Attention-Based Neural Network Equalization in Fiber-Optic Communications

Abtin Shahkarami, Mansoor Yousefi, Yves Jaouën

Attention-Based Neural Network Equalization in Fiber-Optic Communications

Abtin Shahkarami, Mansoor I. Yousefi, and Yves Jaouen

Department of Communications and Electronics, Polytechnic Institute of Paris, Telecom Paris, 91120 Palaiseau, France.

(email: abtin.shahkarami@telecom-paris.fr, yousefi@telecom-paris.fr, yves.jaoeun@telecom-paris.fr)

Abstract:
An attention mechanism is integrated into neural network-based equalizers to prune the fully-connected output layer. For a 100 GBd 16-QAM 20×100 km SMF transmission, this approach reduces the computational complexity by ∼15% in a CNN+LSTM model.

1. Introduction

Nonlinear effects are a major liming factor in fiber-optic communications. The complexity of receivers designed to equalize these impairments must be reduced to enable higher data rates. Non-model-driven deep learning-based equalization methods have shown promise in this area.

In this category of methods, a neural network that does not incorporate the channel model (in contrast to model-driven architectures [1, 2]) is trained using pairs of the sampled waveform at RX (or typically the waveform after chromatic dispersion compensation) and the vector of transmitted symbols. Several such models have been proposed, including multi-layer perceptrons (MLP) [3], convolutional neural network (CNN) [4], bidirectional long short-term memory (bi-LSTM) [5], gated recurrent unit (GRU) [6], and CNN+LSTM models [7]. It is shown that the CNN+LSTM model outperforms the other models in terms of bit-error-rate (BER). [7].

In these models, the output layer is often a fully-connected layer with one hidden unit per symbol in PAM systems, or two hidden units per symbol in QAM systems (one for the real and one for the imaginary part). This results in $2 \times n^{[L-1]}$ real multiplication per symbol in the output layer, where $n^{[L-1]}$ is the number of hidden units in the layer $L-1$. Although dense layers could yield good BER performance, their complexity is high. Furthermore, the signal is already almost equalized in the previous layers, especially in the CNN+LSTM model; thus, paying the same attention to all the nodes in different temporal distances might not be optimal.

In this paper, we address this problem using an attention mechanism [8]. We consider a state-of-the-art CNN+LSTM model and reduce the number of floating-point operations (FLOPs) in it (for the inference mode) by learning and removing the low-impact connections in the output fully-connected layer. We also show that leveraging this technique facilitates replacing the LSTM units with simple RNN cells without performance deterioration. We present a cumulative FLOPs reduction of ∼27% using this approach, broken down into around 15% for connections-dropping and around 12% for replacing LSTMs.

2. Fiber-optic system model

Nonlinear Schrödinger (NLS) equation models the evolution of signal with one polarization in the standard single-mode fiber (SMF)

$$\frac{\partial q(t,z)}{\partial z} = -\frac{\alpha}{2} q(z,t) - \frac{j\beta_2}{2}\frac{\partial^2 q}{\partial t^2} + j\gamma |q(t,z)|^2 q(t,z), \quad 0 \le z \le \mathscr{L}, \tag{1}$$

where $q(t,z)$ is the complex envelope of the signal propagating in fiber as a function of time t and distance z. Here, γ is the nonlinearity parameter, β_2 and α are respectively chromatic dispersion and attenuation coefficients, and $\mathscr{L}$ is the fiber length.

The optical link is split into several spans of equal length. Split-step Fourier method (SSFM) is applied to solve NLS equation in order to calculate the propagation of the signal in each span. To compensate for fiber loss, amplification is performed using Erbium-doped fiber amplifiers (EDFA) after each span, followed by ASE noise addition.

At the transmitter (TX) first the input bit-stream $\mathbf{m} = (m_1, m_2, \ldots, m_{N_b})$, $m_i \in \{0,1\}$, is translated to a sequence of symbols $\mathbf{S} = (s_1, s_2, .., s_{N_s})$, where the symbols are drawn from a QAM-constellation. Following this, $\mathbf{S}$ is mapped to the waveform $q(t,0) = \sum_{i=0}^{N_s} s_i p(t - i/R_s)$, where $p(t)$ is the pulse shape and R_s is the baud rate. The signal $q(t,0)$ is then transmitted and propagated in the optical fiber according to (1). Following sampling the received waveform $q(t,\mathscr{L})$ at the receiver (RX), and typically after applying chromatic dispersion compensation (CDC), the resulting vector $\mathbf{x}$ is passed to an non-model-driven end-to-end neural network-based equalizer to retrieve the original sequence of symbols. The retrieved sequence of symbols and the corresponding bit-stream at RX are denoted by $\hat{\mathbf{S}}$ and $\hat{\mathbf{m}}$, respectively.

The goal of the neural network-based equalizer is to maximize the performance at fixed computational complexity assessed by the number of FLOPs. In this work, the performance is evaluated by the effective signal-to-noise ratio (SNR) $||\mathbf{S}||_2^2/||\mathbf{S}-\hat{\mathbf{S}}||_2^2$, where $||\mathbf{S}||_2 = (\Sigma|s_i|^2)^{1/2}$ is the L^2-norm of $\mathbf{S}$.

3. Attention mechanism and the proposed model

The motivation underlying our approach is to reduce the number of FLOPs in fiber-optic neural network-based equalizers by learning and removing the unnecessary connections in the output fully-connected layer. The idea of our approach is to make this possible using the attention mechanism. Without loss of generality, let us consider the state-of-the-art CNN+LSTM model [7] as the neural network-based equalizer under experiment. In this case, Fig. 1 illustrates the schematic of the attention mechanism integrated with this model. In this approach, an attention unit is placed after each (RNN) time-step t to evaluate its influence on the result of each nodes in the output layer.

The attention unit is implemented, as Fig. 2, with a fully-connected layer with N_s units followed by Softmax activation function $\sigma(\tau(i)) = e^{\tau(i)}/\Sigma_{j=0}^{N_s} e^{\tau(j)}$. This neural network, which outputs N_s values in the range $[0,1]$, learns how much influence the time-step t has on the value of each output symbol $\hat{s}_i$, captured by the attention score $\alpha^{<t,i>}$ ($\alpha^{<t,i>} = 0$ denotes no influence).

We highlight that this process is applied only during the training to obtain the attention scores. Once $\alpha^{<t,i>}$ are obtained, the attention units and the connections with attention scores lower than a given threshold should be removed. Then, the model requires to proceeds with or restart the training in order to update its weights.

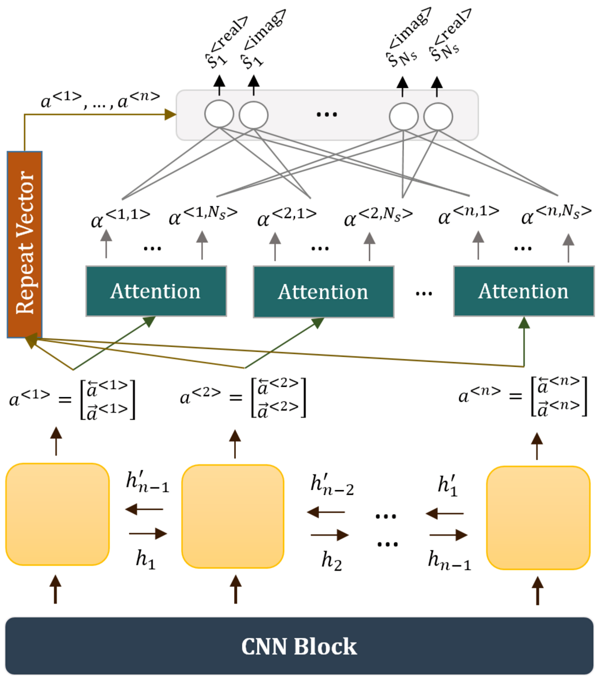

Fig. 1: The schematic of the attention-integrated CNN+RNN model. The CNN block initially processes the input to capture short-temporal dependencies. The output feature map is then passed to a recurrent layer to capture long-term features. There exits one attention unit on top of each RNN time-step.

The adopted attention mechanism also facilitates the replacement of LSTM units with simple RNN cells. Indeed, the attention mechanism could substitute for the primary purpose of using LSTMs in fiber-optics, which is to manage and learn what attention should be given to long-range time-steps. This substitution is done with the difference that the complexity of the attention mechanism does not remain in the inference mode, unlike LSTMs. However, it should be noted that since after the recurrent layer in the CNN+LSTM model, there is a fully connected layer, the advantage of using LSTMs over simple RNNs, in general, is questionable.

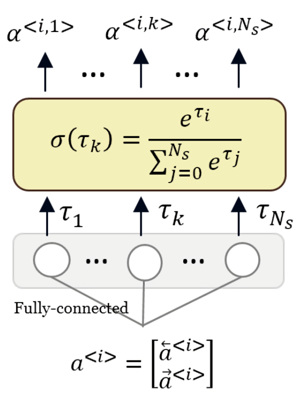

Fig. 2: The schematic of the attention unit.

4. Numerical results

A single-polarization 16-QAM 100 GBd fiber-optic communication system over 20x100km SMF optical-link was considered according to the system model discussed in Section 2. Root-raised cosine (RRC) filters, with roll-off of 0.25, were utilized as pulse shapes. Forward propagation of the optical signal at carrier wavelength of $\lambda_0 = 1550$nm was simulated using 8 sample/symbol and 50 step/span in SSFM (increasing either value did not affect the results). The sampling rate at RX was also adjusted to 8 sample/span.The fiber and noise parameters were assumed as follows: nonlinearity parameter $\gamma = 1.4\ \mathrm{W}^{-1}\mathrm{km}^{-1}$, chromatic dispersion $D = 17 ps/(nm-km)$, fiber loss $a_{\mathrm{dB}} = 0.2$ dB/km, and EDFA noise figure NF = 5 dB.

CNN+LSTM model and the corresponding attention-integrated version were adopted to learn the equalization. The models were created and trained using Tensorflow 2.0 in Python. Logcosh was set as the loss function, and the optimization was conducted using Adam algorithm with the learning rate of 0.001, having the reduce on plateau property with the patience of 7 and the factor of 0.7, $\beta_1 = 0.85$, and $\beta_2 = 0.999$. The hyper-parameters, the same for both the models, were optimized using k-fold cross-validation in such a way as to achieve the maximum performance to the same level as digital back-propagation (DBP). The obtained attention-integrated model was then pruned with the drop-connection threshold of 0.1%.

Fig. 3 demonstrates the effective SNR plot of the resulting models. As expected, both models provide the same level of performance. This is while the number of connections in the output layer of the attention-integrated model is dropped by 25%, resulting in around 15% FLOPs reduction over CNN+LSTM without attention. It is also noteworthy to elaborate that the output of the attention mechanism can be exploited more fruitfully in field-programmable gate array (FPGA) and circuit design frameworks by assigning lower bits to the connections with moderate impacts.

The performance of simple RNN cells in place of LSTMs was also investigated. It was observed the CNN+RNN achieves the same performance as CNN+LSTM in both cases. Considering this observation and the discussion in Section 3, it is highly motivated to use simple RNN cells in lieu of LSTMs. This replacement resulted in $\sim$12% FLOPs reduction in our considered scenario.

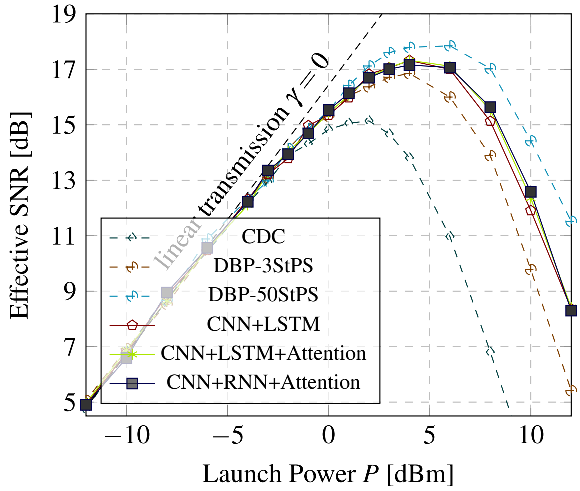

Fig. 3: The effective SNR of the models as a function of the launch power.

5. Conclusion

An attention mechanism was introduced to detect and drop the unnecessary connections in the output layer of non-model-driven equalizers in optical fiber communication systems. For a recently proposed CNN+LSTM architecture, in a 100 GBd 16-QAM 20×100km SMF optical transmission system, it was demonstrated that this technique drops 25% of connections in the output fully-connected layer without affecting the performance, resulting in 15% reduction in the total number of FLOPs. Further, it was shown that the LSTM units in the recurrent layer could be safely replaced with simple RNN cells, bringing an additional 12% FLOPs reduction in the considered system.

Acknowledgement

This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 766115.

References

1. R. M. Butler, C. Hager, H. D. Pfister, G. Liga, and A. Alvarado, "Model-based machine learning for joint digital backpropagation and pmd compensation," J. Light. Technol. pp. 1–1 (2020).
2. O. Sidelnikov, A. Redyuk, S. Sygletos, M. Fedoruk, and S. Turitsyn, "Advanced convolutional neural networks for nonlinearity mitigation in long-haul wdm transmission systems," J. Light. Technol. **39**, 2397–2406 (2021).
3. C. Catanese, R. Ayassi, E. Pincemin, and Y. Jaouën, "A fully connected neural network approach to mitigate fiber nonlinear effects in 200G DP-16-QAM transmission system," in *Proc. Int. Conf. TRN. Opt. Netw. (ICTON),* (2020), pp. 1–4.
4. C. Chuang, L. Liu, C. Wei, J. Liu, L. Henrickson, W. Huang, C. Wang, Y. Chen, and J. Chen, "Convolutional neural network based nonlinear classifier for 112-Gbps high speed optical link," in *Proc. Opt. Fiber Commun. Conf. Expo. (OFC),* (2018), pp. 1–3.
5. S. Deligiannidis, A. Bogris, C. Mesaritakis, and Y. Kopsinis, "Compensation of fiber nonlinearities in digital coherent systems leveraging long short-term memory neural networks," J. Light. Technol. **38**, 5991–5999 (2020).
6. X. Liu, Y. Wang, X. Wang, H. Xu, C. Li, and X. Xin, "Bi-directional gated recurrent unit neural network based nonlinear equalizer for coherent optical communication system," Opt. Express **29**, 5923–5933 (2021).
7. P. J. Freiredecarvalhosouza, Y. Osadchuk, B. Spinnler, A. Napoli, W. Schairer, N. M. S. da Costa, J. Prilepsky, and S. K. Turitsyn, "Performance versus complexity study of neural network equalizers in coherent optical systems," J. Light. Technol. (2021).
8. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS),* (2017), pp. 5998–6008.