Zhuofeng Wu

Ph.D. Candidate at the School of Information, University of Michigan
3442 North Quad, 105 S. State St., Ann Arbor, MI, USA, 48109-1285
(E) zhuofeng@umich.edu (C) (734) 780-9664 (W) https://cserxy.github.io/

RESEARCH INTERESTS

Generative AI
Self-Supervised Representation Learning
Robustness of Machine Learning Models
Historical Language Change

EDUCATION

University of Michigan, Ann Arbor, US Aug 2018 – present
Ph.D. candidate in School of Information
Advisor: V.G. Vinod Vydiswaran
Committee: Rada Mihalcea, Chaowei Xiao, Paramveer Dhillon

Zhejiang University, Hangzhou, China Sept 2013 - Jun 2017
B.E. in the College of Computer Science & Chu Kochen Honors College
Advisor: Fei Wu
Received waiver for the National College Entrance Exam to enter Zhejiang University from **1st Prize in National Olympiad in Informatics in Provinces**

WORKING EXPERIENCE

Apple, Machine Learning Research team Apr 2023 – Aug 2023
ML Research Intern
Mentors: Yizhe Zhang and Navdeep Jaitly

Meta, AI Integrity team May 2021 – Aug 2021
Research Intern
Mentors: Sinong Wang and Hao Ma

Meta, AI Integrity team May 2020 – Aug 2020
Research Intern
Mentors: Sinong Wang and Hao Ma

Alibaba Group May 2019 – Aug 2019
Research Intern
Mentor: Fei Sun

PUBLICATIONS

HiCL: Hierarchical Contrastive Learning of Unsupervised Sentence Embeddings
Zhuofeng Wu, Chaowei Xiao, V. G. Vydiswaran
In Proceedings of Findings EMNLP 2023. (pdf)

PLANNER: Generating Diversified Paragraph via Latent Language Diffusion Model
Yizhe Zhang, Jiatao Gu, Zhuofeng Wu, Shuangfei Zhai, Josh Susskind, Navdeep Jaitly
In Proceedings of NeurIPS 2023. (pdf)

Defending against Insertion-based Textual Backdoor Attacks via Attribution
Jiazhao Li, Zhuofeng Wu, Wei Ping, Chaowei Xiao, V. G. Vydiswaran
In Proceedings of Findings ACL 2023. (pdf)

IDPG: An Instance-Dependent Prompt Generation Method
Zhuofeng Wu, Sinong Wang, Jiatao Gu, Rui Hou, Yuxiao Dong, V. G. Vydiswaran, Hao Ma
In Proceedings of NAACL 2022 (**Oral Presentation**). (pdf) (video)

Identify Shifts of Word Semantics through Bayesian Surprise
Zhuofeng Wu, Cheng Li, Zhe Zhao, Fei Wu, Qiaozhu Mei
In Proceedings of SIGIR 2018 **(Oral Presentation).** (pdf)

PREPRINT

Adversarial Demonstration Attacks on Large Language Models
Jiongxiao Wang, Zichen Liu, Keun Hee Park, Zhuojun Jiang, Zhaoheng Zheng, Zhuofeng Wu, Muhao Chen, Chaowei Xiao
arXiv preprint arXiv:2305.14950 (In submission to ACL 2024) (pdf)

Divide-or-Conquer? Which Part Should You Distill Your LLM?
Zhuofeng Wu, He Bai, Aonan Zhang, Jiatao Gu, VG Vydiswaran, Navdeep Jaitly, Yizhe Zhang
arXiv preprint arXiv:2402.15000 (In submission to ACL 2024) (pdf)

Chatgpt as an attack tool: Stealthy textual backdoor attack via blackbox generative model trigger
Jiazhao Li, Yijin Yang, Zhuofeng Wu, V. G. Vydiswaran, Chaowei Xiao
arXiv preprint arXiv:2304.14475 (In submission to NAACL 2024) (pdf)

Clear: Contrastive learning for sentence representation
Zhuofeng Wu, Sinong Wang, Jiatao Gu, Madian Khabsa, Fei Sun, Hao Ma
arXiv preprint arXiv:2012.15466 (2020). (pdf)

PROJECT EXPERIENCE

Apple, Machine Learning Research Apr 2023 – Aug 2023

Knowledge Distillation from LLM to Small Models: A Perspective from Question Decomposition

- Leverage LLMs such as GPT-4 to decompose a question into several related sub-questions.
- Fine-tune a model on the generated question-subquestions pair initially, and further train it based on the rewards from GPT-4.
- Extensive evaluations on GSM8k and DROP dataset show our proposed method can catch LLMs' question decomposition capability (and sometimes even better, e.g., better than ChatGPT).

Meta, AI Integrity May 2021 – Aug 2021

IDPG: An Instance-Dependent Prompt Generation Method

- First customized prompt for each input rather than one prompt for all inputs.
- Offered comparable performance to Adapter-based methods while using fewer parameters.
- Extensive evaluations on ten natural language understanding tasks show that IDPG consistently outperforms task-specific prompt tuning methods by 1.6–3.1 points.
- This work was presented at **NAACL'22** as **oral**.

Meta, AI Integrity May 2020 – Aug 2020

CLEAR: Contrastive Learning for Sentence Representation

- Proposed to align the representation of different argumentation for same sentence.
- Explored several argumentations and their combinations in the text domain.
- Revealed that different argumentations in pre-training enhance the model's different abilities.
- Outperformed several baselines (including BERT & RoBERTa) on GLUE & SentEval benchmark.

Alibaba Group May 2019 – Aug 2019

Seg-BERT: A Hierarchical Structure for Document Classification

- Applied a hierarchical structure for the long text classification.
- Outperformed the state-of-the-art by a large margin on IMDB.
- Proposed to mask sentence in pre-training to improve the performance.

School of Information, University of Michigan Aug 2018 – Nov 2020
Advisors: Prof. Qiaozhu Mei, Prof. Daniel Romero

Relocation Detection with Extra Information from Online Social Behavior on Twitter

- Proposed to extract extra information from online social behavior to help the relocation detection.

School of Information, University of Michigan Apr 2016 – Apr 2018
Advisor: Prof. Qiaozhu Mei

Identify Shifts of Word Semantics through Bayesian Surprise

- Explicitly established the stable topological structure of word semantics and identified the surprising changes over time.
- Proposed a statistical framework to apply **Bayesian Surprise** in detecting the meaning-changed words in **temporal-based word semantic networks**. This framework can be generalized to finding the change points in many other networks.

- Conducted experiments on ACMDL, DBLP and Google Books Ngram data set for synthetic evaluation which artificially introducing changes to a corpus. Outperformed the state-of-the-art by a large margin.
- This work was presented at **SIGIR'18** as **oral** and was adopted as a part of a **KDD'18 Workshop Keynote Talk** "Identifying Shifts in Evolutionary Semantic Spaces".

A Tool to Visualize the Evolution of Conference Topics

- Visualized a 40-year evolution of data science related communities and embedded papers, keywords, authors in the same space.
- Provided a powerful tool for researchers to model the research focus of different conferences.
- This work was presented in an invited talk in **KDD'18 Deep Learning Day** by Prof. Mei.

Digital Media Computing & Design Lab, Zhejiang University Sept 2014 - Mar 2016
Advisor: Prof. Fei Wu
Explored how to train different embedding models and implemented multiple word representation algorithms.

SKILLS

Programming Languages: Python, C, C++, Verilog, Pascal
Frameworks & Tools: PyTorch, Fairseq, LaTeX, Vim, Git

SERVICE

Student Volunteer

2022 Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL 2022)
The 41st International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2018)

Conference Reviewer

The Forty-first International Conference on Machine Learning (ICML 2024)
Twelfth International Conference on Learning Representations (ICLR 2024)
Thirty-seventh Conference on Neural Information Processing Systems (NeurIPS 2023)
The 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP 2023)
The 61st Annual Meeting of the Association for Computational Linguistics (ACL 2023)
The 46th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2023)

ACL Rolling Reviewer

February 2024 Cycle, December 2023 Cycle, October 2023 Cycle, June 2023 Cycle, April 2023 Cycle, December 2022 Cycle

TEACHING

Graduate Student Instructor

SI 670 Applied Machine Learning (Fall 2020)
SI 630 Natural Language Processing (Winter 2021 & Winter 2024)
SI 650 Information Retrieval (Fall 2021)
LHS 712 Natural Language Processing for Health (Winter 2021)

AWARDS

EMNLP Student Travel Grant from Big Picture Workshop, 2023.
SIGIR Student Travel Grant, 2018.
Outstanding Graduates of Zhejiang Province, 2017.
3rd Prize in Collegiate Programming Contest of Zhejiang University, 2014, 2015.
2nd Prize of Excellent Undergraduate Scholarship, 2014.
1st Prize in National Olympiad in Informatics in Provinces in 2012.
1st Prize in National Olympiad in Mathematics in Provinces in 2010.

MENTORING

Tian Xia, Undergraduate student at Umich May 2023 – present
Jiazhao Li, PhD student at Umich Sept 2021 – present

REFERENCES

V.G. Vinod Vydiswaran | vgvinodv@umich.edu
Associate Professor of Department of Learning Health Sciences, University of Michigan
Associate Professor of School of Information, University of Michigan

Chaowei Xiao | cxiao34@wisc.edu
Assistant Professor at the University of Wisconsin, Madison
Research Scientist at NVIDIA

Fei Sun | sunfei@ict.ac.cn
Associate Professor at the Institute of Computing Technology, Chinese Academy of Sciences