Sample Quiz Paper: CS626- 20/9/20

Preamble:

Here are 4 questions from 4 topics covered so far: Morphology Part of Speech Tagging, Chunking and Deep Parsing (being done now). The quizes will be of this flavour. All questions are MCQs (multiple choice questions). There is NO negative marking.

The questions can be catergorized into: **R** (recall), **C** (comprehension), **A** (application). In choosing the right answer(s) from the options, you will be doing one of these 3 activities. R is for questions like "what is the complexity of Viterbi Decoding?" (Ans: O(S^2L)), where you have to recall and reproduce. Example of C is "Why is Viterbi's complexity linear in the length of the input?" (Ans: Markov Independence Assumption, whereby only one path from amongst all those ending in the same state remains alive). Example of A is "What will be the best possible tag sequence for the sentence: *The dogs ran quickly*" (ans: POS tags, transition probabilities and lexical probabilities will be given. You will have to draw the Viterbi search tree and do the computations).

Questions:

Q1. The correct morpheme grouping for the word "Unproductivity" is:

(a) [[[un] + [[product]+[ive]]]+[ity]]

(b) [[[un] + [[product]]+[ive]+[ity]]

(c) [[un] + [product]+[ive]+[ity]]

(d) [[un] + [[[product]+[ive]]+[ity]]]

Ans: (a). 'Un' as a prefix applies only to adjectives. So the grouping is unproductive+ity.

Q2. Suppose we see the word "unproductive" in a test sentence, and we have not encountered it in the training corpus. This is an example of data sparsity or unknown word situation. However, the morpohological feature of 'un' as prefix helps decide the POS tag as JJ (adjective). This is an examples of which kind of smoothing?

(a) Good Turing

(b) Kneser-Ney

(c) backoff

(d) Laplace

Ans: (c). We are "backing off" by taking a smaller segment of the word ("un"), and using the learnt behaviur from words like "unsuitable", "unworthy" etc.

Q3. In MEMM, the constraint for the optimization problem is

model expectation of a feature= empirical expectation of the feature.

Which of the following is the ADOPTED expression? Also after making the choice give a one line justification.

(a) $\Sigma_{<h,t>} p(h,t) f_j(h,t) = \Sigma_{<h,t>} p\sim(h,t) f_j(h,t)$
(b) $\Sigma_{<h,t>} p\sim(h) p(t|h) f_j(h,t) = \Sigma_{<h,t>} p\sim(h,t) f_j(h,t)$
(c) $\Sigma_{<h,t>} p(h) p(t|h) f_j(h,t) = \Sigma_{<h,t>} p\sim(h,t) f_j(h,t)$
(d) $\Sigma_{<h,t>} p(h,t) f_j(h,t) = \Sigma_{<h,t>} p\sim(h) p(t) f_j(h,t)$

Ans: (b), because p~ (.) values can be simply obtained from counts in the training data. Model probabilities p(h) or p(h,t) cannot be obtained, because model p(h) is very complex to obtain.

Q4. Take the sentence, "The cameraman shot the batsman when he was near the captain". How many parse trees does this have and how many plausible meanings? You do not have to actually draw the parse trees. Just count the phrases and word meanings, and combinations thereof.

(a) (1, 1)

(b) (2, 4)

(c) (3, 4)

(d) (4, 6)

Ans: (b). The basic parse is NP="The cameraman", VP="shot the batsman". This has two meanings: 'shot' in the sense of taking picture and in the sense of firing with a gun. For both meanings the parse tree will be ONE. Now each of these meanings has two possibilities: 'he' refering to 'cameraman' or 'batsman'. In each case there will be a different parse tree. This two parse trees, but 4 meanings.