CSC 585 Algorithms for NLP

Gould-Simpson, 701 Tuesday and Thursday 12:30pm—1:45pm

Course Description

This course covers important algorithms for natural language processing (NLP), including traditional NLP algorithms, distributional similarity algorithms such as word embeddings, sequence labeling models, neural networks such as recurrent, recursive and convolutional neural networks, transformers, and parsing algorithms such as shift-reduce. The course will focus on the algorithms that underlie NLP rather than the application of NLP to various problem domains. We will cover syntax, semantics, and pragmatics. Topics will include both classical and cutting-edge algorithms.

Instructor and Contact Information

Eduardo Blanco, Gould-Simpson 739, eduardoblanco@arizona.edu

Office Hours: Tuesday 1:45pm – 3:45pm or by appointment

Web: https://eduardoblanco.github.io/ and https://d2l.arizona.edu for course materials.

Course Format and Teaching Methods

The course will be delivered in-person lectures and discussions. No lab sections will be offered. We will use Piazza for online discussions outside the classroom: https://piazza.com/arizona/fall2024/csc585

Course Objectives

At the conclusion of this course, students should understand key algorithms for the processing of natural language, including traditional NLP algorithms and deep learning algorithms. Students are expected to have an in-depth understanding of these algorithms at the end of the course, including theoretical understanding and the ability to implement and evaluate them.

Expected Learning Outcomes

Students will:

- Demonstrate understanding of the rationale and complexity of key NLP algorithms.
- Implement and evaluate key NLP algorithms from scratch (using only Python) and using existing libraries and repositories (e.g., sklearn, Hugging Face).
- Be proficient in evaluation metrics and realistic evaluation frameworks.
- Describe the limitations of key NLP algorithms.
- Become proficient at reading research papers and link the content covered in the class with cutting-edge research.

Makeup Policy for Students Who Register Late

There will not be an opportunity to make up missed assignments or quizzes except for excused absences. Reach out to the instructor if you register after the first class.

Course Communications

Contact the instructor at eduardoblanco@arizona.edu or on Piazza. You can expect an answer within 24 hours Monday thorugh Friday.

Required Texts or Readings

There are no required Texts or materials, but we will use components of the freely available books below:

- Dan Jurafsky and James H. Martin. 2023. *Speech and Language* Processing (3rd Edition draft). https://web.stanford.edu/~jurafsky/slp3/
- Mihai Surdeanu and Marco A. Valenzuela. 2022. *Deep Learning for Natural Language Processing: A Gentle Introduction*. Cambridge University Press. https://clulab.org/gentlenlp/

Jacob Eisentein. *Introduction to Natural Language Processing*. MIT Press
https://github.com/jacobeisenstein/gt-nlp-class/tree/master/notes

Assignments and Examinations: Schedule/Due Dates

There will be five pop-up quizzes throughout the class. They will cover topics covered in class, research papers assigned as readings, and enhancements of homework assignments.

There will be six homework assignments. They will include both written questions to demonstrate knowledge of the topics covered in class as well as implementation of key NLP algorithms. Students can expect a new homework to be assigned every other week. Homework assignments will be out on August 27, September 10, September 24, October 10, October 31, and November 21. They will be due on September 5, September 24, October 8, October 24, November 14, and December 5.

The dates when homework assignments are out and due are subject to small changes.

Final Examination

The final exam will be on December 18, 2024, 1:00pm—3:00pm, as decided by the Office of the registrar: https://registrar.arizona.edu/faculty-staff-resources/room-class-scheduling/schedule-classes/final-exams

Grading Scale and Policies

The grade distribution for the course will be calculated as follows:

- Quizzes 10%
- Homework 60%
- Midterm exam 15%
- Final exam 15%

There will be 5 quizzes and 6 homework assignments. Some homework assignments may consist of several parts. Many homework assignments will have a substantial programming component. The quiz grade will be calculated with the top-4 highest-graded quizzes. Exams may have a take-home component.

Incomplete (I) or Withdrawal (W):

Requests for incomplete (I) or withdrawal (W) must be made in accordance with University policies, which are available at https://catalog.arizona.edu/policy/courses-credit/grading/grading-system.

Scheduled Topic and Activities

The class will cover the following topics:

- Introduction [1-2 weeks]
  - NLP problems and tasks
  - learning to solve NLP problems
  - baselines and evaluation
- Traditional NLP [3-4 weeks]
  - regular expressions,
  - n-gram language models
  - POS tagging and sequence labeling, HMMs
  - dependency parsing, the shift-reduce algorithm
  - semantic role labeling, argument identification and classification
- Deep learning background [2-3 weeks]
  - the perceptron, logistic regression
  - feed-forward neural networks
  - best practices
- NLP with deep learning [5-6 weeks]
  - word representations
  - recurrent neural networks
  - convolutional neural networks
  - transformers

In addition to lectures, the class will include discussions. Students will read research papers related to the topics covered in class and discuss them with the instructor in class. The slides prepared by the instructor will be the main reading materials. Slides will be shared with students before lectures take place.

Classroom Behavior Policy

To foster a positive learning environment, students and instructors have a shared responsibility. We want a safe, welcoming, and inclusive environment where all of us feel comfortable with each other and where we can challenge ourselves to succeed. To that end, our focus is on the tasks at hand and not on extraneous activities (e.g., texting, chatting, reading a newspaper, making phone calls, web surfing, etc.).

Notification of Objectionable Materials

This course in unlikely to contain material of a mature nature, which may include explicit language, depictions of nudity, sexual situations, and/or violence. If applicable, the instructor will provide advance notice when such materials will be used. Students are not automatically excused from interacting with such materials, but they are encouraged to speak with the instructor to voice concerns and to provide feedback.

Safety on Campus and in the Classroom

For a list of emergency procedures for all types of incidents, please visit the website of the Critical Incident Response Team (CIRT): https://cirt.arizona.edu/case-emergency/overview

Also watch the video available at

https://arizona.sabacloud.com/Saba/Web_spf/NA7P1PRD161/app/me/ledetail;spf-url=common%2Flearningeventdetail%2Fcrtfy000000000003841

University-wide Policies link

Links to the following UA policies are provided here: https://catalog.arizona.edu/syllabus-policies

- Absence and Class Participation Policies
- Threatening Behavior Policy
- Accessibility and Accommodations Policy
- Code of Academic Integrity
- Nondiscrimination and Anti-Harassment Policy

Department-wide Syllabus Policies and Resources

Links to the following departmental syllabus policies and resources are provided here: https://www.cs.arizona.edu/cs-course-syllabus-policies

- Department Code of Conduct
- Class Recordings
- Illnesses and Emergencies
- Obtaining Help
- Preferred Names and Pronouns
- Confidentiality of Student Records
- Additional Resources

- Land Acknowledgement Statement

Subject to Change Statement

Information contained in the course syllabus, other than the grade and absence policy, may be subject to change with advance notice, as deemed appropriate by the instructor.