Explainable Machine Learning-enabled Just-enough Margin Configurations in Dynamic S+C+L-band Optical Networks

Zeyuan Yang, Rentao Gu*, Yuefeng Ji
Beijing Laboratory of Advanced Information Networks, Beijing University of Posts and Telecommunications, Beijing, China
*Email: rentaogu@bupt.edu.cn *Corresponding author*

Abstract: We configure an explainable machine learning-enabled just-enough margin for each lightpath in dynamic S+C+L-band optical networks. Explainable decisions improve blocking ratio and spectral efficiency performances by 35.9% and 17.2% compared with the benchmarks.

1. Introduction

In the era of F5G, driving down the cost per transported bit is critical, which is realized by reducing network capital expenditure (CAPEX) and operating expense (OPEX). Multiband optical network, extending the spectrum by 5~10 times, is envisioned as a competitive near-term network upgrade trend with less CAPEX [1]. In network operations, although the excess signal-to-noise ratio (SNR) margin is imperative to maintain error-free communication under SNR fluctuation, the rough margin configurations obstruct fully utilizing network capacity which harms the OPEX [2]. Current approaches either configure a static margin, e.g., 2~5dB [3] or assume full-filled channel loads for SNR estimations in lightpath operations [4]. However, these methods lose efficiency in the S+C+L-band networks. The nonuniform inter-channel stimulated Raman scattering (ISRS) effect and the large spectrum range make it significantly different for transmission effects in different channels, and thus, applying static margin results in insufficient margin for some lightpaths but too much for others. Meanwhile, filling all the unoccupied channels in a wideband network with amplified spontaneous emission (ASE) loads will severely overestimate the nonlinear interferences (NLI), resulting in unaffordable degradations in spectral efficiency [4]. Therefore, it is indispensable to configure just-enough margins for lightpaths in S+C+L-band networks, preventing transmission error induced by SNR degradation as well as preventing NLI overestimation to improve spectral efficiency.

In this paper, we make the first attempt to configure a just-enough margin for each lightpath in dynamic S+C+L-band optical networks. Faced with challenges of complex couplings among multiple factors, including ISRS effect, network topology, traffic mode, etc., machine learning (ML)-enabled end-to-end SNR degradation prediction is conducted to infer the just-enough margin for each lightpath. The blocking ratio and spectral efficiency performances are improved by 35.9% and 17.2% compared with the benchmarks. Furthermore, to realize white/gray-box ML-enabled margin configurations, we utilize an explainable machine learning technique, i.e., SHapley Additive exPlanations (SHAP), to peek inside latent knowledge between multiple factors and SNR degradations. The training time of the ML model is reduced by 36.8% with SHAP-based feature selection. Note that although the traffic dynamicity is considered to affect the SNR in this paper, the large-timescale effect, e.g., device aging, can be integrated into our proposed XAI-enabled method when actual data is achieved [5].

2. Explainable AI for Just-enough Margin Configurations

Problem Statement: Fig. 1 (a) shows an optical line system in S+C+L-band optical networks. We assume the generalized SNR (GSNR) which considers the NLI noise and ASE noise. In S+C+L-band networks, the transmission characteristics are different for lightpaths in different channels. the ISRS effect induces power transfer from high-frequency channels to the lower ones. The nonlinear coefficient (γ), attenuation factors, and noise figures (NF) of amplifiers are nonuniform for different channels [1]. Moreover, considering the traffic dynamicity, represented in nonuniform traffic load and network topology characteristics, such as node/edge betweenness, the adjacent traffic when the network is steady will also be different for different lightpaths.

Under the different transmission effects, traditional margin configurations are no longer efficient. As is shown in Fig. 1 (b), if considering a static margin, e.g., 1dB, is subtracted from the beginning of life (BoL) GSNR for all

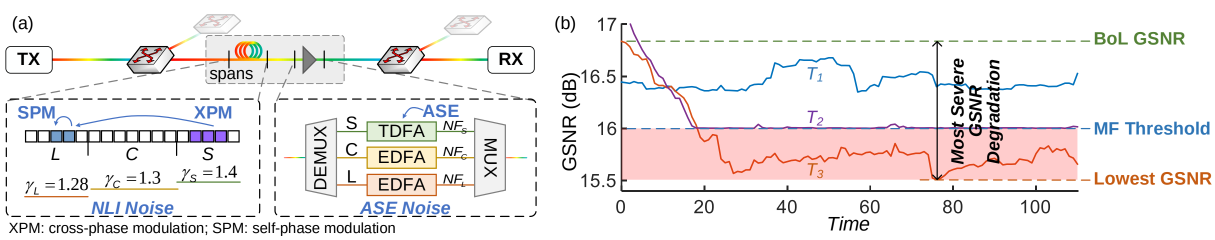


Fig. 1. (a) Different transmission effects in different bands; (b) Illustration of GSNR degradation and just-enough margin configuration.

lightpaths, T_1 will select a lower order modulation format (MF) which wastes the capacity, while T_2 will suffer an insufficient GSNR in its duration. The T_2 will either suffer from transmission interruption or prohibit adding new traffic in its occupied links, which also wastes the potential capacity. Therefore, the just-enough margin should be configured adaptively for each lightpath. As shown in T_3, the most severe GSNR degradation during a lightpath's lifetime can be set as the just-enough margin for this lightpath. Then, just-enough margin configuration is converted to estimate the most severe GSNR degradation for a lightpath. As the GSNR degradations differ significantly among lightpaths, we propose ML-enabled methods to learn the relationships between multiple factors and the GSNR degradations. The framework of the just-enough margin configuration scheme is shown in Fig. 2 (a).

Data Collection: The GSNR degradation estimation is modeled as a regression task. For the *i*th traffic, a lightpath is deployed in the network. Feature $\boldsymbol{x}_i$ and GSNR G^i_{BoL} at BoL are obtained. Initializing $G^i_{\mathrm{Low}} = G^i_{\mathrm{BoL}}$ as the lowest GSNR during the lifetime of the lightpath. At each time point, we recompute GSNR for lightpaths in the network. If the current GSNR of the *i*th traffic is lower than G^i_{Low}, the G^i_{Low} is updated as this GSNR value. When the *i*th lightpath is released, a sample $\{\boldsymbol{x}_i, y_i\}$ is obtained, where $y_i = G^i_{\mathrm{BoL}} - G^i_{\mathrm{Low}}$ is the most severe GSNR degradation. The features are shown in Fig. 2 (b). Notations $(\cdot)_{\min}$, $(\cdot)_{\max}$, $(\cdot)_{\mathrm{mean}}$, $(\cdot)_{\mathrm{med}}$, and $(\cdot)_{\mathrm{std}}$ are the min, max, mean, median, and standard deviation of variable $\cdot$. The LSR, LT, EB, and ND stand for the spectrum utilization of the links, traffic numbers in the links, edge betweenness of the links, and the degrees of the nodes in the lightpaths.

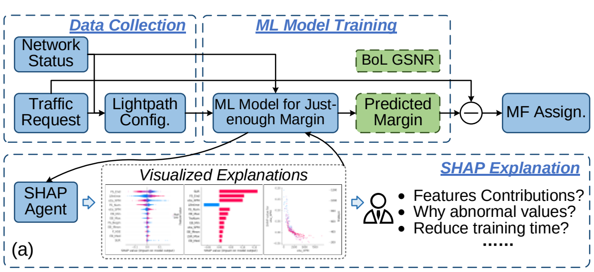
Model Training: The features are normalized with $\hat{x}^d_i = \left(x^d_i - \bar{x}^d\right)/\sigma^d$, where x^d_i is the *d*th feature of the *i*th sample, $\bar{x}^d$ and σ^d are the mean and standard deviation of the *d*th feature. The dataset is divided into training set, validation set, and testing set at the ratio of 70%:15%:15%. We take three typical ML models for GSNR degradation estimation, including random forests (RF), deep neural networks (DNN), and XGBoost [6]. We adopt mean square error (MSE) and coefficient of determination as the evaluation metrics for regression performance. The MSE between the estimated GSNR degradation and the target is in the form of $\mathrm{MSE} = (1/n)\cdot\sum_i \left(\hat{y}_i - y_i\right)^2$, where $\hat{y}_i$ is the ML model output for *i*th sample. Furthermore, the coefficient of determination R^2 is used to evaluate the performance of the regression, which is defined as $R^2 = \sum_i \left(\hat{y}_i - \bar{y}\right)^2 / \sum_i \left(y_i - \bar{y}\right)^2$, where $\bar{y}$ is the mean value of the target output. The more the R^2 approaches 1, the more accurate the regression model is.

Model Explanation: Although ML models efficiently fit the GSNR degradations with input features, their black-box nature prevents operators from explicitly knowing how the output is produced, which hinders the deployment of ML models in the optical networks [7]. Therefore, the model explanation is imperative for white-box just-enough margin configurations. This work adopts SHAP to explain the prediction of the ML model by computing the Shapley value of each feature in a sample, which indicates the contribution of such feature to the model prediction. The following questions are aimed to be addressed: *i) What are the most important features? ii) What are the reasons for the existence of abnormal/extreme values in GSNR degradation? and iii) Can feature contributions be utilized to reduce model training complexity?* Through exploring these questions, we make the just-enough margin decisions more transparent and use explanations in feature selection to reduce training complexity.

3. Performance Evaluations and Explanations Analyses

The DT and CORONET [8] topologies are adopted for evaluations. We consider an S+C+L-band elastic optical network, where spectrum ranges from 187.54 THz to 199.54 THz, with channel granularity of 12.5 GHz. The parameters of the G.652.D fiber, the modulation format thresholds, and the NFs of amplifiers are the same as [1]. The bandwidth request of traffic is between [50, 200] Gbps. Traffic arrival λ is set between 10 and 45, and average lifetime $1/\mu$ is set as 50. For each request, the lightpath is selected in a KSP-FF manner.

AI Model Effectiveness: A dataset that contains 5×10^5 samples are collected under different λ. The best hyperparameters of ML models are selected. The RF model contains 200 trees with a max depth of 10. DNN contains 4 hidden layers with each layer having 100 neurons. There are 100 trees with a max depth of 10 in XGBoost. Table 1 shows the performances of the ML models. Among three ML models, the XGBoost achieves lower training complexity, shorter inference time, and better fitting performance than RF and DNN. Therefore, the XGBoost is selected as the ML model to be evaluated in the following.



(b)

Notation	Description
Lifetime	Lifetime of the request
FS_Num	FS number occupied by the lightpath
FS_Begin/FS_End	The index of the first/last FS in the FS block
P_ASE	ASE noise power
eta_SPM/eta_XPM	NLI coefficient of SPM/XPM
Link_Num	Link number in the lightpath
Traffic_Num	Traffic number in the network at BoL
LP_Len	The length of the lightpath
GSNR_BOL	GSNR of the lightpath at the BoL
$LSR_{\min}, LSR_{\max}, LSR_{\mathrm{mean}}, LSR_{\mathrm{med}}, LSR_{\mathrm{std}}$	$LT_{\min}, LT_{\max}, LT_{\mathrm{mean}}, LT_{\mathrm{med}}, LT_{\mathrm{std}}$
$EB_{\min}, EB_{\max}, EB_{\mathrm{mean}}, EB_{\mathrm{med}}, EB_{\mathrm{std}}$	$ND_{\min}, ND_{\max}, ND_{\mathrm{mean}}, ND_{\mathrm{med}}, ND_{\mathrm{std}}$

Fig. 2. (a) The framework of the explainable machine learning-enabled just-enough margin configurations. (b) Notations of the input features.

Table 1. Performances of ML models for just-enough margins.

	RF	DNN	XGBoost
Training (min)	40.72	120.35	25.05
Testing (ms)	0.053	0.042	0.012
MSE	4.25×10^{-3}	2.63×10^{-3}	1.72×10^{-3}
R^2	0.697	0.812	0.865

Network Performances: We evaluate the blocking ratio and spectral efficiency with benchmarks include static margins [1, 2, 3] dB, and full-filled. The spectral efficiency is presented by the average modulation order of the established lightpaths. As depicted in Fig. 3 (a)-(d), the just-enough margin outperforms the benchmarks in blocking ratio by 35.9% on average. Compared with 2-3dB margin and full-filled, the proposed method overcomes the overestimation of the GSNR degradation and improves the spectral efficiency by 17.2% on average. Compared with 1dB margin, the spectral efficiency is similar but the blocking ratio is still reduced. It is because the just-enough margin is adaptive to each individual lightpath.

Explanations Analyses: The results of applying the SHAP method on the XGBoost model are depicted in Fig. 3 (e)-(g). We can retrieve the significant features from Fig. 3 (e). The lifetime of traffic positively correlates to the margin, which is because with a longer lifetime, there tends to be more adjacent traffics and thus, a more severe XPM interference. The results also reveal some hidden information, such as the FS number having a negative correlation to the margin. It is because the XPM effect is more severe when two frequency is near, but if a lightpath has a larger range of FS block, the center frequency of two FS blocks will be farther, and thus, a weaker XPM effect.

We further explore the explanations of the extreme outputs. The contributions of the features in a maximum/minimum GSNR degradation are shown in Fig. 3 (f)-(g). The percentile of each feature is marked beside each bar. Take Fig. 3 (f) as an example, the lifetime is relatively high while the eta_XPM is low. Low eta_XPM means there are many idle FSs around this lightpath and thus, more requests can be provisioned around these lightpaths. The significant features are usually at the extreme value point and this inspires the network operators to pay more attention to the lightpaths with features at extreme points, because their GSNR degradation may be severe.

Utilizing Explanations for Complexity Reduction: As shown in Fig. 3 (e), the contributions of some features are relatively low, then we further conduct a feature selection following the contributions provided by SHAP. The features are added in the order of their average contributions. As depicted in Fig. 3 (h), the model accuracy increases with adding features and approaches the original model using only the 15 most significant features. The model training time is then reduced by 36.8%. These results demonstrate that the model explanation is effective in reducing the complexity of training with negligible loss in model performance.

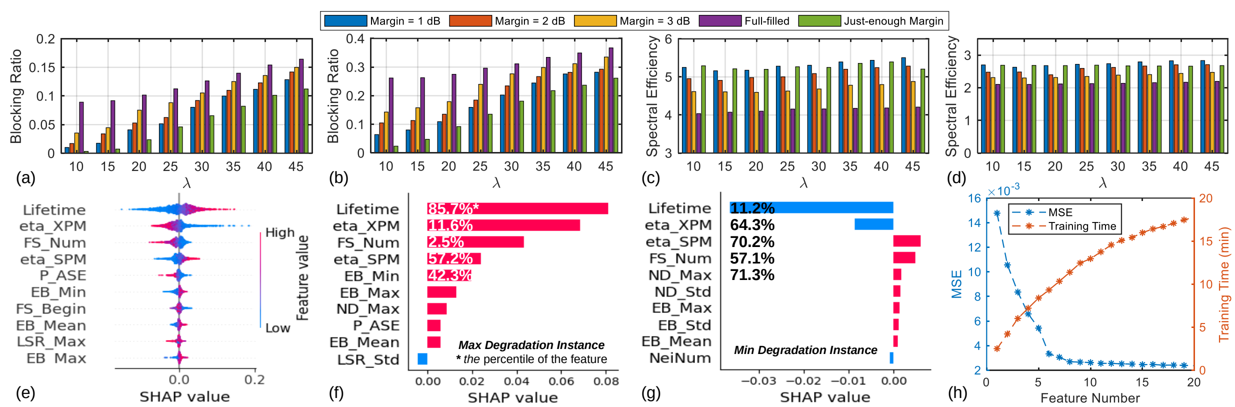


Fig. 3. (a)-(b) The blocking ratio under DT and CORONET. (c)-(d) The spectral efficiency under DT and CORONET. (e) Features contributions. (f)-(g) Features contributions on the maximum/minimum GSNR degradation instance. (h) Performance of XGBoost with feature selection.

4. Conclusions

We proposed an ML-enabled just-enough margin configuration scheme in multiband optical networks. The performances on blocking ratio and spectral efficiency are improved compared with the benchmarks. Furthermore, the explanations on black-box models are investigated for transparent margin configurations and efficient training.

Acknowledgement

This work was supported by National Natural Science Foundation of China (U21B2005), P. R. China.

References

[1] A. Ferrari *et al.*, "Assessment on the Achievable Throughput of Multi-Band ITU-T G.652.D Fiber Transmission Systems," in JLT, 2020.
[2] Y. Pointurier, "Design of low-margin optical networks," in JOCN, 2017
[3] A. Zhang *et al.*, "Field trial of 24-Tb/s (60×400Gb/s) DWDM transmission over a 1910-km G.654.E fiber……," in Optics Express, 2021.
[4] M. Filer *et al.*, "Low-margin optical networking at cloud scale," in JOCN, 2019.
[5] A. W. MacKay *et al.*, "Field Learnings of Deploying Model Assisted Network Feedback Systems," in OFC, 2022.
[6] T. Chen *et al.*, "XGBoost: A scalable tree boosting system," in SIGKDD, 2016.
[7] S. M. Lundberg *et al.*, "A unified approach to interpreting model predictions," in NeurIPS, 2017.
[8] R. Gour *et al.*, "Availability-guaranteed slice composition for service function chains in 5G transport networks," in JOCN, 2021.