$P(\text{izraz} \mid \text{gramatika})$

Urh Primožič
Mentor: Ljupčo Todorovski
Somentor: Matej Petković

Fakulteta za matematiko in fiziko

29. 8. 2022

Gramatika

Definicija

$G = (N, T, R, S)$ *je* ***kontekstno neodvisna gramatika****, kjer*

- N, T ***končni****, disjunktni množici simbolov*
- $S \in N$ *začetni simbol*
- $R \subset N \times (N \cup T)^*$ *celovita relacija*

Za $(A, \alpha) \in R$ pišemo $A \to \alpha$

Primer

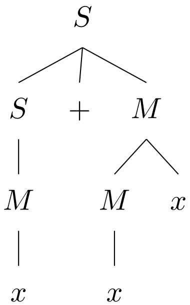
$N = \{S, M\}, T = \{x, +\}$

$$
\begin{aligned}
S &\to S + M \mid M \\
M &\to Mx \mid x
\end{aligned}
$$

Primer

$$N = \{S, M\}, T = \{x, +\}$$

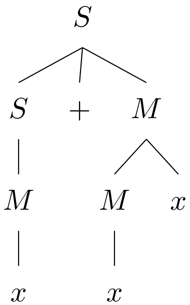
$$\begin{aligned} S &\to S + M \mid M \\ M &\to Mx \mid x \end{aligned}$$



Primer

$$N = \{S, M\}, T = \{x, +\}$$

$$\begin{aligned} S &\rightarrow S + M \mid M \\ M &\rightarrow Mx \mid x \end{aligned}$$

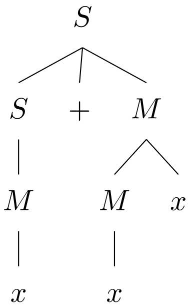


$$S \rightarrow S{+}M \rightarrow M{+}M \rightarrow x{+}M \rightarrow x{+}Mx \rightarrow x{+}xx$$

Primer

$$N = \{S, M\}, T = \{x, +\}$$

$$\begin{aligned} S &\to S + M \mid M \\ M &\to Mx \mid x \end{aligned}$$



$$S \to S{+}M \to M{+}M \to x{+}M \to x{+}Mx \to x{+}xx$$

$L(G)$ so polinomi nad $\mathbb{N}$ brez prostega člena

Verjetnostne gramatike

Definicija

Verjetnostna gramatika G je gramatika skupaj s preslikavo

$$P\colon R \to [0,1]$$

za katero velja

$$\sum_{(A\to\alpha)\in R} P(A \to \alpha) = 1$$

za vsak $A \in N$.

Definicija

Naj bo τ izpeljevalno drevo

$$P(\tau) = \prod_{r \in R} P(r)^{f_\tau(r)}$$

$f_\tau(r)$ *število pojavitev* r *v* τ

Definicija

Naj bo τ izpeljevalno drevo

$$P(\tau) = \prod_{r \in R} P(r)^{f_\tau(r)}$$

$f_\tau(r)$ *število pojavitev r v τ*

Posledica

Naj bo $w \in L(G)$

$$P(w) = \sum_{\tau \text{ izpelje } w} P(\tau)$$

Definicija

Naj bo τ izpeljevalno drevo

$$P(\tau) = \prod_{r \in R} P(r)^{f_\tau(r)}$$

$f_\tau(r)$ *število pojavitev r v τ*

Posledica

Naj bo $w \in L(G)$

$$P(w) = \sum_{\tau \text{ izpelje } w} P(\tau)$$

Rezultat

Algoritem za izračun $P(w)$ obstaja

$$x + x \neq 2.1x$$

$$cx + cx \sim cx$$

Definicija

- *domena za spremenljivke* D
- *domena za konstante* $\mathbb{F}$
- *beseda* $w = w_1 c w_2 \cdots c w_{n+1} \in L(G)$

$$\begin{aligned}\Phi \colon L(G) &\longrightarrow 2^{\{f \colon U \to K \mid U \subseteq D\}}\\ w &\longmapsto \{(x_1, \dots, x_n) \mapsto w_1 c_1 w_2 \cdots c_n w_{n+1} \mid c_1, \dots, c_n \in \mathbb{F}\}\end{aligned}$$

za $K = \{w_1 c_1 w_2 c_2 \cdots c_n w_{n+1} \mid w \in L(G) \wedge c_i \in \mathbb{F}\}$

Definicija

- *domena za spremenljivke* D
- *domena za konstante* $\mathbb{F}$
- *beseda* $w = w_1 c w_2 \cdots c w_{n+1} \in L(G)$

$$\Phi\colon L(G) \longrightarrow 2^{\{f\colon U \to K \mid U \subseteq D\}}$$
$$w \longmapsto \{(x_1, \ldots, x_n) \mapsto w_1 c_1 w_2 \cdots c_n w_{n+1} \mid c_1, \ldots, c_n \in \mathbb{F}\}$$

za $K = \{w_1 c_1 w_2 c_2 \cdots c_n w_{n+1} \mid w \in L(G) \wedge c_i \in \mathbb{F}\}$

Definicija

$$w \sim v \iff \Phi(w) = \Phi(v)$$

Verjetnost izraza

$$P([w]) = \sum_{v \in [w] \cap L(G)} P(v)$$

Rezultat

Problem je neizračunljiv.

Linearna gramatika

$$
\begin{array}{rcl}
E & \to & E + cV \;\; [p] \mid c \;\; [1-p] \\
V & \to & x_1 \;\; [q_1] \mid \cdots \mid x_n \;\; [q_n].
\end{array}
$$

Beseda $w = c + cx_{r_1} + \cdots + cx_{r_k} \in L(G)$

$$
\begin{aligned}
P([w]) &= \sum_{I \subseteq \{1,\ldots,k\}} (-1)^{|I|} \frac{1-p}{1 - p \sum\limits_{i \in \{1,\ldots,k\} \setminus I} q_{r_i}} \\
&= \sum_{i=k}^{\infty} (1-p) p^i \left(\sum_{\substack{l_1 + \cdots + l_k = i \\ l_j \geq 1}} \binom{i}{l_1, \ldots, l_k} q_{r_1}^{l_1} \cdots q_{r_k}^{l_k} \right)
\end{aligned}
$$

Polinomska gramatika

$$\begin{array}{lcl} E & \rightarrow & E + cV \;\; [p] \mid c \;\; [1-p] \\ V & \rightarrow & VF \;\; [q] \mid F \;\; [1-q] \\ F & \rightarrow & x_1 \;\; [q_1] \mid \cdots \mid x_n \;\; [q_n] \end{array}$$

Racionalna gramatika

$$
\begin{array}{rcl}
S & \rightarrow & E/E \;\; [1] \\
E & \rightarrow & E + cV \;\; [p] \mid c \;\; [1-p] \\
V & \rightarrow & VF \;\; [q] \mid F \;\; [1-q] \\
F & \rightarrow & x_1 \;\; [q_1] \mid \cdots \mid x_n \;\; [q_n]
\end{array}
$$