

Xin QIAN

xinq@terpmail.umd.edu · (412) 983 - 3408 · GitHub.com/xeniaqian94

EDUCATION

University of Maryland, College Park 08/2018 - Present

Doctor of Philosophy in Information Studies; Advised by Prof. Joel Chan

- Fundamentals of HCI (current)

Carnegie Mellon University, School of Computer Science 08/2016 - 07/2018

Master of Science in Language Technologies; GPA: 3.69/4.00

- Deep Learning, Machine Learning for Text Mining, Algorithm for Natural Language Processing, Search Engines, Machine Translation, Large-Scale Multimedia Analysis, Deep Reinforcement Learning, Neural Network for NLP

Zhejiang University, Chu Kochen Honors College 10/2012 - 07/2016

Bachelor of Engineering in Software Engineering; GPA: 3.95/4.00;

PROFESSIONAL EXPERIENCE

Alibaba Group, iDST (DAMO Academy) NLP Group China

Algorithm Engineering Intern

05/2018 - 08/2018

- Implemented an attribute-extraction module for the question answering bot at 2.taobao.com. Trained models from weakly supervised consumer dialog data. Experimented with the WOZ 2.0 dialog state tracking dataset. [\[Media\]](#)

Language Technologies Institute, Carnegie Mellon University Pittsburgh, PA

Research and Teaching Assistant

10/2016 - 04/2018

- **Duolingo second language acquisition:** trained a Bi-LSTM-CRF model with syntactic, lemma-based, and user meta features to predict word-level mistakes.
- **Reinforcement LeToR:** Modeled adhoc web search as sequential document selection, optimized with REINFORCE.
- **Crowdsourcing Pipeline:** Parsed and indexed 50 million ClueWeb09 collection web pages with 1K TREC Web/MQ track queries, then rendered and deployed Indri sentence retrieval snippets on CrowdFlower, with random sampling.
- **Data Cleaning and Evaluation:** Processed 20K user sessions with mouse movement behavioral tracking and user-based collaborative filtering for qualification, then analyzed the LeToR performance trained with massive click labels.

Risk and Regulatory Services Innovation Center, Sponsored by PwC Pittsburgh, PA

Research Assistant

10/2016 - 04/2018

- **Topic Tracking with Active Learning:** Community monitoring through training text classifiers with Naive Bayes/SVM, filtering with embedding space matching, geo-fencing with NER, and cost-sensitive annotation sampling.
- **Learning to Rank for Conversation:** Measured LeToR effectiveness on conversations with textual/temporal/metadata feature engineering. Collected a 10K annotation pool sampled from field-based relevance models (BM25F, etc.).
- Co-advised 5 MISM students on the Heinz College Capstone Project, Safe Cities Monitoring.

Data Systems Group, University of Waterloo Canada

Research Assistant; GPA: 88.2/100

10/2015 - 05/2016

- **Push notification:** Built a RESTful client system YoGosling, and integrated into Anserini, an open-source project.
- **A/B Testing vs. Interleaving:** Explored the two evaluation strategies on TREC Microblog track.

PUBLICATION

Xin Qian, Jimmy Lin, and Adam Roegiest. Interleaved evaluation for retrospective summarization and prospective notification on document streams. In *SIGIR'16*, pages 175-184. ACM. Oral presentation. [\[PDF\]](#)

PROJECTS

Data Warehousing: SQL query optimization that manipulates raw RDDs on TPC-H DB benchmark, in Spark and Scala.

Neural Machine Translation: Attention-based Bi-LSTM (RNN) with beam search in DyNet, from German to English.

SKILLS

Languages: Java, Python, HTML/CSS/Javascript, Linux/Unix, SQL, C/C++

Tools: PyTorch, pandas, scikit-learn, Git, Spark, Hadoop, AWS, CrowdFlower, Django, D3.js, seaborn