USC Viterbi School of Engineering

CSCI 544: Applied Natural Language Processing

Syllabus

Units: 4

Term — Day — Time: Fall 2019- Wed. and Fri. 8:00 am to 9:50am
Location: SAL101

Instructor: Anna Farzindar, Ph.D.
Office: GER 202B
Regular Office Hours: Wed. after class
Contact Info: farzinda@usc.edu

Catalogue Course Description

Introduction to key components of human language technologies, including: information extraction, sentiment analysis, question answering, machine translation.

Expanded Course Description

The course will teach Natural Language processing algorithms. It will have an applied focus, in that it is meant for preparing students to utilize topics in NLP to solve real world problems.

This course is designed to introduce some of the problems and solutions of NLP, and their relation to linguistics and statistics. Students need to know how to program and use common data structures.

Recommended Preparation

A basic understanding engineering principles is required, including basic programming skills; familiarity with the Python language is desirable. Most assignments are designed for the Unix environment; basic Unix skills will make programming assignments much easier. Students will need sufficient mathematical background, including probability, statistics, and linear algebra. Some knowledge of machine learning is helpful, but not required.

Course Notes

The course will be run as a lecture class with student participation strongly encouraged. There are weekly readings and students are encouraged to do the readings prior to the discussion in class. All of the course materials, including the readings, lecture slides, home works will be posted online on USC Blackboard.

Technological Proficiency and Hardware/Software Required

Students are expected to know how to program in a language such as Python. Students are also expected to have their own laptop or desktop computer where they can install and run software to do the weekly homework assignments.

Required Readings and Supplementary Materials

- None required; *Speech and Language Processing, 2nd Edition* is optional but out of date https://web.stanford.edu/~jurafsky/slp3/
  - https://www.amazon.com/Speech-Language-Processing-Daniel-Jurafsky/dp/0131873210/
  - The new version is preferred https://web.stanford.edu/~jurafsky/slp3/ed3book.pdf
- Manning & Schütze (recommended)
- Farzindar and Inkpen, Natural Language Processing for Social Media, Second Edition

- All required readings from the textbooks and elsewhere will be listed in the syllabus.

In addition to the textbook, students may be given additional reading materials such as research papers. Students are responsible for all assigned reading assignments.

Description and Assessment of Assignments

Homework Assignments: There will be 4 homework assignments and a final project. The assignments and final project must be done individually. Each assignment is graded on a scale of 0-100 and the specific rubric for each assignment is given in the assignment. Each submission will be checked for plagiarism.

Grading Breakdown

Quizzes: There will be weekly quizzes based on the material from the week before. There is no mid-term for this class.

Comprehensive Exam: There will be an exam at the end of the semester covering all of the material covered in the class.

Grading Schema:

Quizzes	30%
Homework	40%
Comprehensive Exam	15%
NLP Project	15%
Total	100%

Grades will range from A through F. The following is the breakdown for grading:

94 – 100 = A	74 – 77 = C
90 – 94 = A-	70 –74 = C-
87 – 90 = B+	67 – 70 = D+
84 – 87 = B	64 – 67 = D
80 – 84 = B-	60 – 64 = D-
77 – 80 =C+	Below 60 is an F

Assignment Submission Policy

Homework assignments are due at 11:59 pm on the due date and should be submitted in Blackboard. You can submit homework up to one week late, but you will lose 20% of the possible points for the assignment. After one week, the assignment cannot be submitted. Every student has FIVE free late days for the homework assignments. You can use these five days for any reason separately or together to avoid the late penalty. There will be no other extensions for any reason. You cannot use the free late days after the last day of the class.

Schedule

Warning: The schedule below may change. Links to future lectures and assignments are just placeholders and will change.

Week	Wednesday	Friday	Suggested Reading
8/26	**Introduction** ☐ Why is NLP hard? ☐ Levels of language ☐ NLP applications ☐ Random language via n-grams	**Probability concepts** ☐ Joint & conditional prob ☐ Chain rule and backoff ☐ Modeling sequences ☐Cross-entropy and perplexity	• Intro: J&M chapter 1 • Chomsky hierarchy: J&M 16 • Prob/Bayes: M&S 2
9/2	Assignment 1 given **Modeling grammaticality** ☐ What's wrong with n-grams? ☐ Regular expressions, FSAs, CFGs, ...	**N-gram language models** ☐ Language ID ☐ Text categorization ☐ Spelling correction ☐ Segmentation ☐ Speech recognition ☐ Machine translation	• Language models: J&M 3 • Huddleston
9/9	**Bayes' Theorem** and **Smoothing n-grams**	*Assignment 1 Given (Sep 13th)*	☐ Smoothing: J&M 3 4; Rosenfeld (2000)

	☐ Maximum likelihood estimation ☐ Bias and variance ☐ Add-one or add-λ smoothing ☐ Cross-validation ☐ Smoothing with backoff ☐ Good-Turing, Witten-Bell	**Intro to neural language Models** ☐ Conditional log-linear models ☐ Maximum likelihood, regularization ☐ Feedforward neural language Models ☐ Recurrent neural language Models	☐ Log-linear models: Collins (pp. 1-4)
9/16	Guest Lecture or applied NLP (TBD)	Language Models **Context-free parsing** ☐ What is parsing? ☐ Why is it useful? ☐ Brute-force algorithm ☐ CKY and Earley algorithms	☐ Attributes: J&M 15 ☐ Parsing: J&M 13
9/23	*Assignment 1 due (Sep 23rd)* **Earley's algorithm** ☐ Top-down parsing ☐ Earley's algorithm	*Assignment 2 Given (Sep 27th)* **Probabilistic parsing** ☐ PCFG parsing	☐ CCG: Steedman & Baldridge; more ☐ TAG/TSG: Van Noord, Guo, Zhang 1/2/3 ☐ Prob. parsing: J&M 14

		☐Dependency grammar ☐ Lexicalized PCFGs	
9/30	NLP for Social media	*Student project presentation (NLP for Social media)*	Farzindar and Inkpen, Natural Language Processing for Social Media, Second Edition
10/7	*Assignment 2 due (Oct 7th)* **Semantics** ☐ What is understanding? ☐ Lambda terms ☐ Semantic phenomena and representations ☐ More semantic phenomena and representations	**Distributional semantics** (word embeddings) ☐Compositional semantics ☐ Distributional semantics	☐ Semantics: J&M 17-18; up to but not including "denotational semantics" section; lambda calculus for kids ☐ Forward-backward: J&M 6
10/14	*Project proposal due* Assignment 3 given: Semantics **Sequence tagging models** (Excel spreadsheet; Viterbi version; lesson plan; video lecture) ☐ Ice cream, weather, words and tags ☐ Forward and backward probabilities ☐ Inferring hidden states	No class (fall break)	

	☐ Likelihood convergence ☐ Local maxima		
10/21	*Assignment 3 Given(Oct 21)* **Expectation Maximization** ☐ Generalizing the forward-backward strategy ☐ Inside- outside algorithm ☐ Posterior decoding	**Neural sequence tagging models** ☐ Recurrent Neural Networks ☐ Conditional Random Fields ☐ Neural CRFs	☐ Inside-outside and EM: J relation to backprop
10/28	Neural Sequence Tagging Guest Lecture (TBD)	*Assignment 3 due(Nov 1)* **Text classification** ☐ Features ☐ Linear Classifiers	
11/4	*Project proposal revision (if applicable) due* *Assignment 4 Given (Nov 6th)* Guest Lecture or applied NLP (TBD)	**Neural text classification** ☐ CNN ☐ LSTM ☐Contextualized Embeddings	
11/11	**Morphology and phonology** ☐ Stemming ☐ Compounds, segmentation	*Assignment 4 due (Nov 15th)* **Multilinguality and machine translation**	☐ Morphology: R&S 2

	☐ Two-level morphology ☐ Punctuation ☐ Rewrite rules	☐ Intro to MT ☐ Evaluations	
11/18	**Sequence to sequence models** ☐ RNN-based seq-to-seq models ☐ Transformers ☐ Applications to machine translation, summarization.	**Current NLP tasks and competitions** ☐ The NLP research community ☐ Machine Translation ☐ Question Answering ☐ Dialog Systems	☐ MT: J&M 25, M&S 13, statmt.org;tutorial (2003), workbook (1999),introductory essay (1997), technical paper(1993); tutorial (2006) focusing on more recent developments (slides, 3-hour video part 1, part 2)
11/25	**No class** (Thanksgiving break)	**No class** (Thanksgiving break)	
12/2	**Applied NLP continued**	**Comprehensive exam December 6th** *Project final due* Wednesday, December 11	

Statement on Academic Conduct and Support Systems

Academic Conduct

Plagiarism – presenting someone else's ideas as your own, either verbatim or recast in your own words – is a serious academic offense with serious consequences. Please familiarize yourself with the discussion of plagiarism in *SCampus* in Section 11, *Behavior Violating University Standards* https://policy.usc.edu/student/scampus/part-b/. Other forms of academic dishonesty

are equally unacceptable. See additional information in *SCampus* and university policies on scientific misconduct, http://policy.usc.edu/scientific-misconduct.

Discrimination, sexual assault, and harassment are not tolerated by the university. You are encouraged to report any incidents to the *Office of Equity and Diversity* http://equity.usc.edu or to the *Department of Public Safety* http://adminopsnet.usc.edu/department/department-public-safety. This is important for the safety of the whole USC community. Another member of the university community – such as a friend, classmate, advisor, or faculty member – can help initiate the report, or can initiate the report on behalf of another person. *The Relationship and Sexual Violence Prevention Services* http://engemannshc.usc.edu/rsvp/ provides 24/7 confidential support, and the sexual assault resource center webpage http://sarc.usc.edu describes reporting options and other resources.

Support Systems

A number of USC's schools provide support for students who need help with scholarly writing. Check with your advisor or program staff to find out more. Students whose primary language is not English should check with the *American Language Institute* http://dornsife.usc.edu/ali, which sponsors courses and workshops specifically for international graduate students. *The Office of Disability Services and Programs* http://sait.usc.edu/academicsupport/centerprograms/dsp/home_index.html provides certification for students with disabilities and helps arrange the relevant accommodations. If an officially declared emergency makes travel to campus infeasible, *USC Emergency Information* http://emergency.usc.edu will provide safety and other updates, including ways in which instruction will be continued by means of blackboard, teleconferencing, and other technology.

Resources for Online Students

The Course Blackboard page has many resources available for students enrolled in our graduate programs. In addition, all registered students can access electronic library resources through the link https://libraries.usc.edu/.