

Zhiyuan Zeng

Updated June 13, 2022

Email: zhiyuan.zeng@tju.edu.cn

Research interests

Machine Translation, Robustness, Unsupervised Parsing

Education

Tianjin University

MA in Computer Science

Mentors: Professor Deyi Xiong

Tianjin, China

2020 – Present

Nanchang University

BA in Computer Science

Nanchang, China

2016 – 2020

Publications

SCoMoE: Efficient Mixture of Experts with Structured Communication

Zhiyuan Zeng, Deyi Xiong

NIPS Anonymous Submission, 2022

An Empirical Study on Adversarial Attack on NMT: Languages and Positions Matter

Zhiyuan Zeng, Deyi Xiong

ACL, 2021

Unsupervised and Few-Shot Parsing from Pretrained Language Models

Zhiyuan Zeng, Deyi Xiong

Artificial Intelligence, Volume 305, April 2022

Research experience

Efficient Communication for 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 bandwidth structure of communication topology.
- Proposed token clustering and differentiable sorting to compensate the potential performance drop caused by structured communication.
- Speedup MoE to 1.5x with comparable translation performance and to 1.3x with significantly better performance (1.2 BLEU) on massive bilingual and multilingual translation experiments.

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.
- Proposed an algorithm to transform a constituent tree into a distance matrix, which can be used to constrain the attention matrix of transformer and improve the performance.

Industry experience

Global Tone Communication TECHNOLOGY CO., LT Beijing, China

Massive multilingual NMT June 2021 - Present

Training a massive multilingual NMT model with 10 billion parameters based on the Mixtures of Experts.

OPPO Beijing, China

Robustness of NMT July 2020 - May 2021

Investigating the adversarial attack on NMT, especially on the target-side input of NMT, and improving the robustness of NMT.

Skills

Library

Familiar with: Pytorch and Fairseq

Tools

Familiar with: Slurm (for distributed training)