

CS 4120: Natural Language Processing

Professor Lu Wang

Spring 2020

E-mail: luwang@ccs.neu.edu

Web: <http://www.ccs.neu.edu/home/luwang/>

Office Hours: Th 1:30pm-2:30pm (or by appointment)

Office: 2211, 177 Huntington Ave.

Class Hours: Tu 11:45 am - 1:25 pm, Th 2:50pm - 4:30pm

Class Room: Snell Library 037

Course Description

This course aims to introduce fundamental tasks in natural language processing, and its recent advances based on machine learning algorithms (e.g., neural networks) and applications for interdisciplinary subjects (e.g., computational social science). The course materials are mostly delivered as lectures, and accompanied with reading materials. The students will be evaluated based on assignments, a research-driven course project, and an open-book final exam.

Course and Discussion Websites

Course Website: http://www.ccs.neu.edu/home/luwang/courses/cs4120_sp2020/cs4120_sp2020.html

Discussion Forum: <http://piazza.com/class/spring2020/cs4120/home>

Required Materials

- Main textbooks
 - Dan Jurafsky and James H. Martin, "Speech and Language Processing, 2nd Edition", Prentice Hall, 2009.
 - Third edition draft is available at web.stanford.edu/~jurafsky/slp3/.
- Other Reference
 - Chris Manning and Hinrich Schütze, "Foundations of Statistical Natural Language Processing", MIT Press, 1999

- Since many natural language processing problems are driven by machine learning techniques nowadays, we also highly encourage you to read machine learning textbooks:
 - Christopher M. Bishop, “Pattern Recognition and Machine Learning”, Springer, 2006.
 - Tom Mitchell, “Machine Learning”, McGraw Hill, 1997.

Prerequisites/Corequisites

This course is designed for senior undergraduate students majoring in computer science, information science, linguistics, and other related areas. Students who take this course are expected to be able to write code in some programming languages (e.g., Python is recommended) proficiently, and finish courses in algorithms (CS 3000, CS 3800, or CS4810), multivariable calculus, probability, and statistics. Linear algebra is optional, but highly recommended. It would be beneficial if the students have prior knowledge on supervised machine learning.

Course Structure

Grading Policy

Each assignment or report is due by the end of day on the corresponding due date (i.e. 11:59pm, EST). Blackboard is used for electronic submission. Assignment or report turned in late will be charged 20 points (out of 100 points) off for each late day (i.e. every 24 hours). Each student has a budget of 6 days throughout the semester before a late penalty is applied. You may want to use it wisely, e.g. save for emergencies. Each 24 hours or part thereof that a submission is late uses up one full late day. Late days are not applicable to final presentation. Each group member is charged with the same number of late days, if any, for their submission. There is no need to inform the instructors if late days are used; timestamp of the last submission on Blackboard will be used for automatic grade calculation.

Grades will be determined based on three assignments, eight in-class quizzes, one course project, one open-book exam, and participation:

- 45% of your grade will be determined by three assignments, each of 15%.
- 5% of your grade will be determined by eight quick in-class tests, each of 1%; three with lowest grades will be dropped, and no make-up.
- 22% of your grade will be determined by your course project, including proposal (3%), progress report (6%), final report (8%), and presentation (5%) with 1% as bonus if selected as best project by peer student.
- 25% of your grade will be determined by final exam.
- 3% of your grade will be determined by your participation in the class.

Course Policies

Academic Integrity and Honesty

This course follows the Northeastern University Academic Integrity Policy. All students in this course are expected to abide by the Academic Integrity Policy. Any work submitted by a student in this course for academic credit should be the student's own work. Collaborations are allowed only if explicitly permitted. Violations of the rules (e.g. cheating, fabrication, plagiarism) will be reported.

Schedule and Weekly Learning Goals

The schedule is tentative and subject to change. Please check the course website for most recent schedule. The learning goals below should be viewed as the key concepts you should grasp after each week.

Week 01, 01/06 - 01/10: Topic: Introduction, Language Models

- Reading: Ch1, Ch4.1-4.9 (2nd edition by default)
- TODO: start thinking about projects and looking for teammates

Week 02, 01/13 - 01/17: Topic: Text Categorization and Evaluation, Naive Bayes, Part-of-Speech, HMM

- Reading: Ch5.1-5.5, Ch6.1-6.5

Week 03, 01/20 - 01/24: Topic: Word Sense Disambiguation

- Reading: Ch20

Week 04, 01/27 - 01/31: Topic: Machine Learning Basics (Maximum Entropy, Feedforward Neural Networks), Formal Grammars of English, Syntactic Parsing

- Reading: Ch7.1-7.4 in 3rd edition, Ch12.1-12.5
- Course project proposal due on Jan 28.

Week 05, 02/03 - 02/07: Topic: Parsing cont'd, Vector-Space Lexical Semantics

- Reading: Ch12.1-12.5, Ch12.7, Ch 13.4.1, Ch20.7
- Assignment 1 is due on Feb 5.

Week 06, 02/10 - 02/14: Topic: Distributional Semantics

- Reading: Ch20.7

Week 07, 02/17 - 02/21: Topic: Neural Language Models, Sentiment Analysis

- Reading: Ch7 and Ch21 in 3rd edition

Week 08, 02/24 - 02/28: Topic: Sentiment Analysis cont'd, Summarization

- Reading: Ch23.3
- Assignment 2 is due on Feb 28.

Week 09, 03/02 - 03/06: Topic: Spring Break

Week 10, 03/09 - 03/13: Topic: Question Answering

- Reading: Ch23.1-23.2

Week 11, 03/16 - 03/20: Topic: Dialogue Systems, project feedback (see Piazza for arrangement)

- Reading: Ch24.1-24.2
- Course project progress report due on Mar 17.

Week 12, 03/23 - 03/27: Topic: Chatbots

Week 13, 03/30 - 04/03: Topic: Machine Translation, Discourse Analysis

- Reading: Ch21.1-21.2, Ch25.1-25.5
- Assignment 3 is due on Apr 1.

Week 14, 04/06 - 04/10: Topic: Project Presentation

Week 15, 04/13 - 04/17: Topic: Finishing project reports

- Project final report due on Apr 14.

Week 16, 04/20 - 04/24: Topic: Final Exam

- Pending confirmation.