applied sciences

MDPI

Editorial

Special Issue on Unsupervised Anomaly Detection

Markus Goldstein

Department of Computer Science, Ulm University of Applied Sciences, 89075 Ulm, Germany; markus.goldstein@thu.de

1. Introduction to Anomaly Detection

Anomaly detection (also known as outlier detection) is the task of finding instances in a dataset which deviate markedly from the norm. Anomalies are often of specific interest in many real-world analytic tasks, since they can refer to incidents requiring special attention. The detection of suspicious activity can be helpful in both, post-incident investigation, or, in early-warning setups, where anomalies are detected in recent up-to-date datasets [1] or even in streaming data [2].

Among others, intrusion detection [2–5], payment fraud detection, public safety, complex system monitoring [6–10], and medical data analytics are possible application domains.

In the context of *supervised anomaly detection*, a labeled dataset and an established classification algorithm, which can deal well with unbalanced classes, can be used. Since anomalies are often not similar to each other and also often unknown during training, this setup is barely used in practice. In *semi-supervised anomaly detection*, a model is learned with the normal class only. Later, anomalies can be detected using deviations from that model. This setup is also known as one-class machine learning or novelty detection. Lastly, in *unsupervised anomaly detection*, no training is performed at all—the data are solely analyzed according to the intrinsic structure and anomalies are often scored according to their degree of outlierliness. This special issue addresses primarily these algorithms, whereas many of them can also be used in a semi-supervised setup [4].

From an application point of view, an anomaly can be a single record within a multivariate dataset, which is also known as a *point anomaly detection* problem. If the context time needs to be taken into consideration to detect the outliers, the task is also known as *contextual anomaly detection*. Lastly, a *collective anomaly* is a scenario where multiple instances can form altogether an anomaly. A collective anomaly is the most complex scenario and can also be at the same time a contextual anomaly detection problem. Further details about this taxonomy can also be found in Al-amri et al. [5]. To solve a contextual anomaly detection task, the data can be transformed into a point anomaly detection problem [6,9] or a (multivariate) time series anomaly detection algorithm can be applied [3,4,7,10–12].

2. Contributions

Cheng et al. [6] compared different unsupervised point anomaly detection algorithms for detecting outliers among process or product quality profiles. In this context, a profile is a nonlinear relationship between input variables and an output variable, mapping a collective or contextual anomaly detection task to a point anomaly detection problem.

A stochastic Petri net digital twin was used by Lian et al. [7] to detect complex collective outliers for oil and gas station operation behavior based on a multivariate time series anomaly detection using a GAN. According to the authors, the model could also be used to explain anomalies utilizing the reconstructed information.

A comparison of traditional and deep learning unsupervised algorithms for time series anomaly detection was carried out by Rewicki et al. [11], also focusing on the different types of anomalies. Interestingly, the classical machine learning approaches generally outperform the deep learning based algorithms.

Citation: Goldstein, M. Special Issue on Unsupervised Anomaly Detection. *Appl. Sci.* **2023**, *13*, 5916. https://doi.org/10.3390/app13105916

Received: 2 May 2023
Accepted: 5 May 2023
Published: 11 May 2023

A very compelling application is the monitoring of flight parameter data in the article by Jasra et al. [9]. The authors detected anomalous flights after preprocessing using LOF with an automatically determined threshold among recorded and simulated flight data. It is worth noting that the method could also potentially be used in a near-realtime setting during flights.

A new approach for detecting anomalies in multivariate time series data was proposed by Pham et al. [12] using Multi-Scale Temporal convolutional kernels with a Variational AutoEncoder (MST-VAE). As a result, short-scale and long-scale convolutional kernels should be combined to improve the overall model performance.

Furthermore, a new interesting research direction is addressed by Rollón de Pinedo et al. [8]: Functional outlier detection, where anomalies in terms of magnitude and shape are detected based on using h-mode depth and dynamic time warping. The authors also investigate a not very commonly used but interesting application scenario: The detection of anomalies within data originating from costly simulations.

The analysis of monitored KPI data in distributed systems to detect abnormal system states is carried out by Shang et al. [3]. Here, a correlation analysis of multivariate time series is performed using a Hidden Markov Model.

Jiang et al. [4] proposed a new semi-supervised anomaly detection framework for univariate time series entitled Tri-CAD. It categorizes time series into three types and uses different models for each. The different models include statistics, wavelet transforms as well as a deep autoencoder.

Furthermore, an improvement of the well-known ARIMA model to detect anomalies in univariate time series more efficiently and continuously has been proposed by Kozitsin et al. [10]. The focus of the new algorithm includes better performance compared to the original ARIMA as well as the ability to adopt the model over time to changing underlying data distributions.

Anomaly detection utilizing multiple parallel data streams in the form of multiple stochastic processes has been addressed by Qin et al. [2], introducing a low-cost deterministic policy for detecting anomalous processes. The authors point out that their proposed algorithm is an ideal candidate for the challenge of anomaly detection during DOS attacks in intrusion detection systems.

Herskind Sejr et al. [1] developed an application for detecting anomalies in music streaming behavior data, whereby an explanation of the outliers was included such that deviations from the expected time series forecast are more informative. Interestingly, anomalies are presented to the users, e.g., studios or musicians in this application scenario, not only to system administrators as is commonly the case.

The challenges of anomaly detection among IoT data and possible future directions are discussed in the article by Al-amri et al. [5] in a comprehensive review. Besides the problems of common unsupervised anomaly detection tasks, the authors specifically identify feature-evolving data streams as a core point for IoT anomaly detection tasks in the future.

3. Conclusions

In this Special Issue titled "Unsupervised Anomaly Detection" of *Applied Sciences*, a total of 12 papers (11 research articles and one review paper) are published and show recent advances in the area. Besides many contributions addressing particular practical anomaly detection tasks in real-world challenges, some papers also improve well-known unsupervised algorithms. Additionally, it is worth pointing out two articles focusing on current trending research directions: Explainable Anomaly Detection and Functional Anomaly Detection.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: Thanks to all authors of this Special Issue on *Unsupervised Anomaly Detection* and their valuable contributions. Special thanks to all peer reviewers for their relevant and helpful reviews, which improved the quality of the articles significantly. Finally, I would like to extend thanks to the staff and people involved at MDPI.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Herskind Sejr, J.; Christiansen, T.; Dvinge, N.; Hougesen, D.; Schneider-Kamp, P.; Zimek, A. Outlier Detection with Explanations on Music Streaming Data: A Case Study with Danmark Music Group Ltd. *Appl. Sci.* **2021**, *11*, 2270. [CrossRef]
2. Qin, F.; Feng, H.; Yang, T.; Hu, B. Low-Cost Active Anomaly Detection with Switching Latency. *Appl. Sci.* **2021**, *11*, 2976. [CrossRef]
3. Shang, Z.; Zhang, Y.; Zhang, X.; Zhao, Y.; Cao, Z.; Wang, X. Time Series Anomaly Detection for KPIs Based on Correlation Analysis and HMM. *Appl. Sci.* **2021**, *11*, 1353. [CrossRef]
4. Jiang, J.R.; Kao, J.B.; Li, Y.L. Semi-Supervised Time Series Anomaly Detection Based on Statistics and Deep Learning. *Appl. Sci.* **2021**, *11*, 6698. [CrossRef]
5. Al-amri, R.; Murugesan, R.K.; Man, M.; Abdulateef, A.F.; Al-Sharafi, M.A.; Alkahtani, A.A. A Review of Machine Learning and Deep Learning Techniques for Anomaly Detection in IoT Data. *Appl. Sci.* **2021**, *11*, 5320. [CrossRef]
6. Cheng, C.S.; Chen, P.W.; Wu, Y.T. Phase I Analysis of Nonlinear Profiles Using Anomaly Detection Techniques. *Appl. Sci.* **2023**, *13*, 2147. [CrossRef]
7. Lian, Y.; Geng, Y.; Tian, T. Anomaly Detection Method for Multivariate Time Series Data of Oil and Gas Stations Based on Digital Twin and MTAD-GAN. *Appl. Sci.* **2023**, *13*, 1891. [CrossRef]
8. Rollón de Pinedo, Á.; Couplet, M.; Iooss, B.; Marie, N.; Marrel, A.; Merle, E.; Sueur, R. Functional Outlier Detection by Means of h-Mode Depth and Dynamic Time Warping. *Appl. Sci.* **2021**, *11*, 1475. [CrossRef]
9. Jasra, S.K.; Valentino, G.; Muscat, A.; Camilleri, R. Hybrid Machine Learning—Statistical Method for Anomaly Detection in Flight Data. *Appl. Sci.* **2022**, *12*, 261. [CrossRef]
10. Kozitsin, V.; Katser, I.; Lakontsev, D. Online Forecasting and Anomaly Detection Based on the ARIMA Model. *Appl. Sci.* **2021**, *11*, 3194. [CrossRef]
11. Rewicki, F.; Denzler, J.; Niebling, J. Is It Worth It? Comparing Six Deep and Classical Methods for Unsupervised Anomaly Detection in Time Series. *Appl. Sci.* **2023**, *13*, 1778. [CrossRef]
12. Pham, T.A.; Lee, J.H.; Park, C.S. MST-VAE: Multi-Scale Temporal Variational Autoencoder for Anomaly Detection in Multivariate Time Series. *Appl. Sci.* **2022**, *12*, 78. [CrossRef]