

Jun Yan

✉ yanjun@usc.edu 🌐 junyann.github.io

RESEARCH INTERESTS

Trustworthy NLP (Safety, Robustness), Question Answering, Information Extraction

EDUCATION

University of Southern California

Ph.D. in Computer Science

◦ Advisor: Prof. Xiang Ren

Los Angeles, CA, U.S.

08/2019 - 05/2024 (*expected*)

Tsinghua University

B.Eng. in Electronic Engineering

◦ Advisor: Prof. Zhiyuan Liu

Beijing, China

08/2015 - 07/2019

PUBLICATIONS AND PREPRINTS

(* indicates equal contribution)

- Virtual Prompt Injection for Instruction-Tuned Large Language Models
arXiv:2307.16888
Jun Yan, Vikas Yadav*, Shiyang Li*, Lichang Chen, Zheng Tang, Hai Wang, Vijay Srinivasan, Xiang Ren, Hongxia Jin
- Instruction-following Evaluation through Verbalizer Manipulation
arXiv:2307.10558
Shiyang Li, **Jun Yan**, Hai Wang, Zheng Tang, Xiang Ren, Vijay Srinivasan, Hongxia Jin
- AlpGasus: Training A Better Alpaca with Fewer Data
arXiv:2307.08701
Lichang Chen*, Shiyang Li*, **Jun Yan**, Hai Wang, Kalpa Gunaratna, Vikas Yadav, Zheng Tang, Vijay Srinivasan, Tianyi Zhou, Heng Huang, Hongxia Jin
- BITE: Textual Backdoor Attacks with Iterative Trigger Injection
In *Proceedings of ACL'23* (short version in *BANDS@ICLR'23*)
Jun Yan, Vansh Gupta, Xiang Ren
- On the Robustness of Reading Comprehension Models to Entity Renaming
In *Proceedings of NAACL'22*
Jun Yan, Yang Xiao, Sagnik Mukherjee, Bill Yuchen Lin, Robin Jia, Xiang Ren
- RockNER: A Simple Method to Create Adversarial Examples for Evaluating the Robustness of Named Entity Recognition Models
In *Proceedings of EMNLP'21*
Bill Yuchen Lin, Wenyang Gao, **Jun Yan**, Ryan Moreno, Xiang Ren
- AdaTag: Multi-Attribute Value Extraction from Product Profiles with Adaptive Decoding
In *Proceedings of ACL-IJCNLP'21*
Jun Yan, Nasser Zalmout, Yan Liang, Christan Grant, Xiang Ren, Xin Luna Dong
- Learning Contextualized Knowledge Structures for Commonsense Reasoning

In *Findings of ACL-IJCNLP'21* (short version in *KR2ML@NeurIPS'20*)

Jun Yan, Mrigank Raman, Aaron Chan, Tianyu Zhang, Ryan Rossi, Handong Zhao, Sungchul Kim, Nedim Lipka, Xiang Ren

- Scalable Multi-Hop Relational Reasoning for Knowledge-Aware Question Answering
In *Proceedings of EMNLP'20*
Yanlin Feng*, Xinyue Chen*, Bill Yuchen Lin, Peifeng Wang, **Jun Yan**, Xiang Ren
- Learning from Explanations with Neural Execution Tree
In *Proceedings of ICLR'20*
Ziqi Wang*, Yujia Qin*, Wenxuan Zhou, **Jun Yan**, Qinyuan Ye, Leonardo Neves, Zhiyuan Liu, Xiang Ren
- Learning Dual Retrieval Module for Semi-supervised Relation Extraction
In *Proceedings of TheWebConf'19*
Hongtao Lin, **Jun Yan**, Meng Qu, Xiang Ren
- Language Modeling with Sparse Product of Sememe Experts
In *Proceedings of EMNLP'18*
Yihong Gu*, **Jun Yan***, Hao Zhu*, Zhiyuan Liu, Ruobing Xie, Maosong Sun, Fen Lin, Leyu Lin

WORK EXPERIENCE

Research Intern @ Meta AI

Seattle, WA, U.S.

Language Understanding and Question Answering Team

05/2022 – 08/2022

- Mentors: Asish Ghoshal, Scott Wen-tau Yih, Asli Celikyilmaz, Pedro Rodriguez
- I studied the problem of finetuning text-to-text transformers (e.g., BART) for robust generalization to unseen test domains. I proposed a parameter-efficient finetuning method that encourages the model to skip unnecessary adapter layers during adapter tuning. By adaptively limiting the model's capacity, it alleviates the model's overfitting to the training data and achieves better out-of-domain generalization.

Applied Scientist Intern @ Amazon

Seattle, WA, U.S. (Remote)

Product Graph Team

06/2020 – 11/2020

- Mentors: Nasser Zalmout, Yan Liang, Xin Luna Dong
- I studied the problem of extracting attribute values from product profiles. I proposed a model that can handle the value extraction of many attributes. It uses an adaptive decoding mechanism that parameterizes the decoder based on the query attribute. The model achieves improved performance with knowledge sharing across attributes and demonstrates strong scalability with respect to the number of attributes.

HONORS AND AWARDS

- Annenberg Fellowship, University of Southern California. 2019.
- Excellent Graduate, Tsinghua University. 2019.
- Samsung/JJWorld/Evergrande Scholarship, Tsinghua University. 2016/2017/2018.

SERVICES

- PC/Reviewer: ARR, ACL(2021), EMNLP (2020-2022), NAACL(2021), IEEE TNNLS.

KEY SKILLS

Programming Languages

Python, C, C++, Java, MATLAB

Machine Learning Libraries

PyTorch, Scikit-Learn, NumPy, Pandas