Fan **Jiang**

☐ (+61) 469-814-650 | ✉ fan.jiang1@student.unimelb.edu.au | 🌐 fantabulous-j.github.io | ⌨ fantabulous-j | 🎓 Fan Jiang

Educations

The University of Melbourne   *Melbourne, Victoria, Australia*
PhD in Natural Language Processing   *Feb. 2022 - Feb. 2026 (expected)*
- Supervisors: Prof. Trevor Cohn & Prof. Tom Drummond

The University of Melbourne   *Melbourne, Victoria, Australia*
Master of Science (Computer Science) (with Distinction)   *Jul. 2019 - Jun. 2021*
- GPA: 91.9% (3.94/4.0) First Class Honours
- Supervisor: Prof. Trevor Cohn

Fuzhou University   *Fuzhou, Fujian, China*
Bachelor of Engineering in Software Engineering   *Sep. 2015 - Jun. 2019*
- GPA: 4.07/5.0 (Ranking: 1/155) Graduated with Honours

Research Interest

My primary research interests are centered around natural language processing. My research is motivated by the ultimate goal of enhancing information access for individuals across diverse backgrounds. Recently, I have been concentrating on achieving robust and general information access by overcoming linguistic and domain barriers:

- **General knowledge retrieval systems across domains** – Robust systems capable of retrieving reliable information across various formats and domains are crucial for meeting the information requirements of intricate real-world user queries across diverse fields.

- **Intelligent open-domain question-answering systems across languages** – It is crucial to enhance information accessibility for users from diverse linguistic backgrounds, particularly those who speak underrepresented languages. My emphasis is on developing sophisticated systems capable of aiding linguistic minorities in accessing essential information.

Publications

Under Review	**Pre-training Cross-lingual Open Domain Question Answering with Large-scale Synthetic Supervision.** **Fan Jiang**, Tom Drummond and Trevor Cohn. Submitted to ACL ARR 2024 February.
EMNLP-Findings'23	**Boot and Switch: Alternating Distillation for Zero-Shot Dense Retrieval.** **Fan Jiang**, Qiongkai Xu, Tom Drummond and Trevor Cohn. In Findings of the Association for Computational Linguistics: EMNLP 2023.
EMNLP-Findings'23	**Noisy Self-Training with Synthetic Queries for Dense Retrieval.** **Fan Jiang**, Tom Drummond and Trevor Cohn. In Findings of the Association for Computational Linguistics: EMNLP 2023.
EACL'23	**Don't Mess with Mister-in-Between: Improved Negative Search for Knowledge Graph Completion.** **Fan Jiang**, Tom Drummond and Trevor Cohn. Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics.
AAAI'22	**Incorporating Constituent Syntax for Coreference Resolution.** **Fan Jiang** and Trevor Cohn. Proceedings of the 36th AAAI Conference on Artificial Intelligence.
NAACL'21	**Incorporating Syntax and Semantics in Coreference Resolution with Heterogeneous Graph Attention Network.** **Fan Jiang** and Trevor Cohn. Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies.

Research Experience

The University of Melbourne *Melbourne, Victoria, Australia*

Ph.D. Thesis: Towards Robust and General Information Access by Bridging Linguistic and Domain Barriers *March. 2022 - Present*

Student Research Assistant Advisors: Prof. Trevor Cohn &
Prof. Tom Drummond

- Structured knowledge: adapting neural retrievers to knowledge graphs. (EACL 2023)
- Domain adaptation: adapting neural retrievers to zero-resource domains with a novel self-training approach. (EMNLP-Findings 2023)
- General neural retriever: a bootstrapping algorithm for unsupervised retriever training across massive domains. (EMNLP-Findings 2023)
- A unified system for cross-lingual retrieval and multilingual ODQA: a two-stage self-supervised pre-training framework with large-scale synthetic data for strong zero-shot cross-lingual retrieval and multilingual QA performance. (Submitted ACL 2024)

The University of Melbourne *Melbourne, Victoria, Australia*

Master Thesis: Towards Syntax and Semantics-Driven Neural Coreference Resolution *Aug. 2020 - June. 2021*

Student Research Assistant Advisor: Prof. Trevor Cohn

- Improved a strong neural coreference resolution model by incorporating dependency syntax and semantic role labels using heterogeneous graph attention networks. (NAACL 2021)
- Proposed a novel method to effectively incorporate constituent parse trees to enhance a strong neural coreference resolution model by utilising graph attention networks. (AAAI 2022)

Services

Conference Reviewers

- EACL 2023

- EMNLP 2023

- ACL Rolling Review 2024

Selected Projects

End-to-end Speech Synthesis Method Based on Attention Mechanism (Bachelor's Thesis)

- Adopted Tacotron2 model (proposed by Google) as the mel-spectrogram generator, which is a recurrent sequence-to-sequence model with attention mechanism, and it can directly map the input text to spectrum extracted from original audio waveform. Most parts of the Tacotron2 Model were unchanged and the LSTM layers were changed into GRU ones since GRU can achieve almost the same performance as its LSTM counterparts but is computationally cheaper. A GRU layer was added to the decoder which takes current input of the decoder and contexts generated at previous time step as input before the attention layer.
- Adopted WaveNet as the neural vocoder. Use Gaussian Auto-Regressive WaveNet and Gaussian Inverse Auto-Regressive Flow based WaveNet which was trained through knowledge distillation as vocoders, respectively.

Hidden Markov Model based Chinese Singing Voice Synthesis System

- Implemented a Chinese singing voice synthesis system, which could be used to generate an audio in the form of WAV with arbitrary scores in the format of XML (Speech Parameter Generation Algorithm From HMM and Mel Log Spectrum Approximation (MLSA) filter for speech synthesis were implemented in this system).
- Implemented Probability Evaluation Algorithm (included the calculations of Forward and Backward probabilities), Baum-Welch Algorithm (Expectation Maximum Algorithm) and Viterbi Algorithm.
- Used HTK tools and a training set of 34 Chinese songs to train a model.

Skills

Programming Languages	C/C++, Java, Python
Deep Learning Framework	PyTorch, TensorFlow, Keras
Tools	J2EE, LaTeX, Makefile, Gcc
Database	MySQL, Microsoft SQL Server

Honors & Awards

Dean's Honours List 2021, Faculty of Science, The University of Melbourne	Aug. 2022
AAAI-22 Student Scholarship, Association for the Advancement of Artificial Intelligence	Feb. 2022
Outstanding Graduate of 2019 Academic Year, Fuzhou University	Jun. 2019
2nd Prize of 10th China Students Service Outsourcing Innovation and Entrepreneurship Competition	May. 2019
Merit Student of 2017-2018 Academic Year, Fuzhou University (Top 5%)	May. 2018
3rd Prize of Java Development University Group A, 9th Lan Qiao Cup China Software and IT Professionals	Apr. 2018
2nd Prize for Band C in 2016 National English Competition for College Students	May. 2016
Fuzhou University Scholarship (1st #2, 2nd #3, 3rd #1)	2015 - 2019

Language

IELTS   Overall: 7.5 (Listening: 8.5, Reading: 8.5, Writing: 6.5, Speaking: 6.0)
GRE   324 (Verbal: 154, Quantitative: 170)