

Теория автоматов и формальных языков

За пределами контекстно-свободных языков

Автор: Григорьев Семён

Санкт-Петербургский государственный университет

10 декабря 2020

Слегка контекстно-зависимые языки (Mildly context sensitive)

Выйти за пределы КС языков, но сохранить “хорошие свойства”

- Полиномиальное время синтаксического анализа (для фиксированной грамматики)
- Невыразимость “слишком сложных структур”
- Полулинейность
- ...

Multiple Context-Free Grammars

Больше информации в презентациях Sylvain Salvati

Определение

m -MCFG(r) это четвёрка $\langle \Sigma, N, S, P \rangle$

- Σ — терминальный алфавит
- N — нетерминальные символы. Максимальный ранг (арность, местность) равен m
- S — стартовый нетерминальный символ ранга 1
- P — множество правил вида

$$A(s_1, \dots, s_k) \leftarrow B_1(x_1^1, \dots, x_{k_1}^1), \dots, B_n(x_1^n, \dots, x_{k_n}^n)$$

- ▶ A — нетерминал ранга k , B_i — нетерминалы ранга k_i , $n \leq r$
- ▶ Все x_j^i попарно различны (переменные)
- ▶ $s_i \in (\Sigma \cup X)^*$, $X = \bigcup_{i=1}^n \bigcup_{j=1}^{k_i} x_j^i$

Пример (КС грамматики)

$$S \rightarrow aSb$$

$$S \rightarrow \varepsilon$$

$$S(axb) \leftarrow S(x)$$

$$S(\varepsilon) \leftarrow$$

Пример (КС грамматики)

$$S \rightarrow aSb$$

$$S \rightarrow \varepsilon$$

$$S \rightarrow aSbS$$

$$S \rightarrow \varepsilon$$

$$S(axb) \leftarrow S(x)$$

$$S(\varepsilon) \leftarrow$$

$$S(ax_1bx_2) \leftarrow S(x_1), S(x_2)$$

$$S(\varepsilon) \leftarrow$$

$$S(x_1 y_1 x_2 y_2) \leftarrow P(x_1, x_2), Q(y_1, y_2)$$

$$P(ax_1, bx_2) \leftarrow P(x_1, x_2)$$

$$P(\varepsilon, \varepsilon) \leftarrow$$

$$Q(cx_1, dx_2) \leftarrow Q(x_1, x_2)$$

$$Q(\varepsilon, \varepsilon) \leftarrow$$

$$S(x_1 y_1 x_2 y_2) \leftarrow P(x_1, x_2), Q(y_1, y_2)$$

$$P(ax_1, bx_2) \leftarrow P(x_1, x_2)$$

$$P(\varepsilon, \varepsilon) \leftarrow$$

$$Q(cx_1, dx_2) \leftarrow Q(x_1, x_2)$$

$$Q(\varepsilon, \varepsilon) \leftarrow$$

$$L = \{a^n c^m b^n d^m \mid n \in \mathbb{N}, m \in \mathbb{N}\}$$

① PMCFG (parallel MCFG)

$$A(x, ax) \leftarrow B(x)$$

②

$$A(x) \leftarrow B(x), C(x)$$

③ simpleLMG

$$A(x, x) \leftarrow B(x), C(x)$$

1 PMCFG (parallel MCFG)

$$A(x, ax) \leftarrow B(x)$$

2

$$A(x) \leftarrow B(x), C(x)$$

3 simpleLMG

$$A(x, x) \leftarrow B(x), C(x)$$

$$MCFL \subsetneq PMCFL \subsetneq \text{simpleLMG} = P$$

$$\{a^{2^n} \mid n \geq 0\} \in PMCFL - MCFL$$

$$S(xx) \leftarrow S(x)$$

$$S(a) \leftarrow$$

- Неудаляющая —

$\forall i \in \{1, \dots, n\}, j \in \{1, \dots, k_i\} x_j^i$ используется в s_1, \dots, s_k

- **Неудаляющая** —
 $\forall i \in \{1, \dots, n\}, j \in \{1, \dots, k_i\}$ x_j^i используется в s_1, \dots, s_k
- **Непереставляющая** — $\forall i \in \{1, \dots, n\}, j, k \in \{1, \dots, k_i\}$, если $j < k$, то x_j^i встречается в s_1, \dots, s_k перед x_k^i

Разновидности MCFG

- **Неудаляющая** —
 $\forall i \in \{1, \dots, n\}, j \in \{1, \dots, k_i\}$ x_j^i используется в s_1, \dots, s_k
- **Непереставляющая** — $\forall i \in \{1, \dots, n\}, j, k \in \{1, \dots, k_i\}$, если $j < k$, то x_j^i встречается в s_1, \dots, s_k перед x_k^i
- **Well-nested** — неудаляющая, непереставляющая и

$$\forall i, i' \in \{1, \dots, n\}, i \neq i',$$

$$j \in \{1, \dots, k_i - 1\}, j' \in \{1, \dots, k_{i'} - 1\},$$

$$s_1 \cdots s_k \notin (\Sigma \cup X)^* x_j^i (\Sigma \cup X)^* x_{j'}^{i'} (\Sigma \cup X)^* x_{j+1}^i (\Sigma \cup X)^* x_{j'+1}^{i'} (\Sigma \cup X)^*$$

Пример well-nested MCFG

✓ $A(\textcolor{pink}{x_1}, \textcolor{yellow}{z_1, z_2}, \textcolor{pink}{x_2}, \textcolor{green}{y_1, y_2, y_3}, \textcolor{pink}{x_3}) \leftarrow$
 $B(x_1, x_2, x_3), C(y_1, y_2, y_3), D(z_1, z_2)$

✗ $A(\textcolor{yellow}{z_1}, \textcolor{pink}{x_1}, \textcolor{green}{y_1}, \textcolor{pink}{x_2}, \textcolor{yellow}{z_2}, \textcolor{green}{y_2}, \textcolor{pink}{x_3}, \textcolor{green}{y_3}) \leftarrow$
 $B(x_1, x_2, x_3), C(y_1, y_2, y_3), D(z_1, z_2)$

Лемма о накачке

Теорема (general MCFG)

$$\forall L \in m\text{-MCFG} \exists n \geq 1 \exists \underline{z} \in L (|\underline{z}| \geq n)$$

$$\exists \text{ разбиение } z = u_1 v_1 w_1 s_1 u_2 \dots u_m v_m w_m s_m u_{m+1}, \sum |v_j s_j| \geq 1$$

$$\forall i \geq 0 : z_i = u_1 v_1^i w_1 s_1^i u_2 \dots u_m v_m^i w_m s_m^i u_{m+1} \in L$$

Лемма о накачке

Теорема (general MCFG)

$$\forall L \in m\text{-MCFG} \exists n \geq 1 \exists \underline{z} \in L (|z| \geq n)$$

$$\exists \text{ разбиение } z = u_1 v_1 w_1 s_1 u_2 \dots u_m v_m w_m s_m u_{m+1}, \sum |v_j s_j| \geq 1$$

$$\forall i \geq 0 : z_i = u_1 v_1^i w_1 s_1^i u_2 \dots u_m v_m^i w_m s_m^i u_{m+1} \in L$$

Теорема (well-nested MCFG)

$$\forall L \in m\text{-wnMCFG} \exists n \geq 1 \forall \underline{z} \in L (|z| \geq n)$$

$$\exists \text{ разбиение } z = u_1 v_1 w_1 s_1 u_2 \dots u_m v_m w_m s_m u_{m+1}, \sum |v_j s_j| \geq 1$$

$$\forall i \geq 0 : z_i = u_1 v_1^i w_1 s_1^i u_2 \dots u_m v_m^i w_m s_m^i u_{m+1} \in L$$

Теорема

$(m * (k - 1))\text{-MCFL}(r - k) \subseteq m\text{-MCFL}(r)$ если $1 \leq k \leq r - 2$

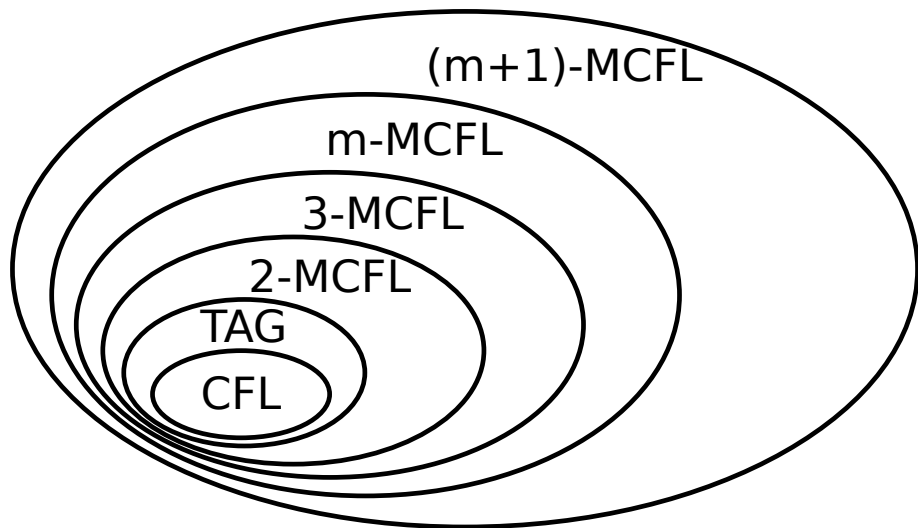
Иерархии внутри MCFL

Теорема

$(m * (k - 1))\text{-MCFL}(r - k) \subseteq m\text{-MCFL}(r)$ если $1 \leq k \leq r - 2$

Теорема (Seki et al)

$L_{m+1} = \{a_1^n b_1^n \cdots a_{m+1}^n b_{m+1}^n \mid n \in \mathbb{N}\}$ является $(m + 1)\text{-MCFL}(1)$, но не является $m\text{-MCFL}(r)$ ни для какого r



Иерархия для $m = 1$

Теорема

$$1\text{-MCFL} = \text{CFL}$$

Иерархия для $m = 1$

Теорема

$$1\text{-MCFL} = \text{CFL}$$

Теорема

$$1\text{-MCFL}(1) \subsetneq 1\text{-MCFL}(2)$$

Иерархия для $m = 1$

Теорема

$$1\text{-MCFL} = \text{CFL}$$

Теорема

$$1\text{-MCFL}(1) \subsetneq 1\text{-MCFL}(2)$$

Теорема

$$1\text{-MCFL}(r) = 1\text{-MCFL}(r + 1), r \geq 2$$

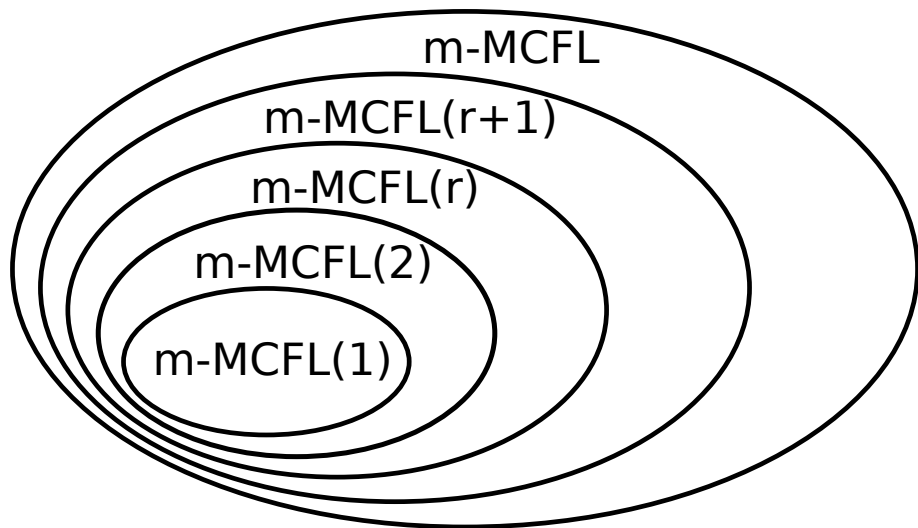
Иерархия для $m = 2$

Теорема (Ramow, Satta)

$$2\text{-MCFL}(2) = 2\text{-MCFL}(3)$$

Теорема

Если $m > 2$ или $r > 2$, то $m\text{-MCFL}(r) \subsetneq m\text{-MCFL}(r+1)$



- $mix = \{\omega \in \{a, b\}^* \mid |\omega|_a = |\omega|_b\}$ — контекстно-свободный язык
- $MIX = \{\omega \in \{a, b, c\}^* \mid |\omega|_a = |\omega|_b = |\omega|_c\}$ — MCFL? Хотелось верить, что нет
 - ▶ MIX is a 2-MCFL and the word problem in \mathbb{Z}^2 is solved by a third-order collapsible pushdown automaton, Sylvain Salvati, 2011

- $mix = \{\omega \in \{a, b\}^* \mid |\omega|_a = |\omega|_b\}$ — контекстно-свободный язык
- $MIX = \{\omega \in \{a, b, c\}^* \mid |\omega|_a = |\omega|_b = |\omega|_c\}$ — MCFL? Хотелось верить, что нет
 - ▶ MIX is a 2-MCFL and the word problem in \mathbb{Z}^2 is solved by a third-order collapsible pushdown automaton, Sylvain Salvati, 2011
- $O_2 = \{\omega \in \{a, \bar{a}, b, \bar{b}\}^* \mid |\omega|_a = |\omega|_{\bar{a}} \wedge |w|_b = |w|_{\bar{b}}\}$
- $O_n = \{\omega \in \{a_1, \bar{a}_1, a_2, \bar{a}_2, \dots, a_n, \bar{a}_n\}^* \mid |\omega|_{a_1} = |\omega|_{\bar{a}_1} \wedge |w|_{a_2} = |w|_{\bar{a}_2} \wedge \dots \wedge |w|_{a_n} = |w|_{\bar{a}_n}\}$
- $MIX_n = \{\omega \in \{a_1, \dots, a_n\}^* \mid |\omega|_{a_1} = |\omega|_{a_2} = \dots = |\omega|_{a_n}\}$

- $mix = \{\omega \in \{a, b\}^* \mid |\omega|_a = |\omega|_b\}$ — контекстно-свободный язык
- $MIX = \{\omega \in \{a, b, c\}^* \mid |\omega|_a = |\omega|_b = |\omega|_c\}$ — MCFL? Хотелось верить, что нет
 - ▶ MIX is a 2-MCFL and the word problem in \mathbb{Z}^2 is solved by a third-order collapsible pushdown automaton, Sylvain Salvati, 2011
- $O_2 = \{\omega \in \{a, \bar{a}, b, \bar{b}\}^* \mid |\omega|_a = |\omega|_{\bar{a}} \wedge |\omega|_b = |\omega|_{\bar{b}}\}$
- $O_n = \{\omega \in \{a_1, \bar{a}_1, a_2, \bar{a}_2, \dots, a_n, \bar{a}_n\}^* \mid |\omega|_{a_1} = |\omega|_{\bar{a}_1} \wedge |\omega|_{a_2} = |\omega|_{\bar{a}_2} \wedge \dots \wedge |\omega|_{a_n} = |\omega|_{\bar{a}_n}\}$
- $MIX_n = \{\omega \in \{a_1, \dots, a_n\}^* \mid |\omega|_{a_1} = |\omega|_{a_2} = \dots = |\omega|_{a_n}\}$
- MIX_n регулярно эквивалентен O_n (существует алгоритм построения грамматики одного языка по грамматике другого)
 - ▶ O_n is an n-MCFL, Sylvain Salvati, 2018

- Уточнение внутренних иерархий
- Сопоставление с другими классами и иерархиями
- Представимость языков
 - ▶ Варианты леммы о накачке
 - ▶ Представимость конкретных языков
 - ★ Многомерный язык Дика: Towards a 2-Multiple Context-Free Grammar for the 3-Dimensional Dyck Language, Konstantinos Kogkalidis, Orestis Melkonian, 2019
 - ★ Шафл языков Дика: Context-sensitive data-dependence analysis via linear conjunctive language reachability, Qirun Zhang, Zhendong Su et al, 2017

- Учёт синтаксической структуры при синтезе: Grammar variational autoencoder, Kusner M. J., Paige B., Hernandez-Lobato J. M., 2017
- Восстановление синтаксической структуры: End-to-end Graph-based TAG Parsing with Neural Networks, Jungo Kasai, Robert Frank et al, 2018
- Извлечение грамматик: Distributional Learning of Context-Free and Multiple Context-Free Grammars, Alexander Clark, Ryo Yoshinaka, 2019