Front. Earth Sci. 2018, 12(2): 253–263
https://doi.org/10.1007/s11707-016-0634-8

RESEARCH ARTICLE

Trip-oriented travel time prediction (TOTTP) with historical vehicle trajectories

Tao XU[1,2], Xiang LI (✉)[1,2], Christophe CLARAMUNT[3]

1 Key Laboratory of Geographic Information Science (Ministry of Education), East China Normal University, Shanghai 200241, China
2 School of Geographic Sciences, East China Normal University, Shanghai 200241, China
3 The French Naval Academy Research Institute, Lanveoc-Poulmic, BP 600, 29240 Brest Naval, France

Abstract Accurate travel time prediction is undoubtedly of importance to both traffic managers and travelers. In highly-urbanized areas, trip-oriented travel time prediction (TOTTP) is valuable to travelers rather than traffic managers as the former usually expect to know the travel time of a trip which may cross over multiple road sections. There are two obstacles to the development of TOTTP, including traffic complexity and traffic data coverage. With large scale historical vehicle trajectory data and meteorology data, this research develops a BPNN-based approach through integrating multiple factors affecting trip travel time into a BPNN model to predict trip-oriented travel time for OD pairs in urban network. Results of experiments demonstrate that it helps discover the dominate trends of travel time changes daily and weekly, and the impact of weather conditions is non-trivial.

Keywords trip-oriented travel time prediction (TOTTP), urban network, Back Propagation Neural Networks (BPNN), historical vehicle trajectories

1 Introduction

For road transportation, accurate travel time prediction is undoubtedly of importance to both traffic managers and travelers, and in the meantime, predicting travel time is never a straightforward task considering the dynamics of traffic situations. Various factors affecting travel time of an individual vehicle may be from traffic infrastructure (e.g., speed limit, network geometry, traffic signals, etc.), traffic flow consisting of all road users, driving behaviors of its driver, or external conditions (e.g., weather, special events, etc.). In free or near-free traffic flow, travel time is mostly determined by the traffic infrastructure which has been planned or programmed, and thus it is possible to obtain reliable prediction results. However, it is not the case in highly-urbanized area, where interrupted traffic flow is the most popular. The factors from traffic flow become dominant and the combined influence among different factors further complicates the prediction.

To improve the quality of travel time prediction for interrupted traffic flow, many research efforts have been made and a detailed review is given in the next section. In terms of the target of prediction, most existing research has focused on isolated traffic environments without interference from intersections, such as an inter-urban freeway or an express overhead road network, because it is easier to collect traffic data and model various factors in a closed and limited system than an open system (Vlahogianni et al., 2004). However, the results are valuable to traffic managers rather than travelers as the latter usually expect to know the travel time of a trip which may cross over multiple road sections (Jiang and Li, 2013). It is usually inapplicable to simply summarize the travel time on individual road sections as the prediction result of a trip unless the influence of intersections can be accurately estimated (Chen and Chien, 2001; Bhaskar et al., 2011), which is always regarded as the most complicated and challenging issue in travel time prediction. As a result, more efforts are still needed to tackle trip-oriented travel time prediction (TOTTP).

TOTTP aims at providing travel time forecasting to travelers for a coming trip between any two points on a road network, which might be scheduled to start off in one minute, next day, or even on a more distant date. There might be two obstacles to the development of TOTTP, including the complexity of representing all influence

Received September 25, 2016; accepted December 6, 2016

E-mail: xli@geo.ecnu.edu.cn

factors in an open traffic environment and the deficiency of real-time or historical traffic data covering multiple road sections or even the whole road network (Vlahogianni et al., 2004).

Benefiting from the dramatic advances of location-based services, the efficiency of traffic data collection is being significantly improved through extracting from widespread vehicle trajectories. Compared to conventional traffic data sources, vehicle trajectories are direct reaction to traffic situations and contain detailed route choice information. As long as enough vehicles can be employed as data providers, a complete picture of a traffic network can be obtained from both spatial and temporal perspectives. Therefore, the data challenge of travel time prediction might be solved to some extent.

Generally, a trip between an origin-destination (OD) pair contains multiple road segments or road links. Most previous research works about TOTTP are interested in modeling traffic flow model by link attributes (speed limit, functional class, etc.) and use the linear or nonlinear combination of link travel time and intersection delays to represent trip travel time, while ignoring part of trip conditions, such as route, date, weather, etc., (Stathopoulos and Karlaftis, 2003; Lee et al., 2009; Li et al., 2011; Jiang and Li, 2013) which compromises the accuracy of prediction (Wu et al., 2004).

In view of this, this research attempts to explore the probability of applying large volume vehicle trajectory data to TOTTP. And based on this, a novel prediction approach is presented. In this approach, besides traffic flow information from vehicle trajectory data, trip-based factors including meteorological conditions, travel date, travel route, etc. are also taken into account. To deal with the complexity of modeling these factors, instead of an exact regression model, a Back Propagation Neural Networks (BPNN) model is adopted. Compared with statistical models, the modeling process in BPNN is more direct, because there is not a specific mathematical relationship between the input and output variables. BPNN can be effective for analyzing such a system containing these complex factors, to establish patterns and characteristics (Goh, 1995). The factor of individual drivers is not included in this model, and without loss of generality, average driving behavior is assumed. Vehicle trajectory data from more than 800 taxis covering a road network of about 371 km^2 in Minhang Distinct of Shanghai, China for more than 13 months are employed to examine the proposed approach. The results show that massive vehicle trajectory data can well feed the proposed BPNN model, and the precision of the proposed approach can be achieved in tolerable range.

The main novelty of this research is that we demonstrate the potential of large amount of historical vehicle trajectory data in travel time prediction to develop a TOTTP approach through integrating multiple factors into a BPNN model. Furthermore, the predicted results based on the proposed approach may implicate the dominate trends of travel time changes along specific routes.

The remainder of this paper is organized as follows. The next section reviews related research. Section 3 presents the BPNN-based prediction approach. Section 4 validates the approach through a series of experiments. And section 5 concludes the paper.

2 Review

In recent years, many research efforts have been made on travel time prediction, and most of them have focused on a closed and relatively isolated traffic system (Mori et al., 2015), such as freeway or highway segments (Chen and Chien, 2001; Dia, 2001; van Lint et al., 2005; Chen et al., 2010; Fei et al., 2011; Dong et al., 2014; Guo et al., 2014), or simple arterial road segments (Lin et al., 2004; Abu-Lebdeh and Singh, 2011). However, in view of the numerous demands from travelers, advanced traffic management systems (ATMS) and advanced traveler information systems (ATIS) service providers normally disseminate average traffic time at link or section level (Long Cheu et al., 2002), and assume that trip travel time is the addition of travel time on its consisting links (Chen and Chien, 2001). Without considering signal and intersection delays, the results are not accurate especially during peak hours.

The main challenges of TOTTP for urban networks are that the acquisition of data of complete spatiotemporal coverage and the construction of the model for all influence factors (Vlahogianni et al., 2004). And then some research attempt to model trip-oriented travel time by complex traffic flow models (Long Cheu et al., 2002; Stathopoulos and Karlaftis, 2003; Hofleitner et al., 2012; Zhan et al., 2013), but drawbacks, such as lacking massive raw data, inaccurately predicting on longer routes, etc., still exist (Jenelius and Koutsopoulos, 2013).

Since a large number of vehicle trajectory data is available, a new opportunity is provided to travel time prediction. For example, taxi trajectory data have been frequently used to predict travel time (Lee et al., 2009; Jenelius and Koutsopoulos, 2013; Jiang and Li, 2013; Zhan et al., 2013); others include bus trajectory (Shalaby and Farhan, 2003), mobile phone data (Bar-Gera, 2007; Herrera et al., 2010), car navigation system data (Jones et al., 2013), etc. Vehicle trajectory data can theoretically provide complete temporal and spatial information of a specific trip, i.e., origin, destination, route, travel time etc. Among of them, closed road systems or independent links are still the targets of most of the aforementioned works. A few research works concerning trajectory-based TOTTP are discussed below.

Lee et al. (2009) propose a real-time model to predict trip-oriented travel time. In this model, trip-oriented travel time for an origin-destination (OD) pair is defined as consisting of links' travel time and intersections' delay. Real-time and historical taxi trajectory data are dynamically combined to modify links' travel time. And intersections' delay is estimated through similar historical trajectories or expert heuristics. This research is limited to real-time or short-term prediction, and the results may vary along the route.

The approach proposed by Jiang and Li (2013) regards trip as a longer road link and uses a time series forecasting method to predict trip-oriented travel time. The approach is applicable to any future trip, and its performance varies with time seriously. The authors advise possible measures to improve the approach and maintain a stable performance such as employing different forecasting methods and considering more influence factors such as weather conditions at a conceptual level.

External factors such as special dates or events (Karl and Trayford, 2000) and weather conditions (Chien and Kuchipudi, 2003; Wu et al., 2004; Tu et al., 2007) have been proven influential to travel time. Therefore, the study of Jenelius and Koutsopoulos (2013) has integrated some of them into trip-oriented travel time prediction. Based on probe vehicle data, they built a maximum likelihood statistics model to represent link attributes (e.g., speed limit, road functional class, etc.) and trip conditions (e.g., day of week, season, and weather, etc.) for TOTTP. Limited by the study area and season, the study mainly emphasizes the influence of snowfall in winter. Besides, the coverage of probe vehicle data is limited, and therefore, its applicability to large volume of trajectory data from general vehicles needs to be examined.

In conclusion, there is much less reported research on TOTTP than predicting travel time on closed traffic sections due to the complexity of models and the shortage of traffic data. Historical vehicle trajectory data, as the direct reflection to traffic flow, have been introduced into the study of trip-oriented travel time prediction, where challenges still exist, including how to develop a prediction model based on large volume of vehicle trajectory data, how to improve the accuracy of prediction results, how to integrate other factors into the model, etc.. With these in mind, we propose a BPNN-based prediction approach, where the BPNN model, as a neuro network method, can represent the dynamic and non-linear processes in traffic (van Lint et al., 2005), and has been employed for travel time prediction with various fashions (Dougherty and Cobbett, 1997; Huisken and Coffa, 2000; Alecsandru and Ishak, 2004; Ishak and Alecsandru, 2004; Elabd and Schlenkhoff, 2009; Wei et al., 2009). In this paper, more factors are considered in the proposed BPNN model, and data of numerous historical trips from vehicle trajectory data as well as other related information are used to train the model. Details are given in the following sections.

3 Methodology

3.1 Principle and overview

The principle of the proposed TOTTP is to develop a BPNN model based on a large number of historical trajectory data and meteorology data. For each travel route, i.e., sharing the same origin, destination, and path, an individual BPNN model is built. This model establishes the linkage from travel information (i.e., departure date and time) and influence factors (i.e., weather conditions and special events) to trips' travel time. Through feeding this model with planned travel information and predictable or known influence factors as the input, users may obtain the travel time of a future trip as the output.

Figure 1 gives an overview of the methodology, which consists of three parts. The first part is to prepare data from historical trajectory data and meteorology data for training BPNN models. The second part is to build BPNN models with training data. The third part is to apply the model to travel time prediction for future trips. Details of data preparation and building BPNN models are given in the following content of this section, and model application is discussed in the next section.

3.2 Data preparation

Each BPNN model is applicable to one specific route. Therefore, the targeted route must be decided before building a BPNN model. The origin, destination, and path of the route are employed as criteria to extract the following training data from historical trajectory and meteorology data.

Usually, historical trajectory data consist of a huge number of location points, such as GPS sampling points, collected from different vehicles. Each location point includes geographic coordinates and sampling time as well as the identification of a vehicle. The first step is to group these points by vehicles and sort points of each vehicle by sampling time. Then, after linking these discrete points one after the other in space, a continuous trajectory of a vehicle is restored, which often includes many trips. We define a trip as a segment of trajectory with origin and destination. For the trajectory of a taxi, it is easy to recognize the origin or destination of a trip if the information whether the taxi is occupied or not can be recorded in its location points. For other vehicles, thematic analysis is needed. From all detected trips, those matching the targeted route are selected and for each selected trip, the following data are extracted to train the BPNN model of the route.

3.2.1 Departure time

Departure time is the time instant when the trip starts, and it is basically the most significant factor of influencing travel

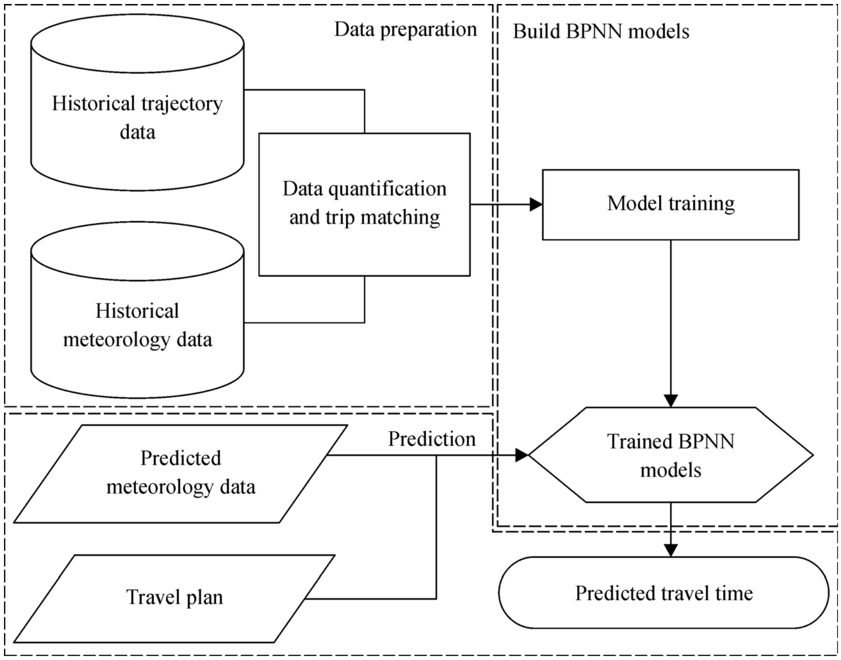

Fig. 1 Overview of the methodology.

time with respect to the periodical traffic congestion in an urban traffic network within one day. We use the number of second intervals that have elapsed since midnight (12:00:00) on the travel day to represent the departure time of a trip.

3.2.2 Travel date

Travel date is the day when the trip starts. Trip travel time is affected by daily, weekly, or occasional events (Nahar and Sultana, 2014). For instance, traffic patterns on Monday and Friday may be very different though both are workdays. We use a 9-bit binary number to represent the features of a travel date as follows.

- Bit 1 is 0 for workdays or 1 for non-workdays.
- Bit 2–5 represents the order of the travel date in continuous workdays or non-workdays.
- Bit 6–9 represents the number of continuous workdays or non-workdays.

For instance, a number "0 0001 0101" means that the travel date is the first day of 5 continuous workdays.

3.2.3 Weather conditions

For each day, weather data are recorded twice, namely daytime weather (8 a.m. to 8 p.m.) and nighttime weather (8 p.m. to 8 a.m. of the next day) and assigned to trips with departure time and date falling into corresponding periods.

According to existing collected data on weather conditions, there are 12 types of weather. They are classified with respect to their influence on traffic flow (Manual, 2000) and quantified in Table 1.

Table 1 Quantified weather conditions

Weather condition	Value
Sunny Cloudy	1
Drizzle Sleet Light Snow	2
Moderate Rain	3
Heavy Rain Rainstorm	4
Moderate Snow Heavy Snow Blizzard Other	5

Air quality and temperature are also considered as infect factors in the proposed BPNN model (Manual, 2000). The level of air quality may influence trip generation. For instance, a travel plan may be cancelled due to highly polluted air. Temperature may facilitate revealing seasonal changes of traffic flow.

We employ daily Air Quality Index (AQI) to quantify air quality since it is widely used by government agencies over the world though its criteria may vary with countries. In general, AQI is a number representing the concentration of air pollution. A big AQI means a large average dose of air pollutant within a unit air volume over a given period, and accordingly, a high adverse health effect. Table 2 shows the six levels of AQI and the quantitative values.

Similar to weather data, temperature is recorded twice every day for daytime and nighttime respectively. The unit

Table 2 Quantified AQI

AQI levels	Value
Excellent	1
Good	2
Light pollution	3
Moderate pollution	4
Severe pollution	5
Serious pollution	6

of measurement for temperature is Celsius. Considering the normal temperature range of the study area, we denote the minimum temperature is minus ten degrees Celsius (–10°C) and the maximum temperature is forty degrees Celsius (40°C).

3.3 Building BPNN models

Inspired by a biological neural network, Artificial Neural Network (ANN) is basically a statistical learning network consisting of a large number of nodes and connections (Dougherty, 1995). Nodes are also known as neurons and joined together by adaptive weighted connections. Data flow along these connections and are scaled during transmissions according to their weights. The process of tuning the weights called model training are carried out by feeding a set of exemplar input-output pairs to the model and adjusting the weights in order to minimize the errors between the answer the network gives and the desired output. Once the training process is completed, the model is able to produce answers for input values from data other than the training data.

Based on the principle of ANN, a group of algorithms sharing the characteristics of massively parallel processing, distributed storage, self-organizing and self-learning have been proposed. Among them, BPNN is a multilayer feedforward ANN trained with a back propagation algorithm. BPNN can learn and store large amounts of input-output mapping based on the gradient descent method which minimizes the sum of the squared errors between actual and desired outputs.

As shown in Fig. 2, a BPNN model includes an input layer, a hidden layer, and an output layer. Stimulation is applied to the input layer, and signals propagate through the hidden layer to the output layer. Each link between a pair of neurons has a unique weighting value. Eqs. (1) and (2) are used to train weights of links, where n denotes the number of iteration, W weight set, η learning rate, α momentum factor, E gradient of error function, and $\Delta W(n-1)$ weight incremental quantity.

$$W(n) = W(n-1) - \Delta W(n), \tag{1}$$

$$\Delta W(n) = \eta \frac{\partial E}{\partial W}(n-1) + \alpha \Delta W(n-1). \tag{2}$$

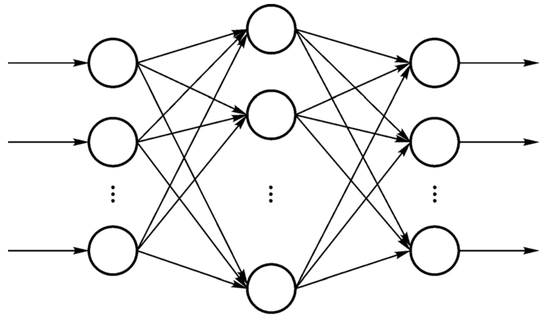

Fig. 2 Structure of BPNN model.

If let t^i be the model output and y^i is the actual output of sample i in training data, then E is given in Eq. (3).

$$E = \frac{1}{2}\sum (t^i - y^i)^2. \tag{3}$$

Learning rate (η), a real number between 0 and 1, applies a greater or less portion of the respective adjustment to existing weights. A large η may generally increase the speed of model learning, but meanwhile, if there is a large variability in the input set, the quality of learning may be low. Usually, η is set to be a small value initially and then increased gradually if learning rate is slow. Momentum factor (α) is also a real between 0 and 1. It basically allows a change to the weights to persist for abundant adjustment cycles. The magnitude of the persistence is controlled by α. The larger α is, the greater the persistence is. Tuning the values of η and α is discussed in our experiments.

According to the Kolmogorov theorem and back propagation fix quantify, a three-layer BPNN model with Sigmoid function as excitation function can approach any continual function in any precision (Kůrková, 1992).

In this research, quantified travel date, departure time, and weather conditions are input variables and travel time is the output. A summary of the training process is briefly given below.

1) Normalize input and output data.

2) Use the current network weights to compute predicted outputs.

3) Compute E based on predicted outputs and desired outputs.

4) If optimal cost ΔW is achieved, exit, and otherwise, go back to step 2.

4 Experiments

4.1 Data

Trajectory data used in this research are from 800 taxies covering a road network of 371 square kilometers in Minhang Distinct of Shanghai, China from June 1, 2013 to

June 30, 2014. Weather data during this period are also collected. The average sampling rate of trajectory data is 0.05 Hz, and the total number of sampling points is about 0.8 billion. Besides location information, occupancy of a taxi is also recorded in sampling points. Therefore, it is possible to segment the taxi's trajectory into individual trips through analyzing the changes of its occupancy.

Figure 3 gives the spatial distribution of the network, and 3 traffic zones are highlighted, where **Zone A** consists of an airport, **Zone B** is a regional transportation hub, and **Zone C** includes a subway station and several bus stations around it. These zones are selected as origins or destinations of trips, and in particular, travel routes from **Zone B** to **Zone A** and from **Zone C** to **Zone A** are targeted by the following experiments to build BPNN models and predict travel time of trips along them.

According to the exploratory analysis on collected trajectory data, the three most-adopted routes from **Zone B** to **Zone A** or from **Zone C** to **Zone A** are extracted. These routes are plotted in Fig. 4 and the information of trips along them are given in Table 3, where trips collected from June 1, 2013 to May 31, 2014 are employed as training data to build BPNN models, while trips collected from June 1, 2014 to June 30, 2014 are used for testing the models. As shown in this table, we can evaluate the applicability and performance of the proposed methodology when the number of historical trips along difference routes varies to a great extent.

4.2 Building BPNN Models

Building BPNN models is a training or learning process including a large number of epochs for adjusting weights. Since there are five input variables, thirteen neuros in a hidden layer are constructed according to the Kolmogorov theorem (Kůrková, 1992). The computational time of training BPNN models mainly relies on the number of iterations. In this research, all experiments are conducted with a computer equipped with Intel i5 3.4 GHz CPU, 4GB RAM and Windows 7 64-bit operation system. It usually takes less than 10 seconds to finish a training process.

During the training process, learning rate (η) and momentum factor (α) are two unknown parameters and needed to be tuned in order to reduce errors.

In our experiments, for each BPNN model, 5 values between 0 and 1, namely 0.1, 0.3, 0.5, 0.7, and 0.9, are assigned to η and α, respectively. Every combination of η and α is applied to the training process and their performances are compared to find the most suitable values for η and α. For instance, Fig. 5 illustrates the relationship between training error E in Eq. (3) and different values of η or α when α or η is fixed for Route 5 defined in Table 3 and Fig. 4. The X axis is the number of epochs during the training process. In Fig. 5(a), $\alpha = 0.3$ while η changes from 0.1 to 0.9, and in Fig. 5(b), $\eta = 0.3$ while α changes from 0.1 to 0.9. According to these figures, when $\alpha = 0.7$ and $\eta = 0.9$, training error quickly

Fig. 3 Road network and 3 selected traffic zones.

Table 3 Trip and route information

Origin and destination	Route No.	Number of trips	Route length/km
From **Zone B** to **Zone A**	Route 1	5206	14.36
	Route 2	2751	11.91
	Route 3	273	18.87
From **Zone C** to **Zone A**	Route 4	205	26.95
	Route 5	169	23.78
	Route 6	45	29.08

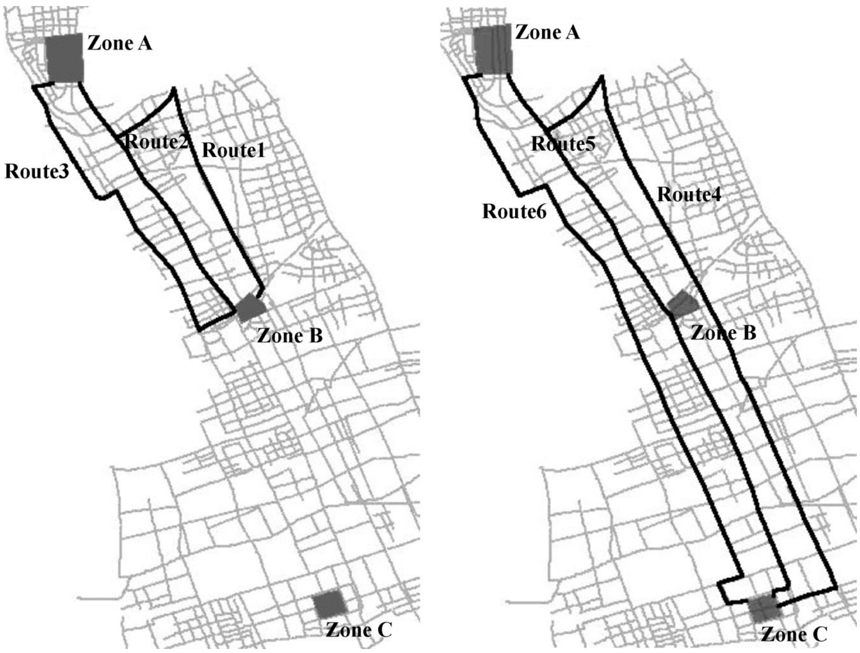

Fig. 4 Selected routes (a) from **Zone B** to **Zone A** and (b) from **Zone C** to **Zone A**.

converges to the lowest value. The same phenomenon can be watched when other combinations of η and α are applied. By this means, values of η and α are determined for each model.

4.3 Predicting travel time

Testing data are applied to the trained BPNN models. As aforementioned, in our experiments, trips in June, 2014, are used to test the models. We feed these models with departure time (from 00:00:00 to 23:59:59), travel date (from day 1 to day 30), weather, air quality, and temperature, and then predicted travel time is obtained. Within this month, all weekends, as well as June 2, which is a national holiday, are non-workdays, and other days are workdays.

Figure 6(a) gives the predicted results for routes from **Zone B** to **Zone A** covering the entire testing month, and Figure 6(b) focuses on one week (from day 9 to day 15) in the month to demonstrate the daily changes of predicted travel time within one week. The X axis denotes the departure time and date, and the Y axis denotes the predicted travel time. Non-workdays are highlighted with grids, and the rest are workdays. Since the BPNN models are able to predict travel time of trips departing at any second, the results are exhibited in a fashion of curves.

According to these figures, periodical characteristics of travel time changes are noted. Patterns on workdays and non-workdays are remarkably different, while similar patterns are observed on each day during continuous workdays or non-workdays. On each workday, two peaks can be found respectively in morning and evening, while in a non-workday, only morning peak exists. The above phenomena are applicable to all the three routes except that their patterns of predicted travel time vary. For example, it always takes the longest time to finish a trip along Route 3 than other routes and its evening peak period lasts to midnight.

The above procedures are further applied to routes from **Zone C** to **Zone A**, and results are given in Fig. 7. It is

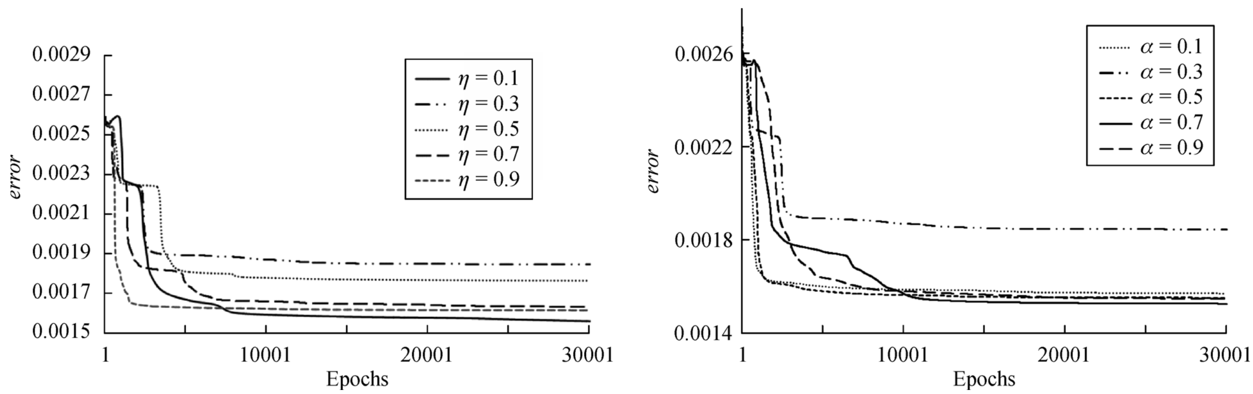

Fig. 5 The relationship between training error E and different values of η or α when α or η is fixed during a training process for Route 4.

interesting to see that weekly periodical changes of travel time still exist, but daily patterns within one week vary from Monday to Friday.

The above analysis indicates that a trained BPNN model is capable of reflecting the features of travel time of trips along a route at a fine scale.

4.4 Evaluating predicted results

To evaluate and compare the reliability of the predicted results, mean absolute error ($\hat{E}$) and error rate (R) are defined in Eqs. (4) and (5), where n denotes the number of trips for testing, l denotes the length of the route, tr_i the actual travel time of i^{th} trip and tp_i its predicted travel time.

$$\hat{E} = \frac{1}{n}\sum\nolimits_{i}{}_{1}^{n}|tp_i - tr_i|, \tag{4}$$

$$R = \frac{\hat{E}}{l}. \tag{5}$$

For comparing the performance of the proposed method, a moving average method is implemented. Its moving window is 30 minutes. With this method, predicted result equals the average travel time of trips within the 30-minute moving window having the same travel date in training data. Error comparison between moving average method and the proposed method is given in Table 4.

As shown in Table 4, the proposed method always

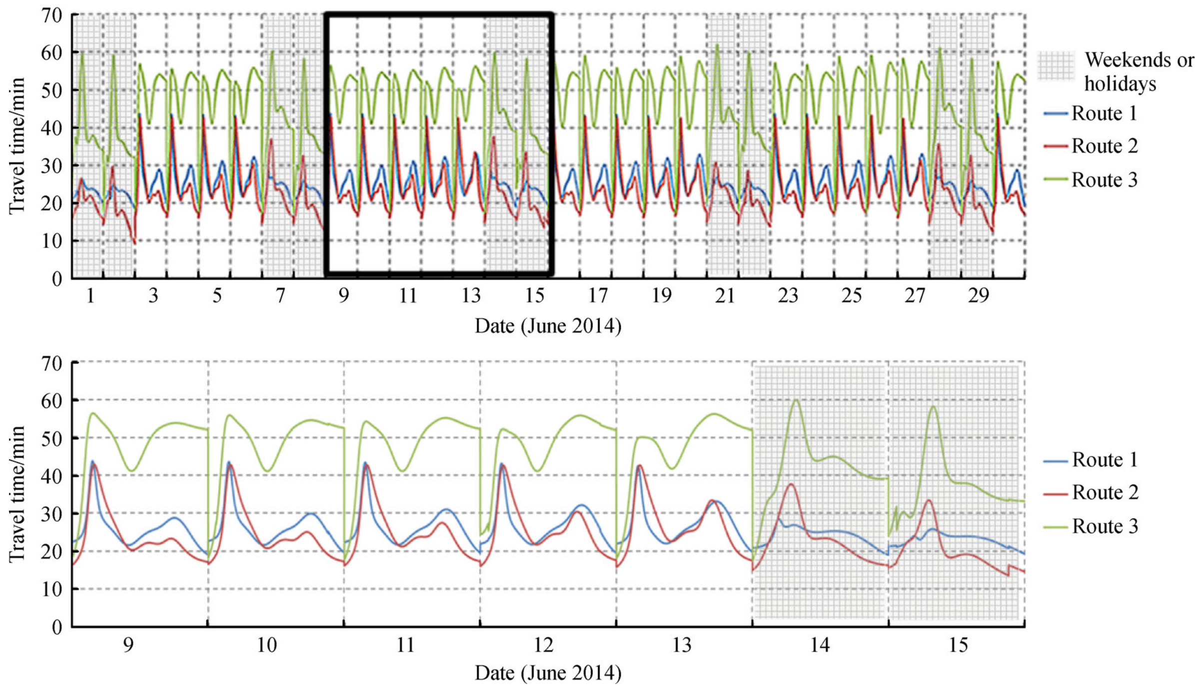

Fig. 6 Predicted travel time for trips in June 2014 from **Zone B** to **Zone A**.

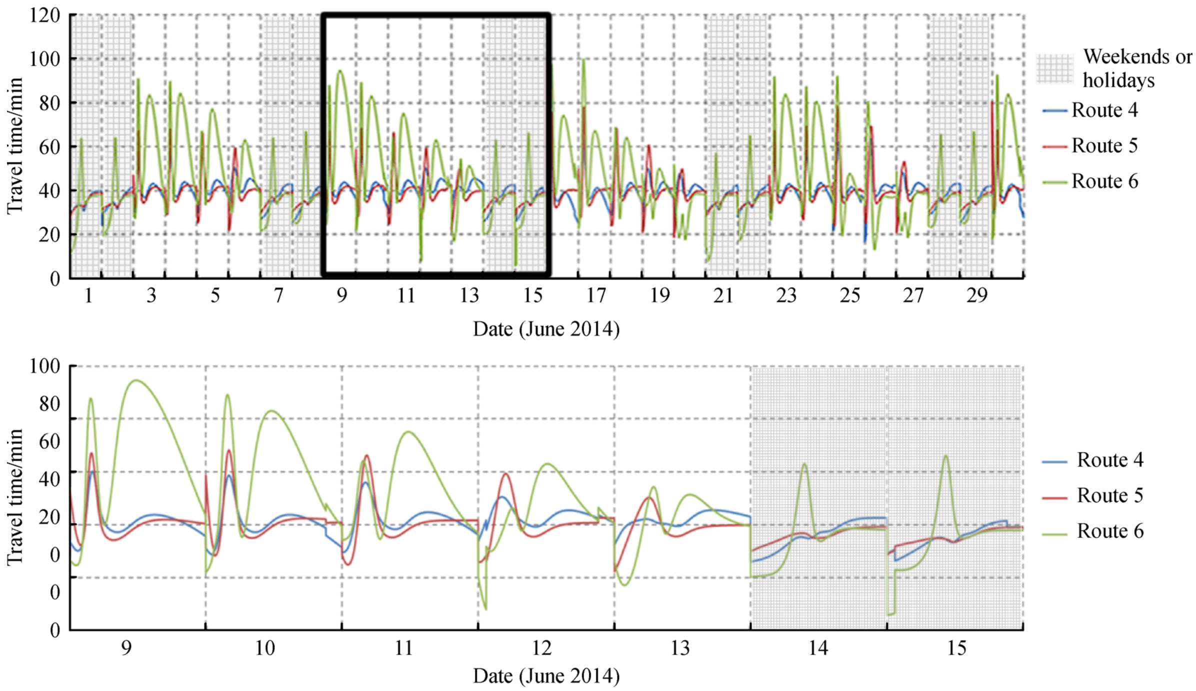

Fig. 7 Predicted travel time for trips in June 2014 from **Zone C** to **Zone A**.

Table 4 Error comparison between moving average method and the proposed method

Route	n	l/km	The proposed method		Moving average method	
			$\hat{E}$ /min	R	$\hat{E}$ /min	R
1	218	14.36	3.64	0.253	5.28	0.368
2	420	11.91	3.28	0.275	4.03	0.338
3	27	18.87	9.87	0.523	14.31	0.758
4	9	26.95	3.03	0.112	9.59	0.356
5	16	23.78	11.47	0.482	13.21	0.556
6	4	29.08	8.63	0.297	19.98	0.687

outperform moving average method, partly because moving average method cannot take into account weather, air quality and temperature when predicting travel time. Furthermore, it is noted that there is no relationship between route lengths, the number of training samples and prediction errors. Comparing Tables 3 and 4, with the same origin and destination, the higher prediction errors are occurs in Routes 3 and 5, while the number of training samples and route length for Route 3 is smaller than Routes 1 and 2, and in contrast, Route 5 is larger than Routes 4 and 6. In addition to the performance of prediction model, the distribution of training samples and commute patterns of route may by important factors for prediction accuracy.

Results listed in Table 4 consider all testing trips as a whole. Besides, we examine the hourly changes of mean absolute error for a certain travel date. Figure 8 gives the error analysis by hour from 5:00 to 22:00 for trips along Route 2 on the first day of 5 continuous workdays. It is noticed that the error varies from 1.99 minutes to 7 minutes and its peaks are consistent with the peaks of traffic flow appearing around 7:00–8:00, 12:00–13:00, and 18:00–19:00.

4.5 Factor analysis

One of the advantages of BPNN models is that different factors influencing travel time can be integrated. To evaluate the significance of meteorological factors, we furthermore repeat the process of building BPNN models but removing some factors and applied the testing data to the built models for Route 2. Accordingly, five groups of experiments are conducted and error analysis is given in Table 5.

According to Table 5, the performance of the proposed

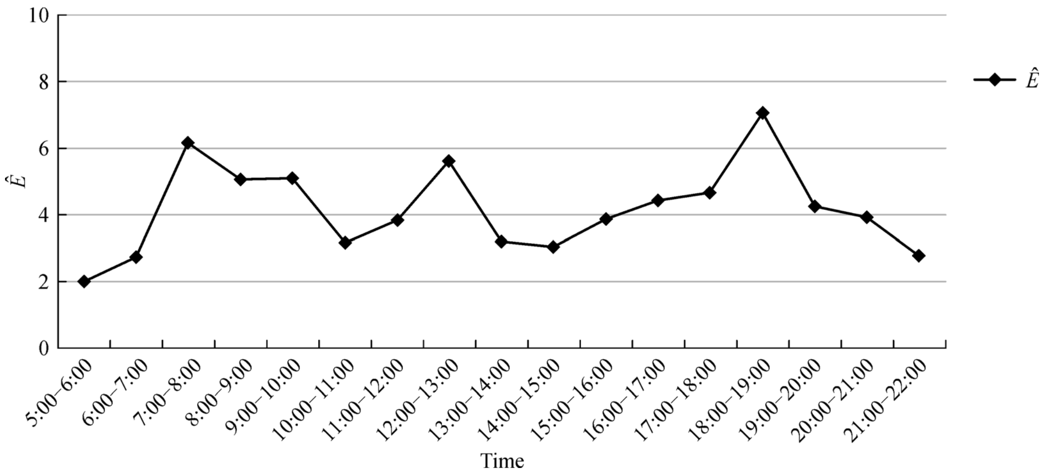

Fig. 8 Error analysis by hour for trips along Route 2 on the first day of 5 continuous workdays.

Table 5 Meteorological factor analysis

Input factors*	$\hat{E}$
DTWAC	3.28
DTAC	4.08
DTWA	4.11
DTWC	4.32
DT	6.23

*D denotes departure time, T denotes travel date, W denotes weather, A denotes air quality and C denotes temperature.

method is the best when all factors are considered, while the performance is the worst when only time and date are considered. In addition to factors about time, meteorological factors also change traffic flow patterns and influence travel time.

5 Conclusions

In this paper, we present an approach to trip-oriented travel time prediction (TOTTP) in highly urbanized area. This approach employs a BPNN model to integrate historical vehicle trajectory data and meteorology data that affect travel time. For each specific route, a BPNN model is built with quantified travel date, departure time, and weather conditions as input variables and travel time as the output. When predicting, feed the model with future travel plan and meteorology data, and then predicted travel time can be produced.

A historical taxi trajectory and meteorology dataset covering 13 months is employed to validate the proposed approach. Results demonstrate that it helps discover the dominate trends of travel time changes daily and weekly, and the impact of weather conditions is non-trivial.

Future efforts will be made in two aspects: At first, the spatial-temporal applicability of the approach will be further examined with other datasets and routes. Secondly, the built BPNN models may have to be updated over time with respect to gradually expending historical database. The update mechanism of BPNN models as well as the maintenance of historical database needs to be explored.

Acknowledgements This research was sponsored by the National Natural Science Foundation of China (Grant No. 41271441).

References

Abu-Lebdeh G, Singh A K (2011). Modeling arterial travel time with limited traffic variables using conditional independence graphs & state-space neural networks. Procedia Soc Behav Sci, 16: 207–217

Alecsandru C, Ishak S (2004). Hybrid model-based and memory-based traffic prediction system. Transp Res Rec, 1879(1): 59–70

Bar-Gera H (2007). Evaluation of a cellular phone-based system for measurements of traffic speeds and travel times: a case study from Israel. Transp Res, Part C Emerg Technol, 15(6): 380–391

Bhaskar A, Chung E, Dumont A (2011). Arterial travel time estimation: revisiting the classical procedure. Australasian Transport Research Forum Adelaide South Australia, 34

Chen D, Zhang K, Liao T (2010). Practical travel time prediction algorithms based on neural network and data fusion for urban expressway. Proceedings 2010 Sixth International Conference on Natural Computation, 4, 1754–1758. IEEE

Chen M, Chien S (2001). Dynamic freeway travel-time prediction with probe vehicle data: link based versus path based. Transp Res Rec, 1768: 157–161

Chien S I, Kuchipudi C M (2003). Dynamic travel time prediction with real-time and historic data. J Transp Eng, 129(6): 608–616

Dia H (2001). An object-oriented neural network approach to short-term traffic forecasting. Eur J Oper Res, 131(2): 253–261

Dong C, Shao C, Richards S H, Han L D (2014). Flow rate and time mean speed predictions for the urban freeway network using state space models. Transp Res, Part C Emerg Technol, 43: 20–32

Dougherty M (1995). A review of neural networks applied to transport.

Transp Res, Part C Emerg Technol, 3(4): 247–260
Dougherty M S, Cobbett M R (1997). Short-term inter-urban traffic forecasts using neural networks. Int J Forecast, 13(1): 21–31
Elabd S, Schlenkhoff A (2009). ANFIS and BP neural network for travel time prediction. World Acad Sci Eng Technol, 57: 116–121
Fei X, Lu C C, Liu K (2011). A bayesian dynamic linear model approach for real-time short-term freeway travel time prediction. Transp Res, Part C Emerg Technol, 19(6): 1306–1318
Goh A T C (1995). Back-propagation neural networks for modeling complex systems. Artif Intell Eng, 9(3): 143–151
Guo J, Huang W, Williams B M (2014). Adaptive Kalman filter approach for stochastic short-term traffic flow rate prediction and uncertainty quantification. Transp Res, Part C Emerg Technol, 43: 50–64
Herrera J C, Work D B, Herring R, Ban X, Jacobson Q, Bayen A M (2010). Evaluation of traffic data obtained via GPS-enabled mobile phones: the mobile century field experiment. Transp Res, Part C Emerg Technol, 18(4): 568–583
Hofleitner A, Herring R, Bayen A (2012). Arterial travel time forecast with streaming data: a hybrid approach of flow modeling and machine learning. Transp Res, Part B: Methodol, 46(9): 1097–1122
Huisken G, Coffa A (2000). Short-term congestion prediction: comparing time series with neural networks. In: Proceedings of Road Transport Information and Control, 2000. Tenth International Conference on (Conf. Publ. No. 472), 66 –69. IET
Ishak S, Alecsandru C (2004). Optimizing traffic prediction performance of neural networks under various topological, input, and traffic condition settings. J Transp Eng, 130(4): 452–465
Jenelius E, Koutsopoulos H N (2013). Travel time estimation for urban road networks using low frequency probe vehicle data. Transp Res, Part B: Methodol, 53(4): 64–81
Jiang Y J, Li X (2013). Travel time prediction based on historical trajectory data. Ann GIS, 19(1): 27–35
Jones M, Geng Y, Nikovski D, Hirata T (2013). Predicting link travel times from floating car data. 16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013), The Hague, 1756–1763
Karl C A, Trayford R S (2000). Delivery of real-time and predictive travel time information: experiences from a Melbourne trial. In: Proceedings of 7th World Congress on Intelligent Systems, 3513
Kůrková V (1992). Kolmogorov’s theorem and multilayer neural networks. Neural Netw, 5(3): 501–506
Lee W H, Tseng S S, Tsai S H (2009). A knowledge based real-time travel time prediction system for urban network. Expert Syst Appl, 36 (3): 4239–4247
Li B, Zhang D, Sun L, Chen C, Li S, Qi G, Yang Q (2011). Hunting or waiting? Discovering passenger-finding strategies from a large-scale real-world taxi dataset. 2011 IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops), Seattle, WA, 63–68
Lin W H, Kulkarni A, Mirchandani P (2004). Short-term arterial travel time prediction for advanced traveler information systems. Journal of Intelligent Transportation Systems, 8(3): 143–154
Long Cheu R, Xie C, Lee D H (2002). Probe vehicle population and sample size for arterial speed estimation. Comput Aided Civ Infrastruct Eng, 17(1): 53–60
Mori U, Mendiburu A, Álvarez M, Lozano J A (2015). A review of travel time estimation and forecasting for Advanced Traveller Information Systems. Transportmetrica A. Transport Sci, 11(2): 119–157
Nahar L, Sultana Z (2014). A new travel time prediction method for intelligent transportation system. IOSR Journal of Computer Engineering, 16(3): 24–30
Shalaby A, Farhan A (2003). Bus travel time prediction model for dynamic operations control and passenger information systems. Transportation Research Board, 2
Stathopoulos A, Karlaftis M G (2003). A multivariate state space approach for urban traffic flow modeling and prediction. Transp Res, Part C Emerg Technol, 11(2): 121–135
Tu H, van Lint H W C, van Zuylen H J (2007). Impact of adverse weather on travel time variability of freeway corridors. In: Proceedings of Transportation Research Board 86th Annual Meeting
van Lint J W C, Hoogendoorn S P, van Zuylen H J (2005). Accurate freeway travel time prediction with state-space neural networks under missing data. Transp Res, Part C Emerg Technol, 13(5–6): 347–369
Vlahogianni E I, Golias J C, Karlaftis M G (2004). Short-term traffic forecasting: overview of objectives and methods. Transp Rev, 24(5): 533–557
Wei L, Fang Z, Luan S (2009). Travel time prediction method for urban expressway link based on artificial neural network. In: Proceedings of 2013 International Conference on Computing, Networking and Communications (ICNC), 1: 358–362
Wu C H, Ho J M, Lee D T (2004). Travel-time prediction with support vector regression. IEEE Trans Intell Transp Syst, 5(4): 276–281
Zhan X, Hasan S, Ukkusuri S V, Kamga C (2013). Urban link travel time estimation using large-scale taxi data with partial information. Transp Res, Part C Emerg Technol, 33(4): 37–49