Transfer Learning from Unbiased Training Data Sets for QoT Estimation in WDM Networks

Jelena Pesic(1), Matteo Lonardi(1), Emmanuel Seve(1), Nicola Rossi(2), Thierry Zami(2)

(1) Nokia Bell Labs, France jelena.pesic@nokia.com
(2) Nokia, France

Abstract *We assess the benefits of transfer learning based on artificial neural networks (ANN) using unbiased training data sets for QoT-estimation of unestablished light paths. This study considers transfer learning from the CONUS topology and from an unbiased training data set to the Germany50 topology.*

Introduction

Machine Learning (ML) is a branch of artificial intelligence and has been applied to many fields. However, its application in optical networks for quality of transmission (QoT) estimation is still in its infancy. Although numerous related research papers have recently appeared [1-6]. Nowadays, designing wavelength division multiplexed (WDM) networks usually leverages an offline tool that estimates QoT from an analytical model calibrated using experiments. With the arrival of ML applied to QoT estimation, two main directions can be found in the literature. The first one uses ML to increase the accuracy of the input parameters for a traditional offline analytical QoT tool [7-8]. The second one relies on ML to partly or completely supersede the existing QoT tool [9].

In this work, we focus on the latter one and, more specifically, on potential benefits of transfer learning based on artificial neural networks (ANN) using unbiased training data sets. Indeed, transfer learning is a promising solution for speeding up the training phase when different topologies need to be learnt [10].

Unbiased data set

Our previous work[9] showed the importance not only of data representation but also of unbiased data set for all feature selections. By unbiased data set we mean a uniform distributed set of training samples at the input. If relying solely on data coming from the deployed network, the data sets will have a built-in bias in terms of the number of fibre spans per light path (LP). Indeed, usually to generate the training data set we use the existing LPs, which are most often routed along the shortest path between the source and destination nodes in the networks. If the shortest path has no available resource, the LP on the 2nd, 3rd, etc. shortest paths will be considered. After a large number of accommodated services, with this allocation method, the number of spans per LP results to be Gaussian distributed, i.e. not uniform and, therefore, biased. Hence, since the training data is normally generated according to this allocation method, which produces a biased distribution, the prediction of a supervised ML approach trained on this set will be less accurate for the least frequent types of LPs in this set, i.e. on the Gaussian distribution tails where the LPs are exhibiting relatively small or high number of spans. To offset this potential bias, in this work we built a synthetic unbiased, i.e., uniform distributed, data set based on Germany50[11] topology.

In this paper, we consider 2 WDM network topologies: CONUS[12] and Germany50[11]. We generate 5 shortest light paths bridging each pair of nodes in these 2 topologies in *go* and *return* directions. In total, we consider 5x N x (N-1) LPs,

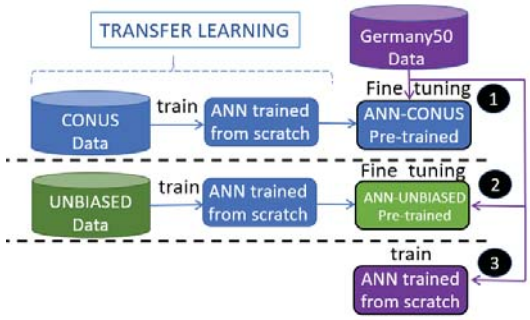

Fig. 1: Three compared cases: (1) Transfer leaning from CONUS topology and fine-tuned with Germany50 (2) Transfer learning from unbiased data set and fine-tuned with Germany50 (3) Training and testing with Germany50 data set with no transfer learning

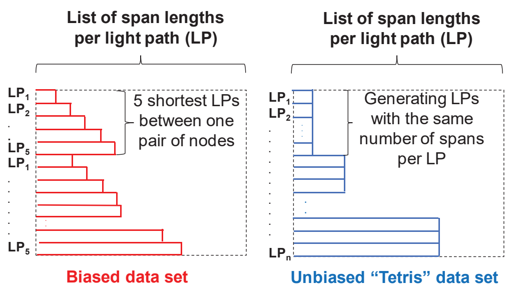

Fig. 2: Data set: (a) Biased and (b) Unbiased (Tetris)

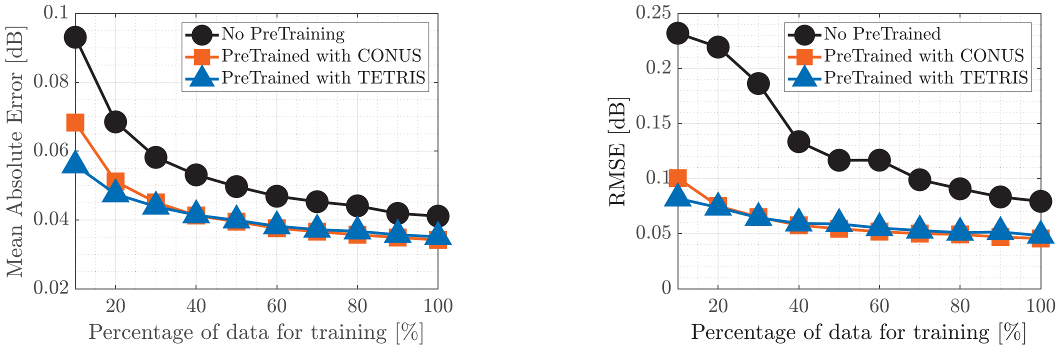

Fig. 3: (a) Mean Absolute Error (MAE) and (b) Root Mean Square Error (RMSE) on the G-OSNR estimation

where N is the number of nodes in each topology. Therefore, for CONUS (where N=75) we consider 5x75x74=27750 LPs and for Germany50 topology (where N=50) we consider 5x50x49=12250 LPs.

We also create a synthetic unbiased data set based on the Germany50 topology data set with the same total number of 12250 LPs and the same maximum number of spans in the longest LP (30). Unlike the previous set based on 5-shortest paths, this data set has uniformly distributed number of LPs in terms of number of spans per LP (Fig.2). This way we avoid having dramatically smaller frequency of LPs with a very small or very high number of spans. Moreover, span lengths range uniformly between 30 km and 80 km.

At the ANN's input, each LP is presented as a list of span lengths and at the output its targeted the generalized optical-signal-to-noise-ratio (G-OSNR) value obtained using a Gaussian noise (GN)-model. We consider both linear noise due to Erbium doped-fiber amplifiers (EDFA) and nonlinear effects along the WDM transmission, as well as optical filtering penalties. Based on the estimated G-OSNR, we allocate the most spectrally efficient modulation format out of the 9 considered 32 GBaud channel modulations from 100 Gb/s to 300 Gb/s, with steps of 25 Gb/s. In this study, we consider homogeneous amplifier noise figure (NF = 4.5 dB) and fiber attenuation (0.22 dB/km) along the transmission. Each transparent node traversal is represented in the data set by a span of 80 km. More details on the physical model used can be found in [9]. A feed-forward shallow artificial neural network (ANN) with 20 fully connected sigmoid-activated neurons for the hidden layer and an identity-activated neuron at the output layer is trained with the Levenberg-Marquardt backpropagation algorithm with 70%/15%/15% division for training/validation/testing.

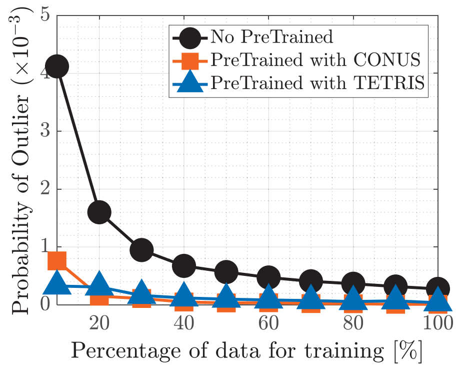

Fig. 4: Probability of outliers

Simulations

In this work, we compare QoT estimation accuracy of LPs belonging to Germany50 using 3 different cases (Fig.1). In the first one, we train the ANN from scratch using CONUS data set and this pre-trained ANN-CONUS is being fine-tuned using data sets from Germany50. In the second one, we also use transfer learning but this time instead of using CONUS data set, as previous knowledge, we pre-train ANN from scratch with the created synthetic unbiased data set and fine tuning is done in the same way as in the first method, using the Germany50 data set. In the third, reference case, we do not apply any transfer learning, i.e., we perform a single-phase training (no dual phase with pre-training). Instead, ANN is being trained from scratch with Germany50 data set. In all 3 cases, for the sake of fair comparison, we test ML-based QoT estimation using the same testing data set from Germany50.

We are interested in the potential benefits of transfer learning in the first two cases and moreover of transfer learning with unbiased data sets compared to using ANN trained from scratch.

We train all ANNs using the same data, which is 70% out of (10%-100%) from Germany50 data set and we test using the non-used data in training phase. This way, for example, in the case of training (fine-tuning) with 50% of Germany50 data set, 70% out of 6125 LPs are used for training, 15% for validation and 15% for testing. Note that since we are considering a list of span lengths as a feature selection when representing a LP as input to ANN, each topology has a different number of input ANN values which corresponds to the max number of spans. In order to pretrain ANN with CONUS topology (max number of spans 106) and to test with Germany50 (max number of spans 30), we had to artificially add 76 (106-30) ANN input values, which were padded to zero, as it is usually done for shorter-than-maximum links.

Results

The results of the ML-based QoT tool are assessed in the form of absolute error, comparing to the ML-estimated G-OSNR to the one given by the analytical GN-model (ground truth).

Mean absolute error (MAE), root mean square error (RMSE) and probability of outliers are used as statistical metrics to measure ANN model performance. Outliers are the outcomes where the absolute error in G-OSNR estimation exceeds 1dB. We define the probability of outliers as the number of outliers divided by the number of total testing estimations. Results are averaged on 1000 runs, where training, validation and testing are done on different ANN realizations.

Fig.3 (a) shows the mean absolute error for G-OSNR estimation for 3 cases: ANN pretrained with CONUS, ANN pretrained with unbiased (Tetris) data set based on Germany50 data set size, and for ANN with no pretraining, in a function of the percentage of data for training (size of the Germany50 training data set). Note that 10% of data for training, in Fig.3, corresponds to 858 training samples (70% out of 10% of 12250 LPs) and 184 testing LPs (15% out of 10% of 12250 LPs). As expected, because of the bias in the Germany50 data set, MAE is the highest for a case with no transfer learning. The smallest MAE is obtained when pretraining ANN with unbiased data set based on Germany50.

It is interesting to notice that for smaller percentages of data for training, when pretrained with CONUS data set, which consists of 70% of total number of CONUS LPs (19425 LPs) MAE is higher than when pretrained with unbiased Germany50 based data set, which consists of 70% of total number of Germany50 LPs (8575 LPs). This means that pretraining with lower number of data can be more beneficial, if chosen carefully. It is useful to recall that CONUS data set is still biased, even though its bias results in a smaller error because of its bigger size. This MAE difference decreases along with the growth of the size of the training data. However, in practice the likelihood to be able to access to a large training data sets is low.

Fig.3 (b) shows the root mean square error for G-OSNR estimation for the same 3 cases presented in Fig.3 (a). Similarly, as in a case of MAE, we show the highest RMSE when no ANN pretraining and the smallest RMSE, for a small training data set when pretrained with the unbiased data based on Germany50. For a higher number of training LPs, the RMSE in case of pretraining with CONUS and unbiased Germany50 based data set converge.

Using pretrained ANN for G-OSNR estimation reduces not only the MAE and RMSE but also decreases the number of outliers. Fig.4. shows the distribution of outliers for all 3 cases. With the increase of training data set, the number of outliers is also decreasing. When pretraining ANN with unbiased data set, especially designed to reduce the number of outliers by having the uniformly distributed number of LPs, in terms of number of spans, we eliminate the possibility of not having sufficient number of samples of LPs with certain number of spans in the training data set.

Conclusions

We have shown the importance of ANN pretraining and transfer learning when using ANN for QoT estimation in WDM networks. We have also shown that the benefits are even higher if ANN is pretrained with the unbiased data set. This benefit is also visible even if the size of the pretrained unbiased data set is more than two times smaller than pretrained biased data set. This is a particularly important information when there is a lack of data for training.

References

[1] Dario Azzimonti, Cristina Rottondi, and Massimo Tornatore, "Reducing probes for quality of transmission estimation in optical networks with active learning," J. Opt. Commun. Netw. 12, A38-A48 (2020)

[2] M. Bouda, S. Oda, Y. Akiyama, D. Paunovic, T. Hoshida, P. Palacharla, and T. Ikeuchi, "Demonstration of Continuous Improvement in Open Optical Network Design by QoT Prediction using Machine Learning," in Optical Fiber *Communication Conference (OFC) 2019*, OSA Technical Digest (Optical Society of America, 2019), paper M3Z.2.

[3] R. Proietti, X. Chen, K. Zhang, G. Liu, M. Shamsabardeh, A. Castro, L. Velasco, Z. Zhu, and S. J. Ben Yoo, "Experimental Demonstration of Machine-Learning-Aided QoT Estimation in Multi-Domain Elastic Optical Networks with Alien Wavelengths," J. Opt. Commun. Netw. 11, A1-A10 (2019)

[4] T. Panayiotou, G. Savva, B. Shariati, I. Tomkos and G. Ellinas, "Machine Learning for QoT Estimation of Unseen Optical Network States," *2019 Optical Fiber*

Communications Conference and Exhibition (OFC), San Diego, CA, USA, 2019, pp. 1-3.

[5] I. Sartzetakis, K. K. Christodoulopoulos and E. M. Varvarigos, "Accurate quality of transmission estimation with machine learning," in IEEE/OSA Journal of Optical Communications and Networking, vol. 11, no. 3, pp. 140-150, March 2019. doi: 10.1364/JOCN.11.000140

[6] Tremblay, C.; Allogba, S.; Aladin, S., "Quality of Transmission Estimation and Performance Prediction of Lightpaths Using Machine Learning," European Conf. on Optical Communication (ECOC'19), Dublin, Ireland, September 22-26, 2019, invited.

[7] E. Seve, J. Pesic, C. Delezoide, S. Bigo, and Y. Pointurier, "Learning Process for Reducing Uncertainties on Network Parameters and Design Margins," J. Opt. Commun. Netw. 10, A298-A306 (2018)

[8] E. Seve, J. Pesic, and Y. Pointurier, "Accurate QoT Estimation by Means of a Reduction of EDFA Characteristics Uncertainties with Machine Learning" *Optical Network Design and Modeling* (ONDM2020)

[9] J. Pesic *et al.,* "How Uncertainty on the Fiber Span Lengths Influences QoT Estimation Using Machine Learning in WDM Networks," in Optical Fiber Communication Conference (OFC) 2020, paper Th3D.5.

[10] J. Pesic et al., “Transfer Learning Using ANN for G-OSNR Estimation in WDM Network Topologies” in Advanced Photonics Congress, paper NeM3B.3, 2020.

[11] http://sndlib.zib.de/home.action

[12] http://monarchna.com/topology.html