Fiber Nonlinearity Equalization with Multi-Label Deep Learning Scalable to High-Order DP-QAM

Koike-Akino, T.; Millar, D.S.; Parsons, K.; Kojima, K.

TR2018-047 July 06, 2018

Abstract

We use deep neural network (DNN) to compensate for Kerr-induced nonlinearity in fiber-optic communications. The proposed DNN is scalable to high-order modulations by employing multi-label classification, achieving greater than 1.2 dB gain in nonlinear regimes

Signal Processing in Photonic Communications (SPPCom)

Fiber Nonlinearity Equalization with Multi-Label Deep Learning Scalable to High-Order DP-QAM

Toshiaki Koike-Akino, David S. Millar, Kieran Parsons, Keisuke Kojima

Mitsubishi Electric Research Laboratories (MERL), Cambridge, MA 02139, USA. *koike@merl.com*

Abstract: We use deep neural network (DNN) to compensate for Kerr-induced nonlinearity in fiber-optic communications. The proposed DNN is scalable to high-order modulations by employing multi-label classification, achieving greater than 1.2 dB gain in nonlinear regimes.

OCIS codes: (060.4510) Optical communications, (060.1660) Coherent communications, (060.4080) Modulation.

1. Introduction

To realize high-speed, reliable, and long-haul optical communications, machine learning techniques have recently been envisioned to mitigate fiber nonlinearity [1], e.g., Gaussian mixture [2], particle method [3], independent component analysis [4], hidden Markov [5], support vector machine (SVM) [6] and artificial neural network (NN) [7]. Modern deep NN (DNN) with rectified linear unit (ReLU) was also adopted to noncoherent optical communications [8]. In this paper, we make an analysis of DNN as well as recurrent NN (RNN) to show significant advantage against the other machine learning approaches. Compared to the previous work [7,8], we tackle the scalability problem in dual-polarization quadrature-amplitude modulation (DP-QAM), which calls for massive training data due to large multinomial classification, by introducing multi-label softmax classification. The multi-label classification is scalable to higher-order DP-QAM, and moreover makes the DNN output directly usable for soft-decision forward error correction (SD-FEC).

2. Deep learning for nonlinear compensation

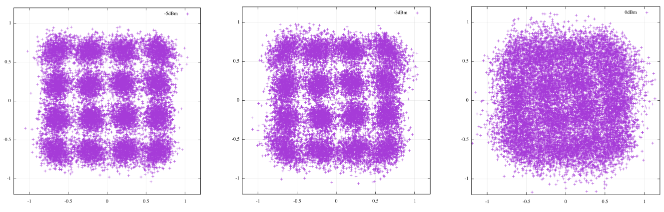
Deep learning [9] has been studied as a breakthrough technique in media processing researches. It uses many-layer many-node NN trained by a large amount of data. Note that big data are available in high-speed optical communications, which can provide terabit-class data in a second. The DNN is massively parallelizable in hardware, which is suited for future optical communications. In modern DNN, various techniques have been introduced, e.g., pre-training, mini-batch, ReLU, dropout, adaptive-moment (ADAM) stochastic gradient, convolutional architecture, and long short-term memory (LSTM) [9]. In this paper, we employ state-of-the-art DNN to cope with fiber nonlinearity.

Fig. 1 shows an example of residual distortion of DP-16QAM constellation after 31-tap linear equalizer (LE) for 15 spans of 80 km non-zero dispersion-shifted fiber (NZDSF) dispersion managed (DM) links with 5 % residual dispersion per span (RDPS). We can see that the constellation is more seriously distorted with the increased launch power due to Kerr fiber nonlinearity. To compensate for the residual nonlinear distortion after LE, we use a three-layer DNN having 1000 ReLU nodes at each layer. For DP-16QAM, there are 8 bits per symbol, leading to $2^8 = 256$ classes to identify. For such multi-class learning, we may use a single nonbinary softmax classification shown in Fig. 2(a), analogous to [8]. However, this single-label DNN does not perform well for a large multinomial classification in DP-QAM with a limited number of training data. This issue is more serious for higher-order DP-QAM, e.g., DP-64QAM requires 4096 classes to identify per symbol, which necessitates unrealistically huge data sets for training to converge.

To be scalable in high-order QAM, we use multi-label classification which employs multiple binary softmax classifiers as shown in Fig. 2(b). The multi-label DNN produces log-likelihood ratio (LLR), which can be directly fed into SD-FEC decoder without external processing such as [8, 10]. This is a great advantage in practice because LLR calculation is cumbersome, especially for high-order and high-dimensional modulation. Note that sum of cross-entropy minimization is equivalent to maximizing the generalized mutual information (GMI), which is used for SD-FEC performance metric. We consider a three-symbol delay line of LE output to feed into the DNN input in parallel. In addition to DNN, we also investigate RNN employing two-layer LSTM [9] to deal with longer-memory nonlinearity.

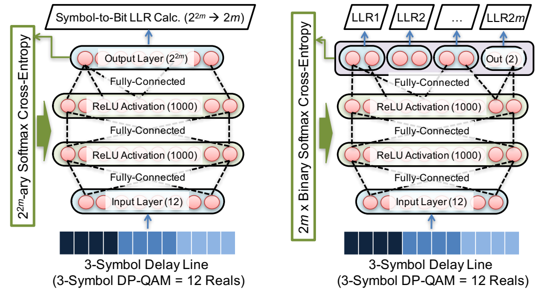
3. Performance results

We compare DNN and LSTM with classical machine learning methods, specifically, linear discriminant analysis (LDA), naïve Bayes (NB), quadratic discriminant analysis (QDA), and SVM. For multi-class SVM, we use one-vs-one rule with linear kernel as it worked best among several variants such as one-vs-all and polynomial kernel. We



(a) −5 dBm Launch (b) −3 dBm Launch (c) 0 dBm Launch

Fig. 1: Residual distortion of DP-16QAM constellation after LE for 15-span NZDSF DM links.



(a) One 2^{2m}-ary softmax (b) $2m\times$ binary softmax

Fig. 2: Single-/multi-label DNN for DP-2^mQAM.

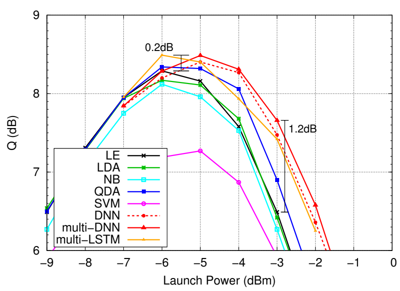


Fig. 3: DP-4QAM 50-span NZDSF.

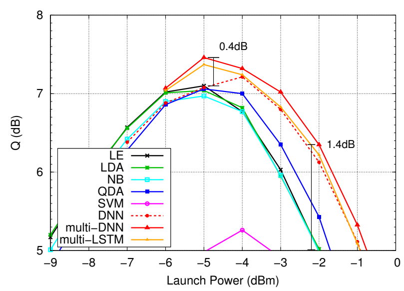


Fig. 4: DP-16QAM 16-span NZDSF.

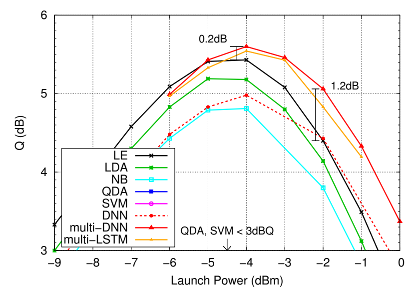


Fig. 5: DP-64QAM 8-span NZDSF.

assume 3-channel DP-QAM transmission for 34 GBd baud rate and 37.4 GHz channel spacing, over NZDSF DM links with 5 % RDPS, having a dispersion parameter of $D = 3.9$ ps/nm/km, a nonlinear factor of $\gamma = 1.6$ /W/km, and an attenuation of 0.2 dB/km. Span loss is compensated by ideal Erbium-doped fiber amplifiers (EDFA) with all amplified spontaneous emission noise added just before the receiver assuming the noise figure of 5 dB. We used digital root-raised cosine filters with 10% rolloff at both transmitter and receiver. The DNN weight is trained by ADAM with a dropout ratio of 0.5 and a batch size of 100 symbols to minimize a sum of softmax cross-entropy loss across all labels, using approximately 5×10^5 training symbols. Figs. 3, 4, and 5 show the Q factor versus launch power of DP-4QAM, DP-16QAM, and DP-64QAM, respectively, for 50, 16, and 8 spans times 80 km fiber configurations. It is observed that DNN can offer greater than 1.2 dB gain in highly nonlinear regimes. Although the improvement at the peak Q factor is not as significant (0.2–0.4 dB gain), it is more important to note that the multi-label classification improves the conventional DNN significantly in particular for higher-order QAM.

4. Conclusions

We made a performance comparison over various machine learning techniques for nonlinear compensation in coherent fiber communications. It was found that modern DNN performs best with 0.2–0.4 dB improvement in peak Q factor, while the Q improvement can be greater than 1.2 dB for highly nonlinear regimes. We also showed a significant advantage of multi-label classification, scalable to higher-order QAMs without the need of an external LLR calculator.

References

1. T. Koike-Akino, "Perspective of statistical learning for nonlinear equalization in coherent optical communications," *SPPCom* (2014), ST2D-2.
2. T. Koike-Akino et al., "High-order statistical equalizer for nonlinearity compensation in dispersion-managed coherent optical communications," *Opt. Expr.* **20** 14 (2012): 15769–15780.
3. N. Irukulapati et al., "Extending digital backpropagation to account for noise," *ECOC* (2013), We.3.C.4.
4. X. Xie et al., "Polarization demultiplexing by independent component analysis," *IEEE PTL* **22** 11 (2010): 805–807.
5. T. Koike-Akino et al., "Cycle slip-mitigating turbo demodulation in LDPC-coded coherent optical communications," *OFC* (2014), M3A-3.
6. E. Giacoumidis et al., "Kerr-induced nonlinearity reduction in coherent optical OFDM by low complexity support vector machine regression-based equalization," *OFC* (2016), Th2A-49.
7. M. A. Jarajreh et al., "Artificial neural network nonlinear equalizer for coherent optical OFDM," *IEEE PTL* **27** 4 (2015): 387–90.
8. R. Rios-Müller et al., "Experimental estimation of optical nonlinear memory channel conditional distribution using deep neural networks," *OFC* (2017): W2A5.1.
9. Y. LeCun et al., "Deep learning," *Nature* **521** 7553 (2015): 436–444.
10. T. Yoshida et al., "Hardware-efficient precise and flexible soft-demapping for multi-dimensional complementary APSK signals," *ECOC* (2016): Th.2.P2.SC3.27.