IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 28, 2018, accepted July 28, 2018, date of publication August 7, 2018, date of current version September 7, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2864222

Spectrum Prediction Based on Taguchi Method in Deep Learning With Long Short-Term Memory

LING YU[1], JIN CHEN[1], GUORU DING[1,2], (Senior Member, IEEE), YA TU[3], JIAN YANG[4], (Member, IEEE), AND JIACHEN SUN[1]

[1]College of Communications Engineering, Army Engineering University of PLA, Nanjing 210007, China
[2]National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China
[3]College of Information and Communication, Harbin Engineering University, Harbin 150001, China
[4]The 63rd Institute, National University of Defense Technology, Nanjing 210014, China

Corresponding authors: Jin Chen (chenjin99@263.net) and Guoru Ding (dr.guoru.ding@ieee.org)

This work was supported in part by the National Natural Science Foundation of China under Grants 61871398, 61501510 and 61631020, in part by the Natural Science Foundation of Jiangsu Province under Grant BK20150717, in part by the China Postdoctoral Science Foundation Funded Project under Grant 2018T110426, and in part by the Natural Science Foundation for Distinguished Young Scholars of Jiangsu Province under Grant BK20160034.

ABSTRACT Spectrum prediction is a promising technology in cognitive radio networks, since it can reduce considerable time and energy consumed in spectrum sensing process. Many spectrum prediction algorithms have achieved good performance, but majority of them with shallow architecture cannot capture the inherent correlations of spectrum data very well. Long short-term memory (LSTM) neural network in deep learning has been validated to have strong capability of solving time series problems. In this paper, we develop a spectrum prediction framework with a deep learning approach on two real-world spectrum datasets. For the first dataset to predict channel occupancy states, we firstly employ the taguchi method to determine the best optimized configuration of neural network for certain spectrum point and then analyze the effect of each design hyper-parameter. Next, we build LSTM neural networks with two perspectives of regression and classification for spectrum prediction. For the second dataset to predict channel quality, we compare the prediction performance of the LSTM neural network and conventional multilayer perceptron (MLP) neural network. For both of our datasets, results show that the prediction performance varies with frequency bands. From the point of statistics, the LSTM neural network has better prediction performance than the MLP neural network and is more stable as well. Furthermore, we find that the performance of the LSTM neural network with classification perspective is slightly better than that with regression perspective in our first dataset.

INDEX TERMS Deep learning, LSTM neural network, spectrum prediction, taguchi method.

I. INTRODUCTION

A. BACKGROUND AND MOTIVATION

Spectrum demand increases along with the development of communication traffic and mobile devices, but a large number of spectrum measurements show that the spectrum band is underutilized due to the static allocation strategy [1]. Cognitive radio network (CRN) is therefore introduced, permitting unlicensed users to communicate in idle time slots with no harmful interference to licensed users [2], [3]. Spectrum sensing and spectrum prediction are both the methods to explore the spectrum access opportunities in CRN. Spectrum sensing determines the current radio spectrum state (RSS) using various signal detection methods [4]–[7] which consume a lot of time and energy, while spectrum prediction infers the future possibly unknown/unmeasured RSS from historically known/measured spectrum data by exploiting the inherent correlations and/or regularities [8]. Spectrum prediction can alleviate the processing delay and energy consumption occurred in spectrum sensing [9], and has wide applications in future wireless networks, such as CRNs, cooperative relay networks [10], [11], and 5G networks [12].

Recently, there have been a few interesting studies on spectrum prediction, and a recent comprehensive survey can be found in our work [13]. However, majority of the existing studies are based on conventional statistical techniques or shallow architecture models to predict channel quality or states. In recent years, deep learning has attracted wide attention in industrial and academic field with the

development of computing power and the arrival of big data era. It is developed from neural network[1] as a branch of machine learning. Its deep architecture and ingenious structural design make deep learning have better power to model and tackle certain problem, achieving outstanding performance in computer vision [14], natural language processing [15] and other fields [16], [17]. One kind of the main architecture in deep learning is convolutional neural network. It behaves very excellent in image classification and target recognition task [14], but the input data are required to be of the same dimension which prevents its direct application on time series of various lengths [18]. Recurrent neural network (RNN), another main architecture, is more suitable for solving problems related to time series problem. Long short-term memory (LSTM) network is an improved version of RNN to overcome the drawback of long-term dependency, introducing several gates into single neuron to better coordinate historical and current information. These observations motivate us in this paper to consider leveraging LSTM network in deep learning techniques for developing powerful spectrum prediction scheme.

B. RELATED WORK

Majority of existing studies on spectrum prediction can be divided into two groups: *i*) prediction of channel quality and *ii*) prediction of channel occupancy. Prediction of channel quality is from the perspective of regression, where prediction models are often performed by an autoregressive integrated moving average (ARIMA) [19] or support vector regression (SVR) [20]. An inconvenience of ARIMA is that its natural tendency to concentrate on the mean values of the past series data, while SVR has a limitation of the absent structured means to determine some key parameters in the model [21]. For the studies on prediction of channel occupancy, binary time series are often considered to represent the channel occupancy state. In [9], spectrum sensing process is modeled as non-stationary hidden markov model and states are predicted through bayesian inference method according to the duration of channel state and accuracy of sensing. Reference [22] proposes partial periodic pattern mining based model to predict real-world Wi-Fi and personal communication service activities, considering that patterns of real-world signals do not repeat perfectly due to noise, sensing errors, and irregular behaviors. In [23], the author analyzes the spectrum occupancy in CRNs using different machine learning algorithms including supervised learning and unsupervised learning techniques. Reference [24] designs artificial neural network with three layers to predict channel status in different traffic scenarios. The biggest disadvantage of binary time series prediction lies in that the conversion of observed data into binary series is largely dependent on the threshold setting, which undoubtedly leads to errors [8]. More related work on spectrum prediction can be found in our recent survey and tutorial work [13].

Though deep learning approach is quite popular in recent years, its applications in communication field are still in its infant stage [25]–[28]. How to take advantage of deep learning algorithm to efficiently and effectively make spectrum prediction is an emerging research point. There are relatively few studies considering the application of deep learning in spectrum prediction. One pioneering study is found in [29], which applies deep learning to predict spectrum availability in cognitive aerospace communications, however, real-world data are transformed into binary channel states like other prediction algorithms. Another study is our recent work [30], which is, to our best knowledge, the first work to apply LSTM for spectrum prediction in frequency hopping communication. However, in our previous work [30], frequency hopping sequence is also the binary time series artificially generated. In addition, the hyper-parameters of the LSTM network haven't been optimized, which have a significant impact on prediction accuracy that will be validated in this paper through real-world spectrum analytics.

C. CONTRIBUTIONS

The main contributions of this paper are summarized as follows:

- We develop a spectrum prediction framework with a deep learning approach, which is the first time to predict on the real-world power spectral density (PSD) values. It is interesting for us to apply LSTM network to tackle time series problems in spectrum prediction because LSTM network can have a memory of historical data when it works.
- We design several main trials, leveraging taguchi method with K-fold cross validation, to determine the optimized structure and analyze the effect of each design hyper-parameter of neural network. Taguchi method is introduced here to effectively reduce requirements of time consumption and computational resources.
- We present extensive experiment results with real-world spectrum data analytics. Several interesting insights are found as follows: *i*) Hyper-parameters of neural network have varying degrees of influence and the depth of network is observed to be more significant; *ii*) LSTM network is found more stable than multilayer perceptron (MLP) network; *iii*) LSTM network with regression and classification perspective both outperforms the MLP network, and LSTM network with classification perspective is observed slightly better than another one.

The remainder of this paper is organized as follows. Section II introduces the process of spectrum prediction. Section III mainly presents the structure of two different network type respectively and the metrics to evaluate network performance. Section IV lists some design hyper-parameters involved in network design and the main trials using taguchi method to determine the optimized architecture.

[1]Hereinafter, if not specifically mentioned, the network means neural network, rather than mobile network or telecommunications network.

Section V shows the experimental results, followed by conclusions in Section VI.

II. SYSTEM MODEL AND PROBLEM FORMULATION

In CRN, the secondary user (SU) perceives the channel at the beginning of time slot. Once the collected PSD data are above certain threshold, SU thinks that the licensed users are communicating and then gives up this time slot. Otherwise, SU starts transmission of information until the end of this time slot. We consider that historical data collected by SU are arranged in a 2-D matrix (f_j, t_i) for each band, where each row represents the measured data at each spectrum point in chronological order, whereas each column represents the data at different time instants of each spectrum point. We have known that each spectrum point has their unique evolution trajectory. It is possible to impose the negative effect on each other if time-frequency combined prediction is carried out. Because of this, data of single spectrum point with temporal correlation are used to make prediction. Spectrum prediction is to predict the spectrum state in next slot by mining the internal relationship of historical spectrum data. That is to say, given historical data of T slots, measured spectrum data in current time slot t and previous slots $x_{t-T+1}, x_{t-T+2}, \ldots, x_t$ are used as input of neural network which has been trained to discover the internal relationship. Then, the output of neural network is the prediction value in next slot $t+1$.

Neural network is trained to capture inherent correlations of spectrum data dependent on the training set before it is used to predict. Fig. 1 shows the dataset construction of single spectrum point for prediction briefly. $q_0, q_1, \ldots\ldots, q_7$ refer to the alternative channel measured value in a time slot. Measured data can be received over time, so dataset is constructed by sliding a window with the fixed length T accordingly for single spectrum point. As illustrated in Fig. 1, fixed window whose length is set five keeps moving a time slot forward to construct multiple samples in the dataset. S_j means the input of the jth sample and L_j means the corresponding label. The spectrum prediction is tackled using the supervised learning, and the neural network is trained to make the predicted value $\hat{L}_j$ as close as possible to the real value L_j. The whole dataset is shuffled in advance and then divided into training set and testing set to avoid sample distribution imbalance.

III. A DEEP LEARNING FRAMEWORK FOR SPECTRUM PREDICTION

Neural network has been well recognized as a powerful tool for spectrum prediction. The neural network usually consists of the input layer, multiple hidden layers and the output layer. By increasing or decreasing the number of hidden layers and the number of hidden layer neurons, any complex nonlinear function can be constructed. In theory, the network with single hidden layer and different number of neurons can fit out any function. Neural network mainly can be classified into two types: feed-forward neural network and RNN. On behalf of these two types, the MLP network and LSTM network are respectively introduced as follows.

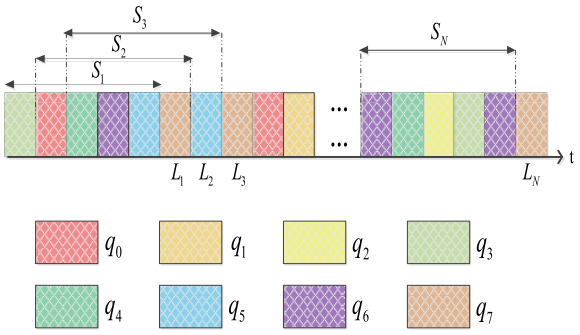

FIGURE 1. Dataset construction of single spectrum point for prediction.

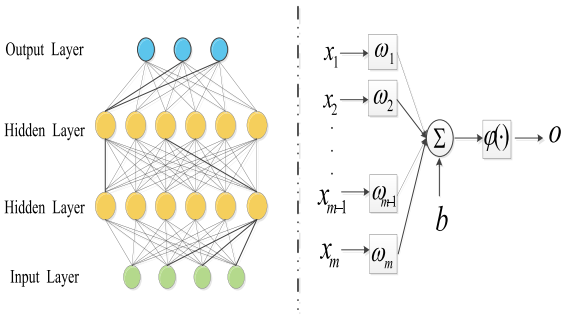

FIGURE 2. Structure of MLP network and its neuron.

A. MLP NETWORK

MLP network is a fully connected feed-forward neural network based on back propagation algorithm. As illustrated in the left of Fig. 2, there are only full connections between adjacent layers, but no connection among the nodes within the same layer. It can only handle the data within the fixed-size window once. It reveals that it is unsuitable for modeling historical dependencies and provides limited capability of temporal modeling and spectrum state prediction. Supposing the size of input data is $C \times T$ which means each one of C channels has data of length T, inputs must be reshaped to a column vector. Fig. 2 also illustrates a mathematical model of one ordinary neuron in MLP network in details [31]. When inputs are $x_1, x_2, \ldots, x_m$, the output of the neuron is calculated as:

$$o = \varphi\left(\sum_{i=1}^{m} \omega_i x_i + b\right) \tag{1}$$

where ω with different subscripts is called weights and b is called bias. $\varphi(\cdot)$ represents the activation function which often employs *sigmoid* or *tanh*.

B. LSTM NETWORK

RNN network is distinguished from feed-forward network where neurons in hidden layers receives feedback, which is from the previous state to current state. Fig. 3 shows a simple RNN architecture with two hidden layers and unfolded in time domain for three time steps. The input vectors are fed into the RNN, one element at one timestep. This kind of

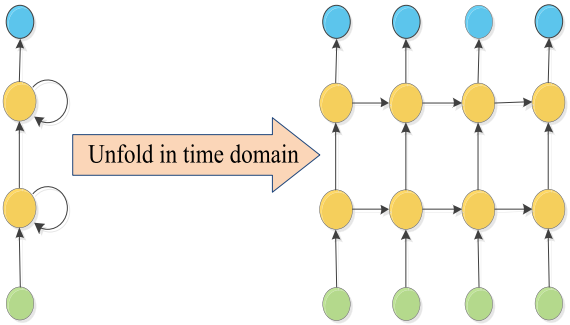

FIGURE 3. Structure of RNN network.

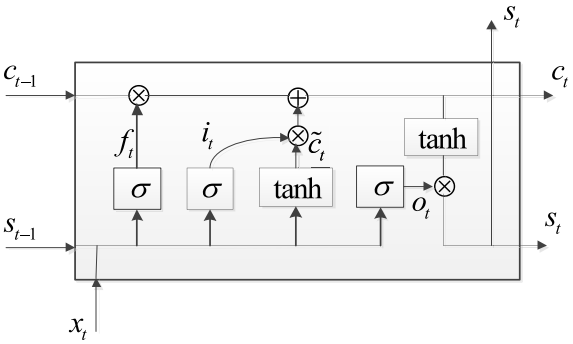

FIGURE 4. Structure of LSTM memory unit.

architecture shows the concept of chronological order while the feed-forward network processes an input vector of fixed length, so it can be seen that RNN is superior in solving time series problems. RNN network allows the information of historical inputs to be stored in the network's internal state, and thereby take advantage of all the available input information up to the current time. In theory, RNN can learn the characteristics of any length of time series. However, it is proved by experiments that the performance of RNN network can be restricted due to gradient vanishment or gradient explosion [32]. To tackle the gradient problems faced by RNN network, LSTM network is designed by introducing a key component named memory unit [33]. As illustrated in Fig. 4, memory unit contains a memory cell and three gates which are named depending on their corresponding practical functionalities. The memory cell has the responsibility to remember the current state of the unit to be used in next timestep. Gates impose controls on the ratio of forgetting and storing related information. The input gate indicted by **i** controls how much much new information flows into the memory cell. The forget gate indicted by **f** decides how much of the memory cell should be abandoned in current memory cell. The last gate denoted as **o**, named by output gate, controls the amount of information to compute the output activation of the memory unit and further flows into the rest of the network. The mathematical calculation of an LSTM unit is as follows [33]:

$$\mathbf{i}_t = \sigma\left(\mathbf{W}_x^i \cdot \mathbf{x}_t + \mathbf{W}_s^i \cdot \mathbf{s}_{t-1} + \mathbf{b}_i\right) \tag{2}$$

$$\mathbf{f}_t = \sigma\left(\mathbf{W}_x^f \cdot \mathbf{x}_t + \mathbf{W}_s^f \cdot \mathbf{s}_{t-1} + \mathbf{b}_f\right) \tag{3}$$

$$\mathbf{o}_t = \sigma\left(\mathbf{W}_x^o \cdot \mathbf{x}_t + \mathbf{W}_s^o \cdot \mathbf{s}_{t-1} + \mathbf{b}_o\right) \tag{4}$$

$$\tilde{\mathbf{c}}_t = \tanh\left(\mathbf{W}_x^c \cdot \mathbf{x}_t + \mathbf{W}_s^c \cdot \mathbf{s}_{t-1} + \mathbf{b}_c\right) \tag{5}$$

$$\mathbf{c}_t = \mathbf{i}_t \odot \tilde{\mathbf{c}}_t + \mathbf{f}_t \odot \mathbf{c}_{t-1} \tag{6}$$

$$\mathbf{s}_t = \mathbf{o}_t \odot \tanh(\mathbf{c}_t) \tag{7}$$

where in the time slot t the input vector is $\mathbf{x}_t$, the hidden state vector is $\mathbf{s}_t$ and the memory cell is denoted as $\mathbf{c}_t$. $\odot$ represents the Hadamard product. $\mathbf{W}_x^i$, $\mathbf{W}_s^i$, $\mathbf{W}_x^f$, $\mathbf{W}_s^f$, $\mathbf{W}_x^o$, $\mathbf{W}_s^o$, $\mathbf{W}_x^c$, $\mathbf{W}_s^c$ are the weights matrices and similarly $\mathbf{b}_i$, $\mathbf{b}_f$, $\mathbf{b}_o$, $\mathbf{b}_c$ are bias vectors.

C. EVALUATION METRICS

It is known that to predict the value of channel quality, the usual perspective is regression. But spectrum prediction can be realized not only by regression but also by classification aspects if measured data are quantized ahead. Accordingly from these two aspects, the evaluation metrics of network performance are root of mean square error (RMSE) and classification accuracy (CA), both of which are mathematically formulated as equation (8) and (9):

$$RMSE = -10 * lg \sqrt{\frac{1}{N_{test}} \sum_{t=1}^{N_{test}} \left(\hat{L}_t - L_t\right)^2} \tag{8}$$

$$CA = \frac{1}{N_{test}} \sum_{t=1}^{N_{test}} flag_t$$

$$flag_t = \begin{cases} 1 & \hat{Q}_t = Q_t \\ 0 & \hat{Q}_t \neq Q_t \end{cases} \tag{9}$$

where N_{test} is the number of testing samples. Q_t and $\hat{Q}_t$ represents the real and predicted category, respectively.

The outputs of neural network should be consecutive values in regression. Therefore, the activation in the output layer is set to be *linear* and network parameters are adjusted to optimize the target function mean square error. In contrast, the outputs of neural network should be discrete category in classification. In addition the labels of samples must be converted to categorical one-hot encoding. The activation in the output layer is set to be *softmax* which is able to transform outputs to probability distribution. Cross-entropy function is often selected as the optimization target for classification task. The expression of cross-entropy is mathematically formulated as:

$$cross\_entropy = -\frac{1}{n} \sum_t y(t) \cdot \log(\hat{y}(t)) \tag{10}$$

where n is the number of training samples. $y(t)$ and $\hat{y}(t)$ represent the probability distribution of true categories and predicted categories, respectively.

IV. TAGUCHI METHOD-BASED HYPER-PARAMETER OPTIMIZATION

As is known, design of the neural network is a quiet challenging task because the performance of the network is influenced

by many hyper-parameters, mainly relying on the rich experience of the engineers. When designing optimized structure of neural network, the hyper-parameters that are considered and their corresponding levels are as follows and are presented in Table 1.

TABLE 1. Design hyper-parameters for optimizing the model structure.

Design Hyper-parameters		Level			
		1	2	3	4
i	Neurons in hidden layers	10	20	30	40
ii	Number of hidden layer	1	2	3	4
iii	Initializer	Uniform	Glorot uniform	Normal	Glorot normal
iv	Activation	Tanh	Sigmoid	Relu	Hard sigmoid
v	Learning rate	0.001	0.005	0.01	0.05

1) DESIGN HYPER-PARAMETER *i*

In terms of the neural network size, the number of neurons in the hidden layer has a great impact on performance in addition to the number of layers. We consider network width of 10, 20, 30, 40 as Levels 1, 2, 3, 4 respectively for all hidden layers.

2) DESIGN HYPER-PARAMETER *ii*

The second design hyper-parameter we consider is the number of hidden layers, which is also very important. The deep architecture of hidden layers is set as 1, 2, 3 and 4. The deeper the network is, and the stronger ability to model the network has. But it may result in over-fitting if the network has too many layers.

3) DESIGN HYPER-PARAMETER *iii*

The initial values of neural network parameters have great influence on network performance. An inappropriate initial value may result in failure to achieve optimal results. So we look at four methods of initialization: uniform distribution, glorot uniform distribution, normal distribution and glorot normal distribution.

4) DESIGN HYPER-PARAMETER *iv*

Activation function which is called *relu* has been widely used due to its outstanding behavior in deep learning. However, *sigmoid* and *tanh* functions which are commonly used may be more effective for different problems. *Tanh*, *sigmoid*, *relu* and *hard sigmoid* are considered as Levels 1, 2, 3 and 4.

5) DESIGN HYPER-PARAMETER *v*

Learning rate decides the span at which the parameters of networks are updated. If learning rate is set too small, then more iterations should be carried out until convergence.

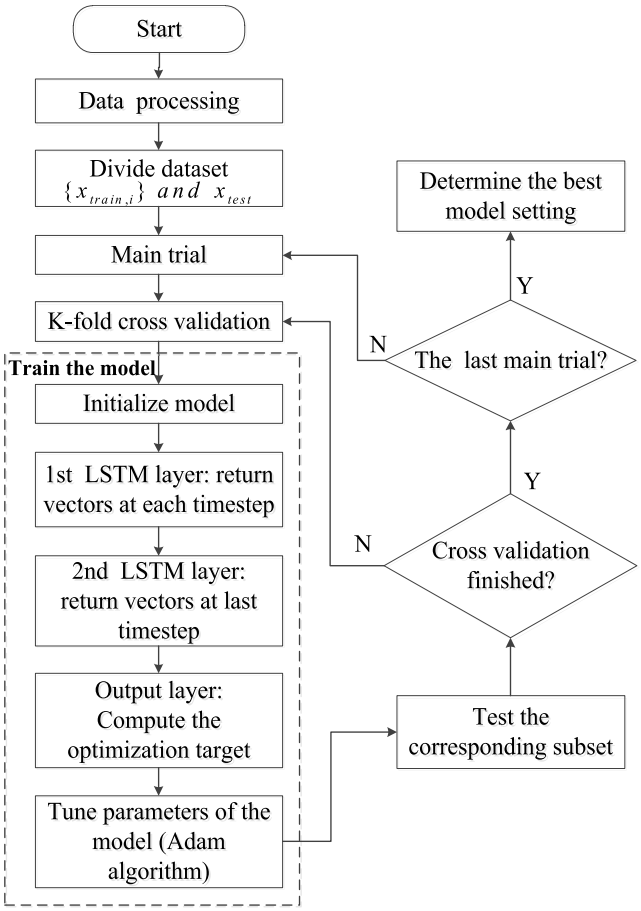

FIGURE 5. Process of taguchi method combined with cross validation.

But being too large will result in missing of the optimal parameters. We consider four levels: 0.001, 0.005, 0.01, 0.05.

These involved hyper-parameters have a combined effect on the performance of neural network and cannot be considered separately. Grid search method is often used to select the optimized structure. It is a kind of violent search method and lists all the possible combinations of the considering hyper-parameters. The training set is divided into K subsets which are denoted as $x_{train,i}$, $i = 1, 2, \ldots, K$ at first. Then, for each possible combination of the hyper-parameters, one certain subset is used to test and evaluate this model while others are used for training. Finally, the mean value of K testing performances is considered as evaluation performance of network with this combination. This is what we call the K-fold cross validation, which can find the relatively superior hyper-parameter settings of the network.

Five design hyper-parameters which can influence the performance of network are set as four levels respectively in this paper. Totally, there are 1024 (4^5) possible combinations to be analyzed if we use grid search method. It is conceivable that so many experiments will consume much time and require more computational resources. The taguchi method [34] is introduced to find the optimized structure through the orthogonal combination of influencing

TABLE 2. Main orthogonal array L_{16} trial and experimental results.

Main Trial	Design Hyper-parameters					$x_{train,0}$ CA	$x_{train,1}$ CA	$x_{train,2}$ CA	Mean Value
	i	ii	iii	iv	v				
1	1	1	1	1	1	0.7133	0.6997	0.7273	0.7134
2	**1**	**2**	**2**	**2**	**2**	**0.7183**	**0.6997**	**0.7236**	**0.7139**
3	1	3	3	3	3	0.7065	0.6892	0.7267	0.7075
4	1	4	4	4	4	0.7065	0.6898	0.7106	0.7023
5	2	1	2	3	4	0.7065	0.6898	0.7106	0.7023
6	2	2	1	4	3	0.7065	0.7040	0.7106	0.7070
7	2	3	4	1	2	0.7071	0.7022	0.7223	0.7105
8	2	4	3	2	1	0.7065	0.6898	0.7106	0.7023
9	3	1	3	4	2	0.7133	0.7040	0.7217	0.7130
10	3	2	4	3	1	0.7096	0.6997	0.7236	0.7110
11	3	3	1	2	4	0.7065	0.6898	0.7106	0.7023
12	3	4	2	1	3	0.7065	0.6898	0.7106	0.7023
13	4	1	4	2	3	0.7133	0.7022	0.7205	0.7120
14	4	2	3	1	4	0.7065	0.6898	0.7174	0.7046
15	4	3	2	4	1	0.7090	0.6898	0.7223	0.7070
16	4	4	1	3	2	0.7158	0.6898	0.7106	0.7054

hyper-parameters, saving a lot of time. It is an orthogonal experimental design method, which originates from quality control method in the engineering field. When evaluation metric is influenced by c design factors and each design factor has b levels, the experimental arrangements can be represent by the orthogonal table $L_a(b^c)$, whose subscript a means the quantity of trials. Actually, $L_a(b^c)$ is a table with a rows and c columns, where all levels of each design factor for each column appear at the same frequency. Meanwhile, occurrence frequency of a certain level between any two columns is the same. These two characteristics can be found in Table 2. It skillfully uses the orthogonal experimental arrangements to deal with the non-linear relation between design factors and evaluation metric, and determines the optimized combination of these factors by selecting the level of design factors. An orthogonal array $L_{16}(4^5)$ is used for the trial design combining the cross validation, listed in Table 2. The whole process of taguchi method combined with cross validation is shown in Fig. 5.

V. EXPERIMENT EVALUATION

In this section, we use two real-world spectrum datasets to demonstrate the effectiveness of the proposed scheme. One dataset is from the terrestrial networks and the other is from the satellite networks.

A. CASE STUDY I: TERRESTRIAL DATA ANALYTICS

In this case, we try to make predictions from the perspective of classification. We use spectrum data from the RWTH Aachen University spectrum measurement campaign [35].[2] After analysis in our previous work [36], the correlation of time domain in GSM1800 downlink band is strong. Therefore, spectrum data in this frequency band which ranges from 1820.0MHz to 1875.5MHz serve as the dataset in this part. Fig. 6 shows the evolution trajectories of a thirteen day RSS, that is, measured PSD values [37].

The total amount of original dataset is so large that it will consume long time to fully process the data. Based on this consideration, the original data are sampled in the same interval. The new time series achieved by sampling should be tested for stationarity. Stationarity test is necessary due to that supposing a time series is stationary, it is easier for us to model and therefore to predict. If the series doesn't satisfy the requirements of stationarity, differential operations must be included in data processing. The augmented dickey-fuller (ADF) test is employed here which belongs to root test methods [38]. When p value is above 0.05, null hypothesis test is received. So, the time series has unit root and it is non-stationary. Conversely, the alternate hypothesis test is accepted, indicating that the time series is stationary. After our analysis, the evolution trajectory of RSS for each spectrum point in the GSM1800 downlink band is stationary and that's what we expected. Through quantization the continuous spectrum data are transformed into finite discrete values, that is to say, quantization interval is determined

[2]In the original datasets [35], the resolution bandwidth of each individual spectrum band is 200 kHz and the inter-sample time is 1.8 seconds, which results in 48000 samples one day [36].

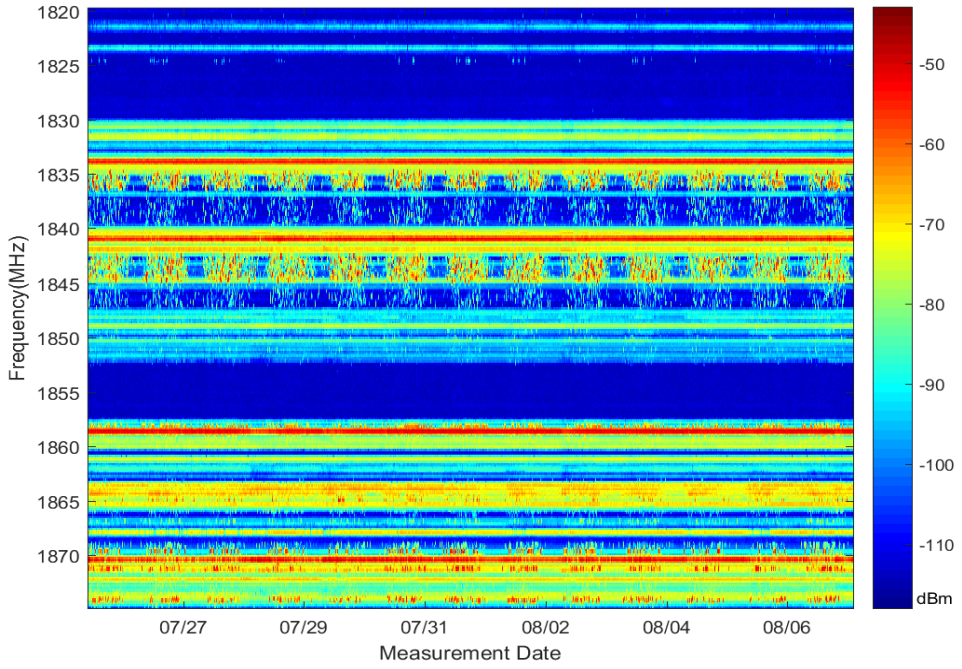

FIGURE 6. Original spectrum data in case I.

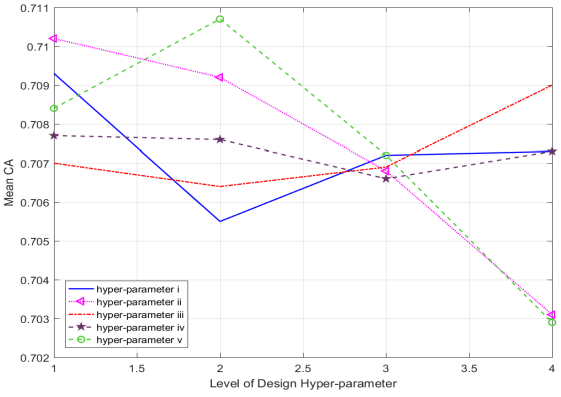

FIGURE 7. Effects of each design hyper-parameter at each four levels.

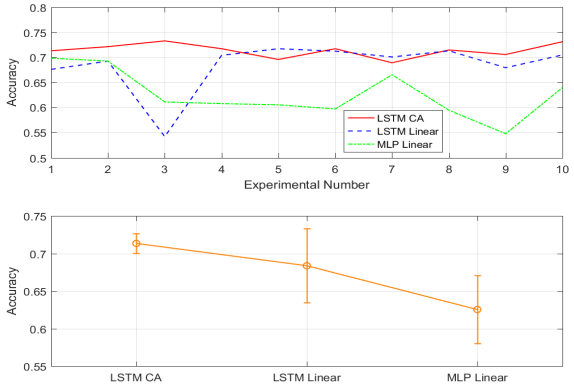

FIGURE 8. Stability of the three networks.

according to the maximum and minimum values of the measured data as well as quantization level which is chosen as 8. The reason why quantization level equals 8 is a tradeoff between prediction accuracy and prediction error. Too large quantization level will increase prediction difficulty and lead to poor prediction accuracy due to quantized spectrum values are close to each other, while too small quantization level may bring out the large discrepancy between quantized value and measured value, which means loss of spectrum information.

Next, we should determine the optimized structure of network. Taking the spectrum point of 1863.6MHz from the perspective of classification as an example, the training set is split into 3 subsets $x_{train,0}$, $x_{train,1}$, $x_{train,2}$ at first. For each main trial in Table 2, 3-fold cross validation is carried out and the corresponding prediction accuracy is listed in the same table. Because the model is a classification one, only the classification accuracy is shown in the table. The last column lists the mean accuracy of cross validation for each trial. We can see that the top two of accuracy are achieved by the 1st and 2nd trial, whose mean accuracy are 0.7139 and 0.7134. The setting of all considered hyper-parameters is in bold when neural network achieves the best performance and as follows: the neurons of the hidden layers are 10, the number of layers is two, learning rate should be 0.005, activation function is chosen *sigmoid* and the initialization of parameters would better to obey glorot uniform distribution. Finally, the accuracy of testing set achieves 0.7145 with the optimized hyper-parameters setting.

Next, we analyze what effect single design hyper-parameter may have on the prediction performance, still using

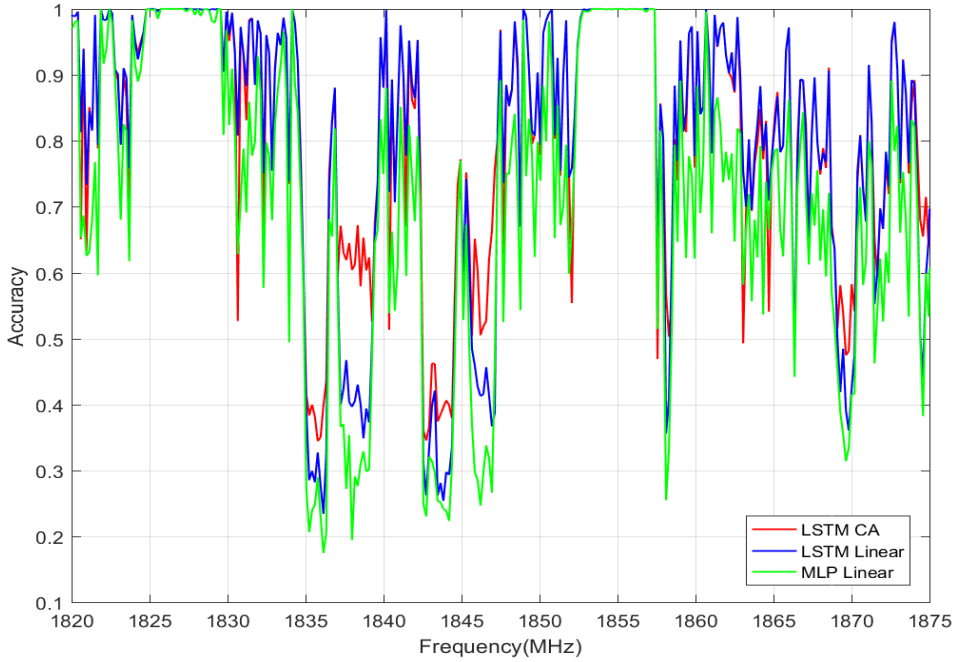

FIGURE 9. Accuracy performance of three networks.

experimental results above. The effects of each design hyper-parameter can be separated since for each trial the combinations of design hyper-parameters are orthogonal [39]. The effects of each design hyper-parameter at each level are calculated by taking the corresponding average from Table 2. For example, the Level 3 of the design hyper-parameter *iii* is in the 3th, 8th, 9th and 14th main trials, and the average accuracy of this level is 0.7069. Here we introduced range analysis [40] to find the sensitivity of the design hyper-parameters for the spectrum point. Sensitivity of one design hyper-parameter means the difference between the maximum mean accuracy and the minimum mean accuracy, and the results are shown in Table 3. The effects of each design hyper-parameter at each of the four levels are displayed in Fig. 7. We find that among these design hyper-parameters, the learning rate has the most important influence on the prediction performance. What's more, the depth of network has more effects on the network performance than the width of network. Both of initializer and activation function have few influences on prediction performance of single spectrum point. So when we design LSTM network for spectrum prediction of all spectrum points in the GSM1800 downlink band, more attention is paid to learning rate and the depth of neural network. The depth of neural network should be one or two layers and the learning rate should be less than 0.005.

Due to the different parameters of neural network during initialization, the results of each experiment will be different. We further compare the stability of three neural networks. LSTM CA model represents the LSTM network with classification perspective and LSTM Linear model represents network from the regression aspect. MLP Linear model is a traditional one as a baseline. The configuration of three models is according to the optimized combination of those design hyper-parameters, and total ten experiments are conducted at the above spectrum point. The experimental and statistical results are shown in Fig. 8, showing results of each experiment and statistical stability respectively. As can be seen from the figure, the prediction performance of two LSTM models is higher and more stable than MLP model.

As mentioned above, RSS varies from spectrum point to spectrum point. Some spectrum points do not have very obvious tidal effects while some spectrum points are always stationary with PSD value essentially unchanged. It can be assumed that the prediction performance of each spectrum point may not be consistent. Fig. 9 shows the accuracy performance of three models on the same testing set in the GSM1800 downlink bands. For total spectrum points, a little small portion of prediction performance of the MLP network is superior to the LSTM network, which is possible due to

TABLE 3. Influence of each design hyper-parameter and sensitivity.

Design Hyper-parameters	Levels				Sensitivity
	1	2	3	4	
i	0.7093	0.7055	0.7072	0.7073	0.38%
ii	0.7102	0.7091	0.7068	0.7031	0.71%
iii	0.7070	0.7064	0.7069	0.7090	0.26%
iv	0.7077	0.7076	0.7066	0.7073	0.11%
v	0.7084	0.7107	0.7072	0.7029	0.78%

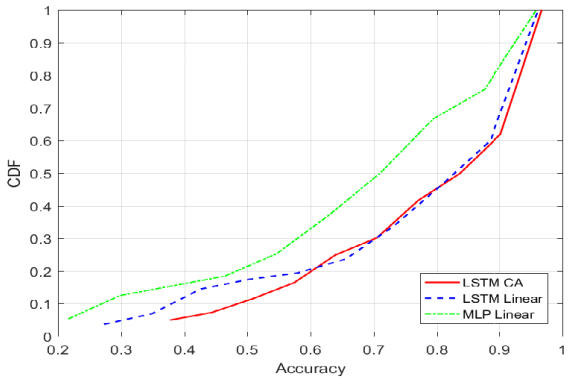

FIGURE 10. CDF of accuracy performance of three networks.

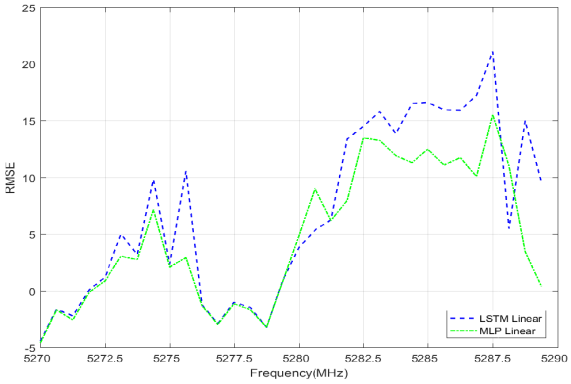

FIGURE 11. RMSE performance of two networks.

different initialization values or optimization targets. In some spectrum points, the performance difference is large while in others the performance is close to each other. It is related to the features of RSS which can be further studied. In the sense of statistics, Fig. 10 shows the cumulative distribution function (CDF) curves of CA performance. We can also conclude that LSTM network with classification aspect is a little better than that with regression aspect.

B. CASE STUDY II: SATELLITE DATA ANALYTICS

In this case, we're going to do the spectrum prediction from a regression aspect. The spectrum data used in this part of experiments are not from open source and are collected from satellite. There are in total 86 frequency bins whose bandwidth is all 20MHz and each frequency bin is split into 32 spectrum points. Satellite measurement campaign starts on 2017-12-01 and ends on 2017-12-11, which can sweep whole frequency band in only 0.8s. Due to the measuring equipment, missing of spectrum data in the whole band occurs in some time periods, so the average value of one spectrum point is filled into those blank slots of that point.

The spectrum data are predicted from the perspective of regression, therefore the evaluation metric of network performance is equation (8). The RMSE function in this paper is monotonically decreasing. In other words, the larger the prediction error of certain frequency point is, the smaller the corresponding RMSE value is.

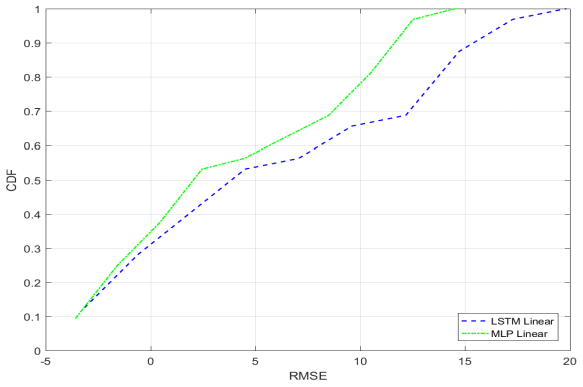

FIGURE 12. CDF of rmse performance of two networks.

Here we employ LSTM Linear and MLP Linear network to make prediction from 5270MHz to 5290MHz. Fig. 11 shows the prediction performance of these 32 frequency points and Fig. 12 shows the corresponding CDF curve. As shown in Fig. 11, the performance advantage of LSTM Linear network is obvious at some frequency points, while the prediction performance at other frequency points is similar to that of MLP Linear network or even worse. As with case study I, the performance difference of these two networks is related to the using rule of frequency point and initialization value of network. In Fig. 12, RMSE of LSTM Linear network is less than 15.2 for about 90% of the frequency points, and that of MLP Linear network is less than 11.8. On this dataset from the perspective of regression, we further verify the superiority of the LSTM network.

VI. CONCLUSION

In this paper, we develop a spectrum prediction framework with a deep learning approach. We predict real-world spectrum data through the emerging LSTM neural network and the conventional MLP neural network. It is widely known that determining the architecture of a neural network is a very complex problem. Here, we introduce taguchi method instead of grid search method to reduce time consumption and computational resources when we design the best optimized configuration of neural network. For the first dataset, the LSTM network is constructed from classification and regression aspects to evaluate prediction accuracy. For the second dataset, the LSTM network is constructed from regression to evaluate prediction error. The experimental results have demonstrated that LSTM network has some advantages in time series problems and better prediction performance than MLP network. For both LSTM network, network with classification aspect performs slightly better than the one with regression aspect.

REFERENCES

[1] "Shared spectrum company general survey of radio frequency bands (30 Hz to 3 GHz)," Vienna, VA, USA, Tech. Rep. 20100323, Sep. 2009. [Online]. Available: http://www.sharedspectrum.com/papers/spectrum-reports/

[2] J. Mitola, III, "Cognitive radio for flexible mobile multimedia communications," *Mobile Netw. Appl.*, vol. 6, no. 5, pp. 435–441, Sep. 2006.
[3] Q. Wu, G. Ding, J. Wang, and Y. D. Yao, "Spatial-temporal opportunity detection for spectrum-heterogeneous cognitive radio networks: Two-dimensional sensing," *IEEE Trans. Wireless Commun.*, vol. 12, no. 2, pp. 516–526, Feb. 2013.
[4] G. Ding, Q. Wu, Y.-D. Yao, J. Wang, and Y. Chen, "Kernel-based learning for statistical signal processing in cognitive radio networks: Theoretical foundations, example applications, and future directions," *IEEE Signal Process. Mag.*, vol. 30, no. 4, pp. 126–136, Jul. 2013.
[5] V. Jamali, A. Ahmadzadeh, and R. Schober, "On the design of matched filters for molecule counting receivers," *IEEE Commun. Lett.*, vol. 21, no. 8, pp. 1711–1714, Aug. 2017.
[6] H. Urkowitz, "Energy detection of unknown deterministic signals," *Proc. IEEE*, vol. 55, no. 4, pp. 523–531, Apr. 1967.
[7] X. F. Zhang, L. Y. Xu, L. Xu, and D. Z. Xu, "Direction of departure (DOD) and direction of arrival (DOA) estimation in MIMO radar with reduced-dimension MUSIC," *IEEE Commun. Lett.*, vol. 14, no. 12, pp. 1161–1163, Dec. 2010.
[8] G. Ding *et al.*, "On the limits of predictability in real-world radio spectrum state dynamics: From entropy theory to 5G spectrum sharing," *IEEE Commun. Mag.*, vol. 53, no. 7, pp. 178–183, Jul. 2015.
[9] X. Xing, T. Jing, Y. Huo, H. Li, and X. Cheng, "Channel quality prediction based on Bayesian inference in cognitive radio networks," in *Proc. IEEE INFOCOM*, Turin, Italy, Apr. 2013, pp. 1465–1473.
[10] Z. Zhang, J. Shi, H. H. Chen, M. Guizani, and P. Qiu, "A cooperation strategy based on Nash bargaining solution in cooperative relay networks," *IEEE Trans. Veh. Technol.*, vol. 57, no. 4, pp. 2570–2577, Jul. 2008.
[11] G. Zhu, C. Zhong, H. A. Suraweera, G. K. Karagiannidis, Z. Zhang, and T. A. Tsiftsis, "Wireless information and power transfer in relay systems with multiple antennas and interference," *IEEE Trans. Commun.*, vol. 63, no. 4, pp. 1400–1418, Apr. 2015.
[12] Z. Feng, C. Qiu, Z. Feng, Z. Wei, W. Li, and P. Zhang, "An effective approach to 5G: Wireless network virtualization," *IEEE Commun. Mag.*, vol. 53, no. 12, pp. 53–59, Dec. 2015.
[13] G. Ding *et al.*, "Spectrum inference in cognitive radio networks: Algorithms and applications," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 1, pp. 150–182, 1st Quart., 2018.
[14] L. Wang, T. Liu, G. Wang, K. L. Chan, and Q. Yang, "Video tracking using learned hierarchical features," *IEEE Trans. Image Process.*, vol. 24, no. 4, pp. 1424–1435, Apr. 2015.
[15] Y. Xiang, Q. Chen, X. Wang, and Y. Qin, "Answer selection in community question answering via attentive neural networks," *IEEE Signal Process. Lett.*, vol. 24, no. 4, pp. 505–509, Apr. 2017.
[16] T. O'Shea and J. Hoydis, "An introduction to deep learning for the physical layer," *IEEE Trans. Cogn. Commun. Netw.*, vol. 3, no. 4, pp. 563–575, Dec. 2017.
[17] J. Li, H. Bu, and J. Wu, "Sentiment-aware stock market prediction: A deep learning method," in *Proc. IEEE Int. Conf. Syst. Service Manage.*, Dalian, China, Jun. 2017, pp. 1–6.
[18] Y. Tang, J. Xu, K. Matsumoto, and C. Ono, "Sequence-to-sequence model with attention for time series classification," in *Proc. IEEE 16th Int. Conf. Data Mining Workshops (ICDMW)*, Barcelona, Spain, Dec. 2016, pp. 503–510.
[19] Z. Wang and S. Salous, "Spectrum occupancy statistics and time series models for cognitive radio," *J. Signal Process. Syst.*, vol. 62, no. 2, pp. 145–155, Feb. 2011.
[20] C. J. Yu, Y. Y. He, and T. F. Quan, "Frequency spectrum prediction method based on EMD and SVR," in *Proc. 8th Int. Conf. Intell. Syst. Design Appl.*, Kaohsiung, Taiwan, Nov. 2008, pp. 39–44.
[21] Y. Hua, Z. Zhao, R. Li, X. Chen, Z. Liu, and H. Zhang. (2017). "Traffic prediction based on random connectivity in deep learning with long short-term memory." [Online]. Available: https://arxiv.org/abs/1711.02833
[22] P. Huang, C.-J. Liu, X. Yang, L. Xiao, and J. Chen, "Wireless spectrum occupancy prediction based on partial periodic pattern mining," *IEEE Trans. Parallel Distrib. Syst.*, vol. 25, no. 7, pp. 1925–1934, Jul. 2014.
[23] F. Azmat, Y. Chen, and N. Stocks, "Analysis of spectrum occupancy using machine learning algorithms," *IEEE Trans. Veh. Technol.*, vol. 65, no. 9, pp. 6853–6860, Sep. 2016.
[24] V. K. Tumuluru, P. Wang, and D. Niyato, "Channel status prediction for cognitive radio networks," *Wireless Commun. Mobile Comput.*, vol. 12, no. 10, pp. 862–874, Jul. 2012.
[25] C. Zhang, P. Patras, and H. Haddadi. (2018). "Deep learning in mobile and wireless networking: A survey." [Online]. Available: https://arxiv.org/abs/1803.04311
[26] Y. He *et al.*, "Deep-reinforcement-learning-based optimization for cache-enabled opportunistic interference alignment wireless networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 11, pp. 10433–10445, Nov. 2017.
[27] Y. He, F. Richard Yu, N. Zhao, V. C. M. Leung, and H. Yin, "Software-defined networks with mobile edge computing and caching for smart cities: A big data deep reinforcement learning approach," *IEEE Comm. Mag.*, vol. 55, no. 12, pp. 31–37, Dec. 2017.
[28] T. Wang, C.-K. Wen, H. Wang, F. Gao, T. Jiang, and S. Jin, "Deep learning for wireless physical layer: Opportunities and challenges," *China Commun.*, vol. 14, no. 11, pp. 92–111, 2017.
[29] L. Yu, Q. Wang, Y. Guo, and P. Li, "Spectrum availability prediction in cognitive aerospace communications: A deep learning perspective," in *Proc. IEEE Cogn. Commun. Aerosp. Appl. Workshop (CCAA)*, Cleveland, OH, USA, Jun. 2017, pp. 1–4.
[30] L. Yu, J. Chen, and G. Ding, "Spectrum prediction via long short term memory," in *Proc. 3rd IEEE Int. Conf. Comput. Commun. (ICCC)*, Chengdu, China, Dec. 2017, pp. 643–647.
[31] D. C. Plaut, S. J. Nowlan, and G. E. Hinton, "Experiments on learning by back propagation," Carnegie Mellon Univ., Pittsburgh, PA, USA, Tech. Rep. CMU-CS-86-126, Jun. 1986.
[32] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.
[33] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
[34] G. Taguchi and Y. Yokoyama, *Taguchi Methods: Design of Experiments(Taguchi Methods Series)*, vol. 4. Tokyo, Japan: American Supplier Institute, Nov. 1993.
[35] M. Wellens, "Empirical modelling of spectrum use and evaluation of adaptive spectrum sensing in dynamic spectrum access networks," Ph.D. dissertation, Inst. Netw. Syst., RWTH Aachen Univ., Aachen, Germany, May 2010.
[36] G. Ding *et al.*, "Robust online spectrum prediction with incomplete and corrupted historical observations," *IEEE Trans. Veh. Technol.*, vol. 66, no. 9, pp. 8022–8036, Sep. 2017.
[37] J. Sun, L. Shen, G. Ding, R. Li, and Q. Wu, "Predictability analysis of spectrum state evolution: Performance bounds and real-world data analytics," *IEEE Access*, vol. 5, pp. 22760–22774, Oct. 2017.
[38] S. E. Said and D. A. Dickey, "Testing for unit roots in autoregressive-moving average models of unknown order," *Biometrika*, vol. 71, no. 3, pp. 599–607, Dec. 1984.
[39] M. S. Phadke, *Quality Engineering Using Robust Design*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1995.
[40] H.-F. Yang, T. S. Dillon, and Y.-P. P. Chen, "Optimized structure of the traffic flow forecasting model with a deep learning approach," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2371–2381, Oct. 2017.

LING YU received the B.S. degree in information engineering from Southeast University, Nanjing, China, in 2016. She is currently pursuing the M.S. degree with the College of Communications Engineering, Army Engineering University of PLA, Nanjing. Her research interests include deep learning, wireless communications, and cognitive radio networks.

JIN CHEN received the B.S. degree in wireless communication and the M.S. and Ph.D. degrees in communications engineering and information system from the Institute of Communications Engineering, Nanjing, China, in 1993, 1996, and 1999, respectively. She is currently a Professor with the Institute of Communications Engineering, Army Engineering University of PLA, Nanjing. Her current research interests include wireless communications networks and cognitive radio networks.

GUORU DING (S'10–M'14–SM'16) received the B.S. degree (Hons.) in electrical engineering from Xidian University, Xi'an, China, in 2008, and the Ph.D. degree (Hons.) in communications and information systems from the College of Communications Engineering, Nanjing, China, in 2014. From 2014 to 2017, he was an Assistant Professor with the College of Communications Engineering and a Research Fellow with the National High Frequency Communications Research Center of China, where he is currently an Associate Professor. Since 2015, he has been a Post-Doctoral Research Associate with the National Mobile Communications Research Laboratory, Southeast University, Nanjing. His research interests include cognitive radio networks, massive MIMO, machine learning, and big data analytics over wireless networks.

Dr. Ding has acted as a Technical Program Committee (TPC) member for a number of international conferences, including the IEEE Global Communications Conference, the IEEE International Conference on Communications, and the IEEE Vehicular Technology Conference (VTC). He is a Voting Member of the IEEE 1900.6 Standard Association Working Group. He was a recipient of the Best Paper Award from the EAI MLICOM 2016, the IEEE VTC 2014, and the IEEE WCSP 2009. He received the Alexander von Humboldt Fellowship in 2017 and the Excellent Doctoral Thesis Award of the China Institute of Communications in 2016. He has served as a Guest Editor for the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS (Special issue on spectrum sharing and aggregation in future wireless networks). He is currently an Associate Editor of the *Journal of Communications and Information Networks*, the *KSII Transactions on Internet and Information Systems*, and the *AEU-International Journal of Electronics and Communications*.

YA TU received the B.S. degree from the College of Computer Science and Technology, Taiyuan University of Technology, Taiyuan, China, in 2016. He currently holds a doctoral position with the College of Information and Communication Engineering, Harbin Engineering University, Harbin, China. His current research interests include signal processing, machine learning, and data analysis.

JIAN YANG (S'14–M'17) received the B.S. degree in information countermeasure technology and the M.S. degree in communications and information systems from the Nanjing University of Science and Technology, Nanjing, China, in 2006 and 2010, respectively, the Ph.D. degree in information and communication engineering from the PLA University of Science and Technology, Nanjing, in 2017. Since 2017, he has been a Post-Doctoral Research Associate with the 63rd Institute, National University of Defense Technology, Nanjing. His research interests include cognitive radio networks, spectrum prediction, and intelligent spectrum management.

JIACHEN SUN received the B.S. degree in information engineering from Southeast University, Nanjing, China, in 2016. She is currently pursuing the M.S. degree with the College of Communications Engineering, Army Engineering University of PLA, Nanjing. Her research interests include data analytics, wireless communications, and cognitive radio networks.