Wireless Channel Pattern Recognition Using k-Nearest Neighbor Algorithm for High-Speed Railway

Lei Xiong[1(✉)], Huayu Li[1], Zhengyu Zhang[1], Bo Ai[1], and Pei Tang[2]

[1] State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China
lxiong@bjtu.edu.cn

[2] China Railway Siyuan Survey and Design Group Co., Ltd., Wuhan 430063, Hubei, China

Abstract. Channel is important for the wireless communication system. The channel in high-speed railway is rapid time-variation and non-stationary. This papers discusses the channel characteristic in open space scenario, and defines 4 patterns. Furthermore a channel pattern recognition algorithm is proposed using k-nearest neighbor method. Simulation results show that the proposed method performs well with high accuracy and robust.

Keywords: Channel · Pattern recognition · k-nearest neighbor (kNN) · High-speed railway

1 Introduction

In recent years, the high-speed railway achieves good development, especially in China. By the end of 2018, China has the longest high speed railway network in the world over 29, 000 km, accounting for about two-thirds of the world's high-speed railway tracks in commercial service. Millions of passengers travel by high-speed railway every day, and the demand of high network capacity and reliable communication services has been growing rapidly. Thus, high-speed railway scenarios are the most important scenarios for wireless communication system, such as LTE and fifth generation (5G) communication systems, and dedicated wireless communication system, such as GSM-R and LTE-R [1–4].

Channel is of vital importance for a communication system. In high-speed railway, users have usually experienced that the throughput declines rapidly and the quality of transmission deteriorates seriously, which have become urgent problems to be solved. The reliable knowledge of the wireless channel characteristic is the foundation of the design and optimization of wireless communication system [5]. Due to the significance of this topic, channel characteristic in the high-speed railway have attracted more and more attention.

The wireless communication in high-speed railway has the unique channel characteristic, as shown below:

H. Gao et al. (Eds.): ChinaCom 2019, LNICST 312, pp. 579–588, 2020.
https://doi.org/10.1007/978-3-030-41114-5_43

(1) Due to the large Doppler shift is caused by high speed moving, and the channel state is rapidly time-varying [6–8].
(2) Due to the switch of the propagation condition, the channel doesn't satisfy the stationary assumption any more, which means the channel parameters vary with time, thus it cannot improve estimation accuracy of channel parameter with time average [9, 10].
(3) The moving direction and velocity of the train don't change rapidly, thus the wireless channel has obvious pattern characteristic, and can be predicted in a certain extent.

According to the characteristic mentioned above, the channel pattern recognition method based on the k-nearest neighbor (kNN) algorithm is proposed in the paper for the open space scenario. By the channel pattern recognition, we can make the channel prediction in further.

The remainder of the paper is organized as follows. Section 2 introduces the 3GPP high speed train (HST) channel model and discuss the patterns of Doppler shift. Section 3 discusses the kNN algorithm for channel patterns recognition. The simulation results are shown in Sect. 4. Finally, we draw some conclusions in Sect. 5.

2 Wireless Channel in High-Speed Railway

For the high-speed railway, 3GPP RAN4 have defined 4 scenarios, including scenario 1 (open space SFN), scenario 2 (tunnel), scenario 3 (relay in open space) and scenario 4 (traditional public network scenario), shown as Table 1 [11, 12].

Table 1. Summary of high-speed railway scenarios.

Scenarios	Cell diameter	Hop
1: Open space SFN	2 km−3 km	1 hop
2d: Tunnel SFN	1 km	1 hop
2e: Tunnel multi-antenna	–	1 hop
2f: Tunnel SFN CPE	2 km	1 hop
2g: Tunnel: Leaky cable to UE	–	1 hop
4: Public network	3 km	1 hop
2a: Tunnel SFN- RP; RP –UE with leaky cable	6 km	2 hops
2b: Tunnel: RRH with different id- RP; RP –UE with leaky cable	3 km	2 hops
2c: Leaky cable in tunnel- RP; RP –UE with leaky cable	1.5 km	2 hops
3: Open space eNB- RP; RP –UE with leaky cable	5 km	2 hops

Many channel models have been proposed for each scenario. For example, single-tap and two-tap HST channel model has been approved by 3GPP for open space scenario, and Ricean channel model for the tunnel with leaky cable scenario. In this paper, we focus on the open space scenario. The single-tap HST channel model is defined as below [13].

Due to the high speed moving of the train, the Doppler shift has severely impact on the performance of communication system. Thus, only the Doppler shift of the signal from the nearest Remote Radio Head (RRH) is considered in the single-tap channel model. The Doppler shift is calculated as:

$$f_d(t) = f\frac{v}{c}\cos\theta(t) \tag{1}$$

where f is carrier frequency, v is the velocity of the train, c is light speed, θ is the angle between the LoS path and moving direction, and

$$\cos(\theta(t)) = \begin{cases} \frac{0.5D_s - vt}{\sqrt{D_{\min}^2 + (0.5Ds - vt)^2}}, 0 < t \leq \frac{D_s}{v} \\ \frac{-1.5D_s + vt}{\sqrt{D_{\min}^2 + (-1.5D_s + vt)^2}}, \frac{D_s}{v} < t \leq \frac{2D_s}{v} \\ \cos\theta(t \bmod (\frac{2D_s}{v})), t > \frac{2D_s}{v} \end{cases} \tag{2}$$

where D_s is the distance between two neighbor RRHs, $D_{\min}$ is RRH railway track distance, shown as Fig. 1.

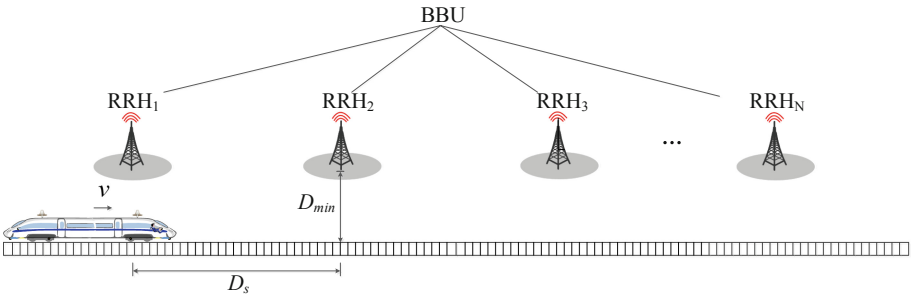

Fig. 1. Single-tap HST channel model

As shown in the Fig. 2, there are 4 variation patterns of the Doppler shift. In the pattern 1, the Doppler shift keeps almost constant positive value; In the pattern 2, the Doppler shift hops from negative value to positive value (the change rate of Doppler shift up to 2000 Hz/s); In the pattern 3, the Doppler shift keeps almost constant negative value; In the pattern 4, the Doppler shift hops from negative value to positive value.

The patterns are very important in the channel parameters estimation. For example, in the pattern 1 and 3, the receiver can take the time average to reduce measurement error of Doppler shift, but in the pattern 2 and 4, it doesn't work. Thus, the pattern recognition

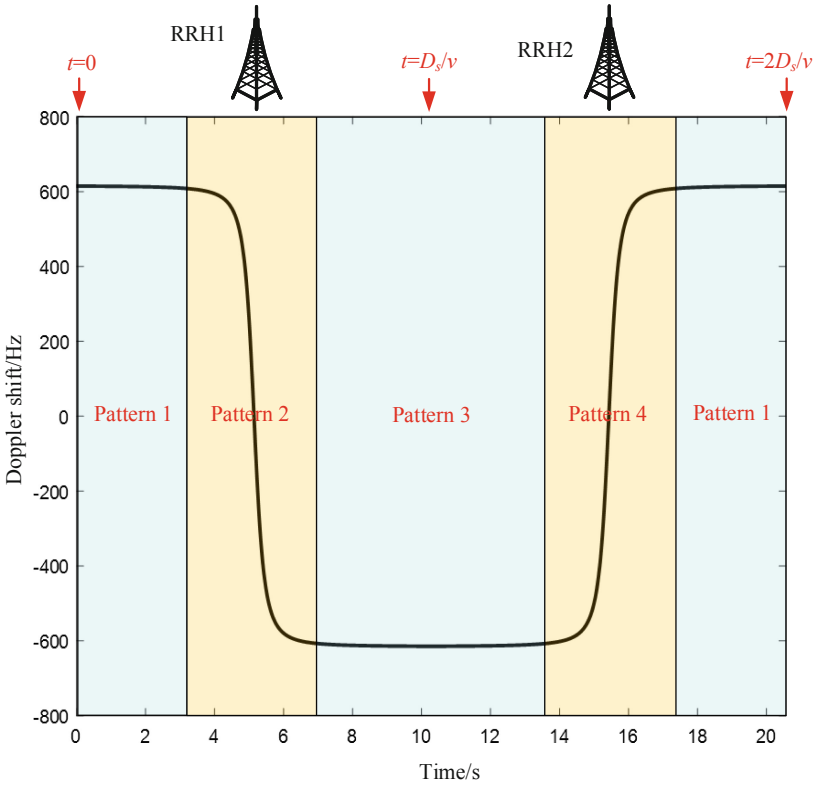

Fig. 2. Doppler shift of single-tap HST channel model ($f = 1890$ MHz, $v = 350$ km/h, $D_s = 1000$ m, $D_{\min} = 30$ m)

is essential for the Doppler shift estimation. Clearly, the pattern recognition between pattern 1 and 3 is a piece of cake. And the pattern recognition between pattern 2 and 4 is similar. Thus, the paper focuses on the recognition between pattern 1 and 2, and define the zones of pattern 1–4 as follow:

Zone of pattern 1: $[0, 0.5D_s - 4D_{\min})$ and $[1.5D_s + 4D_{\min}, 2.0D_s]$
Zone of pattern 2: $[0.5D_s - 4D_{\min}, 0.5D_s + 4D_{\min})$
Zone of pattern 3: $[0.5D_s + 4D_{\min}, 1.5D_s - 4D_{\min})$
Zone of pattern 4: $[1.5D_s - 4D_{\min}, 1.5D_s + 4D_{\min})$

3 Pattern Recognition Based on KNN Algorithm

Pattern recognition is an important topic and widely used in many areas such as speech and character recognition, medical diagnosis, remote sensing, and etc. KNN is one of the most popular classification algorithms in pattern recognition, machine learning, and data mining, which assume that data which are close together based upon some metrics,

such as Euclidean distance, more likely belong to the same category. Thus KNN searches for the group of K objects in the closest training data (similar) to objects in new data or data testing [14–16].

(1) **Training data set**

In the KNN algorithm, suppose that a training data set $\{(\bar{X}_i, p_i)\}$ with M elements, where each element consists of an object vector $\bar{X}_i$ with its classification label p_i, where $p_i = -1$ and $+1$ corresponding to the pattern 1 and pattern 2.

The training data set is generated as Eqs. (1) and (2). The parameter f is known to the receiver, and D_s only changes the cycle of the Doppler shift instead of the pattern recognition. And the v and $D_{\min}$ will be set according to the rail, shown as Table 2. The Each object vector $\bar{X}_i$ consist of L continuously Doppler shift measurement values with measurement interval Δt.

Table 2. The parameters of the training set.

Parameter	Value
f(MHz)	1890
v(km/h)	40, 60, 80, 100, 120, 160, 200, 250, 300, 350
$D_{\min}$(m)	20, 30, 40, 50
D_s(m)	1000
Δt (ms)	10

(2) **The distance between test vector and training vector**

We also have a test set $\left\{\left(\bar{Y}_j, p'_j\right)\right\}$, $j = 1 \ldots N$ of N test vector $\bar{Y}_j$ with unknown pattern label p'_j. Our goal is to calculate the Euclidean distance between the test vector $\bar{Y}_j$ and the training vector $\bar{X}_i$.

Generally the Euclidean distance formula is used to define the distance between the training data and testing data. However, the test vector may be shorter than the training vector, the modified Euclidean distance with sliding is applied,

$$d(\bar{X}_i, \bar{Y}_j) = \min_{\tau} \sqrt{\sum_{n=1}^{L} (x_{i,n+\tau} - y_{j,n})^2} \tag{3}$$

where $x_{i,n}$ and $y_{j,n}$ are the elements of $\bar{X}_i$ and $\bar{Y}_j$.

(3) **Pattern recognition**

Find the K closest vectors of the training set, namely $\bar{X}_i'$, and adopt the distance-weighted criterion

$$C_j = \sum_K \frac{1}{d(\bar{X}_i', \bar{Y}_j)} p_i \tag{4}$$

Thus, the label p_j' of the test vector $\bar{Y}_j$ is determined by

$$p_j' = \begin{cases} 1 & if\ C_j \geq 0 \\ -1 & if\ C_j < 0 \end{cases} \tag{5}$$

Now, we have achieved the pattern recognition for the test vector.

4 Simulation Results

To evaluate the performance of proposed pattern recognition algorithm, we conduct several experiments with the simulation parameters shown in Table 3. And we assume the error of Doppler shift measurement is the Gaussian distribution with zero mean and standard deviation σ_{fd}.

Table 3. Simulation parameter.

Parameter	Value
Frequency f (MHz)	1890
v (km/h)	350
Number of the testing data N	10000
Doppler Measurement interval Δt (ms)	10
K	3, 4, 5, 6
σ_{fd}(Hz)	0, 10, 20, 40, 70, 100
L	20, 30, 40, 50, 70
$D_{\min}$(s)	30, 60, 70, 80, 90, 100

The recognition error rate with K is shown as Fig. 3. We can see the recognition error rate varies with the K. When $K = 3$ and 4, the recognition error rate of pattern 1 is higher, whereas the recognition error rate of pattern 2 is higher when $K = 5$ and 6. Furthermore, there is a smaller gap between the recognition error rate of pattern 1 and pattern 2 when $K = 3$. Thus, the K is set to be 3 in the following simulation.

The recognition error rate with various σ_{fd} is shown as Fig. 4. We can see the recognition error rate is approximated to 0 when σ_{fd} is no more than 40 Hz, and increases with σ_{fd}, in which the recognition error rate of pattern 1 increases faster than pattern 2.

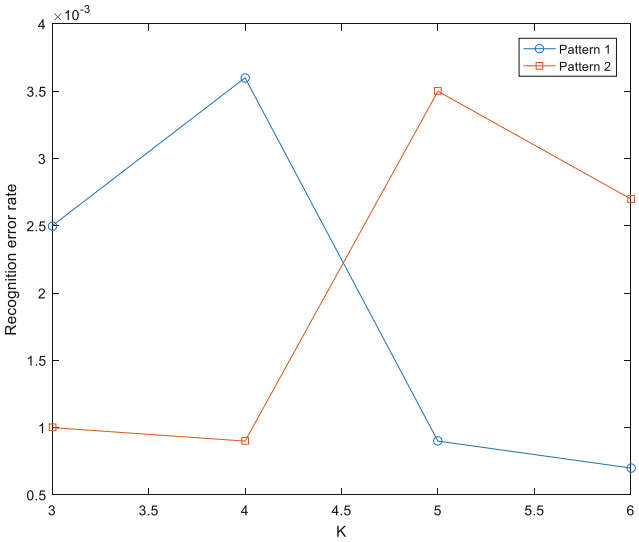

Fig. 3. The recognition error rate with K ($L = 50$, $\sigma_{\text{fd}} = 70$ Hz, $D_{\min} = 30$ m).

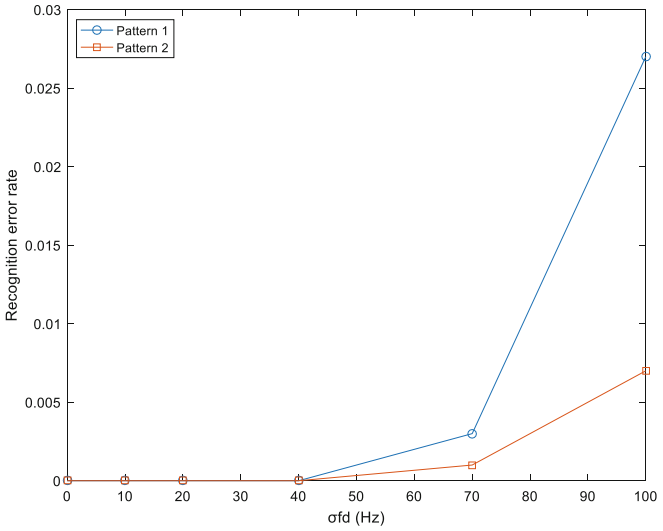

Fig. 4. The recognition error rate with various σ_{fd} ($L = 50$, $D_{\min} = 30$ m).

For example, when $\sigma_{\text{fd}} = 70$ Hz the recognition error rate of pattern 1 is 0.25%, but the recognition error rate of pattern 2 is only 0.1%. Since the σ_{fd} of LTE system is below 40 Hz in practice, the presented algorithm will work very well.

As shown in Fig. 5, the recognition error rate is close to zero when L is greater than 50. And the recognition error rate increase with the decease of the L. For example the recognition error rate of pattern 1 is 8.3, and the recognition error rate of pattern 2 is

0.1% when the L is 30. Clearly the length of the testing data L is higher, the recognition error rate is smaller, but the complex and processing delay are higher. Thus, the L is recommended as 50.

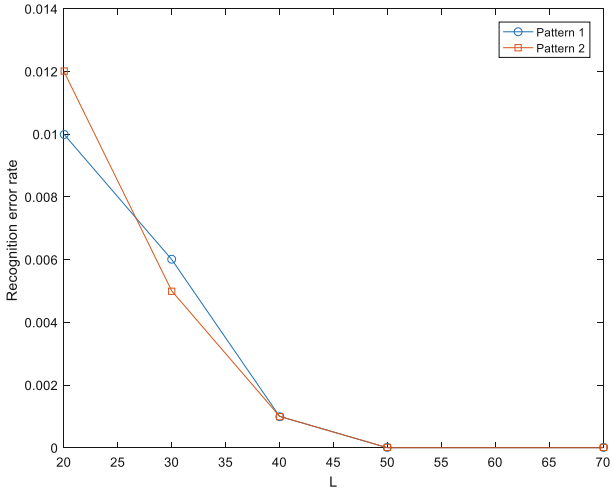

Fig. 5. The recognition error rate with various length of the testing data ($\sigma_{\text{fd}} = 40$ Hz, $D_{\min} = 30$ m).

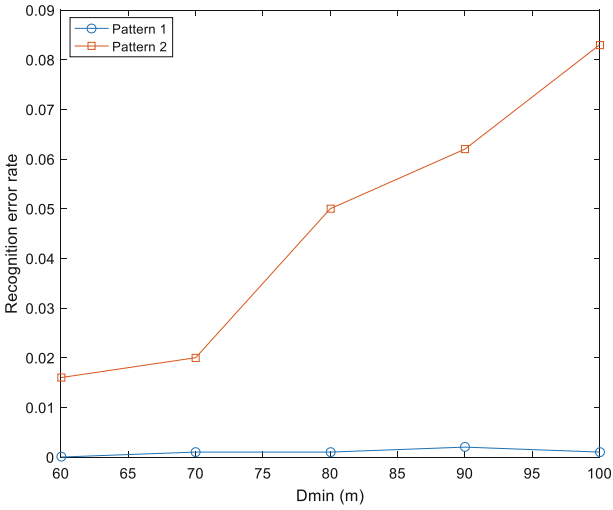

Fig. 6. The recognition error rate with various $D_{\min}$ ($\sigma_{\text{fd}} = 40$ Hz, $L = 50$).

In the rail, the $D_{\min}$ is usually 20–50 m, and the training set is generated in this range. However, the $D_{\min}$ may be greater than 50 m in some case. The recognition error rate

is shown as Fig. 6 for such a case. It can be seen, the recognition error rate of pattern 1 increase faster than that of pattern 1. For example, when $D_{\min} = 100$ m, the recognition error rate of pattern 1 is 0.3%, but the recognition error rate of pattern 2 is 8.3%. It shown the algorithm requires further optimization for the high $D_{\min}$.

5 Conclusion

The channel pattern recognition is crucial to the communication system. In the paper, we discuss the channel pattern in the open space scenario, and propose a channel pattern recognition method based on the KNN algorithm for channel pattern recognition. The simulation results show the pattern recognition method has high accuracy and robustness, can satisfy the demand of the receiver in high speed rail.

Acknowledgments. This work was supported the Fundamental Research Funds for the Central Universities (2018JBM079).

References

1. Zhou, Y., Pan, Z., Hu, J., Shi, J., Mo, X.: Broadband wireless communications on high speed trains. In: Proceedings of the WOCC, pp. 1–6 (2011)
2. Chen, C., Li, C.: Review of high speed rail communication systems. Comput. Technol. Appl. **46**, 24–26 (2010)
3. Wang, C.-X., et al.: Cellular architecture and key technologies for 5G wireless communication networks. IEEE Commun. Mag. **52**(2), 122–130 (2014)
4. Gonzalez-Plaza, A., et al.: 5G communications in high speed and metropolitan railways. In: 11th European Conference on Antennas and Propagation (EUCAP 2017), pp. 658–660. IEEE, Paris (2017)
5. Guan, K., et al.: Towards realistic high-speed train channels at 5G millimeter-wave band—Part I: paradigm, significance analysis, and scenario reconstruction. IEEE Trans. Veh. Technol. **67**(10), 9112–9128 (2018)
6. Zhou, Y., Wang, J., Sawahashi, M.: Downlink transmission of broadband OFCDM systems Part II: effect of doppler shift. IEEE Trans. Commun. **54**(6), 1097–1108 (2006)
7. Hui, B., et al.: Efficient Doppler mitigation for high-speed railway communications. In: 18th International Conference on Advanced Communication Technology (ICACT 2016), pp. 634–638. IEEE, Paris (2016)
8. Wang, S., et al.: Doppler shift and coherence time of 5G vehicular channels at 3.5 GHz. In 2018 IEEE International Symposium on Antennas and Propagation & USNC/URSI National Radio Science Meeting (AP-S 2018), pp. 2005–2006. IEEE, Boston (2018)
9. Ghazal, A., et al.: A generic non-stationary MIMO channel model for different high-speed train scenarios. In: Symposium on Wireless Communications Systems of the International Conference on Communications in China (ICCC 2015), pp. 1–3. IEEE/CIC, Shenzhen (2015)
10. Wang, C.-X., Ghazal, A., Ai, B., Liu, Y., Fan, P.: Channel measurements and models for high-speed train communication systems: a survey. IEEE Commun. Surv. Tutor. **18**(2), 974–987 (2016)
11. Huawei, HiSilicon: UE demodulation performance evaluation under the new scenarios (R4-152601), pp. 1–3. 3GPP, Fukuoka, Japan (2015)

12. Huawei, HiSilicon: Scenario Summary (R4-153904). pp. 1–4. 3GPP, Fukuoka, Japan (2015)
13. Samsung: Discussion on High speed train scenarios (R4-152277). pp. 1–2. 3GPP, Rio de Janeiro, Brazil (2015)
14. Zhang, S.: Nearest neighbor selection for iteratively KNN imputation. J. Syst. Softw. **85**(11), 2541–2552 (2012)
15. Zhang, S., Li, X., Zong, M., Zhu, X., Wang, R.: Efficient kNN classification with different numbers of nearest neighbors. IEEE Trans. Neural Netw. Learn. Syst. **29**(5), 1774–1885 (2018)
16. Cui, L., Zhu, H., Zhang, L., Luan, R.: Improved k nearest neighbors transductive confidence machine for pattern recognition. In: International Conference on Computer Design and Applications (ICCDA 2010), pp. 172–175. IEEE, Qinhuangdao (2010)