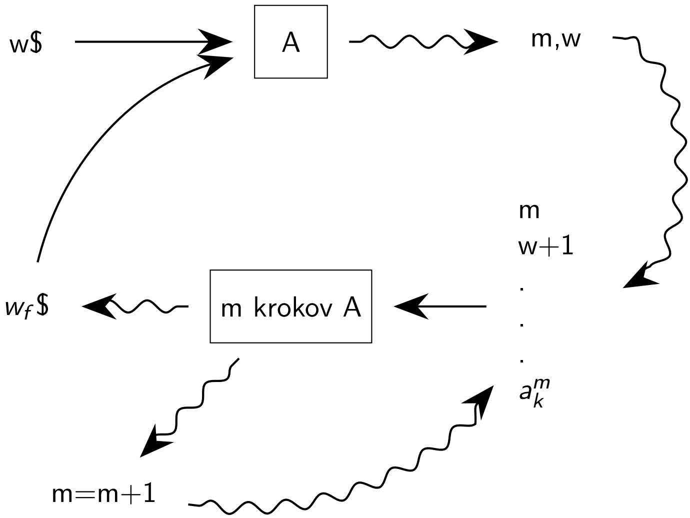
$\mathcal{L}_{\$DG} = \mathcal{L}_{RE}$ - riešenie

- simulovať *m* krokov A
- poradie odvodených slov - podľa počtu krokov, na ktorý sú akceptované A
- pre každé *m* treba simulovať A na všetkých slovách

$\mathcal{L}_{\$DG} = \mathcal{L}_{RE}$ - riešenie

- simulovať m krokov A
- poradie odvodených slov - podľa počtu krokov, na ktorý sú akceptované A
- pre každé m treba simulovať A na všetkých slovách
  - predpoklad, že A vždy prečíta celý vstup $\rightsquigarrow$ stačí prehľadávať slová $\epsilon, ..., a_k^m$ (pre abecedu $\Sigma = \{a_1, ..., a_k\}$)

$\mathcal{L}_{\$DG} = \mathcal{L}_{RE}$ - práca G na vetnej forme $w\$$



Miery nedeterminizmu: podľa počtu krokov odvodenia

- $stepDec(G, m) = max\#$ nedeterministických rozhodnutí za m krokov odvodenia v G
- všeobecný prípad: ľubovoľný $L \in \mathcal{L}_{RE}$
- priamočiara simulácia \$dgs $\rightarrow$ $stepDec(G, m) \leq m$
- vylepšenie: pridať veľa deterministických krokov

Miery nedeterminizmu: podľa počtu krokov odvodenia

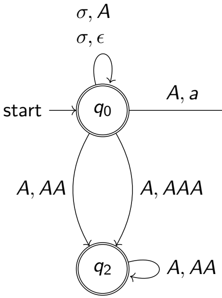
- $stepDec(G, m) = max\#$ nedeterministických rozhodnutí za m krokov odvodenia v G
- všeobecný prípad: ľubovoľný $L \in \mathcal{L}_{RE}$
- priamočiara simulácia \$dgs $\rightarrow$ $stepDec(G, m) \leq m$
- vylepšenie: pridať veľa deterministických krokov
- counter na začiatku vetnej formy
  - v každom kroku sa inkrementuje
  - ak nepretečie - zvyšok sa iba skopíruje
  - ak pretečie - nedeterministicky sa rozhodne či skončí alebo pokračuje
  - ak pokračuje, simuluje jeden krok \$dgs a zväčší counter
  - pre exponencálne zväčšovanie countera dostávame $stepDec(G, m) \leq log^*(m)$

Miery nedeterminizmu: podľa dĺžky slova I

- $lengthDec(G, w) =$ minimálny počet nedeterministických rozhodnutí pre odvodenie w v G
- $lengthDec(G, n) = max(lengthDec(G, w) \mid |w| = n)$
- všeobecný prípad - G vygeneruje ľubovoľné slovo zo Σ^* a simuluje na ňom TS - $lengthDec(G, n) \leq |\Sigma| n$
- zdá sa, že vo všeobecnosti to lepšie nepôjde (treba vygenerovať jedno z $|\Sigma|^n$ slov)

Miery nedeterminizmu: podľa dĺžky slova II

- ľubovoľný počet znakov n vieme dostať pomocou $2log(n)$ rozhodnutí
- pre jazyky Σ^* a $L \subseteq \{a\}^*$ vieme zostrojiť G, pre ktorý $lengthDec(G,n) \leq 2log(n)$



Miery nedeterminizmu: podľa dĺžky slova III

- pre $L \in \mathcal{L}_{REC}$ vieme dosiahnuť $lengthDec(G, n) \leq 2log(n) + 2log(|W_n|)$, kde $W_n = \{w | w \in L, |w| = n\}$

log(n)+log(m) rozhodnutí

$\sigma \rightsquigarrow$ n,m $\rightsquigarrow$ m, a_1^n, ..., a_k^n $\rightsquigarrow$ w_m

simulácia A

Ďakujem za pozornosť

Otázka 4: paralelná vs. sekvenčná simulácia

- dôkaz vety $\mathcal{L}_{RE} = \mathcal{L}_{\$DG}$
- fungovala by aj sekvenčná simulácia

Otázka 5: (ne)generovanie endmarkera

- simulácia \$dgs G' pomocou nedeterministického g-systému G
- takto sa endmarker nevyskytne vo vetnej forme
- simulácia zlyhá, keď G' použije
  $h \in H' : pr_2(h) = \$, pr_3(h) = v\$$, pre $v \neq \epsilon$
- oprava: nahradiť v H prechod (p, σ, u, q) prechodmi
  $(p, \sigma, uv\$, r), (p, \sigma, uv, r)$, kde $(p, \sigma, u, q), (q, \$, v\$, r) \in H'$

Otázka 6: (2)log(n) nedeterministických krokov

- počet nedeterministických krokov pri generovaní slova dĺžky n, pre ľubovoľné n

Otázka 7: rozšírenie vety 5 pre rekurzívne vyčísliteľné jazyky

- pre $L \in \mathcal{L}_{REC}$ vieme dosiahnuť $lengthDec(G, n) \leq 2log(n) + 2log(|W_n|)$, kde $W_n = \{w | w \in L, |w| = n\}$
- použitá konštrukcia simuluje TS A na všetkých slovách dĺžky n a následne m-té akceptované urobí terminálnym
- simulácia prebieha až kým nedospeje buď k akceptovaniu alebo zamietnutiu
- ak na nejakom slove výpočet A neskončí, žiadne z nasledovných slov nemôže byť zterminálnené

Otázka 8: dolné odhady

- iba v rovine intuície
- rádovo $log(n)$ pre jazyk Σ^*
- rádovo $n|\Sigma|$ vo všeobecnom prípade