CS-431 Hands On Part-of-Speech tagging (part 1)

J.-C. Chappelier   M. Rajman

v. 20211020 – 1

QUESTION I   [1 pt]

(from Fall 2018 quiz 2)

Assume that, in the word sequence "*iron shaped cloth*", the word "*shaped*" is replaced by an Out-of-Vocabulary (OoV) form that your spell checker was not able to correct, nor your morphological analyzer to analyze. Select among the following options <u>the most adequate</u> one to decide which possible PoS tags should be associated with the OoV form:

[ ] All the PoS tags

[ ] All the PoS tags corresponding to open grammatical categories

[ ] The most frequent PoS tag

QUESTION II   [2 pt]

(from Fall 2018 quiz 2)

For this question, *one or more* assertions can be correct. Tick only the correct assertion(s). There will be a penalty for wrong assertions ticked.

When using a probabilistic approach to find the optimal tagging for the sentence "*young birds fly*", what does the conditional probability

$$P(\text{``birds''}, \text{``fly''} \mid t_2\, t_3)$$

represent, provided that no additional specific hypotheses are made:

[ ] the probability that the word *"birds"* appears at position 2 conditioned by the tag t_2 only, and the word *"fly"* appears at position 3 conditioned by tag t_3 only;

[ ] the probability that the word *"birds"* and the tag t_2 appear at position 2, and the word *"fly"* and the tag t_3 appear at position 3;

[ ] the probability that the sequence (*"birds"*, *"fly"*) appears at the end of the sentence, conditioned by the tag pair (t_2, t_3).

QUESTION III [2 pt]

(from Fall 2018 quiz 2)

For this question, *one or more* assertions can be correct. Tick only the correct assertion(s). There will be a penalty for wrong assertions ticked.

Indicate which of the following formulas are strictly equal to the conditional probability

$$P(\text{``young''}, \text{``birds''}, \text{``fly''}|t_1, t_2, t_3)$$

provided that no specific additional hypotheses are made:

[ ] $P(\text{``young''}|t_1) \cdot P(\text{``birds''}|t_2, \text{``young''}, t_1) \cdot P(\text{``fly''}|t_3, \text{``young''}, t_1, \text{``birds''}, t_2)$

[ ] $P(\text{``young''}|t_1, t_2, t_3) \cdot P(\text{``birds''}|\text{``young''}, t_1, t_2, t_3) \cdot P(\text{``fly''}|\text{``young''}, \text{``birds''}, t_1, t_2, t_3)$

[ ] $P(\text{``young''}|t_1) \cdot P(\text{``birds''}|t_2) \cdot P(\text{``fly''}|t_3)$

[ ] $P(\text{``young''}|t_1) \cdot P(\text{``birds''}|t_2) \cdot P(\text{``fly''}|t_3) \cdot P(t_1) \cdot P(t_2|t_1) \cdot P(t_3|t_2)$

QUESTION IV [2 pt]

(from Fall 2018 quiz 2)

For this question, *one or more* assertions can be correct. Tick only the correct assertion(s). There will be a penalty for wrong assertions ticked.

When using an HMM to find the optimal tagging of a 3 word sentence "$w_1\, w_2\, w_3$", what probability should be maximized?

[ ] $P(w_1, w_2, w_3 | t_1, t_2, t_3)$

[ ] $P(t_1, t_2, t_3 | w_1, w_2, w_3)$

[ ] $P(w_1, w_2, w_3 | t_1, t_2, t_3) \cdot P(t_1, t_2, t_3)$

[ ] $P(w_1 | t_1, t_2, t_3) \cdot P(w_2 | w_1, t_1, t_2, t_3) \cdot P(w_3 | w_1, w_2, t_1, t_2, t_3)$

QUESTION V [2 pt]

(from Fall 2018 quiz 2)

① Cross out the elements that can be ignored in the following conditional probability, when tagging the sentence "*young birds fly*" under the "*limited lexical conditioning*" hypothesis:

$$P(\text{``fly''} \mid \text{``young''}, \text{``birds''}, t_1, t_2, t_3)$$

② Cross out the elements that can be ignored in the following conditional probability, when tagging the sentence "*young birds fly fast*" under the "*limited scope for syntactic dependencies (1 neighbor)*" hypothesis:

$$P(t_3 \mid t_1, t_2, t_4)$$

QUESTION VI [4 pt]

(from Fall 2018 quiz 2)

For this question, *one or more* assertions can be correct. Tick only the correct assertion(s). There will be a penalty for wrong assertions ticked.

① When using an HMM to find the optimal tagging of the sentence "*young ducks fly fast*", where all the words are considered as potentially ambiguous, indicate which of the following assertions are true under the "*limited lexical conditioning*" and "*limited scope for syntactic dependencies (1 neighbor)*" hypotheses

[ ] The tagging of "*fast*" depends on the one of "*fly*".

[ ] The tagging of "*fly*" depends on the one of "*fast*".

[ ] The tagging of "*young*" depends on the one of "*fast*".

[ ] The tagging of "*fast*" depends on the one of "*young*".

3

② Same question, but for the sentence "*young **birds** fly fast*", where all the words <u>but</u> "*birds*" are considered as potentially ambiguous:

    [ ] The tagging of "*fast*" depends on the one of "*fly*".

    [ ] The tagging of "*fly*" depends on the one of "*fast*".

    [ ] The tagging of "*young*" depends on the one of "*fast*".

    [ ] The tagging of "*fast*" depends on the one of "*young*".