Rare Events Detection and Forecasting in Dynamic Systems

Eugeniu Catlabuga, Nicolae Drabcinschi, Viorel Munteanu, Viorica Sudacevschi
Technical University of Moldova, eugeniu.catlabuga@doctorat.utm.md, nicolae.drabcinski1@lt.utm.md, viorel.munteanu@lt.utm.md, viorica.sudacevschi@calc.utm.md, ORCID: 0009-0001-0286-0667, 0009-0008-4381-836X, 0000-0002-4133-5945, 0000-0003-0125-3491

Keywords: Rare events detection, dynamic systems, SVM kernels, Predictive models, Banking system, Forecasting

Abstract. Rare events, such as financial crises, banking failures and epidemics outbreaks can significantly impact economies and people's lives [1]. Early detection in the banking sector is crucial for mitigating financial losses and maintaining operational stability. Traditional models often fail to predict these events due to their unpredictable nature and complex dynamics. Existing systems based on methods like decision trees and neural networks struggle with scalability and accuracy, limited by specific resources of information systems and real datasets. This work aims to develop and evaluate a machine learning-based software pipeline for detecting and forecasting rare events in dynamic systems, using heterogeneous data sources such as Loghub from the banking system, which contains operations logs, transaction data, and system performance metrics [2]. The pipeline employs advanced feature selection techniques, including Wavelet Transform for multi-scale analysis and Fenchel Transform to extract meaningful patterns. We utilize Support Vector Machines (SVMs) equipped with linear, polynomial, radial basis function (RBF), and sigmoid kernels to capture diverse data characteristics [3]. To ensure training on balanced data and mitigate overfitting, we integrate oversampling (SMOTE) with undersampling strategies. Hyperparameters are refined via grid search and cross-validation, enhancing the SVM classifier's ability to adapt to the unique features of the data. The probabilistic outputs from these models are

further processed using a Binary Logistic Regression, which obtains the final prediction for each event.

The pipeline was tested in the context of banking systems, which involve server-based environments and logs from operating systems, databases, and applications. The study specifically utilizes publicly available logs from Loghub, sourced from lab servers, to validate the model's effectiveness in identifying rare events in complex real-world scenarios. The results demonstrate the pipeline's efficiency, achieving a precision of 94%, recall of 92%, and a low MSE of 0.03, indicating its robustness in predicting rare events [4]. Furthermore, the model achieved an overall accuracy of 96% in classifying and forecasting rare events, highlighting its strong predictive performance in dynamic systems.

References

1. Abubakar, Yahaya Idris, Alice Othmani, Patrick Siarry, and Aznul Qalid Md Sabri. 2024. A Systematic Review of Rare Events Detection Across Modalities using Machine Learning and Deep Learning. https://www.authorea.com/users/693917/articles/683317-a-systematic-review-of-rare-events-detection-across-modalities-using-machine-learning-and-deep-learning.
2. Zhu, Jieming, Shilin He, Pinjia He, Jinyang Liu, and Michael R. Lyu. 2023. Loghub: A Large Collection of System Log Datasets for AI-driven Log Analytics. arXiv. https://doi.org/10.48550/arXiv.2008.06448.
3. Cristianini, Nello, and Elisa Ricci. 2008. Support Vector Machines. In *Encyclopedia of Algorithms*, ed. Ming-Yang Kao, 928–932. Boston, MA: Springer US. https://doi.org/10.1007/978-0-387-30162-4_415.
4. Bioinformatics-Lab-TUM/Rare_events_detection. Bioinformatics Lab, Technical University of Moldova. https://github.com/Bioinformatics-Lab-TUM/Rare_events_detection