Synergistic approaches to mobile intelligent transportation systems considering low penetration rate

T.M. Quang *, Muhammad Ariff Baharudin, Eiji Kamioka

Graduate school of Engineering and Science, Shibaura Institute of Technology, Tokyo, Japan

A R T I C L E   I N   F O

Article history:
Received 24 November 2011
Received in revised form 12 July 2012
Accepted 17 July 2012
Available online xxxx

Keywords:
Mobile probe
Traffic state estimation
Low penetration rate
Mobile intelligent transportation systems (MIT)

A B S T R A C T

This paper investigates the effect of low penetration rate on mobile phone-based traffic state estimation (M-TES) models. Synergistic approaches, including an appropriate genetic algorithm (GA) based velocity–density estimation model and a notable artificial neural network (ANN) based prediction method for unacceptably low penetration rate, are proposed. The GA-based traffic state estimation model not only improves the effectiveness but also reduces the critical penetration rate required in the M-TES model. When the critical penetration rate is reduced the error-tolerance and the scalability of the estimation model can be significantly improved. The ANN-based prediction approach is introduced to overcome the weakness remaining in the GA-based traffic state estimation model when the penetration rate becomes unacceptably low or unknown. In addition, the effect of related road segments on the prediction effectiveness is thoroughly discussed. This work, therefore, provides practical instructions in narrowing the search space for finding prediction rules of the ANN model, thus improving the computational performance without compromising the prediction accuracy. The experimental evaluations confirm the effectiveness as well as the robustness of the proposed approaches. As a result, this research contributes to accelerating the realization of mobile phone-based intelligent transportation systems (M-ITS) or, of the M-TES systems in specific, since the essential issue of low penetration rate has been solved.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Transportation [1] and road traffic [2] are important parts of any economy all over the world. Therefore, research in Intelligent Transportation Systems (ITS) [3,4] has attracted a numerous number of researchers in various fields including vehicular technology, transportation, civil engineering, statistical study, computational science, communication engineering, and so forth [5–7]. However, besides advanced achievements in ITS, traffic congestion still remains as a serious issue in almost every big city across the world. A traffic jam is not only the cause of economic loss but also the source of pollution (air, noise pollution, etc.), violence and other social issues [8,9]. The Urban Mobility Report [10] reported in 2007 that traffic congestion causes 4.2 billion hours of extra travel time requiring 2.9 billion extra gallons of fuel, which cost the United State tax-payers an additional $78 billion [11]. The Ministry of Land Infrastructure and Transport of Japan reported in 2006 that the economic loss caused by a traffic jam is around $100 billion annually [12]. In addition, the situations where ambulances are hopelessly stuck on the way to hospitals; shops along the road sides have to be closed; students, teachers, workers, officers cannot go to school/work in time because of traffic jams, are not unusual in the modern cities. Such uncomfortable and even dangerous traffic environments are declining the citizens’ quality of life (QoL).

* Corresponding author.

E-mail addresses: m709504@shibaura-it.ac.jp (T.M. Quang), m110422@shibaura-it.ac.jp (M.A. Baharudin), kamioka@shibaura-it.ac.jp (E. Kamioka).

1574-1192/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.pmcj.2012.07.008

Please cite this article in press as: T.M. Quang, et al., Synergistic approaches to mobile intelligent transportation systems considering low penetration rate, Pervasive and Mobile Computing (2012), doi:10.1016/j.pmcj.2012.07.008
Traffic state estimation is one of the most important fields in ITS researches [13,3] aiming at alleviating traffic congestion. Obviously, correct updated traffic information helps commuters to avoid unconsciously entering the heavy traffic areas so that traffic congestion can be avoided. A reliable traffic state estimation system must provide not only accurate but also real-time traffic state information at any place (ubiquity). Existing traffic state estimation systems mainly employed traditional data collection methods which relied on road-side fixed sensors such as loop detectors [14,15], RFID readers [16,17], video cameras [18], and so forth. These techniques, however, disclosed their essential weakness in terms of coverage limitation as well as investment cost since it is impractical to install a huge number of road-side fixed sensors on every street [19]. To solve these issues, mobile devices have been utilized as traffic probes to collect traffic data [20,21]. Since mobile phones are available everywhere and the mobile phone network has already been deployed, the issues of coverage limitation, real-time effect, investment and maintenance cost mentioned above can be overcome [22–24]. Consequently, the mobile phone based ITS (M-ITS) [25] research is entering a new stage aiming at realizing a SAFE and GREEN (no traffic congestion) traffic environment [26,27] to improve the citizens’ QoL.

In spite of the aforementioned advantages, the mobile phone based traffic state estimation (M-TES) approach faces several issues related to uncertain penetration rate [11,28,29]. The penetration rate is the fraction of the number of vehicles (i.e. the number of mobile phones carried by vehicles) that report data to the estimation server out of the total number of vehicles traveling through the considered road segment. Intuitively, the higher the penetration rate the more effective the system is. However, there is no way to compel each mobile phone user to report traffic data to the server. In addition, when the system has just been launched, the penetration rate is usually lower than the one required which significantly distorts the estimation’s effectiveness. Therefore, the penetration rate issue is considered as an essential barrier of the M-ITS realization. This barrier should be removed by appropriate approaches ensuring traffic estimation accuracy even in cases of low penetration rate.

This paper investigates and proposes suitable solutions for the aforementioned penetration rate related issues. More concretely, the effect of low penetration rate on traffic state estimation effectiveness will be thoroughly investigated. According to this investigation, a notable genetic algorithm (GA) based velocity–density estimation model will be proposed. The proposed GA mechanism is expected not only to improve the traffic state estimation accuracy but also to minimize the critical penetration rate (the penetration rate at which the estimation accuracy is still good enough but it will drastically decrease if the penetration rate becomes lower). This approach enhances the reliability as well as the scalability of the estimation system. Moreover, a practical Artificial Neural Network (ANN) based prediction model is proposed to ensure the effectiveness of the M-TES model when the penetration rate becomes unacceptably low, namely just several percent or even 0%. Last but not least, the effect of different order/level related road segments on the prediction accuracy is investigated.

As a result, the search space of the ANN-based prediction model can be minimized, thus improving the computational performance without compromising the prediction effectiveness.

This article is organized as follows: Section 2 reviews the related work revealing the necessity of this research. The problem formulation and related definitions are presented in Section 3. Section 4 introduces an appropriate GA-based mechanism to optimize the velocity–density inference model. The ANN-based prediction approach dealing with unacceptably low or unknown penetration rate is discussed in Section 5. This section also investigates the effect of different order related road segments on the prediction accuracy by which the search space for finding prediction rules in the ANN model can be reduced significantly without compromising the prediction effectiveness. The evaluations are thoroughly discussed in Section 6. Section 7 concludes this work and draws out future research directions.

2. Related work

From the past decade, several researches have been dedicated to utilizing mobile phones as traffic probes for traffic state estimation [30–32]. Although several types of sensors such as GPS, accelerometers, cameras, etc., have been equipped in commercialized mobile phones [33,34], not all of these sensors can be effectively utilized for traffic data collection due to the requirements of autonomy and invisibility of this system [35]. Consequently, traffic data a mobile phone can report to the estimation is primarily the GPS data which consists of time stamp (in seconds), position (longitude, latitude), current velocity, direction and vehicle ID of the vehicle that sent data. This data is useful in intelligent transportation related systems such as recognizing pedestrians [36], classifying vehicles [37], monitoring safe drives [38], estimating traffic state [39–41], and so forth. The mobile phone based approach, however, retains several issues which can be categorized as follows: (1) the limitations of the sensors equipped on the commercial “on-the-shelf” mobile phones [42,43]; (2) the limitations in the mobile phone technology resources such as the computational capacity of mobile phones and the bandwidth of the mobile phone network [44]; and (3) the difficulties rooted from low and uncertain penetration rate (the portion of the number of vehicles that report data to the estimation server out of the total number of vehicles traveling in the considered road segment) [45,46].

In 2008, a mobile phone based traffic state estimation project named Mobile Century Project was started in the University of California, Berkeley. This project was renamed Mobile Millennium Project (MMP) in 2010 [47] and is currently in its evaluating stage. Here, GPS enabled mobile phones are utilized as traffic probes for real-time data collection. The estimation server processes data, estimates traffic states, and disseminates traffic state information to drivers via a mobile phone network or via the Internet.
In 2008, Hoh et al. [48] proposed the so-called Virtual Trip Lines (VTLs) approach to collection in order to minimize the data transmission load in the mobile phone network. The VTLs are preset on the road network by which data is collected and reported only when the vehicle passes a VTL. Intuitively, this approach reduces the number of data collections, and thus the data transmission load is also reduced. However, the essential issue here is that useful data may be missed since there is no relationship between VTLs and the places where traffic congestion actually occurs. Moreover, the VTL setting criteria such as where on the road a VTL should be set, how far should two consecutive VTLs be, how to set the VTLs in different road types (e.g. arterial road, highway, urban street, etc.), and so forth, are matters of argument.

Our study in 2010 proposed the so-called “pinpoint” approach [24] by which a vehicle (or a mobile phone carried on the vehicle) reports traffic data to the estimation server only when its velocity changes (increases/decreases) significantly. Obviously, the data reported under this approach closely reflects the change of traffic state, and thus it directly contributes to the traffic state estimation. Therefore, not only the data transmission load but also the data redundancy was reduced significantly. However, that work assumed that every vehicle (mobile phone) reports data to the estimation server. This assumption is difficult to be satisfied in real-world applications.

In addition to controlling data transmission load, effectively estimating and comprehensively representing a traffic state based on less informative GPS data is also an essential requirement for an M-TES system. Existing researches primarily relied on the travel time estimated using data reported by representative vehicles (the GPS-equipped ones) to estimate traffic state [22]. This approach, however, reveals estimation biases since the density of the considered traffic flow was not taken into account. Meanwhile, density itself is considered as an important factor affecting the traffic state level [49]. To solve this issue, both the average velocity and density should be estimated independently before being integrated in an appropriate traffic state estimation model as proposed in [19,50]. In these researches, traffic state was granularly quantified by a continuous value in the range of \([-1, 1]\) representing from the worst to the best traffic conditions. Nevertheless, the essential issue here is that it is sensitive to the penetration rate. Intuitively, when the penetration rate becomes lower the accuracy of the traffic state estimation model will be significantly reduced [51]. However, the effect of low penetration rate on the estimation effectiveness was not discussed in these researches. Instead, they still assumed that the penetration rate must be always relevant.

The work in [11] is closely related to this article which aims at ensuring the traffic estimation effectiveness even if the penetration rate becomes unacceptably low. The authors proposed to apply a statistical learning model to estimate traffic state in terms of travel time and congestion state. Historical data was employed to train the statistical learning model so that it can estimate/forecast the travel time and congestion state of the considered road segment when the current observed data is applied. This work claimed that the logistic regression model works effectively even if the penetration rate is quite low, namely at 5% of penetration rate the estimation error is lower than 30%. However, there are several issues remaining that need to be thoroughly discussed and clarified. First, this work extended the work in [48] by which the VTL concept was applied. As mentioned before, the effectiveness of the VTL itself is still a matter of argument. Second, the work could not estimate density of the traffic flow, and thus this important factor could not be taken into account when estimating congestion state. Third, congestion state was defined as a “binary” indicator which accepts only two states, namely, “congested” and “not-congested”. Obviously, this setting biases the estimation accuracy since even a “blind” guessing approach also has the opportunity to reach the accuracy of 50%. Fourth, the work employed the Paramics simulator [52] to generate synthetic data, which gave information about every vehicle, for evaluations. To imitate a low penetration rate dataset, namely 5% for example, a large portion of data (95%) was removed and only a subset of data (5%) was kept. In fact, this process could not generate the appropriate low penetration rate dataset as it was defined in their work. Therefore, the estimation error versus the penetration rate obtained in that research should be clarified.

Studies in [29,46] thoroughly investigated the effect of low penetration rate on the estimation effectiveness, and then proposed appropriate solutions. The authors proposed two major solutions, namely, the “velocity–density inference circuit” and the ANN-based prediction model. The velocity–density inference circuit is depicted as in Fig. 1.

Here, the inputs of the inference circuit are velocity and density, denoted as \(V^i_{\text{sensed}}\) and \(D^i_{\text{sensed}}\), respectively. These factors were directly calculated from the sensed data collected in the road segment \(i\), during the estimation time interval \(k\). The outputs of the inference circuit are, of course, the estimated velocity and density, namely \(V^i_{\text{est}}\) and \(D^i_{\text{est}}\). The “heart” of the circuit is the estimation model which is fed by the three types of values, namely, the sensed velocity/density mentioned above, the Greenshields-based [53] inferred velocity/density, namely \(V^i_{\text{infer}}\) and \(D^i_{\text{infer}}\), and the moving average values of...
and maximum density of the considered road segment, respectively.

\[ V_{est}^{k,i} = \alpha \cdot \text{Avg}(V_{sensed}^{k,i}, MV^{k,i}) + (1 - \alpha)V_{infer}^{k,i} \]  
\[ D_{est}^{k,i} = (1 - \alpha)D_{sensed}^{k,i} + \alpha \cdot \text{Avg}(D_{infer}^{k,i}, MD^{k,i}). \]  

The velocity–density estimation model (the “heart” of the inference circuit) is described in Eqs. (1) and (2). Here, \( V_{sensed}^{k,i}, D_{sensed}^{k,i}, MV^{k,i} \) and \( MD^{k,i} \) are calculated in Eqs. (3)–(6), respectively. In these equations, \( V_{\text{max}} \) and \( D_{\text{max}} \) are limit speed and maximum density of the considered road segment \( i \), and \( \xi \) is the sliding window for the moving average calculations. The sliding window can be set by domain experts or by using simulation data and it was set to \( 3 (\xi = 3) \) in [29, 46]. In Eq. (1), \( \alpha (0 < \alpha < 1) \) is the coefficient representing the impact of velocity calculated directly from the sensed data, \( V_{sensed}^{k,i} \), on the whole average velocity estimation model. Therefore, the impact of velocity inferred from estimated density, \( V_{infer}^{k,i} \), must be \( 1 - \alpha \). Since \( D_{infer}^{k,i} \) in Eq. (2), is inferred from \( V_{sensed}^{k,i} \) as in Eq. (4), its impact on the whole density estimation model should be \( \alpha \), thus \( 1 - \alpha \) is the impact of \( D_{sensed}^{k,i} \).

\[ V_{infer}^{k,i} = V_{\text{max}} \left(1 - \frac{D_{sensed}^{k,i}}{D_{\text{max}}}\right) \]  
\[ D_{infer}^{k,i} = D_{\text{max}} \left(1 - \frac{V_{sensed}^{k,i}}{V_{\text{max}}}\right) \]  
\[ MV^{k,i} = \frac{\sum_{j=k-\xi}^{k-1} V_{est}^{j,i}}{\xi} \]  
\[ MD^{k,i} = \frac{\sum_{j=k-\xi}^{k-1} D_{est}^{j,i}}{\xi}. \]  

The essential issue in the velocity–density estimation model is how to effectively determine the coefficient \( \alpha \). This coefficient was proposed to be approximated by mapping the current “sensed” situation, namely \( V_{sensed}^{k,i}/V_{\text{max}} \), to the rules revealed from statistical analysis on historical data. As reported in [29, 46], \( \alpha \) was approximated to one of the two values, namely 0.6 or 0.8. Therefore, the coefficients \( (\alpha \) and \( 1 - \alpha) \) of Eqs. (1) and (2) represent only two combinations of \( \{0.6 \text{ and } 0.4\} \) and \( \{0.8 \text{ and } 0.2\} \). These coefficient combinations seem to be so empirical while there may exist other combinations such as \( \{0.7 \text{ and } 0.3\} \) leading to optimal performance. It would be more effective if \( \alpha \) was granularly determined. In this article, an appropriate GA model will be introduced to optimize the coefficient \( \alpha \), thus optimizing the velocity–density estimation model. In addition, this work also enhances the ANN-based prediction model proposed in [29, 46] to predict traffic state of road segments where the penetration rate is unacceptably low. Furthermore, the effect of different order related road segments on the prediction accuracy is also thoroughly investigated. Consequently, appropriate related road segments can be well selected by which the search space can be significantly reduced. As a result, this study provides a reliable way to evaluate the accuracy of the prediction model when the traffic state data of related road segments is missed.

3. Problem formulation

Let \( V = \{i|i = 1, \ldots, N\} \) denote a set of road segments divided from a road network. For any road segment \( i \in V \), a collection of GPS data reported by mobile phones is available at any time \( t \). Since the GPS data is an event-based data, it cannot be directly transformed into traffic state. Therefore, traffic state should be aggregated in predefined time intervals, namely in \( t\text{-second} \) windows [11]. More concretely, traffic state should be estimated at times \( k = 0, t, 2t, \ldots, \) where \( t \) is the aggregation time mentioned above.

Obviously, average velocity and density of the traffic flow reflect traffic state of the considered road segment. These factors can be independently and directly estimated using the GPS data reported by mobile phones.

**Definition 1.** Average velocity of a traffic flow in the road segment \( i \) during time \( k \), denoted as \( V_{\text{Avg}}^{k,i} \), is the average velocity of all vehicles traveling in the considered road segment.

Eq. (7) is the mathematical calculation of average velocity, where \( V_{m,j}^{k,i} \) is the velocity of any individual vehicle \( j \) \((j = 1, \ldots, q)\) detected at time \( t_m \) \((m = 1, 2, \ldots, r)\) within the time interval \( k|k - 1|t \leq t_m \leq k|t \). Here, \( r \) is the total number...
of data (GPS data) detection times during time interval $k$.

$$V_{Avg}^{k,i} = \frac{\sum_{j=1}^{q} V_{tm,j}^{k,i}}{qr}, \quad (k-1)t \leq t_m < kt.$$  \hspace{1cm} (7)

Average velocity reflects the traffic state of the considered road segment in terms of travel time. More concretely, the higher the average velocity is the better the traffic state is, and vice versa. However, this factor cannot be used to compare traffic states of different road segments since the limited speed varies from road segment to road segment. Our previous work [19,50] proposed a new term, namely the mean speed capacity, to better quantify the travel time of a traffic flow at a particular road segment. This term is defined as follows:

**Definition 2.** Mean speed capacity of the road segment $i$ during time $k$, denoted as $M_{V}^{k,i}$, is the fraction of the average velocity and the limit speed of the considered road segment.

The mean speed capacity can be calculated in Eq. (8), where $V_{Avg}^{k,i}$ is the average velocity defined in Definition 1 and $V_{max}^{i}$ is the limit speed of road segment $i$.

$$M_{V}^{k,i} = \frac{V_{Avg}^{k,i}}{V_{max}^{i}}.$$  \hspace{1cm} (8)

**Definition 3.** Density of a traffic flow in the road segment $i$ during time $k$, denoted as $D^{k,i}$, is the fraction of the number of the vehicles traveling through the considered road segment during time $k$ out of the capacity of the considered road segment.

Eq. (9) describes the density calculation, where $q^{k,i}$ is the total number of vehicles traveling through road segment $i$ during time $k$, and $C^{k,i}$ is the flow capacity of the road segment $i$.

$$D^{k,i} = \frac{q^{k,i}}{C^{k,i}}.$$  \hspace{1cm} (9)

**Definition 4.** The flow capacity, $C^{k,i}$ shown in Eq. (10), is the maximum number of vehicles which can pass through the road segment $i$ during time $k$ under the best traffic condition.

$$C^{k,i} = Q_{0}^{i} + Q^{k,i}.$$  \hspace{1cm} (10)

Here, $Q_{0}^{i}$ is the maximum number of vehicles that can be arranged (without moving) in the road segment $i$, and $Q^{k,i}$ is the maximum number of vehicles that move out the road segment $i$ during time $k$ in the best traffic condition. These parameters are calculated in Eqs. (11) and (12), respectively.

$$Q_{0}^{i} = \frac{m}{1.5l_{c}},$$  \hspace{1cm} (11)

$$Q^{k,i} = \frac{m}{t} = mk \frac{V_{max}^{i}}{1.5l_{c}}.$$  \hspace{1cm} (12)

In these equations, $m$ and $l$ are the number of lanes and the length of the road segment $i$; and $l_{c}$ is the average length of a car [50,54]. Here, the space between two cars must be at least 0.5$l_{c}$ (can be reached in the congested areas). In Eq. (12), $t$ is the average elapse time between two consecutive vehicles, namely vehicles $i$th and $(i+1)$th, moving out of the considered road segment. Fig. 2 illustrates these parameters.

As mentioned in the beginning of this section, both the average velocity (represented by the mean speed capacity) and density of the traffic flow represent the traffic state of the considered road segment. Therefore, they should be integrated in an appropriate way to accurately quantify (not merely qualify) traffic state. In this work, a term called the Goodness value is defined to quantify a traffic state.
Fig. 3. Vehicles that report data are denoted as the car-shape ones. Here, \( p = 8 \), \( q = 15 \) and \( \rho = 8/15 \).

**Definition 5.** Goodness value of the road segment \( i \) during the estimation time \( k \), denoted as \( S^{k,i} \), is estimated as a function of the mean speed capacity \( M^{k,i}_{V} \) and the density \( D^{k,i} \) as described in Eq. (13).

\[
S^{k,i} = f(M^{k,i}_{V}, D^{k,i}) = M^{k,i}_{V} - M_{V0} + D_0 - D^{k,i}. \tag{13}
\]

Here, \( M_{V0} \) and \( D_0 \) are the thresholds of the mean speed capacity and density, respectively, by which the traffic state is considered as good enough. These thresholds can be identified by the transportation experts or via numerous evaluations on simulation data [50].

Obviously, Goodness value is a continuous value, ranging from \(-1\) to \(1\), representing from the worst to the best traffic states. It is quite adequate for granularly comparing traffic state levels. However, an essential issue in this approach is that both the mean speed capacity and density are estimated directly from the sensed data reported by mobile phones while there is no way to enforce every vehicle (mobile phone) to report data to the estimation server. In real-world applications, especially when the system has just been launched, the rate of mobile phones that report data to the server is commonly unacceptably low. The lower this rate is, the higher average velocity and density estimation errors are. Finding a suitable solution by which traffic state estimation accuracy is ensured to be as higher than or equal to the expected accuracy level, namely \(70\%\) for example [46] is a challenging issue.

The rate of mobile phones that report data to the server is termed the penetration rate which is formally defined as follows:

**Definition 6.** Penetration rate at road segment \( i \) during time \( k \), denoted as \( \rho^{k,i} \), is the fraction of vehicles that report data to the estimation server out of the total number of vehicles traveling through the considered road segment. The penetration rate can be expressed in Eq. (14), and depicted in Fig. 3. Here, \( p \) is the number of vehicles that report data to the estimation server, and \( q \) is the total vehicles traveling through the road segment \( i \) during time \( k \).

\[
\rho^{k,i} = \frac{p}{q}. \tag{14}
\]

With a given penetration rate \( \rho^{k,i} \), the average velocity estimation model described in Eq. (7) is replaced by Eq. (15). Here, \( V_{km,j}^{k,i} \), \( q \) and \( r \) were defined in Eq. (7).

\[
V_{Avg}^{k,i,\rho^{k,i}} = \frac{\sum_{j=1}^{q} V_{km,j}^{k,i}}{\rho^{k,i}qr}, \quad (k - 1)t \leq t_m < kt. \tag{15}
\]

Under this condition of penetration rate, the average velocity estimation error can be expressed in Eq. (16), where \( V_{Avg}^{k,i} \) is the “actual” average velocity estimated when every vehicle reports data Eq. (7), and \( V_{Avg}^{k,i,\rho^{k,i}} \) is the average velocity estimated under the given penetration rate \( \rho^{k,i} \) Eq. (15).

\[
E_{V}^{k,i} = \left| 1 - \frac{V_{Avg}^{k,i,\rho^{k,i}}}{V_{Avg}^{k,i}} \right| = \left| 1 - \frac{\rho^{k,i}qr}{q} \sum_{j=1}^{q} V_{km,j}^{k,i} \right|. \tag{16}
\]

Similar to average velocity, density estimation is also affected by the penetration rate. According to **Definition 3** and Eq. (9), the density estimation error, denoted as \( E_{D}^{k,i} \), is directly affected by the penetration rate \( \rho^{k,i} \) as expressed in Eq. (17).

\[
E_{D}^{k,i} = (1 - \rho^{k,i}) 100\%. \tag{17}
\]

Obviously, low penetration rate significantly affects both the average velocity and density. For example, if the penetration rate is 20% the density estimation error will be 80%, an unacceptable error for any estimation model. Meanwhile, the penetration rate of 20% or lower is common in practice. Therefore, it is essential to find out suitable methods to ensure traffic state estimation accuracy even when the penetration rate becomes inadequate. In this article, an appropriate GA mechanism will be proposed to optimize the velocity–density estimation model [46]. The GA-based velocity–density estimation model...
can significantly improve the traffic state estimation accuracy when the penetration rate becomes low. The effectiveness of this approach will be shown in the evaluation section (Section 6.3). The GA-based approach, however, may not properly work when the penetration rate becomes unacceptably low, namely just several percent or even 0%. To complement the weakness of the GA-based approach in this point, a notable ANN-based prediction approach was proposed. This approach can effectively predict traffic state of unacceptably low penetration rate road segments. The details of these proposed approaches will be expressed in the remainder of this article.

4. GA-based velocity–density inference model

This section proposes an appropriate genetic algorithm (GA) mechanism to optimize the performance of the velocity–density estimation model.

4.1. Overview of the genetic algorithm

Genetic algorithm is a technique that mimics the natural selection (Darwin’s theory of survival) which is effectively applied for the optimization and global search [55,56]. The process of a GA can be described as the pseudo code depicted in Fig. 4. An initial population, \( P \), (a set of chromosome) with a suitable size is randomly generated. Each chromosome consists of one or several genes. A gene is a string of bit or a string of real-value numbers. Each chromosome represents a solution of the problem that the GA is used to solve. In each population, candidates (chromosomes) are selected based on their fitness values so that those individuals that are more competitive have a larger chance to survive and keep the genetic information to their offspring. High competitive chromosomes (the survived ones) are duplicated with a corresponding probability based on their fitness to replace the low competitive candidates and keep the size of population unchanged. After that, any two selected candidates are mated randomly producing two alternative children under the crossover procedure. The mutation helps to prevent falling GA into local extreme, so that the global solution may be found earlier. The GA would not stop until the fitness of the best chromosome in the current population satisfies a present threshold or when the iterations pass the preset times.

4.2. Modeling the GA mechanism

This section presents in detail the proposed GA mechanism applied to optimize the velocity–density estimation model. This modeling includes coding the chromosome schema, proposing the evaluation and selection mechanism as well as the crossover and mutation method.

4.2.1. Coding

One of the most important steps in applying the GA based optimization model is to “code” the chromosome schema by which a chromosome can appropriately describe a solution that the GA used to solve. In this work, the GA is designed to optimize the coefficient \( \alpha \) in the velocity–density estimation model Eqs. (1) and (2). However, the original velocity–density estimation model Eq. (1) equally treats \( V_{\text{sensed}}^k, i \), \( M_{\text{sensed}}^k, i \), and \( V_{\text{infer}}^k, i \), in terms of their impact on the whole velocity estimation model. Similarly, the impact of \( D_{\text{sensed}}^k, i \) and \( M_{\text{sensed}}^k, i \) on the whole density estimation model (Eq. (2)) is also considered to be equal. This approach reduces the complexity of the estimation model but it distorts the estimation accuracy. Therefore, before applying the GA, the velocity–density estimation model should be modified by which each factor, namely \( V_{\text{sensed}}^k, i \), \( M_{\text{sensed}}^k, i \), etc., should be appropriately assigned a different impact coefficient to make it more flexible for optimization. Here we propose to modify the original velocity–density estimation model Eqs. (1) and (2) as described by Eqs. (18) and (19), respectively.

\[
V_{\text{est}}^k, i = \omega V_{\text{sensed}}^k, i + \beta M_{\text{sensed}}^k, i + \gamma V_{\text{infer}}^k, i \quad (18)
\]

\[
D_{\text{est}}^k, i = \gamma D_{\text{sensed}}^k, i + \beta D_{\text{sensed}}^k, i + \alpha D_{\text{infer}}^k, i \quad (19)
\]
Here, $V^{k}_{\text{sensed}}$, $D^{k}_{\text{sensed}}$, $V^{k}_{\text{infer}}$, $D^{k}_{\text{infer}}$, $MV^{k}$, and $MD^{k}$ are defined as they were in Eqs. (1) and (2). The major difference is that $V^{k}_{\text{sensed}}$, $MV^{k}$, and $V^{k}_{\text{infer}}$ are assigned with different impact coefficients $\alpha$, $\beta$, and $\gamma$, respectively, in the velocity estimation model Eq. (18). Since $D^{k}_{\text{infer}}$, in Eq. (19), is inferred from $V^{k}_{\text{sensed}}$ as in Eq. (4), its impact coefficient should be $\alpha$, and $\gamma$ is the impact of the $D^{k}_{\text{sensed}}$. It should be noted that there are two constraints for these coefficients as described in Eqs. (20) and (21).

\begin{align}
\alpha, \beta, \gamma & \in [0, 1] \\
\alpha + \beta + \gamma & = 1.
\end{align}

Coefficients $\alpha$, $\beta$, and $\gamma$ are real-value numbers in the range of $[0, 1]$. Conventionally, a real-value number is discretized into intervals of an acceptable degree of detail before modeling. For example, the range of $[0, 1]$ can be discretized into 256 intervals by which each interval (i.e. each discretized value) can be described by a string of 8 bits. However, the discretization may contain some drawbacks in the manner of granularity which affects the optimal performance. One may argue that why the range of $[0, 1]$ should not be discretized into 512 intervals or more. Therefore, in this work an appropriate GA mechanism is proposed so that it can work with chromosomes (solutions) modeled by real-value numbers [57]. Concretely, the schema of a chromosome in the proposed GA mechanism is coded as $g = \{\alpha, \beta, \gamma\}$ (the chromosome of 3 genes described in real-value numbers).

### 4.2.2. Evaluation and selection

Candidates (chromosomes) are evaluated to examine their competitiveness. Therefore, the fitness function must be designed so that the coefficients $\alpha$, $\beta$, $\gamma$ are motivated to be optimal. Here, the fitness function is defined in Eq. (22).

\[
f(g_i) = \frac{e(g_i)}{\bar{e}(g_i), \forall g_i \in \text{population}}
\]

where, $e(g_i)$ is the evaluation of candidate $g_i (g_i = \{\alpha, \beta, \gamma\})$ and $\bar{e}(g_i)$ is the average evaluation of all individuals $g_i$ in the current population. The evaluation $e(g_i)$ is the estimation error, namely the velocity estimation error in Eq. (18) caused by selecting $g_i$ as the set of coefficients. The evaluation is defined in Eq. (23), where $V_{\text{est},g_i}$ is the estimated velocity with the set of coefficient $g_i$, and $V_{\text{act}}$ is the “actual” velocity.

\[
e(g_i) = \frac{|V_{\text{est},g_i} - V_{\text{act}}|}{V_{\text{act}}}.
\]

After obtaining the fitness of all chromosomes in a population, the GA selects only the ones whose fitness values $f(g_i)$ are relevant to the next crossover process.

### 4.2.3. Crossover

Since the chromosome schema is built on real-value numbers, the arithmetic crossover [58,59] is applied in Eq. (24), where $g_1$, $g_2$ are the two parent chromosomes and $g'_1$, $g'_2$ are their children; and $\lambda$ is a random coefficient in the range of $[0, 1]$.

\[
\begin{align}
\{g'_1 = \lambda g_1 + (1 - \lambda)g_2 \\
g'_2 = (1 - \lambda)g_1 + \lambda g_2.
\end{align}
\]

### 4.2.4. Mutation

In this work, the non-uniform mutation approach [60] was applied. Assuming that a gene, namely $\alpha_i$, of the chromosome $g''_i = \{\alpha, \beta, \gamma\}$ is mutated, the mutation is described in Eq. (25).

\[
\alpha''_i = \begin{cases} 
\alpha_i + \Delta(t, UB - \alpha_i), & \text{if a random } \zeta \text{ is } 0 \\
\alpha_i - \Delta(t, \alpha_i - LB), & \text{if a random } \zeta \text{ is } 1.
\end{cases}
\]

Here, $UB$ and $LB$ are the upper and lower bounds of the variable $\alpha_i$. The function $\Delta(t, x)$, where $t$ describes the current generation, returns a value in the range of $(0, x)$ which is defined in Eq. (26) [57].

\[
\Delta(t, x) = x \left(1 - r^{(1 - \frac{t}{T})^b}\right).
\]

In Eq. (26), $r$ is a uniform random number ranging in $[0, 1]$, $T$ is the maximal generation number, and $b$ is a system parameter determining the degree of dependency on the iteration number.

Currently, the GA mechanism aiming at optimizing the velocity–density inference model has been proposed. The effectiveness of this approach will be confirmed in the evaluation section. The next section discusses the issue of ensuring the traffic state estimation effectiveness when the penetration rate becomes unacceptably low.
5. Predicting traffic state under unacceptably low penetration rate

In general, the GA approach discussed in Section 4 improves the performance of the velocity–density inference circuit significantly when the penetration rate becomes low. However, the penetration rate must be still relevant enough, namely around 20%, for example, as shown in the evaluation section (Section 6.3). Obviously, this approach cannot work properly if the penetration rate becomes unacceptably low, namely just several percent or even 0%. This issue can be addressed by applying a data mining technique on historical data to predict traffic state of unacceptably low penetration rate road segments [29,46]. Nevertheless, the influence of nearby road segments on traffic state of the considered road segment, and thus on the prediction accuracy has not been investigated. These issues will be thoroughly discussed in this section.

Field studies revealed that traffic state of a road segment is affected by traffic state of the nearby road segments. In addition, the current traffic state of each road segment has a close relationship with its previous states. If these spatial–temporal relationships (rules) are known in advance, the traffic state of the considered road segment can be predicted. These rules can be learned by any machine learning technique using historical traffic state data. In this work, a neural network (ANN) with multilayer perceptron (MLP) [61,62] is employed to predict the average velocity and density of considered road segments when their penetration rates become unacceptably low or unknown. This approach is depicted in Fig. 5.

As shown in Fig. 5, both the current traffic state of any related road segment \( j \) \((j \in V)\) and the historical traffic state of considered road segment \( i \) are applied as the input data for the ANN model. The current (i.e. at time \( t \)) velocity and density of a particular related road segment \( j \) are denoted as \( V^{k,j} \), \( D^{k,j} \), respectively. The previous (at time \( t \), where \( t < k \)) velocity and density of the considered road segment \( i \) are denoted as \( V^{i,j} \), \( D^{i,j} \), respectively. The proposed ANN-based prediction model can be formulated in Eq. (27). Here, the velocity and density of the considered road segment \( i \) at time \( t \), denoted as \( V^{k,i} \), \( D^{k,i} \), are predicted by the so-called \( \text{predict}() \) function where the current traffic state of related road segments \( (V^{k,j}, D^{k,j}) \) and the previous traffic state of the considered road segment \( (V^{i,j}, D^{i,j}) \) mentioned above are served as input parameters.

\[
(V_{\text{predict}}^{k,i}, D_{\text{predict}}^{k,i}) = \text{predict}(V^{k,j}, D^{k,j}, V^{i,j}, D^{i,j}), \quad \text{where } t < k, \ j \neq i, \ j \in V. \tag{27}
\]

One of the essential issues here is how to identify the related road segments of the considered road segment in a wide road network. In fact, not only the directly connected road segments but the ones that indirectly connect to the considered road segment may also affect the considered road segment’s traffic state. However, it is impractical to take all road segments in the whole road network into account. Therefore, the effect of related road segments should be thoroughly investigated in order to reduce the search space while keeping the accuracy of the prediction model.

Let \( j \) be a road segment that directly connects to the road segment \( i \), and \( E \) is the set of directed connect link \( ij \). The road network can be defined as a directed graph \( G = (V, E) \), where \( V \) is the set of all road segments as mentioned in Section 3. The order (level) of related road segments to the road segment \( i \) is defined in Eq. (28). The 1st order related road segments, denoted as \( N^1(i) \), are all the road segments that directly connect to the considered road segment \( i \) (i.e. \( ij \in E \) or \( ji \in E \)). The \( n \)th order related road segments, denoted as \( N^n(i) \), are recursively defined as the union of all the \( j \)'s 1st order related road segments, where \( j \) is any road segment which belongs to \( N^{n-1}(i) \). For example, in Fig. 5, \( \{c, d, e, k, l\} \) (the ones that cut the dotted circle) are the 1st order and \( \{a, b, h, j, m, g, f, o\} \) are the 2nd order related road segments of the given road segment \( i \).

\[
\begin{align*}
N^1(i) &= \{j : j \neq i, j \in V \land (ij \in E \lor ji \in E)\} \\
N^n(i) &= \bigcup_{j \in N^{n-1}(i)} N^1(j), \quad \forall j \in N^{n-1}(i).
\end{align*} \tag{28}
\]

Intuitively, the 1st order related road segments may affect traffic state of the road segment \( i \) more than those of higher order related road segments, namely, 2nd, 3rd orders, and so forth. Consequently, the effect of the \( w \)th order related road segments, namely \( N^w(i) \), on the prediction accuracy becomes minor which can be ignored when \( w \) is large enough. In this article, it is assumed that the traffic state of the road segment \( i \) is spatially independent of any related road segment whose order is equal or larger than \( w \). In addition, the ANN model “learns” the prediction rules from historical data at any time \( t \) prior to the current time \( k \). However, field studies also revealed that the relationship between the current state of the
network with 17 connected road segments, including the considered one, road traffic stated as should be evaluated. In this work, the TSF [63], was utilized to generate simulation data. A road considered road segment and its previous states becomes looser if the time \( t \) is quite far in the past (i.e., the data is quite old). Therefore, the traffic state of the road segment \( i \) is temporally independent of any state which is older than \( r \) time intervals in the past. As a result, the prediction model described in Eq. (27) can be modified in Eq. (29). This modification can significantly minimize the search space temporally and spatially. In this work, the temporal independency \( r \) is kept to 3 [29,46] when considering historical data for training and evaluating the ANN-based prediction model. The remainder of this section investigates the effect of different order related road segments on the considered road segment.

\[
\{v_{\text{pre}}^{k,i}, v_{\text{pre}}^{k,i}\} = \text{predict}(V^{k,j}, D^{i,j}, V^{i,i}, D^{i,i}), \quad \text{where } k - r \leq t < k, \ j \neq i, j \in \bigcup_{s=1}^{w=1} N^s(i). \tag{29}
\]

To investigate the effect of different order related road segments on the considered road segment, a huge amount of road traffic state data should be evaluated. In this work, the TSF [63], was utilized to generate simulation data. A road network with 17 connected road segments, including the considered one, \( r_0 \), was created. Four \( N^1(r_0) \), namely \( r_1, r_2, r_3, r_4 \), six \( N^2(r_0) \), namely \( r_5 \cdots r_{10} \), and other six \( N^3(r_0) \), namely \( r_{11} \cdots r_{16} \), related road segments were marked. A 10 h simulation was conducted in which the average velocity of all the 17 road segments (i.e., from \( r_0 \) to \( r_{16} \)) was recorded. This dataset was divided into 2 parts with the portion of 75%-25% for the training and testing datasets, respectively. To evaluate the effect of different level related road segments, data of some segments in a certain order related road segments was removed in the testing dataset to simulate the missing data in real-world applications. For example, the data of \( r_2, \ldots, r_4 \) (in \( N^1(r_0) \)) was removed to investigate the impact of the 1st order related road segments on the prediction accuracy. The impact of the combined road segments, namely \( r_1, r_2, r_3, r_2, r_3 \), etc., was also studied. The same procedure was performed for road segments in \( N^2(r_0) \) (i.e., \( r_5 \cdots r_{10} \)) and in \( N^3(r_0) \) (i.e., \( r_{11} \cdots r_{16} \)). Some of the experimental results are shown in Table 1.

Table 1 shows that the 1st order related road segments are much more effective concerning prediction accuracy compared to those in the higher order ones (i.e., \( N^2(r_0) \) and \( N^3(r_0) \)). This table presents that the missing data of the indirectly connected road segments does not significantly affect the prediction accuracy. Although all the data of the \( N^2(r_0) \) related road segments is missed, the error is still as low as 43%. Moreover, it seems that the missing data in the \( N^3(r_0) \) related road segments does not affect the prediction accuracy compared to the base-line error. It should be noted that the base-line error of the ANN-based prediction model is around 27% [29,46] when the data of all the \( N^1(r_0) \), \( N^2(r_0) \) and \( N^3(r_0) \) related road segments is available (the third row). As a brief conclusion, the data in \( N^1(r_0) \) should not be missed to guarantee the prediction effectiveness while the data in higher order, namely, in \( N^2(r_0) \) and \( N^3(r_0) \), and so forth, can be missed without any significant effect on prediction accuracy. This finding serves as an instruction in validating related road segments before predicting the traffic state of an unacceptably low penetration rate road segment using the proposed ANN-based prediction model.

### 6. Evaluation

This section evaluates the effectiveness of the proposed GA-based velocity–density inference model and the robustness of the ANN-based prediction approach. Moreover, the effect of different order related road segments on the ANN-based prediction accuracy is also confirmed using numerous simulation data.

### 6.1. Experimental environment

In this work, the TSF [63] was utilized to generate synthetic data for evaluations. In each evaluation, several road segments were selected randomly as shown in Fig. 6. For each selected road segment, two kinds of data were created concurrently as follows:

(a) The GPS data reported by individual vehicles were recorded in the “cars.csv” file. Each record contains time stamp (in seconds), road segment Id, position (longitude, latitude), current velocity, and vehicle Id of the vehicle that reported

<table>
<thead>
<tr>
<th>Table 1: The influence of different order related road segments on the prediction accuracy.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Missing data in related road segments</td>
</tr>
<tr>
<td>( N^1(r_0) )</td>
</tr>
<tr>
<td>All the data of ( N^1(r_0) ), ( N^2(r_0) ), and ( N^3(r_0) ) are available</td>
</tr>
<tr>
<td>( r_1 )</td>
</tr>
<tr>
<td>( r_1r_2 )</td>
</tr>
<tr>
<td>( r_1r_2r_3 )</td>
</tr>
<tr>
<td>( r_1r_2r_3r_4 )</td>
</tr>
<tr>
<td>( r_1r_2r_3r_4r_5 )</td>
</tr>
<tr>
<td>N/A</td>
</tr>
<tr>
<td>( r_5r_6r_7 )</td>
</tr>
<tr>
<td>( r_5r_6r_7r_8 )</td>
</tr>
<tr>
<td>N/A</td>
</tr>
<tr>
<td>( r_{11}r_{12}r_{13}r_{14}r_{15}r_{16} )</td>
</tr>
</tbody>
</table>
data. Different penetration rates, namely 10%, 20%, 25%, 30%, and so forth, were configured by the TSF. With a settled penetration rate, only such a percentage of random vehicles, namely 10% of vehicles, for example, reported data to the server. The frequency of data report timing was set to every 3 s which can mimic the GPS signal frequency.

(b) The summarized traffic state data of selected road segments were recorded in the “averageVelocity.csv” file. Each record contains information of time interval Id (in min), road segment Id, average velocity, and density. The time interval for recording 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).

6.2. Effect of low penetration rate on the estimation accuracy

As discussed in Section 3, the penetration rate significantly affects the effectiveness of velocity and density estimation. This section evaluates this effect by which the so-called acceptable penetration rate is recognized. The acceptable penetration rate is the penetration rate at which the estimation accuracy is good enough but drastically decreases if the penetration rate becomes lower, even in a small value. The recognition of acceptable penetration rate is important since it helps to ensure that the estimated traffic state information is not wrong so much and still be reliable.

To evaluate the effect of penetration rate on estimation effectiveness, several simulations were performed with different penetration rates. For each penetration rate, namely 20%, 30%, and so forth, the detailed GPS 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 using the GPS data in the “car.csv” file, applying the conventional estimation models described in Section 3 for velocity (Definition 1, Eq. (7)) and density (Definition 3, Eq. (9)), respectively. Both the average velocity and average density were estimated in every time interval, namely every minute in this work, by which they could be compared with the “actual” average velocity and density provided 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, 5 one-hour simulations with different density levels were performed. Average differences between the estimated and the “actual” average velocity were drawn out and shown in Fig. 7.

Fig. 7 shows that the lower the penetration rate is, the higher the estimation error thrown out, and vice versa. Especially, if the penetration rate becomes unacceptably low such as lower than 20%, the estimation error increases significantly. At the same time, the deviations are also high. For example, when the penetration rate is 20%, the velocity and density estimation errors are around 55% and 80%, while their deviations are around 20% (i.e. ±10%) and 12% (i.e. ±6%), respectively. The difference of the deviations for velocity and density estimations, namely \( D_v \) and \( D_d \), respectively, represents the different effect of low penetration rate on their estimation. As discussed in Section 3, density estimation is directly affected by the penetration rate, thus deviation of density estimation is more stable compared to that of the velocity estimation. The shape of this chart confirms that the density estimation error \( D_{err} \) is more sensitive to the low penetration rate than that of...
the velocity estimation error ($V_{\text{err}}$). This figure also reveals that if the expected accuracy is set to 70%, the acceptable penetration rate must be around 37%. 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 the predefined expected estimation accuracy. Furthermore, alternative solutions which can ensure the estimation accuracy at some accepted level even with low penetration rate should be investigated before bringing the mobile phone based traffic state estimation into the realization.

6.3. Effectiveness of the GA-based velocity–density estimation model

This section evaluates the effectiveness of the proposed GA-based velocity–density estimation model. The evaluation consists of two parts, namely, (1) clarifying the capacity of the proposed GA mechanism in optimizing the set of coefficients $g = \{\alpha, \beta, \gamma\}$, and thus optimizing the estimation effectiveness; and (2) evaluating the overall effectiveness of the proposed GA mechanism using a huge amount of simulation data.

6.3.1. The detailed optimization capacity

To clarify the capacity of the proposed GA mechanism, a road segment was selected randomly. It was a 2-lane, 335 m, and 85 km/h speed limit road segment. Table 2 shows 10 randomly selected records in the evaluation data generated when the penetration rate was set to 40%. As shown, $\alpha_{\text{Cir}}$ (the coefficient determined by the original velocity–density inference circuit [29,46]) falls into only one of the two values, namely 0.8 or 0.6. On the other hand, the coefficients $\alpha$, $\beta$, and $\gamma$ determined by the proposed GA model are much more granular. Consequently, the estimation accuracy of the GA approach is significantly improved compared to that of the original model in all of the cases. Both the estimation error (the last column) and the standard deviation (the last line) of the GA mechanism are significantly decreased compared to that of the original velocity–density inference circuit. Another interesting point here is that the standard deviation revealed by the GA approach is almost larger than the average estimation error. This is because the GA-based deviation itself is quite small and the estimation error in the proposed GA mechanism is definitely low.

6.3.2. The overall effectiveness of the proposed GA mechanism

To evaluate the overall effectiveness of the proposed GA-based velocity–density estimation model, the same datasets used to evaluate the effect of low penetration rate on traffic state estimation accuracy mentioned in Section 6.2 were reused. Figs. 8 and 9 show the effectiveness of the proposed GA mechanism compared to its counterparts, namely, the conventional estimation model and the original velocity–density estimation model, respectively. The velocity and density estimation errors in the GA-based mechanism are denoted as $GA_{\text{Circuit\_V}}$ and $GA_{\text{Circuit\_D}}$, respectively. These corresponding errors in the conventional model (i.e. the sensed data is applied directly to the conventional estimation model described in Eqs. (7) and (9), respectively) are denoted as $Normal\_V$ and $Normal\_D$. Similarly, $Circuit\_V$ and $Circuit\_D$ represent velocity and density estimation errors in the original velocity–density estimation circuit [29,46]. Both the two figures show that among of these approaches, the proposed GA-based velocity–density inference model is prominent. In all of the cases when the penetration rate is low but still relevant, namely higher than 30%, the estimation model is completely optimized so that the estimation errors are almost lower than 5%. In addition, the proposed GA mechanism also reduces the acceptable penetration rate required for the expected accuracy of 80% from 50% (in the normal approach) or from 30% (in the original velocity–density estimation model) to 25%. This improvement seems to be minor in terms of numeric figure but actually it is important in practice since the system will be more tolerant with low penetration rates, and thus be more scalable.

Table 2

<table>
<thead>
<tr>
<th>The actual facts extracted from the sensed data</th>
<th>The original velocity–density inference model</th>
<th>The GA-based velocity–density estimation model</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D_{\text{sensed}}$</td>
<td>$V_{\text{sensed}}$</td>
<td>$V_{\text{actual}}$</td>
</tr>
<tr>
<td>2.18</td>
<td>82.42</td>
<td>79.57</td>
</tr>
<tr>
<td>3.67</td>
<td>79.64</td>
<td>75.38</td>
</tr>
<tr>
<td>6.42</td>
<td>54.56</td>
<td>64.84</td>
</tr>
<tr>
<td>15.42</td>
<td>23.97</td>
<td>43.65</td>
</tr>
<tr>
<td>26.25</td>
<td>14.43</td>
<td>24.3</td>
</tr>
<tr>
<td>28.92</td>
<td>9.24</td>
<td>19.61</td>
</tr>
<tr>
<td>36.00</td>
<td>8.67</td>
<td>13.27</td>
</tr>
<tr>
<td>38.83</td>
<td>10.25</td>
<td>13.89</td>
</tr>
<tr>
<td>37.83</td>
<td>11.54</td>
<td>11.77</td>
</tr>
<tr>
<td>35.92</td>
<td>9.21</td>
<td>11.06</td>
</tr>
<tr>
<td>AVG</td>
<td>17.78</td>
<td>AVG</td>
</tr>
<tr>
<td>STDEV</td>
<td>13.62</td>
<td>STDEV</td>
</tr>
</tbody>
</table>

Please cite this article in press as: T.M. Quang, et al., Synergistic approaches to mobile intelligent transportation systems considering low penetration rate, Pervasive and Mobile Computing (2012), doi:10.1016/j.pmcj.2012.07.008
6.4. Effectiveness of the ANN-based prediction model

This section evaluates the effectiveness of the ANN-based prediction model. Five road segments were randomly selected as the desired ones whose average velocity and density are required to be predicted. In addition, the three different, namely the 1st, 2nd, and 3rd, order related road segments of each desired road segment were also identified so that their corresponding traffic state data (average velocity and density) were recorded. For each desired road segment $i$, 101 h simulations were performed by which the average velocity and density of not only road segment $i$ but also of any related road segment $j$ ($j \in \bigcup_{i=1}^{3} N(i)$) were obtained. Concretely, at each time interval $k$, the velocity and density of the considered road segment $i$ and its related road segment $j$, denoted as $V^{k;i}$, $D^{k;i}$, $V^{k;j}$, $D^{k;j}$, respectively, were recorded. In addition, regarding the temporally independent concept discussed in Section 5, $r (r = 3$ in this work) previous velocities and densities of road segment $i$, namely $V^{t-r;i}$, $D^{t-r;i}$ ($k - r \leq t < k$) were also extracted. Finally, each record in the evaluation dataset consists of (1) $V^{k;i}$, $D^{k;i}$ serving as the target elements (the outputs of the ANN model); and (2) $V^{k;j}$, $D^{k;j}$, $V^{t-i}$, $D^{t-i}$ ($k - r \leq t < k, j \in \bigcup_{i=1}^{3} N(i)$) serving as the inputs for the ANN model as mentioned in Eq. (29). The summarized traffic state information mentioned above was recorded in each minute, thus in each simulation (1 h), 60 data patterns were generated. Therefore, totally a dataset of 600 patterns (in 101 h simulations) was created for each desired road segment. This dataset was then divided into two parts as the portion of 75% and 25% for the training and the testing datasets, respectively. An ANN with 6 hidden nodes were trained and evaluated by the training and testing datasets mentioned above. Here, the average predicted errors for both the velocity and density estimations of the 5 aforementioned randomly selected road segments were calculated to evaluate the effectiveness of the ANN-based prediction model. The results of these evaluations are shown in Fig. 10 and described as follows.

Fig. 10 reveals that in the ANN-based prediction method, errors for both the velocity, denoted as $ANN_V$, and density, denoted as $ANN_D$, are around 27% regardless of the penetration rate. This figure also shows that even though the ANN-based prediction model properly works without any requirement on the penetration rate, its accuracy is limited as around 73% (the prediction error is about 27%). Therefore, this model is suited in the cases of unacceptably low penetration rate,
namely lower than the critical one (i.e. around 20%). This point is considered as the “even” point of the GA-based prediction model. That means, when the penetration rate is relevant, namely larger than the critical one, the GA-based velocity–density estimation approach (denoted as \( \text{GA}_\text{Circuit}_V \) and \( \text{GA}_\text{Circuit}_D \) for the velocity and density estimation errors, respectively) is dominant.

6.5. Effect of related road segments on prediction accuracy

To evaluate the effect of different order/level related road segments on the accuracy of the ANN-based prediction model, the dataset of the 17 related road segments mentioned in Section 5 (Table 1) was reused. The ANN model was trained by the training dataset. The effect of missing data in each order related road segments represents the effect of the corresponding order related road segments on the prediction effectiveness. Therefore, the testing dataset (25% of the original dataset) was modified to imitate the missing data in different order related road segments as follows. The data of each order, namely 1st, 2nd and 3rd order related road segments was randomly removed by a so-called “data missing rate”, namely 10%, 20%, and so forth, before applying to the trained ANN. Concretely, considering the missing data rate in the 1st order related road segments as 10%, for instance, 10% of the data in \( N_1(r_0) \) (i.e. data in road segments \( r_1, r_2, r_3, r_4 \)) of the original testing dataset was randomly removed. This “new” dataset was applied to the trained ANN prediction model to estimate the average velocity of the considered road segment, \( r_0 \). This process was repeated with different data missing rates, namely 20%, 30%, and so forth and the estimation errors were recorded. Similar processes were performed with different data missing rates set to the data in \( N^2(r_0) \) and \( N^3(r_0) \), respectively. It should be noted that the data missing rate was set to \( N^1(r_0) \), \( N^2(r_0) \) and \( N^3(r_0) \) independently. That means, when the data missing rate was set to \( N^1(r_0) \), all the data in other sets, namely in \( N^2(r_0) \) and \( N^3(r_0) \) remained. The effect of different order related road segments on the prediction accuracy is depicted in Fig. 11.

Fig. 11 confirms that the data in the 1st order related road segments play an important role in predicting traffic state of the considered road segment. The prediction error, denoted as \( \text{Err}_L1 \), increases drastically even if the data missing rate in \( N^1(r_0) \) slightly increases. If this missing rate increases to 10% the error increases from 27% (the base-line error of the ANN-based prediction model [29,46] which was confirmed in Fig. 10) to 35%. When the missing rate slightly increases to 20% the error will be unacceptably as high as 43%. On the other hand, the estimation error seems to be stable, or just slightly changes with the data missing rate in both \( N^2(r_0) \) and \( N^3(r_0) \). Moreover, the prediction error corresponding to the missing data in \( N^3(r_0) \), namely \( \text{Err}_L3 \), seems to be minor (around 30%) compared to the base-line error (27%) of the ANN-based prediction model. This figure reveals that, practically, the traffic state of the 1st order related road segments should be assured to be available before applying the ANN-based prediction model. Meanwhile, the availability of the traffic state of the 2nd and
3rd related road segments may help to improve the accuracy of the ANN-based prediction model but the absence of such data will not affect the prediction effectiveness as much.

7. Conclusions and future work

This paper proposed an appropriate GA mechanism to optimize the velocity–density estimation model. This approach not only significantly improves the effectiveness of the traffic state estimation model but also reduces the requirement for the critical penetration rate, thus enhancing the reliability as well as the scalability of the traffic state estimation system. This work also introduced a notable ANN-based prediction model to deal with the issues of an unacceptably low or unknown penetration rate. In addition, the effectiveness of the ANN prediction model with regard to the missing rate of traffic state data in different level related road segments was also thoroughly discussed. The experimental results reveal that the ANN-based prediction model can ensure the prediction error as low as around 30% if traffic state data of the 1st order related road segments is available. Meanwhile, the availability of the traffic state of higher order, namely 2nd, 3rd orders, related road segments may help to improve the accuracy of the ANN–based prediction model but the absence of such data just slightly affects the prediction effectiveness. This discovery is practically useful since it provides instructions in verifying the related road segments’ traffic state data to ensure the accuracy of the ANN-based prediction model. In addition, following this discovery, the search space can be reduced significantly by considering only the 1st order related road segments in estimating unacceptably low or unknown penetration rate road segments without compromising the prediction accuracy.

As shown in the evaluation section, the proposed GA-based mechanism can optimize the performance of the velocity–density estimation model when the penetration rate is low but still relevant, namely larger than the critical penetration rate (20% in this work). However, when the penetration rate becomes unacceptably low the GA-based mechanism cannot work properly. In this case, the ANN-based prediction model would be an appropriate alternative. Concretely, the GA-based velocity–density estimation model should be employed when the penetration rate is relevant, while the ANN-based prediction model is suited in the case of unacceptably low penetration rate. Nevertheless, an essential issue here is that there is no way to obtain the “actual” penetration rate at the estimation time. Therefore, selecting the “right” estimation model (i.e. the ANN-based prediction method or the GA-based estimation model) is still a challenge which is deferred to the future work. Another interesting thing here is that even though the ANN approach can properly work without any requirement on the penetration rate of the considered road segment, its accuracy is limited to around 73% (the prediction error is around 27%). Therefore, combining both the GA-based velocity–density estimation model and the ANN-based prediction method would be a reasonable approach to improve the effectiveness of the whole traffic state estimation model. This is also a relevant research direction which should be thoroughly considered in future work.

References


W. Schneider, E. Makotsky, Mobile phones as a basis for traffic state information, Proceeding of Intelligent Transportation Systems (2005).


G. Pengfei, W. Xuezhi, H. Yingshi, The enhanced genetic algorithms for the optimization design, in: 3rd International Conference on Biomedical Engineering and Informatics, BMEI, Yantai, China, October 2010, pp. 2989–2994.


