Yuting Yang (杨玉婷)

: +86 15600818200 | ✉: yangyuting@ict.ac.cn | : https://yang-yuting.github.io

Beijing, China

EDUCATION

Institute of Computing Technology, Chinese Academy of Sciences — *Beijing China*
Ph.D. in Computer Science — *2017.09 ~ Present*
- Advisor: *Professor Jintao Li*
- Research interests: *Trustworthy AI, Dialogue System*

NExT++ Research Centre, National University of Singapore — *Singapore*
Visiting Research Scholar — *2021.03 ~ 2022.03*
- Advisor: *Professor Tat-Seng Chua*

Jilin University (985/211 University) — *Jilin, China*
B.S. in Computer Science — *2013.09 ~ 2017.06*
- GPA rank: top 2% (7/300)

RESEARCH EXPERIENCE

Trustworthy AI — *2019.01~ Present*

Funded by National Natural Science Foundation of China

- Rethought the robustness of deep neural networks (DNNs) from a quantification perspective in both NLP [5] and image [3] fields. Proposed weak robustness to evaluate DNN's capability of resisting perturbation and establish the concept of "sufficiently safe". It mines an interesting property of DNNs: adversarial examples occupy a very small ratio in the input space. The property provides insight into figuring out the relationship between robustness and generalization.
- Based on the weak robustness, proposed a model-agnostic method for enhancing the word-level robustness of deep NLP models. Via input perturbation, the method can significantly decrease the rate of successfully attacking and maintain generalization to a great extent. [2]
- Further applied weak robustness combined with ensemble learning to resist a variety of text attacks. Proposed attention-based diversity to promote model diversity, which can consistently improve the defense ability against many types of adversarial attacks and presents good interpretability via visualizing diverse attention patterns. [6]
- Utilized prompting to explore robustness defects of deep NLP models. By malicious prompts construction and mask-filling process, the proposed prompt-based attack algorithm can generate more diverse and natural adversarial examples.

Dialogue System — *2021.03~2022.03*

Funded by Scholarship of University of Chinese Academy of Sciences

- Designed a dual prompt learning framework for few-shot dialog state tracking (DST). The framework can probe DST-related knowledge from pre-trained language models, address low-resource DST efficiently and generate unseen slots. The few-shot method also helps to build a more trustworthy model adaptive to dynamic environments. [1]

Text Generation — *2017.09~2019.01*

Funded by National Natural Science Foundation of China

- Realized a news quality assessment model. The model provides comprehensive analyses of social media news considering eight types of linguistic features [7]. The assessment model is patented and applied in practical applications of official news media (*Xinhua Net* and *People's Daily*).
- Realized guideline-based news headline generation model. The model incorporated nondifferentiable writing guidelines into automatic generation via reinforcement learning. It can generate headlines containing key information and style in writing guidelines, and is applied in practical applications of official news media (*Xinhua Net* and *People's Daily*).

TEACHING EXPERIENCE

- Teaching Assistant of *Multimodal Learning*, University of Chinese Academy of Sciences (2018 spring)

SELECTED HONORS AND AWARDS

President Scholarship, Institute of Computing Technology, Chinese Academy of Sciences, 2020
(Top honor for students in Institute of Computing Technology.)

Merit Student, University of Chinese Academy of Sciences, 2020

Academic Scholarship, University of Chinese Academy of Sciences, 2017~2022

National Scholarship, Ministry of Education of China, 2015
(Top 0.2% of Chinese undergraduate students.)

COMPETITION

- Fourteenth Place Award in Security AI Challenger Program, CVPR2021 Security AI Challenger Phase VI, White-box Adversarial Attacks. (14/1681, top 1%)

PUBLICATIONS

[1] **Yuting Yang**, Wenqiang Lei, Pei Huang, Juan Cao, Jintao Li and Tat-Seng Chua. A Dual Prompt Learning Framework for Few-Shot Dialogue State Tracking, *The Web Conference, **WWW'23***, Texas, USA 2023. (**Regular paper, Research track**)

[2] Pei Huang *, **Yuting Yang***, Fuqi Jia, Minghao Liu, Feifei Ma and Jian Zhang. Word Level Robustness Enhancement: Fight Perturbation with Perturbation, *Thirty-Sixth AAAI Conference on Artificial Intelligence, **AAAI'22***, Vancouver, BC, Canada 2022. **(*Co-First Author, Regular paper, Research track, Acceptance Rate: 1349/9020=15.0%)**

[3] Pei Huang*, **Yuting Yang***, Minghao Liu, Fuqi Jia, Feifei Ma, and Jian Zhang. ε-weakened Robustness of Deep Neural Networks, *Thirty-First ACM SIGSOFT International Symposium on Software Testing and Analysis, **ISSTA'22***, Daejeon, South Korea, 2022. **(*Co-First Author, Regular paper, Research track, Top conference on software analysis)**

[4] **Yuting Yang**, Pei Huang, Juan Cao, Jintao Li, Yun Lin and Feifei Ma. A Prompt-based Approach to Adversarial Example Generation and Robustness Enhancement, *Frontier of Computer Sciences,* 2023. (**SCI Journal**)

[5] **Yuting Yang,** Pei Huang, Feifei Ma, Juan Cao, Meishan Zhang, Jian Zhang and Jintao Li. Quantifying Robustness to Adversarial Word Substitutions. CoRR abs/2201.03829 (2022). (Submitted to ECML/PKDD)

[6] **Yuting Yang**, Pei Huang, Juan Cao, Danding Wang and Jintao Li. PAD: A Robustness Enhancement Ensemble Method via Promoting Attention Diversity. (Submitted to ARR)

[7] **Yuting Yang**, Juan Cao, Mingyan Lu, Jintao Li, Chia-Wen Lin, How to Write High-quality News on Social Network? Predicting News Quality by Mining Writing Style. *CoRR abs/1902.00750*, 2019.

PATENTS

[1] An evaluation system for the vulnerability of social media. Jianfeng Shangguan, Juan Cao, **Yuting Yang**, Jintao Li. CN107 886441A.

[2] A system for modeling news style and evaluating news quality. Juan Cao, **Yuting Yang**, Tian Xie and Junbo Guo. CN1115 53146A.

FUNDING APPLICATIONS

Joined in applying for funding and writing research proposals for the following projects:

- Privacy Protection, Special Foundation for State Major Basic Research Program of China, 2017-2018
- Research on Internet Rumor Detection and Public Opinion Guidance, National Natural Science Foundation of China, 2019-2020
- Research on Reliability Evaluation of User-Generated Content, National Natural Science Foundation of China, 2019-2022

ACADEMIC ACTIVITIES

- PC member: AAAI 2022
- Reviewer: AAAI 2022, ACL Rolling Review (2021-2022), WWW 2022, EAAI 2022, KDD 2023
- Conference Volunteer: ICDM 2019