Dealing with High Cardinality of Network Management System Data for Machine-Learning-Based Alarm Classification

Lareb Zar Khan
Scuola Superiore Sant'Anna
Pisa, Italy
larebzar.khan@santannapisa.it

Ahmed Triki
Orange Innovation
Lannion, France
ahmed.triki@orange.com

Maxime Laye
Orange Innovation
Lannion, France
maxime.laye@orange.com

Nicola Sambo
Scuola Superiore Sant'Anna
Pisa, Italy
nicola.sambo@santannapisa.it

***Abstract*—Failure management in optical networks usually deals with the processing of alarms, including alarm classification. The alarms data obtained from network management systems often include a high number of categorical features (e.g., name of the alarm, cause), which can make machine learning (ML) training a complex and data-intensive process. To overcome this issue, this paper proposes a pre-processing technique for alarms data, enabling the use of smaller datasets for ML-based alarm classification. The results obtained using three different ML models suggest the effectiveness of the proposed approach.**

***Index Terms*—Alarm classification, Network management system, Data dimensionality reduction.**

I. Introduction

Failure management has emerged as a key application of machine learning (ML) in optical networks [1]–[3]. A failure in an optical network is normally associated with a large number of alarms [4], [5] in the network management system (NMS). The classification of alarms according to their root causes can simplify the analysis of alarms for the operational teams and accelerate their intervention to clear failures. The complexity of alarm classification process is not only related to the large number of existing alarms in operational networks but also the context that should be taken into account. Given the complexity of the problem, large datasets are needed to train the ML model [6] to deal with different type of failures and to overcome the high cardinality issue of categorical features in the alarms dataset. This complexity could make the proposed solutions greedy in terms of required computational resources and training time. Authors in [7] propose to consider the overall classification problem as a succession of binary classification sub-problems where the ML-based model has to distinguish between alarms generated by a specific failure and all the other alarms (hereinafter referred to as *noise* alarms). This approach simplifies the complexity of the problem but does not address the high cardinality issue efficiently.

In this paper, we used a small experimental dataset with many noise alarms to perform a binary classification. We proposed an alarms pre-processing approach in which high cardinality of categorical features has been dealt using a more sophisticated and reversible encoding method. The effectiveness of encoded data was evaluated on three different ML models trained for alarm classification in two optical subnetworks, with results on test dataset indicating high F1-scores for alarm classification. Using the available small dataset, our approach has been shown to significantly reduce the training and inference time without having a significant impact on the classification performance.

II. Experimental Testbed Setup

The experimental testbed shown in Fig.1 was used to collect the alarms data. It consisted of two subnetworks managed by the real field-like NMS. The subnetwork 1 included four main nodes labeled N1, N2, N4, and N5, as well as an inline amplifier (N3) considered as a node. Similarly, subnetwork 2 had three nodes marked as N6, N7 and N8 in Fig.1. The optical power attenuation co-efficient for utilized single-mode-fibers was 0.25 dB/km, with an average distance of almost 50 km between two adjacent nodes. All nodes were capable of communicating with the NMS to update the status of their cards and raise alarms in case of any failure. The considered optical services in both the subnetworks are also indicated in Fig.1. In order to artificially introduce failures in the networks, variable optical attenuators (VOAs) were placed at various positions labeled as P1-P8 in Fig.1. The attenuation was increased from its lowest possible level to the point where we observed complete service disruption. This attenuation of an optical signal was considered as a failure, and the resultant raised alarms were used for this investigation. The alarms generated because of other failures or uncleared alarms from previous failures were considered as noise alarms.

For this study, we assumed that failure events (i.e., different levels of attenuation of an optical signal) occur one at a time and are independent of each other. In our future work, we intend to consider other types of failures as well.

III. Alarms Pre-Processing

We started by removing redundant features from the raw alarms data during the feature engineering process by utilizing domain expertise. After that, we were left with five

978-3-903176-54-6 © 2023 IFIP

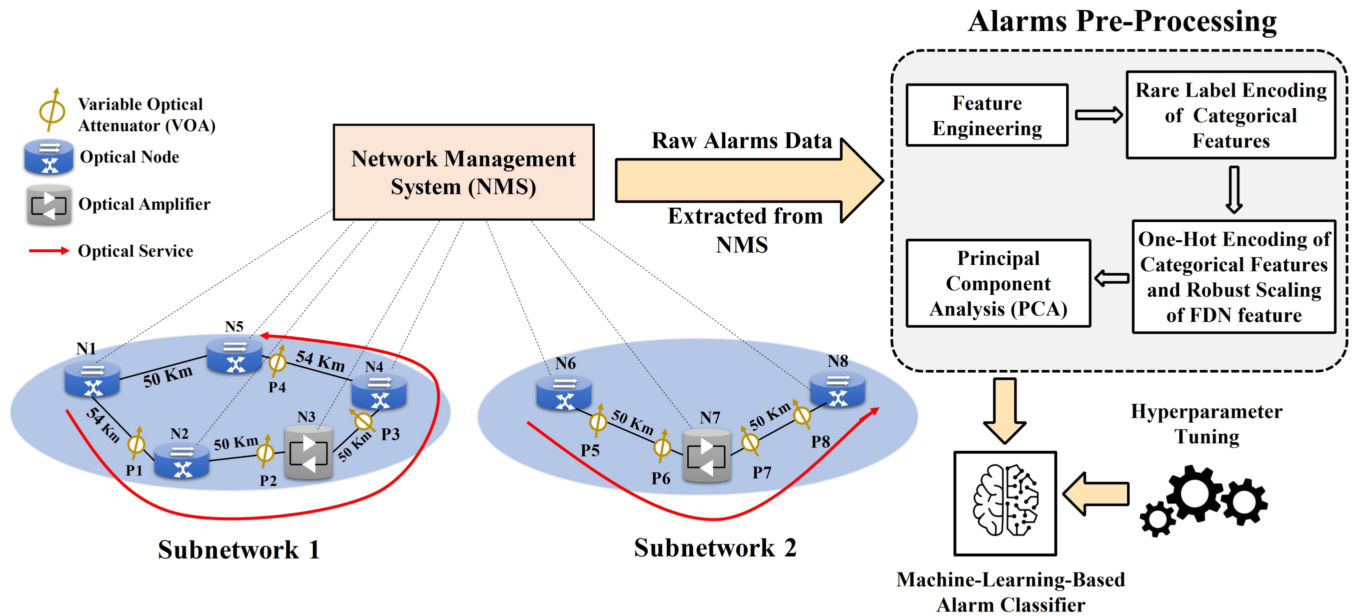

Fig. 1. Experimental testbed setup and alarms pre-processing

categorical features (PROBABLE_CAUSE, ALARM_NAME, NE_NAME, AFFECTED_OBJECT_NAME, ADDITIONAL_TEXT) and one numerical feature (FDN). There were many unique labels in each categorical feature, and each label ideally maps to a new dimension in the dataset after one hot encoding (OHE). Table. I shows the training dataset size, $s = (a, d)$, at various stages of alarms pre-processing for alarm classification. In s, a represents the number of alarms and d represents the dataset dimensions, which are equal to the total number of unique labels across all categorical features and the number of numerical features (i.e., only FDN feature in this case that was scaled using Robust Scaler).

TABLE I
ALARMS PRE-PROCESSING INSIGHTS (TRAINING DATASET)

	Subnetwork 1	Subnetwork 2
Before rare-label encoding	(1400, 866)	(955, 623)
After rare-label encoding	(1400, 94)	(955, 109)
After PCA (98%)	(1400, 57)	(955, 56)

With the proposed pre-processing, to address the high cardinality of the alarms dataset, we assigned a new label *rare* to all the less frequent labels in each categorical feature, determined on the basis of a threshold. We used a 1% threshold for alarm classification. In order to further reduce the dimensions of the dataset, we used the principal component analysis (PCA) technique [8], which linearly maps data from N to K dimensions where K<N, while preserving the specified amount of information. In this case, PCA was used to preserve 98% of the information; which means that we lose 2% of the information at this stage. It should be noted that the dimensions indicated against "Before rare-label encoding" in Table. I are the dimensions obtained with basic pre-processing (in which neither rare-label encoding nor PCA was performed).

IV. Results and Discussion

After pre-processing the alarms data, we evaluated the performance of the encoded dataset using three ML models: gradient boosting classifier (GBC) [9], extreme gradient boosting (XGBoost) [9], and light gradient boosting machine (LightGBM) [9]. The achieved performance from encoded data obtained using the proposed method was compared to the basic pre-processing scenario. The metrics selected for comparison were the training and inference time, as well as the F1-score, which is a reliable measure of classification performance even when class imbalance is present.

Figs. 2(a, b, and c) present the results achieved from basic data pre-processing, whereas Figs. 2(d, e, and f) show the results for proposed pre-processing. As shown in Figs. 2(a) and 2(d), for GBC, training time was reduced by 41.04% and 51.4% for subnetwork 1 and subnetwork 2, respectively. Similarly, inference time was reduced by 58.9% and 18.75% for subnetwork 1 and subnetwork 2, respectively as indicated by Figs. 2(b) and 2(e). Figs. 2(c) and 2(f) demonstrate that the information loss (i.e., 2% during PCA) during dimensionality reduction process, has a minimal impact on the classification performance of the ML models. For GBC, F1-score was reduced by only 0.2% and 3.1% in subnetworks 1 and 2, respectively.

The proposed dimensionality reduction framework has been found to achieve similar performance improvements for the other two considered ML models as well. The proposed pre-processing technique significantly reduced the training and inference time for XGBoost (i.e., up to 47%) while maintaining comparable classification F1-score. Similarly, for LightGBM,

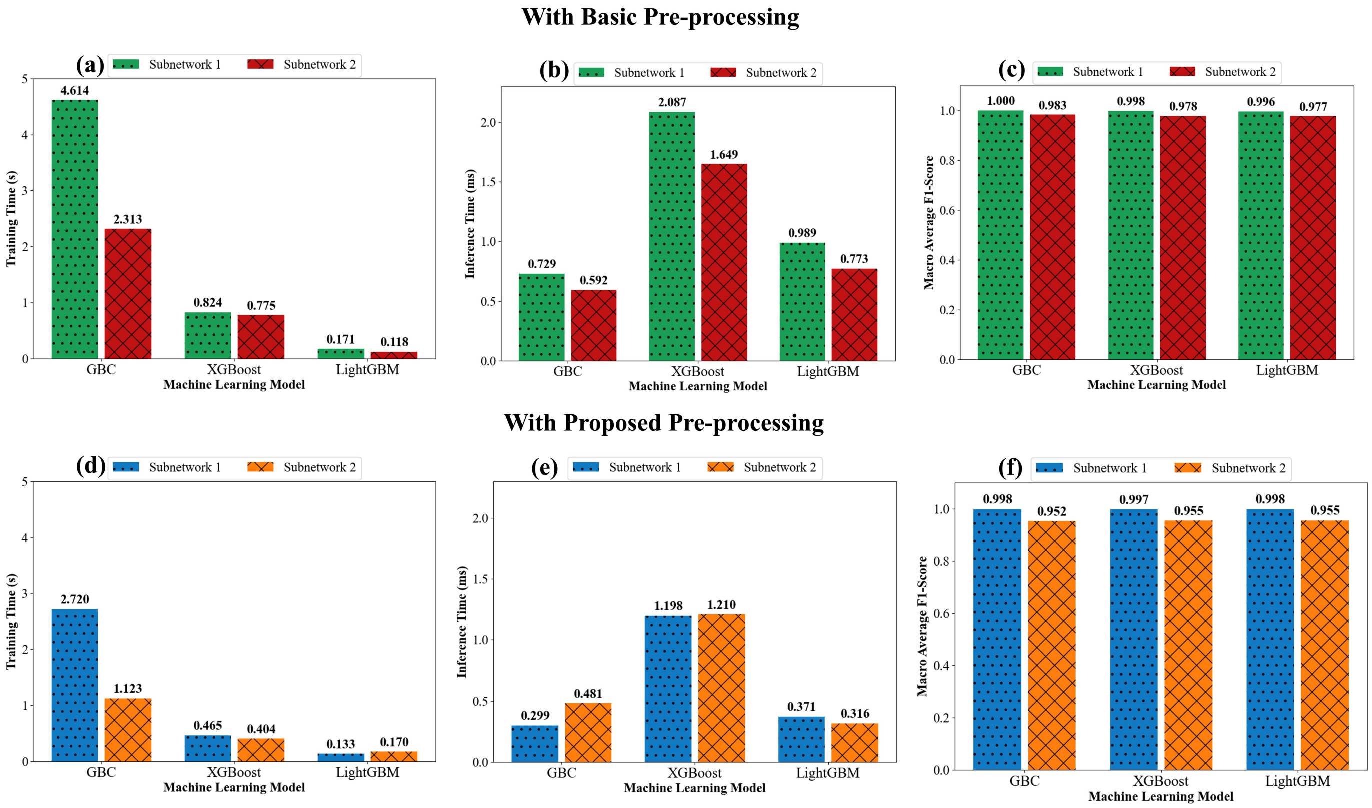

Fig. 2. Performance comparison of proposed and basic pre-processing in terms of training time, inference time, and macro-average F1-Score

we observed a reduction of up to 60% in inference time, with only a slight impact on classification performance. This demonstrates the effectiveness of the proposed framework in reducing computational costs while maintaining performance. The trade-off between reduction in inference/training time and classification accuracy can be considered based on the specific requirements and scenarios.

V. Conclusion

We investigated an alarms pre-processing method to improve the quality of the small training alarms dataset for machine learning (ML) models. The effectiveness of the processed dataset was assessed for a typical use-case i.e., ML-based alarm classification in optical networks. Three different ML models were used for this purpose, and the proposed alarm pre-processing has shown to significantly reduce training and inference time for these ML models, with only a slight impact on F1-score (up to 3%) due to the loss of information during dimensionality reduction.

Acknowledgements

This work has been funded by EU H2020 Marie Skłodowska-Curie Actions ITN project MENTOR (GA 956713).

References

[1] F. N. Khan, Q. Fan, C. Lu, and A. P. T. Lau, "An optical communication's perspective on machine learning and its applications," Journal of Lightwave Technology **37**, 493–516 (2019).

[2] L. Z. Khan, A. Triki, M. Laye, and N. Sambo, "Optical network alarms classification using unsupervised machine learning," in *2022 27th OptoElectronics and Communications Conference (OECC) and 2022 International Conference on Photonics in Switching and Computing (PSC),* (2022), pp. 1–3.

[3] L. Z. Khan, J. Pedro, N. Costa, L. De Marinis, A. Napoli, and N. Sambo, "Data augmentation to improve performance of neural networks for failure management in optical networks," Journal of Optical Communications and Networking **15**, 57–67 (2023).

[4] N. Sambo, F. Cugini, A. Sgambelluri, and P. Castoldi, "Monitoring plane architecture and oam handler," Journal of Lightwave Technology **34**, 1939–1945 (2016).

[5] Q. Pham-Van, V. López, A. M. Lopez-de Lerma, R. Szwedowski, K. Mrówka, S. Auer, H.-T. Thieu, Q.-H. Tran, D. Verchere, G. Atkinson, A. Autenrieth, S. Neidlinger, and L. Tancevski, "Demonstration of alarm correlation in partially disaggregated optical networks," in *2020 Optical Fiber Communications Conference and Exhibition (OFC),* (2020), pp. 1–3.

[6] B. Zhang, Y. Zhao, S. Rahman, Y. Li, and J. Zhang, "Alarm classification prediction based on cross-layer artificial intelligence interaction in self-optimized optical networks (soon)," Optical Fiber Technology **57**, 102251 (2020).

[7] J. Babbar, A. Triki, R. Ayassi, and M. Laye, "Machine learning models for alarm classification and failure localization in optical transport networks," Journal of Optical Communications and Networking **14**, 621–628 (2022).

[8] S. Shalev-Shwartz and S. Ben-David, *Understanding Machine Learning: From Theory to Algorithms* (Cambridge University Press, 2014).

[9] C. Bentéjac, A. Csörgő, and G. Martínez-Muñoz, "A comparative analysis of gradient boosting algorithms," Artif. Intell. Rev. **54**, 1937–1967 (2021).