JIAZHAO LI

(+1)734-604-1596, jiazhaol@umich.edu
Personal Website, Google Scholar

EDUCATION

University of Michigan, Ann Arbor U.S.
Ph.D. in Informatics (Natural Language Processing) *Sept.2020 – Apr.2025*
M.S. in Electrical Computer Engineering (Computer Vision) *Sept.2017 – May.2019*
Nankai University China
B.S. in Electrical Engineering *Sept.2013 – June.2017*

RESEARCH INTEREST

Natural Language Processing & CyberSecurity & Health Informatics
Backdoor Attack and Defense on NLP applications, Few-shot learning, Neural Machine Translation.

RESEARCH EXPERIENCE

BTAttack: Stealthy Textual Backdoor Attacks via Back-Translation
Under ACL Review *Nov 2022 - Jan 2023*

- Propose a stealthy, input-dependent backdoor attack method to mislead textual classifiers utilizing translation models as the trigger, making the generated backdoor examples less noticeable compared with baseline methods.
- BTAttack achieves higher semantic similarity by 0.23 and lower sentence perplexity by 41.65 and lower grammatical errors by 1.38 compared with baseline.
- BTAttack is easily accessible and achieves significant improvement in time efficiency when generating the poisoned sample, being 14.28 faster than syntax-based attacks.

Defending against Insertion-based Textual Backdoor Attacks via Attribution
Under ACL Review *Feb 2022 - Sep 2022*

- Build a defense framework against backdoor attacks on text classifier (pre-training and post-training)
- Apply a poisoned sample detector ELECTRA to identify poisoned samples.
- Identify triggers by calculating attribution score of tokens (trigger word contributes most to mislabeling)
- Achieve SOTA performance, an average accuracy of 79.97% (56.59%↑) and 48.34% (3.99%↑) on 4 benchmarks against pre-training attack and post-training attack respectively.
- Our defense method is more time-efficient, 3.13x faster than the baseline.

PharmMT: A Neural Machine Translation Approach to Simplify Prescription Directions.
In Findings of EMNLP'20 *Sept 2019 - Feb 2020*

- Built Nerual Network-based MT model between Prescription and Pharmacy directions corpus.
- Augmented model using MIMIC-III domain-specific pre-trained word embedding, external information from Drug/ Strength.
- Applied ensemble learning and numerical checking to improve accuracy and avoid fictitious generations.
- Applied BLEU score and SARI score to do automatic evaluation on MT performance and developed web app to do manual evaluation by pharmacists.

Open-domain Aspects Exploration for Qualitative Analysis via Active Learning
Under JAMIA review *Feb 2020 - Sep 2022*

- Build a framework to explore diverse aspects of selected theme (open-domain classification task)

- Use keyword-based filtering and binary text-classifier to collect the relevant sentence-level corpus.
- Select 'difficulty' samples (on classifier decision boundary) to the label instead of random sampling to accelerate diverse aspect exploration.

Re-ranking biomedical literature for precision medicine with pre-trained neural models.
ICHI'20 *Jan 2019 - May 2019*

- TREC precision medicine information retrieval challenge on ontology topics.
- calculating the relevant score using lexical-matching based iterate information retrieval method.
- calculating the relevant score using domain-adaptive contextual word embedding model BioBERT . Combining two relevant score using Rank Fusion.
- 6.2% improvement on inferred NDCG and 6.8% improvement on R-precision against SOTA models .

Video Segments Retrieval System based on Attentive CNN [Report] *Sep.2018 - Nov.2018*

- Enhanced video clip embedding with attentive-weighted contextual video segments embedding.
- Generated cross latent feature between video clip embedding and corresponding video content description text embedding through outer product.
- Trained ACNN model on TACoS dataset with loss function on video-text similarity and offset of video clips achieved 0.347 (IoU=0.5) and 0.719 (IoU=0.1) in Top10.

CONFERENCE PAPER

Jiazhao Li, Yijin Yang, Zhuofeng Wu, V.G.Vinod Vydiswaran, Chaowei Xiao. BTAttack: Stealthy Textual Backdoor Attacks via Back-Translation *(Under ACL23' Review)*
Jiazhao Li, Zhuofeng Wu, Wei Ping, Chaowei Xiao, V.G.Vinod Vydiswaran. Defending against Insertion-based Textual Backdoor Attacks via Attribution *(Under ACL23' Review)*
Jiazhao Li, Corey Lester, Xinyan Zhao, Yuting Ding, Yun Jiang, and V.G.Vinod Vydiswaran. PharmMT: A Neural Machine Translation Approach to Simplify Prescription Directions. *In Findings of EMNLP, the 2020 Conference on Empirical Methods in Natural Language Processing. Pages:2785–2796.*
Jiazhao Li, Adharsh Murali, Qiaozhu Mei, V.G.Vinod Vydiswaran. Re-ranking biomedical literature for precision medicine with pre-trained neural models. *Proceedings of the IEEE International Conference of Healthcare Informatics (ICHI), 2020.*

JOURNAL PAPER

Lester, C.A., **Li, J.**, Ding, Y. et al. Performance evaluation of a prescription medication image classification model: an observational cohort. *npj Digit. Med. 4, 118 (2021).*
Lester CA, Ding Y, **Li J**, Jiang Y, Rowell B, Vydiswaran VGV, Comparing Human versus Machine Translation of Electronic Prescription Directions *Journal of the American Pharmacists Association (2021)*
Chang T, DeJonckheere M, Vydiswaran VGV, **Li J**, Buis L, Guetterman T. Accelerating Mixed Methods Research with Natural Language Processing of Big Text Data. *Journal of Mixed Methods Research (2021).*
Zhao X, **Li J**, Lester C, Jiang Y, Vydiswaran VGV *Focused representation models for transcribing prescription instructions.* (Poster MIDAS 2019 Symposium)

WORK EXPERIENCE

Graduate Student Instructor (LHS 712: NLP for Health) Jan.2023 - Present
Graduate Student Research Assistant Sep.2020 - Dec. 2022
Research Associate Aug.2019 - Aug.2020

PUBLIC SERVICE

- Reviewer: ACL 23', EMNLP 22'21', EACL 22', NAACL 21'
- External Reviewer: Frontiers in Big Data, section Cybersecurity and Privacy.