Transfer Learning across Different Lightpaths for Failure-Cause Identification in Optical Networks

Francesco Musumeci(1)*, Virajit G. Venkata(1), Yusuke Hirota(2), Yoshinari Awaji(2), Sugang Xu(2), Masaki Shiraiwa(2), Biswanath Mukherjee(3), Massimo Tornatore(1,3)

(1) Politecnico di Milano, Milan, Italy, *Corresponding author: francesco.musumeci@polimi.it,
(2) National Institute of Information and Communications Technology, Tokyo, Japan,
(3) University of California, Davis, CA-USA.

Abstract *We perform transfer learning across different lightpaths for failure-cause identification using OSNR traces collected over NICT's Sendai optical-network testbed. Results suggest that limited additional data on the target lightpath allow to achieve satisfactory accuracy.*

Introduction

One of the most attractive applications of Machine Learning (ML) in optical networks is the automation of failure management and troubleshooting. Several studies have already demonstrated the potential of ML in performing failure detection, failure-cause identification, failure localization and even failure prediction[1]. Employed ML techniques come typically from the field of supervised learning and are based on historical monitoring of signal-quality data, e.g., Optical Signal-to-Noise Ratio (OSNR) and/or Bit Error Rate (BER), made available by modern receivers or by Optical Spectrum Analysers (OSAs). The basic idea is that ML models learn the "signature" of past failures from historical data, then this signature is recognized in future occurrences of similar failures.

Although the ability of ML to recognize such failure signatures on a single lightpath has been confirmed in several studies, a common objection to practical ML deployments for failure management is: *does a ML model trained to perform, e.g., failure-cause identification, on a lightpath still work on a different lightpath?* In fact, if complete re-training of the ML model is necessary each time the model is applied to a different lightpath, the amount of training overhead to make this system operational in a large network will be heavy.

We investigate on the use of *transfer learning (TL)*[2] techniques to reduce the amount of additional data required to re-train a failure-management model (originally intended for a lightpath) in order to apply it to a different lightpath. In many ML applications, ML models are developed and validated exploiting a dataset, i.e., the source domain (SD), which is substantially different from the target domain (TD), where the trained models are deployed. Consider, e.g., failure-cause identification, i.e., understanding, by observing a lightpath's OSNR at the receiver, what is the cause of a failure on a lightpath traversing a set of optical fiber links and devices, such as optical amplifiers, filters and Reconfigurable Optical Add/Drop Multiplexers (ROADMs). Although extremely accurate classifiers can be designed to distinguish between distinct failure causes, the same model might not be accurate when applied to a new lightpath having, e.g., different path length, number of ROADMs traversed, types/number of optical amplifiers along the route, used wavelength, etc. Hence, collecting new data in the TD to (re)train the ML model is the most common, though costly, alternative. In particular, this aspect is critical in failure management, as purposely introducing lightpath malfunctioning to collect real training data from the field is neither practical nor desirable for network operators. Conversely, it might be desirable to leverage existing knowledge (i.e., very accurate ML models) obtained on the SD and exploit only a small amount of data collected from the TD to fine-tune the original models. This approach would enable faster and more efficient adaptation of ML models to various lightpaths as it would allow not only to reduce the amount of data (and consequently, e.g., storage and computing resources) needed, but also to substantially reduce the time required for new training when lightpaths characteristics change, due to the much smaller datasets used for model fine-tuning.

TL has been already investigated in other optical networking contexts at both physical and network layers[3]–[6]. To the best of our knowledge, for the first time we focus on TL for *failure-cause identification*. We collect historical data on a lab testbed to perform TL across different lightpaths, considering as source and target domains lightpaths with different number of traversed ROADMs (i.e., number of hops) and optical fiber links. Our results suggest that, using TL from an already trained lightpath (source domain), satisfactory accuracy in the target domain/lightpath can be achieved even with a limited amount of additional data on the target domain/lightpath.

Data Collection over NICT's Sendai Testbed

We perform ML-based failure-cause identification using real data obtained on a testbed of the National Institute of Information and Communications Technology (NICT) located in Sendai (Japan). The testbed (see Fig. 1) consists of 4

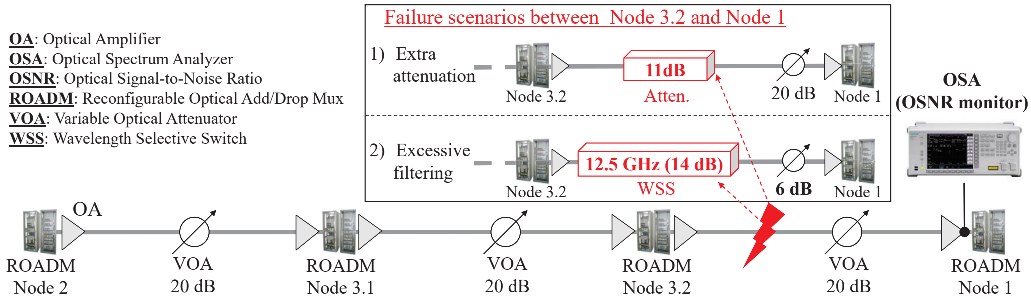

Fig. 1: High-level scheme of NICT's Sendai Testbed setup and emulated failure scenarios.

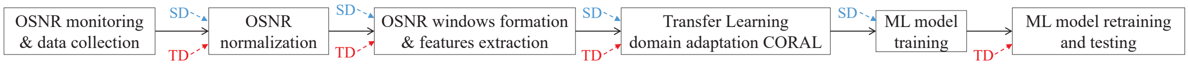

Fig. 2: Steps of the TL-based failure identification (SD: Source Domain; TD: Target Domain).

ROADMs, identified as Node 2, Node 3.1, Node 3.2 and Node 1, interconnected through optical fibers, and equipped with one pre-amplifier and one booster (OA in the figure) at their input and output, respectively. Each fiber link can emulate fiber spans of up to 80 km using a Variable Optical Attenuator (VOA) with maximum 20 dB attenuation. Two failure scenarios, i.e., 1) *Extra attenuation* and 2) *Excessive filtering*, are emulated by including in the last fiber span (i.e., between nodes 3.2 and 1) either an attenuator with extra 11 dB attenuation (emulating the "Extra attenuation" scenario) or a Wavelength Selective Switch (WSS) with passing bandwidth of 12.5 GHz (emulating the "Excessive filtering" scenario). In the latter case, due to the WSS insertion loss of 14 dB, the VOA attenuation is reduced to 6 dB to compensate the effect of the WSS on the last span overall attenuation.

We consider three lightpaths, deployed one at a time, all having Node 1 as the receiver node. At Node 1 an OSA is placed after the pre-amplifier to monitor OSNR every $T_{OSNR} = 1s$. The three lightpaths are characterized by different hop lengths, i.e., lightpath 1 (LP1) between Node 3.2 and Node 1, LP2 between Node 3.1 and Node 1, and LP3 between Node 2 and Node 1, have length of 1, 2 and 3 hops, respectively. The three lightpaths are transmitted at the same wavelength with central frequency 194.8 THz and 100 GHz bandwidth. A 10 Gbit/s signal is transmitted using OOK modulation format.

Our goal is to show that we can train ML models on one lightpath (e.g., LP1) and make it work, with minimal re-training, also on the other lightpath (e.g., LP3).

Problem Definition and TL Methodology

We model the failure-cause identification as a binary classification problem, where we are given the OSNR monitored at the lightpath receiver and discriminate among two failure-cause classes, i.e., *Extra attenuation* vs *Filtering*.

A summary of our methodology is shown in Fig. 2. After collecting OSNR traces for the two failure scenarios (attenuation and filtering) and the available lightpaths LP1, LP2 and LP3, source and target domains (SD and TD, respectively) are identified. As an example, data in SD can be extra-attenuation and filtering OSNR data for LP1 and data in TD can be extra-attenuation and filtering OSNR data for LP3. To make our ML-based classifiers independent from the specific OSNR values obtained in the various cases, which may vary according to system settings (e.g., OA gain, span length, central wavelength, etc.), OSNR values are normalized to lie in the [0-1] range.

Classification is performed considering OSNR *windows* of duration W seconds, and collected at $T_{OSNR} = 1s$ sampling period. Therefore, after OSNR normalization, we form windows and create our dataset of training and testing samples. For each OSNR window, 16 features are extracted as in[7], i.e., (1-4) minimum, maximum, mean and standard deviation of OSNR values in the window, (5) peak-to-peak, i.e., difference between maximum and minimum OSNR in the window, (6) OSNR root mean square value in the window, and finally (7-16) the ten strongest spectral components in the window, obtained by applying Fast Fourier Transform on the OSNR window.

The core part of the TL procedure is then applied to SD and TD data. Several TL approaches exist[8]. Here we adopt the feature-based correlation algorithm[9], called CORrelation ALignment (CORAL), where, before performing model training, the feature values in SD are modified with the objective of minimizing the distance between the covariance of the SD and TD features. After CORAL is applied to TD and SD data, ML model selection and training is performed. We consider ML classifiers based on artificial neural networks (ANN) where the hyperparameters (i.e., number of hidden layers, number of hidden neurons and activation function) are selected by applying 5-fold crossvalidation on SD data. After testing several different combinations of hyperparameters, the best performing ANNs have 2 hidden layers, each with 9 hidden neurons and Relu activation function, whereas sigmoid activation functions has been used in the output layer.

Finally, the trained model is fine-tuned with TD

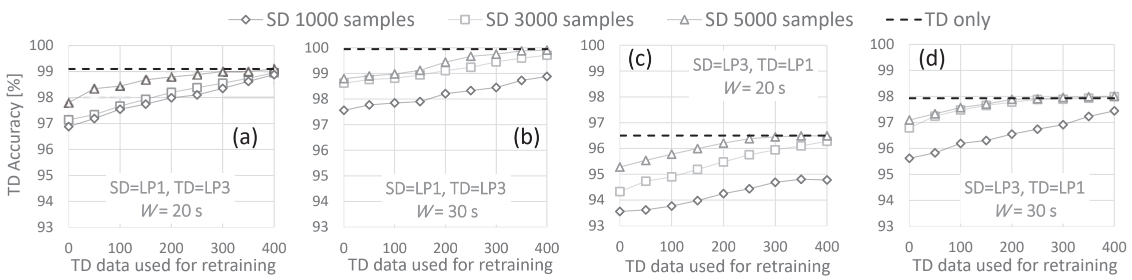

Fig. 3: TD accuracy for increasing amount of TD data used for ML model retraining.

data, and the accuracy of failure identification applied to transfer the knowledge from SD to TD is evaluated considering an independent test set (not used for model fine-tuning) extracted from the TD data. As ANNs are as ML classifiers for our CORAL-based TL approach, fine-tuning of the ANN model through TD data consists of performing additional steps of the ANN training to update ANN weights to fit the TD data distribution, starting from the knowledge (i.e., the ML model) obtained after the initial training using SD data. Our objective is to find under which conditions it is convenient to apply TL instead of performing algorithm training with TD only.

Numerical results

We perform our numerical analyses considering the testbed in Fig. 1, and collect OSNR data for 6 cases (i.e., three lightpaths (LP1, LP2 and LP3) each of which can get the two failed state, due to either excessive attenuation or filtering). In all cases, OSNR samples are collected for 6 hours at a sampling period of $T_{OSNR} = 1s$, so the entire dataset is constituted by 36 hours of OSNR monitoring. We vary the window duration, and alternatively consider LP1 and LP3 as different lightpaths to represent the SD and TD. Similar results, which we do not show for space constraints, have been obtained considering also LP2.

Figure 3 shows failure-identification accuracy obtained using TL. Two distinct window sizes have been considered, i.e., $W = 20s$ and $W = 30s$. In each subfigure, different curves represent different amount of SD training data (namely, 1000, 3000 and 5000 training samples), while the x-axis represents increasing amount of TD data (from 0 and 400 samples) used for retraining after knowledge transfer. Note that the case with 0 data from the TD (i.e., the origin of x-axis) represents the situation when we assume no knowledge of the TD, therefore, in these cases, we do not pre-process SD and TD features with CORAL algorithm. In each case we compare the TL cases with the case when the ML classifiers are trained using only 5000 samples from the TD and no data from the SD ("TD only" in the figure, i.e., the horizontal lines) which can be used as benchmark classification accuracy. As a first observation, looking at the benchmark cases in the four subfigures, the cases TD=LP3 (Figs. 3a,b) achieve higher accuracy compared to the cases TD=LP1 (Figs. 3c,d), independently of the window size. This is due to the difference between OSNR variations over time in the two failure cases (attenuation vs filtering), which is exacerbated when noise accumulates over a longer lightpath.

When applying TL, in all cases, accuracy increases with the number of TD data used for retraining and, in most cases, it is sufficient to use 100-200 TD data points to achieve a satisfactory retraining of the original model. This is noticeable when comparing this result with the much larger (i.e., up to 50 times higher) amount of data of the "TD only" scenarios, where 5000 samples of the TD are necessary to achieve high accuracy. As expected, accuracy is higher for increasing number of SD data used for initial learning, and this aspect is more critical for smaller window size (see Figs. 3a,c), when it is important to have a solid starting accuracy provided by the SD (i.e., corresponding to the values of 0 in the *x* axes) before fine-tuning with TD data. This demonstrates that, in practical network deployments, most of the effort in developing accurate ML-based failure-cause identification can be made offline by using historical data collected from one or more different SDs, instead of retraining from scratch in the new deployment (i.e., the TD). In other words, the original ML models can be directly adopted in the TD during the initial phase of its deployment, and, as few new historical data is available also in the TD, the ML can be fine-tuned to improve its performance.

Conclusion

We studied the effectiveness of TL for failure-cause identification, generating OSNR traces emulating normal and failed lightpath states (due to excess filtering or attenuation) in a 10 Gbit/s testbed. In our case study, by re-training the source ML model with only few hundreds seconds of additional OSNR data coming from the target domain, TL achieves the same accuracy in discriminating between filtering and attenuation obtained when using a much larger (up to 50 times higher) dataset using the target domain alone.

Acknowledgments: The work leading to these results has been supported by the European Community under grant agreement no. 761727 Metro-Haul project and by USA National Science Foundation Award no. 1818972 (JUNO2 program).

References

[1] B. Shariati *et al.*, "Learning From the Optical Spectrum: Failure Detection and Identification", *IEEE/OSA J. Lightwave Technol.*, vol. 37, no. 2, pp. 433–440, 2019.

[2] S. J. Pan and Q. Yang, "A Survey on Transfer Learning", *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.

[3] C.-Y. Liu *et al.*, "Evol-TL: Evolutionary Transfer Learning for QoT Estimation in Multi-Domain Networks", in *Proc. OFC 2020*, Mar. 2020.

[4] Z. Xu *et al.*, "Transfer Learning Aided Neural Networks for Nonlinear Equalization in Short-Reach Direct Detection Systems", in *Proc. OFC 2020*, Mar. 2020.

[5] X. Chen *et al.*, "Exploiting Multi-Task Learning to Achieve Effective Transfer Deep Reinforcement Learning in Elastic Optical Networks", in *Proc. OFC 2020*, Mar. 2020.

[6] J. Pesic *et al.*, "Transfer Learning Using ANN for G-OSNR Estimation in WDM Network Topologies", in *Proc. OFC 2020*, Mar. 2020.

[7] F. Musumeci *et al.*, "A Tutorial on Machine Learning for Failure Management in Optical Networks", *IEEE/OSA J. Lightwave Technol.*, vol. 37, no. 16, pp. 4125–4139, 2019.

[8] C. Tan *et al.*, "A Survey on Deep Transfer Learning", in *Proc. Artificial Neural Networks and Machine Learning â ICANN 2018*, Oct. 2018.

[9] B. Sun *et al.*, "Return of frustratingly easy domain adaptation", in *Proc. AAAI 2016*, Feb. 2016.