Tong Zheng

✉: zhengtong12356@gmail.com | ☺: kidzheng.github.io | ☎: +86 18602457703

EDUCATION

Northeastern University — **Shengyang, China**
Bachelor of Science in Computer Science — Aug. 2017 - July. 2021
Cumulative GPA: 4.001/5 (90.001/100)

RESEARCH INTERESTS

My primary research interest revolves around addressing inefficiencies in current deep learning methodologies. This interest spans two distinct eras:

- Pre-LLM era: Optimizing the design of deep learning backbone networks to enhance overall efficiency, including inefficient modeling of relationship among tokens (ICML 2022), inefficiency among attention head interaction, parameter inefficiency of Transformer architecture and inefficient utilization of visual representation in MMT.
- LLM era: Finding natural inefficiency among both LLMs and autoregressive generation and further build efficient inference method (work in progress).

In addition to these primary research objectives, my secondary research interest lies in the field of foundational AI. Recently, I am also interested in parameter-efficient fine-tuning.

PUBLICATIONS (* DENOTES EQUAL CONTRIBUTION)

1. Bei Li*, **Tong Zheng***, Yi Jing*, Chengbo Jiao, Tong Xiao, Jingbo Zhu. Learning Multiscale Transformer Models for Sequence Generation. International Conference on Machine Learning. 2022. (First ICML in NEU, China.)
2. Yuxin Zuo, Bei Li, Chuanhao Lv, **Tong Zheng**, Tong Xiao, JingBo Zhu. Incorporating Probing Signals into Multimodal Machine Translation via Visual Question-Answering Pairs. Findings of EMNLP2023.
3. Guangqi Wen, Peng Cao, Huiwen Bao, Wenju Yang, **Tong Zheng**, Osmar Zaiane. MVS-GCN: A prior brain structure learning-guided multi-view graph convolution network for autism spectrum disorder diagnosis. Computers in Biology and Medicine. 2022.
4. **Tong Zheng***, Bei Li*, Huiwen Bao*, Yi Jing, Tong Xiao, JingBo Zhu. EIT: Enhanced Interactive Transformer. Arxiv. (Submitted to EMNLP2023, Soundness: 3443, Excitment: 4344, Status: Reject).
5. **Tong Zheng***, Huiwen Bao*, Bei Li*, Weiqiao Shan, Tong Xiao, JingBo Zhu. PartialFormer: Modeling Part Instead of Whole for Machine Translation. Arxiv. (Submitted to EMNLP2023, Soundness: 343, Excitment: 333, Status: Reject).
6. **Tong Zheng**, Bei Li, Huiwen Bao, Jiale Wang, Can Zhao, Weiqiao Shan, Tong Xiao, JingBo Zhu.Bridging the Gap between NMT and LLM: A Prompting Approach for Integrating NMT Knowledge into LLM.

RESEARCH EXPERIENCE

Natural Language Processing Laboratory at NEU **Shengyang, CN**
Research Assistant (AI, NLP) *Seq. 2021 - Now*

- **Project 1:** Learning Multiscale Transformer Models for Sequence Generation.
  - First, this study re-defined the concept of scale for NLP, including scales of sub-word, word and phrase. The intention was to leverage the word boundaries and phrase-level prior knowledge to compensate for the sub-word features. Then this study established the relationships among different scales. Ultimately, this study built a multiscale Transformer model via making full use of the relationships (accepted at ICML2022).
  - I come up the idea of utilizing GCNs to modeling dependency among each scale and defined transformation matrices to enable information flow among different scales.
  - I built the whole framework via Fairseq.
  - I conducted all the experiments about machine translation and all the ablation experiments.
  - I wrote part of this paper (Method part) and drew all the Figures in this paper mainly via tikz.
  - I was glad to severe as co-first author to publish the first ICML paper in our school.
- **Project 2:** EIT: Enhanced Interactive Transformer.
  - This study looked at the age-old problem of attention mechanisms from a new perspective: how the attention heads should behave among themselves. Guided by the complementary theory of multi-view learning, this study proposed a many-to-many mapping module to increase the information capacity. On top of this, the study proposed a fine-grained hierarchical interaction module to promote consensus among different attention heads, which is guided by consensus theory in multi-view learning.
  - I led this project. I found this problem and came up the solution.
  - I wrote all the codes in this project.
  - I conducted all the experiments in this paper.
  - I wrote the paper and drew all Figures in this paper.
- **Project 3:** PartialFormer: Modeling Part Instead of Whole for Machine Translation.
- **Project 4:** Bridging the Gap between NMT and LLM: A Prompting Approach for Integrating NMT Knowledge into LLM.
- **Project 5:** Incorporating Probing Signals into Multimodal Machine Translation via Visual Question-Answering Pairs.
- **Project 6:** Participate in editing a book 'Natural Language Processing– Representation Learning and Neural Models'
  - I drew some of figures in this book via Latex (tikz).
  - I proofread the document.

Key Laboratory of Medical Image Intelligent Computing at NEU **Shengyang, CN**
Research Assistant (AI) *Oct. 2020 - Aug. 2021*

- **Project 1:** BrainTGL:Temporal Graph representation learning for brain network by Exploiting Graph Temporal Information. *Oct. 2020 - Aug. 2021*
  - This study introduces BrainTGL, addressing the challenge of capturing spatio-temporal dynamics in resting-state fMRI data. It achieves state-of-the-art accuracy on ABIDE and HCP datasets and provides consistent clustering results with ASD biomarker evidence, enhancing our understanding of cognitive brain processes.

  - I led this project. I found this problem and came up the solution.
  - I wrote all the codes in this project.
  - I conducted all the experiments in this paper.
  - I wrote the paper and drew all Figures in this paper.
  - This project also servers as my undergraduate thesis. I also wrote the English manuscript and previously submitted to JBHI, PR and CIBM. The manuscript can be found here.

- **Project 2:** MVS-GCN: A prior brain structure learning-guided multi-view graph convolution network for autism spectrum disorder diagnosis *Apr. 2021 - Aug. 2021*
  - This study aims to address the challenges posed by subject heterogeneity and noise correlations in brain networks using the novel MVS-GCN approach. The results demonstrate improved classification performance and alignment with existing neuroimaging evidence on ASD biomarkers (accepted at CIBM2022).
  - **My contributions:** 1) I provided the basic codebase which is derived from my previous work BrainTGL; 2) I participated in the discussion of the idea. For example, inspired the phenomenon: too diverse temporal graphs lead to poor performance, which is derived from my previous work BrainTGL, I suggested the consistency regularization among different views to enhance alignment of views. In fact, this idea was essential for this project. 3) I drew all figures in this paper.

OTHER WORK EXPERIENCE

Shenyang Sulianda Technology Co., Ltd. **Shengyang, CN**
AI Intern *Jun. 29, 2020 - Jul. 29, 2020*

- I finished my undergraduate field practice here.

TEACHING EXPERIENCE

Northeastern University, Natural Language Processing Laboratory **Shengyang, CN**
Instructor for Summer Internship - two direct interns *Jul. 2023 - Aug. 2023*

- Initially, I provided them with a comprehensive introduction to fundamental concepts in deep learning, encompassing topics like Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Multilayer Perceptrons (MLP), and Transformer models. Furthermore, I guided them through practical exercises utilizing the Fairseq framework to reinforce their understanding. Specifically, I required them to extend one of my previous work to a linear version.
- Additionally, I taught essential knowledge into the field of prompting engineering, spanning from elementary prompts to more advanced techniques such as Chain-of-Thoughts, Self-Consistency, Tree-of-Thoughts and so on. Beyond theory, I also taught them how to use OpenAI api to do specific task, such as machine translation.
- Furthermore, I extended an invitation for them to actively engage in my current research initiative, which focuses on exploring the utilization of Large Language Models (LLMs) in the domain of Machine Translation. Specifically, I required them to reproduce results of some baselines for comparison.

PROFESSIONAL SERVICES

Secondary Reviewers: ICML 2022, NeurIPS 2022, NLPCC 2023, TALLIP ((journal), TACL ((journal), Pattern Recognition ((journal).

RELEVANT COURSES

- Programming: C Programming (95), C++ Programming (97).
- AI related: Artificial Intelligence (98), Introduction of Machine Learning (95), Intelligent Computing System (100), Field Practice (95), Introduction of Bioinformatics (95), Introduction of Image Processing (91), Information Processing and Machine Translation (95).

SKILLS

Personalities: Persistent, passionate, and adaptable. (As evidenced by overcoming up to 10 rejections in my research journey.)
Programming: Python, LaTeX, C, C++, Java, Assembly
Tools: PyTorch, Fairseq, Tikz, Numpy, etal.

ENGLISH PROFICIENCY

TOEFL iBT (Best Score): 87 (*Reading:* 27, *Listening:* 19, *Speaking:* 19, *Writing:* 22)

TOEFL iBT (Single Test): 85 (*Reading:* 27, *Listening:* 19, *Speaking:* 19, *Writing:* 20)

PROJECTS & COMPETITIONS

Dynamic Brain Functional Network Analysis based Autism Diagnosis. **Shengyang, CN**
Undergraduate Thesis at NEU *Oct. 2020 - May. 2021*

- This research investigated the impact of dynamic patterns in fMRI data on the performance of deep neural networks (DNN) for Autism diagnosis. A key focus of the study was to devise an efficient method for integrating dynamic information into neural networks. We employed a sliding window strategy to generate dynamic data form and introduced a novel architecture that combines Graph Convolutional Networks (GCN) with Long Short-Term Memory (LSTM) to encode both spatial and dynamic patterns within fMRI data. Experimental results demonstrated the significant advantages of incorporating dynamic patterns for fMRI-based Autism diagnosis.
- The work has previously been submitted to a peer-reviewed journal.

Field Practice (Rank: 1 in this practice) **Shengyang, CN**
Undergraduate Field Practice at NEU *Jul. 2020 - Aug. 2020*

- I excellently completed the implementation of basic machine learning algorithms (logistic regression, K-means, SVM algorithm, neural network, CNN, GAN, Bayes algorithm, text hierarchical clustering algorithm) and solved some practical problems perfectly.
- I also accomplished the work of target detection and recognition based on MTCNN algorithm.

Assignment for Machine Learning Course (Rank: 1 in this Course) **Shengyang, CN**
Assignment for Introduction of Machine Learning *Nov. 2020 - Dec. 2020*

- I chose computer-aided autism diagnosis, which is a crucial application area in machine learning that is not yet mature.
- I examined the advantages of node attention, edge attention, and GCNs and created a novel unified framework integrating them all.
- I collected fMRI time series data using the ni-learn toolkit, dividing the brain into regions with three templates: CC200, CC400, and AAL.
- I implemented the pipeline using PyTorch and re-implemented some baselines with PyTorch.
- I employed a 10-fold cross-validation to validate the effectiveness of the proposed framework.

BANGC Programming Practice (Rank: 1, outstanding student award) **Shengyang,CN**
Practice of Intelligent Computing System *Nov. 2020 - Dec. 2020*

- During this practice, I conducted two experiments: 1) Implementing BANGC power difference operation and integrating it with TensorFlow; 2) BANGC operation quantization and offline inference. With these two experiments, I learnt the basic knowledge of quantization of deep neural networks and how to implement a operation and integrate it to TensorFlow.

Chinese College Student Computer Game Competition (3rd Prize) **Chongqin, CN**
Hosted by Chinese Society for Artificial Intelligence *Jul. 2020 - Aug. 2020*

- I participated in the Surakat chess track.
- I modified an open-source Surakat Chess AI algorithm based on MCTS to use the PVS algorithm and added a timing feature.

Design and Implementation of Graphic Editing Program **ShengYang, CN**
Assignment of Java Programming Course *Jul. 2019 - Aug. 2019*

- I designed and implemented a graphic editing program, which has functions of basic drawing.

Implementation of Virtual Scene **ShengYang, CN**
Assignment of Introduction to Image Processing *May. 2020 - Jun. 2020*

- I designed and implemented Sky box drawing, dynamic texture generation, model file reading, mouse and keyboard interaction.

FPGA-Based Eight-Bit Model Aircraft Design and Implementation **ShengYang, CN**
Digital Electronic Technology Course Project *Jun. 2019 - Jul. 2019*

- I designed and implemented an 8-Bit Model Aircraft System Using VHDL, including: 1) Instruction System and Instruction Formats, 2) Register Structure and Addressing Modes, 3) Memory System and Data Representation, 4) Arithmetic Logic Unit and 5) Controller.

New Instruction Design and Validation in COP2000 (Rank: 1) **ShengYang, CN**
Project Work for Computer Organization *Oct. 2020 - Nov. 2020*

- I designed a new instruction set in COP2000 IDE.
- I designed 4-bit and 8 bit multiplication and division programs based on above instructions.

Sports Event Management System **ShengYang, CN**
Basic Programming (C Language) Course Design *Mar. 2019 - Apr. 2019*

- I designed and implemented a sports event management system with essential features, including user authentication, match searching, and more.

AWARDS

- The Second Award Scholarship of Northeastern University Excellent Student of the Second Term, Northeastern University, China, 2017-2018.
- Northeastern University Chuanglian Industrial Scholarship, 2018-2019.
- The Third Award Scholarship of Northeastern University Excellent Student of the Second Term, Northeastern University, China, 2018-2019.
- The Third Award Scholarship of Northeastern University Excellent Student of the Second Term, Northeastern University, China, 2019-2020.
- The Second Award Scholarship of Northeastern University Excellent Student of the Second Term, Northeastern University, China, 2020-2021.
- The Third Award of 2020 "Competitive World Cup" Chinese College Student Computer Gaming Competition and the 14th China Computer Gaming Championship, Chinese Society of Artificial Intelligence and the Ministry of Education higher education computer teaching Steering Committee, in 2020.
- Outstanding student award of 2020 Intelligent Computing Systems Course, Institute of Computing Technology, Chinese Academy of Sciences, 2021.
- Cambrian best developers, Cambrian Industry, 2021.
- Third place in the school sports tug-of-war competition, 2019.
- Third place in the Computer Science Department soccer competition 2019.

EXTRACURRICULAR ACTIVITIES

- School Sports Meeting Opening Dance Participant.
- School Sports Meeting Tug of War Participant (3rd Prize).
- Participated in College Soccer Competition (3rd Prize).
- Returning to Promote Northeastern University at My Alma Mater.

CONFERENCES ATTENDED

- The Thirty-ninth International Conference on Machine Learning. (remote)Baltimore. 2022.
- CIPS ATT Issue 37&38. Beijing. 2023.