Natural Language Processing

Dan Klein, GSI: Rudy Corona
UC Berkeley

Logistics

Enrollment
- Class is "full" but we’re expecting space
- Won’t know capacity for sure until P0
- We’ll announce as we go

Course expectations
- Readings, lectures, ~4 projects
- No sections, no exams
- Engagement with the course
- Workload will be high, self-direction
- Patience: class is under construction

Requirements
- ML: A-level mastery, e.g. CS189
- PL: Ready to work in Python/Torch (on Colab)
- NL: Care a lot about natural language

Resources and Readings

Resources
- Webpage (syllabus, readings, slides, links)
- Piazza (course communication)
- Gradescope (submission and grades)
- Compute via Colab notebooks
- Limited webcast policy

Readings (see webpage)
- Individual papers will be linked
- Optional text: Jurafsky & Martin, 3rd (more NL)
- Optional text: Eisenstein (more ML)
Projects and Compute

- **Projects**:
  - P0: Warm-up and calibration
  - P1: Language Models
  - P2: Machine Translation
  - P3: Syntax and Parsing
  - P4: Semantics and Grounding

- **Infrastructure**:
  - Python / PyTorch
  - Compute via Colab notebooks
  - Grading via Gradescope

What is NLP?

- **Goal: Deep Understanding**
  - Requires context, linguistic variation, meanings...

- **Reality: Shallow Matching**
  - Requires robustness and scale
  - Amazing successes, but fundamental limitations

NLP History

- 1950: Neural nets?
- 1960: Weaver on MT
- 1970: Bell Labs ASR
- 1980: ALPAC kills MT
- 1990: Rule-based MT
- 2000: Statistical MT
- 2010: Neural MT
- 2020: Pretraining

Natural Language Processing

- **NLP**
  - Goal: Deep Understanding
  - Reality: Shallow Matching

- grep
Transforming Language

Speech Systems

- **Automatic Speech Recognition (ASR)**
  - Audio in, text out
  - SOTA: <1% error for digit strings, 6% for conversational speech, still >20% hard accents

  "Speech Lab"

- **Text to Speech (TTS)**
  - Text in, audio out
  - SOTA: nearly perfect aside from prosody

Machine Translation

- Translate text from one language
- Challenges:
  - What’s the mapping? (learning to translate)
  - How to make it efficient? (fast translation search)
  - Fluency (next class) vs fidelity (later)

Example: 

Machine Translation

- Disney décide de changer le nom du légendaire studio de cinéma Fox
- Disney's decision to change the name of the legendary film studio

Example: 

Google Translate 2020
Spoken Language Translation

Summarization

- Condensing documents
  - Single or multiple
  - Extractive or synthetic
  - Aggregative or representative

- Very context-dependent!
- An example of analysis with generation

Search, Questions, and Reasoning

Understanding Language
Jeopardy!

US Cities: its largest airport is named for a World War II hero, its second largest, for a World War II battle.

Question Answering: Watson

Language Comprehension?

“"The rock was still wet. The animal was glowing, like it was still breathing," revealed Bao Xingang. Bao discovered the animal today, while surveying fossils on a field trip. Before he could identify the rock, his field trip leader rushed over and shouted "You're on the right track!". Bao's instinct told him it was a member of the genus Homo, but he wasn't sure which Homo. He quickly named it Homo hominid. However, for weeks it was 15 million years old than in Canadian remains.

It can be inferred that Bao Xingang's "hands began to shake", because he was:

(a) afraid that he might lose the fossil
(b) worried about the implications of his finding
(c) concerned that he might not get credit for his work
(d) uncertain about the authenticity of the fossil
(e) excited about the magnitude of his discovery
Interactive Language

Example: Virtual Assistants

- Virtual assistants do
- Speech recognition
- Language analysis
- Dialog processing
- Text to speech
- ... and back-end actions!

Conversations with Devices?

ELIZA

Social AIs and Chatbots

Microsoft’s XiaoIce

Chatbot Competitions!

- Alexa Prize competition to build chatbots that keep users engaged
- First winner in 2017: UW’s Sounding Board (Fang, Cheng, Holtzman, Ostendorf, Sap, Clark, Choi)
- Compare to the Turing test (eg Loebner Prize) where the goal is to fool people

SoundingBoard Example

<table>
<thead>
<tr>
<th>User</th>
<th>Sounding Board</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hey!</td>
<td>Hi! How about a chat?</td>
</tr>
<tr>
<td>Can I help you?</td>
<td>Yeah, let’s chat!</td>
</tr>
<tr>
<td>What do you want to talk about?</td>
<td>Let’s talk about interesting things!</td>
</tr>
<tr>
<td>Okay. What would you like to talk about?</td>
<td>Right now, I’m interested in interesting things.</td>
</tr>
<tr>
<td>Do you have any particular interests?</td>
<td>Well, I don’t really have a particular interest.</td>
</tr>
<tr>
<td>Do you have any particular interests in this project?</td>
<td>I don’t really have any particular interest in this project.</td>
</tr>
<tr>
<td>What do you want to talk about?</td>
<td>I don’t really have any particular interest in this project.</td>
</tr>
<tr>
<td>Okay. What would you like to talk about?</td>
<td>Right now, I’m interested in interesting things.</td>
</tr>
<tr>
<td>Can I help you?</td>
<td>Yeah, let’s chat!</td>
</tr>
<tr>
<td>What do you want to talk about?</td>
<td>Let’s talk about interesting things!</td>
</tr>
<tr>
<td>Okay. What would you like to talk about?</td>
<td>Right now, I’m interested in interesting things.</td>
</tr>
<tr>
<td>Do you have any particular interests?</td>
<td>Well, I don’t really have a particular interest.</td>
</tr>
<tr>
<td>Do you have any particular interests in this project?</td>
<td>I don’t really have any particular interest in this project.</td>
</tr>
<tr>
<td>What do you want to talk about?</td>
<td>I don’t really have any particular interest in this project.</td>
</tr>
<tr>
<td>Okay. What would you like to talk about?</td>
<td>Right now, I’m interested in interesting things.</td>
</tr>
<tr>
<td>Can I help you?</td>
<td>Yeah, let’s chat!</td>
</tr>
<tr>
<td>What do you want to talk about?</td>
<td>Let’s talk about interesting things!</td>
</tr>
<tr>
<td>Okay. What would you like to talk about?</td>
<td>Right now, I’m interested in interesting things.</td>
</tr>
<tr>
<td>Do you have any particular interests?</td>
<td>Well, I don’t really have a particular interest.</td>
</tr>
<tr>
<td>Do you have any particular interests in this project?</td>
<td>I don’t really have any particular interest in this project.</td>
</tr>
<tr>
<td>What do you want to talk about?</td>
<td>I don’t really have any particular interest in this project.</td>
</tr>
</tbody>
</table>

Sounding Board’s Architecture

Convolution
Fully connected
L1 256x256
L2 128x128
L3 64x64
L4 32x32
L5 1
F6 (Output)
What is Nearby NLP?

- **Computational Linguistics**
  - Using computational methods to learn more about how language works
  - We end up doing this and using it

- **Cognitive Science**
  - Figuring out how the human brain works
  - Includes the bits that do language
  - Humans: the only working NLP prototype!

- **Speech Processing**
  - Mapping audio signals to text
  - Traditionally separate from NLP, converging

Example: NLP Meets CL

- Example: Language change, reconstructing ancient forms, phylogenies
  - Just one example of the kinds of linguistic models we can build
Why is Language Hard?

Problem: Ambiguity

- Headlines:
  - Enraged Cow Injures Farmer with Ax
  - Teacher Strikes Idle Kids
  - Hospitals Are 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

- Why are these funny?

What Do We Need to Understand Language?
We Need Representation: Linguistic Structure

Example: Syntactic Analysis

Hurricane Emily howled toward Mexico’s Caribbean coast on Sunday packing 135 mph winds and torrential rain and causing panic in Cancun, where frightened tourists squeezed into musty shelters.

We Need Data

SOURCE: Cela constituait une solution transitoire qui permettrait de conduire à terme à une charte à valeur contraignante.

HUMAN: That would be an interim solution which would make it possible to work towards a binding charter in the long term.

1x DATA: That would be an interim solution which would eventually lead to a legally binding charter.

10x DATA: That would be an interim solution which would eventually lead to a legally binding charter.

100x DATA: That would be an interim solution which would eventually lead to a legally binding charter.

1000x DATA: That would be an interim solution which would eventually lead to a legally binding charter.

We Need Lots of Data: MT

SOURCE: That would be an interim solution which would make it possible to work towards a binding charter in the long term.

HUMAN: That would be an interim solution which would eventually lead to a legally binding charter.

1x DATA: That would be an interim solution which would eventually lead to a legally binding charter.

10x DATA: That would be an interim solution which would eventually lead to a legally binding charter.

100x DATA: That would be an interim solution which would eventually lead to a legally binding charter.

1000x DATA: That would be an interim solution which would eventually lead to a legally binding charter.
We Need Models: Data Alone Isn’t Enough!

CLASSIC SOUPS

<table>
<thead>
<tr>
<th>#</th>
<th>Soup</th>
<th>Price</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>57</td>
<td>Hominy Chicken Soup (Chicken, Celery,</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hominy)</td>
<td>1.30</td>
<td>2.75</td>
</tr>
<tr>
<td>58</td>
<td>Chicken Rice Soup</td>
<td>1.85</td>
<td>2.75</td>
</tr>
<tr>
<td>59</td>
<td>Chicken Noodle Soup</td>
<td>1.85</td>
<td>2.75</td>
</tr>
<tr>
<td>60</td>
<td>卡通面汤 (Miso Soup)</td>
<td>1.30</td>
<td>2.75</td>
</tr>
<tr>
<td>61</td>
<td>Tomato Cream Soup</td>
<td>1.35</td>
<td>2.50</td>
</tr>
<tr>
<td>62</td>
<td>Regular World Soup</td>
<td>1.10</td>
<td>2.50</td>
</tr>
<tr>
<td>63</td>
<td>Beef &amp; Barley Soup</td>
<td>1.10</td>
<td>2.50</td>
</tr>
<tr>
<td>64</td>
<td>UX Vegetable Soup</td>
<td>1.10</td>
<td>2.50</td>
</tr>
<tr>
<td>65</td>
<td>UX Vegetable Soup</td>
<td>1.10</td>
<td>2.50</td>
</tr>
<tr>
<td>67</td>
<td>Chicken Corn Chowder Soup</td>
<td>NA</td>
<td>3.50</td>
</tr>
<tr>
<td>68</td>
<td>Cash Meat Corn Chowder Soup</td>
<td>NA</td>
<td>3.50</td>
</tr>
<tr>
<td>69</td>
<td>Seafood Soup</td>
<td>NA</td>
<td>3.50</td>
</tr>
</tbody>
</table>

Proper Nouns (NNP)

<table>
<thead>
<tr>
<th>#</th>
<th>Noun</th>
<th>Abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>Oct</td>
<td>Oct.</td>
</tr>
<tr>
<td>14</td>
<td>Nov</td>
<td>Nov.</td>
</tr>
<tr>
<td>14</td>
<td>Sept</td>
<td>Sept.</td>
</tr>
<tr>
<td>12</td>
<td>John</td>
<td>John</td>
</tr>
<tr>
<td>12</td>
<td>Robert</td>
<td>Robert</td>
</tr>
<tr>
<td>12</td>
<td>James</td>
<td>James</td>
</tr>
<tr>
<td>2</td>
<td>J.</td>
<td>E.</td>
</tr>
<tr>
<td>2</td>
<td>E.</td>
<td>L.</td>
</tr>
<tr>
<td>1</td>
<td>Bush</td>
<td>Bush</td>
</tr>
<tr>
<td>1</td>
<td>Noriega</td>
<td>Noriega</td>
</tr>
<tr>
<td>1</td>
<td>Peters</td>
<td>Peters</td>
</tr>
<tr>
<td>3</td>
<td>York</td>
<td>York</td>
</tr>
<tr>
<td>3</td>
<td>Francisco</td>
<td>Francisco</td>
</tr>
<tr>
<td>3</td>
<td>San</td>
<td>San</td>
</tr>
<tr>
<td>3</td>
<td>Wall</td>
<td>Wall</td>
</tr>
</tbody>
</table>

Learning Latent Syntax

Personal Pronouns (PRP)

<table>
<thead>
<tr>
<th>#</th>
<th>Pronoun</th>
<th>Abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>it</td>
<td>them</td>
</tr>
<tr>
<td>1</td>
<td>it</td>
<td>It</td>
</tr>
<tr>
<td>2</td>
<td>they</td>
<td>he</td>
</tr>
<tr>
<td>2</td>
<td>they</td>
<td>He</td>
</tr>
</tbody>
</table>

We Need World Knowledge

- World knowledge: have access to information beyond the training data
- Department of Justice
- "approaches"
- What is a green light? How do we understand what “green lighting” does?
- Need commonsense knowledge

Data and Knowledge

- Classic knowledge representation worries: How will a machine ever know that...
  - Ice is frozen water?
  - Beige looks like this:
  - Chairs are solid?

Answers:

- 1980: write it all down
- 2000: get by without it
- 2020: learn it from data
Knowledge from Pretraining?

Language Modeling

We Need Grounding

Grounding: linking linguistic concepts to non-linguistic ones

Example: Grounded Dialog

When is my package arriving?

Friday!

Example: Grounded Dialog

What’s the most valuable American company?

Apple

Who is its CEO?

Tim Cook
Why is Language Hard?

- We Need:
  - Representations
  - Models
  - Data
  - Machine Learning
  - Scale
  - Efficient Algorithms
  - Grounding
- ... and often we need all these things at the same time

What is this Class?

- Three aspects to the course:
  - Linguistic Issues
    - What are the range of language phenomena?
    - What are the knowledge sources that let us disambiguate?
    - What representations are appropriate?
    - How do you know what to model and what not to model?
  - Modeling Methods
    - Increasingly sophisticated model structures
    - Learning and parameter estimation
    - Efficient inference: dynamic programming, search, sampling
  - Engineering Methods
    - Issues of scale
    - Where the theory breaks down (and what to do about it)
- We'll focus on what makes the problems hard, and what works in practice...

Class Requirements and Goals

- Class requirements:
  - Uses a variety of skills / knowledge:
    - Probability and statistics, graphical models (parts of cs281a)
    - Basic linguistics background (ling100)
    - Strong coding skills (Python, ML libraries)
    - Most people are probably missing one of the above
    - You will often have to work on your own to fill the gaps
- Class goals:
  - Learn the issues and techniques of modern NLP
    - Build realistic NLP tools
    - Be able to read current research papers in the field
    - See where the holes in the field still are!
- This semester: new projects, new topics, lots under construction!