Yuqing. **Xie**

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

yuqing.xie@uwaterloo.ca | amyxie361.github.io | 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

Distant Supervision for Multi-Stage Fine-Tuning in Retrieval-Based Question Answering — *TheWebConf 2020, Short paper, Oral*

Yuqing Xie, Wei Yang, Luchen Tan, Kun Xiong, Nicholas Jing Yuan, Baoxing Huai, Ming Li and Jimmy Lin — *2020*

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 2019, Demo track*

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*

- Constructed an end-to-end question answering system that integrates BERT (**Tensorflow**) with the open-source Anserini (a **Lucene** IR toolkit) information retrieval toolkit both in English and Chinese and create new state of the art.
- Transferred code from **Tensorflow** implementation to **PyTorch**.
- Improved the system's performance by 10% exact match rate on SQuAD 1.1 under open-domain setting using text augmentation and established new baselines on WebQuestion, CMRC and DRCD datasets.
- 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 bi-directional LSTM, VAE and GPT-2 as 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 and compared the framework against **MapReduce** on **Hadoop**.
- 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 with **PyTorch**.
- 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.

Yitu-Tech *Shanghai, China*

Machine Learning Algorithm Intership *Feb. 2018 - June 2018*

- Worked on the car detection project.
- 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.

BioASQ: Question Answering Based on Biomedical Paper Database *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 and introduced attention mechanism based on questions.
- Adjusted the model and achieved average Factoid MRR of 0.1615, List F measure of 0.1353.

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 (Top 30%)**, COMAP's Mathematical Contest In Modeling *Shanghai, China*

2011-2013 **First Prize (Best Female Participant in 2013) (Top 0.1%)**, National Olympiad in Informatics *Jiangsu, China*

2013 **First Prize (Top 1%)**, Chinese Physics Olympiad *Jiangsu, China*

2012 **First Prize (Top 1%)**, Chinese Mathematical Olympiad *Jiangsu, China*

Teaching

University of Waterloo *Waterloo, ON, Canada*

Teaching Assistant

- 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 and Tools PyTorch, Tensorflow, Keras, Spark, Hadoop, Lucene, Elastic Search, Git, Vim, MapReduce

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