Yuqing. **Xie**

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

yuqing.xie@uwaterloo.ca | amyxie361.github.io | amyxie361 | yuqing-xie-8bb213170 | 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, including Question answering, Information retrieval, Generation, Pretrain models

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)

Skills

DevOps	AWS, Docker, Spark, Hadoop, Git, Vim, MapReduce
Frameworks and Tools	PyTorch, Tensorflow, Keras
Programming Languages	Python, Bash, LaTeX, MATLAB, C, Java, Scala
Languages	Chinese, English, Japanese

Publication

Approximate Nearest Neighbor Search and Lightweight DenseVector Reranking in Multi-Stage Retrieval Architectures — *ACM SIGIR - ICTIR 2020*

Zhengkai Tu*,Wei Yang*,Zihang Fu*, **Yuqing Xie**, Luchen Tan, Kun Xiong, Ming Li, Jimmy Lin (*All authors contributed equally) — *September 2020.*

Distant Supervision for Multi-Stage Fine-Tuning in Retrieval-Based Question Answering — *WWW' 2020, pages 2934-2940.*

Yuqing Xie*, Wei Yang*, Luchen Tan, Kun Xiong, Nicholas Jing Yuan, Baoxing Huai, Ming Li and Jimmy Lin (*Both authors contributed equally) — *April 2020.*

Rapid Adaptation of BERT for Information Extraction on Domain-Specific Business Documents. — *arXiv:2002.01861*

Ruixue Zhang, Wei Yang, Luyun Lin, Zhengkai Tu, **Yuqing Xie**, Zihang Fu, Yuhao Xie, Luchen Tan, Kun Xiong, and Jimmy Lin — *February 2020.*

End-to-End Open-Domain Question Answering with BERTserini — *NAACL 2019, pages 72-77*

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

Industry Experience

Amazon Lex Team — *Seattle, U.S. (Remote)*

Applied Scientist Intern — *Aug. 2020 - Present*

- Research on applying a new method which have good performance on CV dataset to NLP tasks. Related to online learning and transfer learning.

RSVP.ai — *Waterloo, Canada*

Machine Learning Engineer — *Sept. 2019 - June. 2020*

- Implemented paraphrase generation models and conduced human evaluation for paraphrase evaluation.
- Conducted human labeling experiment comparing different models and confirm the effectiveness of BLEU evaluation.
- Improve question answering system with paraphrase-augmented queries and named-entity-filtered examples.

RSVP.ai

Waterloo, Canada

Machine Learning Engineer

Dec. 2018 - Sept 2019

- 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.
- Tested the system's performance with **Elastic Search API** and provided real-time online service.
- Applied the system to domain specific document information retrieval.

Yitu-Tech

Shanghai, China

Machine Learning Algorithm Intern

Feb. 2018 - Aug. 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.

Projects

Pertaining Sequence to Sequence Model for Chinese Conversation

Advisor: Ming Li, & Jimmy Lin

University of Waterloo

June 2020 - Present

- Trained sequence to sequence transformer models on Chinese forum datasets with **PyTorch**.
- Modified positional embedding with sentence and paragraph information.

Introduce tree-structure information into question generation

Advisor: Ming Li

University of Waterloo

Jan 2020 - April 2020

- Researched on add parsing tree into sentence generation to help generate more logic meaningful questions.
- Parsing tree encoder outperforms universal sentence encoder under the encoder-decoder framework.

Data Augmentation for Open Domain Question Answering.

Advisor: Ming Li, & Jimmy Lin

University of Waterloo

June 2019 - Sept 2019

- Researched on improving open domain question answering system's performance using distantly supervised data under the **TensorFlow** framework.
- Improved the system's performance by 10% exact match rate on SQuAD 1.1 and established new baselines on Chinese reading comprehension datasets: CMRC and DRCD.

Paper Recommendation using GraphX

Advisor: Adam Roegiest

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 recommend papers, taking users' interest into account.

Contextual Decomposition for Rationalizing LSTM Predictions

Advisor: 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 named 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.

BioASQ: Question Answering Based on Biomedical Paper Database

Advisor: Shanfeng Zhu & Yimin Wei

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.

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).