k-Limited Erasing Performed by Regular-Controlled Context-Free Grammars

Petr Zemek

(Alexander Meduna)

Brno University of Technology
Faculty of Information Technology

2010-03-10

Regular-Controlled CFG

A *regular-controlled context-free grammar* is a pair, $H = (G, R)$, where

- $G = (V, T, S, P)$ is a context-free grammar and
- $R \subseteq P^*$ is a regular language (*control language*).

$$L(H) = \{w \in T^* \mid S \Rightarrow^* w\ [\alpha] \text{ with } \alpha \in R\}$$

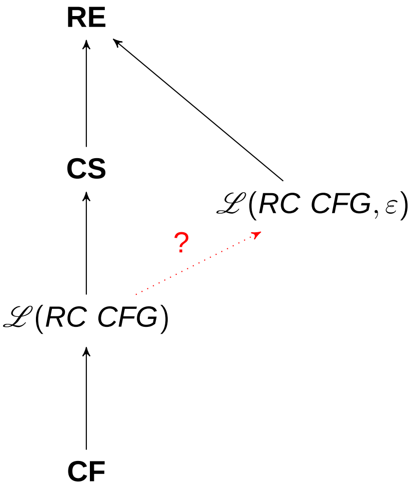
Example

Let $H = (G, R)$ be a regular-controlled context-free grammar, where P contains the following rules

$1: S \rightarrow ABC,$
$2: A \rightarrow aA,$
$3: B \rightarrow bB,$
$4: C \rightarrow cC,$
$5: A \rightarrow \varepsilon,$
$6: B \rightarrow \varepsilon,$
$7: C \rightarrow \varepsilon,$

and $R = \{1\}\{234\}^*\{567\}$.

Generative Power



k-Limited Erasing (Idea)

Regular-controlled context-free grammar H *erases its nonterminals in a k-limited way* provided that in all $S \Rightarrow^* x \Rightarrow^* y$, where $y \in L(H) - \{\varepsilon\}$,

for every symbol in x, which is not erased, x contains at most k nonterminals which are erased.

Example 1

Let $H = (G, R)$ be a regular-controlled context-free grammar, where P contains the following rules

$$1: S \rightarrow ABC, \quad 5: A \rightarrow \varepsilon,$$
$$2: A \rightarrow aA, \quad 6: B \rightarrow \varepsilon,$$
$$3: B \rightarrow bB, \quad 7: C \rightarrow \varepsilon,$$
$$4: C \rightarrow cC,$$

and $R = \{1\}\{234\}^*\{567\}$.

Example 2

Let $H = (G, R)$ be a regular-controlled context-free grammar, where P contains the following rules

$$1: S \to SS,$$
$$2: S \to a,$$
$$3: S \to \varepsilon,$$

and $R = \{1\}^*\{2\}^*\{3\}^*$.

k-Limited Erasing (Definition)

Regular-controlled context-free grammar H *erases its nonterminals in a k-limited way* provided that it satisfies this implication:

- if $S \Rightarrow^* y$ is a derivation of the form $S \Rightarrow^* x \Rightarrow^* y$, where $x \in V^+$ and $y \in L(H) - \{\varepsilon\}$, then in $\Delta(S \Rightarrow^* y)$, the $|x|$ subtrees rooted at all the symbols of x contain $|x|/(k+1)$ or more $+$-subtrees.

Main Result

Theorem. For every regular controlled context-free grammar I that erases its nonterminals in a k-limited way, there is a regular controlled context-free grammar M without ε-rules such that $L(M) = L(I) - \{\varepsilon\}$.

Algorithm – Input and Output

Input: A context-free grammar, $G = (V_G, T_G, S_G, P_G)$, and a regular grammar, $H = (V_H, T_H, S_H, P_H)$, such that regular controlled context-free grammar $I = (G, L(H))$ erases its nonterminals in a k-limited way.

Output: A context-free grammar without ε-rules, $O = (V_O, T_O, P_O, S_O)$, and a regular grammar, $Q = (V_Q, T_Q, P_Q, S_Q)$, such that $L(M) = L(I) - \{\varepsilon\}$ for a regular controlled context-free grammar $M = (O, L(Q))$.

Algorithm – Method 1/4

Method: Initially, set:

$$k' = k + max(\{|rhs(r)| \mid r \in P_G\});$$

$$\begin{aligned}
T_O &= T_G; \\
V_O &= T_O \cup \{\langle X, y\rangle \mid X \in V_G, y \in N_G^*, 0 \leq |y| \leq k'\}; \\
S_O &= \langle S_G, \varepsilon\rangle; \\
P_O &= \{\langle a, \varepsilon\rangle \rightarrow a \mid a \in T_G\}; \\
T_Q &= P_O; \\
V_Q &= T_Q \cup N_H \cup \{Z\}; \\
S_Q &= S_H; \\
P_Q &= \{Z \rightarrow \lfloor\langle a, \varepsilon\rangle \rightarrow a\rfloor Z \mid a \in T_G\} \cup \\
&\quad\ \{Z \rightarrow \lfloor\langle a, \varepsilon\rangle \rightarrow a\rfloor \mid a \in T_G\}.
\end{aligned}$$

Algorithm – Method 2/4

Method: Repeat (1) through (3), given next, until none of the sets P_O, P_Q can be extended in this way.

(1)

If $r: A \to x_0 X_1 x_1 X_2 x_2 \dots X_n x_n \in P_G$ and $\langle A, w \rangle, \langle X_1, w x_0 x_1 \dots x_n \rangle \in N_O$, where $X_i \in V_G$, for all $1 \leq i \leq n$, $x_j \in N_G^*$, for all $0 \leq j \leq n$, $w \in N_G^*$, for some $n \geq 1$

Then add $s: \langle A, w \rangle \to \langle X_1, w x_0 x_1 \dots x_n \rangle \langle X_2, \varepsilon \rangle \dots \langle X_n, \varepsilon \rangle$ to P_O;
for each $B \to r \in P_H$, add $B \to sZ$ to P_Q;
for each $B \to rC \in P_H$, $C \in N_H$, add $B \to sC$ to P_Q

Algorithm – Method 3/4

Method: Repeat (1) through (3), given next, until none of the sets P_O, P_Q can be extended in this way.

(2)

If $r : A \rightarrow w \in P_G$ and $\langle X, uAv \rangle, \langle X, uwv \rangle \in N_O$, where $X \in V_G, u, v, w \in N_G^*$

Then add $s : \langle X, uAv \rangle \rightarrow \langle X, uwv \rangle$ to P_O;
for each $B \rightarrow r \in P_H$, add $B \rightarrow sZ$ to P_Q;
for each $B \rightarrow rC \in P_H$, $C \in N_H$, add $B \rightarrow sC$ to P_Q.

Algorithm – Method 4/4

Method: Repeat (1) through (3), given next, until none of the sets P_O, P_Q can be extended in this way.

(3)

If $\langle X, uAv \rangle, \langle Y, w \rangle, \langle Y, wA \rangle \in N_O$, where $X, Y \in V_G$, $A \in N_G$, $u, v, w \in N_G^*$

Then add $r\colon \langle X, uAv \rangle \to \langle X, uv \rangle$ and $s\colon \langle Y, w \rangle \to \langle Y, wA \rangle$ to P_O; for each $B \in N_H$, add a new (unique) nonterminal C to N_Q and add $B \to rC$ and $C \to sB$ to P_Q.

Proof of Correctness (Idea)

Claim.

$$S_G \Rightarrow^m {}^{\varepsilon}x_0 X_1 {}^{\varepsilon}x_1 X_2 {}^{\varepsilon}x_2 \dots X_h {}^{\varepsilon}x_h \; [\alpha] \text{ in } G$$

if and only if

$$\langle S_G, \varepsilon \rangle \Rightarrow^n \langle X_1, u_1 \rangle \, \langle X_2, u_2 \rangle \dots \langle X_h, u_h \rangle \; [\gamma] \text{ in } O,$$

where $X_i \in V_G$, for all $1 \leq i \leq h$, ${}^{\varepsilon}x_j \in N_G^*$, for all $0 \leq j \leq h$, $u_1 u_2 \dots u_h \in perm({}^{\varepsilon}x_0 {}^{\varepsilon}x_1 \dots {}^{\varepsilon}x_h)$, $\alpha \in P_G^+$, $\gamma \in P_O^+$, and $h, m, n \geq 1$.

References

- A. Meduna and P. Zemek
  On Regular-Controlled Context-Free Grammars that Erase Nonterminals in a k-Limited Way.
  In: Fundamenta Informaticae (to appear).

- G. Rozenberg and A. Salomaa, editors
  Handbook of Formal Languages, Vol. 2: Linear Modeling: Background and Application.
  Springer, 1997. ISBN 3540614869

- C. Martín-Vide and G. Păun and V. Mitrana, editors
  Formal Languages and Applications.
  Springer, 2004. ISBN 3-540-20907-7.