Accurate Data-driven Fiber Channel Modeling Based on BiLSTM and Conditional GAN

Jiaming Liu
School of Communication and Information Engineering
University of Electronic Science and Technology of China
Chengdu, China
15583206681@163.com

Xiatao Huang
School of Communication and Information Engineering
University of Electronic Science and Technology of China
Chengdu, China
201711010116@std.uestc.edu.cn

Xuecheng Ren
School of Communication and Information Engineering
University of Electronic Science and Technology of China
Chengdu, China
r18317862976@163.com

Jinjiang Li
School of Communication and Information Engineering
University of Electronic Science and Technology of China
Chengdu, China
L15583202986@163.com

Jing Zhang*
School of Communication and Information Engineering
University of Electronic Science and Technology of China
Chengdu, China
zhangjing1983@uestc.edu.cn

Qianwu Zhang
Key Laboratory of Specialty Fiber Optics and Optical Access Networks
Shanghai University
Shanghai, China
zhangqianwu@shu.edu.cn

***Abstract*—In this paper, we compare the performance of two data-driven fiber channel modeling methods based on Bi-LSTM and CGAN waveform modeling schemes. The results show that the two methods can learn the accurate transfer function of the fiber channel, with only 0.6% average MSE compared with SSFM. The running times are less than 0.2 seconds and 0.25 seconds for 200-km transmission, versus 400 seconds using the SSFM under the same hardware conditions, which shows the reduction in complexity of the fiber channel modeling.**

Keywords—Data-driven; channel modeling; Bidirectional Long Short-Term Memory; Generative Adversarial Networks

I. INTRODUCTION

In the field of optical fiber communications, the evaluation of the transmission performance is time consuming when there are fiber nonlinearities. The existence of various nonlinear effects and random noise in fiber, which brings great difficulties to information detection and processing. A reliable channel model is needed to simulate fiber nonlinear effects and evaluate the performance of different nonlinearity mitigation strategies. The conventional channel modeling is based on the split-step Fourier method (SSFM), which is carried out by solving the nonlinear Schrödinger equation (NLSE) approximately[1]. However, the iteration steps of SSFM result in high complexity of computation. To avoid such computational complexity, many fiber channel models are proposed, such as the Gaussian noise (GN) model[2] and nonlinear-interference noise model. Although these models can predict the SNR of the signal with minor inaccuracies, they are not available to model the waveforms during transmission. The waveform modeling can provide both time and frequency information, which can be used for waveform observation, and optical system optimization. Fast and accurate channel modeling is still an open issue for further discussions.

Neural network (NN) is a powerful interdisciplinary tool for promoting the development of artificial intelligence (AI) in various emerging areas. Recently, the data-driven NN has ignited massive applications both in industry and academic researches[3]. For optical fiber communication systems, NNs have also been investigated in fiber nonlinearity equalization, carrier recovery and so on. Compared with the conventional model-driven method, the data-driven approach prevents complex mathematical theories and expert knowledge. The calculation operations are mainly multiplications and additions avoiding complicated operations.

In this work, we explore a comparative analysis of two different types of data-driven neural networks for modeling the fiber channels with the characteristics of chromatic dispersion (CD), self-phase modulation, attenuation, and amplified spontaneous emission (ASE) noise induced by erbium-doped fiber amplifier (EDFA). One is Bi-LSTM[4], a special type of efficient RNN, which can deal with sequential data effectively. The other is the Generative Adversarial Network (GAN)[5], a competitive learning framework consisting of generators and discriminators which can generate data from random noise. The ability of data-driven NNs to estimate the transfer function of the fiber channel is demonstrated from constellations, optical waveforms and the normalized mean square errors (MSEs). Moreover, we compare with the waveform generated by SSFM method as a benchmark both in terms of accurately and complexity. We clearly show that with the use of many-to-many training, the complexity of two models is only 0.6% average MSE compared with SSFM whilst the efficiency is much higher than SSFM.

II. CHANNEL MODELING METHOD

A. Traditional fiber channel modeling based on SSFM

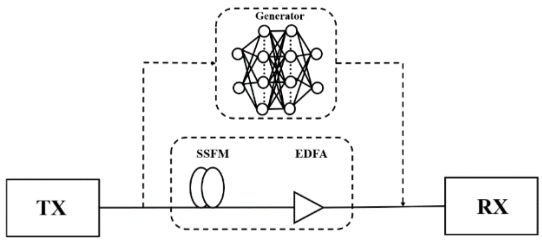


Fig. 1. Fiber-optic transmission system.

As shown in Fig. 1, we simulate an optical fiber communication system to test the versatility of two data-driven NN methods. The transmitter consists of modulation, oversampling, pulse shaping, and power normalization. We assume all the symbols and samples are expressed as complex values in this system. Following the two times up-sampling, root raised cosine (RRC) filter is used for signal shaping, which satisfies the Nyquist criterion. Power normalization controls the transmission power of the optical signal, and then the signal is transmitted to the fiber link.

The propagation model of optical pulse through a single mode fiber can be expressed by the NLSE:

$$\frac{\partial A}{\partial z} = -\frac{\alpha}{2}A + \frac{i}{2}\beta_2\frac{\partial^2 A}{\partial t^2} + \frac{1}{6}\beta_3\frac{\partial^3 A}{\partial t^3} + i\gamma|A|^2 \cdot A \tag{1}$$

where A is the complex envelope of the optical field and z is the propagation distance. The α, β_2, β_3, and γ, represent the propagation attenuation, group velocity dispersion slope, and nonlinear coefficient, respectively. There is no analytical solution for NLSE. By solving the Eq.(1), the most commonly used scheme is SSFM, which divides optical fiber into several small segments, adds nonlinear effect and ASE noise to the end of each segment. The SSFM is an effective and relatively accurate method, with high accuracy, but high computational complexity, requiring a lot of time and memory, especially in the case of long transmission and highly nonlinear fiber transmission.

B. *Bidirectional Long Short-Term Memory*

The LSTM is a type of NN that is used to process sequence data with memory. The perfect internal memory ability of the LSTM motivates us for channel modeling with a time memory. Because past and future samples both affect the current samples, we use the Bi-LSTM to capture the features from two directions. The LSTM is a variant of RNN that can effectively solve the problem of gradient vanishing and gradient explosion. It is composed of an input gate, a forget gate, an output gate and a memory unit, which can control the storage and update of information according to the input data and historical status. The structure of LSTM network is shown in Fig. 2.

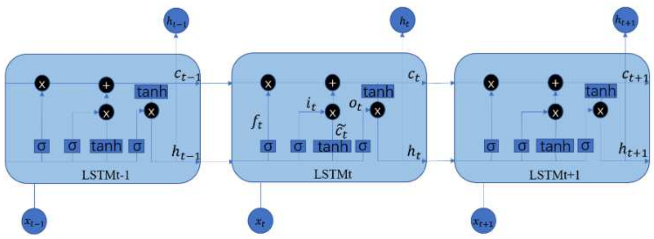


Fig. 2. The structure of LSTM network.

Figure 2 illustrates the LSTM units that indicate how the output h_t is calculated in each case,

$$\begin{aligned} i_t &= \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i) \\ f_t &= \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f) \\ o_t &= \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o) \\ c_t &= f_t * c_{t-1} + i_t * \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \\ h_t &= o_t * \tanh(c_t) \end{aligned} \tag{2}$$

where W *strands for* matrix that contains the weights of connection. The f, i, o and c are forget-, input-, output-gate and cell state. The x_t, h_t, h_{t-1} and b are input-, hidden-output, previous hidden output and bias vectors, respectively. The '$*$' operator denotes the element wise product, σ is the logistic sigmoid function and *tanh* is the hyperbolic tangent activation function.

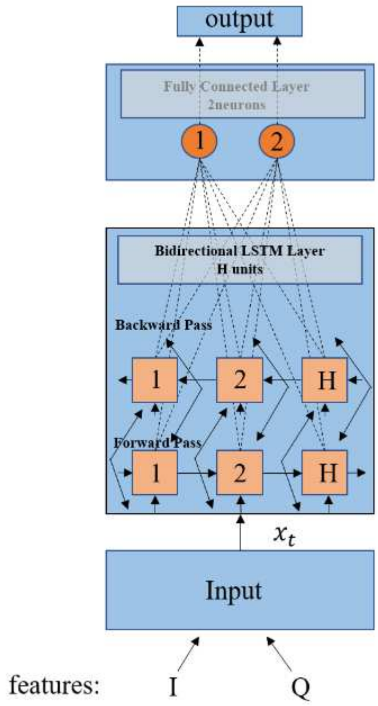


Fig. 3. The structure of Bi-LSTM network.

The sequential neural model is demonstrated in Fig. 3. The input x_t is the transmitted symbol sequence which has the following form $X_{t,L} = [x_{t-H}, \dots x_t, \dots x_{t+H}]$, where L stands for the overall length which is equal to $L = 2H+1$. At time t, we also launch H preceding and H succeeding symbols so as to track the dependencies of the adjacent symbols. The length of L is strictly related to total accumulated chromatic dispersion. Each symbol in each window contains two features (I and Q datas) as the input. We drive the Bi-LSTM output to a fully connected layer with 2 neurons corresponding to the transmitted of the channel waveforms with I and Q components.

C. *Generative Adversarial Networks*

GAN is a competitive framework consisting of generators and discriminators that can generate data that is similar to the real data from random noise. By adding some extra conditional information to the GAN to control the outputs, which is called conditional GAN (CGAN). We adopt CGAN to model the optical fiber channel waveform[7], and the structure of CGAN is shown in Fig. 3. Generator aims to capture the training data distribution and generate new data with the same distribution to fool the discriminator. A vector of noise z with a certain prior distribution and the condition vector x are sent into the generator and mapped to the generative fake data. Different noise vectors can map to different generative data. The condition vector x defines the characteristics of the generative data. Discriminator classifies the real data and fake data with the addition of the condition vector x. The output of discriminator $D(x)$ represents the probability that x is the real data. The total optimization loss function can be represented as,

$$\begin{aligned} \min_G \max_D V(D,G) = E_{y P data(y)}\left[\log D(y|x)\right] + \\ E_{z p_z(z)}\left[\log\left(1 - D\left(G(z|x)\right)\right)\right] \end{aligned} \tag{3}$$

The goal of the generator is to deceive the discriminator as much as possible so that it cannot distinguish between real data and generated data. The goal of the discriminator is to identify the real data and the generated data as much as possible. By constantly training the two networks alternately, a Nash equilibrium[6] is eventually reached, allowing the generator to produce high-quality data, while the discriminator cannot distinguish between true and false. In the end, the ultimate goal is to make the discriminator unable to estimate whether the output of the generator is true, i.e., the output of discriminator approximating to 0.5. The structure of CGAN is shown in Fig. 4.

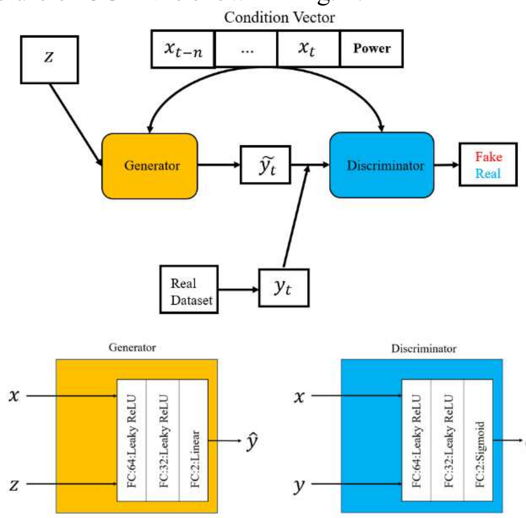


Fig. 4. The structure of GAN.

As shown in Fig. 4, the condition vector structure is customized for fiber channel modeling to improve modeling accuracy, flexibility, and training speed. Considering ISI caused by CD, the condition vector needs to contain several samples, which can help GAN learn the relevance between sequences. The current time-step transmitted samples are centered in the condition vector $X_{t,L} = [x_{t-H}, \ldots x_t, \ldots x_{t+H}]$, where x_t represents the current transmitted samples, x_{t-i} denotes the previous samples, and x_{t+i} are the subsequent samples to x_t. The number of adjacent samples is proportional to the delay caused by CD, relating to the transmission length and signal rate. Next, we use two samples estimation, generates two symbols each time, the result is more accurate than one sample estimation. It is also feasible to generate more symbols each time, but the longer training time is required. Moreover, considering that the input of the neural network must be a real number, we concatenate the real and imaginary part of two complex numbers to form a four-dimensional array. To train the NN, the channel input and output should be unit power normalized firstly before input to the generator and discriminator as,

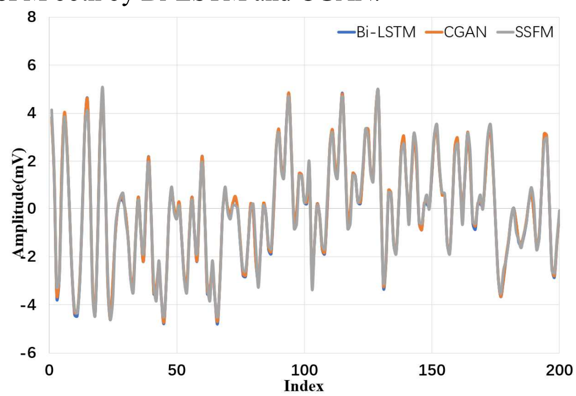
$$\tilde{x}_i = \frac{x_i}{\sqrt{\frac{1}{S}\sum_{i=1}^{S}|x_i|^2}} \tag{4}$$

III. Simulation Research and Verification

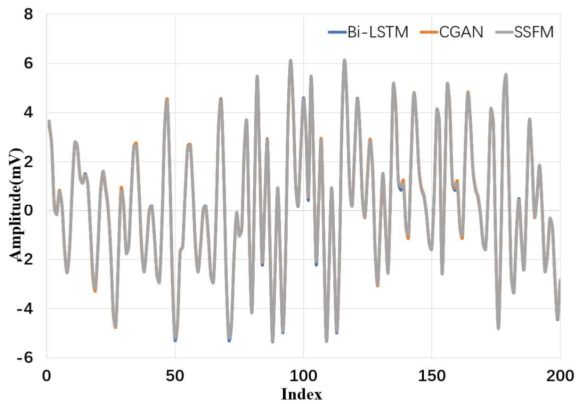
To collect the training data and analyze the signal modeling performance, we simulate a coherent optical transmission system, including the transmitter, the fiber channel, and the receiver. At the transmitter side, one 32-GBaud 16-QAM sequence with 65536 symbols mapped from a random bit source is generated. After the raised-cosine pulse shaping with 0.2 roll-off factor, the signal is launched into the optical I/Q modulator. An EDFA is used to adjust and monitor the signal power before transmit into the fiber link, which is comprised of one span of 100-km standard single mode fibers which has an attenuation factor of 0.2 dB/km, CD parameter 16.9 ps/(nm·km), and nonlinear coefficient 1.3 1/(W·km)). An EDFA with a noise figure of 5 dB is used to compensate the fiber loss. The number of the span is 2. At the receiver, coherent detection is utilized.

To train and test the NN, we allocate all the transmitted symbols into 549 sets and each set contains 120 transmit symbols. The dataset is used to train the NN iteratively with batch size sets 32. The first 75% of the data as a training dataset and the rest 25% of data as a testing dataset. The testing dataset is utilized to guarantee the network reaching the optimized performance. We use the PyTorch framework to build and train both networks, and the Adam optimizer to update the network parameters. The Bi-LSTM network is composed of a layer of bidirectional LSTM units with 32 hidden units, with a fully connected layer with 2 neurons. The CGAN network consists of a generator and discriminator composed with three fully connected layers and using LeakyReLU as the activation function. The discriminator has three fully connected layers, using LeakyReLU and Sigmoid as activation functions.

Fig. 5 illustrates the amplitudes of optical waveforms of SSFM-based channel output, Bi-LSTM-based channel output and CGAN-based channel output in two launch powers, -10 dBm and 2 dBm. From the overall view of the optical waveform, the channel output generated by SSFM and two NNs have similar amplitudes at each sampling point. The consistency for optical waveforms demonstrates that the time-domain characteristics of signals are fitted well with SSFM both by Bi-LSTM and CGAN.



(a)



(b)

Fig. 5. Bi-LSTM and CGAN network generated data with real data simulation waveforms: (a) Transmitted power = -10dBm (b) Transmitted power = 2dBm

We add ASE noise of EDFA after each fiber span in SSFM simulation for a real transmission structure to improve the robustness of NN. Then, we collect the waveform of the transceiver at the end of the system when the transmitting power was -10dBm, -6dBm, -2dBm and 2dBm, and put them into the trained network for testing. We use MSEs under different power as indicators to measure the performance of the two networks.

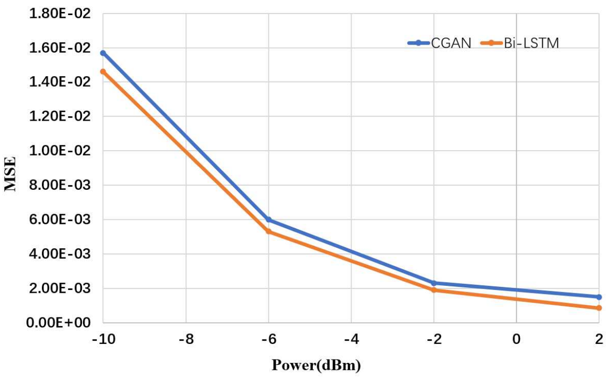


Fig. 6. Normalized MSE at different launch power.

Figure 6 shows the MSEs of two data-driven NNs. First, the values of two MSE decrease as the launch power increases since the ASE noise influence. The average MSE is less than 0.006 for Bi-LSTM and 0.009 for CGAN. The NN can only learn the data characteristics but not the noise distribution. Second, Bi-LSTM structure has a lower MSE than CGAN network, which indicates that Bi-LSTM can capture dynamic changes in fiber communication channels more effectively and generate more realistic data from the noise contaminated data.

IV. CONCLUSIONS

In this paper, we demonstrate the two data-driven methods for optical fiber communication channel modeling. Both results of Bi-LSTM and CGAN show good fitness with the real simulation waveforms with only 0.6% average MSE. Furthermore, we find Bi-LSTM structure can capture more dynamic characteristics of optical fiber communication channel than CGAN. This work acts as a reference for numerical solution of NLSE-like nonlinear partial differential equations. Bi-LSTM and CGAN have the potential to model more devices and even the whole system.

V. ACKNOWLEDGEMENT

This work was supported by National Key Research and Development Program of China (2022YFB2903303), STCSM (SKLSFO2021-01), Fundamental Research Funds for the Central Universities (ZYGX2020ZB043 and ZYGX2019J008).

REFERENCES

[1] Liu X, Lee B. A fast method for nonlinear Schrodinger equation. IEEE Photonics Technology Letters, 2003, 15(11): 1549-1551.

[2] Poggiolini P, Bosco G, Carena A, et al. The GN-model of fiber non-linear propagation and its applications. Journal of lightwave technology, 2013, 32(4): 694-721.

[3] Wang D, Song Y, Li J, et al. Data-driven optical fiber channel modeling: A deep learning approach. Journal of Lightwave Technology, 2020, 38(17): 4730-4743.

[4] Deligiannidis S, Mesaritakis C, Bogris A. Performance and complexity analysis of bi-directional recurrent neural network models versus volterra nonlinear equalizers in digital coherent systems. Journal of Lightwave Technology, 2021, 39(18): 5791-5798.

[5] Yang H, Niu Z, Xiao S, et al. Fast and accurate optical fiber channel modeling using generative adversarial network. Journal of Lightwave Technology, 2020, 39(5): 1322-1333.

[6] Hsieh Y P, Liu C, Cevher V. Finding mixed nash equilibria of generative adversarial networks//International Conference on Machine Learning. PMLR, 2019: 2810-2819.

[7] Ye G, Xiang J, Zhou G, et al. Impact of the input OSNR on data-driven optical fiber channel modeling. Journal of Optical Communications and Networking, 2023, 15(2): 78-86.