Speech and Natural Language Processing **CS60057**
Autumn Semester, 2015 **End-Sem**
Maximum Marks: 90 **Time Limit: 3 Hours**

This exam contains 4 pages (including this cover page) and 12 problems.

You may *not* use your books and notes for this exam. Be *precise* in your answers. All the *sub-parts* of a problem should be answered at *one place* only. On multiple attempts, *cross* any attempt that you do not want to be graded for.

There are no clarifications. In case of doubt, you can take a valid assumption, state that properly and continue.

1. (7 points) Suppose that you are given a PCFG grammar G. Consider the sentence, "The astronomers saw stars with ears". Derive the expressions for the following in terms of inside outside probability
   (a) Probability of this sentence and that there is a consistent spanning from word 3 to word 4 (*saw stars*) given the grammar G. (3 points)
   (b) You are in the learning phase of this grammar and sentence W occurs in training, derive the expression for the expected number of times, rule $A \rightarrow w$ is used in this sentence. (4 points)

2. (6 points) Use the following text to construct distributional thesaurus for *eyes*, *here* and *valley*. Use a 5-word window including open- and closed-class words, ignore case and line-breaks and weight contexts by frequency.

   The eyes are not here
   There are no eyes here
   In this valley of dying stars
   In this hollow valley

   Compute the following
   1. Dice coefficient between (eyes, valley)
   2. Overlap coefficient between (here, valley)

3. (6 points) Consider the following weighted tuple structure as per the Distributional Memory model. What are the semantic spaces that can be constructed from this structure? Give examples of atleast 2 such spaces along with the applications where these can be utilized.

marine	own	book	3.2	teacher	use	bomb	7.0
marine	use	book	3.3	teacher	own	gun	9.3
sergeant	own	bomb	16.7	teacher	use	gun	4.7
sergeant	use	bomb	69.5	teacher	own	book	48.4

4. (10 points) Consider the following sentences, along with the particular class they belong to
   - language independent system data driven dependency parsing – *dependency parsing*

- algorithm deterministic incremental dependency parsing – *dependency parsing*
- transition based techniques projective dependency parsing – *dependency parsing*
- Structured models sentiment analysis – *sentiment analysis*

(a) Build a text classifier using Naïve Bayes algorithm as discussed in the class and using this classifier, classify the following sentence, "dependency sentiment analysis data" (5 points)

(b) Suppose you run your classifier for 10 documents and the table below gives the actual and predicted classes by the classifier. Construct the confusion matrix and compute the micro-averaged and macro-averaged precision values for your classifier. The two classes are denoted as "DP" and "SA" respectively. (5 points)

Doc.	Actual	Predicted	Doc.	Actual	Predicted
1	DP	SA	2	SA	SA
3	DP	DP	4	DP	SA
5	SA	DP	6	SA	SA
7	DP	DP	8	DP	DP
9	SA	SA	10	SA	DP

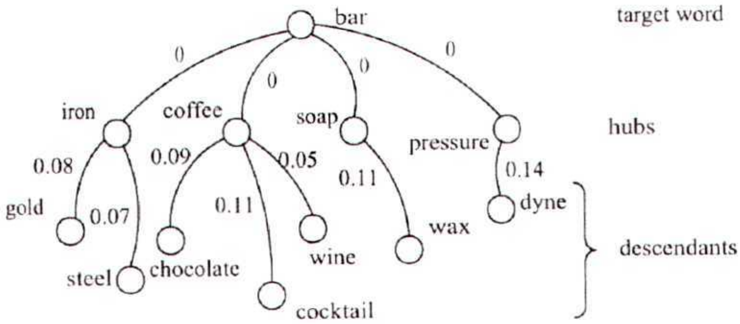
5. (10 points) Suppose you are training MST Parser for dependency and the sentence, "John saw Mary" occurs in the training set. Also, for simplicity, assume that there is only one dependency relation, "rel". Thus, for every arc from word w_i to w_j, your features may be simplified to depend only on words w_i and w_j and not on the relation label.

   Below is the set of features

   - f_1: $\text{pos}(w_i) = \text{Noun}$ and $\text{pos}(w_j) = \text{Noun}$
   - f_2: $\text{pos}(w_i) = \text{Verb}$ and $\text{pos}(w_j) = \text{Noun}$
   - f_3: $w_i = \text{Root}$ and $\text{pos}(w_j) = \text{Verb}$
   - f_4: $w_i = \text{Root}$ and $\text{pos}(w_j) = \text{Noun}$
   - f_5: $w_i = \text{Root}$ and w_j occurs at the end of sentence
   - f_6: w_i occurs before w_j in the sentence
   - f_7: $\text{pos}(w_i) = \text{Noun}$ and $\text{pos}(w_j) = \text{Verb}$

   The feature weights before the start of the iteration are: $\{3, 20, 15, 12, 1, 10, 20\}$. Determine the weights after an iteration over this example.

6. (6 points) Given a new sentence, "chocolate bar wine blood pressure", use HyperLex algorithm to disambiguate the sense of the word 'bar'. The MST obtained from the corpus data is shown below.



7. (6 points) Suppose that you want to use a decision list classifier for Word Sense disambiguation. For the word plant with the senses, {living thing, building}, various collocations as well as the number of times they occur with these senses, are given in the following table. Construct a decision list classifier to disambiguate the sense of 'plant' in a new sentence.

Collocation	living thing	building
plant growth	8	1
equipment (within $\pm$ 2 words)	2	7
fruit (within $\pm$ 2 words	6	1
assembly plant	2	9

8. (7 points) Suppose that you have to extract noun pairs with relation, "film - director".
   (a) Explain how would you use bootstrapping approach for relation extraction. (4 points)
   (b) How would distant supervision approach for this problem would be different from the bootstrapping approach? Explain. (3 points)

9. (6 points) While processing a corpus, you encounter a word, 'insention'. You recognize that this is a spelling error and also that the possible candidate words are 'insertion, inspection, invention, indention, intention'. Assuming that you are using a noisy channel model for spell correction, write down the expression for the correct word in terms of the probabilities to be estimated from a training corpus.

10. (6 points) What is the maximum matching algorithm for Word Segmentation? Use the example 'amiabletogether' to explain the algorithm. What are the main problems with this approach and how do you suggest to modify the basic approach to handle these problems?

11. (15 points) Often during code-switching, two languages can get mixed at the morpheme level. For example,

Original sentence:	***mere***	***friendson****	***me***	***sirf***	***wohi***	***chaluish****	***sa***	***behave***	***karta***	***hai***
Language label:	Hi	En-Hi	Hi	Hi	Hi	Hi-En	Hi	En	Hi	Hi

Translation: Among my friends s/he is the only one who behaves cunningly.

The * marked words here have morpheme level mixing, where the Hindi plural marker *-on* is added to the English word friends, and in the second case the English suffix *-ish* is added to the Hindi word *chalu* (cunning). Your task is to design a system that can do accurate language labelling in presence of morpheme level mixing in written code-switched data. Assume that only two languages L1 and L2 are present in the data. Let's say we decided to build a two stage language labeller:

Stage 1: Word-classifier that takes a word w_i as input and outputs the possible language(s) it can belong to (with some score or probability, if that you think is useful)

Stage 2: Given the output of the classifier for each w_i in the input text $w_1\ w_2 \ldots w_k$, it chooses the final level in the context of the text.

How would you design the word classifier? In particular,

(a) What would be the classes that the classifier outputs? (1 point)

(b) What features will you use to train the classifier? (2 points)

(c) What training data such a classifier would expect? (2 points)

(d) Would you do something special for the words that have morpheme-level mixing? (2 points)

Now for Stage 2,

(e) Is it really necessary? Can you not just rely on the output of stage 1? If not, explain with example why not. (1 point)

(f) What would be the training data for stage 2? (2 points)

(g) What kind of learning problem stage 2 presents? (1 point)

(h) Give examples of two machine learning algorithms that are useful for solving such kind of learning problems. (2 points)

(i) Linguists argue that the use of the word "friendson" is somewhat weird and more natural usages are:
(a) *mere friends me sirf woh ...* or
(b) *mere friendon me sirf woh ...*
What do you think could be the rationale behind such a statement? (1 point)

(j) Also, what can you infer about the status of the word friend in the vocabulary of somebody who uses the variety (a) as opposed to somebody who uses variety (b)? Hint: which user is expected to be more conscious that s/he is mixing an English word in a Hindi sentence? (1 point)

12. (5 points) Match the following columns. *Your answer should only be in terms of the indices used in both the columns, e.g., 1-C etc.*

1. Parts-of-speech tagging 2. Sentiment Detection 3. Transliteration 4. Named entity recognition 5. Language labelling	A. Was thought to be a solved problem earlier but is still unsolved with for social media text. B. Is a sequence transformation problem that can be modelled using the noisy channel approach. C. Is typically modelled as a three-class classification problem. D. Was the first task in NLP that was solved using a statistical learning approach. E. Is typically modelled as a sequence modelling problem with 2, 3, 4 labels.