CS395T: Structured Models for NLP
Lecture 1: Introduction

Greg Durrett

Administrivia

› Lecture: Tuesdays and Thursdays 9:30am - 10:50am
› Course website: http://www.cs.utexas.edu/~gdurrett/courses/fa2017-cs395t.shtml
› Piazza: https://piazza.com/utexas/fall2017/cs395t/home
› My office hours: Wednesday 10am-noon, GDC 3.420
› TA: Ye Zhang; Office hours:
  › Tuesday 2pm-3pm GDC 1.302 Desk 2
  › Thursday 2pm-3pm, GDC 1.302 Desk 1 (until 2:30), Desk 4 (2:30 onwards)

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!
› Priority ordering:
  › CS grad students
  › Other grad students
  › CS undergrads who have satisfied the prerequisites
  › Other undergrads who have satisfied the prerequisites
  › Other undergrads
**What’s the goal of NLP?**

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

**Automatic Summarization**

- Compress text
- Provide missing context
- 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

**Machine Translation**

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

**Textual Entailment**

- 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?
Language is Ambiguous!

- Hector Levesque (2011): “Winograd schema challenge” (named after Terry Winograd, the creator of SHRDLU)

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

- This is so complicated that it’s an AI challenge problem! (AI-complete)
- Can try to use the web to learn pragmatics, but that’s not giving us a deep understanding of text

Language is Really Ambiguous!

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

  Sentence        Syntactic parser        Exponential number
  “fish”          fish                    gh as in tough
  Phish           o as in women          ti as in motion

- Combinatorially many possibilities, many you won’t even register as ambiguities, but systems still have to resolve them!
- So our goal (analyze text) is harder than we thought…how do we do it?

A brief history of (modern) NLP

- “AI winter” rule-based, expert systems
- “AI winter” worked at IBM
- “AI winter” tagger
- Ratnaparkhi parser
- Ratnaparkhi tagger
- Earliest stat MT
- Earliest stat MT
- Ratnaparkhi tagger
- Ratnaparkhi tagger
- “AI winter” tagger
- “AI winter” tagger
- Earliest stat MT

- Unsup: topic models, grammar induction
- Sup: SVMs, CRFs, NER, sentiment
- Semi-sup, structured prediction
- 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.
  
  [Image of a head with annotation (two hours!)]

- Unsupervised learning.
  
  [Image of a head with a better system!]

- Even neural nets can do pretty well!
- Balance tradeoff of data/algorithms/compute

Garrette and Baldridge (2013)

---

Less Manual Structure?

The yield on the benchmark issue rose to 10% from 5%

[Diagram of LSTM]

(S (NP (NP (DT The) (NN yield ...)]

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

---

Less Manual Structure?

Bahdanau et al. (2014)

(a) example word alignment

(b) example phrase alignment

DeNero et al. (2008)
Less Manual Structure?

Translate

| English | French | Spanish | Chinese - detected |

特朗普通家人在白宫阳台观看百年一遇日全食

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

Maybe manual structure would help...

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?

Moosavi and Strube (2017)

Where are we?

- Solving problems with text requires analyzing text
- Many possibilities: rule-based systems, CRFs, neural networks, ...
- Knowing which of these to use requires understanding dataset size, problem complexity, and a lot of tricks!
- What do all of these models have in common? What do they need to capture in order to be successful?

Break!

What’s important?

- High-capacity models + data!

<table>
<thead>
<tr>
<th>SOURCE</th>
<th>Cela constituierait une solution transitoire qui permettrait de conclure 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] [constituera] [solution] [transitoire] [who] [permettra] [licences] [to] [term] [to] [a] [charter] [to] [value] [contrainante]</td>
</tr>
<tr>
<td>10x DATA</td>
<td>[this] [would be] [a transitional solution] [which] [would] [eventually] [lead to] [a] [charter] [legally binding]</td>
</tr>
<tr>
<td>100x DATA</td>
<td>[this] [would be] [a transitional solution] [which] [would] [eventually] [lead to] [a] [binding charter]</td>
</tr>
</tbody>
</table>

slide credit: Dan Klein
What's important?

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

On Sept. 1, 1715 Louis XIV died in this city, site of a fabulous palace he built.

Answer: What is Versailles?

What's important?

- 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's important?

- Multitask interactions: recognize constraints to be more statistically efficient (and humanlike!) in our reasoning

Dell is headquartered just outside Austin. The company...

Durrett and Klein (2014)

What's important?

- 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.
  \( C_0 = \text{John}; C_I = \{\text{John}\} \)

b. He cannot find anyone to take over his responsibilities. (he = John)
  \( C_0 = \text{John}; C_I = \{\text{John, Mike}\} \) (CONTINUE)

c. He called up Mike yesterday to work out a plan. (he = John)
  \( C_0 = \text{John}; C_I = \{\text{John, Mike}\} \) (CONTINUE)

d. Mike has annoyed him a lot recently.
  \( C_0 = \text{John}; C_I = \{\text{Mike, John}\} \) (RETAIN)

e. He called John at 5 AM on Friday last week. (he = Mike)
  \( C_0 = \text{Mike}; C_I = \{\text{Mike, John}\} \) (SHIFT)

Centering Theory
Grosz et al. (1995)
How do we build systems to do all this?

- **Structured statistical models**
- **Structured:** lets us incorporate cross-task constraints, inductive biases from linguistics, knowledge, etc.
- **Statistical:** harness the power of data to do really large-scale pattern recognition and learn from labeled + unlabeled data + interaction with the world

Outline of the Course

- First half: structured prediction
  - Machine learning basics
  - Sequences, trees
  - Inference, learning
- Second half: deep learning
  - RNNs/LSTMs, convolutional networks
  - Word representations
  - Inference, learning

NLP vs. Computational Linguistics

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

---

Hamilton et al. (2016)

---

Bamman, O’Connor, Smith (2013)
Course Goals

- Cover structured machine learning approaches to NLP
- Show connections between structured algorithms: generative and discriminative, margin and likelihood, neural and linear, etc.: these are all closely related
- Dissect the pieces of these structured models: modeling, inference, learning
- Make you a “producer” rather than a “consumer” of NLP tools
- Expose you to classic problems in NLP

Assignments

- Three projects (16.6% each = 50%)
  - Implementation-oriented, open-ended component to each
  - First will be out on 9/12
  - 2-page writeup with statement of what you did
  - ~2 weeks per project, 7 “slip days” for automatic extensions
- Grading: 10-point scale
  - 6 points for minimal code completion
  - 1 point for minimal extension
  - 1 point for minimal 2-page writeup
  - 2 points for better extension, better writeup

Assignments

- Final project (50%)
  - Groups of 1-2
  - (Brief!) proposal to be approved by me
  - Written in the style and tone of an ACL paper
  - Same 10-point grading scheme, 8 points for minimal completion of proposed work

Survey

1. Fill in: I am a [CS / linguistics / other] [grad / 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. HMMs
   4. EM
   5. Part-of-speech tagging
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%