Anomaly Mining: Past, Present and Future

Leman Akoglu
Heinz College of Information Systems and Public Policy
Carnegie Mellon University
Pittsburgh, PA 15213-3890
lakoglu@andrew.cmu.edu

ABSTRACT

Anomaly mining finds high-stakes applications in various real-world domains such as cybersecurity, finance, environmental monitoring, to name a few. Therefore, it has been studied widely and a large body of detection techniques exists [1]. Today, many real-world settings necessitate detection at speed for streaming/evolving data, and/or detection at scale for massive datasets stored in a distributed environment [2].

Despite the plethora of detection algorithms, selecting an algorithm to use on a new task as well as setting the values for its hyperparameter(s), known as the model selection problem, is an open challenge for unsupervised anomaly detection. This issue is only to be exacerbated with the recent advent of detectors based on deep neural networks that exhibit a long list of hyperparameters. The challenge stems from two main factors: the lack of labeled data and the lack of a widely accepted anomaly loss function. Toward automation, one can explore internal evaluation strategies [3], or capitalize on the experience from historical detection tasks through meta-learning [4]. However, the problem remains far from solved.

In deployment, many real-world use cases of anomaly detection require the flagged anomalies from a detector to be screened or audited by a human expert, typically for vetting purposes, where taking automatic actions can be costly (e.g. directly charging a flagged medical provider with fraud). While a vast majority of the literature focuses on novel detection algorithms, as humans are often involved with(in) the process, anomaly mining also concerns various human-centric problems that are beyond mere detection, namely explanation [5, 6], human interaction [7], and fairness [8]. These aspects of the field are under-studied and pose many open challenges.

CCS Concepts/ACM Classifiers

• Computing methodologies~Machine learning~Learning paradigms~Unsupervised learning~Anomaly detection

CIKM '21, November 1–5, 2021, Virtual Event, Australia.

ACM ISBN 978-1-4503-8446-9/21/11.
https://doi.org/10.1145/3459637.3482495

Author Keywords

anomaly mining; anomaly detection for data streams and distributed data; unsupervised model selection; human-centric anomaly detection

BIOGRAPHY

Leman Akoglu is the Heinz College Dean's Associate Professor. Her research broadly spans machine learning and data mining, and focuses on graph mining, pattern discovery and anomaly detection, with applications to fraud and event detection in many real-world domains. Dr. Akoglu is a recipient of the SDM/IBM Early Career Data Mining Research Award (2020), National Science Foundation CAREER Award (2015) and the US Army Research Office Young Investigator Award (2013). Her research has won several publication awards, most recently The Most Influential Paper (PAKDD 2020), Best Research Paper (SIAM SDM 2019), and Best Student Machine Learning Paper Runner-up (ECML PKDD 2018). Her research has been supported by the NSF, US ARO, DARPA, Adobe, Facebook, Northrop Grumman, PNC Bank, and PwC.

ACKNOWLEDGEMENTS

This talk is based on work funded in part by the National Science Foundation CAREER 1452425. We also thank PwC Risk and Regulatory Services Innovation Center at Carnegie Mellon University. Any conclusions expressed in this talk are those of the author and do not necessarily reflect the views, expressed or implied, of the funding parties.

REFERENCES

[1] C. C. Aggarwal "Outlier Analysis, Second Edition.", Springer. 2016.

[2] E. Manzoor, H. Lamba, L. Akoglu. "xStream: Outlier detection in feature-evolving data streams." 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2018.

[3] M. Q. Ma, Y. Zhao, X. Zhang, L. Akoglu. "A Large-scale Study on Unsupervised Outlier Model Selection: Do Internal Strategies Suffice?" arXiv preprint arXiv:2104.01422. 2021.

[4] Y. Zhao, R. A. Rossi, L. Akoglu. "Automating outlier detection via meta-learning." arXiv preprint arXiv:2009.10606. 2020.

[5] M. Macha, L. Akoglu. "Explaining anomalies in groups with characterizing subspace rules." Data Mining and Knowledge Discovery 32, No. 5, pp. 1444-1480. 2018.

[6] N. Gupta, D. Eswaran, N. Shah, L. Akoglu, C. Faloutsos. "Beyond outlier detection: Lookout for pictorial explanation." Joint European Conference on Machine Learning and Knowledge Discovery in Databases. 2018.

[7] H. Lamba, L. Akoglu. "Learning on-the-job to re-rank anomalies from top-1 feedback." SIAM International Conference on Data Mining. Society for Industrial and Applied Mathematics. 2019.

[8] S. Shekhar, N. Shah, L. Akoglu. "FairOD: Fairness-aware outlier detection." AAAI/ACM Conference on AI, Ethics, and Society. 2021.