CS 585: Natural Language Processing
Fall 2021
Location IIT Tower 1F6–1
Tuesdays, 6:45–9:25

Instructor: Derrick Higgins
dhiggins1@iit.edu
Office: Stuart 228A

TAs: Zhenghao Zhao, Juanyan Wang
zzhao48@hawk.iit.edu, jwang245@hawk.iit.edu

Overview

This course is about how to build systems that analyze unstructured natural language texts and extract useful information from them. Students should expect to gain familiarity with the most common types of natural language processing (NLP) tasks, including text classification, sequence labeling, and structure prediction—and to learn appropriate frameworks for performing these tasks. The course will cover the technical methodology in sufficient detail to allow students to apply these frameworks in an informed way, and to make current research accessible.

Prerequisite knowledge

To succeed in the course, you will need to have knowledge of programming, probability theory, algorithm design and linear algebra. Previous knowledge of natural language processing and machine learning will be helpful. Practical programming exercises will be done in Python 3. Some Unix shell commands will be introduced, but shell-scripting experience is not a prerequisite.

Course expectations

You are expected to do the reading for each lecture before class. The course will involve taking an NLP project from end to end: creating a labeled dataset, analyzing the statistical properties of the data and partitioning it appropriately for modeling, and creating a machine learning model using linguistic features to make new predictions. There will also be a midterm and a final exam, which will be open book/notes.

Grading

Course project (Python; in three installments)	50%
Midterm exam	20%
Final exam	30%

Course reading

- **M&S**: *Foundations of Statistical Natural Language Processing*, by Christopher Manning and Hinrich Schutze, MIT Press, 1999. `https://github.com/won1k/cs287/blob/master/Manning_Schuetze_StatisticalNLP.pdf`
- **E-NLP**: *Introduction to Natural Language Processing*, by Jacob Eisenstein, MIT Press, 2019. `https://github.com/jacobeisenstein/gt-nlp-class/tree/master/notes`
- **UNIXWiki**: `https://en.wikibooks.org/wiki/Guide_to_Unix/Commands/Text_Processing`

Communication

- **Slack:** *Do* post (and answer) questions about the course in the `#general` channel; *don't* DM me and expect immediate and comprehensive answers
- **Instructor office hours:** Tuesdays 5:40–6:40 PM, or by special arrangement
- **TA office hours:** Juanyan — Tuesdays 10–12 AM, SB011; Zhenghao — Wednesdays and Fridays 2–3 PM, SB019

Course resources

- **Course Website:** `http://www.cs.iit.edu/~cs585/`
- **Slack Channel:** `https://iitcs585fall2021.slack.com/`
- **Blackboard**

Academic integrity

- Please read IIT's Code of Academic Honesty: `https://web.iit.edu/student-affairs/handbook/fine-print/code-academi`
- All work you turn in must be done by you alone (except where assignments are explicitly given to groups)
- You may not look at the solution of any other student prior to the due date

Lecture schedule

Date		Topic	Reading	Project
8/24/2021	1 2	Welcome, linguistic concepts Mathematics review 1: probability and linear algebra	M&S 2.1, 3	
8/31/2021	3 4	Mathematics review 2: information theory and frequency distributions Practical text processing	M&S 2.2; UNIXWiki	Phase 1 (annotation) out
Words				
9/7/2021	5 6	Words and pattern matching Lexical representations for NLP	M&S 1.4, 4.2, E-NLP 4.3	
9/14/2021	7 8	Neural nets 1: neural word embeddings Word sense disambiguation	M&S 7.1–7.3, E-NLP 14	Phase 1 due
Texts				
9/21/2021	9 10	Text categorization and naïve Bayes Generalized linear models	E-NLP 2.1–2.2, 2.5–2.6, 4.4	Phase 2 (analysis) out
9/28/2021	11 12	Neural nets 2: feedforward networks Sentiment analysis	E-NLP 3.1–3.3.3, 4.1	
10/5/2021	13 14	Unsupervised methods in NLP Midterm Review	E-NLP 5.1	Phase 2 due
10/12/2021		**Midterm**		
Sequences				
10/19/2021	15 16	Language models Parts of speech and sequence tagging	E-NLP 6.1, 6.2.1, 6.2.2, 6.4, 6.5, 7.1, 7.2	Phase 3 (modeling) out
10/26/2021	17 18	Hidden Markov models and the Viterbi algorithm Unsupervised sequence labeling (EM)	E-NLP 7.3, 7.4, 7.7	
11/2/2021	19 20	Structured prediction Neural nets 3: neural models for sequence labeling	E-NLP 7.5, 3.4, 6.3, 7.6	
Trees				
11/9/2021	21 22	Context-free grammars and syntax CKY parsing	E-NLP 9.2, 10.1–10.2	
11/16/2021	23 24	Probabilistic CFG parsing Dependency grammar	E-NLP 10.3–10.4	
Tasks				
11/23/2021	25 26	Semantic role labeling Machine Translation	E-NLP 13.1–13.2, 18	Phase 3 due
11/30/2021	27 28	Other tasks Final review	E-NLP 16, 19.2	
12/7/2021		**Final Exam**		