SCIENCE CHINA
Information Sciences

June 2020, Vol. 63 160305:1–160305:13
https://doi.org/10.1007/s11432-020-2873-x

• RESEARCH PAPER •
Special Focus on Artificial Intelligence for Optical Communications

Overfitting effect of artificial neural network based nonlinear equalizer: from mathematical origin to transmission evolution

Zheng YANG[1], Fan GAO[2], Songnian FU[1*], Ming TANG[1] & Deming LIU[1]

[1] *Wuhan National Laboratory for Optoelectronics, School of Optical and Electronic Information, Huazhong University of Science and Technology, Wuhan* 430074*, China;*
[2] *Alibaba Infrastructure Service, Alibaba Group, Hangzhou* 311121*, China*

Received 26 February 2020/Revised 2 April 2020/Accepted 13 April 2020/Published online 13 May 2020

Abstract Overfitting effect of artificial neural network (ANN) based nonlinear equalizer (NLE) leads to a trap of bit error ratio (BER) overestimation in optical fiber communication system, especially when the performance is evaluated by the commonly-used pseudo-random binary sequence (PRBS). First, we mathematically investigate the PRBS generation and Gray code mapping rules, in comparison with the use of Mersenne Twister random sequence (MTRS). Under the condition of a symbol erasure channel, we identify that ANN can recognize both the PRBS generation and symbol mapping rules, by increasing the weights of NLE at specific positions, whereas the MTRS is currently safe owing to the limited input length of current ANN based NLE. Then, we design four channel models of fiber optical transmission to experimentally examine various impairments on the evolution of overfitting effect. When both the additive white Gaussian noise (AWGN) channel and the bandwidth limited channel are considered, the mitigation of overfitting becomes possible by the use of pruned PRBS (P-PRBS) training set with removing the generation and mapping rules determined input symbols. However, as for both the chromatic dispersion (CD) uncompensated channel and the CD managed channel, the overfitting effect becomes serious, because both CD and fiber nonlinearity induced inter-symbol interference (ISI) is beneficial for ANN to identify the PRBS symbol rules. Finally, possible solutions to mitigate the overfitting effect are summarized.

Keywords artificial neural network, nonlinear equalizer, pseudo-random binary sequence, overfitting

Citation Yang Z, Gao F, Fu S N, et al. Overfitting effect of artificial neural network based nonlinear equalizer: from mathematical origin to transmission evolution. Sci China Inf Sci, 2020, 63(6): 160305, https://doi.org/10.1007/s11432-020-2873-x

1 Introduction

Owing to the improvement of machine learning algorithm and hardware computing capability, artificial neural network (ANN), as an effective tool to model and predict the nonlinear relationship between input features and output responses, is widely adopted in many areas such as image classification [1], speech recognition [2] and language translation [3]. Particularly in fiber optical communication systems, it shows a powerful strength in channel equalization [4–11], modulation format recognition [12], optical performance monitoring [13] and fault diagnosis [14]. The ANN based nonlinear equalizer (NLE) was first adopted in wireless communication to combat the nonlinear impairments [15, 16]. Then ANN is proved capable to learn the fiber optical transmission channel. When ANN based NLE is employed in coherent optical orthogonal frequency-division multiplexing (CO-OFDM) transmission, 2 dB Q factor

[*] Corresponding author (email: songnian@mail.hust.edu.cn)

enhancement is experimentally reported than the use of traditional Volterra filter equalizer (VFE) for the 40 Gb/s 16QAM CO-OFDM transmission over 2000 km standard single mode fiber (SSMF) [4, 5]. As for the intensity modulation direct detection (IM-DD) transmission, the ANN based NLE realizes more than 2 dB receiver sensitivity improvement than VFE for the 4 × 50 Gb/s 4-levels pulse amplitude modulation (PAM-4) transmission over 80 km SSMF [6]. Meanwhile, several kinds of ANN based NLEs with stronger fitting ability were proposed to improve the NLE performance, including Radial basis function neural network (RBFNN) [7], convolutional neural network (CNN) [8] and recurrent neural network (RNN) [9]. Recently, ANN enabled end-to-end channel modeling was proposed [10,11], where the ANN based NLEs are implemented to recognize the overall fiber optical communication system including the transmitter, the transmission channel, and the receiver. All those ANN based NLEs can achieve significant improvements of bit error ratio (BER) than that of traditional NLEs. Unfortunately, such BER improvements may be overestimated, when the commonly-used pseudo-random binary sequences (PRBS) are applied [17,18]. Besides the transmission channel model, the ANN prefers learning the PRBS generation rules, leading to a BER trap that when the truly random network traffic is applied, the BER performance is severely degraded [19, 20]. Such BER performance trap arising in the ANN defined as overfitting effect becomes an obstacle for the cooperation of ANN based NLE and the PRBS. However, the origin and evolution of overfitting effect are still ambiguous, especially for the PRBS generation rules and the bit-to-symbol mapping rules. In particular the evolution of overfitting effect over the fiber optical transmission has not been investigated previously. For the sustainable application of PRBS in fiber optical communication system it is critical to investigate the overfitting effect from mathematical origin to transmission evolution.

In current submission, the origin of overfitting effect is firstly mathematically investigated by comparing the PRBS with Mersenne Twister random sequence (MTRS) [21–23]. With the aid of an symbol erasure channel, the L-∞ weight distributions of well-trained ANN are obtained to illustrate the relationship between the symbol rules and the ANN training process. We identify that ANN can learn the PRBS generation and mapping rules by increasing the weights of ANN based NLE at specific positions. Next, in order to experimentally investigate the evolution of overfitting effect we design an additive white Gaussian noise (AWGN) channel and carry out fiber optical transmissions under conditions of three typical channels. The pruned PRBS (P-PRBS) training set is helpful to mitigate the overfitting effect for both AWGN channel and bandwidth-limited channel. However, the overfitting effect gets worsened for both the chromatic dispersion (CD) uncompensated channel and the CD managed channel, owing to the occurrence of inter-symbol interference (ISI). Finally, possible suggestions to mitigate the overfitting effect are presented.

2 Mathematical origin

2.1 PRBS symbol rules

As standardized by ITU-T [17], PRBS is generated by a linear shift register with specific initialization and feedback. Except for the initialized bits from 1 to $N-1$, each bit is generated after the exclusive OR (XOR) operation through previous two bits. Assuming that the number of linear shift register is N, the period of PRBS is 2^N, and such sequence is called as PRBS-N sequence. For the ease of discussion, we take the PRBS-20 sequence as an example, the recursive relation is expressed as

$$x(n) = \begin{cases} 1, & 1 \leqslant n \leqslant 20, \\ x(n-17) \oplus x(n-20), & n \geqslant 21, \end{cases} \tag{1}$$

where $x(n)$ refers to the n-th bit of the sequence. Owing to such recursive property, there exists a correlation among specific bits of PRBS-20 sequence. By defining the distance as the length spacing between the rule-related bits and current bit arising in the PRBS-20, we can analyze the specific rule within a limited distance, which may affect the length of input vector for the ANN based NLE. In the typical distance of 20 between the related bit and current bit, typical bit generation rules to determine current bit can be shown as

$$x(n) = x(n-20) \oplus x(n-17) = x(n-3) \oplus x(n+17) = x(n+3) \oplus x(n+20). \tag{2}$$

With the growing distance, there are more couples of rule-related bits with single XOR operation as

$$x(n) = x(n-40) \oplus x(n-34) = x(n-6) \oplus x(n+34) = x(n+6) \oplus x(n+40) = \cdots . \tag{3}$$

Meanwhile, multiple XOR operations can be considered. For example,

$$x(n-3) = x(n-23) \oplus x(n-20), \tag{4}$$

$$x(n+17) = x(n+20) \oplus x(n+37). \tag{5}$$

Thus, current bit can be determined by

$$\begin{aligned} x(n) &= [x(n-23) \oplus x(n-3)] \oplus x(n-17) \\ &= [x(n-23) \oplus x(n-3)] \oplus [x(n+20) \oplus x(n+37)]. \end{aligned} \tag{6}$$

Generally, optical signals are transmitted with symbols during the fiber optical transmission, where all PRBS bits are coded at the transmitter side. For the on-off keying (OOK) symbols, the bit-to-symbol mapping rule is

$$X(n) = x(n), \tag{7}$$

where $X(n)$ refers to the n-th OOK symbol. As for the mapping rules for higher-order modulation formats, Gray codes are preferred than the binary counting natural codes. For Gray code rules, adjacent code words differ only at one-bit position, and a slight bit displacement to be coded may give only a small encoding variation [24]. Taking the typical PAM-4 symbols as an example, the mapping rules between Gray coded PAM-4 symbols and corresponding bits are expressed as

$$Y(n) = 2 \times x(2n-1) + x(2n-1) \oplus x(2n), \tag{8}$$

where $Y(n)$ refers to the n-th PAM-4 symbol. And the demapping rules are

$$x(2n-1) = \mathrm{floor}(Y(n)/2), \tag{9}$$

$$x(2n) = (Y(n) - 3 \times \mathrm{floor}(Y(n)/2))/(1 - 2 \times \mathrm{floor}(Y(n)/2)). \tag{10}$$

After substituting (8) and (10) into (2), we can get the Gray coded PRBS PAM-4 mapping rules as

$$\begin{aligned} Y(n) &= \phi[Y(n-10), Y(n-9), Y(n-8)] \\ &= \phi[Y(n-2), Y(n-1), Y(n+8), Y(n+9)] \\ &= \phi[Y(n+1), Y(n+2), Y(x+10)], \end{aligned} \tag{11}$$

where for the ease of expression, we take the operation $\phi[\cdot]$ to denote the combination of Gray code mapping functions, demapping functions and XOR operations of PRBS generation rules. From (1) to (11), we can conclude that current PRBS bit or symbol can be predicted from the bits or symbols with a certain distance. For PRBS OOK symbols, the minimum distance is 17, while for PAM-4 symbols it is just 9.

2.2 MTRS symbol rules

MTRS is recommended as an alternative [21], especially for the training stage of ANN based NLE [22]. MTRS is generated according to the Mersenne Twister algorithm, where a twisted linear feedback shift register of rational normal form is used with state bit reflection and tempering [23]. The analytical expression of the MTRS generation rules can be derived as

$$\begin{cases} z(i) = \{p \oplus [(p \ll t)\ \ \&\ \ c]\} \oplus \{\{p \oplus [(p \ll t)\ \ \&\ \ c]\} \gg l\}, \\ p = [q \oplus (q \gg u)] \oplus \{\{[q \oplus (q \gg u)] \ll s\}\ \ \&\ \ b\}, \\ q = z((i+m) \bmod n) \oplus \{[(z(i)\ \ \&\ \ um)|(y((i+1) \bmod n)\ \ \&\ \ lm)] \gg 1\}, \end{cases} \tag{12}$$

where $z(i)$ refers to the i-th MTRS number with a bit length of 32 or 64, the initialization stage and all parameters including t, c, l, u, s, b, m, n, um and lm were defined in [23]. Taking the commonly used MT-19932 as an example, the initialization stage of MTRS requires a seed to generate 623 numbers, each number has a bit length of 32, leading to 19936 initialized bits. Then, the recursion process of MT-19932 is based on the XOR operations among current number, the later 1-st number and 397-th number, together with the shift operations and other logical operations, in order to realize the twist algorithm. The distance between the feedback bits and current bit is about 12704, which is much longer than that of PRBS. The period of MT-19932 is $2^{19937}-1$, while PRBS-20 is $2^{20}-1$. The long distance, large period and complicated calculation rules make it extremely challenging for ANN to identify the MTRS symbol rules.

2.3 Learning process under a symbol erasure channel

To verify whether ANN is able to learn the symbol rules, including bit generation and mapping rules, we establish a symbol erasure channel to force the ANN to predict current symbol from adjacent symbols. The simulation setup is shown in Figure 1. We set colorful blocks to distinguish different symbols and parts of the ANN. Firstly, a PRBS-20 bit sequence and a MT-19937 bit sequence with the same length of 2^{19} are generated, and mapped into two OOK symbol sequences and two PAM-4 symbol sequences, respectively. Then an ANN with two hidden layers is utilized to learn the symbol rules. The input vector of ANN has a length of $2 \times k$, from the previous k symbols to the later k symbols relative to current symbol which is the same as traditional NLEs. Current symbol is estimated at the ANN output, so that it is null at the input vector. Two hidden layers have 41 and 21 neurons, respectively, while the activation functions are all set as ReLU. At the output layer, the outputs after fully connected operations are transformed into classification probabilities of 2 or 4 categories by the softmax function. The ANN is trained to optimize the cross-entropy cost function with Adam method. During the training process, the batch size is 200 and the number of total iterations is 200000. For each symbol sequence, 50% is used as the training set, while the other 50% is used as the testing set. All parameters are optimized in order to avoid the underfitting effect and decrease the complexity simultaneously. The recovered symbols are decoded into bits for the BER counting. If current symbol cannot be learned from adjacent symbols, the output bit may be randomly recovered as either 0 or 1, leading to a BER of 0.5. Otherwise, the BER will be significantly lower than 0.5, indicating that ANN can understand the symbol generation and mapping rules.

The calculated BER results are shown in Figure 2. When the length of input OOK symbols is less than 34 ($k < 17$) the BER of PRBS symbols is 0.5, as shown in Figure 2(a). However, once $k \geqslant 17$ is satisfied, the BER is almost 0, indicating that the ANN can learn and recover current symbol accurately through previous 17 symbols and later 17 symbols or more. As for the PAM-4 symbols in Figure 2(b), the ANN can learn and recover current PAM-4 symbol accurately through previous 9 symbols and later 9 symbols or more. An interesting phenomenon is a BER of 0.25 when the length of input symbols is 16 ($k = 8$). Considering the two-bit mapping rules, the reason is that only one bit of current PAM-4 symbol is understood, when previous 8 symbols and later 8 symbols are introduced. According to (2) and (8), PRBS-20 PAM-4 symbol rules are shown in Figure 2(c) where the signs of '+' and '−' separately represent the later and the previous position relative to current symbol. It is obvious that with the input length of 16, the previous bit of current symbol can be predicted, whose information is involved in the '−2' symbol and the '+8' symbol. The later bit is determined by the '−1' symbol and the '+9' symbol. As for the MT-19937 sequence, the cross-entropy cost function hardly decreases, and the ANN fails to

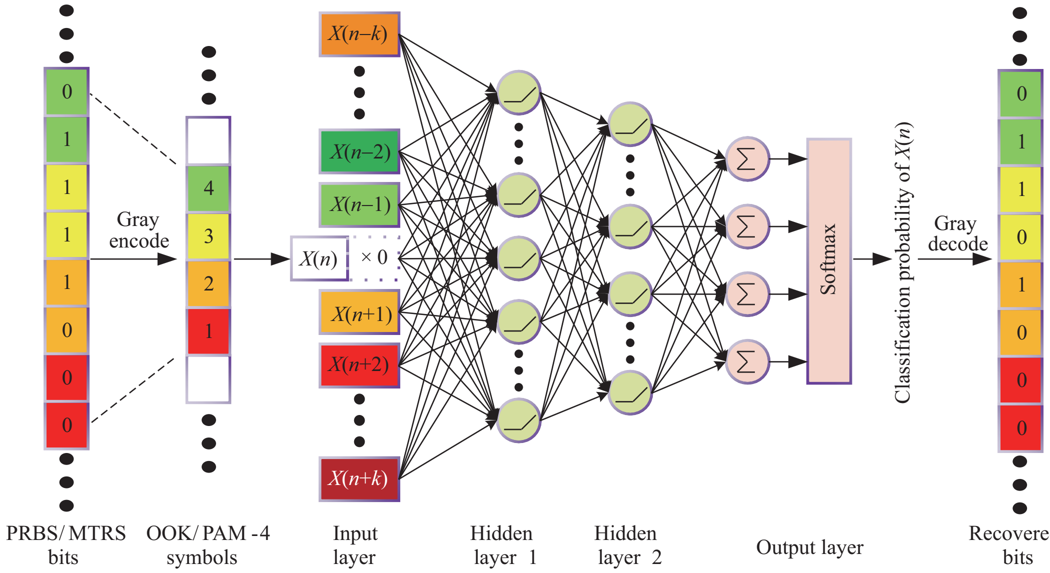

Figure 1 (Color online) ANN learning process under the symbol erasure channel.

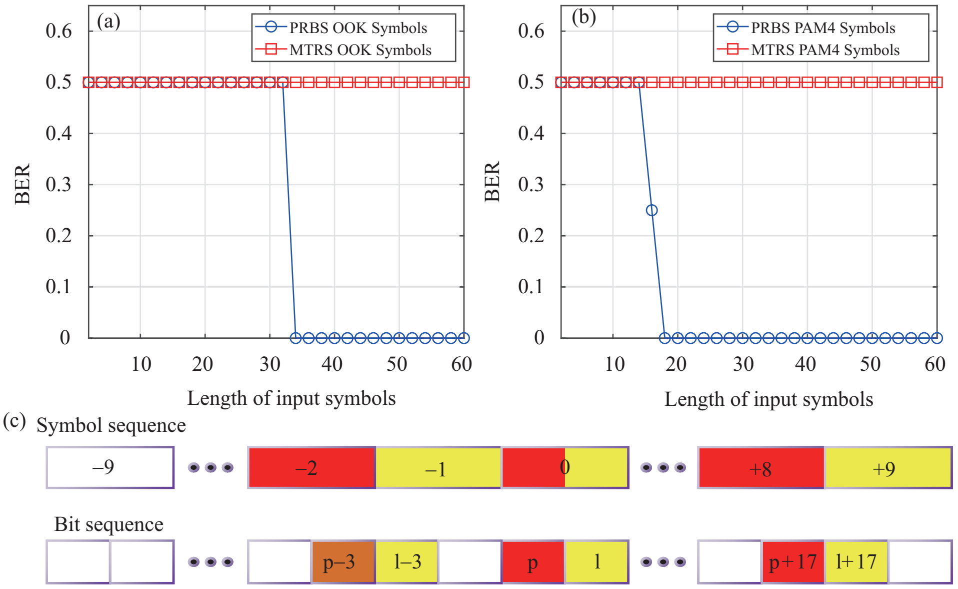

Figure 2 (Color online) BER results under the symbol erasure channel with (a) OOK symbols, (b) PAM-4 symbols and (c) PRBS-20 PAM-4 symbol generation rules.

converge, and the BERs of both OOK and PAM-4 symbols are always 0.5, indicating that input symbols are inadequate to predict the current MTRS symbol.

Next, in order to figure out how the ANN learns the PRBS symbol rules, we statistically investigate the weight distributions from input layer to the first hidden layer. We denote w_{ij} as the weight of i-th input neuron to the j-th hidden neuron, and define the L-∞ weight W_i as

$$W_i = \max\{|w_{ij}|, j = 1, 2, \dots, n_1\}, \quad i = -k, -k+1, \dots, -1, 1, \dots, k-1, k, \tag{13}$$

where $n_1 = 41$ is the number of neurons in the first hidden layer. Since w_{ij} represents the influence of the i-th input symbol on the j-th hidden neuron, the larger the $|w_{ij}|$ is, the more important the i-th input symbol is for the j-th hidden neuron. Moreover, since W_i is the L-∞ norm of $|w_{ij}|$, W_i represents the influence of i-th input symbol on the hidden layers and the output. We set the length of input symbols as 60 ($k = 30$) and train the ANN for 100 times independently and repeatedly, for the purpose of avoiding the random weight distribution. The results of L-∞ weight distributions are presented in

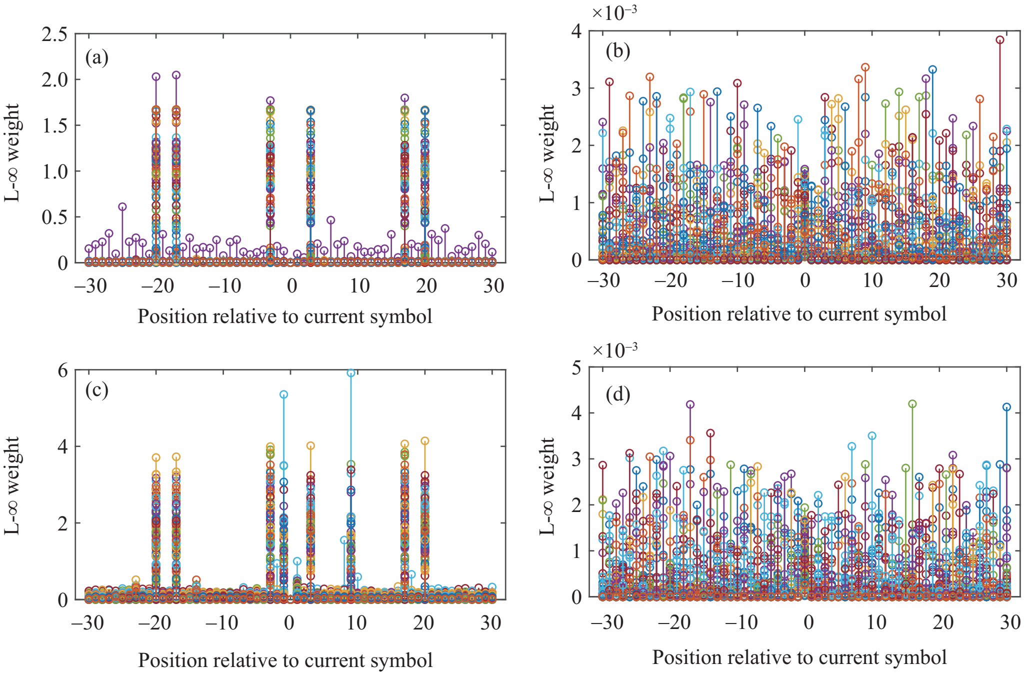

Figure 3 (Color online) L-∞ weight distributions for 100 independent trainings for (a) PRBS OOK symbols, (b) MTRS OOK symbols, (c) PRBS PAM-4 symbols, and (d) MTRS PAM-4 symbols.

Figure 3 with each color denoting one training result. For the PRBS symbols in Figures 3(a) and (c), the weights are significantly enhanced at specific positions, while at other positions the weights are nearly zero. The consistency between such certain positions and (1) to (11) confirms that ANN can learn the PRBS symbol rules by increasing the L-∞ weights at these rules determined positions. As for MTRS symbols in Figures 3(b) and (d), since the symbol rules cannot be recognized through the limited input length, the training losses barely go down, and the final weights remain the same distribution as the random initialization.

3 Transmission evolution

3.1 AWGN channel

To investigate the evolution of overfitting effect under various fiber optical transmission channels, we intend to analyze the L-∞ weight distributions of ANN based NLE. Owing to the two-bit mapping rule, there exists more rule-related symbols for PAM-4 than OOK when both the PAM-4 and OOK input vectors have the same symbol length. Thus, we must take the combination of two bits into account. For the ease of intuitively observing the L-∞ weight distributions, we choose two OOK symbol sequences with the same length of 2^{18} to be transmitted over various channels as shown in Figure 4(a). For each sequence, 5/8 of total symbol sequence is chosen as the training set, and the rest is used as the testing set including both PRBS and MTRS symbols with a ratio of 50%:50%. As a result, we can fairly evaluate the overfitting effect. Firstly, we use an additive white Gaussian noise (AWGN) channel as a reference, as shown in Figure 4(b). The used ANN keeps the same as the ANN in Figure 1, except that the current symbol is remained at the input vector of the ANN, leading to an input length of $2 \times k + 1$ instead of $2 \times k$.

The BER results are presented in Figure 5(a), where the direct decision results are presented as a reference. Three BER curves are overlapped when the length of input symbols is less than 33 ($k \leqslant 16$). Once the length is equal to or larger than 35 ($k \geqslant 17$), by using the PRBS training set, a BER gap occurs between the curve of PRBS testing set and the curve of MTRS testing set. The reason is that once the

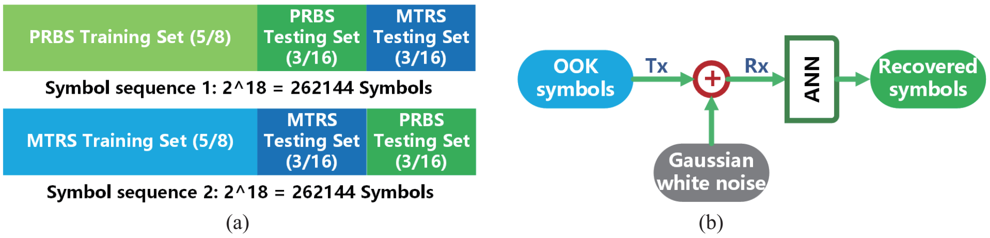

Figure 4 (Color online) (a) Structure of OOK symbol sequences to be transmitted; (b) AWGN channel model.

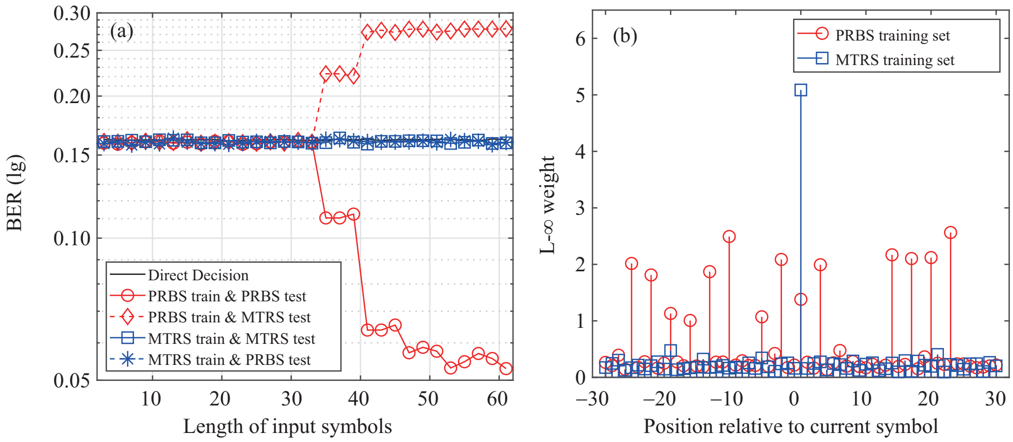

Figure 5 (Color online) (a) BER results of the AWGN channel; (b) L-∞ weight distributions of the ANN with the input length of 61.

PRBS symbol rules are learned by the ANN, the PRBS testing set can be predicted precisely but the MTRS testing set leads to a BER degradation. Therefore, the BER gap between two different testing sets after the application of PRBS training set represents the extent of overfitting effect. The BER gap becomes wider when the input length increases to 41 or more ($k \geqslant 20$), indicating that the longer input length is helpful for the ANN to identify the PRBS symbol rules.

However, there occurs no BER gap, when the MTRS training set is applied. In particular, its BER curve keeps the same as the direct decision result, indicating that ANN is unable to recognize the MTRS symbol rules. Thus, BER results based on MTRS training set can be taken as a benchmark. To illustrate the overfitting effect from the perspective of ANN training, we calculate the L-∞ weight distributions of the ANN with an input length of 61, as shown in Figure 5(b). For the MTRS training set, only the weight of current input symbol is much higher. However, several weights are pretty higher for the PRBS training set, and the positions are consistent with (2) and (6). It is obvious that the ANN can recognize the PRBS symbol rules by increasing the weights of input symbols at specific positions.

To further address the overfitting effect, we utilize P-PRBS training set where some symbols at the input vectors are removed or set as null to mitigate the overfitting effect. Generally, P-PRBS training set is still a PRBS training set with the symbol rules broken. The removed symbols are defined as the Ruleset, which is a complete non-redundant set. The completeness means as long as the symbols in the Ruleset are all removed in the training set, corresponding symbol rules cannot be learned at all. The non-redundancy means that partially removing Rulesets cannot break the overfitting effect, if only one symbol in the Ruleset is kept, the symbol rules can be partially or entirely recognized. Since the removed symbols may in turn degrade the capability of channel equalization, we prefer the searching method with less equalization penalty. The detailed Rulesets for PRBS-20 OOK are listed in Table 1. In particular, the Rulesets for PAM-4 symbols can be derived according to the corresponding mapping rules. Then the BER results based on the P-PRBS training set are shown in Figure 6, where the BER gap vanishes, indicating the use of P-PRBS training set is valid to avoid learning the PRBS rules and mitigate the overfitting effect.

Table 1 The Ruleset of PRBS-20 OOK symbols

k [a]	Ruleset	k	Ruleset	k	Ruleset
0–16	Null	38–39	$\{\text{ans}, X(n+38)\}$	53	$\{\text{ans}, X(n53)\}$
17–19	$\{X(n+17)\}$	40	$\{\text{ans}, X(n-40), X(n+40)\}$	54–55	$\{\text{ans}, X(n+54)\}$
20–22	$\{\text{ans}$[b]$, X(n-20), X(n+20)\}$	41–43	$\{\text{ans}, X(n41)\}$	56	$\{\text{ans}, X(n56)\}$
23–25	$\{\text{ans}, X(n23)\}$	44–45	$\{\text{ans}, X(n44)\}$	57	$\{\text{ans}, X(n+57)\}$
26–28	$\{\text{ans}, X(n26)\}$	46	$\{\text{ans}, X(n46)\}$	58	$\{\text{ans}, X(n58)\}$
29–31	$\{\text{ans}, X(n29)\}$	47–49	$\{\text{ans}, X(n47)\}$	59	$\{\text{ans}, X(n59)\}$
32–33	$\{\text{ans}, X(n32)\}$	50	$\{\text{ans}, X(n50)\}$	60	$\{\text{ans}, X(n-60), X(n+60)\}$
34	$\{\text{ans}, X(n+34)\}$	51	$\{\text{ans}, X(n+51)\}$	...	...
35–37	$\{\text{ans}, X(n35)\}$	52	$\{\text{ans}, X(n52)\}$		

a) The k means the position relative to current symbol at the input vector, the input length is $2\times k+1$.
b) The ans means the Ruleset for current k includes the Ruleset for the smaller k above.

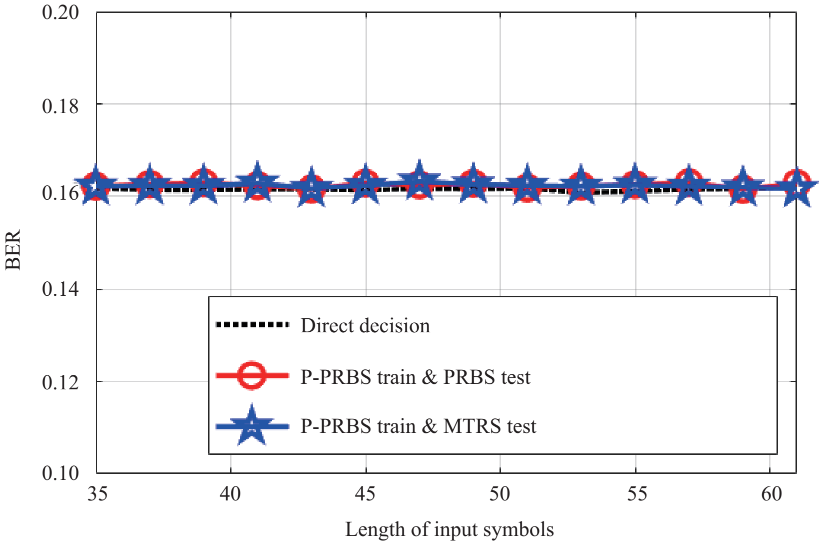

Figure 6 (Color online) BER results under the AWGN channel with the help of P-PRBS training set.

3.2 Bandwidth-limited channel

Next, we carry out a typical IM-DD transmission to experimentally investigate the evolution of the overfitting effect. The experimental setup is shown in Figure 7. At the transmitter side, the OOK symbol sequences in Figure 4(b) are preprocessed and then introduced into an arbitrary waveform generator (AWG) operated at a sampling rate of 92 GSa/s. The electronic signals amplified by an electrical amplifier (EA) drive the directly modulated laser (DML) with 3 dB bandwidth of 18 GHz at the operation wavelength of 1550 nm to generate optical signals with the power of 7.6 dBm. Three different channels are designed to distinguish various impairments, including the back to back (B2B) channel 20 km SSMF without CD compensation, and the 100 km SSMF with a dispersion compensation module (DCM) for full CD pre-compensation. At the receiver side, a variable optical attenuator (VOA) is employed to adjust the received optical power (ROP) at the photodetector (PD) with 3 dB bandwidth of 20 GHz. The signals are sampled by a real-time oscilloscope (Tektronix DPO73304D) operated at 80 GSa/s and processed offline. The input length is 41 ($k=20$) so that we can easily analyze the L-∞ weight distributions when the overfitting effect occurs.

Firstly the major impairments of B2B channel consists of the electical noise and the constraint of limited bandwidth from both DML and PD. The nonlinearity of optoelectronic devices is ignored. We set the OOK baudrate as 25 GB and adjust the ROP from -1 to -7 dBm to highlight the influence of electical noise, the BER results are shown in Figure 8(a). When the ROP is more than -3 dBm, the quality of received signals is pretty good so that the ANN-based NLE is unnecessary, leading to a

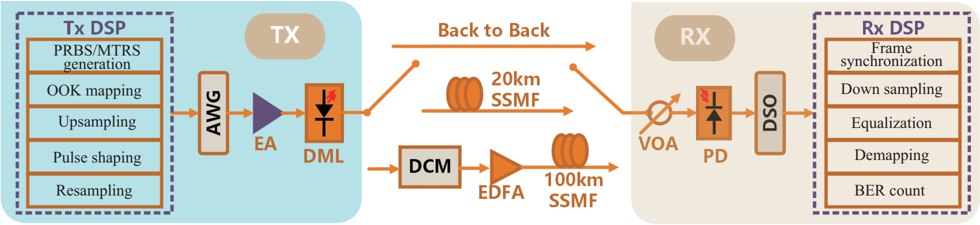

Figure 7 (Color online) Experimental setup of typical IM-DD transmission system.

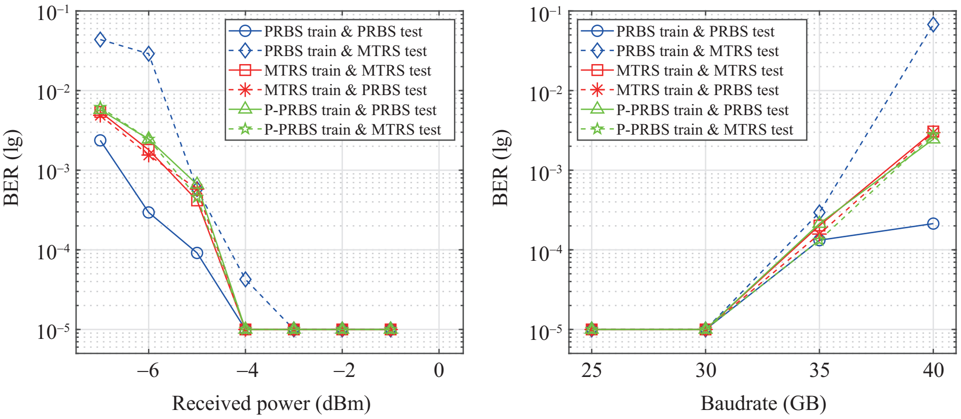

Figure 8 (Color online) BER results of B2B transmission channel under conditions of (a) different ROPs with the baudrate of 25 GB and (b) different baudrates with the ROP of −3 dBm.

convergence of all BER curves at 1E−5. Meanwhile, the finite symbols in the testing set can result in the minimum BER of 1E−5. With the further reduction of ROP, all BER performances become worse leading to the enhancement of BER gap between the PRBS and MTRS testing sets after the application of PRBS training set. After the symbols in the Ruleset including '−20', '+17' and '+20' are removed to generate the P-PRBS training set, the BER gap disappears and the results are almost the same as that of the MTRS training set. Then we set the ROP as −3 dBm and adjust the baudrate from 25 to 40 GB to aggravate the impairment of bandwidth limitation, the BER results are shown in Figure 8(b). All BER curves are convergent to 1E−5 when the baudrate is less than 30 GB, because the ANN-based NLE is unnecessary to compensate the impairment of bandwith limitation. With the growing baudrate, the BER gap occurs for the PRBS training set while it disappears by using the P-PRBS training set, indicating that the overfitting effect arising in the B2B channel is similar to that of the AWGN channel.

3.3 CD uncompensated channel

When 20 km SSMF is introduced, the major impairment becomes the CD effect. We increase the baudrate from 25 to 40 GB under the condition of −1 dBm ROP. Figure 9 indicates the distinctive BER performances of different testing sequences with the same training sequences owing to the overfitting effect. As shown in Figure 9(a), the CD brings obvious penalty when the baudrate is higher than 35 GB. Thus under the baudrate of 40 GB, we vary the ROP from 1 to −5 dBm, and the BER results are summarined in Figure 9(b). The curves of PRBS training set are almost the same under conditions of different received powers, indicating the ANN can always learn the PRBS symbol rules within the range of received powers. For the results of PRBS training set, the BER gap occurs. However, unlike the AWGN channel and the B2B channel, the BER gap narrows a little but still occurs for the case of P-PRBS training set, indicating that removing the Ruleset can only mitigate the overfitting effect partially. To clarify such issue, we calculate the L-∞ weight distributions, when the CD uncompensated channel is operated at 40 GB and −1 dBm ROP, as shown in Figure 10. For the MTRS training set, the input symbols close to current symbol have higher L-∞ weights, indicating that the symbols at the central

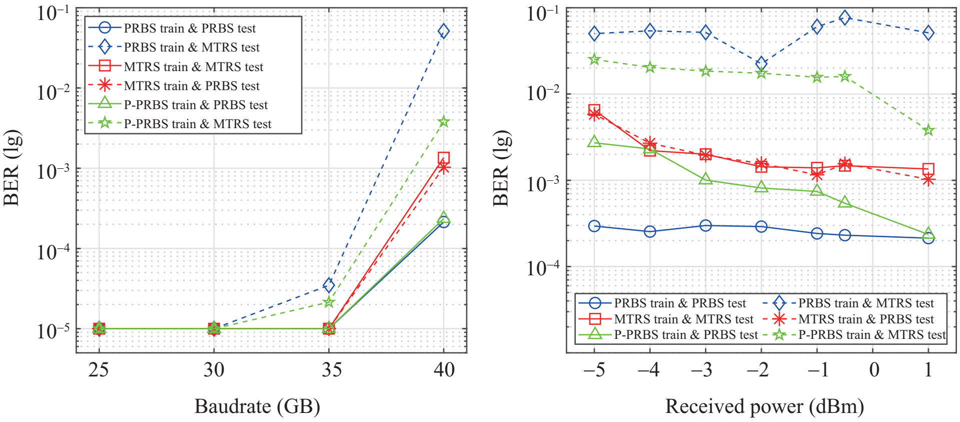

Figure 9 (Color online) BER results of the 20 km SSMF channel under the conditions of (a) different baudrates with the ROP of −1 dBm and (b) different ROPs with the baudrate of 40 GB.

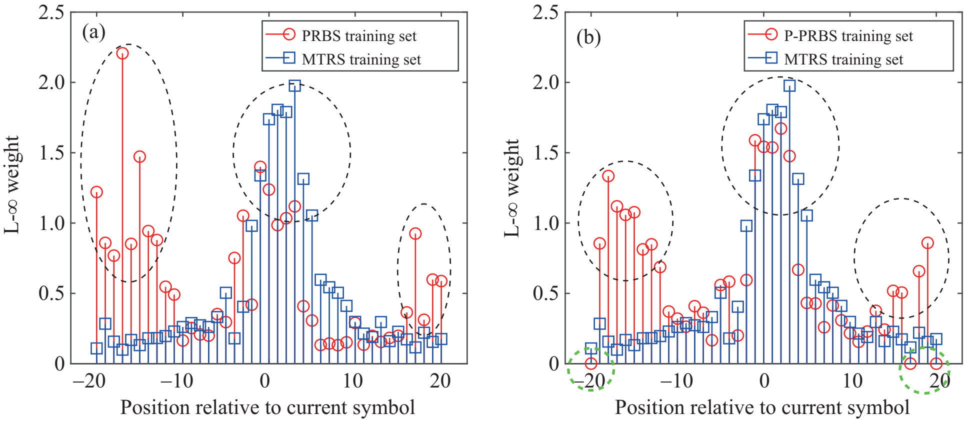

Figure 10 (Color online) L-∞ weight distributions of ANNs for the 20 km SSMF channel under condition of 40 GB and −1 dBm ROP. (a) Using the PRBS training set, and (b) using the P-PRBS training set.

position play more important role for the channel equalization. Thus the ANN recognizes the channel model and acts as the NLE by increasing weights at the central position. However, for the PRBS training set in Figure 10(a), besides the central symbols, the weights at two ends are pretty high, which agrees well with the PRBS symbol rules instead of the channel model. Therefore, the ANN recognizes both channel model and PRBS symbol rules. Even for the P-PRBS training set in Figure 10(b) the symbols in the Ruleset are all zero, but the weights around the Ruleset rise up significantly. It can be explained as that the CD induced ISI enables the ANN to learn the PRBS symbol rules from an extended range of input symbols, leading to the enhancement of oveffiting effect. Another possible explanation is that the ANN can predict the neighboring symbols such as $X(n-1)$ and $X(n+1)$ through the symbols around the Ruleset in the input vector, which is helpful to mitigate the ISI and recover current symbol leading to the possible occurrence of overfitting effect. Although both explanations are put forward by different views, they verify that the CD induced ISI will enhance the overfitting effect. In such case with obovious ISI, an extended Ruleset is requisite to mitigate the overfitting effect.

3.4 CD managed channel

When 100 km CD compensated channel is implemented with a baudrate of 25 GB, the major impairment becomes the fiber nonlinearity. As shown in Figure 7, the CD of 100 km SSMF is pre-compensated by a DCM at the transmitter side. Then the launched power into the SSMF is increased from 13 to 18 dBm by employing an erbium-doped fiber amplifier (EDFA), strenghthening the fiber nonlinearity. The BER results are shown in Figure 11. For the PRBS training set, the BER gap increases with the enhancement

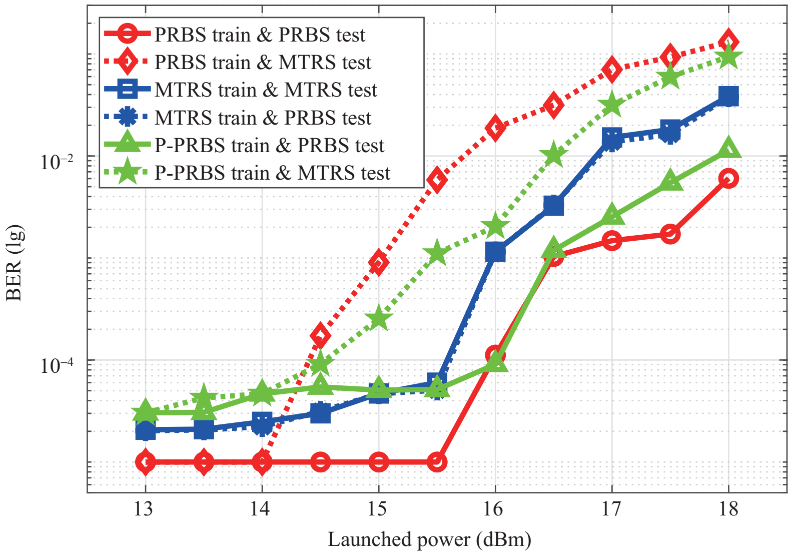

Figure 11 (Color online) BER results of 100 km SSMF channel with the CD pre-compensation.

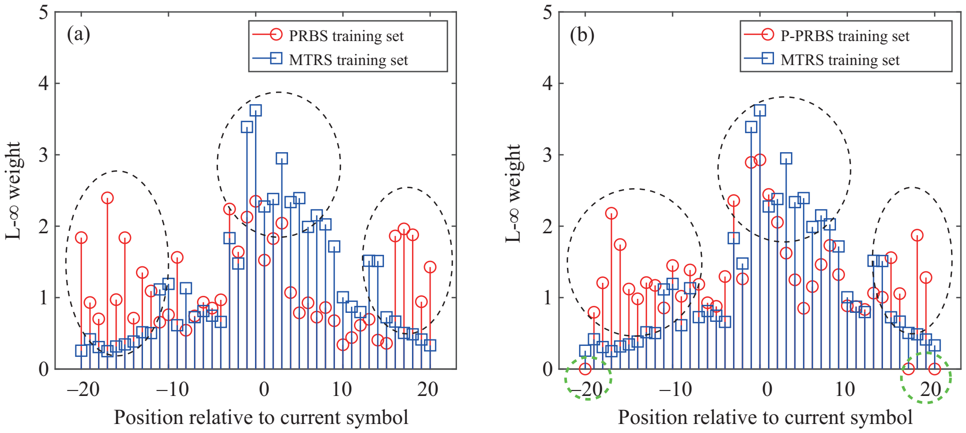

Figure 12 (Color online) L-∞ weight distributions of ANNs for the 100 km SSMF channel at the 18 dBm launch power. (a) Using the PRBS training set, and (b) using the P-PRBS training set.

of the fiber nonlinearity. For the P-PRBS training set, BER gap only partially mitigates. Please note that the performances are almost the same at launched powers from 13 to 15 dBm, because the fiber nonlinearity is not strong enough to bring the transmission impairment. The L-∞ weight distributions at the launched power of 18 dBm is shown in Figure 12. For the P-PRBS training sets, the symbols around the Ruleset still have much higher weights than those of the MTRS training set, which is similar with that under the CD uncompensated channel. That is because self-phase modulation (SPM), the major fiber nonlinearity of single channel fiber optical transmission, induces the chirp which interacts with the disributed CD and finally leads to a pulse broadening [25]. Therefore, the ISI still occurs and consequently enhances the overfitting effect. Another interesting fact is that, by analyzing the L-∞ weight distributions under the MTRS training sets in Figures 10 and 12, the weights of later symbols relative to the current symbol are obvious higher than those of previous symbols, indicating that the channel model may be unsymmetric with respect to current symbol.

4 Conclusion and remarks

We investigate the overfitting effect from mathematical origin to transmission evolution. By comparing the PRBS with MTRS under the condition of a symbol erasure channel, we identify that ANN can learn the PRBS symbol generation and mapping rules by increasing the weights of ANN-based NLE at specific positions, whereas MTRS symbol rules cannot be recognized owing to the limited input length of current ANN-based NLE. Then three transmission channels are experimentally implemented and the BER performances are compared with that of the AWGN channel, for the ease of clarifying the evolution of overfitting effect. The P-PRBS training set is effective for the AWGN channel and the bandwidth limited channel. However, for the CD uncompensated channel and the CD managed channel, both CD and fiber nonlinearity induced ISI is beneficial for ANN to learn the PRBS symbol rules from the extended input symbols.

According to our investigation, we have four suggestions to mitigate the overfitting effect arising in the ANN based NLE. Firstly, instead of PRBS, the MTRS can be used for current ANN based NLE. Secondly, although removing the Ruleset to obtain the P-PRBS training set is theoretically valid, the Ruleset needs to be expanded a lot owing to the introduction of both CD and fiber nonlinearity. Thirdly, after identifying the distinctions of the L-∞ weight distribution under both the PRBS and the MTRS training sets, we can extract different features of PRBS symbol rules with respect to the transmission channel, in order to mitigate the overfitting effect. Finally, from the view of ANN implementation, we can optimize the ANN structure and parameters to keep the training and testing losses at the same level, especially for the joint use of PRBS training set and MTRS testing set.

Acknowledgements This work was supported by National Key R&D Program of China (Grant No. 2018YFB1801301) National Natural Science Foundation of China (Grant No. 61875061), and Key Project of R&D Program of Hubei Province (Grant No. 2018AAA041).

References

1 Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks. In: Proceedings of Advances in neural information processing systems (NIPS), 2012. 1097–1105

2 Hinton G, Deng L, Yu D, et al. Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. IEEE Signal Process Magaz, 2012, 28: 82–97

3 Sagiroglu S, Yavanoglu U, Guven E N. Web based machine learning for language identification and translation. In: Proceedings of the 6th International Conference on Machine Learning and Applications, Cincinnati, 2007. 280–285

4 Jarajreh M A, Giacoumidis E, Aldaya I, et al. Artificial neural network nonlinear equalizer for coherent optical OFDM. IEEE Photon Technol Lett, 2015, 27: 387–390

5 Giacoumidis E, Le S T, Ghanbarisabagh M, et al. Fiber nonlinearity-induced penalty reduction in CO-OFDM by ANN-based nonlinear equalization. Opt Lett, 2015, 40: 5113–5116

6 Luo M, Gao F, Li X, et al. Transmission of 4×50-Gb/s PAM-4 signal over 80-km single mode fiber using neural network. In: Proceedings of Optical Fiber Communication Conference, 2018. M2F.2

7 Yang Z, Gao F, Fu S, et al. Radial basis function neural network enabled C-band 4×50-Gb/s PAM-4 transmission over 80 km SSMF. Opt Lett, 2018, 43: 3542–3545

8 Chuang C, Liu L, Wei C, et al. Convolutional neural network based nonlinear classifier for 112-Gbps high speed optical link. In: Proceedings of Optical Fiber Communication Conference, 2018. W2A.43

9 Ye C, Zhang D, Hu X, et al. Recurrent neural network (RNN) based end-to-end nonlinear management for symmetrical 50 Gbps NRZ PON with 29 dB+ loss budget. In: Proceedings of European Conference on Optical Communication, 2018. 1–3

10 Karanov B, Chagnon M, Thouin F, et al. End-to-end deep learning of optical fiber communications. J Lightw Technol, 2018, 36: 4843–4855

11 Karanov B, Lavery B, Bayvel P, et al. End-to-end optimized transmission over dispersive intensity-modulated channels using bidirectional recurrent neural networks. Opt Express, 2019, 27: 19650–19663

12 Wang D, Zhang M, Li Z, et al. Modulation format recognition and OSNR estimation using CNN-based deep learning. IEEE Photon Technol Lett, 2017, 29: 1667–1670

13 Dong Z, Khan F N, Sui Q, et al. Optical performance monitoring: a review of current and future technologies. J Lightw Technol, 2016, 34: 525–543

14 Chen X, Li B, Shamsabardeh M, et al. On real-time and self-taught anomaly detection in optical networks using hybrid unsupervised/supervised learning. In: Proceedings of European Conference on Optical Communication, 2018. 1–3

15 Charalabopoulos G, Stavroulakis P, Aghvami A H. A frequency-domain neural network equalizer for OFDM. In: Proceedings of IEEE Global Telecommunications Conference, 2003. 571–575

16 Rajbhandari S, Ghassemlooy Z, Angelova M. Effective denoising and adaptive equalization of indoor optical wireless channel with artificial light using the discrete wavelet transform and artificial neural network. J Lightw Technol, 2009, 27: 4493–4500

17 ITU-T. Digital test patterns for performance measurements on digital transmission equipment. CCITT Recommendation O.150. https://www.itu.int/rec/T-REC-O.150-199210-S/en

18 IEEE Standards Association. IEEE Standard for Ethernet Amendment 10: Media Access Control Parameters, Physical Layers, and Management Parameters for 200 Gb/s and 400 Gb/s Operation. IEEE Std 802.3bs. https://standards.ieee.org/standard/802_3bs-2017.html

19 Eriksson T A, Bülow H, Leven A. Applying neural networks in optical communication systems: possible pitfalls. IEEE Photon Technol Lett, 2017, 29: 2091–2094

20 Shu L, Li J, Wan Z, et al. Overestimation trap of artificial neural network: learning the rule of PRBS. In: Proceedings of European Conference on Optical Communication, 2018. 1–3

21 Chuang C, Liu L, Wei C, et al. Study of training patterns for employing deep neural networks in optical communication systems. In: Proceedings of European Conference on Optical Communication, 2018. 1–3

22 Yi L, Liao T, Huang L, et al. Machine learning for 100 Gb/s/λ passive optical network. J Lightw Technol, 2019, 37: 1621–1630

23 Matsumoto M, Nishimura T. Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. ACM Trans Model Comput Simul, 1998, 8: 330

24 Doran R W. The Gray code. J Univ Comput Sci, 2007, 13: 1573–1597

25 Agrawal G P. Nonlinear Fiber Optics. 4th ed. San Diego: Academic Press, 2001