Natural Language Processing

Introduction
Spring 2017

Based on slides from Chris Manning, Michael Collins, and Yoav Artzi
What is NLP?

- **Goal**: Develop methods for processing, analyzing and understanding the **structure** and **meaning** of language.

  ![Diagram](Diagram.png)

  - Language → NLP → Structure
  - Language → NLP → Meaning

- Many languages

- Many “levels”: transcribed, newswire, social media...
What is NLP?

• **Application:** Build systems that help people do stuff (with text):
  
  • Question answering, virtual assistants, translation, etc.
  
  • There is a lot language out there…
Applications are taking off!

• Search
• Ad matching
• Translation
• Sentiment analysis
• Speech recognition
• Chatbots
• Virtual assistants
• …
Levels of analysis in NLP

- **Phonology**: sounds that make up language
  - *daˈvɪd ben guˈrjoːn*
- **Morphology**: internal structure of words
  - ובבתיהם
- **Syntax**: structure of phrases, how words modify one another
  - *pretty little girl's school*
- **Semantics**: meaning of language in the world
  - *students on the back row*
- **Discourse**: relations between clauses and sentences
  - *I registered to the class on NLP because it is fascinating*
What’s so special about language

• Invented by humans for communication

• Learned from experience (!!)

• A symbolic/discrete system:
  
  • *table*: 📑
  
  • *piano*: 🎹

• that pre-dates logic…
Language is special

- Encoded via continuous signals:
  - Sounds
  - Gestures
  - Image (writing)
- Brain encodings are continuous
Why is it hard?

- ambiguity
- variability
- sparsity
- grounding
Ambiguity

“Finally, a computer that understands you like your mother”
(Ad, 1985)

1. The computer understands you as well as your mother understands you.
2. The computer understands that you like your mother.
3. The computer understands you as well as it understands your mother.

“Finally, a computer that understands your lie cured mother”
Syntactic ambiguity

Even short sentences have hundreds of analyses
Semantic ambiguity

- There’s a river that crosses every town

  - Quantification ambiguity (not syntactic)

- תולצת מטיגלת בואדי

  - Lexical, syntactic and semantic ambiguity
Ambiguity

• Headlines:
  • *Enraged Cow Injures Farmer with Ax*
  • *Ban on Nude Dancing on Governor’s Desk*
  • *Teacher Strikes Idle Kids*
  • *Hospitals Are Sued by 7 Foot Doctors*
  • *Iraqi Head Seeks Arms*
  • *Stolen Painting Found by Tree*
  • *Kids Make Nutritious Snacks*
  • *Local HS Dropouts Cut in Half*
Variability

• Crucial in semantics
  
  • *Dow ends up 255 points*
  
  • *Dow climbs 255*
  
  • *All major stock markets surged*
  
  • *Dow gains 255 points*
  
  • *Stock market hits a high record*
  
  • *The Dow Jones Industrial Average close up 255*
Sparsity

“I ate an apple”

• Unigrams: “I”, “ate”, “an”, “apple”

• Bigrams: “I ate”, “ate an”, “an apple”
Humans do not learn language by observing an endless stream of text
Interdisciplinary field

Linguistics

CS

NLP

ML/statistics

AI
Related fields

• Computational Linguistics
  • Use computational models to learn about language (e.g., meaning change)

• Cognitive Science
  • How does the human brain process language?
NLP applications

- Text categorization
- Information Extraction
- Search
- Question Answering
- Virtual assistants
- Machine translation
- Summarization
- Reading comprehension
Text categorization

Sports
Politics
Science
New York Times Co. named Russell T. Lewis, 45, president and general manager of its flagship New York Times newspaper, responsible for all business-side activities. He was executive vice president and deputy general manager. He succeeds Lance R. Primis, who in September was named president and chief operating officer of the parent.

<table>
<thead>
<tr>
<th>Person</th>
<th>Company</th>
<th>Post</th>
<th>State</th>
</tr>
</thead>
<tbody>
<tr>
<td>Russell T. Lewis</td>
<td>New York Times newspaper</td>
<td>president and general manager</td>
<td>start</td>
</tr>
<tr>
<td>Russell T. Lewis</td>
<td>New York Times newspaper</td>
<td>executive vice president</td>
<td>end</td>
</tr>
<tr>
<td>Lance R. Primis</td>
<td>New York Times Co.</td>
<td>president and CEO</td>
<td>start</td>
</tr>
</tbody>
</table>
Question answering

What's the capital of Wyoming?

Cheyenne
Question answering

How many US states’ capitals are also their largest cities?

About 982,000,000 results (0.67 seconds)

State Capitals and Largest Cities - Infoplease
www.infoplease.com › United States › States
State Capitals and Largest Cities. The following table lists the capital and largest city of every state in the United States. State, Capital, Largest city.

State Capitals and Largest Cities - Fact Monster
www.factmonster.com › United States › States
State Capitals and Largest Cities. The following table lists the capital and largest city of every state in the United States. State, Capital, Largest city.

List of capitals in the United States - Wikipedia, the free ...
https://en.wikipedia.org/.../List_of_capitals_in_the_United_Stat...
Wikipedia
Austin is the largest state capital that is not also the state's largest city. .... The Confederate States of America had two capitals during its existence. The first ... In many cases, former capital cities of states are outside the current state borders.
State capitals - Insular area capitals - Former national capitals
Question answering

PR video
Critique

• Douglas Hofstadter: “just a text search algorithm connected to a database, just like Google search. It doesn’t understand what it’s reading.”
Virtual assistants

Move all my Wed meetings in April with John to 5pm
Machine translation
Summarization

WASHINGTON (CNN) -- President Obama's inaugural address was cooler, more measured and reasonable than that of other presidents making it, perhaps, the right speech for the times.

Some inaugural addresses are known for their soaring, inspirational language. Like John F. Kennedy's in 1961. "Ask not what your country can do for you. Ask what you can do for your country.

Obama's address was less stirring, perhaps. It was also more candid and down-to-earth.

"Starting today," the new president said, "we can heal these wounds, calm these fears, and harvest the fruits of a new and better era." Obama's speech was a cool speech, not a hot one, Schneider says.

President Obama renewed his call for a massive plan to stimulate economic growth.

Obama, too, offered reassurance.

"We gather because we have chosen hope over fear, unity of purpose over conflict and discord," Obama said.

Obama's call to unity after decades of political division echoed Abraham Lincoln's first inaugural address in 1861. Even though he delivered it at the onset of a terrible civil war, Lincoln's speech was not a call to arms. It was a call to look beyond the war, toward reconciliation based on what he called "the better angels of our nature."

Some presidents used their inaugural address to set out a bold agenda.
"The rock was still wet. The animal was glistening, like it was still swimming," recalls Hou Xianguang. Hou discovered the unusual fossil while surveying rocks as a paleontology graduate student in 1984, near the Chinese town of Chengjiang. "My teachers always talked about the Burgess Shale animals. It looked like one of them. My hands began to shake." Hou had indeed found a Naraoia like those from Canada. However, Hou's animal was 15 million years older than its Canadian relatives.

It can be inferred that Hou Xianguang'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
Some history

A. **Colorless green ideas sleep furiously**

B. **Furiously sleep ideas green colorless**

“It is fair to assume that neither sentence (1) nor (2) … had ever occurred in an English discourse. Hence, in any statistical model for grammaticalness, these sentences will be ruled out on identical grounds as equally “remote” from English. Yet (1), though nonsensical, is grammatical, while (2) is not.”

(Chomsky, Syntactic structures, 1957)
History: rule-based

- 70s and 80s:
  - Grammars (rules) of English syntax
  - Small domains
  - Substantial engineering effort

\[
\begin{align*}
S &\rightarrow NP \ VP \\
NP &\rightarrow DET \ N \\
VP &\rightarrow VT \ N \\
VP &\rightarrow VI
\end{align*}
\]

\[
\begin{align*}
N &\rightarrow \text{man} \\
N &\rightarrow \text{woman} \\
DET &\rightarrow \text{the} \\
DET &\rightarrow \text{a} \\
VT &\rightarrow \text{hugged} \\
VI &\rightarrow \text{slept}
\end{align*}
\]
Empirical revolution

“Whenever I fire a linguist, our system performance improves.”

Jelinek (?), 1988

When I look at an article in Russian, I say: 'This is really written in English, but it has been coded in some strange symbols. I will now proceed to decode.'

Weaver, 1955

Of course, we must not go overboard and mistakenly conclude that the successes of statistical NLP render linguistics irrelevant (rash statements to this effect have been made in the past, e.g., the notorious remark, “Every time I fire a linguist, my performance goes up”). The information and insight that linguists, psychologists, and others have gathered about language is invaluable in creating high-performance broad domain language understanding systems; for instance, in the speech recognition setting described above, better understanding of language structure can lead to better language models.” (Lilian Lee, 2001)
Empirical Revolution

- 1990: corpus-based statistical methods
- 2000s: rich structural formalisms
- 2010s: representation learning with neural nets
Class goals

• Appreciate the challenges of NLP
• Learn about NLP problems and their solution
• Learn general purpose methods
• Be able to read modern research in NLP
Disclaimer

- I am not inventing the wheel in this class

- Will heavily use materials from statistical NLP by Michael Collins

- Will also use materials from Stanford’s 224n class

- You are welcome to watch videos, read those if you feel more comfortable with that

- This is a new class. We will change things based on experience, so please be patient.
Plan

- Primary focus: structured prediction
- Secondary focus: deep learning

- **Class 1**: introduction + representing words (word vectors)
- **Class 2**: Language models, feed-forward networks, back-propagation
- **Class 3**: Tagging, HMMs, Viterbi
- **Class 4**: Log-linear models for tagging
- **Class 5**: PCFGs, CKY
- **Class 6**: Log-linear models for parsing, Lexicalized parsing, dependency parsing
- **Class 7**: Globally normalized log-linear models (structured perceptron, CRFs)
- **Class 8**: Semantic parsing
- **Class 9**: Semantic parsing
- **Class 10**: TBD (RNNs?)
- **Class 11**: TBD (QA?)
- **Class 12**: Project presentation + TBD
- **Class 13**: Project presentation
Administration

- **Instructor**: Jonathan Berant
- **TA**: Dor Muhlgay
- **Time**: Tue, 13-16
- **Office hours**: by e-mail
- **Grades**: 50% final project, 50% homework assignments
Prerequisites

• Introduction to machine learning

• Proficiency in programming (some assignments require python)

• Implied from intro to ML:
  • Probability
  • Linear algebra
  • Calculus
  • Algorithms
Assignments

• Four or five homework assignments

• All involve coding

• Submit in pairs

• Submit by e-mail to Dor
  
  • E-mail title: Assignment <number>: <firstname1> <lastname1> <firstname2> <lastname2>

• A few simple written questions (that will help you code your solution)
  
  • We will randomly grade some fraction of it (if we do, then the written part will be 25% of the total grade for the assignment)
Final project

• Groups of exactly three (one group can be smaller)

• Options:
  • Choose a research paper, implement it, extend it
  • Build a model that does something you care about
  • Define a research problem and work on it

• Must submit a project proposal (that needs to be approved) by 11.5

• If you are taking Prof. Globerson’s class (advance ML) it is advised that you have one project for both classes (it is then expected to be good…)

• Student projects at Stanford
Textbooks

• Martin and Jurafsky:
  • Speech and Language Processing: An Introduction to Natural Language Processing

• Schutze and Manning:
  • Statistical Foundations of NLP

• Noah Smith:
  • Linguistic structure prediction

• More links on the webpage