CS 584-A Natural Language Processing

Department of Computer Science

Fall 2023

Instructor: Ping Wang
Canvas Course Address: https://sit.instructure.com/courses/68054
Course Schedule: Monday, 3:00 PM-5:30 PM, 2023 Fall
Course Location: Burchard 103
Contact Info: ping.wang@stevens.edu
Virtual Office Hours: Tuesday, 11AM-12PM. Zoom: https://stevens.zoom.us/j/97789938972

Teaching Assistant:

- Bharath Beeravelly (bbeerave@stevens.edu) Office hours: Wednesday, 1PM-2PM. Zoom: https://stevens.zoom.us/j/94713156470.
- Chetna Agarwal (cagarwal@stevens.edu) Office hours: Friday, 1PM-2PM. Zoom: https://stevens.zoom.us/j/93947263945, passcode: 251118.

COURSE DESCRIPTION

Natural language processing (NLP) is one of the most important technologies referring to automatic computational processing of human languages. This includes algorithms that take human-produced text as input or produce text as output. People communicate almost everything in language: emails, phone calls, language translation, web searches, reports, books, social media, etc. Human language is symbolic in nature and also highly ambiguous and variable. Comprehending human language is a crucial and challenging part of artificial intelligence. There are a large variety of underlying tasks and machine learning models behind NLP applications. Recently, deep learning approaches have been studied and achieved high performance in many NLP tasks. The course provides an introduction to machine learning and deep learning research applied to NLP. We will cover topics including word vector representations, neural networks, recurrent neural networks, convolutional neural networks, seq2seq models, attention-based models, transformers, pre-training, fine-tuning, prompting, and many other advanced NLP techniques and applications.

Prerequisites:

This course is fast-paced and covers a lot of ground, so it is important that you have a solid foundation on both the theoretical and empirical fronts. **You should have background in python programming, probability theory, linear algebra, calculus, and foundations of machine learning.**

STUDENT LEARNING OUTCOMES

After successful completion of this course, students will be able to

- Understand the neural networks models and backpropagation optimization. Implement neural network models in TensorFlow.
- Apply word2vec models in real-world text corpora.
- Understand convolutional neural networks and their application in NLP.
- Understand sequence to sequence models and attention in NLP deep neural networks.
- Understand dependency parsing, recurrent neural networks and their application in NLP.
- Implement gradient descent (GD) and stochastic gradient descent (SGD) techniques for learning problems and understand the theory behind them.
- Understand the recent advances in NLP, such as transformers, pre-training, fine-tuning, prompting, etc.

COURSE FORMAT AND STRUCTURE

This course is on-campus. To access the course, please visit stevens.edu/canvas. For more information about course access or support, contact the Technology Resource and Assistance Center (TRAC) by calling 201-216-5500.

Course Logistics

Some important course logistics are as follows:

- When assignments are due, they are due by **11:59 am EST** on the due date listed in the course schedule. Any changes of the due dates will be announced in class or sent as announcements on Canvas.
- Deadlines are an unavoidable part of being a professional, and this course is no exception. Course requirements must be completed and posted or submitted on or before the specified due date and delivery time deadline. Due dates and delivery time deadlines are in Eastern Time (as used in Hoboken, NJ). Please note that students living in distant time zones or overseas must comply with this course time and due date deadline policy. Avoid any inclination to procrastinate. Due dates have been established for each assignment to encourage you to stay on schedule.
- Any late submission within 24 hours will be penalized 10%. Any late submission within 24-48 hours will be penalized 40%.
- The 48 hours after the deadline will be the hard deadline for each assignment. Assignments submitted after this hard deadline will not be graded and get no points for the assignment.
- An assignment file should be appended by your username, such as "hw1_FirstName_LastName.pdf". This makes it easier for managing assignment files during grading.
- Each student needs to submit your own answers or codes. Similar or the same answers/codes will not be graded.
- Use generative AI technology with care. More details can be found in the section about generative AI technologies.

Virtual Office Hours

Virtual Office Hours are a synchronous session through Zoom to discuss questions related to the course. Office hours will be held Tuesday from 11:00 am-12:00 pm EST. You can join the virtual office hours at https://stevens.zoom.us/j/97789938972.

TENTATIVE COURSE SCHEDULE

Week	Date	Topic	HW
1	Sep 4	Labor Day, No class	
2	Sep 11	Introduction to NLP	HW1 Out
3	Sep 18	Machine Learning Basics	
4	Sep 25	Vector Semantics	
5	Oct 2	Language Modeling; N-gram; RNN	HW2 Out; HW1 Due
6	Oct 9 Oct 10	Fall Recess (Columbus Day): No class Monday class schedule More RNN; CNN	 Project proposal due
7	Oct 16	Machine Translation; seq2seq	
8	Oct 23	Self-attention; transformers	HW3 Out; HW2 due
9	Oct 30	Pretraining	
10	Nov 6	Prompting	Project midterm report due
11	Nov 13	Chain-of-thought; Dependency Parsing	HW4 Out; HW3 Due
12	Nov 20	Midterm Exam	Midterm exam
13	Nov 27	Interpretability; Ethics; Applications	
14	Dec 4	Project Presentation 1	HW4 Due
15	Dec 11	Project Presentation 2	Project final report and code due

Important Dates:

- Oct 10: Project proposal due
- Nov 6: Project midterm report due
- Nov 20: Midterm exam
- Dec 11: Project final report and codes due

COURSE MATERIALS

- Yoav Goldberg. Neural Network Methods for Natural Language Processing. 2017.
- Dan Jurafsky and James H. Martin. Speech and Language Processing. 2018.
- Goodfellow et al., Deep Learning, MIT, 2016
- Lewis Tunstall, Leandro von Werra, and Thomas Wolf. Natural Language Processing with Transformers, revised edition, O'Reilly Media, Inc., 2022.

COURSE REQUIREMENTS

Grading Policy: The course will use the following grading scale: A (90-100), A- (85-90), B+ (80-85), B (75-80), B- (70-75), C+ (65-70), C (60-65), F (<60).

- **Homework (30%):** There will be four homework assignments with both written and programming problems. The assignments are designed to help you deepen your understanding of the theoretical concepts.
- **Midterm Exam (30%):** The midterm exam will be an in-class written exam to evaluate your understanding of the course so far.

- **Course project (35%):** More details of the course project can be found in the project guidelines.
    - **Project proposal: 5%**
    - **Midterm report: 10%**
    - **Final report and codes: 15%**
    - **Presentation: 5%**
- **Participation (5%):** Attending classes, participating in classes. You are also encouraged to initiate or participate in discussions on Canvas, which will also be considered as class participation.

Submission: You are encouraged to work and discuss in a group, but you have to write down and **submit your OWN answers and codes. Any similar or the same answers/codes will not be graded.**

GRADING PROCEDURES

Details of the grading can be found in grading policy in the section of course requirements.

Late Policy
All the homework assignments must be **submitted on Canvas before 11:59 AM** on the due date.

- Any late submission within 24 hours will be penalized 10%.
- Any late submission within 24-48 hours will be penalized 40%.
- The 48 hours after the deadline will be the hard deadline for each assignment. Assignments submitted after this hard deadline will not be graded and get no points for the assignment.

Academic Integrity
All students in this course should strictly follow the Honor System at Stevens specified below.

Generative AI Technologies
You may use AI programs e.g. ChatGPT to help generate ideas and brainstorm. However, you should note that the material generated by these programs may be inaccurate, incomplete, or otherwise problematic. Beware that use may also stifle your own independent thinking and creativity. Therefore, it is important that you use generative AI technology with care.

You may not submit any work generated by an AI program as your own. If you include material generated by an AI program, it should be cited like any other reference material (with due consideration for the quality of the reference, which may be poor).

Any plagiarism or other form of cheating will be dealt with under relevant Stevens policies.

Undergraduate Honor System
Enrollment into the undergraduate class of Stevens Institute of Technology signifies a student's commitment to the Honor System. Accordingly, the provisions of the Stevens Honor System apply to all undergraduate students in coursework and Honor Board proceedings. It is the responsibility of each student to become acquainted with and to uphold the ideals set forth in the Honor System Constitution. More information about the Honor System including the constitution, bylaws, investigative procedures, and the penalty matrix can be found online at http://web.stevens.edu/honor/.

The following pledge shall be written in full and signed by every student on all submitted work (including, but not limited to, homework, projects, lab reports, code, quizzes and exams) that is assigned by the course instructor. No work shall be graded unless the pledge is written in full and signed.

"I pledge my honor that I have abided by the Stevens Honor System."

Students who believe a violation of the Honor System has been committed should report it within ten business days of the suspected violation. Students have the option to remain anonymous and can report violations online at www.stevens.edu/honor.

Graduate Student Code of Academic Integrity
All Stevens graduate students promise to be fully truthful and avoid dishonesty, fraud, misrepresentation, and deceit of any type in relation to their academic work. A student's submission of work for academic credit indicates that the work is the student's own. All outside assistance must be acknowledged. Any student who violates this code or who knowingly assists another student in violating this code shall be subject to discipline.

All graduate students are bound to the Graduate Student Code of Academic Integrity by enrollment in graduate coursework at Stevens. It is the responsibility of each graduate student to understand and adhere to the Graduate Student Code of Academic Integrity. More information including types of violations, the process for handling perceived violations, and types of sanctions can be found on the Office of Graduate Academics web page.

Special Provisions for Undergraduate Students in 500-level Courses
The general provisions of the Stevens Honor System do not apply fully to graduate courses, 500 level or otherwise. Any student who wishes to report an undergraduate for a violation in a 500-level course shall submit the report to the Honor Board following the protocol for undergraduate courses, and an investigation will be conducted following the same process for an appeal on false accusation described in Section 8.04 of the Bylaws of the Honor System. Any student who wishes to report a graduate student may submit the report to the Senior Vice Provost for Graduate Education or to the Honor Board, who will refer the report to the senior vice provost. The Honor Board Chairman will give the Senior Vice Provost for Graduate Education weekly updates on the progress of any casework relating to 500-level courses. For more information about the scope, penalties, and procedures pertaining to undergraduate students in 500-level courses, see Section 9 of the Bylaws of the Honor System document, located on the Honor Board website.

ACCOMMODATIONS

Stevens Institute of Technology is dedicated to providing appropriate accommodations to students with documented disabilities. The Office of Disability Services (ODS) works with undergraduate and graduate students with learning disabilities, attention deficit-hyperactivity disorders, physical disabilities, sensory impairments, psychiatric disorders, and other disabilities to help students achieve their academic and personal potential. They facilitate equitable access to the educational programs and opportunities offered at Stevens and coordinate reasonable accommodations for eligible students. These services are designed to encourage independence and self-advocacy with support from the ODS staff. The ODS staff will facilitate the provision of accommodations on a case-by-case basis.

For more information about Disability Services and the process to receive accommodations, visit https://www.stevens.edu/student-diversity-and-inclusion/disability-services. If you have any questions please contact the Office of Disability Services at disabilityservices@stevens.edu or by phone: 201.216.3748.

Disability Services Confidentiality Policy
Student Disability Files are kept separate from academic files and are stored in a secure location within the Office of Disability Services. The Family Educational Rights Privacy Act (FERPA, 20 U.S.C. 1232g; 34CFR, Part 99) regulates disclosure of disability documentation and records maintained by Stevens Disability Services. According to this act, prior written consent by the student is required before our Disability Services office may release disability documentation or records to anyone. An exception is made in unusual circumstances, such as the case of health and safety emergencies.

INCLUSIVITY

Stevens Institute of Technology believes that diversity and inclusiveness are essential to excellence in academic discourse and innovation. In this class, the perspective of people of all races, ethnicities, gender expressions and gender identities, religions, sexual orientations, disabilities, socioeconomic backgrounds, and nationalities will be respected and viewed as a resource and benefit throughout the semester. Suggestions to further diversify class materials and assignments are encouraged. If any course meetings conflict with your religious events, please do not hesitate to reach out to your instructor to make alternative arrangements.

You are expected to treat your instructor and all other participants in the course with courtesy and respect. Disrespectful conduct and harassing statements will not be tolerated and may result in disciplinary actions.

Name and Pronoun Usage
As this course includes group work and class discussion, it is vitally important for us to create an educational environment of inclusion and mutual respect. This includes the ability for all students to have their chosen gender pronoun(s) and chosen name affirmed. If the class roster does not align with your pronouns and/or name, please inform the instructor of the necessary changes.

Religious Holidays
Stevens is a diverse community that is committed to providing equitable educational opportunities and supporting students of all ethnicities and belief systems. Religious observance is an essential reflection of that rich diversity. Students will not be subject to any grade penalties for missing a class, examination, or any other course requirement due to religious observance. In addition, students will not be asked to choose between religious observance and academic work. Therefore, students should inform the instructor at the beginning of the semester if a requirement for this course conflicts with religious observance so that accommodations can be made for students to observe religious practices and complete the requirements for the course.

MENTAL HEALTH RESOURCES

Part of being successful in the classroom involves a focus on your whole self, including your mental health. While you are at Stevens, there are many resources to promote and support mental health. The Office of Counseling and Psychological Services (CAPS) offers free and confidential services to all enrolled students who are struggling to cope with personal issues (e.g., difficulty adjusting to college or trouble managing stress) or psychological difficulties

(e.g., anxiety and depression). Appointments can be made by phone (201-216-5177), online at https://stevensportal.pointnclick.com/confirm.aspx, or in person on the 2nd Floor of the Student Wellness Center.

EMERGENCY INFORMATION

In the event of an urgent or emergent concern about your own safety or the safety of someone else in the Stevens community, please immediately call the Stevens Campus Police at 201-216-5105 or on their emergency line at 201-216-3911. These phone lines are staffed 24/7, year-round. For students who do not reside near the campus and require emergency support, please contact your local emergency response providers at 911 or via your local police precinct. Other 24/7 national resources for students dealing with mental health crises include the National Suicide Prevention Lifeline (1-800-273-8255) and the Crisis Text Line (text "Home" to 741-741). If you are concerned about the wellbeing of another Stevens student, and the matter is *not* urgent or time sensitive, please email the CARE Team at care@stevens.edu. A member of the CARE Team will respond to your concern as soon as possible.