A Hand Over and Call Arrival Cellular Signals-based Traffic Density Estimation Method

Yizhuo Zhang
zhangyzy86@hotmail.com
College of Computer and Data Science, Fuzhou University, Fuzhou, Fujian, China

Hao Fang
hfang7@163.com
College of Computer and Data Science, Fuzhou University, Fuzhou, Fujian, China

Chi-Hua Chen
chihua0826@gmail.com
College of Computer and Data Science, Fuzhou University, Fuzhou, Fujian, China

ABSTRACT
The growing number of vehicles has put a lot of pressure on the transportation system. Intelligent Transportation System (ITS) faces a great challenge of traffic congestion. Traffic density displays the congestion of current traffic which reflects explicitly about traffic status. With the development of communication technology, people use mobile stations (MSs) at any time and cellular signals are everywhere. Different from traditional traffic information estimation methods based global positioning system (GPS) and vehicle detector (VD), this paper resorts to Cellular Floating Vehicle Data (CFVD) to estimate the traffic density. In this paper, Hand over (HO) and call arrival (CA) cellular signals are essentials to estimate traffic flow and traffic speed. In addition, mixture probability density distribution generator is adopted to assist estimating the probabilities HO and CA events. Through accurate traffic flow and traffic speed estimations, precise traffic density is achieved. In the simulation experiments, the proposed method achieves estimation MAPEs 11.92%, 13.97% and 16.47% for traffic flow, traffic speed and traffic density, respectively.

CCS CONCEPTS
• Computing methodologies → Neural networks; Machine learning.

KEYWORDS
hand over, call arrival, cellular signals, traffic density, traffic speed, traffic flow

ACM Reference Format:

1 INTRODUCTION
With the increasing number of private cars, traffic management, traffic congestion and other traffic problems are emerging. The prerequisite of solving these traffic issues is collecting the traffic information such as traffic flow, traffic speed and traffic density. Traditional approaches to obtain traffic information are based on Global Positioning System (GPS) and Vehicle Detectors (VDs). Vehicles equipped GPS-enabled devices will report their real time positions, and traffic information (e.g., traffic speed and traffic flow) [12, 16] is precisely estimated based on these GPS data. VD-based traffic information acquiring method resorts to the VD devices installed underground [1] to detect the passed vehicles. VDs are able to capture the detailed data of the passing vehicles, which includes vehicle types [8, 10], passing time and vehicle speed. However, above two methods require external devices installed inside vehicles or road segments and cause privacy issues. In addition, GPS-enabled devices and vehicle detectors require high maintenance fees. Recently, the advanced communication technology promotes the widespread uses of mobile stations (MSs). MSs rely on Base Stations (BSs) to communicate with others. Therefore, a lot of cellular signals are generated between MSs and BSs, which are regarded as Cellular Floating Vehicle Data (CFVD). When vehicles are traveling, communications of MSs in vehicles happen and cellular signals (i.e., Handover (HO), Call Arrival (CA) and et al.) are collected by mobile operator through BSs. Based on the statistics of CFVD, traffic information can be estimated in a cheaper and more convenient way [7, 11]. Specifically, HO and CA events occur when switching the base station during the call and call arriving respectively, and the BS will record the time and location of the HO and CA signal. In other words, temporal and spatial characteristics of the vehicle with MSs can be recorded in the statistics of CFVD. In this paper, a novel approach based on HO and CA is proposed to estimate traffic density. The proposed method contains three phrases: HO-based traffic flow estimation, CA-based traffic speed estimation and traffic density estimation. Specially, each vehicle is assumed to carry only one MS [5]. In a Location Area (LA), there are several cells. HO event is triggered when a MS passes through a cell while maintaining the communication. When there is a call arrival, a CA signal is collected by mobile operator. The proposed method adopts the statistics of HOs and CAs to estimate traffic flow and traffic speed in the cell, respectively. The time between two consecutive HO events is regarded as call holding time. Call inter-arrival time represents the time interval between two CA events. The distributions of call holding time and call inter-arrival time determine the probabilities of HO events and CA events in a cell, respectively. This paper resorts to the probability density distribution generator [6] to precisely generate the Cumulative Distribution Functions (CDFs) of both call holding time and call inter-arrival time. The estimation of traffic density is on the basis of traffic flow and traffic
speed. The main contributions of this paper are summarized as follows:

- This paper proposes novel CFVD-based statistical approaches to accurately estimate traffic flow, traffic speed and traffic density.
- The experiments adopt VISSIM software to generate highway simulation environment in Taicang, China and CFVD is also generated randomly. The experimental results have achieved high accuracy on traffic flow, traffic speed and traffic density estimations.

2 RELATED WORK

In recent researches, GPS-based traffic information estimation approaches [2, 17] can achieve high accuracy. However, the collected traffic information is reliable when there are certain amount of vehicles installed GPS-enable devices [4]. Moreover, GPS requires detailed information of the running vehicles which causes privacy issues. VD-based traffic information collection method will record vehicle types, time and speed, which to some extent avoids privacy issues. VDs are always installed underground to detect the passing vehicles [10, 15]. With the aging VDs, the accuracy of detection cannot be proved and replacements of VDs require substantial investment. Currently, MSs are everywhere and produce cellular signals such as HO, CA, Normal Location Update (NLU) and Periodic Location Update (PLU). Researchers analyzed the relationship between the variation of CFVD and people communicating to infer the traffic status [3]. CFVD-based traffic information estimation spends less resources and achieves low estimation errors. Demissie et al. used HO signals to construct a multinomial logit (MNL) model and artificial neural network (ANN) to estimate road traffic status [9]. HO signals were used to obtain traffic flow level in real-time [13]. Besides, the authors adopted a support vector machine to extract the relations between HO events data and traffic flow levels. Chen et al. combined NLU and CA signals to effectively estimate traffic speed [5]. They also applied machine learning techniques to improve the estimation accuracy. PLU signals are sensitive to the position change in certain period. Thus, Lin et al. [14] extracted features from communication behaviors and utilized PLU signals to estimate traffic density. CFVD-based estimation methods requires less resources and performs well on estimating traffic information. In addition, these methods are privacy friendly since estimating traffic information is not direct from vehicle data.

3 A HAND OVER-BASED TRAFFIC FLOW ESTIMATION METHOD

In cellular network, the signal range of BS is finite. If the MS moves out of signal range of the BS, the MS will connect to the new BS to maintain the call. As shown in Figure 1, there is a call arrival at $t_0$ and the call holds for $t_h$. Since MS moves from $Cell_{i-1}$ to $Cell_i$ in $t_h$, a HO event happens at $t_1$ which denotes the change of BS. This study utilizes the number of HO event happened in the cell to estimate traffic flow under assumption that each vehicle carries one MS. Figure 2 illustrates two scenarios of HO event at the cell boundary. In the first scenario, the call arrives at $t_0$ and shut down before the MS leaves $Cell_{i-1}$. In other words, the time difference $x_h = t_1 - t_0$ is larger than the call holding time $t_h$. In the second scenario, HO event occurs at $t_1$ when $x_h < t_h$. The probability and number of HO events at $t_1$ are used to estimate traffic flow $q_i$ in $Cell_i$. To obtain the probability of HO event (i.e. $Pr(x_h < t_h)$) at $t_1$, the probability density distribution of $t_h$ is necessary. This paper adopts mixture probability density distribution to approximate the practical probability density distribution of $t_h$. In addition, the mixture probability density distribution is consist of exponential distribution (ED), normal distribution (ND) and lognormal distribution (LD). The mixture CDF of $t_h$ is represented as $F_{th}(x_h)$, which is shown in Equation 1. $\frac{1}{\lambda_1}$ is the expectation of ED, $\mu_1$ and $\sigma_1$ denote mean and variance of
ND, respectively. Similarly, $\mu_2$ and $\sigma_2$ represent mean and variance for LD. $w_{11}, w_{12}$ and $w_{13}$ are weights for ED, ND and LD. Moreover, CDF is generated by Neuron Network-based probability density distribution generator from [6]. Since mixing more probability density distributions will increase the computational complexity, this paper only adopts three common distributions from the original paper [6]. The structure of NN is displayed in Figure 3. The parameters to train in NN are the same in Equation 1. The activation functions of neurons are CDF functions of ED, ND and LD, respectively. The final output is the CDF value of $F_{\text{Cell}_i}(x)$. The number of HO events at $\text{Cell}_i$ is denoted as $h_i$. Although not all vehicles crossing the cell boundary will trigger HO event, there are more HO events at the cell boundary with the increase of traffic flow. The relationships of traffic flow and the probability and the number of HO events are illustrated in Equation 2. Based on Equation 2, traffic flow $q_i$ is obtained in Equation 3.

$$F_{\text{Cell}_i}(x) = w_{11} \times e^{\lambda_1 x_h} + w_{12} \times \frac{1}{2} \left(1 - \text{erf} \left( \frac{x_h - \mu_1}{\sigma_1 \sqrt{2}} \right) \right) + w_{13} \times \left(1 - \text{erf} \left( \frac{\ln(x_h) - \mu_2}{\sigma_2 \sqrt{2}} \right) \right).$$  

(1)

$$h_i = q_i \times \int_{x_h=0}^{\infty} Pr(x_h < t_h) dx_h = q_i \times \int_{x_h=0}^{\infty} F_{\text{Cell}_i}(\infty) - F_{\text{Cell}_i}(x_h) dx_h = \frac{q_i}{w_{11} + w_{12} + w_{13}} \times \int_{x_h=0}^{\infty} (w_{11} \times e^{\lambda_1 x_h} + w_{12} \times \frac{1}{2} \left(1 - \text{erf} \left( \frac{x_h - \mu_1}{\sigma_1 \sqrt{2}} \right) \right) + w_{13} \times \left(1 - \text{erf} \left( \frac{\ln(x_h) - \mu_2}{\sigma_2 \sqrt{2}} \right) \right)) dx_h = \frac{q_i}{w_{11} + w_{12} + w_{13}} \times \left( \frac{w_{11}}{\lambda_1} + \frac{1}{2} w_{12} \times (\mu_1 - \mu_1 \text{erf} \left( \frac{\mu_1}{\sigma_1 \sqrt{2}} \right)\right) + \frac{\sqrt{2}}{\pi} \sigma_1 e^{\left( -\frac{\mu_1^2}{2\sigma_1^2} \right)}\right) + \frac{\sqrt{2}}{\pi} \sigma_1 e^{\left( -\frac{\mu_1^2}{2\sigma_1^2} \right)} + w_{13} \times e^{\frac{\sigma_2^2}{2} + \mu_2}.$$  

(2)

$$q_i = h_i \times (w_{11} + w_{12} + w_{13}) \times \frac{w_{11}}{\lambda_1} + \frac{1}{2} w_{12} \times (\mu_1 - \mu_1 \text{erf} \left( \frac{\mu_1}{\sigma_1 \sqrt{2}} \right)\right) + \frac{\sqrt{2}}{\pi} \sigma_1 e^{\left( -\frac{\mu_1^2}{2\sigma_1^2} \right)} + \frac{\sqrt{2}}{\pi} \sigma_1 e^{\left( -\frac{\mu_1^2}{2\sigma_1^2} \right)} + w_{13} \times e^{\frac{\sigma_2^2}{2} + \mu_2}.$$

(3)

4 A CALL ARRIVAL-BASED TRAFFIC SPEED ESTIMATION METHOD

When there is a call arrival to a MS, the connected BS will record the CA event. Therefore, the number of CA events in a cell is available in the statistics of CFVD. When a vehicle passing the cell with high speed, the probability of CA events is very small, and vice versa. Therefore, this paper resorts to the probability and the number of CA events to estimate traffic speed in the cell. In Figure 4, $t_c$ is call inter-arrival time and $x_c = t_1 - t_0$ represents the time difference from the first call arriving to entering the next cell. The length of the $\text{Cell}_i$ is $l_i$ and the estimated traffic speed is denoted as $v_i$. Therefore, the traveling time in $\text{Cell}_i$ is calculated as $\frac{l_i}{v_i}$. When $x_c < t_c < x_c + \frac{l_i}{v_i}$, CA events happen inside $\text{Cell}_i$. The probability of CA events triggered in $\text{Cell}_i$ is $Pr(x_c < t_c < x_c + \frac{l_i}{v_i})$. The mixture probability density distribution of $t_c$ consists of ED, ND and LD, $F_{t_c}$ denotes the CDF of $t_c$. $F_{t_c}$ is trained by NN-based CDF generator with the same structure mentioned in Section 3. Furthermore, NN is trained by the practical call-inter-arrival time data. The calculation of $F_{t_c}$ is shown in Equation 4. $\lambda_2, \mu_3$ and $\sigma_3$ are parameters for ED, ND and LD, respectively. $w_{21}, w_{22}$ and $w_{23}$ are the corresponding weights for ED, ND and LD. Based on the traffic flow $q_i$ and the probability of CA events in $\text{Cell}_i$, the amount of CA events $c_i$ in $\text{Cell}_i$ is calculated in Equation 5. Since it is hard to derive $v_i$ directly from Equation 5, this paper adopts Newton down-hill method to obtain an approximate solution which satisfies the error $\epsilon < 0.00001$. The specific process is demonstrated in Equation 6-9. In Equation 7, $w_{3}$ is the iterative parameter of Newton down-hill method. After several rounds of iteration, the estimated traffic speed $v_i$ is obtained.

$$F_{t_c}(x) = w_{21} \times e^{\lambda_2 x} + w_{22} \times \frac{1}{2} \left(1 - \text{erf} \left( \frac{x - \mu_3}{\sigma_3 \sqrt{2}} \right) \right) + w_{23} \times \left(1 - \text{erf} \left( \frac{\ln(x) - \mu_4}{\sigma_4 \sqrt{2}} \right) \right).$$  

(4)

$$c_i = q_i \times \int_{x_c=0}^{\infty} Pr(x_c < t_c < x_c + \frac{l_i}{v_i}) dx = q_i \times \int_{x_c=0}^{\infty} F_{t_c}(x_c + \frac{l_i}{v_i}) - F_{t_c}(x_c) dx = \frac{q_i}{2(w_{21} + w_{22} + w_{23})} \times \left( \frac{2w_{21}}{\lambda_2} e^{\lambda_2 \frac{l_i}{v_i}} - 1 \right) + w_{22} \sqrt{\frac{2}{\pi}} \sigma_3 e^{-\left( \frac{\mu_3}{\sigma_3 \sqrt{2}} \right)^2} \left( e^{-\left( \frac{\mu_3}{\sigma_3 \sqrt{2}} \right)^2} - 1 \right) + w_{22} \left(\frac{\mu_3}{v_i} + \frac{\mu_3}{v_i} \text{erf} \left( \frac{\mu_3}{\sigma_3 \sqrt{2}} \right) + \mu_3 \text{erf} \left( -\frac{\mu_3}{\sigma_3 \sqrt{2}} \right) \right) + w_{23} \left(\frac{\mu_3}{v_i} + \frac{\sigma_4^2}{2} + \mu_4 - \ln \left( \frac{\mu_4}{v_i} \right) \right) - 1) + \frac{\mu_4}{v_i} \text{erf} \left( \frac{\mu_4}{v_i} \right) \right).$$  

(5)

$$f(u_k) = c_i - q_i \times \int_{x_c=0}^{\infty} Pr(x_c < t_c < x_c + \frac{l_i}{v_i}) dx,$$  

(6)

$$u_{k+1} = u_k - w_k \frac{f(u_k)}{f'(u_k)},$$  

(7)

$$\epsilon = |f(u_{k+1}) - 0|,$$  

(8)

$$v_i = u_{k+1}.$$  

(9)
vehicles. The highway is shown in Figure 5. For each lane, there are 11 VDs installed with 1 kilometers interval underground to obtain the practical traffic information. Besides, the length of the cell is also set 1 kilometers. The length of traffic information collection time is 24 hours. There is only one MS in each vehicle for collecting CFVD. Call holding time and call inter-arrival time are randomly sampled from CDFs. In the highway, the HO events are free to change their lanes and follow other vehicles. In the simulation traffic environment generated by VISSIM software, the HO events are recorded when \( h_k < t_k, x_k \) is randomly sampled from \( P_{HO} \). The BS recording call arrival signals when there are calls coming. When last call arrives, \( t_c \) will be sampled from \( P_{CA} \) and the time for the next call is set.

In the simulation traffic environment generated by VISSIM software, the HO events are recorded when \( h_k < t_k, x_k \) is randomly sampled from \( P_{HO} \). The BS recording call arrival signals when there are calls coming. When last call arrives, \( t_c \) will be sampled from \( P_{CA} \) and the time for the next call is set.

\[
P_{HO}(x) = w_{31} \times e^{\lambda_{HO}x} + w_{32} \times \left(1 - erf\left(\frac{x - \mu_{HO}}{\sigma_{HO} \sqrt{2}}\right)\right) + w_{33} \times \left(1 - erf\left(\frac{\ln(x) - \mu_{HO}}{\sigma_{HO} \sqrt{2}}\right)\right).
\]

\[
P_{CA}(x) = w_{41} \times e^{\lambda_{CA} x} + w_{42} \times \left(1 - erf\left(\frac{x - \mu_{CA}}{\sigma_{CA} \sqrt{2}}\right)\right) + w_{43} \times \left(1 - erf\left(\frac{\ln(x) - \mu_{CA}}{\sigma_{CA} \sqrt{2}}\right)\right).
\]

\[
MAPE = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{\hat{y} - y}{y} \right| \times 100\%
\]

### 6.2 Results Analysis

Based on the collected CFVD, two NN-based probability density distribution generators are trained for generating precise CDFs of call holding time and call inter-arrival time. In Figure 5, there are 10 cells in total. Experimental results of Cell2 are displayed as examples to show the performance of traffic information estimation in 24 hours. As shown in Figure 6, the practical traffic flow rises rapidly at 7 a.m and declines at 5 p.m. Furthermore, in Figure 7, traffic speed is relative slow from 7 a.m to 10 p.m while the traffic flow is high at the same time period. Both HO-based traffic flow estimation and CA-based traffic speed estimation have learned the variation trend of the practical traffic. According to the Equation 11, traffic density is obtained and Figure 8 shows that the estimated traffic density is approximate to the practical traffic density in Cell2. In the experiments, MAPEs of proposed traffic information estimation methods in Cell2 are displayed in Table. Besides, the average MAPEs of all cells are shown in Table. HO-based traffic flow estimation method achieves MAPE 10.09% in Cell2 and average MAPE 11.92% of all cells. The MAPE of CA-based estimated traffic speed in all cells is 12.47% while 13.97% in Cell2. For the MAPE of traffic density estimation, the proposed method achieves 17.75% in all cells and 16.47% in Cell2.

### 7 CONCLUSIONS

This paper proposes a novel traffic density estimation method based on CFVD. Compared with the traditional GPS-based and VD-based traffic information estimation method, the proposed method costs less and achieves accurate estimation. Specifically, traffic density is composed of traffic flow and traffic speed. Traffic flow estimation utilizes quantiative relations among the number and probability of HO events in the cell and traffic flow to derive the specific formulation. Similarly, CA signals recorded by BS are adopted to estimate the traffic speed. The derivation of traffic speed is based on the number and probability of CA events as well. The probabilities...
A Hand Over and Call Arrival Cellular Signals-based Traffic Density Estimation Method

CFVD is randomly sampled from the mixture probability density distribution generator to precisely fit the practical distributions. The proposed methods have achieved MAPE 11.92% and 13.97% of traffic flow and traffic speed estimations, respectively. In traffic density estimation, MAPE is 16.47%. The experimental results have shown the outstanding performance of CFVD-based traffic density estimation method. In the future, the proposed traffic information estimation method can be applied in more complicated traffic environment.

8 ACKNOWLEDGMENT
This work is partly supported by the Natural Science Foundation of China under Grant No. 61906043.

REFERENCES


