List of Changes and Answers to Reviewers

Adaptable and Interpretable Framework for Anomaly Detection in SCADA-based Industrial Systems
M. Wadinger, M. Kvasnica

November 23, 2023

1 List of Changes

List of main changes:

1. The manuscript was shortened to 7 pages.
2. The leading paragraph of Section 4 was reworded to address reviewers' comments on SVMs.
3. Section 4.1 was removed, linear separation is now tackled as a special case in the *Polynomial Separation* section (labeled as Section 4.1 in the revision).
4. Big-O notation in Section 5 was replaced by hard figures indicating required computation and consumed memory.
5. Section 6.2 was enriched to address reviewers' comments.
6. Conclusions was modified according to reviewers' requests.
7. Figure 2 was added to illustrate the definition of anomalies.
8. Captions in Figures 5,6,7 (labeled as Figures 4,5,6 in the previous manuscript) were reworded to make them more clear.
9. Table 2 was transposed to allow for new metrics to be added.
10. False Alarm Rate, AUC, and Mean of Rolling AUC were added to the results in Table 2.
11. Added reference to [DSXS21] in the Introduction.

2 Answers to Reviewers and to the Associate Editor

We would like to thank all reviewers and to the associate editor for encouraging comments and hints. We have tried to address all of them appropriately.

Associate Editor

Reviewers have now commented on your paper. You will see that they are advising that you revise your manuscript. If you are prepared to undertake the work required, I would be pleased to reconsider my decision.

For your guidance, reviewers' comments are appended below.

If you decide to revise the work, please submit a list of changes or a rebuttal against each point raised by the reviewers. You can upload this as the 'Detailed Response to Reviewers' when you submit the revised manuscript.

Response: We would like to thank the associate editor for his/her evaluation. We believe that the modifications, described in more detail below, address all issues pointed out by the reviewers.

Reviewer 1

This paper proposes a new online anomlay detection method and verifies its effectiveness on real-world datasets. However, there are some limitations as follows:

1. *There is no clear definitions of point anomaly, collective anomalies and concept changes. Figure 2 does not illustrate their differences either.*

   Response: While we tried to establish the notation of anomalies and their categorization used throughout our paper in Section 2.6, the visualization would be more illuminating.

   Indeed, the work would benefit from more clarity in the definition of point anomaly, collective anomalies, and concept changes.

   Therefore, we have added Figure 2 to illustrate the definition of anomalies.

2. *The captions of Figures 4,5,6 are similar but the labels are different. It is a bit confused as which one is the ground truth.*

   Response: Figures 4,5,6 were indeed carrying identical captions. We realize that our effort to describe elements of the figure in the caption compromised the delivery of the key messages they represented.

   Thanks to the reviewer's comment, we have reworded the captions to make them more clear. Please note that due to the addition of Figure 2, Figures 4,5,6 in question are now labeled as Figures 5, 6, and 7 in the revision.

   However, the ground truth information remains undisclosed in the Figures. This is due to the fact that operators did not provide us with information about the exact time of abnormal events, which has shown ambiguity.

Therefore, in Section 4.1, we refer to the dates of the events. Moreover, picking the time of the anomaly based on observation for plotting purposes would be arbitrary, which would compromise the objectivity of the results.

3. *There are other self-supervised change-point detection method, such as [DSXS21].*

   Response: Our aim was to provide a comprehensive review of the state-of-the-art methods for self-supervised adaptive detection methods with interpretability. Though making attempt to cover the most relevant self-supervised change-point detection methods, the review was not exhaustive in this matter.

   We would like to thank the reviewer for pointing out this reference, which, after examination, proved to be relevant in the paragraph concerning the need for early change point detection. We have added it to the Introduction section of the thesis.

4. *It seems that using ARIMA or moving average can easily detect the anomalies or change points on the real-world datasets.*

   Response: While we agree that the impression from the figures may be that the anomalies are easy to detect by ARIMA or moving average, we would like to point out that our proposed method considers interactions between the variables while providing diagnostic capabilities.

   As we broadened the scope of our research on the current state of research, we found that Vector Autoregression, the multivariate extension of ARIMA, is a promising choice for anomaly detection in multivariate time series, capturing interactions. We found at least two papers dealing with offline trained anomaly detection methods based on vector autoregression [MBMO16, ZZQ23]. Our paper, on the other hand, deals with online anomaly detection. Therefore, we believe that our method provides a unique combination of features relevant to real-world scenarios. Nevertheless, we are aware of the existence of online trained ARIMA methods, which gives a promise of extension to its multivariate counterpart.

   Fortunately, ARIMA could be effectively combined with our method during feature engineering, which could further enhance the performance of the proposed method, similar to the utilization of the physics-based model in Section 4.1.

5. *It would be better to use AUC rather than F1 to measure the anomaly detection performance. Moreover, range-based AUC is even better and more fair for streaming or sliding window-based method.*

   Response: We agree with the reviewer that AUC, in general, is a better metric for imbalanced datasets. However, due to the poor convergence of the reference methods on benchmark data during hyperparameter optimization, we decided to use F1 score, which showed better convergence for all three compared methods. Nevertheless, we have added AUC to the results in Table 2. as it might be of interest to the reader.

To our surprise, we were not aware of the range-based AUC metric during the time of paper writing and result collection. We have added it to the results in Table 1. using implementation from [BS17].

Moreover, we tried to use the range-based AUC metric for hyperparameter optimization. Due to little improvement in convergence, we decided to keep the F1 score as the main metric and not include the results obtained using the range-based AUC metric in the cost function of hyperparameter optimization in the paper. As proof, the results are provided in this response in Table 1.

Table 1: Evaluation of models optimized for Rolling AUC score on SKAB dataset. The best-performing model is highlighted in bold. Values in brackets represent macro values of the metric.

Algorithm	AID	HS-Trees	OC-SVM
Precision [%]	**47** (60)	30 (47)	32 (48)
Recall [%]	**55** (61)	4 (50)	3 (50)
F1 [%]	**51** (60)	7 (42)	6 (42)
AUC [%]	**61**	50	50
Mean Rolling AUC [%]	**60**	50	49
FPR [%]	**38**	37	37
Avg. Latency [ms]	1.45	**0.05**	**0.05**

Reviewer 2

This paper presents an interesting and potentially useful framework called AID for anomaly detection and root cause diagnosis in industrial internet-of-things (IoT) systems. It incorporates dynamic conditional probability distribution modeling to adapt to non-stationary data streams, which is crucial for industrial systems. And industrial case studies demonstrate capabilities on real systems. However, i still have following concerns.

1. *More analysis of computational complexity and scalability limitations for high-dimensional industrial systems would strengthen the work.*

   Response:

2. *While the paper mentions comparisons with other methods, it lacks detailed benchmarking data, such as false positive rates.*

   Response: We added a False Positive Rate to the results in Table 2. Please note that thanks to the point of the other reviewer, we also added AUC and Mean of Rolling AUC, which are both relevant metrics for imbalanced datasets and might illuminate the performance of the proposed method.

3. *Comparing diagnosis accuracy for root causes against other interpretable methods could better highlight capabilities.*

   Response:

References

[BS17] Dariusz Brzezinski and Jerzy Stefanowski. Prequential auc: properties of the area under the roc curve for data streams with concept drift. *Knowledge and Information Systems*, 52(2):531–562, Aug 2017.

[DSXS21] Shohreh Deldari, Daniel V. Smith, Hao Xue, and Flora D. Salim. Time series change point detection with self-supervised contrastive predictive coding. In *Proceedings of the Web Conference 2021*, WWW '21, pages 3124–3135, New York, NY, USA, 2021. Association for Computing Machinery.

[MBMO16] Igor Melnyk, Arindam Banerjee, Bryan Matthews, and Nikunj Oza. Semi-markov switching vector autoregressive model-based anomaly detection in aviation systems. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, page 1065–1074, New York, NY, USA, 2016. Association for Computing Machinery.

[ZZQ23] Ruiyao Zhang, Ping Zhou, and Junfei Qiao. Anomaly detection of nonstationary long-memory processes based on fractional cointegration vector autoregression. *IEEE Transactions on Reliability*, pages 1–12, 2023.