Review for Final Exam

Adam Meyers
New York University
Outline

• Administrative Details
• Annotation & Descriptive Linguistics
• Rules used by Automated Procedures
• Algorithms Discussed in Class
• Evaluation Metrics
• Topics: Formal Languages, Syntax, Parsing, POS tagging, Information Extraction, Terminology Extraction, Reference Resolution, Semantic Role Labeling, Feature Structures, Machine Translation
Administrative Details

• Determines 1/4 of grade
• Inclusive: topics over the course of the entire term
• Time and Place:
  – Thursday, December 17, 8:00—9:50, Room 317
• Open Book, Open Notes
  – You can bring materials
  – Search the web
  – Do simple calculations with a calculator
  – DO NOT:
    • communicate with others (texting, email, phone)
    • write/run actual programs
• Put your name and ID number on all test materials
Annotation

• You should be able to write usable specifications

• You should be able to annotate based on specifications

• You should understand some of the mechanics
  – Character offsets
  – A Markup language
  – BIO tags

• You should understand the difference between training and test corpora
Descriptive Linguistics

• The basic parts of speech and phrasal categories.
  – The difference between a determiners and an adjective
  – forms of verbs

• You know how to manually divide sentences into tokens

• You should know how to identify the head of a phrase

• You should be able to draw a phrase structure tree modeling the linguistic analysis of a sentence
Phrase Structure Rules and Tree

• Lets mark Parts of Speech, Draw a Phrase Structure Tree, and List the Rules for the following sentence from Wikipedia:
  – *Parodia tenuicylindrica is a small species of cactus native to the Rio Grande do Sul region of Brazil*
  – We will assume that:
    • species names consist of proper nouns
    • all the words in proper noun phrases are proper nouns
  – This sentence has other things about it that make it slightly more difficult than I would choose for the test
Rules: Regular Expressions

• You should know how to write a basic regular expression
• You should know how to write a phrase structure rule including at least:
  – Context free rules
  – Left (or right) regular rules
• For a regular expression, you should be able to identify a set of phrase structure rules that describe the same language (set of strings)
A Sample Regular Expression: In Class

- A regular expression for a date that will include the following expressions:
  - January 3, 2012
  - January, 2012
  - January 3
  - Jan. 3, 2012

- Specific Constraints:
  - Allow all the months of the year, as well as abbreviations that consist of the first 3 letters of the month and a period
  - The day should be a one or two digit number
  - The year should be a four digit number
  - The day and year are optional
  - A comma and a space precedes the year.
Algorithms: Deterministic Finite State Machine

• Given:
  – Finite State Machine (FSM)
  – Input String

• Would the FSM recognize the string?

• Which sequence of states would be entered before recognition was complete?

• How would the FSM on the next slide process:
  – AababAB
  – AABB
DFSA for Regexp: $A(ab)^*ABB$?

<table>
<thead>
<tr>
<th>State</th>
<th>Input</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q0</td>
<td>Q1</td>
</tr>
<tr>
<td>Q1</td>
<td>Q3</td>
</tr>
<tr>
<td>Q2</td>
<td>Q1</td>
</tr>
<tr>
<td>Q3</td>
<td>Q4</td>
</tr>
<tr>
<td>Q4</td>
<td>Q5</td>
</tr>
<tr>
<td>Q5</td>
<td>Q5</td>
</tr>
</tbody>
</table>
Algorithms: Context-Free Generator

• Show the steps for randomly generating a sentence given:
  – A lexicon and a context-free grammar with start symbol S
• The algorithm expands each non-terminal into a randomly chosen right hand side.
• Going left to right, the first non-terminal symbol is always expanded first.
• The mechanism (as discussed in class) is to place each right hand side on top of the stack with the left-most symbol at the top of the stack.
Chomsky Normal Form (required by CKY)

- Context Free Grammars can be converted to CNN
  - 3 types of rules:
    - XP → YP ZP  Binary Branching
    - XP → x  NonTerminal → terminal
    - XP → ε  NonTerminal → empty string

- Conversion
  - Replace $VP \rightarrow VG$ & $NP \rightarrow NG$ with nonbranching rules expanding VP and NP to whatever VG and NG matched to:
    - Not good examples: $VP \rightarrow ate$, $VP \rightarrow had$, $NP \rightarrow food$, ...
    - Usually V and N are assumed to be nonterminals, and the POS of words are looked up
  - Replace $VP \rightarrow V NP PP$ with 2 rules:
    - $VP \rightarrow VG PP$
    - $VG \rightarrow V NP$
  - Replace: $NP \rightarrow POSSP N PP$ with 2 rules:
    - $NP \rightarrow NG PP$
    - $NG \rightarrow POSSP N$
Algorithms: The CKY parsing algorithm

• Fill in the triangular chart given a (short) sentence and a set of context free rules

• Remember
  – How the chart encodes start and end positions
  – That each rule is in Chomsky Normal Form
    • i.e., is binary branching

• See the next slide
# 6th Iteration of CKY Algorithm

<table>
<thead>
<tr>
<th>The</th>
<th>clam's group</th>
<th>had knowledge</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>0</td>
<td>D [0,1]</td>
<td>NP [0,2]</td>
</tr>
<tr>
<td>1</td>
<td>N, NP [1,2]</td>
<td>POSSP [1,3]</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>POSS [2,3]</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Review for Final Exam 2015
Viterbi Decoding of HMM for *rose pickles*

- **Likelihood:**
  - *rose*: NNP .01, NN .02, VBD .05
  - *pickles*: NNP .001, NNS .03, VBZ .05

- **Transition Probabilities:**
**Rose Pickles**

- **Likelihood:**
  - *rose*: NNP .01, NN .02, VBD .05
  - *pickles*: NNP .001, NNS .03, VBZ .05

- Fill in: max (previous X transition X likelihood)

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1: Rose</th>
<th>2: Pickles</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Start</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NNP</td>
<td>*.42 *.01</td>
<td></td>
<td><em>.3</em>.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>*.01</td>
<td>*.01</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>.23</em>.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NNS</td>
<td>*.03</td>
<td></td>
<td><em>.5</em>.03</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>.24</em>.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NN</td>
<td>*.20 *.02</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VBZ</td>
<td><em>.3</em>.05</td>
<td></td>
<td><em>.1</em>.05</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>.1</em>.05</td>
<td>*.05</td>
<td></td>
</tr>
<tr>
<td>VBD</td>
<td>*.05 *.05</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>End</td>
<td></td>
<td></td>
<td>*.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>*.15</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>*.45</td>
<td></td>
</tr>
</tbody>
</table>

- Likelihood:
  - *rose*: NNP .01, NN .02, VBD .05
  - *pickles*: NNP .001, NNS .03, VBZ .05

- Fill in: max (previous X transition X likelihood)
**Rose/NNP Pickles/VBZ**

- **Likelihood:**
  - *rose*: NNP .01, NN .02, VBD .05
  - *pickles*: NNP .001, NNS .03, VBZ .05

- **Fill in**: max (previous X transition X likelihood)

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1: Rose</th>
<th>2: Pickles</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Start</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NNP</td>
<td></td>
<td>.42 * .01</td>
<td><em>.3</em>.001=1.26*10^-6</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><em>0</em>.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><em>.23</em>.001</td>
<td></td>
</tr>
<tr>
<td>NNS</td>
<td></td>
<td></td>
<td><em>0</em>.03</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><em>.5</em>.03=6*10^-5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><em>.24</em>.03</td>
<td></td>
</tr>
<tr>
<td>NN</td>
<td></td>
<td>.20 * .02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VBZ</td>
<td></td>
<td></td>
<td><em>.3</em>.05=6.3*10^-5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><em>.1</em>.05</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><em>.0</em>.05</td>
<td></td>
</tr>
<tr>
<td>VBD</td>
<td></td>
<td>.05 * .05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>End</td>
<td></td>
<td></td>
<td></td>
<td>(.2=2.52*10^{-7})</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(.15 = 9.0*10^{-6})</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(.45 = 2.85*10^{-5})</td>
</tr>
</tbody>
</table>

Review for Final Exam
2015
Common Evaluation Metrics

• If all instances are classified
  – Accuracy = Correct/All-Instances

• If only some instances are classified
  – Precision = Correct/Instances in System Output
  – Recall = Correct/Instances in Answer Key
  – F-measure = Mean of Precision and Recall
    • Actually Harmonic Mean of Precision and Recall
      – \[ \frac{2}{\frac{1}{\text{precision}} + \frac{1}{\text{recall}}} \]
Sample Precision and Recall

- System for finding holiday names
- Exactly 10 correct holiday names in hand-coded corpus (the answer key)
- The system marks 12 holiday names, 8 of which match the ones in the answer key.
  - Precision = 8/12 = .67
  - Recall = 8/10 = .80
  - F-measure = 2/(1/.80+1/.67) = .73
TFIDF

- TFIDF – Property of Term with respect to a document
  - keyword suitability, representativeness of a topic, etc.
  - Uses: Doc Retrieval, Term Extraction, etc.
- TF = frequency in a document
- IDF = number of documents in sample divided by number of documents containing word
- TFIDF = TF * log(IDF)
- Example: “rock” occurs 10 times in document X. It occurs in 100 out of 3000 documents in collection. TFIDF = 10*log(3000/100) = 34.01
Cosine Similarity Between Query and Document

\[
\text{Similarity}(A, B) = \frac{\sum_i a_i \times b_i}{\sqrt{\sum_i a_i^2 \times \sum_i b_i^2}}
\]

- Example:
  - the terms in the vectors include: animal, vegetable, mineral, monkey, golf enthusiast
  - The vector for the query is: [0, 0, 0, 34, .8]
  - The vector for a given document is: [1, 2, 3, 4, 5]
  - What is the similarity?
Binding Theory for English 3rd Pers Prons

• Case 1: If the pronoun \( p \) is inside an NP premodified by a possessive, the antecedent needs to be outside of this NP
  – John likes Mary's drawing of him
  – John likes his drawing of Mary

• Case 2: Otherwise, the antecedent must be outside the immediate tensed clause containing the personal pronoun.
  – John said that he liked pizza.
  – John wanted for him to like pizza.
  – John liked him.

• Theories of binding vary about how these (and similar) constraints are encoded, but the differences in the final result (quality of system output) is minimal. While these 2 rules cover most cases, there are also some exceptions:
  – John always carries a slice of pizza with him.
Binding Theory for English Reflexives/Reciprocals

• The antecedent of a reflexive/reciprocal **must be** the closest subject or possessive such that:
  – The antecedent precedes and “commands” the pronoun
    • A commands B if A is the sibling of a phrase that dominates B.
  – There is no possessive or subject for phrases in the path in the phrase structure tree between antecedent and pronoun

• Examples:
  – *Mary saw herself* vs. **Mary said that John would meet herself soon**
  – *Mary's picture of herself* vs. **Mary saw John's picture of herself**

• These rules covers most cases.
  – Exception: *Pictures of themselves made the actors nervous.*
Hobbs Search Algorithm to Find Antecedent of Anaphors

1. Go to NP immediately dominating pronoun

2. Go up to 1st dominating NP or S node. This node = X path to X = p.

3. Traverse branches below X to the left of p, left-to-right and and breadth first. Propose each NP n as an antecedent if there is an NP or S between n and X.

4a. Is X the highest S in the sentence? (Recursive)

4b. Search previous sentences in order from right ot left. Search each tree from left to right, breadth 1st, proposing each NP as antecedent

5. From Node X, go up to the 1st NP or S. Call this node X and the path to X = p.

6. If X = NP and there is no N’ in p, propose X as antecedent

7. Search for antecedent in branches below X, left-to-right, breadth first. Propose each NP.

8. If X is an S node, traverse all branches of X preceding p, left-to-right, breadth first, but not going below any S or NP node found Propose each NP.

End
Hobbs Search Example

1. Mary saw the chicken.
2. Jim said that she laughed.
Evaluation

- Recall/Precision/F-Score given answer key for a task
  \[
  \text{Recall} = \frac{\text{Correct}}{\text{Answer Key}} \quad \text{Precision} = \frac{\text{Correct}}{\text{System Output}} \quad F - \text{Score} = \frac{2}{\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}}
  \]

- Kappa – How much better is interannotator agreement than chance?
  \[
  \text{Kappa} = \frac{\text{Percent (Actual Agreement)} - \text{Prob (Chance Agreement)}}{1 - \text{Prob (Chance Agreement)}}
  \]

- Coreference with the B-cubed algorithm
  - Precision: For each partition of Coreferential NPs in system output: average the percentage of links in the answer
  - Recall: For each partition of Coreferential NPs in the answer, average the percentage of links in the system output
  - Example:
    - 2 system entities: \{A_1,A_2,A_3,A_4,C_1\} and \{B_1,B_2\}
    - 2 answer key entity: \{A_1,A_2,A_3,A_4,B_1,B_2\} and \{C_1\}
    - Precision = \((4 \times 4/5) + 1/5 + 2\) \times 1/7 = .77
    - Recall = \((4 \times 2/3) + (2 \times 1/3) + 1\) \times 1/7 = .62
Example Feature Structure Question

• What (Typed) Feature Structure would result if FS1 was unified with the value of the path Quant in FS2?
  – Note: orth is short for orthography and cat is short for category
Machine Translation Questions

• Decoding Question (like the sample test)
• Walk through part of Maximization/Estimation
  – Given a set of translation probabilities, calculate the probability of a particular alignment
  – Given a set of alignments and their probability scores, calculate the new translation probabilities
Simplified Example of EM model

• Given
  – 4 French words: *la, maison, bleu*, and *fleur*
  – 4 English words: *the, house, blue* and *flower*
  – We only allow 1 to 1 alignments

• Starting assumption
  – Each French word has a .25 chance of being translated as a given English word
Initial Alignment Probs for 3 E/F pairs

• *la maison → the house*
  – *la/the* (.25), *maisson/the* (.25), *la/house* (.25), *maisson/house* (.25)
  – *la/the* X *maisson/house* = .0625
  – *maisson/the* X *la/house* = .0625

• *la maison bleu → the blue house* (all possible alignments)
  – *la/the* X *maisson/house* X *bleu/blue* = .25^3 = .015625
  – *la/the* X *maisson/blue* X *bleu/house* = .015625
  – *la/house* X *maisson/the* X *bleu/blue* = .015625
  – *la/house* X *maisson/blue* X *bleu/house* = .015625
  – *la/blue* X *maisson/house* X *bleu/the* = .015625
  – *la/blue* X *maisson/the* X *bleu/house* = .015625

• *La fleur → the flower*
  – *la/the* X *fleur/flower* = .0625
  – *fleur/the* X *la/flower* = .0625
Maximum Liklihood Estimates (MLE)

- For each e/f pair and for each sentence, add up the probabilities of alignments that contain that pair and regularize to 1 (initially: all prob=.25)
- Sum these scores and divide by the number of instances of f.

Translations from X to the
- \text{la/the}: .5 of the first set of alignments, .33 of the second set and .5 of the 3\textsuperscript{rd}
  - \((.5 + .33 + .5) / 3 = .44\)
- \text{maisson/the}: .5 of the 1\textsuperscript{st} + .33 of the 2\textsuperscript{nd}, 0 in the 3\textsuperscript{rd}
  - \((.5 + .33)/3 = .42 = .29\)
- \text{bleu/the}: 0 in the 1\textsuperscript{st} + .33 of the 2\textsuperscript{nd} + 0 in the 3\textsuperscript{rd}
  - \(.33/3 = .11\)
- \text{fleur/the}: 0 in the 1\textsuperscript{st} and 2\textsuperscript{nd}, .5 in the 3\textsuperscript{rd}
  - \(.5/3 = .17\)

- \text{house}: \text{la/house}=.42, \text{maisson/house}=.42, \text{bleu/house}=.17, \text{fleur/house}=0
- \text{blue}: \text{la/blue}=.33, \text{maisson/blue}=.33, \text{bleu/blue}=.33, \text{fleur/blue}=0
- \text{flower}: \text{la/flower}=.5 \text{maisson/flower}=0, \text{blue/flower}=0, \text{fleur/flower}=.5
Expectation: Rescore Alignments

- **la maison → the house**
  - la/the (.44), maison/the (.29), la/house (.42), maison/house (.42)
  - la/the X maison/house = .1848
  - maison/the X la/house = .1276

- **la maison bleu → the blue house** (all possible alignments)
  - la/the X maison/house X bleu/blue = .06098
  - la/the X maison/blue X bleu/house = .02468
  - la/house X maison/the X bleu/blue = .04019
  - la/house X maison/blue X bleu/house = .02356
  - la/blue X maison/house X bleu/the = .045274
  - la/blue X maison/the X bleu/house = .016269

- **La fleur → the flower**
  - la/the X fleur/flower = .22000
  - fleur/the X la/flower = .08500

Review for Final Exam 2015
Translating sample sentence

- Input: *La maison bleu*
- Translation probabilities (hypothetical):

<table>
<thead>
<tr>
<th></th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><em>the</em></td>
</tr>
<tr>
<td><em>la</em></td>
<td>.70</td>
</tr>
<tr>
<td><em>maison</em></td>
<td>.24</td>
</tr>
<tr>
<td><em>bleu</em></td>
<td>.25</td>
</tr>
<tr>
<td><em>fleur</em></td>
<td>.19</td>
</tr>
</tbody>
</table>

- Unigram probabilities (count in WSJ ÷ 1 million)
- The most probable translation would be:
  - *the house blue* = translation-prob X language prob = 4.37 X 10^-10
    - translation-prob = .7 X .5 X .41 = .1435
    - Lang-prob = .035 X 6.7 X 10^-4 X 1.3 X 10^-4 = 3.05 X 10^-9

Review for Final Exam 2015
Sample Annotation Task

- **APPOSITION**: if the comma joins to NPs that are in an is-a relation, e.g., *Barack Obama, president of the U.S.*
- **AFFILIATED**: if the comma joins to NPs that are related such that the second provides an affiliation or a larger entity that the first is a part of, e.g., a) *Adam Meyers, NYU*; b) *Columbus, Ohio*
- **CONJUNCTION**: if the comma is between constituents that are conjoined together, e.g., *big cats, little cats, cats with fangs, or regular cats*
- **ADVERBIAL**: if comma immediately precedes or follows an adverb or adverbial phrase, e.g., *Probably, that will not work.*
- **PARENTHEtical**: if commas surround a phrase that comments on the surrounding text or indicates who is speaking, thinking or observing – the parenthetical text should be deletable without changing the grammaticality of the surrounding text. It should comment on the text but not change its meaning, e.g., a) *Mary, I believe, is twelve feet tall*; b) *Mary, oh yeah I mean her, is twelve feet tall.*
The beauty of automatic replenishment is that the buyer is really the customer. She is telling us what she wants and needs in the future. Quite frankly, of all the buying we do, letting our customer make the choice seems to make the most sense.

—Tom Cole, Chairman and CEO, Federated Logistics and Operations

Our goal is to replace the product on the retail shelf as quickly as possible, because that’s where the consumer buys it.

—Jeff Kernodle, Vice President for Replenishment, VF Corp

Many of the popular accounts of quick response, rapid replenishment, and supply-chain management assume that all parties—consumers, retailers, and suppliers—win as a result of these policies. Consumers have definitely benefited because these practices afford them a greater choice of products at lower average prices.1 It is safe to say that lean retailers have also come out ahead, given their rapid growth in relation to, and at the expense of, traditional retailers in many different retail channels. But have suppliers benefited from entering into relations with lean retailers? Have such firms improved their competitive position along with the retailers they supply?
Sample Final

- Structured like the Actual Final
- Approximately the same difficulty
- Both the sample and real final will cover topics discussed in this talk
- Proviso: The topics of the actual final and the sample overlap, but are not exactly the same – some topics are only in the final and some are only in the sample