

ECE59500NL Lecture 22: Probabilistic Context-Free Grammars (PCFGs)—I

Jeffrey Mark Siskind

School of Electrical and Computer Engineering

Spring 2021



© 2021 Jeffrey Mark Siskind. All rights reserved.

Size Hyperparameters and Indices

- ▶ L : number of sentences
- ▶ l : sentence index
- ▶ $I(l)$: number of words in sentence l
- ▶ i : word position in sentence l
- ▶ K : number of words in dictionary
- ▶ k : word index in dictionary
- ▶ J : number of nonterminals
- ▶ j : nonterminal index

Random Variables

- ▶ $T_{li_1i_2} \in \{1, \dots, J\}$: nonterminal index that generates substring from position i_1 through i_2 of sentence l
- ▶ $W_{li} \in \{1, \dots, K\}$: word index at position i of sentence l

Model Parameters

- ▶ $b_j = \Pr(T_{l|l(l)} = j) \ (\forall l)$
- ▶ $a_{j_1 j_2 j_3} = \Pr(T_{li_1 i_2} = j_2, T_{li_2+1 i_3} = j_3 | T_{li_1 i_3} = j_1) \ (\forall l)(\forall i_1)(\forall i_2 \geq i_1)(\forall i_3 \geq i_2 + 1)$
- ▶ $c_{jk} = \Pr(W_{li} = k | T_{lii} = j) \ (\forall l)(\forall i)$

To compute $s(T)$.

- 1 Sample either W from c_T or $T'T''$ from a_T .
- 2 If sampled W , return W .
- 3 If sampled $T'T''$, return $s(T')s(T'')$.
- 1 Sample T from b .
- 2 Return $s(T)$.

► $\alpha_{li_1 i_2 j} = \Pr(W_{li_1} = k_{li_1}, \dots, W_{li_2} = k_{li_2}, T_{li_1 i_2} = j | a, b, c)$

Outside Probabilities

► $\beta_{li_1 i_2 j} = \Pr(W_{l1} = k_{l1}, \dots, W_{li_1-1} = k_{li_1-1}, W_{li_2+1} = k_{li_2+1}, \dots, W_{ll(l)} = k_{ll(l)}, T_{li_1, i_2} = j | a, b, c)$

Inside Algorithm

- ▶ $\alpha_{liij} = c_{jw_{li}} (\forall l)(\forall i)(\forall j)$
- ▶ $\alpha_{li_1i_2j} = \sum_{j_1, j_2, i_1 \leq i, i+1 \leq i_2} a_{jj_1j_2} \alpha_{li_1ij_1} \alpha_{li+1i_2j_2} (\forall l)(\forall i_1)(\forall i_2)(\forall j)$

Outside Algorithm

- ▶ $\beta_{l1I(l)j} = b_j (\forall l)(\forall j)$
- ▶ $\beta_{li_1i_2j} = \sum_{j_1, j_2, i < i_1} a_{j_1j_2j} \alpha_{lii_1-1j_2} \beta_{lii_2j_1} + \sum_{j_1, j_2, i_2 < i} a_{j_1jj_2} \alpha_{li_2+1ij_2} \beta_{li_1ij_1}$

Likelihood Estimation

- ▶ $\Pr(W_{l1} = k_{l1}, \dots, W_{lI(l)} = k_{lI(l)} | a, b, c)$
- ▶ $= \sum_j b_j \alpha_{l1I(l)j}$
- ▶ $= \sum_j \beta_{liij} c_{lW_{li}} \ (\forall i)$

Inside-Outside Algorithm

► $\gamma_{li_1i_2j} = \Pr(W_{l1} = k_{l1}, \dots, W_{l(l)} = k_{l(l)}, T_{li_1i_2} = j | a, b, c) \propto \alpha_{li_1i_2j} \beta_{li_1i_2j}$

Baker-Lari-Young Reestimation Procedure

- ▶ $b_j : \propto \sum_l \gamma_{lI(l)j} \ (\forall j)$
- ▶ $a_{j_1 j_2 j_3} : \propto \sum_{l, i_1 \leq i_2, i_2+1 \leq i_3} \gamma_{li_1 i_3 j_1} \gamma_{li_1 i_2 j_2} \gamma_{li_2+1 i_3 j_3} \ (\forall j_1)(\forall j_2)(\forall j_3)$
- ▶ $c_{jk} : \propto \sum_{l, i, W_{li}=k} \gamma_{liij} \ (\forall j)(\forall k)$