# Yuqing **Xie**

PH.D. STUDENT IN COMPUTER SCIENCE (NATURAL LANGUAGE PROCESSING)

yuqing.xie@uwaterloo.ca | amyxie361.github.io | amyxie361 | yuqing-xie-8bb213170 | Google Scholar


## 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: NLP applications, including Question Answering, Information Retrieval, and Model Regression Analaysis.

**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 BiDAF (Supervisor: Yiming Wei and Shanfeng Zhu)

## Publication

**Regression Bugs Are In Your Model! Measuring, Reducing and Analyzing Regressions In NLP Model Updates** — *ACL 2021*

YUQING XIE, YI-AN LAI, YUANJUN XIONG, YI ZHANG, STEFANO SOATTO — *May 2021.*

**Segatron: Segment-Aware Transformer for Language Modeling and Understanding** — *AAAI 2021*

HE BAI, PENG SHI, JIMMY LIN, **YUQING XIE**, LUCHEN TAN, KUN XIONG, WEN GAO, MING LI — *Feb 2021.*

**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

**RSVP.ai** — *Waterloo, Canada*

MACHINE LEARNING ENGINEER — *May 2021 - Present*

- Research on topics related to **language model pre-training**, question answering, and paraphrasing.

**Amazon AWS AI** — *Seattle, U.S. (Remote, Parttime)*

APPLIED SCIENTIST INTERN — *Aug. 2020 - April 2021*

- Explored **regression-free model update problem**, which aims to reduce the newly introduced bug by the new neural network systems.
- Proposed methods including constraint optimization learning, meta adaptor, data augmentation, to effectively reduce regression in NLU tasks.
- Published a paper in ACL 21' and will submit another to EMNLP 21' that relates to the project.

**RSVP.ai** *Waterloo, Canada*

Machine Learning Engineer *Dec. 2018 - June 2020*

- Implemented paraphrase generation models and conducted human labelling experiment comparing different models to verify the rationality of current evaluation methods.
- Improved the performance of open domain question answering system with paraphrase-augmented queries, retriever augmented context, and named-entity-filtered examples.
- 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.
- Deployed the QA system's **Elastic Search API** and provided real-time online service.

**Yitu-Tech** *Shanghai, China*

Machine Learning Algorithm Intern *Feb. 2018 - Aug. 2018*

- Improved the accuracy of the **Single Shot MultiBox Detector** model for car detection.
- Implemented the HOG-SVM for digital recognition in car license identification.

## Projects

**Tree-structure Information Integration for Question Generation** *Advisor: Ming Li*

University of Waterloo *Jan 2020 - April 2020*

- Explored integrate parsing tree knowledge for more logical question generation.
- The proposed method qualitatively introduces structure information compared with universal sentence encoder.

**Paper Recommendation via 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 recommend according to users' interest.

**Contextual Decomposition for Rationalizing LSTM Predictions** *Advisor: Yaoliang Yu*

University of Waterloo *Sept. 2018 - Nov 2018*

- Decomposed and analyzed LSTM models in token level to understand the behaviour on named entity detection task.
- Implemented and modified multi-view learner to analyze the difference of the two-directions of LSTM model.

## Honors & Awards

2018-2021 **Scholarship,** UW Grad Scholarship *Waterloo, Canada*

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*

## Skills

**DevOps** AWS(EC2, EFS, S3), 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

## Teaching

**University of Waterloo** *Waterloo, ON, Canada*

Teaching Assistant

- CS 651/451 (Data-Intensive Distributed Computing) - 19Fall, 20Winter, 20Fall, 21Winter
- CS 245 (Logic and Computation) - 19Spring, 21Spring
- CS 136 (Elementary Algorithm Design and Data Abstraction) - 19Winter
- CS 246 (Object-Oriented Software Development) - 18Fall, 20Spring