NOISE-CONTRASTIVE LEARNING FOR SYSTEM LOG ANOMALY DETECTION WITH PATTERN FEATURES

Author's: K. Vignesh[1], N. Vanjulavalli[2], K. Sujith [3]

Abstract

System logs serve as a critical source of information for monitoring system behavior, diagnosing failures, and detecting security threats in modern enterprise environments. Traditional log anomaly detection approaches primarily rely on next-event prediction models using deep learning architectures such as LSTM and Transformer-based frameworks. However, these approaches suffer from an inherent objective mismatch, as event prediction does not directly optimize anomaly detection performance. Additionally, many deep learning models incur high computational costs, making them unsuitable for CPU-based enterprise deployments. This proposes a context-aware log anomaly detection framework that reformulates anomaly detection as a direct binary classification task using a Noise-Contrastive Learning (NCL) strategy. Instead of predicting the next log event, synthetic noise samples are generated by injecting controlled perturbations into normal event sequences. This enables the model to learn a discriminative decision boundary between normal and abnormal patterns without requiring labeled anomaly data.To effectively capture contextual dependencies among log events, a novel feature engineering method termed Pattern Feature is introduced. This feature highlights potential anomalies within event sequences through masked event prediction. Furthermore, a computationally efficient ensemble learning model, STrees (Feature Representation Decision Trees combined with Extremely Randomized Trees), is employed to enhance detection performance while maintaining CPU-friendly inference.Experimental validation on real-world log datasets demonstrates improved anomaly recall and balanced detection capability compared to traditional LogDeep and Deep Anomaly Detection (DeepAD) methods. The proposed framework achieves robust performance, improved anomaly sensitivity, and superior computational efficiency, making it suitable for large-scale enterprise deployment.