

Yue Zhao

CONTACT INFORMATION Hamburg Hall 2005
4800 Forbes Ave Pittsburgh, Pennsylvania
United States, 15213 Carnegie Mellon University

zhaoy@cmu.edu
github.com/yzhao062
linkedin.com/in/yzhao062
andrew.cmu.edu/user/yuezhao2
CMU Expert Finder

Highlight: I am a seasoned ML software/system architect with 10 Python machine learning libraries (10,000 GitHub stars, and 300,000 total downloads). I am an EX PwC Consultant.

RESEARCH KEYWORDS

<input type="checkbox"/> Outlier & Anomaly Detection	<input type="checkbox"/> Machine Learning Systems
<input type="checkbox"/> Automated Machine Learning	<input type="checkbox"/> Information Systems
<input type="checkbox"/> Healthcare AI & Therapeutic for ML	<input type="checkbox"/> Scalable Machine Learning
<input type="checkbox"/> Ensemble Learning	<input type="checkbox"/> Human Machine Interaction

EDUCATION

Carnegie Mellon University Aug. 2019 - May. 2024 (expected)
Heinz College of Information Systems and Public Policy
Ph.D. in Machine Learning and Information Systems

- **Research Topics:**
 - **data mining** topics related to outlier detection (anomaly detection)
 - **machine learning systems (MLSys)** that can speed up and/or scale up data mining and machine learning algorithms
- **Mentors:** at CMU, I work with Prof. Leman Akoglu (Heinz) on anomaly detection, and Prof. Zhihao Jia (CSD) on machine learning systems (MLSys).

University of Toronto Sep. 2015 - Dec. 2016
Department of Computer Science
Master of Science in Computer Science (Applied Computing)

University of Cincinnati Sep. 2010 - May. 2015
College of Engineering and Applied Science
Bachelor of Science in Computer Engineering
GPA: *Magna Cum Laude*
Student Marshall

- **Minor:** *Computer Science and Mathematics*

Shanxi Experimental Secondary School Sep. 2007 - Jul. 2010
Experimental Class (Honor Class)

- **Concentration:** *Science*

PUBLICATIONS See my [Google Scholar](#), [DBLP](#), [ORCID](#), and [ResearchGate](#).

Preprints & Working Papers

1. Kexin Huang, Tianfan Fu, Wenhao Gao, [Yue Zhao](#), Yusuf Roohani, Jure Leskovec, Connor W. Coley, Cao Xiao, Jimeng Sun, Marinka Zitnik
Therapeutics Data Commons: Machine Learning Datasets and Tasks for Therapeutics
arXiv preprint arXiv:2102.09548
2. [Yue Zhao](#), Ryan A. Rossi, Leman Akoglu
Automating Outlier Detection via Meta-Learning
Submitted to a major CS conference, under review.
arXiv preprint arXiv:2009.10606

3. Yue Zhao, Zhi Qiao, Cao (Danica) Xiao, Lucas M. Glass, Jimeng Sun
PyHealth: A Python Library for Health Predictive Models
arXiv preprint arXiv:2101.04209

Peer-reviewed Journal Papers

1. Yue Zhao, Zain Nasrullah, Zheng Li
PyOD: A Python Toolbox for Scalable Outlier Detection
Journal of Machine Learning Research (JMLR), 2019.

Peer-reviewed Conference & Workshop Papers (with proceedings)

1. Yue Zhao*, Xiyang Hu*, Cheng Cheng, Cong Wang, Changlin Wan, Wen Wang, Jianing Yang, Haoping Bai, Zheng Li, Cao Xiao, Yunlong Wang, Zhi Qiao, Jimeng Sun, Leman Akoglu
SUOD: Accelerating Large-scale Unsupervised Heterogeneous Outlier Detection
Conference on Machine and Learning Systems (MLSys), 2021.
Acceptance rate 23.5% (52/221). (*equal contribution)
2. Kwei-Herng Lai*, Daochen Zha*, Guanchu Wang, Junjie Xu, Yue Zhao, Devsh Kumar, Yile Chen, Purav Zumkhwaka, Minyang Wan, Diego Martinez and Xia Ben Hu
TODS: An Automated Time Series Outlier Detection System (Demo paper)
Thirty-Fifth AAAI Conference on Artificial Intelligence (AAAI), 2021.
(*equal contribution)
3. Meng-Chieh Lee, Yue Zhao, Aluna Wang, Pierre Jinghong Liang, Leman Akoglu, Vincent S. Tseng, Christos Faloutsos
AutoAudit: Mining Accounting and Time-Evolving Graphs
IEEE International Conference on Big Data (Big Data), 2020
4. Changlin Wan, Dongya Jia, Yue Zhao, Wennan Chang, Sha Cao, Xiao Wang, and Chi Zhang
A Data Denoising Approach to Optimize Functional Clustering of Single Cell RNA-sequencing Data
2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2020
5. Zheng Li, Yue Zhao, Nicola Botta, Cezar Ionescu, Xiyang Hu
COPOD: Copula-Based Outlier Detection
IEEE International Conference on Data Mining (ICDM), 2020.
6. Zheng Li, Yue Zhao, Jialin Fu
SYNC: A Copula based Framework for Generating Synthetic Data from Aggregated Sources
IEEE International Conference on Data Mining Workshops (ICDMW), 2020.
7. Yiqun Mei, Yue Zhao, Wei Liang
DSR: An Accurate Single Image Super Resolution Approach for Various Degradations
IEEE International Conference on Multimedia and Expo (ICME), 2020, London, UK.
8. Yue Zhao, Xuejian Wang*, Cheng Cheng*, Xueying Ding*
Combining Machine Learning Models and Scores using combo Library (Demo paper)
Thirty-Fourth AAAI Conference on Artificial Intelligence (AAAI), 2020.
(*equal contribution)
9. Zain Nasrullah, Yue Zhao
Music Artist Classification with Convolutional Recurrent Neural Networks
IEEE International Joint Conference on Neural Networks (IJCNN), 2019, Hungary.
10. Yue Zhao, Zain Nasrullah, Maciej K. Hryniewicki, Zheng Li
LSCP: Locally Selective Combination in Parallel Outlier Ensembles
SIAM International Conference on Data Mining (SDM), 2019, Calgary, Canada.
Acceptance rate 22.7% (90/397).

11. Yue Zhao, Maciej K. Hryniewicki
XGBOD: Improving Supervised Outlier Detection with Unsupervised Representation Learning
IEEE International Joint Conference on Neural Networks (IJCNN), 2018, Rio, Brazil.
12. Yue Zhao, Maciej K. Hryniewicki, Francesca Cheng, Boyang Fu, Xiaoyu Zhu
Employee Turnover Prediction with Machine Learning: A Reliable Approach
Intelligent System Conference (Intellisys), 2018, London, UK.
Acceptance rate 34% (194/568).
13. Yue Zhao*, Zhongtian Qiu*, Yiqing Yang*, Weiwei Li*, Mingming Fan
An Empirical Study of Touch-based Authentication Methods on Smartwatches
ACM International Symposium on Wearable Computers (ISWC), 2017, Maui, USA.
Acceptance rate 25.6% (23/90). (*equal contribution)

Peer-reviewed Workshop Papers (without proceedings)

1. Yue Zhao, Xueying Ding, Jianing Yang, and Haoping Bai.
SUOD: Toward Scalable Unsupervised Outlier Detection
Workshops at the Thirty-Fourth AAAI Conference on Artificial Intelligence, 2020.
2. Colin Wan, Zheng Li, Alicia Guo, Yue Zhao
SynC: A Unified Framework for Generating Synthetic Population with Gaussian Copula
Workshops at the Thirty-Fourth AAAI Conference on Artificial Intelligence, 2020.
3. Yue Zhao, Maciej K. Hryniewicki
DCSO: Dynamic Combination of Detector Scores for Outlier Ensembles
ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD Workshop on Outlier Detection De-constructed), 2018, London, UK.

COMMUNITY ACTIVITIES & MEMBERSHIP

Program Committee

- AAAI 2021: The Thirty-Fifth AAAI Conference on Artificial Intelligence (AAAI-21)
- AI4AN 2020: The 1st Workshop on Artificial Intelligence for Anomalies and Novelties, co-located with IJCAI-PRICAI 2020

Journal/Conference Reviewer

- IEEE Transactions on Knowledge and Data Engineering (TKDE)
- IEEE Transactions on Neural Networks and Learning Systems (TNNLS)
- IEEE Intelligent Systems
- Data Mining and Knowledge Discovery (DMKD)
- ACM Transactions on Management Information Systems (TMIS)
- Knowledge and Information Systems (KAIS)
- Big Data
- IEEE Transactions on Systems, Man, and Cybernetics: Systems
- IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)
- IEEE Computational Intelligence Magazine (CIM)
- The Journal of Open Source Software (JOSS)
- AAAI 2021, KDD 2020, ICDM 2020, MICCAI 2020, 2021

Membership: *ACM, IEEE, SIAM, AAAI, SIAG/SDM*

PROFESSIONAL EXPERIENCE

IQVIA, Analytics Center of Excellence

Machine Learning Research Intern

May. 2020–Aug. 2020

- Designed new machine learning models in healthcare.
- Supervised by Dr. Cao (Danica) Xiao (IQVIA) and Prof. Jimeng Sun (UIUC).

PwC Canada, Consulting & Deals

Senior Consultant (Data Scientist)	Aug. 2017–Jun. 2019
Consultant (Data Scientist)	Feb. 2017–Jul. 2017
Research Associate (Intern)	May. 2016–Jan. 2017
<ul style="list-style-type: none">• Designed fraud analytic solutions for major Canadian banks and insurance firms.• Led various applied data mining projects, e.g., client segmentation and churn analysis.• Developed multiple pricing optimization models with statistical methods.	

Siemens PLM Software USA

Software Engineer (Intern & Contract)	Mar. 2012–Dec. 2014
<ul style="list-style-type: none">• Managed a Java project to transition the LabManager system to vCloud Director.• Refactored outdated automation code and added new modules and JUnit test cases.• Led a C++ Code Coverage project on Teamcenter platform to strengthen its stability.	

TEACHING EXPERIENCE

Carnegie Mellon University

Teaching Assistant (grading & lectures on AutoML and MLSys) S 2021
Intro to Artificial Intelligence (Prof. David Steier)

Teaching Assistant (grading & lecture on AutoML) F 2020
Intro to Artificial Intelligence (Prof. David Steier)

University of Toronto & University of Cincinnati

Teaching Assistant, Embedded Systems (Prof. Philip Anderson, Toronto) F 2015

Teaching Assistant, Intro to Programming (Prof. George Purdy, Cincinnati) F 2014

FUNDS AND AWARDS

CMU GSA/Provost Conference Funding		2019, 2020
AAAI Student Travel Grant		Dec. 2019
Mitacs-Accelerate Research and Development Funding	\$30,000	2016–2017
University Global Award and Scholarship	\$32,000	2010–2015
Mantei/Mae Award & Scholar (https://manteimaeawards.com/)	\$40,000	2012–2015
Engineer of the Month (University of Cincinnati)		Jun. 2014

RELEVANT SKILLS

Technical: Python (expert), C++ (advanced), Database (proficient)
Languages: English (fluent), Mandarin (native)

GRADUATE COURSEWORK

Courses at Carnegie Mellon University (all Ph.D. level courses)

<input type="checkbox"/> Machine Learning Systems (CSD)	<input type="checkbox"/> Ph.D. Seminar I, II, III (Heinz)
<input type="checkbox"/> Machine Learning for Large Datasets (MLD)	<input type="checkbox"/> Ph.D. Microeconomics (Heinz)
<input type="checkbox"/> Advanced Intro to Machine Learning (MLD)	<input type="checkbox"/> Econometric Theory and Methods (Heinz)
<input type="checkbox"/> Probability and Statistics (Stats)	<input type="checkbox"/> Econometrics I (Tepper)
<input type="checkbox"/> Intermediate Statistic (Stats)	<input type="checkbox"/> Human and Algorithmic Bias (Tepper)
<input type="checkbox"/> Multimodal Machine Learning (LTI)	<input type="checkbox"/> Human Judgment and Decision Making (SDS)

Courses at University of Toronto (all Ph.D. or Master level courses)

<input type="checkbox"/> Machine Learning and Data Mining	<input type="checkbox"/> Human Computer Interaction
<input type="checkbox"/> Systems Thinking for Global Problems	<input type="checkbox"/> Technical Entrepreneurship
<input type="checkbox"/> Communications for Computer Scientists	<input type="checkbox"/> Big Data Analytics in Healthcare