Syllabus

CS 478

Natural Language Processing (NLP)

Instructor

Ziyu Yao (ziyuyao [at] gmu [dot] edu)
Office Hours: by appointments (Virtual or in person at ENGR4415).

Teaching Assistant

TBD
Office Hours: TBD

Meets

Monday and Wednesday, 3:00 to 4:15 PM, Exploratory Hall L102.

Course Web Page

https://nlp.cs.gmu.edu/course/cs478-fall24/.
We will use **Blackboard** for course materials/assignments/grading, and **Piazza** for Q&A (sign up link here).

Course Description

Massive amounts of information in our daily life are expressed in natural language. In this class, we will study building computing systems that can process, understand, and communicate in natural language. This field is called natural language processing, or NLP. This class will focus on introducing foundamental concepts in NLP, and will cover techniques and necessary programming skills for building machine learning/deep learning-based NLP models. In the last several classes, we will further study cutting-edge research problems in NLP, including text generation, question answering, neural network interpretation, interactive learning, multilingual NLP, and so on.

Prerequisites

CS310 (Data Structure), CS330 (Formal Methods and Models), and proficiency in Python programming. Please contact the instructor if you have questions about the necessary background.

Class Format

The class will be in-person. Each class will take the following format:

- *Reading:* Before the class, you will be pointed to some reading materials *(see "Reading Materials" in the course schedule)* . Reading is not required but highly recommended to do before the class.
- *Summary/Elaboration/Q&A:* In the class, the instructor will summarize important points from the reading material, elaborating on details that were not included in the reading while fielding any questions. New material on cutting-edge methods, or a deep look into one salient method will also be covered.
- *Quiz:* In some classes, there will be in-class quizzes evaluating student learning performance. The quiz will be based on the content in the previous lectures.

Grading

There will be no midterm or final exam. Your final grade will be dependent on:
Quizzes and Class Participation: 30%. We will have weekly in-class quizzes, each containing 2-3 short-answer questions covering contents introduced in previous lectures. The instructor will specify the range one class before the quiz. Each quiz question is worth 1% of the total grade, with the remaining credits for class participation (based on student attendance and ad-hoc presentation).

Assignments: 70%. There will be six assignments in this class. Each assignment must be completed independently.

- A1: Introduction (5%). The assignment will ask you a few questions about the class logistic and help you get familiar with classic NLP topics.
- A2: Understand n-gram language models (10%).
- A3: Build a feedforward neural net for sentiment classification (20%). This assignment will have three submission checkpoints.
- A4: Build a BERT-based sentiment classifier (10%).
- A5: Explore Large Language Models (15%).
- A6: Analyze a BERT-based sentiment classifier (10%).

Your final letter grade will be given based on:

Letter Grade	Points (out of 100)
A+	100+ (w/ extra credits)
A	95-100
A-	90-94
B+	85-89
B	82-84

B-	78-81
C+	74-77
C	72-73
C-	70-71
D	60-69
F	0-59

Late Day Policy for Assignments:
In case there are unforeseen circumstances that don't let you turn in your assignment on time, 5 late days *total* over the six assignments will be allowed. Notes: (1) The late days cannot be used fractionally, e.g., submitting the assignment 1 hour late will incur 1 late day; (2) You may want to save the late days for coding assignments, which are generally harder than the others. Assignments that are late beyond the allowed late days will be graded down by 5% per day. In the case of a serious illness or other excused absence, as defined by university policies (including providing necessary evidence), coursework submissions will be accepted late by the same number of days as the excused absence. However, students should inform the instructor as soon as they can, ideally before the due date.

Class Attendance Policy:
As we will have quizzes for some classes, attendence is highly suggested. If you miss a quiz, you lose the credits (i.e., no make ups for quiz). However, in the case of a serious illness or other excused absence, as defined by university policies (including providing necessary evidence), students will be excused and the quiz can be taken home. However, students should inform the instructor as soon as they can, ideally before the quiz date.

Readings

Students should be able to understand the course content just by following the lecture and by doing the readings. However, the following textbooks serve as good references.

- Jurafsky and Martin, Speech and Language Processing, 3rd edition [online] **(Referred to as "JM")**;
- Jacob Eisenstein, Natural Language Processing [online] **(Referred to as "Eisenstein")**;
- Yoav Goldberg, Neural Network Methods in Natural Language Processing [publisher] [online primer pdf] **(Referred to as "Goldberg-Publisher/Primer")**; Note that the "publisher" version can be downloaded if you use the school VPN.

Tentative Schedule

#	Date	Topic	Reading Materials	Assignment Dates
1	08/26	Introduction and Class Outline		A1 Out
2	08/28	Text Processing and Probability Basics	JM Ch2	
3	09/02	Labor Day - No Class		
4	09/04	N-gram Language Models	JM Ch3.1-3.4	A1 Due; A2 Out
5	09/09	ML/NLP Experimental Design	JM Ch4.7-4.9, Ch11.5	
6	09/11	Classification 1	JM Ch5.1-5.5, Ch4.1-4.4, Eisenstein Ch2.3-2.4, 2.6	
7	09/16	Classification 2		
8	09/18	Classification 3 & Intro to Neural Nets	JM Ch5.6	A2 Due; A3 Out
9	09/23	Neural 1: Feedforward Neural Networks	JM Ch7.1-7.4 & Goldberg-Primer Ch6.1-6.3 Blog by Michael Nielsen, DL book. PyTorch basics	
10	09/25	Neural 1: Feedforward Neural Networks (cont.)	JM Ch6; Mikolov et al., 2013a&b	
11	09/30	Neural 2: Word Embeddings	JM Ch6; Mikolov et al., 2013a&b	A3.1 Due
12	10/02	Neural 3: RNN-based Neural Language Models	JM Ch9-9.3; "understand LSTM" blog by Olah; "gradient vanishing" blog by Nielsen; Karpathy et al. 2015	
13	10/07	Neural 4: Seq2Seq & Attention, Transformers	Eisenstein 18.3.1; Attention-based NMT; Transformer paper and Alammar's blog	
14	10/09	Mid-term Wrap-up		A3.2 Due
15	10/14	Spring Recess - No Class		
16	10/16	Contextual Representations	Peters et al., 2018 (ELMo); Devlin et al., 2019 (BERT); OpenAI GPT2	A3.3 Due; A4 Out
17	10/21	Prompt-based Learning	Liu et al. survey 2021, Jiang et al., 2019, Li and Liang, 2021	
18	10/23	Sequence: POS tagging (HMM) & NER (CRF)	JM Ch8.1-8.2, 8.4 JM Ch8.5, Eisenstein Ch7.5.3	
19	10/28	Parsing 1: Dependency Parsing	JM Ch14	
20	10/30	Parsing 2: Constituency Parsing	JM Ch12.1-12.6.1, 13.1-13.2, 13.4, Ch15-15.3	A4 Due; A5 Out
21	11/04	Parsing 3: Semantic Parsing	Eisenstein Ch13	

22	11/06	Machine Translation	Eisenstein 18.1-18.2	
23	11/11	Interpreting and Debugging NLP Models 1	LIME, e-SNLI, EMNLP20 tutorial	
24	11/13	Interpreting and Debugging NLP Models 2 & A6 Preview	Hewitt&Liang'19, CheckList	A5 Due; A6 Out
25	11/18	Text Generation	JM Ch24; Holtzman et al., 2020; Ranzato et al., 2016; Maynez et al., 2020; Sellam et al., 2020; See et al., 2017	
26	11/20	Question Answering 1	JM Ch23; ACL20 tutorial; QA over text: Chen et al., 2017 (DrQA); Lee et al., 2019 (ORQA); Zhu et al., 2021 (survey); QA over structured data:Pasupat&Liang, 2015 (Table QA); Yih et al., 2015 (KBQA); Rajpurkar et al., 2016	
27	11/25	Question Answering 2		
28	11/27	Thanksgiving - No Class		
29	12/02	Human-AI Interaction	Wang et al., 2016, Hancock et al., 2019, guidelines for human-AI interaction; Materials by Daume III	A6 Due
30	12/04	AI/NLP Ethics	Zhao et al., 2017, Rudinger et al., 2018, Gebru et al., 2018	
31	12/09	Select A5 Presentations		

Honor Code

The class enforces the GMU Honor Code, and the more specific honor code policy special to the Department of Computer Science. You will be expected to adhere to this code and policy.

Note to Students

Take care of yourself! As a student, you may experience a range of challenges that can interfere with learning, such as strained relationships, increased anxiety, substance use, global pandemics, feeling down, difficulty concentrating and/or lack of motivation. All of us benefit from support during times of struggle. There are many helpful resources available on campus and an important part of having a healthy life is learning how to ask for help. Asking for support sooner rather than later is almost always helpful. GMU services are available, and treatment does work. You can learn more about confidential mental health services available on campus at: https://caps.gmu.edu/. Support is always available (24/7) from Counseling and Psychological Services: 703-527-4077.

Disabilities

If you have a documented learning disability or other condition which may affect academic performance, make sure this documentation is on file with the Office of Disability Services and come talk to me about accommodations. I will work with you to ensure that accommodations are provided as appropriate. If you suspect that you may have a disability and would benefit from accommodations but are not yet registered with the Office of Disability Services, I encourage you to contact them at ods@gmu.edu.

Diversity and Inclusion

GMU seeks to create a learning environment that fosters respect for people across identities. We welcome and value individuals and their differences, including gender expression and identity, race, economic status, sex, sexuality, ethnicity, national origin, first language, religion, age and ability. We encourage all members of the learning environment to engage with the material personally, but to also be open to exploring and learning from experiences different than their own. Check out the Mason Non-Discrimination Policy and the Mason Diversity statement.

Name and Pronouns Statement

If you wish, please share your name and gender pronouns with me and indicate how best to address you in class and via email. I use "she/her/hers" for myself and you may address me as “Ziyu”, “Dr./Prof. Yao” in email and verbally.

Sexual or Interpersonal Violence

As a faculty member, I am designated as a “Non-Confidential Employee,” and must report all disclosures of sexual assault, sexual harassment, interpersonal violence, stalking, sexual exploitation, complicity, and retaliation to Mason’s Title IX Coordinator per University Policy 1202. If you wish to speak with someone confidentially, please contact one of Mason’s confidential resources, such as Student Support and Advocacy Center (SSAC) at 703-993-3686 or Counseling and Psychological Services (CAPS) at 703-993-2380. You may also seek assistance or support measures from Mason’s Title IX Coordinator by calling 703-993-8730, or emailing titleix@gmu.edu.

Student Privacy

Student privacy is governed by the Family Educational Rights and Privacy Act (FERPA). For this reason, students must use their Mason email account to receive important University information, including communications related to this class. I will not respond to messages sent from or send messages to a non-Mason email address.

Recording and/or sharing class materials

Some kinds of participation in online study sites violate the Mason Honor code: these include accessing exam or quiz questions for this class; accessing exam, quiz, or assignment answers for this class; uploading of any of the instructor's materials or exams; and uploading any of your own answers or finished work. Always consult your syllabus and your professor before using these sites.

Undergraduate Course Repetition

Please see AP. 1.3.4 in the University Catalog and consult with your academic advisor if you have any questions regarding repeating an undergraduate class for credit.

Last updated on Jan 1, 2023