ZHANG Xu

Email: xux.zhang@mail.utoronto.ca

Tel.: (647) 675-7767

Education Background

2012-2016 **Huazhong University of Science and Technology**

Naval Architecture and Ocean Engineering (B.Eng.)   GPA: 3.41/4.0

2016-2018   **University of Toronto**

Industrial Engineering (M.Eng.)

Project Experience

Course project: **Identifying user roles based on user behavior of Github**

MIE 1512: Data anlytics

➢ Download and merge historical event log in Jan 2017 of Github using Github Archive. More than 80 GB raw data is processed in total.

➢ Clean the raw data, convert and normalize it to the feature vectors that represent each user.

➢ User vectors are fed into the K-means|| model and the number of K are determined by the metrics of WSSSE(Within Set Sum of Squared Error) and Beta-Coefficient of variance.

➢ Four different user roles are identified by averaging features of each cluster produced by K-means. The user role and its composition are analyzed regarding to Github's further development.

Course project: **Image Scene Classification**

CSC411: Introduction to machine learning

➢ Used various techniques for the preprocessing of images including PCA, GIST(For feature extraction) and some small techniques like flipping, rotating the images as supplement to data and transfer RGB images to L mode images.

➢ Computed the results of different training models including KNN, MOG, Neural Network and deep learning model inception-v3.

➢ Tunned the parameters of each model to enhance their performance and compare different models.

➢ Used Adaboost to ensemble different classifiers into a weighted sum and obtained a very high classify accuracy at 85%.

Course project: **Data analytics of universities in Canada**

MIE 1624: Introduction to data science and analytics

➢ Scrap reviews and ratings of Canadian universities on niche.com with Beautifulsoup.

➢ The reviews are cleaned, tokenized with NLTK and transferred to Tf-idf matrix.

➢ Perform Non-negative Matrix Factorization on the Tf-idf matrix to generate topics of each university.

➢ Using the topics generated and the ratings on the website, the performance of Canadian universities are analyzed and compared with top universities in USA.

Technical skills

Programming: Python(Pyspark, Numpy, Pandas, Tensorflow) , SQL, Java, R

Data Engineering: Web scraping, API(eg. Twitter api, Facebook api etc.), Spark.

Machine learning: Supervised learning (Regression, Neural Network, KNN, Naive Bayes Classifier, SVM), Unsupervised learning(Clustering, Topic modeling, Mixture models), etc.