CSE 587: Deep Learning for Natural Language Processing

Lecture 1. Introduction

Rui Zhang Spring 2023



To share thoughts with conspecifics

- to **communicate** with others
- to store and pass knowledge

To share thoughts with conspecifics

- to **communicate** with others
- to store and pass knowledge

To think more complex thoughts

- reasoning
- interpretation

To share thoughts with conspecifics

- to **communicate** with others
- to store and pass knowledge



Almost all of your use of language is internal. Virtually all of the use of language has nothing to do with communication. The idea that language has evolved as a system of communication, or designed for communication, makes no sense.

Noam Chomsky

The systems of thought use linguistic expressions for reasoning, interpretation, organizing action, and other mental acts.

To think more complex thoughts

- reasoning
- interpretation

To share thoughts with conspecifics

- to **communicate** with others
- to store and pass knowledge



Almost all of your use of language is internal. Virtually all of the use of language has nothing to do with communication. The idea that language has evolved as a system of communication, or designed for communication, makes no sense.

Noam Chomsky

The systems of thought use linguistic expressions for reasoning, interpretation, organizing action, and other mental acts. To think more complex thoughts

- reasoning
- interpretation

The limits of my language mean the limits of my world.



Ludwig Wittgenstein

Natural Language Processing (NLP)

It deals with **natural languages** that human speak (e.g., English, Chinese, Spanish, Arabic, etc), not programming languages.

• NLP is also referred to as **Human Language Technology**

Natural Language Processing (NLP)

It deals with **natural languages** that human speak (e.g., English, Chinese, Spanish, Arabic, etc), not programming languages.

• NLP is also referred to as **Human Language Technology**

It aims to use computational approaches to analyze human languages.

• morphology, lexical analysis, syntax, semantics, discourse

Natural Language Processing (NLP)

It deals with **natural languages** that human speak (e.g., English, Chinese, Spanish, Arabic, etc), not programming languages.

• NLP is also referred to as **Human Language Technology**

It aims to use computational approaches to analyze human languages.

• morphology, lexical analysis, syntax, semantics, discourse

It aims to build computer systems that can **understand** and **generate** natural language.

• It can be used for human-human communication (e.g., machine translation, grammatical error correction) or human-machine communication (e.g., dialog systems, question answering).

Natural Language Processing vs Computational Linguistics

A lot of people use them interchangeably.

They are very similar, but also different.

The focus is a bit different:

- Computational Linguistics focus on scientific questions in "Linguistics". It aims to study scientific questions about human languages using computational tools.
- Natural Language Processing focus on engineering problems of building practical systems and applications that can mimic human usage of languages.

Machine Translation

DETECT LANGUAGE	CHINESE	ENGLISH	SPANISH	\sim	÷	CHINESE (SIMPLIFIED)	ENGLISH	SPANISH	\sim		
到目前为止,还没有生物标志物可以预测个体对COVID-19的敏感性。 × 南方科技大学、上海交大、武汉中南医院、武汉金银潭医院等8家单位 的最新研究显示: A、B、O、AB血型与新冠肺炎易感性存在关联。这 是该领域的首份研究。 Dào mùqián wéizhǐ, hái méiyǒu shēngwù biāozhì wù kěyǐ yùcè gètǐ duì COVID-19 de mǐngǎnxìng. Nánfāng kējì dàxué, shànghǎi jiāodà, wǔhàn zhōngnán yīyuàn, wǔhàn jīnyíntán yīyuàn děng 8 jiā dānwèi de zuìxīn yánjiū xiǎnshì:A,B,O,AB xiěxíng yǔ xīnguān					×	To date, there are no bid sensitivity to COVID-19. Southern University of S University, Wuhan Zhon etc. showed that blood susceptibility to new co this field.	omarkers that The latest res Science and To gnan Hospita types of A, B, ronary pneum	can predict an search from 8 u echnology, Sha I, Wuhan Jinyin O, and AB are a nonia. This is th	n individual's units includin Inghai Jiaoto Itan Hospita associated v Ine first study	ng ong I, vith vin	\$
↓ ↓				107/5000 拼	•	•()			D	I	Ş

Question Answering

About 3,340,000 results (0.86 seconds)

Nittany Lion

The Nittany Lion is the mascot of the Penn State Nittany Lions—the athletic teams of the Pennsylvania State University, located in University Park, Pennsylvania, USA. It is an eastern mountain lion, the "Nittany" forename referring to the local Mount Nittany, which overlooks the university.

en.wikipedia.org > wiki > Nittany_Lion **v** Nittany Lion - Wikipedia



Conversational Personal Assistants



Apple Siri



Google Assistant



Microsoft Cortana

Natural Language Understanding is Hard



Natural Language Understanding is Hard

"The old man the boat."

Natural Language Understanding is Hard

"The old man the boat."





16

OpenAI GPT-3 Writes Entire Articles

Q. Mario Klingemann 🤣 @quasimondo

Another attempt at a longer piece. An imaginary Jerome K. Jerome writes about Twitter. All I seeded was the title, the author's name and the first "It", the rest is done by #gpt3

...

Here is the full-length version as a PDF: drive.google.com/file/d/1qtPa1c...

The importance of being on twitter by Jerome K. Jerome

London, Summer 1897

It is a curious fact that the last remaining form of social life in which the people of London are still interested is Twitter. I was struck with this curious fact when I went on one of my periodical holidays to the sea-side, and found the whole place twittering like a starling-cage. I called it an anomaly, and it is.

I spoke to the sexton, whose cottage, like all sexton's cottages, is full of antiquities and interesting relics of former centuries. I said to him, "My dear sexton, what does all this twittering mean?" And he replied, "Why, sir, of course it means Twitter." "Ah!" I said, "I know about that. But what is Twitter?"

"It is a system of short and pithy sentences strung together in groups, for the purpose of conveying useful information to the initiated, and entertainment and the exercise of wits to the initiated, and entertainment and the exercise of wits to the rest of us."

"Very interesting," I said. "Has it a name?" "It has," he said; 'It is called Twitter. "Yes," I said, "I know that, but what is it?" "It is a system of information," he said. "Oh, yes," I replied; "but what is it?"

GPT-3 as A Universal Language Model

https://beta.openai.com/?demo=1

Even for Writing Code...



This is mind blowing.

With GPT-3, I built a layout generator where you just describe any layout you want, and it generates the JSX code for you.

...

WHAT



NLP Tasks

- <u>Automatic speech recognition</u>
- <u>CCG</u>
- <u>Common sense</u>
- <u>Constituency parsing</u>
- <u>Coreference resolution</u>
- Data-to-Text Generation
- Dependency parsing
- <u>Dialogue</u>
- Domain adaptation
- Entity linking
- Grammatical error correction
- Information extraction
- Intent Detection and Slot Filling
- Language modeling
- Lexical normalization
- Machine translation

http://nlpprogress.com/

- <u>Missing elements</u>
- <u>Multi-task learning</u>
- <u>Multi-modal</u>
- Named entity recognition
- Natural language inference
- Part-of-speech tagging
- Paraphrase Generation
- Question answering
- Relation prediction
- Relationship extraction
- <u>Semantic textual similarity</u>
- Semantic parsing
- Semantic role labeling
- Sentiment analysis
- Shallow syntax
- Simplification
- Stance detection
- Summarization
- <u>Taxonomy learning</u>
- <u>Temporal processing</u>
- <u>Text classification</u>
- Word sense disambiguation

NLP Tasks

Foundational Technologies

- Language Modeling
- Part-of-speech Tagging
- Syntactic Parsing
- Dependency Parsing
- Named Entity recognition
- Coreference resolution
- Word Sense Disambiguation
- Semantic Role Labelling
-

High-Level Tasks and Applications

- Sentiment Analysis
- Information Extraction
- Machine Translation
- Question Answering
- Semantic Parsing
- Summarization
- Dialogue systems
- Language and Vision
- Data-to-Text Generation
-

Natural Language Processing Systems

Input X	Output Y	Task
Text	Label	Text Classification (e.g., Sentiment Analysis)
Text	Linguistic Structure	Structured Prediction (e.g., Part-of-Speech Tagging)
Text	Text	Text Generation (e.g., Translation, Summarization)

Natural Language Processing Systems

Input X	Output Y	Task
Text	Label	Text Classification (e.g., Sentiment Analysis)
Text	Linguistic Structure	Structured Prediction (e.g., Part-of-Speech Tagging)
Text	Text	Text Generation (e.g., Translation, Summarization)

To create systems for X -> Y, we can

- Expert Systems: Manually create rules mapping X to Y based on our knowledge of languages
- Data-Driven Methods: Collect Data and Use Machine Learning Techniques to learn the mapping function

Data Driven Methods and Machine Learning

"Machine Learning" is a data-driven problem solving strategy.

Data Driven Methods and Machine Learning

"Machine Learning" is a data-driven problem solving strategy.

The ultimate goal of machine learning is **generalization**! Making future prediction on unseen examples as accurate as possible!



Common theme is to solve a prediction problem:

Given an input $oldsymbol{x}$,

Predict an "appropriate" output y (label, decision, action, etc).

Machine Learning Pipeline

Can I say something about prediction error at testing?



Training, Validation, Test



Data Hygiene

Training Data: Use this and only this to train your neural networks.

Development Data: Use this to decide when to stop training your neural networks, and do hyperparameter tuning.

Test Data: Use test data after picking your model. You should use test data only once.

Neural Networks and Deep Learning





Figure 1: The Transformer - model architecture.

Deep Learning for NLP



From "Real-Time Social Media Analytics with Deep Transformer Language Models: A Big Data Approach" by Ahmet and Abdullah

Course Topics

- Part 1: Introduction and Neural Network Basics
 - 1. Introduction to NLP
 - 2. Text Classification and Language Modeling
 - 3. Neural Networks and Backpropagation

- Part 3: NLP Applications
 - 7. Question Answering
 - 8. Semantic Parsing
 - 9. NLP for Programming
 - 10. Natural Language Generation
 - 11. Summarization
 - 12. Multilingual NLP
 - 13. Language and Vision
 - 14. Commonsense Knowledge and Reasoning

Part 2: NLP Models

- 4. Sequence-to-Sequence, Attention, Transformers
- 5. Pretraining Language Models and Transfer Learning
- 6. Text-to-Text and Prompt-based Learning

- Part 4: Advanced Techniques
 - 15. Variational Autoencoders and Latent Variable Models
 - 16. Structured Prediction
 - 17. Robustness in NLP
 - 18. Interpretability, Explainability, Model Analysis
 - 19. Efficient NLP Methods
 - 20. Ethics in NLP



Course Information

Lecture

MoWe 2:30PM - 3:45PM @ Walker Building 124

Instructor

https://ryanzhumich.github.io/ Office Hour: Wednesdays 4-6pm @ W329 Westgate Building and Zoom

Check our research!

TA

Wenjian Miao Office Hour: Thursdays 4-6pm @ W300

How to Contact us

Rui Zhang

- For most questions, please use Piazza.
- Please only use Canvas Inbox for personal issues. •
- Don't use psu emails: It's hard for me to monitor class status using emails.

Canvas Page

Canvas Inbox Piazza GradeScope

Textbooks

All Available online! You are strongly encouraged to read beyond slides.

- <u>Speech and Language Processing</u>. Dan Jurafsky and James H. Martin.
- <u>Dive into Deep Learning</u>. Aston Zhang, Zack C. Lipton, Mu Li, Alex J. Smola.
- <u>Natural Language Processing</u>. Jacob Eisenstein.
- Foundations of Statistical Natural Language Processing. Chris Manning and Hinrich Schütze.
- <u>Linguistic Structure Prediction</u>. Noah A. Smith.
- <u>Machine Learning: A Probabilistic Perspective</u>. Kevin Murphy.
- Linguistic Fundamentals for NLP. Emily M. Bender.

Prerequisites

- CMPSC 448: Machine Learning
- Programming in Python
- CSE 582: Natural Language Processing

Previous experience in NLP/Deep Learning is preferred but not required.

Programming Assignment

One assignment on coding neural networks models for NLP.

- Minimalist Neural Networks for Sentiment Analysis
- January 27 February 13

Python

GradeScope

Not late assignment accepted unless there is a compelling reason. Send me an Canvas Inbox if you know you are going to be late.

Paper Review and Presentation

Each student will do one lecture-style presentation (2 presentations in 1 class).

Each student will submit 5 paper reviews before the class and participate in the discussion during presentations.



No Exam. This course is a research project oriented course.

Project Format

Project Topics. This project aims to conduct original and independent research over **any NLP-related topics**.

- Must deal with natural language data
- Must include some degree of model implementations and experiments

Group Policy. You can work on the course project in a group of **1** - **3** people. You are allowed to combine this project with your research projects or projects from other courses.

Project Deliverables

- **Project Proposal**. Write a 3-page proposal that outlines your plan including what problem or task you want to address, what dataset(s) you want to work on, what metrics you need to employ, what baselines you would like to compare with. You should also cite a few relevant prior papers.
- Final Report. Your final report should use our Latex template with at least 8-page plus references. Your report should begin with an abstract and introduction to clearly state the problem you want to solve and contributions you have made. It should also have a section on related work, a section on your methodology, a section on your experimental settings and results, and a section on conclusions.
- **Code and Data**. Please submit your data and runnable code with a detailed instruction.

Project Timeline

Group Formation: January 27 Project Proposal: February 17 Final Report/Code and Data: May 5 (during the final exam week)

Grading	ii • Assignment			
0	H B Assignment 20 pts	◎ :		

ii • Paper Review and Presentation	25% of Total +	:
Paper Presentation 10 pts	\odot	:
Paper Review 1 3 pts	\otimes	:
Paper Review 2 3 pits	\odot	:
Image: Paper Review 3 Paper Review 3 3 pts 3	\odot	:
Paper Review 4 3 pts	\otimes	:
Paper Review 5 3 pts	\otimes	:

ij + Pr	roject	(50% of Total) + :
	Project Proposal 10 pts	0 :
II 🖻	Final Report 30 pts	0 :
:: P	Code and Data 10 pts	0 :

Eventse Feedback	5% of Total +	- :	
ii 📴 SRTE 5 pts	\odot	:	

Final letter grades will be curved and the cut scores will be determined after all grades are finalized.