CS4650: Natural Language Processing

 Wei Xu
(many slides from Greg Durrett)
Administrivia

- Course website:
  https://www.cc.gatech.edu/classes/AY2021/cs4650_spring/

- Piazza/Gradescope: link on the course website

- My office hour: TBA

- TAs:
  
<table>
<thead>
<tr>
<th>Jingfeng Yang</th>
<th>Kaige Xie</th>
<th>Sarmishta Velury</th>
</tr>
</thead>
</table>

- TA office hours: Mon 6:00-7:00pm, Wed 6:00-7:00pm, Thu 10:00-11:00pm
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!
Enrollment and Prereq

- Background Test (5%) is out now (due tomorrow 1/21):
  - Designed to help you determine whether you have enough math and programming background to succeed in this class.
  - If the background test is not enough for calibrating, you may read Chapters 2~4 of the textbook by Jacob Eisenstein. We will cover these content in the first few weeks of the semester.
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

Introduction to Natural Language Processing
By Jacob Eisenstein
Published by The MIT Press
Oct 01, 2019 | 536 Pages | 7 x 9
| ISBN 9780262042840
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 (or four) programming assignments should teach you what you need to know to understand nearly any system in the literature
Assignments

- Four Homework Assignments (55%)
  - Implementation-oriented
  - Homework 1 will be out soon
  - ~2 weeks per assignment, 3 “slip days” for **up to 2** homework (3 days each)
    - Homework 1: 15% (written + 1st programming)
    - Homework 2: 10% (written)
    - Homework 3: 15% (written + 2nd programming)
    - Homework 4: 15% (written + 3rd programming)

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!**
Final Project, etc.

- No Midterm

- Final project (20%)
  - Groups of 2-3 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.
  - **Alternatively,** you may choose to complete the Homework 5 (written+programming) individually

- Quizzes (10%)
- Participation (10%)
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.
Trump and Pope family watch a hundred years a year in the White House balcony.
Trump and his family watched a 100-year total solar eclipse on the balcony of the White House.
Textual Entailment

<table>
<thead>
<tr>
<th>Text</th>
<th>Judgments</th>
<th>Hypothesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>A man inspects the uniform of a figure in some East Asian country.</td>
<td>contradiction C C C C</td>
<td>The man is sleeping</td>
</tr>
<tr>
<td>An older and younger man smiling.</td>
<td>neutral N N E N N</td>
<td>Two men are smiling and laughing at the cats playing on the floor.</td>
</tr>
<tr>
<td>A black race car starts up in front of a crowd of people.</td>
<td>contradiction C C C C</td>
<td>A man is driving down a lonely road.</td>
</tr>
</tbody>
</table>

SNLI (Bowman et al., 2015)

- Text is connected to intelligence and knowledge in a fundamental way!
- Goal of NLP (solving problems with text) requires *analyzing* and *understanding* text
- What makes this analysis hard?
NLP Analysis Pipeline

Text

- Syntactic parses
- Coreference resolution
- Entity disambiguation
- Discourse analysis

Annotations

Applications

- Summarize
- Extract information
- Answer questions
- Identify sentiment
- Translate

- 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 stars in the new Mission Impossible film

**Trees**

I eat cake with icing

\[ \lambda x. \text{flight}(x) \land \text{dest}(x)=\text{Miami} \]

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?)
Students Cook & Serve Grandparents

On Thursday, September 9, Gorman School hosted the first annual Grandparent's Day. All Grandparents were invited to a school-wide pancake breakfast. Upper grade students served as excellent chefs, as well as taking responsibility for serving the food and the clean up after-
Language is Ambiguous!

Other 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 Ambiguous!

- 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 **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

```
DOJ greenlights Disney - Fox merger
```

Department of Justice

What is a green light? How do we understand what “green lighting” does?

- metaphor; “approves”
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 02?

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)
     \[ C_b = \text{John}; C_f = \{\text{John, Mike}\} \]
  d. Mike has annoyed him a lot recently.
     \[ C_b = \text{John}; C_f = \{\text{Mike, John}\} \]
  e. He called John at 5 AM on Friday last week. (he = Mike)
     \[ C_b = \text{Mike}; C_f = \{\text{Mike, John}\} \]

Centering Theory
Grosz et al. (1995)
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: earliest stat MT work at IBM
- 1990: Penn treebank (S, NP, VP)
- 2000: Collins vs. Charniak parsers
- 2000: Ratnaparkhi tagger (NNP, VBZ)
- 2010: Unsup: topic models, grammar induction
- 2010: Sup: SVMs, CRFs, NER, Sentiment
- 2010: Semi-sup, structured prediction
- 2018: Pretraining

Pretraining

Neural

Semi-sup, structured prediction

Sup: SVMs, CRFs, NER, Sentiment

Unsup: topic models, grammar induction

Ratnaparkhi tagger (NNP, VBZ)

Collins vs. Charniak parsers

Penn treebank (S, NP, VP)

earliest stat MT work at IBM

“AI winter” rule-based, expert systems
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?

(a) example word alignment

(b) example phrase alignment

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><strong>Newswire</strong></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><strong>Wikipedia</strong></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)
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 to 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)
Outline of the Course

### ML and structured prediction for NLP

- Neural Networks
- Semantics

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

<table>
<thead>
<tr>
<th>Date</th>
<th>Topics (tentative and subject to change)</th>
<th>Readings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/14/2021</td>
<td>first day of class</td>
<td></td>
</tr>
<tr>
<td>1/18/2021</td>
<td>No class - MLK national holiday</td>
<td></td>
</tr>
<tr>
<td>1/20/2021</td>
<td>Course Overview - 1st lecture</td>
<td>J+M 1</td>
</tr>
<tr>
<td>1/25/2021</td>
<td>Binary Classification (naive bayes and logistic regression)</td>
<td>J+M 4, Eisenstein 2.0-2.5, 4.1-4.3-4.5,</td>
</tr>
<tr>
<td>1/27/2021</td>
<td>Multiclass Classification</td>
<td>J+M 5, Eisenstein 4.2</td>
</tr>
<tr>
<td>2/1/2021</td>
<td>Neural Networks (feedforward networks)</td>
<td>Eisenstein 3.1-3.3, J+M 7.1-7.4</td>
</tr>
<tr>
<td>2/3/2021</td>
<td>Neural Networks (back propagation)</td>
<td>Eisenstein 3.1-3.3, J+M 7.1-7.4</td>
</tr>
<tr>
<td>2/8/2021</td>
<td>PyTorch Tutorial, Sequence Models</td>
<td>J+M 8</td>
</tr>
<tr>
<td>2/10/2021</td>
<td>Viterbi Algorithm</td>
<td>Eisenstein 7.0-7.4</td>
</tr>
<tr>
<td>2/15/2021</td>
<td>Conditional Random Fields</td>
<td>Eisenstein 7.5, 8.3</td>
</tr>
<tr>
<td>2/17/2021</td>
<td>N-gram Language Models</td>
<td></td>
</tr>
<tr>
<td>2/22/2021</td>
<td>Word Embeddings</td>
<td>Eisenstein 3.3-4, 14.5, 14.6, J+M 6</td>
</tr>
<tr>
<td>2/24/2021</td>
<td>Recurrent Neural Networks</td>
<td>J+M 9, Goldberg 10.11</td>
</tr>
<tr>
<td>3/1/2021</td>
<td>Convolutional Neural Networks</td>
<td>Goldberg 9, Eisenstein 3.4, 7.6</td>
</tr>
<tr>
<td>3/3/2021</td>
<td>Statistical Machine Translation</td>
<td>Eisenstein 18.1, 18.2</td>
</tr>
<tr>
<td>3/8/2021</td>
<td>(Guest Lecture)</td>
<td></td>
</tr>
<tr>
<td>3/10/2021</td>
<td>Sequence-to-Sequence Model</td>
<td>J+M 10</td>
</tr>
<tr>
<td>3/15/2021</td>
<td>Attention and Copy Mechanism</td>
<td>Eisenstein 18.3, 18.4</td>
</tr>
<tr>
<td>3/17/2021</td>
<td>Question Answering / Reading Comprehension</td>
<td>SQuAD, BiDAF</td>
</tr>
<tr>
<td>3/17/2021</td>
<td>Withdrawal deadline</td>
<td></td>
</tr>
<tr>
<td>3/22/2021</td>
<td>Parsing</td>
<td></td>
</tr>
<tr>
<td>3/24/2021</td>
<td>No class - mid-semester break</td>
<td></td>
</tr>
<tr>
<td>3/29/2021</td>
<td>Neural Machine Translation</td>
<td>Google NMT</td>
</tr>
<tr>
<td>3/31/2021</td>
<td>Transformer Model</td>
<td>Attention is all you need</td>
</tr>
<tr>
<td>4/5/2021</td>
<td>Generation (Guest Lecture)</td>
<td></td>
</tr>
<tr>
<td>4/7/2021</td>
<td>Information Extraction</td>
<td>Eisenstein 13, 17</td>
</tr>
<tr>
<td>4/12/2021</td>
<td>Dialog (Guest Lecture)</td>
<td></td>
</tr>
<tr>
<td>4/14/2021</td>
<td>Pre-trained Language Models / BERT</td>
<td></td>
</tr>
<tr>
<td>4/19/2021</td>
<td>Computational Social Science (Guest Lecture)</td>
<td></td>
</tr>
<tr>
<td>4/21/2021</td>
<td>Speech Recognition</td>
<td></td>
</tr>
<tr>
<td>4/26/2021</td>
<td>Final class day</td>
<td></td>
</tr>
</tbody>
</table>
ACL 2019 conference

<table>
<thead>
<tr>
<th>Area</th>
<th>Long submissions</th>
<th>Accepts</th>
<th>Accept rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Applications</td>
<td>65</td>
<td>14</td>
<td>28.8</td>
</tr>
<tr>
<td>2. Dialogue and Interactive Systems</td>
<td>126</td>
<td>38</td>
<td>30.2</td>
</tr>
<tr>
<td>3. Discourse and Pragmatics</td>
<td>33</td>
<td>7</td>
<td>21.2</td>
</tr>
<tr>
<td>4. Document Analysis</td>
<td>48</td>
<td>8</td>
<td>16.7</td>
</tr>
<tr>
<td>5. Generation</td>
<td>96</td>
<td>32</td>
<td>33.3</td>
</tr>
<tr>
<td>6. Information Extraction and Text Mining</td>
<td>155</td>
<td>37</td>
<td>23.9</td>
</tr>
<tr>
<td>7. Linguistic Theories, Cognitive Modeling and Psycholinguistics</td>
<td>39</td>
<td>9</td>
<td>23.1</td>
</tr>
<tr>
<td>10. Multidisciplinary and Area Chair COI</td>
<td>69</td>
<td>21</td>
<td>30.4</td>
</tr>
<tr>
<td>11. Multilinguality</td>
<td>43</td>
<td>11</td>
<td>25.6</td>
</tr>
<tr>
<td>12. Phonology Morphology and Word Segmentation</td>
<td>26</td>
<td>7</td>
<td>26.9</td>
</tr>
<tr>
<td>13. Question Answering</td>
<td>99</td>
<td>32</td>
<td>32.3</td>
</tr>
<tr>
<td>14. Resources and Evaluation</td>
<td>70</td>
<td>26</td>
<td>37.1</td>
</tr>
<tr>
<td>15. Sentence-level semantics</td>
<td>69</td>
<td>14</td>
<td>20.3</td>
</tr>
<tr>
<td>16. Sentiment Analysis and Argument Mining</td>
<td>91</td>
<td>24</td>
<td>26.4</td>
</tr>
<tr>
<td>17. Social Media</td>
<td>51</td>
<td>14</td>
<td>27.5</td>
</tr>
<tr>
<td>18. Summarization</td>
<td>48</td>
<td>11</td>
<td>22.9</td>
</tr>
<tr>
<td>19. Tagging chunking Syntax and Parsing</td>
<td>50</td>
<td>17</td>
<td>34.0</td>
</tr>
<tr>
<td>20. Textual Inference and Other Areas of Semantics</td>
<td>44</td>
<td>16</td>
<td>36.4</td>
</tr>
<tr>
<td>21. Vision Robotics Multimodal Grounding and Speech</td>
<td>56</td>
<td>20</td>
<td>35.7</td>
</tr>
<tr>
<td>22. Word-level Semantics</td>
<td>78</td>
<td>20</td>
<td>25.6</td>
</tr>
</tbody>
</table>
ACL’19 at a Glance