

Yuchen Shen

yuchens3@cs.cmu.edu | website | [Google Scholar](#)

Education

Carnegie Mellon University 2023/08-2024/12 (Expected)
Master of Science in Intelligent Information Systems, Language Technologies Institute
Pittsburgh, PA, USA
Courses: Probabilistic Graphical Models, Multimodal Machine Learning, Advanced Deep Learning

University of Electronic Science and Technology of China 2019/08-2023/06
Bachelor of Engineering in Software Engineering, School of Information and Software Engineering
Chengdu, China
Courses: Program Design and Algorithm Foundation, Probability and Mathematical Statistics, Calculus

Publications

- [1] Zhichao Hou, Weizhi Gao, **Yuchen Shen**, and Xiaorui Liu. [Protransformer: Robustify transformers via plug-and-play paradigm](#). ICLR 2024 Workshop on Reliable and Responsible Foundation Models, 2024.
- [2] **Yuchen Shen**, Xiaojun Wan. [OpinSummEval: Revisiting Automated Evaluation for Opinion Summarization](#). Arxiv:2310.18122.
- [3] Fei Zhao*, **Yuchen Shen***, Zhen Wu, and Xinyu Dai. [Label-driven denoising framework for multi-label few-shot aspect category detection](#). Findings of the Association for Computational Linguistics: EMNLP 2022.
- [4] Mengjuan Liu, Xiaoming Bao, Jiang Liu, Pei Zhao, and **Yuchen Shen**. [Generating emotional response by conditional variational auto-encoder in open-domain dialogue system](#). Neurocomputing, 460:106–116, 2021.

Research Experience

Prior-Data Fitted Network for Anomaly Detection, Carnegie Mellon University 2024/05 – present
Advisor: Leman Akoglu, Associate Professor, Heinz College of Information Systems and Public Policy.

- Proposed to pre-train Prior-Data Fitted Network (PFN) on synthetic data and detect anomalies without further fine-tuning.
- Achieved an AUCROC of 99.72 and an AUCPR of 98.74 on synthetic anomalies drawn from Gaussian Mixture Models.

Graph Generative Pre-training, Carnegie Mellon University 2024/05 – present
Advisor: Leman Akoglu, Associate Professor, Heinz College of Information Systems and Public Policy.

- Proposed to pre-train on molecules with auto-regressive generative models for 2D graphs.
- Investigated how dataset sizes, model architectures, and pre-training methods affect the quality of pre-training on molecules.

Geometry Optimized Molecule Generation with Chemistry Guidance, Carnegie Mellon University 2024/01 – 2024/05
Advisor: Barnabás Póczos, Associate Professor, Machine Learning Department, School of Computer Science.

- Proposed to optimize the stability of molecules with a latent diffusion model and [xtb](#), a semi-empirical quantum mechanical method for accurate and efficient quantum chemistry calculations, such as the force on the atoms.
- Developed a method to estimate the gradient from the non-differentiable [xtb](#) to guide the generation process of 3D molecules.
- Achieved a 7.22%–9.19% decrease in L-1 norm for forces and improved the validity of generated molecules by 5.0%–10.0%.

Graph Tokenization for Molecules, Carnegie Mellon University 2024/01 – 2024/05
Advisor: Barnabás Póczos, Associate Professor, Machine Learning Department, School of Computer Science.

- Proposed to mine common sub-graphs based on the topological structures and node types of molecules from a large corpus.
- Generalized Byte-Pair Encoding (BPE) algorithm to 2D graphs and developed a non-parametric tokenization method that partitions 2D graphs into subgraphs, which can transform the current graph into a new 2D graph and a hypergraph.
- Improved the performance for different GNNs and HGNNs (for hypergraphs) on both classification and regression datasets across various hyperparameters. Findings demonstrated the importance of data preprocessing in molecular machine learning.

Controllable Toxicity Generation for Plant Molecules, Carnegie Mellon University 2023/09 – 2023/12
Advisor: Barnabás Póczos, Associate Professor, Machine Learning Department, School of Computer Science.

- Constructed a dataset consisting of 1535 toxic and 39576 non-toxic plant molecules.
- Achieved a classification accuracy of 80.8% on a balanced test set (1:1) for an imbalanced training set (~1:35). Proposed to achieve controllable toxicity generation via diffusion models with contrastive learning at both molecule-level and model-level.

Convergence of Decentralized Machine Learning Algorithms, North Carolina State University 2023/02 – 2023/12
Advisor: Xiaorui Liu, Assistant Professor, Department of Computer Science, College of Engineering.

- Analyzed convergence rates for different decentralized algorithms under a (strongly-)convex setting.
- Studied and generalized the previous method to analyze lower-bound for decentralized optimization with (strongly-)convex objectives. Improved the tightness of previous proofs on the convergence of decentralized algorithms such as [LEAD](#).

Automated Metric Evaluation for Opinion Summarization, Peking University

2022/12 – 2023/05

Advisor: Xiaojun Wan, Professor, Wangxuan Institute of Computer Technology.

- Proposed to evaluate automated metrics with aspect relevance, self-coherence, sentiment consistency, and readability.
- Constructed a dataset with annotated outputs from 14 popularly used models in opinion summarization.
- Analyzed 26 popularly used automatic metrics, showing neural-based metrics correlate better with human annotations.

Zero-Shot Unsupervised Opinion Summarization with Prefix-Tuning, Peking University

2022/02 – 2022/12

Advisor: Xiaojun Wan, Professor, Wangxuan Institute of Computer Technology.

- Proposed a zero-shot setting for unsupervised opinion summarization with unseen aspects to summarize in the test stage.
- Proposed to control the number of aspects and sentiment coherency in the generated summary with two kinds of prefixes.
- Improved the zero-shot performance on the Space dataset compared with strong baselines (e.g., on the “service” aspect, ROUGE-1 score raised from 33.56 to 35.94 compared with previous fine-tuned SOTA model AceSum).

Label-enhanced Few-shot Learning for Multi-label Aspect Category Detection, Nanjing University

2021/08 – 2022/01

Advisor: Xinyu Dai, Professor, Department of Computer Science and Technology.

- Identified the generic and noisy features as the bottleneck for multi-label few-shot aspect category detection (FS-ACD).
- Proposed to denoise the feature of each category with label texts. Designed a label-guided attention module and a label-weighted contrastive loss for FS-ACD to learn representative features and to distinguish semantically close categories.
- Improved the performance of existing models on FS-ACD (e.g., F1 score for the 5-way 5-shot setting (5 classes with 5 instances to learn per class) raised from 75.37 to 78.27 on the FewAsp dataset for model Proto-AWATT).

Projects

Edge Weighting Algorithm with Ollivier-Ricci Curvature for Graph Classification

2023/09 – 2023/12

Carnegie Mellon University, 10-708 Probabilistic Graphical Models

- Proposed to weight edges based on Ollivier-Ricci curvature to overcome over-smoothing and over-squashing.
- Proposed to optimize the idleness in the computation of the Ollivier-Ricci curvature and achieved a classification accuracy of 70.90 ± 0.047 (with a baseline accuracy of 69.59 ± 0.048) on the Proteins dataset by weighting edges.

Multimodal Low-Rank Adaptation

2023/09 – 2023/12

Carnegie Mellon University, 11-777 MultiModal Machine Learning

- Proposed to share $n\%$ of the Low-Rank Adaptation (LoRA) matrices for texts and images in vision language models.
- Developed a regularization loss based on singular value decomposition (SVD) that promotes the similarity of LoRA matrices if more information is shared between the input language and vision modalities.
- Improved the performance of LoRA on 6 out of 7 subtasks of the VALSE benchmark (e.g., our method improved the pairwise rank accuracy of vanilla LoRA from 44.2% to 49.0% on the coreference-clean data for the CLIP model).

Deep Neural Models in Aspect-Based Sentiment Analysis

2021/07 – 2021/08

2021 Nanjing University NLP Summer Camp

- Implemented sentiment analysis models such as interactive attention network (IAN) and aspect-specific graph convolutional network (ASGCN).
- Improved the accuracy by 14.86% and 7.37% on SemEval 2014 laptop and restaurant datasets respectively using IAN augmented with the pre-trained BERT model. Validated the effectiveness of graph information in ASGCN with 2.8% accuracy improvement over IAN on the laptop dataset.

Skills

Language: Chinese (native), English (fluent)

Programming: Python (proficient), C, Java