Efficient Classification of Polarization Events Based on Field Measurements

[1]Kyle Guan, [1]Jesse E. Simsarian, [2]Fabien Boitier, [3]Dan Kilper, [2]Jelena Pesic, and [4]Michael Sherman

[1]*Nokia Bell Labs, 791 Holmdel Road, Holmdel, NJ 07733, USA,*
[2]*Nokia Bell Labs, Route de Villejust, 91620, Nozay, France*
[3]*College of Optical Sciences, University of Arizona, Tucson, AZ 85721, USA*
[4]*Electrical and Computer Engineering, Rutgers University, Piscataway, NJ 08854, USA*
jesse.simsarian@nokia-bell-labs.com

Abstract: We present rare-event classification of polarization transients based on field measurements with data augmentation combined with robot-generated fiber-disturbance data. We compare machine learning methods for accuracy and required number of training sample traces. © 2020 The Author(s)

OCIS codes: 060.4510, 060.2370.

1. Introduction

Continuing on the rapid growth of cloud services, the next generation of cloud applications such as industrial control over ultra-reliable 5G networks will require strict bounds on wireless-wireline network performance metrics such as packet loss, particularly with fiber deployed in more disruption-prone metro environments. We investigate methods to monitor the underlying optical transport network to enable proactive action to avoid packet loss from fiber breaks that are preceded by state of polarization (SOP) transient events. Previous work examined low-cost embedded polarization sensors at optical amplifier sites [1] that localize fiber disturbances to a span. Alternatively, utilizing the SOP recovered from the digital signal processing logic of dual-polarization coherent receivers allows for fiber path monitoring without additional hardware. In recent work, SOP transients measured by a coherent receiver and created by robotic arm movements of a fiber were detected and analyzed by machine-learning (ML) naive Bayes event classification [2, 3]. While fiber breaks are a major source of network outages, they are also relatively rare and anomalous events, motivating the need to determine the most accurate methods of event classification training on realistic SOP data in an efficient manner, i.e., requiring the fewest number of event training samples. In this work we initially benchmark different ML methods on the SOP transient dataset from Ref. [2, 3] to compare the relative performance of the methods. To get more realistic training data, we add additional events to the dataset based on the SOP monitoring of a field-deployed fiber network in Manhattan, COSMOS [4]. Due to the rarity of the events observed in the field, we augment the dataset with polarization-rotated versions. We expect similar data augmentation techniques to be broadly applicable to efforts in the optical communications industry to find sources of data for ML models gathered from operational networks [5]. Finally, we report the accuracy and efficiency of the ML classification methods on the enlarged dataset to compare their performance.

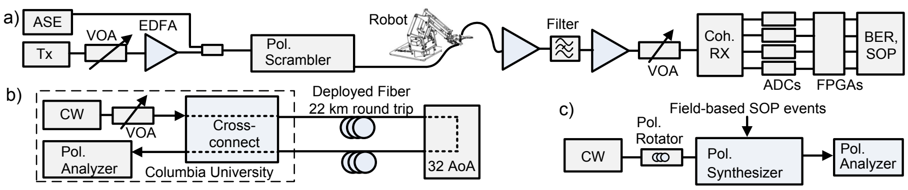

Fig. 1. a) Detection of robot arm movements of a fiber with coherent a receiver b) SOP field measurements on the COSMOS network (c) Data augmentation of field-based measurements.

2. Data Collection and Processing

In this work, we use three different experimental setups (Fig. 1) to collect two datasets of SOP traces for ML event classification. The first "benchmark dataset", used to benchmark ML model performances, contains the SOP traces resulting from mechanical manipulations using a robot arm [2], as shown in the experiment diagram of Fig.1a. The robot arm simulates four types of events, each generating multivariate time series of Stokes parameters S_1, S_2, and S_3. The benchmark dataset has a total of 10,000 SOP traces of four classes/labels (an average of 2,500 traces per class), with all traces polarization rotated to achieve polarization-independent event classification.

To evaluate the performance of ML methods in classifying realistic SOP traces and more event types (labels), we construct a second "enhanced dataset", which includes a new collection of SOP traces for four types of events generated by a robot arm, labeled as "Mvt1/Label1" to "Mvt4/Label4" in Fig. 2a-d. Significantly, we include in this dataset SOP traces based on data collected from monitoring a field-deployed fiber link of the COSMOS network

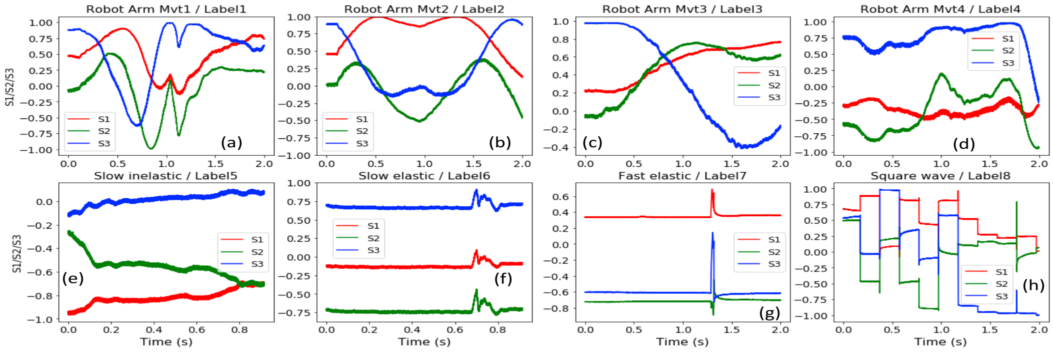

Fig. 2. Eight types of SOP events in the enhanced dataset. (a)-(d) are traces generated by robot arm movements, (e)-(f) are traces related to the COSMOS field link, and (g)-(h) are additional traces

[4], which is 22 km round trip from Columbia University to 32 Avenue of the Americas (AoA) in Manhattan, as shown in the experiment setup in Fig. 1(b). Over 9 days of monitoring, we observed periods of polarization activity on the link, as shown by the plot of total polarization angle travel during mostly 0.9 s data captures (Fig. 3a). We observed "slow inelastic" (Label5) and "slow elastic" (Label6) events, similar to those reported in [6] as shown in Fig. 2 (e) and (f), respectively. We collected 175 slow inelastic traces and added all of them into the dataset. We also added "fast elastic" events (Label7, Fig. 2g), previously observed [6, 7] and square-wave like scrambling SOP traces (Label8, Fig. 2h) to stress test the ML methods. One goal of the field experiments is to find polarization signatures that precede a fiber break, which we did not observe, motivating future monitoring work [5]. The relative rarity of such events requires data augmentation of the training traces, which we accomplish by programming similar fast and slow elastic traces into a polarization synthesizer and using a polarization rotator (Fig. 1c) to generate traces that cover the Poincare sphere. Additionally, when synthesizing slow and fast elastic traces, we ensure that the transients occur uniformly along the time axis for time-translation invariance. The enhanced dataset, by design, is more difficult to train/classify. It has a total of 2,000 SOP traces for eight labels (shown in Fig. 2, an average of 250 traces per label), twice the number of events to classify compared to the baseline dataset, but with ∼10x fewer traces per classification label - recall that we want accurate classification of more labelled event types with fewer required training samples.

We use two simple pre-processing steps on the traces (in Stokes space) before the ML classification. First, to reduce the complexity and possible overfitting, we down-sample the SOP traces to 100 points per dimension. We observe that SOP traces of the same event, though different in starting/ending position on the Poincare sphere, share similar trajectory structures that could be exploited by ML methods (Fig. 3b) [8]. As such, we create a new dataset by rotating every SOP trace to the same staring point on the Poincare sphere (Fig. 3c) [8].

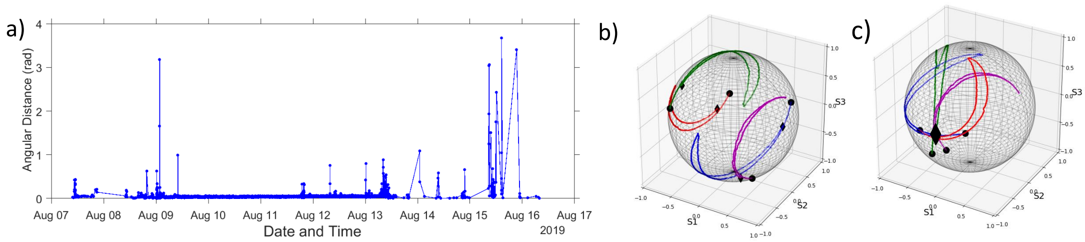

Fig. 3. (a) SOP angular travel during traces captured over 9 days of field monitoring (b) SOP traces of the same type of event (c) SOP traces of the same type of event with back rotations to the same staring point. Diamond and circle markers indicate starting and ending positions, respectively.

3. Machine Learning Models

We consider a total of five ML models for classifying SOP events: linear support vector machine (LSVM), logistic regression (LR) classifier, kernel SVM, neural network (NN), and a long-short-term-memory (LSTM) network. Linear SVM and logistic regression only work well with linearly separable data, and are therefore expected to have worse performance. In comparison, kernel SVM maps the inputs into a higher dimension space, thus allowing more complex decision boundaries [9]. In this work, we use a Gaussian radial basis function as the kernel for SVM and optimize over parameters C (the penalty for classification error) and γ (defines the influence of a single training trace). Similarly, the nonlinear activation functions used at each NN layer greatly improves the classification capabilities compared to the linear models. For the NN, we use two (hidden) layers (each with a size of 512 neurons) with dropouts at each layer to reduce overfitting. We optimize over the parameters such as dropout rate, batch size, and training epochs. Since LSTM is a deep learning model that uses the long and short-term temporal correlations within a time series to extract features for classification, we tune the parameters such as dropout rate, batch size, and training epochs, similar to the NN. All the training and testing are implemented via ML libraries such as sckit-learn [9] or TensorFlow with Python.

4. Results

We evaluate the classification accuracy, defined as the percentage of correct classification to the 8 classes of Fig. 2, with and without back rotation. For each dataset, we set aside 20% of the samples as the test set. Of the 80% of the samples used for the training set, we vary the number of training samples to evaluate the impact of training set size on the accuracy. For a chosen sample size, we ran 10 tests, each with different random seeds used for selecting training and testing samples. We report the average accuracy of these 10 tests.

When we use all the 8,000 training samples of the benchmark dataset [2] (without back rotation), the accuracy of the results using kernel SVM, NN, and LSTM are all above 99% (Fig. 4a), similar to those achieved by the quaternion sequence and naive Bayes approach [2]. As expected, LSVM and LR perform poorly in comparison, due to their deficiency in classifying non-linearly separable data. When the samples are projected on the higher dimension (as in kernel SVM) or processed by non-linear activation functions in NN or LSTM, they become linearly separable. Note that both kernel SVM and LSTM are efficient and robust to training set size, showing a drop of 1% in accuracy with one-order magnitude reduction of training set size and a drop of only 5% in accuracy with two orders of magnitude reduction of training set size, as shown in Fig. 4. This is critical for rare-event scenarios and motivates our usage of smaller dataset size for the enhanced dataset. The plot also shows the benefit of back rotating SOP traces – it not only improves the performance of both LSVM and LR, it also improves the accuracy for kernel SVM, NN, and LTSM for small training sets. We also compute the precision and recall for each test. The results are consistent with accuracy performance and omitted here due to space constraints.

The results of the enhanced dataset are shown in Fig. 4b). When we use all 1600 training samples (no back rotations), the accuracy of the results of kernel SVM, NN, and LSTM are all above 96%, despite the larger number of event labels to classify and fewer training samples. As baseline models, LSVM and LR perform poorly, due to the problem of the non-linearly separable dataset being exacerbated by a larger number of labels (eight vs. four in the benchmark dataset). Similar to the benchmark dataset, kernel SVM is efficient in training set size – a drop of only 4% with a 75% reduction of the training set. Similar to the benchmark dataset, SOP back-rotations improve the performance of both linear SVM and logistic regression. However, in contrast to the benchmark, the back rotation degrades the accuracy for kernel SVM, NN, and LTSM. This is likely attributed to the fact that the back rotations can make certain temporal features less distinct among some types of SOP events.

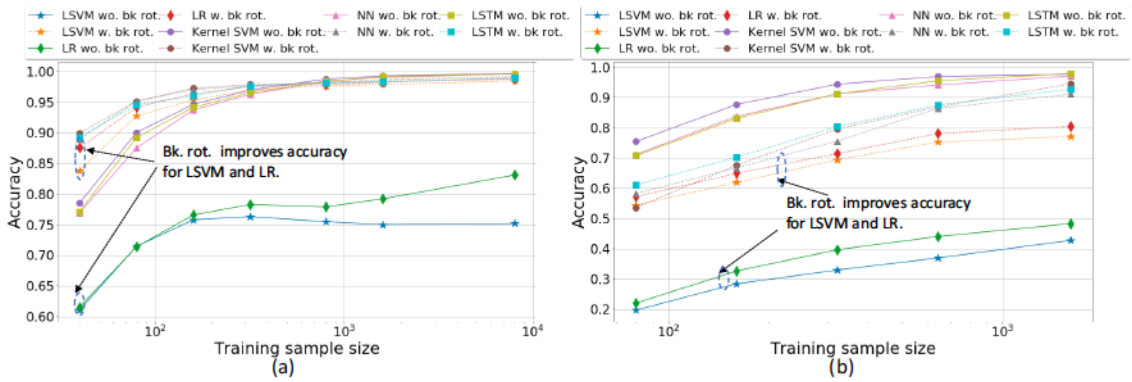

Fig. 4. a) Classification accuracy for the benchmark dataset, and b) the enhanced dataset.

5. Conclusion

In this work, we evaluated ML methods for SOP event classification in accuracy and sample size efficiency including field measurements. We show that kernel SVM and LSTM have the best accuracy and are robust to reduced training sets. We also show that data augmentation and simple pre-processing can improve rare event classification.

References

1. J. E. Simsarian and P. J. Winzer, "Shake before break: Per-span fiber sensing with in-line polarization monitoring," in Proc. OFC 2017, paper M2E.6.
2. V. Lemaire, *et al.*, "Proactive fiber break detection based on quaternion time series and automatic variable selection from relational data," https://project.inria.fr/aaltd19/files/2019/08/AALTD_19_Lemaire.pdf.
3. F. Boitier *et al.*, "Proactive fiber damage detection in real-time coherent receiver," in Proc. ECOC 2017, paper Th.2.F.
4. J. Yu, *et al.*, "COSMOS: Optical architecture and prototyping," in Proc. OFC 2019, paper M3G.3.
5. www.nist.gov/news-events/events/2019/08/machine-learning-optical-communication-systems
6. M. Boroditsky, *et al.*, "Polarization dynamics in installed fiberoptic systems," in Proc. LEOS 2005, 413–414.
7. D. Charlton *et al.*, "Field measurements of SOP transients in OPGW, with time and location correlation to lightning strikes," Optics Express, **25**, 9689-9696 (2017).
8. J. Pesic, *et al.*, "Proactive restoration of optical links based on the classification of events," in Proc. ONDM 2011, 1–6.
9. F. Pedregosa *et al.*,"Scikit-learn: Machine learning in Python," Journal of Machine Learning Research, **12**, 2825–2830 (2011).