Journal of Network Intelligence
Taiwan Ubiquitous Information

Volume **7**, Number **3**, August **2022**

Deep Learning-Based Probability Model for Traffic Information Estimation

Zhaoshan Sun

College of Computer Science and Mathematics
Fujian University of Technology
Fuzhou, 350118, China
sunzhaoshan1003@qq.com

Jeng-Shyang Pan*

Collage of Computer Science and Engineering
Shandong University of Science and Technology, China
Chaoyang University of Technology, Taiwan
Fujian University of Technology, China
jengshyangpan@gmail.com

Tien-Szu Pan

Department of Electronic Engineering
National Kaohsiung University of Science and Technology
415 Chien-Kung Road, Kaohsiung, 807, Taiwan
tpan@nkust.edu.tw

Chi-Hua Chen

College of Computer and Data Science/College of Software
Fuzhou University
Fuzhou, 350116, China
chihua0826@gmail.com

*Corresponding author: Jeng-Shyang Pan

Received September 4, 2021, revised December 5, 2021, accepted March 9, 2022.

Abstract. *This study proposes a novel traffic information estimation method based on deep learning and cellular floating vehicle data (CFVD). In this paper, a probabilistic analysis model based on deep learning is proposed to consider the relationship between vehicle speed and communication behaviors. Then, a vehicle speed estimation method based on the proposed probabilistic analytical model is proposed to estimate vehicle speed. For traffic flow estimation, normal location update is adopted to estimate traffic flow. The estimated vehicle speed and the estimated traffic flow can be gathered to estimate traffic density. The proposed method is verified by the simulation tool and the experiment results revealed that the accuracies of estimated vehicle speed, estimated traffic flow, and estimated traffic density are 96.36%, 99.80%, and 96.45%, respectively. Thus, this method estimates traffic information accurately and helps improve the performance of the intelligent transportation system (ITS).*
Keywords: Traffic information estimation, Probabilistic analysis model, Cellular floating vehicle data, Deep learning, Intelligent transportation system.

---

1. **Introduction.** Intelligent transportation system (ITS) is getting smarter in improving urban cleanliness and residents' happiness such as relieving traffic congestion, saving energy, and increasing travel safety, etc. [1, 2]. Because of advances in technology and the maturity of hardware, the ITS is building the foundation in applications as time goes on [3, 4]. For example, the expansion of physical memory and the improvement of data transmission efficiency can make researchers resolve previously unsolvable problems [5,6]. Acquiring actual traffic information such as vehicle speed and traffic flow is extraordinarily significant in ITS [7–9]. For traffic service, real-time traffic information collection helps the government make decisions more effectively and traffic information estimation can assist people accurately when they travel on vacation [10, 11]. Thus, several studies discussed the enhancement methods for the effectiveness of intelligent transportation system.

Recent research papers show three types of methods that collect real-time traffic information: vehicle detectors (VDs), global position system (GPS)-based probe cars, and cellular floating vehicle data (CFVD).

Traditionally, stationary VDs are utilized to assemble actual traffic information. They are installed on road network and acquire real-time data of vehicle speed and traffic flow [12, 13]. Often, they are sensors based on active infrared/laser, magnetic, radar, or video technologies [14–16]. However, accurate real-time traffic information gathering based on VDs is paid for a large amount of money. The reasons involve the short service life of the equipment, high maintenance costs, and high probability to be affected by climate factors such as temperature fluctuation, humidity, or natural disasters such as typhoons and earthquakes [17, 18].

GPS receivers installed on vehicles can have wireless communication capability and collect real-time traffic information. Test vehicles' locations, speeds and time information can be reported and total traffic information can be gathered periodically by GPS receivers [19–21]. However, a 2%-3% penetration of GPS-based probe cars would be sufficient to collect enough accurate measurements of space mean speed [22, 23]. Otherwise, generating enough actual traffic data is too tough. Furthermore, the expensive facility cost is a big problem for most researchers [24, 25].

Contrasted with the above two types of mediums, CFVD has the qualities of low data fee, large data volume, and the ability to effectively avoid privacy issues. Several cellular network signals such as call arrivals (CAs) and handovers (HOs), are utilized to estimate traffic information while the mobile station (MS) moves across the road [12]. For instance, large patented technologies of CFVD were created by a famous company and put them into use for estimating and predicting actual traffic information [26,27]. Thus, CFVD can easily obtain a large amount of data without paying a lot of fees for devices [28, 29]. Also, CFVD collects the anonymous locations and time information of CFVD-based cars in road segments covered by the specific cell, which does not pose a threat to users' privacy [12].

The combination of call arrival and vehicle speed were considered and study for traffic information estimation [30,31]. However, the exponential probability density function may not correspond to the actual probability density distribution of practical data. Thus, a probabilistic analysis model is proposed and the vehicle speed is estimated by the amount of CAs. The traffic flow is estimated by the amount of normal location updates. The combination of estimated vehicle speed and traffic flow can generate the traffic density. The proposed deep learning model allows the traffic information estimation method to adapt to different probability distributions such as exponential distribution, log-normal distribution and normal distribution. From experimental results, the proposed method decreases the error of traffic information estimation.

This paper has the contributions below.

- A deep learning-based probabilistic analysis model is devised to illustrate the probability of communication behavior.
- The CFVD based on the proposed probabilistic analysis model is devised to reckon the traffic information.

The following section is organized below. Section 2 briefly introduces relevant methods for traffic information estimation. Section 3 details how traffic information can be reckoned. Section 4 explains the simulation model and simulated results. Section 5 summaries the study and introduces the plan for the next step.

2. **Literature reviews.** For cellular network, the service area is composed of base stations (BSs). The cell is the radio coverage of a BS which is honeycombed or fanned. The HO event is performed while a communicating mobile station (MS) is moving to another cell and the MS is assigned to another channel. Several cells consist of a location area (LA) which collects and transmits communication behaviors from cells. The normal location update (NLU) event can be considered to reckon vehicle volumes. Two consecutive NLU procedures can be gathered and analyzed for estimating vehicle speed and travel time [32, 33]. According to the statistical data, the mobile phone penetration rate can be regarded as 100%. Therefore, the location of vehicles can be traced.

In this section, several methods based on cellular network signals (e.g., handover (HO), received signal strength indicator (RSSI), and call arrival (CA)) are provided to analyze the development and progress of traffic information estimation.

2.1. **HO-based method for traffic information estimation.** Many researches based on handover procedures have been published for traffic information estimation. Wu et al. estimated vehicle speed by recording the HO events of MS [28]. Lin et al. improved a traffic flow estimation approach by considering the combination of traffic flow and the amount of HO events [34]. Caceres et al. reckoned traffic flow on road network by employing inter-cell handover signaling [35]. Liou et al. used two consecutive HO events to reckon traffic information [36]. The accuracies of reckoning vehicle speed in congested flow circumstance and free flow circumstance were 97.63% and 93.89% separately. Lai et al. combined the HO and CA to estimate vehicle speed and the experiment result showed that this method was available [17]. Chen devised a cell probe (CP)-based approach which proposed a regression model to enhance vehicle speed estimation approach based on HO and NLU [37].

2.2. **CA-based method for traffic information estimation.** It is feasible and effective to apply CA-based methods for traffic information estimation. Lin et al. proposed an approach to reckon traffic flow by calculating the amount of call arrivals [36]. Chen et al. devised a method to reckon traffic information by considering the relationship between the amount of call arrivals and traffic density [17, 31]. The experimental results indicated that the exponential distribution of communication behavior (e.g., call inter-arrival time) had a probabilistic relation with traffic density. However, different communication behaviors did not exactly correspond to a single distribution such as exponential distribution so that the estimation accuracy had potential to be improved. Lai et al. devised a vehicle positioning approach by collecting and analyzing CFVD signals [38].

2.3. **Fingerprint-based method for traffic information estimation.** Many researches have contributed to measuring the radio signal strengths (RSS) by tracking users' MS location for estimating traffic information. Lin et al. devised a vehicle speed estimation method, fingerprint positioning algorithm (FPA), which determined the location of MS by measuring the radio signal strengths [39]. The average error of vehicle speed estimation by using FPA was 3.39%. Ergen et al. proposed a MS localization method by measuring the received signal strength indicator (RSSI) and locating the MS [40]. Cheng et al. devised a estimation model that analyzed two connected fingerprint positioning locations for estimating vehicle speed [41]. Chitraranjan et al. proposed a RSS fingerprints model based on probability [42]. This method had a good performance in estimating travel time.

2.4. **The limitation of current related work.** Many researchers have discussed the methods of traffic information estimation from CFVD. The limitation of current related work is generalized and organized below: CFVD mostly analyzed traffic by tracking the movement of the same phone, and there are privacy issues; The single exponential distribution cannot completely conform to the probability distribution of real communication behavior, resulting in low estimation accuracy. Thus, a deep learning-based probabilistic analysis model is devised to illustrate the probability of communication behavior. And the proposed probabilistic analysis model is devised to reckon the vehicle speed.

3. **Traffic information estimation.** In this section, a novel traffic information estimation approach based on deep learning and CFVD is devised. The amount of CAs is utilized to estimate the probabilistic analytic model by deep learning. The vehicle speed can be estimated based on the proposed model. Then, traffic flow is estimated in line with the amount of NLUs. Furthermore, the traffic density can be acquired by the combination of estimated vehicle speed and traffic flow.

3.1. **Vehicle speed estimation.** The sequence diagram that the MS performs CA is implied as Fig. 1. The MS is moving along the road and performs one call(at time t_0). At time t_1, it enters the road covered by $Cell_i$. Then, it performs another call(at time t_2). A probabilistic analysis model is devised and help to propose a vehicle speed estimation approach.

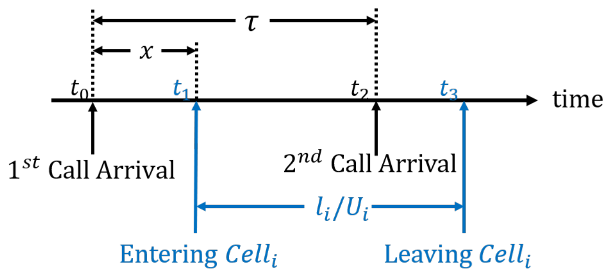

FIGURE 1. The MS performs CA for vehicle speed estimation

3.1.1. *A probabilistic analysis model for the consideration between CA and vehicle speed.* This study devised a probabilistic analysis model for the consideration between CA and vehicle speed. The proposed probabilistic analysis model establishes a deep learning model to learn the cumulative distribution function (CDF) of effective communication data. Then differentiated learned model can be utilized to reckon the probability density function (PDF) of actual communication data.

As Fig. 2 shows, in the training phase, practical data of the MS communication is gathered from CFVD and the cumulative probabilities is collected. Then the proposed model is trained and the trained model can be exploited and reckon the CDF of practical communication data. In the performing phase, the trained model is differentiated and the PDF of actual communication data is estimated.

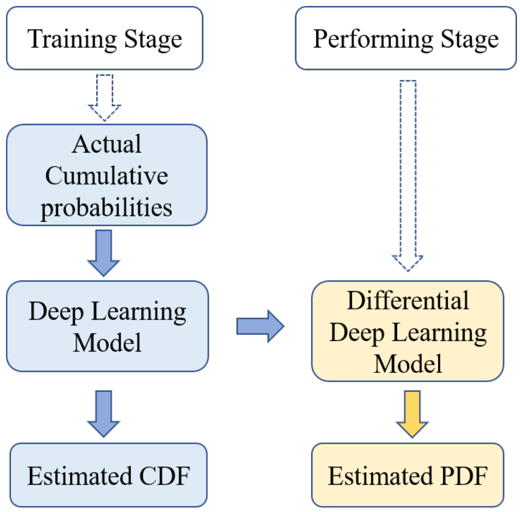

FIGURE 2. The suggested deep learning method for PDF

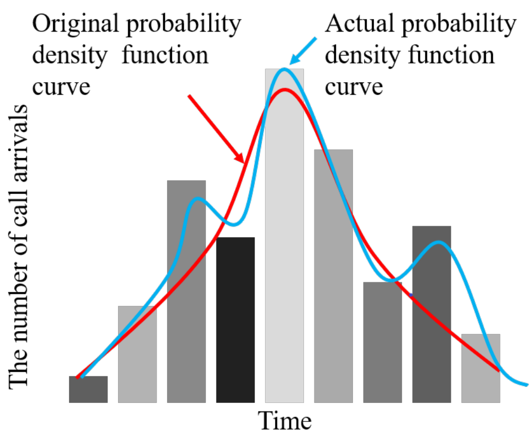

FIGURE 3. The difference between original probability density function curve and actual probability density function curve

This study utilizes some common probability distributions to simulate the proposed model. The MS communication data collected from CFVD was distributed exponentially [17, 20, 22, 30]. Whereas, the actual distribution of communication data should be complicated. As Fig. 3 indicates, the red curve is the normal distribution, and the blue curve is the actual probability distribution. Obviously, two curves are generally similar, but there are specific differences which can be distinguished. Therefore, in order to enhance the precision of practical probability distribution estimation , this study utilizes deep learning method to construct probabilistic analytic model and estimate corresponding CDF and PDF.

As Fig. 4 shows, the cumulative probabilities of actual communication data is acquired as the input of neural network model. In the hidden layer, several common distributions are proposed to be the activation functions. For example, the CDFs of normal distribution (ND), log-normal distribution (LD) and exponential distribution (ED) can be appropriate activations. The output proposed model is the CDF of actual data. Furthermore, the PDF of actual communication data can be estimated from the derivation of neural network model [43]. The notation in this study is outlined in Table 1. Equation 1 depicts the correlation between vehicle speed and the amount of CAs.

$$
\begin{aligned}
r_i &= Q_i \times \Pr\left[t_1 < t_2 < t_3\right] \\
&= \mu_1 \times \left(1 - e^{-\frac{1}{\mu_1}\left(\frac{l}{u_i}\right)}\right) + \frac{l}{2u_i} \\
&- \frac{1}{2} \times \left(\frac{l}{u_i} - \mu_2\right) \times \operatorname{erf}\left(\frac{\frac{l}{u_i} - \mu_2}{\sigma_2\sqrt{2}}\right) \\
&+ \left(-\frac{\mu_2}{2}\right) \times \operatorname{erf}\left(\frac{-\mu_2}{\sigma_2\sqrt{2}}\right) \\
&- \frac{\sigma_2\sqrt{2} \times e^{-\left(\frac{\frac{l}{u_i}-\mu_2}{\sigma_2\sqrt{2}}\right)^2} - \sigma_2\sqrt{2} \times e^{-\left(\frac{-\mu_2}{\sigma_2\sqrt{2}}\right)^2}}{\sqrt{\pi}} + \frac{l}{2u_i} \\
&- \frac{1}{2} \times e^{\mu_3 + \frac{\sigma_2^2}{2}} \times \left(\operatorname{erf}\left(\frac{-\ln\left(\frac{l}{u_i}\right) + \mu_3 + \sigma_3^2}{\sigma_3\sqrt{2}}\right) - 1\right) \\
&- \left(\frac{l}{2u_i}\right) \times \operatorname{erf}\left(\frac{\ln\left(\frac{l}{u_i}\right) - \mu_3}{\sigma_3\sqrt{2}}\right)
\end{aligned} \tag{1}
$$

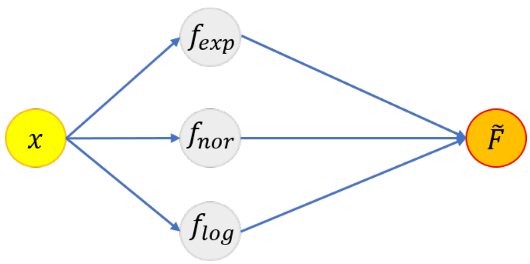

FIGURE 4. The neural network approach for reckoning the CDF

TABLE 1. The definition of summarized notation

Parameter	Description
i	The cell's number
Q_i(car/h)	The practical traffic flow of road segment covered by $Cell_i$
K_i(car/km)	The practical traffic density
U_i(km/h)	The practical vehicle speed
q_i(car/h)	The estimated traffic flow of road segment covered by $Cell_i$
k_i(car/km)	The estimated traffic density
u_i(km/h)	The estimated vehicle speed
r_i(event/h)	The amount of CAs
l_i(km)	The length of road segment covered by $Cell_i$
μ_1	The expected value of the exponential probability density function
μ_2	The expected value of the normal probability density function
μ_3	The expected value of the log-normal probability density function
σ_2	The variance of the normal probability density function
σ_2	The variance of log-normal probability density function

3.1.2. *A vehicle speed estimation approach.* The proposed model illustrates that CA and vehicle speed has interesting correlation. Thus, a single neural network is adopted to learn the proposed probabilistic analysis model and reckon the vehicle speed. The input is the amount of CAs and output is estimated vehicle speed of MSs. Thus, vehicle speed can be reckoned by the probabilistic analysis model.

3.2. **Traffic flow estimation.** Fig. 5 introduces the process of NLU when vehicle is moving on the road. The introduction of NLU events for traffic flow estimation is described below: NLU event is executed while location area (LA) changes [12]. With only one MS per vehicle, real traffic flow is Q_i. Thus, the estimated traffic flow (q_i) can be described as Equation 2.

$$q_i = Q_i \tag{2}$$

3.3. **Traffic density estimation.** The unit of traffic flow is vehicles per hour. The unit of vehicle speed is kilometers per hour. And the unit of traffic density is vehicles per kilometer. According to the unit of traffic flow and vehicle speed, traffic density (k_i) can be expressed as Equation 3.

$$k_i = \frac{q_i}{u_i} \tag{3}$$

4. **Simulation results.** This section considers a suitable experimental environment for traffic information estimation method and evaluates the feasibilities of traffic flow estimation, vehicle speed estimation and traffic density estimation.

4.1. **Experimental environment and methodology architecture.** This study designs a simulation experiment to sustain the feasibility of the proposed approach. Fig. 6 introduces the architecture of the proposed method which includes vehicle movement trace generation, MS communication trace generation, and deep learning model. The inputs of vehicle trace generator consist of the road conditions (e.g., traffic flow, the number of lanes, the length of the road, and the locations of handover points) and the vehicle movement behaviors (e.g., desired vehicle speeds, car-following model, and lane-changing model). The output of vehicle movement trace generation is a traffic trace file that documents vehicles' information (e.g., actual vehicle speed, location, time information and vehicle's

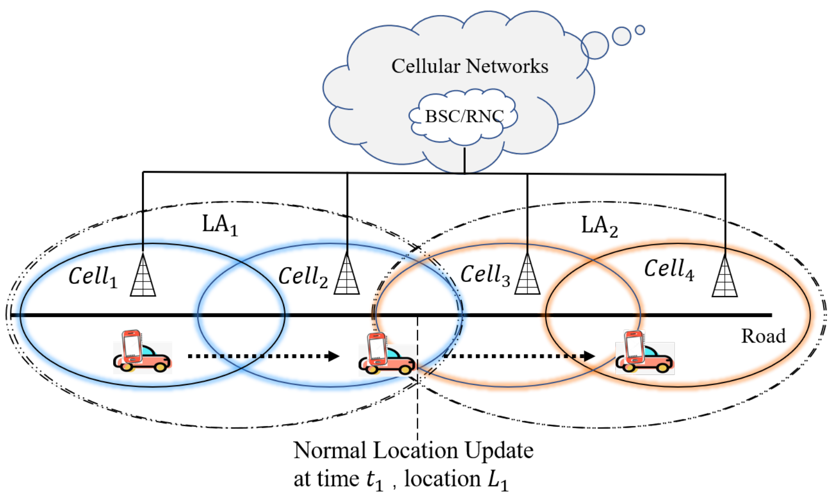

FIGURE 5. The spatial graph for NLU event

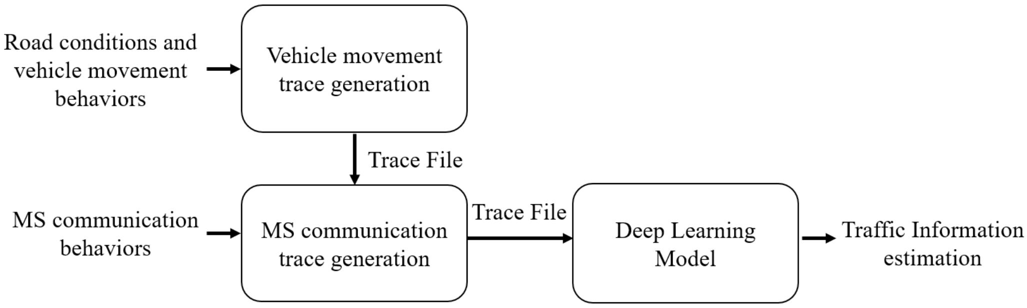

FIGURE 6. Trace-driven simulation for the traffic information estimations from CFVD and deep learning

number). The inputs of MS communication generator consist of the MS communication behaviors (e.g., normal location update, call inter-arrival time, and call holding time) and the vehicle movement behaviors. The output of MS communication trace generation is a trace file which includes each MS's information (e.g., vehicle's number, desired vehicle's speed, call arrival time, and call departure time). MS communication information includes CA's number. The estimated traffic flow is collected from vehicle movement trace generation.

This study generates traffic simulation data from traffic simulation tool VISSIM [44,45]. The situation of freeway is considered as which the vehicle speeds range from 85 km/h to 120 km/h and the length of a 3-lane freeway is 10 km. Fig. 7 indicates that there are 11 handover points and 10 cells distributed in four location areas on a freeway segment. Actual traffic information is gathered by VDs. It includes vehicle's ID, vehicle's speed, and the time information of each vehicle moving into each cell. The simulation data refers to reference [20]. The simulation time is set to 24 hours for objective results and high accuracy. The MS communication behaviors (e.g., call inter-arrival time and call holding time) are generated by a random number generator. The data of MS communication behaviors are both normally distributed. Their mean values are 1 h/call and 1 min/call, respectively.

The assumptions adopted in the experiments are concluded below.

- The actual vehicle speed U_i km/h, actual traffic flow Q_i cars/h, and traffic density K_i cars/km can be gathered from VDs on the road segment covered by $Cell_i$.
- The estimated vehicle speed u_i km/h, estimated traffic flow q_i cars/h and traffic density k_i cars/km can be gathered from CFVD on the road segment covered by $Cell_i$.
- The desired value of call inter-arrival time is 1 h/call according to the statistical results from [12].

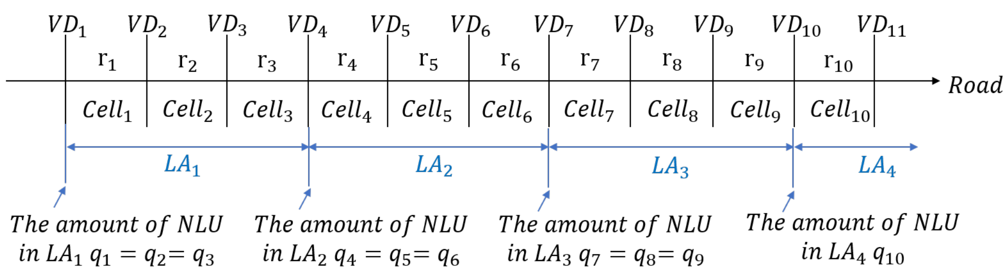

FIGURE 7. The spatial graph of simulation experiment

- The desired value of call holding time is 1 min/call according to the statistical results from [12].
- The accuracies of vehicle speed estimation, traffic flow estimation and traffic density estimation are described as $1 - \frac{|U_i - u_i|}{U_i}$, $1 - \frac{|Q_i - q_i|}{Q_i}$ and $1 - \frac{|K_i - k_i|}{K_i}$, respectively.

4.2. **The assessment of traffic information estimation Method.** This section introduces the assessment of traffic information estimation: the assessment of vehicle speed estimation based on probabilistic analysis model, the assessment of traffic flow estimation based on NLU and the assessment of traffic density estimation.

4.2.1. *The assessment of vehicle speed estimation.* Equation 1 indicates that the relationship between the vehicle speed and the number of call arrivals. The number of call arrivals had a mapping relationship with the inverse of the vehicle speed. Based on the mapping relationship, the feasible neural network model was trained and acquired. The model adopted designed activation functions which include the CDF of normal distribution, exponential distribution and log-normal distribution. From simulation and data analysis, the proposed neural network model generated better performance. The simulated amounts of CAs from $Cell_i$ are acquired from CFVD to estimate vehicle speed. Parameter settings were as follows: l_i=1.0km and μ_i=1 call/h. The accuracy of vehicle speed estimation is described as $1 - \frac{|U_i - u_i|}{U_i}$. From Table 2, the accuracy of estimated vehicle speed on road section wrapped by $Cell_i$ is 96.13%.

4.2.2. *The assessment of traffic flow estimation.* As Fig. 5 showed that a MS was moving from one LA to the next LA, the NLU event was performed. The simulated amounts of NLUs from $Cell_i$ were gathered by CFVD to generate q_i. The NLU event occurs when MS enters a new LA, so the estimated traffic flow covered by $Cell_i$ is consistent in the same LA, as Fig. 7 shows. The accuracy of traffic flow estimation was $1 - \frac{|Q_i - q_i|}{Q_i}$, as shown in Table 3. The experiment result indicated that the average accuracy of estimated traffic flow for 1 hour was 99.43%.

4.2.3. *The assessment of traffic density estimation.* For the assessment of traffic density estimation, the estimated traffic density was generated from the estimated vehicle speed based on the probabilistic analytical model and the estimated traffic flow based on NLU. This paper calculates the accuracy of traffic density estimation as $1 - \frac{|K_i - k_i|}{K_i}$. Table 4 indicates that the average accuracy of traffic density estimation covered by $Cell_i$ is 95.98%.

4.2.4. *The assessment of traffic information estimation.* From simulation results, shown in Table 5, the average accuracies of traffic information estimation for 24 hours was calculated. For the assessment of vehicle speed estimation, traffic flow estimation, and traffic density estimation, the mean values are 96.36%, 99.80%, and 96.45%, respectively.

4.3. **The comparison analysis with other methods.** This study proposed a probabilistic analysis model-based vehicle speed estimation algorithm for traffic information estimation. Compared with existing methods, this method uses deep learning to train a probabilistic analysis model. The model adopts the PDF of normal distribution, log-normal distribution and exponential distribution as the activation function. The PDF of actual communication data can be estimated accurately compared with linear function or sigmoid function. The simulation consequences implied that the proposed approach performs well with low computation cost in estimating vehicle speed. Compared with current related work, the proposed method has advantages in limitation of communication data. This method could learn the PDF of complex communication data and has a good performance. [12, 28]

TABLE 2. The estimation accuracies of vehicle speed.

Time	U_i	r_i	u_i	$1 - \frac{\lvert U_i - u_i \rvert}{U_i}$
1	85.38	49	84.97	99.51%
2	85.31	50	84.88	99.49%
3	85.19	45	85.25	99.93%
4	88.33	54	84.41	95.56%
5	88.78	57	83.9	94.5%
6	85.12	57	83.9	98.56%
7	75.56	77	75.15	99.46%
8	68.6	80	73.67	92.6%
9	67.88	103	69.13	98.15%
10	70.1	72	77.95	88.8%
11	84.08	49	84.97	98.94%
12	83.8	77	75.15	89.68%
13	84.01	52	84.67	99.22%
14	84.39	58	83.69	99.17%
15	83.71	50	84.88	98.6%
16	82.52	96	69.69	84.46%
17	70.47	88	70.99	99.27%
18	69.87	77	75.15	92.44%
19	70.32	98	69.49	98.83%
20	86.33	67	80.61	93.38%
21	87.91	60	83.2	94.64%
22	87.37	42	85.39	97.73%
23	85.25	44	85.3	99.94%
24	90.27	47	85.12	94.3%
Average				96.13%

TABLE 3. The estimation accuracies of traffic flow for 1 hour

Cell	LA	Q_i	q_i	$1 - \frac{\lvert Q_i - q_i \rvert}{Q_i}$
1	1	4001	4001	100.00%
2	1	4035	4001	99.16%
3	1	4059	4001	98.57%
4	2	4049	4049	100.00%
5	2	4075	4049	99.36%
6	2	4109	4049	98.54%
7	3	4140	4140	100.00%
8	3	4156	4140	99.62%
9	3	4179	4140	99.07%
10	4	4216	4216	100.00%
Average				99.43%

However, comparing with reference [37] can emerge some weak spots. Reference [37] proposed a cell probe-based algorithm to consider cellular network signals (i.e., NLUs, HOs, and CAs) for reckoning vehicle speed and the simulation results revealed that the accuracy of vehicle speed estimation by cell probe-based algorithm is 97.63%. Thus, deep

TABLE 4. The estimation accuracies of traffic density of the road section wrapped by $Cell_i$

Time	K_i	k_i	$1 - \frac{\lvert K_i - k_i \rvert}{K_i}$
1	46.86	47.09	99.51%
2	48.06	48.3	99.49%
3	50.5	50.47	99.93%
4	45.26	47.37	95.35%
5	43.94	46.5	94.18%
6	49.29	50.01	98.54%
7	66.1	66.45	99.46%
8	80.24	74.71	93.11%
9	98.68	96.89	98.18%
10	88.45	79.54	89.93%
11	64.29	63.61	98.95%
12	66.86	74.56	88.49%
13	65.42	64.91	99.22%
14	61.61	62.12	99.17%
15	65.69	64.79	98.62%
16	76.32	90.37	81.59%
17	80.93	80.34	99.27%
18	87.34	81.2	92.97%
19	86.74	87.76	98.82%
20	61.42	65.77	92.91%
21	52.3	55.26	94.34%
22	52.65	53.87	97.68%
23	50.47	50.45	99.94%
24	46.53	49.34	93.96%
Average			95.98%

TABLE 5. The estimation accuracies of vehicle speed, traffic flow and traffic on each road section

$Cell$	$1 - \frac{\lvert U_i - u_i \rvert}{U_i}$	$1 - \frac{\lvert Q_i - q_i \rvert}{Q_i}$	$1 - \frac{\lvert K_i - k_i \rvert}{K_i}$
1	96.13%	100.00%	95.98%
2	96.71%	99.73%	96.95%
3	96.66%	99.54%	96.73%
4	95.46%	100.00%	95.59%
5	96.73%	99.77%	96.72%
6	96.74%	99.62%	96.93%
7	96.42%	100.00%	96.51%
8	96.50%	99.77%	96.66%
9	95.35%	99.57%	95.62%
10	96.85%	100.00%	96.80%
Average	96.36%	99.80%	96.45%

learning-based vehicle speed estimation should be improved and vehicle speed estimation accuracy will be higher.

5. **Conclusions and future work.** At present, the car ownership is still increasing, which puts forward new requirements for ITS. Meanwhile, traffic congestion has wasted large amounts of common society resources. Using CFVD to acquire real-time traffic information has become a significant technique. It is cost-effective and convenient compared with VDs and GPS-based probe cars for ITS. In addition, CFVD-based method can avoid privacy issues. Thus, CFVD is becoming more and more popular. Some studies discussed the various methods of CFVD (e.g., HO-based method, CA-based method, and fingerprint-based method) for traffic information estimation. Experiments results indicated that the method based on CFVD had plenty potential for improvement. For instance, it is tough to apply HO-based method and CA-based method for urban road networks because of the complex environment. For traffic flow estimation, the CA-based method is convenient and accurate compared with other approaches.

A novel approach using deep learning and cellular floating vehicle data was proposed to estimate traffic information for ITS. For vehicle speed estimation, a probabilistic analysis model was proposed to deduce the amount of CAs. A neural network model was used to explore the correlation between vehicle speed and CA. The NLU programmes was adopted to acquire traffic flow data. The estimated vehicle speed and the estimated traffic flow generate traffic density estimation. For conclusion, deep learning and CFVD had a good cooperation effect. In simulations, the estimated data was obtained by VDs. Experimental results implied that the averaged accuracies of vehicle speed estimation and traffic density estimation are 96.36% and 96.45%, respectively. Therefore, this study provided a feasible approach for enhancing the capability of the ITS.

However, there are some limitations in the study. In the current stage, communication data may have different CDFs due to different time periods. Because the simulation site is a section of highway, the influence of complex factors brought by the environment is as low as possible. Thus, it is difficult to simulate in the urban road network [46–48]. Some improvements can be implemented in the future [49–51]. More wireless communication techniques (e.g., WiFi, LoRaWAN, and other wireless networks) could be considered for reducing estimated location errors [52–54]. There are many discussions about collecting traffic information while the cost is as low as possible. The popular method of gathering traffic information is VD such as magnetic or video technologies and the cost is very high. Furthermore, in this study, traffic flow estimation is accurate in a location area, but not in each cell. Some studies have contributed to estimating traffic flow from handover events and the estimated value was available. The next objective may combine the CFVD and semantic web methods [55–57] for improving estimation accuracy and qualifying new methods in urban road segments. Furthermore, the security techniques [58–60] could be considered to be applied for improving the privacy.

Acknowledgment. This work was partially supported by the National Natural Science Foundation of China (Nos. 61906043, 62002063, 61877010, 11501114, and 11901100), Fujian Natural Science Funds (Nos. 2020J05112, 2020J05111, 2019J01243), Funds of Education Department of Fujian Province (No. JAT190026), and Fuzhou University (Nos. 0330/50009113, 510872/GXRC-20016, 510930/XRC-20060, 510730/XRC-18075, 510809/GXRC-19037, 510649/XRC-18049, and 510650/XRC-18050).The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers,which have improved the presentation.

REFERENCES

[1] K. K. Bhardwaj, A. Khanna, D. K. Sharma and A. Chhabra (eds.), *Designing Energy-Efficient IoT-Based Intelligent Transport System: Need, Architecture, Characteristics, Challenges, and Applications*, Springer, Singapore, 2019.

[2] A. Ometov, S. Bezzateev, V. Davydov, A. Shchesniak, P. Masek, E. S. Lohan and Y. Koucheryavy, "Positioning information privacy in intelligent transportation systems: An overview and future perspective," *Sensors*, vol. 19, no. 7, 1603, 2019.

[3] N. Dadashzadeh, M. Ergun, S. Kesten and M. Žura, "An automatic calibration procedure of driving behavior parameters in the presence of high bus volume," *Promet-Traffic & Transportation*, vol. 31, no. 5, pp. 491-502, 2019.

[4] J. S. Pan, Z. Y. Meng, S. C. Chu and H. R. Xu, "Monkey King Evolution: an enhanced ebb-tide-fish algorithm for global optimization and its application in vehicle navigation under wireless sensor network environment," *Telecommunication Systems*, vol. 65, no. 3, pp. 351-364, 2017.

[5] Z. L. Ning, K. Y. Zhang, X. J Wang, M. S. Obaidat, L. Guo, X. P. Hu, B. Hu, Y. Guo, B. Sadoun and R. Y. K. Kwok, "Joint computing and caching in 5g-envisioned internet of vehicles: A deep reinforcement learning-based traffic control system," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1-12, 2020.

[6] J. S. Pan, C. Wang and T. W. Sung, "Road network modeling with layered abstraction for path discovery in vehicle navigation systems," *Journal of Information Hiding and Multimedia Signal Processing*, vol. 7, no. 4, pp. 791-801, 2016.

[7] M. A. Amin, S. Hadouej and T. S. J. Darwish, "Big data role in improving in-telligent transportation systems safety: A survey," in *International Conference on Emerging Internetworking, Data & Web Technologies*, pp. 187–199, Springer, Cham, 2019.

[8] V. Astarita, V. P. Giofr, G. Guidoet and A. Vitale, "A single inter-section cooperative-competitive paradigm in real time traffic signal settings based on floating car data," *Energies*, vol. 12, no. 3, 409, 2019.

[9] N. Formosa, M. Quddus, S. Ison, M. Abdel-Aty and J. Yuan, "Predicting real-time traffic conflicts using deep learning," *Accident Analysis & Prevention*, vol. 136, 105429, 2020.

[10] C. D. Yan, X. B. Wei, X. Liu, Z. G. Liu, J. X. Guo, Z. W. Li, Y. Lu and X. H. He, "A new method for real-time evaluation of urban traffic congestion: a case study in Xi'an, China," *Geocarto International*, vol. 35, no. 10, pp. 1033–1048, 2019.

[11] G. B. Sun, Y. J. Chiu, G. W. Lu and M. Xiong, "The Study of Dynamic Programming with Fuzzy Logic Energy Design and Simulation of Gear Shift for Electric Vehicles," *Journal of Network Intelligence*, vol. 4, no. 3, pp. 88–99, 2019.

[12] W. K. Lai, T. H. Kuo and C. H. Chen, "Vehicle speed estimation and forecasting methods based on cellular floating vehicle data," *Applied Sciences*, vol. 6, no. 2, 47, 2016.

[13] J. W. Zhuang, H. Luo, T. S. Pan and J. S. Pan, "Improved Flower Pollination Algorithm for the Capacitated Vehicle Routing Problem," *Journal of Network Intelligence*, vol. 5, no. 3, pp. 141–156, 2020.

[14] S. Ishida, M. Uchino, C. Li, S. Tagashira and A. Fukuda, "Design of Acoustic Vehicle Detector with Steady-Noise Suppression," *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, Auckland, New Zealand, pp. 2848–2853, 2019.

[15] L. L. Lou, J. Y. Zhang, Y. Xiong and Y. L. Jin, "A novel vehicle detection method based on the fusion of radio received signal strength and geomagnetism," *Sensors*, vol. 19, no. 1, 58, 2019.

[16] T. Ringwald, L. Sommer, A. Schumann, J. Beyerer and R. Stiefelhagen, "UAV-Net: A fast aerial vehicle detector for mobile platforms," *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 544–552, 2019.

[17] W. K. Lai, T. H. Kuo, C. H. Chen and D. R. Lee, "A vehicle speed estimation mechanism using handovers and call arrivals of cellular networks," *12th International Conference on Advances in Mobile Computing and Multimedia*, pp. 19–26, 2014.

[18] J. S. Pan, X. P. Wang, S. C. Chu and T. T. Nguyen, "A Multi-group Grasshopper Optimisation Algorithm for Application in Capacitated Vehicle Routing Problem," *Data Science and Pattern Recognition*, vol. 4, no. 1, pp. 41–56, 2020.

[19] Y. J. Byon, C. E. Cortés, Y. S. Jeong, F. J. Martínez, M. A. Munizaga and M. Zúñiga , "Bunching and headway adherence approach to public transport with gps," *International Journal of Civil Engineering*, vol. 16, no. 6, pp. 647–658, 2018.

[20] Z. H. Kan, L. L. Tang, M. P. Kwan, C. Ren, D. Liu and Q. Q. Li, "Traffic congestion analysis at the turn level using taxis' gps trajectory data," *Computers, Environment and Urban Systems*, vol. 74, pp. 229–243, 2019.

[21] Q. L. Ma and K. Kockelman, "A low-cost gps-data-enhanced approach for traffic network evaluations," *International Journal of Intelligent Transportation Systems Research*, vol. 17, no. 1, pp. 9–17, 2019.

[22] M. F. Chang, C. H. Chen, Y. B. Lin and C. Y. Chia, "The frequency of cfvd speed report for highway traffic," *Wireless Communications and Mobile Computing*, vol. 15, no. 5, pp. 879–888, 2015.

[23] H. H. Chen, H. Jin and S. L. Wu, "Minimizing Inter-server Communications by Exploiting Self-similarity in Online Social Networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 27, no. 4, pp. 1116–1130, 2016.

[24] J. C. Herrera, D. B. Work, R. Herring, X. J. Ban, Q. Jacobson and A. M. Bayen, "Evaluation of traffic data obtained via gps-enabled mobile phones: The mobile century field experiment," *Transportation Research Part C: Emerging Technologies*, vol. 18, no. 4, pp. 568–583, 2010.

[25] H. Long, Y. N. Chung and J. D. Li, "Automatic Vehicle Speed Estimation Method for Unmanned Aerial Vehicle Images," *Journal of Information Hiding and Multimedia Signal Processing*, vol. 9, no. 2, pp. 442–451, 2018.

[26] S. Maerivoet and S. Logghe, "Validation of travel times based on cellular floating vehicle data," *Proc. of the 6th European Congress and Exhibition on Intelligent Transport Systems and Services*, Aalborg, Denmark, pp. 18–20, 2007.

[27] P. C. Song, J. S. Pan and S. C. Chu, "A Parallel Compact Cuckoo Search Algorithm for Three-Dimensional Path Planning," *Applied Soft Computing*, vol. 94, 106443, 2020.

[28] C. I. Wu, C. H. Chen, B. Y. Lin and C. C. Lo, "Traffic information estimation methods from handover events," *Journal of Testing and Evaluation*, vol. 44, no. 1, pp. 656–664, 2016.

[29] G. B. Sun, Y. J. Chiu, J. H. Cao and Y. Wang, "A Dynamic Programming based Fuzzy Logic Energy Management Strategy for Series-parallel Hybrid Electric Vehicles," *Journal of Information Hiding and Multimedia Signal Processing*, vol. 10, no. 2, pp. 422–433, 2019.

[30] C. H. Chen, H. C. Chang, C. Y. Su, C. C. Lo and H. F. Lin, "Traffic speed estimation based on normal location updates and call arrivals from cellular networks," *Simulation Modelling Practice and Theory*, vol. 35, pp. 26–33, 2013.

[31] J. S. Pan, P. C. Song, S. C. Chu and Y. J. Peng, "Improved Compact Cuckoo Search Algorithm Applied to Location of Drone Logistics Hub," *Mathematics*, vol. 8, no. 3, 333, 2020.

[32] A. Janecek, D. Valerio, K. A. Hummel, F. Ricciato and H. Hlavacs, "The cellular network as a sensor: From mobile phone data to real-time road traffic monitoring," *IEEE transactions on intelligent transportation systems*, vol. 16, no. 5, pp. 2551–2572, 2015.

[33] Z. G. Du, J. S. Pan, S. C. Chu, H. J. Luo and P. Hu, "Quasi-Affine Transformation Evolutionary Algorithm With Communication Schemes for Application of RSSI in Wireless Sensor Networks," *IEEE Access*, vol. 8, pp. 8583–8594, 2020.

[34] B. Y. Lin, C. H. Chen, C. H. Lei and C. C. Lo, "The analysis of communication behaviors based on using real-time traffic information in personal communication systems," in *2011 International Conference in Electrics, Communication and Automatic Control Proceedings*, Springer, New York, NY, pp. 453–460, 2012.

[35] N. Caceres, L. M. Romero, F. G. Benitez and J. M. del Castillo, "Traffic flow estimation models using cellular phone data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 3, pp. 1430–1441, 2012.

[36] R. H. Liou, Y. B. Lin, Y. L. Chang, H. N. Hung, N. F. Peng and M. F. Chang, "Deriving the vehicle speeds from a mobile telecommunications network," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 3, pp. 1208–1217, 2013.

[37] C. H. Chen, "A cell probe-based method for vehicle speed estimation," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. 103, no. 1, pp. 265–267, 2020.

[38] W. K. Lai and T. H. Kuo "Vehicle positioning and speed estimation based on cellular network signals for urban roads," *ISPRS International Journal of Geo-Information*, vol. 5, no. 10, 181, 2016.

[39] B. Y. Lin, C. H. Chen and C. C. Lo "A novel speed estimation method using location service events based on fingerprint positioning," *Advanced Science Letters*, vol. 4, no. 11-12, pp. 3735–3739, 2011.

[40] S. C. Ergen, H. S. Tetikol, M. Kontik, R. Sevlian, R. Rajagopal and P. Varaiya, "Rssi-fingerprinting-based mobile phone localization with route constraints," *IEEE Transactions on Vehicular Technology*, vol. 63, no. 1, pp. 423–428, 2013.

[41] D. Y. Cheng, C. H. Chen, C. H. Hsiang, C. C. Lo, H. F. Lin and B. Y. Lin, "The optimal sampling period of a fingerprint positioning algorithm for vehicle speed estimation," *Mathematical Problems in Engineering*, vol. 2013, 306783, 2013.

[42] C. D. Chitraranjan, A. M. Denton and A. S. Perera, "A complete observation model for tracking vehicles from mobile phone signal strengths and its potential in travel-time estimation," *2016 IEEE 84th Vehicular Technology Conference (VTC-Fall)*, IEEE, pp. 1–7, 2016.

[43] C. H. Chen, F. Y. Song, F. J. Hwang and L. Wu, "A probability density function generator based on neural networks," *Physical A: Statistical Mechanics and its Applications*, vol. 541, 123344, 2020.

[44] M. Fellendorf and P. Vortisch, "Microscopic traffic flow simulator VISSIM," in *Fundamentals of traffic simulation*, Springer, New York, NY, pp. 63–93, 2010.

[45] C. Chen, X. Zhao, H. Liu, G. Ren, Y. Zhang and X. Liu, "Assessing the influence of adverse weather on traffic flow characteristics using a driving simulator and VISSIM," *Sustainability*, vol. 11, no. 3, 830, 2019.

[46] X. S. Xue and J. W. Lu, "A Compact Brain Storm Algorithm for Matching Ontologies," *IEEE Access*, vol. 8, pp. 43898–43907, 2020.

[47] T. K. Dao, J. S. Pan, T. S. Pan and T. T. Nguyen, "Optimal path planning for motion robots based on bees pollen optimization algorithm," *Journal of Information and Telecommunication* , vol. 1, no. 4, pp. 351–366, 2017.

[48] X. S. Xue, "A Compact Firefly Algorithm for Matching Biomedical Ontologies," *Knowledge and Information Systems*, vol. 62, pp. 2855–2871, 2020.

[49] C. C. Shih, M. F. Horng, T. S. Pan, J. S. Pan and C. Y. Chen, "A genetic-based effective approach to path-planning of autonomous underwater glider with upstream-current avoidance in variable oceans," *Soft Comput*, vol. 21, no. 18, pp. 5369–5386, 2017.

[50] X. S. Xue and J. F. Chen, "Optimizing Sensor Ontology Alignment through Compact co-Firefly Algorithm," *Sensors*, vol. 20, no. 7, pp. 1–15, 2020.

[51] J. S. Pan, N. Liu and S. C. Chu, "A hybrid differential evolution algorithm and its application in unmanned combat aerial vehicle path planning," *IEEE Access*, vol. 8, pp. 17691–17712, 2020.

[52] J. Q. Wu, Q. Gang and K. Pengbin, "Emerging 5G Multicarrier Chaotic Sequence Spread Spectrum Technology for Underwater Acoustic Communication," *Complexity*, vol. 2018, 3790529, 2018.

[53] X. M. Yu, H. Q. Wang, H. W. Lv, X. B. Liu and J. Q. Wu, "A fusion optimization algorithm of network element layout for indoor positioning," *EURASIP Journal on Wireless Communications and Networking*, vol. 2019, no. 1, 284, 2019.

[54] X. S. Xue and Y. P. Wang, "Optimizing ontology alignments through a Memetic Algorithm using both MatchFmeasure and Unanimous Improvement Ratio," *Artificial Intelligence*, vol. 223, pp. 65–81, 2015.

[55] X. M. Yu, H. Q. Wang and J. Q. Wu, "A method of fingerprint indoor localization based on received signal strength difference by using compressive sensing," *EURASIP Journal on Wireless Communications and Networking*, vol. 2020, no. 1, 72, 2020.

[56] C. Liu, J. Yang and J. Q. Wu, "Web intrusion detection system combined with feature analysis and SVM optimization," *EURASIP Journal on Wireless Communications and Networking*, vol. 2020, no. 1, 33, 2020.

[57] X. S. Xue and Y. P. Wang, "Using Memetic Algorithm for Instance Coreference Resolution," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 2, pp. 580–591, 2016.

[58] L. Lyu and C. H. Chen, "Differentially Private Knowledge Distillation for Mobile Analytics," *Proc. of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2020)*, Xi'an, China, pp.1809–1812, 2020.

[59] L. Lyu, J. Yu, K. Nandakumar, Y. Li, X. Ma, J. Jin, H. Yu and K.S. Ng, "Towards Fair and Privacy-Preserving Federated Deep Models," *IEEE Transactions on Parallel and Distributed Systems*, vol. 31, no. 11, pp. 2524–2541, 2020.

[60] L. Lyu, J.C. Bezdek, X. He and J. Jin, "Fog-embedded deep learning for the internet of things," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 7, pp. 4206–4215, 2019.