

Lightpath Length and Launch Power Estimation with Machine Learning

Shayan Hajipour

Abstract—The lightpath distance between transmitter and constellation profile observation points and launch power are predicted with linear regression and support vector machine.

I. INTRODUCTION

Digital twins (DTs) with proper abstractions could be used for deriving insights, diagnosis, planning, and carrying out drills on complex systems. An analytical DT called GNP_y is already developed for signal propagation in optical networks [1]. However, sometimes analytical DTs are hard to develop. Machine learning (ML) approach is the algorithmically simple candidate for building DTs with minimal knowledge of the subject matter.

An ML-based DT could be used to estimate lightpath (LP) attributes (e.g., launch power and distance) given In-Phase and Quadrature (IQ) optical constellation diagrams. For example, LP launch power estimation could be used to diagnose potential discrepancies between planned, telemetry, and actual launch power estimated by DT. Such attribute estimations could be performed at the pass-through observation points via optical spectrum analyzers to give a better resolution of the LP status as well as the receiver site [2]. After discrepancy detection, the adopted remediation measures could be initiated.

In this report, we estimate launch power and the distance between transmitter and observation point of lucid lightpaths.

II. DATASET

Two datasets are gathered for lightpath attributes estimation. The datasets are collected by numerical simulations of two scenario configurations:

- 1) *Single link scenario*, where LPs travel up to 25 spans deployed between a transmitter and receiver. The observation point is the receiver, where the IQ constellation profiles consisting of 2048 symbol samples are measured as datapoints. The number of spans and the length of the first span are the labeled variables of the dataset. The length of the first span could be 80 km, 60 km, or 40 km, which are called *optimal*, *sub-optimal*, and *degradation* modes, respectively. The length of other spans is 80 km. All simulation modes are combined and a dataset with the total number of 2250 datapoints of IQ constellation profiles is created. Each datapoint consists of $2048 \times 2 + 1 + 1$ columns that represent in-phase and quadrature components of symbol samples, the distance between the transmitter and the observation point, and datapoint mode.

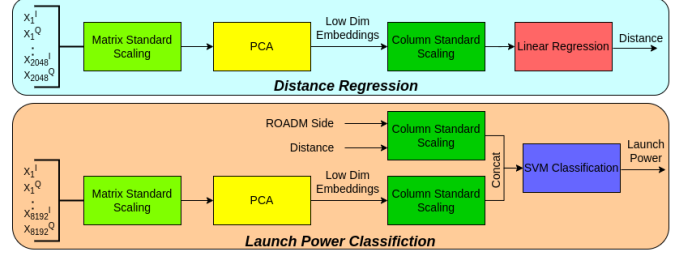


Fig. 1: Lightpath attribute estimation pipelines

- 2) *multiple links scenario*, where LPs travel 5 reconfigurable optical add-drop multiplexers (ROADMs) and 4 equal links. The number and length of spans in each link, launch power, and the observation point location are the labeled variables of the dataset. The observation point location candidates are either ingress or egress ports of ROADMs, where the IQ constellation profile consisting of 8192 symbol samples are measured as datapoints. A total number of 800 scans (datapoints) of IQ constellation profiles are collected. Each datapoint consists of $8192 \times 2 + 1 + 1 + 1$ columns that represent in-phase and quadrature components of symbol samples, the distance between the transmitter and the observation point, the ROADM side of the observation point, and the launch power. The ROADM side could be either ingress or egress, which is represented by a binary number. The launch power is either 1 or 2 dBm, which is also represented by a binary number.

The data generation simulation configurations are further elaborated in [3].

III. MACHINE LEARNING MODELS

We estimate the distance between the transmitter and the observation point in the single link scenario. We also classify the launch power in the multiple links scenario since it is the lightpath configuration attribute and could be vulnerable to potential misconfiguration. The ML pipeline for distance regression and launch power classification is shown in Fig. 1. The IQ constellation profile is compressed with the principal component analysis (PCA) method. PCA projects the original data to low dimension directions that maximize the variance. Before PCA, the data is scaled by a custom standard scaler that operates with $\frac{(\mathbf{D} - \mu)}{\sigma}$ formula, where \mathbf{D} , μ , and σ are the IQ constellation profile, mean, and standard deviation of \mathbf{D} elements. A linear regression model estimates the distance for the single link scenario after another standard scaling

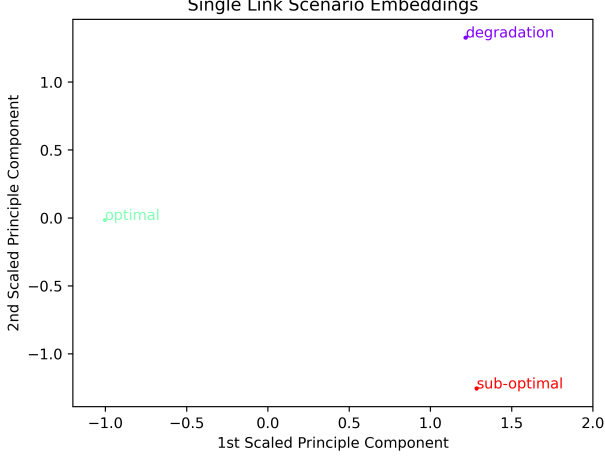


Fig. 2: Single link scenario embeddings

across columns of low dimensional embeddings resulted from PCA. The column standard scaler operates with $\frac{\mathbf{D} - \mathbf{M}}{\Sigma}$ formula, where \mathbf{D} , \mathbf{M} , and Σ are the IQ constellation profile, mean vector, and standard deviation vector of each column of \mathbf{D} . For launch power classification, the ROADM side and distance data are concatenated with the PCA embeddings after standard scaling across columns. Then, a support vector machine (SVM) classification model classifies the launch power as depicted in Fig. 1.

IV. RESULTS

The proposed pipelines in Fig. 1 are implemented with Python and available in [4]. PCA models compress IQ constellation profiles to the embeddings with size 50 for either pipeline. The datasets are split into train-test subsets with 67%-33% ratio. Modules in Fig. 1 are trained with the training data subset. The testing data subset is used for validating the pipelines' performance reported in the following.

A. Single Link Scenario

The single link scenario dataset is used for distance estimation. First, the whole dataset is compressed to embeddings of 2 dimensions with PCA. The embeddings are depicted in Fig. 2. It seems there is a clear distinction between the 3 modes in the 2D embeddings. Yet, the distance-related embeddings variation is concentrated in small areas in Fig. 3. The same compression process is applied to datapoints of *optimal* mode, which is shown in Fig. 3 along with the distance between the transmitter and the observation point. The locations of distance numbers on Fig. 3 are the centeriods of distance clusters. The distance is distinguishable with the 2D embeddings.

We train the lightpath distance regression pipeline in Fig. 1. The compression metrics of testing data subset after PCA is reported in Table. I.

The coefficient of determination (R^2)-score of distance estimation with linear regression is 0.99916 for the testing data subset.

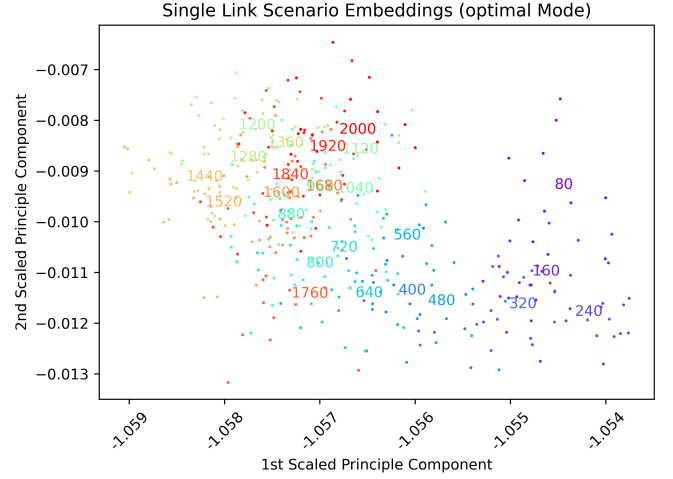


Fig. 3: Single link scenario embeddings for the *optimal* mode

TABLE I: Compression Metrics

Scenario	Compression Rate	MAE	MAPE
Single Link	98.78%	0.0475	3.5%
Multiple Links	99.69%	1.1959	67.22%

B. Multiple Links Scenario

The multiple links scenario dataset is used for launch power classification. We compress the IQ constellation profile with PCA and depict the 2D embeddings in Fig. 4. There is no clear distinction between the 2D embeddings of 1 and 2 dBm constellation profiles. However, increasing the size of embeddings might create clusters associated with launch power.

We train the launch power classification pipeline in Fig. 1. The confusion matrix of the testing data subset is reported in

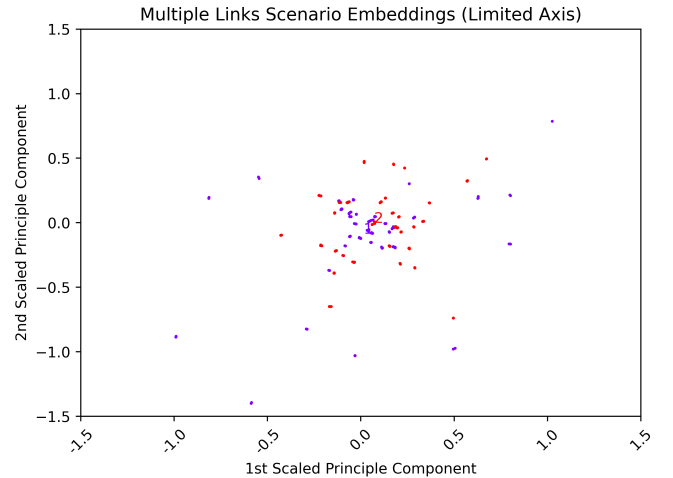


Fig. 4: Multiple links scenario embeddings. The 1 and 2 dBm launch power datapoint embeddings are shown with purple and red colors, respectively. Several outlier embeddings reside outside the plot scope for better visualization.

TABLE II: Launch Power Classification Confusion Matrix

	1 dBm (Prediction)	2 dBm (Prediction)
1 dBm (Target)	132	0
2 dBm (Target)	34	98

Table. II. The accuracy of classification is 0.87121 based on Table. II

V. CONCLUSION

The IQ constellation profile could be used to estimate lightpath attributes. The report investigates launch power and distance between transmitter and observation point estimation use cases with ML. The use cases do not require information about the dynamics of optical signal propagation in optical networks. Estimations are derived with the raw dataset.

The study could be extended in several directions, including: investigation of more efficient methods for compression (e.g., Gaussian mixture models), using more complex ML models for classification and regression, hyperparameter search for finding the best ML configurations that optimize the metrics, and error analysis of dataset cohorts.

REFERENCES

- [1] TIP. GNPY: Optical Route Planning and DWDM Network Optimization. *GitHub Repository*. (2023), <https://github.com/Telecominfraproject/oopt-gnpy>
- [2] Ruiz, M., Sequeira, D. & Velasco, L. Deep learning-based real-time analysis of lightpath optical constellations [Invited]. *Journal Of Optical Communications And Networking*. **14**, C70-C81 (2022)
- [3] Ruiz Ramírez, M., Velasco Esteban, L. & Sequeira, D. Optical Constellation Analysis (OCATA). (CORA.Repositori de Dades de Recerca,2022), <https://doi.org/10.34810/data146>
- [4] Hajipour, S. Lightpath Length and Launch Power Estimation with Machine Learning. *GitHub Repository*. (2023), <https://github.com/hprshayan/lightpath-attribute-estimation>