

Ontology-enhanced deep learning framework for anomaly detection in oil and gas production plants

Gustavo Alessandro de Lima¹, Mara Abel¹

¹Instituto de Informática– Universidade Federal do Rio Grande do Sul (UFRGS)– 91501-970– Porto Alegre– RS– Brazil

Abstract

This proposal presents the creation of a framework that combines deep learning-based anomaly detection with ontology-driven knowledge representation to improve fault diagnosis in oil and gas production plants. The framework aims to leverage the strength of both techniques to reduce false alarm rates and to provide operators with more comprehensive information for decision-making.

Keywords

anomaly detection, ontology, oil and gas, time-series, framework

1. Introduction

Technological advancements pave the way for increased productivity and security in industrial process plants. Smart factories, brought by Industry 4.0, are characterized by their outstanding use of cutting-edge technology, with automation, monitoring, and artificial intelligence playing a significant role in operation efficiency [1].

These technological advancements apply not only to traditional manufacturing industries but also to various industrial processes, including the oil and gas sector, which is the focus of this proposal. An important improvement resulting from these advancements is the installation of sensor devices for constant information monitoring.

Despite the benefits, the vast amount of data generated by these sensors can prove challenging to analyze, creating the need for automated processes that verify that continuous stream of information in search of anomalies [2] that can indicate equipment failures, safety hazards, or inefficiencies in production.

The detection of these failures is of utmost importance to the sector. Plant shutdowns caused by failures can bring significant economic problems for companies. Moreover, safety hazards in this industry can have catastrophic consequences due to the hazardous nature of the industry, posing severe risks to worker safety and environmental integrity.

While traditional anomaly detection models can bring good results in specific areas, they still fail to understand the semantic characteristics of an oil and gas production plant, creating false results that can make it harder for an operator to address potential issues.

This work aims to create a framework that uses machine learning anomaly detection methods with a layer of ontology for semantic analysis of the oil and gas industry anomalies.

The paper is structured in the following manner: firstly, an analysis of the current state-of-the-art research will be done, focusing on works on the anomaly detection and ontology front, then the research proposal will be specified, showing improvements of the study and potential challenges. After that, the following steps in the research will be presented.

Proceedings of the 17th Seminar on Ontology Research in Brazil (ONTOBRAS 2024) and 8th Doctoral and Masters Consortium on Ontologies (WTD0 2024), Vitória, Brazil, October 07-10, 2024.

✉ gustavo.lima@inf.ufrgs.br (G. A. d. Lima); marabel@inf.ufrgs.br (M. Abel)

🌐 <https://www.petwin.org/ontology/> (G. A. d. Lima); <https://www.petwin.org/ontology/> (M. Abel)

🆔 0000-0002-9589-2616 (M. Abel)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

2. Related Work

Anomaly detection is a critical field in data analysis, with applications in various industries and domains. This section will explore state-of-the-art research in the area, focusing on machine-learning techniques and the use of ontologies.

2.1. Anomaly Detection

Anomaly detection is used to expose data points that differ from typical values. It can be used in network intrusion detection, fraud detection, fault diagnosis, and many other areas that require data mining applications [3].

Traditional anomaly detection relies on rules and statistical methods, often struggling to keep up with extensive dynamic and heterogeneous data [4]. Advanced techniques leverage machine learning algorithms for efficient and scalable detection to cope with that.

The context of fault diagnosis, especially in the oil and gas sector, sees anomaly detection widely used and studied for its role in maintaining operational efficiency and safety.

In [5], the authors propose a methodology considering seven fault types in production wells and lines, using a classifier based on random forest, and tuning their hyperparameters with a Bayesian non-convex optimizer. The authors achieved an accuracy of above 94% on the used data.

The authors at [6] propose a methodology that uses dynamic time warping and k-means clustering of the time series data to improve the performance of one-class classifiers, achieving increased performance metrics after using the proposed method.

When focusing more on neural network works, the authors at [7] propose a methodology that uses a generative adversarial network that is driven by a digital twin to conduct multivariate time series data anomaly detection that was able to increase by 2.6% the detection of anomalous data compared to other methods.

2.2. Ontologies

Computer science uses ontologies to model a system's relevant entities and relations. They can be used to infer information for all the explicitly represented knowledge [8].

The oil and gas industry employs ontologies to provide domain-specific knowledge and modeling concepts such as, but not limited to, wells, reservoirs, and production facilities. This representation can enable data integration and interoperability between different systems.

In the context of anomaly detection in the oil and gas sector, a noticeable research gap presents an opportunity for further exploration. Existing studies in this field have predominantly utilized ontology to enhance the visualization and integration of time series data.

Other areas use ontologies' rules and logic framework to represent and specify anomaly information [9].

In [10], the authors introduce NORIA-O, an ontology that represents network infrastructures, incidents, and maintenance that can be used to model complex Information and Communications Technology systems situations and can be used as the basis for anomaly detection.

The authors at [11] present expert knowledge on the maritime field using an ontology expressed in description logic, using automated reasoning tools for the context of anomaly detection. While the proposed approach was validated, the authors believe that further research is needed to suit the high processing demands of real maritime environments.

2.3. Anomaly Detection with Ontologies and Machine Learning

Combining ontologies and machine learning algorithms for anomaly detection is also an area with opportunities for further research, with very few works exploring the combination of both.

In [12], the authors explore this possibility by combining a long short-term memory network for the mathematical search of anomalies with a fuzzy web ontology for a second stage that filters the results

for anomalies that only affect a specific subject area. Their experiments focused on facies logs of nine drilling wells and achieved good efficiency.

The authors at [13] propose a new methodology, FLAGS, that combines data and knowledge-driven techniques, using semantic filters to classify anomalies as known behavior, reducing the load on operators to verify real alerts. This methodology was tested on the railway domain on the topic of predictive maintenance.

In [14], the authors introduce a method that uses a semantic approach to reduce the number of features used for the anomaly detection process, then integrates the proposed model on IBM's Mape-K loop, which combines inference rules and a Hierarchical Temporal Memory algorithm. The authors applied their methodology to cellular vehicular communication systems and achieved encouraging results.

3. Research Proposal

The main goal of this proposal is to develop a modular framework that combines a deep-learning anomaly detection model layer with an ontology knowledge representation layer to enhance the identification and interpretation of anomalies in oil and gas production plants. The framework aims to leverage the strength of both machine learning and semantic technologies to provide a more accurate fault diagnosis system.

The framework's modularity comes from the first layer, which is designed to be agnostic as to which deep learning model is implemented, creating the ability to plug different trained models as needed.

One of the side goals of this research is the creation of the ontology itself, which will contain the representation of a certain area inside an oil and gas production plant, with the creation of semantic rules to insert the semantic knowledge into the framework. The creation of the ontology will follow the NeOn methodology [15].

The other side goal is the choice and training of machine learning models, focusing on novel and recent advancements in the area, such as graph neural networks or the use of transformer-based architectures.

The proposed framework is shown in Figure 1.

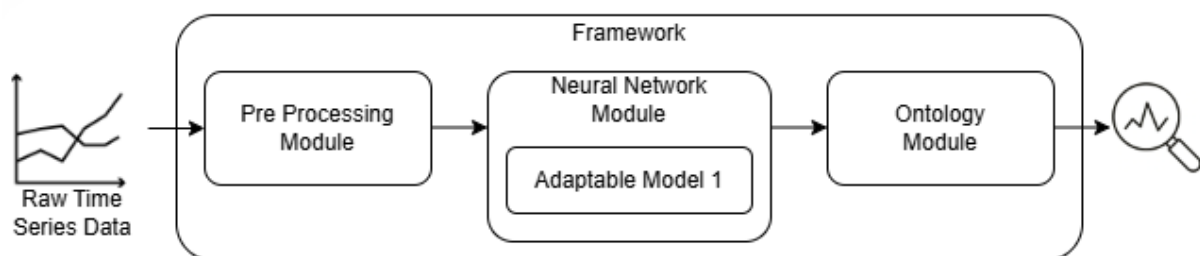


Figure 1: Proposed framework.

3.1. Expected Contributions

The proposed framework can offer advantages over existing methods:

- Increased accuracy in anomaly detection by combining data-driven insights and domain expertise.
- Facilitate the operator's ability to understand anomalous data and quickly verify affected systems.
- Reduced false alarm rates.
- Deepen research on the combination of ontologies and machine learning applications.

3.2. Potential Challenges of the Proposal

Combining two different methods can increase the number of challenges the research faces.

On the side of anomaly detection, it is required that the technique chosen is adequate for the process and the data provided. It is also worth noting the difficulty in gathering data for training, which might require extra work on data gathering and creation.

Creating an ontology is also a complex task, requiring the assistance of domain experts and knowledge of industry standards to ensure the ontology's validity and completeness.

After these challenges are overcome, there's also the need to validate the framework in real-world settings and check if the system is scalable enough to handle the large amount of data a real oil and gas production plant can generate.

3.3. Research Benchmarks

In order to evaluate the framework and compare it with other applications, a standard for anomaly detection introduced in [16] will be utilized. This benchmark involves performing multiple training and testing rounds and computing precision, recall, and F1 scores for each round. The average F1 score will be used to assess the framework's validity.

4. Research Methodology

The methodology used to guide this research will be the Design Science Research Methodology, which has roots in engineering and is fundamentally a problem-solving paradigm seeking innovative solutions to real-world problems [17].

The steps of this methodology are as follows:

- Step 1: Problem identification and motivation.
- Step 2: Define the objectives for a solution.
- Step 3: Design and development.
- Step 4: Demonstration.
- Step 5: Evaluation.
- Step 6: Communication.

4.1. Current steps

The research is still in its first step. The problem was first identified via interviews with oil and gas operators, introducing the everlasting anomaly detection problem in specific areas and the burden it can create on operators that need to verify large amounts of time series data to verify the validity of alarms in a production plant.

4.2. Next Steps

More profound research will be done for the next steps to gather more specific objectives. This includes defining a particular area of interest in applying the framework to incorporate a case study approach.

With a specific area defined, it's time to start planning the development of the ontology, which will include interviews with domain experts in the area to gather expert knowledge.

There's also the need for data gathering to evaluate and train the machine-learning models to use in the modular layer of the framework to check the model's effectiveness and the increased accuracy of including a semantic layer.

After the ontology has been planned and the data for the machine learning models has been gathered, the bulk of the research will start, which includes the design and then the implementation of the framework, with benchmarks of the validity of the solution happening in parallel with the development.

5. Conclusion

This research proposal aims to develop a framework that combines the strengths of deep learning-based anomaly detection models with ontology-driven knowledge to improve fault diagnosis in the oil and gas industry, refining the accuracy of detection and interpretability of anomalies.

The research will follow the Design Science Research Methodology to find a solution to improve false alarm rates and provide operators with more information to enable better decisions, addressing the research gap in the area and contributing to the continued advancement of Industry 4.0 technologies in the oil and gas sector.

Acknowledgments

The authors acknowledge CAPES-Brazil Finance Code 001, the Brazilian Agency CNPq, and the Petwin Project, supported by FINEP and Libra Consortium (Petrobras, Shell Brasil, Total Energies, CNOOC, CNPC).

References

- [1] M. Soori, B. Arezoo, R. Dastres, Internet of things for smart factories in industry 4.0, a review, *Internet of Things and Cyber-Physical Systems* 3 (2023) 192–204. URL: <https://www.sciencedirect.com/science/article/pii/S2667345223000275>. doi:<https://doi.org/10.1016/j.iotcps.2023.04.006>.
- [2] D. Miodutzki, C. Tacla, L. Gomes-Jr, Outlier detection with ontology-driven fault contextualization in the industry 4.0, in: *Anais do XXXVII Simpósio Brasileiro de Bancos de Dados*, SBC, Porto Alegre, RS, Brasil, 2022, pp. 267–278. URL: <https://sol.sbc.org.br/index.php/sbbd/article/view/21812>. doi:[10.5753/sbbd.2022.224309](https://doi.org/10.5753/sbbd.2022.224309).
- [3] D. Samariya, A. Thakkar, A comprehensive survey of anomaly detection algorithms, *Annals of Data Science* 10 (2023) 829–850.
- [4] B. Dhamodharan, Beyond traditional methods: A novel approach to anomaly detection and classification using ai techniques, *Transactions on Latest Trends in Artificial Intelligence* 3 (2022).
- [5] M. A. Marins, B. D. Barros, I. H. Santos, D. C. Barrionuevo, R. E. Vargas, T. d. M. Prego, A. A. de Lima, M. L. de Campos, E. A. da Silva, S. L. Netto, Fault detection and classification in oil wells and production/service lines using random forest, *Journal of Petroleum Science and Engineering* 197 (2021) 107879.
- [6] A. P. F. Machado, C. J. Munaro, P. M. Ciarelli, R. E. V. Vargas, Time series clustering to improve one-class classifier performance, *Expert Systems with Applications* 243 (2024) 122895.
- [7] Y. Lian, Y. Geng, T. Tian, Anomaly detection method for multivariate time series data of oil and gas stations based on digital twin and mtad-gan, *Applied Sciences* 13 (2023). URL: <https://www.mdpi.com/2076-3417/13/3/1891>. doi:[10.3390/app13031891](https://doi.org/10.3390/app13031891).
- [8] N. Guarino, D. Oberle, S. Staab, *What Is an Ontology?*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2009, pp. 1–17. URL: https://doi.org/10.1007/978-3-540-92673-3_0. doi:[10.1007/978-3-540-92673-3_0](https://doi.org/10.1007/978-3-540-92673-3_0).
- [9] J. Baumeister, D. Seipel, Anomalies in ontologies with rules, *Journal of Web Semantics* 8 (2010) 55–68. URL: <https://www.sciencedirect.com/science/article/pii/S1570826809000778>. doi:<https://doi.org/10.1016/j.websem.2009.12.003>.
- [10] L. Tailhardat, Y. Chabot, R. Troncy, Noria-o: an ontology for anomaly detection and incident management in ict systems, in: *European Semantic Web Conference*, Springer, 2024, pp. 21–39.
- [11] J. Roy, M. Davenport, Exploitation of maritime domain ontologies for anomaly detection and threat analysis, in: *2010 International WaterSide Security Conference*, IEEE, 2010, pp. 1–8.
- [12] V. Moshkin, D. Kurilo, N. Yarushkina, Integration of fuzzy ontologies and neural networks in the

detection of time series anomalies, *Mathematics* 11 (2023). URL: <https://www.mdpi.com/2227-7390/11/5/1204>. doi:10.3390/math11051204.

- [13] B. Steenwinckel, D. De Paepe, S. V. Haute, P. Heyvaert, M. Bentefrit, P. Moens, A. Dimou, B. Van Den Bossche, F. De Turck, S. Van Hoecke, et al., Flags: A methodology for adaptive anomaly detection and root cause analysis on sensor data streams by fusing expert knowledge with machine learning, *Future Generation Computer Systems* 116 (2021) 30–48.
- [14] Q. Ricard, P. Owezarski, Ontology based anomaly detection for cellular vehicular communications, in: *10th European Congress on Embedded Real Time Software and Systems (ERTS 2020)*, 2020.
- [15] M. C. Suárez-Figueroa, Neon methodology for building ontology networks: Specification, scheduling and reuse, 2010. URL: <https://oa.upm.es/3879/>. doi:10.20868/UPM.thesis.3879, ontology Engineering Group.
- [16] R. E. V. Vargas, C. J. Munaro, P. M. Ciarelli, A. G. Medeiros, B. G. do Amaral, D. C. Barrionuevo, J. C. D. de Araújo, J. L. Ribeiro, L. P. Magalhães, A realistic and public dataset with rare undesirable real events in oil wells, *Journal of Petroleum Science and Engineering* 181 (2019) 106223.
- [17] J. Vom Brocke, A. Hevner, A. Maedche, Introduction to design science research, *Design science research. Cases* (2020) 1–13.