CS695-002 Natural Language Processing (Special Topics)

Instructor
Ziyu Yao (https://ziyuyao.org/)
Office Hour: TBD

Meets
AU2021, Thursdays, 4:30 – 7:10 PM, Location TBD

Course Web Page
TBD

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. The class will start with an introduction to the foundations of natural language processing (NLP), and then focus on cutting-edge research problems in NLP. Each section will introduce a particular problem or phenomenon in natural language, describe why it is difficult to model, and demonstrate recent models that were designed to tackle this problem. In the process of doing so, the class will cover different techniques that are useful in creating neural network models, including handling variably sized and structured sentences, semi-supervised and unsupervised learning, structured prediction, and so on. The class will include assignments culminating in a final project.

Prerequisites
Ideally, (a) Algorithms and Data Structures, (b) Artificial Intelligence or Data Mining, and (c) Probability and Statistics (STAT 344) or equivalent.
Students should be experienced with writing substantial programs in Python. The ideal set of prerequisites includes Algorithms and Data Structures, Artificial Intelligence or Data Mining, and linear algebra and calculus, but please contact the instructor if you have questions about the necessary background.

Class Format
The class will be in-person. As the class aims to provide skills necessary to familiarize the students with, and to do cutting-edge NLP research, the classes and assignments will be at least partially implementation-focused. In general, each class will take the following format:
- *Reading:* Before the class, you will be given a reading assignment that you should read before coming to class that day. This will usually be an important paper that tackles a specific NLP problem.
- *Quiz:* At the beginning of class, there will be a short quiz that tests your knowledge of the reading assignment. (These quizzes should be easy if the reading assignment has been completed and understood.)
- *Summary/Elaboration/Questions:* The instructor will summarize the important points of the reading material, elaborate on details that were not included in the reading while fielding any

questions. Finally, new material on cutting-edge methods, or a deep look into one salient method will be covered.
- *Code Walk:* In some classes we will walk through some demonstration code that implements a simple version of the main concepts presented in the reading material.
- *Presentation:* In some classes, students will be asked to present their project progress.

Grading

There will be no midterm or final exam. Your final grade will be dependent on:

Quizzes (15%): Your lowest 2 quiz grades will be dropped. If you are sick or traveling on business (e.g. to a conference, for a job interview, or delayed in return due to visa issues), send a doctor's note or evidence of the reason for being away to the instructor within a week of the absence, and you will be excused. I expect excused quizzes to be relatively rare, and if you'll be away for more than, e.g. 2 classes over the semester, please consult in advance.

Presentation (10%): You will give two presentations in class.
- Project Proposal Presentation: In the middle of the semester, you will present your project proposal in class and will receive feedback from your classmates. Similarly, you will be asked to provide feedback to your classmates' proposals. The purpose is to (1) help you learn to evaluate others' proposals and (2) allow you to further improve your own proposal by learning from your peers.
- Final Project Presentation: In the last class (before the Assignment 4 Due), you will present your final project. Requirements on the presentation will be provided by the instructor.

Assignment and Final Project (75%): There will be 4 assignments, covering one programming assignment (which must be completed *independently*) and one open-ended final project (which could be done independently or in a group with *no more than 3 students*).
- **Assignment 1: Implementation and Initial Interest Survey (10%)**
  You will be asked to implement a neural network-based NLP model almost from scratch. The purpose is to help you get familiar with basic concepts and skills for building neural network systems. In the second part of this assignment, there will also be an initial questionnaire regarding what task you are interested in tackling for the final project. Some suggested topics and resources will be provided for reference.
- **Assignment 2: Project Proposal and Literature Survey (10%)**
  This is Checkpoint 1 of your project: it involves a proposal of a project topic and a literature survey regarding this topic. In the survey, explain the task that you would like to tackle in concrete terms, and also cover all of the relevant recent research on the topic. You will also need to include a concrete plan and timeline towards accomplishing the final project.
  Note that you should actively communicate with the instructor throughout the project:
  - Before and After Assignment 2 Due: Make sure you have discussed your project idea with the instructor before finalizing it and submitting the proposal. After the submission, you will receive second-round feedback from the instructor, based on which you may revise your proposal and reflect the changes in your proposal presentation.
  - Changing Topics: Although it is not recommended, you are still allowed to change topics after the Assignment 2 Due and the proposal presentation. However, you should confirm with the instructor first and will need to adjust your proposal accordingly.

- **Assignment 3: Project Baseline Implementation (20%)**
  Checkpoint 2 will involve reproducing the evaluation numbers of a state-of-the-art baseline model for the task of interest with code that you have implemented (mostly from scratch, dependent on the project). In other words, you must get the same numbers as the previous paper on the same dataset.
  Students need submit the source code implementation with a clear README documentation, such that the instructor can easily run and check the outputs.
- **Assignment 4: Final Project Report (35%)**
  The final project work will be expected to be a novel research contribution that either (1) introduces new techniques for one of the existing tasks in the assignment utilizing one of the more advanced techniques introduced in the class, or (2) tackles a new NLP task (potentially with a neural network model that is motivated by the unique problems posed by the application domain), or (3) presents a novel, meaningful analysis of existing methods and their potential failures. The final project submission should include your report and your source code implementation. In the last class before the Assignment 4 Due, you will present your project.

Late Day Policy: In case there are unforeseen circumstances that don't let you turn in your assignment on time, 5 late days total over the first three assignments will be allowed (late days may not be applied to the final project, assignment 4). Note that the third assignment is harder than the first one, so it'd be a good idea to try to save your late days for the third assignment if possible. Assignments that are late beyond the allowed late days will be graded down one half-grade per day late.

Readings
For each topic/class the instructor will provide a list of papers as suggested readings. One paper will be required reading and will be tested with a quiz (see above). 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, 2nd edition
  *Make sure you get the purple 2nd edition of J+M, not the white 1st edition.
- Jurafsky and Martin, Speech and Language Processing, 3rd edition
- Goldberg, Neural Network Methods in Natural Language Processing

Tentative Schedule

Week	Date	Topic	Due on Same Date (Unless Specified)
1	08/26	Introduction and Class Outline	
2	09/02	Language Models, Smoothing, and Recurrent Neural Networks	
3	09/09	Distributional Semantics and Word Vectors	
4	09/16	Contextual Representations and Text Classification	Assignment 1 Due
5	09/23	Sequence Labeling: HMM & CRF	
6	09/30	Syntactic Parsing	
7	10/07	Semantic Parsing	Assignment 2 Due

8	10/14	Project Proposal Presentation	
9	10/21	Language Generation	
10	10/28	Question Answering	
11	11/04	Interactive Systems in NLP	Assignment 3 Due
12	11/11	Emerging Research: Language and Code	
13	11/18	Ethical Considerations in NLP	
14	11/25	Thanksgiving (no class)	
15	12/02	Final Project Presentation	Assignment 4 Due 12/09

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.

Notes 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.