

Manuscript version: Author's Accepted Manuscript

The version presented in WRAP is the author's accepted manuscript and may differ from the published version or Version of Record.

Persistent WRAP URL:

<http://wrap.warwick.ac.uk/131786>

How to cite:

Please refer to published version for the most recent bibliographic citation information. If a published version is known of, the repository item page linked to above, will contain details on accessing it.

Copyright and reuse:

The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions.

Copyright © and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable the material made available in WRAP has been checked for eligibility before being made available.

Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

Publisher's statement:

Please refer to the repository item page, publisher's statement section, for further information.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk.

Machine Learning for Fiber Nonlinearity Mitigation in Long-Haul Coherent Optical Transmission Systems

Yifan Liu¹, Bowei Yang², Tianhua Xu^{1,3,4*}

¹School of Engineering, University of Warwick, Coventry, United Kingdom, CV4 7AL

²Institute of Intelligent Communication Network & Security, Zhejiang University, Hangzhou, China, 310027

³School of Precision Instruments and Opto-Electronics Engineering, Tianjin University, Tianjin, China, 300072

⁴Optical Networks Group, University College London, London, United Kingdom, WC1E 7JE

e-mail: tianhua.xu@ieee.org

(Invited Paper)

Abstract—Fiber nonlinearities from Kerr effect are considered as major constraints for enhancing the transmission capacity in current optical transmission systems. Digital nonlinearity compensation techniques such as digital backpropagation can perform well but require high computing resources. Machine learning can provide a low complexity capability especially for high-dimensional classification problems. Recently several supervised and unsupervised machine learning techniques have been investigated in the field of fiber nonlinearity mitigation. This paper offers a brief review of the principles, performance and complexity of these machine learning approaches in the application of nonlinearity mitigation.

Keywords—machine learning; nonlinearity compensation; optical communications and networking;

I. INTRODUCTION

Optical fiber communication systems, as a core part of telecommunications, are expected to operate at high data rate with maximal throughput and robust error resilience under ultra large bandwidth [1]. Fiber nonlinearities, as a result of Kerr effect where the refractive index varies with the intensity of the optical signals, are considered as the primary challenges in enhancing the overall capacity of the optical transmission systems. Fiber nonlinear impairments including self-phase modulation, cross-phase modulation and four-wave mixing cooperatively affect the evolution of optical spectrum, signal phase and pulse shape along the propagation, which will significantly degrade the performance of optical transmission systems [2].

Digital backpropagation (DBP) approach has been widely investigated due to its satisfactory performance in cancelling deterministic fiber nonlinearities by reversely operating the signal propagation [3]. However, there still exist stochastic impairments e.g. polarization mode dispersion, laser phase noise and signal-to-noise interactions that cannot be well compensated by DBP. More seriously, the complexity of DBP prevents its potential real-time implementation in FPGA and related hardware [4].

The exploding amount of data manifests the application of machine learning (ML) techniques on optical communication systems and networks. Due to the strength of ML techniques in recognizing underlying connections and hidden patterns from data without the need to acquire complicated

mathematical models, it is prone to compensating stochastic distortions therefore an optimal solution for nonlinearity mitigation. Moreover, by learning characteristics of nonlinear impairments from the collected data at the receiver (Rx), ML techniques have great potential to compensate for stochastic nonlinearity-induced signal distortions while reducing the massive computing resource that DBP demands [5].

In this paper, we review both established and emerging ML methods which have been applied to fiber nonlinearity mitigation in long-haul optical transmission systems. Neural networks (NN) and support vector machine (SVM) as supervised learning methods and K-means clustering as an unsupervised algorithm are discussed. We illustrate how these ML algorithms have been applied for nonlinearity mitigation in optical transmission systems and compare the performance of their applications under the metrics of bit error rate (BER), Q-factor and computational complexity.

II. MACHINE LEARNING METHODS AND APPLICATIONS IN NONLINEARITY COMPENSATION

In this section, we review NN, SVM for supervised learning and K-means for unsupervised learning applied in point-to-point optical communication systems.

ML, when applied in nonlinearity compensation, is similar to digital compensation methods that equalizes the nonlinear effects at Rx side based on received symbols. Two major approaches are conducted: 1) to treat the received symbols as ordinary data samples and develop a ML model for symbol detection without considering system parameters; 2) to integrate fiber parameters into ML modeling, particularly input fiber parameters into neural network models [5]. The second approach utilizes more comprehensive knowledge of optical fibers and transmission systems.

Here we consider an optical transmission system with coherent detection described as follows. Pseudo random bit sequence (PRBS) data are converted to quadrature amplitude modulation (QAM) optical signals using the in-phase and quadrature (IQ) modulator. These optical signals are then transmitted in the transmission fiber loop with erbium doped fiber amplifier (EDFA) to compensate for the fiber loss in each span. Coherent detection is employed at the Rx, and symbol detection is realized after the signal equalization e.g. CD compensation, NLC, ML, carrier phase estimation (CPE).

Figure 1 illustrates a schematic diagram of the optical transmission system where blocks A, B and C denote ML applications for nonlinearity compensation at corresponding interfaces. Block B denotes the most prevalent scheme which blindly operates on received symbols as training data to generate the ML classifiers for symbol detection at the Rx end

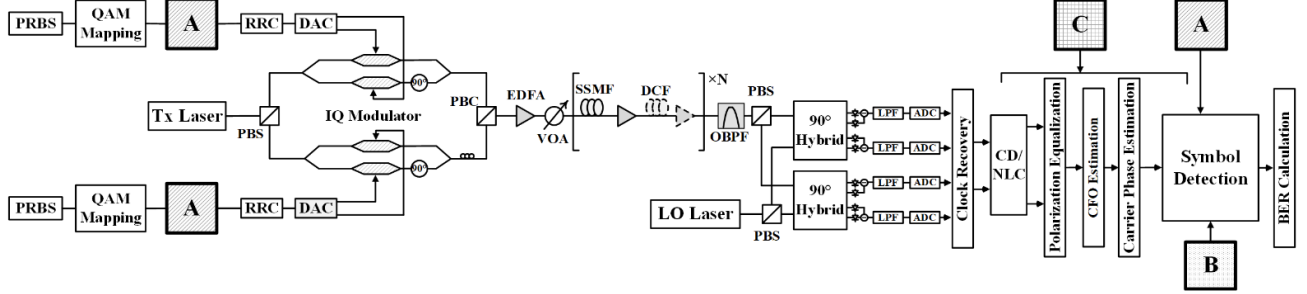


Figure 1. Schematic of machine learning-assisted optical transmission system: Block A, B and C denote different ML application types.

A. Neural Networks

Artificial neural networks (ANNs) that simulate neurons for information processing has demonstrated widespread applicability during the past decade with its unique structure of multiple layer perceptron, hidden layer and activation functions [6]. In communication systems, they are beginning to demonstrate potential in improving the performance of sub-systems. A typical deep neural network layout is depicted in Fig. 2: input nodes, output nodes with many intermediate hidden layers that consist of linear and nonlinear functions.

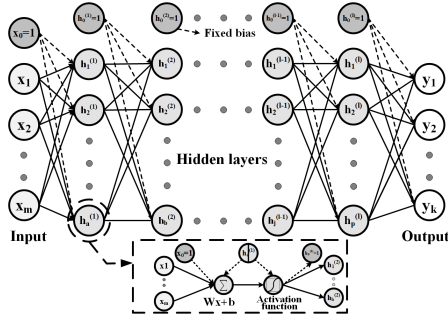


Figure 2. Deep neural network structure.

Compared to the DBP, neural networks can significantly reduce the computational complexity [7]. NN models are usually generated as a black box with aforementioned two major ML approaches for NLC that are applicable for neural networks. For the first equalization approach, the NN design is independent of the optical physics and the architecture can be considered sub-optimal either from a physics or a neuroevolution sense, while the second builds transmission system parameters into the construction of the NN model. Even though a functional NN could be built to serve the classification purpose from the former approach, the simulation and experimental results of the latter has shown significant improvement. Kamalov et al. proposed an

through training and execution period, the operation of which is also part of A and C. Block A represents the type of ML equalization where the trained model is developed based on received symbols and applied at either transmitter (Tx) side or Rx side. Block C represents the type of ML technique applied at multiple stages of the digital equalization process.

artificial-intelligence NLC (AI-NLC) algorithm considering “triplets” generated based on intra-channel cross-phase (IXPM) modulation and intra-channel four-wave mixing (IFWM) in time-domain perturbation pre/post-distortion (PPD) algorithm [9, 10]. Based on a similar concept of interpreting Tx symbols into input features fed into neural networks during training period, a field and lab experiment of nonlinearity compensation was conducted, and the obtained neural network model demonstrates the capability of being applied without the necessities of acquiring information of link parameters hence transparent to transmission systems. Also, a deep neural network structure was proposed by Häger and Pfister where linear and nonlinear functions of the deep neural networks were modeled based on the split-step Fourier method (SSFM) which is similar to a DBP structure by employing hidden layers as a substitution of the reverse SSFM for optical fibers which negate the nonlinear process with numerous steps for the sake of precision however with much less computing resource [7].

Due to the specific linear, nonlinear and layered indirect characteristics of NN, choosing appropriate activation function and loss function is also vital in achieving better performance for corresponding neural network models. However, the selection of the model appears to be case by case. Leaky RELU as an experimentally demonstrated example, gives the best performance under the NN configuration of [11], for which case the generated model can also be employed at either Tx or Rx with around 1dB Q-factor gain at Tx side compared to it employed at Rx side.

B. Support Vector Machine

SVM as a classic classification ML method that derives the largest margin i.e. largest support vector and finds a hyperplane to separate complex low-dimensional data where the distance from the separating hyperplane corresponds to the “confidence” of prediction [12]. Hinge loss and kernel method distinguish SVM from other ML techniques. Kernel method avoids complicated calculations when transforming data from

low to high dimension and equivalency has been illustrated between data transformation and the inner product of the input vector with all support vectors [13].

Figure 3(a) and Fig. 3(b) demonstrate the application of binary and multi-class SVM classifications on received symbols with nonlinear distortions under the modulation format QPSK at Rx side. The seemingly inseparable data is usually kernel-mapped to feature space to be separable and separated in high dimensions. Due to its robustness, SVM is believed to separate received symbols to correct classifications well during nonlinearity equalization process. An SVM classifier was first trained and applied in nonlinearity compensation for combating nonlinear phase noise in amplitude phase-shift keying (APSK) system [14].

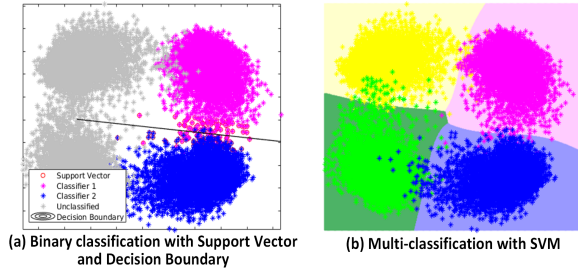


Figure 3. Schematic diagram of binary and multi-class SVM classification.

Recently, five SVM methods including: 1) the one versus rest (OvR) where the multi-classifiers are built one by one considering the rest belonging to the other class with the concept of binary SVM; 2) the symbol encoding; 3) the binary encoding (BE) is based on whether each bit of label feature is 0 or 1; 4) the constellation rows and columns (RC); and 5) the in-phase and quadrature components (IQC) were investigated and IQC indicates the optimal results among all five in terms of computing resource and hardware storage [15].

C. Clustering

Clustering method as unsupervised learning groups unlabeled data and K-means algorithm is considered the most commonly used clustering algorithm by obtaining hard boundary after minimizing the cost function according to cluster assignments from received symbols with three steps: 1) centroids initialization with randomized centroids of clusters; 2) cluster assignment based on minimal-distance principle; 3) moving centroids for updating and finalization with the goal of minimizing the objective function. Since the constellation pattern of transmitted symbols is determined with modulation format, centroids initialization for optimal performance of the algorithm can be achieved with the number of centroids consistent with constellation points. The application of K-means algorithm in optical transmission system under QPSK is illustrated through Figure 4. As stated, the initialization of centroids is fixed according to modulation format and its constellation shape in Fig. 4(a). The centroids are then easily updated as seen in Fig. 4(b), and the updating process will be terminated after certain times of iteration while

achieving a minimized objective function in Fig. 4(c). As can be seen K-means may not performed as supervised learning algorithm due to lacking training labels, its simplicity may bring practical benefits with real-time communications.

A novel density-centroid-tracking (DCT) algorithm was then proposed to enhance the performance by centroids initialization where the centroids can be initially tracked by the density of received symbols due to the fact of the non-convex squared error objective function of K-means [16]. A modified density-based spatial clustering of applications with noise (DBSCAN) algorithm was proposed [17]. This novel algorithm combines K-means clustering on the noisy un-clustered symbols and surpasses conventional K-means and fuzzy logic C-means algorithms. Advanced algorithms e.g. Hierarchical and Fuzzy-logic C-means (FLC) demonstrate the applicability to practical single- and multichannel optical communications after being successfully applied inter-disciplinarily [18]. Clustering algorithms can also be used in multiple stages of equalization introduced in [19], where K-means and Gaussian mixture model (GMM) algorithms are distributed at polarization equalization, CPE and symbol detections stages jointly validating clustering method in enhancing nonlinearity mitigation at system-level.

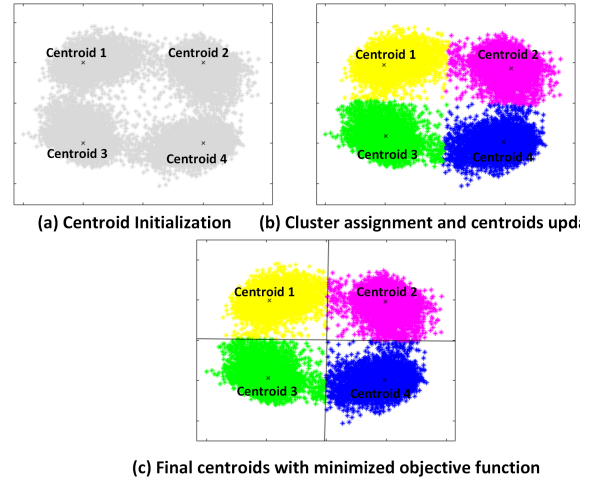


Figure 4. Schematic diagram of K-means algorithm for clustering.

III. PERFORMANCE AND DISCUSSION

Table I summarizes the transmission performance improvement with ML methods introduced in section III and considers transmission rate, transmission distance, number of channels under the metrics of BER, Q-factor and complexity for nonlinearity mitigation.

The performance of a technique based on supervised algorithm, Parzen window (PW), which classifies symbols at Rx side based on the labeled training data generated before by associating a label to each symbol, is also presented. It provides slightly better than SVM, as a two-step nonlinearity mitigation method is employed by first applying DBP for equalizing deterministic nonlinear effects and then combating stochastic nonlinearities with the PW algorithm [20].

TABLE I. PERFORMANCE OF ML METHODS APPLIED FOR NONLINEARITY MITIGATION

ML Algorithm		Parameters					
		Transmission data rate	Transmission distance	Modulation Format	No. of channels	Performance metric	Performance Improvement
Supervised	NN-LDBP [7]	20Gbaud	3200km	16QAM	1	Q-factor and complexity	3 steps per span (StPS) for LDBP is slightly (~0.1dB) better than 50 StPS in DBP
	NN-NLC [11]	32Gbaud	2800km	16QAM	1	Q-factor	Pre-distort symbols at transmitter side improves with ~0.6dB
	SVM [15]	100.3Gb	Back-to-back	64QAM	1	BER and complexity	IQC method shows that BER improves to the order of 2 with the least complexity among other SVM methods
	PW [20]	224Gb	1600km	16QAM/64 QAM	1	Q-factor	Dispersion unmanaged system and dispersion managed system with ~0.4dB improvement compared to systems without compensation techniques
Unsupervised	K-means [16]	75Gb	80km	64QAM	1	BER	Power penalty can be improved by ~2dB with DCT equalizer at BER 10^{-3}

IV. CONCLUSION

In this paper, the recent progress on ML based nonlinearity compensation is reviewed. Even though ML methods have shown performance improvements at blind symbol detection at Rx side, but do not necessarily have an advantage over physics-based DSP approaches. Nonetheless, there is reason to be hopeful that ML models will bring its superiority into full play when developed by integrating the optical impairment physics. We believe that the benefits ML brings to optical communications will become more and more significant with the rapid growth in exploding data volume. Its potential in detecting inherent connections of data is undeniable when traditional signal processing analysis is not capable in some scenarios such as combating stochastic impairments in optical communication systems.

ACKNOWLEDGMENT

This work is supported by EU H2020 RISE Program (No. 778305) and UK EPSRC Program Grant TRANSNET. The authors would like to thank Prof Weisi Guo from Cranfield University for project supervisions and helpful discussions.

REFERENCES

- [1] R. Ramaswami, K. Sivarajan, and G. Sasaki, Optical networks: a practical perspective. Morgan Kaufmann, 2009.
- [2] A. Ellis, M. McCarthy, M. Al Khateeb, M. Sorokina, and N. Doran, "Performance limits in optical communications due to fiber nonlinearity," Adv. Opt. Photon., vol. 9, no. 3, pp. 429–503, 2017.
- [3] P. Bayvel, R. Maher, T. Xu, G. Liga, N. A. Shevchenko, D. Lavery, A. Alvarado, and R. I. Killey, "Maximizing the optical network capacity," Phil. Trans. R. Soc. A, vol. 374, no. 2062, p. 20140440, 2016.
- [4] F. Musumeci, C. Rottondi, A. Nag, I. Macaluso, D. Zibar, M. Ruffini, and M. Tornatore, "An overview on application of machine learning techniques in optical networks," IEEE Commun. Surv. Tut., vol. 21, no. 2, pp. 1383–1408, 2018.
- [5] F. N. Khan, Q. Fan, C. Lu, and A. P. T. Lau, "An optical communication's perspective on machine learning and its applications," J. Lightw. Technol., vol. 37, no. 2, pp. 493–516, 2019.
- [6] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, p. 436, 2015.
- [7] C. Häger and H. D. Pfister, "Nonlinear interference mitigation via deep neural networks," in 2018 Opt. Fiber Commun. Conf. and Expo. (OFC). IEEE, 2018, pp. 1–3.
- [8] M. A. Jarajreh, E. Giacoumidis, I. Aldaya, S. T. Le, A. Tsokanos, Z. Ghassemloooy, and N. J. Doran, "Artificial neural network nonlinear equalizer for coherent optical ofdm," IEEE Photon. Technol. Lett., vol. 27, no. 4, pp. 387–390, 2014.
- [9] Z. Tao, L. Dou, W. Yan, L. Li, T. Hoshida, and J. C. Rasmussen, "Multiplier-free intrachannel nonlinearity compensating algorithm operating at symbol rate," Journal of Lightwave Technology, vol. 29, no. 17, pp. 2570–2576, 2011.
- [10] V. Kamalov, L. Jovanovski, V. Vusirikala, S. Zhang, F. Yaman, K. Nakamura, T. Inoue, E. Mateo, and Y. Inada, "Evolution from 8qam live traffic to ps 64-qam with neural-network based nonlinearity compensation on 11000 km open subsea cable," in Opt. Fiber Commun. Conf. Optical Society of America, 2018, pp. Th4D–5.
- [11] S. Zhang, F. Yaman, K. Nakamura, T. Inoue, V. Kamalov, L. Jovanovski, V. Vusirikala, E. Mateo, Y. Inada, and T. Wang, "Field and lab experimental demonstration of nonlinear impairment compensation using neural networks," Nat. Commun., vol. 10, no. 1, p. 3033, 2019.
- [12] N. M. Nasrabadi, "Pattern recognition and machine learning," J. Electron Imaging, vol. 16, no. 4, p. 049901, 2007.
- [13] S. Shalev-Shwartz and S. Ben-David, Understanding machine learning: From theory to algorithms. Cambridge university press, 2014.
- [14] Y. Han, S. Yu, M. Li, J. Yang, and W. Gu, "An svm-based detection for coherent optical apsk systems with nonlinear phase noise," IEEE Photonics J., vol. 6, no. 5, pp. 1–10, 2014.
- [15] C. Wang, J. Du, G. Chen, H. Wang, L. Sun, K. Xu, B. Liu, and Z. He, "Qam classification methods by svm machine learning for improved optical interconnection," Opt. Commun., vol. 444, pp. 1–8, 2019.
- [16] J. Zhang, W. Chen, M. Gao, B. Chen, and G. Shen, "Novel low-complexity fully-blind density-centroid-tracking equalizer for 64-qam coherent optical communication systems," in Opt. Fiber Commun. Conf. Optical Society of America, 2018, pp. M1G–4.
- [17] E. Giacoumidis, Y. Lin, and L. P. Barry, "Dbscan for nonlinear equalization in high-capacity multi-carrier optical communications," arXiv preprint arXiv:1902.01198, 2019.
- [18] E. Giacoumidis, A. Matin, J. Wei, N. J. Doran, L. P. Barry, and X. Wang, "Blind nonlinearity equalization by machine-learning-based clustering for single-and multi-channel coherent optical ofdm," J. of Lightw. Technol., vol. 36, no. 3, pp. 721–727, 2018.
- [19] M. Xu, J. Zhang, H. Zhang, Z. Jia, J. Wang, L. Cheng, L. A. Campos, and C. Knittle, "Multi-stage machine learning enhanced dsp for dp-64qam coherent optical transmission systems," in 2019 Opt. Fiber Commun. Conf. and Exhib. (OFC). IEEE, 2019, pp. 1–3.
- [20] A. Amari, X. Lin, O. A. Dobre, R. Venkatesan, and A. Alvarado, "A machine learning-based detection technique for optical fiber nonlinearity mitigation," IEEE Photon. Technol. Lett., vol. 31, no. 8, pp. 627–630, 2019.