Yuqing. **Xie**

Ph.D. Student in Computer Science (Natural Language Processing)

amyxie361@outlook.com | amyxie361 | Yuqing (Sherry) Xie

Education

Cheriton School of Computer Science, University of Waterloo

Waterloo, ON, Canada

Ph.D. Student in Computer Science — *Sept. 2018 - Present*

- Supervisor: Prof. Ming Li and Prof. Jimmy Lin.
- Research interest: Natural language processing, Question answering, Information retrieval, Generation

School of Mathematical Sciences, Fudan University

Shanghai, China

B.S. in Mathematics and Applied Mathematics — *Sept. 2014 - June. 2018*

- GPA:3.6/ Major 3.5/4.0(Overall)
- In the Honor Class of the National Basic Subject Top-notch Talent.
- Thesis: Biological Question Answering System based on Neural Network (Supervisor: Yiming Wei and Shanfeng Zhu)
- Selected Courses: Abstract Algebra, Real Analysis, Functional Analysis, Time Series, Information Theory
- Academic Seminars: Set Theory, Galois Theory, Communicative Algebra, Functional Analysis, Neural Network and Deep Learning, Quantitive Trading.

Publication

Data Augmentation for BERT Fine-Tuning in Open-Domain Question Answering

arXiv:1904.06652

Wei Yang*, **Yuqing Xie***, Luchen Tan, Kun Xiong, Ming Li, Jimmy Lin (*Both authors contributed equally) — *2019*

End-to-End Open-Domain Question Answering with BERTserini

NAACL, demo, arXiv:1902.01718

Wei Yang*, **Yuqing Xie***, Aileen Lin, Xingyu Li, Luchen Tan, Kun Xiong, Ming Li, Jimmy Lin (*Both authors contributed equally) — *2019*

Research Experiences

Chatbot Module Development - Research Assistant

Advisor:Prof. Ming Li, and Prof. Jimmy Lin

University of Waterloo & RSVP.ai — *Sept. 2018 - Present*

- Developed an open domain question answering module for the chatting robot.
- Constructed an end-to-end question answering system that integrates BERT with the open-source Anserini information retrieval toolkit both in English and Chinese and create new state of the art.
- Transferred code from tensor-flow implementation to pytorch.
- Improved the system's performance by 10% exact match rate on SQuAD 1.1 under open-domain setting using text augmentation.
- Tested the system's performance with Elastic Search API and provided real-time online service.
- Applied the system to domain specific document information retrieval.
- Applied a BERT based named entity recognition model to contract key information extraction.
- Implemented and compared different paraphrase generation models.

Understanding How Human Generate Questions

Advisor: Prof. Ori Friedman

University of Waterloo — *Sept. 2019 - Presernt*

- Surveyed the source of curiosity from psychology perspective.
- Surveyed the theory on linguistic perspective on question language.
- Analyzed the language features on two datasets: SQuAD and Quora Quesition Pair.

Paper Recommendation using GraphX

Advisor: Prof. Jimmy Lin

University of Waterloo — *Jan. 2019 - April 2019*

- Applied GraphX to build an academic paper recommendation system.
- Implemented PageRank, keyword filtering, and pattern finding algorithms in GraphX.
- Applied the algorithm on a citation network to give recommendations of papers, taking users' interest into account.

Contextual Decomposition for Rationalizing LSTM Predictions

Advisor: Prof. Yaoliang Yu

University of Waterloo — *Sept. 2018 - Nov 2018*

- Decomposed and analyzed LSTM model in token level to understand the effectiveness source of the model based on entity detection task.
- Implemented and modified multi-view concept to understand the learned weight in the two-directions of LSTM model.
- Traced the source of the effective source of LSTM models and concluded the most effectiveness comes from embedding.

Car Detection project - Yitu-Tech

Shanghai, China

Machine Learning Algorithm Inter-ship

Feb. 2018 - Aug. 2018

- Improved the Single Shot MultiBox Detector model for object detection on car detection task.
- Implemented the HOG-SVM for digital recognition in car license detection.

Question Answering Based on Biomedical Paper Data base - BioASQ

Advisor: Prof . Shanfeng Zhu

Shanghai Key Laboratory of Intelligent Information Processing, Fudan University

July. 2017 - April. 2018

- Introduce occurrence possibility of words as the representation of answers for question answering.
- Adapted source code from TensorFlow for text processing including applying skip gram algorithm for word embedding matrix training.
- Constructed Bidirectional LSTM Recurrent Neural Networks under TensorFlow framework.
- Introduce attention mechanism based on questions for better performance in question answering.
- Adjusted the model and achieved average Factoid MRR of 0.1615, List F measure of 0.1353.

Place of Interest Prediction on Weibo Dataset - NLPCC 2017 shared task

Advisor:Prof Zho ngyu Wei

Fudan University

April. 2017 - June. 2017

- Introduced bipartite graph structure for similarity measuring according to user visiting records. Clustered users and spots based on semantic similarity of tweets, Euclidean distance of average visit spots and cosine similarity of other extracted features using K-means, KNN
- Developed a user info based recommendation system for future visiting prediction. Applied recsys package to train and achieved a top 20% F1 score of 0.021 in POI recommendation.
- Predicted user gender profile according to heterogeneous data using models like SVM, Naive Bayes, Logistic Regression, and further analyzed and compared ensemble results of Random Forest, AdaBoost an d Voting and Bagging.
- Processed heterogeneous data from Weibo (Twitter of China) including basic user tweets contents and site access records into semi structured data by feature extraction.
- Applied packages of FudanNLP, jieba for Chinese text segmentation, word frequency statistics; PCA for dimension reduction; word2vec for embedding.

Spread Trading Strategy Using Math Models

Advisor: Prof . Donghua Zhao

Fudan University

Mar. 2016 - May. 2016

- Implemented a framework of programmed trading system and simulated spread trading strategy.
- Forecasted future trend of short period stock prices of Copper-Zinc spread trading using models including Autoregressive Moving Average Model , Kalman Filtering and Hidden Markov Model.
- Applied forecasting model for new trading strategy and achieved annual profit rate of 21%

Honors & Awards

In Canada

2018-2020 **Scholarship**, UW Grad Scholarship — *Waterloo, Canada*

In China

2015-2017 **Scholarship**, Outstanding Students of Fudan University — *Shanghai, China*

2016 **Honorable Mention**, COMAP's Mathematical Contest In Modeling — *Shanghai, China*

2011-2013 **First Prize (Best Female Participant in 2013)**, National Olympiad in Informatics — *Jiangsu, China*

2013 **First Prize**, Chinese Physics Olympiad — *Jiangsu, China*

2012 **First Prize**, Chinese Mathematical Olympiad — *Jiangsu, China*

Teaching

University of Waterloo

Waterloo, ON, Canada

Teaching Asistant

- Teaching assistant for CS 651/451(Data-Intensive Distributed Computing), CS 245 (Logic and Computation),CS 136 (Elementary Algorithm Design and Data Abstraction), CS 246 (Object-Oriented Software Development).

Skills

Frameworks Pytorch, Tensorflow, Keras, Spark, Hadoop

Programming Python, C, Java, LaTeX, MATLAB, Scala

Languages Chinese, English, Japanese