FALI WANG

fqw5095@psu.edu ⋄ (+1) 814-325-1526
fairyfali.github.io ⋄ Github: FairyFali

EDUCATION

College of Information Sciences and Technology (IST), Penn State University (PSU)
Third-year Infomatics Ph.D. Candidate *08/2022-*

· Advisor: Dr. Suhang Wang
· Lab: Data Mining Lab

School of Cyber Security (SCS), University of Chinese Academy of Sciences (UCAS)
Master of Engineering in Software Engineering *09/2018-07/2021*

· Major GPA: 3.90/4
· Developed expertise in machine learning theory and applications.
· Developed skills in programming (Python, PyTorch).

School of Information and Computer Engineering, Northeast Forestry University (NEFU)
Bachelor of Engineering in Software Engineering *09/2014-07/2018*

· Major GPA: 4.09/5
· Ranking: No.1 in the grade
· Developed expertise in probability and statistics, and algebra.

RESEARCH INTERESTS

NLP	Customized LLMs, Knowledge Integration, LLM RAG
Graph	Graph Self-Training, Knowledge Graph

SELECTED PUBLICATIONS

· **Fali Wang**, Zhiwei Zhang, Xianren Zhang, Zongyu Wu, Tzuhao Mo, Qiuhao Lu, Wanjing Wang, Rui Li, Junjie Xu, Xianfeng Tang, Qi He, Yao Ma, Ming Huang, Suhang Wang✉. *A Comprehensive Survey of Small Language Models in the Era of Large Language Models: Techniques, Enhancements, Applications, Collaboration with LLMs, and Trustworthiness.* arXiv 2024.
· **Fali Wang**, Runxue Bao, Suhang Wang, Yanchi Liu, Wenchao Yu, Wei Cheng, Haifeng Chen. *InfuserKI: Infuser-Guided Knowledge Integration for Enhancing Large Language Models with Knowledge Graphs.* In EMNLP 2024.
· **Fali Wang**, Tianxiang Zhao, Junjie Xu, Suhang Wang✉. *HC-GST: Heterophily-aware Distribution Consistency based Graph Self-training.* In CIKM 2024.
· **Fali Wang**, Tianxiang Zhao, Suhang Wang✉. *Distribution Consistency-based Self-Training for Graph Neural Networks with Sparse Labels.* In WSDM 2024.
· **Fali Wang**, Zheng Lin✉, Zhengxiao Liu, Minyu Zheng, Lei Wang, Daren Zha. *MACROBERT: Maximizing Certified Region of BERT to Adversarial Word Substitutions.* In International Conference on Database Systems for Advanced Applications (DASFAA), 2021.
· Zhengxiao Liu, Bowen Shen, Zheng Lin✉, **Fali Wang**, Weiping Wang *Maximum Entropy Loss, the Silver Bullet Targeting Backdoor Attacks in Pre-trained Language Models.* In Findings of the Association for Computational Linguistics: ACL 2023.
· Zhengxiao Liu, **Fali Wang**✉, Zheng Lin, Lei Wang, Zhiyi Yin. *DE-CO_A Two-Step Spelling Correction Model for Combating Adversarial Typos.* In International Symposium on Parallel and Distributed Processing with Applications (ISPA), 2020.

PROJECT EXPERIENCE

Integrating Knowledge into LLMs *08/2023-now*

Enhancing Large Language Models with Knowledge Graphs through Infuser-Guided Knowledge Integration @ NEC Labs

- **Motivation:** We study a novel problem of effectively integrating unknown knowledge from KGs into LLMs without affecting known knowledge.
- **Method:** We propose InfuserKI, which enables the adaptive selection of additional information for known and unknown knowledge, effectively mitigating knowledge forgetting.
- **Experiments:** Evaluations on UMLS and MetaQA reveal InfuserKI's effective knowledge integration with less forgetting, sustained performance on large-scale data, and superior generality across unseen templates and downstream tasks.

Targeting Backdoor Attacks in Pre-trained Language Models *08/2022-12/2023*

Maximum Entropy Loss, the Silver Bullet Targeting Backdoor Attacks in Pre-trained Language Models

- **Motivation:** Backdoor attacks create a distribution gap between pre-trained and victim models, despite varying triggers.
- **Method:** Our approach counters attackers' minimum cross-entropy loss fine-tuning with maximum entropy loss on clean data.
- **Experiments:** On SST-2 and AG's News, within outsourcing attack scenarios, our method effectively eliminates backdoors, balancing backdoor removal and clean data accuracy.

Maximizing Certified Region of BERT *06/2020-12/2021*

MACROBERT: Maximizing Certified Region of BERT to Adversarial Word Substitutionss

- **Motivation:** Aiming to secure DNNs against adversarial examples, we focus on training models with certified robustness.
- **Method:** Our strategy, MACROBERT, enhances BERT models by replacing the base classifier with a differentiable soft smoothed classifier, thus expanding certified robustness.
- **Experiments:** Tests on IMDB and SNLI datasets reveal MACROBERT not only retains original performance on clean data but also outperforms existing defense methods in robustness.

SELECTED CERTIFICATES & SCHOLARSHIPS

National Scholarships

- Awarded Top 1 National Scholarship by China's MOE in both November 2015 and 2016.

Mathematical Contest in Modeling

- Honorable Mention at COMAP's Interdisciplinary Contest in Modeling, April 2018.
- Second Prize at CSIAM's National Undergraduate Math Contest in Modeling, December 2016.

Programming

- First Prize in Provincial and Third Prize in National at LAN QIAO International Programming Contest, May 2017.