Machine Learning Techniques for Optical Performance Monitoring from Directly Detected PDM-QAM Signals

Thrane, Jakob; Wass, Jesper; Piels, Molly; Medeiros Diniz, Júlio César; Jones, Rasmus Thomas; Zibar, Darko

Published in:
Journal of Lightwave Technology

Link to article, DOI:
10.1109/JLT.2016.2590989

Publication date:
2017

Document Version
Peer reviewed version

Machine Learning Techniques for Optical Performance Monitoring from Directly Detected PDM-QAM Signals

Jakob Thrane, Jesper Wass, Molly Piels *Member IEEE*, Julio C. M. Diniz, Rasmus Jones and Darko Zibar *Member IEEE*

***Abstract*—Linear signal processing algorithms are effective in dealing with linear transmission channel and linear signal detection, while the nonlinear signal processing algorithms, from the machine learning community, are effective in dealing with nonlinear transmission channel and nonlinear signal detection. In this paper, a brief overview of the various machine learning methods and their application in optical communication is presented and discussed. Moreover, supervised machine learning methods, such as neural networks and support vector machine, are experimentally demonstrated for in-band optical signal to noise ratio (OSNR) estimation and modulation format classification, respectively. The proposed methods accurately evaluate optical signals employing up to 64 quadrature amplitude modulation (QAM), at 32 Gbaud, using only directly-detected data.**

***Index Terms*—optical communication, machine learning, performance monitoring, neural networks, support vector machines**

I. Introduction

The field of machine learning offers many powerful techniques to: estimate parameters from noisy measurement data, determine complex mapping between input and output data, infer probability distributions, predict the output based on the past input data and perform classification [1], [2]. Choosing the right machine learning algorithm strongly depends on the problem that needs to be solved.

The challenges associated with the optical communication is that, the optical fibre channel is nonlinear, due to the Kerr nonlinearity, and also the optical signal detection may be nonlinear as in the case of direct detection. However, many of the tasks addressed by the machine learning community are of the nonlinear nature and we therefore believe that machine learning techniques may prove useful to combat optical fibre channel nonlinearities and also extract useful information about the optical signal after direct detection. An overview of the machine learning methods applied to optical communication, together with the references is shown in Fig. 1. In the following, a brief description of the methods and their application to optical communication is discussed.

Recently, several methods within the framework of nonlinear state-space based Bayesian filtering, (extended Kalman and particle filter), have been employed for time-varying parameter estimation such as: amplitude and phase noise, cross-polarization and cross-phase modulation induced polarization scattering and polarization mode dispersion [3]–[8]. The advantages of the state-space based Bayesian filtering for time-varying parameter estimation are that: 1) the framework is very well suited for joint parameter estimation, 2) it allows for the inclusion of the underlying physics of optical components and optical fibre channel into the estimation algorithms and 3) it allows for more complicated models of the time-varying parameters.

Machine learning methods have also been employed for optical performance monitoring for optical communication systems employing advanced modulation formats. It has been shown that various algorithms such as: neural networks, k-means, variational Bayesian methods for mixture models and statistical methods using cumulants, can be used to perform blind modulation format classification [9]–[12]. Neural networks have also been employed for optical channel parameters estimation such as: chromatic dispersion, differential group delay, baud rate and optical signal to noise ratio [13]–[15]. An interesting application of well-known and widely used machine learning algorithms such as independent and principal component analysis have also been demonstrated for signal demodulation, modulation format and bit rate identification [12], [16], [17].

One of the main applications of the machine learning techniques is to perform optimum classification. The

J. Thrane, J. Wass, M. Piels, J. C. M. Diniz, R. Jones and D. Zibar are with DTU Fotonik, Department of Photonics Engineering, Technical University of Denmark, Kgs. Lyngby, Denmark, (email: dazi@fotonik.dtu.dk)

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JLT.2016.2590989, Journal of Lightwave Technology

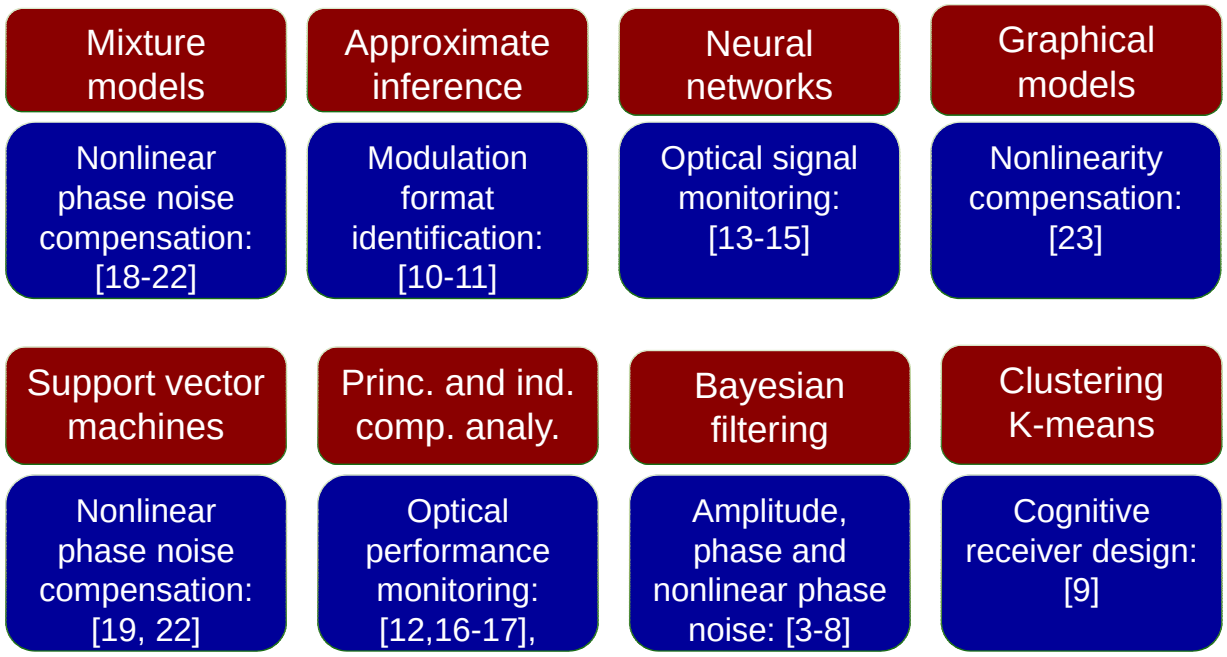

Fig. 1. An overview of the various machine learning methods applied to optical communication

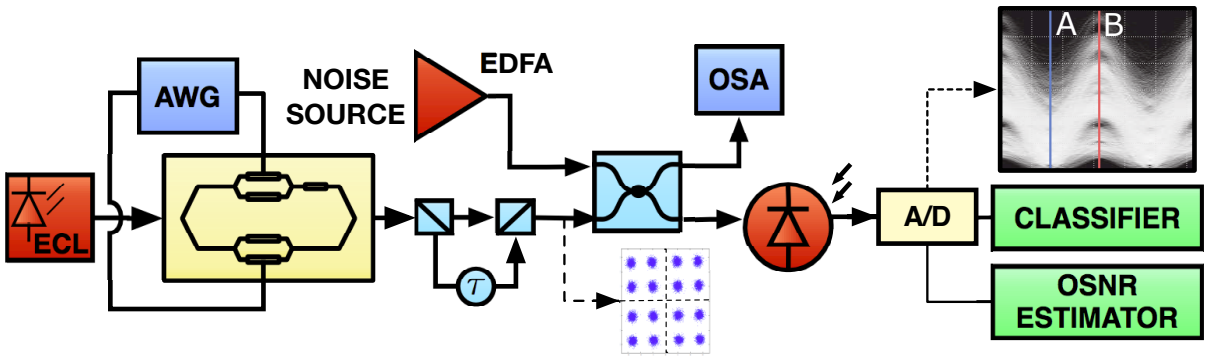

Fig. 2. Diagram showing the experimental back-to-back setup. The power eyediagrams produced from photodetection are sampled to extract minimum and maximum variance features (A and B, respectively) for signal evaluation.

optimum classification directly translates into optimum symbol detection for the optical communication systems. For the case of memoryless nonlinearity such as: nonlinear phase noise, I/Q modulator and driving electronics nonlinearity, Euclidean distance metric, resulting in linear decision boundaries, is no longer optimum. For those particular cases, optimum symbol detection and thereby decision boundaries, can be obtained by using machine learning techniques such as: support vector machines, kernel density estimator and Gaussian mixture models [18]–[22].

Finally, for the nonlinearity mitigation, factor graphs and message passing algorithms have been applied to include the effects of noise in the digital backpropagation (DBP) and thereby perform optimum symbol detection. This has resulted in significant gains compared to the deterministic DBP [23].

In this paper, the application of machine learning techniques for optical performance monitoring are investigated. Optical performance monitoring is vital to ensure robust and reliable networks [15], [24], [25]. The modulation format and the optical signal to noise

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JLT.2016.2590989, Journal of Lightwave Technology

ratio (OSNR) are key parameters for assessing the performance of optical transmission links. In-band OSNR estimation is especially important in next generation wavelength division multiplexed (WDM) systems where the channel spacing approaches signal baud rate. To this end, systems based on back-to-back transmission characteristics and pilot sequences have been demonstrated [24]–[26]. However, such approaches require complete demodulation and hence are too complex and costly for mid-span monitoring points. One promising approach in this area is to use intensity information coupled with advanced signal processing to extract the relevant information using only one photodiode and ADC [15], but this has yet to be demonstrated for polarization division multiplexed signals or higher-order QAM formats.

To address those challenges classification scheme based on Support Vector Machines (SVM) is investigated for modulation format recognition and nonlinear regression technique, employing neural network, is investigated for in-band optical-signal-to-noise-ratio monitoring. Both of the techniques are applied to directly detected polarization multiplexed quadrature amplitude modulated optical signals. In short, this means that modulation format can be classified and in-band OSNR can be estimated by employing a single photodiode followed by a analogue-to-digital converter. It is envisioned that the proposed techniques can be very useful for mid-span monitoring systems.

The remainder of this paper is organized as follows. In section II, the main principles on how to perform modulation format classification and in-band OSNR monitoring from directly detected polarization multiplexed QAM modulated optical signals are outlined. Moreover, main operating principles behind neural networks and support vector machines are presented. The theory on how to optimize neural networks is briefly discussed as well. In section III, the experimental set-up is presented and described. In section IV, the effectiveness of support vector machines and neural networks for modulation format classification and in-band OSNR estimation is investigated using the experimental results. The conclusions of the investigations are summarized in Section V.

II. Performance monitoring using intensity information

The reason why the focus is on the direction detection and thereby intensity information is because if OSNR monitoring is going to be widely employed along the link then the monitoring unit needs to be relatively simple and cost effective. We would like to avoid solutions that rely on coherent detection and full signal demodulation as this is quite complex and costly. One of the simplest approaches to perform OSNR monitoring is therefore to employ a single photodiode and then estimate the OSNR directly after the photodetection without prior signal demodulation.

In this paper, supervised learning, based on support vector machines and neural networks, is considered. This means that the machine learning algorithms need to infer a mapping function from the labeled training data. The training data consists of training examples pairs, each containing input features and targets (modulation format and OSNR). The crucial part is to extract the relevant features, from the directly detected optical signal, that are representative for the target data. For instance, we need to extract the features that vary as the modulation format and OSNR are varied.

As already states, the modulation format classifier and OSNR estimator operate on directly-detected (DD) data. This is illustrated along with the complete experimental setup, in Fig. 2. From the power eyediagram after the photodetector, we extract eight features. The OSNR estimator only considers one feature but the modulation format classification considers all of them. The mean values and variances of the eye diagram at points A and B are four of the eight extracted features used by the machine learning blocks. Obtaining the maximum and minimum variance has been performed using a simple method based on a sliding window principle: The signal was upsampled to 10 samples/symbol and split up into 50 frames consisting of 80000 symbols, where the minimum and maximum variance of each frame, along with the mean was computed. Averaging over the results from all the frames obtains an approximation of the variance at the maximum and minimum eye opening points. The four additional features, only used by the modulation format classifier, are another mean value, the difference and the ratio of the variances of the two distinct eyediagram positions as well as a last mean value of the two positions' mean values. The various features are chosen to vary with and, therefore, be dependent on different modulation formats or OSNRs. Both methods follow the same general structure: 1) sample the eyediagram of different modulation formats or OSNRs and extract dependent features, 2) train the classifier or the neural network on various instances of the extracted features for the given modulation format or OSNR, and 3) use the trained classifier or neural network to classify the modulation format or predict the OSNR of new observed eyediagrams/features.

The task of the classifier is to learn a mapping function from the extracted features to the used modulation for-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JLT.2016.2590989, Journal of Lightwave Technology

mat. During training, we pass several 8-D input vectors of features, paired up with their corresponding target class, here the modulation format, to a linear support vector machine classifier. The goal is not to reproduce the mapping of the just mentioned training cases, but let the SVM generalize the mapping between input vector of features and modulation format. Hence, the goal is to correctly classify the modulation format for new input vectors of yet unseen eyediagrams. For this reason, validation is performed on traces that are not used during the training phase.

The OSNR estimator, unlike the classifier, is designed to learn a continuous mapping function between the input feature and the OSNR. It also differs in that only one feature, the minimum variance (A), is used. To accomplish this, several instances of this feature with its corresponding OSNR are used to train the neural network. Comparable to the classification, which predicts the modulation format, the trained neural network attempts to predict the correct OSNR for new inputs extracted from so far unseen eyediagrams.

A. Neural networks

The structure of the employed neural network to perform the nonlinear regression is shown in Fig. 3. The type of network presented in Fig. 3 is a feed-forward neural network, also known as the multilayer perceptron as described by [1]. The neural network can be seen as a series of functional transformations of the input, some of which are performed by non-linear activation functions $h(\cdot)$. For the considered case, the non-linear activation function is $\tanh(\cdot)$. The feed-forward, one hidden layer, neural network shown in Fig. 3 can be expressed as:

$$y(x, \mathbf{w}) = \sum_{j=1}^{3} w_{1j}^{(2)} h(w_{j1}^{(1)} x) \tag{1}$$

where x is the input variable, i.e. variance extracted from the eyediagram and $\mathbf{w} = [w_{11}^{(1)}, w_{21}^{(1)}, w_{31}^{(1)}, w_{11}^{(2)}, w_{12}^{(2)}, w_{13}^{(2)}]$ is the weight vector. The superscript of each weight denotes the layer of the network. The weights are adaptive and to solve the regression problem, they have to be adjusted to fit the data during a training phase. Since this function depends on adaptive parameters, the transformation will also be adaptive even though the function itself is fixed.

To find the optimal weight vector $\mathbf{w}$, the neural network has to be trained in a supervised manner. For the training of the neural network, a dataset S of N observations is used: $S = \{(\sigma_n^2, OSNR_n)|n = 1, ..., N\}$. For the considered case, the training points consist of a variance value (input) and a corresponding known OSNR values (target). Since the neural network is used for regression, we will assume that the target values $t_n = OSNR_n$ have a Gaussian distribution with an input dependent mean given by the output of the neural network. This means that the output of the neural network can be expressed as:

$$t_n = y(\sigma_n^2, \mathbf{w}) + \epsilon \tag{2}$$

where ϵ is a zero mean Gaussian random variable with variance β^{-1}. By assuming that the target values t_n have Gaussian distribution, an analytical expression for the likelihood function of the target variables is available. The availability of the likelihood function eases the implementation of the learning algorithm to optimize the network parameters $\mathbf{w}$. However, the distribution after the photodetection is best described by non-central chi distribution and in order to approach the Gaussian distribution traces are superimposed. According to the central limit theorem, the distribution after direct detection approaches the Gaussian distribution as the number of superimposed traces tend to infinity. The results presented in the section IV which show the ability to accurately estimate the OSNR indicate that we are well within the limits of the approximation.

Considering a data set of inputs $\mathbf{X} = [\sigma_1^2, ..., \sigma_N^2]$ and targets $\mathbf{t} = [t_1, ..., t_n]$, the likelihood function of the neural network parameters, $\mathbf{w}$, is expressed as:

$$\begin{aligned} p(\mathbf{t}|\mathbf{X}, \mathbf{w}, \beta) &= \prod_{n=1}^{N} \mathcal{N}(t_n|y(\sigma_n^2, \mathbf{w}), \beta^{-1}) \\ &= \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi}\beta^{-1}} e^{-\frac{\beta}{2}\left(y(\sigma_n^2, \mathbf{w}) - t_n\right)^2} \end{aligned} \tag{3}$$

To optimize the network parameters, $\mathbf{w}$, we need to maximize the likelihood function consisting of N independent, identically distributed observations. From equation (3), the negative log likelihood function takes the form:

$$\begin{aligned} -\ln p(\mathbf{t}|\mathbf{X}, \mathbf{w}, \beta) &= \frac{\beta}{2} \sum_{n=1}^{N} \{y(\sigma_n^2, \mathbf{w}) - t_n\}^2 \\ &- \frac{N}{2} \ln \beta + \frac{N}{2} \ln(2\pi) \end{aligned} \tag{4}$$

Maximizing the negative likelihood function is equivalent to minimizing the sum-of-squares error function given by:

$$E(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^{N} \{y(\sigma_n^2, \mathbf{w}) - t_n\}^2 \tag{5}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JLT.2016.2590989, Journal of Lightwave Technology

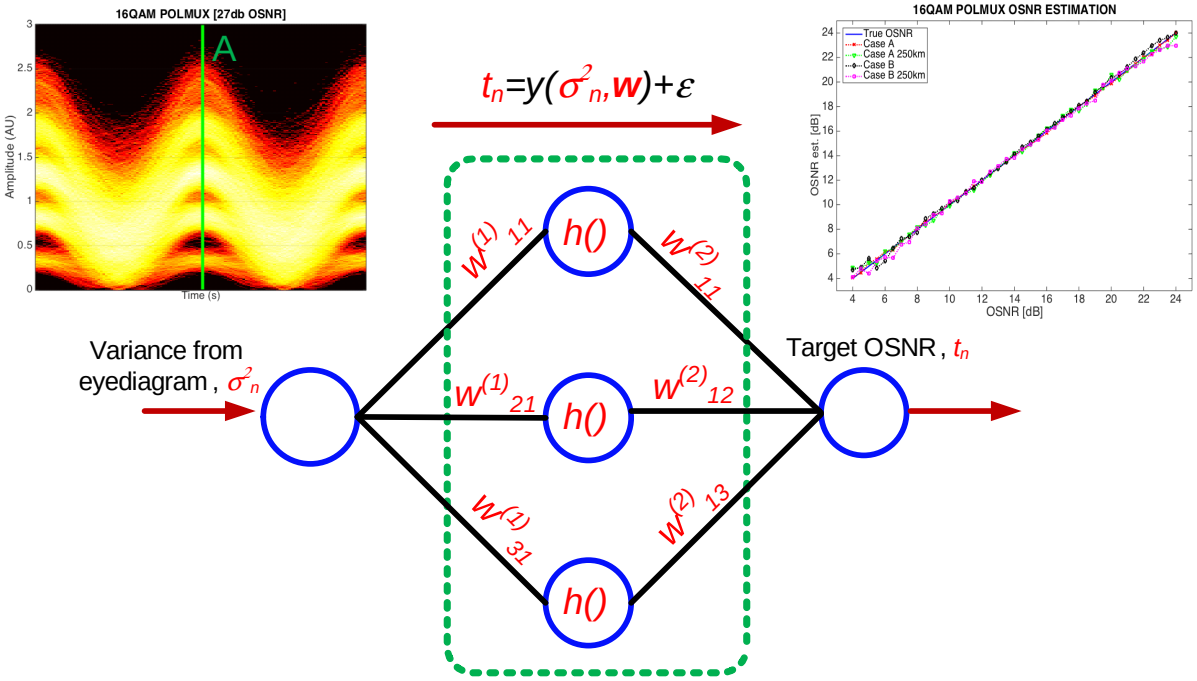

Fig. 3. Illustration of the application of the neural network for OSNR estimation. For the upper right figure, case A and B represent different sampling points in the eyediagram used to extract the variance for the back-to-back case and after 250 km of uncompensated transmission. y : non-linear mapping function that needs to be inferred from the training data.

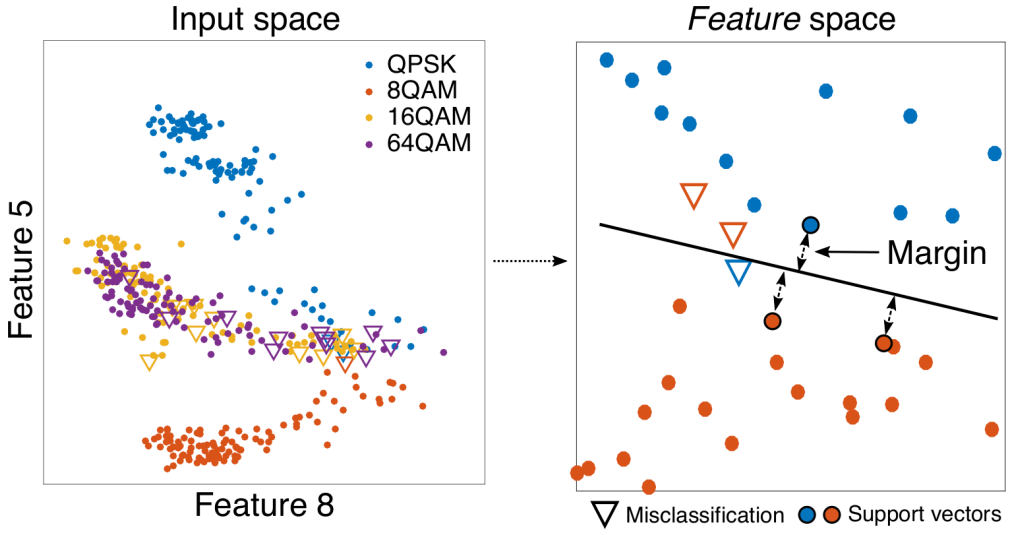

Fig. 4. Classification results shown in data input space consisting of feature 5 and 8, corresponding to the mean and the absolute difference in variance respectively. Shown by triangles are classification errors preformed by a trained linear SVM. Feature space is conceptually illustrated with a probable decision boundary between two arbitary classes.

Since it is not possible to find the optimal solution to the error function with an analytical expression, the solution is found iteratively using the *gradient descent* technique which is built on the idea of *error backpropagation* as described by [1]. When the training phase has been completed, the neural network has been optimized to the training data, and can now estimate OSNR from variance measures. It should be noted that the number of

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JLT.2016.2590989, Journal of Lightwave Technology

training points is vital to the accuracy of the regression solution, and to avoid severe over-fitting. The number of training points should be taken into consideration when determining the complexity of the neural network.

B. Support vector machine

Support Vector Machine (SVM) became popular solving problems in classification and regression due to it's kernel-based algorithms that have a sparse solution. Kernel methods use kernel functions to express similarities over all pairs of raw input data. This results in a large training matrix consisting of metrics of similarities between all input data points. Usually, kernel-based methods store the entire training matrix for predictions and classification of new inputs, SVM however only store a subset and these are known as support vectors. For this particular property SVM is considered computational efficient compared to many other algorithms that take use of kernel functions, as making predictions based on a large training matrix can be computationally infeasible. In general, the SVM is based on a geometric mindset which aims to maximize the margin. The margin is a measure of how well classes can be separated by a given decision boundary. The optimization problem of SVM consists of finding a decision boundary where the margin is maximized i.e. where the classes are most separated.

Being a kernel-based method, SVM seeks to take advantage of the so called "kernel trick" which enables operation in high-dimensional feature space without the computational cost of computing the actual coordinates. Typically, the kernel function is given as an inner product in a feature space, $k(\mathbf{x}, \mathbf{x}') = \phi(\mathbf{x})^T \phi(\mathbf{x}')$ where nonlinear feature space mapping is $\phi(\mathbf{x})$ and $\mathbf{x}$ is an input vector and $\phi(\cdot)$ is a mapping function. The Kernel trick essentially is to define the kernel $k(\mathbf{x}, \mathbf{x}')$ in terms of original input vector $\mathbf{x}$ without even defining or even knowing, the transformation function $\phi(\cdot)$. The task is then to find the kernel function $k(\mathbf{x}, \mathbf{x}')$. For that purpose, in this paper, K-fold was used as cross-validation, more specifically we used a 5-fold split on a data set of the size 53×4. This results in 32 training points per fold. So in short, the results are obtained training 5 different kernels (5-fold cross-validation) and picking the best preforming one which for the considered case is a linear kernel. The illustration of SVM for modulation format classification is shown in Fig. 4.

The main idea behind Fig. 4 is to illustrate the feature space for modulation format classification. The reason why we use features 5 and 8 is because the classes (modulation formats) are most distinguishable for those two features in the two dimensional space. The reviewer is correct that by only employing those two features (5 and 8) it is not possible to distinguish between 16QAM and 64QAM. Increasing the features space to three dimensions does not help either and that was the reason why we did not plot higher dimensional feature space. It is only by employing all 8 features that the modulation formats become distinguishable.

III. Experimental set-up

To verify the proposed method, an optical channel is set up as illustrated in Fig. 2. An external cavity laser (100 kHz linewidth) is used as a transmitter. The laser output is modulated by an electrical signal using an I/Q optical modulator. The 3-dB bandwidth of the I/Q modulator is 22 GHz. The electrical driving signal for the I/Q modulator is generated by an electrical arbitrary waveform generator with 64 GS/s and 20 GHz bandwidth. The symbol rate is 32 Gbd with 2 samples per symbol. Polarization multiplexing of the I/Q modulated signal is emulated by introducing a delay. Noise is added to the generated modulated PDM signal using an EDFA and an optical coupler. One output of the coupler is sent to a photodiode, sampled at 80 GS/s and connected to machine learning block for modulation format classification and OSNR estimation. The other output of the coupler is connected to an optical spectrum analyzer to obtain an OSNR reference value used for training during the supervised learning. The OSNR reference values are obtained for both NRZ and RC pulseshapes using rolloff factors at 0.01, 0.1 and 0.2. Overall, we obtained 3 measurements at every OSNR value for each of the 16 combinations of modulation format and pulseshape. The OSNR values range from 4-30 dB with steps of 0.5 dB.

The parameters of the neural network were found using the MATLAB toolbox which optimizes the weights given in eq. 5. The training data is provided to the toolbox and we only need to select the optimization routine. The MATLAB toolbox then finds the NN parameters. The methods we selected for finding the parameters is gradient descent based on error backpropagation in combination with Bayesian Regularization to reduce overfitting. This was additionally cross-validated by the toolbox which takes use of the entire available data set (K-fold) as to further reduce the chance of overfitting. The number of hidden nodes in the single hidden-layer NN construction has been determined by model selection

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JLT.2016.2590989, Journal of Lightwave Technology

methods using the above explained optimization process to find the appropriate model complexity. The support vectors and the parameters required for the linear SVM classifier were obtained by employing the similar approach. The training data is provided to the toolbox and the model selection of different kernels is performed using cross-validation to reduce overfitting. The toolbox determines automatically the kernel parameters using the provided training data. The best performing kernel is then selected.

IV. Results

The modulation format classification is trained for different OSNRs, thus it does not require knowledge of the OSNR. For each considered pulseshape we trained one classifier with four target classes: QPSK, 8QAM, 16QAM and 64QAM. Training is done using the MATLAB R2015b classification toolbox. The training set consists of two feature measurements for each OSNR value. The classification model is evaluated using a third independent feature measurement. Fig. 5 displays the classification results of a trained linear SVM classifier using 8-D input feature vector. An average classification accuracy of 94 is obtained. Similar classification accuracy has been achieved for less aggressive rolloff factors for RC pulseshaping as well as NRZ. As seen in Fig. 5, no uncertainty is present when classifying signals employing 8QAM and almost no uncertainty when classifying QPSK regardless of the present OSNR. However, signals employing 16QAM or 64QAM involve a small risk (approximately 8-13 on average) of misclassification towards each other. Investigating this misclassification rate further, we observed that the misclassification rate is reduced to less than 3 when operating above 11 dB OSNR. However, below 11 dB OSNR, it increases to around 35. Hence, the classification is more challenging below 11 dB OSNR as features exhibit a high degree of similarity.

The OSNR estimation requires knowledge of the pulseshape and the modulation format, thus one neural network is trained per combination of pulseshape/modulation format. Training is done using MATLAB R2015b neural networks toolbox. The training set consists of two feature measurements for each OSNR value. In Fig. 6, the feature measurements obtained from DD eyediagrams are plotted as a function of the known OSNR. The measurements are plotted for optical PDM signals employing various modulation formats using 0.01 RC pulseshaping. The feature and target value (OSNR) have a clear nonlinear relationship and depict the mapping function, which the corresponding neural

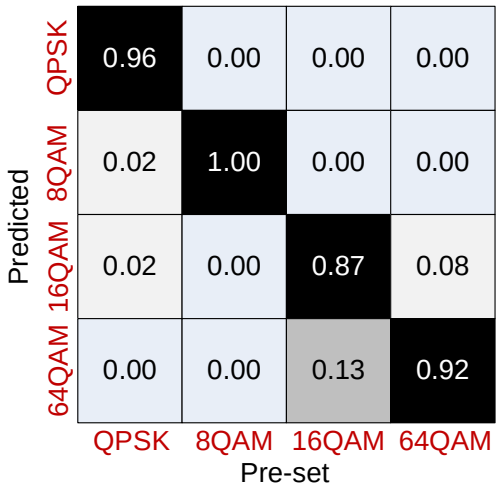

Fig. 5. Mean misclassification for 4-30 dB OSNR using a linear SVM classifier with 8 features and RC 0.01 pulseshaping.

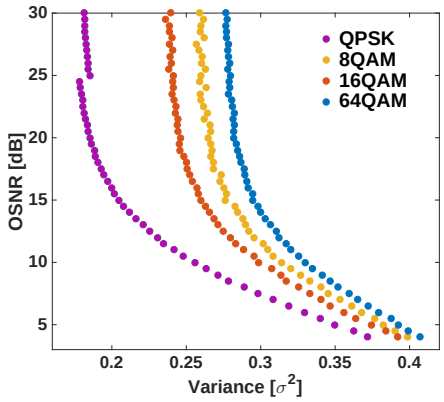

Fig. 6. Minimum variance measurements for QPSK, 8 QAM, 16 QAM and 64 QAM with RC 0.01 pulseshaping.

network learns by adjusting its weights. This makes the OSNR prediction possible. [1] Fig. 7 displays the OSNR estimation as function of the OSNR reference values for PDM-64QAM signals using RC pulseshaping with rolloff 0.01.

Fig. 8 illustrates the corresponding squared training and test (estimation) error as function of target OSNR to assess potential model overfitting. As seen, estimation is most accurate from 4-17 dB OSNR with a mean estimation error of 0.2 dB and worst-case error of 1.0 dB. When the signal quality exceeds this interval, estimation is less accurate with a mean estimation error of 1.2 dB

[1]The fact that very simple neural network consisting of one hidden layer and three hidden neuron, is employed demonstrates that the regression is relatively simple. Indeed, also polynomial regression could also have been used.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JLT.2016.2590989, Journal of Lightwave Technology

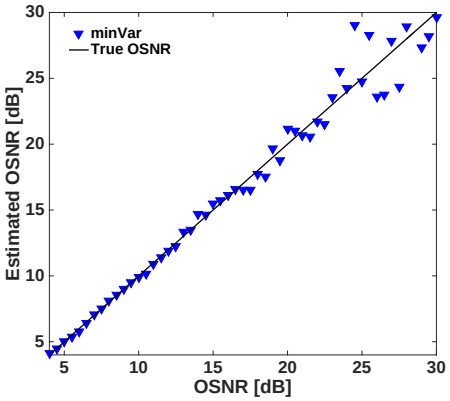

Fig. 7. OSNR estimation using minimum variance feature with trained NN for DP-64 QAM and RC 0.01 pulseshaping.

and 3.5 dB worst-case error. As seen in Fig. 6, the feature does not vary much from 17-30 dB, which increases the uncertainty on the final OSNR estimation. A total mean OSNR estimation error of 0.7 dB is achieved with a worst-case estimation error of 3.5 dB. Results with similar estimation accuracy have been obtained for less aggressive rolloffs for RC pulseshaping as well as NRZ.

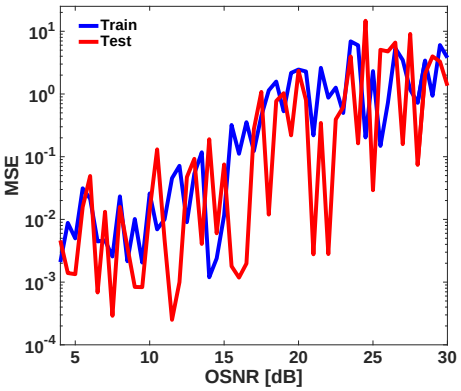

Fig. 8. Squared error of OSNR estimation for training and test data using DP-64 QAM and RC 0.01 pulseshaping.

The presented results, so far, have only been verified for the back-to-back case and thereby only in the presence of the white Gaussian noise. For a real transmission scenario, linear optical fibre impairments (chromatic dispersion, polarization mode dispersion, PMD, and polarization dependent loss, PDL) and nonlinear impairments (intra- and interchannel nonlinear distortions) must also be considered.

To investigate the impact of the chromatic dispersion, we have run a numerical simulation and tested the

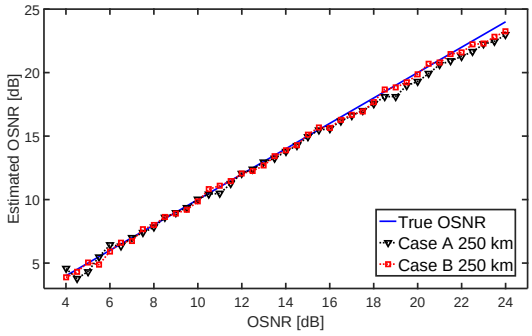

Fig. 9. OSNR estimation after 250 km of dispersion uncompensated transmission.

method for up to 250 km of transmission through a dispersion uncompensated link. The modulation format is still dual polarization 64 QAM at 32 Gbaud and the results are shown in Fig. 9. For the considered transmission distance, the method still works and is not impacted by the chromatic dispersion. Going beyond 250 km of uncompensated transmission distance is more challenging as the extracted features from the eyediagram flatten out. Therefore, to go beyond 250 km we will need to explore different features than the ones we are using now. We have not tested the method for dependence on the polarization dependent loss and polarization mode dispersion. However, it can be said that if the extracted features from the eyediagram are affected by PMD and PDL then the method will suffer in accuracy. One way around it is to find features that do vary with OSNR but are independent to PMD and PDL or to design training stage where the effects of PMD and PDL are taken into consideration. The same argument can be used for the nonlinear optical fibre impairments. In conclusion, more investigations are need to determine the accuracy of the proposed method in the presence of linear and nonlinear optical fibre impairments and these tasks remain for the future work.

A. *Feasibility for the real-time implementation*

Typically, optical performance monitoring unit does not necessarily have to run in real-time. It can run in quasi real-time similar to digital sampling oscilloscopes. When running in quasi real-time the requirements on the latency are not as stringent as in the case of the real-time operation. This of means that higher degree of complexity and latency can be considered. We have not done any evaluation on the feasibility of the method for the real and -quasi-real time implementation. However,

this important topic will be considered for the next publication.

V. Conclusion

The presented results show that techniques from machine learning such as neural networks based nonlinear regression and support vector machine classifiers are beneficial for optical-signal-to-noise ratio estimation and modulation format classification. The method for the OSNR estimation is independent of the modulation format, however, the prerequisite is that the training stage is performed using the modulation format that is going to be employed for the transmission. It has been shown experimentally that estimation of in-band OSNR and modulation format classification is achievable from directly detected PDM signals employing advanced modulation formats of up to 64QAM with varying pulseshapes.

VI. Acknowledgments

Research leading to these results has received funding from the Villum Foundation Young Investigator program.

References

[1] C. M. Bishop, Pattern recognition and machine learning, 2006.

[2] Z. Ghahramani, "Probabilistic machine learning and artificial intelligence," Nature, vol. 521, no. 7553, pp. 452–459, 2015.

[3] D. Zibar, L. Henrique, H. D. Carvalho, M. Piels, A. Doberstein, J. Diniz, B. Nebendahl, C. Franciscangelis, J. Estaran, H. Haisch, N. G. Gonzalez, J. C. R. F. D. Oliveira, and I. T. Monroy, "Application of Machine Learning Techniques for Amplitude and Phase Noise Characterization," Journal of Lightwave Technology, vol. 33, no. 7, pp. 1333–1343, 2015.

[4] B. Szafraniec, T. S. Marshall, and B. Nebendahl, "Performance monitoring and measurement techniques for coherent optical systems," Journal of Lightwave Technology, vol. 31, no. 4, pp. 648–663, 2013.

[5] L. Barletta, M. Magarini, and a. Spalvieri, "The information rate transferred through the discrete-time Wiener's phase noise channel," Journal of Lightwave Technology, vol. 30, no. 10, pp. 1480–1486, 2012.

[6] M. Piels, M. I. Olmedo, S. Member, W. Xue, X. Pang, C. Sch, R. Schatz, G. Jacobsen, I. T. Monroy, S. Member, J. Mørk, and S. Popov, "Laser Rate Equation-Based Filtering for Carrier Recovery in Characterization and Communication," Journal of Lightwave Technology, vol. 33, no. 15, pp. 3271–3279, 2015.

[7] D. Zibar, M. Piels, R. Jones, and C. G. Schaeffer, "Machine Learning Techniques in Optical Communication," Journal of Lightwave Technology, vol. 34, no. 6, pp. 1442–1452, 2016.

[8] L. Barletta, M. Magarini, S. Pecorino, and A. Spalvieri, "Upper and Lower Bounds to the Information Rate Transferred Through First-Order Markov Channels With Free-Running Continuous State," IEEE Transactions on Information Theory, vol. 60, no. 7, pp. 3834–3844, 2014.

[9] N. G. Gonzalez, D. Zibar, and I. T. Monroy, "Cognitive digital receiver for burst mode phase modulated radio over fiber links," European Conference on Optical Communication, ECOC, vol. 1-2, no. 1, pp. 23–25, 2010.

[10] R. Borkowski, D. Zibar, A. Caballero, V. Arlunno, and I. T. Monroy, "Stokes space-based optical modulation format recognition for digital coherent receivers," IEEE Photonics Technology Letters, vol. 25, pp. 2129–2132, 2013.

[11] P. Isautier, J. Pan, R. Desalvo, and S. E. Ralph, "Stokes space-based modulation format recognition for autonomous optical receivers," Journal of Lightwave Technology, vol. 33, no. 24, pp. 5157–5163, 2015.

[12] M. C. Tan, F. N. Khan, W. H. Al-Arashi, Y. D. Zhou, and a. P. T. Lau, "Simultaneous Optical Performance Monitoring and Modulation Format/Bit-Rate Identification Using Principal Component Analysis," Journal of Optical Communications and Networking, vol. 6, no. 5, pp. 441–448, 2014.

[13] X. Wu, J. Jargon, R. Skoog, L. Paraschis, and A. Willner, "Applications of Artificial Neural Networks in Optical Performance Monitoring," Journal of Lightwave Technology, vol. 27, no. 16, pp. 3580–3589, 2009.

[14] X. Wu, J. a. Jargon, L. Paraschis, and A. E. Willner, "ANN-based optical performance monitoring of QPSK signals using parameters derived from balanced-detected asynchronous diagrams," IEEE Photonics Technology Letters, vol. 23, no. 4, pp. 248–250, 2011.

[15] Z. Dong, F. Khan, Q. Sui, K. Zhong, C. Lu, and A. Lau, "Optical Performance Monitoring: A Review of Current and Future Technologies," Journal of Lightwave Technology, vol. 8724, no. c, pp. 1–1, 2015.

[16] X. Xie, F. Yaman, X. Zhou, and G. Li, "Polarization demultiplexing by independent component analysis," IEEE Photonics Technology Letters, vol. 22, no. 11, pp. 805–807, 2010.

[17] A. N. Tait, J. Chang, B. J. Shastri, M. A. Nahmias, and P. R. Prucnal, "Demonstration of WDM weighted addition for principal component analysis," Optics Express, vol. 23, no. 10, p. 12758, may 2015.

[18] S. Tan, H. Wymeersch, P. Johannisson, E. Agrell, P. Andrekson, and M. Karlsson, "An ML-based detector for optical communication in the presence of nonlinear phase noise," IEEE International Conference on Communications, vol. 9, 2011.

[19] D. Wang, M. Zhang, Z. Li, Y. Cui, J. Liu, Y. Yang, and H. Wang, "Nonlinear decision boundary created by a machine learning-based classifier to mitigate nonlinear phase noise," European Conference on Optical Communication, ECOC, vol. 2015-Novem, no. 1, pp. 2–4, 2015.

[20] T. Koike-Akino, C. Duan, K. Parsons, K. Kojima, T. Yoshida, T. Sugihara, and T. Mizuochi, "High-order statistical equalizer for nonlinearity compensation in dispersion-managed coherent optical communications." Optics express, vol. 20, no. 14, pp. 15 769–80, 2012.

[21] D. Zibar, O. Winther, and N. Franceschi, "Nonlinear impairment compensation using expectation maximization for dispersion managed and unmanaged PDM 16-QAM transmission," Optics ..., vol. 20, no. 26, pp. 181–196, 2012.

[22] M. Li, S. Yu, J. Yang, Z. Chen, Y. Han, and W. Gu, "Nonparameter Nonlinear Phase Noise Mitigation by Using M-ary Support Vector Machine for Coherent Optical Systems," IEEE Photonics Journal, vol. 5, no. 6, pp. 7 800 312–7 800 312, 2013.

[23] N. V. Irukulapati, S. Member, D. Marsella, P. Johannisson, E. Agrell, S. Member, M. Secondini, and H. Wymeersch, "Stochastic Digital Backpropagation with Residual Memory Compensation," Journal of Lightwave Technology, vol. 34, no. 2, pp. 566–572, 2016.

[24] L. Dou, Z. Tao, Y. Zhao, S. Oda, Y. Aoki, T. Hoshida, and C. Jens, "Differential Pilots Aided In-Band OSNR Monitor with Large Nonlinear Tolerance," OFC/NFOEC 2015 - Optical Fiber Communication and the National Fiber Optic Engineers Conference 2015, vol. 1, pp. 4–6, 2015.

[25] L. Dou, T. Yamauchi, X. Su, and Z. Tao, "An Accurate Nonlinear Noise Insensitive OSNR Monitor," Optical Fiber Communication (OFC),, vol. 1, no. 1, pp. 5–7, 2016.

[26] D. Dahan and U. Mahlab, "In-Band OSNR System Margin and Overall Link Impairment Level Monitoring Techniques Using an Optical Coherent Receiver," Journal of Lightwave Technology, vol. 33, no. 19, pp. 4063–4070, 2015.