

TWO-WAY K-LINEAR N-PC GRAMMAR SYSTEMS

Zbyněk KŘIVKA, Master Degree Programme (5)
Dept. of Information Systems, FIT, BUT
E-mail: xkrivk01@stud.fit.vutbr.cz

Supervised by: Dr. Alexandr Meduna

ABSTRACT

This paper introduces 2-way k-linear n-PC grammar systems, which can be used for parallel and cooperating parsing. Components use context free grammars (using LR parsing method). Co-operation is ensured by control set of n-tuples of nonterminals. In conclusion are discussed practical applications and introduction to proof of Turing machine's power.

1 ÚVOD

V moderní teoretické informatice posledních let je stále více populární využívat systémy obsahující více gramatik ve vzájemné interakci; říkáme jim gramatické systémy. Ty nám umožňují kombinovat několik gramatik (komponent), pro něž máme například efektivní syntaktické analyzátoři, a obdržet jazyk s větší popisnou silou.

V tomto článku je zkoumán gramatický systém pracující s komponentami konstruovanými z bezkontextových gramatik a se spolupracujícím mechanismem, který využívá kontrolní (či řídicí) množinu n-tic neterminálů. Díky tomu získáme prostředek pro příjem jazyků vyšší třídy než je třída bezkontextových jazyků podle Chomského hierarchie.

2 GRAMATICKÝ SYSTÉM

Definice bezkontextové gramatiky, přímé derivace a levé resp. pravé derivace (značme \Rightarrow a $l_m \Rightarrow$ resp. $r_m \Rightarrow$) naleznete v [1]. Použijeme-li pro přímou derivaci $uAv \Rightarrow uxv$ pravidlo $A \rightarrow x$, tak píšeme za lomítkem neterminál, jenž byl přepsán:

$$uAv/A \Rightarrow uxv \text{ nebo } uAv/A \text{ } l_m \Rightarrow uxv \text{ respektive } uAv/A \text{ } r_m \Rightarrow uxv$$

2.1 DEFINICE NOVÝCH POJMŮ

Nechť n je kladné celé číslo. **Neterminály-řízený k-lineární n-komponentní gramatický systém** (zkráceně GS) je $(n + 1)$ -tice Γ , kde:

$$\Gamma = (G_1, G_2, \dots, G_n, Q),$$

kde pro všechna $i = 1, 2, \dots, n$ máme bezkontextovou gramatiku $G_i = (N_i, T_i, P_i, S_i)$

(tzv. komponenta), pro níž lze sestrojít LR(k) syntaktický analyzátor a všechny množiny N_i jsou navzájem disjunktní, a podmnožinu kartézského součinu $Q \subseteq N_1 \times N_2 \times \dots \times N_n$. Množinu n -tic neterminálů značenou Q budeme nazývat **řídící množinou gramatického systému**.

Konfigurace gramatického systému je n -tice tvaru (x_1, \dots, x_n) , kde $x_i \in (N_i \cup T_i)^*$, $1 \leq i \leq n$.

Počáteční konfigurace σ je definována jako $\sigma = (S_1, \dots, S_n)$.

Nechť Θ označuje množinu všech konfigurací gramatického systému Γ a necht' Φ značí množinu všech koncových konfigurací tvaru (x_1, \dots, x_n) , kde $x_i \in T_i^*$, $1 \leq i \leq n$.

Mějme konfigurace $y, x \in \Theta$ tvaru (x_1, \dots, x_n) a (y_1, \dots, y_n) . Pak píšeme $y \Rightarrow x$ v Γ (**přímá derivace**), jestliže existuje v řídící množině Q taková n -tice $(A_1, A_2, \dots, A_n) \in Q$, že gramatiky G_i umožňují provést přímé derivace $y_i/A_i \Rightarrow x_i$ v G_i , pro všechna i , kde $1 \leq i \leq n$.

Analogicky definujeme **levou a pravou derivaci** v gramatickém systému:

$y \xrightarrow{lm} x$ v Γ , jestliže existuje $(A_1, A_2, \dots, A_n) \in Q$ a $y_i/A_i \xrightarrow{lm} x_i$ v gramatice G_i , $1 \leq i \leq n$.

$y \xrightarrow{rm} x$ v Γ , jestliže existuje $(A_1, A_2, \dots, A_n) \in Q$ a $y_i/A_i \xrightarrow{rm} x_i$ v gramatice G_i , $1 \leq i \leq n$.

Nechť \Rightarrow^* , \xrightarrow{lm}^* , \xrightarrow{rm}^* označují tranzitivní a reflexivní uzávěr operací \Rightarrow , \xrightarrow{lm} a \xrightarrow{rm} .

Pro každou konfiguraci $\xi = (x_1, \dots, x_n) \in \Theta$ polož $\mu(\xi) = \{x_1, \dots, x_n\}$.

Jazyk nad Γ značený jako $L(\Gamma)$ je definován následovně:

$$L(\Gamma) = \{z \mid z \in \mu(\xi), \text{ kde } \sigma \Rightarrow^* \xi \text{ v } \Gamma \text{ a } \xi \in \Phi\}$$

Analogicky varianty jazyků využívající pouze levou resp. pravou derivace řetězců:

$$\xrightarrow{lm} L(\Gamma) = \{z \mid z \in \mu(\xi), \text{ kde } \sigma \xrightarrow{lm}^* \xi \text{ v } \Gamma \text{ a } \xi \in \Phi\}$$

$$\xrightarrow{rm} L(\Gamma) = \{z \mid z \in \mu(\xi), \text{ kde } \sigma \xrightarrow{rm}^* \xi \text{ v } \Gamma \text{ a } \xi \in \Phi\}$$

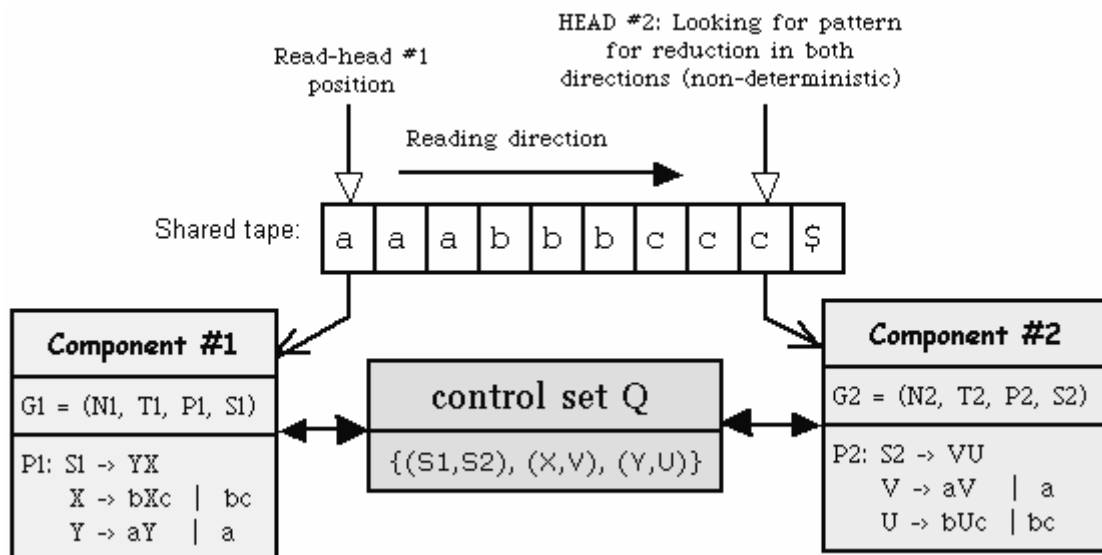
Poznamenejme, že jazyk nad takovýmto gramatickým systémem lze definovat i několika dalšími způsoby, které mohou být v praxi výhodné.

2.2 PŘÍKLAD

Na následujícím obrázku je zobrazen GS přijímající jazyk $L = \{a^i b^i c^i \mid i > 0\}$, který není bezkontextový. Tento GS je mírně modifikován, takže používá jednu sdílenou pásku místo dvou různých pásek se stejným obsahem, ale to je pouze implementační detail.

Lze provést i další praktické modifikace GS, na které zde není dostatek prostoru. Například modifikovat některé komponenty, aby přijímaly místo jazyka $L_k = \{w\}$, jazyk $L_{\text{revk}} = \{w^c \mid w^c \text{ je reverzní řetězec z řetězce } w \in L_k\}$. Toho lze jednoduše dosáhnout algoritmickou změnou pravidel gramatiky pro danou komponentu. S využitím takovéto řeckně reverzní komponenty lze sestrojít 3-komponentní GS, který přijímá jazyk $L_{ww} = \{ww \mid w \in \Sigma^*\}$, kde Σ je vstupní abeceda} (viz. [4]).

Bohužel zvýšením mocnosti přijímaných jazyků narůstá i míra nedeterminismu při praktické syntaktické analýze složitějších gramatik. Nedeterminismus daného GS lze však studovat a v mnoha případech se mu vyhnout využitím některých běžných mechanismů teoretické informatiky. Např. místo silně nedeterministického jazyka L_{ww} analyzovat mírně upravený jazyk $L_{w\#w} = \{w\#w \mid w \in (\Sigma \setminus \{\#\})^*\}$.



Obr. 1: Příklad gramatického systému pro jazyk $a^i b^j c^i$ (není bezkontextový)

2.3 MOCNOST TURINGOVA STROJE

Již intuitivně vidíme, že pro takto navržený model lze bez problémů sestavit 2-komponentní GS přijímající například jazyk $L_{AB} = \{ w \mid w \in L_A \wedge w \in L_B, \text{ kde } L_A, L_B \text{ jsou bezkontextové jazyky} \}$, což dokáže pouze prostředek s mocností Turingova stroje.

Rigorózní důkaz tvrzení, že model gramatických systémů řízených množinou neterminálů je k nalezení v [4] i s podrobným popisem a slovním rozebráním.

3 ZÁVĚR

GS tedy značně rozšiřuje mocnost přijímaných jazyků výše než L_2 (bezkontextové) nejen díky schopnosti přijímat průniky bezkontextových jazyků. Pro svůj složitý a někdy nedeterministický mechanismus spolupráce komponent se však nehodí pro syntaktickou analýzu bezkontextových jazyků, protože pro ty již máme dostatečně efektivní nástroje i pro paralelní překlad [3]. Zajímavé by však bylo studovat využitelnost pro analýzu jazyků třídy L_1 a L_0 .

LITERATURA

- [1] Češka, M.: Gramatiky a jazyky, VUT FEI Brno, 1992
- [2] Meduna, A. : Automata and Languages, Springer, London, 2000, ISBN 1-85233-074-0
- [3] Elbl, S.: Parallel Translation Based On Grammar Systems, In Proceedings of the International Conference and Competition STUDENT EEICT 2003, Brno, 2003, ISBN-80-214-2401-X, str. 269-273
- [4] Křivka, Z.: Syntaktická analýza založená na multigenerování [diplomová práce], VUT Brno FIT, 2004