Maximal and Minimal Scattered Context Rewriting

Jiří Techet Tomáš Masopust Alexander Meduna

Department of Information Systems
Faculty of Information Technology
Brno University of Technology
Božetěchova 2, Brno 61266, Czech Republic

Modern Formal Language Theory, 2007

Scattered Context Grammar

Scattered context grammar (SC grammar)

G = (V, T, P, S), where

V is a finite alphabet

T is a set of terminals, $T \subset V$

S is the start symbol, $S \in V - T$

P is a finite set of productions of the form

$$(A_1, \dots, A_n) \to (x_1, \dots, x_n),$$

where $A_1, \dots, A_n \in V - T$, $x_1, \dots, x_n \in V^*$

Propagating scattered context grammar (PSC grammar)

- each $(A_1, \dots, A_n) \to (x_1, \dots, x_n)$ satisfies $x_1, \dots, x_n \in V^+$

SC Grammar—Derivation Step

Derivation step

For $(A_1, \ldots, A_n) \rightarrow (x_1, \ldots, x_n) \in P$ and

$$u = u_1 A_1 \ldots u_n A_n u_{n+1}$$
$$v = u_1 x_1 \ldots u_n x_n u_{n+1}$$

we write $u \Rightarrow v \; [(A_1, \ldots, A_n) \rightarrow (x_1, \ldots, x_n)]$

Generated language

$L(G) = \{x \in T^* : S \Rightarrow^* x\}$

Generative power

- $\mathscr{L}(SC) = \mathscr{L}(RE)$
- $\mathscr{L}(CF) \subset \mathscr{L}(PSC) \subseteq \mathscr{L}(CS)$

SC Grammar—Example

Example

SC grammar $G = (\{A, B, C, S, a, b, c\}, \{a, b, c\}, P, S)$ with

$$\begin{aligned} P = \{&(S) \rightarrow (ABC), \\ &(A, B, C) \rightarrow (aA, bB, cC), \\ &(A, B, C) \rightarrow (\varepsilon, \varepsilon, \varepsilon)\} \end{aligned}$$

$S \Rightarrow ABC \Rightarrow aAbBcC \Rightarrow aaAbbBccC \Rightarrow aabbcc$
$L(G) = \{a^n b^n c^n \, : \, n \geq 0\}$

Maximal and Minimal Derivation

Production length

- $\text{len}((A_1, \ldots, A_n) \rightarrow (x_1, \ldots, x_n)) = |A_1 \ldots A_n| = n$

Maximal derivation step

$u \;_{\max}\!\Rightarrow v\ [p]$, $p \in P$ if and only if

1. $u \Rightarrow v\ [p]$,
2. there is no $r \in P$, $\text{len}(r) > \text{len}(p)$, such that $u \Rightarrow w\ [r]$

Minimal derivation step

$u \;_{\min}\!\Rightarrow v\ [p]$, $p \in P$ if and only if

1. $u \Rightarrow v\ [p]$,
2. there is no $r \in P$, $\text{len}(r) < \text{len}(p)$, such that $u \Rightarrow w\ [r]$

Maximal and Minimal Languages

Maximal and minimal languages

- $L_{\max}(G) = \{x \in T^* : S \ {}_{\max}\Rightarrow^* x\}$
- $L_{\min}(G) = \{x \in T^* : S \ {}_{\min}\Rightarrow^* x\}$
- language families denoted by $\mathscr{L}(PSC, \max)$ and $\mathscr{L}(PSC, \min)$

Example

SC grammar $G = (\{A, B, C, S, a, b, c\}, \{a, b, c\}, P, S)$ with

$$\begin{aligned} P = \{&(S) \rightarrow (ABC), \\ &(A) \rightarrow (a), (B) \rightarrow (b), (C) \rightarrow (c), \\ &(A, B, C) \rightarrow (aA, bB, cC), \\ &(A, B, C) \rightarrow (\varepsilon, \varepsilon, \varepsilon)\} \end{aligned}$$

$L_{\max}(G) = \{a^n b^n c^n : n \geq 0\}$,
$L_{\min}(G) = \{abc\}$

Main Result I

Theorem
$\mathscr{L}(CS) = \mathscr{L}(PSC, \max)$

State Grammar

State grammar (ST grammar)

$G = (V, T, K, P, S, p_0)$, where

V is a finite alphabet

T is a set of terminals, $T \subset V$

K is a finite set of states

S is the start symbol, $S \in V - T$

p_0 $p_0 \in K$

P is a finite set of productions of the form

$$(A, p) \rightarrow (x, q),$$

where $p, q \in K$, $A \in V - T$, $x \in V^+$

State Grammar—Derivation Step

Derivation step

For $(A, p) \rightarrow (x, q) \in P$,

$$u = (rAs, p),$$
$$v = (rxs, q),$$

where $r, s \in V^*$, and for every $(B, p) \rightarrow (y, t) \in P$, $B \notin \mathrm{alph}(r)$, we write

$$u \Rightarrow v \; [(A, p) \rightarrow (x, q)]$$

Generated language

$L(G) = \{x \in T^* : (S, p_0) \Rightarrow^* (x, q) \text{ for some } q \in K\}$

Generative power

$\mathscr{L}(ST) = \mathscr{L}(CS)$

Basic Idea I

CS language representation

- let $L \in \mathscr{L}(CS)$
- let $\bar{G} = (\bar{V}, T, K, \bar{P}, \bar{S}, p_0)$ be a state grammar such that $L = L(\bar{G})$
- let $(x_1 B x_2 A x_3 C x_4, p)$ be a sentential form of $\bar{G}$
- let $(x_1 B x_2 A x_3 C x_4, p) \Rightarrow (x_1 B x_2 x x_3 C x_4, q)$ $[(A, p) \rightarrow (x, q)]$ in $\bar{G}$

Construction of maximal PSC grammar

- $(x_1 B x_2 A x_3 C x_4, p)$ corresponds to $x_1 B x_2 \langle A, p\rangle x_3 C x_4$ in G
- simulate $(A, p) \rightarrow (x, q)$ by
  1. $(B, \langle A, p\rangle, C) \rightarrow (X, X, X)$ for every $(B, p) \rightarrow (y, r) \in P$
  2. $(B, \langle A, p\rangle) \rightarrow (\langle B, q\rangle, x)$

$$x_1 B x_2 \langle A, p\rangle x_3 C x_4 \Rightarrow x_1 \langle B, q\rangle x_2 x x_3 C x_4$$

Definitions

1. let $\Delta(t)$ be the set of all permutations of $\{1, \dots, t\}$
2. let permute$(n, m) =$
   $\{(i_1, \dots, i_{n+m}) \in \Delta(n+m) \, : \, 1 \leq i_k < i_l \leq n \text{ implies } k < l\}$
3. let reorder$((x_1, \dots, x_n), (i_1, \dots, i_n)) = (x_{i_1}, \dots, x_{i_n})$
   for $x_1, \dots, x_n \in V^*$, $(i_1, \dots, i_n) \in \Delta(n)$
4. let $\Pi((x_1, \dots, x_n), i) = x_i$

Example

1. $\Delta(3) = \{(1,2,3), (1,3,2), (2,1,3), (2,3,1), (3,1,2), (3,2,1)\}$
2. permute$(2, 1) = \{(1,2,3), (1,3,2), (3,1,2)\}$
3. reorder$((x_1, x_2, x_3), (3,1,2)) = (x_3, x_1, x_2)$
4. $\Pi((x_3, x_1, x_2), 3) = x_2$

Construction I

- let $\bar{G} = (\bar{V}, T, K, \bar{P}, \bar{S}, p_0)$
- set $Y = \{\langle A, q\rangle : A \in \bar{V} - T, q \in K\}$
- set $Z = \{\bar{a} : a \in T\}$
- define the homomorphism α, form $\bar{V}^*$ to $((\bar{V} - T) \cup Z)^*$ as
  - $\alpha(A) = A$ for all $A \in \bar{V} - T$ and
  - $\alpha(a) = \bar{a}$ for all $a \in T$
- set $V = \bar{V} \cup Y \cup Z \cup \{S, X\}$
- define the propagating scattered context grammar $G = (V, T, P, S)$ with P defined as follows:

Step 1

For every $x \in L(\bar{G})$, $|x| \leq 2$, add
$(S) \rightarrow (x)$

Construction II

Step 2

For every

$$(x, q) \in \{(x, q) : (\bar{S}, p_0) \Rightarrow^{+}_{\bar{G}} (x, q) \text{ for some } q \in K \\ \text{and } 3 \leq |x| \leq \min\{3, \max\{|\Pi(\text{rhs}(p), 1)| : p \in \bar{P}\}\}\},$$

where

1. $x \in T^*$, add
   $(S) \rightarrow (x)$
2. $x = x_1 A x_2$, $A \in \bar{V} - T$, $x_1, x_2 \in \bar{V}^*$, add
   $(S) \rightarrow (\alpha(x_1)\langle A, q\rangle\alpha(x_2))$

Construction III

Step 3

For every $(A, p) \rightarrow (x, q)$, $(B, p) \rightarrow (y, r) \in \bar{P}$, $C \in \bar{V}$,
$\Gamma_{21} \in \text{permute}(2, 1)$,

$$z = \text{reorder}((B, \langle A, p \rangle, \alpha(C)), \Gamma_{21}),$$

add
$z \rightarrow (X, X, X)$

Construction IV

Step 4

For every $(A, p) \rightarrow (x, q) \in \bar{P}$, $B \in \bar{V} - T$, $C \in \bar{V}$, $\Gamma_{11} \in \text{permute}(1, 1)$,

$$y = \text{reorder}((\langle A, p \rangle, \alpha(C)), \Gamma_{11}),$$

add

1. $(B, \langle A, p \rangle) \rightarrow (\langle B, q \rangle, \alpha(x))$,
2. $(\langle A, p \rangle, B) \rightarrow (\alpha(x), \langle B, q \rangle)$
3. If $x = vBw$, $v, w \in \bar{V}^*$, for every

   $$z = \text{reorder}((\alpha(v)\langle B, q \rangle \alpha(w), \alpha(C)), \Gamma_{11}),$$

   add $y \rightarrow z$
4. For every $u = \text{reorder}((\alpha(x), \alpha(C)), \Gamma_{11})$,
   add $y \rightarrow u$

Construction V

Step 5

For every $a \in T$, add
$(\bar{a}) \rightarrow (a)$

Main Result II

Theorem

$\mathscr{L}(CS) = \mathscr{L}(PSC, \min)$

Basic Idea II

CS **language representation**

- let $L \in \mathscr{L}(CS)$
- let $\bar{G} = (\bar{V}, T, K, \bar{P}, \bar{S}, p_0)$ be a state grammar such that $L = L(\bar{G})$
- let $(x_1 B x_2 A x_3 C x_4, p)$ be a sentential form of $\bar{G}$
- let $(x_1 B x_2 A x_3 C x_4, p) \Rightarrow (x_1 B x_2 x x_3 C x_4, q)$ $[(A, p) \rightarrow (x, q)]$ in $\bar{G}$

Construction of minimal *PSC* **grammar**

- $(x_1 B x_2 A x_3 C x_4, p)$ corresponds to $x_1 B x_2 \langle A, p \rangle x_3 C x_4$ in G
- simulate $(A, p) \rightarrow (x, q)$ by
  1. $(B, \langle A, p \rangle) \rightarrow (X, X)$ for every $(B, p) \rightarrow (y, r) \in P$
  2. $(B, \langle A, p \rangle, C) \rightarrow (\langle B, q \rangle, x, C)$

$$x_1 B x_2 \langle A, p \rangle x_3 C x_4 \Rightarrow x_1 \langle B, q \rangle x_2 x x_3 C x_4$$

Summary

$$\mathscr{L}(CF) \subset \mathscr{L}(PSC) \subseteq \mathscr{L}(CS)$$

By restricting derivations to

- the longest applicable productions, we obtain

$$\mathscr{L}(PSC, \max) = \mathscr{L}(CS)$$

- the shortest applicable productions, we obtain

$$\mathscr{L}(PSC, \min) = \mathscr{L}(CS)$$