

CS447: Natural Language Processing

<http://courses.engr.illinois.edu/cs447>

Lecture 12

Midterm Review

Julia Hockenmaier

juliahmr@illinois.edu

3324 Siebel Center

Midterm Exam

When: Next Thursday, Oct 12, 6:30pm

Where: DCL 1320

What: Closed book exam:

- You are not allowed to use any cheat sheets, computers, calculators, phones etc.
(you shouldn't have to anyway)
- Only the material covered in lectures
- Bring a pen (black/blue) or pencil
- **Short questions — we expect short answers!**

Today's lecture

Quick run through the material we've covered, with some example questions.

(Not an exhaustive list of possible questions!)

Question types

Define X:

Provide a mathematical/formal definition of X

Explain X; Explain what X is/does:

Use plain English to define X and say what X is/does

Compute X:

Return X; Show the steps required to calculate it

Draw X:

Draw a figure of X

Show/Prove that X is true/is the case/...:

This may require a (typically very simple) proof.

Discuss/Argue whether ...

Use your knowledge (of X,Y,Z) to argue your point

Fundamentals

Basics about language

Explain Zipf's law and why it makes NLP difficult.

Explain why we often use statistical models in NLP.

Give two examples of ambiguity and explain how they make natural language understanding difficult.

Basics about language

Explain Zipf's law and why it makes NLP difficult.

Zipf's law says that a few words are very frequent, and most words are very rare. This makes NLP difficult because we will always come across rare/unseen words.

Explain why we often use statistical models in NLP.

To handle ambiguity (and make NLP systems more robust/to deal with the coverage problem).

Give two examples of ambiguity and explain why we have to resolve them.

POS ambiguity: back = noun or verb? Need to resolve this to understand the structure of sentences.

Word sense ambiguity: bank = river bank or institution. Need to resolve this to understand the meaning of sentences.

Morphology and finite-state transducers

Morphology

Explain what we mean by derivational morphology, and given an example in a language of your choice.

Draw a finite-state automaton for the language $\{a^n b^m\}$

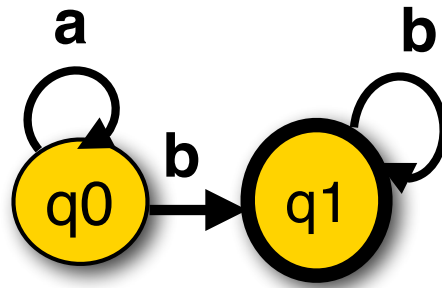
Define what a finite-state transducer is.

Explain what a finite-state transducer is.

Explain how we can use finite-state transducers for the morphological analysis of irregular verbs in English.

Morphology

Draw a finite-state automaton for the language $\{a^n b^m\}$



Morphology: Example answers

Define what a finite-state transducer is.

A **finite-state transducer** $T = \langle Q, \Sigma, \Delta, q_0, F, \delta, \sigma \rangle$ consists of:

- A finite **set of states** Q ,
- A finite alphabet Σ of **input symbols**
- A finite alphabet Δ of **output symbols**
- A designated **start state** $q_0 \in Q$
- A set of **final states** $F \subseteq Q$
- A **transition function** $\delta: Q \times \Sigma \rightarrow 2^Q$
- An **output function** $\sigma: Q \times \Sigma \rightarrow \Delta^*$

Explain what a finite-state transducer is

An FST defines a relation between two regular languages.

It maps each string in the input language to one (or more) strings in the output language. An FST is similar to a finite-state automaton, except that each transition is labeled with an input symbol and $a[n]$ (sequence of) output symbol(s).

Language modeling

Language modeling

Explain: What is a language model?

Explain and define: What is an n-gram language model?

Discuss the advantages and disadvantages of bigram language models over unigram models?

Explain and define how to estimate the parameters of a bigram model

Explain and define how evaluate the quality of a language model

Smoothing

Explain what smoothing is, and why it is necessary.

Define add-one smoothing and explain when it can be used.

Discuss the advantages/disadvantages of add-one smoothing.

Define how smoothing can be done via linear interpolation and explain when this technique can be used.

Hidden Markov Models and POS tagging

POS tagging

Discuss how you would define a POS tag set.

Explain the differences between open and closed word classes.

Explain how to do a quantitative evaluation of a POS tagger.

HMMs

Give the mathematical definition of a bigram HMM.

Explain how to estimate the parameters of a bigram HMM from labeled data.

Explain how the Viterbi algorithm is used for POS tagging with an HMM.

Find the most likely tag sequence for the following sentence (given some HMM).

Sequence labeling

Sequence labeling

Define the BIO encoding for NP chunking.

Define Maximum Entropy Markov Models.

Explain why MEMMs may be more suitable for named entity recognition than HMMs.

Draw the graphical model of MEMMs

Distributional similarities and Brown clusters

Dist. similarities & Brown clusters

Define pointwise mutual information.

Explain how pointwise mutual information can be used to find semantic clusters.

Explain how a class-based bigram model can be used to induce word classes.

Define what we mean by distributional similarities.

Explain how to construct a model of distributional similarities.