Haozhe Ji (计昊哲)

Email: jihaozhe@gmail.com
Web: haozheji.github.io
Phone: +86 15652655811

Research Interests

My research is driven by the goal of developing **theoretically grounded and scalable methods** to improve neural language models in the areas of **natural language generation** and **language model alignment**. Specifically, my work aims to develop practical algorithms and systems that address the fundamental limitations of the standard paradigm of language modeling in a principled manner.

Firstly, in terms of the **choice of modeling**, my research explores model families beyond the auto-regressive models (ARMs) which possess a strong local inductive bias, to facilitate more accurate modeling of the growing volume of data. This includes practical realization of **theoretically more expressive model families**, e.g., energy-based models [2], latent variable models [6], and semi-parametric models [8].

Secondly, in terms of the **problem of learning**, my research advocates for **quality-aware learning objectives** beyond maximum likelihood estimation (MLE) which is biased towards coverage. These new objectives are theoretically grounded in probability metrics that facilitate quality assessment, including reverse KL divergence [1] and total variation distance [3] to accommodate the growth of high-quality data annotations in various forms.

Education

Tsinghua University, Beijing, China — September 2020 - Present
Ph.D. Student, Computer Science and Technology
Advisor: Minlie Huang

Tsinghua University, Beijing, China — September 2016 - July 2020
B.E., Electronic Engineering

Research Highlight

Alignment with principled and scalable policy optimization and evaluation

Efficient Exact Policy Optimization [1]

- Policy optimization as probability matching under reverse KL divergence.
- Theoretically **unbiased** estimation of the PPO objective, while more **efficient** and stable.
- **Consistently outperforms PPO and DPO** across academic benchmarks to realistic user instruction data, and models scaled up to 32B.

Principled Multi-Objective Aggregation [2]

- Automatic aggregation of multiple objectives to improve on all evaluation dimensions.
- In language model decoding, outperforms all heuristic decoding methods with **guaranteed perplexity improvement** in aligning with human texts.
- In reward aggregation, greatly improves the efficiency and **instance-level performance on all dimensions** over manual interpolation in Best-of-N evaluation.

Talk

Beyond the Theoretical Limits of Language Modeling — June 2024
ByteDance, Seed Team

Towards Efficient Exact Optimization of Language Model Alignment — March 2024
ByteDance, RAI Group

PREPRINTS

PUBLICATIONS

[1] **Towards Efficient Exact Optimization of Language Model Alignment**
Haozhe Ji, Cheng Lu, Yilin Niu, Pei Ke, Hongning Wang, Jun Zhu, Jie Tang, Minlie Huang
International Conference on Machine Learning (ICML), 2024.

[2] **Language Model Decoding as Direct Metrics Optimization**
Haozhe Ji, Pei Ke, Hongning Wang, Minlie Huang
International Conference on Learning Representations (ICLR), 2024.

[3] **Tailoring Language Generation Models under Total Variation Distance**
Haozhe Ji, Pei Ke, Zhipeng Hu, Rongsheng Zhang, Minlie Huang
International Conference on Learning Representations (ICLR), 2023.
(***Oral / Notable top 5%***)

[4] **Curriculum-Based Self-Training Makes Better Few-Shot Learners for Data-to-Text Generation**
Pei Ke, **Haozhe Ji**, Zhenyu Yang, Yi Huang, Junlan Feng, Xiaoyan Zhu, Minlie Huang
International Joint Conference on Artificial Intelligence (IJCAI), 2022.

[5] **LaMemo: Language modeling with look-ahead memory**
Haozhe Ji, Rongsheng Zhang, Zhenyu Yang, Zhipeng Hu, Minlie Huang
North American Chapter of the Association for Computational Linguistics (NAACL), 2022. (***Oral***)

[6] **DiscoDVT: Generating Long Text with Discourse-Aware Discrete Variational Transformer**
Haozhe Ji, Minlie Huang
Empirical Methods in Natural Language Processing (EMNLP), 2021. (***Oral***)

[7] **Jointgt: Graph-text joint representation learning for text generation from knowledge graphs**
Pei Ke, **Haozhe Ji**, Yu Ran, Xin Cui, Liwei Wang, Linfeng Song, Xiaoyan Zhu, Minlie Huang
Findings of the Association for Computational Linguistics (Findings of ACL), 2021.

[8] **Language generation with multi-hop reasoning on commonsense knowledge graph**
Haozhe Ji, Pei Ke, Shaohan Huang, Furu Wei, Xiaoyan Zhu, Minlie Huang
Empirical Methods in Natural Language Processing (EMNLP), 2020. (***Oral***)

[9] **Generating commonsense explanation by extracting bridge concepts from reasoning paths**
Haozhe Ji, Pei Ke, Shaohan Huang, Furu Wei, Minlie Huang
Asia-Pacific Chapter of the Association for Computational Linguistics (AACL), 2020.

[10] **Sentilare: Linguistic knowledge enhanced language representation for sentiment analysis**
Pei Ke*, **Haozhe Ji***, Siyang Liu, Xiaoyan Zhu, Minlie Huang
Empirical Methods in Natural Language Processing (EMNLP), 2020.

[11] **Denoising distantly supervised open-domain question answering**
Yankai Lin, **Haozhe Ji**, Zhiyuan Liu, Maosong Sun
Annual Meeting of the Association for Computational Linguistics (ACL), 2018.

RESEARCH EXPERIENCE

CoAI Lab, **Tsinghua University** — September 2020 - Present
Ph.D. Candidate (Supervisor: Minlie Huang)

Alignment Group, **Zhipu AI** — March 2024 - Present
Research Intern (Supervisor: Hongning Wang)

Natural Language Comupting group, **Microsoft Research Asia** July 2019 - July 2020
Research Intern (Supervisors: Shaohan Huang, Furu Wei)

AWARDS

Tang Junyuan (唐君远) Scholarship, Tsinghua University 2023
Sohu Scholarship, Tsinghua University 2022
Yang Huiyan (杨惠妍) Scholarship, Tsinghua University 2021
Comprehensive Merit Scholarship, Tsinghua University 2019
Comprehensive Merit Scholarship, Tsinghua University 2017
Gold Medal, 32nd China Physics Olympiads (CPhO) 2015
Distinguished Honor Roll (Top 1%), American Mathematics Contest 12 (AMC12) 2015

SERVICES

Reviewer/Program Committee: ACL, EMNLP, NAACL, ARR

TEACHING

I was the Head TA of the undergraduate course Artificial Neural Network, instructed by Minlie Huang (2021 Fall, 2022 Fall, 2023 Fall).

I coauthored the NLP textbook *Modern Natural Language Generation* (in Chinese). Specifically, I mainly drafted the fourth chapter *Transformer-based Language Generation Model*.