ENHANCING TRAVEL TIME FORECASTING WITH TRAFFIC CONDITION DETECTION

Yingying Duan, Feng Lu, Jun Ouyang, Chuanbin Chen

State Key Laboratory of Resources and Environmental Information System,
Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences,
No. 11A, Datun Road, Chaoyang District, Beijing 100101, P. R. China - (duanyy, luf, ouyj,chencb)@lreis.ac.cn

KEY WORDS: Travel time forecasting, traffic condition detection, natural language understanding, spatial-temporal data mining

ABSTRACT:

Short-term traffic forecasting aims to provide more reliable travel information service, so as to assist people in making more reasonable travel decisions. With the increasing availability of traffic data along with the development of communication technology, both the capability and accuracy of travel time forecasting have been significantly enhanced in real-time conditions and a great number of forecasting methods have been carried out during recent years. However, they are inadequate when confronted with the real world traffic problems, since the real-time traffic condition can be affected easily and changed constantly. In our study, a hybrid forecasting approach is presented from a more practical perspective, based on a traffic condition detection method which monitors the real-time traffic condition and performs the travel time forecasting according to different traffic conditions.

In particular, we first build a traffic conditions evaluation system to detect different sorts of traffic conditions. In this study, the traffic conditions are divided into four types including light, stable, congested and abnormal traffic condition according to travel time cost. We use a clustering tool to obtain traffic flow patterns of different traffic conditions. And the process characterize the state of the system with respect to the deviation of current conditions from an expected ones based on historical data as a definition for abnormalities in the traffic stream. Then the hybrid forecasting approach, in which several methods are used to deal with different traffic conditions, is trained to judge with certain confidence which method performs the best according to the certain traffic condition with historical traffic data. Then the travel time forecasting is taken out after the detection of real-time condition by the hybrid forecasting approach with fixed historical data and received real-time traffic information. Case studies are carried out using a real-time traffic dataset in downtown Beijing.

1. INTRODUCTION

Travel information services develop quickly all around the world and make great efforts in the area of intelligent transportation systems (ITS). With the increasing public travel demands, traditional travel information services such as shortest path search on static maps are out of requirements. Since the accuracy and practicality of dynamic navigation highly depends on the short term traffic forecasting, a lot of short term travel time cost forecasting models on the roadways have been developed during recent years. The short term travel time cost forecasting research can be categorized as follows: direct and indirect model-based approaches, data-driven approaches (Van Lint et al., 2005) , and hybrid approaches.

Model-based approaches solve the travel time forecasting problem by forecasting traffic conditions from a number of time periods ahead and then subsequently deduce mean travel times from these forecasted traffic conditions. Examples include METANET (Smulders et al., 1999) , DynaSMART (Hu, 2001) , and DynaMIT (Ben-Akiva et al., 2001) . Data driven models are able to directly learn the complex traffic dynamics from the data and to forecast traffic flows and speeds. Many successful efforts have been reported including ARIMA models (Williams, 2001) ; artificial neural networks (Mark et al., 2004); fuzzy neural networks (Yin et al., 2002);and support vector regression models (Wu et al., 2004)and so on. All the data driven methods have in common that they correlate mean travel times or traffic conditions to current and past traffic data, and they do not require extensive expertise on traffic flow modelling. The easily use and the accuracy in progressing non-linear problem make the data driven models highly developed. A lot of forecasting models have been developed during recent years but none of them could consistently outperform the others. In the real-world applications, traffic forecasting conditions can be affected by a lot of factors. That leads to the research on hybrid approaches on travel time forecasting. The hybrid approaches combine with many possibly applicable candidate models for various traffic conditions (Zhu, 2009).

Studies on short-term traffic forecasting have shown that, one of the key functions of traffic management systems is to monitor traffic conditions and detect the presence of conditions that are abnormal or may not be expected (Turochy, 2006) . The traffic condition detection problem can be viewed as recognizing the non-recurrent congestion patterns from observed data series obtained from loop detectors. In recent years, computational intelligence approaches including neural-computing, evolutionary computing; wavelet analysis and fuzzy logic have been employed to solve the complex and mathematically intractable incident detection problems (Jin et al., 2006a).

However, most of them are based on single roadway pattern, geospatial neighbourhoods relationships between roadways do not involve in these pattern detection methods. And the real traffic data used in those researches are obtained from fixed sensors such as cameras and magnetic loops. These traffic data can easily be used in the freeway, but in the urban road network, the fixed sensors are most settled at the intersections in arterial roads. That causes the lack of detecting on the roadway segments and on the other roads.

Aiming at such problems, we propose a travel time cost forecasting approach based on traffic condition detection. Although traffic incidents are not predicable, we can enhance the travel time forecasting by detecting the abnormal traffic conditions and using different forecasting strategies under certain situation. Here we use two ways to get the abnormal traffic information. One is detecting the traffic condition; the other is collect traffic events described in natural language. The major contribution of this traffic condition detection method proposed is considering both spatial and temporal information in detection using the traffic velocity collected by floating vehicles since it is an appropriate measurement to indicate the congestion. When the real time traffic speeds on the roadways are received, the traffic detection module is working on detecting abnormalities in the traffic. And a traffic events collection module is also used to catch the traffic incidents described in natural language at the same time. After that, two types of neural networks are built to forecast the speeds in the future time intervals according to the traffic condition is recurrent or not. The method presented in this paper is argued to provide a practical solution for real-time public travel information service.

2. TRAFFIC CONDITION DETECTION

The real-time traffic data updated every five minutes is collected from the floating vehicles. We use a new algorithm outperformed the California algorithm consistently under various scenarios. The proposed approach includes spatial-temporal data mining and online detecting using California Algorithm. The California Algorithm is one of the earliest and most popular algorithms which based on the logical assumption that a traffic incident increases the traffic occupancy upstream of the incident and decreases the traffic occupancy downstream of the incident significantly(Jin et al., 2006b) . The incident detection process can be divided into four steps as listed below.

Step 1 Pre-processing
In this task, the raw velocity data obtained from float vehicles are transformed in the format needed for the algorithm. Common pre-processing approaches include calculating the cumulative values of time-series data and interpolating for the missing data.

Step 2: Traffic Model Generation
The second stage analyzes the traffic data on different weekdays to construct traffic models respectively. We de-noising and enhancement of the signal output obtained from the pre-processing and using statistics methods to confirm the confidence interval for each time interval on each roadway. This is an important stage because noise corruption is one of the primary reasons for poor reliability of the incident detection algorithms. Each dataset contains the velocity collected on one roadway at the same time on the same day of the week. Test the datasets follow which kind of distributions, and calculate the confidence interval by the cleaned data. The final traffic model will be generated as the confidence interval of the remaining historical data.

Step 3 Non-recurrent conditions confirmation
The detection will be performed by calculating the distance between the real-time data and the limit of the confidence interval. Once the real-time data is outlier of the confidence interval, the traffic data will be considered as a suspect abnormal condition. Then the traffic conditions on this roadway and on the adjacent roadways in the past time intervals should be taken into consideration. If the past time intervals are confirmed abnormal conditions or suspect condition, the current time interval can also be confirmed as non-recurrent condition. Or if the relational roadways such as the upstream and downstream roadways have been confirmed abnormal, the current time interval on the current roadway can be recognized as non-recurrent condition. If the traffic condition on current time interval can not be confirmed, it remains as a suspect condition.

3. TRAFFIC EVENTS COLLECTION

3.1 Traffic Events Reported in Nature Language

Neither of the magnetic loops, float vehicle or cellular phone signal analysis technologies can obtain the abrupt traffic events on spots or road cross turns. Once the abrupt traffic events happen, the traffic policemen, onlookers or people concerned will report the events and resulted influence (on the spot or monitor viewing) to the information center via cellular phones, short messages or other instant message systems. Since these traffic events are detected by human, those reported massages are most described in natural language. It has been a time consuming task to translate the messages into the valuable information that suit the applications and requires artificial work. The bottleneck focuses on understanding the natural language describing traffic and matching understood traffic information in LRS forms with the underlying road network spatial dataset, including the matching of address with the geometrical information, matching of multi-source LRS, and LRS and GIS positioning manner.

Although the natural language can not be truly understood automatically nowadays, the key words could be caught by the describing rules in some specific field. Thus the automatic understanding of traffic events reported by natural language should focus on the characteristic rules of traffic information. Those messages are always described briefly in certain forms combined with "<Address/Landmark> + {Direction} + {Offset} + <Traffic Events>".

According to the description rules, we proposed a cross-step word segmentation algorithm to process real-time traffic events represented in natural Chinese in this approach. Four libraries are built to process the word segmentation. The Address word library contains the addresses, landmarks as well as point of interests; The Direction word library is filled with the directions such as "由南向西"(from south to west) and so on; The Event word library stores the traffic events like "车行缓慢", "两车刮蹭", "追尾"; And the Ext Location word library used to save the offset value.Considering the record length distribution of the word libraries depicting real-time traffic information, this algorithm sets corresponding steps of word segmentation for address, direction and event libraries, and improves the one step running of the string pointer in classical Chinese word segmentation to flexible multiple steps running, so as to aggregate possible Chinese words efficiently. Then the addresses should be matched with the roadways and the influence of the events should be quantified for the velocity forecasting.

3.2 Traffic events processing module

Understanding the traffic events only concerns how to rapidly and accurately process the key words representing addresses, directions and events. Two sequential stages of the traffic events processing are identified: (1) natural language segmentation; (2) semantic understanding of the events. Figure 1 shows the flow of the processing.

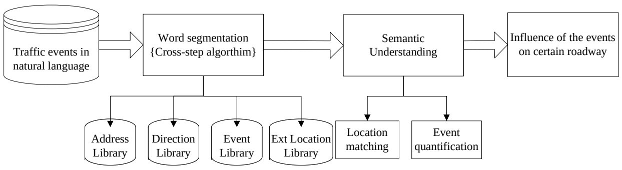

Figure 1 Technical work flow for traffic events processing

Here we use a novel MM algorithm to identify the key words. The classical maximum matching (MM thereafter) algorithm is working as follows: denote D as the word library, Max as the longest length in D , Str as the string needed segment. The MM algorithm searches in D with the substring from Str at the length of Max . If the substring matches with the word, the substring is denoted as a word and the pointer moves on the whole sentence. If no word matches the substring, the length of the substring should be reduce and search again (Feng et al., 2006).

Since the Max is longer than most of the words length, the classical MM algorithm may cause low efficiency. Aiming at this specific issue, we calculate the length in the word libraries and use the appropriate length as the initial length of the substring. Taking the address library as an example:

Record Length	Number	Ratio/%
2	29	0.71
3	797	19.48
4	1 480	36.18
5	1 218	29.77
6	427	10.44
>=7	140	3.42
total	4 091	100

Table 1 Length Distribution of Address Records in traffic events

We can see that more than 90% of the records have a length more than two characters. So here we use two as the initial length for the algorithm. A brief work flow of the improved MM algorithm is shown in figure 2:

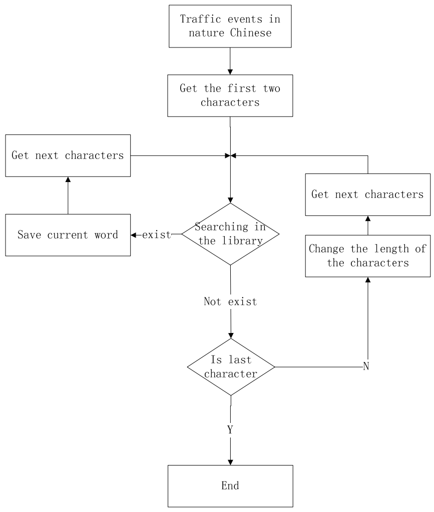

Figure 2 Work flow for word segmentation

This algorithm sets corresponding steps of word segmentation for address, direction and event libraries, and improves the one step running of the string pointer in classical Chinese word segmentation to flexible multiple steps running, so as to aggregate possible Chinese words efficiently. The proposed algorithm runs 10 times faster than an improved MM algorithm, while keeps similar accuracy and robustness.

The semantic understanding of the traffic events is about how to locate the address to the roadway and quantify the influence of the event. The traffic events reported in nature language most represented by the LRM(Linear Reference Methods). The representation can be defined by two ways shown in table 2.

Formula	Instance
Position Point + Direction＋Offset	北沙滩桥往东 100 米
Position Line + Direction + Offset	大屯路往东 100 米

Table 2 The standard LRM Formula

After the word segmentation, the key words can be used to match the roadway. The geographical positions of the addresses are used to confirm the position point/line, the direction and offset are transformed into distance in geographical coordinate system. The matching process is carried out by four steps.

1. Finding position point/line

According to different address types (Roadway, Intersection, POI, Overpass), using the geographical information such as the starting position, end position to confirm the position of the event.

2. Detecting the position type

Figure out the event is posited by point or line.

3. Matching the roadways

If the event is described by position point, find the start roadway which nearest the position and has the same direction with the offset. If the event is described by position line, the roadway matches the position line is the start roadway. Then search the roadways connected with start roadway to identify the end roadway. The searching considers both direction and offset distance which means the angle between the roadways should follows the direction and the cumulative length of the

roadways should larger than the offset. The traffic condition on the matched roadways will be effect by the traffic event.
The influence that the traffic events brings is reflected through the possible driving speed lost on underlying roadways, which is argued to be correlative with the event types and degrees.

The real-time traffic and events undoubtedly pose important influence on the turn costing. In our study, the famous HCM is utilized to model the influence of the real-time driving speed on the turn costing, with a precondition that the driving speed has a positive relation with the traffic flow.

4. FORECASTING METHOD

4.1 Travel time forecasting

A real-time traffic forecasting approach based on BP (Back Propagation) neural network is utilized in our study. If the traffic condition is been assessed as normal condition, which contain the recurrent congestion and the smooth traffic, the input will consider the spatial-temporal characteristics of travel time cost, the change pattern of city traffic is identified and adopted to obtain the rules of travel time cost forecasting using BP neural network. Otherwise, if the traffic condition is been assessed as abnormal condition which refers to non-recurrent traffic condition, the input is the speed in the past time intervals on the roadway as well as the speeds on the upstream and downstream roadways.

It is well known that the dynamic characteristics of traffic are very complicated and restricted by the statistics granularity of the traffic time series. When the granularity of the traffic time series is too small(less than 2 minutes), the randomness plays the key role which makes the time series too complex. If the granularity of the traffic time series are too large(more than 15 minutes), the real time traffic influence in the future time segment can be ignored. Based on the empirical results, we use 5~15 minutes as the granularity of the traffic time series.

4.2 Forecasting module

The BP neural networks are trained independently for each roadway. In normal traffic condition, two types of traffic data are required for the BP neural networks training when forecasting the travel time cost on roadway L_i at time $T+t$:

(1) Historical data:
The historical data includes the historical speed of the roadway L_i at time $T+t$ on the same day of the week.

(2) Real-time data:
The real-time data includes the real time collected speed of the roadway L_i before time $T+t$ which means the time intervals before T .

The real travel time cost of the roadway L_i at time $T+t$ is used as the teacher to supervise the training.
After well trained, the travel time cost on each roadway in the future time segment can be forecasted.

In non-recurrent congestion, two types of traffic data are required for the BP neural networks training when forecasting the travel time cost on roadway L_i at time $T+t$:

(1) Traffic environment data:
The real-time collected speeds at time T on the roadways which are the upstream and downstream to roadway L_i

(2) Real-time data:
The real-time data includes the real time collected speed of the roadway L_i before time $T+t$ which means the time intervals before T .

The real travel time cost of the roadway L_i at time $T+t$ is used as the teacher to supervise the training.

After well trained, the travel time cost on each roadway in the future time segment can be forecasted.

5. CASE STUDY

Traffic network of Beijing, China is selected as test bed for this study. The traffic information is gathered by the floating car data in Beijing. Floating car data (FCD) is a method to determine the traffic speed on the road network. It is based on the collection of localization data, speed and direction of travel and time information from driving vehicles.

5.1 Choice of testing datasets and data pre-processing

The raw data is the mean speed data of over 2000 major roadways collected every five minutes in Beijing from July 1st to September 30th, 2007. Since the day of the week effect the traffic, our experiments are taken separately on each day of the week.

Rough data caused by the data collecting, transmitting and matching with the road network can not be used directly since the outliers and missing data are too numerous. Hence, we provide a inverse time weighted interpolation and completion algorithm including two steps:

1). Outliers detection is based on the following two criteria:
(a) Detection of two low speed measures, below 5km/h for more than one hour;
(b) Detection of inconstant speed measures, less than half of the average speeds for less than 10 minutes.

2). Missing data completion is carried out by taking the weighted mean of the available data.

5.2 Schematic System Structure

The system framework is presented as shown in Figure 3. This figure illustrates the logical sequence of how proposed forecasting system operates.

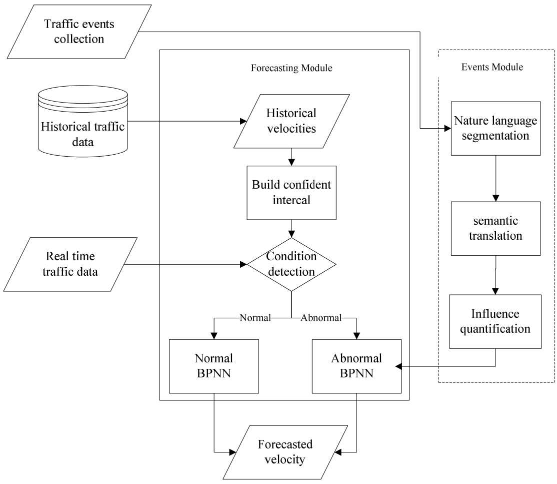

Figure 3 Schematic structure of system design

5.3 Procedure:

The template model is carried out for each roadway by statistical the historical data. Here we take the FuChengRd as an example. The velocities collected on this roadway around 8:00 AM on weekdays are shown in Figure 4.

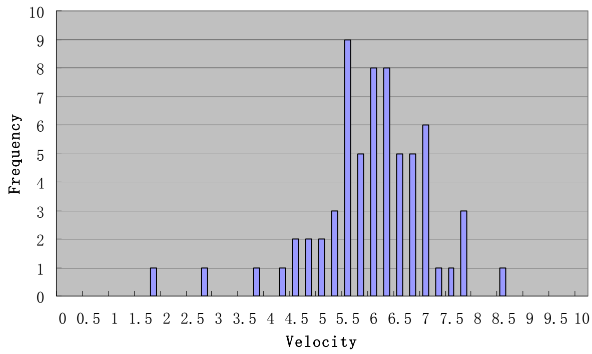

Figure 4 Velocity collected on FuChengRd

The distribution is tested as Gaussian distribution. The confident interval at 8:00 AM on weekday can be calculated according to the formula in Gaussian distribution. And the other time intervals are calculated in the same way. In Figure 5, the cyan line shows the template model on FuChengRd on weekdays.

The forecasting results on FuChengRd are shown in figure 5. It shows the velocity on this roadway from 5 AM to 8 PM on one Tuesday. The box in the figure highlight the non-recurrent congestion detected by the proposed approach and forecasted by the BP neural network for abnormal traffic condition.

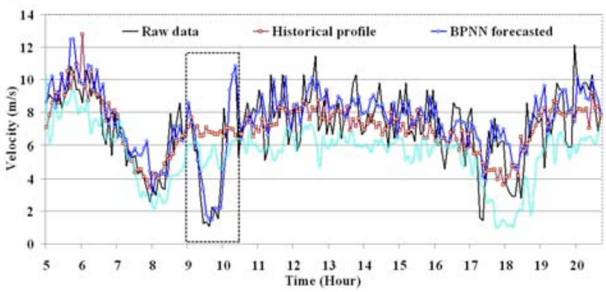

Figure 5 Velocity on roadway

The traffic events are got from the Beijing Traffic Radio in Chinese nature language at first, and then be syncopated and parsed into segmented words by semantic rules. The segmented words are translated into address, direction, event and location by the four word libraries using the improved maximum matching algorithm that presented in this paper. After that, the address is matched with map by geometrical information.

The traffic events are collected from Beijing Traffic Radio in Chinese nature language firstly (as shown in figure 4), and then be syncopated and parsed into segmented words with semantic rules. The segmented words are translated into address, direction, event and location libraries using a cross-step word segmentation algorithm. Then the resulted traffic influence is matched with road network maps for further applications
When the BP neural network based traffic forecasting server receives the real-time traffic and events information, the real-time traffic condition is detected and the forecasted driving speed in next 5~15 minutes for every roadways is obtained from the well trained network, and the network is adjusted at the same time.

Fig.6 shows an example for natural Chinese understanding. By the use of the improved MM algorithm proposed above, the sentence “阜成门桥北向南方向中间车道现在发生两车事故造成后车行驶缓慢” is segmented into five key Chinese words, namely, “阜成门桥” as an address, “北向南” as a direction, and “事故” and “行驶缓慢” as two events.

ID	路况信息
386	西长安街东西双向现在采取临时的管控措施
385	受西客站的影响，莲西路西向东方向车多一直到了南沙窝桥
384	阜成门桥北向南方向中间车道现在发生两车事故，造成后车行驶缓慢
383	十八里店桥到大洋坊桥北向南方向车行缓慢
382	十八里店北桥和南桥北向南方向的辅路车行缓慢
381	天宁寺桥内环方向最外侧车到有事故，请候车小心
380	八达岭高速出京方向沙河桥上行驶非常缓慢
379	保福寺桥的辅路由东往西方向行驶缓慢

Figure 6 Traffic events described in natural language

Then the address is matched with the roadway “阜成门北大街” in the spatial data, and the events effect the traffic by changing the velocity of this matched roadway.Fig.7 shows the how this event effects the traffic. The roadway “阜成门北大街” is turned red since a congestion occurred here and the velocity is very slow.

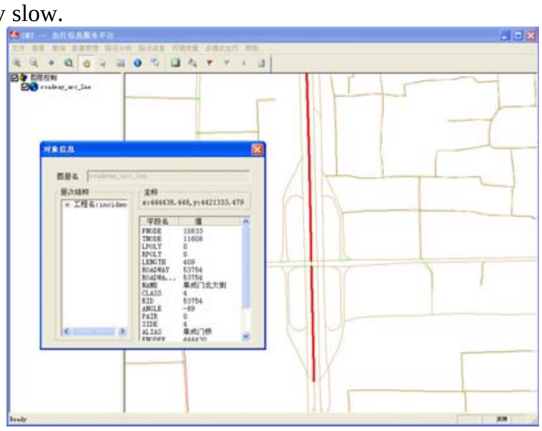

Figure7 Influence caused by a traffic event

The forecasting is taken by different BPNN models according to the certain traffic condition. Here we use historical profile as the comparative study. Historical profile is calculating the average of historical traffic data as the forecasted traffic data. The forecasting is taken independently with the forecasting time interval as 5 minutes, 10minutes and 15minutes. The results carried out at normal conditions are shown in table 3.

Day	AVG	BP5	BP10	BP15
Mon	0.122	0.121	0.124	0.125
Tue	0.105	0.11	0.111	0.114
Wed	0.115	0.118	0.119	0.121
Thu	0.115	0.115	0.118	0.120
Fri	0.113	0.108	0.111	0.114
Sat	0.125	0.123	0.125	0.126
Sun	0.128	0.126	0.129	0.130

Table 3 Forecasting errors on normal conditions

The forecasting errors on detected abnormal condition are shown in table 4.

Method	Error
Average	1.27
BPNN in next 5 min	0.546
BPNN in next 10 min	0.689
BPNN in next 15 min	0.995

Table 4 Forecasting errors on abnormal conditions

6. CONCLUSION AND DISCUSSION

In this paper, we proposed a new method to forecasting travel time costs on roadways by identify the traffic condition in real time. The traffic conditions are detected by a spatial-temporal data mining and differentiating as recurrent and non-recurrent conditions. Different input data and neural networks are used according to the certain condition. The approach can take advantages both in historical and real-time traffic information. The changes in traffic patterns can be well taken into consideration to enhance the accuracy of the forecasting. The results show that the process can enhancing short-term travel time cost forecasting especially in the abnormal conditions.

Further exploration of this topic may include other detection methods with other distance metrics and other forecasting equations as well。Since the point here is to detect the different traffic condition and to recognize the traffic incident, the influence of the traffic incidents learned in the paper are only used to forecasting the changes of velocity in a very short time. For the future work, the estimation of the influence both in spatial and temporal should be well discussed.

ACKNOWLEDGEMENTS

This research is supported by the Hi-Tech Research and Development Program of China under grant No. 2006AA12Z209 and 2007AA12Z241; Chinese NSFC program under grant No. 40871184. Their supports are gratefully acknowledged.

REFERENCES

Ben-Akiva, M., Bierlaire, M., Burton, D., Koutsopoulos, H. and Mishalani, R., 2001. Network state estimation and prediction for real-time transportation management applications. Networks and Spatial Economics, 1(3/4): 293-318.

Feng, L., Zhu, L. and Zhang, J., 2006. A Model for Ontology driven Web Information Retrieval and Its Realization. Journal of the China Society for Science and Techenical Information, 25(003): 276-281.

Hu, T.Y., 2001. Evaluation framework for dynamic vehicle routing strategies under real-time information. Transportation Research Record: Journal of the Transportation Research Board, 1774(-1): 115-122.

Jin, Y., Dai, J. and Lu, C.T., 2006a. Spatial-Temporal Data Mining in Traffic Incident Detection, SIAM DM 2006 Workshop on Spatial Data Mining. Citeseer, Bethesda.

Jin, Y., Dai, J. and Lu, C.T., 2006b. Spatial-Temporal Data Mining in Traffic Incident Detection, SIAM Conference on Data Mining. Citeseer, Maryland, USA.

Mark, C.D., Sadek, A.W. and Rizzo, D., 2004. Predicting experienced travel time with neural networks: a PARAMICS simulation study, pp. 906-911.

Smulders, S.A., Messmer, A. and Knibbe, W., 1999. Real-time application of metanet in traffic management centres.

Turochy, R.E., 2006. Enhancing Short-Term Traffic Forecasting with Traffic Condition Information. Journal of Transportation Engineering, 132(6): 469-474.

Van Lint, J.W.C., Hoogendoorn, S.P. and van Zuylen, H.J., 2005. Accurate freeway travel time prediction with state-space neural networks under missing data. Transportation Research Part C (13): 347-369.

Williams, B.M., 2001. Multivariate vehicular traffic flow prediction: Evaluation of ARIMAX modeling. Transportation Research Record: Journal of the Transportation Research Board, 1776(-1): 194-200.

Wu, C.H., Ho, J.M. and Lee, D.T., 2004. Travel-time prediction with support vector regression. IEEE transactions on intelligent transportation systems, 5(4): 276-281.

Yin, H., Wong, S.C., Xu, J. and Wong, C.K., 2002. Urban traffic flow prediction using a fuzzy-neural approach. Transportation Research Part C, 10(2): 85-98.

Zhu, J., 2009. A self-learning short-term traffic forecasting system through dynamic hybrid approach, 11th International Conference on Computers in Urban Planning and Urban Management, Hong Kong, China.