CHALMERS
UNIVERSITY OF TECHNOLOGY

Towards Better QoT Estimation: An ML Architecture with Link-Level Embedding Layers

Citation for the original published paper (version of record):
Lechowicz, P., Natalino Da Silva, C., Arpanaei, F. et al (2025). Towards Better QoT Estimation: An ML Architecture with Link-Level Embedding Layers. IEEE Networking Letters, 7(2): 122-125. http://dx.doi.org/10.1109/LNET.2025.3561336

(article starts on next page)

Toward Better QoT Estimation: An ML Architecture With Link-Level Embedding Layers

Piotr Lechowicz, Carlos Natalino, *Member, IEEE*, Farhad Arpanaei, *Member, IEEE*, Stefan Melin, Renzo Diaz, Anders Lindgren, David Larrabeiti, and Paolo Monti, *Senior Member, IEEE*

***Abstract*—Machine learning (ML) is emerging as a promising tool for estimating the Quality of Transmission (QoT) in optical networks, especially for unestablished lightpaths where traditional methods are limited. However, inaccuracies in ML-based QoT predictions—typically expressed in terms of generalized signal-to-noise ratio (GSNR)—can significantly affect network operation. Overestimation may lead to retransmissions due to overly aggressive modulation format choices, while underestimation results in underutilized spectral resources. To address this, we propose a novel ML architecture that incorporates an embedding layer for link-level features alongside path- and service-level inputs. Using data generated from an accurate analytical model, we show that our approach reduces prediction error by up to 34% compared to standard architectures. Simulated deployment scenarios further demonstrate operational benefits, with a 15.9% decrease in incorrect and a 34.8% reduction in overly conservative modulation format selections.**

***Index Terms*—Quality of transmission, machine learning, artificial neural networks, network embedding layer, generalized signal-to-noise ratio.**

I. Introduction

QUALITY of transmission (QoT) estimation of unestablished lightpaths (LPs) is crucial in planning and operating low-margin optical networks. Analytical models can perform accurate estimations but require rich and precise information about the network. QoT estimation becomes more challenging in disaggregated optical networks due to the diversity of equipment. Keeping track of physical layer parameters becomes more challenging over time, as network equipment aging, replacement, and/or reconfiguration may increase the uncertainties. Machine learning (ML) methods have gained popularity in evaluating QoT as they are characterized by robustness for uncertain or missing parameters that analytical models require [1], [2]. Artificial neural networks (ANNs) are among the most accurate and adopted models. Despite their high accuracy, even minor estimation errors can result in significant transmission inefficiencies, such as incorrect modulation format selection, unnecessary network reconfigurations, suboptimal spectral efficiency, and degraded network performance. These inefficiencies impact connection setup times and the network overall throughput. Therefore, further reducing estimation errors leads to more robust QoT estimation models.

Traditionally, the input to ANNs for QoT estimation includes various path and LP information, such as endpoint nodes, maximum/minimum/average link length, and number of spans [3], [4], [5], [6]. However, these feature sets do not account for detailed link-level input, e.g., the specific links traversed by a LP [7]. Recent findings show that incorporating link-level details into ML models improves their accuracy [5], as well as the accuracy of joint ML/analytical models used for generalized signal to noise ratio (GSNR) estimation [7]. However, providing such information as input to ANNs is challenging. One way to represent link-level information is through a binary vector [5]. The result is a sparsely populated vector, which ANNs are known not to handle well [8]. Another way is to use embedding layers, which can represent sparse vectors as dense real-valued numerical vectors that can be used more efficiently by ANNs [9]. This approach has been used in graphs, where the *node2vec* embedding has been proposed to learn a continuous representation of nodes, aiding the decision to add new links to improve connectivity, i.e., planning network upgrades [10], [11].

In this letter, we propose the use of an embedding layer to input detailed link-level information to ANNs that predicts the QoT value of unestablished LPs. We generate a dataset using an accurate analytical model to train and assess the performance of the proposed ANN architecture. The dataset contains, for each LP, a set of descriptive features and the corresponding achieved GSNR. We benchmark the proposed ANN architecture, leveraging embedding against two traditional architectures from the literature while maintaining the same complexity for all ANNs. Results show that the proposed use of link-level inputs combined with an embedding layer reduces the GSNR error by up to 34%. Then, we assess the impact of the estimation error reduction provided by our proposed architecture on the long-term network performance. This reduces the chances of selecting incorrect and conservative modulation formats by nearly 16% and 35%, respectively.

Received 27 March 2025; accepted 8 April 2025. Date of publication 16 April 2025; date of current version 9 June 2025. This work was supported in part by ECO-eNET funded by the Smart Networks and Services Joint Undertaking (SNS JU) under Grant 10113933; in part by Fun4date-Redes under Grant PID2022-136684OB-C21; and in part by FLEX-SCALE under Grant 101096909. The associate editor coordinating the review of this article and approving it for publication was L. Foschini. *(Corresponding author: Piotr Lechowicz.)*

Piotr Lechowicz, Carlos Natalino, and Paolo Monti are with the Department of Electrical Engineering, Chalmers University of Technology, 412 96 Gothenburg, Sweden (e-mail: piotr.lechowicz@chalmers.se; carlos.natalino@chalmers.se; mpaolo@chalmers.se).

Farhad Arpanaei and David Larrabeiti are with the Department of Telematic Engineering, Universidad Carlos III de Madrid, 28911 Leganés, Spain (e-mail: farhad.arpanaei@uc3m.es; david.larrabeiti@uc3m.es).

Stefan Melin, Renzo Diaz, and Anders Lindgren are with Telia Company, 169 94 Solna, Sweden (e-mail: stefan.melin@teliacompany.com; renzo.z.diaz@teliacompany.com; anders.x.lindgren@teliacompany.com).

Digital Object Identifier 10.1109/LNET.2025.3561336

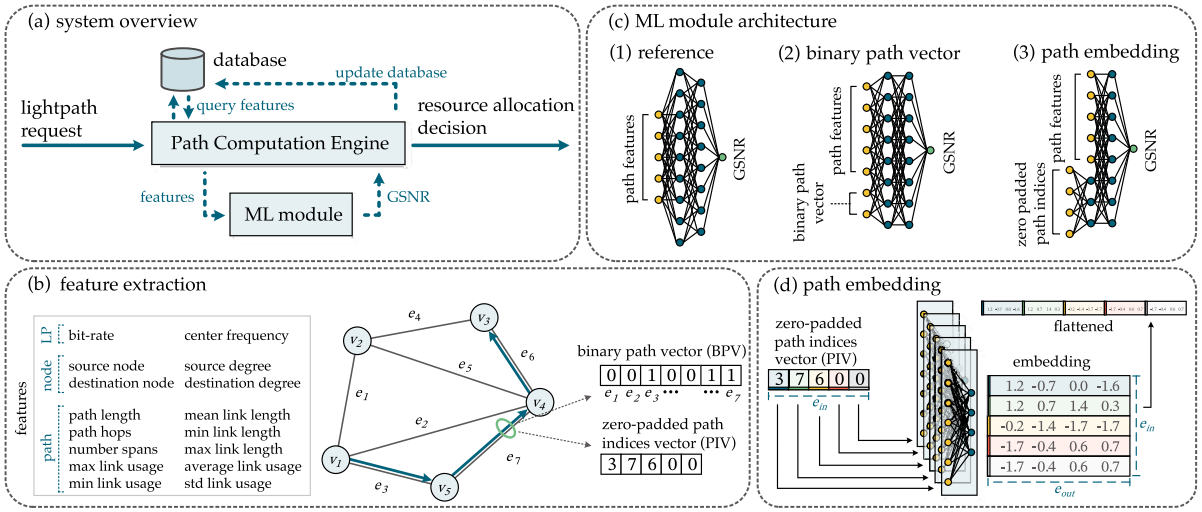

Fig. 1. Summary of the proposed approach, including *(a)* system overview, *(b)* feature extraction, *(c)* ML module architecture, and *(d)* link-level path embedding.

II. QoT Regression With Link-Level Embedding

Consider an optical network with a control plane architecture that uses a path computation engine [12]. In an elastic optical network (EON), the path computation engine solves the routing, modulation, and spectrum assignment problem upon receiving a LP request. Fig. 1(a) illustrates the scenario under consideration, where the path computation engine leverages a ML model to estimate the GSNR of potential (unestablished) LP configurations. The ML model is composed of an ANN that takes as an input a set of observed network parameters (*features*) and returns the estimated GSNR.

Traditionally, ML-based QoT estimators are fed end-to-end features related to the LP request, including summarised lightpath, node, and path features, as illustrated in Fig. 1(b) [6]. For instance, the path (links and spans therein) is commonly summarised with features such as total path length, number of spans composing the path, and their length expressed as average, minimum, maximum, and standard deviation. Fig. 1(c.1) illustrates an ANN architecture using such a feature set. However, previous works showed that providing link- [5] and span-level [7] information, in addition to the classic end-to-end path features, can improve the accuracy of ML models. For instance, two significantly different paths may have similar *average* feature values [5], making it difficult for the ML model to distinguish between them. One possible approach to address this issue is illustrated in Fig. 1(c.2), where link-level information is encoded as a binary path vector (BPV). Each bit in the vector corresponds to a particular link in the network, where a "1" indicates a specific link is included in the path [5]. In this case, the features' cardinality scales with the network topology's size. This translates into two main challenges. Firstly, the feature cardinality varies with network topology, potentially requiring a new ANN architecture for each topology. Secondly, since each LP traverses only a fraction of the links/spans, the BPV is sparsely populated with non-zero elements. ANNs are notoriously inefficient with sparse features, thus making BPV not an ideal solution. ANN embeddings can help overcome these challenges by representing each unique word, i.e., a path described as a sequence of links, as a real-valued dense numerical vector. The intuition is that similar paths should have a smaller Euclidean distance of created vectors [9]. In our approach, we convert the BPV into a path indices vector (PIV), as illustrated in Fig. 1(b), and show that accounting for PIV feature improves GSNR prediction compared to the method without link-level information. The maximum size of the PIV, i.e., e_{in}, is defined by the maximum number of hops among the considered paths. As ANNs require a fixed number of input features, the PIV is zero-padded when the number of links in the path is lower than e_{in}. As shown in Fig. 1(d), each input link of the routing path produces a vector of dimension e_{out}. The output of the embedding process is a matrix of dimension $e_{in} \times e_{out}$ (e_{in} links in a path, each resulting in e_{out} vector). Such matrix is flattened to $1 \times e_{in} \cdot e_{out}$ vector, then fed to the first fully connected ANN layer.

III. Artificial Neural Network (ANN) Architectures

To investigate the benefits of using ANN embedding, three ANN architectures are compared, namely, *ref*, *bpv*, and *emb*. The proposed architectures' parameters are selected based on the architecture in [6], which is effective for QoT estimation. We benchmark them in terms of input feature encoding while maintaining the same number of trainable parameters and, thus, the same complexity. First, a set of common assumptions for all architectures is provided. The ANNs are composed of two hidden layers, where the number of neurons in the first one is adjusted depending on the encoding used. All ANN architectures use the same set of 16 features listed in Fig. 1(b) as input. Each ANN is trained using the RMSprop optimizer with a 0.01 learning rate and mean squared error as the loss function. The output is the estimated GSNR value. The 16 features and the GSNR are normalized by removing the mean and scaling to the unit variance of the feature. The two benchmark architectures are set up as follows. The *ref*

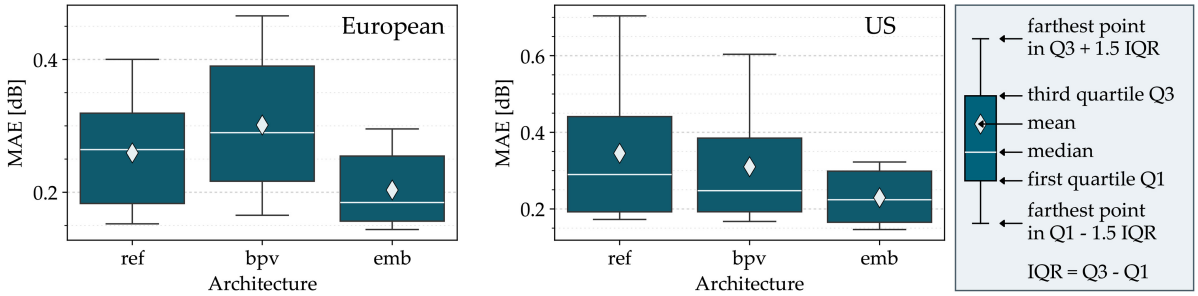

Fig. 2. Mean absolute error (MAE) distribution for the reference (ref), binary path vector (bpv), and proposed path embedding (emb) architectures for the European (left) and U.S. (right) network topology.

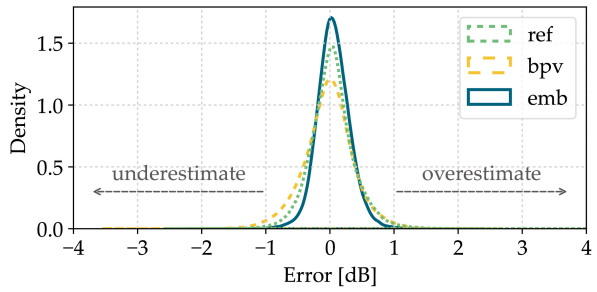

Fig. 3. Kernel density estimation of the GSNR prediction error considering the European topology.

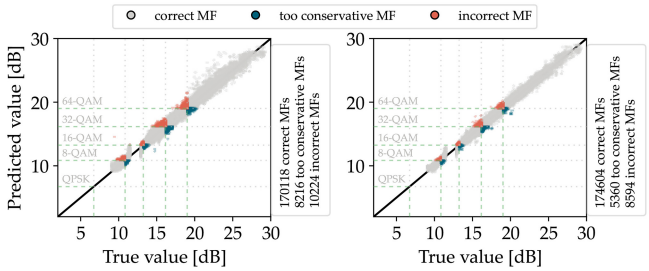

Fig. 4. Predicted vs. true values of GSNR for reference (left) and embedding (right) architectures for the European topology.

architecture in Fig. 1(c.1) is created based on an architecture from the literature [6], with one additional hidden layer. The input contains only the 16 basic features, and the first hidden layer contains a parameterized number of neurons n_{ref} and a hyperbolic tangent activation function. The second hidden layer contains 256 neurons and a hyperbolic tangent activation function. The output is activated using a linear function. The *bpv* architecture in Fig. 1(c.2) receives as input, in addition to the 16 basic features, the BPV. The first hidden layer contains n_{bpv} neurons. The proposed *path embedding (emb)* architecture in Fig. 1(c.3) uses 16 basic features plus the PIV as input. Based on the network scenario and embedding output size, it is possible to calculate the number of ANN trainable parameters (i.e., weights and biases) in the *emb* scenario and calculate the number of neurons in the first hidden layer, i.e., n_{ref} and n_{bpv}, for the *ref* and *bpv* scenarios, respectively.

IV. Dataset Collection

We used the analytical model from [13] to generate a dataset used as the ground truth (GT) for this letter. The GSNR of a channel under test (CUT) can be calculated as a function of the set of spans S along the path:

$$GSNR_{\text{CUT}}{}^{-1} = \left(\sum_{s \in S} \frac{P^s_{\text{CUT}}}{P^s_{\text{ASE}} + P^s_{\text{NLI}}} \right)^{-1}, \qquad (1)$$

where, for each span $s \in S$, P^s_{CUT} is the launch power, P^s_{ASE} is the noise incurred by amplified spontaneous emissions, and P^s_{NLI} is the noise incurred by non-linearity. The dataset is created by simulating the dynamic provisioning of LP requests in an EON. Two network topologies are considered to draw more general conclusions. The European topology (28 nodes and 41 links) and the U.S. continental topology (26 nodes and 42 links) are identified as *nobel-eu* and *janos-us* in the SNDlib [14]. These topologies allowed us to collect datasets with different feature distributions, such as GSNR, path length, and number of hops/spans. Ultimately, these features determine the datasets' modulation format (MF) distribution [15]. Spans within a link have equal length with a maximum of 80 km. We compute the 5 shortest paths for each node pair, which results in a maximum of 9 links per path. Starting from an empty network, we simulate 200,000 LP requests and provision them using a random path, best modulation format, and first-fit spectrum assignment, assuming a launch power of 0 dBm. We considered five modulation formats: QPSK, 8-, 16-, 32-, and 64-QAM, with their GSNR threshold set to {6.72, 10.84, 13.24, 16.16, and 19.01} dB, respectively. The best modulation format (i.e., in terms of spectral efficiency) is decided by checking the expected LP GSNR value (i.e., computed via the analytical model described earlier) against the GSNR thresholds of the five options, considering a 2 dB margin. Of the 200,000 requests, 188,557 and 172,319 LPs are accepted for the European and U.S. topology, respectively, and their features and resulting GSNR are recorded. Fig. 1(b) shows the complete set of features recorded for each LP in the dataset. For the European topology, the input size for the ANN is 16, 57, and 25 for *ref*, *bpv*, and *emb* architectures, respectively. The corresponding input sizes for the U.S. topology are 16, 58, and 25. The input size of embedding e_{in} is 9, and the output size e_{out} is 6. The number of neurons in the first hidden layer is chosen to ensure all three tested architectures have the same number of trainable parameters, i.e., the same complexity (n_{bpv} is 58 and n_{ref} is 67).

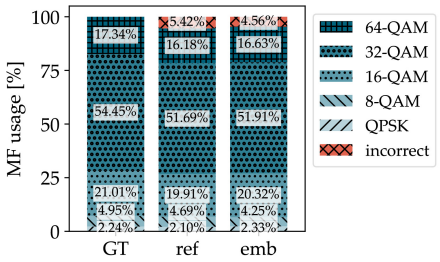

Fig. 5. Percentage of MFs selected for LPs using the ground truth (GT), reference, and embedding architectures for the European topology.

V. Results and Discussion

The results in this section are obtained by training each ANN architecture using the k-fold strategy over the samples collected for the European and U.S. network topology, respectively, with $k = 10$ folds. Fig. 2 shows the mean absolute error (MAE) statistics for each ANNs. For the European topology, the average MAE is 0.26, 0.30, and 0.20 dB for *ref*, *bpv*, and *emb*, respectively. The *emb* architecture reduces the MAE by 23.1%, and the median of the absolute error by 30.8% when compared to *ref*, while *bpv* worsens the mean and median by 15,4% and 11,5%, respectively. For the U.S. topology, the average MAE is 0.35, 0.31, and 0.23 dB. The mean and median MAE differences between the *emb* and the *ref* architectures are 34.3% and 24.1%, respectively.

Fig. 3 shows the kernel density estimation of the absolute GSNR prediction error for the tested architectures from all test samples in all folds, considering the European topology. The bandwidth of the density estimation is set according to the Silverman formula. It can be observed that the *emb* architecture has a higher density of around 0 dB, narrower tails than the other architectures, and no strong bias towards under- or overestimation. *Ref* and *bpv* tend to underestimate the GSNR more than the *emb* architecture. In the remainder of this section, we focus on the two best-performing architectures, i.e., *ref* and *emb*.

Fig. 4 presents the actual vs. predicted GSNR values for *ref* and *emb*. Grey points depict a correct modulation format selection, i.e., the selection based on the predicted GSNR matches the one based on the GT. Points in blue refer to GSNR estimations lower than the GT (underestimation), resulting in a conservative modulation format selection. Points in red refer to GSNR estimations higher than the GT (overestimation), resulting in a failed LP establishment. With *emb* we have fewer incorrect and conservative modulation format selections (i.e., 15.9% and 34.8% fewer, respectively). This leads to fewer failed LP establishments and more spectrally efficient LPs.

Fig. 5 shows the percentage of modulation formats selected when using the considered architectures. Compared to *ref*, the better accuracy provided by *emb* results in higher-order modulation formats (16-, 32-, 62-QAM) being used more often (88.8% vs. 87.8%). More importantly, the number of incorrectly selected MFs decreased by 18%. Regarding inference time, irrespective of the model, the average value varied between 30 and 40 ms. We repeated the experiments with the U.S. national network topology, achieving similar results, which are not shown in this letter due to space limitations.

VI. Conclusion

This letter analyzed how embedding link-level information may assist the performance of ANNs for QoT regression of unestablished lightpaths. With embedding, the categorical data is represented as multi-dimensional numerical vectors. According to our findings, embedding link-level information decreased the MAE by up to 34% when compared to the reference scenario for an EU network topology. A similar approach can be adopted in more advanced architectures and strategies, such as convolutional/time-series ANNs and deep reinforcement learning. We plan to continue investigating the benefits of embedding other network properties, check the method scalability with different network sizes and traffic loads, and investigate the pre-training of embedding layers.

References

[1] M. Lonardi et al., "The perks of using machine learning for QoT estimation with uncertain network parameters," in *Proc. OSA Adv. Photon. Congr. (AP)*, 2020, Art. no. NeM3B.2.

[2] O. Karandin et al., "Probabilistic low-margin optical-network design with multiple physical-layer parameter uncertainties," *J. Opt. Commun.*, vol. 15, no. 7, pp. C129–C137, 2023.

[3] T. Panayiotou et al., "Performance analysis of a data-driven quality-of-transmission decision approach on a dynamic multicast-capable metro optical network," *J. Opt. Commun.*, vol. 9, no. 1, pp. 98–108, Jan. 2017.

[4] C. Rottondi et al., "Machine-learning method for quality of transmission prediction of unestablished lightpaths," *J. Opt. Commun.*, vol. 10, no. 2, pp. A286–A297, Feb. 2018.

[5] I. Sartzetakis et al., "Accurate quality of transmission estimation with machine learning," *J. Opt. Commun.*, vol. 11, no. 3, pp. 140–150, Mar. 2019.

[6] G. Bergk et al., "ML-assisted QoT estimation: A dataset collection and data visualization for dataset quality evaluation," *J. Opt. Commun.*, vol. 14, no. 3, pp. 43–55, Mar. 2022.

[7] F. Arpanaei et al., "A novel approach for joint analytical and ML-assisted GSNR estimation in flexible optical network," in *Proc. Eur. Conf. Opt. Commun. (ECOC)*, 2022, pp. 1–4.

[8] J. T. Hancock and T. M. Khoshgoftaar, "Survey on categorical data for neural networks," *J. Big Data*, vol. 7, no. 1, p. 28, 2020.

[9] R. Patil et al., "A survey of text representation and embedding techniques in NLP," *IEEE Access*, vol. 11, pp. 36120–36146, 2023.

[10] A. Ahuja et al., "Topology-driven edge predictions with graph machine learning for optical network growth," in *Proc. Opt. Fiber Commun. Conf. (OFC)*, 2024, pp. 1–3.

[11] A. Grover and J. Leskovec, "Node2vec: Scalable feature learning for networks," in *Proc. Int. Conf. Knowl. Discov. Data Mining (ACM SIGKDD)*, 2016, pp. 855–864.

[12] N. Sambo et al., "Lightpath establishment assisted by offline QoT estimation in transparent optical networks," *J. Opt. Commun.*, vol. 2, no. 11, pp. 928–937, Nov. 2010.

[13] M. Ranjbar Zefreh et al., "Accurate closed-form real-time EGN model formula leveraging machine-learning over 8500 thoroughly Randomized full C-band systems," *J. Light. Technol.*, vol. 38, no. 18, pp. 4987–4999, 2020.

[14] S. Orlowski et al., "SNDlib 1.0–survivable network design library," *Networks*, vol. 55, no. 3, pp. 276–286, 2010.

[15] P. Lechowicz et al., "Elastic optical network lightpath GSNR dataset." Mar. 2025. [Online]. Available: https://doi.org/10.5281/zenodo.15097597