Qiyu Wu

Email: wuqiyu576@gmail.com · Personal Website: qiyuw.github.io · LinkedIn: qiyuw

Research Interests

My research focus lies in **semantic representation learning for NLP**, mainly encompassing better representing word/phrase semantics [C3, C5, C8], sentence semantics [C8, C7, C4, J1], in both monolingual and multilingual [C7, C5] contexts. Recently I also broaden my scope to visual-language representation [M1, M4] and entity representation [C6].

Education

The University of Tokyo, Tokyo, Japan.   Oct. 2021 – Sep. 2024
Ph.D. student at Tsuruoka Lab; **JSPS DC2 Fellowship**; **JST SPRING Fellowship**.

Peking University, Beijing, China.   Sep. 2018 – Jul. 2021
M.Sc. in Computer Software & Theory.

Sichuan University, Chengdu, China.   Sep. 2014 – Jun. 2018
B.Eng. in Computer Science & Technology (GPA top 3%).

Research Experiences

Research Intern at Sony, Creative AI Lab   Jun. 2023 – Sep. 2023
Mitigate reporting bias in visual-language datasets with bimodal augmentation [M1].

Visiting Student at UC San Diego   Jan. 2023 – Feb. 2023
Explore deep generative models [M4].

Research Intern at Microsoft, Software Technology Center Asia, NLP Group   Oct. 2021 – Feb. 2022
Leverage diverse augmentation for sentence representation to mitigate the bias in training data [C4].

Research Intern at Microsoft Research Asia, Machine Learning Group   Apr. 2020 – Jun. 2021
Utilize the cross-sentence links to mitigate noisy representation of long-tail words in language pre-training [C3].

R&D Intern at Baidu Research   Oct. 2019 – Feb. 2020
Transportation prediction for different communities [C2]; route mode recommendation in Baidu Maps.

Publications

∗ indicates equal contribution, † indicates students mentored by me.

[M4] Guangyi Liu, Yu Wang, Zeyu Feng, **Qiyu Wu**, Liping Tang, Yuan Gao, Zhen Li, Shuguang Cui, Julian McAuley, Eric P. Xing, Zichao Yang, Zhiting Hu. Generating, Reconstructing, and Representing Discrete and Continuous Data: Generalized Diffusion with Learnable Encoding-Decoding, *manuscript, under review*

[M3] **Qiyu Wu**, Masaaki Nagata, Zhongtao Miao, Yoshimasa Tsuruoka. "Word Alignment as Preference for Machine Translation", *manuscript, under review*.

[M2] Yuting Tang, Xin-Qiang Cai, Yao-Xiang Ding, **Qiyu Wu**, Guoqing Liu, Masashi Sugiyama. "Reinforcement Learning from Bagged Reward: A Transformer-based Approach for Instance-Level Reward Redistribution", *manuscript, under review*.

[M1] **Qiyu Wu**, Mengjie Zhao, Yutong He, Lang Huang, Junya Ono, Hiromi Wakaki, Yuki Mitsufuji. "Bi-modal Data Augmentation by Decoupling Object-attribute Association.", *manuscript, under review*.

[C8] Zhongtao Miao†, **Qiyu Wu**, Kaiyan Zhao, Zilong Wu, Yoshimasa Tsuruoka. "Enhancing Cross-lingual Sentence Embedding for Low-resource Languages with Word Alignment", in *the 2022 Conference of the North American Chapter of the Association for Computational Linguistics* **(NAACL 2024, findings)**.

[C7] Kaiyan Zhao†∗, **Qiyu Wu**∗, Xin-Qiang Cai, Yoshimasa Tsuruoka. "Leveraging Multi-lingual Positive Instances in Contrastive Learning to Improve Sentence Embedding.", in *the 18th Conference of the European Chapter of the Association for Computational Linguistics.* **(EACL 2024, main conference long paper)**.

[C6] Shiwen Wu, **Qiyu Wu**, Honghua Dong, Wen Hua, Xiaofang Zhou. "Blocker and Matcher Can Mutually Benefit: A Co-Learning Framework for Low-Resource Entity Resolution.", *Proc. VLDB Endow. (2024). Will appear at Very Large Databases Conf.* **(VLDB 2024, research track paper)**.

[C5] Qiyu Wu, Masaaki Nagata, Yoshimasa Tsuruoka. "WSPAlign: Word Alignment Pre-training via Large-Scale Weakly Supervised Span Prediction", in *the 61st Annual Meeting of the Association for Computational Linguistics (**ACL 2023, main conference long paper**)*.

[C4] Qiyu Wu, Chongyang Tao, Tao Shen, Can Xu, Xiubo Geng, Daxin Jiang. "PCL: Peer-Contrastive Learning with Diverse Augmentations for Sentence Embeddings", in *Conference on Empirical Methods in Natural Language Processing (**EMNLP 2022, main conference long paper**)*.

[J1] Yutao Xie[†], **Qiyu Wu**, Wei Chen and Tengjiao Wang. "Stable Contrastive Learning for Self-Supervised Sentence Embeddings With Pseudo-Siamese Mutual Learning", in *IEEE/ACM Transactions on Audio, Speech, and Language Processing (**TASLP**), vol. 30, pp. 3046-3059, 2022*.

[C3] Qiyu Wu[*], Chen Xing[*], Yatao Li, Guolin Ke, Di He, Tie-Yan Liu. "Taking Notes on the Fly Helps Language Pre-Training", in *the Ninth International Conference on Learning Representations (**ICLR 2021**)*.

[C2] Hao Liu[*], **Qiyu Wu**[*], Fuzhen Zhuang, Xinjiang Lu, Dejing Dou, Hui Xiong. "Community-Aware Multi-Task Transportation Demand Prediction", in *Thirty-Fifth AAAI Conference on Artificial Intelligence (**AAAI 2021**)*.

[C1] Dechuan Yang[*], **Qiyu Wu**[*], Wei Chen, Tengjiao Wang, Zhen Qiu, Di Liu. "Stance Detection with Stance-wise Convolution Network", in *CCF International Conference on Natural Language Processing and Chinese Computing (**NLPCC 2020**), pp. 555-567, 2020*.

[J]: Journal paper; [C]: Conference paper; [M]: Manuscript.

Awards and Scholarships

• Research Fellowship for Young Scientists (DC2), JSPS	2024 – 2026
• Support for Pioneering Research Initiated by the Next Generation (SPRING) Fellowship, JST	2021 – 2024
• Excellent Research Award, Peking University	2020
• National Scholarship, China	2016, 2017
• First Place Scholarship, Sichuan University	2016, 2017

Teaching and Academic Services

- Student I mentored

– Zhongtao Miao, Ph.D. student at The University of Tokyo, on [C8].	2023 – Present
– Kaiyan Zhao, Ph.D. student at The University of Tokyo, on [C7].	2022 – Present
– Yutao Xie, Ph.D. student at Peking University, on [J1].	2020 – 2021

- Teaching Assistant, Peking University Summer School International: "Social Implications of Computing".   2019
- Reviewer of ACL Rolling Review 2023, 2024; ICLR 2022, 2023, 2024; EMNLP 2023; NeurIPS 2021, 2022, 2023; ICML 2022.