Zhuofeng Wu

3442 North Quad, 105 S. State St., Ann Arbor, MI, USA, 48109-1285
(C) (734) 780-9664 (E) zhuofeng@umich.edu

EDUCATION

University of Michigan, Ann Arbor, US Aug 2018 – present
Ph.D. candidate in School of Information (Advisor: V.G. Vinod Vydiswaran)
Natural Language Processing, Machine Learning

Zhejiang University, Hangzhou, China Sept 2013 - Jun 2017
B.E. in Computer Science (Overall GPA: 3.82/4.0, Top 5% among all 215 students)
Pursuit Science Class, Chu Kochen Honors College (CKC College)
Received waiver for the National College Entrance Exam to enter Zhejiang University from **1st Prize in National Olympiad in Informatics in Provinces** (top 1.8% over 60,000 participants)

PUBLICATIONS

Zhuofeng Wu, Chaowei Xiao, V. G. Vydiswaran. *HiCL: Hierarchical Contrastive Learning of Unsupervised Sentence Embeddings.* Findings of The 2023 Conference on Empirical Methods in Natural Language Processing. (**EMNLP'23**)

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

Jiazhao Li, **Zhuofeng Wu**, Wei Ping, Chaowei Xiao, V. G. Vydiswaran. *Defending against Insertion-based Textual Backdoor Attacks via Attribution.* Findings of the Association for Computational Linguistics: ACL 2023. (pdf)

Zhuofeng Wu, Sinong Wang, Jiatao Gu, Rui Hou, Yuxiao Dong, V. G. Vydiswaran, Hao Ma. *IDPG: An Instance-Dependent Prompt Generation Method*. Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. **(NAACL'22 oral)** (pdf)

Zhuofeng Wu, Cheng Li, Zhe Zhao, Fei Wu, Qiaozhu Mei. *Identify Shifts of Word Semantics through Bayesian Surprise*. Proceedings of the 41st International ACM SIGIR conference on Research and Development in Information Retrieval. ACM, 2018 **(SIGIR'18 oral)** (pdf)

PREPRINT

Jiazhao Li, Yijin Yang, **Zhuofeng Wu**, V. G. Vydiswaran, Chaowei Xiao. *Chatgpt as an attack tool: Stealthy textual backdoor attack via blackbox generative model trigger.* arXiv preprint arXiv:2304.14475 (In submission to EMNLP 2023) (pdf)

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

EXPERIENCE

Apple Machine Learning Research Apr 2023 – Aug 2023
Research Intern, Advisor: Dr. Yizhe Zhang
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).
- This work is an ongoing project and targets at **ICLR'24**.

Facebook AI
Research Intern, Advisor: Dr. Sinong Wang 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**.

Research Intern, Advisor: Dr. Sinong Wang 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
Research Intern, Advisor: Dr. Fei Sun

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
Research Assistant, 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
Research Intern, 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
Research Assistant, Advisor: Prof. Fei Wu

Explored how to train different embedding models and implemented several word representation algorithms in C++.

SKILLS

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

---

SERVICE

Conference Reviewer: ACL'23, SIGIR'23, EMNLP'23, NeurIPS'23.
ACL Rolling Reviewer: Dec 22, Apr 23, Jun 23.
Student volunteer for SIGIR'18, NAACL'22.
GSI for SI 670 Applied Machine Learning, SI 630 Natural Language Processing, SI 650 Information Retrieval, LHS 712 Natural Language Processing for Health.

AWARDS

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