CIS*4500 (Fall 2019) Special Topics

Machine Learning for Natural Language Processing

Instructor: *Fei Song* (Reynolds 2225, ext. 58067)
Email: fsong@uoguelph.ca
Office Hours: TBA
Course Website: https://courselink.uoguelph.ca/

Overview

Natural Language Processing is an interdisciplinary area among Mathematics, Information Theory, Linguistics, and Computer Science. It has been successfully applied to a wide range of problems such as Speech Recognition, Information Retrieval and Extraction, Text Classification, Sentiment Analysis, Automatic Summarization, and Machine Translation. Due to the sequence nature of language data, similar techniques have also been applied to other domains such as Biological and Time Series data analysis. This course provides an introduction to this emerging field, with emphasis on the applications of machine learning techniques. The students should have proficiency in Java and working experience in Python, as well as good understanding of the related mathematics such as Probability and Statistics. Through warmup assignments and a term project, the students will build good development skills in applying machine learning techniques to real world problems. The understanding of key concepts and techniques will be tested through exams. Listed in the following is a set of topics that we intend to cover in the course, subject to changes based on the availability of time:

- Introduction to Natural Language Processing
- Language Modeling and N-gram Models
- Information Retrieval Models and Implementations
- Machine Learning for Text Classification and Sentiment Analysis
- Introduction to Deep Learning Models
- Unsupervised Machine Learning Methods

Prerequisites: CIS*2750 and STAT*2040

Evaluation

- Warmup Assignments: 2 x 12 = 24% (due on Oct. 6 and Nov. 3, respectively)
- Term Project: 26% (due on Nov. 28)
- Midterm: 20% (scheduled on Oct. 24)
- Final Exam: 30% (scheduled on Dec. 10 by the Registrar's office)

Recommended References

Chris Manning and Hinrich Schültz. *Foundations of Statistical Natural Language Processing*. The MIT Press, 1999.

Ian H. Witten, Eibe Frank, Mark A. Hall, and Christopher J. Pal. *Data Mining: Practical Machine Learning Tools and Techniques*. Fourth Edition. Morgan Kaufmann, 2016.

Sebastian Raschka and Vahid Mirjalili. *Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow*. Second Edition. Packt Publishing, 2017.

Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Francis Bach. *Deep Learning*. The MIT Press, 2016.