Zuoyu Yan

Ph.D. Student ·

Wangxuan Institute of Computer Technology, Peking University

yanzuoyu3@pku.edu.cn | https://pkuyzy.github.io/ | https://github.com/pkuyzy

Research Interest

My research interest lies in the intersection of machine learning and topological data analysis. In particular, I am interested in:

- Enhancing modern machine learning frameworks, e.g., graph neural networks, with topology/geometry.
- Applying these models in real-world scenarios, e.g., biomedicine, and recommendation systems.

Education

Peking University

Ph.D. Student, Computer Science *2019.09 till now*

Peking University

Undergraduate Degree, Data Science *2015.09 - 2019.06*

Intern Experience

Alibaba DAMO Academy

2021.05-2021.10

- Multimodal entity linking on Wikipedia and Wikinews data.

Publications

Zuoyu Yan, Tengfei Ma, Liangcai Gao, Zhi Tang, Yusu Wang, Chao Chen. Learning on Graphs Conference (**LoG**), 2023. (**Oral**)

Zuoyu Yan, Junru Zhou, Liangcai Gao, Zhi Tang, Muhan Zhang. Efficiently Counting Substructures by Subgraph GNNs without Running GNN on Subgraphs. arXiv, 2023.

Zuoyu Yan, Tengfei Ma, Liangcai Gao, Zhi Tang, Yusu Wang, Chao Chen. Neural Approximation of Graph Topological Features. Advances in Neural Information Processing Systems (**NeurIPS**, Top conference in machine learning.), 2022. (**Spotlight, acceptance rate 4%-5%**)

Zuoyu Yan, Tengfei Ma, Liangcai Gao, Zhi Tang, Chao Chen. Cycle Representation Learning for Inductive Relation Prediction. International Conference on Machine Learning (**ICML**, Top conference in machine learning.), 2022.(**Short talk, acceptance rate 21.9%**)

Zuoyu Yan, Tengfei Ma, Liangcai Gao, Zhi Tang, Chao Chen. Link Prediction with Persistent Homology: An Interactive View. International Conference on Machine Learning (**ICML**, Top conference in machine learning.), 2021.(**Short talk, acceptance rate 21.5%**)

Zuoyu Yan, Xinpeng Zhang, Liangcai Gao, Ke Yuan, Zhi Tang. ConvMath: A Convolutional Sequence Network for Mathematical Expression Recognition. International Conference on Pattern Recognition (**ICPR**), 2020.

Wenqi Zhao, Liangcai Gao, **Zuoyu Yan**, Shuai Peng, Lin Du, Ziyin Zhang. Handwritten mathematical expression recognition with bidirectionally trained transformer. International Conference on Document Analysis and Recognition (**ICDAR**), 2021. (**Best poster award, 2 out of 340 submissions**).

Liangcai Gao, Xiaohan Yi, Yuan Liao, Zhuoren Jiang, **Zuoyu Yan**, Zhi Tang. A deep learning-based formula detection method for PDF documents. International Conference on Document Analysis and Recognition (**ICDAR**) 2017.

Honors

- Won the 2021-2022th ”Outstanding Research Award” of Peking University. (15% out of 27 PhD students of Wangxuan Institute of Computer Technology of Peking University (**WICT**))
- Won the 2021-2022th ”Outstanding Student Award” of WICT. (7% out of 70 students of WICT)
- Won the 2018-2019th Excellent Project of Peking University President’s Fund (10% out of 97 projects)
- Won the third prize in the ”Schlumberger Cup” programming competition held in Peking University.
- Ranked 8th(out of 836 teams) in the Plant Seedlings Classification competition held on Kaggle website

Skills

- Languages: C/C++, Matlab, Python
- Tools: Torch, Tensorflow, PyTorch, OpenCV

Past projects

In the domain of machine learning, classic models often utilize data samples without considering the underlying topology or geometry, which restricts their representational power. Conversely, advanced machine learning models and topological features, though powerful, are frequently associated with high computational demands. This aspect renders them less feasible for application in complex and extensive graphs, a common occurrence in real-world scenarios. My research, as outlined below, have been primarily concentrated on overcoming these limitations.

Graph Machine Learning. In graph data, structures such as rings and the triadic closure property constitute meaningful components, thus are essential to graph machine learning. Classic graph machine learning models, however, are unable to detect various structures, including cycles and cliques, leading to failures in many relevant tasks. To address the limitation, I have developed robust features that contain crucial structural information and organically integrated them with graph learning models (**ICML’ 2021, ICML’2022, arXiv’ 2023, LoG’ 2023**). These works have demonstrated both theoretical and empirical enhancements in the representational power across diverse benchmarks.

Topological Representation Acceleration. To accelerate the computation of these powerful topological features, I propose a provably more efficient algorithm (**ICML’ 2021**) that significantly accelerates their computation. Furthermore, I have developed a machine learning model with strong algorithmic alignment designed to approximate these features (**NeurIPS’ 2022**). The model not only achieves strong approximation accuracy but also runs nearly 100 times faster compared to conventional algorithms. Given its proven transferability, the model holds substantial potential for application across various real-world datasets.

Professional Service

Conference Reviewer for:

- Conference on Neural Information Processing Systems (NeurIPS)
- International Conference on Machine Learning (ICML)
- International Conference on Learning Representations (ICLR)
- Learning on Graphs Conference (LoG)

Reference

- **Zhi Tang**
  Professor, Peking University
  tangzhi@pku.edu.cn
  https://www.wict.pku.edu.cn/cpdp/kydw/ggcy/1297369.htm
- **Chao Chen**
  Associate Professor, Stony Brook University
  chao.chen.1@stonybrook.edu
  https://chaochen.github.io/
- **Yusu Wang**
  Professor, University of California, San Diego
  yusuwang@ucsd.edu
  http://yusu.belkin-wang.org/
- **Tengfei Ma**
  Assistant Professor, Stony Brook University
  feitengma0123@gmail.com
  https://sites.google.com/site/matf0123/
- **Muhan Zhang**
  Assistant Professor, Peking University
  muhan@pku.edu.cn
  https://muhanzhang.github.io/