

He Bai

200 Ring Rd, Waterloo
ON, Canada N2L 3G1
☎ (+1) 548 333 6369
✉ he.bai@uwaterloo.ca
📄 www.richardbaihe.github.io

Education

- Start from **University of Waterloo**,
09/2019 *Computer Science*, PhD,
Supervisor: Ming Li.
- 09/2017–
07/2019 **Institute of Automation, Chinese Academy of Sciences**,
Pattern Recognition and Intelligent System, Master of Science,
Supervisor: Yu Zhou, Chengqing Zong.

Research

Segment-aware Transformer, *Ming Li*, University of Waterloo.

- Introducing a segment-aware transformer for language modeling and understanding.
- Achieving the SOTA results in language modeling, and a new pre-trained model SegBERT.

Multi-turn Spoken Language Understanding, *Chengqing Zong*, Chinese Academy of Sciences.

- Proposing a self-supervised dialogue coherence task for multi-turn spoken language understanding.
- Model trained with our proposed method get significant improvement on KVRET dataset.

Cross-lingual Spoken Language Understanding, *Mei-yuh Hwang*, Mobvoi.

- Transferring Chinese spoken language understanding dataset into English with reinforcement learning controlled machine translation method.

Internship

- 07/2018–
10/2018 **Research Intern**, IBM RESEARCH.
◦ Chinese Openai-GPT model pre-training.
◦ Matching based neural dialogue generation with GAN.
- 06/2018–
07/2018 **Summer Intern**, ALIBABA GROUP, ALIME .
◦ Query rewriting and contextual queries assignment.
- 09/2017–
03/2018 **Research Intern**, MOBVOI.
◦ Language transferring for spoken language understanding.
◦ Paraphrase generation.
- 07/2016–
09/2016 **Summer Intern**, UNISOUND.
◦ Implementing the seq2seq based chit-chat bot with C++.

Publications

- He Bai, Peng Shi, Jimmy Lin, Luchen Tan, Kun Xiong, Wen Gao, Ming Li. **Segatron: Segment-aware Transformer for Language Modeling and Understanding**. *arXiv preprint arXiv:2004.14996*. (under review)
- He Bai, Peng Shi, Jimmy Lin, Luchen Tan, Kun Xiong, Wen Gao, Jie Liu, Ming Li. **Semantics of the Unwritten: The Effect of End of Paragraph and Sequence Tokens on Text Generation**. *arXiv preprint arXiv:2004.02251*. (under review)
- Peng Shi, He Bai, Jimmy Lin. **Cross-Lingual Training of Neural Models for Document Ranking**. To appear in *Findings of EMNLP (Conference on Empirical Methods in Natural Language Processing) 2020*. Online.
- He Bai, Yu Zhou, Jiajun Zhang and Chengqing Zong. **Memory Consolidation for Contextual Spoken Language Understanding with Dialogue Logistic Inference**. In *Proceedings of ACL (Annual Meeting of the Association for Computational Linguistics) 2019*. Florence, Italy, July.

- **He Bai**, Yu Zhou, Jiajun Zhang, Liang Zhao, Mei-Yuh Hwang and Chengqing Zong. **Source Critical Reinforcement Learning for Transferring Spoken Language Understanding to a New Language.** *In Proceedings of COLING (International Conference on Computational Linguistics) 2018. Santa Fe, New Mexico, USA, August.*

Service

Reviewer EMNLP 2019, ACL2020, AAAI 2020, COLING 2020

Interests

Research Natural language processing, Representation Learning

Personal Traveling, Music