Activation functions of artificial-neural-network-based nonlinear equalizers for optical nonlinearity compensation

Yuki Miyashita, Takeru Kyono, Kai Ikuta, Yuichiro Kurokawa, and Moriya Nakamura

School of Science and Technology, Meiji University, 1-1-1 Higashimita, Tama-ku, Kawasaki, Kanagawa 214-8571, Japan
a) m_naka@meiji.ac.jp

Abstract:

We investigated the performance of artificial neural network (ANN)-based nonlinear equalizers for optical nonlinearity compensation by comparing activation functions, including a sigmoid function, ReLU, and Leaky ReLU. We compared the learning speeds and compensation performances by evaluating the resulting error vector magnitudes of the compensated signals. The performance was investigated using simulated 100-km optical fiber transmission of 10-GSymbol/s 16QAM signals. When the number of hidden-layer units in the ANN was small, the sigmoid function showed better performance in learning speed than ReLU and Leaky ReLU. This point is important because the number of ANN units has to be reduced in order to improve the computational complexity of the ANN-based nonlinear equalizer.

Keywords: optical fiber communications, optical nonlinearity, digital signal processing, artificial neural network, activation function

Classification: Fiber-optic transmission for communications

References

[1] E. Ip, and J. M. Kahn, “Compensation of dispersion and nonlinear impairments using digital backpropagation,” J. Lightw. Technol., vol. 26, no. 20, pp. 3416-3425, Oct. 2008. DOI: 10.1109/JLT.2008.927791

[2] Y. Gao, F. Zhang, L. Dou, Z. Chen, and A. Xu, “Intra-channel nonlinearities mitigation in pseudo-linear coherent QPSK transmission systems via nonlinear electrical equalizer,” Opt. Commun., vol. 282, issue 12, pp. 2421-2425, Jun. 2009. DOI: 10.1016/j.optcom.2009.03.002

[3] S. Owaki, and M. Nakamura, “Equalization of optical nonlinear waveform distortion using neural-network based digital signal processing,” Optoelectronics and Communications Conference (OECC2016), WA2-40, Jul. 2016.
Introduction
Nonlinear waveform distortion caused by optical nonlinearities such as self-phase modulation (SPM) and cross-phase modulation (XPM) is one of the factors that limit the transmission distance of long-haul optical fiber communication systems. Some nonlinear compensation methods based on digital signal processing...
Fig. 1. (a) ANN unit. (b) ANN-based nonlinear equalizer. (c) Sigmoid function. (d) ReLU. (e) Leaky ReLU.

(DSP) have been investigated, including digital back propagation (DBP) and the Volterra series transfer function (VSTF) [1,2]. However, these methods need an enormous amount of calculations, which increases the power consumption and the delay time at the receiver. On the other hand, artificial neural network (ANN)-based nonlinear equalizers are attracting attention because of their lower computational complexity [3-7]. In recent years, the rectified linear unit (ReLU) has often been employed as an activation function of the ANN-unit instead of the conventional sigmoid function in deep neural networks (DNNs) used for, e.g., speech recognition and image recognition [8,9]. ReLU has some advantages compared to the sigmoid function, including fewer vanishing gradient problems and efficient computation [10]. Some investigations of ANN-based nonlinear equalizers also employed ReLU as the activation function [11,12]. In this paper, we investigate the performance of three-layer ANN-based nonlinear equalizers by comparing the sigmoid function and ReLU as the activation functions. Besides our former study in [13], we have also investigated the performance of Leaky ReLU [14]. We also investigated how the performance depends on the number of hidden-layer units in the three-layer ANN. Learning speed and compensation performance were compared by numerical simulations of 100-km standard single-mode fiber (SSMF) transmission of a single-channel 16-ary quadrature amplitude modulation (16QAM) signal.

2 Activation functions of ANN-based nonlinear equalizers

Figure 1(a) shows the construction of a unit in a complex-valued ANN [4]. The complex inner potential, $u$, is described as

$$u = \sum_{k=1}^{n} w_k x_k + b,$$  

(1)

where $x_k$ is the complex input signal, $w_k$ is the complex weight, and $b$ is the complex bias. The complex output of the unit, $y$, is expressed as

$$y = f(\text{Re}[u]) + j \times f(\text{Im}[u]),$$  

(2)

where $f$ is the activation function of the unit. Figure 1(b) shows an ANN-based
nonlinear equalizer composed of a three-layer complex-valued ANN with a feedforward tapped delay line. The weights and biases are updated in the learning process using the error back propagation (EBP) algorithm. The units in the input and output layers have linear activation functions. The units in the hidden layer have nonlinear activation functions. Figures 1(c), (d), and (e) show three nonlinear activation functions: a sigmoid function, ReLU, and Leaky ReLU, respectively. The sigmoid function has an S-shaped curve that is expressed as

\[ f(x) = \frac{1}{1 + e^{-x}}. \tag{3} \]

One of the advantages of the sigmoid function is that the derivative can be expressed by the output of the function itself as

\[ f'(x) = f(x)\{1 - f(x)\}, \tag{4} \]

which is used in the EBP algorithm. ReLU and Leaky ReLU are kinds of piecewise-linear functions described as

\[ f(x) = \begin{cases} 
  x & (x \geq 0) \\
  ax & (x < 0)
\end{cases}, \tag{5} \]

where \(a\) is a coefficient that controls the slope of the negative part of the function. If \(a\) is zero, the function becomes a standard ReLU. If \(a\) is a small fixed value, typically about 0.01, the function is called Leaky ReLU [14]. Leaky ReLU can avoid zero gradients over its entire domain, unlike the standard ReLU. This simple calculation contributes to a lower computational cost in both compensation and training processes. However, we can employ a lookup table to reduce the computational cost of the sigmoid function. The comparison of the computational cost of the activation functions depends on the structure of the DSP.

3 System setup

Figure 2 shows the optical transmission system used in our simulations. A 10-Gsymbol/s 16QAM optical signal was transmitted by 100-km SSMF. The optical fiber had a dispersion of 16.75 ps/nm/km. The input power to the optical fiber was 10 dBm. The noise figure of the Er-doped fiber amplifiers (EDFAs) was 3 dB. The optical signal was received by optical homodyne detection using an optical 90°-hybrid and balanced photodetectors (BPDs). Here, we assumed that the local oscillator (LO) was ideally synchronized to the optical signal. The transmitted signal was distorted by chromatic dispersion and optical nonlinearity. The distorted signal was compensated using the ANN-based nonlinear equalizer in the DSP. We
used PRBS $2^{15}-1$ data in the training of the ANN, and PRBS $2^{11}-1$ to evaluate the performance of the system and the equalizer to eliminate the possibility of overfitting. The number of input-layer units was 7, which is the same as the number of taps of the tapped delay line. The number of hidden-layer units, $N_{\text{hidden}}$, was varied from 2 to 20. The signal quality after the compensation was evaluated by using the error vector magnitude (EVM). Figure 2 also shows the constellations obtained with a back-to-back (BtB) setup and after 100 km transmission. The EVMs of the constellations were 1.25 % and 43.7 %, respectively. The nonlinear equalizer works to reduce the degraded EVM, and the target EVM is 1.25 %.

4 Results and discussion

Figure 3(a) shows the error versus the number of iterations of the learning steps, where $N_{\text{hidden}}$ was as small as 3. We performed the learning process 10 times, while

![Figure 3](image-url)
changing the random initial values of the weight each time. In this case, ReLU and Leaky ReLU showed large variations in convergence of the learning process. Figures 3(b), (c), and (d) show the constellations after the nonlinear equalization, where $N_{\text{hidden}}$ was 3, using the sigmoid function, ReLU, and Leaky ReLU, respectively. Each constellation shows the worst case of the compensation results evaluated using the EVM. In the cases of ReLU and Leaky ReLU, we observed serious residual distortion, as shown in Figs. 3(c) and (d). When $N_{\text{hidden}}$ was 20, however, this variation in convergence was improved to some extent, as shown in Fig. 3(e). Figures 3(f), (g), and (h) show the worst constellations after the compensation using the sigmoid function, ReLU, and Leaky ReLU, respectively. When $N_{\text{hidden}}$ was 20, we could not find any significant difference in the compensation performance using the different activation functions. Figure 3(i) shows the EVM after the compensation versus the number of hidden-layer units, $N_{\text{hidden}}$. We plotted the average of the 10 trials of the learning process. The error bars show the standard deviation. When $N_{\text{hidden}}$ was small, the compensation performance using ReLU or Leaky ReLU was degraded in comparison with the case of the sigmoid function. Additionally, EVM showed larger variation. We could not find any significant difference in the compensation performance between ReLU and Leaky ReLU. The simple calculation of ReLU and Leaky ReLU contributes to a lower computational cost. However, we also have to decrease the number of ANN units to lower the computational cost. It is known that ReLU shows good performance in DNNs which include a huge number of hidden-layer units [8,9]. However, it should be noted that ReLU showed different aspects when we used it in the ANN-based nonlinear equalizer with the smaller number of hidden-layer units.

5 Conclusion

We investigated the performance of ANN-based nonlinear equalizers for optical nonlinearity compensation, using a sigmoid function, ReLU, and Leaky ReLU as the activation functions of the hidden-layer units. The sigmoid function showed better performance than ReLU and Leaky ReLU when the number of hidden-layer units was small. The sigmoid function requires higher computational cost in both compensation and training processes. One possible way to avoid this problem is to use a lookup table.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Number JP20K05367 and The Fujikura Foundation.