Towards Reproducible, Automated, and Scalable Anomaly Detection

Yue Zhao

Thomas Lord Department of Computer Science
University of Southern California
yzhao010@usc.edu

Anomaly detection (AD), often termed outlier detection, is a key machine learning (ML) task, aiming to identify uncommon yet crucial patterns in data. With the increasing complexity of the modern world, the applications of AD span wide—from NASA's spacecraft monitoring to early patient prioritization at the University of Pittsburgh Medical Center. Tech giants like Google and Amazon also leverage AD for prompt service disruption identification. Here, I will traverse my AD works with promising new directions, particularly emphasizing *reproducible benchmarks* (Part 1), *automated algorithms* (Part 2), and *scalable systems* (Part 3).

Part 1: Reproducible Benchmarks. The efficacy and characteristics of AD algorithms remain understudied due to adhoc evaluations on subsets of public datasets, resulting in questionable advancements and conclusions. In my recent works, I conducted extensive AD analyses on various data types—tabular (ADBench (Han et al. 2022)), time-series (TSOD (Lai et al. 2021)), graph data (BOND (Liu et al. 2022)), and deep AD (Jiang et al. 2023). For instance, in ADBench, we performed a comprehensive benchmark on 57 datasets using 30 detection algorithms, encompassing various settings, and totaling 98,436 experiments. Our rigorous statistical analysis revealed that *no unsupervised AD method stood out statistically*. These extensive benchmarks offer critical insights into AD, guiding the development of optimized algorithms suitable for real-world applications. Thus, I will briefly introduce two of my recent AD benchmarks, specifically addressing tabular and graph datasets: ADBench (Han et al. 2022) and BOND (Liu et al. 2022). They identify two pivotal challenges/needs in current AD research: (1) automatic AD model selection for underlying applications/-tasks (which we will dive into in Part 2), and (2) fast and scalable AD systems (our focus in Part 3).

Part II: Automated Algorithms. Despite the large number of existing AD algorithms, the pressing question arises: *how do we automatically determine the optimal detector for various applications*? Given the absence of ground truth in unsupervised AD, this challenge is important, where random model selection may yield inadequate accuracy for tasks like fraud detection and network intrusion identification. In this section, I will introduce our progress in automating AD model selection and composing deep AD models. Our spotlight will be on MetaOD (Zhao, Rossi, and Akoglu 2021) and ADGym (Jiang et al. 2023)—marking both the advent and the latest innovations in this trajectory. The key of our discussion will center around meta-learning– leveraging insights from similar historical datasets for fast model selection and more automated ML tasks.

Part III: Scalable Detection Systems. From a system perspective, applying AD algorithms in real-world scenarios with large, high-dimensional data remains challenging. I have designed scalable ML systems (MLSys) for various AD types to mitigate this, integrating distributed learning, approximation, and more. I will first cover Python Outlier Detection (PyOD) (Zhao, Nasrullah, and Li 2019). It stands out for its comprehensive nature, housing 50+ methods, and unparalleled efficiency, with advanced techniques such as JIT optimization. Our main focus, however, will be on introducing TOD (Zhao, Chen, and Jia 2022), the pioneering GPU system accommodating a broad spectrum of AD algorithms. TOD can abstract complex AD algorithms into simple tensor operations for multi-GPU computation, offering an $11\times$ speed improvement to PyOD.

References

Han, S.; Hu, X.; Huang, H.; Jiang, M.; and Zhao, Y. 2022. ADBench: Anomaly detection benchmark. *NeurIPS*, 35: 32142–32159.

Jiang, M.; Hou, C.; Zheng, A.; Han, S.; Huang, H.; Wen, Q.; Hu, X.; and Zhao, Y. 2023. ADGym: Design Choices for Deep Anomaly Detection. *NeurIPS*.

Lai, K.-H.; Zha, D.; Xu, J.; Zhao, Y.; Wang, G.; and Hu, X. 2021. Revisiting time series outlier detection: Definitions and benchmarks. In *NeurIPS*.

Liu, K.; Dou, Y.; Zhao, Y.; Ding, X.; Hu, X.; Zhang, R.; Ding, K.; Chen, C.; Peng, H.; Shu, K.; et al. 2022. BOND: Benchmarking unsupervised outlier node detection on static attributed graphs. *NeurIPS*, 35: 27021–27035.

Zhao, Y.; Chen, G. H.; and Jia, Z. 2022. TOD: GPU-Accelerated Outlier Detection via Tensor Operations. *VLDB*, 16(3): 546–560.

Zhao, Y.; Nasrullah, Z.; and Li, Z. 2019. PyOD: A Python Toolbox for Scalable Outlier Detection. *JMLR*, 20: 1–7.

Zhao, Y.; Rossi, R.; and Akoglu, L. 2021. Automatic unsupervised outlier model selection. *NeurIPS*, 34: 4489–4502.