5525: Speech and Language Processing

Wei Xu

(many slides from Greg Durrett)
Administrivia

- Course website: https://cocoxy.github.io/courses/5525_spring20.html
- Piazza: link on the course website
- My office hours: Thursday 4:30-5:30pm DL 495
- TA: Ashutosh Baheti;
- TA Office hours: Tuesday 4:00-5:00pm DL 190
Course Requirements

- Probability (e.g., conditional probabilities, Bayes Rule, etc.)
- Linear Algebra (e.g., multiplying vectors and matrices, matrix inversion)
- Calculus (e.g., calculating gradients of functions with several variables)
- Programming / Python experience
- Prior exposure to machine learning algorithms very helpful

There will be a lot of math and programming!
Homework 1 is out now (due in 2 weeks):

- Please look at the assignment well before then

- If this seems like it’ll be challenging for you, come and talk to me (this is smaller-scale than the later assignments, which are smaller-scale than the final project)
Two great textbooks for NLP

- There will be assigned readings from both
- Both freely available online

**Speech and Language Processing** (3rd ed. draft)
*Dan Jurafsky* and *James H. Martin*

Draft chapters in progress, October 16, 2019

**Natural Language Processing**

*Jacob Eisenstein*

November 13, 2018
What’s the goal of NLP?

- Be able to solve problems that require deep understanding of text
- Example: dialogue systems

Siri, what’s the most valuable American company?

Apple

Who is its CEO?

Tim Cook

- recognize marketCap is the target value
- do computation
- recognize predicate
- resolve references
Google Critic Ousted From Think Tank Funded by the Tech Giant

WASHINGTON — In the hours after European antitrust regulators levied a record $2.7 billion fine against Google in late June, an influential Washington think tank learned what can happen when a tech giant that shapes public policy debates with its enormous wealth is criticized.

But not long after one of New America’s scholars posted a statement on the think tank’s website praising the European Union’s penalty against Google, Mr. Schmidt, who had been chairman of New America until 2016, communicated his displeasure with the statement to the group’s president, Anne-Marie Slaughter, according to the scholar.

Ms. Slaughter told Mr. Lynn that “the time has come for Open Markets and New America to part ways,” according to an email from Ms. Slaughter to Mr. Lynn. The email suggested that the entire Open Markets team — nearly 10 full-time employees and unpaid fellows — would be exiled from New America.

One of New America’s writers posted a statement critical of Google. Eric Schmidt, Google’s CEO, was displeased.

The writer and his team were dismissed.

paraphrase to provide clarity

provide missing context

compress text
Trump Pope family watch a hundred years a year in the White House balcony
NLP is about building these pieces!

All of these components are modeled with statistical approaches trained with machine learning.
How do we represent language?

Labels
- the movie was good
- Beyoncé had one of the best videos of all time (subjective)

Sequences/tags
- PERSON
  - Tom Cruise
- WORK_OF_ART
  - Mission Impossible film

Trees
- S
  - VP
    - NP
      - VBZ: eat
      - NN: cake
    - PP: with icing
- λx. flight(x) ∧ dest(x)=Miami
- flights to Miami
Main question: What representations do we need for language? What do we want to know about it?

Boils down to: what ambiguities do we need to resolve?
Why is language hard?
(and how can we handle that?)
Hector Levesque (2011): “Winograd schema challenge” (named after Terry Winograd, the creator of SHRDLU)

The city council refused the demonstrators a permit because they ______ violence

they advocated

they feared

This is so complicated that it’s an AI challenge problem! (AI-complete)

Referential/semantic ambiguity
Language is Ambiguous!

- Headlines
  - Teacher Strikes Idle Kids
  - Hospitals Sued by 7 Foot Doctors
  - Ban on Nude Dancing on Governor’s Desk
  - Iraqi Head Seeks Arms
  - Stolen Painting Found by Tree
  - Kids Make Nutritious Snacks
  - Local HS Dropouts Cut in Half

- Syntactic/semantic ambiguity: parsing needed to resolve these, but need context to figure out which parse is correct
Language is **Really** Ambiguous!

- There aren’t just one or two possibilities which are resolved pragmatically

  - It is really nice out
  - It’s really nice
  - The weather is beautiful
  - It is really beautiful outside
  - He makes truly beautiful
  - He makes truly boyfriend
  - It fact actually handsome

- Combinatorially many possibilities, many you won’t even register as ambiguities, but systems still have to resolve them
What do we need to understand language?

- Lots of data!

<table>
<thead>
<tr>
<th>SOURCE</th>
<th>Cela constituerait une solution transitoire qui permettrait de conduire à terme à une charte à valeur contraignante.</th>
</tr>
</thead>
<tbody>
<tr>
<td>HUMAN</td>
<td>That would be an interim solution which would make it possible to work towards a binding charter in the long term.</td>
</tr>
<tr>
<td>1x DATA</td>
<td>[this] [constituerait] [assistance] [transitoire] [who] [permettrait] [licences] [to] [terme] [to] [a] [charter] [to] [value] [contraignante] [.]</td>
</tr>
<tr>
<td>10x DATA</td>
<td>[it] [would] [a solution] [transitional] [which] [would] [of] [lead] [to] [term] [to] [a] [charter] [to] [value] [binding] [.]</td>
</tr>
<tr>
<td>100x DATA</td>
<td>[this] [would be] [a transitional solution] [which would] [lead to] [a charter] [legally binding] [.]</td>
</tr>
<tr>
<td>1000x DATA</td>
<td>[that would be] [a transitional solution] [which would] [eventually lead to] [a binding charter] [.]</td>
</tr>
</tbody>
</table>
What do we need to understand language?

- World knowledge: have access to information beyond the training data

**Diagram:**
- DOJ greenlights Disney-Fox merger
- Department of Justice
- metaphor; “approves”

- What is a green light? How do we understand what “green lighting” does?
What do we need to understand language?

- Grounding: learn what fundamental concepts actually mean in a data-driven way

**Question:** What object is right of O2?

Golland et al. (2010)

McMahan and Stone (2015)
What do we need to understand language?

- Linguistic structure
- ...but computers probably won’t understand language the same way humans do
- However, linguistics tells us what phenomena we need to be able to deal with and gives us hints about how language works

a. John has been having a lot of trouble arranging his vacation.

b. He cannot find anyone to take over his responsibilities. (he = John)
   \[ C_b = \text{John}; \ C_f = \{\text{John}\} \]

c. He called up Mike yesterday to work out a plan. (he = John)
   \[ C_b = \text{John}; \ C_f = \{\text{John, Mike}\} \ (\text{CONTINUE}) \]

d. Mike has annoyed him a lot recently.
   \[ C_b = \text{John}; \ C_f = \{\text{Mike, John}\} \ (\text{RETAIN}) \]

e. He called John at 5 AM on Friday last week. (he = Mike)
   \[ C_b = \text{Mike}; \ C_f = \{\text{Mike, John}\} \ (\text{SHIFT}) \]
What techniques do we use?
(to combine data, knowledge, linguistics, etc.)
A brief history of (modern) NLP

1980

“AI winter” rule-based, expert systems

1990

Penn treebank

S
NP VP

Ratnaparkhi tagger

NNP VBZ

1990

Collins vs. Charniak parsers

2000

Unsup: topic models, grammar induction

Sup: SVMs, CRFs, NER, Sentiment

Semi-sup, structured prediction

2010

Pretraining

2018

Neural Pretraining

early stat MT work at IBM

[Image of Ratnaparkhi]
Structured Prediction

- All of these techniques are data-driven! Some data is naturally occurring, but may need to be labeled.
- Supervised techniques work well on very little data.
- Even neural nets can do pretty well!

“Learning a Part-of-Speech Tagger from Two Hours of Annotation”
Garrette and Baldridge (2013)
Less Manual Structure?

<table>
<thead>
<tr>
<th>(a) example word alignment</th>
<th>(b) example phrase alignment</th>
</tr>
</thead>
</table>

Thank you, I shall do so.  
Gracias, lo haré de muy buen grado.

DeNero et al. (2008)  
Bahdanau et al. (2014)
Does manual structure have a place?

- Neural nets don’t always work out of domain!
- Coreference: rule-based systems are still about as good as deep learning out-of-domain
- LORELEI: transition point below which phrase-based systems are better
- Why is this? Inductive bias!
- Can multi-task learning help?

<table>
<thead>
<tr>
<th></th>
<th>CoNLL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Avg. F₁</td>
</tr>
<tr>
<td>Newswire</td>
<td></td>
</tr>
<tr>
<td>rule-based</td>
<td>55.60</td>
</tr>
<tr>
<td>berkeley</td>
<td>61.24</td>
</tr>
<tr>
<td>cort</td>
<td>63.37</td>
</tr>
<tr>
<td>deep-coref [conll]</td>
<td>65.39</td>
</tr>
<tr>
<td>deep-coref [lea]</td>
<td>65.60</td>
</tr>
<tr>
<td>Wikipedia</td>
<td></td>
</tr>
<tr>
<td>rule-based</td>
<td>51.77</td>
</tr>
<tr>
<td>berkeley</td>
<td>51.01</td>
</tr>
<tr>
<td>cort</td>
<td>49.94</td>
</tr>
<tr>
<td>deep-coref [conll]</td>
<td>52.65</td>
</tr>
<tr>
<td>deep-coref [lea]</td>
<td>53.14</td>
</tr>
<tr>
<td>deep-coref⁻</td>
<td>51.01</td>
</tr>
</tbody>
</table>

Moosavi and Strube (2017)
Does manual structure have a place?

- Trump, Pope, family watch a hundred years a year in the White House balcony

- Maybe manual structure would help...
Where are we?

- NLP consists of: analyzing and building representations for text, solving problems involving text

- These problems are hard because language is ambiguous, requires drawing on data, knowledge, and linguistics to solve

- Knowing which techniques use requires understanding dataset size, problem complexity, and a lot of tricks!

- NLP encompasses all of these things
NLP vs. Computational Linguistics

- NLP: build systems that deal with language data
- CL: use computational tools to study language

Hamilton et al. (2016), KulKarni et al. (2015)
NLP vs. Computational Linguistics

- Computational tools for other purposes: literary theory, political science...

Bamman, O’Connor, Smith (2013)
ML and structured prediction for NLP

Neural Networks semantics

Applications: MT, IE, summarization, dialogue, etc.

Outline of the Course

<table>
<thead>
<tr>
<th>Date</th>
<th>Topics (tentative and subject to change)</th>
<th>Readings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/8/2020</td>
<td>Course Overview</td>
<td>J+M, 3rd Edition Chapter 1</td>
</tr>
<tr>
<td>1/10/2020</td>
<td>Binary Classification</td>
<td>Eisenstein 2.0-2.5, 4.1-4.3-4.5, CILM, 4.1-4.4, 4.6-4.7</td>
</tr>
<tr>
<td>1/15/2020</td>
<td>Binary Classification</td>
<td>Eisenstein 2.0-2.5, 4.1-4.3-4.5, CILM, 4.1-4.4, 4.6-4.7</td>
</tr>
<tr>
<td>1/17/2020</td>
<td>Multiclass Classification</td>
<td>J+M Chapter 5</td>
</tr>
<tr>
<td>1/22/2020</td>
<td>Multiclass Classification</td>
<td>J+M Chapter 5</td>
</tr>
<tr>
<td>1/24/2020</td>
<td>Neural Networks</td>
<td>Eisenstein 3.1-3.3, J+M 7.1-7.4</td>
</tr>
<tr>
<td>1/29/2020</td>
<td>Sequence Tagging</td>
<td>Eisenstein 7.0-7.4, J+M Chapter 8</td>
</tr>
<tr>
<td>1/31/2020</td>
<td>Viterbi Algorithm</td>
<td>J+M Chapter 8</td>
</tr>
<tr>
<td>2/5/2020</td>
<td>Conditional Random Fields</td>
<td>Eisenstein 7.5, 8.3,</td>
</tr>
<tr>
<td>2/7/2020</td>
<td>Conditional Random Fields</td>
<td>Eisenstein 7.5, 8.3,</td>
</tr>
<tr>
<td>2/12/2020</td>
<td>Word Embeddings</td>
<td>Eisenstein 3.3.4, 14.5, 14.6, J+M 6</td>
</tr>
<tr>
<td>2/14/2020</td>
<td>Semantics</td>
<td>Eisenstein 3.3.4, 14.5, 14.6, J+M 6</td>
</tr>
<tr>
<td>2/19/2020</td>
<td>Recurrent Neural Networks</td>
<td>Goldberg 10.11,</td>
</tr>
<tr>
<td>2/21/2020</td>
<td>Convolutional Neural Networks, Neural CRFs</td>
<td>Eisenstein 3.4, 7.6</td>
</tr>
<tr>
<td>2/26/2020</td>
<td>Machine Translation</td>
<td>Eisenstein 18.1, 18.2</td>
</tr>
<tr>
<td>2/28/2020</td>
<td>Encoder-Decoder Networks</td>
<td>Seq2Seq</td>
</tr>
<tr>
<td>3/4/2020</td>
<td>Spring Break</td>
<td></td>
</tr>
<tr>
<td>3/6/2020</td>
<td>Spring Break</td>
<td></td>
</tr>
<tr>
<td>3/11/2020</td>
<td>Information Extraction</td>
<td>Eisenstein 13, 17</td>
</tr>
<tr>
<td>3/13/2020</td>
<td>Information Extraction</td>
<td>Eisenstein 13, 17</td>
</tr>
<tr>
<td>3/18/2020</td>
<td>Possible Midterm Date (tentative)</td>
<td>Eisenstein 18.3</td>
</tr>
<tr>
<td>3/20/2020</td>
<td>Neural Machine Translation</td>
<td></td>
</tr>
<tr>
<td>3/25/2020</td>
<td>Reading Comprehension</td>
<td>E2E Memory Networks, CBT, SQAD, BiDAF</td>
</tr>
<tr>
<td>3/27/2020</td>
<td>Summarization</td>
<td>Eisenstein 19, MMR, Gillick,</td>
</tr>
<tr>
<td>4/1/2020</td>
<td>Generation</td>
<td>Sentence compression, SummaRuNNER, Pointer</td>
</tr>
<tr>
<td>4/3/2020</td>
<td>Dialogue</td>
<td>J+M Chapter 24</td>
</tr>
<tr>
<td>4/8/2020</td>
<td>Unsupervised Learning in NLP</td>
<td>Painless unsup, VAE, ELMo, BERT</td>
</tr>
<tr>
<td>4/10/2020</td>
<td>Unsupervised Learning in NLP</td>
<td>Painless unsup, VAE, ELMo, BERT</td>
</tr>
<tr>
<td>4/15/2020</td>
<td>Ethics/Wrapup</td>
<td></td>
</tr>
<tr>
<td>4/17/2020</td>
<td>No Class</td>
<td></td>
</tr>
</tbody>
</table>

final exam date  Final Project Presentation
Course Goals

- Cover fundamental machine learning techniques used in NLP
- Understand how to look at language data and approach linguistic phenomena
- Cover modern NLP problems encountered in the literature: what are the active research topics in 2018~2020?
- Make you a “producer” rather than a “consumer” of NLP tools
  - The three assignments should teach you what you need to know to understand nearly any system in the literature
Assignments

- Three Homework Assignments
  - Implementation-oriented, with an open-ended component to each
  - Homework 1 (Naive Bayes for sentiment classification) is out NOW
  - ~2 weeks per assignment, 3 “slip days” for automatic extensions

These projects require understanding of the concepts, ability to write performant code, and ability to think about how to debug complex systems. They are challenging, so start early!
Midterm Exam & Final Project

- Midterm — in class, date TBD (20%)

- Final project (20%) — presentation on the final exam day
  - Groups of 3-4 preferred, 1 is possible.
  - Good idea to talk to run your project idea by me in office hours or email.
  - 4 page report + final project presentation.

- Participation (10%)