CS388: Natural Language Processing
Lecture 1: Introduction

Greg Durrett

TEXAS
The University of Texas at Austin
Lecture: Tuesdays and Thursdays 12:30pm - 1:45pm

Course website:  

Piazza: link on the course website

My office hours: Office hours: Wednesday 4pm, Thursday 2pm

TA: Uday Kusupati. Office hours: Monday 12pm-1pm, Tuesday 11am-12pm, 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
We’ll get as many people in as we can

Mini1 is out now (due September 10), please look at it soon

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)

Other NLP offerings:

CS378 (ugrad course, taught by me in the spring)

LIN 393 (taught by Jessy Li): NLP with minimal supervision
What’s the goal of NLP?

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

Example:
- Dialogue systems: go to meet

Siri, what’s the most valuable American company?

Apple

Who is its CEO?

Tim Cook
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.

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.
Trump Pope family watch a hundred years a year in the White House balcony
NLP Analysis Pipeline

- NLP is about building these pieces!
- All of these components are modeled with statistical approaches trained with machine learning
How do we represent language?

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

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

\[ \text{flights to Miami} \]
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 advocated violence

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

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

>5 datasets in the last two years examining this problem and commonsense reasoning

Referential ambiguity
Language is Ambiguous!

- Syntactic and semantic ambiguities: parsing needed to resolve these, but need context to figure out which parse is correct

Example credit: Dan Klein
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
  - 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

  - Various diagrams and words:
    - DOJ greenlights Disney - Fox merger
    - Department of Justice
    - Metaphor; "approves"

- What is a green light? How do we understand what "green lighting" does?

- Need commonsense knowledge
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 \( O_2 \)?

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.
   \( C_b = \text{John}; \ C_f = \{\text{John}\} \)

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

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

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

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

Centering Theory
Grosz et al. (1995)
What techniques do we use?
(to combine data, knowledge, linguistics, etc.)
A brief history of (modern) NLP

- **1980**: Largely rule-based, expert systems
- **1980**: Earliest statistical MT work at IBM
- **1990**: Penn treebank
- **1990**: Ratnaparkhi tagger
- **2000**: Collins vs. Charniak parsers
- **2000**: Unsup: topic models, grammar induction
- **2000**: Sup: SVMs, CRFs, NER, Sentiment
- **2010**: Semi-sup, structured prediction
- **2019**: Pretraining

Diagram notes:
- Early work at IBM
- Ratnaparkhi tagger: NNP VBZ
- Sentence structure: S (NP VP)
Supervised vs. Unsupervised

- Supervised techniques work well on very little data (even neural networks)
  - annotation (two hours!)
  - unsupervised learning
  - better system!

- Fully unsupervised techniques have fallen out of favor

“Learning a Part-of-Speech Tagger from Two Hours of Annotation”
Garrette and Baldridge (2013)
Training is supervised but models still rely less on manual structure

- Klein and Manning (2003)
  Manually-constructed grammars

- Petrov et al. (2006)
  Induced grammars

- Hall, Durrett, Klein (2014)
  Basic grammar + features
The yield on the benchmark issue rose to 10% from 5%

No grammars at all!

Sutskever et al. (2015), Bahdanau et al. (2014)
Interpretability

- Hard to analyze why these errors happen in neural models (but people are trying)
- Models with more manual structure might be more interpretable
Pretraining

- Language modeling: predict the next word in a text $P(w_i | w_1, \ldots, w_{i-1})$

$P(w | \text{I want to go to}) = 0.01 \text{ Hawaiʻi}$

$0.005 \text{ LA}$

$0.0001 \text{ class}$

: use this model for other purposes

$P(w | \text{the acting was horrible, I think the movie was}) = 0.1 \text{ bad}$

$0.001 \text{ good}$

- Model understands some sentiment?

- Train a neural network to do language modeling on massive unlabeled text, fine-tune it to do {tagging, sentiment, question answering, ...}

Peters et al. (2018), Devlin et al. (2019)
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

<table>
<thead>
<tr>
<th>Aug 29</th>
<th>Introduction [4pp]</th>
<th>Mini1 out</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sept 3</td>
<td>Binary classification</td>
<td>Eisenstein 2.0-2.5, 4.2-4.4.1, JM 4, JM 5.0-5.5</td>
</tr>
<tr>
<td>Sept 5</td>
<td>Multiclass classification</td>
<td>Eisenstein 4.2, JM 5.6, Structured SVM secs 1-2</td>
</tr>
<tr>
<td>Sept 10</td>
<td>Sequence Models 1: HMMs (Guest Lecture: Ray Mooney)</td>
<td>Eisenstein 7.0-7.4, 8.1, JM 8, Manning POS, Viterbi algorithm lecture note</td>
</tr>
<tr>
<td>Sept 12</td>
<td>Sequence Models 2: CRFs</td>
<td>Eisenstein 7.5, 8.3, Sutton CRFs 2.3, 2.6.1, Wallach CRFs tutorial, Illinois NER</td>
</tr>
<tr>
<td>Sept 17</td>
<td>NN1: Feedforward</td>
<td>Eisenstein 3.0-3.3, Goldberg 1-4, 6, NLP with FFNNs, DANs</td>
</tr>
<tr>
<td>Sept 19</td>
<td>NN2: Word embeddings</td>
<td>Eisenstein 3.3.4, 14.5-14.6, JM 6, Goldberg 5, word2vec, Levy, GloVe, Dropout</td>
</tr>
<tr>
<td>Sept 24</td>
<td>NN3: RNNs</td>
<td>JM 9.1-9.4, Goldberg 10-11, Karpathy</td>
</tr>
<tr>
<td>Sept 26</td>
<td>NN4: Language Modeling and Pretraining</td>
<td>Eisenstein 6, JM 9.2.1, ELMo</td>
</tr>
<tr>
<td>Oct 1</td>
<td>NN5: Interpretability/CNNs/Neural CRFs/etc.</td>
<td>Mini2 out</td>
</tr>
</tbody>
</table>

**ML and structured prediction for NLP**

**Neural nets (this part is still in flux)**
# Outline: Syntax + Semantics

<table>
<thead>
<tr>
<th>Date</th>
<th>Topic</th>
<th>Reading Material</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oct 3</td>
<td>Trees 1: Constituency, PCFGs</td>
<td>Eisenstein 10.0-10.5, JM 12.1-12.6, 12.8, Structural, Lexicalized, State-split</td>
<td></td>
</tr>
<tr>
<td>Oct 10</td>
<td>Trees 3: Dependency Parsers</td>
<td>Eisenstein 11.3, JM 13.4, Parsey, Huang 2</td>
<td></td>
</tr>
<tr>
<td>Oct 15</td>
<td>Semantics 1</td>
<td>Eisenstein 12, Zettlemoyer, Berant</td>
<td>FP proposal due</td>
</tr>
<tr>
<td>Oct 17</td>
<td>Semantics 2 / Seq2seq 1</td>
<td>Seq2seq, Jia</td>
<td>Proj 2 out</td>
</tr>
<tr>
<td>Oct 22</td>
<td>Seq2seq 2: Attention and Pointers</td>
<td>Attention, Luong Attention, Transformer</td>
<td></td>
</tr>
<tr>
<td>Date</td>
<td>Topic</td>
<td>Meetings/Activities</td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>-------------------------------------------------</td>
<td>----------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Oct 24</td>
<td>Machine Translation 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oct 29</td>
<td>Machine Translation 2 / Transformers</td>
<td>BERT, RoBERTa</td>
<td></td>
</tr>
<tr>
<td>Oct 31</td>
<td>Pretrained Transformers / BERT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nov 5</td>
<td>Information Extraction / SRL</td>
<td>Proj2 due</td>
<td></td>
</tr>
<tr>
<td>Nov 7</td>
<td>Question Answering 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nov 12</td>
<td>Question Answering 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nov 14</td>
<td>Dialogue</td>
<td>RNN chatbots, Diversity, Goal-oriented, Latent Intention, QA-as-dialogue</td>
<td></td>
</tr>
<tr>
<td>Nov 19</td>
<td>Summarization</td>
<td>Eisenstein 19, MMR, Gillick, Sentence compression, SummaRuNNER, Pointer</td>
<td></td>
</tr>
<tr>
<td>Nov 21</td>
<td>Multilinguality and morphology</td>
<td>Xlingual POS, Xlingual parsing, Xlingual embeddings</td>
<td></td>
</tr>
<tr>
<td>Nov 26</td>
<td>Wrapup + Ethics</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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 2019?
- 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
  - 1 week for minis, ~2 weeks per project, 5 “slip days” for automatic extensions

- Grading:
  - Minis: largely graded based on code performance
  - Projects: graded on a mix of code performance, writeup, 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 by the midpoint of the semester (October 15)
  - Written in the style and tone of an ACL paper
A climate conducive to learning and creating knowledge is the right of every person in our community. Bias, harassment and discrimination of any sort have no place here. If you notice an incident that causes concern, please contact the Campus Climate Response Team: diversity.utexas.edu/ccrt

The College of Natural Sciences is steadfastly committed to enriching and transformative educational and research experiences for every member of our community. Find more resources to support a diverse, equitable and welcoming community within Texas Science and share your experiences at cns.utexas.edu/diversity
Survey (Optional)

1. Name
2. Fill in: I am a [CS / _____] [PhD / masters / undergrad] in year [1 2 3 4 5+]
3. Write one reason you want to take this class or one thing you want to get out of it
4. One interesting fact about yourself, or what you like to do in your spare time