2016 IEEE International Congress on Big Data

Urban Area Vehicle Number Estimation based on RTMS data

Yue Hu⋆, Yuanchao Shu†, Peng Cheng⋆, Jiming Chen⋆
⋆State Key Laboratory of Industrial Control Technology
Zhejiang University, Hangzhou, China
†Microsoft Research Asia, Beijing, China
hyuezju@gmail.com, yushu@microsoft.com, {pcheng, jmchen}@iipc.zju.edu.cn

***Abstract*—Along with the increase of vehicle ownership, the traffic problem has a serious impact on people's daily life. Not only the traffic congestion, but also the parking problem troubles urban daily traveling. Therefore it is important to obtain the parking demand to help the government to make a rational decision on traffic planning and management. This paper focuses on estimating the vehicle number in a certain area (i.e., the spaces surrounded by the arterial roads) in each time slot to analyze the area parking demand, using RTMS (Remote Traffic Microwave Sensor) data. We first propose a basic method to calculate the AVN (Area Vehicle Number) based on the inflow and outflow traffic of the area. In order to correct the error caused by minor roads without RTMS data, we propose an advanced method to improve the estimation accuracy by exploiting the road traffic correlation from a network perspective. Comprehensive evaluation is conducted to verify our design based on large amount of RTMS data from the Hangzhou city during one month. The estimation results also demonstrate interesting human behaviors among various urban areas.**

I. Introduction

The number of vehicles has been growing tremendously in various cities which makes the traffic congestion a global issue. According to the report of the Texas A&M Transportation Institute at 2015, the traffic congestion has cost 960 dollars and 42 hours per person annually in the U.S. [1]. As a result, studies on moving vehicles (e.g., traffic speed prediction [2], traffic flow prediction [3], and etc) and public transportations (e.g., bike sharing system [4], express system [5]) have been widely presented in literature.

At the same time, as a critical element of traffic, parking problem has been attracting increasing attention in many countries. For example, there were more than 2.5 million vehicles in Hangzhou by the end of 2013 [6], and the gap between parking space and parking demand was more than 650 thousand [7]. Note that such gap in city scale does not even accurately reveal the situation of temporary hotspots where the gap may be much larger. With the fast urbanization in many developing countries, such parking difficulty becomes more obvious, resulting in unauthorized parking or long-time search for parking slots. Therefore, in order to accurately reflect the area traffic status, it is crucial to estimate numbers of both moving vehicles and parked vehicles within the corresponding area, which we define as Area Vehicle Number (AVN) in this paper.

The estimation of AVN is of significance from two aspects. First, AVN including both the moving vehicles and the parked ones reflects the vehicle capacity of the area, and can also estimate the parking space capacity by subtracting the number of moving vehicles on the roads which may be inferred from the traffic speed [8]. Second, the dynamic AVN helps to better reveal the human mobility patterns that are closely related to the functionality of the area. It can be further exploited for better civil infrastructure planning such as parking lots. For instance, if we can provide enough parking space to meet the actual demand, the on-street parking will be reduced substantially.

Although it is possible to recover the moving vehicle number through the traffic speed, the number of parked vehicles is difficult to infer due to the lack of the measurement data. For example, the vehicles may be parked at various lots where the occupancy status is not well observed, like the street lots without occupancy sensors, or open space of residential areas. The existing parking demand prediction methods [9] only provide a prediction of the number of parked vehicles with coarser granularity through costly onsite survey.

In this work, we aim to establish a general framework for accurate AVN estimation by exploiting the available large amount of traffic data, especially the Remote Traffic Microwave Sensor (RTMS) data. Intuitively, it is direct to obtain the AVN by adding up the directed traffic flow for the specific area. However, the problem is very challenging due to the incomplete RTMS measurements in both time and space domains. Furthermore, the estimation error may accumulate along with the time.

In this paper, we propose a systematic work to estimate the AVN using the large amount of RTMS data and road network data. Specifically, we explore the temporal and spatial correlations among the RTMS measurements and the road segments for completing the missing data. Meanwhile, we correct the accumulating estimation error dynamically by exploiting the periodic human behavior. The intellectual contributions of this paper can be summarized as follows:

- To the best of our knowledge, we first propose a systematic method to estimate the area vehicle number using RTMS data.
- We infer the parking demand based on the inflow and

978-1-5090-2622-7/16 $31.00 © 2016 IEEE
DOI 10.1109/BigDataCongress.2016.59

outflow traffic, and improve the results by connecting roads balance the traffic flow from a network perspective. At last, we combine the statistic information of the area to obtain the gap between parking space and people regular demand.

- We perform comprehensive evaluation based on more than 3.5 million RTMS data from the city of Hangzhou during one month and compare the estimated parking demand to the vehicle ownership data, and the mean relative error is about 10% in the cases.

The rest of this paper is organized as follows. We first describe our datasets and preliminary data processing in Section II. In Section III, we introduce our method to estimate the AVN based on RTMS data including both basic design and advanced design. Then we evaluate the method using three typical areas in Hangzhou in Section IV. Related works are discussed in Section V and we make a conclusion of this paper in Section VI.

II. Preliminary

In this section, we briefly introduce the traffic flow data collected by RTMS and the road network data from Hangzhou. After that, we propose methods to compensate the incomplete RTMS measurements of road segments.

A. Background

Remote Traffic Microwave Sensor (RTMS) is widely installed on roads to count real-time traffic nowadays. For example, there are more than 570 such sensors deployed in Hangzhou at the end of 2013. Comparing to the other traffic flow sensors, the RTMS produce traffic counts that are up to 97 percent accurate [10]. Furthermore, RTMS is a low-cost, general-purpose, all-weather traffic sensor which detects presence and measures traffic parameters in multiple independent zones [11].

1) RTMS Dataset: The RTMS dataset is collected from Dec. 1st, 2013 to Dec. 31st, 2013, and the total number of the records is about 25 million where the sampling period is 1 minute. The sensors (RTMS is deployed on the middle of the roads) mainly measure the traffic flow space (i.e., the number of vehicles passing the sensors per minute) of 515 arterial road segments (the number of sensors deployed in some important arterial roads is more than one). Moreover, each record consists of five fields, including ID of the road segment, ID of the microwave sensor, number of the lane of the road segment, total traffic flow of the lane of the road segments and the record time.

2) Road Network Dataset: We use the road network dataset of urban area of Hangzhou from the government, which includes 1759 arterial road segments. Each record consists of road code, start of the road, end of the road and nodes (one complete road is divided into individual road segments without intersection, the node is the point of division) with longitude and latitude.

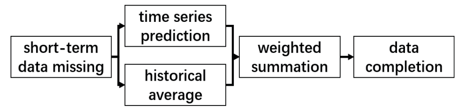


Figure 1: Traffic flow completion methodology

B. Data Preprocessing

We consider the traffic data completion in this part. We first convert the sampling time into 15 min (the number of records is 96 per day). In the following, we show how to compensate the missing RTMS data (missing means some of the sensor measurement records are missing).

We combine a grey model (GM(1,1)) with historical data to design a strategy to fill the missing values as Figure 1 shows. Grey forecasting model is a time series forecasting model and using a small amount of incomplete information to establish a prediction model. Grey forecasting model has been used in many applications, like traffic flow/speed prediction, power demand forecasting, stock price prediction and so on. The GM(1,1) is one of the most frequently used grey forecasting models [12].

Note that the grey model (GM(1,1)) is able to characterize the short-term correlation of RTMS measurements for estimating the missing value, while the historical average value during the same time of the day and day of the week is able to represent the periodic pattern of traffic flow which captures the long-term correlation of RTMS measurements. Thus, in order to improve the accuracy, a weighted summation method is adopted to fuse the two value as Figure 1 shows. The variables we use in the grey model is as follows: 1) $f(t)$ is the value to be estimated. $f(t)$ demonstrates true traffic flow of time t, and we define $f(\hat{t})$ as the estimated value of $f(t)$. 2) The training sample is as $f(t-1), f(t-2), f(t-3), ..., f(t-n)$, which represents the traffic flow at time $t-1, t-2$, and etc.

In order to calculate the result, the length of the sliding window (n) needs to be optimized. Therefore, we choose the whole day flow data as the training set without records missing to find an appropriate n. Specifically, the estimation $f(\hat{t})$ is obtained by a sliding window (from $f(t-1)$ to $f(t-n)$) to get the training sample with a certain length of n, and the ground truth is $f(t)$. Then, by calculating the absolute error between the estimation $f(\hat{t})$ and the ground truth $f(t)$ with different n, we can determine the optimal n with the minimum absolute error.

Additionally, because the traffic flow data has strong cyclical characteristics [13], we also calculate the historical value of the flow data at the same time of the day and the day of the week to improve the GM(1,1) estimation. Specifically, we propose the following weighted linear combination for the estimated value $f(\hat{t})$: $f(\hat{t}) = \alpha x_1 + \beta x_2 \quad (\alpha + \beta = 1)$, where x_1 is the result by direct GM (1,1) prediction, and x_2 is the historical average value. In order to obtain the

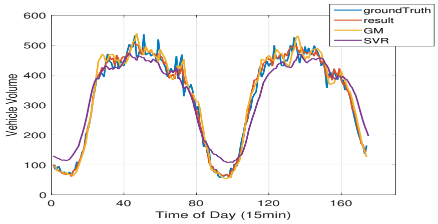


Figure 2: the RTMS data completion

optimized value of weights α and β, we define the objective function $J(\alpha)$ to minimize the RMSE of the estimation:

$$J(\alpha) = \frac{1}{2}\sum_{i=1}^{m}(\alpha x_1^{(i)} + \beta x_2^{(i)} - y^{(i)})^2 \tag{1}$$

By taking the derivative, we have

$$\frac{\partial J(\alpha)}{\partial \alpha} = -\sum_{i=1}^{m}(\alpha x_1^{(i)} + (1-\alpha)x_2^{(i)} - y^{(i)})x_1^{(i)}x_2^{(i)} = 0 \tag{2}$$

which means $\alpha = \frac{Y - X_2}{X_1 - X_2}$. Thus the optimal α and β can be calculated by the training samples.

Furthermore, since GM (1,1) estimation is close to the historical average due to the cyclicity of the traffic flow, the outliers can be identified if the deviation between the GM (1,1) estimation and the historical average value is quite larger, therefore we change the weights to decrease the influence of the outliers.

To briefly demonstrate the efficiency of our proposed method, we show the data completion performance in Figure 2 where we choose the rest as the training set, and the two-day data of a road segment as the testing set. Then the MAE (mean absolute error) and the MRE (mean relative error) are defined as

$$MAE = \frac{1}{n}\sum|\hat{f(t)} - g(t)| \tag{3}$$

$$MRE = \frac{1}{\sum g(t)}\sum|\hat{f(t)} - g(t)| \tag{4}$$

Figure 2 shows the completing result of one road segment base on our method, GM (1,1) and SVR respectively. The MAE and the MRE are (14.36, 4.36%), (22.79, 6.92%) and (36.7, 6.92%) for three approaches respectively.

III. AVN Estimation Design

In this section, we focus on the AVN estimation problem. First we propose a basic method to infer the AVN based on the principle of in/out vehicle approximated balance. Then by taking into account more information, we propose an advanced design to improve the AVN estimation accuracy.

In order to get more accurate AVN, the targeted area should be surrounded by minimum number of intersections. The road network can be represented as a directed graph $G = (N, E)$, where N denotes nodes (the intersection in road network) and E denotes links (each road segments) in road network topology. The arbitrary selected area is denoted by A, which is a subgraph of G. For each targeted area, the critical nodes are defined as the nodes representing the interaction points of links entering/exiting the area.

A. Basic Design of Area Vehicle Number Estimation

1) Basic Design Idea: The basic idea of AVN estimation is the approximated balance of the vehicle number moving in and out of the area for each day. More specifically, the vehicle number in the area is controlled by the amount of entrance and exit traffic for the selected area [14], and the AVN of the start and the end of a day should be relatively stable and approximately equivalent normally. Therefore we may infer the AVN by utilizing the flow data of links which run in/ out of the area.

(a) road network topology (b) real road network

Figure 3: The basic method

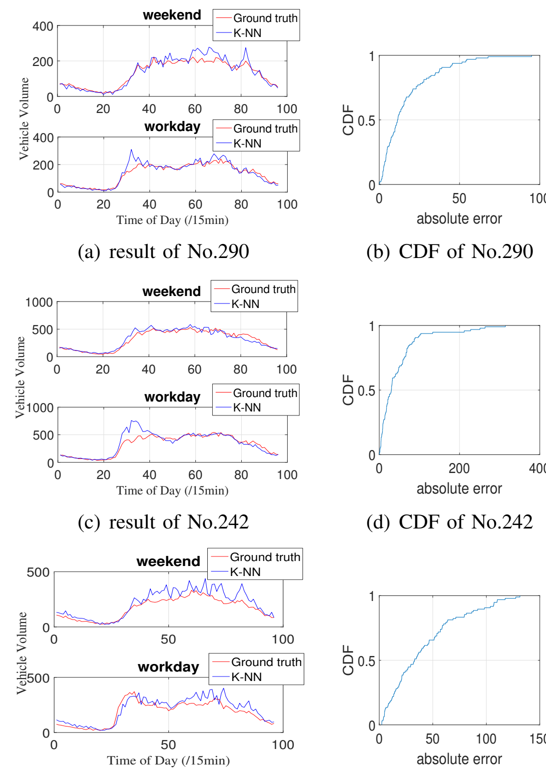
We explain the basic design idea through the example shown in Figure 3(a), where the black blocks represent the remote traffic microwave sensor, the dotted oval represent the area selected, black arrows indicate the flow direction of the key links. The basic idea is as follows. After selecting a certain area, we find out the key links which enter/exit the area, then set the initial AVN of the area to 0 and sum over all directional flow data of these links in each time slot of a day (from 00:00 to 23:59), which obtain the area vehicle number at last.

2) Compensate the RMTS data for arterial roads: One difficulty for the basic design is that we may not have the RTMS data for some key arterial links (in/out of the area) which are not equipped with the sensors. This problem is different from the previous data preprocessing where only part of the RTMS data is missing. In order to handle this difficulty, we utilize spatial correlation of links based on the road network topology, and adapt the weighted-KNN approach to estimate the missing data of key links from their neighbor links [15]. Specifically, we select the road segments connected to the arterial links directly as the neighbor links, and decide the weights according to the distance between them (length of the shortest path form the middle of neighbor segment to the middle of target segment). The road segment more close to the target road segment has the larger weight by inverse of distance [15].

Algorithm 1: the Basic Method of AVN Estimation

Input: $F \leftarrow$RTMS dataset, $G \leftarrow$road network graph, $A \leftarrow$selected area
Output: area vehicle number V

initialize $V \leftarrow$ zero vector
for *($r(i) \in G.E$)* **do**
  if *($r(i) \in A.E$)* **then**
    if *($r(i)! = null$)* **then**
      $R' \leftarrow r(i)$
    else
      $r(i) \leftarrow weighted - KNN$
      $R' \leftarrow r(i)$
    end
  end
end
for *($r(i) \in R'$)* **do**
  $\Delta V \leftarrow [\sum_{t=1}^{1} f(r(i),t), \sum_{t=1}^{2} f(r(i),t), ..., \sum_{t=1}^{end} f(r(i),t)]$
  if *(r(i) leave A)* **then**
    $V = V - \Delta V$
  end
  if *(r(i) enter A)* **then**
    $V = V + \Delta V$
  end
end
return V



(a) result of No.290 (b) CDF of No.290

(c) result of No.242 (d) CDF of No.242

(e) result of No.291 (f) CDF of No.291

Figure 4: KNN based Road RMTS Data Compensation

In order to demonstrate the effectiveness of our compensation, we choose three road segments (No.242, No.290, No.191), and show the estimation performance in Figure 4 where a typical estimation trace as well as the CDF results is demonstrated. The MAE and RMSE of the results are (18, 26.0953), (49, 76.1078), (43, 57.5097). We can see that the estimation is close to the ground truth for most time slots. However, for some time instance, the estimated traffic flow deviates relatively large from the ground truth. One important reason is that there exist some minor roads (as red curves shows in Figure 3(b)) which run in/out of the area and steal/input some traffic flow. Moreover, the estimation error at each time slot may be accumulated to increase the error along with time. Therefore, in the following part, we propose our advanced design for more accurate AVN estimation.

B. Advanced Design of Area Vehicle Number Estimation

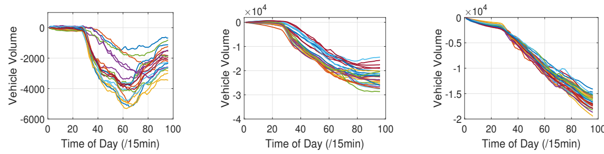
In this part, we aim to compensate the error caused by unknown minor road segments and the accumulated error. The proposed advanced design is based on two observations: 1) The number of vehicles which belong to the area should be equal during a period (24 hours for example), due to the regular daily behavior of most people living or working in the area. 2) Vehicles running on arterial roads can either end up at the next node or leave the road halfway (i.e., go to the parking place or leave the area through the minor road segments).

Based on the first observation, we set 24 hours as a period, and start from 0:00 to 23:59 each day for most areas, and check the differences between the initial vehicle number and the end vehicle number of the area for different days. If the RTMS data is accurate and complete for all the roads and time slots, such a sequence of gap should be relatively stable and close to zero since the difference mainly represents the number of vehicles which enter the area without leaving until 23:59 (or vice versa) each day. However, due to the stealing/input flow from minor roads as well as the error caused by other reasons, such as data preprocessing, the difference sequence fluctuates as shown in Figure 5. where we testify the results for three typical areas in Hangzhou city. The details of the three areas will be explained in Evaluation part.

Therefore, in the following part, we utilize the observations to correct the error. From Observation 1), we have the following equation for each day

$$\sum_t [f_{in,a}(t) + f_{in,m}(t)] = \sum_t [f_{out,a}(t) + f_{out,m}(t)] \quad t = 1, 2, ..., 96 \tag{5}$$

where $f_{in,a}(t)$ and $f_{in,m}(t)$ in Equation 5 represent the traffic flow entering the area through the arterial roads



(a) Area1 (b) Area2 (c) Area3

Figure 5: basic design of three areas

and the minor roads at time slot t respectively. $f_{out,a}(t)$ and $f_{out,m}(t)$ in Equation 5 means the traffic flow leaving the area through the arterial roads and the minor roads at time slot t respectively. We have converted the time slot into 15 min before, so the value of t is from 1 to 96 per day. We consider $\sum_t[f_{in,a}(t)] = \sum_t[f_{out,a}(t)]$ in our basic method based on the balance of traffic flow. But because of traffic flow of the minor roads, the error is $0 - (\sum_t[f_{in,a}(t)] - \sum_t[f_{out,a}(t))]$, which is equal to $(\sum_t[f_{out,m}(t) - f_{in,m}(t)])$. The error means the traffic flow in/out of the area through the minor roads. We use the connection of the roads and the balance of traffic flow to correct the error.

It should be note that from Figure 5, the difference varies largely for all three areas, which proves that the error is not a constant value but indeed correlated with the traffic flow of its neighbor roads. Therefore, a static correction over the whole area is not enough to compensate such a dynamic error. In order to effectively reduce the error, we propose a partition-and-correction approach.

We first partition the area into several sub-areas depending on the critical nodes because the traffic flow exits/enters the area through the critical nodes. And the points where the sensors are located on the arterial links are the boundaries of each sub-area. For each sub-area, the traffic flow should be balanced, and the unbalanced part calculated by accumulating the traffic flow over key arterial links (the arterial roads connecting to the critical node) should correspond to the error caused by the minor links which connected directly to the arterial links. Therefore the traffic flows between the arterial links and the minor links should be highly correlated which can help us to correct the sub-error caused by the minor links (in/out of the sub-area) without RTMS. At last, we add up each sub-area's modification to obtain the modification of the target area.

Without the exact priori knowledge about the traffic flow over minor roads, we make the following assumptions: (1) the traffic flow of the arterial road i ($flow_i$) is proportional to the traffic flow of the minor roads (e_i) connecting to it. But if an arterial road has no minor roads, we think this link doesn't cause any error. Because the arterial roads has strong correlation with the minor roads connected to it directly and the traffic flow obey the principle of continuity [14], the traffic flow of the minor roads should be closely correlated to the traffic flow of the arterial road.

According to the supposition, we can assign the sub-error to the each arterial links in the sub-area as Equation 6.

$$e_i = \frac{flow_i}{\sum_i flow} e \tag{6}$$

where e is the total error of the critical node. e_i is the total error caused by the neighbor minor roads of the arterial road i (we think the error is caused by the arterial road i for convenience).

After that, we normalize the traffic flow of the arterial road i ($flow_i$) and use it to replace the traffic flow curve of the minor link to obtain the modification in each time slots of a day as Equation 7.

$$e_i(t) = \frac{flow_i(t)}{\sum_t flow_i(t)} e_i \tag{7}$$

where $e_i(t)$ is the error of the arterial road i at time slot t.

From the above equations, we can first calculate the correction error for all arterial links of each sub-area. Then by adding the correction errors up, we are able to obtain the total correction error for the whole targeted area, which can further be utilized to correct the estimated AVN obtained by the basic method. The algorithm of the advanced method is shown in Algorithm 2. The time slot is 15 min in algorithms. The time complexity of the basic design is $O(mt)$. The time complexity of the advanced design is $O(mnt)$, where n is the number of critical nodes, m is the number of links connected to the critical node, and t is the number of the time samples.

Algorithm 2: the Advanced Method of AVN Estimation

Input: $F \leftarrow$RTMS dataset, $R \leftarrow$road network dataset, $A \leftarrow$selected area
Output: area vehicle number V

initialize $V \leftarrow$ zero vector
$V \leftarrow$ the basic method
$E \leftarrow$ the error of A
for *(node(j) $\in$ A)* **do**
  for *(link(i) $\in$ node)* **do**
    $e_j \leftarrow$ balance of flow
    $e_i \leftarrow [\frac{flow_i}{\sum_i flow_i}] e_j$
    $e_i(t) = \frac{flow_i(t)}{\sum_t flow_i(t)} e_i$
  end
  $e_j \leftarrow [\sum e_i(1), \sum e_i(2), ..., \sum e_i(t)]$
end
$V \leftarrow V + \sum e_j$
return V

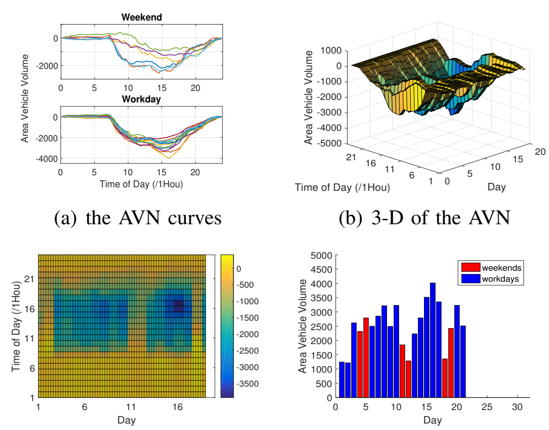
IV. Evaluations

We evaluate the effectiveness of our design through three typical cases. Specifically, we select three representative areas in Hangzhou city, China. Area 1 is mainly composed by residential buildings and some schools. Area 2 is a mixed area including residential buildings, business buildings for work, and Area 3 is a bit complicated including residential buildings, restaurants and entertainment places. The difference of major functionalities among the areas makes the AVN of each area appears differently.

Area 1 contains 8 arterial roads (6 of them have RTMS data) and 3 minor roads, totally 725398 records. Area 2 contains 10 arterial roads (7 of them have RTMS data) and 9 minor roads, totally 1591786 records. Area 3 contains 12 arterial roads (9 of them have RTMS data) and 8 minor roads, totally 1847081 records.

Note that at 00:00 am every morning, we reset the initial value of AVN as 0, then estimate the AVN at each time slot (every 1 hour) by the advanced design approach. The positive AVN means that the number of vehicles in the area is larger than the initial value, and more vehicles come in and park in the area, and vice versa. By adding up the absolute values of the maximum and the minimum AVN estimation, we can directly obtain the Maximum AVN Volume of each day.

A. Analysis of Area 1



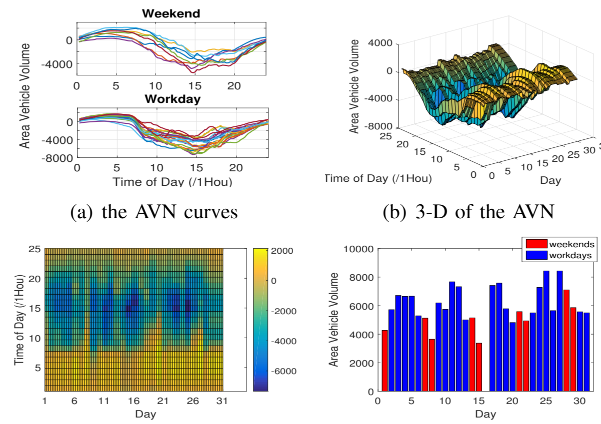
(a) the AVN curves (b) 3-D of the AVN

(c) heat map of the AVN (d) Maximum of the AVN

Figure 6: AVN of the Area 1

Figure 6(a) depicts the AVN estimation of Area 1 for each day during a month. We can find that the curves are similar to each other and the curve of a day present a 'U' shape obviously. For instance, AVN (workday) decreases rapidly start at 7:00 am and slows down at 10:00 am, then keeps stable. At 4:00 pm, the AVN begins to increase and stays at 0 at the end of a day. Furthermore, the pattern of curves reflects the mobility of people in the area (people go to work in the morning and back home in the evening). At weekend, many people travel by public transport instead of cars so that the curves of the weekends are different from those of the workdays, as the steep curves and the flat curves in the Figure 6(a) (the steep curves indicate the workdays, and the flat curves indicate the weekends). According to the AVN curves, it is clear that the change of AVN for Area 1 coincides with its type, residential area. Figure 6(b) and Figure 6(c) show the three dimension description and the heat map of the AVN respectively.

B. Analysis of Area 2



(a) the AVN curves (b) 3-D of the AVN

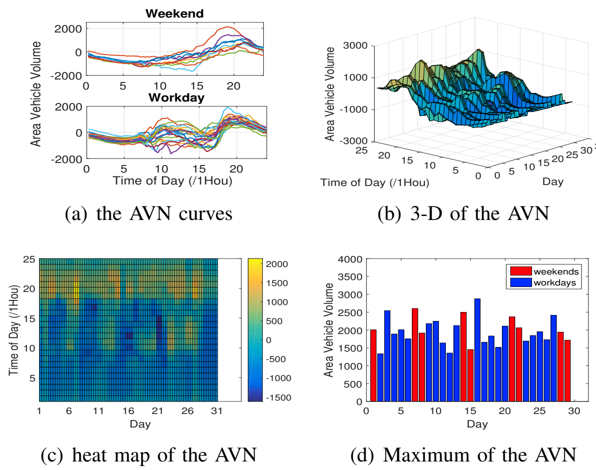
(c) heat map of the AVN (d) Maximum of the AVN

Figure 7: AVN of the Area 2

Figure 7(a) shows the AVN estimation of Area 2 for each day during a month. This area is similar to the Area 1 according to the AVN curves. The curves of Area 2 (workday) between 9:00 am and 10:00 am have a small peak which means traffic flow coming into the area during the time (e.g., people come to work). And we also can find the the segment between 4:00 pm and 5:00 pm is stable which means the traffic flow in/out of the area is equal (e.g., number of people come into the area equals to the people left the area). After that, we can infer some people work and some people live in this area. In fact, Area 2 is a mixed area with residential and business function. Figure 7(b) and Figure 7(c) show the three dimension description and the heat map of the AVN respectively.

C. Analysis of Area 3

Figure 8(a) shows the AVN change of a day in Area 3, which is different from that in the former areas. Combining Figure 8(b) and Figure 8(c), we can find there are two peaks and two valleys in most days (workdays). The first valley appears at the early morning rush-hour because people leave the area to work. After that, the first peak at the 11:00 am where there are many people coming into Area 3. And the second valley demonstrates the evening rush-hour. But the second peak appears between 7:00 pm and 8:00 pm, which means after the evening rush-hour, the AVN of the area increase first and then decrease until to the start of the next day. By contrasting to the area, there are some



(a) the AVN curves (b) 3-D of the AVN

(c) heat map of the AVN (d) Maximum of the AVN

Figure 8: AVN of the Area 3

Table I: The information of areas

area	function	household number	parking space
1	residential, education	8950	1800
2	residential, business	6400	700
3	residential, restaurant	2760	840

restaurants and hotels in it, so people come to have dinner or entertainment during the time which causes the second peak. Moreover, we also can find this area varies every day, the cyclicity is not as strong as the former areas because the components of the area are more complex than the other two areas and the vehicles enter and leave the area more often.

D. Parking Demand Analysis

We collect the planning data of housing estate (like household number, parking space number) in each area from the website ([16–18] and etc.) as the ground truth, only statistical data, which do not contain any personal private or sensitive information. From the detailed data, we can calculate the vehicle ownership of each area (number of household times vehicle ownership per capita) as the ground truth of the parking demand. After that, we can analyze the gap between people regular demand and the parking space.

According to the estimation as Figure 6(d), Figure 7(d) and Figure 8(d) show, we can obtain the average maximum of AVN and use it as the parking demand of each area. Table. II shows the average parking demand of three areas during a month and the ground truth we infer from the statistic data. We can find that the maximum AVN of weekend is smaller than that of workdays in Area 1 and Area 2, because

Table II: Parking Demand Estimation Result

area	workdays demand	weekends demand	ground truth
1	2964	2340	2455
2	6425	4742	4465
3	1935	2062	1926

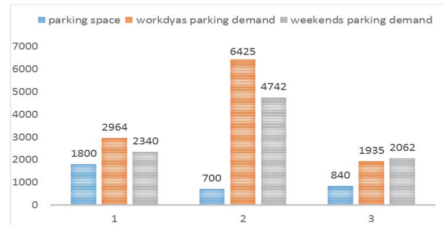


Figure 9: Area Parking Space and Demand

many people don't drive out on weekend. For Area 3, the maximum AVN of weekends and workdays is nearly equal although the curve of AVN is quite different. One possible reason is that many people come to the restaurants and hotels in Area 3 to enjoy their spare time.

Comparing to the vehicle ownership, we can find that the vehicle ownership is smaller than the workdays parking demand but similar to the weekends parking demand in Area 2 and Area 3. For Area 1, the vehicle ownership is approximate to the demand all in workdays and weekends. We can use the weekends demand estimation to obtain the regular parking demand of the people living in the area. The relative error between our AVN and the ownership data is 0.06, 0.07 and 0.05 separately in these areas.

Moreover, we can easily obtain the relation between the parking demand and parking space in the Figure 9. From the bar chart, we can clearly know the supply-demand gap of these areas. The parking demands of three case areas are all larger than the parking space. Especially, in the Area 2, the demand is almost 7 or 9 times bigger than the parking space, which causes to on-street parking (more than 1/3 roads of the Urban core blocks has been occupied [7]). Our estimated AVN can guide authorities to make the traffic infrastructure planning according to the different situation in different areas to relieve the parking problem.

V. Related Work

Conventional parking demand estimation methods need a lot of data which usually is gained by the survey [19] at the cost of a lot of resources and money. In order to obtain the parking demand in an easier way, we estimate the AVN based on the RTMS data which can be collected online and use the AVN to obtain the dynamic parking demand. Our method mainly consists of missing flow data completion, AVN estimation, and parking demand inferring.

The urban black holes detection is related to the AVN estimation in some degree because these questions (black hole detection and AVN estimation) should consider the accumulation of the data, such as flow data, human mobility data. In [20], the authors use the human mobility data to detect urban black holes. But our work consider the whole cycle of AVN to estimate the unbalance of the traffic flow in and out of area and estimate the AVN with high granularity. Furthermore, the detection of urban black holes focuses on the accumulation of human in a period of time and finds out regions that have objects higher than a threshold, which is quite different from ours.

There are some other related works including the urban traffic commuting analysis based on mobile phone data [21] and inferring gas consumption and pollution of vehicles [22]. There approaches along with the corresponding data may further be utilized to improve our AVN estimation accuracy.

VI. Conclusion

This paper focuses on the AVN estimation in urban cities. By utilizing the large amount of RTMS data, we propose the basic method for inferring the AVN based on the inflow and outflow traffic. Then we improve the estimation accuracy by exploiting the roads balance and the traffic network flow. Extensive evaluations are conducted for different roads and areas based on more than 3.5 million RTMS data from the city of Hangzhou during one month. It is shown that our approach is able to achieve around 10% mean relative error by comparing the estimated parking demand with the public vehicle ownership data.

VII. Acknowledgement

The paper was partially supported by the 973 Program under Grant 2015CB352500, and National Program for Special Support of Top Notch Young Professionals.

References

[1] T. transportation institute, “Traffic gridlock sets new records for traveler misery,” https://tti.tamu.edu/2015/08/26/traffic-gridlock-sets-new-records-for-traveler-misery/.

[2] R. Zhang, Y. Shu, Z. Yang, P. Cheng, and J. Chen, “Hybrid traffic speed modeling and prediction using real-world data,” in *BigData Congress, 2015 IEEE International Congress on*. IEEE, 2015, pp. 230–237.

[3] B. Pan, U. Demiryurek, and C. Shahabi, “Utilizing real-world transportation data for accurate traffic prediction,” in *Data Mining, 2012 IEEE 12th International Conference on*. IEEE, 2012, pp. 595–604.

[4] Z. Yang, J. Hu, Y. Shu, P. Cheng, J. Chen, and T. Moscibroba, “Mobility Modeling and Prediction in Bike-Sharing Systems,” in *ACM MobiSys*, 2016.

[5] X. Tan, Y. Shu, X. Lu, P. Cheng, and J. Chen, “Characterizing and modeling package dynamics in express shipping service network,” in *Big Data (BigData Congress), 2014 IEEE International Congress on*. IEEE, 2014, pp. 144–151.

[6] H. gov, “Assessment report of hangzhou city construction and measures for administration of motor vehicle parks,” http://www.hangzhoufz.gov.cn/Html/201412/23/6111.html.

[7] H. Daily, http://z.hangzhou.com.cn/2014/2014hzlh/content/2014-02/09/content_5121142.htm.

[8] M. Treiber, *Traffic Flow Dynamic*. Springer, 2013.

[9] D. B. Ottosson, C. Chen, T. Wang, and H. Lin, “The sensitivity of on-street parking demand in response to price changes: A case study in seattle, wa,” *Transport Policy*, vol. 25, pp. 222–232, 2013.

[10] N. A. Weber and S. Dakota, “Verification of radar vehicle detection equipment,” South Dakota Department of Transportation, Office of Research, Tech. Rep., 1999.

[11] D. Manor, “Multiple zone radar detection by rtms,” *Traffic Technology International’96. Annual Review Issue*, 1996.

[12] C.-C. Hsu and C.-Y. Chen, “Applications of improved grey prediction model for power demand forecasting,” *Energy Conversion and Management*, vol. 44, no. 14, pp. 2241–2249, 2003.

[13] Y. Zhang, Y. Zhang, and A. Haghani, “A hybrid short-term traffic flow forecasting method based on spectral analysis and statistical volatility model,” *Transportation Research Part C: Emerging Technologies*, vol. 43, pp. 65–78, 2014.

[14] S.-G. Wang, L. Bai, X. Bao, Y. Liu, and G. Cao, “Novel estimation method of community aadt and vmt via circuit network models and simulation,” in *Decision and Control and European Control Conference, 2011 50th IEEE Conference on*. IEEE, 2011, pp. 5305–5310.

[15] B. L. Smith, B. M. Williams, and R. K. Oswald, “Comparison of parametric and nonparametric models for traffic flow forecasting,” *Transportation Research Part C: Emerging Technologies*, vol. 10, no. 4, pp. 303–321, 2002.

[16] http://baike.baidu.com/view/2092957.htm.

[17] http://baike.baidu.com/view/2093100.htm.

[18] http://baike.baidu.com/view/2092965.htm.

[19] D. A. Hensher and J. King, “Parking demand and responsiveness to supply, pricing and location in the sydney central business district,” *Transportation Research Part A: Policy and Practice*, vol. 35, no. 3, pp. 177–196, 2001.

[20] L. Hong, Y. Zheng, D. Yung, J. Shang, and L. Zou, “Detecting urban black holes based on human mobility data.” ACM Association for Computing Machinery, November 2015.

[21] H. Dong, X. Ding, M. Wu, Y. Shi, L. Jia, Y. Qin, and L. Chu, “Urban traffic commuting analysis based on mobile phone data,” in *2014 IEEE 17th International Conference on Intelligent Transportation Systems*. IEEE, 2014, pp. 611–616.

[22] 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 ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2014, pp. 1027–1036.