Exact component parameter agnostic QoT estimation using spectral data-driven LSTM in optical networks

Lars E. Kruse*, **Sebastian Kühl and Stephan Pachnicke**
Chair of Communications, Kiel University, 24143 Kiel, Germany
**E-mail: lars.kruse@tf.uni-kiel.de*

Abstract: We propose the use of spectral data-driven LSTM-based machine learning to improve generalized signal-to-noise ratio (gSNR) quality-of-transmission estimation in component parameter-agnostic network scenarios. We show gSNR estimation improvements up to 1.1 dB for unseen networks. © 2021 The Author(s)

1. Introduction

Over the last decade, optical networks underwent enormous changes to fulfill the ever-growing demand for capacity. In this regard, the networks have become more sophisticated, configurable, and adaptable. New flexible add-drop multiplexers allow for a more versatile network operation, as well as the implementation of flexible frequency grids. Due to the significantly increased complexity of the networks, monitoring and optimizing the performance and the quality of transmission (QoT) is of increasing importance for maximizing the capacity and enabling possible self-management of the networks. While QoT estimation has been around for a long time in optical networks, the above-mentioned changes as well as the possibility of applying data-driven, autonomous artificial intelligence (AI) techniques renewed the interest in reliable QoT estimation.

Over the last years, many efforts have been put into the development of analytical [1], machine learning-based techniques [2,3], and hybrid approaches [4,5,6] for evaluating the performance of a certain lightpath in the system, based on its generalized-SNR (gSNR) [4]. All the mentioned works require exact knowledge of the necessary component parameters for QoT estimation. However, in practice, the network operator in many cases does not have an exact knowledge of all parameters of the deployed network elements (for example gain and noise figure of all amplifiers, fiber parameters, etc.); we call such a scenario *exact component parameter agnostic* or short: *agnostic*. A GN-model based estimator would require exact knowledge of the parameters due to its analytical nature or otherwise would demand for high margins, while machine learning based QoT estimation has shown to be fast and accurate [2,3,4]. Therefore, we focus on the QoT estimation in agnostic network scenarios with machine learning. The challenge arises from the deviations of the assumed component parameters as well as only sparsely available monitoring data at the intermediate nodes of a complex meshed network, at which full demodulation of the signals to obtain exact telemetry data is not a viable option. We use an extensive simulation with heuristically varying input parameters based on realistic assumptions and margins to obtain a comprehensive data set for the training of the machine learning algorithm. The dataset consists of transmission-based features and spectral-based features with a total number of $45 \cdot 10^7$ feature sets representing also different channel assignments throughout the network. We use optical spectrum analyzers (OSAs) to obtain spectral data, which are fed into an LSTM to improve the estimation performance with a novel approach interpreting a link as a series of data. The trained machine learning algorithm is tested on the unseen COST266 European network for performance evaluation and compared to a QoT estimator, which is not using spectral data. For the first time of our knowledge, this spectral data-driven LSTM is used for estimating the QoT in exact component parameter agnostic network scenarios.

2. Machine learning based gSNR estimation

The simulation setup (see Fig. 1 (a)) consists of a database of the obtained feature vectors and the optical transmission system, which defines the environment. A set N_c of coherent channels, $\boldsymbol{c} = [c_1, \ldots, c_{Nc}]$, are to be transmitted over a WDM link using a fixed channel spacing and equal launch powers per channels. Representing the missing knowledge of the exact component parameters in an agnostic network, the transmission parameters used for the simulations are calculated following a heuristic approach with a certain mean value and a standard deviation based on realistic assumptions and margins. The links are analyzed for various modulation formats (MF) i.e., QPSK, 8-QAM, 16-QAM, 32-QAM, and 64-QAM. The baudrate (b) is changed between 32 and 64 Gbaud according to the channel spacing Δf of 50 or 100 GHz. The launch power is varied between -3 and 1 dBm per channel and as the center wavelength of the WDM transmission 1550 nm is considered. Uncertainties are considered in the span lengths (L_S) by choosing randomly a length with a mean of 80 km and a standard deviation σ of 5 km. In the same way, the EDFA gain (G): $\overline{G}$: 16 dB; σ: 0.5 dB and noise figure (NF): $\overline{NF}$: 5 dB; σ: 0.5 dB are randomly chosen for each amplifier. The linear fiber

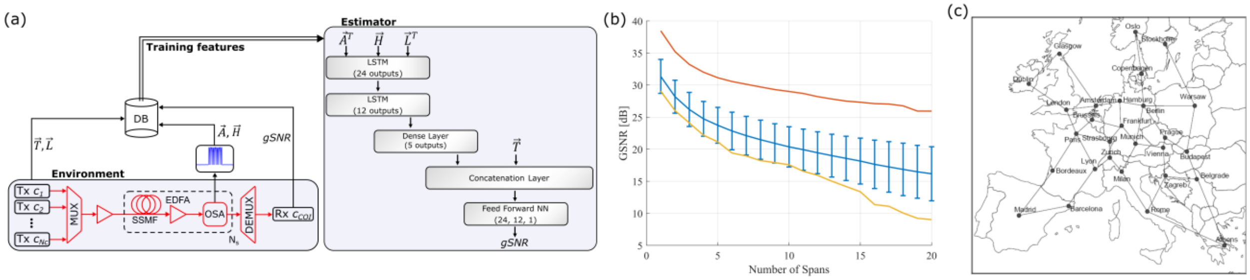

Fig. 1. (a) Simulation setup with the database (DB) containing the obtained feature vectors and gSNR estimator structure built out of LSTM and FF-NN layers, (b) COI gSNR measurements for 64 Gbaud/100 GHz spacing dependent on the number of spans. The blue curve represents the mean values of the entire sample; the error bars are equal to the standard deviation. The orange and yellow curves show the maximum and the minimum for each span number, respectively, (c) COST266 core reference network [7].

parameters α and D are set to be $\overline{\alpha}$: 0.2 dB/km; σ: 0.02 dB/km and $\overline{D}$: 17 ps/(nm·km); σ: 0.2 ps/(nm·km), respectively. Therefore, the parameters differ per span according to the random distribution. The nonlinear coefficient γ is assumed to be 1.295 $(\mathrm{W{\cdot}km})^{-1}$. It is assumed that the transmission is done over N_s spans, which are composed out of standard single-mode fibers (SSMF) followed by an EDFA with a flat gain characteristic in the C-band and an optical spectrum analyzer (OSA) with a resolution of 13 pm (according to specs of commercially available OSAs) at every intermediate node. The nonlinearities are calculated with the split-step Fourier method with a maximum phase shift of 5mrad per step. N_s is increasing with every iteration step of the simulation, i.e., in every step we add one span to the simulation setup until we reach a total of 20 spans. By this, we ensure to represent a network structure in the simulation, since we can interpret the link in different ways. For example, if the link contains 4 spans, we represent a set of links with 0, 1, 2, and 3 intermediate nodes and their different possible distance variations for the distances to and from the intermediate nodes. We use the received power (P_R), the linear noise power, i.e., the ASE noise power (P_{ASE}), and the NLI noise power (P_{NLI}) to compute the gSNR of the channel of interest (COI) with $gSNR = P_R/(P_{ASE} + P_{NLI})$ at the receiver or, in other words, at every intermediate node during the simulation. During the simulation, N_c is varied from 1 to 11 and different scenarios of adding and dropping channels are considered. Up to a total of 10 neighboring channels to the COI are dropped or added at all the intermediate nodes. For the drop scenario, the 2, 4, 6, 8, and 10 neighboring channels are dropped at every possible node in the network. For the add scenario, on the other hand, these neighboring channels are added to the channel configurations in which they are free. We restrict the power of channels to add not to be higher than 1.2 times the current mean channel powers to reduce the overall simulation effort.

We have set up this scenario for the data generation in our Matlab-based simulation tool. The data extracted from the simulation can be divided into two different categories, the transmission-related features and the features extracted from the spectrum obtained by the OSA. The transmission-related features are composed out of the vector $\vec{T} = [MF, P_L, \Delta f, b, \ L]$ with L being the total link length and the vector $\vec{L} = [L_{N_S-k,N_S-(k-1)}, \dots,\ L_{N_S-1,N_S}]$ with k as the number of intermediate nodes. $\vec{L}$ contains the information on the lengths between the intermediate nodes themselves and the end of the link. The spectrum-related features are the vector $\vec{A} = [A_{N_S-k,N_S-(k-1)}, \dots,\ A_{N_S-1,N_S}]$ that contains the area under the envelope of the PSD obtained by the OSA and the vector $\vec{H} = [H_{c_1}, \dots, H_{c_{11}}]$ which is composed out of the hights of the PSD of the channels $\boldsymbol{c}$. In $\vec{H}$ the channel usage, the influences of uncertainties in the gain, noise figure, attenuation, and the influences of the non-linearities on the power of the channels can be seen. As a result of analyzing the different modulation formats, launch powers, baudrates, and channel configurations, we obtain a dataset with around $30 \cdot 10^6$ feature sets each of which contains up to 71 features. To represent a larger number of uncertainties, the simulation was repeated 15 times, resulting in a large dataset of $45 \cdot 10^7$ feature sets. As the number of considered intermediate nodes increases, the second dimension of $\vec{L}, \vec{A}$ and $\vec{H}$ increases, too. Fig. 1 (b) shows the gSNR distribution over the different lengths of the links for 64 Gbaud, 100 GHz spacing in the dataset. It can be seen that the dataset spans a wide range of gSNR values. The changing dimensions of the input vectors lead to a different challenge for choosing a suitable ML algorithm. Most algorithms operate on a fixed number of features, therefore, we interpret every link as a series of values that can be fed into a recurrent neural network, for example, a long-short term memory network (LSTM). However, $\vec{T}$ is not changing in size so it is fed into the gSNR estimator in a later stage. The estimator structure can be seen on the right side of Fig. 1 (a). The LSTM layers consist of 24 and 12 outputs, respectively. These parameters were hand-selected and could be further improved using common methods such as grid search or Baysian hyper parameter optimization, which however is not the focus of this work. The output is followed by a dense layer, i.e., a fully connected layer with 5 neurons, to match the size of $\vec{T}$ for the input in the concatenation layer. The feed-forward neural network (FF-NN) consists of 2 hidden layers with ReLU activation

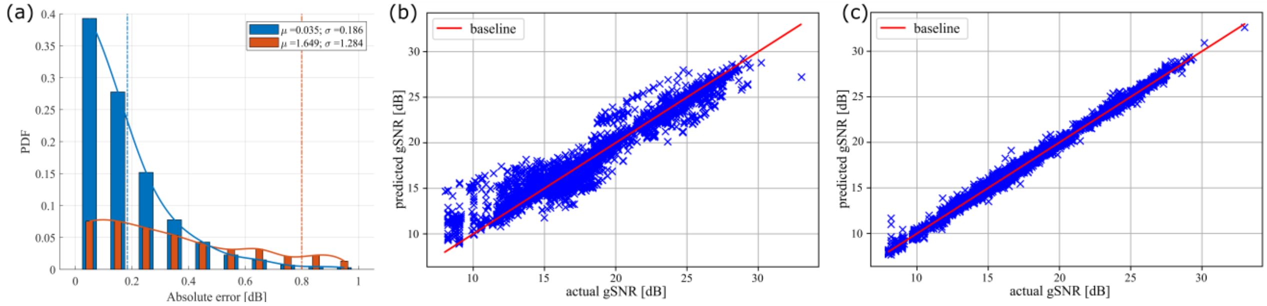

Fig. 2. (a) Probability density function over the absolute error in the estimated values. The blue bars show the occurrence of the errors in the case of using spectral data. The orange bars show the case with only the transmission-based features. The corresponding-colored dashed vertical lines show the mean absolute error, (b) predicted gSNR by the estimator for the non-spectral feature case over actual gSNR, (c) predicted gSNR by the estimator for the spectral feature case over actual gSNR.

functions composed out of 24 and 12 neurons and an output layer with one output which is the gSNR. Before the training, the obtained dataset is split into 60% training, 20% validation, and 20% test data. The LSTM/FF-NN structure is trained over 1000 epochs and optimized with the Adam optimizer. With our simulation tool, we modeled the COST266 European network topology [7] (Fig. 1 (c)) with 28 nodes and 41 bidirectional edges with 378 traffic scenarios. The links for these 378 scenarios reach from 0 to 10 crossed intermediate nodes with link lengths from ranging 300 to 1500 km. There is no information contained in the dataset on the fiber types, lengths, or other properties of the links or network, so we call this an agnostic network. We take values for 10,000 links in the network from the heuristic parameter spaces mentioned before for the simulation. At the three nodes with the highest rank (Berlin, Zürich, and Milan) a random number of channels (1 to 10) are dropped. Furthermore, during the simulation, the gSNR values are obtained using coherent receivers to generate a ground truth for the evaluation of the gSNR estimator. However, these values are not necessary for the operation of the estimator; they are obtained to show the performance. Our goal is to show that the deployment of OSAs and the usage of spectral data for QoT estimation in agnostic network scenarios is beneficial. So, for comparison, we train another LSTM/FF-NN structure with the transmission-based feature vectors $\vec{T}$ and $\vec{L}$ as well as a channel allocation vector containing a 0 for a free slot in the frequency grid and a 1 for an occupied slot. The training is done in the same way.

3. Results

Both estimators are tested on the unseen COST266 European network topology. Fig. 2 (a) shows the probability density function (PDF) of the gSNR estimation. When using only the transmission-based features $\vec{T}$ and $\vec{L}$ (orange bars and curve), the absolute error of the gSNR in dB is widely spread, since the algorithm is only learning the dependency between the length and the nonlinearities. However, the usage of spectral features obtained from the OSA narrows the occurrence of the absolute error towards zero. The mean absolute error (MAE) is 0.18 dB and 0.81 dB while the standard deviation is 0.186 dB and 1.284 dB for the spectral and non-spectral data case, respectively; this shows an improvement of 1.1 dB. The difference in estimation accuracy is also represented in the regression (R^2-) scores of 0.995 and 0.879 as it can be seen in Fig. 2 (b) and (c). It shows the plots of predicted gSNR against the actual gSNR of the scenarios (b) without and (c) with training on spectral features. The difference in estimation performance can be seen in the occurrence of outliers from the baseline in the case without the spectral features.

4. Conclusion

We show that the usage of spectral data obtained from sparsely deployed optical spectrum analyzers in agnostic network scenarios is beneficial. We propose representing such an agnostic network by heuristically varying component parameters. Our LSTM and FF-NN-based estimator - trained with spectral data obtained by OSAs - shows a 1.1 dB lower mean absolute error in gSNR estimation than the case without employing monitors at intermediate nodes and exhibits a low deviation of only 0.18 dB from the theoretically obtained baseline. The spectral data driven estimator finally demonstrates superior overall regression performance.

References

[1] Poggiolini, "The GN model of non-linear propagation in uncompensated coherent optical systems.", in JLT, Vol. 30.24 (2012): 3857-3879.
[2] Rottondi, et al., "Machine-learning method for quality of transmission prediction of unestablished lightpaths.", in JOCN, Vol. 10.2 (2018): A286-A297.
[3] Aladin, et al., "Quality of transmission estimation and short-term performance forecast of lightpaths.", in JLT, Vol. 38.10 (2020): 2807-2814.
[4] Sartzetakis, et al., "Accurate quality of transmission estimation with machine learning.", in JOCN, Vol. 11.3 (2019): 140-150.
[5] Seve, et al., "Associating machine-learning and analytical models for quality of transmission estimation: combining the best of both worlds.", in JOCN, Vol. 13.6 (2021): C21-C30.
[6] Müller, et al. "Estimating Quality of Transmission in a Live Production Network using Machine Learning." in OFC, 2021.
[7] http://sndlib.zib.de.