Tien-Tso Ning

14 November, 2019

Linguistics TP5

1. Language Modeling

1. Formulate the language model problem for the sentence: $\{We/PRP\ want/VB\ to/TO\ get/VB\ feedback/NN\ on/IN\ their/PRP\ knowledge/NN\ ./.\}$

$$\sum_{x \in V} p(x) = 1, \qquad p(x) \geq 0, \forall x \in V$$

2. Decompose the language model for sentence in (2) using the chain rule.

$$P(X_1 \dots X_n) = P(X_1)P(X_2|X_1)P(X_3|X_1^2) \dots P(X_n|X_1^{n-1})$$

$$P(We\ want\ to\ get\ feedback\ on\ their\ knowledge.) =$$

$$P(We)P(want|We)P(to|We\ want)P(get|We\ want\ to)P(feedback|We\ want\ to\ get)$$

$$P(on|We\ want\ to\ get\ feedback)P(their|We\ want\ to\ get\ feedback\ on)$$

$$P(knowledge|We\ want\ to\ get\ feedback\ on\ their)P(.|We\ want\ to\ get\ feedback\ on\ their\ knowledge)$$

3. Decompose the language model for the sentence in (2) using the Markov-Assumption

$$P(X_1 \dots X_n) = \prod_{k=1}^{n} P(x_k|x_{k-1})$$

$$P(We\ want\ to\ get\ feedback\ on\ their\ knowledge.) =$$

$$P(*|We)P(want|We)P(to|want)P(get|to)P(feedback|get)$$

$$P(on|feedback)P(their|on)P(knowledge|their)P(.|knowledge)$$

4. Estimate the probability of the sentence in (2) using the Markov decomposition, maximum likelihood estimates and corpus in (1) for training.

25 total words in (2)

Word, Frequency, Maximum Likelihood Estimate

We 2 0.08

identify 1 0.04

remaining 2 0.08

gaps 2 0.08

in 2 0.08

knowledge 3 0.12

. 2 0.08

want 1 0.04

to 1 0.04

boost 1 0.04

their 2 0.08

level 1 0.04

, 1 0.04

get 1 0.04

feedback 1 0.04

on 1 0.04

the 1 0.04

(reminder that the probability of $P(* \,|We) = 1.0$)

$$P(We\ want\ to\ get\ feedback\ on\ their\ knowledge.) =$$

$$P(x_1^n) \approx \prod_{k=1}^{n} P(x_k|x_{k-1})$$

$$P(x_n|x_{n-1}) = \frac{count(x_{n-1}x_n)}{count(x_{n-1})} =$$

$$\frac{2}{2} * \frac{1}{2} * \frac{1}{1} * \frac{0}{1} * \frac{1}{1} * \frac{1}{1} * \frac{0}{2} * \frac{2}{3} * \frac{2}{2} = 0$$

5. Estimate the probability of the sentence in (2) using the Markov decomposition, maximum likelihood estimate with Jelinek-Mercer smoothing (assume $\lambda = 0.5$) and corpus (1) for training.

In the smoothing, we just back-off to unigrams with 0.5, amending our previous calculation

$$(\frac{2}{2}0.5 + \frac{2}{25}0.5) * (\frac{1}{2}0.5 + \frac{1}{25}0.5) * (\frac{1}{1}0.5 + \frac{1}{25}0.5) * (\frac{0}{1}0.5 + \frac{1}{25}0.5) * (\frac{1}{1}0.5 + \frac{1}{25}0.5) * (\frac{1}{1}0.5 + \frac{1}{25}0.5) * (\frac{0}{2}0.5 + \frac{2}{25}0.5) * (\frac{2}{3}0.5 + \frac{3}{25}0.5) * (\frac{2}{2}0.5 + \frac{2}{25}0.5) = 0.00000665204$$

2. POS Tagging

1. Formulate the POS tagging model for the sentence in (2)

We/PRP want/VB to/TO get/VB feedback/NN on/IN their/PRP knowledge/NN ./.

The problem is expressed as learning a mapping f that maps x to labels $f(x)$,

$$f(x) = \arg max_y(p(y)p(x|y))$$

2. Apply the Hidden Markov Model for the tagging problem.

$$p(x_1, \dots x_n, y_1, \dots, y_n) = \prod_{i=1}^{n+1} q(y_i|y_{i-1}) \prod_{i=1}^{n} e(x_i|y_i)$$

$$p(We, want, to, get, feedback, on, their, knowledge, \quad .) = p(PRP, VB, TO, VB, NN, IN, PRP, NN, .)$$

$$p(PRP|*)p(VB|PRP)p(VB|TO)p(TO|VB)p(VB|TO)p(NN|VB)p(IN|NN)$$
$$p(PRP|IN)p(NN|PRP)p(.|NN)p(We|PRP)p(to|TO)p(get|VB)p(feedback|NN)$$
$$p(on|IN)p(their|PRP)p(knowledge|NN)p(.|.)$$

3. Estimate the tagging probability of the sentence in (2) using Hidden Markov Model, maximum likelihood estimate and the corpus in (1) for training.

Separated calculations of readability:

$$p(We|PRP)p(to|TO)p(get|VB)p(feedback|NN)p(on|IN)p(their|PRP)p(knowledge|NN)p(.|.)$$

$$\frac{2}{4}*\frac{1}{1}*\frac{1}{4}*\frac{1}{4}*\frac{1}{2}*\frac{2}{4}*\frac{3}{4}*\frac{2}{2}$$

$$p(PRP|*)p(VB|PRP)p(TO|VB)p(VB|TO)p(NN|VB)p(IN|NN)p(PRP|IN)p(NN|PRP)p(.|NN)$$

$$\frac{2}{2}*\frac{2}{4}*\frac{2}{2}*\frac{1}{4}*\frac{1}{4}*\frac{2}{4}*\frac{1}{4}*\frac{2}{4}*\frac{2}{2}$$

= 0.00001144409

3. Syntax

1. Define a grammar that generates the trees in (1).

S': S . | S S 0.5, 0.5

S: PRP VP | S , | VP . | VGB PP 0.4, 0.2, 0.2, 0.2

PRP: We | their 0.5, 0.5

VP: VB NP | VB VP | VP NP | TO VB 0.4, 0.2, 0.2, 0.2

VB: identify | boost | get | want 0.25, 0.25, 0.25, 0.25

NP: VBG NP | PRP NP | NN NN | NN PP | DT NP | NNS S | PRP NN | NNS PP 0.125, 0.125, 0.125, 0.125, 0.125, 0.125

VBG: remaining 1.0

NNS: gaps 1.0

PP: IN NN | IN NP 0.33, 0.66

IN: in | on 0.5, 0.5

NN: knowledge | feedback | level 0.6, 0.2, 0.2

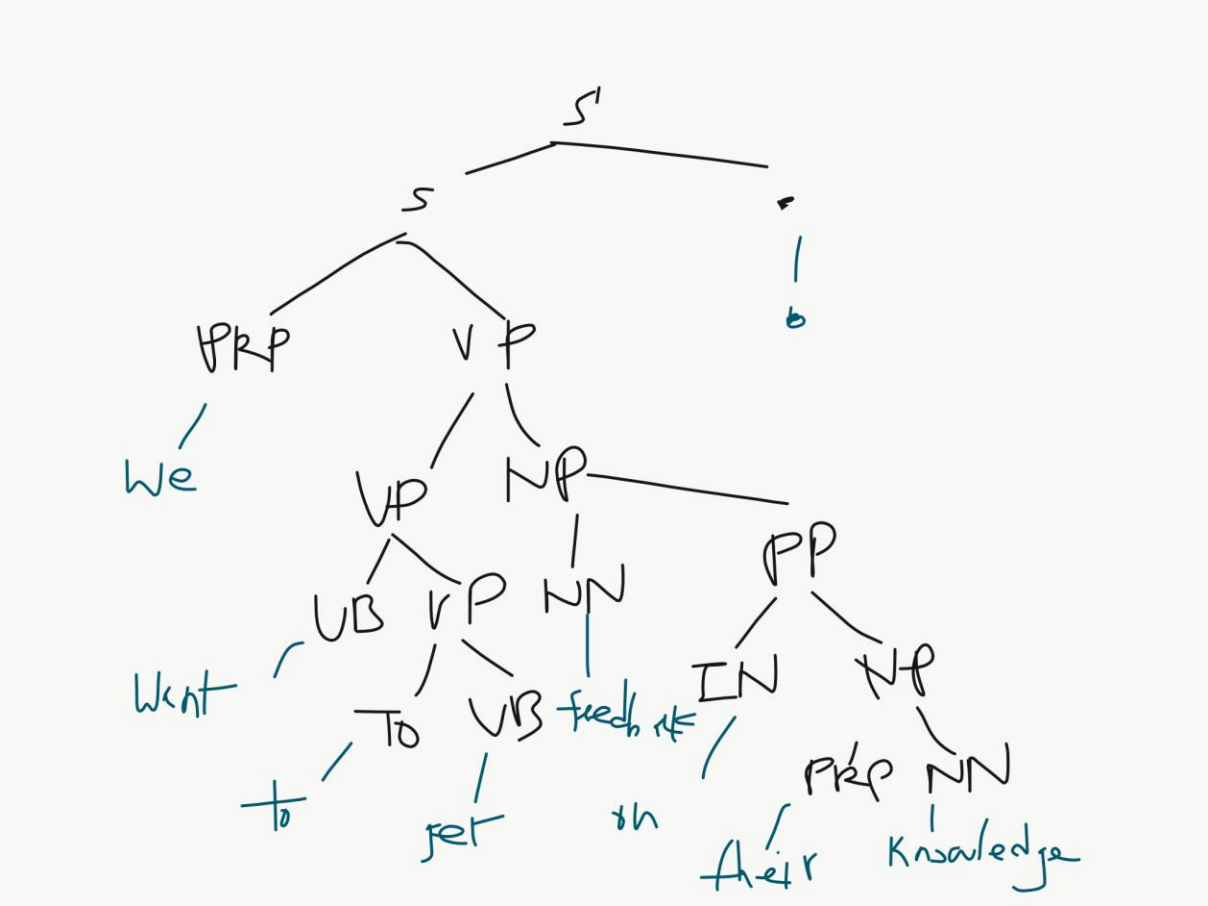
TO: to 1.0

DT: the 1.0

, : , 1.0

. : . 1.0

2. Draw a tree for the sentence (2) using the same grammar in (1).



3. Estimate the probability of the tree in (2) using maximum likelihood estimate and the corpus in (1) for training.

3.09375e-8