Anomaly Detection on Experimental Chemical Process Data

Justus Arweiler[1], Aparna Muraleedharan[2], Fabian Hartung[3,4], Indra Jungjohann[1], Dennis Wagner[3], Marius Kloft[3], Jakob Burger[2], Hans Hasse[1], and Fabian Jirasek[1]

[1] Laboratory of Engineering Thermodynamics, RPTU Kaiserslautern
[2] Chemical Process Engineering, Technical University of Munich
[3] Machine Learning Group, RPTU Kaiserslautern
[4] Gas Treatment Technology, BASF SE Ludwigshafen

Abstract. The reliable detection of faults and anomalies in chemical processes is vital for safe operation and high product quality in chemical industries. To date, human experts must evaluate process data to fulfill this task, so automatizing is highly desirable. Over the past decade, many anomaly detection (AD) methods based on deep machine learning have appeared. These methods are usually developed, assessed, and compared using artificial data from the Tennessee-Eastman Process (TEP). This work, for the first time, presents findings on deep learning-based AD methods applied to experimental process data. The results revealed an excellent performance of the AD methods on the synthetic TEP data while struggling with experimental data, particularly for dynamic processes. These findings underscore the need for new, tailored approaches for AD detection to discriminate abnormal from normal process behavior.

Keywords: Machine Learning · Anomaly Detection · Distillation · Data Generation.

1 Introduction

Anomalies in chemical plants can have disastrous consequences for humans and the environment. Therefore, reliable methods for detecting faults and anomalies in the process as early as possible are paramount in chemical industries to ensure safety and product quality. While anomaly detection (AD) in chemical processes to date mainly relies on human experts, the machine-learning (ML) community has developed powerful AD methods for time-series data. Advancements in time-series AD have prominently featured the evolution of unsupervised deep learning methods that operate independently without needing labeled anomaly instances [14]. These unsupervised deep learning models are trained on patterns representing standard behavior within time-series data, allowing them to internalize fundamental structures and representations of regular sequences. Consequently, anomalies reflecting deviations from the learned norms are effectively detected within the data.

However, although these deep AD methods are generally suited for the time-series data from chemical processes, e.g., the sensor data continuously collected in a chemical plant, they have yet to find their way into chemical industries. The main obstacle, thereby, is the lack of available experimental process data for training the AD methods since chemical industries, where most of the processes are run, do not share their data. Specifically, the only data set for training and testing AD methods for chemical processes is the so-called Tennessee Eastman Process (TEP) data set [2], which has, consequently, been the benchmark dataset for time-series AD, including previous works on AD in chemical processes [9]. However, the TEP is not a real but hypothetical process. Hence, the TEP data set contains no experimental but only simulated data that does not cover the full spectrum of possible operating scenarios or abnormalities encountered in real-world industrial settings. Consequently, the results of AD methods on the TEP data do not allow conclusions on their applicability to actual chemical processes.

In the present work, we, for the first time, introduce deep AD methods in real chemical processes, focusing on distillation processes as the prime example. For this purpose, we generate suitable, large data sets with a continuous distillation plant, as it is part of many large-scale industrial production processes, and with a batch distillation plant, which is more flexible and, therefore, established for small-scale production ventures. We performed anomaly-free experiments and experiments with various enforced anomalies with both plants. State-of-the-art AD methods from the ML literature were systematically trained and tested on the experimental data, and the results were compared to the synthetic TEP data set [13].

2 Methods

2.1 Experimental Data Generation

A continuous distillation mini-plant with a capacity of five t/y was utilized to generate stationary process data. Multivariate time-series data from 17 sensors were collected. The distillation column has one pressure controller, two flow controllers, two level controllers, and two temperature controllers. Experiments lasting 30 days were carried out with pure water.

A laboratory-sized plant with a two-litre glass still and a glass column with three sections was used for the batch distillation experiments. Multivariate time-series data from 20 sensors were collected. The batch distillation column has a pressure controller, six temperature controllers, and one level controller. The collected dynamic process data span 25 distillation experiments with ternary mixtures of n-butanol, 2-propanol, and water as feeds.

Experimental data with and without anomalies were generated with both plants; all data were labeled accordingly. The anomalies were introduced in a clearly defined manner. The anomaly's location was defined as either a control loop, sensor, or component of the plant. For a short time window, a fault, e.g., leakage or sealing failure, or a control setting deviating from normal behavior

in the respective experiment was introduced into the process. Literature on the most common process anomalies from real-life processes was reviewed to support the optimal selection of recreated anomalies [7]. The faults included in the dataset are listed in 2.2. The labeled data were used to train, test, and systematically compare different AD methods from the literature.

2.2 Anomaly Detection

Using the synthetic TEP data, 27 unsupervised deep AD methods, taken from contemporary literature for time-series AD, were tested and compared [5]. The tested methods include reconstruction-based methods [6], forecasting-based methods [10], generative methods [16], and hybrid approaches [15,18]. Seven of the AD methods were randomly chosen and also applied and tested using the experimental process data from the continuous and batch plant generated in this work. All methods were trained equally on a training data set of five runs with about 1000 time steps each for the continuous and batch distillation data. The periods of the plant's start-up and cool-down phases were not considered, respectively. All training data sets were normalized and anomaly-free. The hyperparameters were not fixed but optimized with cross-validation to adjust them to this new data set for a better comparison. For this cross-validation, we split the training set into folds of equal size. The training excluded one of these folds, on which we validated the trained model. This process was repeated by switching the folds to have each fold used once for validation. Afterward, all optimal trained models were evaluated on a test data set containing ten other runs than the training set with 1,000-11,000 time steps each. These time-series data contain anomalies of varying type and effect duration. As types of anomalies, faults in the supplied heat, internal pressure, or leaking were considered, which can also be found in the TEP dataset [13]. The types of anomalies are thereby distributed uniformly in the test cases.

In all cases, we use the overall F1-score and the area under the precision-recall curve (AUPRC) as anomaly scores [11] for each time step for evaluating and comparing the AD methods, with an anomaly being detected if the score surpasses a threshold, that is set as a hyperparameter for the AD methods. These scores were used to rank the AD methods according to their average performance in all three settings.

3 Results and Discussion

All studied 27 AD methods perform very well on the synthetic TEP data set with F1-scores above 0.9 and AUPRC above 0.96 throughout, cf. our paper [5] for details.

Table 1 summarizes the results of seven of these AD methods on the experimental process data generated in this work and compares them to those for the TEP data. The ranking of the AD methods on the different data sets is based on the AUPRC score. Among the experimental data, the performance of the

Method	Method Type	AUPRC			Ranking		
		TEP	continuous	batch	TEP	continuous	batch
DONUT[17]	Generative VAE based	0.9829	0.4111	0.5761	1	4	2
GMM_VAE[4]	Generative VAE based	0.9815	0.3353	0.5164	2	6	4
USAD[1]	Reconstruction based	0.9779	0.3583	0.5974	3	5	1
LSTM_VAE_GAN[12]	Generative VAE based	0.9735	0.6445	0.4900	4	3	5
TADGAN[3]	Generative GAN based	0.9690	0.6723	0.3545	5	1	7
MADGAN[8]	Generative GAN based	0.9621	0.3296	0.5282	6	7	3
THOC[15]	Hybrid	0.9618	0.6702	0.4330	7	2	6

Table 1. Results of seven AD methods tested on synthetic TEP data and experimental process data using a continuous and a batch distillation plant.

AD methods on the dynamic data is lower than for the stationary data. This finding can be explained by the nature of dynamic data, where even the normal operating point is object to transitions at all times. Hence, detecting faults in the dynamic data is a much more challenging task than in stationary data from a continuous plant.

By comparing the results for the experimental data with those for the TEP data, it is particularly noticeable that the AUPRC scores are significantly lower for the experimental data, indicating much poorer performance regarding correctly identifying anomalies from actual process data. This result is observed for the stationary data from the continuous distillation plant and the dynamic data from the batch distillation plant. Noteworthy, even the least effective method for the TEP data set, which is THOC [15], exhibits a thirty percent higher AUPRC score than the best-performing methods on the experimental data sets. Possible explanations for the significant differences in performance with the different data sets are that the available experimental data sets for training the AD methods are too small and that the nature of the experimentally introduced anomalies is more complex than that of the synthetic anomalies. Furthermore, AD methods have so far only been developed and tested on synthetic data, so they are likely tailored for synthetic scenarios rather than real-world applications.

4 Conclusion

Deep learning-based AD techniques perform excellently on synthetic TEP data but largely fail to detect abnormal behavior in experimental process data sets correctly. We conclude that developing, testing, and comparing the performance of AD methods on synthetic process data does not lead to suitable methods for AD in real process plants. The outcomes of our research underscore the pressing need for the development of new, tailored approaches for AD in real-world applications within the chemical industries. To achieve this, more experimental process data, including normal operation and data sets with anomalies, are required to develop suitable AD methods. These data sets must include dedicated experiments with anomalies not easily detectable by a human expert to maxi-

mize the impact of deep AD methods. The dataset used in this work will also be made available in future publications to support this.

Acknowledgments. The authors gratefully acknowledge financial support by Deutsche Forschungsgemeinschaft (DFG) in the frame of the DFG research unit FOR 5359 "Deep Learning on Sparse Chemical Process Data".

References

1. Audibert, J.: Usad: Unsupervised anomaly detection on multivariate time series. In: Proc. of the 26th Int. Conf. on Knowledge Discovery and Data Mining. p. 3395–3404. https://doi.org/10.1145/3394486.3403392
2. Downs, J., Vogel, E.: A plant-wide industrial process control problem. Computers and Chemical Engineering **17**(3), 245–255,. https://doi.org/10.1016/0098-1354(93)80018-I
3. Geiger, A.: Tadgan: Time series anomaly detection using generative adversarial networks. In: Proc. of the 8th Int. Conf. on Big Data, IEEE, Piscataway. p. 33–43. https://doi.org/10.1109/BigData50022.2020.9378139
4. Guo, Y., Liao, W., Wang, Q., Yu, L., Ji, T., Li, P.: Multidimensional time series anomaly detection: A gru-based gaussian mixture variational autoencoder approach. Proc. Mach. Learn **95**, 97–112
5. Hartung, F., Franks, B.J., Michels, T., Wagner, D., Liznerski, P., Reithermann, S., Fellenz, S., Jirasek, F., Rudolph, M., Neider, D., Leitte, H., Song, C., Kloepper, B., Mandt, S., Bortz, M., Burger, J., Hasse, H., Kloft, M.: Deep anomaly detection on tennessee eastman process data. https://doi.org/10.1002/cite.202200238
6. Kim, S.: Towards a rigorous evaluation of time-series anomaly detection. In: Proc. of the 36th AAAI Conf. on Artificial Intelligence. p. 7194–7201. https://doi.org/10.1609/aaai.v36i7.20680
7. Kister, H.: What caused tower malfunctions in the last 50 years? Chemical Engineering Research and Design **81**(1), 5–26 (2003). https://doi.org/https://doi.org/10.1205/026387603321158159
8. Li, D.: Mad-gan: Multivariate anomaly detection for time series data with generative adversarial networks. In: Proc. of the 28th Int. Conf. on Artificial Neural Networks. p. 703–716. Springer. https://doi.org/10.1007/978-3-030-30490-4_56
9. Madakyaru, M., Kini, K.: A novel anomaly detection scheme for high dimensional systems using kantorovich distance statistic. Int. J. Inf. Technol. **14**, 3001–3010,. https://doi.org/10.1007/s41870-022-01046-0
10. Malhotra, P., Vig, L., Shroff, G., Agarwal, P.: Long short term memory networks for anomaly detection in time series. In: Proc. of the 23rd European Symposium on Artificial Neural Networks. p. 2015–56
11. Mirza, A., Cosan, S.: Computer network intrusion detection using sequential lstm neural networks autoencoders. In: Proc. of the 26th Signal Processing and Communications Applications Conf., IEEE. pp. ,1–4. https://doi.org/10.1109/SIU.2018.8404689
12. Niu, Z., Yu, K., Wu, X.: Lstm-based vae-gan for time-series anomaly detection. Sensors **20**(13), 3738. https://doi.org/10.3390/s20133738
13. Rieth, A., Amsel, B., Tran, R., Cook, M.: Additional tennessee eastman process simulation data for anomaly detection evaluation. Harvard Dataverse . https://doi.org/10.7910/DVN/6C3JR1
14. Ruff, L., Vandermeulen, R., Goernitz, N., Deecke, L., Siddiqui, S.A., Binder, A., Mueller, E., Kloft, M.: Deep one-class classification. Proc. Mach. Learn. Res **80**, 4393–4402
15. Shen, L., Li, Z., Kwok, J.: Timeseries anomaly detection using temporal hierarchical one-class network. In: Proc. of the 34th Int. Conf. on Neural Information Processing Systems. 13016–13026, Article 1092.
16. Soelch, M., Bayer, J., Ludersdorfer, M., van der Smagt, P.: Variational inference for on-line anomaly detection in high-dimensional time series. https://doi.org/10.48550/arXiv.1602.07109

17. Xu, H., Chen, W., Zhao, N., Li, Z., Bu, J., Li, Z., Liu, Y., Zhao, Y., Pei, D., Feng, Y., Chen, J., Wang, Z., Qiao, H.: Unsupervised anomaly detection via variational auto-encoder for seasonal kpis in web applications. In: Proc. of the 27th Int. World Wide Web Conf. https://doi.org/10.1145/3178876.3185996
18. Zhao, H., Wang, Y., Duan, J., Huang, C., Cao, D., Tong, Y., Xu, B., Bai, J., Tong, J., Zhang, Q.: Multivariate time-series anomaly detection via graph attention network. In: 2020 IEEE International Conference on Data Mining (ICDM). pp. 841–850. IEEE (2020). https://doi.org/10.1109/ICDM50108.2020.00093