QoT Estimation for Unestablished Lighpaths using Machine Learning

Luca Barletta[1], Alessandro Giusti[2], Cristina Rottondi[2], and Massimo Tornatore[1]

[1]*Dept. of Electronics, Information, and Bioengineering, Politecnico di Milano, Italy*
[2]*Dalle Molle Institute for Artificial Intelligence, Lugano, Switzerland*
{*luca.barletta,massimo.tornatore*}*@polimi.it,*{*cristina.rottondi,alessandro.giusti*}*@supsi.ch*

Abstract: We investigate a machine-learning technique that predicts whether the bit-error-rate of unestablished lightpaths meets the required threshold based on traffic volume, desired route and modulation format. The system is trained and tested on synthetic data.

OCIS codes: 060.4250 Fiber optics and optical communications, Networks; 060.4510 Fiber optics and optical communications, Optical communications

1. Introduction

Estimating the lightpath Quality of Transmission (QoT) prior to deployment is essential for an optimized design and planning of optical networks. Sophisticated analytical models capturing different physical layer impairments often impose too high computational requirements for real-time prediction and are thus not scalable to large network topologies and more dynamic network operations. Moreover, nowadays a wide range of design parameters are available to system and network engineers (e.g. Forward Error Correction (FEC) coding, single/multicarrier transmission, nonlinearity mitigation solutions, adaptive channel spacings, and "flex-grid" network technologies), thus making the number of possible combinations for lighpath route, spectrum, modulation format and baud rate assignment grow dramatically.

On the other hand, approximated formulas introduce higher marginations in the calculation of the ligthpath budget to compensate for model inaccuracies, thus leading to an underutilization of network resources. Therefore, an alternative approach to QoT estimation relies on sensing the QoT of already deployed lightpaths by means of optical performance monitors (OPMs) [1] installed at the receiver side and on exploiting the knowledge extracted from field data to predict the QoT of unestablished lightpaths. To this aim, different data mining techniques can be applied, ranging from network kriging [2] to machine learning (ML) [3]. In this paper, we apply a ML-based classifier to predict the probability that the Bit-Error-Rate (BER) of a candidate lightpath will not exceed the system tolerance threshold, using as features the traffic volume to be served, the modulation format, the lightpath total length, the length of its longest link and the number of lighpath links. To train the classifier, we assume that either BER measurements over already deployed lightpaths are provided by field OPMs[1] or that, in absence of real field data, a BER Estimation Tool (ETool) is used to generate synthetic data. In the remainder of this paper we opt for the latter approach due to the difficulty of retrieving field data. The classification output is meant to be provided to a Routing and Spectrum Assignment (RSA) algorithm that will take the final deployment decision.

The remainder of the paper is structured as follows: Section 2 describes our ETool, Section 3 describes the proposed ML binary classifier, and Section 4 discusses results and conclusions.

2. Bit Error Rate Estimation Tool for Synthetic Data Generation

For the generation of synthetic data we propose an ETool that, on input of a candidate lightpath and modulation format, calculates the BER as a function of the signal-to-noise ratio (SNR) measured at the input of the channel decoder, assuming that the cascade of channel and detection/non-linear compensation stage at the receiver is statistically equivalent to an AWGN channel. Hence, once the forward error correction (FEC) threshold T is fixed, we can compute the required SNR. For a specific lightpath, the pre-FEC SNR can be estimated by a link budget that takes into account the transmit power P_{in}, gains, and losses. If the pre-FEC SNR exceeds the required SNR, then the lightpath can be established. We consider the approximated AWGN model of a dispersion uncompensated (DU) transmission over standard single mode fiber, a flexible grid with standard slice width of 12.5 GHz and elastic transceivers operating at 28 Gbaud with optical bandwidth of 37.5 GHz (i.e., 3 slices). Superchannels with multiple adjacent transceivers are used to serve

[1]The proposed classifier is agnostic to the number of input features and could rely on multiple field-measured parameters, if available.

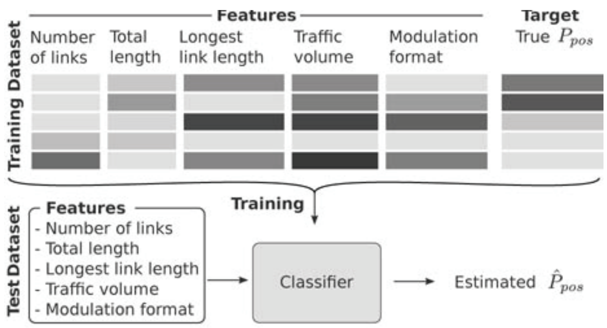

Fig. 1: The classifier structure

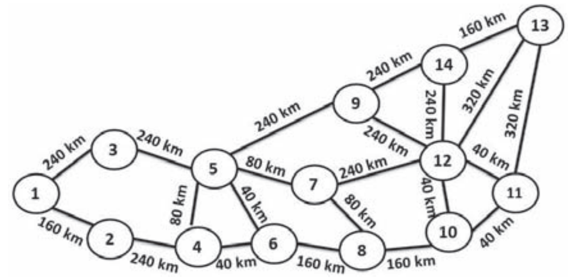

Fig. 2: Japan network topology

traffic demands exceeding the capacity of a single transceiver. The gains are provided by identical optical amplifiers equally spaced over the links (100 km), with gain $G = 20\,\text{dB}$ and noise figure $F = 5\,\text{dB}$.

The considered losses are: back-to-back penalties for each modulation format in a 37.5 GHz grid network, whose values are extrapolated from [4, Table I] [5, Fig. 7]; a penalty of 2 dB for any modulation format due to non-linear propagation phenomena, evaluated in correspondence of the optimum input power P_{in} [4, Fig. 8] (note that for DU systems P_{in} is virtually independent of the modulation format and can be estimated from [4, Fig. 9]); a system margin which is a random parameter uniformly drawn in the range $[0,2]$ dB. The randomization of the latter parameter accounts for the unpredictability of fast time-varying penalties (such as polarization effects [6]).

3. Classifier Description, Dataset Generation and Performance Metrics

Our proposed classifier considers the following features: *i*) the number of links of the lightpath; *ii*) the lightpath total length; *iii*) the length of its longest link; *iv*) the traffic volume it serves; *v*) the modulation format. The target variable that the classifier tries to predict is a binary variable, which is True if and only if the ligthpath BER is lower than the system threshold $T = 4 \cdot 10^{-3}$, chosen as in [4]. Note that the BER value is affected by other factors than those captured by the classification features (e.g., time-varying penalties). Therefore, it may occur that two dataset entries whose set of features are exactly the same exhibit different BER values and in turn different values of the target variable, i.e. the association between feature values and BER value is not deterministic. The classifier is trained on a *training dataset* and quantitatively evaluated on a separate and disjoint *testing dataset*, as depicted in Fig. 1.

We generate the *training set* by repeating the following procedure L times, where L defines the training test size (i.e., the number of instances it contains): we randomly select a source-destination node pair within the Japan network topology (see Fig. 2) and associate a traffic request uniformly chosen in the range $[50, 1000]$ Gbps with 50 Gbps granularity. In the following, we name each triplet of source node, destination node and traffic volume as a "scenario". For each scenario, we randomly select a route within a set of 3 candidate paths (including the shortest path), and one out of 6 possible modulation formats (i.e. dual polarization (DP)-BPSK, DP-QPSK and DP-n-QAM, with $n = 8, 16, 32, 64$). Then, we evaluate the BER by means of the ETool, whose output is considered as the *ground truth*. For our experiments we set $L = 90000$ (but in Section 4 we perform a sensitivity analysis on L). Based on the results obtained with the ETool, our training dataset contains 59111 instances with BER$\geq T$ (i.e. the class of *negative* instances) and 30889 instances with BER$< T$ (i.e. the class of *positive* instances).

Conversely, we generate the *testing dataset* by randomly selecting 50 scenarios (i.e. triplets) and for each scenario we consider *all* the 18 possible combinations of 3 routes and 6 modulation formats. We name the combination of scenario, route and modulation format as a "setting" (for a total of 900 settings). For each setting, we repeat the BER calculation 100 times to obtain a statistical estimation of the *true* probability, P_{pos}, that BER$< T$. The testing dataset contains 56183 negative and 33817 positive instances.

Since the classifier requires feature values that are numeric and that have comparable ranges to avoid numerical instability, we pre-process features as follows: *i*) the modulation format feature is replaced by 6 binary values (one for each possible format) using one-hot encoding; *ii*) the values of each feature are offset and rescaled to ensure that their distribution in the whole training set has mean 0 and standard deviation 1.

On input of a test instance, the output produced by our classifier is the estimated probability $\widehat{P}_{pos}$ that the instance belongs to the positive class. The instance is then classified as positive iff such probability is greater than or equal to a threshold $\gamma = 0.5$. We assess the classifier performance by evaluating the following two metrics: *i*) the *accuracy*, i.e. the fraction of the testing instances whose class is correctly predicted. Note that the value of the accuracy depends on

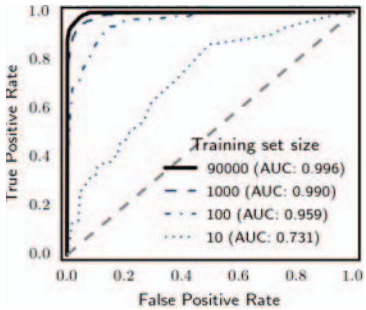

Fig. 3: ROC curve and AUC for different training set sizes

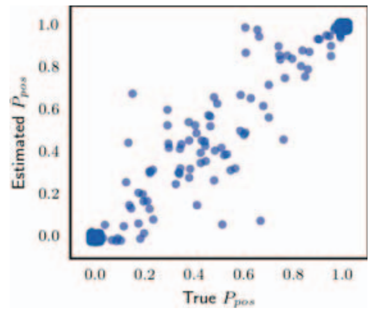

Fig. 4: Scatter plot of predicted values vs. ground truth

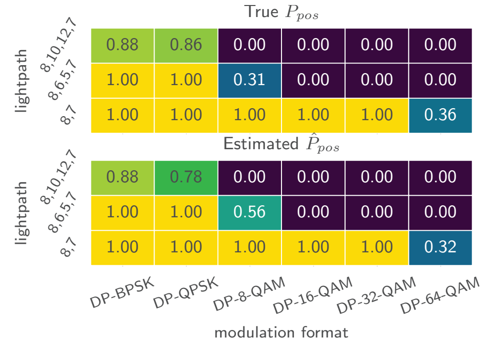

Fig. 5: Example of classification output for a traffic request of 600 Gbps from node 8 to 7.

the choice of γ. Also note that a dummy classifier that ignores the feature values and always predicts the most frequent class has an accuracy equal to the prevalence of such class in the testing dataset. *ii*) the *Area Under the ROC Curve* (AUC) [7], which is not affected by class unbalance and does not depend on the choice of γ. The AUC may vary within the range $[0.5, 1]$, where 0.5 is the performance of a dummy classifier and 1 is the performance of a classifier that always predicts the correct class.

4. Results and Conclusions

For our experiments we use a Random Forest (RF) classifier with 100 estimators [7], as it is one of the most widely used in the applied machine learning literature (as a preliminary step of our analysis, we have compared the performance of 5 RF classifiers and 3 k-Nearest-Neighbor classifiers: the RF with 100 estimators provided the best trade-off between performance and computational time). We first evaluate the impact of the training set size L: Fig. 3 shows the ROC curves and AUC values obtained on testing data when the classifier is trained on subsets of the training set composed by $L = 10, 100$, and 1000 randomly-sampled instances (such amounts are realistic if the dataset consists of field data, depending on the network size), as well as the ROC curve of the classifier trained on the whole training dataset. Results obtained with $L = 1000$ closely approach those obtained using the whole training dataset.

We now consider the classifier trained over the whole training dataset. For each setting in the testing dataset, in Fig. 4 we compare the true P_{pos} (y axis) to the estimated $\widehat{P}_{pos}$ (x axis); each point represents one setting. The points are jittered by 0.02 units on each axis to help resolving overlapping points. Note that most points lie at either $(0,0)$ (e.g., settings with all negative instances, classified with $\widehat{P}_{pos} = 0$) or $(1,1)$ (e.g., settings with all positive instances, classified with $\widehat{P}_{pos} = 1$). The remaining points identify settings with both positive and negative instances: they scatter close to the diagonal, thus showing that the discrepancy between their respective P_{pos} and $\widehat{P}_{pos}$ is limited.

Results for each scenario can be arranged in a table, where each cell corresponds to one setting. As an example, Fig. 5 compares the true P_{pos} of each setting (top) to the corresponding estimated $\widehat{P}_{pos}$ (bottom) for a traffic request of 600 Gbps from node 8 to node 7. Since the $\widehat{P}_{pos}$ of a given setting indicates the estimated probability that, when transmitting over the lightpath with the modulation format indicated by the considered setting, the BER will not exceed the threshold T, the closer $\widehat{P}_{pos}$ approaches 1, the safer would be the choice of that setting from a network design perspective. This output can be exploited to take the final decision about the deployment of the new lightpath by any RSA method. Based on the reported results, our proposed classifier can be considered a useful component to be integrated in RSA decision tools.

References

1. Christodoulopoulos, K., et al. "ORCHESTRA-Optical performance monitoring enabling flexible networking," Proc. ICTON 2015.
2. Sambo, N. et al. "Lightpath establishment assisted by offline QoT estimation in transparent optical networks," Journal of Optical Communications and Networking 2.11 (2010): 928-937.
3. Jimenez, T. et al. "A cognitive quality of transmission estimator for core optical networks," Journal of Lightwave Technology 31.6 (2013): 942-951.
4. G. Bosco et al., "On the performance of Nyquist-WDM Terabit Superchannels based on PM-BPSK, PM-QPSK, PM-8QAM or PM-16QAM subcarriers," Journal of Lightwave Technology 29.1 (2011): 53-61.
5. X. Zhou et al., "High spectral efficiency 400 Gb/s transmission using PDM time-domain hybrid 3264 QAM and training-assisted carrier recovery," Journal of Lightwave Technology 31.7 (2013): 999-1005.
6. Pointurier, Y. "Design of low-margin optical networks," Proc. OFC 2016.
7. Bishop, C. M. "Pattern recognition." Machine Learning 128 (2006).