CSE325/425 Natural Language Processing

Fall 2019

Instructor: Sihong Xie   **Email:** six316@lehigh.edu

Time: Mon/Wed 12:40-1:55 pm   **Place:** BC 115.

Catalog Description Overview of modern natural language processing techniques: text normalization, language model, part-of-speech tagging, hidden Markov model, syntatic and dependency parsing, semantics, word sense, reference resolution, dialog agent, machine translation. Three projects to design, implement and evaluate classic NLP algorithms. Credit will not be given for both CSE 325 and CSE 425.

Prerequisites Probability and statistics (MATH 231 or ECO 045) and programming experience (CSE 017).

Communication

- https://piazza.com/   For questions answering and notifications.

- https://coursesite.lehigh.edu/   For posting grades only.

- http://www.cse.lehigh.edu/~sxie/teaching/2019_fall_nlp/nlp.html   Course website has the most up-to-date information (lecture notes, codes, datasets, references).

Office Hours Monday/Wednesday 4:00 pm - 5:00 pm, or by appointment, BC 326.

Textbooks They are all optional.

- SLP3 = *Speech and Language Processing.* Dan Jurafsky and James H. Martin. 3rd editions.

- FSNLP = *Foundations of statistical natural language processing.* C. Manning and H. Schütze. MIT Press, 2000.

- DLNLP = *Deep Learning for Natural Language Processing.* Stephan Raaijmakers. Manning Publications. 2019.

- DL = *Deep Learning.* Ian Goodfellow and Yoshua Bengio and Aaron Courville. MIT Press. 2016.

- NNMNLP = *Neural Network Methods for Natural Language Processing.* Yoav Goldberg. Morgan & Claypool Publishers. 2017.

Programming Assignments

We will provide datasets, code sketches, the GPU computing facilities through SandBox (PA 112) or Google Colab. Your programs will be written in Python using the PyTorch framework and will be tested on a SandBox machine.

- Word embedding (10%).

- LSTM-CRF POS tagging (25%).

- Relation extraction (25%).

Each student has a quota of 6 late days at his/her disposal for all assignments and the project. When the quota runs up, we start penalize late submissions 20% for each late day (24 hours or part thereof) and no assignment will be accepted more than four days after its due date. The assignment will be graded partly based on your programs' running time and accuracy. Requests for re-grading must be made within 48 hours after the grades are released. **For graduates** There is an extra research project involving proposal (5%), discussion (5%), presentation (10%), and written report with experimental results (20%). **For undergraduates** Only the three programming assignments need to be submitted to get the full credit.

Accommodations for Students with Disabilities If you have a disability for which you are or may be requesting accommodations, please contact both your instructor and the Office of Academic Support Services, Williams Hall, Suite 301 (610-758-4152) as early as possible in the semester. You must have documentation from the Academic Support Services office before accommodations can be granted.

Principles of Our Equitable Community Lehigh University endorses The Principles of Our Equitable Community. We expect each member of this class to acknowledge and practice these Principles. Respect for each other and for differing viewpoints is a vital component of the learning environment inside and outside the classroom.

Academic Integrity The work you submit must be entirely your own. While discussions of basic concepts covered in class with classmates are encouraged, plagiarism is never acceptable, and various methodswill be used to detect unreasonably similar copies of codes submitted. Such cases will be referred to the University Committee on Discipline and, if you are found guilty, you may be given the failing grade WF in the course. If you have questions about this policy at any point, ask me. It is far better to be safe than sorry when your academic career may be on the line.