Towards Universal Paradigm of QoT Estimation over Optical Transport Network through Graph Neural Structure

Cen Wang(1), Yuta Wakayama(1), Noboru Yoshikane(1), Takehiro Tsuritani(1)

(1) KDDI Research, Inc. xce-wang@kddi.com

Abstract *We propose a practical and generalizable PNPAConv-based neural structure to adapt network-scale QoT estimation. The performance* $\log_{10}\mathrm{MSE}$ *achieves* -1.982 *and* -1.506 *over JPN12 and EURO28, respectively.*

Introduction

Network-scale QoT estimation is essential in network planning, transport system evolution, and fault diagnosis, etc. for providers. Analytical models are computationally intensive and limited to scenarios. In contrast, neural networks easily capture implicit functional relationships, thus neural model-based QoT estimation has been extensively studied. Existing methods[1]–[3] mainly use a fully connected neural structure as an estimator, and its essence is to proximate the Gaussian noise (GN) model or the generalized Gaussian noise (GGN) model. In addition, leveraging a simple fully connected structure makes the estimation model skip topology and routing issues, thus lacking universality and practicability.

In this work, we aim to find a paradigm: the inputs are service requests (i.e., lightpaths to be established) of the whole optical transport network (OTN), and the QoT information (such as GSNR) of each node can be directly inferred through the trained model. To this end, a kind of modeling is worth to be considered: the signal propagates along a certain route over a topology. We have found that an advanced message-passing graph convolutional network (MP-GCN)[4] most closely matches the problems we expect to solve[5].

In light of this, we modify the principal neighborhood aggregation convolution (*PNAConv*) layer[6] which is a variant of MP-GCNs achieving state-of-the-art performance in many tasks, and propose principal neighborhood propagation aggregation convolution (*PNPAConv*) layer to adapt the learning logic of the QoT estimation problem over OTNs. The existing GCN-based QoT estimation[7] uses basic GCN and thereby suffers scaling problem. We evaluate the performance of *PNPAConv*-based model with services in C-band. The estimation $\log_{10}\mathrm{MSE}$ can achieve -1.982 and -1.506 (with from 0.01 dB to 0.1 dB errors) over *JPN12*[8] and *EURO28*[9], respectively. We additionally verify the practicability and generalization ability of *PNPAConv* through multiple trainings/testings.

Graph Neural Network and Modeling

The existing QoT estimation neural models can be referred to as the unfolded structure[2] and the cascaded structure[1],[10], which are shown in Fig. 1 (a) and (b), respectively. The unfolded structure requires the feeding status of each edge (i.e., optical multiplex section (OMS)) transmitting side, which is essentially a data leak and not feasible. The cascaded structure is reasonable, however, it requires multiple levels of cascading to emulate a lightpath, and it cannot adapt to topologies with bifurcations.

As shown in Fig. 1 (c), we expect to estimate the whole network QoT in a practical way. The inputs are service requests, and the outputs are GSNRs on each node. A service request can be described by a tensor $R_{n\times n\times *\times *}$, where $R(i,j,:,:)$ refers to the channels transmitted by node i and received by node j, including information such as transmit power p (the consideration of transceivers is omitted for modeling simplicity), center wavelength f, bandwidth bw, and baud rate br. We map the requests into node features and edge features through calculated routes. A node feature v_i consists of $R(i,:,p)$, $R(:,i,p)$, and other necessary information (e.g., ROADM-specific structure and settings). In an edge feature, a spectral channel utility is described as a triad $< f_m, bw_m, br_m >$, and the optical span, fiber information, and noise figure of an amplifier, etc. are also included. Although these features are all numbers, they are not all continuous variables. To make the neural model training stable, we use two tunable encoders to realize the embeddings of node and edge features, to achieve dimension reduction.

The (per-service) GSNR of a node i can be regarded as the function of the transmission status of all the lightpaths along which i is the target[5],[11],[12]. However, it is computationally expensive and prone to overfitting that directly uses

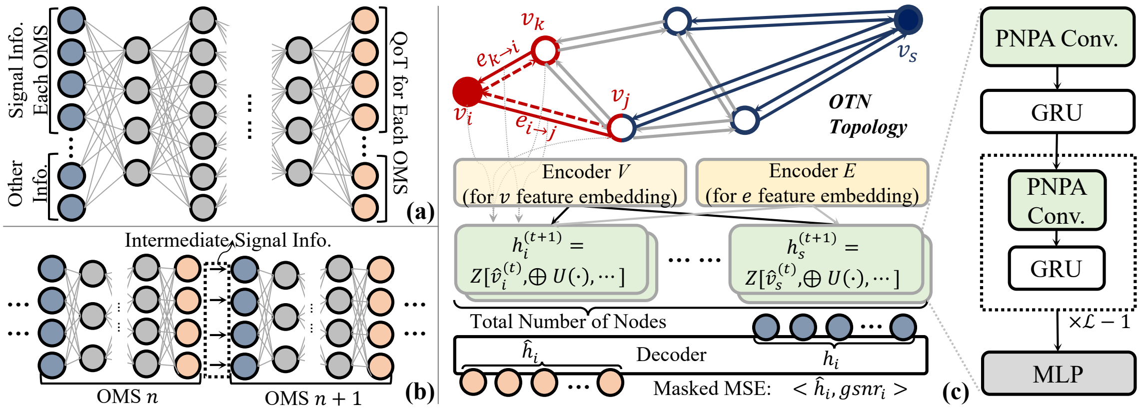

Fig. 1: Recent vs. the proposed *PNPAConv*-based paradigm for OTN QoT estimation .

the transmission status of multiple whole lightpaths for each node in model training. Therefore, inspired by *PNAConv* paradigm, we decompose network-scale training into local training iterations, so that the local transmission status is finally propagated to the entire network. At each iteration, the hidden representation h_i of the node i is updated so that it captures the influence of the learned transmission status[13].

We find the original *PNAConv* paradigm is insufficiently effective. The reason is that *PNAConv* supposes an additive aggregation of the neighbors' influence on node i. In contrast, in a transmission status regression task, the influence of the neighbor nodes $N(i)$ and connected edges on the node i is directional. Considering the difference between the task intentions, we modify the aggregation idea of the original *PNAConv* and propose *PNPAConv*.

Eq. 1 shows the layer expression of *PNPAConv*.

$$h_i^{(t+1)} = Z\left(\hat{v}_i^{(t)}, \bigoplus_{k,j\in N(i)} U\left(\hat{v}_k^{(t)}, \hat{e}_{k\to i}^{(t)}, \hat{v}_i^{(t)}, \hat{e}_{i\to j}^{(t)}, \hat{v}_j^{(t)}\right)\right) \tag{1}$$

where $h_i^{(t+1)}$ is the latent representation at t-th iteration of node i; Z and U are neural networks; $\hat{v}$ and $\hat{e}$ are embedded vertex (i.e., node) and edge features. It can be found that $\left(\hat{v}_k^{(t)}, \hat{e}_{k\to i}^{(t)}, \hat{v}_i^{(t)}, \hat{e}_{i\to j}^{(t)}, \hat{v}_j^{(t)}\right)$ is a series of directed lightpath segments centered at i. The encoders for embeddings are shown in Eq. 2 and Eq. 3.

$$\hat{v}_i^{(t)} = \text{Encoder}_v\left(v_i^{(t)}, h_i^{(t-1)}\right) \tag{2}$$

$$\hat{e}^{(t)} = \text{Encoder}_e\left(e^{(t)}\right) \tag{3}$$

The operation $\oplus$ is used to integrate local influence and guarantee the convergence of h, and it is the tensor product $\otimes$ of standard deviation-based aggregation operator σ (in Eq. 5) and attenuation factor A (in Eq. 6).

$$\oplus = A \otimes \sigma \tag{4}$$

$$\sigma_i(Q) = \sqrt{\text{ReLU}\left(\mu_i\left(Q^2\right) - \mu_i(Q)^2\right) + b} \tag{5}$$

$$A(d,\alpha) = \left[\frac{\log(d+1)}{\delta}\right]^{\alpha}, \delta = \frac{1}{|V|}\sum_{i\in V}\log(d_i+1) \tag{6}$$

where d is the degree of the node receiving signals, and α is a variable parameter.

The loss function is (masked) mean square error (MSE), as shown in Eq. 7. The masked MSE refers to only computing final QoTs of services that terminate at node i.

$$Loss = \underset{\forall i\in V}{\text{MSE}}\left(\hat{h}_i^{(t)} - \text{gsnr}_i\right) \tag{7}$$

where $\hat{h}_i^{(t)}$ is the output of a decoder which is fed with $h_i^{(t)}$ as the input.

Moreover, as shown in the right part of Fig. 1 (c), the *PNPACov* layers can be multiple. The number of layers is $\mathcal{L}$ which is a function of the number of OTN network nodes $|V|$. A gated recurrent unit (GRU) is applied after each layer, to retain information from previous layers.

Training and Testing

We modified `PNAConv` function to implement `PNPAConv` in `torch_geometric.nn`[14]. The encoder/decoder and Z/U are *Transformer*[15] and fully connected structures, respectively, and $\mathcal{L} = |V|/2$. To further reduce the number of parameters (i.e., to improve the model efficiency), we

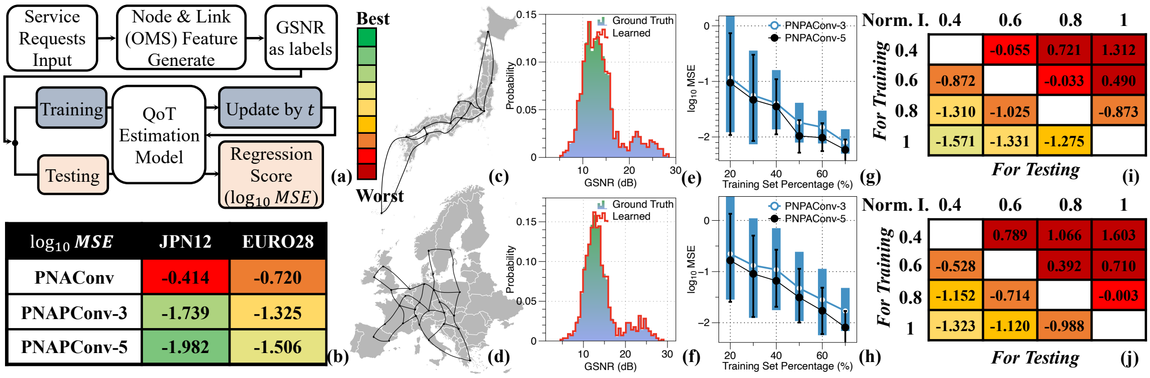

Fig. 2: Training and testing results for GGN model over *JPN12* [(e)(j)(i)] and *EURO28* [(f)(h)(j)] topologies.

replace the five-tuples (referring to *PNPAConv-5*) with three-tuples (i.e., $\left(\hat{e}^{(t)}_{k\to i}, \hat{v}^{(t)}_{i}, \hat{e}^{(t)}_{i\to j}\right)$, referring to *PNPAConv-3*) for the $\oplus$ operation. We compare performance of *PNAConv*, *PNPAConv-3* and *PNPAConv-5*, over *JPN12* (Fig. 2 (c)) and *EURO28* (Fig. 2 (d)).

The datasets preparation and training/testing procedures are shown in Fig. 2 (a). The GSNR calculations follow our previous work[16], using `GNPy` (GGN model)[17]. The services generated over the networks are in C-band 96 channels, with PM-8QAM and PM-16QAM. To analyze the model practicability, we adjust the percentage of the dataset (normalized service intensity I is 1) for training. In addition, to evaluate the model generalization ability, we train a dataset with a specific I and test the scenarios with the other I. We use $\log_{10}\text{MSE}$ as the performance metric.

As shown in Fig. 2 (b), we use different colors to indicate the level of performance. *PNPAConv-5* outperforms *PNPAConv-3* and *PNAConv*, due to more parameters and adaptive learning structure, respectively. *EURO28* with a larger topology deteriorates the model performance. It can be found that the performance of the *PNPAConv*-based model on OTN-scale topologies is effective in the QoT estimation task. The distributions of the learned GSNRs (fitting and inference) and the ground truth are compared in Fig. 2 (e) and Fig. 2 (f). The two distributions almost coincide.

It is more realistic to train the model with fewer data entries, and this additionally verifies the inherent adaptability of the model to the QoT estimation task. As shown in Fig. 2 (g) and Fig. 2 (h), if we consider that $\log_{10}\text{MSE}$ less than -1 is sufficiently good, then for a *JPN12*-scale and a *EURO28*-scale network, the performance with 20% and 30% of the training data is acceptable, respectively.

From Fig. 2 (i) and Fig. 2 (j), it can be seen that if the training is conducted under a larger I, the generalization ability of the model on other I is stronger. The reason is that the model can learn the estimation of more link status under a larger I (i.e., different channel distributions in a link). Conversely, the model performance becomes unacceptable when training under a smaller I. These results also suggest the strong generalization ability of *PNPAConv*.

Discussion. Q1: Does *PNPAConv* really includes any service routes into consideration? The feature designs for a node and an edge have already recorded the source-target pairs of the corresponding services. The directional influence modeling makes *PNPAConv* recognize how the services are transmitted locally, thereby the network-scale routes are recognized through iterations. In addition, *PNPAConv* itself can learn the shortest paths[13]. **Q2:** Can *PNPAConv* be extended to estimate QoT in another scenario? We have proved that the mechanism for *PNPAConv* to complete the QoT estimation task is scenario-adaptive, thus QoT dataset preparation is the only requirement for *PNPAConv* to adapt to a more complex scenario such as multi-band OTN.

Conclusions

Considering the logic of the OTN-scale QoT estimation, we modified the powerful *PNAConv* into *PNPAConv* for adaptability enhancement. Our model takes the service requests and routes of the entire OTN as inputs, and directly infers the QoT values of the services. The estimation performance $\log_{10}\text{MSE}$ achieves -1.982 and -1.506 over *JPN12* and *EURO28*, respectively. The model still achieves acceptable estimation performance with less training data. In addition, we demonstrated the generalization ability of the model under different service intensities. While using advanced neural models to solve OTN problems, this paper also reversely promotes the development of machine learning.

References

[1] R. Yang, S. Shen, H. Li, *et al.*, "986 km field trial of cascaded ANN-based link-penalty models for QoT prediction", in *Optical Fiber Communication Conference (OFC) 2023*, Optica Publishing Group, 2023, W4G.5. [Online]. Available: `https://opg.optica.org/abstract.cfm?URI=OFC-2023-W4G.5`.

[2] Y. Zhang, Y. Song, Y. Shi, *et al.*, "Deep neural network-enabled fast and large-scale QoT estimation for dynamic C+L-band mesh networks", in *Optical Fiber Communication Conference (OFC) 2023*, Optica Publishing Group, 2023, Tu2F.4. [Online]. Available: `https://opg.optica.org/abstract.cfm?URI=OFC-2023-Tu2F.4`.

[3] H. Gao, X. Chen, L. Sun, and Z. Li, "Compqote: Generalizing qot estimation with composable ml and end-to-end learning", in *Optical Fiber Communication Conference (OFC) 2023*, Optica Publishing Group, 2023, W4G.1. [Online]. Available: `https://opg.optica.org/abstract.cfm?URI=OFC-2023-W4G.1`.

[4] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, "Neural message passing for quantum chemistry", in *Proceedings of the 34th International Conference on Machine Learning - Volume 70*, ser. ICML'17, Sydney, NSW, Australia: JMLR.org, 2017, 1263â1272.

[5] K. Rusek, J. Suárez-Varela, P. Almasan, P. Barlet-Ros, and A. Cabellos-Aparicio, "RouteNet: Leveraging graph neural networks for network modeling and optimization in SDN", *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 10, pp. 2260–2270, 2020. DOI: `10.1109/JSAC.2020.3000405`.

[6] G. Corso, L. Cavalleri, D. Beaini, P. Liò, and P. Velickovic, "Principal neighbourhood aggregation for graph nets", in *Proceedings of the 34th International Conference on Neural Information Processing Systems*, ser. NIPS'20, Vancouver, BC, Canada: Curran Associates Inc., 2020, ISBN: 9781713829546.

[7] A. Prakash and S. Kar, "Supervised graph convolution networks for OSNR and power estimation in optical mesh networks", *Journal of Optical Communications and Networking*, vol. 14, no. 6, pp. 469–480, 2022. DOI: `10.1364/JOCN.455539`.

[8] T. Tachibana, Y. Hirota, K. Suzuki, T. Tsuritani, and H. Hasegawa, "Metropolitan area network model design using regional railways information for beyond 5g research", *IEICE Transactions on Communications*, vol. E106.B, no. 4, pp. 296–306, 2023. DOI: `10.1587/transcom.2022EBN0001`.

[9] M. O'Mahony, "Ultrahigh capacity optical transmission network: European research project cost 239", *Information, Telecommunications, Automata Journal*, vol. 12, pp. 33–45, 1993.

[10] J. Müller, F. Slyne, K. Kaeval, *et al.*, "Experimental demonstration of ml-based dwdm system margin estimation", in *Optical Fiber Communication Conference (OFC) 2023*, Optica Publishing Group, 2023, Tu2F.2. [Online]. Available: `https://opg.optica.org/abstract.cfm?URI=OFC-2023-Tu2F.2`.

[11] A. Mestres, E. Alarcón, Y. Ji, and A. Cabellos-Aparicio, "Understanding the modeling of computer network delays using neural networks", in *Proceedings of the 2018 Workshop on Big Data Analytics and Machine Learning for Data Communication Networks*, ser. Big-DAMA '18, Budapest, Hungary: Association for Computing Machinery, 2018, 46â52, ISBN: 9781450359047. DOI: `10.1145/3229607.3229613`. [Online]. Available: `https://doi.org/10.1145/3229607.3229613`.

[12] F. Geyer, "Deepcomnet: Performance evaluation of network topologies using graph-based deep learning", *Performance Evaluation*, vol. 130, pp. 1–16, 2019, ISSN: 0166-5316. DOI: `https://doi.org/10.1016/j.peva.2018.12.003`. [Online]. Available: `https://www.sciencedirect.com/science/article/pii/S0166531618300944`.

[13] P. Veličković, R. Ying, M. Padovano, R. Hadsell, and C. Blundell, "Neural execution of graph algorithms", in *International Conference on Learning Representations*, 2020. [Online]. Available: `https://openreview.net/forum?id=SkgKO0EtvS`.

[14] P. Team, *Torch_geometric.nn*. [Online]. Available: `https://pytorch-geometric.readthedocs.io/en/latest/modules/nn.html`.

[15] A. Vaswani, N. Shazeer, N. Parmar, *et al.*, "Attention is all you need", in *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, *et al.*, Eds., vol. 30, Curran Associates, Inc., 2017. [Online]. Available: `https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf`.

[16] C. Wang, N. Yoshikane, and T. Tsuritani, "Towards more accurate and effective service provision in multiband transport networks", in *2022 European Conference on Optical Communication (ECOC)*, 2022, pp. 1–4.

[17] A. Ferrari, M. Filer, E. Le Rouzic, *et al.*, "Gnpy: An open source planning tool for open optical networks", in *2020 International Conference on Optical Network Design and Modeling (ONDM)*, 2020, pp. 1–6. DOI: `10.23919/ONDM48393.2020.9133027`.