IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Deep Learning Based Channel Prediction for Edge Computing Networks Towards Intelligent Connected Vehicles

GUANGQUN LIU[1], YAN XU[1], ZONGJIANG HE[1], YANYI RAO[1], JUNJUAN XIA[1], AND LISENG FAN[1]

[1]School of Computer Science and Cyber Engineering, Guangzhou University, Guangzhou 510006, China

Corresponding author: Junjuan Xia and Liseng Fan (e-mail: xiajunjuan@gzhu.edu.cn, lsfan2019@126.com).

This work was supported in part by the NSFC (No. 61871139), by the Innovation Team Project of Guangdong Province University (No. 2016KCXTD017), by the Science and Technology Program of Guangzhou (No. 201807010103), by Natural Science Foundation of Guangdong Province (No. 2017A030308006).

ABSTRACT With the development of intelligent connected vehicles (ICVs), there emerge many new services and applications which involve intensive computation. To support the intensive computation in vehicle-to-everything (V2X) communication system, the framework of edge computing networks has been proposed, which exploits the computation ability of edge nodes at the cost of wireless transmission. Hence, it is of vital importance to predict the wireless channel parameters, which can help schedule the system resource management and optimize the system performance in advance. To fulfil this challenge, this paper proposes a novel prediction model based on long short-term memory (LSTM) network, which is powerful in capturing valuable information in the sequence and hence is good at analyzing the spatio-temporal correlation in the channel parameters. To validate the proposed model, we conduct extensive simulations to show that the proposed model is quite effective in the channel prediction. In particular, the proposed model can outperform the conventional ones substantially.

INDEX TERMS Vehicular network, edge computing, channel prediction, LSTM network.

I. INTRODUCTION

In recent years, there has been a tremendous development in the wireless communication [1]–[3], and the wireless data rate has been increasing rapidly. Along with the data rate, there emerge many new wireless services, which requires intensive computation in a short time [4]–[6]. To meet the requirement of both communication and computation, the framework of mobile edge computing (MEC) has been proposed [7]–[9], where the edge nodes are exploited to assist the accomplishment of the computation task. In this area, many researchers have studied to optimize the MEC networks by taking into account the cost by the communication and computation, and pointed out that there should be a trade-off between the computation and communication. In particular, the offloading in the networks can help accomplish the computation task at the cost of wireless communication [10], [11].

The vehicle-to-everything (V2X) communication system is one of the most popular application scenarios of intelligent connected vehicles (ICVs). The typical applications include vehicle-to-vehicle (V2V), vehicle-to-pedestrian (V2P), vehicle-to-infrastructure (V2I), and vehicle-to-network (V2N) [12]. In these applications, the wireless transmission should provide high data rate services, since the nodes in the network will communicate with each other very frequently. Moreover, as the services involve both the communication and computing, the latency in the transmission and computing should be very small, especially in the application scenarios of mobile vehicles with high speed.

In wireless networks including the MEC networks, it is very important to estimate the accurate channel state information (CSI) and predict the CSI in advance, which can help schedule the system resource management and optimize the system performance [13]–[15]. So far, channel prediction has been intensively studied in the literature and researchers have proposed many channel prediction models, such as auto regressive integrated moving average (ARIMA), support-vector machine for regression (SVR). ARIMA is a category

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2935463, IEEE Access

of time series prediction model that figures out the inherent relationship among the data. However, one limitation of ARIMA is that it cannot capture the feature of a rapidly changing data sequence [16]. As one class of machine learning methods, SVR has disadvantages in high algorithmic complexity and the selection of the kernel function parameters. Therefore, we need to find an effective model to implement channel prediction for the MEC networks.

Due to the explosively increasing computational ability, the artificial intelligent (AI) algorithms have been applied in the wireless networks. Different from the conventional model-based algorithms, the AI algorithms are data-driven and require a vast number of data to adaptively capture the features inherented in the data. In this area, deep learning has been widely used in neural networks [17], [18], and the authors in [18] proposed an approach for combining the power of high-level spatio-temporal graphs and sequence learning success of recurrent neural networks (RNN). As a popular algorithm, Q-learning has widely applied in secure physical-layer communication [19]–[21]. The authors in [19] investigated a secure communication system under the smart attack from one unmanned aerial vehicle (UAV), and the simulation results indicated that this strategy could effectively decrease the attack rate of the UAV attacker and increase the system secrecy capacity, regardless of the channel estimation accuracy. Moreover, convolutional neural networks (CNN) and long short-term memory (LSTM) have been used in computer vision, natural language processing and pattern recognition in recent years [22], [23]. In particular, the existing works have demonstrated that the deep CNN and LSTM based networks can explore the local correlation among data very efficiently.

In this paper, we propose a novel channel prediction model based on deep learning approach. Specifically, we apply the LSTM based networks to build an efficient channel prediction model to predict the future channel parameters, based on the past and current channel parameters. In order to validate the efficiency of the proposed model under Raleigh fading channels, we carry out massive simulations to test the model performance. Simulation results are demonstrated to show that the proposed deep learning based model outperforms the conventional models such as ARIMA and SVR ones. The results in this paper can be extended to other fading channels, such as Nakagami-m and Ricean channels.

The paper is organized as follows. Firstly, we introduce the knowledge about Rayleigh fading channel and the structure of LSTM in Section II. Then, we describe the proposed model in detail in Section III, including the architecture of the learning framework and the prediction scheme. In further, we present the method of data preprocessing and the discussions about the simulation results in Section IV. Finally, conclusions are given in Section V.

II. SYSTEM MODEL

A typical V2X communication system is shown in Fig. 1, where there are multiple vehicles which are willing to exchange information with each other. To support the ultra-reliable and low-latency communication among vehicles, channel prediction should be applied to help manage the system resource in advance. The channel prediction in the MEC networks is shown in Fig. 2, where the task of channel prediction is to exploit the historical channel information for training a model, which can be then used to predict the future channel information.

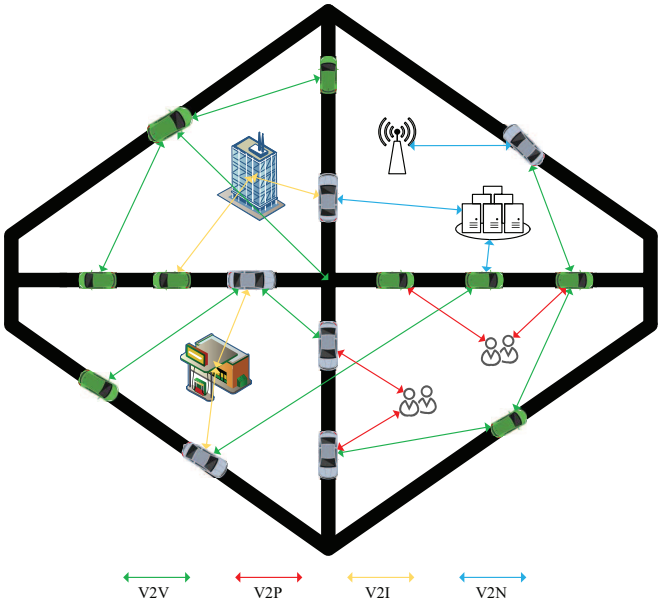


FIGURE 1: A typical V2X communication system.

A. RAYLEIGH FADING CHANNELS

In wireless communication systems, the data is transmitted through wireless channels. Without loss of generality, we select the Rayleigh fading channels in this paper, which is a reasonable model when there are many scatters in the environment that reflect the radio signal before it arrives at the receiver [24]–[26]. The envelope of the channel response is subject to the Rayleigh distributed, and the mutually uncorrelated Rayleigh channel information in the improved Jake's model can be described as,

$$H(t) = H_I(t) + jH_Q(t). \tag{1}$$

The CSI of the channel is complex-valued, and the associated complex valued CSI of $H(t)$ can be divided into real part H_I and imaginary part H_Q, which are given by the following equations,

$$H_I(t) = \frac{1}{\sqrt{N}} \sum_{m=1}^{N} \cos(2\pi f_d \cos a_m t + a_m), \tag{2}$$

$$H_Q(t) = \frac{1}{\sqrt{N}} \sum_{m=1}^{N} \sin(2\pi f_d \cos a_m t + b_m), \tag{3}$$

with

$$f_d = (v/c) f_c \cos a_m, \tag{4}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2935463, IEEE Access

where N is the number of scatters, v is the moving speed of the user, c is the speed of light, f_d is the Doppler frequency shift, and f_c is the carrier frequency. Notation a_m and b_m represent the in-phase and orthogonal-phase angles of signal arrival, which are uniformly distributed over the interval $[0, 2\pi]$. At the k-th symbols, the associated channel parameter $H(kT_s)$ at the discretely sampled time kT_s can be expressed by

$$H(kT_s) = H_I(kT_s) + jH_Q(kT_s), \tag{5}$$

where $H_I(kT_s)$ and $H_Q(kT_s)$ are the real part and imaginary part of $H(kT_s)$, respectively.

B. OVERVIEW OF LSTM NETWORKS

LSTM is a category of recurrent neural networks that can remember features among the data over arbitrary intervals. The LSTM was proposed to solve vanishing gradient problem on RNN by explicitly introducing a memory unit, which is shown in Fig. 3. Specifically, the LSTM network is composed of several LSTM units, where each unit has an input gate, a forget gate, an output gate, and a memory cell. The LSTM can be used to create large recurrent networks that in turn can help address difficult sequential problems in machine learning and achieve state-of-the-art results. The principle of LSTM is made up of forward components and backward components. As to the forward components, the gates are defined by

$$a_t = \phi(W_a x_t + U_a c_{t-1} + b_a), \tag{6}$$
$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i), \tag{7}$$
$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f), \tag{8}$$
$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o), \tag{9}$$

where a_t is the input activation vector, i_t is the input gate's activation vector, f_t is the forget gate's activation vector, o_t is the output gate's activation vector, and x_t is the input vector. The subscript t denotes the time step. The hidden state vector and memory cell state vector are written as,

$$s_t = o_t \odot \phi(c_t), \tag{10}$$
$$c_t = f_t \odot c_{t-1} + i_t \phi(W_c x_t + U_c h_{t-1} + b_c). \tag{11}$$

where the notation $\odot$ denotes the element-wise product. From the above equations, we can write the following matrices,

$$S_t = \begin{bmatrix} a_t \\ i_t \\ f_t \\ o_t \end{bmatrix}, W = \begin{bmatrix} W_a \\ W_i \\ W_f \\ W_o \end{bmatrix}, U = \begin{bmatrix} U_a \\ U_i \\ U_f \\ U_o \end{bmatrix}, b = \begin{bmatrix} b_a \\ b_i \\ b_f \\ b_o \end{bmatrix}, \tag{12}$$

where $W \in \mathbb{R}^{h \times d}$, $U \in \mathbb{R}^{h \times h}$ are the weight matrices, $b \in \mathbb{R}^h$ is the bias vector on the network, and d and h represent the number of input features and number of hidden units, respectively. The functions $\sigma(\cdot)$ and $\phi(\cdot)$ denote the sigmoid activation function and the tanh activation function,

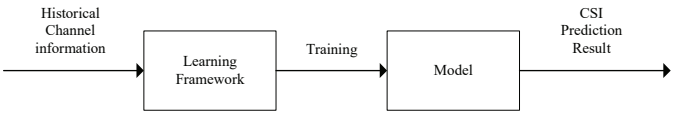


FIGURE 2: CSI prediction for the MEC networks.

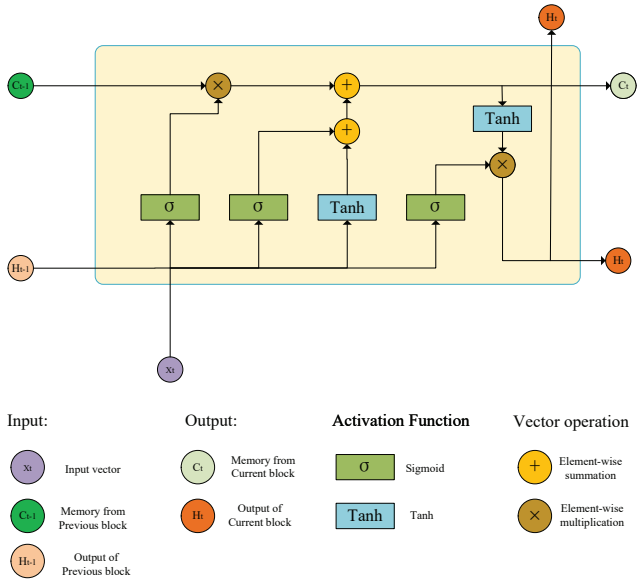


FIGURE 3: The structure of LSTM networks.

respectively. The initial values are set with $c_0 = 0$ and $h_0 = 0$.

As to the backward components, we have:

$$\delta c_t = \Delta t + \Delta c_t, \tag{13}$$
$$\delta s_t = \delta c_t \odot o_t \odot (1 - \phi^2(s_t)) + \delta s_{t+1} \odot f_{t+1}, \tag{14}$$
$$\delta a_t = \delta s_t \odot i_t \odot (1 - a_t^2), \tag{15}$$
$$\delta i_t = \delta s_t \odot h_{t-1} \odot f_t \odot (1 - f_t), \tag{16}$$
$$\delta o_t = c_t \odot \phi(s_t) \odot o_t \odot (1 - o_t), \tag{17}$$
$$x_t = W^T \times \delta S_t, \tag{18}$$
$$\Delta c_{t-1} = U^T \times \delta S_t, \tag{19}$$

where δ denotes the difference between the current and the next layers. The internal parameters are finally updated as,

$$\delta W = \sum_{t=o}^{T} \Delta S_t \otimes x_t, \tag{20}$$
$$\delta U = \sum_{t=o}^{T-1} \Delta S_{t+1} \otimes c_t, \tag{21}$$
$$\delta b = \sum_{t=0}^{T} \Delta S_{t+1}. \tag{22}$$

Through the cooperation between memory cells and gates, the LSTM has a remarkable ability on time series analysis which provides a powerful ability in exploiting the long-term dependence to predict the time series data.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2935463, IEEE Access

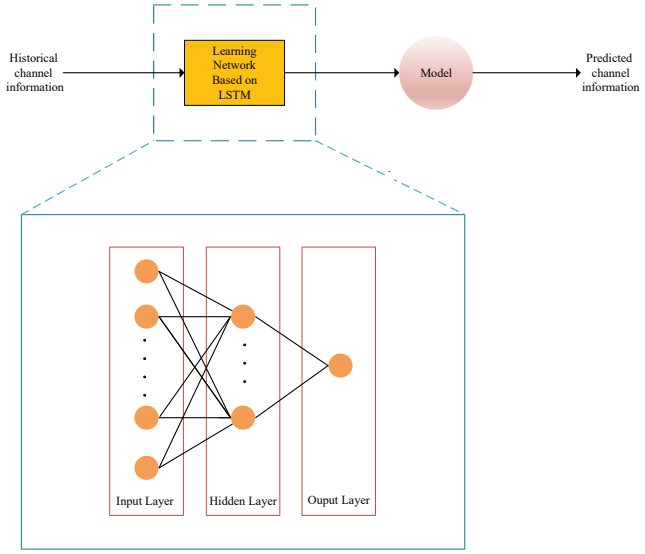


FIGURE 4: **The framework of the learning-based channel prediction.**

III. LSTM BASED ON CHANNEL PREDICTION

In this section, we describe the learning framework of the channel prediction model and the prediction scheme.

A. ARCHITECTURE OF LEARNING FRAMEWORK

We design the learning framework to train the channel prediction model by analyzing the historical channel information, where the architecture of the learning framework is shown in Fig. 4. Specifically, the framework consists of an LSTM network layer and 1-dimensional dense layer. The LSTM network is used for state vector prediction. The dense layer performs the operation: $o = \odot((input, kernel) + bias)$, where $\odot$ is the element-wise activation function passed as the activation argument, the kernel is a weight matrix created by the layer, and the bias is a bias vector created by the layer.

B. PREDICTION SCHEME

1) Single-sample prediction

The historical CSI will be divided into three parts: training, validation and test datasets. The proportions of the three datasets are 0.80, 0.10 and 0.10, respectively. We use the first n samples of the CSI as the input samples and employ the $(n+1)$-th of the channel information at the $(n+1)$-th sample as the output sample, which is also regarded as the target value. We construct a training set in this way. The dimension of the input layer is n, and the dimension of the output layer is equal to 1. The data on the input layer can be represented as,

$$x_k = [H(kT_s), H((k+1)T_s), ..., H(k+d-1)T_s], \quad (23)$$
$$y_k = H((k+d)T_s), \quad (24)$$

where the data pair (x_k, y_k) is the training dataset at the k-th sample. The historical CSI will be constructed as this way to train the network. The size of the training set is set to N, and hence the training set (X_{tra} and Y_{tra}) are represented as,

$$X_{tra} = [x_1, x_2, ..., x_N], \quad (25)$$
$$Y_{tra} = [y_1, y_2, ..., y_N]. \quad (26)$$

Moreover, in order to avoid the overfitting in the training process, we set a proportion of historical CSI as the validation dataset. The sample of the validation set is used to provide an unbiased evaluation of the model fitting on the training dataset while tuning model hyperparameters. The validation dataset X_{val}, Y_{val} can be represented as follows,

$$X_{val} = [x_{N+1}, x_{N+2}, ..., x_M], \quad (27)$$
$$Y_{val} = [y_{N+1}, y_{N+2}, ..., y_M]. \quad (28)$$

The test dataset is used to measure the performance of the model in predicting the channel parameters. We take the last part of the historical channel information as the test dataset. According, X_{test} and Y_{test} can be represented as,

$$X_{test} = [x_{M+1}, x_{M+2}, ..., x_L], \quad (29)$$
$$Y_{test} = [y_{M+1}, y_{M+2}, ..., y_L]. \quad (30)$$

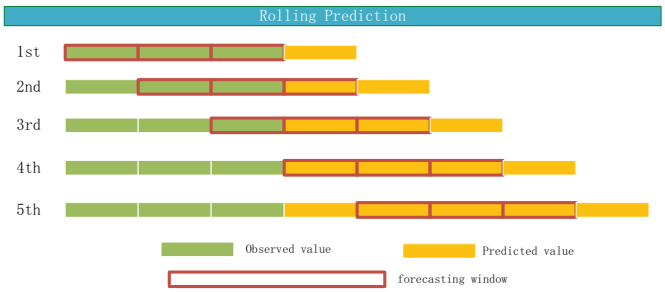


FIGURE 5: **The process of rolling prediction.**

2) Rolling prediction

In the rolling prediction, the size of the forecasting window is fixed, and each predict value needs to be added to the forecasting window. Rolling prediction is an add/drop process for forecasting future data which ensures the forecasting window always maintains the same amount of time series data. The rolling prediction updates the forecasting window by using the rule of FIFO and it is also called continuous prediction. An example of the process of the rolling prediction is shown in Fig. 5, where the window size of the forecast is set to 3.

IV. SIMULATION RESULTS AND DISCUSSIONS

In this section, we evaluate the performance of the proposed model in predicting the channel parameters through several cases. We use the Rayleigh fading channels in this paper, and we describe the data generation and preprocessing, evaluation of performance, and simulation results and analysis as follows. As to the data description and preprocessing, we collect the Rayleigh fading channel information by using the improved Jake's model [27]. Moreover, we add the Gaussian

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2935463, IEEE Access

white noise at the channel state information, which is given by

$$\hat{H}(t) = H(t) + n(t). \tag{31}$$

In further, we normalize the channel parameters into the range of $[0, 1]$, since the normalization can accelerate the speed of the optimal solution of gradient descent in the deep learning networks. The normalized processing can be represented by

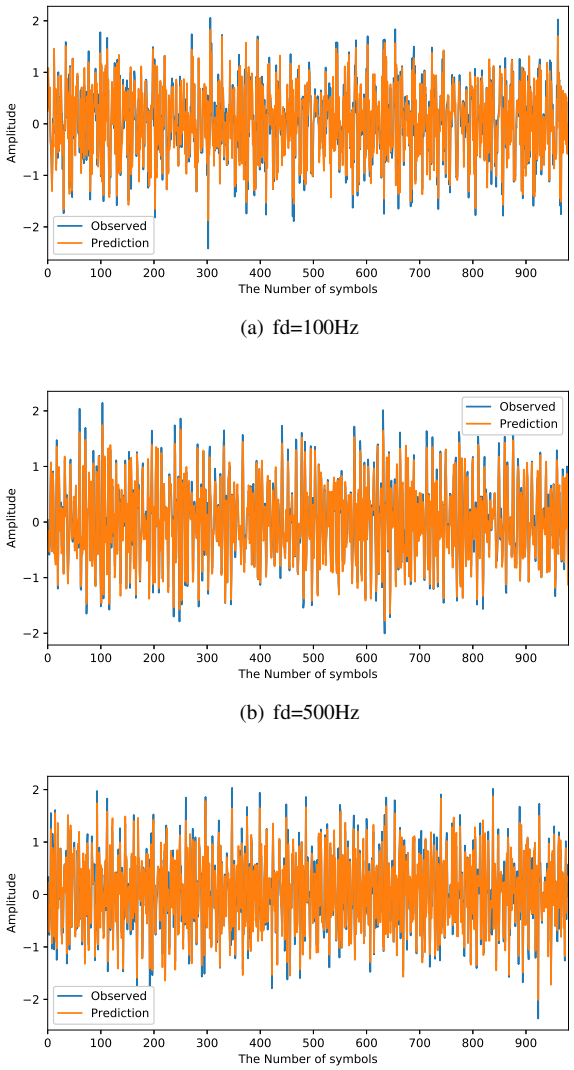
$$X_{scaled} = \frac{X - X_{min}}{X_{max} - X_{min}}, \tag{32}$$

where X_{min} and X_{max} denote the minimum and maximum values of the channel parameters, respectively. After the data normalization, we divide the channel information into training, validation, test dataset. The training set is used as the input data for the model to predict the channel information, and then we calculate the error between the predicted channel information and observed channel information by using the criterion of the normalized mean square error (NMSE), which is given by,

$$\text{NMSE} = \frac{\sum_{i=1}^{n} |y_p(i) - y_o(i)|^2}{\sum_{i=1}^{n} |y_o(i)|^2}, \tag{33}$$

where $y_o(i)$ and $y_p(i)$ denote the observed and predicted values of the channel parameters, respectively.

Fig. 6 depicts the amplitude of the observed and predicted channel parameters, where the signal-to-noise ratio (SNR) is set to 20dB. In this figure, the single-sample prediction scheme is used. The number of scattering components in the Rayleigh fading channels is set to 40, i.e. $N = 40$, and we set the sampling interval T_s to 1×10^{-4}s. In addition, the maximum Doppler frequency shift f_d is set to 100, 500 and 1000Hz in Fig. 6 (a), (b) and (c), respectively. The number of samples in the training, validation and test sets are set to 8000, 1000 and 1000, respectively. The dimension of the input layer is set to 20, so that the length of the trace back window is 20, and the number of the predicted channel samples is 980. We perform the simulations 2000 times, and take the average value to show in Fig. 6. As observed from this figure, we can find that the predicted channel parameter is almost identical to the observation one, which indicates that the proposed model is very effective in the single-sample prediction scheme.



(a) fd=100Hz

(b) fd=500Hz

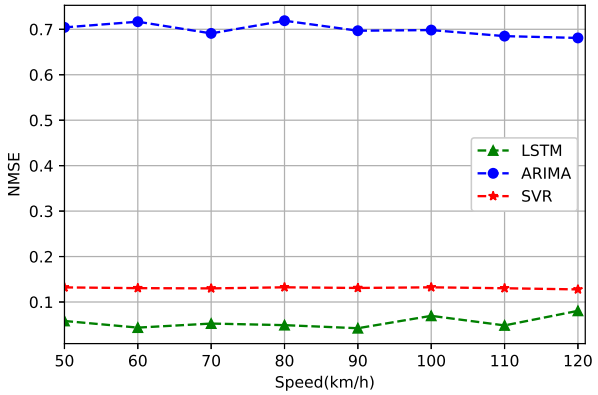
(c) fd=1000Hz

FIGURE 6: The amplitude of observed and predicted channels versus maximum Doppler frequency shift.

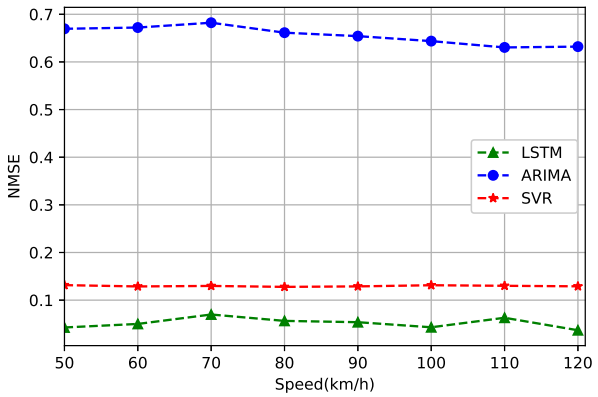
Fig. 7 represents the prediction error of the proposed LSTM based scheme versus the speed of vehicle, where the speed ranges from 50km/h to 120km/h. For comparison, we also plot the prediction error of the conventional ARIMA and SVR schemes, in Fig. 7. According to the relationship of $f_d = (v/c) f_c \cos a_m$, we set $a_m = 0$ and the carrier frequency is set to 1GHz, therefore we can obtain the linear relationship between the maximum Doppler frequency shift f_d and the speed of vehicle v. In addiction, the number of scattering components in the Rayleigh fading channels is set to 40, i.e. $N = 40$, and we set the sampling interval T_s to $1\times$ from 1000, 5000 and 10000s, respectively. The ratio of samples in the training, validation and test sets is set to 0.80, 0.10 and 0.10 in Fig. 7 (a), (b) and (c), respectively. And we obtain the average NMSE by performing the simulations 2000 times. From Fig. 7, we can find that for various values of speed of vehicle, the NMSE of the proposed LSTM scheme is much smaller than those of the ARIMA and SVR schemes, which validates the effectiveness of the proposed scheme in V2V communication system of ICVs.

Fig. 8 shows the prediction error of the proposed LSTM based scheme versus SNR, where SNR varies from 10dB to 40dB. For comparison, we also plot the prediction error of the conventional ARIMA and SVR in Fig. 8. In addition, the number of scattering components in the Rayleigh fading channels is set to 40, i.e. $N = 40$, and we set the sampling interval T_s to $1\times$ from 1000, 5000 and 10000s in Fig. 8 (a),

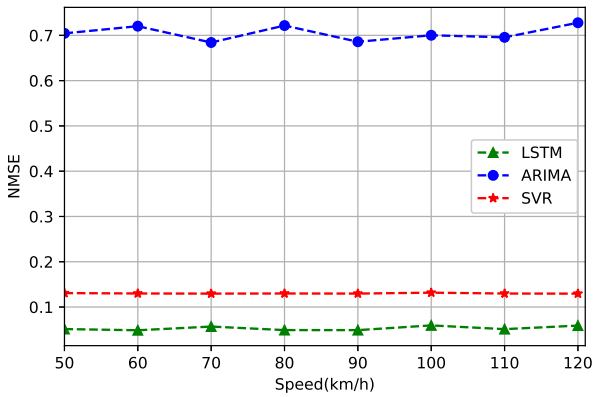
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2935463, IEEE Access



(a) 1/Ts=1000

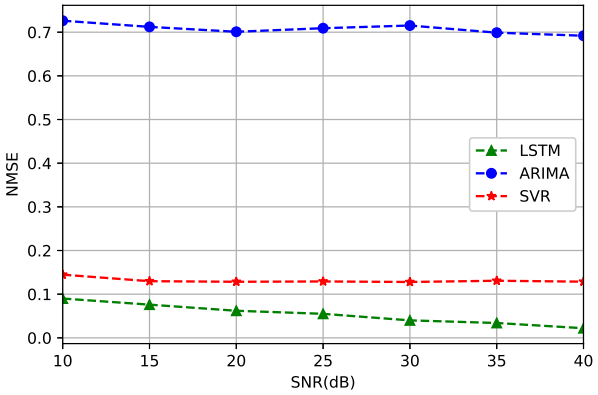


(b) 1/Ts=5000

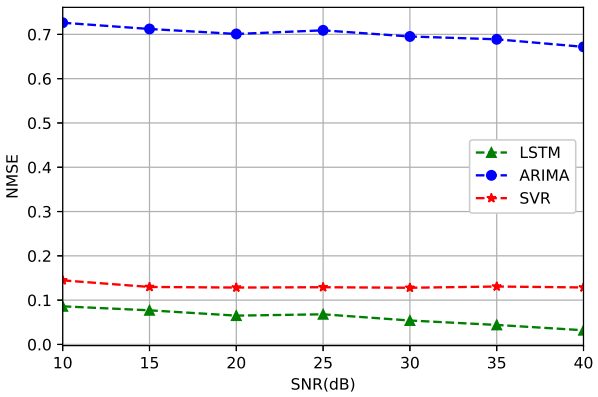


(c) 1/Ts=10000

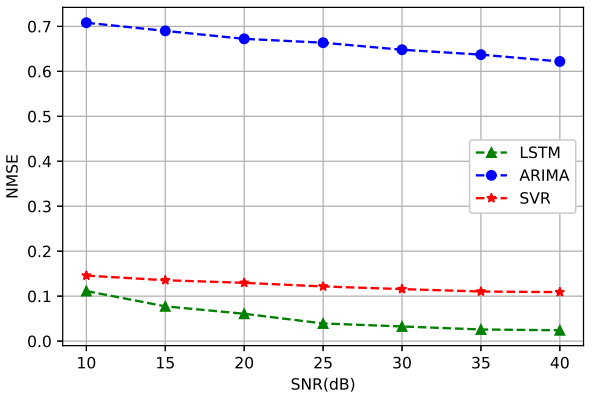
FIGURE 7: The prediction error of the LSTM, ARIMA and SVR versus speed of vehicle.



(a) 1/Ts=1000



(b) 1/Ts=5000



(c) 1/T=10000

FIGURE 8: The prediction error of the LSTM, ARIMA and SVR versus SNR.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2935463, IEEE Access

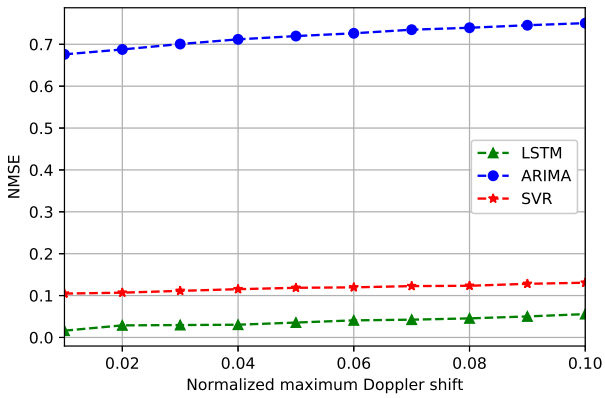
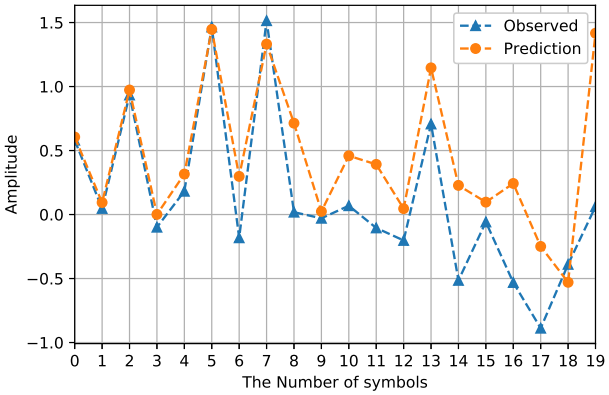


FIGURE 9: The prediction error of the LSTM, ARIMA and SVR versus the normalized maximum Doppler frequency shift.

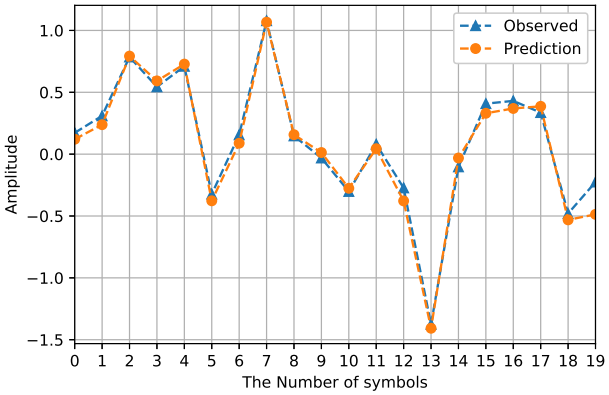
(b) and (c), respectively. The maximum Doppler frequency shift f_d is set to 1000Hz. The ratio of samples in the training, validation and test sets are set to 0.80, 0.10 and 0.10, respectively. And we obtain the average NMSE by performing the simulations 2000 times. From Fig. 8, we can find that for various values of SNR, the NMSE of the proposed LSTM scheme is much smaller than those of the ARIMA and SVR, which validates the effectiveness of the proposed scheme. Moreover, the NMSE of the three schemes decreases when the SNR becomes larger, since larger SNR can help predict the channel parameters.

Fig. 9 illustrates the prediction error of the proposed LSTM-based scheme, where the normalized Doppler frequency shift varies from 0.01 to 0.10. Note that the normalization is performed by multiplying f_d with the sample interval T_s. For comparison, we also plot the prediction error of the conventional ARIMA and SVR in Fig. 9. In addition, the number of samples in the training, validation and test sets are set to $1/T_s \times 0.8$, $1/T_s \times 0.1$ and $1/T_s \times 0.1$, respectively. The dimension of the input layer is set to 20. We perform the simulations 2000 times. We can see from Fig. 9 that the prediction error of the proposed scheme is much smaller than that of the conventional ARIMA and SVR, which further validates the effectiveness of the proposed scheme. Moreover, the NMSE of the three schemes becomes larger with the increasing Doppler frequency shift. This is because that the channels vary much more rapidly when f_d becomes larger, which increases the difficulty in predicting the channel parameters.

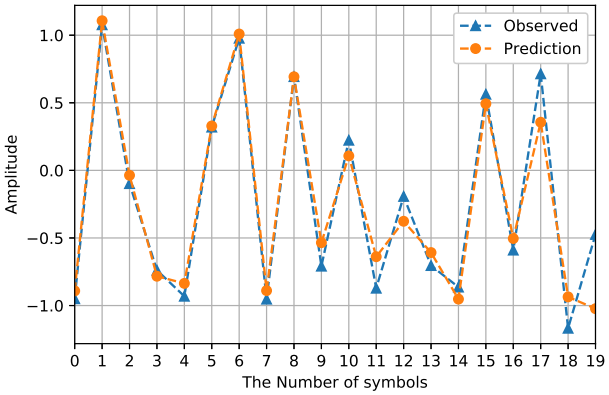
Fig. 10 demonstrates the amplitude of the observed and predicted channel parameters in the rolling prediction scheme. Note that we perform the simulations based on the rolling prediction scheme, in order to verify the robustness of the proposed model. The number of scattering components in the Rayleigh fading channels is set to 40, i.e. $N = 40$. We set the sampling interval T_s to 1×10^{-4}s. In addition, the maximum Doppler frequency shift f_d is set to 100, 500



(a) fd=100Hz



(b) fd=500Hz



(c) fd=1000Hz

FIGURE 10: The amplitude of observed and predicted channels in the rolling prediction scheme versus maximum Doppler frequency shift.

and 1000Hz in Fig. 10 (a), (b) and (c), respectively. The number of samples in the training, validation and test sets are set to 8000, 1000 and 1000, respectively. The dimension of the input layer is set to 20, so that the length of the trace back window is 20, and the number of the predicted channel samples is 20. We perform the simulations 2000 times, and take the average value to show in Fig. 10. We can find from Fig. 10 that in the rolling prediction scheme, the predicted channel parameter is still very close to the observed one, which implies that the proposed model can still work well in the channel prediction for the rolling prediction scheme.

V. CONCLUSIONS

In this paper, a novel channel prediction model based on deep learning approach was proposed for improving the performance of forecasting the CSI in the field of edge computing networks. To be more specific, we firstly investigated the feature of Rayleigh fading channel and the structure of LSTM network. We then proposed the channel prediction model consisting mainly of LSTM network. In order to validate the proposed model efficacy, we designed five case studies of channel prediction and conducted extensive simulations. Simulation results demonstrated that the proposed model could achieve highly efficient channel prediction, and it could outperform the conventional prediction model, such as ARIMA and SVR. In the future works, we will study how to use the predicted channel parameters to enhance the network security [28]–[30], and how to use the predicted channels for the wireless caching [31], [32].

REFERENCES

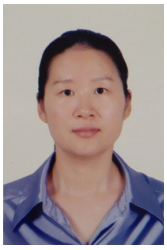
[1] F. Shi, "Secure probabilistic caching in random multi-user multi-uav relay networks," Physical Communication, vol. 32, pp. 31–40, 2019.
[2] L. Fan, N. Zhao, X. Lei, Q. Chen, N. Yang, and G. K. Karagiannidis, "Outage probability and optimal cache placement for multiple amplify-and-forward relay networks," IEEE Transactions on Vehicular Technology, vol. 67, no. 12, pp. 12 373–12 378, 2018.
[3] J. Zhao, T. Yang, Y. Gong, and et al., "Power control algorithm of cognitive radio based on non-cooperative game theory," China Commun., vol. 10, no. 11, pp. 143–154, 2013.
[4] J. Zhao, Y. Liu, and Y. Gong, "A dual-link soft handover scheme for C/U plane split network in high-speed railway," IEEE Access, vol. 6, pp. 12 473–12 482, 2018.
[5] X. Lai and L. Fan, "Distributed secure switch-and-stay combining over correlated fading channels," IEEE Transactions on Information Forensics and Security, vol. 14, no. 8, pp. 2088–2101, August 2019.
[6] X. Liu, F. Li, and Z. Na, "Optimal resource allocation in simultaneous cooperative spectrum sensing and energy harvesting for multichannel cognitive radio," IEEE Access, vol. 5, pp. 3801–3812, 2017.
[7] Z. Zhao, "A novel framework of three-hierarchical offloading optimization for MEC in industrial IoT networks," IEEE Transactions on Industrial Informatics, vol. PP, no. 99, pp. 1–12, 2019.
[8] X. Lai and W. Zou, "DF relaying networks with randomly distributed interferers," IEEE Access, vol. 5, pp. 18 909–18 917, 2017.
[9] X. Ma, J. Zhao, Y. Gong, and X. Sun, "CSMA/CA-aware connectivity quality of downlink broadcast in vehicular relay networks," IET Microwaves, Antennas and Propagation, vol. PP, no. 99, pp. 1–10, 2019.
[10] J. Xia, "When distributed switch-and-stay combining meets buffer in iot relaying networks," Physical Communication, vol. PP, pp. 1–9, 2019.
[11] J. Zhao, X. Guan, and X. Li, "Power allocation based on genetic simulated annealing algorithm in cognitive radio networks," Chinese Journal of Electronics, vol. 22, no. 1, pp. 177–180–, 2013.
[12] H. Guo, J. Zhang, and J. Liu, "Fiwi-enhanced vehicular edge computing networks: Collaborative task offloading," IEEE Vehicular Technology Magazine, vol. 14, no. 1, pp. 45–53, 2019.
[13] X. Lin, "Probabilistic caching placement in uav-assisted heterogeneous wireless networks," Physical Communication, vol. 33, pp. 54–61, 2019.
[14] X. Liu, M. Jia, Z. Na, W. Lu, and F. Li, "Multi-modal cooperative spectrum sensing based on dempster-shafer fusion in 5g-based cognitive radio," IEEE Access, vol. 6, pp. 199–208, 2018.
[15] J. Zhao, S. Ni, Y. Gong, and Q. Zhang, "Pilot contamination reduction in TDD-based massive MIMO systems," China Commun., vol. 13, no. 10, pp. 1425–1432, 2019.
[16] X. Xu, C. Wu, Q. Hou, and Z. Fan, "Gyro error compensation in opto-electronic platform based on a hybrid arima-elman model," Algorithms, vol. 12, no. 1, p. 22, 2019.
[17] J. Schmidhuber, "Deep learning in neural networks: An overview." Neural Netw, vol. 61, pp. 85–117, 2015.
[18] A. Jain, A. R. Zamir, S. Savarese, and A. Saxena, "Structural-rnn: Deep learning on spatio-temporal graphs," 2015.
[19] C. Li and Y. Xu, "Protecting secure communication under UAV smart attack with imperfect channel estimation," IEEE Access, vol. 6, no. 1, pp. 76 395–76 401, 2018.
[20] Y. Xu, "Q-learning based physical-layer secure game against multi-agent attacks," IEEE Access, vol. 7, pp. 49 212–49 222, 2019.
[21] C. Li and W. Zhou, "Enhanced secure transmission against intelligent attacks," IEEE Access, vol. 7, pp. 53 596–53 602, 2019.
[22] G. Li and Y. Yu, "Visual saliency detection based on multiscale deep cnn features," IEEE Transactions on Image Processing, vol. 25, no. 11, pp. 5012–5024, 2016.
[23] A. Abdulnabi, G. Wang, J. Lu, and K. Jia, "Multi-task cnn models for attribute prediction," IEEE Transactions on Multimedia, vol. 17, no. 11, pp. 1–1, 2015.
[24] X. Liu and X. Zhang, "5g-based green broadband communication system design with simultaneous wireless information and power transfer," Physical Communication, vol. 28, pp. 130–137, 2018.
[25] Z. Na, Y. Wang, and X. Li, "Subcarrier allocation based simultaneous wireless information and power transfer algorithm in 5g cooperative OFDM communication systems," Physical Communication, vol. 29, pp. 164–170, 2018.
[26] S. Ni, J. Zhao, H. H. Yang, and Y. Gong, "Enhancing downlink transmission in MIMO HetNet with wireless backhaul," IEEE Transactions on Vehicular Technology., vol. PP, no. 99, pp. 1–10, 2019.
[27] M. K. Simon and M. S. Alouini, Digital Communication over Fading Channels, 2nd ed. John Wiley, 2005.
[28] Z. Na, J. Lv, M. Zhang, and M. Xiong, "GFDM based wiereless powered communication for cooperative relay system," IEEE Access, vol. 7, pp. 50 971–50 979, 2019.
[29] J. Xia, "Intelligent secure communication for internet of things with statistical channel state information of attacker," IEEE Access, vol. PP, no. 99, pp. 1–7, 2019.
[30] J. Yang, D. Ruan, J. Huang, X. Kang, and Y.-Q. Shi, "An embedding cost learning framework using gan," IEEE Transactions on Information Forensics and Security, vol. PP, no. 99, pp. 1–10, 2019.
[31] X. Lin, "MARL-based distributed cache placement for wireless networks," IEEE Access, vol. 7, pp. 62 606–62 615, 2019.
[32] J. Xia and L. Fan, "Secure cache-aided multi-relay networks in the presence of multiple eavesdroppers," IEEE Transactions on Communications, vol. PP, no. 99, pp. 1–10, 2019.

GUANGQUN LIU graduated and received the bachelor's degree from the School of Computer Science and Cyber Engineering, Guangzhou University in June, 2019. He will be a Graduate Student of Xiamen University in September 2019. His main research interests include the fog/edge computing, and the deep Learning.

YAN XU received the bachelor's degree in electronic information science and technology from Hubei Normal University in 2017. She is currently a Graduate Student with the School of Computer Science and Cyber Engineering, Guangzhou University. Her research interests focus on wireless cooperative communications, physical-layer security, and multiagent machine learning algorithm.

ZONGJIANG HE received the bachelor's degree in electronic information science and technology from Huaihua University in 2015. He is currently a Graduate Student with the School of Computer Science and Cyber Engineering, Guangzhou University. His research interests focus on big data, server-side development, and machine learning.

YANYI RAO obtained the doctoral degree from the Institute of Information Science and Technology, Tsinghua University, in 2016. She is now a teacher at the School of Computer Science and Cyber Engineering, Guangzhou University. Her main research interests include the pattern recognition and the artificial Intelligence.



JUNJUAN XIA received the bachelor degree from the department of computer science from Tianjin University in 2003, and obtained the master degree from the department of electronic engineering from Shantou University in 2015. Now she works for the School of Computer Science and Cyber Engineering, Guangzhou University as a laboratory assistant. Her current research interests include wireless caching, physical-layer security, cooperative relaying and interference modeling.

LISENG FAN obtained the doctoral degree from Tokyo Institute of Technology, Tokyo, in the year of 2008. He is now a Professor at the School of Computer Science and Cyber Engineering of Guangzhou University. He has published more than 40 papers on the IEEE journal and IEEE conferences. His main research interests include the information security, wireless networks, and the artificial intelligence. His recent research interest is the application of artificial intelligence into the wireless networks.