Context-Aware Mobile Intelligent Transportation Systems

Quang Tran Minh, Muhammad Ariff Baharudin, Eiji Kamioka
Graduate school of Engineering and Science, Shibaura Institute of Technology, Tokyo, Japan
{m709504, nb11507, kamioka}@shibaura-it.ac.jp

Abstract—**This paper proposes a practical quantification model for mobile phone based traffic state estimation systems (M-TES). The low penetration rate issue, an inherent issue impeding the realization of a mobile phone based application such as the M-TES, is thoroughly discussed. A notable solution framework, namely the intelligent context-aware velocity-density inference circuit (ICIC), is proposed to effectively resolve the low penetration rate issue. In the ICIC model, velocities and densities calculated directly from the *sensed* data and inferred by using different inference models such as the Greeshields or the moving average model are appropriately integrated. In addition, appropriate contexts extracted from data reported by mobile devices are utilized to identify the optimal estimation parameters leading to the optimal estimation effectiveness. The experimental evaluations reveal the effectiveness and the robustness of the proposed solutions.**

Key words: *context-aware, mobile probes, low penetration rate, genetic algorithm, GA, neural network, ANN, ITS, M-ITS*

I. INTRODUCTION

Traffic state estimation is one of the most important fields in Intelligent Transportation Systems (ITS) research. An ideal traffic state estimation system must provide not only accurate but also real-time traffic state information at any place (ubiquity). To collect real-time traffic data, conventional systems utilize road-side fixed sensors such as loop detectors [1], RFID readers [2], video cameras [3] and so on. The essential weakness of these approaches, however, is their coverage limitation since it is impractical to install a huge number of sensors at every street.

In the recent years, with the advances of mobile phone technologies, mobile devices have been utilized as traffic probes to collect real-time traffic data [4], [5]. Since mobile phones are available everywhere and the mobile phone network has already been deployed, the essential issues in the traditional road-side fixed sensor approaches such as coverage limitation, real-time effect, investment and maintenance cost can be overcome. As a result, the mobile phone based ITS (M-ITS) research is entering a new stage accelerating the realization of mobile phone based traffic state estimation systems (M-TES).

In spite of advantages mentioned above, the M-TES faces on several issues ranging from comprehensively estimating traffic state using less informative traffic data reported by mobile devices [6] to effectively removing errors come from low and uncertain penetration rate [7], [8]. Firstly, in the M-TES, traffic data is collected by mobile phones on which GPS (Global Positioning System) receiver is only the common sensor available. However, GPS-based data does not present the detailed traffic state information which drivers expected. Secondly, in the practice the portion of vehicles that report data to the estimation server out of the total number of vehicles traveling in the considered road segment is low. This portion is called the penetration rate. One may imagine that the lower the penetration rate is, the more data is missed, and thus the lower the estimation accuracy is. However, to the best of our knowledge, there is no research discusses these essential issue thoroughly.

This paper aims at proposing a practical traffic state quantification model based on less informative data reported by mobile phones and a notable approach to low penetration rate issues. The major contributions of this paper are as follows:

- Proposing a practical traffic state quantification (TSQ) model by which velocity and density are accurately estimated and then appropriately integrated to granularly identify the traffic state level.
- Proposing an intelligent context-aware velocity-density inference circuit (ICIC) by which contextual data extracted from the data reported by mobile devices is appropriately processed to provide accurate traffic state estimation.

This paper is organized as follows: Section 2 reviews the related work. Section 3 proposes a novel traffic state quantification model. The effects of low penetration rate on traffic state estimation is investigated thoroughly in section 4. Section 5 proposes a notable solution framework, namely the ICIC, for the low penetration rate issue. The feasibility and the effectiveness of the proposed solutions are evaluated in section 6, and section 7 concludes this work.

II. RELATED WORK

Existing traffic state estimation systems such as VICS [9], NAVITIME [10] in Japan, the ITS project at Kansas, USA [11] majorly rely on road-side fixed sensors for traffic data collection. This approach is costly in terms of initial installation and maintenance, thus the aforementioned systems confront the coverage limitation issue. Ad-hoc network technology [12] theoretically helps to improve the coverage but it is not matured enough to be applied in the real-world applications.

Mobile Millennium Project (MMP) [13] is closely related to our project which employs GPS-enable mobile phones as traffic probes for real-time data collection. An estimation server at the system's center processes traffic data, estimates traffic state, and informs drivers the traffic information. However, this system estimates traffic state by employing the dynamical theory to analyze vehicle flows on road networks. The dynamical theory may work effectively in an environment of narrow and short flows but it may reveal serious errors when being applied to a complex environment such as a long road network. Moreover, the essential issue of low penetration rate was not discussed thoroughly.

R. Herring, at el. [7] proposed to apply a statistical learning model to estimate traffic state in terms of *travel time* and *congestion state*. This work employed the historical data to

978-1-4673-1881-5/12/$31.00 ©2012 IEEE

train the statistical learning model and then apply the current observed data into the trained model to estimate/forecast the travel time and congestion state of the considered road segment. However, there still exist several issues that need to be thoroughly discussed and clarified. **First**, this work did not mention the effect of the density of the traffic flow on the congestion state. **Second**, the congestion state was just a "binary" indicator accepting only two states, namely *"congested"* and *"not-congested"*, which may bias the estimation accuracy since even the "blind" guessing approach has an opportunity to reach the accuracy of *50%*.

The study in [8] thoroughly analyzed the effect of low penetration rate on traffic state estimation, and then proposed appropriate solutions. The authors proposed a so-called "velocity-density inference circuit", which employed both the average velocity and density calculated directly from the *sensed* (the real) data and estimated using the Greenshields model [14], to improve the accuracy of the estimation model in cases of low penetration rate. However, the optimization of impact coefficients for corresponding parameters in the whole estimation model was not discussed thoroughly.

III. TRAFFIC STATE QUANTIFICATION

This section proposes a unique traffic state quantification model by which less informative, in terms of traffic state estimation, GPS data reported by mobile phones is effectively processed to granularly quantify traffic state levels.

Let $V=\{i|i = 1..N\}$ is the set of all road segments in a road network. For any road segment $i \in V$, traffic data is available at any time t. However, the obtained data (GPS data) is the event-based data which cannot be directly transformed into traffic state. Traffic state should be aggregated in predefined time intervals, namely in t-second windows. Concretely, traffic state is estimated at times $k = 0, t, 2t, \ldots$, where t is the aggregation time mentioned above. The task here is to effectively estimate traffic state of the considered road segment i at time interval k.

Obviously, *velocity* and *density* of a traffic flow directly reflect traffic state of a road segment. In this work, a notable traffic state quantification model is proposed by which *velocity* and *density* of a traffic flow are independently and directly estimated using the *sensed* data reported by mobile phones before being integrated in an appropriate way to identify traffic state in a granular level.

Definition 1: The *average velocity* of a traffic flow in the road segment i during time k, denoted as $V^{k,i}_{Avg}$, is the average velocity of vehicles traveling in the considered road segment.

The average velocity can be formally expressed in equation *(1)*. Here, $V^{k,i}_{t_m,j}$ is the velocity of any individual vehicle j $(j = 1 \ldots q)$ detected at time t_m $(m = 1, 2, \ldots, r)$ during time interval k $([k-1]t \leq t_m < kt)$, q is the total number of vehicles, and r is the total number of detection times during time interval k.

$$V^{k,i}_{Avg} = \frac{\sum_{j=1}^{q} V^{k,i}_{t_m,j}}{qr}, (k-1)t \leq t_m < kt \qquad (1)$$

Since the limited speed varies from road segment to road segment, the absolute average velocity defined above may not appropriately represent a traffic state in terms of travel time. Here, a new term, namely the *mean speed capacity*, is proposed to present the travel time at any particular road segment.

Definition 2: The *mean speed capacity* of the road segment i during time k, denoted as $M_V^{k,i}$, is defined in equation *(2)*, where, $V_{Max}{}^{i}$ is the limited speed of the road segment i.

$$M_V^{k,i} = \frac{V^{k,i}_{Avg}}{V^{i}_{Max}} \qquad (2)$$

Observation 1: The higher the M_V is, the better the traffic state is, and vice versa. In this work, the threshold of M_V is set to *0.6* for a good traffic condition in term of travel times.

Definition 3: The *density* in the road segment i during time k, denoted as $D^{k,i}$, is the fraction of the number of vehicles traveling through the considered road segment during time k out of the capacity of the considered road segment. The density is defined in equation (3).

$$D^{k,i} = \frac{q^{k,i}}{C^{k,i}} \qquad (3)$$

Here, $q^{k,i}$ is the total number of vehicles traveling through the road segment i during time k which is estimated based on the *sensed* data reported by mobile phones, and $C^{k,i}$ is the *flow capacity* of the road segment i.

Definition 4: The *flow capacity*, denoted as $C^{k,i}$, is the maximum number of vehicles which can pass through the road segment i during time k under the best traffic condition. The *flow capacity* can be calculated in equation *(4)*.

$$C^{k,i} = Q^{i}{}_{0} + Q^{k,i} \qquad (4)$$

where, Q^{i}_{0} is the maximum number of vehicles that can be arranged (without moving) in the road segment i, and $Q^{k,i}$ is the volume of the traffic flow passing the down-stream boundary of the road segment i during time k in the best traffic condition. The static and dynamic volumes, namely Q^{i}_{0} and $Q^{k,i}$, are calculated in equations *(5)* and *(6)*, respectively.

$$Q^{i}{}_{0} = m * \frac{l}{1.5 * l_c} \qquad (5)$$

$$Q^{k,i} = m\frac{k}{\bar{t}} = mk\frac{V^{i}_{Max}}{1.5 l_c} \qquad (6)$$

Here, m and l are the number of the lanes and the length of the road segment i, respectively, and l_c is the average length of a car which is set to *5m* in this work [15]. The value of *1.5* is the coefficient describing the space which must be yielded between two cars in the worst congested area. In equation *(6)*, $\bar{t}$ is the average elapse time between two consecutive vehicles, namely vehicles i^{th} and $(i+1)^{th}$, passing the down-stream boundary. Figure *1* illustrates these parameters.

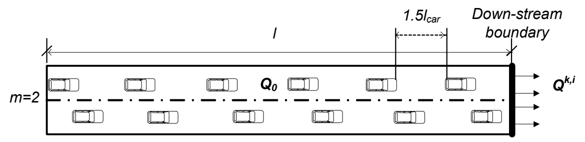

Figure 1. Capacity, $C^{i} = Q^{i}_{0} + Q^{k,i}$, of a 2-lane road segment (segment i)

For convenience in estimating traffic state in the manner of *density*, a new term called the *FREE space ratio* is proposed.

Definition 5: The *FREE space ratio* of the road segment i during time k, denoted as $\sigma^{k,i}_{s}$, is calculated in equation *(7)*, where $D^{k,i}$ and $C^{k,i}$ were defined in definitions *3* and *4,* respectively.

$$\sigma^{k,i}{}_S = \frac{C^{k,i} - D^{k,i}}{C^{k,i}} = 1 - \frac{D^{k,i}}{C^{k,i}} \qquad (7)$$

Observation 2: The higher the *FREE space ratio* is, the better a traffic state is. In this work, the threshold of σ_s is set to *0.4* for a good traffic condition.

After *mean speed capacity* and *FREE space ratio* being obtained, they should be integrated in an appropriate way to quantify the traffic state. In this work, a new term called the *Goodness value* which can be used to quantify traffic state is proposed.

Definition 6: The *Goodness value* of the road segment *i* during the estimation time *k*, denoted as $G^{k,i}(M^{k,i}{}_v, \sigma^{k,i}{}_s)$, is calculated in equation *(8)*, where M_{v0} and σ_{s0} are the thresholds of the *mean speed capacity* and the *FREE space ratio.*

$$G(M^{k,i}{}_v, \sigma^{k,i}{}_s) = (M^{k,i}{}_v - M_{v0}) + (\sigma^{k,i}{}_s - \sigma_{s0}) \qquad (8)$$

It should be noted that the *Goodness value* is continuous value. Therefore, it is quite adequate to be used for granularly comparing traffic state levels.

IV. Effect of Penetration Rate on Traffic State Estimation

In the M-TES, traffic state is estimated based on the traffic data reported by mobile phones carried by vehicles. However, in practice it is not necessary that every vehicle reports data to the estimation server. Therefore, the M-TES must have to overcome the *low* penetration rate related issues.

Definition 7: The *penetration rate* at the road segment *i* during time *k*, denoted as $\rho^{k,i}$ and calculated in equation *(9)*, is the fraction of the number of vehicles that report data *(p)* out of the total number of vehicles *(q)* traveling through the considered road segment during time *k*.

$$\rho^{k,i} = \frac{p}{q} \qquad (9)$$

With a given penetration rate $\rho^{k,i}$, the average velocity estimation model described in equation *(1)* must be replaced with equation *(10)*. Here, $V^{k,i}_{t_m,j}$ is the velocity of an individual vehicle *j* *(j = 1..q)* detected at time t_m *(m = 1, 2,...r)* during time interval *k* *([k-1]t* $\le t_m <$ *kt)*, *q* is the total number of vehicles traveling in the considered road segment, and *r* is the total data reporting times during time *k*.

$$V^{k,i,\rho^{k,i}}_{Avg} = \frac{\sum_{j=1}^{\rho^{k,i}q} V^{k,i}_{t_m,j}}{\rho^{k,i} qr}, (k-1)t \le t_m < kt \qquad (10)$$

Under this condition of penetration rate, the average velocity estimation error, $E^{k,i}{}_V$, can be expressed in equation *(11)*, where $V^{k,i}{}_{Avg}$ is the "actual" average velocity estimated when every vehicle reports data to the estimation server (equation *(1)*), and $V^{k,i,\rho^{k,i}}_{Avg}$ is the average velocity estimated under the given penetration rate $\rho^{k,i}$ (equation *(10)*).

$$E^{k,i}_V = \left| 1 - \frac{V^{k,i,\rho^{k,i}}_{Avg}}{V^{k,i}_{Avg}} \right| = \left| 1 - \frac{1}{\rho^{k,i}} \cdot \frac{\sum_{j=1}^{\rho^{k,i}q} V^{k,i}_{t_m,j}}{\sum_{j=1}^{q} V^{k,i}_{t_m,j}} \right| \qquad (11)$$

Similar to average velocity, density estimation is also affected by the penetration rate. According to the density definition described in equation *(3),* the density estimation error, denoted as $E^{k,i}_D$, is directly affected by $\rho^{k,i}$ and expressed in equation *(12)*.

$$E^{k,i}_D = (1 - \rho^{k,i}) * 100\% \qquad (12)$$

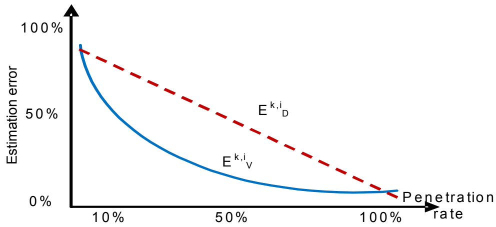

Figure 2. Effect of penetration rate on *velocity* and *density* estimations

Figure 2 shows the effect of penetration rate on velocity and density estimations. The $E^{k,i}{}_D$ is linearly affected by the penetration rate, while the $E^{k,i}{}_V$ is not at such direct effect but it is also affected by the penetration rate significantly. This issue will be solved in the remainder of this paper.

V. Intelligent Context-Aware Velocity-Density Inference Circuit (ICIC)

In order to alleviate errors rooted from low penetration rate mentioned in the previous section, a novel velocity-density inference mechanism, namely the intelligent context-aware velocity-density inference circuit (ICIC), is proposed. This mechanism is illustrated in Fig.3.

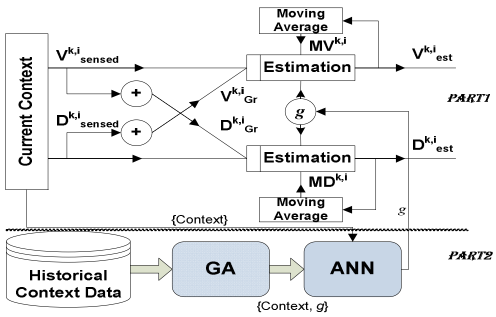

Figure 3. Intelligent context-aware velocity-density inference circuit (ICIC)

The proposed ICIC model consists of 2 parts separated by a dotted line. Part 1 is the velocity-density inference circuit and part 2 is the mechanism that identifies the optimal parameters for the inference model in part 1.

In part 1, *velocity* and *density* calculated directly from the *sensed* data reported by mobile phones, namely $V^{k,i}{}_{sensed}$ and $D^{k,i}{}_{sensed}$, serve as the primary inputs. The outputs of the circuit are the final estimated velocity and density, namely $V^{k,i}{}_{est}$ and $D^{k,i}{}_{est}$. The intermediate velocity and density, namely $V^{k,i}{}_{Gr}$, $D^{k,i}{}_{Gr}$, obtained by applying the Greenshields model [14] are also taken into account. In addition, moving average values of estimated velocity and density at time *k*, namely $MV^{k,i}$ and $MD^{k,i}$, calculated using the corresponding values estimated in the previous phases are *fed* back to the estimation model.

The philosophies behind this inference circuit are as follows: *1) Velocity* and *density* calculated independently from *sensed* data help to void any error propagation. *2)* The Greenshields model [14] used to infer density from estimated velocity and vice versa, can help to avoid the over-error of density estimation when penetration rate is unacceptably low. *3)* Current traffic state contains inherent relations with previous traffic states at the same road segment. *4)* All of the estimation approaches (direct estimation using *sensed* data, inference using the Greenshields model, inference using the previous estimated data) may uphold their advantages while diminishing their inherent disadvantages if being appropriately integrated.

The overall velocity-density inference model is formally presented in equations *(13)* and *(14)*, where α, β, γ are the impact coefficients of the corresponding parameters. It should be noted that α, β, γ are encapsulated in the simplified parameter g, namely $g = \{\alpha, \beta, \gamma\}$, presented in Fig.3.

$$V_{est}^{k,i} = \alpha V_{sensed}^{k,i} + \beta MV^{k,i} + \gamma V_{Gr}^{k,i} \quad (13)$$

$$D_{est}^{k,i} = \gamma D_{sensed}^{k,i} + \beta MD^{k,i} + \alpha D_{Gr}^{k,i} \quad (14)$$

Here, $MV^{k,i}$, $MD^{k,i}$, $V^{k,i}_{Gr}$ and $D^{k,i}_{Gr}$ are calculated in equations *(15)*, *(16)*, *(17),* and *(18).* In these equations, D^{i}_{max} and V^{i}_{max} are the maximum density and the limited velocity of the road segment i; and ξ is the sliding window for moving average calculations which can be set by domain experts or by using simulation data.

$$MV^{k,i} = \frac{\sum_{j=k-\xi}^{k-1} V_{est}^{j,i}}{\xi} \quad (15)$$

$$MD^{k,i} = \frac{\sum_{j=k-\xi}^{k-1} D_{est}^{j,i}}{\xi} \quad (16)$$

$$D_{Gr}^{k,i} = D_{\max}^{i}\left(1 - \frac{V_{sensed}^{k,i}}{V_{\max}^{i}}\right) \quad (17)$$

$$V_{Gr}^{k,i} = V_{\max}^{i}\left(1 - \frac{D_{sensed}^{k,i}}{D_{\max}^{i}}\right) \quad (18)$$

The most important key contributing to the estimation model described above is the optimization of coefficients α, β, γ. In this work, an intelligent machine learning based approach is proposed to optimize these coefficients based on the current contexts. Part 2 in Fig.3 illustrates this component. An appropriate artificial neural network (ANN) [16] model is employed to learn the rules of inferring *g (g = { , , })* given a set of contexts from the historical contextual database. At the estimation time, current contexts are fed to the ANN model, so that it can infer the set of coefficients, *g*, using the knowledge it learnt before.

Dey [17] defines context as *"Context is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves."* The *"situation"* or the *"status"* of a certain entity may be clearly represented by the entity itself. However, in practice, entity's status is still a variable which must be "inferred". The inference model based on contextual data surrounding the entity at the estimation time can be interpreted as "context-aware inference". In this work, the "status" to be inferred is the *traffic state*.

Obviously, it is not necessary that every context is useful and can be sensed (obtained) successfully under the given infrastructural conditions of the proposed application. In the M-TES's circumstance, only GPS data is given. The system has to infer velocity and density from this data for estimating traffic state as discussed in section 3. Therefore, in this work, feasible and useful contexts are selected carefully as follows:

- V_{sensed}/V_{max} represents both the information related to the real-time data, V_{sensed}, and the physical feature of the considered road segment, namely the limited speed, V_{max}.
- V_{sensed}/V_{Gr} reflects the "quality" of the *sensed* data under the effect of penetration rate.
- D_{sensed}/D_{max} represents the current recognition of the density based on the real-time data. This factor is identical to the V_{sensed}/V_{max} but in terms of density.
- MV^{t} (the moving average of velocity at the estimation time t) serves as a factor that transits previous traffic states to the current traffic state.

Table 1 presents a portion data extracted from evaluated data showing that coefficients α, β, γ were successfully inferred from current contextual data.

V_{sensed}/V_{max}	**V_{sensed}/V_{Gr}**	**D_{sensed}/D_{max}**	**MV^{t}**	**α**	**β**	**γ**
0.2349	0.2779	0.1545	66.4811	0.6660	0.0395	0.2945
0.1682	0.2104	0.2004	41.4499	0.6621	0.2881	0.0497
0.1115	0.1454	0.2331	26.7230	0.5543	0.3969	0.0488
0.1172	0.1536	0.2369	14.5153	0.5752	0.3973	0.0275
0.1444	0.1977	0.2697	13.0120	0.8554	0.1446	0.0000
0.5807	0.6252	0.0712	74.7853	0.6468	0.1996	0.1536
0.2935	0.3378	0.1311	60.6069	0.6293	0.2867	0.0840
0.0938	0.1258	0.2547	15.3041	0.7937	0.1498	0.0565
0.4878	0.6231	0.1086	74.3697	0.4981	0.2393	0.2625
0.2031	0.4155	0.2556	56.6439	0.3786	0.1972	0.4242

Table 1. Coefficients α, β, γ were inferred from current contextual data

It should be noted that α, β, γ are real-value numbers ranging in [0, 1] summing to 1. Finding a combination of α, β, γ that satisfies these 2 conditions which is optimal (resulting in the estimated velocity and density that are close to the "actual" velocity and density) is an NP-hard problem. To solve this issue, an appropriate genetic algorithm (GA) [18] based mechanism is proposed to process the historical contextual data and provides the ANN component a set of *{contexts, g}* as shown in Fig.3.

Since α, β, γ are real-value numbers, the proposed GA mechanism must have the ability of working with chromosomes (solutions) modeled by real-value numbers [18]. More concretely, the schema of a chromosome in the GA mechanism is coded as $g = \{\alpha, \beta, \gamma\}$ (the chromosome of *3* genes). Candidates are evaluated using the *fitness function* proposed in equation *(19)*.

$$f(g_i) = \frac{e(g_i)}{\bar{e}(g_j \mid g_j \in population)} \quad (19)$$

Here, $e(g_i)$ is the evaluation of candidate g_i ($g_i = \{\alpha_i, \beta_i, \gamma_i\}$) and $\bar{e}(g_j)$ is the average evaluation of all individuals g_j in the current population. The evaluation $e(g_i)$ is the estimation error, namely the velocity estimation error in equation *(13)* caused by selecting g_i as the set of coefficients. The evaluation $e(g_i)$ is defined in equation *(20),* where V_{est_gi} is the estimated velocity with the set of coefficient g_i, and V_{act} is the *"actual"* velocity.

$$e(g_i) = \frac{|V_{est\_gi} - V_{act}|}{V_{act}} \quad (20)$$

The proposed ICIC model can be summarized as follows: The GA mechanism provides optimal set of *{contexts, g}* to train the ANN model. At the estimation time, the current contextual data is fed to the ANN model to identify the optimal traffic state estimation coefficients. These coefficients are forwarded to the velocity-density inference circuit in the part 1 to effectively estimate traffic state of a certain road segment. The effectiveness of the proposed ICIC model will be evaluated in the next section.

VI. EVALUATIONS

This section evaluates the effectiveness of the proposed traffic state quantification (TSQ) model and the ICIC approach.

1) *Effectiveness of the TSQ Model*

The novelty of the proposed TSQ model is its capability of granularly quantifying traffic state using the real-time traffic data collected by mobile phones. This section evaluates the effectiveness of the proposed model by comparing its quantification results with the human being evaluations. The evaluations were conducted using the 1st prototype of the proposed TSQ model. Three video clips, namely v1, v2, v3 of real traffic flows on the National Route No.16 (Higashi Omiya, Saitama, Japan) were taken. People who are experienced in driving were asked to evaluate traffic states of these traffic flows by watching corresponding video clips. The evaluation results depicted in Table 2 shows the consistency of the proposed TSQ model with the human being evaluations. The human being rated the traffic states in v2, v1, v3 as 6, 7, 8 (in the range of 1 to 10 shorting from the better to the worse traffic states), respectively, which are consistent with the rating provided by the proposed TSQ model as -0.75, -0.82, -0.96 for v2, v1, v3, respectively. In addition, the TSQ model provided more detailed information about the serious level of a certain traffic state via three quantitative indicators, namely the *mean speed capacity* (M_v), the *FREE space ratio* (σ_s), and the *Goodness value* ($G(M_v, \sigma_s)$). These indicators help drivers comprehend a traffic state better. More concretely, $G(M_v, \sigma_s)$ describes the traffic state in general, M_v represents the traffic state in terms of velocity, and σ_s describes the density of the traffic flow.

Traffic flow 1 (v1)	Traffic flow 2 (v2)	Traffic flow 3 (v3)
- 100m length, 2 lanes - Main inflow: 4760 vehicles/h - No. of trucks/ No. of cars = 8.7%	- 100m length, 2 lanes - Main inflow: 3360 vehicles/h - No. of trucks/ No. of cars = 27%	- 130m length, 2 lanes - Main inflow: 4800 vehicles/h - No. of trucks/ No. of cars = 25%
$M_v = 0.1026 \pm 0.017$ $\sigma_s = 0.0769 \pm 0.0132$ $G(M_v,\sigma_s) = -0.82 \pm 0.031$	$M_v = 0.136 \pm 0.022$ $\sigma_s = 0.1154 \pm 0.0188$ $G(M_v,\sigma_s) = -0.75 \pm 0.04$	$M_v = 0.043 \pm 0.0073$ $\sigma_s = 0 \pm 0.0041$ $G(Mv, \sigma_s) = -0.96 \pm 0.014$
Evaluations from 18 people Given a range of 1 to 10 represents from the best to the worst traffic states, people are asked to quantify these traffic state levels by watching the corresponding video clips. Most of them quantified v2-v1-v3 as 6-7-8.		

Table 2. The effectiveness of the TSQ model

2) *Effectiveness of the ICIC Model*

To evaluate the effectiveness of the proposed ICIC model, a large amount of data is required. In this work, the TSF simulator [19] was utilized to generate synthetic data for evaluations. Different road segments were selected randomly as shown in Fig.4. For each selected road segment, two types of data were created concurrently as follows:

a) The GPS data reported by individual vehicles including time stamp (in second), road segment Id, position (longitude, latitude), current velocity, and vehicle Id of the vehicle that report the data. Different penetration rates, namely 20%, 25%, 30%, and so forth, were configured by the TSF. With a settled penetration rate, only such percentage of random vehicles, namely 20% vehicles, for example, reported the data to the server. The frequency of the data report timing was set to every 3s (similar to the common GPS signal frequency).

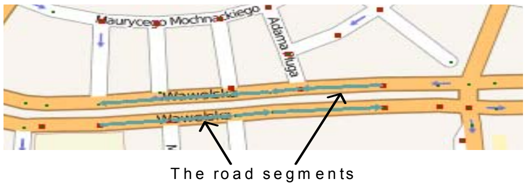

Figure 4. Road segmentation in the TSF

b) The summarized data on the traffic state of the selected road segments was recorded. Each record contains the information of *time interval Id* (in minute), *road segment Id*, *average velocity,* and *density*. The time interval for recoding the summarized information was set to every minute. This summarized traffic state information was used to evaluate the accuracy of the estimation methods applying the GPS data described in *a)*.

In this section the overall effectiveness of the ICIC model will be presented. However, as depicted in Fig.3, the GA component plays an important role in processing a large amount of historical data to provide the ANN component an optimal set of *{contexts, g}*. Therefore, the feasibility (represented by the convergence rate) as well as the computation time of the GA component is evaluated first.

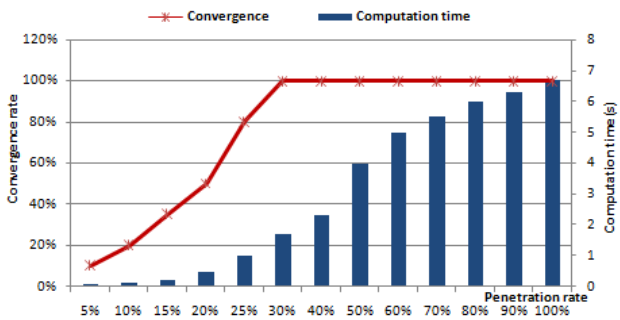

Figure 5. The convergence and the computation time of the proposed GA

The proposed GA mechanism was implemented with the error threshold of 0.1% (i.e. the maximum distance between V_{est} and V_{act} to be accepted is 0.1% of V_{act}), and the maximum number of iterations was set to 500,000. The GA is convergent if the distance between V_{est} and V_{act} reaches the preset error threshold before the maximum iteration number is hit. The convergence and computation time of the proposed GA are affected by the quality and the amount of the evaluated dataset. These qualifiers are affected by the penetration rate. As show in Fig.5, when the penetration rate is low, namely lower than 20%, the convergence rate is quite low, namely lower than 50% and the computation time is small, namely less than a second. This effect is because that when the penetration rate is low a lot of useful data is missed, so that the GA cannot find out the optimal solution (cannot converge). Moreover, the dataset in this case is small thus less computation time is required. When the penetration rate is relevant, namely larger than 25%, the convergence rate reaches 100%. It should be noted that, when the penetration rate is larger, more data is collected revealing that longer computation time is required. However, as shown, the maximum computation time is limited to 7s (when the penetration rate is 100%), which is fast enough.

The overall effectiveness of the proposed ICIC is shown in Fig.6 by comparing with that of the conventional method where the average velocity and density is estimated directly based on the *sensed* data. Here, the accuracy of the velocity and density in the conventional and the ICIC model are denoted as *Cont _V*, *Cont_D, ICIC_V,* and *ICIC_D,* respectively.

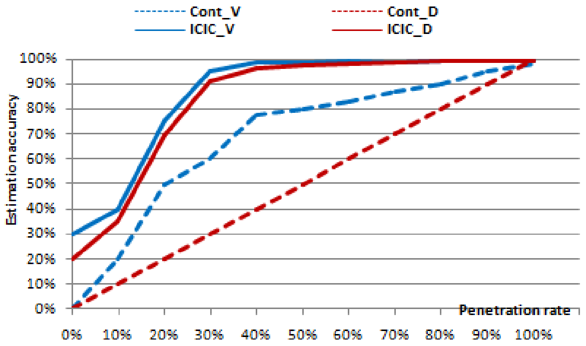

Figure 6. Effectiveness of the ICIC model

Figure 6 shows that the proposed ICIC model is significantly more effective than the conventional model in both velocity and density estimations. For example, if the expected accuracy is set to 80%, the penetration rates required in the conventional method are as large as 50% and 80% for the velocity and density estimation, respectively. Mean while, the ICIC model requires only around 25% of penetration rates for both the velocity and density estimations. In addition, the accuracy of velocity and density estimations in the conventional method, namely the *Cont_V* and *Cont_D*, confirms the different effects of penetration rate on these estimations as discussed in section 4 and depicted in Fig.2. More concretely, the effectiveness of the density estimation effectiveness is more significantly affected by the penetration rate compared to that of velocity estimation. This is also the reason of why the *ICIC_V* seems to be better than its counterpart, the *ICIC_D*. It should be noted that, even though the ICIC model provides high estimation accuracy when the penetration rates are relevant, namely larger than 20%, it cannot work properly if the penetration rate is unacceptably low, namely lower than 20%. The reason is that with unacceptably low penetration rates, the quality of the contexts extracted from the *sensed* data reported by mobile phones is declined.

VII. CONCLUSIONS

This work proposed a unique traffic state quantification (TSQ) model by which both velocity and density of the considered traffic flow is integrated in an appropriate way to provide traffic state information in a granular level. The effectiveness of the TSQ model was evaluated by using the real traffic data and was confirmed by human being evaluations.

This work also discussed the effect of penetration rate on mobile phone based traffic state estimation model thoroughly leading to suitable solutions for real world applications. A notable intelligent context-aware velocity-density inference circuit (ICIC) was proposed to solve the low penetration rate issue, an essential issue in any mobile phone-based application. Therefore, the ICIC model is not only useful in the M-TES but also can be flexibly applied in any mobile phone-based application. The experimental results revealed that the ICIC provides significantly high estimation accuracy.

However, as any inference approach, the proposed ICIC model cannot properly work in cases of unacceptably low penetration rate, namely lower than 20%. Beside that, 20% of penetration rate is still considered to be high in practice for a mobile phone-based application such as the M-TES. Therefore, finding a sound solution for the issues of unacceptably low or even uncertain penetration rate is still a challenging task which is deferred to the future work.

REFERENCES

[1] S. Tang and F.-Y. Wang, *"A PCI-based evaluation method for level of services for traffic operational systems," IEEE Trans. Intell. Transp. Syst.*,vol. 7, no. 4, pp. 494–499, Dec. 2006.

[2] X. Ban, Y. Li, A. Skabardonis, and J.D. Margulici, *"Performance evaluation of travel time methods for real time traffic applications."* In 11th World Conference on Transport Research, Berkeley, CA, June 2007.

[3] Y. Cho and J. Rice, *"Estimating velocity fields on a freeway from lowresolution videos," IEEE Trans. Intell. Transp. Syst.*, vol. 7, no. 4, pp. 463–469, Dec. 2006.

[4] G. Rose, "*Mobile Phones as Traffic Probes: Practices, Prospects and Issues,*" Transport Reviews, vol. 26, no. 3, pp.275-291, May 2006.

[5] K. Sohn and K. Hwang, *"Space-based passing time estimation on a freeway using cell phones as traffic probes," IEEE Trans. Intell. Transp. Syst.*, vol. 9, no. 3, pp. 559–568, Sep. 2008.

[6] T.M. Quang, E. Kamioka, *"Traffic State Estimation with Mobile Phones Based on The "3R" Philosophy,"* IEICE Transactions on Communications, Vol.E94-B, No.12, pp.3447-3458, Dec. 2011.

[7] R. Herring, A. Hofeitner, S. Amin, T. A. Nasr, A. A. Khalek, P. Abbeel, A. Bayen *"Using Mobile Phones to Forecast Arterial Traffic through Statistical Learning,"* The 89th Annual Meeting of the Transportation Research Board, Aug. 2009.

[8] T. M. Quang and E. Kamioka: "*Adaptive Approaches in Mobile Phone Based Traffic State Estimation with Low Penetration Rate,"* Journal of Information Processing (IPSJ) Vol.20, No.1, pp.297-307, Jan.2012.

[9] Ministry of Land Infrastructure and Transport: *"The System Outline of VICS,"* http://www.vics.or.jp.

[10] Route search and location based consumer services in Japan http://www.navitime.co.jp.

[11] Kansas ITS: http://www.ksdot.org/burtransplan/burovr/inttrans.asp.

[12] N. Shibata, T. Terauchi, T. Kitani, K.Yasumoto, M. Ito, T. Higashishino, *"A Method for Sharing Traffic jam Inforamtion Using Inter-Vehicle Communication,"* mobiquitous, 3rd Annual International Conference on Mobile and Ubiquitous Systems – Workshops, pp. 1-7, 2006.

[13] http://traffic.berkeley.edu/theproject.html.

[14] B. D. Greenshields "*A Study in Highway Capacity,"* Highway Research Board Proceedings, pp. 448-477, 1935.

[15] The average size of a car: http://en.wikipedia.org/wiki/Full-size_car.

[16] D. E. Rumelhart, B. Widrow, M. A. Lehr, "*The basic ideas in neural networks*," Communications of the ACM, v.37 n.3, pp.87-92, Mar. 1994.

[17] A.K. Dey, *"Understanding and Using Context,"* Personal and Ubiquitous Computing Journal, Vol. 5, lss. 1, pp. 5-7 , 2001.

[18] Z. Michaelewicz, *"Genetic algorithms + data structures = evolution programs,"* (3rd ed.), Springer-Verlag, Berlin, 1996.

[19] P. Gora. *"Traffic Simulation Framework — a Cellular Automaton-Based Tool for Simulating and Investigating Real City Traffic,"* Recent Advances in Intelligent Information Systems ISBN 978-83-60434-59-8, pp. 641–653, 2009.