Received July 29, 2018, accepted August 28, 2018, date of current version September 21, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2868480

Jointly Optimized Extreme Learning Machine for Short-Term Prediction of Fading Channel

YONGBO SUI[1], WENXIN YU[2], AND QIWU LUO[1], (Member, IEEE)
[1]School of Electrical and Automation Engineering, Hefei University of Technology, Hefei 230009, China
[2]School of Electrical and Information Engineering, Hunan University, Changsha 410082, China

Corresponding author: Qiwu Luo (luoqiwu@hfut.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 51704089, in part by the Anhui Provincial Natural Science Foundation of China under Grant 1808085QF190, in part by the China Postdoctoral Science Foundation under Grant 2017M621996 and Grant 2017M622574, in part by the Fundamental Research Funds for the Central Universities of China under Grant JZ2018YYPY0296, and in part by the Ph.D. Special Research Fund of HFUT under Grant JZ2016HGBZ1030.

ABSTRACT Accurately evaluating channel state information (CSI) is an extremely important precondition for wireless communication to effectively obtain exact sending data. In order to fast obtain CSI, channel prediction tends to get its popularity in obtaining CSI as a result of the advantages of lightweight calculation and negligible feedback delay. In this paper, a jointly optimized extreme learning machine (JOELM) is proposed for the short-term prediction of fading channel. The JOELM scheme consists of two key steps: intelligent optimization and targeted repair. First, in order to gain high prediction accuracy, the firefly algorithm is imported to intelligently optimize the traditional extreme learning machine. Second, the Savitzky–Golay filter is innovatively adopted for reducing potential prediction errors. Extensive experiments about computational complexity, influences of repair coefficient and weight coefficient, contributing degree of two key steps, amplitude, transmission bit/symbol error rates, root-mean-square errors, and four typical statistical properties are given in final simulation section. The analyzed results indicate that the proposed JOELM can more accurately and efficiently deal with the short-term prediction of fading channel.

INDEX TERMS Fading channel, short-term prediction, channel state information (CSI), extreme learning machine (ELM).

I. INTRODUCTION

In wireless communication, fading is variation of a signal with various variables since barriers exist ubiquitously in transmitting path. This phenomenon of strength attenuation of wireless signal is called channel fading, which is divided into long-term (slow) fading and short-term (fast) fading [1]. The former mainly pertains to power average attenuation caused by environmental profile, while the latter is mainly caused by multipath reflections. Think of there were no noise, fast fluctuation of the receiving signal is inevitably due to the existence of time-variant multipath effects. Besides, different attenuations and the random nature of the signal potentially lead to inter-symbol interferences [2]. Short-term fading thus receive more attention for its great importance in obtaining channel state information (CSI), so as to evaluate the performance of wireless communication systems.

Traditionally, CSI in short-term fading can be estimated by adaptive estimation algorithms, and CSI also can be traced when system parameters of CSI vary [3]. However, they are also reported to be suffered with computational complexity and unavoidable feedback delay [4]. Short-term prediction of fading channel emerges as the times require. Due to acceptable complexity and lower feedback delay, it has been applied to analyze CSI in wireless communication. Some preliminary studies have been done to build short-term prediction methods of fading channel. Mainly based on the regressing calculation, Eyceoz *et al.* [5] proposed an autoregressive (AR) algorithm to describe (also predict) channel characteristics. Inspired by AR scheme, Sharma and Chandra built an improved second-order autoregressive (AR-2) to predict CSI [6]. However, the accumulated prediction errors in iteration process limit its applications to some extent [7]. Another technology roadmap of channel prediction is mainly based on nonlinear transformation. Support vector machine (SVM) was successfully applied on path loss prediction in airplane cabin scenarios. The simulation results indicate that higher prediction accuracy was achieved than those of conventional curve fitting methods [8]. Further, a hybrid channel prediction method based on SVM and recurrent least squares was proposed, then fading channel

samples in one-dimensional time domain could be mapped to high dimension by embedding phase space (EPS) theory [9]. Just one year later, Xiang *et al.* [10] found that random characteristic of fading channel was similar to chaos in some respects. Hence, a combinational short-term predictor of fading channel was established based on EPS and chaos theory. But it is difficult to determine the optimal dimension and specific delay for EPS, especially when encountering various complex scenarios. In recent report [4], echo state network (ESN) was first applied to short-term prediction in Ricean fading scenarios, performing better simulation results than those of conventional AR- or SVM-based methods. The aforementioned method needs to handle a common issue, that several key parameters (i.e., size of internal reservoir, sparse degree in internal reservoir, *et al.*) should be carefully determined by repeated training or only by empirics [11].

Compared with the conventional neural networks, extreme learning machine (ELM) exhibits faster learning speed and higher convergence precision, due to its special single hidden layer feed-forward structure [12]. Consequently, ELM is widely applied to solve regression and classification issues in various areas, such as electric power system [13], distribution grid [14] and visual image [15]. There are already some preliminary research reports in the field of wireless communication. ELM has been used for path loss analysis in complex fading environment [16] and outdoor propagation scenarios [17]. It is worth mentioning that, experimental results produced in [17] indicate ELM outperformed BPNN, Okumura–Hata, and COST-231 algorithms. Nevertheless, the room for ELM-based channel prediction is still large, further theoretical and empirical studies for wireless communication are required.

In this paper, a jointly optimized extreme learning machine (JOELM) is proposed for short-term prediction of fading channel. The main contributions are as follows.

1) A swarm intelligent optimization algorithm, firefly algorithm (FA) is imported to enhance the traditional ELM. Then the random generation procedure of weights and biases is replaced by parameter optimization procedure. Hence, more accurate and efficient prediction can be achieved.
2) The Savitzky-Golay (S-G) filter is innovatively adopted for reducing potential prediction errors. Predicted channel data are carefully estimated by S-G filter, the detected abnormal data are repaired in targeted manner.
3) Extensive evaluations (i.e., computational complexity, influences of coefficients, contributing degree of optimization step and repair step, amplitude, transmission bit/symbol error rates, root mean square errors (RMSEs) and four typical statistical properties) are carried out and discussed.

The structure of this paper is as follows. In section II, basic theories of Rayleigh channel are described briefly. Section III outlines JOELM from two perspectives, intelligent optimization and targeted repair. In section IV, simulations and discussions are presented in detail. Finally, section V concludes this paper.

II. CHANNEL MODEL

Using the classic Rayleigh channel as an example, a basic channel model is introduced briefly in this section.

Due to the multipath effect, the receiving signals are superposed by envelope signals with various delays, and local fast fading exists. Therefore, a channel is modeled as follows:

$$R(t) = h(t)S(t) + \lambda(t) \tag{1}$$

where $R(t)$, $h(t)$ and $S(t)$ are the receiving signal, the channel parameters and the transmitting signal, respectively. $\lambda(t)$ is added white Gaussian noise (AWGN). In the classic Rayleigh Jakes model, $h(t)$ is modeled for low-pass fading in a statistical process, which can be defined by

$$h(t) = \sum_{\varsigma=1}^{\Psi} \Gamma_\varsigma e^{j(2\pi f_\varsigma t + \vartheta_\varsigma)} \tag{2}$$

where ς is the transmitting path index number, Ψ is the total transmitting path number, Γ_ς is the amplitude of the channel and ϑ_ς is the phase angle. f_ς is the Doppler frequency shift, which is determined by the relative velocity V of the wireless signals' receiver with a transmitter, that is,

$$f_\varsigma = \frac{V}{C} f_d \cos\phi \tag{3}$$

where C is the transmitting speed of the electromagnetic wave, f_d is the maximum Doppler frequency shift. ϕ is the angle calculated by the straight-line distance D_s and vertical distance D_{in} from transmitter to receiver in Fig. 1.

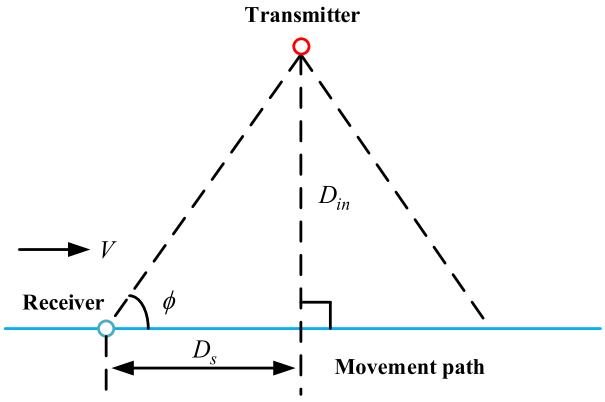


FIGURE 1. Figure for calculation of $cos\phi$.

For a Rayleigh channel, the probability density function (PDF) is

$$f(z) = \frac{z}{\varpi} e^{-\frac{z^2}{2\varpi^2}}, \quad z \geq 0 \tag{4}$$

and the cumulative distribution function (CDF) is

$$F(z) = 1 - e^{-\frac{z^2}{2\varpi^2}} \tag{5}$$

where ϖ^2 is the power of the scatter components in a Rayleigh channel.

III. THE PREDICTOR OF SHORT-TERM FADING CHANNEL BASED ON JOELM

In this section, the short-term predictor of fading channel, JOELM, is introduced in detail from two steps: intelligence optimization and targeted repair. First, the traditional ELM is optimized by the swarm intelligence algorithm FA to obtain the optimal weights and biases. Then, the predicted data of optimal ELM is estimated by the S-G filter and repaired by the repair strategy, which further improves the performance of the prediction mechanism and reduces prediction errors.

A. INTELLIGENT OPTIMIZATION

As an effective and simple machine learning tool, ELM exhibits excellent advantages over traditional neural networks. The ELM algorithm has some special features, such as a simple structure, high learning speed and satisfactory convergence precision. As shown in Fig. 2, the typical structure is a single hidden layer feed forward structure. Hence, output weights are obtained by random initial input weights and biases.

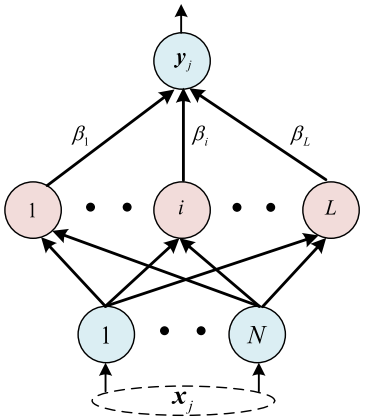


FIGURE 2. Typical structure of ELM.

Therefore, the corresponding expression is

$$\boldsymbol{H}\boldsymbol{\beta} = \boldsymbol{Y} \tag{6}$$

where $\boldsymbol{H}$, $\boldsymbol{\beta}$ and $\boldsymbol{Y}$ are given by

$$\boldsymbol{H} = \begin{bmatrix} g(\mu_1, \nu_1, x_1)g(\mu_2, \nu_2, x_1)\ldots g(\mu_L, \nu_L, x_1) \\ \vdots \\ g(\mu_1, \nu_1, x_o)g(\mu_2, \nu_2, x_o)\ldots g(\mu_L, \nu_L, x_o) \end{bmatrix}_{o\times L}$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_L \end{bmatrix}_{L\times 1}, \quad \boldsymbol{Y} = \begin{bmatrix} y_1 \\ \vdots \\ y_o \end{bmatrix}_{o\times 1}$$

where $(\boldsymbol{X}, \boldsymbol{Y})$ are the N input data and the output data of ELM with $\boldsymbol{X} \in \boldsymbol{R}^{o\times N}$ and $\boldsymbol{Y} \in \boldsymbol{R}^{o\times 1}$. $\boldsymbol{H}$, $\boldsymbol{\beta}$ and $\boldsymbol{Y}$ are the output of the hidden layer, the output weights and the output of ELM, respectively. o and L are the dimension of input data and the neurons number in hidden layer. $\boldsymbol{\mu} \in \boldsymbol{R}^{N\times L}$, $\boldsymbol{\nu} \in \boldsymbol{R}^{o\times L}$ and $\boldsymbol{\beta} \in \boldsymbol{R}^{L\times 1}$ are input weights, biases and output weights.

It is noted that $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ are commonly assigned randomly, and once they are chosen, output weights are confirmed by (6).

However, the random method of generating input weights and biases limits the accuracy of prediction and regression fitting in traditional ELM. To address this drawback, the FA is motivated to be imported for handling the parameter optimization of its weights and biases. It is noted that the intelligent optimization can be implemented offline to reduce computation complexity and can be applied in various scenarios with new optimization processes.

Inspired by swarming of fireflies in summer, FA arrived in 2008 [18], and then becomes a typical biological swarm intelligent algorithm. It has continuously drawn increasing attention in various fields, such as multi-objective optimization [19], image processing [20] and optimization control [21].

The algorithm is presented by the following expressions:

$$\varphi_i(t+1) = \varphi_i(t) + \tau(\varphi_j(t) - \varphi_i(t)) + \alpha(rand - \frac{1}{2}) \tag{7}$$

$$\zeta = \zeta_o \times e^{-\gamma r_{ij}}, \quad \tau = \tau_o \times e^{-\gamma r_{ij}^2} \tag{8}$$

where $\varphi_i(t)$ is the spatial vector of the ith firefly in the tth iteration. In addition, ζ and τ are the lightness and attraction of the firefly. α and *rand* are a random coefficient and a random value in (0,1).

According to the explanations above, the state $\boldsymbol{H}$ of the hidden layer and the ELM predictor are conformed when only input weights and biases are confirmed. Hence, to obtain an optimal short-term predictor, the FA is used to optimize the input weights and biases in ELM. Hence, the intelligent optimization and prediction are shown in Fig. 3. According to (6), N samples in time t are regarded as input data of ELM, and the output data y are defined as $h((t+N+1)T_s)$. Therefore, we can obtain a series sample data as follows:

$$\begin{cases} \boldsymbol{x}(t) = [h(tT_s), h((t+1)T_s), \ldots, h((t+N)T_s)]^T \\ \boldsymbol{y}(t) = h((t+N1)T_s) \end{cases} \tag{9}$$

where $t \in \{1, 2, \ldots, N_T\}$ and N_T is the total number of samples using the equation above. Then, the sample data are divided into two parts, the training sample data $\boldsymbol{x}_c$ and the testing sample data $\boldsymbol{x}_s$, and the ELM is constructed using those data.

Hence, the intelligent optimization and prediction processes are as follows:

Initialization:

Step 1: Obtain the training data $\boldsymbol{x}_c$ and the testing data $\boldsymbol{x}_s$.

Step 2: Initialize the parameters of ELM, such as the number neurons in the input layer N and the number of neurons in the hidden layer L. It is noted that the number of neurons in the input layer N is same as in (9).

Step 3: Initialize related parameters of FA, such as the firefly size $\hbar$, the spatial dimension of fireflies D, the maximum number of iterations n, the random coefficient α and the reference accuracy ε, and generate and update the spatial vectors of all fireflies $\boldsymbol{X}(\mu, \nu)$. Note that the spatial dimension of fireflies D is calculated by

$$D = N * L + L \tag{10}$$

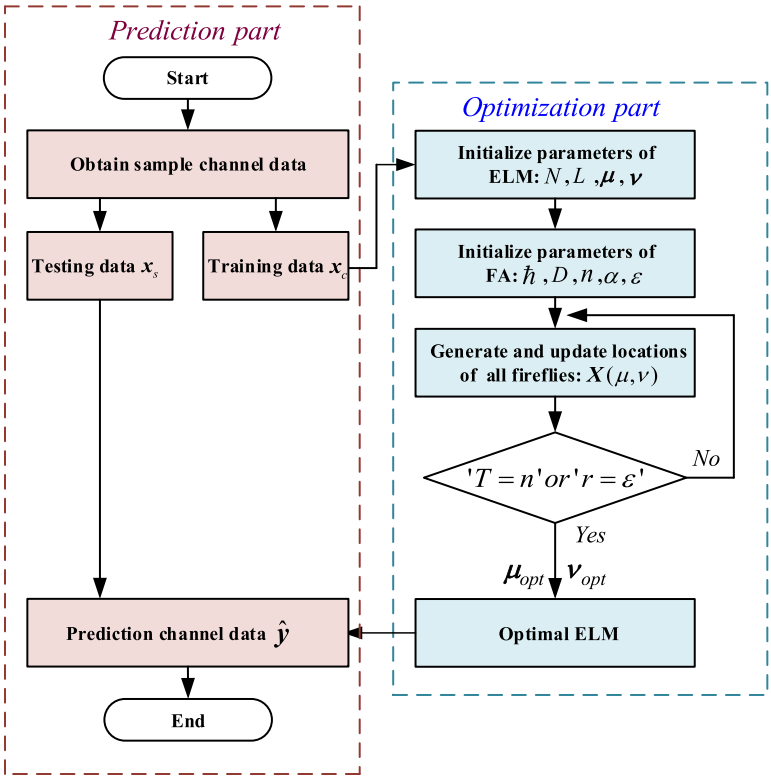


FIGURE 3. Intelligent optimization and prediction in proposed JOELM.

Optimization

Step 4: Calculate the lightness and attraction of all fireflies using (8).

Step 5: Calculate the output training samples $\hat{y}$ according to (6).

Step 6: Obtain the root mean squared error (RMSE) by

$$RMSE = \sqrt{\frac{1}{o}\sum_{i=1}^{o}(\hat{y}_i - y_i)^2}, \quad i = (1, 2, \ldots, o) \tag{11}$$

where o is the length of the output training samples.

Step 7: Update the global optimal firefly.

Convergence

Step 8: Determine whether the current iteration T is equal to the maximum iterations n, and determine whether the current RMSE value r is equal to or less than the defined convergence precisionε. If so, output optimal weights $\boldsymbol{\mu}_{opt}$ and biases $\boldsymbol{\nu}_{opt}$, or $T = T + 1$, and go back to**Step 3.**

Step 9: Construct ELM with optimal weights $\boldsymbol{\mu}_{opt}$ and biases $\boldsymbol{\nu}_{opt}$, and verify the algorithm's performance using the testing samples $\boldsymbol{x}_s$.

B. TARGETED REPAIR

Because prediction errors are unavoidable, approaches to minimize them are considered. In order to further improve the performance, a repair mechanism based on S-G filter is employed to repair the output data of ELM.

Proposed by Savitzky and Golay in 1964, S-G filter has been widely applied for data denoising. Due to its use of polynomial least squares fitting for local data, the filter requires less computation and achieves better performance.

Assuming that the interval of the abscissa in the data is homogeneous, defined as $\Delta\zeta$, and there are n_r data to the left of χ_i and, similarly, n_l data to the right of χ_i, the fitting value is

$$\hat{\chi}_i = \sum_{j=0}^{k} b_j(\frac{\chi - \chi_i}{\Delta\chi})j \tag{12}$$

where $k = n_l + n_r$. Therefore, assuming that an optimal b_j exists, such that

$$\min \sum_{j=i-n_l}^{i+n_r} [p_i(\chi_j) - t_j]^2 \to 0 \tag{13}$$

then $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{T}$ are defined as follows:

$$\boldsymbol{A} = \begin{pmatrix} -n_l^k & \cdots & -n_l & \cdots & 1 \\ \vdots & & & & \\ 0 & \cdots & 0 & \cdots & 1 \\ \vdots & & & & \\ n_r^k & \cdots & n_r & \cdots & 1 \end{pmatrix} \in \boldsymbol{R}^{(n_r+n_l+1)}$$

$$\boldsymbol{B} = \begin{bmatrix} b_k \\ \vdots \\ b_1 \\ b_0 \end{bmatrix} \in \boldsymbol{R}^{(k+1)}, \quad \boldsymbol{T} = \begin{bmatrix} t_{j-n_i} \\ \vdots \\ t_j \\ \vdots \\ t_{j+n_r} \end{bmatrix} \in \boldsymbol{R}^{(n_l+n_r+1)} \tag{14}$$

Therefore, we can obtain $\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{B} = \boldsymbol{A}^T\boldsymbol{T}$. that is,

$$\min \sum_{j=i-n_l}^{i+n_r} [p_i(x_j) - t_j]^2 = \min \|\boldsymbol{A}\boldsymbol{B} - \boldsymbol{T}\| \tag{15}$$

Assuming that the equation above is optimal, $\boldsymbol{A}\boldsymbol{B} = \boldsymbol{T}$. that is, $\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{B} = \boldsymbol{A}^T\boldsymbol{T}$. Based on the positive definite matrix $\boldsymbol{A}^T\boldsymbol{A}$, it follows that

$$\boldsymbol{B} = (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{T} \tag{16}$$

Finally, the estimated value is

$$\boldsymbol{T}_{estimate} = \boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{T} \tag{17}$$

In JOELM, the targeted repair includes four key steps: initialization, preprocessing, detection and processing for abnormal prediction channel data and updating data in Fig. 4. Based on the explanations and introductions above, the targeted repair is followed.

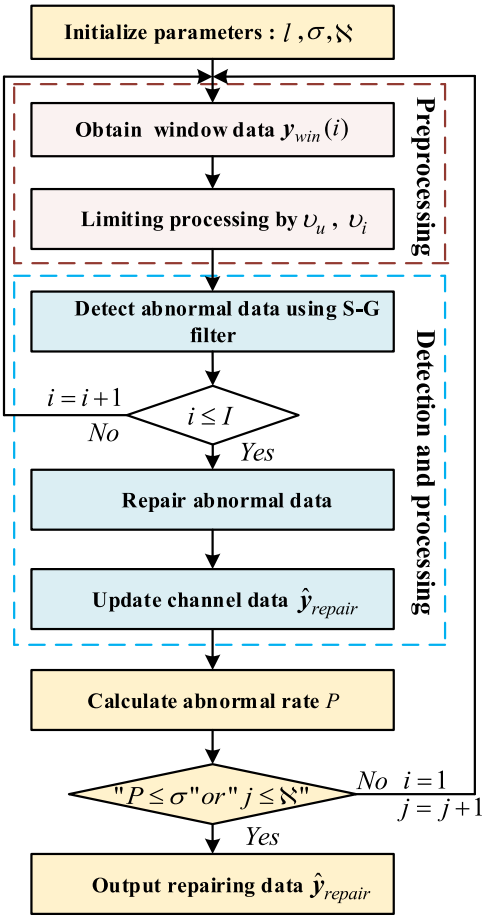


FIGURE 4. The flow chart of targeted repair in proposed JOELM.

Initialization:

Step 1: Obtain predicted data $\hat{\boldsymbol{y}}$ from the modified ELM.

Step 2: Initialize related repair parameters, such as the upper threshold value υ_u and the inferior threshold value υ_i in the preprocessing part. Initialize the sampling window length l, the reference repair mean square error σ, and the maximum number of repairing iterations $\aleph$.

Step 3: Obtain the window samples $\hat{\boldsymbol{y}}_{win}(t)$, $t \in (1, 2, 3, \ldots, I)$, where t contributes the window sampling number and I is the maximum sample number. Note that, here, t is different from its counterpart in the prediction algorithm due to the different meaning in each context, and the maximum sample window number I is determined by the sampling window length l and the obtained prediction data $\hat{y}$ from the modified ELM. that is,

$$I = \Phi[\frac{\hat{y}}{l}]$$

where $\Phi[\bullet]$ is the rounding function.

Preprocessing:

Step 4: Determine whether the number of samples is greater than the upper threshold value υ_u or less than the inferior threshold value υ_i. If so, replace the values υ_u and υ_i accordingly. Conspicuous abnormal predicted channel samples can be preprocessed in this way.

Detection and processing:

Step 5: Calculate the estimated window samples $\hat{\boldsymbol{y}}_{win_esti}$ by S-G filter using (17). Because local fitting samples are calculated by partial least squares, a smooth curve can also be obtained.

Step 6: Determine whether the current window is equal to the maximum I. If not, $t = t + 1$, and go back to **Step 3**.

Step 7: Obtain the estimated prediction samples $\hat{\boldsymbol{y}}_{esti}$.

Step 8: Detect abnormal values.

Because local fitting samples are calculated by partial least squares in (17), a smooth curve can also be obtained. Hence, to detect abnormal values, detection errors can be calculated by

$$\Delta = \frac{|\hat{\boldsymbol{y}}_{esti} - \hat{\boldsymbol{y}}|}{\hat{\boldsymbol{y}}} \tag{18}$$

If some sample errors exceed the standard error, they are regarded as abnormal samples and will be repaired in the next process. The error standard Ξ is an empirical value and can be calculated by

$$\Xi = \frac{\kappa}{\ln(\hat{\boldsymbol{y}} + e)} \tag{19}$$

where κ is the repair coefficient in [0.2,0.67]. Obviously, with the decrease in κ, the error standard Ξ decreases correspondingly.

Step 9: Repair the abnormal values. When abnormal values are identified in **Step 7**, they are repaired in the next process using the repair strategy. Hence, the repair strategy is followed by

$$\tilde{y}(k) = \lambda y(k) + (1 - \lambda)\frac{y(k-1) + y(k+1)}{2} \tag{20}$$

where $\tilde{y}(k)$ is the repair value of the kth sample, $k \in (2, 3, \ldots, l * I)$ and λ is the weight coefficient in [0.2,0.91]. Therefore, the repaired samples are the sum of the sample's original value and the average value of its nearby samples.

In addition, when proper and more training samples $\boldsymbol{x}_c$ are available, the reliability of the original value is higher and the weight of the coefficient λ is greater. Note that only abnormal prediction samples are repaired by the repair strategy, and normal prediction samples are reserved.

Step 10: Update the predicted channel samples $\hat{\boldsymbol{y}}_{repair}$.

Step 11: Calculate abnormal rate P in the current jth repair process.

Step 12: Determine whether P is less than or equal to the reference error σ or whether the current repair number j is less than or equal to the maximum $\aleph$. If not, $i = 1$ and $j = j + 1$, and return to **Step 3**. Note that the abnormal rate P is optional for reducing computational complexity.

Step 13: Output the repair prediction channel samples $\hat{\boldsymbol{y}}_{repair}$.

C. COMPLEXITY ANALYSIS

When to estimate performances of algorithm, some metrics are employed, such as convergence rate, accuracy and robustness. Besides, calculation complexity also is a valid perspective to estimate computational demand for some algorithms. It denotes how many memory capacity are required. Hence, it is vital to analyze complexity relative to that of existing channel prediction algorithms, such as AR, BPNN and SVM in our paper.

To reduce the complexity of JOELM, the intelligent optimization is implemented offline to reduce system overhead and cost. Thus, optimal input weights, biases and output weights are obtained offline. Furthermore, ELM has a simple hidden layer structure, in contrast to traditional back-propagation neural networks. According to (6), the main computation complexity includes two parts, $\boldsymbol{H}$ and $\boldsymbol{Y}$ in the prediction part. For $\boldsymbol{H}$, the computation includes input calculation and active function calculation. The former is $(\boldsymbol{X} * \boldsymbol{\mu} + \boldsymbol{v}) \in \boldsymbol{R}^{N \times o}$ and its computation complexity is $O(oNL)$. For the latter, the calculation number is oL, so its calculation complexity is $O(oL)$. Similarly, when to obtain $\boldsymbol{Y}$, the computation complexity is $O(oL)$. Moreover, because the total repair number is $I * \aleph$ in targeted repair, the computational complexity is $O(I\aleph)$. Here, I and $\aleph$ denotes the maximum sample window number and maximum repair number. Therefore, the total computation complexity is $O(\max(oNL, oL, oL) + I\aleph)$. Due to $oNL > oL$, the total computation complexity is $O(oNL + I\aleph)$. According to [22] and [23], the computational complexity of SVM is $O(o^3)$, where o is the number of input data. Moreover, the computation complexity of AR is $O(N_{AR})$, where N_{AR} is the order number in AR [4]. Because $o \gg N$, the complexity of our proposed short-term prediction method is comparable to the complexity of AR and BPNN and is lower than the complexity of SVM.

IV. SIMULATIONS AND DISCUSSIONS

In this section, the classic Rayleigh channel is used to test and verify the performance of proposed short-term prediction mechanism of fading channel.

A. PARAMETER CONFIGURATIONS AND ROBUSTNESS ANALYSIS

1) COEFFICIENT AND WEIGHT COEFFICIENTS SETTING

As explained above, two vital parameters are drawn into targeted repair of JOELM, the repair coefficient κ and the weight coefficient λ. The former determines how many abnormal prediction channel samples are detected. Obviously, a large repair coefficient κ indicates a high repairing standard, and fewer abnormal prediction samples are detected in turn. Similarly, the latter determines the repair values for abnormal prediction channel samples.

To discuss effects on the performance of targeted repair, the RMSE values with repair coefficient κ and weight coefficient λ are plotted in Fig. 5. In addition, the related parameters are set as follows: classic Rayleigh channel: $f_d = 500\ Hz$, $T_s = 5{*}10^{-5}s$, $f_s = 20\ kHz$, and $C = 3{*}10^8 m/s$, extreme learning machine: $N_T = 3000$, $x_c = 2400$, $x_c = 600$, $N = 5$, and $L = 10$, firefly algorithm: $\hbar = 30$, $\zeta_o = 1$, $\tau_o = 1$, $n = 50$, $\alpha = 10$, and $\varepsilon = 1{*}10^{-4}$, and S-G filter: $1 = 5$, $\sigma = 1{*}10^{-4}$ and $\aleph = 50$.

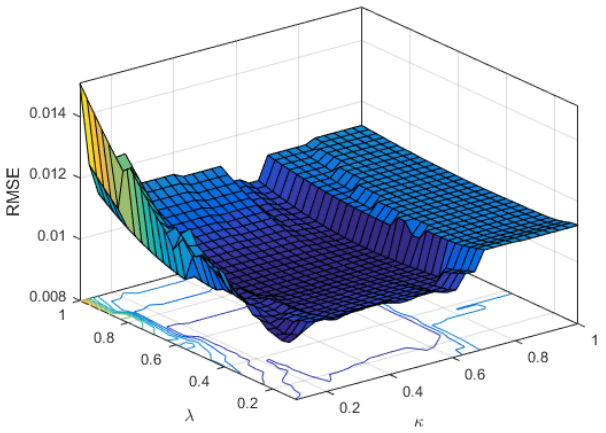


FIGURE 5. RMSE after targeted repair with repair coefficient κ and weight coefficient λ.

As shown, when the repair coefficient κ is 0.1, the RMSE after targeted repair is worse when the weight coefficient λ is increased. This effect occurs because almost of all prediction channel samples are regarded as abnormal values, which causes the normal prediction channel samples to decay. Then, when repair coefficient κ and weight coefficient λ are increased, the repair channel prediction samples mainly depend on their own predicted values according to (20). Hence, the RMSE values are higher than those with the appropriate weight coefficient λ. For example, when $\kappa = 0.4$, the RMSE is 0.009753 at $\lambda = 0.55$, which is less than the value of 0.01129 at $\lambda = 1$ and less than the value of 0.01 at $\lambda = 0.1$. In addition, when repair coefficient κ is greater than 0.7, the error standard Ξ is higher than error value in (18), which indicates few normal prediction channel samples are detected and only the preprocess is applied for predicted data. Hence, according to the explanations above, we can conclude that excellent performances can be obtained

only by proper ranges of the repair coefficient κ and the weight coefficientλ. Their proper ranges are [0.2, 0.67] and [0.2, 0.91] in JOELM.

2) AMPLITUDE

In order to further estimate the performance of JOELM, repair coefficient κ and weight coefficient λ are set to 0.2 and 0.7 in this subsection. The other related parameters are the same as those described in Subsection ***B***.

The related curves are clearly drawn in Fig. 6∼ Fig. 10. As shown, with the continuous iteration process in the FA, the optimal weights and biases are found in the *14th* iteration, and the optimal fitness value is 0.007014. Due to FA and targeted repair, five abnormal samples are detected, that is, the *27*th, *111*th, *112*th, *113*th and *246*th samples (marked by red circles) in Fig. 7. Fig. 8 shows the abnormal prediction samples detected by the repair process. Obviously, the number of detected abnormal channel samples is the same as that shown in Fig. 7. Hence, the abnormal sample curve with the repair iterations also indicates that these five abnormal samples existed in the initial condition before the repair process. Based on Fig. 7 and Fig. 8, we observe

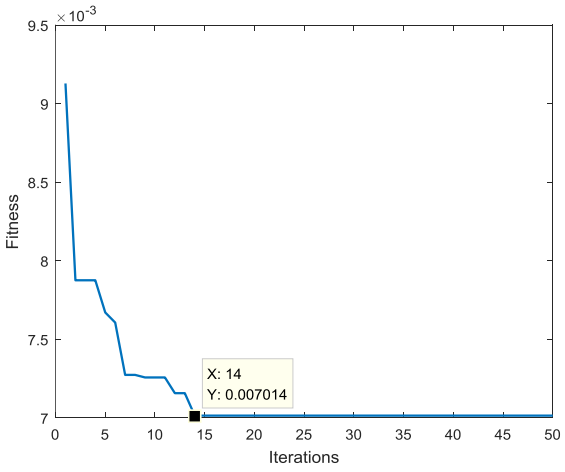


FIGURE 6. Fitness curve in the firefly algorithm.

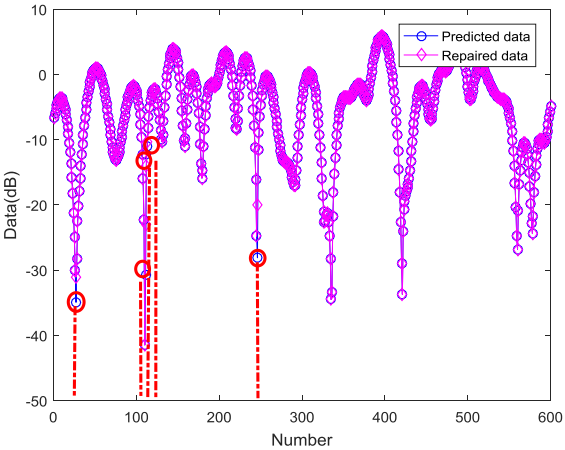


FIGURE 7. Predicted data and repaired data.

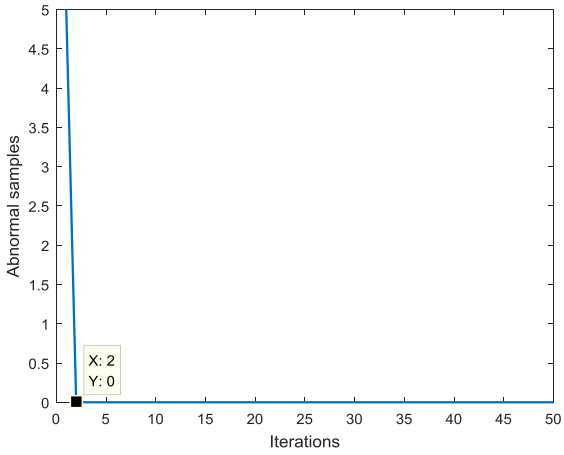


FIGURE 8. Abnormal samples in repair process.

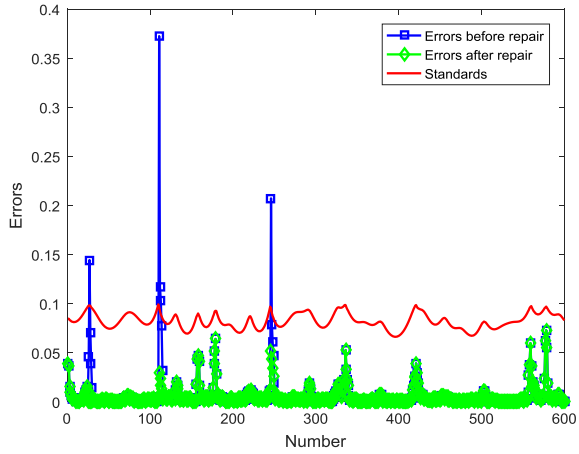


FIGURE 9. Errors in the repair process.

that the number of abnormal samples are reduced drastically, and no abnormal sample are remained in the second repair iteration. Thus, the errors before and after targeted repair and the repair standards indicated by (9) are as shown in Fig. 9. As shown, for 600 predicted channel samples, the *27*th, *111*th, *112*th, *113*th and *246*th values are detected as abnormal prediction samples (marked by red circles in Fig. 7).

Then Fig. 10 shows the RMSE value curve with the repair process in view of the actual testing channel data. As clearly shown, RMSE is reduced to 0.006562. Compared to initial 0.007014 in Fig.6, the prediction performance is improved about 6.44%, demonstrating the validity of the proposed mechanism. It is noted that abnormal predicted sample is a relative concept rather than an absolute concept. Hence, not all abnormal predicted samples are remained in all targeted repair of JOELM, and the number of detected abnormal prediction channel samples is determined by the repair coefficient κ and the weight coefficient λ.

3) ANALYSIS OF INTELLIGENT OPTIMIZATION AND TARGETED REPAIR

As is explained above, JOELM consists two key steps, that is, the intelligent optimization and targeted repair,

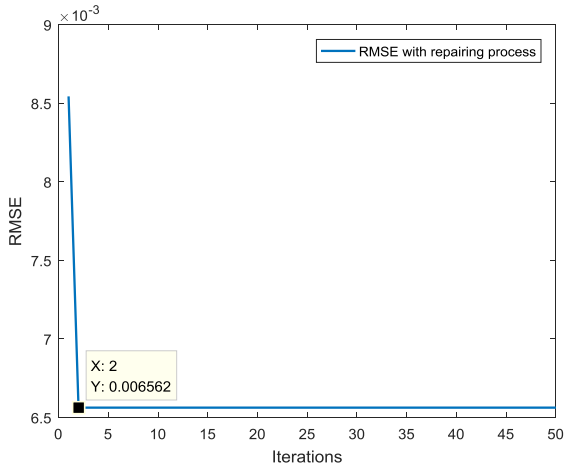


FIGURE 10. RMSE in the repair process.

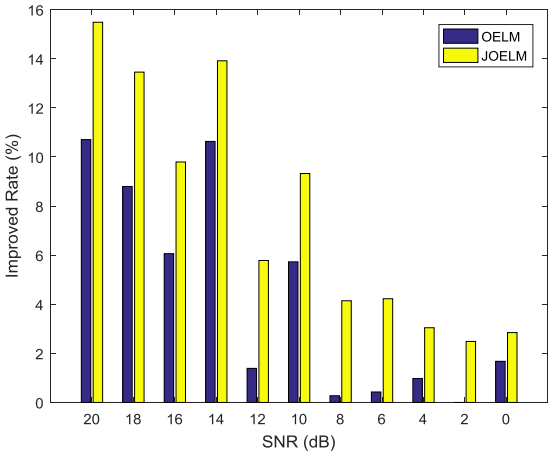


FIGURE 12. Improved rates of OELM and JOELM for ELM.

which contribute the whole prediction performance. Hence, in order to estimate contributions of those two steps, a series of simulations are implemented. The simulation is the performances of RMSE with signal to noise ratio (SNR) under ELM, optimized extreme learning machine (OELM) and JOELM. Repair coefficient κ and weight coefficient λ are set as 0.2 and 0.5, and other defined parameters are same to subsection $\boldsymbol{B}$. It is noted that the RMSE with SNR of ELM can be implemented independently and taking the consistency into considerations, a rand seed is required when to randomly generate weights $\boldsymbol{\mu}$ and biases $\boldsymbol{\nu}$. As for the RMSEs of OELM and JOELM are obtained simultaneously in intelligent optimization and targeted repair.

Hence, RMSEs with SNR of ELM, OELM and JOELM are showed in Fig.11. As we can see, the RMSE is 0.2163 with ELM, while the value is 0.16 after optimization using FA. Then when the targeted repair is added, RMSE is reduced to 0.1437. So the improved rates of intelligent optimization and targeted repair are 10.71% and 4.78%, respectively in Fig.12. Based on this, the contributions of intelligent optimization and targeted repair are 69.14% and 30.86%. It is obvious that

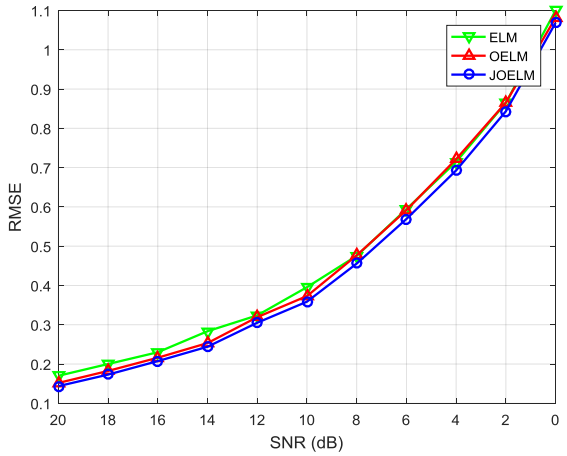


FIGURE 11. RMSEs with SNR using ELM, OELM and JOELM.

the former holds more contributions than the latter. When to decrease SNR, contribution of the former tends to decrease. Such as when SNR = 12dB, the whole improved performance is 5.783% and RMSEs of ELM, OELM and JOELM are 0.3238, 0.3193 and 0.3051, respectively. Hence, the contributions of intelligent optimization and targeted repair are 1.389% and 4.394%, and the contributions are 24.02% and 75.98% in turn. Then when SNR = 4dB, RMSEs of ELM, OELM and JOELM are 0.7216, 0.7146 and 0.6928. So the whole improved rates of two steps are 0.97% and 3.044% and their main contributions are 32.08% and 67.92%. As we can see, when SNR is high, excellent prediction performances are obtained and main contributions of improved performances are given by intelligent optimization, targeted repair plays an assistant role. Since the noise power increases, the whole performances tend to decay and the targeted repair gives main contributions to improve whole performances and the optimization part plays assistant role. Hence, a conclusion is given that when SNR is bigger than 8dB, the intelligent optimization gives more contributions than targeted repair in whole performances, and when SNR is less than 8dB and bigger than 0dB, the targeted repair plays main role to improve prediction performances, the intelligent optimization is an assistant role. Due to RMSE is more than 1 when SNR is less than 0dB, valid channel data are decayed for severe noise power. So in order to obtain better prediction accuracy, necessary denoising algorithms are required in this case.

B. PREDICTED PERFORMANCE EVALUATIONS

1) BIT ERROR RATE AND SYMBOL ERROR RATE

When data are sent through wireless channel, received data often are decayed due to unavoidable noise. In order to evaluate quality of data transmission, bit error rate (BER) and symbol error rate (SER) are often employed, and the mightier noise power is, the higher values of BER and SER are, which indicate communication quality is worse. Based on BER and SER, necessary strategies are required to improve communication quality.

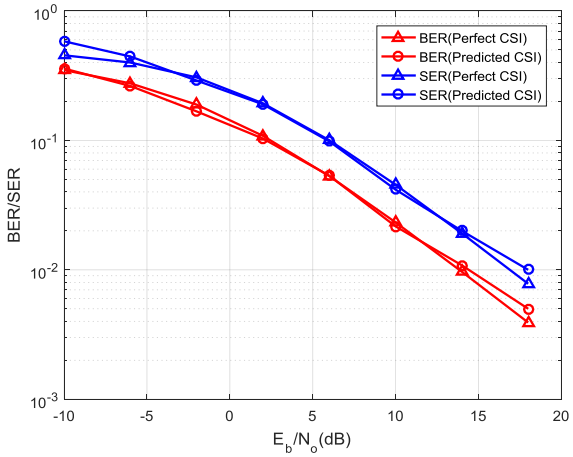


FIGURE 13. BER/SER evaluations in perfect CSI and predicted CSI.

Hence, comparisons of BER and SER in perfect CSI and predicted CSI are given in Fig.13. In this subsection, 2-ary differential phase-shift keying (2-DPSK) is employed and we assume a coherent receiver is adopted for signal detection. Parameters are same to the subsection ***B***, and repair coefficient κ and weight coefficient λ are 0.1 and 0.5, respectively. As we can see, when to increase E_b/N_0(dB) (i.e., bit energy to noise power spectral density ratio), BER and SER are decreasing from 0.3592 to 0.003916 and 0.5833 to 0.007801 in perfect CSI, respectively, which indicate communication quality is more better. Then for predicted CSI, its BER and SER are close to perfect CSI. So we can get a conclusion that valid and exact CSI can be obtained by JOELM.

2) ROBUSTNESS COMPARISONS

Robustness is a vital parameter to estimate performance of algorithm, especially in wireless signal with noise. In order to estimate performance of JOELM, the robustness is added in this paper.

AR is a classic strategy for predicting a short-term fading channel. Thus, in this subsection, AR is implemented with AR filter of order 6. To compare with other existing machine learning tools, BPNN and SVM are also discussed in this subsection. The BPNN is implemented using the neural network toolbox in MATLAB, and its parameters are as follows: the training function is TRAINLM, the adaption learning function is LEANGOM, the performance function is set to MSE, the number of layers in the BPNN is 2, and the transfer function is TANSIG. The SVM is implemented using limsvm-3.22 developed by Chih-Chung Chang, and its parameters are as follows: the type of SVM is epsilon-SVR, the epsilon in the loss function of epsilon-SVR is 0.01, and the kernel function type is radial basis function, where the gamma value is 2.8. The simulation environment is MATLAB 2016a in a Windows 7, Intel(R) CPU E5-1620@3.6 GHz with 8.0 GB of RAM.

Hence, RMSEs with various SNR for AR, SVM, BPNN and JOELM are showed in Fig.14. As we can see, although JOELM has similar RMSEs to BPNN when the SNR is bigger than 12dB, it still precedes AR and SVM. Then when SNR is less than 12dB, JOELM has less RMSE values than BPNN, AR and SVM. Hence, a conclusion is given that when SNR is bigger than 12dB, JOELM has similar robustness to BPNN and has better robustness than AR and SVM. Then SNR is less than 12dB, JOELM holds better robustness than BPNN, SVM and AR.

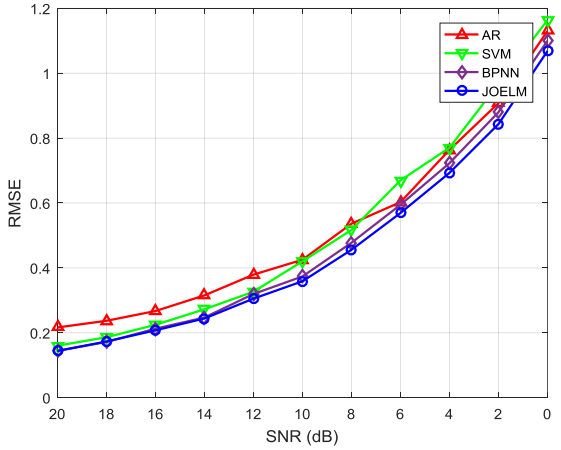


FIGURE 14. RMSEs with various SNR for AR, SVM, BPNN and JOELM.

3) PROBABILITY DENSITY FUNCTION (PDF) AND CUMULATIVE DISTRIBUTION FUNCTION (CDF)

The PDF and CDF are two statistical properties used to estimate the quality of a wireless communication channel. It is known that various channels yield different PDFs and CDFs. Hence, the theoretical PDF and CDF curves of AR, BPNN, SVM and JOELM are shown in Fig. 15. In addition, RMSEs of the PDF and CDF values are calculated in Table 1.

TABLE 1. RMSE of estimated PDF, CDF, LCR and ADF in AR, BPNN, SVM and JOELM.

Algorithm	PDF	CDF	LCR	ADF
AR	0.0307	0.0063	38.9651	$3.84*10^{-4}$
SVM	0.0370	0.0196	26.3857	$7.93*10^{-4}$
BPNN	0.0497	0.0369	38.0607	$1.3*10^{-4}$
JOELM	0.0297	0.0129	68.2691	$2.64*10^{-4}$

As shown, the curves pertaining to AR, BPNN, SVM and JOELM are very close to the theoretical curve in Fig. 15(a), indicating that the four channel short-term prediction methods are valid and accurate. Moreover, according to Table I, JOELM yields the best RMSE value of 0.0297 compared with those of the three methods, i.e., 0.0307 for AR, 0.0370 for SVM and 0.0497 for BPNN with respect to the PDF, indicating that the four channel short-term prediction methods

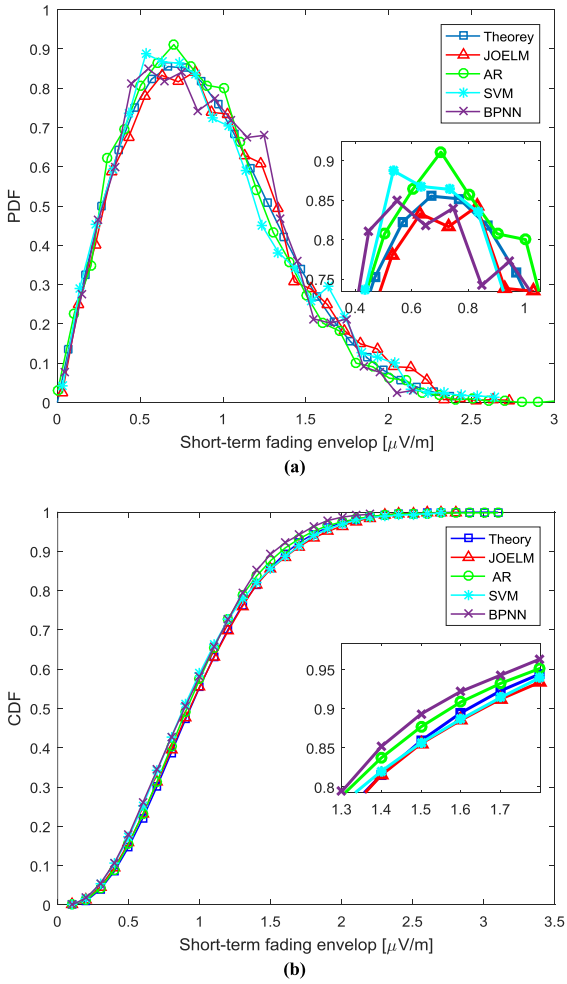


FIGURE 15. Comparisons of PDF (a) and CDF (b) with theory, AR, BPNN, SVM and JOELM.

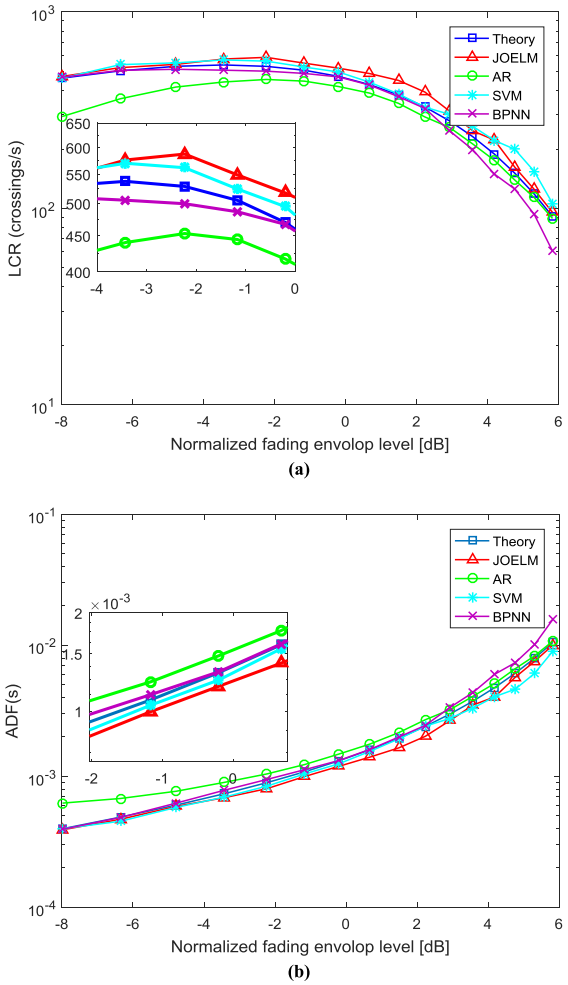


FIGURE 16. Comparisons of LCR (a) and ADF (b) with theory, AR, BPNN, SVM and JOELM.

are valid and accurate. Moreover, the four CDF curves nearly overlap the theoretical curve in Fig. 15(b). Moreover, the RMSE of JOELM is 0.0129, which is better than 0.0196 for SVM and0.0369 for BPNN and only worse than AR. Therefore, we can conclude that for PDF, JOELM yields better performance than AR, BPNN and SVM, whereas for CDF, JOELM is only worse than AR.

4) LEVEL CROSSING RATE (LCR) AND AVERAGE DURATION OF FADES (ADF)

Similar to PDF and CDF, LCR and ADF also are two statistical parameters used to estimate the positive direction in wireless communication. The former represents the expected rate at which the fading envelope crosses a specified signal level in a positive-facing direction, while the latter represents the average duration crossing below a certain threshold in the receiver.

Hence, the LCR and ADF curves of AR, BPNN, SVM and JOELM are illustrated in Fig. 16, and the RMSEs of LCR and ADF for the four channel prediction algorithms are shown in Table. 1. As indicated, the four curves of the LCR in AR, SVM, BPNN and JOELM are close to the theoretical curve in Fig. 16(a). Similarly, the ADF curves for the AR, BPNN, SVM and JOELM are close to the theoretical curve in Fig.16 (b), which indicates the validity of four channel prediction methods with respect to LCR and ADF. Furthermore, although the RMSE value of 68.2691 for JOELM is greater than 38.9651 for AR, 38.0607 in BPNN and 26.3857 in SVM for LCR, RMSE value of $2.64^{*}\ 10^{-4}$ of ADF in JOELM is only inferior to that of the BPNN and is superior to those of the AR and SVM, which also validates JOELM.

V. CONCLUSION

Our paper focuses on short-term prediction for fading channel. In particular, a novel prediction strategy (JOELM) based on ELM, FA and the S-G filter is proposed. The JOELM consists of two key steps, intelligent optimization and targeted repair. In the former, FA aims at the optimization of input weights and biases of ELM, preserving these key parameters from being randomly determined. In the latter, the S-G filter is innovatively adopted for reducing potential prediction errors, so as to further improve the final performance of JOELM.

Comprehensive characteristics of the JOELM framework are evaluated and discussed through performing all-around simulation tests. The recommended parameter settings and implementation processes are also given in detail. The comparison experiments prove that our proposed JOELM performs better than AR, BPNN and SVM, which promise it can be potentially applied in short-term prediction of fading channel for wireless communication systems.

REFERENCES

[1] N. M. Tomasevic, A. M. Neskovic, and N. J. Neskovic, "Short-term fading simulation using artificial neural networks," *AEU—Int. J. Electron. Commun.*, vol. 65, no. 7, pp. 641–649, Jul. 2011.

[2] N. M. Tomasevic, A. M. Neskovic, and N. J. Neskovic, "Short-term fading simulator based on artificial neural networks," in *Proc. IEEE EUROCON*, May 2009, pp. 1681–1688.

[3] T. Eyceoz, A. Duel-Hallen, and H. Hallen, "Deterministic channel modeling and long range prediction of fast fading mobile radio channels," *IEEE Commun. Lett.*, vol. 2, no. 9, pp. 254–256, Sep. 1998.

[4] Y. Zhao, H. Gao, N. C. Beaulieu, Z. Chen, and H. Ji, "Echo state network for fast channel prediction in ricean fading scenarios," *IEEE Commun. Lett.*, vol. 21, no. 3, pp. 672–675, Mar. 2017.

[5] T. Eyceoz, A. Duel-Hallen, and H. Hallen, "Prediction of fast fading parameters by resolving the interference pattern," in *Proc. IEEE Conf. Rec. 31st Asilomar Signals, Syst. Comput.*, Pacific Grove, CA, USA, Jan. 1997, pp. 167–171.

[6] P. Sharma and K. Chandra, "Prediction of state transitions in Rayleigh fading channels," *IEEE Trans. Veh. Technol.*, vol. 56, no. 2, pp. 416–425, Mar. 2007.

[7] T. Ding and A. Hirose, "Fading channel prediction based on combination of complex-valued neural networks and chirp Z-transform," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 9, pp. 1686–1695, Sep. 2014.

[8] N. X. Zhao, C. Hou, and Q. Wang, "A new SVM-based modeling method of cabin path loss prediction," *Int. J. Antennas Propag.*, vol. 2013, Apr. 2013, Art. no. 279070.

[9] J. Sun, T. Zhang, and L. Feng, "Novel nonlinear prediction algorithm for fast fading channel," (in Chinese), *J. Beijing Univ. Aeronaut. Astronaut.*, vol. 31, no. 5, pp. 499–503, Mar. 2005.

[10] Z. Xiang, T. Zhang, and J. Sun, "Prediction algorithm for fast fading channels based on the chaotic attractor," (in Chinese), *J. Xidian Univ.*, vol. 33, no. 1, pp. 145–149, Feb. 2006.

[11] H. Jaeger and H. Haas, "Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication," *Science*, vol. 304, no. 5667, pp. 78–80, Apr. 2004.

[12] M.-B. Li, G.-B. Huang, P. Saratchandran, and N. Sundararajan, "Fully complex extreme learning machine," *Neurocomputing*, vol. 68, nos. 1–4, pp. 306–314, Oct. 2005.

[13] B. Jing, Z. Qian, Y. Pei, and J. Wang, "Ultra short-term PV power forecasting based on ELM segmentation model," *J. Eng.*, vol. 2017, no. 13, pp. 2564–2568, 2017.

[14] M. Shafiullah, M. Abido, and Z. Al-Hamouz, "Wavelet-based extreme learning machine for distribution grid fault location," *IET Gener., Transmiss. Distrib.*, vol. 11, no. 11, pp. 4256–4263, Nov. 2017.

[15] Y. Yang, M. Yang, S. Huang, Y. Que, M. Ding, and J. Sun, "Multifocus image fusion based on extreme learning machine and human visual system," *IEEE Access*, vol. 5, pp. 6989–7000, 2017.

[16] J. Liu, X. Jin, F. Dong, L. He, and H. Liu, "Fading channel modelling using single-hidden layer feedforward neural networks," *Multidimensional Syst. Signal Process.*, vol. 28, no. 3, pp. 885–903, Jul. 2017.

[17] I. S. Popool, S. Misra, and A. A. Atayero, "Outdoor path loss predictions based on extreme learning machine," *Wireless Pers. Commun.*, vol. 99, no. 1, pp. 441–460, Mar. 2018.

[18] X.-S. Yang, "Analysis of algorithms," in *Nature-Inspired Optimization Algorithms*. Oxford, U.K.: Elsevier, 2014, ch. 2, pp. 23–44.

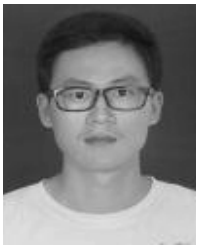
[19] P. Baumgartner *et al.*, "Multi-objective optimization of Yagi–Uda antenna applying enhanced firefly algorithm with adaptive cost function," *IEEE Trans. Magn.*, vol. 54, no. 3, Mar. 2018, Art. no. 8000504.

[20] Y. Guo, B. Z. Li, and N. Goel, "Optimised blind image watermarking method based on firefly algorithm in DWT-QR transform domain," *IET Image Process.*, vol. 11, no. 6, pp. 406–415, Jun. 2017.

[21] D. F. Teshome, C. H. Lee, Y. W. Lin, and K. L. Lian, "A modified firefly algorithm for photovoltaic maximum power point tracking control under partial shading," *IEEE J. Emerg. Sel. Topics Power Electron.*, vol. 5, no. 2, pp. 661–671, Jun. 2017.

[22] A. Abdiansah and R. Wardoyo, "Time complexity analysis of support vector machines (SVM) in LibSVM," *Int. J. Comput. Appl.*, vol. 128, no. 3, pp. 28–34, Oct. 2015.

[23] L. Gupta, M. Samaka, R. Jain, A. Erbad, D. Bhamare, and C. Metz, "COLAP: A predictive framework for service function chain placement in a multi-cloud environment," in *Proc. IEEE 7th Annu. Comput. Commun. Workshop Conf. (CCWC)*, Las Vegas, NV, USA, Jan. 2017, pp. 1–9.



YONGBO SUI received the B.E. degree in electrical engineering from Xiangtan University, Xiangtan, Hunan, in 2014. He is currently pursuing the Ph.D. degree with the Hefei University of Technology, Hefei, Anhui, China. His research interests include intelligent algorithms and wireless communication channels.

WENXIN YU received the B.Sc. degree in applied mathematics from Hebei Normal University, Shijiazhuang, China, in 2005, the M.S. degree in wavelet analysis from the Changsha University of Science and Technology, Changsha, in 2008, and the Ph.D. degree in electrical engineering from Hunan University, Changsha, in 2015. He joined the Hunan University of Science and Technology, where he is currently a Lecturer with the School of Information and Electrical Engineering. He also holds a post-doctoral position with the School of Electrical and Information Engineering, Hunan University. His interests include fault diagnosis, signal processing, and wavelet analysis and its applications.

QIWU LUO (M'17) received the B.S. degree in communication engineering from the National University of Defense Technology, Changsha, China, in 2008, and the M.Sc. degree in electronic science and technology and the Ph.D. degree in electrical engineering from Hunan University, Changsha, in 2011 and 2016, respectively.

He was a Senior Engineer of instrumentation with Wasion Group Co., Ltd., Changsha, and the Deputy Technical Director with Hunan RAMON Technology Co., Ltd., Changsha. He is currently a Lecturer with the School of Electrical Engineering and Automation, Hefei University of Technology, Hefei, China. His research interests include the research of real-time information processing, parallel hardware architecture design and reconfigurable computing, and fault testing and diagnosis of large-scale analog circuits.