Einführung in die Computerlinguistik
Hausaufgabe 5, Abgabe 22.05.2018

Laura Kallmeyer

SoSe 2018, Heinrich-Heine-Universität Düsseldorf

Aufgabe 1 *Consider a toy example where the sentences are built over $\{a, b, c\}$ and our training data consists of the following sentences:*

`<s>` a b c a b c `</s>`
`<s>` a a b b a a b b a a `</s>`
`<s>` b b a a b b a a `</s>`
`<s>` a b c a b a `</s>`

The bigram probabilities we obtain are the following:

$P(a\|\texttt{<s>}) = \frac{3}{4}$	$P(b\|\texttt{<s>}) = \frac{1}{4}$	
$P(a\|a) = \frac{5}{15} = \frac{1}{3}$	$P(b\|a) = \frac{7}{15}$	$P(\texttt{</s>}\|a) = \frac{3}{15} = \frac{1}{5}$
$P(a\|b) = \frac{5}{12}$	$P(b\|b) = \frac{4}{12} = \frac{1}{3}$	$P(c\|b) = \frac{3}{12} = \frac{1}{4}$
$P(a\|c) = \frac{2}{3}$	$P(\texttt{</s>}\|c) = \frac{1}{3}$	

(all other values are 0)

1. *Compute the probabilities and the perplexities of the following sequencies:*

   (1) `<s>` a b c

   (2) `<s>` a b c a a b b

2. *Explain why the perplexity value is better for measuring the quality of a sentence.*

Solution:

1. (1): probability $\frac{3}{4} \cdot \frac{7}{15} \cdot \frac{1}{4} = \frac{21}{240} = \frac{7}{80} = 0.0875$
   perplexity $\sqrt[3]{\frac{80}{7}} \approx 2.252$

   (2): probability $\frac{3}{4} \cdot \frac{7}{15} \cdot \frac{1}{4} \cdot \frac{2}{3} \cdot \frac{1}{3} \cdot \frac{7}{15} \cdot \frac{1}{3} = \frac{3 \cdot 7 \cdot 2 \cdot 7}{4 \cdot 15 \cdot 4 \cdot 3 \cdot 3 \cdot 15 \cdot 3} = \frac{49}{16200} \approx 0.003$
   perplexity $\sqrt[7]{\frac{16200}{49}} \approx 2.29$

2. Due to the shortcomings mentioned in the last homework, the probability of a sequence is not a good candidate for measuring its quality. A problem is in particular that shorter sequences are preferred. This is why (1) has a much higher probability than (2).

   The perplexity value does not have this problem since for a sentence of length n, we have an exponent $\frac{1}{n}$, which balances the multiplication of n probabilities ($\sqrt[n]{x^n} = x$ for all n). (1) and (2) therefore show only a small difference in perplexity while having a large difference in probability.

Aufgabe 2 *Take again the same training data. This time, we use a bigram LM with Laplace smoothing.*

1. *Give the following bigram probabilities estimated by this model:*

   $P(a|c)$ $P(c|a)$ $P(c|\texttt{<s>})$ $P(b|b)$ $P(b|\texttt{<s>})$

   Note that for each word w_{n-1} and for the symbol `<s>`*, we count an additional bigram for each possible continuation w_n. Consequently, as possible continuations, we have to take the words into consideration and also the symbol* `</s>`.

2. *Calculate the perplexities of the following sequences according to this model:*

(3) <s> $c\ a\ c\ a$

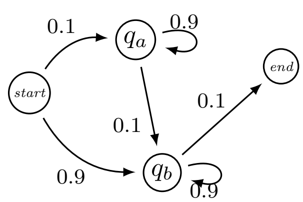
(4) <s> $b\ b\ b\ b$

Which of the two sequences is better according to our LM and the perplexity values?

Solution:

1. If we include </s> (this can also appear as second element of a bigram), we get $|V| = 4$ for our vocabulary ($V = \{a, b, c, \texttt{</s>}\}$).
   $P(a|c) = \frac{2+1}{3+4} = \frac{3}{7}$ $\quad P(c|a) = \frac{0+1}{15+4} = \frac{1}{19}$ $\quad P(c|\texttt{<s>}) = \frac{0+1}{4+4} = \frac{1}{8}$
   $P(b|b) = \frac{4+1}{12+4} = \frac{5}{16}$ $\quad P(b|\texttt{<s>}) = \frac{1+1}{4+4} = \frac{2}{8} = \frac{1}{4}$
2. (3): $\sqrt[4]{8 \cdot \frac{7}{3} \cdot 19 \cdot \frac{7}{3}} = \sqrt[4]{\frac{8 \cdot 7 \cdot 19 \cdot 7}{3 \cdot 3}} \approx 5.36$
   (4): $\sqrt[4]{4 \cdot \frac{16}{5} \cdot \frac{16}{5} \cdot \frac{16}{5}} = \frac{8\sqrt{2}}{\sqrt[4]{125}} \approx 3.38$
   The first sequence has a higher perplexity, consequently the second one is better according to our LM.

Aufgabe 3 *Consider the following HMM, taken from the course slides:*



in q_a, a is emitted with probability 0.9
in q_a, b is emitted with probability 0.1
in q_b, b is emitted with probability 0.9
in q_b, a is emitted with probability 0.1

1. *What is the probability of traversing a path*

   $$start\ q_a\ q_b\ end$$

   according to this HMM?
2. *What is the most probable output when traversing*

   $$start\ q_b\ q_b\ end$$

   according to this HMM?
3. *What is the probability of an output ba given this model?*
4. *Given that we see an output ab, what is the most probable path that has been traversed?*

Solution:

1. This is independent from the output, only transition probabilities matter: $0.1 \cdot 0.1 \cdot 0.1 = 10^{-3}$
2. In q_a, a has a higher probability to be output and in q_b, b has a higher probability.
   Most probable output for *start* $q_b q_b$ *end* is *bb*.
3. We have to sum over the joint probabilities of path and output for all possible paths and the output *ba*:
   (path *start* $q_a q_b$ *end*) $\quad 0.1 \cdot 0.1 \cdot 0.1 \cdot 0.1 \cdot 0.1$
   (path *start* $q_b q_b$ *end*) $\quad +0.9 \cdot 0.9 \cdot 0.9 \cdot 0.1 \cdot 0.1$
   $= (1 + 729)10^{-5} = 73 \cdot 10^{-4} = 0.0073$

4. Here we need the maximum over the joint probabilities of path and output for an output ab.

   Possible paths and the probability of having this path with the output ab:

   $start\ q_a q_b\ end$: $0.1 \cdot 0.9 \cdot 0.1 \cdot 0.9 \cdot 0.1$

   $start\ q_b q_b\ end$: $0.9 \cdot 0.1 \cdot 0.9 \cdot 0.9 \cdot 0.1$

   Consequently, the path $start\ q_b q_b\ end$ is more probable.