

Učenje gramatike na temelju skupa uzoraka za učenje

(engl. Grammatical Inference)

- Učenje gramatike \rightarrow određivanje gramatike na temelju skupa uzoraka za učenje (uzorci su predloženi kao nizovi terminala)

Formalno:

$$U_M \rightarrow \{G_1, G_2, \dots, G_M\},$$

gdje je U_M skup uzoraka za učenje koji se sastoji od podskupova

$$U_i'; \quad i=1, 2, \dots, M$$

- svaki podskup U_i' odgovara skupu uzoraka za učenje (nizovima terminala) $\{x\}_i$

- neka vrijedi $i: L(G_1) \cap L(G_2) \dots \cap L(G_M) = \emptyset$

$$- U_i' \subseteq \{x \mid x \in L(G_i)\}$$

*
komentar

Postupak učenja gramatike iz skupa uzoraka $U_i' \subseteq L(G_i)$ (J. Feldman, 1969; (Tou, Gonzalez, 1974)

PRISTUP:

1. Konstruirati neekurzivnu gramatiku koja generira točno zadane nizove iz skupa za učenje

Ako je $U_i = L(G_i)$ gramatiku G_i možemo dobiti jednostavnim postupkom pregleda nizova $x \in L(G_i)$

↑
obično to u praksi nije slučaj

$$U_i \subset L(G_i)$$

zahtjevamo da je skup uzoraka za učenje STRUKTURNO CIELOVIT što znači da se u skupu uzoraka za učenje nalazi toliko rečenica (nizova x) iz $L(G_i)$ da su u oblikovanju skupa za učenje upotrebljena sva produkcijska pravila P iz G_i .

2. Nakon toga udružujemo nizove znakova tako da dobivamo jednostavniju rekurzivnu gramatiku koja može generirati beskonačan broj nizova (rečenica);
3. Pojednostaviti rekurzivnu gramatiku

Primjer:

zadan je skup uzoraka za učenje U_i' iz razreda C_i :

$$U_i' = \{caaaab, bbaab, caab, bbaab, cab, bbb, cb\}$$

Potražimo gramatiku G_i koja generira nizove iz U_i' !

① Sastavimo nerekurzivnu gramatiku koja generira točno zadani skup nizova.

- Nizove obrađujemo tako da najprije razmatramo niz s najvećim brojem znakova / nizove obrađujemo po padajućoj duljini!
- Sastavljamo produkcijska pravila i pridodajemo ih gramatici.
- Zadnje pravilo koje je upotrebљeno za generiranje najduljeg niza neka ima oblik

$$A \rightarrow a_1 a_2,$$

(naziva se i "residue production") \rightarrow OSTATAK
gdje je $A \in V_N$ i $a_1, a_2 \in V_T$

Opaska: Pretpostavljamo da će ostatak najduljeg niza biti sufiks nekog kraćeg niza iz skupa za učenje.

- Mogli smo izabrati i dulje ostatke

$$A \rightarrow a_1 a_2 \dots a_n$$

npr. $a_1 a_2 a_3$ ili $a_1 a_2 a_3 a_4$, međutim ti primjeri zahtijevaju potpuniji skup zadanih uzoraka.

- Najdulji niz u zadanom skupu je:

caaab

- generiramo ga pomoću sljedećih produkcija:

$$S \rightarrow c A_1$$

$$A_1 \rightarrow a A_2$$

$$A_2 \rightarrow a A_3$$

$$A_3 \rightarrow ab \rightarrow \text{ostatak najduljeg niza}$$

$$/ S \Rightarrow c A_1 \Rightarrow c a A_2 \Rightarrow c a a A_3 \Rightarrow c a a a b /$$

- Sljedeći niz je:

bbaab

Generiramo ga slijedom produkcijskih pravila:

$$S \rightarrow b A_4$$

$$A_4 \rightarrow b A_5$$

$$A_5 \rightarrow a A_6$$

$$A_6 \rightarrow ab$$

ostatak duljine 2 se i tu zahtijeva jer je ovaj niz jednake duljine kao i prethodni niz!

Možemo primjetiti da drugi niz možemo generirati i dodatnim pravilima:

$$S \rightarrow bA_4$$

$$A_4 \rightarrow bA_2$$

$$/ S \Rightarrow bA_4 \Rightarrow bbA_2 \Rightarrow bbaA_3 \Rightarrow bbaab /$$

Međutim, u ovom prvom dijelu postupka bavimo se samo odredjivanjem pravila koja generiraju zadani skup nizova.

- Treći niz: caab

dobivamo tako da dodamo slijedeće pravilo skupu pravila:

$$A_3 \rightarrow b$$

$$/ S \Rightarrow cA_1 \Rightarrow caA_2 \Rightarrow caaA_3 \Rightarrow caab /$$

- Razmatranjem preostalih nizova iz skupa nizova dobivamo konačan skup produkcijskih pravila P :

$$S \rightarrow cA_1 / bA_4$$

$$A_1 \rightarrow b / aA_2$$

$$A_2 \rightarrow b / aA_3$$

$$A_3 \rightarrow b / ab$$

$$A_4 \rightarrow bA_5$$

$$A_5 \rightarrow b / aA_6$$

$$A_6 \rightarrow b / ab$$

② U ovom dijelu dobivamo rekursivnu gramatiku udruživanjem produkcija oblika $A \rightarrow a_1 a_2$ s produkcijama koji taj oblik nemaju.

- Svako pravilo oblika $A_r \rightarrow a_1 a_2$ udružujemo s pravilom $A_n \rightarrow a_1 A_m$, gdje je $A_m \rightarrow a_2$.

To napravimo tako da zamijenimo A_r s A_n i brišemo pravila oblika $A_r \rightarrow a_1 a_2$

U našem slučaju imamo:

$$A_6 \rightarrow b/ab$$

$$A_5 \rightarrow b/aA_6$$

$$A_5 \rightarrow b$$

$$A_5 \rightarrow aA_5$$

$$A_6 \rightarrow ab \text{ brišemo!}$$

$$A_3 \rightarrow b/ab$$

$$A_2 \rightarrow b/aA_3$$

$$A_2 \rightarrow b$$

$$A_2 \rightarrow aA_2$$

$$A_3 \rightarrow ab, \text{ brišemo!}$$

Dobivamo:

$$S \rightarrow cA_1 \mid bA_4$$

$$A_1 \rightarrow b \mid aA_2$$

$$A_2 \rightarrow b \mid aA_2 \mid b$$

$$A_4 \rightarrow b \mid A_5$$

$$A_5 \rightarrow b \mid A_5 \mid b$$

Rekurzivne produkcije su:

$$A_2 \rightarrow aA_2 \quad i \quad A_5 \rightarrow bA_5$$

- ③. Pojednostavljujemo gramatiku dobivenu u dijelu ②.

To postizemo udruživanjem ekvivalentnih produkcijskih pravila.

Ekvivalentna produkcijska pravila:

Dva pravila sa lijevim dijelovima A_i i A_j su **ekvivalentna** ako zadovoljavaju sljedeći uvjet:

- pretpostavimo da pravilo koje na lijevoj strani ima A_i generira skup nizova

$$\{x\}_{A_i}$$

- pravilo koje na lijevoj strani ima A_j neka generira skup nizova

$$\{x\}_{A_j}$$

- ako je $\{x\}_{A_i} = \{x\}_{A_j}$

pravila su **EKVIVALENTNA**

Drugim riječima:

A_i možemo zamijeniti s A_j ako vrijedi:

$$\{x \mid A_i \stackrel{*}{\Rightarrow} x\} \equiv \{x \mid A_j \stackrel{*}{\Rightarrow} x\}$$

U našem slučaju su pravila koja imaju na lijevoj strani A_1 i A_2 ekvivalentna. Udružimo A_1 i A_2 i izbrisimo pravilo koje se više puta pojavljuje u skupu pravila:

$$\begin{aligned} S &\rightarrow cA_1 / bA_4 \\ A_1 &\rightarrow b/aA_2 \\ A_2 &\rightarrow b/aA_2/b \\ A_4 &\rightarrow b/A_5 \\ A_5 &\rightarrow b/aA_5/b \end{aligned}$$

Dobivamo:

$$\begin{aligned} S &\rightarrow cA_1 / bA_4 \\ A_1 &\rightarrow b/aA_1 \\ A_4 &\rightarrow bA_5 \\ A_5 &\rightarrow b/aA_5/b \end{aligned}$$

Pravila koja na lijevoj strani imaju A_1 i A_5 su ekvivalentna zato ih udružujemo te dobivamo:

$$\begin{aligned} S &\rightarrow cA_1 & A_1 &\rightarrow aA_1 \\ S &\rightarrow bA_4 & A_4 &\rightarrow bA_1 & A_1 &\rightarrow b \end{aligned}$$

Na temelju ulaznog skupa nizova odredili smo gramatiku koja generira uzorke iz razreda C_i :

$$G_i = (V_N, V_T, P, S)$$

$$V_N = \{S, A, B\}$$

$$V_T = \{a, b, c\}$$

P:

$$S \rightarrow cA$$

$$S \rightarrow bB$$

$$A \rightarrow aA$$

$$B \rightarrow bA$$

$$A \rightarrow b$$