NLP Endsem 2016

Time: 3 hrs | Total Marks: 70 | Be Precise

==

Important: Answer all questions from PART A, and any 3 out of 4 questions from Part B.

==

Part A (40 marks)

(Answer all questions from this part)

1. Constructing a bilingual dictionary (say English to Hindi) is a non-trivial task. How can machine learning help? Be specific in identifying the algorithm used and explaining how parameters are estimated. [3]
2. Identify the specific problem that the Viterbi algorithm solves in the context of HMMs. Explain the central intuition behind the algorithm using an example. [3]
3. What is CBOW in the context of Word2Vec? Why is it useful? [2]
4. Which of polysemy or synonymy is LSA better at handling, and why? Given a rectangular matrix of size 2 x 3, how would you compute SVD of this matrix by hand? [2+2]
5. You are given a set of sentences from an unknown language, which has never been studied till date. How can you use Expectation Maximization to arrive at the correct parse of these sentences? [3]
6. Are there situations where the parameters learnt from corpus in PCFG are successful in correctly parsing some sentences, but unsuccessful over others? If yes, explain with an example, and suggest a fix. If no, justify (your justification must be accompanied by a proof sketch). [3]
7. We can use distributional models of similarity (KL divergence or its symmetrised version) to estimate document and term relatedness from corpus. What advantage does a method like Latent Semantic Analysis have over this approach? [2]
8. Bottom up filtering is used to improve the efficiency of top down parsers. Is this true? If yes, how? If no, correct the sentence and justify. [2]
9. Apart from decision trees, identify a rule induction technique that addresses a classical problem in NLP. Discuss briefly the central idea behind the approach. [3]
10. How is the success of HMM parameter learning related to an important property of KL divergence? [2]
11. What limitation of Laplace smoothing does Good Turing smoothing overcome? Where can Good Turing smoothing fail? Suggest a repair to overcome this shortcoming. [2+1.5+1.5]
12. Can you think of any NLP task where knowledge of recall can reduce uncertainty about precision (or vice versa)? Explain. [2]
13. Briefly explain the connection between branching factor and perplexity with an example. [3]
14. What are Hearst patterns and what are they used for? Explain briefly with two examples. Identify a limitation of Hearst patterns. [2+2]
15. There are only two words A and B in a corpus, and two (hidden) topics that can generate these words. We have ten documents, each having 20 tokens of type A and B. Suggest an approach to estimate (a) the probabilities with which each topic generates A and B (b) estimate the belongingness of each document to the two topics. Identify clearly all assumptions you may have made. Are there any criteria that the document collection should ideally satisfy? [6]
16. What are rhetorical relations and in which context are they useful? Explain with two examples. [3]

(Please Turn Over)

Part B (30 marks)
(Answer any three questions from this part)

1. (a) Use dynamic programming to compute the edit distance between words WRONG and WINGS, assuming that the costs of insertion and deletion are 2 and the cost of substitution is 1. [6]

   (b) Assume that there are only two senses of the word *bank* (one pertaining to the financial sense, and the other to the "river bank" sense) in WordNet. In a given piece of raw text, the distributional neighbours of bank are {*account, deposit, river*}. Given that each of these neighbours can have multiple senses as well, how would you go about assigning dominance based ranks to the two senses of *bank*? Show the steps in detail. [4]

2. Consider a Machine Translation parallel corpus having three sentence pairs. The first sentence pair is "go there fast"/"jaldi udhar jaao". The second sentence pair is "go there"/"udhar jaao". The third sentence pair is "go"/"jaao". (a) Show how the first few iterations of EM are useful in learning word alignments from this corpus. Make clear any simplifying assumptions on top of IBM Model 3. (b) How is extra knowledge "getting generated" in successive iterations of EM? [8+2]

3. What limitations of the basic parsing techniques does the CYK parser address? Is there an assumption on the grammar rules that CYK can deal with? If yes, what are these? Given the grammar below and the input sentence "w =(()(()))", show the steps in chart parsing using CYK. Alongside your charts showing each step, mention clearly the rule(s) that is(are) used (if any) to advance to this step from the previous one. [1.5 + 1.5 + 7]

   $S \rightarrow SS$
   $S \rightarrow (S_1$
   $S_1 \rightarrow S)$
   $S \rightarrow ()$

4. A PCFG is based on the following rules:
   a. S → A B
   b. B → D A
   c. B → D A C
   d. A → A C
   e. A → a
   f. A → b c
   g. A → b d e
   h. C → f g h
   i. D → i

   The corpus has the following two sentences, the first occurring 15 times and the second 30 times:

   1. a i b c f g h
   2. b c i b d e

   (a) Are the sentences accepted by the grammar? In case both of them are, which of these two sentences is/are ambiguous? Show all possible parse trees of the sentence(s).
   (b) Make an APPROPRIATE initial choice of the rule probabilities. Show the first three steps of the EM algorithm for estimating the parameters of this PCFG. [3+7]

== The End ==