Simon Wachter: Assignment 4

siwachte@ethz.ch, 19-920-198

13/01/2023 - 13:59h

Question 1:

a)

$$\begin{aligned}
\sum_{w\in\Sigma^*}\tilde{p}(w) &= \sum_{w\in\Sigma^*,|w|=1}\tilde{p}(w) + \sum_{w\in\Sigma^*,|w|=2}\tilde{p}(w) + \sum_{w\in\Sigma^*,|w|=3}\tilde{p}(w) + \dots && (1)\\
&= \sum_{w\in\Sigma}\tilde{p}(w) + \sum_{w\in\Sigma^*,|w|=2}\tilde{p}(w) + \sum_{w\in\Sigma^*,|w|=3}\tilde{p}(w) + \dots && (2)\\
&= 1 + \sum_{w\in\Sigma^*,|w|=2}\tilde{p}(w) + \sum_{w\in\Sigma^*,|w|=3}\tilde{p}(w) + \dots && (3)\\
&= 1 + \sum_{w\in\Sigma,|w|=1}\sum_{w'\in\Sigma^*}\tilde{p}(w)\tilde{p}(w') + \sum_{w\in\Sigma^*,|w|=3}\tilde{p}(w) + \dots && (4)\\
&= 1 + \sum_{w\in\Sigma}\tilde{p}(w)\sum_{w'\in\Sigma^*,|w|=1}\tilde{p}(w') + \sum_{w\in\Sigma^*,|w|=3}\tilde{p}(w) + \dots && (5)\\
&= 1 + \sum_{w\in\Sigma}\tilde{p}(w)\sum_{w'\in\Sigma}\tilde{p}(w') + \sum_{w\in\Sigma^*,|w|=3}\tilde{p}(w) + \dots && (6)\\
&= 1 + 1\cdot 1 + \sum_{w\in\Sigma^*,|w|=3}\tilde{p}(w) + \dots && (7)\\
&= 1 + 1\cdot 1 + \sum_{w\in\Sigma}\sum_{w\in\Sigma^*,|w|=2}\tilde{p}(w)\tilde{p}(w') + \dots && (8)\\
&= 1 + 1\cdot 1 + \sum_{w\in\Sigma}\tilde{p}(w)\sum_{w\in\Sigma^*,|w|=2}\tilde{p}(w') + \dots && (9)\\
&= 1 + 1\cdot 1 + 1\cdot 1\cdot 1 + \dots && (10)
\end{aligned}$$

We can see that the series equation 10 diverges to ∞.

b) We first state some equations that are used later:

$$\sum_{w \in \Sigma \cup \{\text{EOS}\}} p(w) = 1 \tag{11}$$

$$\sum_{w \in \Sigma} p(w) < 1 \tag{12}$$

$$\sum_{w \in \Sigma} p(\text{EOS}) p(w) = p(\text{EOS}) \sum_{w \in \Sigma} p(w) \tag{13}$$

$$\sum_{w \in \Sigma^*, |w|=0} p(w) = p(\text{EOS}) \qquad \text{because we have only one EOS symbol} \tag{14}$$

$$\tag{15}$$

We define l as:

$$l = \sum_{w \in \Sigma} p(w) \tag{16}$$

$$l < 1 \tag{17}$$

$$\sum_{w\in\Sigma^*} p(w) = \sum_{w\in\Sigma^*,|w|=0} \tilde{p}(w) + \sum_{w\in\Sigma^*,|w|=1} \tilde{p}(w) + \sum_{w\in\Sigma^*,|w|=2} \tilde{p}(w) + \sum_{w\in\Sigma^*,|w|=3} \tilde{p}(w) + \dots \tag{18}$$

$$= p(\text{EOS}) + \sum_{w\in\Sigma^*,|w|=1} \tilde{p}(w) + \sum_{w\in\Sigma^*,|w|=2} \tilde{p}(w) + \sum_{w\in\Sigma^*,|w|=3} \tilde{p}(w) + \dots \tag{19}$$

$$= p(\text{EOS}) \left(1 + \sum_{w\in\Sigma^*,|w|=1} \tilde{p}(w) + \sum_{w\in\Sigma^*,|w|=2} \tilde{p}(w) + \sum_{w\in\Sigma^*,|w|=3} \tilde{p}(w) + \dots\right) \tag{20}$$

$$= p(\text{EOS}) \left(1 + \sum_{w\in\Sigma^*,|w|=1} \tilde{p}(w) + \sum_{w\in\Sigma^*,|w|=2} \tilde{p}(w) + \sum_{w\in\Sigma^*,|w|=3} \tilde{p}(w) + \dots\right) \tag{21}$$

$$= p(\text{EOS}) \left(1 + \sum_{w\in\Sigma} p(w) + \sum_{w\in\Sigma^*,|w|=2} \tilde{p}(w) + \sum_{w\in\Sigma^*,|w|=3} \tilde{p}(w) + \dots\right) \tag{22}$$

$$= p(\text{EOS}) \left(1 + l + \sum_{w\in\Sigma^*,|w|=2} \tilde{p}(w) + \sum_{w\in\Sigma^*,|w|=3} \tilde{p}(w) + \dots\right) \tag{23}$$

$$= p(\text{EOS}) \left(1 + l + \sum_{w\in\Sigma} \sum_{w'\in\Sigma} p(w)p(w') + \sum_{w\in\Sigma^*,|w|=3} \tilde{p}(w) + \dots\right) \tag{24}$$

$$= p(\text{EOS}) \left(1 + l + \sum_{w\in\Sigma} p(w) \sum_{w'\in\Sigma} p(w') + \sum_{w\in\Sigma^*,|w|=3} \tilde{p}(w) + \dots\right) \tag{25}$$

$$= p(\text{EOS}) \left(1 + l + l^2 + \sum_{w\in\Sigma^*,|w|=3} \tilde{p}(w) + \dots\right) \tag{26}$$

$$= p(\text{EOS}) \left(1 + l + l^2 + l^3 + \dots\right) \tag{27}$$

$$= p(\text{EOS}) \left(\sum_{n=0}^{\infty} l^n\right) \tag{28}$$

$$= p(\text{EOS}) \frac{1}{1-l} \qquad \text{(limit geom. series, because } l < 1\text{)} \tag{29}$$

$$= \frac{p(\text{EOS})}{1-(1-p(\text{EOS}))} \tag{30}$$

$$= \frac{p(\text{EOS})}{p(\text{EOS})} \tag{31}$$

$$= 1 \tag{32}$$

c)

$$\sum_{u\in\Sigma^*} p(wu) = \sum_{u\in\Sigma^*} p(\text{EOS}|wu)p_{pre}(u|w)p_{pre}(w) \qquad \text{by def.} \tag{33}$$

$$= p_{pre}(w)\left(\sum_{u\in\Sigma^*} p(\text{EOS}|wu)p_{pre}(u|w)\right) \tag{34}$$

$$= p_{pre}(w)\left(\sum_{u\in\Sigma^*} p(u|w)\right) \qquad \text{def. p(w)} \tag{35}$$

$$= p_{pre}(w)\left(\sum_{u\in\Sigma^*} \frac{p(w|u)p(u)}{p(w)}\right) \qquad \text{Bayes rule} \tag{36}$$

$$= p_{pre}(w)\left(\frac{p(w)}{p(w)}\right) \qquad \text{Bayes rule} \tag{37}$$

$$= p_{pre}(w) \tag{38}$$

d) We use CKY with the $(+,\times)$-semiring. Further we use $\log p$ insted of p for our score function. These operations will lead to the desired $p(w)$ being calculated at the top of the tree.

e)

$$p(S \overset{*}{\Rightarrow} wv) = \sum_{t\in\mathcal{T}_x(wv)} p(t) \tag{39}$$

$$= \sum_{t\in\mathcal{T}_x(w_0,\ldots,w_k)} p(t) + \sum_{t\in\mathcal{T}_x(w_{k+1},\ldots,v_0,\ldots,v_m)} p(t) + \sum_{t\in\mathcal{T}_x(v_{m+1},\ldots,v_l)} p(t) \tag{40}$$

$$\tag{41}$$

?????

f) Following the hint given in the exercise, we create a matrix $M \in |\mathcal{N}| \times |\mathcal{N}|$ where each entry $M_{Y,X}$ corresponds to the probability of deriving Y from X:

$$M_{Y,X} = p(X \to Y\alpha) \tag{42}$$

We can see that by multiplying in the inside semiring this matrix with itself we get the probability of deriving i from j in two steps and so on. Thus, the kleene star over M will yield the desired property. To derive the kleene star:

$$M^* = \bigotimes_{k=0}^{\infty} \tag{43}$$

$$= I + M\bigotimes_{k=0}^{\infty} \tag{44}$$

$$= I + MM* \tag{45}$$

$$M * -MM* = I \tag{46}$$

$$\rightleftarrows (I - M)M^* \tag{47}$$

$$\rightleftarrows M^* = (I - M)^{-1} \tag{48}$$

Using eq. (48) we can calculate M^* in $\mathcal{O}(|\mathcal{N}|^3)$. The entries in the matrix now correspond to:

$$M_{Y,X} = p_{\text{lc}}(Y|X) \tag{49}$$

From this matrix we further need to derive the $p_{\text{lc}(YZ|X)}$. To get these entries we look at the left side of Figure 1 in the exercise sheet.By taking the sum of of a row X we get the probability of deriving X from any string. Now we iterate over all possible choices for X, X', Y, Z to derive:

$$p_{\text{lc}}(YZ|X) = \sum_{X' \in \Sigma} p_{\text{lc}}(X'|X) p_{\text{lc}}(Y|X') p_{\text{lc}}(Z|X'Y) \tag{50}$$

Which can be done in $\mathcal{O}(|\mathcal{N}|^4)$.

g)

$$p_{pre}(w_i \dots w_k|X) = \sum_{j=1}^{k-1} \sum_{Y,Z \in \mathcal{N}} p(X \overset{*}{\Rightarrow} YZ\alpha) p(Y \overset{*}{\Rightarrow} w_i \dots w_j) p(Z \overset{*}{\Rightarrow} w_{j+1} \dots w_k) \tag{51}$$

$$p(X \overset{*}{\Rightarrow} YZ\alpha) = \text{probability of deriving the two subtrees } Y \text{ and } Z \text{ from } X \tag{52}$$

$$p(Y \overset{*}{\Rightarrow} w_i \dots w_j) = \text{probability of deriving the string } w_i \dots w_j \text{ from } Y \tag{53}$$

$$p(Z \overset{*}{\Rightarrow} w_{j+1} \dots w_k) = \text{probability of deriving the string } w_{j+1} \dots w_k \text{ from } Z \tag{54}$$

$$\sum_{j=1}^{k-1} \sum_{Y,Z \in \mathcal{N}} p(X \overset{*}{\Rightarrow} YZ\alpha) p(Y \overset{*}{\Rightarrow} w_i \dots w_j) p(Z \overset{*}{\Rightarrow} w_{j+1} \dots w_k) \tag{55}$$

$$= \sum_{j=1}^{k-1} \sum_{Y,Z \in \mathcal{N}} p_{\text{lc}}(YZ|X) p_{\text{inside}}(w_i \dots w_j|Y) p_{\text{pre}}(w_{j+1} \dots w_k|Z) \tag{56}$$

$$\tag{57}$$

h) Clarify: We can see that if we have all precomputation done, then the calculation of $p_{pre}(w')$ can be done in $\mathcal{O}(1)$ for a given subsequence w' of w. As $|w| = N$, this has runtime $\mathcal{O}(N)$. Now for the precompution, we first look at $p_{\text{lc}}(YZ|X)$. As shown in exercise f) this can be done in $\mathcal{O}(|\mathcal{N}|^4)$. Equation (15) on the exercise sheet shows that $p_{pre}(w_i \dots w_k|X)$ depends only on $p_{\text{lc}}(YZ|X)$, $p_{inside}(w_i \dots w_j)$ and $p_{\text{pre}(w_{j+1} \dots w_k)}$ for $j < k$. The first of the three we have already calculated. We calculate p_{inside} with the CKY algorithm and save all intermediate values. The size of our grammar is $\mathcal{O}(|\mathcal{N}|^2)$. Thus, CKY runs in $\mathcal{O}(N^3|\mathcal{N}|^2)$ and we need to run it $\mathcal{O}(|\mathcal{N}|)$ times for each $X \in \Sigma$. After these steps we have all our value precomputed in $\mathcal{O}(N^3|\mathcal{N}|^3 + |\mathcal{N}|^4)$. We then invoke the algorithm naively for each subsequence of $w_1 \dots w_i$. This gives us a runtime of $\mathcal{O}(N^4|\mathcal{N}|^3 + N|\mathcal{N}|^4)$.

$$s \tag{58}$$