

Peiran Yao | PhD Student

2-21 Athabasca Hall – Edmonton, AB T6G 2E8 – Canada
☎ +1 (825) 440 3379 ✉ peiran@ualberta.ca 🌐 xavieryao

Education

University of Alberta

PhD student, Computing Science

2019–now

Area: Knowledge representation & NLP

Supervisor: Dr. Denilson Barbosa

Tsinghua University

BEng, Computer Science and Technology

2015–2019

BA, English Language and Literature (2nd Degree)

Advisor: Dr. Jie Tang

Research

WDM Lab, University of Alberta

Knowledge Graph Error Detection and Correction

2020–now

- error detection without large-scale labelling, with 0.72 F1 score
- based on learning-with-noise approaches
- assessment of typing errors in DBpedia
- accepted by WWW'21

Scotiabank & WDM Lab, University of Alberta

Contact Centre Call Topic Dissection

2020–2021

- identify common topics and proportion of customer calls
- predict future call topic based on customer profile
- original phrase mining algorithm

Recurrent AI

Fast Transformer

2019

- incorporate depth-wise separable and multi-grid convolution into Transformer
- reduce FLOPs of Transformer models
- without compromising performance

KEG, Tsinghua University

Cold-starting Task-oriented Dialogue Systems (Thesis Project)

2018–2019

- a dialogue system for academic knowledge graph AMiner
- template-guided NLU training in data-constraint scenarios
- multi-hop reasoning in KG with cognitive graphs

Author Name Disambiguation in AMiner

2017–2018

- name disambiguation algorithm and system with state-of-the-art accuracy
- using metric learning and network embedding
- role: part of the theory, experiments and online system
- accepted by KDD'18

Open Academic Graph

2017

- o integrated 2 knowledge graphs: Microsoft Academic Graph & AMiner
- o both having 150 million+ papers
- o with 99.7%+ accuracy
- o role: framework design and engineering
- o accepted by KDD'19

Publications

- o Peiran Yao and Denilson Barbosa. Typing errors in factual knowledge graphs: Severity and possible ways out. In *Proceedings of the Web Conference 2021*, pages 3305–3313, 2021.
- o Fanjin Zhang, Xiao Liu, Jie Tang, Yuxiao Dong, Peiran Yao, Jie Zhang, Xiaotao Gu, Yan Wang, Bin Shao, Rui Li, et al. OAG: Toward linking large-scale heterogeneous entity graphs. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2585–2595, 2019.
- o Yutao Zhang, Fanjin Zhang, Peiran Yao, and Jie Tang. Name disambiguation in AMiner: Clustering, maintenance, and human in the loop. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1002–1011, 2018.
- o Fanjin Zhang, Xiaotao Gu, Peiran Yao, and Jie Tang. Integration of heterogeneous data from multiple sources. *Journal of Chinese Information Processing*, 32(9):84–92, 2018.

Services

Teaching.....

CMPUT 361: Intro. to IR (W20) "Peiran is the best TA I've had in a compsci course."

CMPUT 497: Intro. to NLP (F20) "Peiran was awesome. He is the best TA ever! I believe you are lucky to have him as your TA."

CMPUT 379: Operating System Concepts (F19) "He's so understanding and caring! Thank you so much!"

Associate Reviewer.....

RL4NLP Workshop @ ACL

EMNLP

Student Association.....

Tsinghua University TUNA Association

President

2018-2019

- o focused on open source and computer network
- o hosting the largest open source software mirror in China
- o role: organization, planning, service maintenance

Awards

Fung Scholarship Oct. 2017

THU Freshman Award: Oct. 2015

Engineering

Internships & Projects.....

Tsinghua University

BT Express: Official Campus Activity Social App 2019

- WeChat mini-program for campus activity browsing, publishing, and reviewing
- backend: Python, Flask, MongoDB
- frontend: JavaScript, Vue.js, Element UI

BizSeer Co., Ltd.

Carpenter: Anomaly Detection from Logs for AIOps 2018

- adaptation and implementation of log template extraction algorithms
- extracting template types and parameters
- 1-D and multi-D time series anomaly detection

Tencent Co., Ltd.

AetosDios: KPI Anomaly Detection System for AIOps 2018

- 1-D time series anomaly detection system from scratch
- supervised model using local, seasonal and regression features and GBDT
- monitoring 10k time series
- role: algorithm & model design and implementation
- complete system available for the Ops team

Inst. for Network Sciences and Cyberspace, Tsinghua University

Campus Network Access Control Using Bio-metrics 2018

- a network access control scheme with face and fingerprint recognition
- identify true user using TOTP, TPM, and asymmetric cryptography

Architecture for Enhancing HTTPS Performance 2016-2017

- foundation of the deployment of HTTPS in campus network services
- comprehensive survey on HTTPS-related cryptography algorithms, implementations and hardware

Dept. of Engineering Physics, Tsinghua University

ESong: Controller App for Wearable Device 2016

- design and implementation of an Android app to control a wearable device remotely
- using Java
- communication protocol based on Bluetooth LE

P.A.P. Dev Team

BlackLight: an Android Weibo Client 2014-2015

- development of a Weibo client on Android platform
- strictly conforming to Material Design
- with all the major features provided by the official client
- 130,000+ downloads and 900+ stars on GitHub

Tech Stack.....

Experienced Python, PyTorch, C, Java

Familiar Swift, Scala, C++, Spark