CS 162: Natural Language Processing
Nanyun (Violet) Peng

Course information:

- **Instructor**
    - Prof. Nanyun (Violet) Peng
        - Office location: Eng VI 397A

- **Course Description:** Natural Language Processing (NLP) is a rapidly developing field, with recent advances of deep neural networks that revolutionize many NLP applications. This course is intended as an introduction to a wide range of NLP tasks, algorithms for effectively solving these problems (including the most recent advances of deep learning models), and methods for evaluating their performance. There will be a focus on statistical and neural-network learning algorithms that train on (annotated) text corpora to automatically acquire the knowledge needed to perform the task. Class lectures will discuss general issues as well as present abstract algorithms. The homework will touch both theoretical foundations of linguistic phenomena and implementation of the algorithms. Implemented versions of some of the algorithms will be provided in order to give a feel for how the systems discussed in class "really work" and allow for extensions and experimentation as part of the course projects.

- **Course Activities:** The instructor will give lectures; TAs will hold demo and review sessions. There will be a team-based course project.

    - Lecture, four hours; discussion, two hours; outside study, six hours.

- **What you will learn:** You will learn the fundamentals, classic models, and recent trends in natural language processing (NLP), as well as learning the skills to build real world NLP systems with teamwork.

- **Tentative Topics:**
    - **Introduction to NLP: What's Important? What's Hard? What's Easy? (1 week)** Introduction to NLP applications, ambiguity in language, different levels of language.
    - **Lexical semantics (1 week):** distributional semantics and word vectors, word-document matrix, LSA, neural network basics
    - **Language models (3 weeks).** N-gram language model, log-linear language model, RNN language models, transformers, neural masked language model.
    - **Midterm (0.5 week)**

    - **Grammars and Parsers (2 weeks).** Word segmentation, chunking, Part-of-speech tagging, constituency parse, dependency parse, sequence tagging models, shift-reduce parser.
    - **Information Extraction (1.5 weeks).** Overview of tasks: named entity recognition, relation extraction, event extraction, etc. Deep dive of named entity recognition
    - **Current NLP tasks and competitions (1 weeks):** introductions to the current frontier of NLP applications, fairness considerations.
- **Prerequisites:**
    - Enforced requisite: Course COM SCI M146 or COM SCI M145 or EC ENGR M146. We will not cover ML background in this course.
    - Recommended: COM SCI 35L.
    - Basic math skill
    - Programming skill. We will mainly use python with PyTorch, but you can use any other libraries for your final project.

Related Courses

- List of related courses:
    - http://web.stanford.edu/class/cs224n/
    - https://courses.cs.washington.edu/courses/cse447/20wi/
    - https://www.cs.jhu.edu/~jason/465/

Course policy

- Grading policy:
    - 30% Homework
    - 15% Course Project
    - 20% Midterm
    - 30% Final
    - 5% Participation
- Accommodations for students with disabilities: To ensure that disability-related concerns are properly addressed from the beginning, students with disabilities who require assistance to participate in this class are asked to see me as soon as possible. We aim to be as accommodating and fair as possible.
- Use of outside resources: When using outside resources, proper citation is required. This includes papers, text books, software libraries, websites, and help from others. For the details of Honor Code, please refer to **UCLA Student Conduct**. If you have any doubt, please check with me in advance. You may get an F in the final letter grade if we detect any cheating.
- Late policy: The submission site will be closed 2 days after the deadline. A total of 2 days of late hours can be used towards homeworks and the course project. We do not further break the 2 days into hours -- even if you only used 1 minute of the late hours, it counts as 1 day.

- Collaboration: For the course project, by default, students in the same team will get the same score unless there are special circumstances. We encourage students to use a version control system (e.g., github, gitlab, etc…). It is important to keep your hard work in a safe place and log the contributions of individuals. If your team members complain about you and you cannot provide evidence of your contribution, we may lower your score.

Homeworks

- Homework 1:
  - Release: beginning of week 2
  - Due: end of week 3
  - Return: end of week 4
- Homework 2:
  - Release: beginning of week 5
  - Due: end of week 6
  - Return: end of week 7
- Homework 3:
  - Release: beginning of week 8
  - Due: end of week 9
  - Return: end of week 10

Final project

- The goal of the course project is to provide students with an opportunity to explore doing research on a related topic. Therefore, the project aims at producing a "deliverable" result, meaning that your project should be self-complete, reproducible (scientifically correct) and related to the course content. A typical (successful) project consists of 1) a novel and sound solution to an interesting research problem, 2) correct and meaningful comparisons among baselines and existing approaches, 3) applying existing techniques to a new application. We will not penalize negative results, as long as the proposed approach is well explored. It is recommended to form a group with a diversity of background, but not required.
- The instructor will provide a list of potential final project topics. Students are welcome to propose other project ideas.
- Final project schedule:
  - Project proposal: each group should provide a 1-page project proposal at the end of **week 3**. The template will be provided.
  - Project midterm report: by **week 7**, each group should provide a 2-page midterm report (introduction/related work/methodology/proposed experiments)
  - A **4-page final project report** is due in **week 10**. Each group is expected to record a 10~15 min presentation (a demo in the presentation session is

    - encouraged, but not required), and the slides should be submitted with the project report for grading.
    - Each student will be assigned to review a project by others. The review is due in **the final week**. A review template will be provided.
    - The final project will be graded by the instructor and the TAs, and detailed grading rubrics will be announced.

Tentative Schedule:

- Week 1: Introduction to NLP: What's Important? What's Hard? What's Easy?
- Week 2: Lexical semantics
- Week 3: Language models (n-gram language models, smoothing)
- Week 4: Language models (neural language models)
- Week 5: Language models (transformers, masked language model)
- Week 6: Midterm + word segmentation/chunking, Part-of-speech tagging, sequence tagging
- Week 7: Named entity recognition
- Week 8: Syntax and constituency parsing
- Week 9: Dependency parsing
- Week 10: Current NLP tasks and competitions
- Final week