MSCI 641 – Text analytics

Spring 2019

Lectures: Tuesday, Thursday 10:00AM-11:20PM (E2 1732)

Instructor: Olga Vechtomova, DWE 2513M, ovechtom@uwaterloo.ca, ext. 32675

TA: Amirpasha Ghabussi, aghabussi@uwaterloo.ca

TA Office hours: TBD

Course outline: With the rapid growth of unstructured natural language data, such as web pages, blogs, product reviews, news articles and enterprise data, there is an increasing need for systems that would retrieve relevant documents, extract specific information from them, mine opinions, summarize and categorize texts. This course provides students with an understanding of the major methods for retrieving, mining and analyzing textual data, with the emphasis on algorithms, techniques and their evaluation.

Syllabus:

Week 1	Introduction to natural language processing (NLP). Linguistics Essentials.
Week 2	Foundations of text processing: tokenization, stemming, stopwords, lemmatization, part-of-speech tagging, syntactic parsing.
Week 3	Word association measures. Distributional word similarity.
Week 4	Probabilistic language modelling. N-grams, perplexity, smoothing. Text classification. Naïve Bayes. Evaluation of text classification systems.
Week 5	Introduction to Neural Networks (NNs) for NLP. Feed-forward NNs, activation functions, cross-entropy loss, Mean Squared Error (MSE) loss, word embeddings.
Week 6	Recurrent neural networks (RNNs) for text classification. Attention mechanisms.
Week 7	Convolutional neural networks (CNNs). Sentence classification with CNNs.
Week 8	Autoencoders and their NLP applications.
Week 9	Sequence-to-sequence models. Machine translation. Dialogue systems.
Week 10	Adversarial and multi-task learning for NLP.
Week 11	Other current research topics in NLP (TBD)
Week 12	Other current research topics in NLP (TBD)

Grading scheme:

25% - assignments
15% - 3 paper reviews
60% - course project

Recommended textbooks:

- Dan Jurafsky and James H. Martin. Speech and Language Processing (3rd ed. draft)
- Jacob Eisenstein. Natural Language Processing
- Yoav Goldberg. A Primer on Neural Network Models for Natural Language Processing
- Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Deep Learning