Zhiyuan Zeng

Email: zhiyuan_zeng@tju.edu.cn

Research interests	Machine Translation, AI for Education

Education

Tianjin University　　　　　　　　　　　　　　　　Tianjin, China
MA in Computer Science　　　　　　　　　　　　　2020 – Present
Mentors: Professor Deyi Xiong

Nanchang University　　　　　　　　　　　　　Nanchang, China
BA in Computer Science　　　　　　　　　　　　　　2016 – 2020

Publications

An Empirical Study on Adversarial Attack on NMT: Languages and Positions Matter
Zhiyuan Zeng, Deyi Xiong
Association for Computational Linguistics (ACL), 2021, [[pdf](.)]

Unsupervised and Few-Shot Parsing from Pretrained Language Models
Zhiyuan Zeng, Deyi Xiong
The journal of Artificial Intelligence (AIJ), Volume 305, April, 2022 [[pdf](.)]

SCoMoE: Efficient Mixture of Experts with Structured Communication
Zhiyuan Zeng, Deyi Xiong
The Eleventh International Conference on Learning Representations (ICLR) , 2023, [[pdf](.)]

Research experience

Empirical Study on Adversarial Attack on NMT
Mentors: Professor Deyi Xiong　　　　　　July 2020 – January 2021
- Empirically found that adversarial attack on the source side is more effective than that on the target side in terms of the performance degradation of NMT models under attack.
- Empirically found that adversarial attacks on front positions are more effective than those on back positions due to the autoregressive translation.
- Propose a new adversarial attack generation approach that samples positions for injecting perturbations according to the attention distribution.

Unsupervised and Few-Shot Parsing from Pretrained Language Models
Mentors: Professor Deyi Xiong　　　　December 2019 – December 2020
- Proposed unsupervised and few-shot parsing models solely based on the self-attention weight matrix in pretrained language models.
- Achieved the state-of-the-art results on most languages in SPMRL in both unsupervised and few-shot parsing.

Efficient Mixture of Experts

Mentors: Professor Deyi Xiong   June 2021 – May 2022

- Proposed an efficient MoE model (SCoMoE) with structured all-to-all communication based on the hierarchical structure of communication topology.
- Proposed the token clustering and differentiable sorting to compensate the potential performance drop caused by structured communication.
- Speedup the baseline (Gshard) to 1.5x with the comparable translation performance and to 1.3x with significantly better performance (1.2 BLEU) on massive bilingual and multilingual translation.

Industry experience   **OPPO**   Beijing, China

Robustness of NMT   July 2020 - May 2021

Empirical study on the adversarial attack on NMT.

Global Tone Communication TECHNOLOGY CO., LT   Beijing, China

Massive multilingual NMT   June 2021 - December 2022

Training a massive multilingual NMT model with 40 billion parameters on a massive multilingual parallel corpus with 1 billion samples.

Skills   **Library**

Familiar with: Pytorch and Fairseq

Tools

Familiar with: Slurm