Homework 4

Markov Models, POS Tagging, and Grammar

Natural Language Processing

Instructor: Arjun Mukherjee

Q1: This problem is based on the classic weather example due to [Lussier, 1998]. Referring to Section 1, given today is rainy, what is the probability that tomorrow is foggy and the day after is sunny? [4 points]

Q2: This problem is based on the classic weather example due to [Lussier, 1998]. Referring to Section 1, given today is sunny, what is the probability it will be foggy after two days from now. i.e., compute $P(w_3 = F|w_1 = S)$. [6 points]

Q3: This problem is based on the classic weather example due to [Lussier, 1998]. Referring to Section 1, what is the probability it will be rainy on day 2 (tomorrow). We don't know what the weather is like today, i.e., we have no prior for w_1. Assume uniform prior, i.e., $P(w_1 = S) = P(w_1 = R) = P(w_1 = F) = 1/3$. [4 points]

Q4: This problem is based on the classic weather example due to [Lussier, 1998]. Referring to Section 1, what is the probability that tomorrow is foggy and day after tomorrow is rainy. Again, We don't know what the weather is like today, i.e., we have no prior for w_1. Assume uniform prior, i.e., $P(w_1 = S) = P(w_1 = R) = P(w_1 = F) = 1/3$. [6 points]

Q5: Recall that under a standard Hidden Markov Model (HMM) with first order Markov property, latent states $w_1 \dots w_n$ and corresponding observations $u_1 \dots u_n$, we know from Bayes'rule that

$$P(w_1, \dots w_n|u_1, \dots u_n) = \frac{\left(\prod_{i=1}^{n} P(u_i|w_i)\right) \times \left(\prod_{i=1}^{n} P(w_i|w_{i-1})\right)}{\prod_{i=1}^{n} P(u_i)}$$

Now, consider a part-of-speech tagger, which accounts for POS bigrams (i.e., a first order HMM) in estimating the most likely tag prediction for a given observed word sequence. Now the states, $w_1 \dots w_n$ are the tags for the observed sequence of length n, $u_1 \dots u_n$ which are nothing but actual words. Also, the POS tagger only needs to find the relative likelihood of a prediction and hence it only needs to compute:

$$P(w_1, \dots w_n|u_1, \dots u_n) \propto \left(\prod_{i=1}^{n} P(u_i|w_i)\right) \times \left(\prod_{i=1}^{n} P(w_i|w_{i-1})\right)$$

Show the equation that would be used to compute the probability $P(w_1 = INF, w_2 = VB, w_3 = CONJ, w_4 = ADV, w_5 = INF, w_6 = VB \mid u_1 = To, u_2 = be, u_3 = or, u_4 = not, u_5 = to, u_6 = be)$ using a part-of-speech tagging model with POS bigrams. [4 points]

Q6: Extending what we learnt from Q5, now, consider a part-of-speech tagger, which accounts for POS trigrams (i.e., a second order HMM) in estimating the most likely tag prediction for a given observed word sequence. [4+2]

(a) Provide the analytical expression for $P(w_1, \dots w_n|u_1, \dots u_n)$ that the trigram POS tagger would employ in estimating the probabilities of tag sequence given observed words. Consider two placeholders φ_1 and φ_2 which mark the start tag and tag before the start tag respectively., i.e, assume the tag sequence as $\varphi_2, \varphi_1, w_1, \dots w_n$.

(b) Show the equation that would be used to compute the probability $P(w_1 = INF, w_2 = VB, w_3 = CONJ, w_4 = ADV, w_5 = INF, w_6 = VB \mid u_1 = To, u_2 = be, u_3 = or, u_4 = not, u_5 = to, u_6 = be)$ using a trigram POS tagger as in (a).

Q7. Referring to the crazy soft drink machine in section 9.2 of FSNLP, the book provides a solution for the probability of observing the sequence {lem, ice_t}. A detailed and more clear solution appears here too. [4+6 points].

(a) Based on the above problem definition, assuming the machine always starts in the ice-tea preferring state, what is the probability of observing {col, lem}?

(b) Continuing further, what is the probability of observing the sequence {col, ice_t, lem} assuming the machine always starts off in the cola preferring state.

Q8. Referring to the classic weather example in Section 2 due to [Lussier, 1998], suppose the first day you were locked in the room it was sunny, the caretaker did not get an umbrella on the second day, but got one on the third day. Assuming the prior probability of the caretaker bringing an umbrella on any day is 0.5, compute the probability that day 3 is rainy [6 points]

Q9. Which of these sentences is a valid derivation of the following grammar? For each sentence write valid or invalid. If you deem a sentence valid, show the parse tree for it. For sentence you deem invalid, justify your answer. How did you compute your answer? [5 + 5 + 5 = 15 points]

Grammar

S → NP VP
S → Aux NP VP
S → VP
NP → Pronoun
NP → Proper-Noun
NP → Det Nominal
Nominal → Noun
Nominal → Nominal Noun
Nominal → Nominal PP
VP → Verb
VP → Verb NP
VP → VP PP
PP → Prep NP

Lexicon

Det → the | a | that | this
Noun → book | flight | meal | money
Verb → book | include | prefer
Pronoun → I | he | she | me
Proper-Noun → Houston | NWA
Aux → does
Prep → from | to | on | near | through

(a) $I\ prefer\ that\ flight$
(b) $I\ prefer\ to\ fly\ through\ Houston$
(c) $does\ the\ flight\ include\ a\ meal$