Introduction

He He

New York University

September 8, 2021

Logistics


He He


Udit Arora


Abed Qaddoumi


Wenqian Ye


Hyejin Kim

- Best way to communicate with us: Campuswire (remember to sign up: link and code on Brightspace).
- Lectures will be in person, but with Zoom sessions and recordings.
- Office hours will be online through Zoom (link on calendar).
- Let us know if you have accessibility needs.

What you'll be able to do by the end of this course

- Understand the core problems and challenges in NLP
- Formalize NLP tasks as statistical learning problems
- Have a toolbox for solving different families of NLP problems
- Gain hands-on experience in building NLP systems
- Pursue advanced study and be prepared for NLP research in academia or industry

What we expect you to know

- **Linear algebra**: vector space, vector norm, dot product, gradient etc.
- **Probability and statistics**: conditional probability, expectation, Bayes rule etc.
- **Basic machine learning**: loss function, gradient descent, logistic regression etc.
- **Programming**: read and write Python code, use Numpy and deep learning libraries

Homework (40%)

Four homeworks (10% each)

- Word vectors and text classification
- Language models and sequence prediction
- Hidden Markov models and part-of-speech tagging
- Constituent parsing

Midterm (30%)

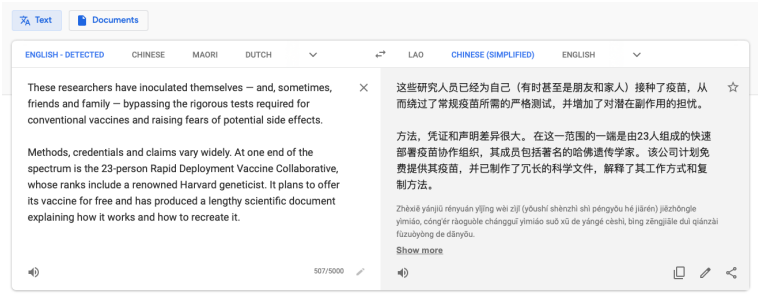
- Format: online or in-person
- Date: Week 7 (October 20)
- MC questions and written questions

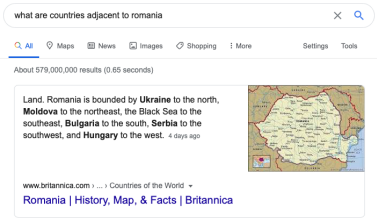
Project (30%)

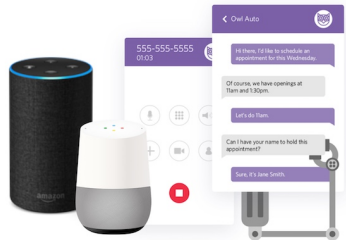
- Related to NLP (doesn't have to be in the scope of this course)
- New algorithms or models for existing problems
- Applications of NLP techniques
- Analysis of well-known approaches that leads to new insight
- ML Reproducibility Challenge 2021 (`https://paperswithcode.com/rc2021`)

Table of Contents

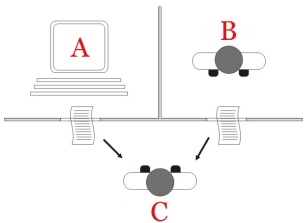
Products powered by NLP technologies

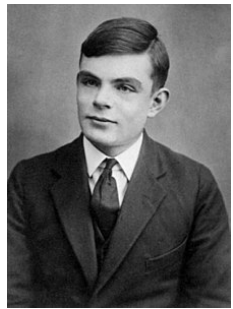






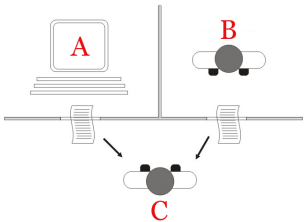
The imitation game

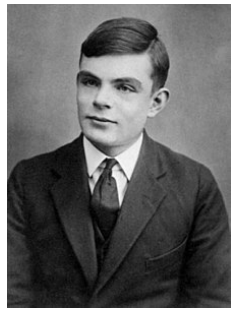




"I believe that in about fifty years' time it will be possible to programme computers, with a storage capacity of about 10^9*, to make them play the imitation game so well that an average interrogator will not have more than 70 percent chance of making the right identification after five minutes of questioning."* Turing (1950)

The imitation game



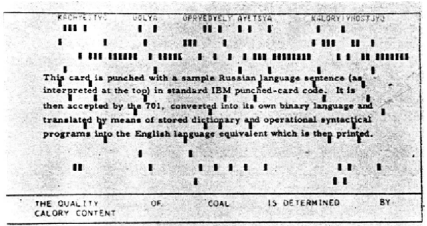


"I believe that in about fifty years' time it will be possible to programme computers, with a storage capacity of about 10^9*, to make them play the imitation game so well that an average interrogator will not have more than 70 percent chance of making the right identification after five minutes of questioning."* Turing (1950)

Caution: Humanlikeness can be easy to fake.

The Georgetown-IBM experiment

- The Russian-English machine translation program:



- A vocabulary of 250 words
- Using 6 grammar rules, e.g.,

  `If first code is 110, is third code associated with preceding complete word equal to 21? If so, reverse order of appearance of words in output (i.e., word carrying 21 should follow that carrying 110)---otherwise, retain order.`

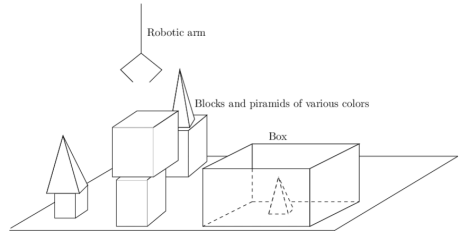
ELIZA

- Built by Joseph Weizenbaum at MIT in 1964 to demonstrate the *superficiality* of human-machine communication.
- Surprisingly, people were convinced that ELIZA had human intelligence.

Human:	Well, my boyfriend made me come here.
ELIZA:	Your boyfriend made you come here?
Human:	He says I'm depressed much of the time.
ELIZA:	I am sorry to hear you are depressed.
Human:	It's true. I'm unhappy.
ELIZA:	Do you think coming here will help you not to be unhappy?

SHRDLU

- Built by Terry Winograd at MIT in 1968.
- A person instructs the computer to build structures in a blocks world.
- Grounding, coreference resolution, question answering, semantic parsing etc.



Person:	Pick up a big red block.
Computer:	OK.
Person:	Grasp the pyramid.
Computer:	I don't understand which pyramid you mean.
Person:	(changing their mind) Find a block which is taller than the one you are holding and put it into the box.
Computer:	By "it", i assume you mean the block which is taller than the one i am holding.

Limitations of early systems

- Optimism in the 50's and 60's
  "Within the very near future—much less than twenty-five years—we shall have the technical capability of substituting machines for any and all human functions in organizations."

- Disappointing results due to
  - **Limited computation**: hardware has limited speed and memory
  - **Combinatorial explosion**: algorithms are intractable in realistic settings
  - **Underestimated complexity**: ambiguity, commonsense knowledge etc.

The rise of statistical learning in the 80's

- Notable progress in MT from IBM (neglected knowlege of linguistics).
- HMMs widely used for speech recognition.
  "Every time I fire a linguist, the performance of the speech recognizer goes up."—Frederick Jelinek.
- The paradigm shift: expert knowledge + rules $\rightarrow$ data + features
- Machine learning is the main driving force of NLP today.

Table of Contents

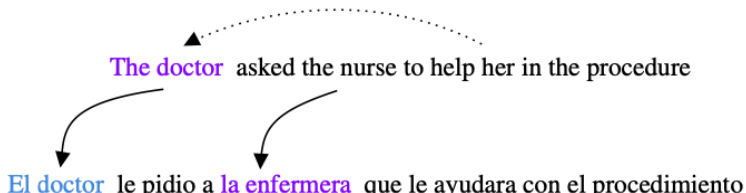
Are we there yet?

Predictions are not robust to benign perturbations [Ribeiro+ 2020].

Test case	Expected	Predicted	Pass?
A Testing **Negation** with ***MFT***	Labels: negative, positive, neutral		
Template: I {NEGATION} {POS_VERB} the {THING}.			
I can't say I recommend the food.	neg	pos	X
I didn't love the flight.	neg	neutral	X
...			
		Failure rate = 76.4%	

Are we there yet?

MT systems are prone to gender-biased translation errors [Stanovsky+ 2019].



Are we there yet?

QA models are easily distracted by irrelevant sentences [Jia+ 2017].

Article: Super Bowl 50
Paragraph: "*Peyton Manning became the first quarterback ever to lead two different teams to multiple Super Bowls. He is also the oldest quarterback ever to play in a Super Bowl at age 39. The past record was held by John Elway, who led the Broncos to victory in Super Bowl XXXIII at age 38 and is currently Denver's Executive Vice President of Football Operations and General Manager. Quarterback Jeff Dean had jersey number 37 in Champ Bowl XXXIV.*"
Question: "*What is the name of the quarterback who was 38 in Super Bowl XXXIII?*"
Original Prediction: John Elway
Prediction under adversary: Jeff Dean

Why is language hard?

Why is language hard?

- **Discrete**
  - How to define metrics?

    I work at NYU. vs I work for NYU.
    This is good. vs This is actually good.

  - How to define transformations?

    The food is okay. → The food is awesome!
    They made a brief return to Cambridge to drop the book. → They returned.

  - In general, hard to represent text as mathematical objects.

Why is language hard?

- **Compositional**
  - The whole is built from parts (chars, words, sentences, paragraphs, documents...)
  - How to generalize when we don't see all possible combinations?
  - An example from [Lake+ 2018]
    Vocabulary:

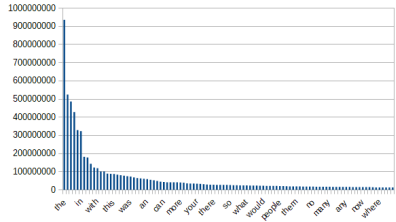
    {jump, walk, turn, once, twice, left, right, before, after, and}

    Sentences:

    jump
    jump left
    jump left and walk right
    jump left after walk right once before turn left twice
    ...

Why is language hard?

- **Sparse**
  - How to handle the long tail?
  - Zipf's law: word frequency $\propto \frac{1}{\text{rank}}$



  - Many linguistic phenomena follow Zipf's law
    BoA's financial assistant Erica:
    The bank "learned [that] there are over 2,000 different ways to ask us to move money."[1]

[1] https://www.aiqudo.com/2019/06/28/voice-success-story-erica-bank-america/

Why is language hard?

- **Ambiguous**
  - How to interpret meaning in context?

    Bass: fish? guitar? frequency? (word sense disambiguiation)

    I shot an elephant in my pajamas: who is in the pajamas? (PP attachment)

    The spirit is willing but the flesh is weak.
    $\rightarrow$ The vodka is strong but the meat is rotten.

Table of Contents

Overview

1. Representation of text: string $\rightarrow$ feature vector
2. Structured prediction: predicting sequences, trees, graphs
3. Neural networks for NLP: the deep learning tsunami [Manning, 2015]

Representation of text

Symbolic: structured objects/concepts

Distributed/Connectionist: patterns of latent properties

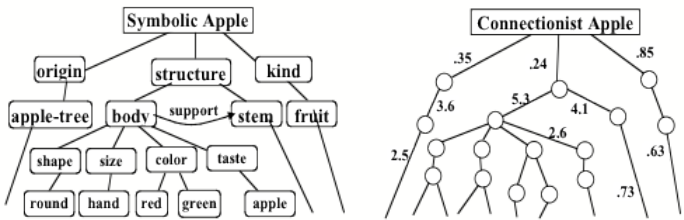


Figure: https://web.media.mit.edu/~minsky/papers/SymbolicVs.Connectionist.html

Pros and cons?

Structured prediction: sequences

- Named entity recognition

$\underbrace{\text{New York University}}_{org}$ is a private research university based in $\underbrace{\text{New York City}}_{loc}$. It is founded in $\underbrace{1831}_{year}$ by $\underbrace{\text{Albert Gallatin}}_{people}$.

Structured prediction: sequences

- Named entity recognition

$\underbrace{\text{New York University}}_{org}$ is a private research university based in $\underbrace{\text{New York City}}_{loc}$. It is founded in $\underbrace{\text{1831}}_{year}$ by $\underbrace{\text{Albert Gallatin}}_{people}$.

$\underbrace{\text{CT of the maxillofacial area}}_{test}$ showed no facial $\underbrace{\text{bone fracture}}_{symptom}$. $\underbrace{\text{CT of the brain}}_{test}$ showed no acute changes.

Structured prediction: sequences

- Semantic role labeling (slot filling)

  I would like to book a ticket from New York to San Francisco on Christmas eve.

  action=book_ticket

  departure city=New York

  destination city=San Francisco

  date=Christmas eve

  time=

Structured prediction: sequences

- Anaphora resolution

  John had a great evening meeting with his high school friends.

Structured prediction: sequences

- Anaphora resolution

  John had a great evening meeting with his high school friends.

  The city councilmen refused the demonstrators a permit because they feared violence.

Structured prediction: trees

- Syntactic parsing

Bob bought a book

The old man the boat

Text generation

- Machine translation

  爱屋及乌 → Love me, love my dog

- Data-to-text

Date	min	max
tomorrow	21°C	29°C

→ Tomorrow's temperature will be between 21 and 29 degrees.

- Summarization

  The Justice Department plans to bring an antitrust case against Google as soon as this month, after Attorney General William P. Barr overruled career lawyers who said they needed more time to build a strong case against one of the world's wealthiest, most formidable technology companies, according to five people briefed on internal department conversations.

  → Justice Dept. plans to file antitrust charges against Google in coming weeks.

Key questions in structured prediction

- **Modeling**

  How to model iteractions among substructures?

  $$\text{score}(\text{sentence}, \text{structure})$$

- **Learning**

  How to efficiently learn the model parameters given data?

  $$\max_{\text{parameters}} \text{score}(\text{sentence}, \text{gold structure})$$

- **Inference**

  How to efficiently find the best structure given a learned model?

  $$\max_{\text{structure}} \text{score}(\text{sentence}, \text{structure})$$

Neural networks for NLP

- Encoder-decoder models

- Pre-training and fine-tuning

Beyond text: grounding

Connect language to the world

"Can you bring me an apple?"