Natural Language Processing

Introduction
Fall 2019/2020
tl;dr: we would like computers to understand human language

2001: A Space Odyssey
In case you didn’t notice

• The class is in English

• You can ask questions in Hebrew if you feel more comfortable and I will translate
Goal of this lecture

• Put the field of NLP in context:
  • Goal
  • Applications
  • Related fields
  • History
  • Challenges
What is NLP?

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

  - For all languages
  - Many “levels”: transcribed, newswire, social media…
What is NLP?

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

- Search
- Translation
- Text generation
- Speech recognition
- Chatbots
- Virtual assistants
- Ad matching
- …
Levels of linguistic analysis

- **Phonology**: sounds that make up language (more focus in the speech community)
  - דוד בן-גוריון: David Ben-Gurion

- **Morphology**: internal structure of words
  - ובסתיים

- **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

- **Pragmatics**: the way we use language in context
  - It's hot in here
What’s so special about language

• מצויר האדם מון הה المتحدة

  • Allows people to pass along abstractions and tools through generations

• Created 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.
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)
- HOLZ<!--HE-->A מתיילה 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

more
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
"I ate an apple"

- Unigrams: "I", "ate", "an", "apple"
- Bigrams: "I ate", "ate an", "an apple"
Sparsity

“*I ate an apple*”

- Unigrams: “*I*”, “*ate*”, “*an*”, “*apple*”
- Bigrams: “*I ate*”, “*ate an*”, “*an apple*”

Compositionality is key!
Grounding

- 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?

About 984,000 results (0.54 seconds)

Wyoming / Capital

Cheyenne
Question answering

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

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_Sta...
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
Donald Trump may have won the election and was sworn in as president of the United States, but who really determines the steps the White House is his fragile ego. Trump ego dictates to sign the decrees rate of fire so that no one would say that we fulfill the promises of his campaign, but also ego complicates it every day new and blatant lies that make now, the teaching of the ego, of course, also acts. The combination of two things - the dramatic and Hifbrokim orders goats - leaving the US in shock.
Donald Trump may have won the election and been sworn in as President of the United States, but the one who really sets the pace in the White House is his fragile ego. The ego dictates to Trump to sign presidential orders at a rate of fire so that no one will say that he does not keep his election promises, but the ego also complicates him every day with new, blunt lies that are now, on the ego's orders, The combination of these two things - the dramatic orders and the mighty fibrokes - leaves the United States shocked.
Donald Trump may have won the election and was sworn in as President of the United States, but the one who really sets the pace in the White House is his fragile ego. The ego dictates to Trump to sign presidential decrees at a rate of fire so that no one will say that he does not keep his election promises, but the ego also complicates him every day with new, blatant lies that are now, on the ego's orders, of course, also actions. The combination of these two things - the dramatic orders and the powerful fibres - leaves the United States shocked.
Summarization

WASHINGTON (CNN) -- President Obama’s inaugural address was cooler, more measured and reas 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 cou
Obama’s address was less stirring, perhaps, it was also more candid and down-to-earth.

“Electors today, the new president said, this is a test of our democracy. It is a test of our resilience, a test of our will to come together.

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 war. 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
I propose to consider the question, "Can machines think?" This should begin with definitions of the meaning of the terms "machine" and "think"... If the meaning of the words "machine" and "think" are to be found by examining how they are commonly used... is to be sought in a statistical survey such as a Gallup poll. But this is absurd... I shall replace the question by another, which is closely related to it and is expressed in relatively unambiguous words.
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

S → NP VP
NP → DET N
VP → VT N
VP → VI

N → man
N → woman
DET → the
DET → a
VT → hugged
VI → slept
Eliza

Please tell me what's bothering you.
Oh, hello, ELIZA. It's been a long time.
Hi. What seems to be your problem?
Lately, I've been feeling a little depressed.
And how do you feel right now?
I feel sad, but relieved to be writing.
Of what does feeling sad remind you?
Well, I'm an alcoholic, and I've been sober for a long time, but I drank yesterday. And then I decided to stop drinking again. But still that makes me feel sad.
Is it because you are an alcoholic that you came to me?
Empirical revolution

“Wellever 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

• This class borrows from NLP classes around the world
  • Statistical NLP by Michael Collins
  • Stanford 224n
  • Others

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

• However - **slides are not self-contained**. You need to be in class to fully understand their meaning

• Classes might be video-taped?
Plan

• Technical foci: structured prediction and deep learning

• **Class 1**: introduction + word embeddings

• **Class 2**: Word embeddings

• **Class 3**: Language models, contextualized word representations

• **Class 4**: RNNs, Transformers

• **Class 5**: TBD

• **Class 6**: Tagging, log-linear models
Plan

- **Class 7**: Globally-normalized linear models, neural tagging models
- **Class 8**: Parsing, CKY
- **Class 9**: Lexicalized PCFGs
- **Class 10**: Neural parsing models
- **Class 11**: Sequence-to-sequence models, Attention, pointer networks
- **Class 12**: Semantic parsing, subword models
- **Class 13**: Projects
Prerequisites

• Introduction to machine learning (or in parallel)
  • Taking it in parallel means you will have to make-up some material

• Proficiency in programming (some assignments require python)

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

- Every two classes. Tentative plan:
  - Assignment 1: Word embeddings
  - Assignment 2: Language modeling
  - Assignment 3: Sequence tagging
  - Assignment 4: Parsing
  - Assignment 5: Sequence-to-sequence
Assignments

• Five homework assignments
• All involve coding
• Submit in pairs
• Submit by e-mail to the 2 TAs:
  • E-mail title: Assignment <number>: <firstname1> <lastname1> <id1> <firstname2> <lastname2> <id2>
• Some simple written questions
Administration

• **Instructor**: Jonathan Berant

• **Grader**: Ohad Rubin

• **Time**: Wed, 09-12

• **Place**: Dach 005

• **Office hours**: coordinate by e-mail


• **Forum**: [http://moodle.tau.ac.il/course/view.php?id=368307701](http://moodle.tau.ac.il/course/view.php?id=368307701)

• Grades: 50% final project, 50% homework assignments
Final project

• Groups of 2-3 people

• There will be a default project - TBD

• Or you can do a research projects
  • Find a research question you would like to answer
  • Build a model for a task you are excited about
  • Come up with a new approach for a problem
  • ...

• If you are taking Prof. Globerson’s class (advance ML) it is advised that you have one project for both classes. Standard for such project will be higher. All members of team must be enrolled in both classes.
Final project

• Research projects:
  • Decide on project by roughly week 10.
  • Get 1-bit approval
  • Present in last class (5 min.)
• 2-3 top-tier papers result from the class per year.
• See class website for examples
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

• Yoav Goldberg:
  • Neural Network Methods for Natural Language Processing