Journal of Information Processing Vol.20 No.1 297–307 (Jan. 2012)

[DOI: 10.2197/ipsjjip.vol.20.297]

Regular Paper

Adaptive Approaches in Mobile Phone Based Traffic State Estimation with Low Penetration Rate

Quang Tran Minh[1,a) Eiji Kamioka[1,b)

Received: March 31, 2011, Accepted: October 3, 2011

Abstract: The penetration rate is one of the most important factors that affects the effectiveness of the mobile phone-based traffic state estimation. This article thoroughly investigates the influence of the penetration rate on the traffic state estimation using mobile phones as traffic probes and proposes reasonable solutions to minimize such influence. In this research, the so-called "acceptable" penetration rate, at which the estimation accuracy is kept as an "acceptable" level, is identified. This recognition is important to bring the mobile phone-based traffic state estimation systems into realization. In addition, two novel "velocity-density inference" models, namely the "adaptive" and the "adaptive **feedback**" velocity-density inference circuits, are proposed to improve the effectiveness of the traffic state estimation. Furthermore, an artificial neural network-based prediction approach is introduced to a the effectiveness of the velocity and the density estimation when the penetration rate degrades to *0%*. These improvements are practically meaningful since they help to guarantee a high accurate traffic state estimation, even in cases of very low penetration rate. The experimental evaluations reveal the effectiveness as well as the robustness of the proposed solutions.

Keywords: mobile probes, low penetration rate, inference circuit, traffic state estimation, intelligent transportation system

1. Introduction

Traffic state estimation is one of the most important fields in Intelligent Transportation Systems (ITS) research [1], [2]. The essential requirement for a reliable traffic state estimation system is that it must provide not only accurate but also real-time traffic state information at any place (ubiquity). The existing traffic state estimation systems majorly employed the conventional data collection methods which relied on the road-side fixed sensors such as loop detectors [3], [4], RFID readers [5], [6], video cameras [7], and so forth. These traditional techniques disclosed their essential weakness in coverage limitation since it is impractical to install a huge number of road-side fixed sensors at every street.

Recently, with the advances of mobile phone technologies, mobile devices such as mobile phones, PDA, etc., have been utilized as traffic probes for collecting real-time traffic data [8], [9], [10], [11], [12]. This approach provides a potential alternative in collecting the real-time traffic data. Since mobile phones are available everywhere and the mobile phone network has already been established, the essential issues such as the coverage limitation, the real-time effect, the investment and maintenance cost can be overcome. Consequently, research in applying mobile phone technologies in ITS is entering a new stage. Several practical issues such as the accuracy of traffic state estimation models [13], the effectiveness of data communication methods [14], the side-effect of low penetration rate on the performance of the whole estimation model [15], [16], and so on, are being investigated to accelerate the realization of the mobile phone-based traffic state estimation systems.

In traffic state estimation using mobile phones as traffic probes, the essential issue is the one that relates to the operational feasibility rather than the technical issues. Several researches revealed that it is technically feasible to establish a mobile phone-based traffic state estimation system in the real world. However, beside the technical issues, the effectiveness of the traffic state estimation is significantly affected by the volume of the sampled data. The more data collected, the higher accuracy the estimation can be achieved, and vice versa. In turn, the volume of the sampled data is affected by the penetration rate of the mobile phones that participate in the traffic state estimation system. In addition, it is difficult to ensure that the penetration rate is always high enough, especially when the system has just been launched. To our best knowledge, there is no relevant research which thoroughly discusses the relationship between the penetration rate and the accuracy of the mobile phone-based traffic state estimation. Consequently, there is no research on assuring an "acceptable" accuracy of the traffic state estimation in cases of low penetration rate. The existing researches were primarily based on the assumption that the penetration rate is always relevant [13]. This article aims at overcoming these issues with the following contributions:

- Thoroughly analyzes the influence of the penetration rate on the traffic state estimation's accuracy. Based on this analysis, the so-called "acceptable" penetration rate will be identified. This recognition helps to set a pre-condition in order to ensure an "acceptable" accuracy of the traffic state estimation. The "accept-

1 Shibaura Institute of Technology, Department of Communications Engineering, College of Engineering, Koto, Tokyo, 135–8548, Japan
a) m709504@shibaura-it.ac.jp
b) kamioka@shibaura-it.ac.jp

able" penetration rate also serves as a trigger to notify the commuters on the confidence of the estimate in cases of low penetration rate.

- Introduces two novel "velocity-density inference" circuits based on the "adaptive" and the "adaptive **feedback**" approaches to improve the traffic state estimation's effectiveness. These models can also minimize the "acceptable" penetration rate. This contribution is practically meaningful since it helps to ensure the effectiveness of the estimation model even with a small portion of mobile phones participating in the system.

- Proposes a reasonable data mining-based approach to ensure the accuracy of the mobile phone-based traffic state estimation model when the penetration rate degrades to zero. This contribution is very important since it helps to ensure an "acceptable" accuracy in the traffic state estimation when the penetration rate becomes unacceptably low such as several percent or even 0%.

This paper is organized as follows: Section 2 reviews the related works and describes the preliminary definitions. The relationship between the penetration rate and the effectiveness of the velocity and density estimation is thoroughly discussed in Section 3. Section 4 proposes the "adaptive" and "adaptive **feedback**" "velocity-density inference" circuits to improve the effectiveness of the traffic state estimation model in cases of low penetration rate. A suitable data mining approach to ensure the estimation's accuracy when the penetration rate becomes unacceptably low is described in Section 5. The effectiveness of the proposed solutions is deeply analyzed in detail in Section 6 using numerous evaluations on the simulation data. Section 7 concludes this work and draws out future research directions.

2. Related Work and Preliminary Definitions

Existing researches on traffic state estimation using mobile phones as traffic probes focus on two major directions as follows: 1) *real-time traffic data collection using mobile devices,* and 2) *effective traffic state estimation using data collected by the mobile devices*.

The first direction focuses on proposing effective techniques on collecting **relevant** traffic data with the lowest cost. Studies in Ref. [17] proposed a traffic data collection method based on virtual lines (VTL) which are predefined on the road network. The traffic data is collected only when the transponder passes a VTL. This approach decreases the amount of data collected, thus reducing the data transmission load. However, this approach may lead to the loss of useful data. The VTL setting (i.e., where on the road a VTL should be created) has been a matter of argument since it may significantly affect the effectiveness of the whole system. The studies in Ref. [14] introduced a more flexible solution, namely the "pinpoint" approach, by which traffic data is collected only when there is change in the traffic state, which is detected by the vehicle's velocity change rate. More concretely, a vehicle (or a mobile phone carried on the vehicle) collects and reports the traffic data to the server when only its velocity changes (increases/decreases) significantly. In addition, the "pinpoint" approach also successfully differentiated mobile phones carried by walkers from the ones on vehicles, order to prevent walkers' mobile phones from reporting the data since their data are not relevant to traffic state estimation. The background philosophy in the "pinpoint" approach is that only the **relevant** data are collected as the **right** time (i.e., when the velocity changes significantly) by the **right** mobile devices (i.e., only the devices carried by vehicles). Therefore, not only the data transmission load but also the data redundancy was reduced significantly.

The second direction focuses on the data processing and information representing models by which traffic states are not only accurately estimated but also comprehensibly presented to the commuters. Conventional researches primarily relied on the travel time to identify the traffic state [10]. The travel time was commonly calculated from the average velocity of the traffic flow which is estimated based on the data reported by some representative vehicles (the GPS-equipped ones). This approach, however, might reveal some estimation biases since the density of a traffic flow was not considered. Meanwhile, density itself is one of the important factors affecting traffic state level [18]. Numerous researchers proposed to integrate both the density and the average velocity to improve the effectiveness of the estimation models [13], [19]. The density has been proposed to be inferred from the estimated average velocity using the velocity-density inference models such as the Greenshields [20], the Greenberg [21], or the Underwood [22] models. These approaches, however, might reveal systematic errors drawn from any error in the average velocity estimation. Such a systematic error can be avoided by estimating both the average velocity and the density independently and directly using the traffic data collected by mobile phones as proposed in Ref. [13]. Nevertheless, the essential issue in this approach is its sensitivity to the penetration rate of the mobile phones which report traffic data to the server. To our best knowledge, no prominent approach to provide accurate average velocity and density estimation considering the penetration rate has been proposed.

Closely related to this work, the study in Ref. [16] proposed to apply a data mining technique on historical data to ensure error-tolerance on traffic state estimation in cases of low penetration rate. However, details on the data mining approach and the effectiveness of the proposed models were not thoroughly discussed. In this paper, the influence of the penetration rate on the effectiveness of the traffic state estimation will be carefully analyzed. In addition, two novel "velocity-density inference" circuits and proposed to apply a data mining technique on historical data to ensure error-tolerancean artificial neural network-based prediction model are proposed to ensure the accuracy of the traffic state estimation in cases of low penetration rate. On top of these proposals, the preliminary definitions on penetration rate and other related terms will be presented in the remainder of this section.

Definition 1: The *penetration rate* of vehicles on a road segment, denoted as ρ, is the fraction of vehicles that collect and report data to the server per the total number of vehicles travel on the considered road segment at the current time.

Definition 2: The *average velocity* of a traffic flow, denoted as V_{avg}, is the average value of the velocities of all vehicles travel in the traffic flow at the considered time.

The average velocity defined above can be mathematically expressed as in Eq. (1), where V_i is the velocity of the vehicle i, and

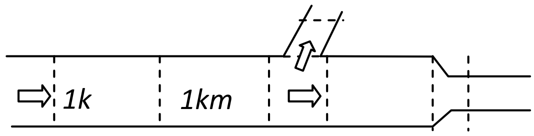

Fig. 1 Road segmentation.

n is the total number of vehicles travelling on the considered road segment.

$$V_{avg} = \frac{\sum_{i=1}^{n} V_i}{n} \quad (1)$$

Since the limitation speed varies from road segment to road segment, the average velocity may not correctly represent the traffic condition in terms of travel time of a road segment. The study in Ref. [13] has proposed a new term, namely the *mean speed capacity*, to better present the travel time of a traffic flow at a specific road segment. This term can be defined as follows:

Definition 3: The *mean speed capacity* of a road segment, denoted as M_v, is defined as the average velocity divided by the limitation speed of the considered road segment.

The mean speed capacity can be expressed as in Eq. (2), where V_{avg} is the average velocity defined in definition *2* and $V_{\max}$ is the limitation speed of the considered road segment.

$$M_v = \frac{V_{avg}}{V_{\max}} \quad (2)$$

Definition 4: The *density* of a traffic flow on a road segment, denoted as D, is the fraction of the number of participating vehicles per the capacity of the considered road segment.

The density can be expressed as in Eq. (3), where n is the number of vehicles on the traffic flow and C is the capacity (the maximum number of vehicles) of the considered road segment.

$$D = \frac{n}{C} \quad (3)$$

The terms defined above are associated with road segments. A road segment is a part of a road by which road-land marks such as the intersection, the crosswalk, the curved place, and so forth, are the end-points of road segments [13]. On the highways where the above characteristics do not change frequently, road segments are divided into each kilometer. It should be noted that the road segments are divided based on the road direction instead of the number of lanes. Therefore, a stretch of *2-way* road consists of two different segments regardless of the number of lanes in each direction. **Figure 1** shows a road segmentation by which road segments are created at the places where the characteristics of a road change.

3. The Influence of the Penetration Rate on the Estimation Accuracy

This section analyzes the relationship between the penetration rate and the effectiveness of the traffic state estimation. As an extension of the studies in Refs. [13] and [19], the traffic state of a considered road segment is estimated as a function of the *mean speed capacity* (M_v) and the *density* (D) as described in Eq. (4). In this equation, M_{v0} and D_0 are the thresholds of mean speed capacity and density, respectively, by which the traffic state of the road segment is considered as good enough. These thresholds can be identified by the transportation experts or via numerous evaluations on simulation data. For example, in Ref. [13] both the thresholds were defined as 0.6. Regardless of what values are set for the thresholds, the accuracy of the traffic state estimation is affected by the effectiveness of the mean speed capacity and the density estimation. Obviously, it is relevant to analyze the relationship between the penetration rate and the effectiveness of the mean speed capacity and the density estimation. As defined in Eq. (2), the accuracy of the mean speed capacity estimation is drawn directly from the accuracy of the average velocity estimation. Therefore, the relationship between the penetration rate and the mean speed capacity estimation degrades to the relationship between the penetration rate and the average velocity estimation.

$$s = f(M_v, D) = M_v - M_{v0} + D_0 - D \quad (4)$$

In the mobile phone-based traffic state estimation, mobile devices carried by vehicles report the traffic data which includes the position (longitude, latitude), the heading, the velocity, etc., of the vehicles to the server. The average velocity of a traffic flow on the considered road segment is estimated based on the aforementioned reported data. If all the vehicles travel on such a traffic flow report the data, the estimated average velocity, denoted as V_{avgest}, and the "actual" average velocity (V_{avg}) must be identical. However, this condition is inappropriate in real world applications for a variety of reasons. For example, the mobile phones have not yet registered to the estimation system, especially when the system has just been launched; the mobile phones are busy serving their conventional functions such as calling, sending message; or even the users do not want any data to be reported from their mobile phones, and so forth. The error on the estimated average velocity (V_{avgest}) occurs in cases of low penetration rate is analyzed as follows.

With a penetration rate ρ (please refer to the definition 1), the *average velocity* is estimated as in Eq. (5), where V_i is the velocity of the vehicle i, and n is the actual number of vehicles travelling on the considered road segment. The average velocity estimation error can be expressed as in Eq. (6), where V_{avg} and V_{avgest} are the "actual" and the estimated average velocities defined in Eqs. (1) and (5), respectively.

$$V_{avgest} = \frac{\sum_{i=1}^{n\rho} V_i}{n\rho} \quad (5)$$

$$E_v = |1 - \frac{V_{avgest}}{V_{avg}}| = |1 - \frac{1}{\rho}.\frac{\sum_{i=1}^{n\rho} V_i}{\sum_{i=1}^{n} V_i}| \quad (6)$$

According to the Eq. (6), the average velocity estimation error is affected by two factors as follows: 1) the penetration rate ρ, and 2) the distribution of the collected velocities (the set of $\rho * n$ velocities) in the set of the actual velocities (the set of n velocities). The penetration rate has been defined as in definition 1. However, the factor in 2) is a complicated phenomenon. If most of the collected velocities (the ones on the numerator of the Eq. (6)) are on the centroid of the distribution of the "actual" velocities (the ones on the denominator of the Eq. (6)), then the error is minor. In contrast, if most of the collected velocities are the outliers in the "actual" velocities' distribution, then the error will drastically increase.

As analyzed previously, the lower penetration rate commonly reveals a lower accuracy in average velocity estimation. However, real-field studies also revealed that the influence of the penetration rate on the average velocity estimation varies with differing traffic situations. More concretely, in the heavy traffic flow, even the penetration rate is low by which a large amount of actual data is missed, there are many vehicles travel in the considered road segment. Therefore, the movements of vehicles in such a traffic flow are highly correlated with each other. This correlation is called the "inter-correlation" by which no individual vehicle can travel significantly faster than the others, neither faster than the actual average velocity of the traffic flow. As a result, the difference between the actual average velocity and the estimated one is still small enough. In contrast, the velocity fluctuation in a light traffic flow is usually large since several vehicles may travel faster or slower than the others (i.e., the outlier). Consequently, the error of the estimated average velocity will drastically increase if more actual data is missed (i.e., low penetration rate). As a brief conclusion, the penetration rate directly affects the effectiveness of velocity estimation with regard to the real traffic state. This effect seems to be minor in heavy traffic situations but it becomes more serious in light traffic states. This statement will be confirmed by the evaluations on Section 6 – the evaluation section.

Similar to the average velocity, the density estimation is also affected by the penetration rate. As defined in definition 4 and Eq. (3), the density estimation error, denoted as (E_D), is directly affected by the penetration rate ρ, as expressed in Eq. (7). In the real world application, especially when the system has just been launched, the penetration rate is usually very low, namely 5%, 10% or less. Therefore, the density estimation error will be drastically large. For example, if the penetration rate is 10% then the density estimation error will be 90%, which cannot be accepted in any estimation model.

$$E_D = 1 - \rho \tag{7}$$

As mentioned in Section 2 (the related work section), one of the possible ways to improve the effectiveness of the density estimation is to apply the velocity-density inference models with a given estimated average velocity. The Greenshields-like velocity-density inference models [20], [21], [22], however, cannot address this issue thoroughly since any error in the estimated average velocity will propagate to the inferred density. In addition, as analyzed above, the lower penetration rate draws a higher error in average velocity estimation. Therefore, the conventional Greenshields-like models cannot effectively work in low penetration rate situations. This paper proposes two novel velocity-density inference models, namely the "adaptive" and the "adaptive **feedback**" velocity-density inferences circuits and a data mining-based approach to solve the aforementioned issues thoroughly. These proposed solutions will be discussed in the following sections.

4. Velocity-Density Inference Circuits

4.1 The Conventional Velocity-Density Inference Models

The Greenshields velocity-density inference model is commonly used to infer the density when it cannot be directly estimated using real traffic data. The Greenshields model is described as in Eq. (8), where V and D are the current velocity and density of a traffic flow, respectively; $V_{\max}$ and $D_{\max}$ are the speed limitation and the maximum density (the density of the congested state) of the considered road segment, respectively. The drawback of this model is that it assumes a linear relation between the density and the velocity. However, the linear model is not appropriately describes the relationship between the velocity and the density in cases of high and low density.

$$V = V_{\max}\left(1 - \frac{D}{D_{\max}}\right) \tag{8}$$

The Greenberg [21] and the Underwood [22] models were proposed to better describe the relationship between the velocity and the density in cases of high and low densities, respectively. These two models are expressed as in Eqs. (9) and (10), where V, D, $V_{\max}$ and $D_{\max}$ are the same parameters as those in Eq. (8). In addition, V_0 and D_0 are the velocity and the density at the time by which the volume of the traffic flow hits the maximum value. The volume of a traffic flow at a current time is calculated as in Eq. (11).

$$V = V_0 \log\left(\frac{D_{\max}}{D}\right) \tag{9}$$

$$V = V_{\max} e^{-\frac{D}{D_0}} \tag{10}$$

$$Q = V.D \tag{11}$$

As mentioned, the Greenberg and the Underwood models better express the velocity-density relationship in special traffic flows such as in dense and sparse density, respectively. However, different to the $V_{\max}$ and $D_{\max}$ in the Greenshields model, V_0 and D_0 are not known in advance. Furthermore, there is no effective way to estimate these values while the density is still a variable. Consequently, neither the Greenberg nor the Underwood model can be suitably applied to the mobile phone-based traffic state estimation.

Turning back to the Greenshields inference model, the density can be inferred from a given velocity, and vice versa. However, as analyzed in the previous section, the density is directly and hence more seriously affected, compared to that of the velocity, by the penetration rate. More concretely, in cases of low penetration rate, the directed density estimation error is usually drastically high as described in Eq. (7). In contrast, at the same time, the velocity estimation error may be much smaller because of the *"inter-correlation"* between vehicles in a traffic flow. Therefore, in prior works, the velocity was usually estimated in advance and then used to infer the density. However, this approach might reveal the systematic error drawn from the estimated velocity error. This work introduces novel velocity-density inference models, namely the "adaptive" and the "adaptive **feedback**" velocity-density inference circuits, to address the drawback of the Greenshields-like estimation models. Here, both the velocity and the density estimated from the real data are applied as inputs of the inference models to alleviate any distortion from outliers (i.e., the ones move too fast or too slow compared to the others). This model will be presented in the remainder of this section.

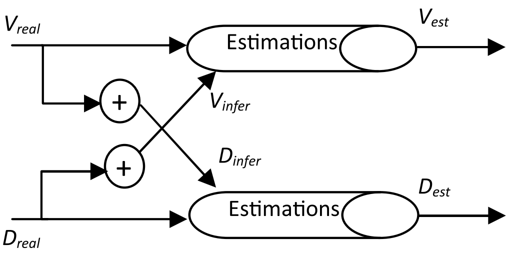

Fig. 2 The adaptive Velocity-Density inference circuit.

4.2 The Adaptive Velocity-Density Inference Circuit

The proposed adaptive "velocity-density inference circuit" is depicted as in **Fig. 2**. In this model, both the velocity and the density estimated/calculated directly from the real data collected by mobile phones, namely V_{real} and D_{real}, respectively, are served as the inputs. The outputs of this circuit are the final estimated values of velocity and density, namely V_{est} and D_{est}, respectively. The intermediate velocity (V_{infer}) and the density (D_{infer}) also contribute to the effectiveness of the inference circuit. The background philosophy of this inference model is that both the velocity calculated directly from real data and inferred from estimated density must contribute to the final velocity estimation. More concretely, the final estimated velocity, V_{est}, is a function of the directly calculated and the Greenshield-based inferred velocities as expressed in Eq. (12). The same rationale is also applied to the density estimation as in Eq. (13). The intermediate velocity and density are inferred using the Greenshields model as described in Eqs. (14) and (15).

$$V_{est} = f(V_{real}, V_{infer}) = \alpha V_{real} + (1 - \alpha)V_{infer} \quad (12)$$

$$D_{est} = f(D_{real}, D_{infer}) = (1 - \alpha)D_{real} + \alpha D_{infer} \quad (13)$$

$$V_{infer} = V_{\max}\left(1 - \frac{D_{real}}{D_{\max}}\right) \quad (14)$$

$$D_{infer} = D_{\max}\left(1 - \frac{V_{real}}{V_{\max}}\right) \quad (15)$$

In Eq. (12), α ($0 \leq \alpha \leq 1$) is the coefficient represents the *impact* of velocity calculated directly from the real data, V_{real}, on the whole average velocity estimation model. Therefore, the impact of the velocity inferred from the estimated density, V_{infer}, must be $1 - \alpha$. Since D_{infer} in Eq. (13), is inferred from V_{real} as in Eq. (15), its impact on the whole density estimation model should be α, thus 1-α is the impact of the D_{real}. One may argue that D_{infer} in Eq. (15) may impractically be estimated as a negative value if some (or all) of the vehicles that report the data illegally travel faster than the limitation speed. To solve this side effect, V_{real} should be clarified as which it is the velocity of the traffic flow at the considered road segment. It is defined as in Eq. (16).

$$V_{real} = \frac{\sum_{i=1}^{k} V_i}{k} \quad (16)$$

where, k is the number of vehicles on the considered road segment that report the data to the server, and V_i is the estimated velocity of any individual vehicle i, which is defined as in Eq. (17).

$$V_i = \min(V_{sensed\_i}, V_{\max}) \quad (17)$$

Equation (17) describes that the estimated velocity of an individual vehicle i (V_i) is the minimum value of the limitation speed ($V_{\max}$) and the detected real-time velocity of the vehicle i (V_{sensed_i}). The rationale of this normalization is that $V_{\max}$ describes the best condition of the road segment. At the same time, vehicles can travel with an illegally high speed ($V_{sensed\_i} > V_{\max}$) in only the best traffic conditions. The pre-processing of the estimated velocity described in Eq. (17) can help to avoid any bias caused by illegally speeding vehicles without compromising the effectiveness of the whole traffic state estimation model. In practice, when V_i is selected as $V_{\max}$, that is to say, V_{sensed_i} is larger than $V_{\max}$, the road segment can be regarded as almost in the best condition. In this case, it is reasonable that D_{infer} is calculated as 0 or close to 0. The difficulty here is to reasonably identify α.

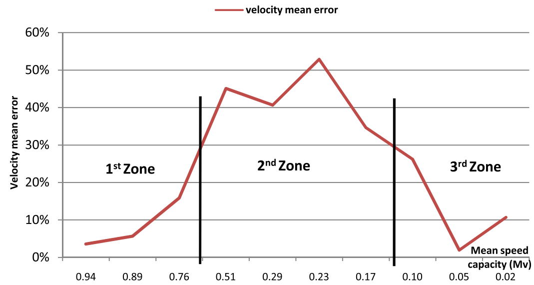

Fig. 3 Relationship between the meanspeed capacity and the velocity estimation error (both are estimated from real data).

Obviously, the impact of the velocity calculated from the real data, referred as V_{real} shortly, on the whole velocity estimation model is affected by the penetration rate. Intuitively, the lower penetration rate the higher error of the V_{real}, thus the lower impact the V_{real} should be. Concretely, the impact of V_{real}, referred as α shortly, can be expressed as a function of the penetration rate. However, the penetration rate cannot be exactly known in advance when the "actual" density of the traffic flow is still a variable (please refer to the penetration rate's definition in definition 1, Section 2). Therefore, another way to express this impact should be investigated. One suitable method is using the statistical studies on historical experimental or simulation data. The statistical study must be based on features which can be estimated at the estimation time. Fortunately, besides the penetration rate, the mean speed capacity estimated from the real data collected by mobile phones also reflects on the accuracy of the V_{real}. As defined in definition 3 and Eq. (2), Section 2, the mean speed capacity can be completely estimated at any time using real data. Therefore, the impact of V_{real}, α, can be expressed as in Eq. (18). The rightmost expression on the Eq. (18) states that the accuracy of V_{real} can be retrieved from historical data given a value of M_{vreal} (the mean speed capacity obtained from the real data collected by mobile phones).

$$\alpha = Accuracy(V_{real}) = 1 - MeanErrFrom(M_{vreal}) \quad (18)$$

Investigations on the historical experimental data and simulation data revealed that the *mean speed capacity* and the *velocity estimation errors* (both estimated from real data collected by mobile phones) are closely related with each other. As shown in **Fig. 3**, with a given *mean speed capacity*, the mean error of the estimated velocity can be retrieved, thus the impact of V_{real}, α, can be inferred using Eq. (18).

In practice, in order to reduce the retrieval time, the mean error of the estimated velocity coming from the estimated mean speed capacity can be approximated. As shown in Fig. 3, the error distribution can be divided into 3 zones. The first zone represents the mean speed capacity which is greater 0.6, the second zone represents the mean speed capacity from 0.6 to 0.15 and the third zone for the remainder. The velocity estimation error of each traffic state belong to a zone is almost similar to the mean error of that zone. The mean errors of all the three zones are extracted from historical data and presented as in Eq. (19).

$$MeanErrFrom(M_{vreal}) = \begin{cases} 16.8\%(\approx 0.2) & M_{vreal} \geq 0.6 \\ 43.1\%(\approx 0.4) & \text{other} \\ 17.2\%(\approx 0.2) & M_{vreal} \leq 0.15 \end{cases} \tag{19}$$

Applying the Eq. (19) into Eq. (18), the impact coefficient α can be simply approximated as in Eq. (20).

$$\alpha = \begin{cases} 1 - 0.2 = 0.8 & M_{vreal} \geq 0.6 \\ 1 - 0.4 = 0.6 & \text{other} \\ 1 - 0.2 = 0.8 & M_{vreal} \leq 0.15 \end{cases} \tag{20}$$

After obtaining the coefficient α, the final velocity and density can be easily estimated using Eqs. (12) and (13) mentioned before. The effectiveness of this "adaptive velocity-density inference" circuit will be thoroughly discussed in Section 6 – the evaluation section.

4.3 The Adaptive Feedback Velocity-Density Inference Circuit

Obviously, traffic state is a phenomenon which does not change drastically in a short time interval, but the change must have some relationship with previous traffic states at the same road segment. More concretely, the velocity and density at time t must relate with those at time $t-1$, $t-2$, and so forth. Therefore, the estimated velocities/densities at some previous phases in the past are useful for estimating velocity/density at the current time. This section introduces the so-called "adaptive **feedback** velocity-density inference" circuit to benefit the estimated values in the previous phases for estimating the current velocity/density as depicted in **Fig. 4** and described as follows.

Figure 4 shows that, the "adaptive **feedback** velocity-density inference" circuit is a modification of the "adaptive velocity-density inference" circuit proposed in the previous section. The velocity and density, which are estimated in every time interval t, namely every minute as in this work, are stored and re-used for the future estimations. The moving averages of the estimated velocity and the density at time t, namely MV_t and MD_t, are calculated within a sliding window ξ, and fed back to the estimation

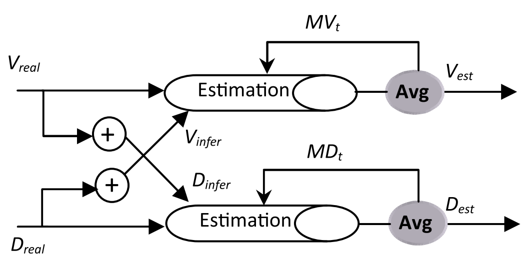

Fig. 4 The adaptive **Feedback** Velocity-Density inference circuit.

components in the circuit for the estimation at time $t + 1$. The moving average of the estimated velocity and density are calculated as in Eqs. (21), (22), respectively. The sliding window can be set by the domain experts and was set to $3(\xi = 3)$ in this work.

$$MV_t = \frac{\sum_{i=t}^{t-\xi} V_{est_i}}{\xi} \tag{21}$$

$$MD_t = \frac{\sum_{i=t}^{t-\xi} D_{est_i}}{\xi} \tag{22}$$

Since MV_t is the moving average of the estimated velocity in ξ previous calculation phases, it must have the same impact of the velocity estimated directly from the real data, V_{real}. Thus the estimation model for estimating the velocity at the current time, which was expressed in Eq. (12), is modified as in Eq. (23). In this equation, the $Avg(V_{real}, MV_t)$ is the average function of the two corresponding parameters. Applying MD_t at the same way, the density estimation model described in Eq. (13) is modified as in Eq. (24). It should be noted that the impact of the MD_t should be the same as that of the inferred density, D_{infer}, which is inferred from V_{real}.

$$\begin{aligned} V_{est} &= f((V_{real}, MV_t), V_{infer}) \\ &= \alpha Avg(V_{real}, MV_t) + (1-\alpha)V_{infer} \end{aligned} \tag{23}$$

$$\begin{aligned} D_{est} &= f(D_{real}, (D_{infer}, MD_t)) \\ &= (1-\alpha)D_{real} + \alpha Avg(D_{infer}, MD_t) \end{aligned} \tag{24}$$

As mentioned above, it is expected that the "adaptive **feedback** velocity-density inference" circuit is more effective than the "adaptive" counterpart introduced in Section 4.2. The effectiveness evaluation of these both "velocity-density inference" circuits will be presented in Section 6.

5. A Data Mining Approach to Error-Tolerance in Velocity/Density Estimation

In addition to improving the traffic state estimation's effectiveness, the proposed "velocity-density inference" circuits can minimize the "acceptable" penetration rate, thus they are adaptively workable in more strict conditions in terms of penetration rate. However, when the penetration rate degrades significantly to a very low value such as several percents or even to zero, they cannot work properly. To address the low penetration rate related issues thoroughly, a data mining technique applying the historical data of traffic states should be investigated.

Field studies reveal that the traffic state not only represents the movement phenomenon of vehicles on the considered road segment but also represents a complicated relation with the movement on the nearby road segments. For example, if the traffic state of the road segment in the front is heavy then the traffic state of the considered road segment is also heavy. Therefore, the traffic state of a road segment can be predicted if the current traffic states of the nearby road segments and the "rules" which represent the relationship between traffic states at different road segments in a region are given. The relation "rules" can be "learned" by a machine-learning technique from the historical traffic state data. In this work, an artificial neural network (ANN) with a multilayer perceptron (MLP) [23] is employed to learn such "rules" in order

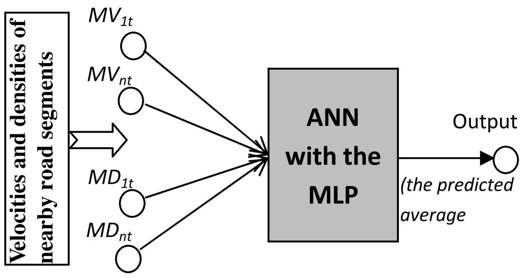

Fig. 5 An ANN-based velocity/density prediction model applied in cases of very low penetration rates.

to effectively predict the average velocity and density of a road segment in cases of very low penetration rate. This ANN-based prediction model is depicted as in **Fig. 5**.

Figure 5 shows that the historical data of the moving averages of the velocities and the densities of all the road segments in a region are fed to the ANN for "learning" the prediction "rules." Assuming that there are n road segments in a region, hence n moving average velocities and n moving average densities are put into the model at the same time, namely t. In Fig. 5, MV_{1t}, MV_{2t}, MD_{1t}, MD_{2t} are the moving averages of the velocity and density at time t, which were defined in Eqs. (21) and (22), Section 4.3, respectively. The experimental evaluations revealed that the proposed ANN-based prediction model can effectively predict the density and average velocity regardless any penetration rate of the considered segment. The effectiveness of this ANN-based prediction model will be analyzed in the next section.

6. Evaluation

This section evaluates the effectiveness of the proposed "velocity-density inference" circuits and the ANN-based prediction model. In addition, the influence of the penetration rate on the velocity/density estimation accuracy will be thoroughly evaluated.

6.1 The Experiment Evironment and the Data Structures

Several simulations were performed using the Traffic Simulation Framework (TSF) [24] to generate synthetic data for evaluations. In each simulation, the desired road segments and the desired regions (i.e., an area with several segments) on which the traffic state related data would be generated were selected as shown in **Fig. 6**. For each selected road segment, two kinds of data were created concurrently as follows:

a) The GPS-based detail data reported by individual vehicles was recorded in the "*cars.csv*" file. Each record in this file contains the *Time stamp* (in seconds) when the data is reported, the *road segment Id*, the *position* (*longitude, latitude*), the *current velocity*, and the *vehicle Id* of the vehicle which reports the data. The penetration rates were also configured by the TSF by which only the "penetration rate" percents of random vehicles, namely 20%, 25%, 30% vehicles, and so forth, reported the data to the server. The frequency of the data report timing was set to every 3s (i.e., similar to the GPS (Global Positioning System) signal frequency).

b) The summarized data on the traffic state of the selected road segments in a predefined time interval were recorded in the "*av-*

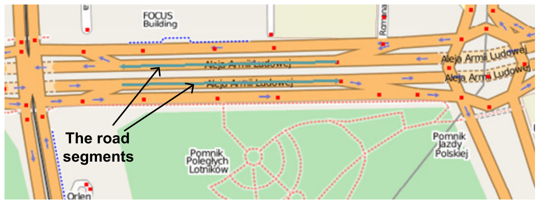

Fig. 6 The road segmentation in TSF.

erageVelocity.csv" file. Each record in this file contains the information of the *Time interval Id* (in minutes), the *road segment Id*, the *average velocity*, and the *density* of the traffic flow on the selected road segment in such a time interval. The time interval for recoding the summarized traffic state information was set to every minute in this work. This summarized traffic state information was used to evaluate the accuracy of the estimation method applying the GPS-based detailed data described in *a*).

It should be noted that the density of the vehicles in all the simulation was also configured by different levels. These settings helped to avoid any bias in evaluations which comes from an arbitrary density. The details of the effectiveness evaluations will be presented in the remainder of this section.

6.2 The Influence of the Penetration Rate on the Estimation Accuracy

As discussed in Section 3, the penetration rate significantly affects the effectiveness of the velocity and density estimation. This section evaluates this relationship using the simulation data. As a result, the so-called "acceptable" penetration rate, where the estimation accuracy is good enough and satisfies the designer, was recognized. The acceptable penetration rate is important since it ensures that the estimated traffic state information is not wrong so much and still be reliable.

To analyze the relationship between the penetration rate and the estimation's effectiveness, several simulations were performed with different penetration rates. For each penetration rate, namely 20%, 30% for instance, the GPS-based detailed data and the summarized data of the selected road segments were recorded in the "*car.csv*" and the "*averageVelocity.csv*" files, respectively. The average velocity and average density were estimated from the GPS-based detailed data in the "*car.csv*" file, applying the conventional estimation model as described in Eqs. (1) and (3), Section 2. Both the average velocity and the density were estimated in each time interval, namely every minute in this work, by which they could be compared with the "actual" average velocity and density generated by the TSF simulator (i.e., in the "*averageVelocity.csv*" file). This comparison was repeated with different simulation data generated from 10 randomly selected road segments. For each road segment, five one-hour simulations with different density levels were performed. After then, the average differences between the estimated and the "actual" average velocity were drawn out and shown in **Fig. 7**.

Figure 7 shows that the higher the penetration rate the better the estimation is, and vice versa. If the penetration rate degrades to 20% or lower, the estimation error increases significantly. At the same time, the deviations in cases of low penetration rate are also higher. For example, when the penetration rate is 20%, the

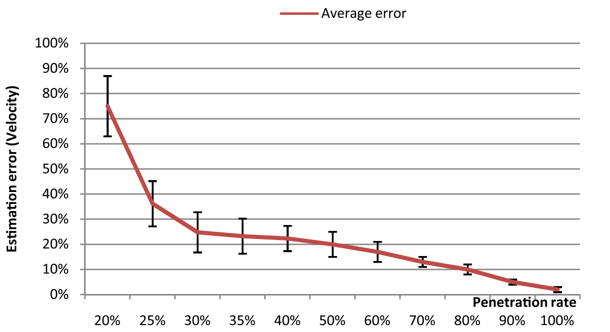

Fig. 7 The relationship between the penetration rate and the everage error on the estimated average velocity.

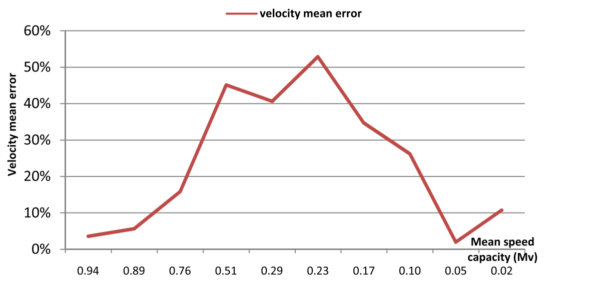

Fig. 8 Relationship between the meanspeed capacity and the mean error of average velocity estimation.

deviation is around 24% (i.e., ±12%). This figure also reveals that if the expected accuracy is set to 70%, the acceptable penetration rate must be around 30%. This recognition of the "acceptable" penetration rate helps researchers and practitioners to prepare the prerequisite on the penetration rate (i.e., the portion of mobile phone users joining the system) for an acceptable estimation accuracy. Otherwise, alternative solutions to avoid the wrong estimation before bring the mobile phone-based traffic state into the realization should be considered.

As mentioned above, the expected accuracy is set to 70%. There are two reasons for that. First, the accuracy of 70% is a quite critical value. Second, the value is useful to compare this research with related research. More concretely, as shown in Fig. 7, the penetration rate must be around 25% when the accuracy is set to 60% (inappropriate accuracy), and the penetration rate must be around 50% when the accuracy is set to 80% (accurate enough). It means that the ratio of accuracy to the penetration rate drastically changes at the point of accuracy of 70%. Generally speaking, it is almost impossible to keep the penetration rate of 50% even though the accuracy of 80% is high enough. However, it is more confident that the penetration rate of 27% which corresponds to the accuracy of 70% could be realized in a real situation and this accuracy is acceptable. In addition, as a related research, Muraki et al. [25] illustrated the traffic prediction accuracy based on travel time estimations using the Nearest Neighbor Method. They evaluated the proposed method in terms of traffic prediction error. The error ranges are from 33% to 38% for Metropolitan Expressway and from 19% to 23% for ordinary roads. The average error for both road types is almost 30% which shows the prediction accuracy is 70%. Therefore, the expected accuracy was set to 70% in this work as a criterion for our evaluation.

In addition to evaluating the relationship between the penetration rate and the estimation's effectiveness, the influence of the "actual" traffic states on the estimation accuracy in cases of low penetration rate was also evaluated. In this evaluation, the "actual" traffic state of a road segment was represented by the *mean speed capacity* (M_v) as defined in Eq. (2), Section 2. Therefore, the relationship between the *mean speed capacity* and the mean error on the average velocity estimation was analyzed. The same dataset used for estimating the relationship between the penetration rate and the estimation's effectiveness was re-used in this evaluation. Here, the detailed information about the error on the average velocity estimation was recorded at different *mean speed capacity*. The mean error of the average velocity estimation with regard to the *mean speed capacity* is shown in **Fig. 8**. This figure reveals that, the mean error of the average velocity estimation is small in cases of very light (i.e., M_v is very high), and very heavy (i.e., M_v is very low) traffic flows. However, the mean error of the average velocity estimation is higher in cases of medium traffic state. The reason of this phenomenon comes from the "*inter-correlation*" of the movements of vehicles in the corresponding traffic flows. In the cases of very heavy traffic state (i.e., M_v is very small), no individual vehicles can travel faster than the others, thus even though the penetration rate is low, the estimated average velocity is almost the same as the "actual" average velocity. However, in the cases of medium traffic states, there may exist several outliers (i.e., the vehicles that move faster or slower than the others), thus the deviation of the velocity is high revealing a high mean error on average velocity estimation. The deviation on velocity at the light traffic states (i.e., M_v is very high) may also high. Nevertheless, the number of vehicles travelling in such a traffic state is usually small, thus the probability for the outliers to be occurred and collected is small. As a result, the mean error on average velocity estimation is still small as shown Fig. 8. This figure also confirms Fig. 3 and the rationale on identifying the coefficients on the proposed "adaptive velocity-density inference" circuit as discussed in Section 4.2.

It should be noted that the penetration rate directly affects the accuracy of the density estimation as expressed in Eq. (7), Section 3. Therefore, it is clear that when the penetration rate degrades, the error on density estimation drastically increases. For example, at 40% of penetration rate, the density estimation error will be 60% (please refer to Eq. (7)), and this error drastically increases to 80% if the penetration rate degrades to 20%. The effectiveness of the proposed "velocity-density inference" cricuits will be analyzed in the remainder of this section.

6.3 Effectiveness of the Velocity-density Inference Circuits

The same datasets used to evaluate the influence of the penetration rate on the effectiveness of the average velocity estimation mentioned in the previous section were reused. In this evaluation, the effectiveness of the proposed "velocity-density inference" circuits was compared with that of the conventional model.

Figure 9 shows the effectiveness of the "adaptive" and the "adaptive **feedback**" velocity-density inference circuits in estimating the average velocity. The effectiveness of the "adaptive" and the "adaptive **feedback**" circuits are represented as *Circuit_V* and *Circuit1_V*, respectively. Both of these models are clearly

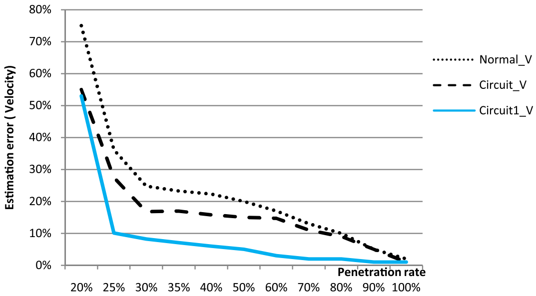

Fig. 9 The effectiveness of the velocity-density inference circuits in estimating the average velocity.

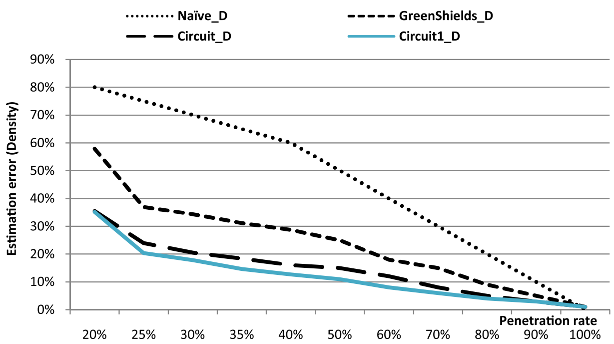

Fig. 10 The effectiveness of the velocity-density inference circuits in estimating the density.

better than the conventional estimation model (*Normal_V*). The effectiveness of the "adaptive velocity-density inference" circuit (without the feedback information from the previous estimation phases) degrades somewhat when the penetration rate increases. For example when the penetration rate is greater than 70%, its effectiveness is not prominent. The reason is that with a relevant penetration rate (e.g., 70%) the conventional estimation method may obtain enough data for a high accurate estimation. In contrast, the "adaptive **feedback**" velocity-density inference circuit shows its power to be far better than both the conventional and the previous adaptive inference counterpart. In addition, this inference circuit can minimize the "acceptable" penetration rate. As mentioned in Section 6.2, the acceptable penetration rate (i.e., the penetration rate to guarantees 70% of accuracy) in the conventional average velocity estimation model is 30%. However, the "adaptive **feedback**" velocity-density inference circuit reduces this value to around 25%. Furthermore, in all cases of penetration rates which are greater than this acceptable value, the estimation errors are always smaller than 10%. This effectiveness is great and far satisfies the requirement in this research.

Figure 10 shows the effectiveness of the "adaptive" and the "adaptive **feedback**" velocity-density inference circuits, represented by *Circuit_D* and *Circuit1_D*, respectively, in estimating the average density. The figure shows that these models are far better than the conventional estimation model (*naive_D*), and the Greenshields (*Greenshields_D*) counterparts. However, difference from the average velocity estimation case, the "adaptive **feedback**" velocity-density inference circuit is not far better than its "adaptive" counterpart. This result also reflects the issue discussed in section 3 that *"the density is directly and hence more seriously affected, compared to that of the velocity, by the penetration rate."* Therefore, the "adaptive" and the "adaptive **feedback**" inference circuits could not improve the effectiveness of the density estimation as good as that in case of average velocity estimation as shown in Fig. 9.

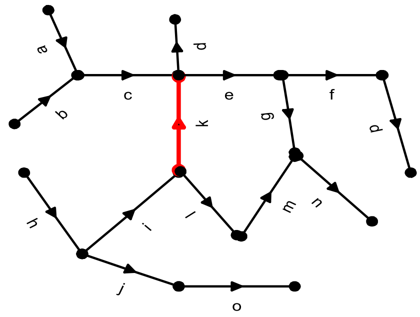

Fig. 11 A region of road segments including the disired segment and its related ones.

It should be noted that, although the proposed "velocity-density inference" circuits improve the effectiveness of the velocity and density estimations significantly, they could not ensure the estimation accuracy when the penetration rate becomes unacceptably low, namely lower than 25% as shown in Fig. 9 and Fig. 10. Meanwhile, in the real world application, the penetration rate may degrade to a very low value such as several percent or even zero. As discussed in Section 5, this issue can be addressed by a suitable ANN-based prediction model. The effectiveness of this model is evaluated in the remainder of this section.

6.4 Effectiveness of the ANN-based Prediction Model

To evaluate the accuracy of the ANN-based prediction model, another dataset was prepared. Five road segments were randomly selected as the desired ones whose average velocity and density are required to be predicted. In addition, the related road segments of the desired ones were also identified so that their corresponding data (average velocity and density) were recorded. The related road segments are those whose traffic states influence the traffic state of the desired road segment. In this work, the related road segments have been divided into two types: the "direct" and "in-direct" connected ones. The "direct" connected road segments are those connect to the desired road segment directly. The "in-direct" connected road segments are the ones that directly connect to the desired road segment's direct connected ones. For legibility, the "direct" connected and the "in-direct" connected road segments are called the *level* 1 (l_1) and *level* 2 (l_2) related ones, respectively, as depicted in **Fig. 11**. Assuming that k is the desired road segment, its l_1 and l_2 related road segments are {*c, d, e, i, l*} and {*a, b, h, j, m, g, f*}, respectively.

For each desired road segment, 10 1-*hour* simulations were performed by which the average velocity and density of the desired road segment and its related counterparts were generated. The summarized traffic state information was recorded in each minute, thus in each simulation, 60 patterns including both the average velocity and density of the desired road segment and its related road segments were generated. Therefore, totally a dataset of 600 patterns (i.e., in 10 1-*hour* simulations) was created for each desired road segment. At each time interval, namely every minute, the moving averages of the velocity and density of each road segment were calculated as described in Eqs. (21) and (22),

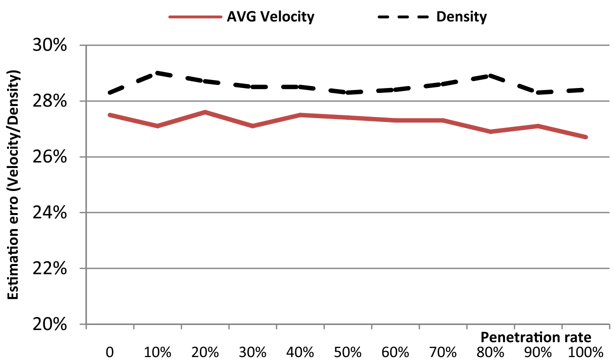

Fig. 12 The effectiveness of the ANN-based prediction model.

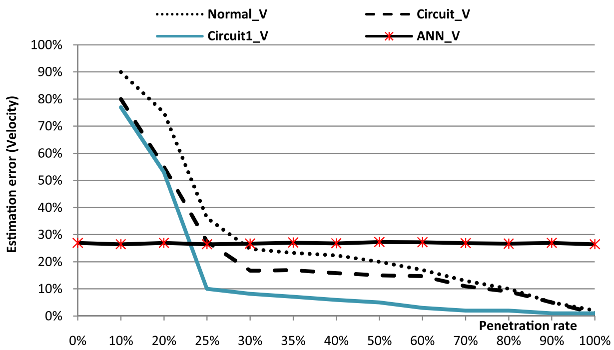

Fig. 13 The effectiveness comparison between the ANN-based prediction and the "velocity-density inference" circuits - The average velocity case.

Section 4.3. This dataset was then divided into two parts as the portion of 75% and 25% for the training and the testing datasets, respectively. The training dataset was used to train an ANN with 6 hidden nodes. The testing dataset was used to evaluate the effectiveness of the ANN-based prediction model. The target values to be estimated were the average velocity and the density of the aforementioned 5 randomly selected desired road segments. The average predicted error of these 5 selected road segments was calculated to evaluate the effectiveness of the ANN-based prediction model. The results of these evaluations are shown in **Fig. 12** and described as follows.

Figure 12 shows the effectiveness of the ANN-based prediction in predicting the average velocity and the density of the road segments. It seems that the prediction accuracy is a little bit better in estimating the average velocity (the average error is around 27%) than in estimating the density (the average error is around 29%). The reason of this difference may be that the velocity of the desired road segment has a better relation with the traffic state of the related road segments than the density does. However, this difference is minor and both of the prediction results are good enough. In addition, both the prediction results are not affected by any penetration rate. The predicted errors are almost identical regardless of the penetration rate (either 100% or 0%). This is because that the penetration rate of the desired road segments has not played any role in the ANN-based prediction model.

Figure 13 and **Fig. 14** show the summarized evaluation results of all the proposed "velocity-density inference" circuits and the ANN-based prediction model in estimating the average velocity and the density, respectively. Both the figures reveal that, the "adaptive **feedback** velocity-density inference" circuit is the best

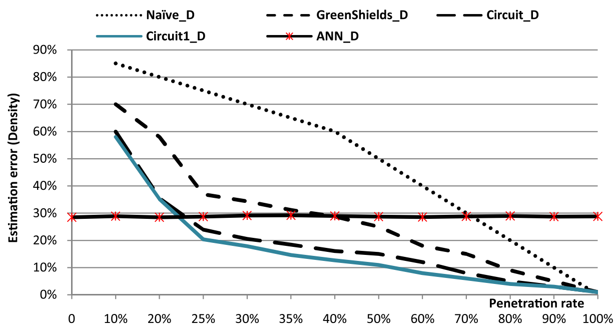

Fig. 14 The effectiveness comparison between the ANN-based prediction and the "velocity-density inference" circuits - The density case.

one if the penetration rate is low but still relevant. When the penetration rate is lower than the so-called "acceptable" penetration rate, namely 25%, the estimation error of the "adaptive **feedback** velocity-density inference" circuits increases drastically, thus the ANN-based prediction model is dominant.

7. Conclusions and Future Work

This article thoroughly analyzed the influence of the penetration rate on the mobile phone-based traffic state estimation. As a result, the so-called "acceptable" penetration rate with regards to a predefined expected accuracy could be identified by statistical studies on a huge amount of simulation data. This recognition is useful for researches and practical investigations in order to bring a traffic state estimation system using mobile phones into the realization. Furthermore, two notable "velocity-density inference" circuits, based on the so-called "adaptive" and "adaptive **feedback**" approaches, were proposed to improve the velocity/density estimation's effectiveness. These "velocity-density inference" circuits not only make the velocity/density estimation more accurate but also help to minimize the "acceptable" penetration rate required. This advantage is very important in practice since it guarantees the effectiveness of a traffic state estimation system even with a small number of vehicles (mobile phones) participating.

This article also introduced a suitable ANN-based prediction model to ensure the effectiveness of the mobile phone-based traffic state estimation when the penetration rate degrades to a very low value, namely several percent or even 0%. At such very low penetration rates, the aforementioned "velocity-density inference" circuits will not properly work. The effectiveness of all the proposed solutions (i.e., the "velocity-density inference" circuits and the ANN-based prediction model) has been confirmed by a plenty number of experimental evaluations.

The results of this research open a new research direction in the traffic state estimation using mobile phones. This research revealed that 100% of the vehicles travelling on the road network are not required to send the data to the server for ensuring an accurate estimation. In contrast, the estimation accuracy is kept to be high enough if the penetration rate passes the minimum threshold, namely the "acceptable penetration rate." Furthermore, the value required for an "acceptable" penetration rate can be reduced by reasonable estimation methods such as the proposed "velocity-density inference" circuits and the ANN-based prediction model. These results brace the confidences of accelerating the realization of mobile phone-based traffic state estimation systems. However, more detailed performance based on real-field experiments needs to be evaluated in order to confirm the robustness of the proposed approaches. In addition, combining both the "velocity-density inference" circuits and the data mining approach in a reasonable way for further effectiveness improvement is also a relevant research direction which should be considered in the future work.

Reference

[1] Papageorgiou, M.: *Application of Automatic Control Concepts to Traffic Flow Modeling and Control*, Springer-Verlag, NY. (1983).

[2] Kumar, P., Singh, V. and Reddy, D.: Advanced traveler information system for Hyderabad City, *IEEE Trans. Intell. Transp. Syst.*, Vol.6, No.1, pp.26–37 (Mar. 2005).

[3] Coifman, B.: Improved velocity estimation using single loop detectors, *Trans. Res. A.*, Vol.35, No.10, pp.863–880 (Dec. 2001).

[4] Tang, S. and Wang, F.Y.: 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).

[5] Wright, J. and Dahlgren, J.: *Using vehicles equipped with toll tags as probes for providing travel times*, California PATH Working Paper UCB-ITSPWP-2001-13, Institute of Transportation Studies, University of California, Berkeley, CA. (2001).

[6] Ban, X., Li, Y., Skabardonis, A. and Margulici, J.D.: Performance evaluation of travel time methods for real time traffic applications, *11th World Conf. on Transport Research*, Berkeley, CA. (June 2007).

[7] Cho, Y. and Rice, J.: Estimating velocity fields on a freeway from lowresolution videos, *IEEE Trans. Intell. Transp. Syst.*, Vol.7, No.4, pp.463–469 (Dec. 2006).

[8] Schneider, W. and Mrakotsky, E.: Mobile Phones as a Basis for Traffic State Information, *Proc. Intelligent Transportation Systems* (2005)

[9] Rose, G.: Mobile Phones as Traffic Probes: Practices, Prospects and Issues, *Transport Reviews*, Vol.26, No.3, pp.275–291 (May 2006).

[10] Yoon, J., Noble, B. and Liu, M.: Surface Street Traffic Estimation, *MobiSys 07*, San Juan, Puerto Rico, USA. (June 2007).

[11] Sohn, K. and Hwang, K.: 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).

[12] Work, D.B. and Bayen, A.M.: Impacts of the Mobile Internet on Transportation Cyberphysical Systems: Traffic Monitoring using Smartphones, *National Workshop for Research on High-Confidence Transportation Cyber-Physical Systems: Automotive*, Aviation and Rail, Washington, DC. (Nov. 2008).

[13] Quang, T.M. and Kamioka, E.: Granular Quantifying Traffic State Using Mobile Probes, *IEEE 72nd Vehicle Technology Conference*: VTC2010-Fall, Ottawa, Canada, CD-edition (Sep. 2010).

[14] Quang, T.M. and Kamioka, E.: Pinpoint: An Efficient Approach to Traffic State Estimation System Using Mobile Probes, *IEEE 6th International Conference on Wireless Communications, Networking and Mobile Computing*, Chengdu, China, CD-edition (Sep. 2010).

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

[16] Quang, T.M. and Kamioka, E.: Error-Tolerance in Quantifying Traffic States Using Mobile Phones, *The 4th International Workshop on Ad Hoc, Sensor and P2P Networks*, Tokyo, Japan (Mar. 2011).

[17] Hoh, B., Gruteser, M., Herring, R., Ban, J., Work, D., Herrera, J.C., Bayen, A., Annavaram, M. and Jacobson, Q.: Virtual Trip Lines for Distributed Privacy-Preserving Traffic Monitoring, *MobiSys'08*, Breckenridge, Colorado, USA. (June 2008).

[18] Liang, Y., Guo, D., Suo, M. and Sun Q.: A New Idea of Identification of Road Traffic Conditions, *International conference on Computer Application and System Modeling* (2010).

[19] Quang, T.M. and Kamioka, E.: Toward An Efficient Mobile Phone Based Traffic Estimation System, IEICE Technical Report, Okinawa, Japan, Vol.110, No.40, pp.19–24 (May 2010).

[20] Greenshields, B.D.: A Study in Highway Capacity, *Highway Research Board Proceedings*, pp.448–477 (1935).

[21] Greenberg, H.: An analysis of traffic flow, *Operation Research*, pp.79–85 (1959).

[22] Underwood, R.T.: Speed, Volume, and Density Relationship, Quality and Theory of Traffic Flow, pp.141–188, Yale University Bureau of Highway Traffic, New Haven, Connecticut (1961).

[23] Rumelhart, D.E., Widrow, B. and Lehr, M.A.: The basic ideas in neural networks, *Comm. ACM*, Vol.37, No.3, pp.87–92 (Mar. 1994).

[24] Gora, P.: Traffic Simulation Framework – A Cellular Automaton-Based Tool for Simulating and Investigating Real City Traffi, *Recent Advances in Intelligent Information Systems*, ISBN 978-83-60434-59-8, pp.641–653 (2009).

[25] Muraki, J. and Takano, D.: Development of Traffic Prediction System Based on Nearest Neighbor Method Considering Information of Connected Road, Special Issue of System Development Research Project, CS Technical Report, University of Tsukuba (2007) (in Japanese).

Tran Minh Quang is a Ph.D. candidate of Functional Control Systems at Shibaura Institute of Technology, Japan. He received his B.Eng. (2001), M.Eng. (2006) in Computer Science and Engineering from Hochiminh City University of Technology, Vietnam, and Ritsumeikan University, Japan, respectively. He has been a Lecturer at Hochiminh City University of Technology since 2003. His current research interest includes advanced information systems such as data mining, intelligent transportation systems, mobile multimedia communications and mobility computing.

Eiji Kamioka is a Professor at Shibaura Institute of Technology. He received his B.S. (1989), M.S. (1991) and D.S. (1997) degrees in physics from Aoyama Gakuin University. He also has other job experiences, working for SHARP Communications Laboratory (1991–1993), Institute of Space and Astronautical Science (ISAS) as a JSPS Research Fellow (1997–1998) and National Institute of Informatics (NII) as an Assistant Professor (1998–2006). His current research interests encompass mobile multimedia communications, ubiquitous computing and context-aware applications.