Interactive Visual Analytics Dashboard for the Paradigm of ML-assisted Autonomous Optical Networking

Behnam Shariati, Wanda Baltzer, Geronimo Bergk, Pooyan Safari, and Johannes Karl Fischer

Fraunhofer Institute for Telecommunications Heinrich Hertz Institute, Einsteinufer 37, 10587 Berlin, Germany
Corresponding author's email address: behnam.shariati@hhi.fraunhofer.de

Abstract: We demonstrate a novel visualization dashboard, compatible with multiple data and telemetry sources, which offers dataset quality evaluation, dataset comparison, ML model error analysis interpretation, and network health monitoring. © 2020 The Author(s)

1. Overview

Machine Learning (ML) is becoming an inevitable part of optical network evolution and has thus far found application in various aspects of the optical transmission and networking ecosystem [1-4]. Most of the work in the literature focus on the ML model development, primarily, relying on in-house generated datasets collected in simulation [5], limited lab/field experiments [3], or in rare occasions, real networks [6]. These efforts have to be complemented with other major developments to allow the realization of carrier-grade ML-assisted network automation.

On the one hand, in addition to the numerous out-of-field [7] model-training studies, investigation and development of different stages of an end-to-end ML pipeline for optical network automation has gained a significant momentum. In this regard, there has been a special focus on the development of telemetry retrieval protocols and data models [8] as well as telemetry brokers [9] to allow data producers (i.e., network elements) and consumers (e.g., different stages of the ML pipeline [10]) to interact with each other. There is also a limited number of work on the development of ML function orchestration [11] and inter-operator ML model sharing and trading platforms [12], which contribute to the end-to-end ML pipeline for optical network automation.

On the other hand, the current unavailability of standard and comprehensive reference datasets as well as the inconsistency in the considered datasets across different studies in the field, have spurred multiple research groups and organizations to focus on reference dataset generations [13][14][15]. However, they are quite limited and cannot address the massive problem space of ML-assisted solutions for optical network automation.

Regardless of all the above-mentioned ongoing developments, a critical aspect has largely remained unexplored, which is the explainability of the underlying datasets and the delivered output of the ML models. As concluded in a whitepaper by European Telecommunications Standards Institute [16], ETSI, the quality of datasets, trustworthiness, and explainability of ML models are key requirements for an ML-assisted solution to be considered certifiable in the field.

In this work, we demonstrate a multi-source multi-purpose *Visual Analytics Dashboard* that offers a multitude of features and interactive visualizations allowing dataset quality evaluation, ML model outcome explanation, visualized error analysis interpretation, and observing network-wide lightpath characteristics. The dashboard consumes data from multiple sources. These sources include: i) an InfluxDB network database [17], which interacts with a Kafka-based telemetry broker in the back-end [9], ii) real-time retrieval and interaction with HHI's network planning tool, PLATON [12], iii) a Traffic Engineering Database (TED) [12], representing the historical database of an Operation Support System (OSS), and iv) the publicly available QoT dataset collection of HHI [13]. We believe our proof-of-concept demonstration will shed light on an unexplored corner of the ML-assisted network automation journey and paves the way for its consideration by the key players in the ecosystem, like operators and vendors.

We provide the details of the demonstration in section 2 and present its relevance to OFC in section 3.

2. Innovation

We aim at a proof-of-concept demonstration of a visual analytics dashboard with novel features and capabilities making the development of ML-assisted solutions explainable. The dashboard offers various interactive visualizations (e.g., network topologies, violin plots, Sankey diagrams, bar charts, heat maps, etc.) and tabular boxes (e.g., to report characteristics of the network links, inline amplifiers, active lightpaths, etc.). In the demonstration, we target specifically the following use-cases and applications in the context of explainable ML-assisted network automation:

I. *Dataset Quality Evaluation*: The dashboard offers highly informative two-dimensional interactive plots with multi-categorical feature distributions revealing nontrivial aspects of the dataset; such as, whether and how one class might be biased towards certain feature values of the dataset. The visualizations are accompanied by spatial mapping on an interactive network topology. The spatial mapping can be configured to show a single lightpath or all active lightpaths at a specific time. The dashboard also reveals easy-to-grasp insights of the lightpath characteristics across the links and nodes of the network, thus offering a visualized interpretation of the lightpath features. In this scenario, the dashboard consumes data from a publicly available QoT dataset collection [13].

II. *Dataset Comparison and Augmentation*: Even though datasets might look similar when considering only some basic statistics of their features (e.g., mean, variance, median), they could be significantly different as shown in [18]. The dashboard allows comparing datasets and makes it possible to explain the performance of the designed ML models while trained on different datasets. Furthermore, it allows deriving guidelines for data augmentation by revealing the poor regions of the datasets with respect to the underlying features. This is quite useful for inter-network ML model transfer and (re-)training. In this scenario, the dashboard consumes data from a publicly available QoT dataset collection [13].

III. *Error Analysis Interpretation*: ML models may perform badly when facing unseen samples. This can be common in optical networks, due to the numerous degrees of freedom and involved parameters (e.g., characteristics of the selected routes, fiber types, amplifiers characteristics, transceiver type and configurations such as symbol rate and modulation format) that were not fully captured in the training dataset. The dashboard provides interactive visualization of the wrongly predicted samples to allow exploring the root-cause of the wrong decision of the ML model. The provided insights can help the ML model designers to improve the ML models by incorporating the discovered root-causes. In this scenario, the dashboard reads the report of the ML model and visualizes multi-categorical plots as explained in use-case I, but only for the false positive and false negative samples.

IV. *Visualized Lightpath Provisioning and Monitoring*: The dashboard provides interactive visualizations of the lightpath establishment procedure, including end-to-end commissioning, performance monitoring (in terms of OSNR, SNR, and pre-FEC BER), and eventual provisioning of the service. In this scenario, the dashboard interacts with the Path Computation Element (PCE) of PLATON [12], as an optical network emulator. The performance metrics are calculated, using a Gaussian Noise based QoT estimator similar to [12], stored in a Kafka-based telemetry broker, and then consumed by the dashboard. All other network status related details (e.g., route, transceiver configuration, or already active lightpaths in the network, etc.) are retrieved from the TED of PLATON. In the next development phase, our plan is to integrate our solution with OSS layer of open transport SDN architecture, such the one envisioned in TIP [19], instead of our in-house network emulator. This use-case of the dashboard is beneficial for explainable optical network monitoring and operation, regardless of exploiting ML-assisted solutions.

V. *Network Health Monitoring*: The dashboard allows interactive visualization of online OSS databases (e.g., TEDs) and topology related characteristics. The benefits include: i) revealing feature-wise utilization of the links, ii) identifying strategic links/nodes that act as the arteries of the network, iii) anticipating the impact of a potential line failure (e.g., fiber cut) on the running lightpaths and the eventual loss, and iv) monitoring *beginning-of-life* to the *end-of-life* lightpath performance.

Fig. 1 and Fig. 2 provide a glimpse of the visual analytics dashboard. Fig. 1 shows how the visual analytics dashboard looks like and what it offers when in health monitoring mode. Fig. 2 presents some other interactive visualizations that are incorporated to perform dataset quality evaluation and comparisons.

3. OFC Relevance

We believe our demonstration proposes a novel methodology and a set of practical tools to make the monitoring and operation of optical networks explainable. The demonstrated tools will eventually become an inevitable part of the development journey towards the carrier-grade ML-assisted solutions for the realization of network automation, as they bring huge benefits for optical network practitioners, ML model designers, vendors, and operators. Regardless of their tangible technical benefits, there might be even regulatory reasons that force the operators and vendors to offer explainability of their employed ML-assisted solutions to comply with the required acts and mandates. OFC is the most proper medium to perform the planned demonstration, as most of the stockholders and the relevant technical community are present.

Fig. 1. A screenshot of the visual analytics dashboard in the health monitoring mode utilizing TED databases. The topology diagram is interactive and illustrates feature-wise link utilization to reflect the actual network status and the characteristics of the active lightpath simultaneously. Each individual link and node can be chosen to visualize various aspects of its status using feature-wise visualization such as Sankey diagrams and bar charts. In addition, floating text boxes can be activated to provide infrastructure characteristics, lightpath configurations, etc. The other modes, such as dataset evaluation mode, has other types of interactive diagrams to investigate different aspects of network datasets. Some of those plots are provided in Fig. 2.

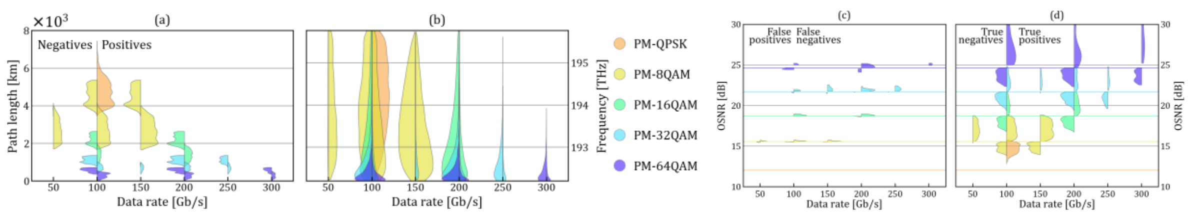

Fig. 2. One example of feature-wise multi-categorical illustration of an QoT dataset is shown for (a) path length with respect to linerate, and (b) central frequency with respect to linerate. One example of visualized error analysis for ML model performance explanation is shown representing (c) wrongly classified samples, and (d) correctly classified samples [18].

4. Acknowledgment

This work has partially received funding from the German ministry of education and research (BMBF) in the framework of the project AI-NET PROTECT (Ref. KIS8CEL010).

5. References

[1] F. Musumeci, et al., “An overview on application of machine learning techniques in optical networks,” IEEE S&T, 21(2), 2018.
[2] F. N. Khan, et al., “An optical communications’ perspective on machine learning and its applications,” JLT, 37(2), pp. 493-516, 2019.
[3] Z. Gao, et al., “ANN-based multi-channel QoT-prediction over a 563.4-km field-trial testbed,” JLT, 38(9), pp. 2646-2655, 2020.
[4] J. Yu, et al., “Machine-learning-based EDFA gain estimation,” JOCN, 13(4), pp. B83-B91, 2021.
[5] P. Safari, et al., “Deep convolutional neural network for network-wide QoT estimation,” in Proc. OFC, Jun 2021.
[6] C. Delezoide, et al., “Leveraging field data for the joint optimization of capacity and availability in low-margin optical networks,” JLT, 2020.
[7] L. Velasco, et al., “Learning life cycle to speed up autonomic optical transmission and networking adoption,” JOCN, 11(5), 2019.
[8] R. Vilalta, et al., “Experimental evaluation of control and monitoring protocols for optical SDN networks and equipment,” JOCN, 2021.
[9] A. Sgambelluri, et al., “Reliable and scalable Kafka-based framework for optical network telemetry,” JOCN, 13(10), 2021.
[10] ITU-T, “Architectural framework for machine learning n future networks including IMT-2020,” Recommendation ITU-T Y.3172, Jun 2019.
[11] A. Wassington, et al., “Implementing a machine learning function orchestration,” in Proc. ECOC, Bordeaux, France, Sep 2021.
[12] B. Shariati, et al., “Inter-operator machine learning model trading over acumos ai federated marketplace,” in Proc. OFC, June 2021.
[13] G. Bergk, et al., “QoT Dataset Collection, ” [online – accessed Oct 2021] (https://www.hhi.fraunhofer.de/networkdata)
[14] M. Ghobadi, et al., “Optical layer failure in a large backbone,” in Proc. IMC. 2016
[15] D. Ives, et al, “Dataset: Distributed abstraction and verification of an installed optical fiber network,” (https://www.repository.cam.ac.uk/handle/1810/309241)
[16] ETSI Whitepaper No. #34, “Artificial intelligence and the future directions for ETSI,” 1st Edition, Jun 2020.
[17] InfluxDB (https://www.influxdata.com/)
[18] G. Bergk, et al., “ML-assisted QoT estimation: a dataset collection and data visualization for dataset quality evaluation,” JOCN, 2022.
[19] Telecom Infra Project Whitepaper, “Open transport SDN architecture whitepaper.”