CS 6370 (NLP) – Endsem – Nov 21, 2012

Time: 3 hours

(Be concise. Marks may be deducted for answers that are unnecessarily verbose.)

1. Explain the central idea behind dominance based Word Sense Disambiguation using a concrete example. [2]

2. What is Jensen's inequality? Using Jensen's inequality, show that K-L divergence is always non-negative. [1+2]

3. Provide a Bayesian setting for Statistical Machine Translation by giving an interpretation for the likelihood and prior terms. How are these terms estimated in practice? [2+2]

4. Identify one central limitation of the EM algorithm, and one way to address this limitation. [1]

5. What is the idea behind transfer-based Machine Translation? Give an example. [2]

6. Context sensitive spelling correction can be viewed as a classification problem. Explain briefly one technique for performing feature selection in this context. [2]

7. Propose one path based measure of Wordnet based similarity. Show why it works and the properties it satisfies. How can Information Theoretic approaches be used to better path based approaches? [3]

8. Explain precisely the most important reason why (a) parameter estimation in PCFGs uses EM instead of Maximum Likelihood estimation using direct counts from corpus (b) finite state transducers are used for lemmatization when dictionaries carrying derivation/inflection information are available (c) empirical NLP has dominated over rationalistic (classical) NLP over the last two decades (d) interlingua is an attractive option when machine translators have to be built between several pairs of languages. [2]

9. List properties of hanger, stretcher and aligner matrix. What is the interpretation of these matrices in the context of LSI? Explain clearly the geometry of SVD using these matrices. [5]

10. Define rank of a matrix. In LSI, how does rank reduction correspond to concept extraction? Explain using two limiting cases, one of a full rank matrix and another of a maximally rank deficient one. [2]

11. Draw a picture to illustrate a situation where LSA can fail to extract concepts, but PLSA may succeed. Justify in a single sentence your answer. [1]

12. Discuss very briefly a bootstrapping approach to Word Sense Disambiguation. [2]

13. In the context of PLSA, identify the parameters that need to be estimated, the objective function and the constraints. Instead of using a conventional optimization technique like hill climbing, why is EM used? [3]

14. Explain using an example, the following ideas: (a) Lexical Chains (b) Explicit Semantic Analysis (c) smoothing in the context of Language Models (d) inferencing step of Information Extraction (e) HMMs for sequence modeling (f) dynamic programming in parsing. [6]

15. Consider a Machine Translation parallel corpus having three sentence pairs. The first sentence pair is "come here fast"/"jaldi idhar aao". The second sentence pair is "come here"/"idhar aao". The third sentence pair is "come"/"aao". (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? [5+1]

16. 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. [4]

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

17. A PCFG is based on the following rules:
    a. $S \rightarrow A\ B$
    b. $B \rightarrow D\ A$
    c. $B \rightarrow D\ A\ C$
    d. $A \rightarrow A\ C$
    e. $A \rightarrow a$
    f. $A \rightarrow b\ c$
    g. $A \rightarrow b\ d\ e$
    h. $C \rightarrow f\ g\ h$
    i. $D \rightarrow i$

    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. [5]

END