Transportation Research Part C 71 (2016) 108–121

Contents lists available at ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

Multi-step prediction of experienced travel times using agent-based modeling ☆

Hao Chen [a], Hesham A. Rakha [b,*]

[a] Virginia Tech Transportation Institute, 3500 Transportation Research Plaza, Blacksburg, VA 24061, United States
[b] Charles E. Via, Jr. Department of Civil and Environmental Engineering, Virginia Polytechnic Institute and State University, 3500 Transportation Research Plaza, Blacksburg, VA 24061, United States

ARTICLE INFO

Article history:
Received 15 May 2014
Received in revised form 9 April 2016
Accepted 15 July 2016
Available online 25 July 2016

Keywords:
Experienced travel time
Travel time prediction
Agent-based model
Agent interaction rule
Probe data

ABSTRACT

This paper develops an agent-based modeling approach to predict multi-step ahead experienced travel times using real-time and historical spatiotemporal traffic data. At the microscopic level, each agent represents an expert in a decision-making system. Each expert predicts the travel time for each time interval according to experiences from a historical dataset. A set of agent interactions is developed to preserve agents that correspond to traffic patterns similar to the real-time measurements and replace invalid agents or agents associated with negligible weights with new agents. Consequently, the aggregation of each agent's recommendation (predicted travel time with associated weight) provides a macroscopic level of output, namely the predicted travel time distribution. Probe vehicle data from a 95-mile freeway stretch along I-64 and I-264 are used to test different predictors. The results show that the agent-based modeling approach produces the least prediction error compared to other state-of-the-practice and state-of-the-art methods (instantaneous travel time, historical average and *k*-nearest neighbor), and maintains less than a 9% prediction error for trip departures up to 60 min into the future for a two-hour trip. Moreover, the confidence boundaries of the predicted travel times demonstrate that the proposed approach also provides high accuracy in predicting travel time confidence intervals. Finally, the proposed approach does not require offline training thus making it easily transferable to other locations and the fast algorithm computation allows the proposed approach to be implemented in real-time applications in Traffic Management Centers.

1. Introduction

Tackling congestion (both recurrent and non-recurrent) has proven to be a challenge for highway agencies. Adding capacity in response to congestion is becoming less of an option for these agencies due to a combination of financial, environmental, and social issues. Therefore, the main focus has been on improving the performance of existing facilities through continuous monitoring and dissemination of traffic information. The minimum that can be accomplished is to inform the public or, specifically, the potential users of what they should expect on the roadways before and during their trips. Additionally, this information can be applied to provide alternatives to users so that they may make informed decisions

☆ This article belongs to the Virtual Special Issue on: Agent-Based Modeling.
* Corresponding author.
E-mail addresses: hchen@vtti.vt.edu (H. Chen), hrakha@vt.edu (H.A. Rakha).

http://dx.doi.org/10.1016/j.trc.2016.07.004

about their trips. This is the essence of Advanced Traveler Information System (ATIS) applications such as 511 that have been implemented nationwide. In many states, relevant traffic information is also posted on variable message signs (VMSs) that are strategically positioned along highways. Consequently, there is a need to provide predicted travel times to road users for better planning their trips and choosing their route of travel, further reducing congestion.

Various traffic sensing technologies have been used to collect traffic data for use in computing travel times, including point to point travel time collection (e.g. license plate recognition systems, automatic vehicle identification systems, mobile, Bluetooth, probe vehicle, etc.) and station based traffic state measuring devices (e.g. loop detector, video camera, remote traffic microwave sensor, etc.). Private companies such as INRIX integrate different sources of measured data to provide section-based traffic speed or travel time, which can be used to construct traffic speed matrix over spatial and temporal and thus is used in this paper. The benefit of using temporal-spatial speed data is that travel time can be easily estimated afterward (van Lint and van der Zijpp, 2003). More importantly, such data provides the flexibility for scalable applications on traffic networks. By providing section-based traffic state data, generally there are two approaches to compute travel time depending on the trip experience, which are instantaneous and experienced travel time (Mazaré et al., 2012; Tu, 2008).

Previous research has demonstrated that prediction accuracy typically deteriorates quickly with the increase in prediction horizon (Chen et al., 2012). In order to demonstrate the discrepancy between instantaneous and experienced travel times, especially the errors of using instantaneous information for multi-step prediction of experienced travel time, a spatiotemporal traffic speed data provided by INRIX is presented in Fig. 1. The traffic data was collected along I-64 from Richmond to Norfolk during afternoon peak hours on June 22, 2013. The trip trajectories are plotted on the contour speed map. According to the black trajectory, the instantaneous travel time is calculated as 40 min for time interval at 4 p.m. Although the traffic on the selected route is uncongested at 4 p.m., two bottlenecks rapidly form afterward. Consequently, the instantaneous travel time at 4 p.m. underestimates the experienced travel time by 28 min, 50 min and 60 min for the prediction horizon of 0 min, 30 min and 60 min, respectively. These results demonstrate that the instantaneous travel time may not be a good predictor of experienced travel time, especially for multi-step prediction.

During the past decades, many studies have been conducted attempting to predict travel times. According to the manner of modeling, these methods can be classified into parametric methods (e.g. linear regression models (Rice and Van Zwet, 2004; Zhang and Rice, 2003), Kalman filter methods (Chen and Chien, 2001; Chien and Kuchipudi, 2003; Nanthawichit et al., 2003), Auto-Regressive Integrated Moving Average (ARIMA) models (Billings and Yang, 2006; Guin, 2006; Xia et al., 2011)) and non-parametric methods (e.g. *k*-Nearest Neighbor (*k*-NN) (Bustillos and Chiu, 2011; Myung et al., 2011; Qiao et al., 2012), artificial neural network (ANN) models (Lint et al., 2005; Park and Rilett, 1998; van Lint, 2006) and support vector regression (SVR) methods (Vanajakshi and Rilett, 2007; Wu et al., 2004)). These techniques are implemented through direct or indirect procedures to predict travel times using different types of state variables (Chen and Rakha, 2013). Travel time is directly used as the state variable in parametric or non-parametric methods to predict travel times. Indirect procedures are performed using other variables (such as traffic speed, density, flow, and occupancy) as the state variable

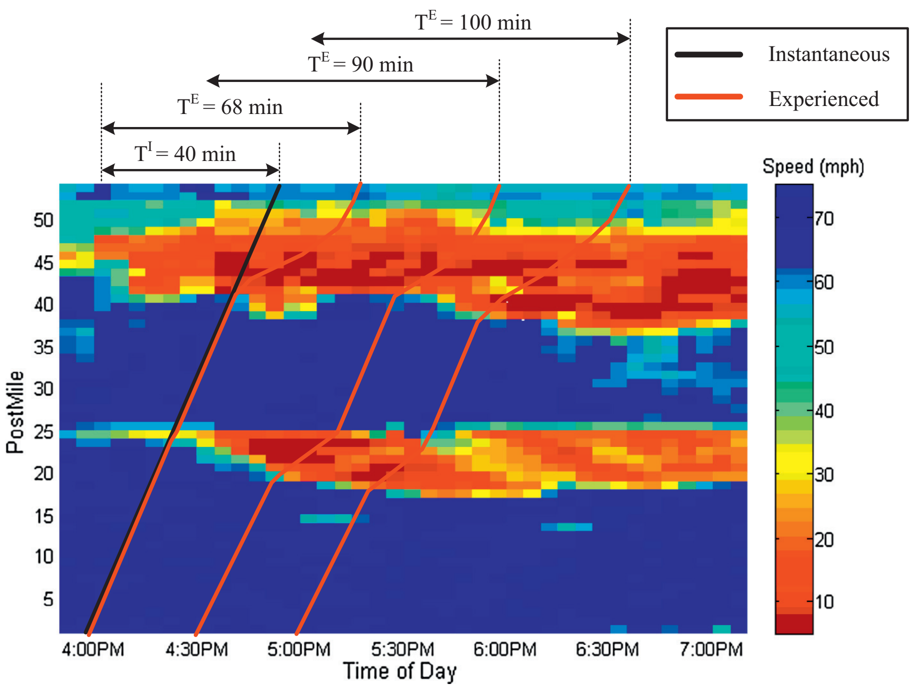

Fig. 1. Spatiotemporal traffic speed map and trip trajectories on I-66 (June 22, 2013).

to predict the future traffic speed over space and time, and then travel times can be calculated based on the spatiotemporal speed map (van Lint and van der Zijpp, 2003). This paper attempts to predict experienced travel times for departures at current or future time intervals. For real-time application, instantaneous travel time can be obtained as the summation of section travel times at every time interval. Nevertheless, experienced travel time can only be obtained after the completion of the trip, because the spatiotemporal evolution of speed should be considered. In this case, the experienced travel time for the previous time interval usually is not available for predicting travel time in the next interval, especially for long trips. Consequently, many existing methods cannot work well for predicting experienced travel times (Yildirimoglu and Geroliminis, 2013).

Other than real-time information, historical data provide a pool of experienced traffic patterns that can be used to predict travel times. ANN methods are widely used to generate the predictor by training on a large historical dataset. However, the same problem exists, namely the prediction accuracy deteriorates rapidly for multi-step predictions (Boné and Crucianu, 2002; Parlos et al., 2000). Considering the stochastic nature of traffic behavior, it is very difficult to predict travel time for multi-step time horizons accurately. For instance, a time-delayed state-space neural network (TDSSNN) approach was recently developed and demonstrated to outperform other popular ANN methods for travel time prediction. However, the prediction error for the proposed TDSSNN method on incident-free data increased from 5.4% to 15.1% for a prediction horizon of 5 min to 25 min (Zeng and Zhang, 2013). In addition, there are several other deficiencies for ANNs, such as high computational costs for data the training process, a lack of the flexibility to deal with non-recurrent traffic patterns, and difficulty to implement on large-scale traffic networks or different sites. Consequently, there is a need to develop a robust method for multi-step prediction of experienced travel time, yet is still easily transferable to other sites without the need for a data training process.

Considering the aforementioned problems, the concept of agent-based modeling is used in this paper to address travel time prediction problems. Agent-based modeling has been widely used for problems of decision making, complex social system and, etc. (Bui and Lee, 1999). The advantages lie on the feature that each agent can behave as an individual expert decision system, so that individual agent has the ability to analyze data input and produce its own decision output by constructing rules. More importantly, different groups of agents can cooperate to model complex social systems. In the past decades, agent-based modeling has been successfully applied to various transportation problems, because of the flexibility and computational advantages of modeling complex transportation systems (Zhang and Levinson, 2004). Although the direct application to predict travel time using agent-based modeling has not been developed yet, several similar applications have already been attempted to deal with time series prediction problems in other application fields. For instance, a group of individual cooperating agents are used to simulate different components of the stock trading process and tested to provide accurate prediction for stock buying/selling decisions (Luo et al., 2002). Similar approaches have been developed to predict the evolution of market shares for electric vehicles (Shafiei et al., 2012) and the price change of the US wholesale power market (Sueyoshi and Tadiparthi, 2008). These examples demonstrate the agent-based modeling methods can efficiently and accurately solve time-series prediction problems in complex systems. It should be noted that the state-transition of travel times over neighboring time intervals also has strong nonlinear trends as the aforementioned problems in other fields. In addition, a set of guidelines to use agent-based models for data forecasting problems are developed in (Hassan et al., 2013), and the related problems of building a predictor using agent-based model for different categories of forecasting problems have been discussed. Consequently, there is a need to explore the potential of using the concept of agent-based model for travel time prediction.

In this paper, an agent-based modeling approach is developed to predict experienced travel times using real-time and historical traffic data. At the microscopic level, each agent acts as an expert and a set of agent interactions are developed to produce a recommendation for future experienced travel time with a measurement of recommendation confidence. Consequently, the aggregation of each agent's recommendation (predicted travel time with associated weight) provides a macroscopic level of output – a predicted travel time distribution. The INRIX probe data from Richmond to Virginia Beach along I-64 and I-264 in 2012 are used to test the performance of the proposed method. The results show that the agent-based modeling approach produces the least prediction error compared with other state-of-practice and state-of-art methods (instantaneous travel time, historical average and *k*-nearest neighbor), and maintains less than 9% error for future trip departures from the current time to 60 min later.

The remainder of this paper is organized as follows. The framework of the proposed agent-based modeling method is provided together with descriptions of the microscopic agent interaction rules. This is followed by an implementation on a selected test site and a comparison with other predictors to estimate experienced travel times considering different prediction horizons (0–60 min). The last section includes the summary conclusions of the proposed method and recommendations for future research.

2. Agent-based model

The concept of correlating real-time and historical traffic measurement data using an agent-based model to predict travel time and the details of agent interactions are described in this section.

In this research, we assume that the traffic speed data for each time interval is updated along all roadway segments from the trip origin to destination. In this way, the daily traffic measurement data can be represented as a matrix, in which each cell is an average speed for the corresponding time interval and roadway segment. Here, different colors are used to represent the speed value - the dark blue denotes free flow speed and the bright red corresponds to congestion. Therefore, the traffic data matrix is demonstrated as a color map. At the same time, the experienced travel time can be computed by providing the spatiotemporal traffic speed map (van Lint and van der Zijpp, 2003). Consequently, a traffic speed map and an experienced travel time curve are included for each day.

An illustration of the agent-based modeling approach is presented in Fig. 2. Each agent corresponds to a specific time interval on a historical day. In this example, i and j are the day and time interval indices of the sample agent. Assume the current traffic pattern for the testing day is denoted by the speed matrix from time $t - L + 1$ to t across all the segments. Thus the agent is used to provide a prediction of experienced travel time at $t + p$. The prediction result includes a value of travel time T^E and a corresponding weight value w. The former value is obtained by finding the experienced travel time at time interval $j + p$ on historical day i. The latter value is calculated by comparing the dissimilarity between two matrixes relative to the current day and the historical day, represented by dotted rectangle windows in Fig. 2. The details of the framework of the proposed agent-based modeling approach and the interactions between agents are described below.

2.1. The framework of agent-based modeling approach

From a traditional expert system perspective, each expert makes a recommendation based on its own experience of the target problem. The proposed agent-based model adopts the same logic in order to predict travel times using real-time and historical traffic data. Each agent represents an expert, who is responsible for providing a travel-time prediction estimate at each time interval. The framework of the proposed agent-based model is presented in Fig. 3. At a microscopic level, each agent interacts individually according to the real-time and historical traffic status. Different interaction rules are constructed in order to simulate the process of choosing and updating individual experts according to its performance (similarity to real-time traffic information) for each time interval. The aggregation of each agent's recommendation (predicted travel time with associated weight) provides a macroscopic level of output – predicted travel time distribution.

Assume the current time is t, the available measurement data $u_{Nseg \times L}$ is the speed matrix from short past $t - L + 1$ to t along all the freeway segments (total segment number is N_{seg}). Here, the speed matrix is denoted by the tail time as variable z_t, which also represents the real-time traffic status and is updated every time interval. The real-time traffic status and historical data will be used to conduct a data mining process to predict the experienced travel time T^E_{t+p} which departs at time $t + p$. Each agent represents an expert who can provide a prediction estimate based on the experience of a specific historical day. Consequently, the ith agent, denoted by $x_t^{(i)}$, corresponds to a day index $d_t^{(i)}$ from the historical dataset Ω and a time index $j_t^{(i)}$ on that day. The corresponding speed matrix from time interval $d_t^i - L + 1$ to d_t^i along N_{seg} segments can be obtained

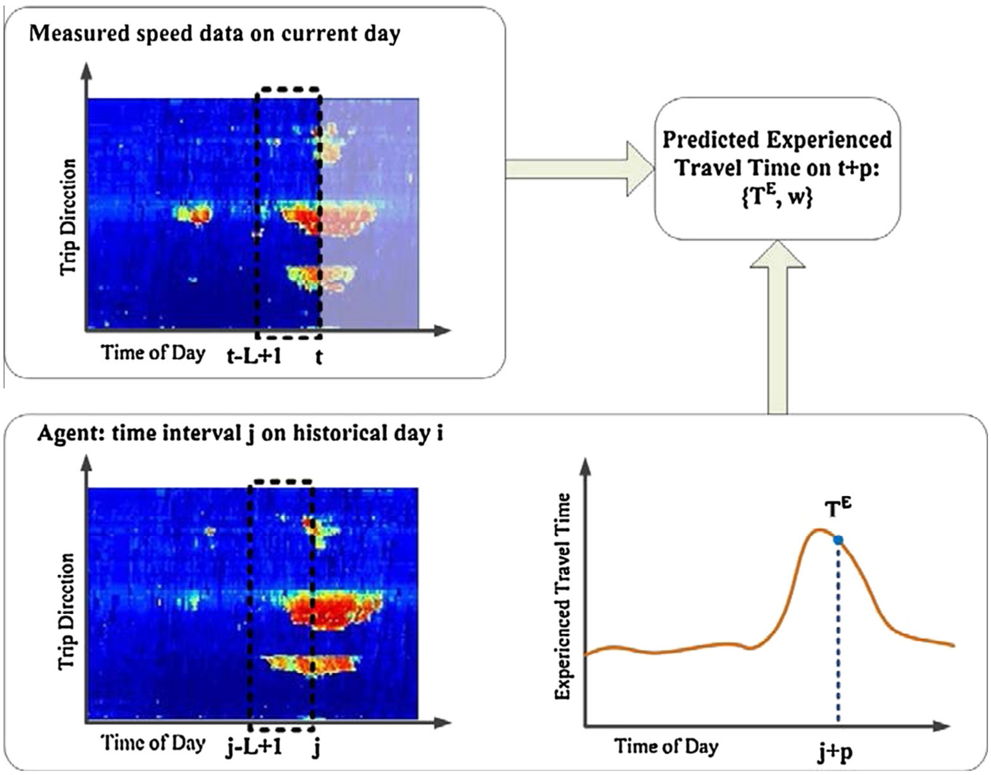

Fig. 2. Illustration of travel time prediction by a single agent.

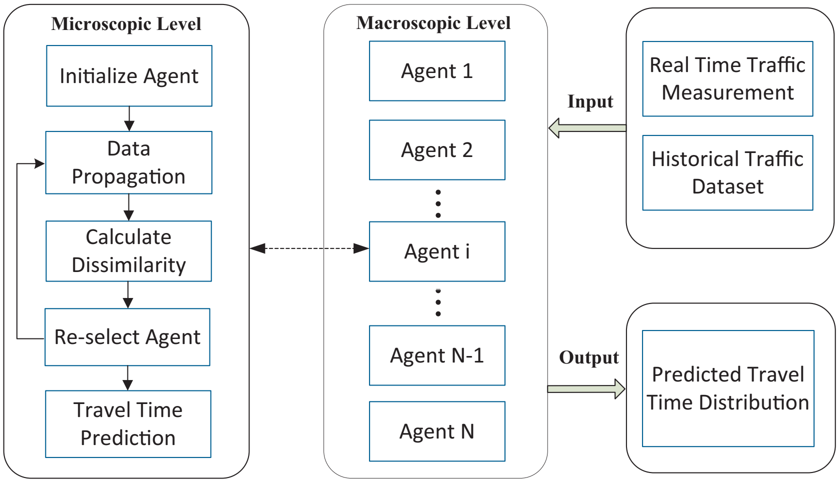

Fig. 3. The proposed agent-based model framework.

as $\Omega(d_t^i, j_t^i)$. The difference between two speed matrices from real-time measurement z_t and historical experience $\Omega(d_t^i, j_t^i)$ corresponding to the ith agent can be used to calculate the confidence level of an agent, denoted by weight $w_t^{(i)}$. In addition, the experienced travel time can be calculated in the historical dataset, so that the ith agent can produce a recommendation of travel time $T^E(d_t^i, j_t^i + p)$ which departures at time $j_t^i + p$ for historical day d_t^i. Finally, the prediction output is denoted by the integration of each agent's recommendation with the corresponding travel time $T^E(d_t^i, j_t^i + p)$ and weight $w_t^{(i)}$. The details of agent interaction rules are presented as blow.

2.2. Initialize agent

At beginning, each agent should be assigned an initial correspondence on the historical dataset. The index of day $d_0^{(i)}$ is randomly selected from the historical dataset Ω (a total of D days) and then the index of time interval $j_0^{(i)}$ is randomly selected for that day. The initial agent set x_0 can be represented as

$$x_0 : \left\{ x_0^{(i)} \middle| x_0^{(i)} = \Omega\left(d_0^{(i)}, j_0^{(i)}\right), i \in [1:N] \right\} \tag{1}$$

2.3. Data propagation

In order to match with the new incoming measurement data, the corresponding speed matrix for each agent needs to propagate along the time domain. This process is conducted by maintaining the same day index and increasing the time index by an additional time interval as

$$d_t^{(i)} = d_{t-1}^{(i)}, \quad j_t^{(i)} = j_{t-1}^{(i)} + 1, \quad i \in [1:N] \tag{2}$$

In the proposed algorithm, each daily traffic data is considered as a separate dataset from the adjacent days, even though the end of one day is followed by the start of next day. The reason for using each daily traffic dataset separately is based on two considerations. First, the adjacent day's traffic data may not be available in the historical dataset. Second, the measured traffic data may not provide full coverage for 24 h. For instance, it is possible that the traffic data is only measured during the day time or peak hours. Consequently, a process to identify valid agents is developed to examine if the data propagation reaches the boundary of the same day. Here, the last time interval of the historical traffic speed matrix on day $d_t^{(i)}$ is denoted by $Hd_t^{(i)}$. Considering the prediction horizon p, the collection of the valid agent is identified as

$$\Psi_t = \{i | j_t^{(i)} \leqslant H_{d_t^{(i)}} - p, i \in [1:N]\} \tag{3}$$

2.4. Calculate dissimilarity measure

This process aims to calculate the weight of each valid agent and then find the top N_{th} number of agents associated with the largest weight values. The average absolute error between the speed matrices for the current and historical times is computed using Eq. (4) to represent the dissimilarity $s_t^{(i)}$ between the current traffic status z_t and each valid agent $x_t^{(i)}$. The small value of dissimilarity represents the data matrices are more similar to each other. Here, a likelihood function which follows a Gaussian distribution $N(\mu, \sigma^2)$ is used to transfer the value of dissimilarity into weight $w_t^{(i)}$ as Eq. (5).

$$s_t^{(i)} = \left| z_t - x_t^{(i)} \right| \Big/ (L \times N_{seg}), \quad i \in \Psi_t \tag{4}$$

$$w_t^{(i)} = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{\left(s_t^{(i)} - \mu\right)}{2\sigma^2} \right\}, \quad i \in \Psi_t \tag{5}$$

Thereafter, the value of the weight for each valid agent is sorted in descending order and the top N_{th} number of agents with large weight values are maintained to use in the next iteration. This process is described in Eq. (6), in which the index of preserved agents is denoted by j and $j = 1{:}N_{th}$. In this way, the agents are divided into two groups. The first group includes the agents with large weights. The second group includes the invalid agents that cannot provide prediction values (exceed data boundary) or the agents with negligible weights. The second group of agents will be re-selected in the next process so that new agents with similar traffic status to the current time interval can be selected.

$$x_t^{(j)} = x_t^{(i)}, \quad w_t^{(j)} = w_t^{(i)}, \quad when\ i = \underset{i \in \Psi_t}{\arg\max}\ w_t^{(i)}, \quad \Psi_t = \Psi_t - \{i\} \tag{6}$$

2.5. Re-select agent

Considering the top N_{th} numbers of valid agents with large weights are maintained but the rest $N - N_{th}$ number of agents are disregarded, a re-selection algorithm is developed here to fill the gap with agents associated with similar traffic patterns to the current traffic speed matrix. Here, each historical day can be selected to represent the new agent. Therefore, the probability to select each historical day is calculated. Firstly, the index of time interval which corresponds to the traffic speed matrix with minimum dissimilarity to the current traffic status is computed as σ_t^n for each historical day n by Eq. (7). Thereafter, the dissimilarity between current traffic status z_t and the historical speed matrix $\Omega(n, \sigma_t^n)$ is calculated and then the same likelihood function as Eq. (5) is used to obtain the selection probability λ_t^n of historical day n by Eq. (8).

$$\sigma_t^n = \underset{k \in [L, H_n - p]}{\arg\min} (z_t - \Omega(n, k)), \quad n \in [1 : D] \tag{7}$$

$$\lambda_t^n = p_{e_t}\left(z_t - \Omega\left(n, \sigma_t^n\right)\right), \quad n \in [1 : D] \tag{8}$$

After the above calculation, the remaining $N - N_{th}$ number of agents (the index is denoted by $j = N_{th} + 1{:}N$) can be re-selected according to the probability of λ_t^n, which represents the re-selection probability of historical day n under the condition of current time t. Therefore, the corresponding traffic speed matrix and weight can be located according to Eq. (9).

$$\begin{gathered} d_t^{(j)} = \left[n | P(n|t) = \lambda_t^n \right], \quad n \in [1 : D] \\ x_t^{(j)} = \Omega\left(d_t^{(j)}, \sigma_t^{d_t^{(j)}} \right), \quad w_t^{(j)} = \lambda_t^{d_t^{(j)}} \end{gathered} \tag{9}$$

2.6. Travel time prediction

Finally, the total N number of agents are located for time interval t. Since each agent corresponds to a historical day with a certain time index, the ith agent can produce a recommendation of travel time $T^E(d_t^{(i)}, j_t^{(i)} + p)$ which departures on time $j_t^{(i)} + p$ at historical day $d_t^{(i)}$. Therefore, the predicted travel time distribution can be obtained by aggregating the recommendations from all agents as Eq. (10). And the average predicted value is calculated by the weighted average travel time using Eq. (11).

$$T_{t+p}^E = \left\{ T^E\left(d_t^{(i)}, j_{t+p}^{(i)}\right), w_t^{(i)} \right\}, \quad i \in [1 : N] \tag{10}$$

$$\overline{T_{t+p}^E} = \sum_{i=1}^{N} w_t^{(i)} \cdot T^E\left(d_t^{(i)}, j_{t+p}^{(i)}\right) \Big/ \sum_{i=1}^{N} w_t^{(i)}, \quad i \in [1 : N] \tag{11}$$

3. Case study

3.1. Test environment

A freeway stretch from Richmond to Virginia Beach (95 miles long) connected by I-64 and I-264 is selected as the test site in this study. This test site usually experiences high traffic volumes and serious congestion during the summer season, since Virginia Beach is a famous tourist destination and the selected freeway stretch serves as the main route heading to the beaches. The evaluation of travel time prediction on the test site is conducted based on probe data from INRIX. The data provided by INRIX are mainly collected by GPS-equipped vehicles and supplemented with traditional road sensor data, as well as mobile devices and other sources (INRIX, 2012). The probe data on the test site covers 96 freeway segments with a total length of 95 miles. The average segment length is 0.65 miles long, and the length of each segment is unevenly divided in the raw data from 0.1 to 6.36 miles. The location of the study site and the deployment of segments are presented in Fig. 4. The raw data provides the average speed for each segment and is collected at one-minute interval. However, the raw data cannot be directly used considering the issues of geographically inconsistent segments, irregular time intervals of data collection, missing data and measurement error over spatial and temporal. In order to transfer raw data into daily spatiotemporal speed matrices, the data reduction processes are conducted to solve the above mentioned problems. Initially, the daily speed data are sorted by time and location from the raw data into a 2-dimensional matrix. In order to reduce the stochastic noise and measurement error, the speed matrices are aggregated by five-minute intervals. The missing data in the aggregated speed matrices are estimated using the moving average from a 3 by 3 window. In this way, the daily spatiotemporal speed matrices are generated and the corresponding instantaneous and experienced travel times can be calculated. More detailed information for the data reduction can be found in (Rakha et al., 2013). In this study, the aggregated traffic data from May 16, 2012 to September 15, 2012 and the corresponding afternoon time periods between 2 p.m. and 8 p.m. for each day are considered to test the algorithm, since most congested periods are observed during this time frame.

Given the fact that the limited summer data (totally 123 days) on the selected freeway stretch are used in this study, the leave-one-out cross-validation method is considered to quantify the prediction accuracy. Leave-one-out cross-validation is a classic model validation technique for assessing how the results of a statistical analysis will generalize to an independent data set. This method has also been used in many application fields including traffic prediction problems (Kwon et al., 2000). It should be noted that the data of previous two weeks are required for historical average method. Therefore, the leave-one-out testing starts from May 30 to September 15, 2012 (totally 109 days). The testing is conducted by using four different prediction methods for each day and the remaining 122 days serve as the historical dataset. Finally, the average performance over the 109 test days is used to compare the prediction accuracy of different methods. The parameters in the proposed method are pre-defined for the test. The width of the matching window to measure the dissimilarity between historical and real-time traffic status is chosen to be 6 time intervals (30 min), which entails the traffic speed matrix along all freeway segments over half an hour being used as an input variable. The number of agents N is selected to be 100. The re-selection threshold N_{th} is 80. The likelihood function to calculate the agent weight factor follows a Gaussian distribution $N(0,2)$. It should be noted that a sensitivity analysis is conducted in the case study to quantify the impacts of various parameters on the algorithm performance.

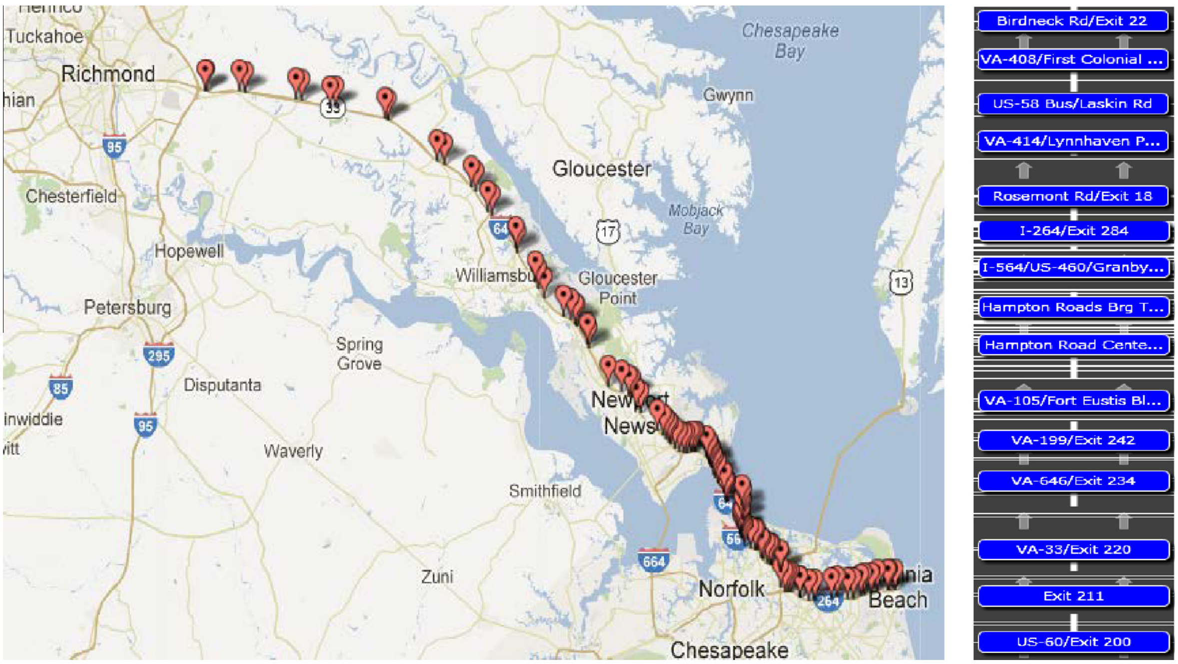

Fig. 4. The selected freeway stretch from Richmond to Virginia Beach.

3.2. Comparison methods and performance indices

In order to evaluate the performance of the proposed predictor, three other prediction methods are also tested on the same dataset. The instantaneous travel time method is the easiest alternative to predict future travel times by assuming the current traffic speed along all the segments will remain constant until the completion of the trip as Eq. (12). This method is currently used by the Virginia Department of Transportation (VDOT) to display travel time information on variable message signs. Therefore, instantaneous travel times are considered as the state-of-practice method and used to quantify the tradeoff between simplicity and prediction accuracy.

$$T^E_{t+p} = T^I_t \tag{12}$$

Historical average data is a common indicator for recurrent traffic conditions and can also be used for travel time prediction. A previous study demonstrates that simple historical average of long term periods (e.g. one year) has large variations, and is not a good predictor for experienced travel times (Yildirimoglu and Geroliminis, 2013). Therefore, the historical average from the past two weeks is used as an alternative. Since the current study only includes the summer season dataset from May 16 to September 15, 2012 and the variations of travel times between different seasons are non-existent, the historical average is anticipated to provide better prediction accuracy when compared to instantaneous methods. It should be noted that the traffic pattern on Monday and Friday are usually different from other weekdays (Rakha and Van Aerde, 1995), therefore we divide daily traffic data into four groups - Monday, Friday, weekday (Tuesday, Wednesday and Thursday), weekend (Saturday and Sunday). The average experienced travel times at the same time interval from the same group of days in the previous two weeks, which is denoted by, are used as the historical average to predict travel times on the current day using Eq. (13).

$$T^E_{t+p} = \overline{T}^E_{t+p} \tag{13}$$

The k-NN is a widely used state-of-art method for real-time travel time prediction problems (Bustillos and Chiu, 2011; Myung et al., 2011; Qiao et al., 2012). In order to conduct an objective comparison between k-NN and the proposed approach, the same spatial-temporal traffic speed matrix $u_{Nseg \times L}$ from short past to current time along all the freeway segments is used for k-NN. Thereafter, m numbers of similar data matrix with tail time $\{h_1, h_2, \ldots, h_m\}$ can be selected as the candidates from the historical dataset Ω as given in Eq. (14). For each candidate with index i, a weight w_i is calculated using the normalized matching error. Moreover, the corresponding experienced travel time departure at h_{i+p} on the selected candidate day i can be obtained from the historical data set. Consequently, the experienced travel time T^E_{t+p} on the current day can be predicted as the weighted average of the travel times from all candidates using Eq. (15).

$$\begin{aligned} & H_c = \{h_1, h_2, \ldots, h_m\} \\ & For\ i = 1 : m \\ & \qquad h_i = \arg\min_{h \in \Omega} |z_t - h|, \quad \Omega = \Omega - h_i \\ & End\ For \end{aligned} \tag{14}$$

$$w_i = |z_t - h_i| \Big/ \sum_{i=1}^{m} |z_t - h_i|$$

$$T^E_{t+p} = \sum_{i=1}^{m} T^E_{h_i+p} \cdot w_i \tag{15}$$

Different combinations of parameters were tested and the optimum parameters were selected as $L = 6$ and $m = 20$, which corresponds to the least prediction error (Qiao et al., 2012). These parameters are used to test the k-NN method using the same data set and to serve as a comparison with other methods. It should be noted that the k-NN method is different from the proposed approach even if the number of candidates in the k-NN method is equal to the number of particles. The reason lies in the fact that there is no data propagation process in the k-NN algorithm, and all candidates are blindly selected from each time interval based on its similarity measure (shortest Euclidean distance) to the current travel time sequence.

Both relative and absolute prediction errors are used to evaluate the performance of predictors. The absolute error is denoted by the mean absolute error (MAE) using Eq. (16), which represents the average absolute deviations between the predicted and the ground truth values. The corresponding relative error is represented by the mean absolute percentage error (MAPE) of Eq. (17), which denotes the absolute proportional deviations between the predicted and the ground truth values.

$$MAE = \frac{1}{I \times J} \sum_{j=1}^{J} \sum_{i=1}^{I} \left| y_i^j - \hat{y}_i^j \right| \tag{16}$$

$$MAPE = \frac{100}{I \times J} \sum_{j=1}^{J} \sum_{i=1}^{I} \frac{\left| y_i^j - \hat{y}_i^j \right|}{y_i^j} \tag{17}$$

where J is the total number of days in the testing dataset (109 days in our case study); I is the total number of time intervals in one day (i.e., 72 intervals occurring every five minutes between 2 p.m. and 8 p.m.); and y_i^j and $\hat{y}_i^j$ denote the ground truth and the predicted value, of the experienced travel time for the ith time interval on the jth day in the testing dataset.

3.3. Test results

The prediction results for four methods are summarized in Table 1. The least prediction errors are located at the bottom of the table using the proposed agent-based method (ABM). The relative absolute errors for the four predictors are presented in Fig. 5. The figure clearly demonstrates a significant degradation in the prediction accuracy for longer prediction horizons for the instantaneous and historical average methods. By comparing the historical average method with the k-NN method, the latter method outperforms the former for short-term prediction (prediction horizons from 0 to 30 min) and historical average method outperforms the k-NN method for longer prediction horizons ranging from 40 to 60 min. Moreover, the MAPE by the k-NN method increases from 9.24% to 13.18% (43% increase), which is higher than the error associated with the proposed method from 6.75% to 8.57% (25% increase). More importantly, the relative errors produced by the proposed ABM approach is always below 9% for the different prediction horizons within one hour, which indicates the prediction performance of the proposed method is much more reliable compared to the other three methods. Based on the observations of daily prediction results, the historical average and instantaneous methods produce large variations in performance compared to the k-NN and ABM methods. The historical average method can accurately predict travel times in recurrent days, but the performances for non-recurrent days are very low. The instantaneous method works well for uncongested days, but produces large errors for congested days. Comparatively, both of k-NN and ABM methods generate reliable performance on different days, however the ABM method outperforms the k-NN method especially for long prediction horizons (e.g. 60 min).

The daily average prediction errors by the ABM method on June 2012 are presented in Fig. 6. Generally, no obvious trends are observed in the 30 days-worth of results. Specifically, occasionally the algorithm performs better on weekdays and occasionally it performs better on weekends. Maybe this is caused by the fact that special conditions such as inclement weather or incidents are not filtered from our dataset. In order to further analyze the prediction errors for different methods, the predicted travel time curves for the four methods are compared with the ground truth data for two sample weekdays, as

Table 1
Prediction results by different methods.

		Prediction horizon (min)						
		0	10	20	30	40	50	60
Instantaneous	MAE (min)	11.52	13.06	14.40	15.78	17.10	18.28	19.38
	MAPE (%)	10.64	12.12	13.47	14.85	16.12	17.29	18.31
Historical average	MAE (min)	13.01	13.01	13.01	13.01	13.01	13.01	13.01
	MAPE (%)	11.46	11.46	11.46	11.46	11.46	11.46	11.46
k-NN	MAE (min)	10.48	11.12	12.10	12.84	13.62	14.24	15.06
	MAPE (%)	9.24	9.95	10.68	11.31	11.98	12.61	13.18
ABM	MAE (min)	7.69	7.92	8.14	8.33	8.62	8.97	9.49
	MAPE (%)	6.75	6.98	7.21	7.53	7.86	8.18	8.57

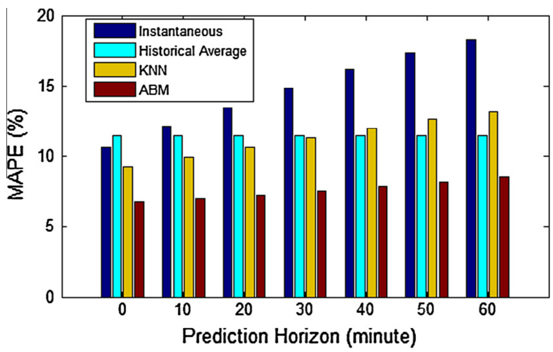

Fig. 5. MAPEs using four predictors under various prediction horizons.

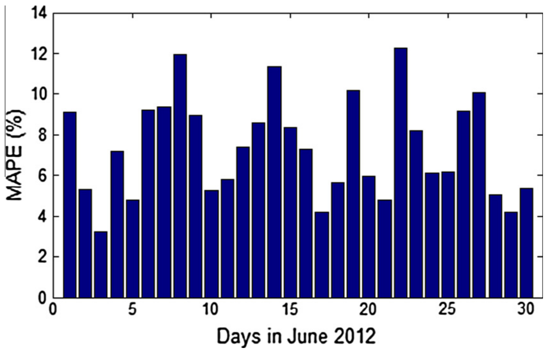

Fig. 6. Daily prediction errors by ABM in June 2012.

illustrated in Fig. 7. Both the instantaneous and the *k*-NN method experience some time lag relative to the ground truth data, especially during the onset and dissipation of congestion. Specifically, the instantaneous method highly underestimates the travel time when congestion is forming, and overestimates the travel time when congestion is dissipating. It should be noted that the historical average method overestimates the congestion for June 21 between 3 and 5 p.m., and underestimates the ground truth data when congestion is dissipating. According to ground truth data on June 21, the congestion forms between 2 and 4 p.m. and dissipates between 4 and 8 p.m., and the corresponding average MAPEs by the ABM are 3.8% and 5.6%, respectively. On the other hand, the prediction errors using the historical average method are even worse by highly overestimating the travel times on June 29. However, the proposed method improves the prediction performance by producing

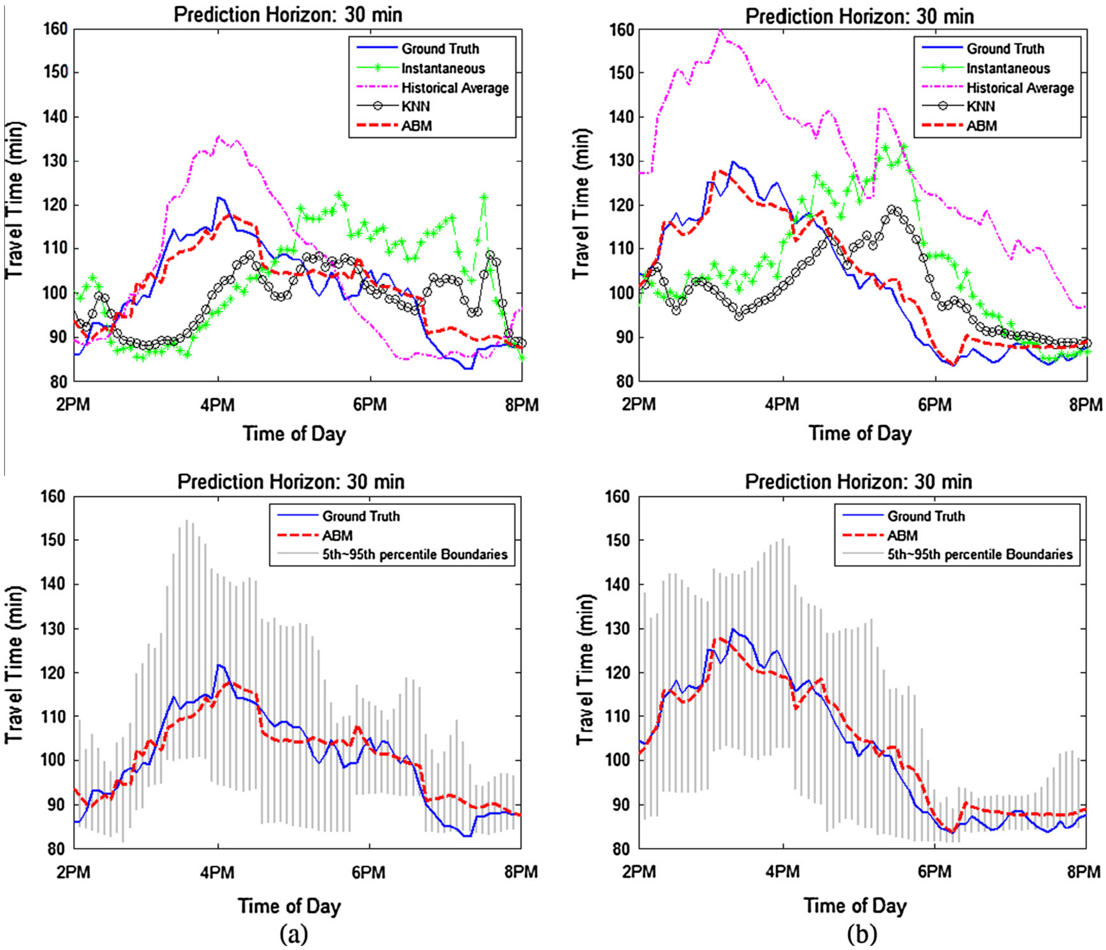

Fig. 7. Travel time prediction results by four methods and the confidence intervals by ABM on (a) June 21, 2012 (Thursday); (b) June 29, 2012 (Friday).

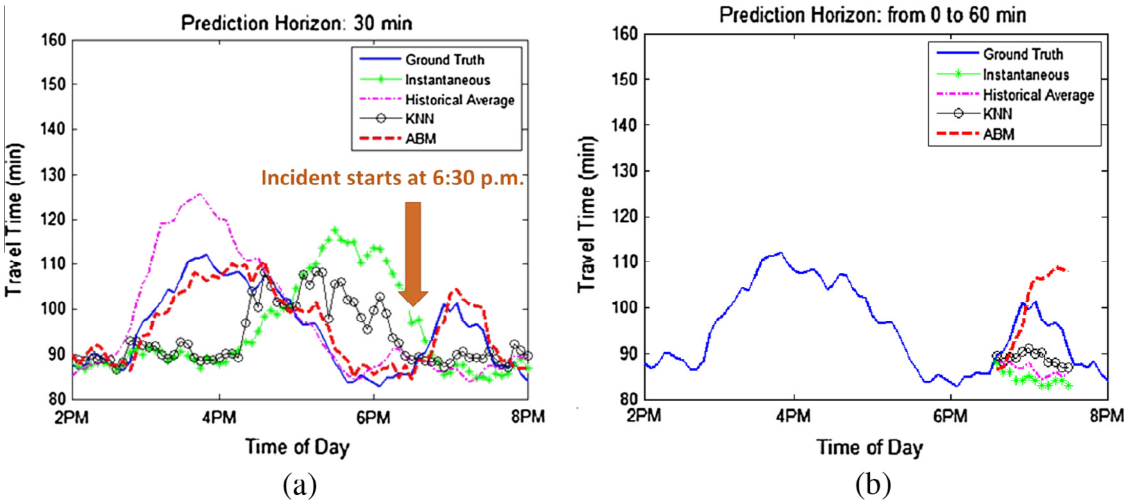

Fig. 8. Illustration of travel time prediction when incident occurs on June 11, 2012; (a) Prediction results by four methods; (b) Prediction up to 60 min by four methods.

3.3% and 4.5% of errors during congestion forming (2–3:30 p.m.) and dissipation (3:30–8 p.m.) periods. Moreover, the proposed method provides confidence intervals of predicted travel times. The 5th and 95th percentile of the predicted travel times are selected as the upper and lower boundaries in Fig. 7. The gray shadow area between the boundaries covers most of the ground truth data temporal variation, which demonstrates that the proposed approach provides good accuracy in predicting travel time confidence intervals. There is a trade-off between the width of the confidence intervals and the accuracy of prediction results. Consequently, the use of a narrow confidence interval does not represent higher prediction accuracy, since the ground truth curve has less chance to be included within the confidence interval. More tests of different upper and bottom percentiles are needed to find the optimum choice of confidence intervals with narrow coverage but also providing dependable predictions to cover most of the ground truth curve.

During the case study tests, the days with and without traffic incidents are mixed together in both the historical and test datasets. In order to demonstrate the performance of predictors when incidents occur, a special day with an incident is selected, and the travel time prediction results using the four methods are illustrated in Fig. 8. An incident occurs at the location directly upstream of the Hampton Roads Bridge Tunnel between 6:30 and 7:30 p.m. on June 11, 2012. The proposed ABM method outperforms the other three methods by producing the least errors to ground truth data as shown in Fig. 8 (a). When the incident occurs, the multi-step predicted travel times from prediction horizon 0 to 60 min using the four methods are presented in Fig. 8(b). The traffic congestion builds up quickly at 6:30 p.m., thus the ground truth travel time also increases fast. Apart from the ABM method, all the other methods cannot capture the growing trend and the predicted travel times deviate from the ground truth data. Although there is a slight delay, the ABM method can still predict the congestion-forming trend caused by the incident and thus it produces much higher prediction accuracy than other methods.

The prediction accuracy of the agent-based method is highly related to the number of historical days. In order to investigate this relationship, we run multiple times of tests using the same amount of test days and different amount of historical days. During these tests, the same parameters in the initial test are used and the prediction horizon is 30 min. The original 123 days of data are divided into two groups, in which the test dataset include the last two weeks (14 days) of data during September 2–15, 2012 and the historical dataset include the rest 109 days of data. The tests use different numbers of historical days including 20, 40, 60, 80, 100 and 109 days. For the tests with historical days less than 109, a bootstrap resampling strategy is used to select historical days (Varian, 2005). For instance, if the test uses 60 historical days, we randomly select 60 days from the total 109 historical days to run the test, and then the same test is repeated 20 times and each time a new subset of 60 days are randomly generated. Eventually the average result from the 20 repetitions is calculated to capture the performance of the agent-based method using 60 historical days. Fig. 9 presents the prediction accuracy using different number of historical days. Generally, the prediction error decreases as the number of historical days increases, since more historical days means that more traffic patterns are included in the learning dataset to improve the prediction accuracy. The test results indicate that a 40-day historical dataset can still produce an acceptable prediction accuracy with a relative error of 9.5%.

3.4. Sensitivity analysis

A sensitivity analysis is conducted to quantify the impact of parameter L and agent number N on the prediction accuracy of the proposed agent-based method. The parameter L determines the matching window width in the agent propagation and re-selection processes, so a larger value of L results in a wider matching window and vice versa. Here, the agent number

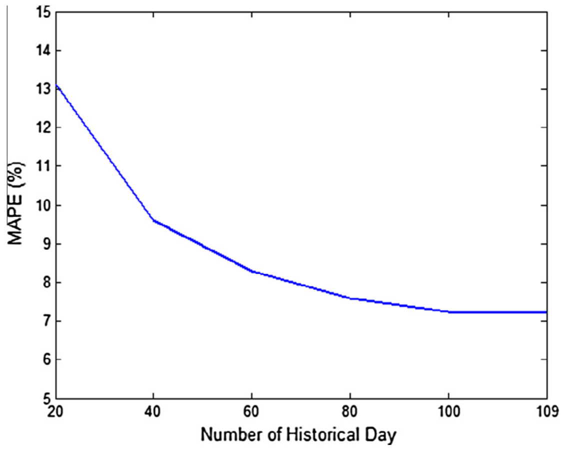

Fig. 9. Compare prediction accuracy using different number of historical days.

maintains the same value of 100 as in the previous testing and the values of L vary from 2 to 10 time intervals to calculate the average MAPE for different prediction horizons, as presented in Fig. 10(a). Although the least MAPE for 60 min prediction horizon corresponds to the L value of 10, the optimum L value is selected by considering the best performance for prediction horizon from 0 to 60 min. The minimum errors for prediction horizon from 0 to 50 can be obtained when L equal to 6, so 6 time intervals is selected as the optimum value of L. Consequently, the best matching window is a half-hour of traffic speed data along all the freeway segments. Moreover, under the condition that L value is 6, different agent numbers are also investigated to calculate the relative errors as shown in Fig. 10(b). Generally, the prediction error decreases with larger agent number. However, the error reaches the minimum value when agent number is 100, and then prediction error slightly increases with agent number greater than 100. Therefore, the agent number of 100 is the best choice to predict travel times in the case study. The same analysis can be conducted on different sites or roadway compositions to find the optimum model parameters.

3.5. Computation speed

In order to investigate the potential to use the proposed method for real-time application, the computation speed of the algorithm on the cast study needs to be calculated. The testing of the ABM travel time predictor was performed on a personal computer with Intel dual core CPU, 2.40 GHz and 4 GB of random-access memory within the MATLAB 2012b environment. It should be noted that the procedure of re-selecting agents in ABM method is very time consuming since many iterations are needed to calculate the probabilities of selecting each historical day. Considering matrix computation is very fast in MATLAB software, the iterations in Eq. (7) to compare the dissimilarity between current traffic pattern and historical traffic pattern on each time interval was coded by matrix calculation to expedite computation speed. Under the scenario of setting $L = 6$,

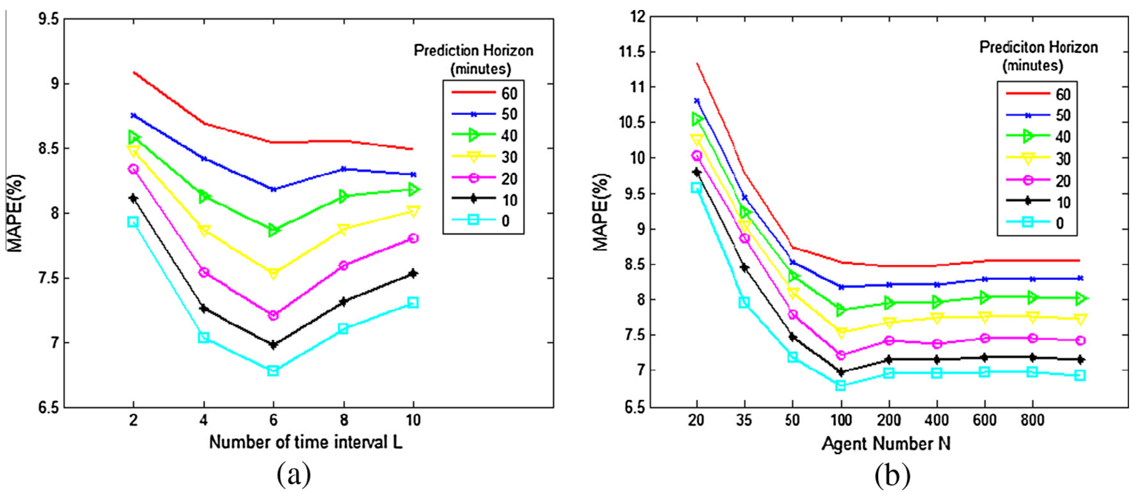

Fig. 10. Sensitivity analysis of ABM.

N_{th} = 80 and N = 100, the total average computation time for one day between 2 and 8 p.m. was 6.20 s. Totally 72 time intervals are included in this time period, so the calculation of a single prediction by every 5-min only requires 0.086 s. Therefore, the computational efficiency allows the agent-based model approach to be implemented in real-time in Traffic Management Center (TMC).

4. Conclusions

In this paper, an agent-based modeling approach is proposed that performs multi-step travel time predictions. Although agent-based models have been widely used in various transportation problems, the algorithm developed in this paper is the first attempt to use the concept of agent-based modeling to predict travel times. Similar to the traditional expert decision systems, each agent represents an expert and can provide a recommendation of future experienced travel time. At the same time, a set of agent interaction rules are developed to update agents for future time intervals and also calculate the weight of each agent, which reflects the confidence level of the predicted travel time value. In this way, the average predicted travel time and the corresponding confidence boundaries can be calculated by the algorithm. A 95-mile freeway stretch from Richmond to Virginia Beach along I-64 and I-264 is used to test the performance of the proposed method. The test results indicate that the proposed ABM produces the least absolute and relative error when compared to instantaneous travel time, the historical average and the *k*-NN methods, and maintains less than a 9% prediction error for trip departures up to 60 min later. Moreover, the confidence boundaries of the predicted travel times indicate that the proposed approach also provides high accuracy in predicting travel time confidence intervals. In addition, a sensitivity analysis was conducted to quantify the impact of different model parameters to the prediction accuracy. Lastly, the computation time of the proposed algorithm was tested and the results demonstrates the ABM can provide accurate and efficient travel time predictions for real-time applications.

Although probe data from the private sector are used in the case study, the proposed method is data source independent as long as the spatiotemporal speed measurements are available. For instance, loop detector data, Bluetooth or cell phone data can also be used in the proposed method. Considering the proposed predictor provides more than 90% accuracy in predicting travel times with departures up to 60 min into the future, the proposed agent-based prediction algorithm can be extended to make a recommendations on the optimum departure time in addition to providing the expected travel time. In the future, we could also categorize historical days by different conditions such as weather, incident, and holiday/special event. Using these categories the algorithm can use a sub-dataset with to predict travel times in each category to improve prediction accuracy.

Acknowledgements

This research effort was funded partially by Virginia Department of Transportation (VDOT) and partially by the Mid-Atlantic Universities Transportation Center (MAUTC). The authors also acknowledge Catherine McGhee, Ralph L. Jones and Philomena Lockwood for their assistance and feedback.

References

Billings, D., Yang, J.-S., 2006. Application of the ARIMA models to urban roadway travel time prediction – a case study. In: IEEE Conference on Systems, Man and Cybernetics. IEEE, pp. 2529–2534.
Boné, R., Crucianu, M., 2002. Multi-step-ahead prediction with neural networks: a review. 9emes rencontres internationales: Approches Connexionnistes en Sciences 2, 97–106.
Bui, T., Lee, J., 1999. An agent-based framework for building decision support systems. Decis. Support Syst. 25 (3), 225–237.
Bustillos, B.I., Chiu, Y.-C., 2011. Real-time freeway-experienced travel time prediction using N-curve and k nearest neighbor methods. Transport. Res. Rec.: J. Transport. Res. Board 2243 (2011), 127–137.
Chen, H., Rakha, H.A., 2013. Forecasting freeway dynamic travel times by constructing trip trajectories. In: Transportation Research Board 92nd Annual Meeting Washington D.C.
Chen, H., Rakha, H.A., Sadek, S.A., Katz, B.J., 2012. A particle filter approach for real-time freeway traffic state prediction. In: Transportation Research Board 91st Annual Meeting, Washington D.C.
Chen, M., Chien, S.I.J., 2001. Dynamic freeway travel-time prediction with probe vehicle data: link based versus path based. Transport. Res. Rec.: J. Transport. Res. Board 1768, 157–161.
Chien, S.I.-J., Kuchipudi, C.M., 2003. Dynamic travel time prediction with real-time and historic data. Transport. Res. Rec.: J. Transport. Res. Board 129 (6), 608–616.
Guin, A., 2006. Travel time prediction using a seasonal autoregressive integrated moving average time series model. In: IEEE Conference on Intelligent Transportation Systems. IEEE, pp. 493–498.
Hassan, S., Arroyo, J., Galán, J.M., Pavón, J., 2013. Asking the oracle: introducing forecasting principles into agent-based modelling. J. Artif. Soc. Soc. Simulat. 16 (3), 13.
INRIX, 2012. <http://www.inrix.com/trafficinformation.asp>.
Kwon, J., Coifman, B., Bickel, P., 2000. Day-to-day travel-time trends and travel-time prediction from loop-detector data. Transport. Res. Rec.: J. Transport. Res. Board 1717 (1), 120–129.
Lint, J.W.C.V., Hoogendoorn, S.P., Zuylen, H.J.v., 2005. Accurate freeway travel time prediction with state-space neural networks under missing data. Transport. Res. Part C: Emer. Technol. 13 (5–6), 347–369.
Luo, Y., Liu, K., Davis, D.N., 2002. A multi-agent decision support system for stock trading. IEEE Netw. 16 (1), 20–27.
Mazaré, P.-E., Tossavainen, O.-P., Bayen, A., Work, D.B., 2012. Trade-offs between inductive loops and GPS vehicles for travel time estimation: a mobile century case study. In: Transportation Research Board 91st Annual Meeting, Washington, D.C.

Myung, J., Kim, D.-K., Kho, S.-Y., Park, C.-H., 2011. Travel time prediction using k nearest neighbor method with combined data from vehicle detector system and automatic toll collection system. Transport. Res. Rec.: J. Transport. Res. Board 2256, 51–59.
Nanthawichit, C., Nakatsuji, T., Suzuki, H., 2003. Application of probe-vehicle data for real-time traffic-state estimation and short-term travel-time prediction on a freeway. Transport. Res. Rec.: J. Transport. Res. Board 1855, 49–59.
Park, D., Rilett, L.R., 1998. Forecasting multiple-period freeway link travel times using modular neural networks. Transport. Res. Rec.: J. Transport. Res. Board 1617 (1), 163–170.
Parlos, A.G., Rais, O.T., Atiya, A.F., 2000. Multi-step-ahead prediction using dynamic recurrent neural networks. Neural Netw. 13 (7), 765–786.
Qiao, W., Haghani, A., Hamedi, M., 2012. Short term travel time prediction considering the weather impact. In: Transportation Research Board 91st Annual Meeting, Washington D.C.
Rakha, H.A., Chen, H., Haghani, A., Farokhi, K., 2013. Assessment of data quality needs for use in transportation applications. US Department of Transportation, ISBN-13: 978-1495413780.
Rakha, H.A., Van Aerde, M., 1995. Statistical analysis of day-to-day variations in real-time traffic flow data. Transport. Res. Rec.: J. Transport. Res. Board 1510, 26–34.
Rice, J., Van Zwet, E., 2004. A simple and effective method for predicting travel times on freeways. IEEE Trans. Intell. Transport. Syst. 5 (3), 200–207.
Shafiei, E., Thorkelsson, H., Ásgeirsson, E.I., Davidsdottir, B., Raberto, M., Stefansson, H., 2012. An agent-based modeling approach to predict the evolution of market share of electric vehicles: a case study from Iceland. Technol. Forecast. Soc. Change 79 (9), 1638–1653.
Sueyoshi, T., Tadiparthi, G.R., 2008. An agent-based decision support system for wholesale electricity market. Decis. Support Syst. 44 (2), 425–446.
Tu, H., 2008. Monitoring Travel Time Reliability on Freeways. Department of Transport and Planning, Technische Universiteit Delft.
van Lint, J.W.C., 2006. Reliable real-time framework for short-term freeway travel time prediction. J. Transport. Eng. 132 (12), 921–932.
van Lint, J.W.C., van der Zijpp, N.J., 2003. Improving a travel time estimation algorithm by using dual loop detectors. Transport. Res. Rec.: J. Transport. Res. Board 1855 (1), 41–48.
Vanajakshi, L., Rilett, L.R., 2007. Support vector machine technique for the short term prediction of travel time. In: IEEE Proceedings of Intelligent Vehicles Symposium. IEEE, pp. 600–605.
Varian, H., 2005. Bootstrap tutorial. Math. J. 9 (4), 768–775.
Wu, C.-H., Ho, J.-M., Lee, D.T., 2004. Travel-time prediction with support vector regression. IEEE Trans. Intell. Transport. Syst. 5 (4), 276–281.
Xia, J., Chen, M., Huang, W., 2011. A multistep corridor travel-time prediction method using presence-type vehicle detector data. J. Intell. Transport. Syst.: Technol. Plan. Oper. 15 (2), 104–113.
Yildirimoglu, M., Geroliminis, N., 2013. Experienced travel time prediction for congested freeways. Transport. Res. Part B: Methodol. 53, 45–63.
Zeng, X., Zhang, Y., 2013. Development of recurrent neural network considering temporal-spatial input dynamics for freeway travel time modeling. Comput.-Aided Civ. Infrastruct. Eng. 28 (5), 359–371.
Zhang, L., Levinson, D., 2004. Agent-based approach to travel demand modeling: exploratory analysis. Transport. Res. Rec.: J. Transport. Res. Board 1898 (1), 28–36.
Zhang, X., Rice, J.A., 2003. Short-term travel time prediction. Transport. Res. Part C: Emer. Technol. 11 (3), 187–210.