Tianjian Li

Github: github.com/tianjianl
Webpage: tianjianl.github.io

Email: tli104@jhu.edu
Mobile: +1-(667)7709373

Education

- **Johns Hopkins University** — Baltimore, MD
  M.S.E. in Computer Science — *Aug 2022 - Present*
- **New York University** — New York, NY
  B.A. in Computer Science and Mathematics — *Aug 2017 - Sept 2021*

Research Experience

- **Johns Hopkins University - Center for Language and Speech Processing (CLSP)** — Baltimore, MD
  Advisors: Kenton Murray and Philipp Koehn — *Sept 2022 - Present*
  - **Improving Robustness of Text Generation Models (ICLR 24)**: Developed and implemented Error Norm Truncation to enhance robustness in text generation models, significantly improving generation quality and increasing model robustness against noise in language modeling, machine translation, and text summarization tasks.
  - **Parameter Importance Guided Optimization**: Proposed a unified view of methods that harnesses importance of individual parameters. Designed a novel parameter importance guided optimizer that achieve comparable performance with a 2.4 $\times$ speed boost.
  - **Cross-lingual Transfer in Language Generation (ACL 23)**: Carried out analysis of why cross-lingual transfer works for text classification but fails in generation. We found that multilingual LMs transfers supervision from language to another by increasing their representation similarity, which is beneficial for classification but harmful for generation.
- **Tsinghua University - Knowledge Engineering Group (KEG)** — Beijing, China
  Advisor: Jie Tang — *Mar 2022 - Aug 2022*
  - **Multilingual LM Pre-training**: Trained and open-sourced a multilingual language model with 1B parameters based on a novel autoregressive blank infilling objective.
  - **Cross-Lingual Summarization**: Collected multilingual summarization data across different domains and fine-tuned our multilingual model to perform cross-lingual summarization in any language.

Publications

- *Error Norm Truncation: Robust Training in the Presence of Data Noise for Text Generation Models.*
  Tianjian Li, Haoran Xu, Philipp Koehn, Daniel Khashabi, and Kenton Murray.
  ICLR 2024 (Spotlight - Top 5%). Link
- *Why Does Zero-Shot Cross-Lingual Generation Fail? An Explanation and a Solution.*
  Tianjian Li and Kenton Murray.
  Findings of ACL 2023 (long). Link

Industrial Experience

- **Baidu Inc.** — Beijing, China
  Machine Learning Engineer - Intern — *Aug 2021 - Feb 2022*
  - Built a classification model on the influence of small paths on customers' driving experience with XGBoost.
  - Optimized route ranking model by experimenting with two strategies: 1D-CNN and multi-head self-attention in modeling sequential trajectory data.
  - Designed a Spatial-Temporal Graph Neural Network model further to improve the performance of the route ranking model to anticipate and dodge traffic jams. Reimplemented STGCN (IJCAI '18) and STFGNN (AAAI '21) in PaddlePaddle (Baidu's DL Framework).

Honors, Awards and Services

- Reviewer: ACL 2023, EMNLP 2023, EACL 2024
- New York University College of Arts and Sciences (CAS) Scholarship 2020
- First Prize in National Olympiad in Informatics Provincial (NOIP)

Skills Summary

- **Programming Languages**: Python, Java, C/C++, SQL, JavaScript, Shell Scripting, Unix Commands(grep, sed)
- **DL Frameworks**: PyTorch, TensorFlow, Keras, PaddlePaddle, Huggingface(Accelerate), Fairseq, DeepSpeed, Jax
- **Tools**: Docker, git, Hadoop streaming, Spark, Vim, LaTex
- **Natural Languages:**: Chinese (Native), English, French