COSI 134: Statistical (and Neural) approaches to Natural Language Processing

Fall 2018

Instructor: Nianwen Xue

- [http://www.cs.brandeis.edu/~xuen](http://www.cs.brandeis.edu/~xuen)
- Email: xuen@brandeis.edu
- Office: 253 Volen
- Office hours: Thursday, 3-5pm or by appointment

Teaching Assistants:

Yuchen Zhang

- Email: yuchenz@brandeis.edu
- Office: 110 Volen
- Office hours: Wed 3-5 or by appointment

TBA:

- Email: TBA
- Office: TBA
- Office hours: TBA

Time: 12:30am – 1:50pm, Tue/Fri

Location: Gerstenzang 122

Class website: [http://latte.brandeis.edu](http://latte.brandeis.edu)

Course Learning Objectives:

The objective of this course is to provide an introduction to statistical and neural network based approaches to Natural Language Processing. Upon completion of the course, the student is expected to 1) have a general understanding of how to design statistical and neural network models and algorithms to solve common Natural Language Processing problems such as text classification (e.g., document classification, sentiment analysis, word sense disambiguation, semantic role labeling), sequence labeling (e.g., part-of-speech tagging, named entity recognition, word segmentation, text chunking), as well as structured prediction (e.g. syntactic, semantic, and discourse parsing), 2) be able to gain the necessary knowledge to understand the technical literature in NLP that is published in conference proceedings such as the Annual Meetings of ACL and computational linguistics journals such as *Computational Linguistics* and *Transactions of ACL*. Statistical and neural network based approaches currently dominate the field of NLP and it is hard to understand the cutting edge research without having a good understanding of these statistical and neural network based approaches.
Course Description:

The field of Natural Language Processing is moving in a very fast pace, thanks primarily to the advancement in statistical and neural network based models. Statistical methods dominated the field of NLP in the last two decades (roughly from early 1990s), and now neural network based approaches (also called “deep learning”) are making a massive comeback. Major NLP conferences such as the Annual Meeting of the Association of Computational Linguistics are dominated by technical presentations using neural network based approaches to NLP problems. In this course we will first introduce the statistical models and algorithms that have been the working horse for the last two decades, and then transition into an overview of the neural networks models that are currently in use today. We will provide a quick overview of the mathematical background that is needed for the course (probability theory, information theory, basic calculus, and linear algebra), but knowledge of these areas is generally assumed. We will then embark on an examination of the fundamental statistic models such as Naïve Bayes, Maximum Entropy, Perceptron as well as their adaptations to sequence labeling and structured prediction problems. Finally we will provide an overview of neural network models such as Multi-level Perceptron (MLP) and representation learning techniques such Convolutional Networks (CNNs) and Recurrent Neural Networks (RNNs). The programming assignments will be in Python as the popular neural network libraries are implemented in Python. If you don’t know Python already, it is recommended that you pick it up quickly before the course starts.

Tentative schedule:

Week 1: Introduction and preliminaries

Weeks 2-6: Simple classification problems

Topics: Naïve Bayes, logistical regression (binary & multiclass), perceptron (binary & multiclass), neural network models (Multi-level Perceptron, Convolutional Networks, Recurrent Networks, distributed representations)

Weeks 7-9: Sequence labeling

Topics: Hidden Markov Models (HMM), Conditional Random Fields (CRF), Perceptron for sequence labeling, RNN-CRF, Unsupervised learning (EM and backward-forward)

Weeks 10-12: Structured Prediction

Topics: Generative models (CKY parsing, lexicalized and latent PCFG), Discriminative (shift-reduce with Perceptron learning, shift-reduce with neural network learning), transition-based and graph-based dependency parsing, neural dependency parsing, Unsupervised parsing (Inside-outside)

Weeks 13: Misc. and Wrap-up
Textbooks:

1. Required: Yoav Goldberg, *Neural Network Methods for Natural Language Processing*
2. Recommended: Manning and Schultze, *Foundations of Statistical Natural Language Processing*
3. Recommended: Daniel Jurafsky and James Martin. *Speech and Language Processing*

Grading:

- Participation: 10%
- Quizzes (mid-term 10% and final 10%): 20%
- Written homework Assignments (4-6): 20%
- Programming assignments (4): 50%

Academic Integrity:

You should finish homework assignments, exams, and project reports on your own unless a project is explicitly stated as a collaborative project. Late projects are subject to grade deduction.