Yuting Yang

ytyang17@gmail.com | +86 15600818200 | Google Scholar | Personal Website

Beijing, China

EDUCATION

University of Chinese Academy of Sciences *Beijing, China*
Ph.D. in Computer Science *2017.09 ~ 2023.07*
- Supervisor: *Professor Jintao Li*
- Research interests: *natural language processing*

NExT++ Research Centre, National University of Singapore *Singapore*
Visiting Research Scholar *2021.03 ~ 2022.03*
- Supervisor: *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

Robustness *2019.01~ Present*

Make intelligent systems adaptive to dynamic changes in environments, funded by National Natural Science Foundation of China

- Rethought the robustness of deep neural networks (DNNs) from a quantification perspective in both Natural Language Processing (NLP) [5] and image [3] fields. Proposed "weak robustness" to evaluate DNN's capability of resisting perturbation and establish the concept of "sufficiently stable". It mines an interesting property of DNNs: adversarial examples occupy a very small ratio in the input space, which provides a new 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 perturbing and maintain generalization to a great extent. [2]
- Further applied weak robustness combined with ensemble learning to achieve perturbation-agnostic robustness enhancement. Proposed attention-based diversity to promote model diversity, which can consistently improve the stability against various types of adversarial perturbation and presents good interpretability. [7]
- Among first to propose that prompt can be maliciously constructed to arise robustness issues of pre-trained language models, which later became a popular research topic (LLM alignment) with the popularity of large models and prompt learning. [4]
- Proposed a lightweight model that excels in detection accuracy and demonstrates resilience against adversarial prompts for large language models [9].

Dialogue System *2021.03~2022.03*

Funded by Scholarship of University of Chinese Academy of Sciences

- Pioneered the idea that prompt learning could be used to understand dialogue states in few-shot or zero-shot settings. This idea was later proven to be one of the keys to realizing intelligent dialog systems by the success of ChatGPT. [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. The assessment model is patented and applied in practical applications.
- 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.

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.)

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, ***WWW*** *2023*.

[2] Pei Huang *, **Yuting Yang***, Fuqi Jia, Minghao Liu, Feifei Ma and Jian Zhang. Word Level Robustness Enhancement: Fight Perturbation with Perturbation, ***AAAI*** *2022*. **(*Co-First Author, 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, ***ISSTA*** *2022*. **(*Co-First Author, 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, ***ECML-PKDD*** *2023*.

[6] Pei Huang, Haoze Wu, **Yuting Yang**, Ieva Daukantas, Min Wu, Yedi Zhang, Clark Barrett, Towards Efficient Verification of Quantized Neural Networks, ***AAAI*** *2024*. (**Oral**)

[7] **Yuting Yang**, Pei Huang, Juan Cao, Danding Wang and Jintao Li. PAD: A Robustness Enhancement Ensemble Method via Promoting Attention Diversity, ***COLING*** *2024*.

[8] Pei Huang, **Yuting Yang**, Haoze Wu, Ieva Daukantas, Min Wu, Fuqi Jia and Clark Barrett. Parallel Verification for δ-Equivalence of Neural Network Quantization, ***SAIV*** *2024.*

[9] **Yuting Yang**, Tianyu Pang, Chao Du, Mohan Kankanhalli and Min Lin. An Efficient Adversarial Prompt Shield for Large Language Models. ***COLM*** 2024 (under revision).

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.

ACADEMIC ACTIVITIES

- PC member/Reviewer: AAAI 2022-2024, ACL Rolling Review (2021-2023), WWW 2022, EAAI 2022, KDD 2023
- Conference Volunteer: ICDM 2019

TECHNICAL SKILLS

- Programming Languages: Python, C, C++, SQL, Java, MATLAB, Latex
- Tools & Libraries: Pytorch, Keras, TensorFlow, Hugging Face…