Big Data-Driven Vehicle Mobility Analysis and Design for 5G

Ruoxi Sun(✉), Kai Zhang, and Yong Ren

Department of Electronic Engineering, Tsinghua University, Beijing 100084, China
xysrx@163.com, zhang-k16@mails.tsinghua.edu.cn, reny@tsinghua.edu.cn

Abstract. Each generation of communication technology has a subversion, 5G will have a greater bandwidth, high carrier frequency, extreme base station and device densities, especially in vehicular network. Mobility models play a pivotal role in vehicular network, especially for routing policy evaluation. Relying on big data technology, the big data aided vehicle mobility analysis and design gets a lot of attentions. In this paper, we commerce with introducing the data set, i.e., a big GPS data set in Beijing. Then, a novel vehicle and location collaborative mobility scheme is proposed relying the GPS data set. We evaluate its performance based on degree distribution, duration distribution and interval time distribution. Our works may help the mobility design in vehicular networks.

Keywords: Big data · Vehicle mobility · 5G

1 Introduction

Each generation of communication technology has a subversion, 5G will have a greater bandwidth, high carrier frequency and extreme base station and device densities, especially in vehicular network [4,6]. The growing number of vehicles and the peoples in city, resulting in traffic congestion and automobile exhaust, which greatly reduces people's life travel experience. To address these issue, many schemes have been proposed, and intelligent transportation system (ITS) is one of them. Thanks to the accumulation of vehicle data and the maturer big data technology, such as parallel computing, machine learning and deep learning, the vehicular network analysis and design based on big data has got more and more attention.

As we all known, vehicular network is a dynamic network [1,5,8]. Therefore, the mobility models are of great importance for evaluating the performance of upper-layer protocol, and incorrect mobility model even lead to wrong conclusions. In present stage, GPS data for the analysis of the mobility model has got widespread concern. For vehicular network, mobility models can be roughly divided into four categories [1]: synthetic models, survey-based models, traffic simulator-based models and the trace-based models. Synthetic models, which is based on mathematical models, is capable of reflecting a realistic physical

K. Long et al. (Eds.): 5GWN 2017, LNICST 211, pp. 238–245, 2018.
https://doi.org/10.1007/978-3-319-72823-0_23

effect, such as random way point [7] and weighted waypoint model [9]. Survey based Models get the models property by surveys, and the agenda-based mobility model [11] is a typical example. And the traffic simulator-based models is obviously, which are generated form traffic simulator, such as SUMO [13], VISSIM [14] and TraNS [15]. And the big data contributes to the trace-based models [16].

In fact, the big data for mobility model analysis is not limited in the race-based models. Specifically, the model based on social network belong to the first category [1–3,10,12,18]. Gonzalez *et al.* [17] analyzed the data collected tracing mobile phone users, and found that most people traveled around their familiar place. Song *et al.* [18] analyzed the distribution of time interval, i.e., how much time a user stayed at a location, and found the a truncated power law distribution. Based on real trajectory data, Musolesi *et al.* [19] modeled social relationships by interaction matrix, and the value of matrix elements represented the relationship between the two users.

Relying on the vehicle GPS data, we conduct the relevant research and our original contributions are as follows:

- Mobility model design: Inspired form the recommended system, we proposed three novel vehicular mobility models, which approximates real data and is easy to interpret.
- Real-world dataset evaluation: Relying on the vehicle GPS data in Beijing, we evaluate our scheme.

The remainders of this article are outlined as follows. In Sect. 2, we introduce the mobility models and relevant indicators. Section 3 establishes the performance comparison in degree distribution, duration distribution and interval time distribution, followed by the conclusion in Sect. 4.

2 System Model

In this section, we commerce with introducing the mobility model designed in Sect. 2.1. Then, in Sect. 2.2, we specify the evaluating indicator, which is utilized for mobility scheme comparison in this paper.

2.1 Mobility Model

We first need to explain the reasons for introducing social attributes. For a specific vehicle, we plot all positions in the given time zone. The corresponding results are presented in the Fig. 1.

As can be seen from the diagram, the driver has a preference for position. And this property is a traditional random walk mobility model can not be reflected.

Typical social features based mobility model need a interaction matrix, i.e., R, to reflect the relationship between two vehicles or a vehicle to a location. In this paper, we also consider the distance martrix, i.e., D. And the relationship

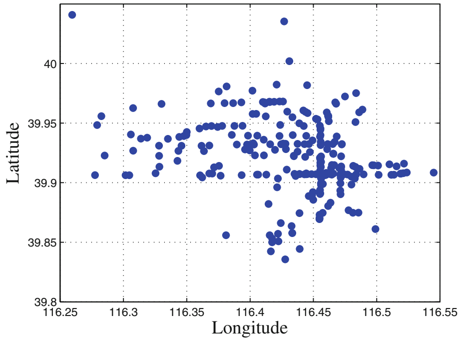


Fig. 1. Moving track of a vehicle.

and the distance of i and j is represented by R_{ij} and D_{ij}. The movement of vehicle could be modeled by the Markov process, which can be defined as:

$$Q_{ij} = \frac{R_{ij} D_{ij}}{\sum_{j=1}^{m} R_{ij} D_{ij}} \tag{1}$$

Based on the different definitions of social properties in R, the related solutions include the random initialization scheme (RIS), the global critical location assessment scheme (GCLAS) and the personalized important location assessment scheme (PCLAS).

Specifically, if the values in R are random given, then the random initialization scheme are achieved. In the random initialization scheme (RIS), the R_{ij} is given by:

$$R_{ij} = rand(1), \tag{2}$$

where $rand(1)$ means a random number between 0 and 1.

Based on the global critical location assessment, i.e., the probability of each site being accessed is proportional to its frequency, we get the global critical location assessment scheme. In global critical location assessment scheme (GCLAS), the R_{ij} is given by:

$$R_{ij} = \frac{\sum\limits_{v} \sum\limits_{t} transit_{i \to j}}{\sum\limits_{v} \sum\limits_{t} \sum\limits_{k} transit_{i \to k}}, \tag{3}$$

where v reflects a vehicle, t means a specific time and k denotes any position that connects with position i.

Considering the individual characteristics of the vehicle, we evaluate the importance of the location based on the specific vehicle history data, which lead to the personalized important location assessment scheme. In the personalized important location assessment scheme (PCLAS), the R_{ij} is given by:

$$R_{ij}^{k} = \frac{\sum\limits_{t} transit_{i \to j}^{k}}{\sum\limits_{t} \sum\limits_{k} transit_{i \to j}^{k}}. \tag{4}$$

And the Q_{ij} should be modified as follows:

$$Q^k_{ij} = \frac{R^k_{ij} D_{ij}}{\sum_{j=1}^{m} R^k_{ij} D_{ij}} \tag{5}$$

Besides, we adopt the real track data, simplified trajectory data and random walk as a comprising. Specifically, we map the real track data (RTD) into a 25$*$25 grid network, then we get the simplified trajectory data (STD). The mobility in random walk scheme (RWS) assumes all the vehicles' movements are random.

2.2 Evaluating Indicator

Specifically, VANETs can be viewed as a time variant graph $\mathcal{G} = (No, E, \mathcal{T}, \rho)$. Vehicles compose the entity in No set and the relationship between them is the E set. In this paper, the relationship represents a communication link. In dynamic network, this relationship may chance over time, so T represents the survival time and T represents the time domain, which satisfy $T \subseteq \mathcal{T}$. $\rho : E \times \mathcal{T} \rightarrow \{0, 1\}$ indicates survival function, which reflects whether a given edge exists at a given time slot. For time vary graph, there are two important parameters, the duration of the connection and the length of time interval. In other words

- the duration of the connection: starting from the entity i and entity j connection, to the first break time point, $\rho_{ij}(T) = 1$ was established in this time range;
- the length of time interval: starting from the entity i and entity j break, to the first connection time point, $\rho_{ij}(T) = 0$ was established in this time range.

And these two parameters will be utilized later to evaluate our mobility model.

3 Scheme Comparison

In this section, we conduct some simulation analysis on mobility models on the vehicular networks relying on the dataset introduced in Subsect. 3.1. We consider three typical indicators, i.e., degree distribution, duration distribution and interval time distribution, in mobility model assessment. And the relevant results and analysis are presented in Sects. 3.2, 3.3 and 3.4, respectively.

3.1 Dataset Analysis

This dataset records around ten thousand vehicles' GPS data in Beijing. For each vehicle, the travel track of the vehicle is recorded for several days.

And the Fig. 2 depicts the whole vehicle distribution in a specific time. This figure is not only a good reflection of the spatial distribution of vehicles, while reflecting the characteristics of Beijing's road network structure, i.e., the grid network topology.

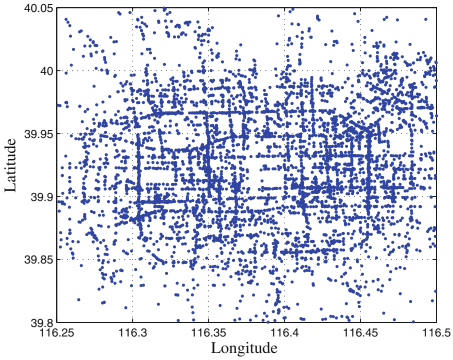


Fig. 2. System model.

3.2 Degree Distribution Comparison

In this subsection, we evaluate the mobility models based on degree distribution. Actually, the degree distribution is a very vital indicator in complex network, especially in distinguishing network types.

Based on the above five mobility schemes specified in Sect. 3.2 and the real data, we plot their degree distribution in Fig. 3. From the results, we can easily find that the real data network shows the characteristics of scale-free. However, both the simplified trajectory data and other four mobility schemes are more like a Gaussian distribution network.

3.3 Duration Comparison

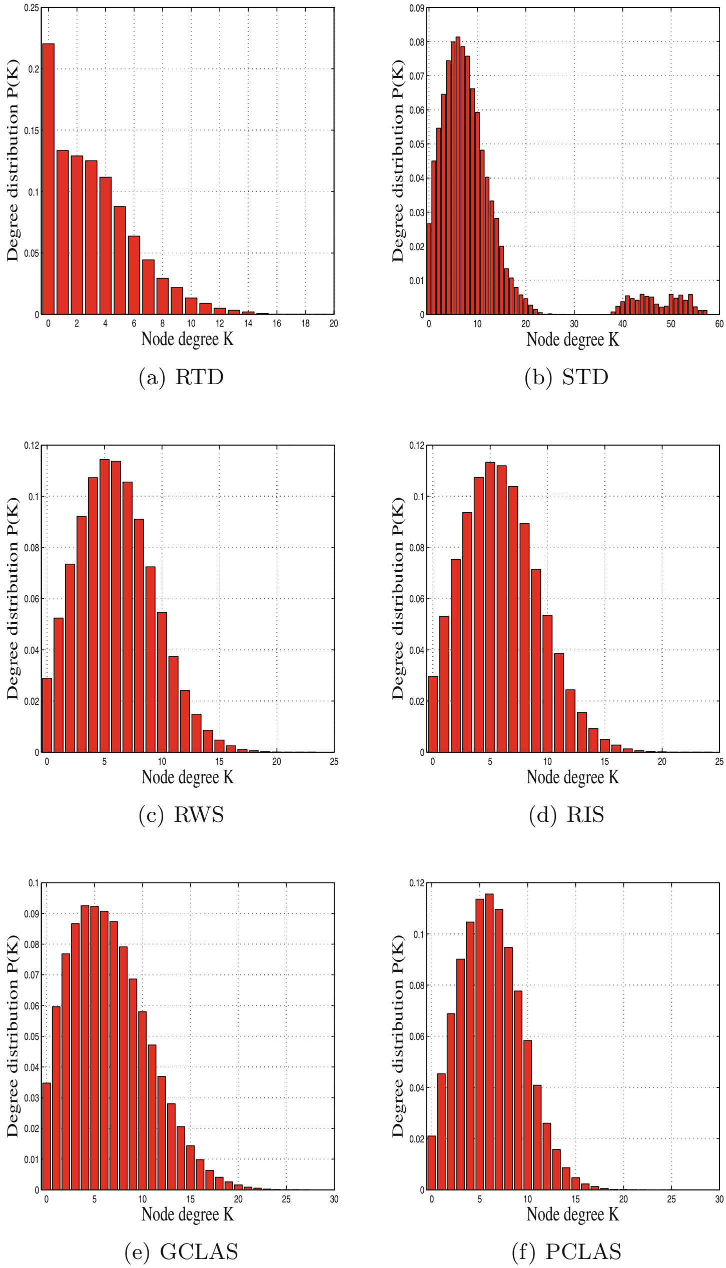
In network communication field, we may pay more attention to the latter two indicators. So in this section, we explore their performance in duration distribution.

The Fig. 4 represents their comparison in duration distribution. In addition to the simplified trajectory data being closer to the real data, the other four schemes converge. As can be seen from the local magnification in Fig. 4, although the latter four programs close to each other, GCLAS and PCLAS is closer to the real data.

3.4 Interval Time Comparison

Similarly, in this section we explore their performance in interval time distribution.

The results are revealed in Fig. 5. We can see that the GCLAS and PCLAS scheme is closer to the simplified trajectory data, which can verify its superiority.



(a) RTD (b) STD (c) RWS (d) RIS (e) GCLAS (f) PCLAS

Fig. 3. Degree distribution comparison

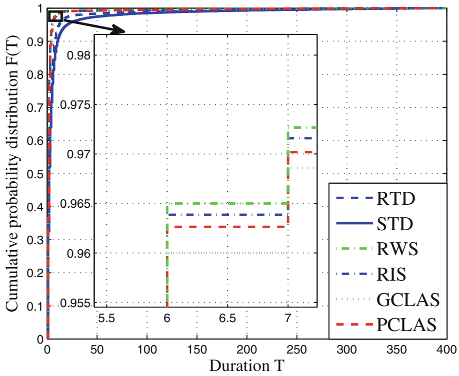


Fig. 4. Duration comparison

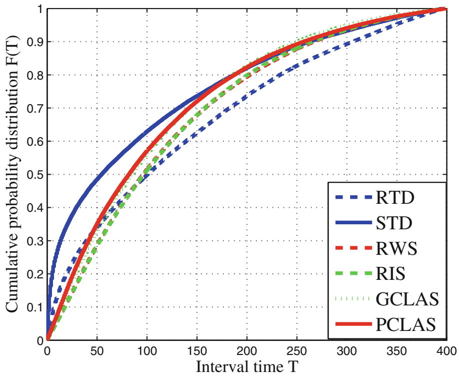


Fig. 5. Interval time comparison

4 Conclusion

To sum, inspired by user to product collaboration scheme in the recommended system, we propose the corresponding vehicle to location collaboration scheme in vehicular network. Based on degree distribution comparison, duration distribution comparison and interval time distribution comparison, the performances of the vehicle to location collaboration scheme are verified.

References

1. Harri, J., Filali, F., Bonnet, C.: Mobility models for vehicular ad hoc networks: a survey and taxonomy. IEEE Commun. Surv. Tutor. **11**, 19–41 (2009). Fourth Quarter
2. Jiang, C., Chen, Y., Liu, K.J.R.: Graphical evolutionary game for information diffusion over social networks. IEEE J. Sel. Top. Sig. Process **8**(4), 524–536 (2014)

3. Jiang, C., Chen, Y., Liu, K.J.R.: Evolutionary dynamics of information diffusion over social networks. IEEE Trans. Sig. Process **62**(17), 4573–4586 (2014)
4. Andrews, J.G., Buzzi, S., Choi, W., Hanly, S.V., Lozano, A., Soong, A.C., Zhang, J.C.: What will 5G be? IEEE J. Sel. Areas Commun. **32**(6), 1065–1082 (2014)
5. Jiang, C., Zhang, H., Ren, Y., Chen, H.: Energy-efficient non-cooperative cognitive radio networks: micro, meso and macro views. IEEE Commun. Mag. **52**(7), 14–20 (2014)
6. Wang, J., Jiang, C., Han, Z., Ren, Y., Hanzo, L.: Network association strategies for an energy harvesting aided super-wifi network relying on measured solar activity. IEEE J. Sel. Areas Commun. **34**(12), 3785–3797 (2016)
7. Bettstetter, C., Resta, G., Santi, P.: The node distribution of the random waypoint mobility model for wireless ad hoc networks. IEEE Trans. Mob. Comput. **2**(3), 257–269 (2003)
8. Wang, J., Jiang, C., Zhi, B., Quek, T.Q.S., Ren, Y.: Mobile data transactions in device-to-device communication networks: pricing and auction. IEEE Wirel. Commun. Lett. **5**(3), 300–303 (2016)
9. Hsu, W., Merchant, K., Shu, H., Hsu, C., Helmy, A.: Weighted waypoint mobility model and its impact on ad hoc networks. ACM SIGMOBILE Mob. Comput. Commun. Rev. **9**(1), 59–63 (2005)
10. Wang, J., Jiang, C., Quek, T.Q.S., Wang, X., Ren, Y.: The value strength aided information diffusion in socially-aware mobile networks. IEEE Access **4**, 3907–3919 (2016)
11. Zheng, Q., Hong, X., Liu, J.: An agenda based mobility model21. In: Proceedings of the 39th Annual Symposium on Simulation, pp. 188–195. IEEE Computer Society, April 2006
12. Wang, J., Jiang, C., Quek, T.Q.S., Ren, Y.: The value strength aided information diffusion in online social networks. In: IEEE Global Conference on Signal and Information Processing (GlobalSIP), pp. 1–6, December 2016
13. Sommer, C., German, R., Dressler, F.: Bidirectionally coupled network and road traffic simulation for improved ivc analysis. IEEE Trans. Mob. Comput. **10**(1), 3–15 (2011)
14. Fellendorf, M., Vortisch, P.: Microscopic traffic flow simulator vissim. In: Barceló, J. (ed.) Fundamentals of Traffic Simulation. ISOR, vol. 145, pp. 63–93. Springer, New York (2010). https://doi.org/10.1007/978-1-4419-6142-6_2
15. Piorkowski, M., Raya, M., Lugo, A.L., Papadimitratos, P., Grossglauser, M., Hubaux, J.-P.: Trans: realistic joint traffic and network simulator for vanets. ACM SIGMOBILE Mob. Comput. Commun. Rev. **12**(1), 31–33 (2008)
16. Kim, M., Kotz, D., Kim, S.: Extracting a mobility model from real user traces. In: INFOCOM, vol. 6, pp. 1–13 (2006)
17. Gonzalez, M.C., Hidalgo, C.A., Barabasi, A.-L.: Understanding individual human mobility patterns. Nature **453**(7196), 779–782 (2008)
18. Song, C., Koren, T., Wang, P., Barabasi, A.-L.: Modelling the scaling properties of human mobility. Nat. Phys. **6**(10), 818–823 (2010)
19. Musolesi, M., Mascolo, C.: Designing mobility models based on social network theory. ACM SIGMOBILE Mob. Comput. Commun. Rev. **11**(3), 59–70 (2007)