CS388: Natural Language Processing
Lecture 1: Introduction

Greg Durrett
Lecture: Tuesdays and Thursdays 9:30am - 10:50am


Piazza: link on the course website

My office hours: Wednesday 10am-noon, GDC 3.420

TA: Jifan Chen; Office hours:
  - Monday + Tuesday, 1pm-2pm GDC 1.302
Course Requirements

- 391L Machine Learning (or equivalent)
- 311 or 311H Discrete Math for Computer Science (or equivalent)
- Python experience
- Additional prior exposure to probability, linear algebra, optimization, linguistics, and NLP useful but not required
Enrollment

- I want everyone to be able to take this class!

- Mini1 is out now (due September 11):
  - 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 projects, which are smaller-scale than the final project)
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

resolve references
do computation
recognize predicate
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.

provide missing context
compress text
paraphrase to provide clarity
Trump Pope family watch a hundred years a year in the White House balcony
NLP Analysis Pipeline

Text Analysis
- 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 text?

Text

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

WORK_OF_ART

Mission Impossible film

Trees

λx. flight(x) ∧ dest(x)=Miami

flights to Miami

I eat cake with icing
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?)
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 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

slide credit: Dan Klein
Language is **Really** Ambiguous!

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

  - **il fait vraiment beau**

    - 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

- 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?

- O2
- O1
- O3

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)
     \[ C_b = \text{John}; \quad C_f = \{\text{John}\} \]

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

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

  e. He called John at 5 AM on Friday last week. (he = Mike)
     \[ C_b = \text{Mike}; \quad C_f = \{\text{Mike, John}\} \quad \text{(SHIFT)} \]

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
2000: Ratnaparkhi tagger
2010: Neural
2018: Semi-sup, structured prediction

Penn treebank

Collins vs. Charniak parsers

Unsup: topic models, grammar induction

Sup: SVMs, CRFs, NER, Sentiment

Neural
Structured Prediction

- All of these techniques are data-driven! Some data is naturally occurring, but may need to label.
- 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)
Manually constructed grammars -> EM-induced grammars -> basic grammars + features -> ...

- Klein and Manning (2003)
- Petrov et al. (2006)
- Hall, Durrett, Klein (2014)
The yield on the benchmark issue rose to 10% from 5%

- No grammars!

Sutskever et al. (2015), Bahdanau et al. (2014)
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>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)
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 nets

Syntax/semantics

Applications: MT, IE, summarization, dialogue, etc.
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?
- Make you a “producer” rather than a “consumer” of NLP tools
  - The four assignments should teach you what you need to know to understand nearly any system in the literature (e.g.: state-of-the-art NER system = project 1 + mini 2, basic MT system = project 2)
Assignments

- Two minis (10% each), two projects (20% each)
  - Implementation-oriented, with an open-ended component to each
  - Mini 1 (classification) is out NOW
  - ~2 weeks per assignment, 5 “slip days” for automatic extensions

- Grading:
  - Minis: 80% for reaching the performance threshold, 20% writeup
  - Projects: 60% for reaching the performance threshold, 20% writeup, 20% extension

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!
Assignments

- Final project (40%)
  - Groups of 2 preferred, 1 is possible
  - (Brief!) proposal to be approved by me
  - Written in the style and tone of an ACL paper
1. Fill in: I am a [CS / ____] [PhD / masters / undergrad] in year [1 2 3 4 5+]
2. Which of the following have you learned in a class?
   1. Bayes’ Rule
   2. SVMs
   3. Expectation maximization
   4. RNNs
3. Which of the following have you used?
   1. Python
   2. numpy/scipy/scikit-learn
   3. Tensorflow/(Py)Torch/Theano
4. Fill in: Assuming I can enroll, my probability of taking this class is X%
5. One interesting fact about yourself, or what you like to do in your spare time