

Research Article

A New Approach to Regional Traffic Estimation for Intelligent Transport Systems

Minghe Yu ¹, Jianhua Feng,² Tianmiao Zhang,³ and Tiancheng Zhang ³

¹School of Software, Northeastern University, Shenyang 110169, China

²Department of Computer Science, Tsinghua University, Beijing 100084, China

³School of Computer Science and Engineering, Northeastern University, Shenyang 110169, China

Correspondence should be addressed to Minghe Yu; yuminghe@mail.neu.edu.cn

Received 30 January 2022; Accepted 1 June 2022; Published 28 June 2022

Academic Editor: Sang-Bing Tsai

Copyright © 2022 Minghe Yu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the great development of urban transportation systems, immediate urban traffic information has become an essential resource for the public. Traffic estimation is to predict current or future traffic situation (traffic speed and/or volume) in a road or a region of a city, and can benefit our daily life from many aspects, such as routing planning and traffic management. Existing works focus on estimating future traffic for individual road segments from a perspective of fine-grained level. This paper presents a new approach to estimating future traffic from a perspective of coarse-grained level, by which we estimate the traffic situation of a region, instead of an individual road segment. We propose a new concept about regional traffic named Ω -region, which aims to reflect the traffic situation of a region precisely. Two challenges in the regional traffic estimation problem are how to partition the road network into reasonable regions and how to estimate the regional traffic effectively. To address these challenges, first we define reasonable regions Ω -regions with traffic situations so that the all the road segment in the region has similar traffic. Then, we propose a three-phase partition method to divide the road network into Ω -regions based on historical trajectory data. Thirdly, we propose an effective linear-based model to estimate regional traffic. Experimental results on real-world dataset show that our proposed method achieves high performance.

1. Introduction

With the great development of urban transportation systems, immediate urban traffic information has become an essential resource for the public. Many human activities, such as path planning, traffic management, and city's infrastructure construction plan, are related to the awareness of urban traffic. Nowadays, many large metropolises in the world suffer from constant traffic congestion with their undergoing rapid economic growth. To mitigate the burden of the underlying road networks, efficient traffic management is of great importance, and metropolitan-scale traffic estimation is valuable to traffic management and to build smart city. To obtain precise traffic information for future plans, many existing studies [1–7] have been focused on the future traffic estimation, which aim to predict the urban traffic situations for the nearby time

periods. These works provide very fine-grained level of traffic estimation; i.e., they try to predict the future traffic for each individual road segments. However, these works are quite difficult to realize and thus they are hard to put into practical applications. This is because (1) their approaches are often developed based on the foundation that they are aware of the details of current traffic which, in fact, are difficult to retrieve. Existing methods, including obtaining the current traffic through the deployed traffic cameras and probing buses/taxis, can only retrieve sparse part of the current traffic. (2) Sometimes, the traffic situation of an individual road segment is irregular, which means the variation of its traffic flow is not similar with its adjacent roads. Therefore, the methods developed from general observations do not work effectively on regions, since we cannot assure that the traffic of every road segment in a region is uniform.

In this paper, we present an approach to estimating future traffic situation from a new perspective of coarse-grained level, regional traffic. The regional traffic is considered as the sum of vehicles appears in a region which consists of several adjacent road segments. A road can be one-way or two-way, and the traffic flows on the two directions in a two-way road usually are different. But they impact the traffic situation in a region together since the vehicles on two ways of a road pass the same junctions. Therefore, we combine the traffic flows of a two-way road into a single traffic for regional traffic estimation. There are many approaches to defining regions in a road network, for example, by grid, by population region, or by administration district. We propose a new way to define the region by traffic situation and name it as Ω -region, which has the property: all the traffic situations of road segments within it must be similar (see Section 3). There are certain advantages using the regional traffic. First, it can reflect both the regional traffic and the individual traffic since all the traffic situation within an Ω -region is similar. And the individual traffic can be easily deduced by the regional traffic as we can obtain the traffic ratio of each road segment in the region. Second, it can avoid the sparse problem of retrieving the current traffic since it is estimated in a coarse-grained level. Meanwhile, there are many practical applications of utilizing regional traffic estimation. Here, we give two examples. The regional traffic model can help our routing plan and is more expressive than using individual traffic estimation because in our daily life, when we are planning a path and trying to avoid traffic congestions, we usually consider to avoid congested regions rather than to remember all the individual congest road segments. We can also benefit traffic management with regional traffic estimation. Traffic polices in a city are usually assigned to certain regions to maintain traffic order. By detecting the congested region, the traffic polices can be made more judiciously.

However, there are several challenges to estimate regional traffic. First, we need to divide the road network into reasonable regions Ω -regions. To guarantee that the road segments within a region have similar traffic situation, we consider the historical trajectory data over the road segments and proposed a three-phase partition method to find those regions correctly. Second, we need to estimate the regional traffic precisely. To this end, we propose a linear-based model by considering the traffic of the region as well as the traffic of its neighborhood regions. To the best of our knowledge, it is the first study to estimate the reasonable regional traffic. In this paper, we make the following contributions.

- (1) We propose the concept of reasonable regional traffic and formulate the problem of regional traffic estimation.
- (2) We propose a three-phase partition method to divide the road network into Ω -regions by clustering historical trajectory data.
- (3) We propose an effective linear-based model to estimate regional traffic in Ω -regions.

- (4) We conduct experiments on real metropolitan traffic datasets, and the results show that our proposed methods achieve high performance.

The rest of the paper is organized as follows. We present the reviews of related works in Section 2. The preliminaries are defined in Section 3, and then we introduce how to generate Ω -regions in Section 4. We discuss how to estimate the future traffic for Ω -regions in Section 5. Experimental results are reported in Section 6. Finally, we conclude the paper in Section 7.

2. Related Works

Traffic estimation or modeling has been attracted many attentions from the researching community in recent years. Existing studies mainly focus on two categories:

Current traffic estimation. Current traffic estimation is based on partial observed traffic information from monitoring cameras and probing vehicles [4, 8, 9]. Traffic situations of a road network are usually modeled as a road-time matrix, where each entry stands by a traffic situation of a road segment in a specific time period. By assuming the adjacent road network may have similar traffic situation, matrix factorization [4, 8–11]-based methods are proposed to estimate the missing value of the matrix. In addition, Liu et al. [12] present GPTE, which utilizes nonlinear correlation modeling to represent real-time road network for traffic speed prediction. H-ARIMA [13] provided by Pan et al. utilizes both historical traffic patterns and current traffic speed for hybrid traffic prediction.

Future traffic estimation. Plenty approaches have been put into future traffic estimation [2, 3, 14–21]. A general method is developed over the basic hidden Markov model, which focuses on reflecting the evolvement of traffic through time. Other features (e.g., the scale of road segments and the traffic signals) that will affect traffic are also considered and incorporated into their models. Typically, future traffic estimation is used to predict traffic speed and volume. Ma et al. [22] study traffic speed prediction with deep learning theory. They represent network traffic as images with an image-based method and utilize convolutional neural network to extract spatio-temporal features. Wang et al. [23] utilize error feedback recurrent convolutional neural network (eRCNN) to capture complicated interactions of traffic speeds for prediction, which introduces separate error feedback neurons into the model and captures prediction errors to achieve high accuracy. Zhang et al. [24] study city crowd flow prediction. They consider both spatial dependency and temporal property and employ the residual neural network framework to dynamically aggregate them to predict the final traffic of crowds. There are also some works that focus on vehicular network to prediction future location of vehicle [25–27]. These methods provide essential and novel strategies, including vehicle coordinate normalization and context feature mining, to

improve the accuracy of prediction and efficiency of delivery ratio.

Recently, some works focus on regional traffic [28–31]. Zhu et al. [29] design a spatio-temporal attention mechanism to capture dynamic impact, including the number of vehicles in the accident and the amount of the injured people, of both local and global regions to predict traffic accident risk. Wang et al. [30] study regional traffic volume of the highway network in holidays and propose a holiday traffic growth model to predict holiday traffic based on seasonality, holiday, and trend components. Liu et al. [31] and Kang et al. [28] work on regional traffic prediction. In their works, a road network is partitioned into a grid map based on the longitude and latitude, and a region is represented by a grid.

These works are different from our work. We propose a new concept about regional traffic, the reasonable region Ω -region. An Ω -region is a subgraph of road network, which is consisted of several connected road segments with similar traffic situation. As far as we know, this is the first study on reasonable regional traffic estimation. To solve the problem, we first study how to generate those regions based on the historical traffic information. Next, we study how to predict future regional traffic effectively. We find the regional traffic in a coarse-grained level follows a linear relationship; thus, we propose an effective linear-based model to estimate regional traffic.

3. Preliminaries

Road Network. We model the road network as an undirected graph $G=(V, E)$, where V is the vertex set and E is the edge set. Each vertex of V is composed by a pair of geo-coordinates, i.e., latitude, longitude. Figure 1(a) shows an example of road network.

Region. We define the region based on the road network. A region is a disjointed subgraph of G . The road network G consisted of k disjoint subgraphs $G_1=(V_1, E_1)$, $G_2=(V_2, E_2)$, \dots , $G_k=(V_k, E_k)$, where $E_1 \cap E_2 \cap \dots \cap E_k = \emptyset$. We denote $|\mathcal{E}_i|$ as the number of edges in G_i .

Regional Traffic. We focus on traffic volume for regional traffic in this paper. To measure the traffic volume in a region, we utilize the vehicle trajectories. A trajectory T is a series of geo-coordinates which is generated by the GPS devices of moving vehicles, i.e., $T=(ts_i, lat_i, lng_i)$ ($1 \leq i \leq T$), where ts_i is its timestamp, lat_i is the latitude of the vehicle, and lng_i is the longitude. Given a region G_i and a time interval $P_t=[P_s, P_e]$, the regional traffic $R(G_i, P_t)$ is defined as the average number of vehicles which appear within the edges of subgraph G_i between the time P_s and P_e . Formally,

$$\mathcal{R}(\mathcal{G}_x, \mathcal{P}_t) = \frac{|\{\mathcal{T} | ts_i \in [P_s, P_e], (lat_i, lng_i) \in \mathcal{G}_x\}|}{|\mathcal{G}_x|} \quad (1)$$

Problem Statement. In this paper, we study two problems. (1) Given a road network work G , we first study how to divide the road network into reasonable regions G_1, \dots, G_k , which is named Ω -region. All the road segments within the Ω -region have similar traffic situations as the traffic(s) on a one-way road or a two-way road are considered as a single one (we will introduce the details in Section 4). (2) Given a time interval P_t and an Ω -region G_i , we predict the regional traffic $R(G_i, P_t)$ based on previous traffics P_{t-1}, P_{t-2}, \dots of the road network.

Example 1. Figure 1(b) shows the regional traffic on the road network of Figure 1(a). The points shown on the edge are the trajectories of the vehicles for a given time period P_t . We divide the original road network G into nine regions, in which some edges are clustered into four regions, i.e., G_1, G_2, G_3 , and G_4 . Within these regions, we use four colors, i.e., dark red, red, orange, and yellow, to represent different traffic situations. In particular, $R(G_1, P_t) = 1$, $R(G_2, P_t) = 2$, $R(G_3, P_t) = 3$, and $R(G_4, P_t) = 7$.

4. Regional Traffic

4.1. Basic Idea. Intuitively, the traffic situations of two close locations are usually similar. If one place is suffered from heavy traffic jams, the nearby places will be jammed too. For example, in Beijing, the West Railway Station is a place where traffic congestions happen frequently, while the nearby place Gongzhufen is also always suffered from heavy traffic. This is because that the traffic situations in one place can easily affect the nearby place on the road network, as the vehicles have to move from one road segment to the adjacent one. In this section, we study how to divide the road network into separate regions with the property that the road segments within a region have similar traffic situations. A naive method is to divide the road network according to the spatial proximity, e.g., to partition the road network into grids. However, such method does not consider the traffic conditions on the road segments. Apparently, the traffic situations vary for some connected road segments (e.g., segments connected through crosses or main roads); thus, it cannot guarantee the consistency of traffic situations in those divided partitions by gridding.

To this end, we devise a novel method to find Ω -region named REGION, which considers both the graph topology and the trajectories together in the region partitioning. We first attach each trajectory to the nearest edge on the road network, and then we construct a new graph by assigning the weight of edges as the amounts of trajectories appear in the corresponding segments. Finally, we iteratively merge the edges that are connected both on road network and with similar amounts of trajectories by utilizing hierarchical clustering technique. Next, we formally introduce our method.

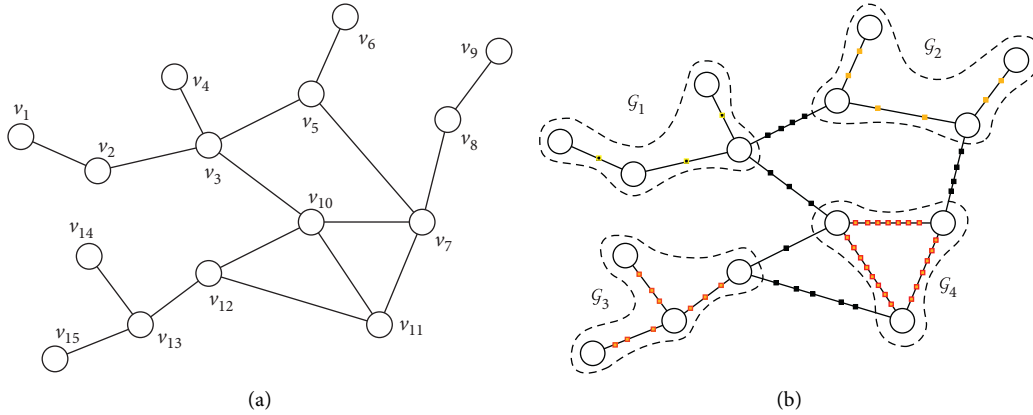


FIGURE 1: An example of the road network with trajectories. (a) A road network. (b) Regional traffic.

4.2. Region Finding

REGION involves three phases: (1) trajectory positioning; (2) graph building; and (3) hierarchical clustering. In this section, we discuss these three phases in detail.

Trajectory Positioning. For each trajectory $T = (ts_i, lat_i, lng_i)$, we find an edge $e \in E$ that is closest to (lat_i, lng_i) . We can use the linear scan to find the closest edge. To speed up this procedure, we build the quadtree to hierarchically index all the edges of the road network and employ the best-first search algorithm to find the nearest edge. We omit the details of this, as it is not the focus of this paper.

Graph Construction. Based on the results of trajectory positioning, we construct a new graph $G' = (V, E)$. For each edge (segment) $e \in E$, we assign a weight, which is denoted by $w(e)$. The weight is assigned as the amount of trajectories which lies on edge e . The $w(e)$ is defined by

$$\omega(e) = |\{\mathcal{T} | \mathcal{T} \in e\}|. \quad (2)$$

The weight can reflect the historical traffic situations. If the amounts of two close edges differ greatly, they are very likely to have different traffic situations. Therefore, those two edges should not be partitioned into the same region.

Hierarchical Clustering. We adopt a hierarchical clustering to generate the regions. First, we initialize each edge $e \in E$ as an independent group. Then, we iteratively merge two groups together by two rules. The first rule is that two groups, i.e., G_1 and G_2 , are topologically adjacent. The second rule is that the difference of the average weight between two groups is less than a threshold λ , i.e., $|\bar{w}(\mathcal{E}_1) - \bar{w}(\mathcal{E}_2)| \leq \lambda$, where $\bar{w}(\mathcal{E}_x) = \sum \omega(e \in \mathcal{E}_x) / |\mathcal{E}_x|$. We stop this process until the difference of the group weight is larger than λ . λ reflects the granularity of the regions we retrieve. If λ is small, we can obtain many small consistent regions; otherwise, we will divide the road network into less larger regions. λ is decided and tuned by applications

and experiences. We demonstrate the pseudo-code of this procedure in Algorithm 1 in Figure 2.

In Algorithm 1 in Figure 2, we first assign a weight to each edge (line 4) and count adjacent edges (line 6–7). Meanwhile, we initialize a queue Q to store the edges as independent groups (line 8). For each group G_x in Q , we calculate the difference of the average weight between its adjacent group and it, respectively (line 12). If there is an adjacent group $G_{x'}$ satisfying the condition, we merge it with G_x as G_y , and update adjacent groups of G_y and all the other groups containing G_x and $G_{x'}$ (line 13–18). Finally, we push G_y into queue Q (line 20). Otherwise, we consider G_x as a generated region, that is, Ω -region, and push it into set T (line 22). We repeat these steps until Q becomes empty, and all the regions in T are the Ω -regions from the graph of road network (line 23).

Example 2. Figure 3 shows the process of the hierarchical clustering. First, we use different colors to label the edges which initially belong to different groups. Then, we find two groups which are connected on the road network and have the average weight within the threshold λ . We combine these two groups together. We repeat this procedure until we cannot find any two groups that the difference of their average weight is smaller than λ . In Figure 3, we finally divide the road network into nine regions.

5. Traffic Estimation

In this section, we discuss how to estimate the traffic situation of a given region for next period of time.

5.1. Fundamental. In daily life, vehicles always move along a path of road segments. Since we have already divided the road network into regions, the vehicles have only three options to get the destination: (1) stay in the current region, (2) go in, or (3) go out to other regions. For example, in Figure 1, suppose a car is heading to v_8 from v_{13} . The path from v_{13} to v_8 is $v_{12} \rightarrow v_{10} \rightarrow v_7 \rightarrow v_8$. It will leave the region G_3 at v_{12} , and then it has to bypass several regions like G_4 to arrive at the region G_2 . Therefore, given a specific

Algorithm 1: Region (\mathcal{G}, λ)

Input: \mathcal{G} : the graph of road network
 λ : the threshold to tune the granularity of regions

Output: T : the set containing the generated regions $\{G_1, \dots, G_k\}$

```

1  Begin
2  Queue  $Q = \emptyset$ ;
3  for  $e \in \mathcal{E}$  do
4     $w(e)$  = number of trajectories on  $e$ ;
5     $G_e = \{e\}$ ;
6    for each adjacent edge  $e'$  of  $e$  do
7       $\mathcal{A}(G_e).add(G_{e'})$ 
8     $Q.push(G_{e'})$ ;
9  while  $Q \neq \emptyset$  do
10    $G_x = Q.pop()$ ;  $G_y = \emptyset$ ;
11   for each  $G_{x'} \in \mathcal{A}(G_x)$  do
12     if  $Abs\left(\frac{\sum_{e \in G_x} w(e)}{|G_x|} - \frac{\sum_{e \in G_{x'}} w(e)}{|G_{x'}|}\right) < \lambda$  then
13        $G_y = G_x \cup G_{x'}$ ;
14        $\mathcal{A}(G_y) = \mathcal{A}(G_x) \cup \mathcal{A}(G_{x'})$ ;
15       for each  $G_{x''} \in \mathcal{A}(G_x) \cup \mathcal{A}(G_{x'})$  do
16          $\mathcal{A}(G_{x''}).remove(G_x, G_{x'})$ ;
17          $\mathcal{A}(G_{x''}).add(G_y)$ ;
18       break;
19   if  $G_y \neq \emptyset$  then
20      $Q.push(G_y)$ ;
21   else
22      $T.add(G_x)$ ;
23 return  $T$ ;
24 end

```

FIGURE 2: Region finding algorithm.

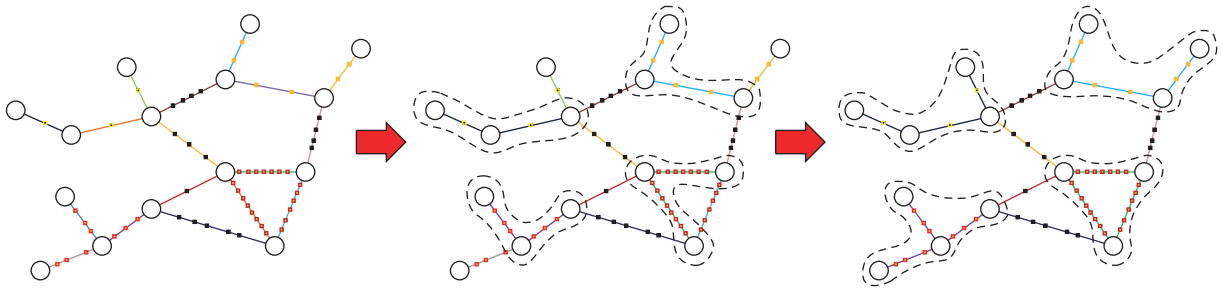


FIGURE 3: The process of the hierarchical clustering.

region, the traffic condition of such region will only be affected by the traffic situation of adjacent regions in the previous period of time. If one region is suffering from the traffic jams, it will then give an impact on the adjacent regions sometime later. Moreover, according to the closure property of the regions on road network, the vehicles that go out from a region must go in another adjacent region, while the total amount of transitions is unchanged.

The above facts indicate that we can use the linear model to address this problem. Next, we formally define our model. We denote our method by LINEST.

5.2. Estimation Model. First of all, we divide the daily time into m intervals, i.e., P_1, P_2, \dots, P_m . We denote the adjacent regions of a given region G_x by $\mathcal{A}(G_x)$, where the adjacent

regions of G_x are the regions that connected to G_x with at least one edge in G . Given a specific region G_x at the time interval P_t , our goal is to estimate the regional traffic $R(G_x, P_{t+1})$, $R(G_x, P_{t+2})$, etc., based on $\mathcal{R}(\mathcal{G}'_{\mathcal{G}' \in \mathcal{A}(G_x)}, \mathcal{P}_t)$.

According to the previous analysis, we use a linear model to depict the relationship among this regional traffic, which is shown as

$$\mathcal{R}(G_x, \mathcal{P}_t) = \alpha_0 + \sum_{\mathcal{G}' \in \mathcal{A}(G_x)} \alpha_{\mathcal{G}'} \cdot \mathcal{R}(G', \mathcal{P}_{t-1}). \quad (3)$$

Equation (2) has good interpretability. The parameters $\alpha_1, \alpha_2, \dots, \alpha_{G'}$ in equation (2) indicate the proportions of vehicles that transport between the regions G_x and $\mathcal{A}(G_x)$ from the previous period of time to the current period. The parameter α_0 is a constant factor which suggests the usual traffic situation at the current period of a day.

Notice that, as we have divided the daily time into m intervals P_1, P_2, \dots, P_m , thus, for a single region G_x , we can use m distinct models for a single day. This is because that the underlying patterns of vehicles' moving are different within a whole day. For example, in the morning, the cars are prone to move to downtown area, while in the evening, those cars tend to drive off from downtown. Based on this observation, we separately treat m models for region G_x within a day.

Equation (2) also suggests that we can employ the technique of linear regression to solve the parameters $\alpha_1, \alpha_2, \dots, \alpha_{G'}$. Our problem can be formulated by the linear regression as follows:

$$\mathcal{R}(G_x, \mathcal{P}_t) = X_{t-1} \beta_t + \varepsilon, \quad (4)$$

where $\beta_t = [\alpha_{t_0}, \alpha_{t_1}, \alpha_{t_2}, \dots, \alpha_{t_{|\mathcal{A}(G_x)|}}]^T$ and X is a $n \times (|\mathcal{A}(G_x)| + 1)$ matrix, where n rows describe the traffic situations of region G_x and its neighbors in $\mathcal{A}(G_x)$ at time period P_{t-1} during n different days (n can be understood as the number of training data used to solve β_t). To minimize the overall sum of squared estimating errors, i.e., $\sum \varepsilon$ (given a region, we can compute a standard variation of vehicles' amounts over edges within the region; the average standard variation is the mean of the standard variations over all regions), we use its close-form solution, which is shown below.

$$\hat{\beta}_t = (X_{t-1}^T X_{t-1})^{-1} X_{t-1}^T \mathcal{R}(G_x, \mathcal{P}_t). \quad (5)$$

We take the learned parameters into equation (2) to estimate the regional traffic $R(G_x, P_{t+1})$.

6. Experiments

In this section, we first introduce the datasets and experimental settings. Then, we present the experimental results and analysis of comparative performance.

6.1. Experiment Setup

6.1.1. Dataset. To evaluate our method, we used the detailed road network of Beijing in 2012. The vertex size is 1,278,984,

and the edge size is 2,402,784. We also obtained the trajectory data of taxis in Beijing from October 2012 to December 2012. The total number of trajectories is 3.05 billion. We divided a daytime into 96 intervals; i.e., each interval lasts 15 minutes.

6.1.2. Baseline Methods. As we discuss in Section 2, regions in the existing methods are defined as a grid, which is constructed by directly dividing map based on geo-coordinates. Therefore, these works are not comparable with ours. To prove the effectiveness of our methods, we collect the ground truth data and compare our methods REGION and LINEST with two baseline methods GRID and HISTORY, respectively.

- (i) GRID. GRID is a straightforward method to divide road network. It utilized geo-coordinates to divide road network directly.
- (ii) HISTORY. We collected historical data of road network in Beijing as the ground truth data. And HISTORY is computed as the average amounts of vehicles based on only historical data.

6.1.3. Evaluation Metrics. We implemented REGION and LINEST with Python 3.4. All experiments were run on a Linux 10.0.4 machine with a 3.2 GHz CPU and 8 GB RAM.

6.2. The Effectiveness of REGION. We evaluated the effectiveness of REGION. First, we tested the number of generated regions by varying the threshold λ as 32, 64, 128, and 256. Figure 4(a) shows the results. We can see that the number of regions decreases greatly with the increase of λ ; this is because the region finding algorithm will merge more road segments into regions when λ is large. We generated total 274 K regions when $\lambda = 32$, and if $\lambda = 512$, we obtained 40 K regions.

Next, we tested the average standard variations of the amounts of vehicles which appear on the edges within corresponding regions (given a region, we can compute a standard variation of vehicles' amounts over edges within the region; the average standard variation is the mean of the standard variations over all regions). If the road segments within a region have similar traffic situation, it will have a small standard variation. Otherwise, the standard variation will be large. Figure 4(b) shows the average standard variations when we varied λ . We compared REGION with GRID, which divided road network into regions based on their geo-coordinates. The number of grids is determined by the corresponding regions of REGION. For example, if $\lambda = 32$, we generate 274 K grids. We can see that REGION outperforms GRID with a much smaller standard variation, which means that the traffic situations in regions generated by REGION are more consistent than the regions obtained from GRID. In addition, we also show the evolvment of average standard variation over recent 20 days in

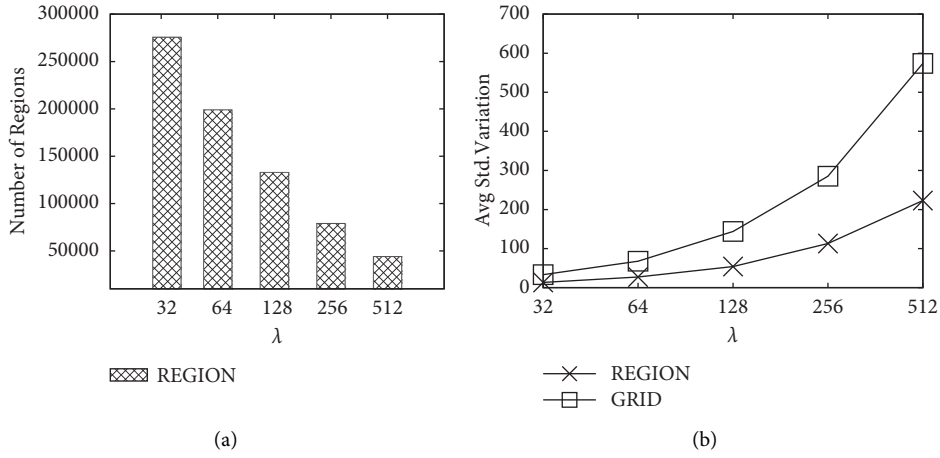


FIGURE 4: Varying threshold λ . (a) Number of generated regions. (b) Average standard variations.

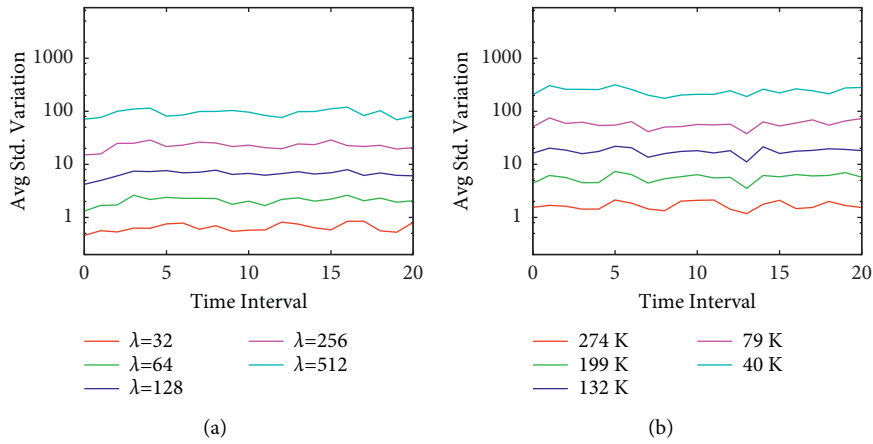


FIGURE 5: Average standard variations over 20 days. (a) REGION. (b) GRID.

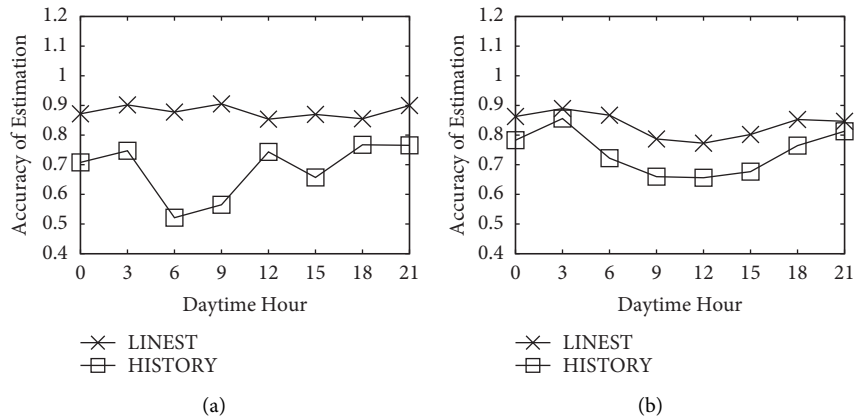


FIGURE 6: Regional traffic estimation. (a) Downtown. (b) Uptown.

Figure 5. It shows that REGION can reach a very stable performance over different days.

6.3. Regional Traffic Estimation. We evaluated the effectiveness of LINEST by using the accuracy ratio computed as follows:

$$\text{Accuracy} = 1 - \frac{\sum \text{Abs}(\mathcal{R}(\mathcal{E}_x, \mathcal{P}_t) - \mathcal{R}(\mathcal{E}_{\bar{x}}, \mathcal{P}_t))}{\sum \mathcal{R}(\mathcal{E}_x, \mathcal{P}_t)}, \quad (6)$$

where $\mathcal{R}(\mathcal{E}_{\bar{x}}, \mathcal{P}_t)$ is the estimating value of regional traffic and $\text{Abs}(\mathcal{R}(\mathcal{E}_x, \mathcal{P}_t) - \mathcal{R}(\mathcal{E}_{\bar{x}}, \mathcal{P}_t))$ is the estimating error. The trajectories within the last week are extracted as test

datasets. We further divide the test data into two categories: *Downtown* and *Uptown*. *Downtown* is the region located within the fourth ring road of Beijing (the center area of Beijing). *Uptown* is the other region of Beijing. We reported the accuracy by varying the hours of daytime. We compared LINEST with HISTORY, which computed as the average amounts of vehicles from only historical data. The results are shown in Figure 6. LINEST achieved very high performance comparing with HISTORY. The average accuracy reaches 90%. For example, the accuracy of downtown at 9 am can reach 91.2%. We also found that the performance increases from the uptown area to the downtown area, which indicates the linear model is more fit to the traffic situations of downtown.

7. Conclusion

In this paper, we study the regional traffic estimation problem. We first discussed how to divide road network into reasonable regions. We propose a three-phase region generating algorithm by clustering historical trajectory data. Next, we studied how to estimate future regional traffic effectively. By considering the linear relationship between the regional traffics, we proposed a linear-based model. Experimental results on read dataset show our method achieves high performance. In the next step, we will further optimize the algorithm to adapt to dynamic traffic estimation. In the future, we can consider on how to improve our model for granular-level transportation by constructing regions in higher dimension on the premise of ensuring the effectiveness of the existing model.

Data Availability

The data used to support the findings of this study have not been made available because these data are from our co-operative company. They only authorize us to use data for analysis, but not publish data.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the Fundamental Research Funds for the Central Universities (N2117001).

References

- [1] H. Yuan and G. Li, "A survey of traffic prediction: from spatio-temporal data to intelligent transportation," *Data Science and Engineering*, vol. 6, no. 1, pp. 63–85, 2021.
- [2] Y. Qi and S. Ishak, "Stochastic approach for short-term freeway traffic prediction during peak periods," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 2, pp. 660–672, 2013.
- [3] B. Yang, C. Guo, and C. S. Jensen, "Travel cost inference from sparse, spatio-temporally correlated time series using Markov models," *Proceedings of the VLDB Endowment*, vol. 6, no. 9, pp. 769–780, 2013.
- [4] Y. Wang, Y. Zheng, and Y. Xue, "Travel time estimation of a path using sparse trajectories," in *Proceedings of the 20th SIGKDD conference on Knowledge Discovery and Data Mining (KDD 2014)*, pp. 25–34, ACM, New York, NY, USA, August 2014.
- [5] J. Snowdon, O. Gkoutouna, A. Zufle, and D. Pfoser, "Spatiotemporal traffic volume estimation model based on GPS samples," in *Proceedings of the the Fifth International ACM SIGMOD Workshop*, pp. 1–6, ACM, New York, NY, USA, 2018.
- [6] L. Lin, J. Li, F. Chen, J. Ye, and J. Huai, "Road traffic speed prediction: A probabilistic model fusing multi-source data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 7, pp. 1310–1323, 2018.
- [7] T. Liebig, N. Piatkowski, C. Bockermann, and K. Morik, "Route planning with real-time traffic predictions," *Information Systems*, vol. 1226, pp. 83–94, 2014.
- [8] J. Shang, Y. Zheng, W. Tong, E. Chang, and Y. Yu, "Inferring gas consumption and pollution emission of vehicles throughout a city," in *Proceedings of the 20th SIGKDD conference on Knowledge Discovery and Data Mining*, pp. 1027–1036, ACM, New York, NY, USA, 2014.
- [9] Y. Zhu, Z. Li, H. Zhu, M. Li, and Q. Zhang, "A compressive sensing approach to urban traffic estimation with probe vehicles," *IEEE Transactions on Mobile Computing*, vol. 12, no. 11, pp. 2289–2302, 2013.
- [10] T. Darwish and K. Abu Bakar, "Traffic density estimation in vehicular ad hoc networks: A review," *Ad Hoc Networks*, vol. 24, pp. 337–351, 2015.
- [11] A. Kaminska-Chuchmala, "Spatial internet traffic load forecasting with using estimation method," *Procedia Computer Science*, vol. 35, pp. 290–298, 2014.
- [12] Z. Liu, P. Zhou, Z. Li, and M. Li, "Think like A graph: Real-time traffic estimation at city-scale," *IEEE Transactions on Mobile Computing*, vol. 18, no. 10, pp. 2446–2459, 2019.
- [13] B. Pan, U. Demiryurek, and C. Shahabi, "Utilizing real-world transportation data for accurate traffic prediction," in *Proceedings of the 2012 IEEE 12th International Conference on Data Mining*, pp. 595–604, IEEE, Brussels, Belgium, December 2012.
- [14] L. Chen and C. L. P. Chen, "Ensemble learning approach for freeway short-term traffic flow prediction," in *Proceedings of the 2007 IEEE International Conference on System of Systems Engineering*, pp. 1–6, IEEE, San Antonio, TX, USA, April 2007.
- [15] R. Herring, A. Hofleitner, P. Abbeel, and A. Bayen, "Estimating arterial traffic conditions using sparse probe data," in *Proceedings of the 13th International IEEE Conference on Intelligent Transportation Systems*, pp. 929–936, IEEE, Funchal, Portugal, September 2010.
- [16] H. Yuan, G. Li, Z. Bao, and L. Feng, "Effective travel time estimation: When historical trajectories over road networks matter," in *Proceedings of the SIGMOD/PODS '20: International Conference on Management of Data*, pp. 2135–2149, ACM, New York, NY, USA, 2020.
- [17] R. Shi, Z. Mo, and X. Di, "Physics-informed deep learning for traffic state estimation: A hybrid paradigm informed by second-order traffic models," *IEEE Transactions on Intelligent Transportation Systems*, pp. 540–547, 2021, AAAI, AAAI Press.
- [18] M. Sederlin, X. Ma, and J. Jin, "A hybrid modelling approach for traffic state estimation at signalized intersections," in *Proceedings of the 2021 IEEE International Intelligent*

- Transportation Systems Conference (ITSC)*, pp. 3604–3609, IEEE, Indianapolis, IN, USA, September 2021.
- [19] M. X. Wang, W. C. Lee, T. Y. Fu, and G. Yu, “Learning embeddings of intersections on road networks,” in *Proceedings of the the 27th ACM SIGSPATIAL International Conference*, pp. 309–318, 2019.
- [20] X. Zhan, Y. Zheng, X. Yi, and S. V. Ukkusuri, “Citywide traffic volume estimation using trajectory data,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 2, pp. 272–285, 2017.
- [21] H. Yuan, G. Li, Z. Bao, and L. Feng, “An effective joint prediction model for travel demands and traffic flows,” in *Proceedings of the 2021 IEEE 37th International Conference on Data Engineering (ICDE)*, pp. 348–359, IEEE, New York, NY, USA, 2021.
- [22] X. Ma, Z. Dai, Z. He, J. Ma, Y. Wang, and Y. Wang, “Learning traffic as images: A deep convolutional neural network for large-scale transportation network speed prediction,” *Sensors*, vol. 17, no. 4, p. 818, 2017.
- [23] J. Wang, Q. Gu, J. Wu, G. Liu, and Z. Xiong, “Traffic speed prediction and congestion source exploration: A deep learning method,” in *Proceedings of the 2016 IEEE 16th International Conference on Data Mining (ICDM)*, pp. 499–508, IEEE, Barcelona, Spain, December 2016.
- [24] J. Zhang, Y. Zheng, and D. Qi, “Deep spatio-temporal residual networks for citywide crowd flows prediction,” *AAAI*, pp. 1655–1661, 2017.
- [25] L. Zhao, Y. Liu, A. Y. Al-Dubai, A. Y. Zomaya, G. Min, and A. Hawbani, “A novel Generation-Adversarial-network-based vehicle trajectory prediction method for intelligent vehicular networks,” *IEEE Internet of Things Journal*, vol. 8, no. 3, pp. 2066–2077, 2021.
- [26] L. Zhao, Z. Li, A. Y. Al-Dubai et al., “A novel prediction-based temporal graph Routing Algorithm for Software-defined vehicular networks,” *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–16, 2021.
- [27] L. Zhao, T. Zheng, M. Lin, A. Hawbani, J. Shang, and C. Fan, “SPIDER: A Social computing Inspired predictive Routing Scheme for Softwarized vehicular networks,” *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–12, 2021.
- [28] Y. Kang, J. Li, S.-J. Lee, and H. Li, “Generative adversarial network-based regional epitaxial traffic flow prediction,” *Advances in Natural Computation, Fuzzy Systems and Knowledge Discovery*, vol. 1075, pp. 804–814, 2019.
- [29] L. Zhu, T. Li, and S. Du, “TA-STAN: A deep spatial-temporal attention learning framework for regional traffic accident risk prediction,” in *Proceedings of the 2019 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, IEEE, Budapest, Hungary, September 2019.
- [30] Z. Wang, Y. Chen, J. Su et al., “Measurement and prediction of regional traffic volume in holidays,” in *Proceedings of the 2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, pp. 486–491, IEEE, Auckland, New Zealand, October 2019.
- [31] Z. Liu, M. Huang, Z. Ye, and K. Wu, “Deeprtp: A deep spatio-temporal residual network for regional traffic prediction,” in *Proceedings of the 2019 15th International Conference on Mobile Ad-Hoc and Sensor Networks (MSN)*, pp. 291–296, IEEE, Shenzhen, China, December 2019.