J. Meas. Sci. Instrum., 2025, 16(1): 47-57. http://jmsi.publish.founderss.cn/
DOI: 10.62756/jmsi.1674-8042.2025005 jmsi@nuc.edu.cn

Handoff algorithm of high-speed railway wireless communication based on CNN-WaveNet

ZHAO Rongrong[1], LI Cuiran[1*], XIE Jianli[1], ZHANG Zepeng[2]

1. School of Electronics and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China;
2. Key Laboratory of Optoelectronic Technology and Intelligent Control of Ministry of Education, Lanzhou Jiaotong University, Lanzhou 730070, China

***Corresponding author:** LI Cuiran (licr@mail.lzjtu.cn)

Received: September 27, 2023 **Revised:** November 30, 2023 **Accepted:** December 19, 2023

Abstract: In order to ensure the uninterrupted communication between high-speed train and base station, driving safety and satisfying online experience of passengers, a dual-link switching algorithm based on CNN-WaveNet decision parameter multi-step prediction model is proposed to establish a two-hop relay communication system model between the high-speed train and the base station. Firstly, the switching algorithm uses convolution neural network (CNN) to extract the time sequence characteristics of decision parameters. Then, it learns the mapping relationship between feature information and decision parameters based on WaveNet and combining with rolling prediction method to realize multi-step prediction of decision parameters. Finally, dual-antenna communication mode is adopted to realize dual-link communication. The simulation results show that the proposed handover algorithm can improve handover trigger rate and handover success rate.

Key words: high-speed railway (HSR); handover; wireless communication; convolution neural network (CNN)

0 Introduction

With the rapid development of high-speed railway (HSR), the wireless communication system for HSR requires more reliable transmission and real-time sending and receiving of information without interruption, wherein handover is a key technology of HSR wireless communication. When the train runs from the coverage of the source base station to the coverage of the destination base station, the communication link is changed to ensure that the train receives scheduling information in real time and provides high-quality communication services for passengers. Therefore, it is necessary to optimize the handover of HSR wireless communication.

The scheme of handover optimization can be based on decision parameter optimization, decision algorithm optimization, network architecture optimization and signaling flow optimization. The optimization scheme based on decision parameters can make use of the characteristics of high-speed rail trajectory for tracking and prediction, and the information such as geographical location and train speed is beneficial to judge whether the train reaches pre-handover position. Thus, the scheme can carry out signaling interaction in advance and improve switching efficiency[1,2]. However, the optimization scheme has strong dependence on train position information. Since the handover decision conditions in the optimization scheme based on decision algorithm jointly determine the factors such as speed, power, and communication satisfaction, the deficiency of single decision algorithm will be reduced[3-5]. In addition, the optimization of network architecture depends on the characteristics of HSR environment, such as the large demand for passengers to surf the Internet, the airtight compartment space, and large number of surveillance videos. The optimization scheme is beneficial to improve resource utilization and system capacity and reduce compartment penetration loss. However, the change of network architecture makes the handoff complicated[6].

In recent years, the research on improving handover performance based on signaling flow optimization has been widely concerned. The traditional handover process is divided into three stages: handover measurement, handover decision, and handover execution[7,8]. The handover

prediction process can be added between the handover measurement and the handover decision to realize early handover by predicting decision parameters. Hao et al.[9] proposed a prediction model based on grey system theory, which realizes the prediction of decision parameters before handover decision. Since the train reports the decision parameters every T, the $(n+1)$ th decision parameter can be predicted according to the decision parameters obtained from previous n decision parameters. Li et al.[10,11] combined machine learning with handover technology to build a prediction model based on neural network. The input of neural network is n decision parameters from measurement report, and the output is the prediction of the $(n+1)$ th decision parameter. The algorithm optimizes handover success rate by improving prediction accuracy. Li[12] proposed a prediction model based on K-Trend-LSTM algorithm, which uses K-Trend to cluster the dataset and then uses long short term memory (LSTM) neural network model to predict parameters. The algorithm improves the handover success rate on the basis of improving the prediction accuracy of decision parameters.

However, in the above literatures, multi-step prediction is not considered when the model is used to predict decision parameters, but only single-step prediction is realized. The single-step prediction is to use n data from the previous n times to predict one future data $x_{\text{predict}}=\{x(n+1)\}$ at the $(n+1)$ th time. If x_{predict} meets the decision conditions, it is necessary to complete tasks such as prediction, handover judgment, and handover resource preparation within T time interval. The multi-step prediction model uses rolling prediction to predict the next value based on the predicted value. After the prediction step size k is set, the multi-step prediction model predicts a total of k future data ($x_{\text{predict}}=\{x(n+1), x(n+2), \cdots, x(n+k)\}$) by using the data from previous n times. If the gth predictive value in x_{predict} satisfies the handover judgment condition, it is necessary to complete tasks such as prediction, handover judgment and handover resource preparation within the gT time interval. In contrast, the multi-step prediction model is more practical than the single-step prediction model in the handover preparation time.

The above research uses single-step decision parameter prediction model and the prediction accuracy of the model can be improved. In order to improve the prediction accuracy and handover success rate, we propose a dual-link handover algorithm based on CNN-WaveNet multi-step prediction model. Firstly, the antennas are installed on the front and rear carriages of the train, for dual-link communication mode, which can reduce the probability of communication interruption. Secondly, handover prediction process is added to handover process for the prediction of decision parameters based on CNN-WaveNet multi-step prediction model. In this model, n data is input into convolution neural network (CNN) to extract features, and then WaveNet learns the features to obtain the prediction of the $(n+1)$ th parameter. To realize multi-step prediction, the prediction of decision parameter is added to the end of the input and then put into the prediction model. If the predicted value satisfies the handover decision condition, the handover resource is prepared. Finally, the handover trigger probability and handover success rate are calculated by simulation.

1 System model

1.1 High-speed rail wireless communication network architecture base on C/U plane split

At present, HSR technology is developing towards intelligence. The diversification of passenger services and the emergence of a large number of real-time monitoring require railway wireless communication systems to expand system capacity and improve communication efficiency[13]. However, The C/U decoupling architecture was recorded in the 5G high-speed mobile white paper jointly released by Wireless World Research Forum (WWRF) and Future Mobile Communication Forum (FuTURE) in 2016[14]. This architecture separates control plane (C-plane) from user plane (U-plane). Macro evolved NodeB (eNB) and phantom eNB realize the functions of C-plane and U-plane, respectively, which makes the network more flexible. Macro/phantom eNBs do not use the same spectrum, which solves the problem of co-channel interference between base stations, low spectrum utilization rate and small system capacity[15]. Fig. 1 shows the HSR wireless communication network architecture based on C/U plane split [16]. There are two types of handoff in this architecture:

1) The handover within the macro base station. In the process of handover from the phantom eNB$a1$ to the phantom eNB$a2$ when the train runs along the track (phantom-phantom eNB handover), eNB$a1$ and eNB$a2$ are controlled by macro eNBi.

2) The handover between macro base stations. In the process of handover from phantom eNB$a2$ to phantom eNB$b1$ after the handover of from macro eNBi to macro eNBj, the handover between macro eNBs occur twice (macro-macro eNB handover and phantom-phantom eNB handover), so the possibility of handover outage increases.

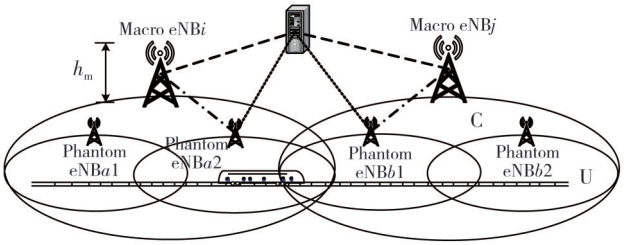

Fig. 1 HSR wireless communication network architecture based on C/U plane split

1.2 Two-hop relay communication system model

Because the compartment of the train is closed and made of metal materials, the penetration loss of the compartment may be about 24 dB when the user communicates directly with the base station. If train relay station (TRS) is installed on the train, the user can connect with TRS through access point (AP) and TRS is connected with the base station to realize two-hop relay communication. The two-hop relay communication mode can effectively reduce the penetration loss caused by closed compartments. In order to solve the problem of high outage probability of handover between macro eNBs, as shown in Fig. 2, antennas are installed at the front and rear of the train respectively and connected with the TRS to form two communication links[17]. When the train is running, the TRS compares and selects the antenna with good signal quality for communication. In the process of handover, the train establishes a communication link with the target base station when the front antenna meets the condition of handover. At this time, the rear antenna keeps connected with the source base station all the time. Since the rear antenna is synchronized with the target base station after the successful handover of the front antennas, the communication will not be interrupted during the handover. In addition, if the handover of the front antenna fails, the rear antenna can also be triggered once, which improves the handover success rate.

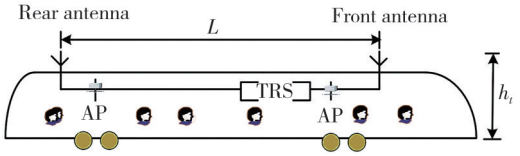

Fig. 2 Two-hop relay communication system model

2 Handoff algorithm based on CNN-WaveNet

2.1 Algorithmic process

The execution steps of dual-link handover algorithm based on CNN-WaveNet decision parameter multi-step prediction model are as follows:

1) Handover measurement. As seen from Figs. 1 and 2, the front antenna of the train keeps connected with macro eNB*i* and phantom eNB*a*1 (or phantom eNB*a*2). As the train moves, the train receives the signal of the target macro/phantom eNBs and reports it to macro eNB*i*. Firstly, it is necessary to judge the handover type and then judge whether the target phantom base station is under macro eNB*i* control or under macro eNB*j* control. If the target phantom base station is under macro eNB*i* control, the handover within the macro base station will occur, otherwise, the handover between macro base stations will occur.

2) Handover prediction. The prediction step is k. The input of CNN-WaveNet has n decision parameters as $\{x(1), x(2), \cdots, x(n)\}$, and the output has k predicted decision parameters as $\{x(n+1), x(n+2), \cdots, x(n+k)\}$.

3) Handover decision. In this study, signal to interference ratio (SIR) acts as the decision parameter, which more realistically reflects the signal strength of the base station without ignoring interference. The SIR of the source/target base station varies with the movement of train, and the calculation process of SIR is shown in later. The decision condition is that the SIR difference between the target base station and the source base station reaches u. According to handover type, handover decision condition is divided into two types. The first decision condition is that the handover is within macro base station, and the calculation formula is

$$R_{a2} - R_{a1} \geqslant u, \tag{1}$$

where R_{a1} and R_{a2} are the SIR values of eNBa1 and eNBa2, respectively. The second decision condition is that the handover is within the macro base stations, and the formula is

$$\begin{cases} R_j - R_i \geqslant u, \\ R_b - R_a \geqslant u, \end{cases} \tag{2}$$

where R_i and R_j are the SIR values of eNBi and eNBj, respectively; R_a and R_b are the SIR values of eNBa2 and eNBb1, respectively. If the predicted values do not meet the handover condition, return to step 2 (handover prediction). After receiving data of the $(n+1)$th time, the input of model is $\{x(2), x(3), \cdots, x(n+1)\}$ and the output is $\{x(n+2), x(n+3), \cdots, x(n+1+k)\}$ and so forth. Until the predicted value meets the handover condition, step 4 (handover execution) is executed.

4) Handover execution. The handover resource is prepared first and then the handover is performed. If the handover of the front antenna to the target base station is

successful, the rear antenna synchronizes with the target base station.

2.2 CNN-WaveNet switching decision parameter multi-step prediction model

2.2.1 CNN and WaveNet

CNN can extract features from multi-dimensional matrix data. It can reduce the complexity of the network and the number of network training parameters, and alleviate the overfitting of the model by means of its local receptive field and weight sharing and pooling layer characteristics. The structure of CNN mainly includes convolution layer, pooling layer, fully connected layer, and output layer. Usually, convolution layer and pooling layer take multiple and alternate settings[18]. After the data pass through several layers of convolutional layer and pooling layer, the whole features are obtained by integrating local features through full connected layer.

WaveNet will not lead to the explosion of model complexity when it correctly handles time sequence and is in long-term dependence. The structure of WaveNet includes causal convolution, extended convolution, and residual network[19]. In the causal convolutional network, $x=\{x(0), x(1), \cdots, x(t-1)\}$ is the input of WaveNet. $x(t)$ is the output of WaveNet. WaveNet can express $x=\{x(0),x(1),\cdots,x(t-1),x(t)\}$ with the conditional probability as

$$p(x)=\prod_{t=1}^{T} p(x_t|x_1,\cdots,x_{t-1}). \tag{3}$$

It can be seen from Eq. (3) that the prediction of $x(t)$ completely depends on the data from the previous t times and does not depend on the future data, which avoids the cheating behavior of the model using the future value to participate in the prediction. Expanding the network receptive field in the extended convolution network can make the model use the distant historical data and improve the reliability of the model. By setting different expansion factor d for each layer, the receptive field of the network increases exponentially with the depth. Fig.3 is a dilated convolutional network with the number of network layers $L=3$ and filter size $M=2$, and its network receptive field size is $2^{L-1}M=8$. The core of the residual network is to realize the mapping of nonlinear transformation function $f(x)$ and the objective function $h(x)$ in network layer. The calculation formula is

$$h(x)=f(x)+x, \tag{4}$$

where x is the input to the network layer.

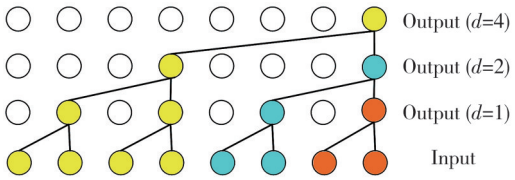

Fig. 3 Dilated convolution network

2.2.2 Edge feature extraction

The values of decision parameter (SIR) will be reported at a certain interval during the train running, and the decision parameter will form a set of time series $x=\{x(1),x(2),\cdots,x(N)\}$. In this study, the SIR time series of each base station in different environments are collected as the dataset, and then it is divided into S clusters $\{X_1,X_2,\cdots,X_\alpha,\cdots,X_S\}$ by k-means clustering method based on Euclidean distance. The k-means clustering method is mainly used to deal with the dataset. Because the data of the dataset come from different environments, the SIR sequence changes are different and the dataset is complex and huge. If the dataset is directly used to train the model, the operation efficiency may be affected. Therefore, different clusters obtained by using k-means clustering method represent the original datasets of different base stations. The dataset partition follows the similarity measurement method between sequences, which is calculated by Euclidean distance. Then, the sequences with the same sequence changes are divided into the same cluster according to similarity. The cluster contains sequences from different base stations, and the cluster data is used as the original dataset of the prediction model of the base stations in the corresponding cluster. The original dataset is used to train and test the model. If the SIR sequence of the source macro eNB belongs to cluster X_α, X_α will be used as the original dataset of the prediction model of the source macro eNB. Therefore, the k-means clustering method can reduce the workload and maintain better prediction accuracy. The number of prediction models varies linearly with the number of base stations. Sliding window method is used to preprocess a plurality of time series in X_α. Assuming that X_α^q is the qth time series in X_α, X_α^q is converted into ρ sequences $\{x_1^q,x_2^q,\cdots,x_\rho^q\}$, where $x_1^q=\{x_1(1), x_1(2), \cdots, x_1(n)\}$, whose sequence length is n, and the predicted value $\{x_1(n+1)\}$ is the label of x_1^q. The sequence training model with a length of n can be used to predict the future data according to the decision parameters of the previous n times, thus the decision parameter switching prediction process is realized.

Fig.4 shows the process of transforming a time series

$x=\{x(1), x(2), \cdots, x(N)\}$ into c time series with a length of n after preprocessing.

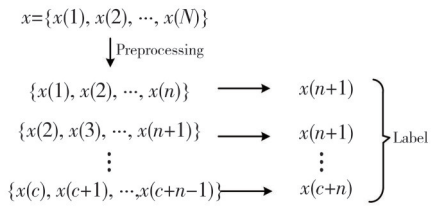

Fig. 4 Sliding window preprocessing process

Fig.5 is the CNN-WaveNet model, which shows the prediction process of putting the time series of size $1 \times n$ into CNN-WaveNet. When the time series sequence with the size of $1 \times n$ is used as the input of the model, there are z convolution kernels with the size of $1 \times N_0$ in the model for convolution processing. Finally, the output of model is z feature surfaces, with the size of $1 \times N_{cov}$ satisfying the relationship as

$$N_{cov}=\frac{n-N_0}{l}+1, \tag{5}$$

where l is the sliding step size of the convolution kernel window. In CNN, if z is larger, more features are extracted, but the corresponding model parameters are also more, which makes the model more complex. Eq. (6) is the calculation formula for the convolution layer of CNN. The input is weighted and accumulated, then an offset value is added, and finally, the extracted features are output through the activation function. The feature matrix after the sequence passes through the convolutional layer is

$$y(i,z)=f\left(\sum_{n_0=0}^{N_0-1}(x(il+n_0)w_z(n_0))+b_z\right), \tag{6}$$

where $i=0,1,2,\cdots,(N_{cov}-1)$; $\boldsymbol{X}$ is the input matrix; $y(i,z)$ denotes the ith element of the feature matrix $\boldsymbol{Y}$ obtained by the zth convolution kernel; $\boldsymbol{W}$ is the convolution kernel (weight matrix) and $\boldsymbol{b}$ represents the bias, so b_z and $w_z(n_0)$ are the bias and weight of the zth convolution kernel, respectively. After convolution operation, the characteristics are output by nonlinear activation function $f(\cdot)$. At present, the activation function commonly used in the convolution layer in CNN is ReLU [20], and the nonlinear activation function will not reach saturation due to the increase of input. In this study, the maximum pooling method is used in the pooling layer to down-sample the input feature plane. The maximum pooling method takes the maximum value of the elements in the sliding window of the pooling kernel as the calculation result on the input feature plane. The size of the pooled core is $1 \times N_1$, and the size of the feature surface after pooling is $1 \times N_{pooling}$, where $N_{pooling}$ is calculated by

$$N_{pooling}=\frac{N_{cov}}{N_1}. \tag{7}$$

Notably, reducing the size of feature surface can effectively reduce the workload. After extracting features from several convolution layers and pooling layers, the feature surfaces are integrated by using full connected layers, and then the time series prediction is carried out by using WaveNet. In Fig. 5, the circles marked with serial numbers represent the predicted values. Using the rolling prediction method, the prediction value obtained by CNN-WaveNet is put into the sequence and input into the model again to obtain the next predicted value. For example, the circle marked with sequence number 1 is put into the original sequence and input into CNN-WaveNet model to predict the circle marked with sequence number 2 again, and in turn, the circle marked with sequence number 3 can be obtained.

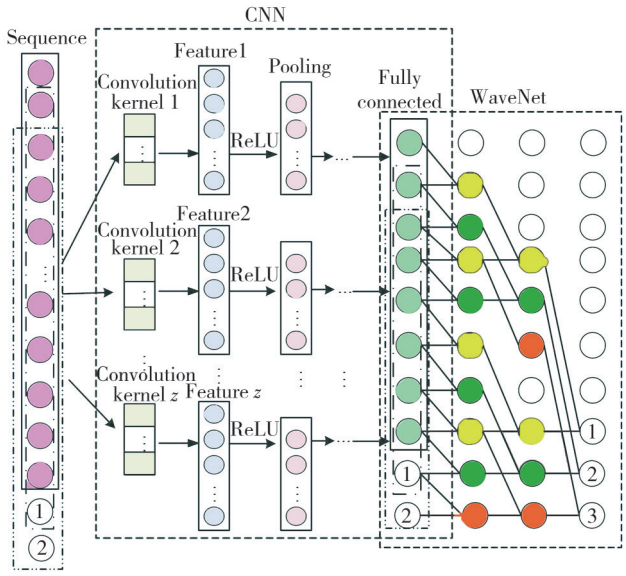

Fig. 5 CNN-WaveNet model

After the predicted value is obtained through model prediction, the predicted value is substituted into the handover decision condition for judgment. If the decision condition is met, handover resources need to be prepared. Since the time consumed by handover judgment is much less than T (T is the time interval for train cycle reporting), it is ignored. The time consumed by the whole process is

$$T_{total}=T_{predict}+T_{prepare}, \tag{8}$$

where $T_{predict}$ is the time consumed by the prediction model when predicting the SIR value, and $T_{prepare}$ is the time for preparing handover resources after reaching the decision condition. Fig.6 is a comparison between single-step prediction and multi-step prediction. It can be seen that if the single-step prediction model is adopted, the prediction and resource switching preparation work need

to be completed within T time interval. From $T=T^{\mathrm{s}}_{\mathrm{predict}}+T^{\mathrm{s}}_{\mathrm{prepare}}$, the single-step prediction handover resource preparation time is

$$T^{\mathrm{s}}_{\mathrm{prepare}}=T-T^{\mathrm{s}}_{\mathrm{predict}}. \tag{9}$$

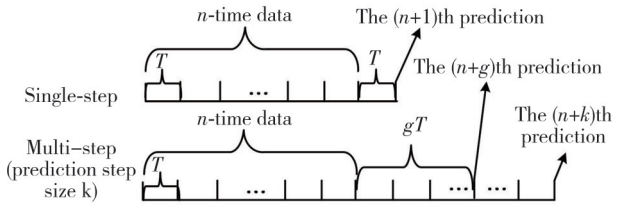

Fig. 6 Comparison of single-step prediction and multi-step prediction

In the multi-step prediction model with the prediction step size of k, if the gth prediction value satisfies the decision condition ($g\leqslant k, g\in\mathbf{N}^{+}$), the prediction and handover resource preparation work need to be completed within gT. Therefore, $gT=T^{\mathrm{m}}_{\mathrm{predict}}+T^{\mathrm{m}}_{\mathrm{prepare}}$ can be obtained, then the multi-step prediction handover resource preparation time is

$$T^{\mathrm{m}}_{\mathrm{prepare}}=gT-T^{\mathrm{m}}_{\mathrm{predict}}. \tag{10}$$

Because the time consumed by the multi-step prediction model based on rolling prediction is k times of the time consumed by the single-step prediction model ($T^{\mathrm{m}}_{\mathrm{predict}}=kT^{\mathrm{s}}_{\mathrm{predict}}$), $T^{\mathrm{m}}_{\mathrm{prepare}}=gT-kT^{\mathrm{s}}_{\mathrm{predict}}$. Usually, $T>T^{\mathrm{s}}_{\mathrm{predict}}$, according to Eq. (9), $T^{\mathrm{s}}_{\mathrm{prepare}}>0$. To ensure the significance of multi-step prediction, Eq. (10) should be greater than 0 ($gT-kT^{\mathrm{s}}_{\mathrm{predict}}>0$), so $g>\dfrac{kT^{\mathrm{s}}_{\mathrm{predict}}}{T}$.

According to Eqs. (9) and (10), the difference of switching resource preparation time between multi-step prediction model and single-step prediction model is

$$T^{\mathrm{m}}_{\mathrm{prepare}}-T^{\mathrm{s}}_{\mathrm{prepare}}=(g-1)T-(k-1)T^{\mathrm{s}}_{\mathrm{predict}}. \tag{11}$$

When the result of Eq. (11) is greater than 0, the multi-step prediction model is better than the single-step prediction model, and $g>\dfrac{(k-1)T^{\mathrm{s}}_{\mathrm{predict}}+T}{T}$ is obtained. Based on the above conditions, there are

$$\begin{cases} g\leqslant k,\ g\in\mathbf{N}^{+}, \\ g>\dfrac{kT^{\mathrm{s}}_{\mathrm{predict}}}{T}, \\ g>\dfrac{(k-1)T^{\mathrm{s}}_{\mathrm{predict}}+T}{T}, \\ \dfrac{kT^{\mathrm{s}}_{\mathrm{predict}}}{T}<\dfrac{(k-1)T^{\mathrm{s}}_{\mathrm{predict}}+T}{T}. \end{cases} \tag{12}$$

Therefore, according to Eq. (12), when g and k meet the condition $\dfrac{(k-1)T^{\mathrm{s}}_{\mathrm{predict}}+T}{T}<g\leqslant k\ (g\in\mathbf{N}^{+})$, the multi-step prediction model is better than the single-step prediction model, with the preparation time of handover resources of $(g-1)T-(k-1)T^{\mathrm{s}}_{\mathrm{predict}}$ more than that of the single-step prediction model.

2.3 Performance analysis

In this study, the handover trigger probability and handover success rate are used to analyze the handover performance[21]. In Fig. 1, it is assumed that the vertical point from the macro eNBi to the rail is the origin of coordinates, the vertical line from the macro eNBi to the rail is the Y axis, and the rail is the X axis. Since x is the position of the front antenna of the train on the track, the distance from the macro eNBi to the front and rear antennas is represented by $d_i(x)$ and $d_i(x-L)$ respectively as

$$d_i(x)=\sqrt{(h_{\mathrm{m}}-h_{\mathrm{t}})^2+d_{\mathrm{m}}+x^2}, \tag{13}$$

$$d_i(x-L)=\sqrt{(h_{\mathrm{m}}-h_{\mathrm{t}})^2+d_{\mathrm{m}}+(x-L)^2}, \tag{14}$$

where h_{m} is the height of macro eNBi antenna; h_{t} is the height of the train antenna, d_{m} is the vertical distance from macro eNBi to the rail, and L is the train length. When the train is running, the received signal quality changes with the distance between the train and the base station. The signal strengths of macro eNBi and phantom eNBa receiving the antenna in front of the train are respectively calculated by

$$P_i(x)=P_{\mathrm{t,m}}-L_{\mathrm{path,m}}(d_i(x)), \tag{15}$$

$$P_a(x)=P_{\mathrm{t,s}}-L_{\mathrm{path,s}}(d_a(x)), \tag{16}$$

where $P_{\mathrm{t,m}}$ is the transmission power of the macro eNB; $P_{\mathrm{t,s}}$ is the transmission power of the phantom eNB; $d_i(x)$ and $d_a(x)$ are the distances from macro eNBi and phantom eNBa to the antenna in front of the train, respectively; $L_{\mathrm{path}}(\cdot)$ is the path loss of the signal transmitted by the base station. Assuming that each base station has two co-frequency base stations, with P_{i1} and P_{i2} respectively denoting the signal quality of these two co-frequency base stations of macro eNBi, the co-frequency interference of macro eNBi can be got by

$$I_i(x)=10\lg\left(10^{\frac{P_{i1}}{10}}+10^{\frac{P_{i2}}{10}}\right). \tag{17}$$

Therefore, the SIR of the macro eNBi is

$$R_i(x)=P_i(x)-I_i(x)-\varepsilon_i, \tag{18}$$

where ε_i is a shadow fading subject to mean 0 and variance δ_i. In the process of handover, when the SIR value of the target base station and the SIR value of the original base station reach the threshold u, the handover will be triggered. The trigger probability of front antenna of macro eNBi handover to macro eNBj is

$$p_{i,j}^{\mathrm{f}}(x)_{\mathrm{ho}}=p\{R_j(x)-R_i(x)\geqslant u\}=$$
$$p\left\{P_{\mathrm{t,m}}-L_{\mathrm{path,m}}[d_j(x)]-\varepsilon_j-I_j-(P_{\mathrm{t,m}}-L_{\mathrm{path,m}}[d_i(x)]-\varepsilon_i-I_i)\geqslant u\right\}=$$
$$p\left\{P_{\mathrm{t,m}}-L_{\mathrm{path,m}}[d_j(x)]-\varepsilon_j-I_j-P_{\mathrm{t,m}}+L_{\mathrm{path,m}}[d_i(x)]+\varepsilon_i+I_i\geqslant u\right\}=$$
$$p\left\{\varepsilon_i\geqslant u+L_{\mathrm{path,m}}[d_j(x)]-L_{\mathrm{path,m}}[d_i(x)]+I_j-I_i+\varepsilon_j\right\}=$$
$$p\left\{\varepsilon_i\geqslant u+L_{\mathrm{path,m}}[d_j(x)]-L_{\mathrm{path,m}}[d_i(x)]+I_j-I_i+\frac{\varepsilon_0}{\varepsilon_j}=\varepsilon_0\right\}p\{\varepsilon_j=\varepsilon_0\}=$$
$$\int_{-\infty}^{+\infty}Q\left[\frac{u+L_{\mathrm{path,m}}[d_j(x)]-L_{\mathrm{path,m}}[d_i(x)]+I_j-I_i+\varepsilon_0}{\delta_i}\right]\times\frac{1}{\sqrt{2\pi{\delta_j}^2}}\exp\left(-\frac{{\varepsilon_0}^2}{2{\delta_j}^2}\right)\mathrm{d}\varepsilon_0, \tag{19}$$

where ε_j is the shadow fading of macro eNBj, obeying Gaussion distrabution with a mean of 0 and a variance of δ_j; Q is the probability that a normal random variable will have a value greater than x; and I_j is the co-frequency interference of macro eNBj. Since the front and rear antennas are installed on the train, if the front antenna handover fails, the rear antenna can be triggered for handover, so the trigger probability of handover is improved. If $p_{i,j}^{\mathrm{r}}(x-L)_{\mathrm{ho}}$ is used to represent the trigger probability of the rear antenna, the trigger probability of handover from the macro eNBi to macro eNBj is

$$p_{i,j}(x)_{\mathrm{ho}}=p_{i,j}^{\mathrm{f}}(x)_{\mathrm{ho}}+(1-p_{i,j}^{\mathrm{f}}(x)_{\mathrm{ho}})\,p_{i,j}^{\mathrm{r}}(x-L)_{\mathrm{ho}}. \tag{20}$$

Similarly, the trigger probability of handover from phantom eNB$a2$ to phantom eNB$b1$ is

$$p_{a,b}(x)_{\mathrm{ho}}=p_{a,b}^{\mathrm{f}}(x)_{\mathrm{ho}}+(1-p_{a,b}^{\mathrm{f}}(x)_{\mathrm{ho}})\,p_{a,b}^{\mathrm{r}}(x-L)_{\mathrm{ho}}. \tag{21}$$

A handover of macro eNBi to macro eNBj and a handover of phantom eNB$a2$ to phantom eNB$b1$ occur during the handover between macro base stations. Only when both handovers are successfully triggered can the handover between macro base stations be realized. Therefore, the trigger probability of handover between macro base stations is

$$p(x)_{\mathrm{ho}}=p_{i,j}(x)_{\mathrm{ho}}\,p_{a,b}(x)_{\mathrm{ho}}. \tag{22}$$

In the handover process, when the SIR value is lower than the threshold γ_{m}, there will be an outage. The outage probability of the front antenna of the macro eNBi is

$$p_i^{\mathrm{f}}(x)_{\mathrm{off}}=p[R_i(x)<\gamma_{\mathrm{m}}]=$$
$$p[P_{\mathrm{t,m}}-L_{\mathrm{path,m}}[d_i(x)]-\varepsilon_i-I_i<\gamma_{\mathrm{m}}]=$$
$$p[\varepsilon_i>P_{\mathrm{t,m}}-L_{\mathrm{path,m}}[d_i(x)]-\varepsilon_i-I_i-\gamma_{\mathrm{m}}]=$$
$$Q\left(\frac{P_{\mathrm{t,m}}-L_{\mathrm{path,m}}[d_i(x)]-\varepsilon_i-I_i-\gamma_{\mathrm{m}}}{\delta_i}\right). \tag{23}$$

Base station outage means that both the front and rear antennas are interrupted. The outage probability of the rear antenna is expressed as $p_i^{\mathrm{r}}(x-L)_{\mathrm{off}}$, and the outage probability of the macro eNBi is expressed as

$$p_i(x)_{\mathrm{off}}=p_i^{\mathrm{f}}(x)_{\mathrm{off}}\frac{1}{R_{\mathrm{m}}+L-x}\int_{x-L}^{R_{\mathrm{m}}}p_i^{\mathrm{r}}(x-L)_{\mathrm{off}}\,\mathrm{d}x, \tag{24}$$

where R_{m} is the radius of macro eNBs.

Using $p_j(x)_{\mathrm{off}}$ to represent the outage probability of macro eNBj, the outage probability of macro eNBi to macro eNBj is

$$p_{i,j}(x)_{\mathrm{off}}=p_i(x)_{\mathrm{off}}\,p_j(x)_{\mathrm{off}}+p_i(x)_{\mathrm{off}}\times(1-p_j(x)_{\mathrm{off}})+p_j(x)_{\mathrm{off}}(1-p_i(x)_{\mathrm{off}}). \tag{25}$$

The outage probability of phantom eNB$a2$ to phantom eNB$b1$ is

$$p_{a,b}(x)_{\mathrm{off}}=p_a(x)_{\mathrm{off}}\,p_b(x)_{\mathrm{off}}+p_a(x)_{\mathrm{off}}\times(1-p_b(x)_{\mathrm{off}})+p_b(x)_{\mathrm{off}}(1-p_a(x)_{\mathrm{off}}). \tag{26}$$

In the process of handover between macro base stations, any interruption will lead to failure, so the outage probability of handover between macro base stations is

$$p(x)_{\mathrm{off}}=p_{i,j}(x)_{\mathrm{off}}\,p_{a,b}(x)_{\mathrm{off}}+p_{i,j}(x)_{\mathrm{off}}\times(1-p_{a,b}(x)_{\mathrm{off}})+p_{a,b}(x)_{\mathrm{off}}(1-p_{i,j}(x)_{\mathrm{off}}). \tag{27}$$

When macro eNBi and macro eNBj are not interrupted and successfully triggered, the handover between macro eNBi and macro eNBj is successful, and the handover success rate is

$$p_{i,j}(x)_{\mathrm{suc}}=(1-p_i(x)_{\mathrm{off}})\,p_{i,j}(x)_{\mathrm{ho}}(1-p_j(x)_{\mathrm{off}}). \tag{28}$$

Similarly, the handover success rate between phantom eNB$a2$ and phantom eNB$b1$ is

$$p_{a,b}(x)_{\mathrm{suc}}=(1-p_a(x)_{\mathrm{off}})\,p_{a,b}(x)_{\mathrm{ho}}(1-p_b(x)_{\mathrm{off}}). \tag{29}$$

The success of handover between macro base stations means that the handover between macro eNBi and macro eNBj is successful, and the handover between phantom eNB$a2$ and phantom eNB$b1$ is successful, too. Thus, the success probability of handover between macro base stations is

$$p(x)_{\mathrm{suc}}=p_{i,j}(x)_{\mathrm{suc}}\,p_{a,b}(x)_{\mathrm{suc}}. \tag{30}$$

3 Simulation

3.1 Parameters settings

An HSR wireless communication system model based on Python 3.6 simulation software under the Tensorflow framework was established. The HSR wireless

communication network architecture based on C/U plane split is shown in Fig.1, the two-hop relay system model is shown in Fig.2, and the model parameters are listed in Table 1[15]. The propagation model of the HSR wireless communication system is COST231-Hata[22], which can modify the basic formula according to the field data collected in different scenarios. Then, the COST231-Hata model was modified to simulate different environments (viaducts, U-shaped grooves, suburbs, towns, cities.) by changing the path loss of the macro base stations, and SIR time series set of each base station in different environments is used as dataset.

Table 1 Model parameters of HSR wireless communication system

Parameter	Value	Parameter	Value
Frequency of macro eNBs f_{c1}/MHz	900	Frequency of phantom eNBs f_{c2}/GHz	5
Transmission power of macro eNBs $P_{t,m}$/dBm	43	Transmission power of phantom eNBs $P_{t,s}$/dBm	33
Path loss model of macro eNBs	COST231-Hata	Path loss model of phantom eNBs	WINNER Ⅱ D2a
Radius of macro eNBs R_m/km	3	Radius of phantom eNBs R_s/km	1.4
Overlapping area of macro eNBs A_m/km	1.2	Overlapping area of phantom eNBs A_s/km	1.2
Distance between macro eNBs D_m/km	4.8	Distance between phantom eNBs D_s/km	1.6
Distance between macro eNBs and railway d_m/km	0.4	Distance between phantom eNBs and railway d_s/km	0.02
Antenna height of macro eNBs h_m/km	0.03	Antenna height of phantom eNBs h_s/km	0.005
Co-channel macro eNBs distance I_m/km	14.4	Co-channel phantom eNBs distance I_s/km	9.6
Signal quality threshold of macro eNBs γ_m/dBm	16	Signal quality threshold of phantom eNBs γ_s/dBm	22
Handover trigger threshold u/dB	3	Reporting cycle of train T/s	0.1
Train length L/km	0.4	Antenna height of train h_t/km	0.002 5

When the train runs from the coordinate origin (the center of the macro eNBi) to the coverage edge of the macro eNBi, the SIR value of each base station is simulated and calculated 300 times for each time series as $x=\{x(1), x(2), \cdots, x(300)\}$. After acquiring the dataset, the time series are clustered using the k-means clustering algorithm based on Euclidean distance. The cluster containing the SIR time series of macro eNBi is used as the original data set of the macro eNBi prediction model. After preprocessed, 80% is used as the training set and 20% is used as the test set. The preprocessing steps are shown in Fig.4. The processed series with a length of n are input into CNN-WaveNet for training, and the model parameters are continuously debugged according to the training and testing results. Therefore, when the train is running, the SIR series with a length of n reported by the train cycle are selected and input into the trained prediction model, and the prediction model will output k prediction values with a prediction step size of k. The loss function used in the training of the CNN-WaveNet prediction model is mean absolute error loss (MAE), and its calculation formula is

$$L_{\mathrm{MAE}}=\sqrt{\frac{\sum_{i=1}^{k}(y_i^{\mathrm{predict}}-y_i)}{k}}, \tag{31}$$

where k represents the number of predicted values, y_i^{predict} represents the predicted value of the model, and y_i represents the theoretical value. The optimization algorithm used in the training of the prediction model is the Adam algorithm[23], which can adaptively adjust learning rate. After a large number of experimental tests, the parameter values of the CNN-WaveNet prediction model are obtained, as shown in Table 2.

Table 2 CNN-WaveNet model parameters

Parameter	Value	Parameter	Value
CNN convolution kernel size	5	Learning rate	0.001
Number of CNN convolution kernels	64	Iterative training times	200
WaveNet convolution kernel size	5	Dropout	0.2
Number of WaveNet convolution kernels	64	WaveNet expansion factor d	[1,2,4,8,16,32]

3.2 Simulation results and analysis

Fig.7 shows the relationship between the SIR value of each base station and the position of the train when the train runs along the positive direction of x axis. According to the parameters set of the system model in Table 1, the abscissas of the source/target macro base stations are $x_i=0$ km and $x_j=4.8$ km, respectively; and the abscissas of the source/target phantom base stations are $x_{a2}=1.6$ km and $x_{b1}=3.2$ km, respectively. The positions of the base stations are consistent with that in Fig.7.

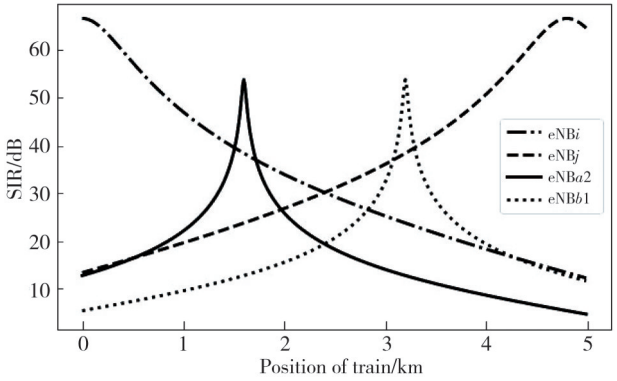

Fig. 7 Change in SIR values of base stations

It can be seen that when the train travels from the source base station to the target base station, the SIR value of the target base station gradually increases, while the SIR value of the source base station gradually decreases, especially the source phantom base station. The SIR attenuation of the source phantom base station (eNBa2) is large, which is mainly due to the high path loss of the phantom base station using 5 GHz high frequency band.

In order to further verify that the multi-step prediction model is more practical than the single-step prediction model, the model prediction time $T^{\mathrm{s}}_{\mathrm{predict}}=0.046\ 1$ s and decision time $T^{\mathrm{s}}_{\mathrm{decision}}=0.000\ 007$ s are obtained after averaging the data recorded by multiple simulations. Because $T^{\mathrm{s}}_{\mathrm{decision}}$ is much less than T (T=0.1 s), $T^{\mathrm{s}}_{\mathrm{decision}}$ is ignored. According to Eq. (9), for the single-step prediction model, the preparation time for handover of resources can be calculated as 0.053 9 s. According to Eqs. (11) and (12), the relationship between the prediction step size k and the preparation time for handover of resources is shown in Table 3. With the increase of prediction step size k, the multi-step prediction model has more and more preparation time for handover of resources than the single-step prediction model on the basis that g satisfies the conditions. Therefore, the multi-step prediction model is beneficial to make up for the time consumed by adding signaling flow in the preparation process for handover of resources, such as adding the dual-casting mechanism to reduce the probability of interruption[24], with a certain amount of time consumed. To sum up, the multi-step prediction model will have more practical significance.

Table 3 Relationship between k and preparation time for handover of resources

k	$g(g\in\mathrm{N}^+)$	$(T^{\mathrm{m}}_{\mathrm{prepare}}-T^{\mathrm{s}}_{\mathrm{prepare}})$/s
2	$1.461<g\leqslant 2$	0.053 9
3	$1.922<g\leqslant 3$	0.007 8−0.107 8
4	$2.383<g\leqslant 4$	0.061 7−0.161 7
⋮	⋮	⋮
k	$\dfrac{((k-1)T^{\mathrm{s}}_{\mathrm{predict}}+T)}{T}<g\leqslant k$	$(g-1)T-(k-1)T^{\mathrm{s}}_{\mathrm{predict}}$

In order to verify the validity of the model, two indicators are selected for evaluation. The first is mean absolute percent error (MAPE), which is calculated as

$$E_{\mathrm{MAP}}=\frac{\sum_{i=1}^{k}\left|y_i^{\mathrm{predict}}-y_i\right|}{ky_i},\tag{32}$$

where k represents the number of predicted values. The smaller the E_{MAP}, the better the prediction performance of the model. The second indicator is R^2, which is calculated as

$$R^2=1-\frac{\sum_{i=1}^{k}(y_i^{\mathrm{predict}}-y_i)^2}{\sum_{i=1}^{k}(y_i-\bar{y}_i)^2},\tag{33}$$

where $\bar{y}_i$ represents the mean of theoretical true values. The closer the R^2 score is to 1, the better the model prediction effect is.

LSTM neural network prediction model and CNN are commonly used as time series prediction models. In this study, LSTM with the same depth is used to compare the prediction performance with CNN and CNN-WaveNe. As shown in Table 4, the E_{MAP} and R^2 of three prediction model are compared, and the results are calculated by the mean value of 10 simulation data. The data show that compared with the LSTM prediction model, the E_{MAP} of CNN-WaveNet is decreased by 16.54%, and R^2 is increased by 10.13%; Compared with CNN model, E_{MAP} of CNN-WaveNet is decreased by 17.76%, and R^2 is increased by 17.87%. The prediction performance of CNN-WaveNet model is better than LSTM and one-dimensional CNN model.

Table 4 Comparison of predictive performance

Model	E_{MAP}/%	R^2
LSTM	4.05	0.859 4
CNN	4.11	0.803 0
CNN-WaveNet	3.30	0.946 5

Fig.8 shows the results of SIR prediction using multi-step prediction model based on CNN-WaveNet decision parameter. It can be seen that the predicted value of CNN-WaveNet is close to the historical data, and the prediction effect is better.

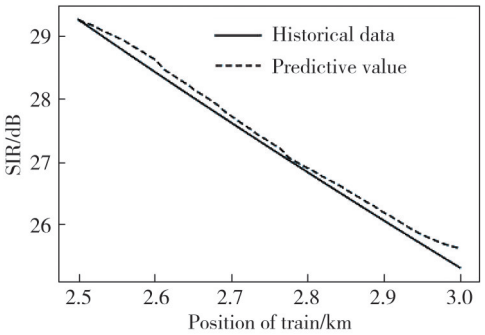

Fig. 8 CNN-WaveNet prediction results

Figs. 9 and 10 will compare the handover trigger probability and handover success rate of the following three algorithms in the HSR wireless communication system based on C/U decoupling architecture: the first algorithm is the traditional handover algorithm (single link), the second algorithm is the dual-link handover algorithm, and the third algorithm is the dual-link handover algorithm based on our prediction model. The handover trigger

probability based on the prediction model can be obtained according to Eq. (22), and the handover success rate can be obtained according to Eq. (30). Assuming that the train travels along the track in the positive direction of x axis, the range of x from 1.5 km to 3.0 km is the position where the train travels to the overlap area of the base station. Figs.9 and 10 show the changes of handover performance when the train is in the overlapping area.

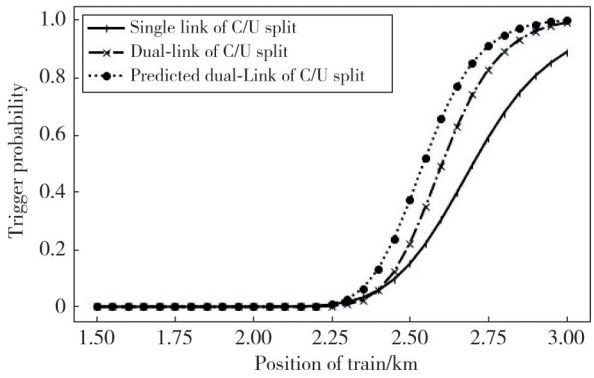

Fig. 9 Comparison of handover trigger probability

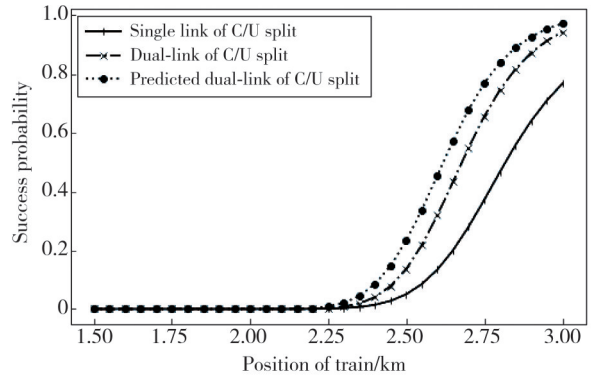

Fig. 10 Comparison of handover success probability

Fig.9 shows the performance comparison of handover trigger probability with different algorithms. It can be seen that the trigger probability of the traditional handover algorithm (single link) is lower than that of the dual-link handover algorithm. The handover trigger probability of the dual-link handover algorithm based on the decision parameter prediction model is the highest.

Fig. 10 shows the performance comparison of handover success rate under different algorithms. It can be seen that the traditional handover algorithm (single link) has the lowest handover success rate. Since the handover times between macro base stations in HSR wireless communication system based on C/U decoupling architecture become more and the possibility of interruption becomes higher; The handover success rate of the dual-link handover algorithm is significantly improved since the front and rear antennas are installed. If one antenna is interrupted, the other antenna can still ensure communication, thus reducing the probability of interruption; Dual-link handover algorithm based on decision parameter prediction model has the highest handover success rate because handover can be triggered in advance after the prediction value is obtained.

4 Conclusions

The optimization of wireless communication handover in HSRs environment is of great significance to ensure the safety of train operation and satisfying Internet experience of users. In this study, the handover prediction process is added after handover measurement and before handover decision. The decision parameters are predicted by CNN-WaveNet prediction model, and the handover decision is made in advance by using the predicted values of decision parameters to realize the preparation for handover of resources. At the same time, the antenna is installed on the top of the front and rear compartments of the train to reduce the handover outage probability and improve the handover trigger probability and handover success rate. Simulation results show that the prediction accuracy of decision parameters in the dual-link handover algorithm based on prediction model is high, and the algorithm can effectively improve the handover trigger probability and handover success rate.

Acknowledgement

This work was supported by National Natural Science Foundation of China (Nos. 62161016, 61661025), and Gansu Provincial Science and Technology Plan (No.20JR10RA273).

Declaration of conflicting interests

The authors have no conflict of interests related to this publication.

References

[1] MI G, MA S. Advance trigger handover algorithm based on the speed in LTE-R. Journal of Electronics & Information Technology, 2015, 37(12): 2852-2857.

[2] CHEN M M, YAN Y, ZHONG Z D. Location-based handover decision algorithm in LTE networks under high-speed mobility scenario//79th Vehicular Technology Conference, May 18-21, 2014, Seoul, South Korea. New York: IEEE, 2014.

[3] CHO H, SHIN S, LIM G, et al. LTE-R handover point control scheme for high-speed railways. IEEE Wireless Communications, 2017, 24(6): 112-119.

[4] ACHHAB T A, ABBOUD F, ASSALEM A. A robust self-optimization algorithm based on idiosyncratic adaptation of handover parameters for mobility management in LTE-A

heterogeneous networks. IEEE Access, 2021(9): 154237-154264.
[5] CHEN Y, NIU K, WANG Z. Adaptive handover algorithm for LTE-R system in high-speed railway scenario. IEEE Access, 2021(9): 59540-59547.
[6] HSIEH P J, LIN W S, LIN K H, et al. Dual-connectivity prevenient handover scheme in control/user-plane split networks. IEEE Transactions on Vehicular Technology, 2018, 67(4): 3545-3560.
[7] EL BANNA R, ELATTAR H M, ABOU EL-DAHAB M M. Fast adaptive handover using fuzzy logic for 5G communications on high speed trains//2021 16th International Conference on Telecommunications, June 30-July 2, 2021, Zagreb, Croatia. New York: IEEE, 2021: 10-17.
[8] WEN Q, LI J Y, YANG Y H, et al. The research of LTE-R handoff algorithm based on fuzzy prediction//2019 11th International Conference on Measuring Technology and Mechatronics Automation, April 28-29, 2019, Qiqihar, China. New York: IEEE, 2019: 493-497.
[9] HAO S, FANG X, MEMBER, et al. Handover scheme for 5G C/U plane split heterogeneous network in high-speed railway. IEEE Transactions on Vehicular Technology, 2014, 63(9): 4633-4646.
[10] LI D H, LI D P, XU Y Y. Machine learning based handover performance improvement for LTE-R//IEEE International Conference on Consumer Electronics, May 20-22, 2019. YILAN, Taiwan, China. New York: IEEE, 2019: 1-2.
[11] LI D H, LI D P, XU Y Y, et al. RNN-based C/U-plane decoupled HSR wireless Network handover mechanism for QoS improvement on control plane//2019 IEEE 8th Global Conference on Consumer Electronics, October 15-18, 2019. Osaka, Japan. New York: IEEE, 2019: 119-120.
[12] LI D H. Research on optimizatttion technology of high-speed railway wireless communication handover based on machine learning. Nanjing: Nanjing University of Posts and Telecommunications, 2020.
[13] AI B, MOLISCH A F, RUPP M, et al. 5G key technologies for smart railways. Proceedings of the IEEE, 2020, 108(6): 856-893.
[14] ZHAO J, LIU J, YANG L, et al. Future 5G-oriented system for urban rail transit: opportunities and challenges. China Communications, 2021, 18(2): 1-12.
[15] LI Y, FANG X. Reliability evaluation of 5G C/U-plane decoupled architecture for high-speed railway. Eurasip Journal on Wireless Communications & Networking, 2014, 2014(1): 127.
[16] ZHAO J, LIU Y, GONG Y, et al. A dual-link soft handover scheme for C/U plane split network in high-speed railway. IEEE Access, 2018(6): 12473-12482.
[17] YU X, YUAN L, CHEN X. An optimized seamless dual-link handover scheme for high-speed rail. IEEE Transactions on Vehicular Technology, 2016, 65(10): 8658-8668.
[18] WANG W L, YANG R F, GUO C X, et al. CNN-based hybrid optimization for anomaly detection of rudder system. IEEE Access, 2021(9): 121845-121858.
[19] RETHAGE D, PONS J, SERRA X, et al. A wavenet for speech denoising//2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), April 15-20, 2018, Calgary, Canada. New York: ACM, 2018: 5069-5073.
[20] SUN X, LIU L, LI C, et al. Classification for remote sensing data with improved CNN-SVM method. IEEE Access, 2019(7): 164507-164516.
[21] LIU Y Y. Research on handover technology in LTE-R for high-speed railway environment. Beijing: Beijing Jiaotong University, 2018.
[22] XIA L Y. Research on mobile communication networking strategy in high-speed railway scenario. Xi'an: Xidian University, 2018.
[23] KINGMA D, BA J. Adam: A method for stochastic optimization. Computer Science, 2015.
[24] YAN L, FANG X M, FANG Y G. A novel network architecture for C/U-plane staggered handover in 5G decoupled heterogeneous railway wireless systems. IEEE Transactions on Intelligent Transportation Systems, 2017, 18(12): 3350-3362.

基于CNN-WaveNet的高铁无线通信越区切换算法

赵容瑢[1], 李翠然[1*], 谢健骊[1], 张泽鹏[2]

1. 兰州交通大学 电子与信息工程学院, 甘肃 兰州 730070;
2. 兰州交通大学 光电技术与智能控制教育部重点实验室, 甘肃 兰州 730070

摘　要： 为保证高速列车与地面之间通信不中断、提高行车安全和乘客上网体验，提出了一种基于CNN-WaveNet判决参数多步预测模型的双链路切换算法，以建立列车与基站间的两跳中继通信系统模型。首先，该切换算法利用卷积神经网络(Convolution neural network，CNN)提取判决参数的时序序列特征。其次，基于WaveNet学习特征信息与判决参数之间的映射关系，并结合滚动预测方式实现判决参数的多步预测。最后，采用双天线通信模式实现双链路通信。仿真结果表明，所提出的切换算法可 实现切换触发率和切换成功率的优化。

关键词： 高铁；越区切换；无线通信；卷积神经网络

引用格式： ZHAO Rongrong, LI Cuiran, XIE Jianli, et al. Handoff algorithm of high-speed railway wireless communication based on CNN-WaveNet. Journal of Measurement Science and Instrumentation, 2025, 16(1): 47-57. DOI: 10.62756/jmsi.1674-8042.2025005