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

Urh Primožič
Mentor: Ljupčo Todorovski
Somentor: Matej Petković

Fakulteta za matematiko in fiziko

22. 11. 2021

Gramatika

Definicija

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

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

Za $(A, X_1 \cdots X_n) \in R$ pišemo $A \to X_1 \cdots X_n$

Množica besed

Definicija

Gramatika G generira besedo $w \in T^*$, če jo lahko dobimo iz S z zaporedno uporabo ***končnega*** *števila prepisovalnih pravil iz R. $L(G)$ je množica besed, ki jih generira G.*

Primer

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

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

Verjetnostne gramatike

Definicija

Verjetnostna gramatika G je gramatika skupaj s preslikavo

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

če velja

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

za vsak $A \in N$.

Definicija

Naj bo τ izpeljevalno drevo. Definiramo

$$P(\tau) = \prod_{(A\to\alpha)\in R} P(A \to \alpha)^{f(A\to\alpha)},$$

kjer je $f(A \to \alpha)$ število pojavitev $A \to \alpha$ v τ.

Definicija

Naj bo τ izpeljevalno drevo. Definiramo

$$P(\tau) = \prod_{(A \to \alpha) \in R} P(A \to \alpha)^{f(A \to \alpha)},$$

kjer je $f(A \to \alpha)$ število pojavitev $A \to \alpha$ v τ.

Definicija

Naj bo $w \in L(G)$:

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

Primer od prej

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

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

Izraz

- G gramatika, ki generira smiselne izraze
- $T = \{c, x_1, \ldots, x_n, +, -, *, /\}$
- $\mathbb{F}$ domena za konstante
- $D = (D_1, \ldots, D_n)$ domena za $(x_1, \ldots, x_n)$

Izraz

- G gramatika, ki generira smiselne izraze
- $T = \{c, x_1, \ldots, x_n, +, -, *, /\}$
- $\mathbb{F}$ domena za konstante
- $D = (D_1, \ldots, D_n)$ domena za $(x_1, \ldots, x_n)$

Definicija

Naj bo $w = w_0\mathbf{c}w_1 \cdots \mathbf{c}w_m \in L(G) \implies w_0\mathbf{c_0}w_1 \cdots \mathbf{c_m}w_m$. *Definiramo*

$$\Phi(w) = \{\varphi_{c_1,\ldots,c_m}(w) \colon D \to \mathbb{F} \mid c_i \in \mathbb{F} \wedge \varphi_{c_1,\ldots,c_m}(w) \textit{ je definirana}\}.$$

$$\varphi_{c_1,\ldots,c_m}(w)(y_1, \ldots, y_n) = w_0c_0w_1 \cdots c_mw_m \,|_{x_i=y_i}$$

Izraz

- G gramatika, ki generira smiselne izraze
- $T = \{c, x_1, \ldots, x_n, +, -, *, /\}$
- $\mathbb{F}$ domena za konstante
- $D = (D_1, \ldots, D_n)$ domena za $(x_1, \ldots, x_n)$

Definicija

Naj bo $w = w_0 \mathbf{c} w_1 \cdots \mathbf{c} w_m \in L(G) \implies w_0 \mathbf{c_0} w_1 \cdots \mathbf{c_m} w_m$.
Definiramo

$$\Phi(w) = \{\varphi_{c_1,\ldots,c_m}(w) \colon D \to \mathbb{F} \mid c_i \in \mathbb{F} \wedge \varphi_{c_1,\ldots,c_m}(w) \textit{ je definirana}\}.$$

$$\varphi_{c_1,\ldots,c_m}(w)(y_1, \ldots, y_n) = w_0 c_0 w_1 \cdots c_m w_m \,|_{x_i = y_i}$$

Za $u, v \in L(G)$

$$w \sim v \Leftrightarrow \Phi(w) = \Phi(v).$$

Izraz

- G gramatika, ki generira smiselne izraze
- $T = \{c, x_1, \ldots, x_n, +, -, *, /\}$
- $\mathbb{F}$ domena za konstante
- $D = (D_1, \ldots, D_n)$ domena za $(x_1, \ldots, x_n)$

Definicija

Naj bo $w = w_0 \mathbf{c} w_1 \cdots \mathbf{c} w_m \in L(G) \implies w_0 \mathbf{c_0} w_1 \cdots \mathbf{c_m} w_m$.
Definiramo

$$\Phi(w) = \{\varphi_{c_1,\ldots,c_m}(w) \colon D \to \mathbb{F} \mid c_i \in \mathbb{F} \wedge \varphi_{c_1,\ldots,c_m}(w) \textit{ je definirana}\}.$$

$$\varphi_{c_1,\ldots,c_m}(w)(y_1, \ldots, y_n) = w_0 c_0 w_1 \cdots c_m w_m \mid_{x_i = y_i}$$

Za $u, v \in L(G)$

$$w \sim v \Leftrightarrow \Phi(w) = \Phi(v).$$

Definicija

Množica izrazov $I(G) = L(G)/_\sim$ *je množica ekvivalenčnih razredov za* $\sim$.

Definicija

Za $[w] \in I(G)$ *definiram*

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

Cilj

Algoritem za izračun verjetnosti za neko družino gramatik.