"Detection and Root Cause Analysis of Performance Degradation in Optical Networks Using Machine Learning"

Christine Tremblay(1), Ali Mahmoudialami(1), Philippe Alain Ngani Sigue(1), Yinqing Pei(2), David Côté(2), David Boertjes(2), Dacian Demeter(3), Christian Desrosiers(1)

(1) École de technologie supérieure, Montréal, Canada, christine.tremblay@etsmtl.ca
(2) Ciena Corp., Ottawa, Canada
(3) TELUS Corp., Edmonton, Canada

Abstract *In this paper, we show how multiclass classifiers based on supervised learning can be used for identifying the root cause of performance degradations observed in a production optical network. ©2023 The Author(s)*

Introduction

To ensure that the network operates continuously and to maintain the desired level of quality of service (QoS), it is imperative that network faults are resolved as quickly as possible. Fault management methods based on machine learning (ML) have been explored extensively in the last years to serve this purpose [1-16].

The identification and localization of failures through a smart fault analysis system can lead to preventive maintenance and fast network recovery. Currently, fault management methods rely on predefined thresholds to detect anomalies and human expertise to determine fault resolution [3,4]. A threshold-based system that is covering the entire network is not trivial to establish since different vendors provide different equipment and technologies for different layers of the network (known as heterogeneity factors). Moreover, a narrow threshold may result in false alarms, while a broad threshold can reduce fault detection rates [3].

Previous threshold-based methods have high cost and limited scalability. Due to the unpredictable nature of network behavior, proactive measures must be implemented rather than reactive measures. The remaining challenge of failure recovery is time consuming and prone to human interpretation errors due to lack of end-to-end expertise. The goal of this study is to develop root cause analysers in a software-defined networking (SDN) context that translates the optical domain knowledge into software tools that automate the fault localization and identification tasks and enable the post failure classification study either with full, or, more importantly, incomplete network monitoring data.

In this paper, we show how to identify circuit failures in optical networks using multiclass classifiers trained with field data.

Network faults and performance degradation

An anomaly or fault can be defined as an anomalous behavior that causes a system to deviate from its normal operating conditions or states in an unacceptable way [5].

Two types of methods are used in root cause analysis (RCA): knowledge-based models and data-driven models. Knowledge-based methods try to formalize the extensive knowledge of the domain expert into a set of rules [4]. Data-driven models, also known as data mining methods, can extract unknown useful information from a very large volume of data [6].

The proposed pipeline for fault classification is shown in Fig. 1. Using labelled field data, supervised learning models are trained to classify the failure types observed in lightpaths (circuits) carried in an optical network. The building blocks are described in the following sections.

Data collection and pre-processing

The knowledge base (KB) in this study is

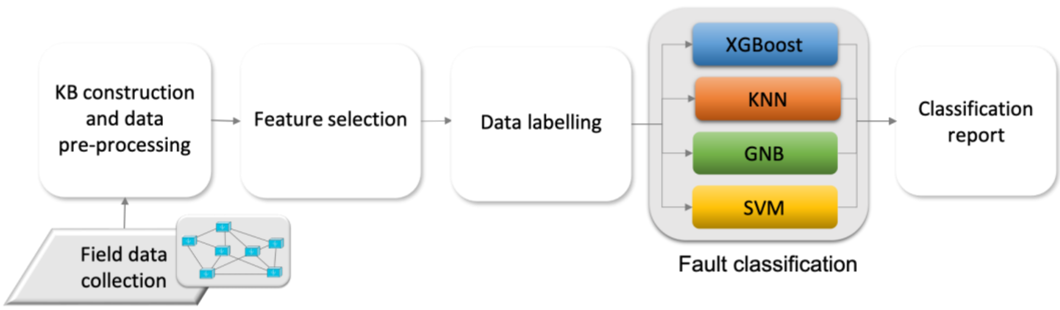
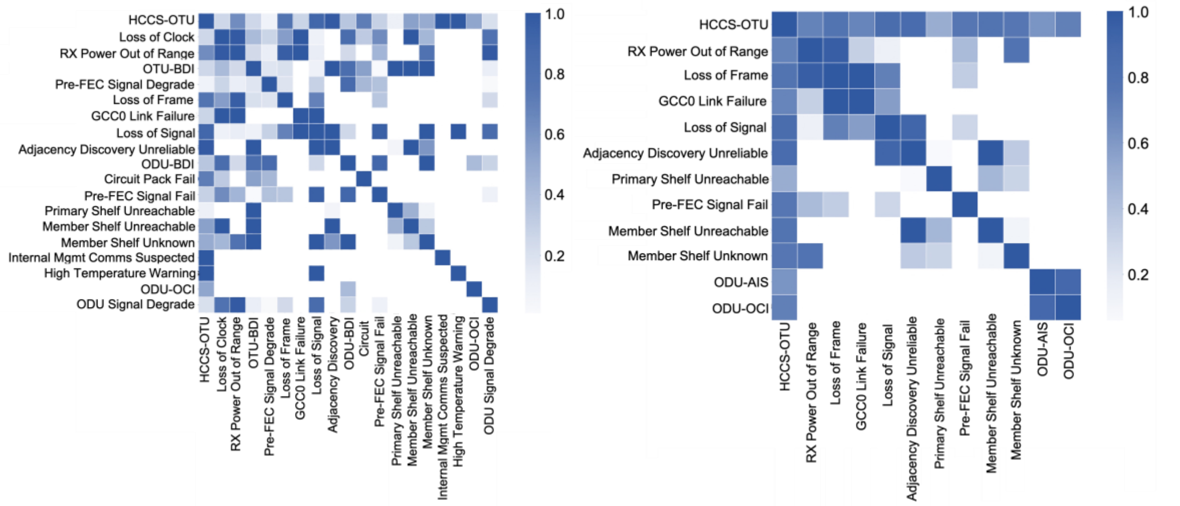


Fig. 1: Proposed pipeline for fault classification. XGBoost: Extreme Gradient Boosting; KNN: K-Nearest Neighbors; GNB: Gaussian Naive Bayes; SVM: Support Vector Machine



(a) **(b)**

Fig. 2: Correlation analysis: (a) TX alarms against HCCS; (b) RX alarms against HCCS. OTU: Optical Transport Unit; GCC0: General Communications Channel 0; ODU: Optical Channel Data Unit; OCI: Optical Channel Interface; AIS: Alarm Indication Signal.

composed of field data collected at 15-min sampling rate in a production network for 7 months. Two sources of information were utilized to build the dataset: performance monitoring (PM) and alarm information. Topology stitching was performed on the raw data collected for 348 optical tributary signals. Using only the receiver's PMs as input, appropriate failure classification tasks can be defined to train ML models to classify the failure types of the circuits.

Feature selection and data labelling

Feature selection was performed using the Pearson correlation measure to find the relationship between the network alarms and the target variable. The high correction count seconds (HCCS) performance metric at the receiver was considered the target variable, as its value specifies a circuit's health status. Fig. 2 and Fig. 3 show the correlation analysis for the transmitter (TX) and receiver (RX) alarms, and path alarms, respectively.

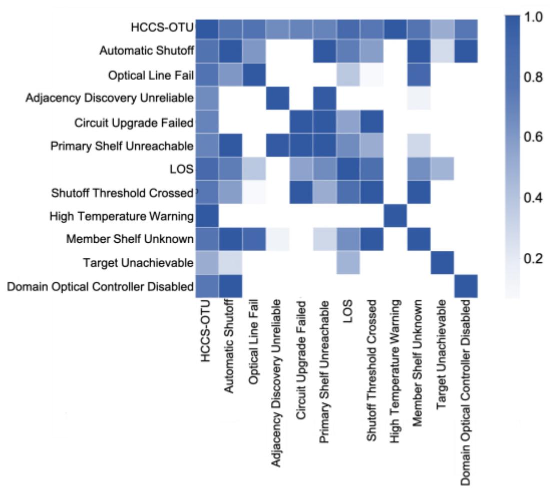


Fig. 3: Correlation analysis of path alarms against HCCS.

A rule-based method is used in the labelling process of our data. We considered a circuit faulty when it registers HCCS. The labelling mechanism was determined using domain knowledge on PMs and the relationship between the alarm against HCCS Pearson correlation graphs shown in Figures 2-3. According to the domain knowledge, four alarm types raised by network equipment, which are highly correlated to the HCCS value, were considered for the data labelling: Optical line fail (OLF), Automatic shutoff, Loss of signal (LOS), Backward Defect Identifier (BDI). During data labelling, a governing assumption was that the first power drop along the circuit is the root cause, with some exceptions imposed by domain experts.

The labels that are correlated with one another are further clustered into five sub-categories which constitute the final class of data for classification task: loss of signal at RX (LOS RX), loss of signal in the path (LOS Path), power drop in the link (Power drop), transient, and "Other". The label "transient" was manually created for HCCS=1 cases without relevant alarm raised. The label "Other" indicates cases where the root cause is not found. Table 1 summarizes the resulting labelled dataset.

Table 1: Train and test datasets for multiclass classification.

Label	Number of samples	
	Training (70%)	Test (30%)
LOS RX	1,977	780
LOS Path	625	266
Power drop	7,040	3,055
Transient	47	25
Other	2,061	911
Total	11,750	5,037

Table 2: Classification reports of the multiclass fault classifiers

Failure Type	XGBoost			KNN			GNB			SVM		
	Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
LOS Rx	0.53	0.14	0.22	0.44	0.35	0.39	0.40	0.73	0.52	0.33	0.72	0.46
LOS Path	0.36	0.02	0.03	0.20	0.14	0.17	0.17	0.12	0.14	0.09	0.56	0.16
Power Drop	0.73	0.85	0.78	0.72	0.76	0.74	0.85	0.13	0.22	0.69	0.33	0.45
Other	0.58	0.80	0.67	0.54	0.58	0.56	0.30	0.96	0.46	0.36	0.06	0.10

The 3044 (18%) samples of failures that could not be identified were mostly not power related or they were not reported by the equipment PMs. These failures could be polarization-related issues, non-power-related SNR degradation and other non-power related transients.

Multiclass fault classifiers

Four ML models, namely, Extreme Gradient Boosting (XGBoost), K-Nearest Neighbors (KNN), Gaussian Naive Bayes (GNB), and support vector machines (SVMs), were implemented for the multiclass classification task.

The 16,787 labelled samples are forwarded to ML models using a train-test split ratio of 70/30 as shown in Table 1. The class "transient" was excluded due to lack of samples for training, leaving only four classes.

The training dataset was used to train the models and to determine the appropriate hyperparameters to be used in the models. The grid search method was applied to find the best performing combination of hyperparameters. Using the same methodology as in [3], the hyperparameters corresponding to learning rate, number of estimators, max depth, as well as gamma and C values, were optimized by testing over different values (0.01, 0.02, 0.03, 0.05, 0.10), (100, 300, 500,1000), (1, 3, 5, 7), (1×10^{-3}, 1×10^{-4}) and (0.1, 1, 10), respectively.

Results

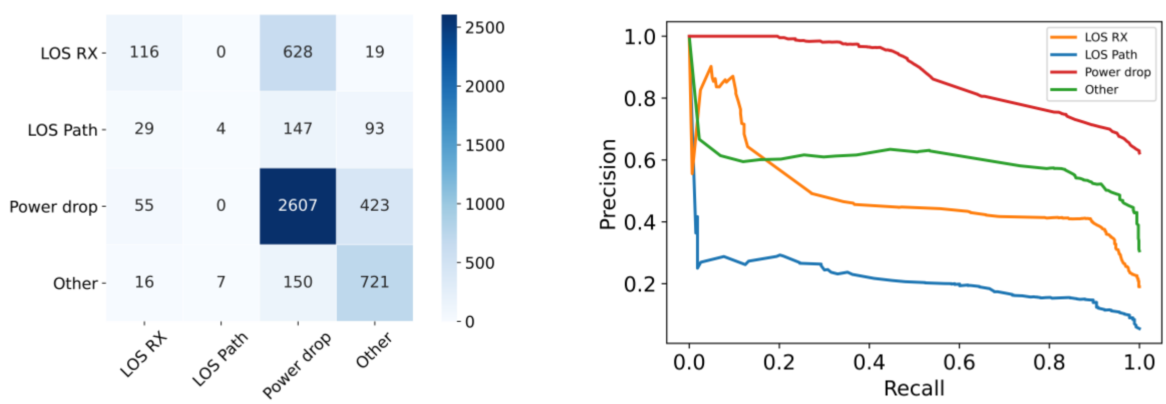
Table 2 summarizes the main classification metrics of each classifier for each class of data. The results in Table 2 demonstrate that all the classifiers except GNB perform well in distinguishing the "power loss" class, whereas all models have poor results on "LOS path" class. From the normalized F1-score measure which combines the precision and recall scores of a model, it can be seen that the XGBoost and KNN classifiers exhibit superior overall performance. GNB has the worst performance. KNN as a distance-based model has relatively similar performance to XGBoost but at the expense of a longer execution time. The confusion matrix and the precision-recall curve for the best performing XGBoost model are shown in Fig. 4.

Conclusions

In this work, we have shown how multiclass classifiers trained with field data can be used to identify circuit failure in optical networks. The XGBoost model outperformed the KNN, GNB and SVM models with a significantly lower misclassification rate in failure classes. The ultimate goal of this research is to address the fault management problem in optical networks by using the proposed XGBoost-based failure identification method in concert with failure localization methods.

Acknowledgements

This research was supported by Mitacs under grant IT16352 and by Ciena Corp.



(a) (b)

Fig. 4: XGBoost model: (a) Confusion matrix; (b) Precision-recall curve.

References

[1] D. Wang, C. Zhang, W. Chen, H. Yang, M. Zhang, and A P. T. Lau, "A review of machine learning-based failure management in optical networks," *Sci. China Inf. Sci.*, vol. 65, no. 211302, pp. 1-19, November 2022, doi: 10.1007/s11432-022-3557-9.

[2] F. Musumeci, C. Rottondi, G. Corani, S. Shahkarami, F. Cugini, M. Tornatore, “A Tutorial on Machine Learning for Failure Management in Optical Networks,” *Journal of Lightwave Technology,* vol. 37, no. 16, pp. 4125–4139, August 2019, doi: 10.1109/JLT.2019.2922586.

[3] D. Rafique, T. Szyrkowiec, H. Grieser, A. Autenrieth, and J.-P. Elbers, “Cognitive Assurance Architecture for Optical Network Fault Management,” *Journal of Lightwave Technology*, vol. 36, no. 7, pp. 1443–1450, April 2018, doi: 10.1109/JLT.2017.2781540.

[4] D. Rafique, T. Szyrkowiec, A. Autenrieth, and J.-P. Elbers, “Analytics-Driven Fault Discovery and Diagnosis for Cognitive Root Cause Analysis,” *2018 Optical Fiber Communications Conference (OFC)*, 2018, paper W4F.6.

[5] C. Natalino, A. di Giglio, M. Schiano, and M. Furdek, “Root Cause Analysis for Autonomous Optical Networks: A Physical Layer Security Use Case,” *2020 European Conference on Optical Communications (ECOC)*, Brussels, Belgium, 2020, pp. 1–4, doi: 10.1109/ECOC48923.2020.9333141.

[6] M. Riazi, O. Zaïane, T. Takeuchi, A. Maltais, J. Günther, and M. Lipsett. (2019). “Detecting the Onset of Machine Failure Using Anomaly Detection Methods,” Ordonez, C., Song, IY., Anderst-Kotsis, G., Tjoa, A., Khalil, I. (eds) Big Data Analytics and Knowledge Discovery. DaWaK 2019. *Lecture Notes in Computer Science*, vol. 11708. Springer, Cham, doi:10.1007/978-3-030-27520-4_1.

[7] M. Ruiz, F. Fresi, A. P. Vela, G. Meloni, N. Sambo, F. Cugini, L. Poti, L. Velasco, and P. Castoldi, M. Ruiz et al., "Service-triggered Failure Identification/Localization Through Monitoring of Multiple Parameters," 42nd European Conference on Optical Communication (ECOC), Dusseldorf, Germany, 2016, pp. 1-3.

[8] A. P. Vela *et al.*, “BER Degradation Detection and Failure Identification in Elastic Optical Networks,” *Journal of Lightwave Technology*, vol. 35, no. 21, pp. 4595–4604, November 2017, doi: 10.1109/JLT.2017.2747223.

[9] F. N. Khan, C. Lu, and A. P. T. Lau, “Optical Performance Monitoring in Fiber-Optic Networks Enabled by Machine Learning Techniques,” *Optical Fiber Communication Conference*, San Diego, California, 2018, paper M2F.3, doi: 10.1364/OFC.2018.M2F.3.

[10] S. Shahkarami, F. Musumeci, F. Cugini, M. Tornatore, “Machine-Learning-Based Soft-Failure Detection and Identification in Optical Networks,” *2018 Optical Fiber Communication Conference*, San Diego, California, 2018, paper M3A.5, doi: 10.1364/OFC.2018.M3A.5.

[11] C. Zhang et al., "Interpretable Learning Algorithm Based on XGBoost for Fault Prediction in Optical Network," *2020 Optical Fiber Communications Conference (OFC)*, San Diego, California, 2020, paper Th1F.3.

[12] W. Du, D. Côté, C. Barber, and Y. Liu, “Forecasting loss of signal in optical networks with machine learning,” *J. Opt. Commun. Netw.*, vol. 13, no. 10, pp. E109-E121, October 2021, doi: 10.1364/JOCN.423667.

[13] F. Musumeci, V. V. Garbhapu, Y. Hirota, Y. Awaji, S. Xu, M. Shiraiwa, B. Mukherjee, and M. Tornatore, "Domain adaptation and transfer learning for failure detection and failure-cause identification in optical networks across different lightpaths [Invited]," *J. Opt. Commun. Netw.*, vol. 14, no. 2, pp. A91-A100, February 2022, doi: 10.1364/JOCN.438269.

[14] C. Zhang, D. Wang, J. Jia, L. Wang, K. Chen, L. Guan, Z. Liu, Z. Zhang, X. Chen, and M. Zhang, "Potential failure cause identification for optical networks using deep learning with an attention mechanism," *J. Opt. Commun. Netw.,* vol. 14, no. 2, pp. A122-A133, February 2022, doi: 10.1364/JOCN.438900.

[15] S. Allogba, B. L. M. Yaméogo and C. Tremblay, "Extraction and Early Detection of Anomalies in Lightpath SNR Using Machine Learning Models," *Journal of Lightwave Technology*, vol. 40, no. 7, pp. 1864-1872, April 2022, doi: 10.1109/JLT.2021.3134098.

[16] Y. Pei, D. Côté, P. A. Ngani Sigue, A. Mahmoudialami, C. Tremblay, and C. Desrosiers, "Identifying root causes of network service degradation," U.S. Patent 11,477,070, October 18, 2022.