

CS4740/5740
Introduction to Natural Language Processing

Final Exam
May 13, 2014

Name:

Netid:

The Dec 9, 2015 final is CLOSED BOOK, CLOSED NOTES.

Instructions: You have 2 hours and 30 minutes to complete this exam. ~~The exam is open book, open notes. You can use your laptops for calculations and for accessing notes, book, etc. BUT ABSOLUTELY NO INTERNET ACCESS OR COMMUNICATION WITH OTHERS DURING THE EXAM. Either of the latter will result in an immediate failure in the course.~~

| # | description | score | max score |
|---|---------------------------------|-------|------------|
| | Parsing | _____ | 30 |
| | HMMs and Part-of-speech Tagging | _____ | 30 |
| | Information Extraction | _____ | 22 |
| | Grab Bag | _____ | 18 |
| | Total | _____ | out of 100 |

1 Parsing (30 points)

1. Bottom-up Chart Parsing (8 pts)

Given the grammar and lexicon below, show the **final chart** for the following phrase after applying the bottom-up chart parser from class:

The final chart

Remember that the final chart contains all edges added during the parsing process. You may use either the notation from class or the notation from the book to depict the chart.

$S \rightarrow \text{adj adj n}$

$S \rightarrow \text{det adj n}$

$\text{det} \rightarrow \text{the}$

$\text{n} \rightarrow \text{final} \mid \text{chart}$

$\text{adj} \rightarrow \text{final}$

2. Earley Algorithm (6 pts)

Using the same grammar and lexicon as the previous question (provided again here for your convenience), show the chart produced by the Earley algorithm for the sentence:

The final chart

$S \rightarrow \text{adj adj n}$

$S \rightarrow \text{det adj n}$

$\text{det} \rightarrow \text{the}$

$\text{n} \rightarrow \text{final} \mid \text{chart}$

$\text{adj} \rightarrow \text{final}$

3. (3 pts) What is the primary advantage of the Earley algorithm over the bottom-up chart parser?

4. CFGs and PCFGs. **Consider the following CFG for problems (a)–(c).** (No lexicon is provided (or needed).)

| production rule |
|-----------------------------|
| $S \rightarrow NP VP$ |
| $VP \rightarrow Verb NP$ |
| $VP \rightarrow Verb NP PP$ |
| $NP \rightarrow NP PP$ |
| $NP \rightarrow Det Noun$ |
| $PP \rightarrow Prep Noun$ |

- (a) (3 pts) Modify the CFG to show one possible probabilistic version of the it (i.e. make it into a PCFG).
- (b) (4 pts) One common problem with CFGs and PCFGs is their inability to properly perform *prepositional phrase attachment*. The (P)CFG above, for example, will make the same attachment decision for “with the binoculars” for each of the following sentences:

The officer saw the man with the binoculars.
The officer called the man with the binoculars.

Why is this not the correct behavior for the parser?

- (c) (6 pts) Describe how you would lexicalize the given PCFG in order to address this problem (from (b)). Then show specifically how the grammar rules below should be modified according to your lexicalization scheme.

production rule

$VP \rightarrow \text{Verb NP}$

$VP \rightarrow \text{Verb NP PP}$

2 HMMs (30 points)

1. Mary/N had/V a/DET little/ADJ lamb/N ,/, little/ADJ lamb/N ,/, little/ADJ lamb/N ./ . Mary/N had/V a/DET little/ADJ lamb/N ./ . The/DET fleece/N was/V snow/N white/N ./ .

Assume that the above text is provided as the (entire) training corpus for a Hidden Markov Model (HMM) part-of-speech tagger. Further, assume that the tagger does **no stemming** and is a **bigram**-based tagger.

- (a) (6 pts) Using Maximum Likelihood Estimation (MLE) and the above training data, show **the fractions** that correspond to the bigram probability estimates for **all bigrams involving nouns (in either position in the bigram)**. That is, show the numerator and denominator of each fraction; i.e. you do **not** need to show the decimal equivalent. (Nouns are indicated via the N tag.)

- (b) (6 pts) Using Maximum Likelihood Estimation and the above training data, show the fractions that correspond to the **lexical-generation probability estimates** (i.e., the emission probabilities) for the **noun** state.

2. (3 pts) Part-of-speech tagging is considered an “easy” task — what is a reasonable simple baseline to compare an HMM approach to?

3. (5 pts) If a bigram HMM pos-tagger, using Viterbi for inference (i.e. to determine the correct tag sequence during testing), takes 10 seconds to predict tags for a set of documents, about how long would a similarly-coded 4-gram HMM tagger take? Assume there are 40 tags, and no optimizations (such as beam search) are done.

4. (10 pts) Suppose you are doing Viterbi inference with a bigram HMM pos-tagger on this sentence: *All cows eat grass*

Suppose the algorithm had already progressed through the first two words (i.e. the program had “looked at” ‘All’ and ‘cows’). Assume there are three parts of speech (N(oun), V(erb), D(eterminer)) — what calculations will occur for the next word? For your answer, please use the following notation:

the words of the sentence are w_1, w_2, w_3, w_4 ; corresponding tags are t_1, t_2, t_3, t_4 , the scores are $s_{something}$ and the backpointers (or chains) are $b_{something}$.

Just as a hint, you should end up with 3 scores and 3 backpointers (or chains) for the word ‘eat’.

3 Information Extraction (22 pts)

For the questions below, consider the following paragraph in the context of an information extraction system for identifying mergers and acquisitions:

AT&T is in talks to buy DirecTV for at least \$50 billion, and the two sides are actively working toward an announcement. If completed, a deal would give AT&T, the countrys second-largest wireless carrier, control of the countrys largest satellite television provider. AT&T has grown interested in DirecTV in recent months because of its 20 million U.S. subscribers.

1. (3 pts) **Underline** each named entity in the text and put a label above each, indicating its named entity type.
2. (3 pts) Put a **square box** around each element in the coreference chain for ATT. Put a **circle** around each element in the coreference chain for DirecTV.
3. (8 pts) Show one useful extraction pattern that might be proposed by the AutoSlog system for identifying the name of a company that is involved in a merger or acquisition.

We did not specifically cover AutoSlog this semester, but we did talk about syntactico-semantic extraction patterns.

4. (4 pts) Why is noun phrase coreference resolution important for information extraction systems?
5. (4 pts) Provide one example from the sample text where accurate noun phrase coreference resolution could be important for the information extraction task. Be sure to explain why this is so.

5. (4 pts) Provide one example from the sample text where accurate noun phrase coreference resolution could be important for the information extraction task. Be sure to explain why this is so.

4 Grab Bag (18 pts)

- (a) (4 pts) (True or False. Explain your answer.) It is Zipf's law that accounts for the fact that the most frequent words in one corpus might be rare words in another corpus.
- (b) (4 pts) (True or False. Explain your answer.) WordNet's entries are lexeme-based.

- (c) (10 pts) Language Modeling Explain how one could use language modeling for *language identification*, i.e. the task of determining the language that a document is written in. Your answer should make clear how the trained language model(s) is/are to be used and should discuss the kind of data that you might use for training.