Advances in Reliability, Safety and Security

ESREL 2024
Collection of Extended Abstracts

Robust Anomaly Detection In Unmanned Ship Systems Based On Large Language Models

Dun Li, Huan Wang, Yan-Fu Li

Department of Industrial Engineering, Tsinghua University, Beijing, China

Keywords: anomaly detection, Unmanned Ship Systems, Large Language Models, maritime automation

In the increasingly complex realm of maritime automation, the rise of Unmanned Ship Systems (USS) ushers in a transformative era marked by unparalleled advancements in safety, efficiency, and reliability. This progress urgently calls for the creation of anomaly detection mechanisms that are not only precise but also versatile enough to adapt to the distinct operational settings of autonomous maritime endeavors. Drawing inspiration from the triumphant application of Human Reliability Analysis (HRA) across various fields of automation and the pioneering use of health metrics in high-speed trains, our research introduces a pioneering anomaly detection framework for USS (Sethu, 2023). This framework, by harnessing the power of Large Language Models (LLMs) such as BERT and GPT, seeks to offer an exhaustive, interdisciplinary tool that adeptly mirrors operational conditions through semantic analysis, thus boosting survivability and reliability in the face of the marine environment's dynamic challenges (Devlin, 2018; Wu, 2023). The proposed methodology enhances anomaly detection processes by incorporating insights from successful health metric implementations across industries and integrating HRA within autonomous frameworks. We begin by preprocessing operational data—system logs and sensor readings — and transforming them into structured formats amenable to language model analysis. Pretrained LLMs, specifically BERT and GPT, further refine this data, identifying operational deviations through their advanced semantic understanding capabilities. This process involves two critical stages: the conversion of operational data into text format for in-depth analysis, and the subsequent use of LLMs to pinpoint anomalies based on this semantic comprehension.

Building on this foundation, this method extends to converting raw log data into structured log templates and variables via efficient log parsing techniques. This structured data is then formatted into text, setting the stage for deeper language model processing. The employment of trained LLMs to translate this textual data into vectorized representations is a key step, capturing the nuanced semantic information contained within log texts and providing essential context for anomaly detection. Employing the vectorized log data, we leverage a Bidirectional Long Short-Term Memory (Bi-LSTM) model, adept at handling time-series data and discerning dependencies within log data. Training the Bi-LSTM model with data from standard operations enables it to recognize patterns of log sequences under normal conditions, laying the groundwork for anomaly detection when deviations occur. Upon training, the model is capable of processing new log data vectors in real time, assessing anomalies by comparing deviations from the norm. To mitigate the impact of software updates or configuration changes that introduce data variations, we've designed a model transfer strategy that allows for swift adaptation to new operational contexts, ensuring the continuous accuracy of anomaly detection. Furthermore, a feedback mechanism, informed by the discovery of new anomalies and the accuracy of model predictions during actual operations, facilitates the model's constant refinement and optimization. This ensures sustained learning and adaptation, underpinning our approach's capacity to address the evolving challenges of maritime automation with agility and precision. Specifically, the model contains the following:

- Data Preprocessing and Standardization;
- Vectorization with Pre-Trained Large Language Models;
- Serialized Analysis through Bidirectional LSTM Networks;
- Real-Time Anomaly Detection and Assessment;
- Model Optimization and Adaptation Strategies;
- Continuous Learning and Feedback Mechanisms;

- Experimental Evaluation and Performance Comparison.

As Table 1 shows, the process of anomaly detection within USS is underpinned by a comprehensive workflow, beginning with data preprocessing and transformation, and culminating in sophisticated model training and real time anomaly assessment. Initially, original log data generated by USS undergoes standardization through efficient log parsing technology. This step extracts meaningful structured data from unstructured log information, which is then formatted into text for subsequent language model processing. Then, the preprocessed text data is converted into vectorized representations using pre-trained LLMs such as BERT and GPT. The next phase involves serialized analysis using a Bi-LSTM network. Through training with data collected during normal operational periods, the Bi-LSTM model learns to recognize the patterns of log sequences under standard conditions, setting a baseline for anomaly detection. Once the model is trained, new vectorized log data is fed in real time for anomaly detection. The model assesses whether anomalies exist by comparing the deviations between the input data and the learned normal patterns. To ensure the model remains effective despite potential data drifts due to USS software updates or configuration changes, a model transfer strategy has been designed. This allows the model to rapidly adapt to new operational environments, continuing to provide accurate anomaly detection services. Moreover, a feedback loop is established to continuously refine and optimize the model based on newly discovered anomalies and the accuracy of model predictions, enabling ongoing learning and adaptation.

Table 1. Predictive Health Indicators for Autonomous Ships.

Stage	Description	Tools/Models Used
Data Preprocessing and Transformation	Convert raw USS operational data into structured formats for analysis, involving log parsing to extract meaningful data and format it into text.	Log Parsing Techniques
Vectorization with Pre-Trained Large Language Models (LLMs)	Transforms text-formatted log data into vectorized representations, capturing semantic information crucial for anomaly detection.	BERT, GPT
Serialized Analysis through Bidirectional LSTM Networks	Employ a Bi-LSTM model to process vectorized log data, recognizing patterns and dependencies within time series data for anomaly detection.	Bi-LSTM
Real-Time Anomaly Detection and Assessment	Feed new log data vectors into the trained Bi-LSTM model for real-time anomaly detection, comparing deviations from the learned normal patterns.	Bi-LSTM
Model Optimization and Adaptation Strategies	Implement a model transfer strategy to adapt to new operating environments and data variations, ensuring continuous provision of accurate anomaly detection.	Model Transfer Strategy
Continuous Learning and Feedback Mechanisms	Establish a feedback loop for ongoing model learning and adaptation based on the discovery of new anomalies and the accuracy of predictions.	Feedback Mechanism
Experimental Evaluation and Performance Comparison	Evaluat the method in a simulated environment, comparing its performance with traditional machine learning methods in anomaly detection tasks.	Simulation Environment

Experimental evaluation of this method in a simulation environment demonstrates that the performance of LLMs combined with the Bi-LSTM model in anomaly detection tasks significantly surpasses traditional machine learning methods, not only in terms of detection accuracy but also in achieving more efficient real-time performance. The successful application of this approach not only showcases the enormous potential of LLMs in enhancing anomaly detection capabilities but also provides robust technical support for safe operation, bringing new research directions and solutions to the field of maritime automation. Through the evaluation of simulated operation data, incorporating various operational anomalies, the effectiveness of our proposed method has been assessed. Compared to traditional machine learning methods, LLMs exhibit significant advantages in the accuracy and timeliness of anomaly detection. Our findings highlight the potential of LLMs to significantly enhance the reliability and safety of operations. Utilizing LLMs for anomaly detection offers an innovative and effective solution to ensure operational safety.

Acknowledgements

This manuscript is supported by the National Natural Science Foundation of China under a key project Grant 71731008 and the Beijing Municipal Natural Science Foundation-Rail Transit Joint Research Program (L191022).

References

Devlin, J., Chang, M. W., Lee, K. et al. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv, 1810.04805.

Sethu, M., Kotla, B., Russell, D. et al. 2023. Application of artificial intelligence in detection and mitigation of human factor errors in nuclear power plants: a review. Nuclear Technology 209(3), 276-294.

Wu, T., He, S., Liu, J. et al. 2023. A brief overview of ChatGPT: The history, status quo and potential future development[J]. IEEE/CAA Journal of Automatica Sinica 10(5), 1122-1136.