TRAVEL TIME PREDICTION BASED ON PATTERN MATCHING METHOD

Jiandong Zhao¹, Feifei Xu¹, Wenhui Liu¹, Jigen Bai² and Xiaoling Luo²

¹School of Mechanical, Electronic and Control Engineering, Beijing Jiaotong University, Beijing, 100044, China
²Beijing Boyotod Technology Co., Ltd, Beijing, 100078, China
Emails: zhaojd@bjtu.edu.cn

Abstract- The microwave detection technology has become an effective tool for monitoring highway traffic flow in China. The cross section traffic data collected provides models opportunity for travel time prediction. However, the sparseness of data somewhat constrains the prediction accuracy. To tackle this problem, the paper presents a highway travel time prediction algorithm based on pattern matching method. First, a pattern library is established by choosing traffic volume and speed as its certain state components and time as its uncertain state component. Then, the space-time two-dimensional linear interpolation method is used to calculate the mean speed and subsequently the travel time. Finally, similar patterns are obtained using K Nearest Neighbor approach and predicted travel time is calculated by the Weighted Average method. The case study shows that the pattern matching method for travel time prediction based on microwave detection data produces sufficient accuracy, which solves the problem of sparse detectors effectively.

Index terms: Travel time prediction, Pattern matching method, Microwave detection data, Two-dimensional linear interpolation method.
I. INTRODUCTION

Travel time prediction is crucial to the development of intelligent transportation systems (ITS) and advanced traveler information systems (ATIS). Travel time is an important indicator for traffic operation and the degree of congestion [1]. Real-time traffic information can help travelers to make decisions about route choices, as well as traffic managers implementing better traffic control strategies. In recent years, traffic congestions are becoming more frequent on most roadways in China, which increases the uncertainty of vehicle travel time. Therefore, accurately predicting travel time in real-time has become a critical issue for advanced traveler information systems (ATIS) [2].

Highway travel time data [3] mainly come from fixed detectors and toll stations. In China, the well-developed toll system can provide real-time and complete data. Since 2011, based on the toll data, the authors carried out series of research about travel time prediction by using the Kalman filter algorithm. At the beginning, using the entry toll station time and export toll station time, we created historical travel time series. Then, the Kalman filter model was build based on these series. At the same time, a piecewise linear interpolation method was introduced to help to reduce the prediction error. In 2012, a system of travel time prediction was developed and applied. Case study results showed that in the normal, the accident and holiday states, the APE of all prediction periods is less than 10% and the APE of accident prediction periods is less than 13% [1].

In the subsequent study, when the traffic flow between two stations was unstable, we found that the Kalman filtering model can result in poor adaptability, even inaccurate prediction. Therefore, the adaptive algorithm was introduced to construct the Sage-Husa adaptive Kalman filter prediction model. Case study results showed that, in the normal, the accident and the holiday state, the APE of all prediction periods is less than 7.5% and the APE of accident prediction periods is less than 10% [4].

However, there are still some shortcomings in predicting travel time with the toll data and the Kalman filtering model:

1. The distance between two toll stations is more than 10 kilometers, which often result in a large error;
2. The uncertain parking payment time can lead to large non-linear prediction errors.
For highway travel time prediction, another data mainly comes from fixed detection equipment installed 150-800 meters apart [1]. For highways in China, fixed detection equipment generally includes loop detectors, video detectors and microwave detectors. Among those techniques, loop detectors are only installed on ramps of expressways 10 kilometers apart, which cannot meet the data requirement for travel time prediction. Video detectors can replace loop detectors for wider area coverage, and provide real-time collection of traffic parameters. However, video detectors are greatly influenced by poor weather conditions, and can’t work well in poor visibility conditions.

In recent years, to solve the above problems, the microwave detection technology has been continuously developed and widely applied. Microwave detectors have the advantages of working under all weather conditions, high accuracy and low malfunction rate. What’s more, those devices can be quickly installed without destroying road surface and blocking existing traffic flow. The microwave detection technology is becoming another effective highway monitoring means following video monitoring, which provides opportunities for predicting highway travel time.

So, we want to use microwave detection data to predict the travel time. But, we should consider the following three questions:

1. The distance between microwave detectors is 3-4 kilometers apart in China. Can the Linear interpolation solve the problem of sparse data?
2. Is the Kalman filter algorithm appropriate for the microwave detection data? And should an accurate mode be built?
3. How to evaluate the prediction errors, and are the APE and the MAPE suitable?

II. LITERATURE REVIEW

Many studies on travel time prediction models have been conducted in previous decades. Tooru and Kana (1999) [5] collected occupancy, traffic volume and speed data by setting loop vehicle detectors 150 meters apart. Then, they used the pattern matching method to predict travel time, and found the most similar historical traffic state to apply to the real-time status. Steven (2003) [6] collected data by setting roadside testing equipment, and predicted the actual highway travel time using the Kalman filtering algorithm. He pointed out that the Kalman filtering algorithm had
the advantage of updating state variables in real-time. On the other hand, real-time and historical travel time data affected the travel time prediction accuracy. Lelitha (2007) [7] obtained occupancy, traffic volume and speed data from loop vehicle detectors with intervals of 250 meters. To resolve the problem that there was few sample data, they used support vector machines for short-term travel time prediction. Sun Lijun (2008) [8] developed a simulating model for forecasting traffic state. He obtained density and speed data from vehicle detectors with intervals of 300-600 meters, and applied the extended Kalman filtering theory into the macroscopic dynamic traffic flow model. Finally, travel time was calculated using the method of virtual vehicle. Wosyka et al. (2012) [9] built up the Decision Tree Model to estimate travel time on urban roads. They obtained actual travel time by utilizing the License Plate Recognition system and traffic flow data from loop vehicle detectors. Yeon et al. (2013) [10] used the Markov Chain Algorithm to simulate and forecast travel time when traffic was congested. The authors pointed out that the quantity of vehicles related to travel time prediction accuracy.

Reviewing the literature, travel time prediction methods include generalized linear regression [11, 12], pattern matching, neural network models[13], Kalman filters approach[14], and support vector machines [15, 16] et al.

Linear Regression methods have a higher accuracy in stable state of traffic flow. However, this method has poor real-time performance and a lower accuracy in an unstable state of traffic flow. Neural Network Models can identify complex non-linear systems with strong self-learning skills, but model training needs a large number of samples. And furthermore Neural Network Models cannot always guarantee the convergence speed and global convergence. The Kalman Filtering approach has a high prediction accuracy and flexible selection of predictors. The Kalman Filtering model is modified recursively according to actual travel time. However, this method has poor nonlinear performance, and needs adjusting weights each time in calculation.

Unlike the general analytic method, the pattern matching method is an intelligent approach without a fixed model. This method has the following advantages:

1. It does not require prior knowledge or much parameter identification;
2. It can give good results as long as there is enough historical data;
3. It has a strong ability to respond to emergencies, better prediction accuracy and error distribution. Last but not least, the pattern matching principle is clear and suitable for nonlinear and uncertain systems.
The uncertainties of traffic incidents such as traffic accidents and congestion have a serious effect on travel time prediction. Therefore, in this paper, in order to reduce these influences, the pattern matching method is used to predict travel time. The space-time two-dimensional linear interpolation method is used to solve the problem of data sparseness. The Absolute Percentage Error (APE) and the Mean Absolute Percentage Error (MAPE) are adopted to evaluate the prediction performance.

III. TRAVEL TIME DATA

3.1 Microwave Detector Data Sources

In order to collect sufficient data, we choose a section of the G1 Expressway in Beijing as the experiment section. The section is 27.13 kilometers long with 6 toll stations and 11 microwave detectors. Figure 3.1 shows the deployment of toll stations and microwave detectors. Bailu is a mainline toll station and others are interchange toll stations.

![Figure 3.1 Deployments of Toll Stations and Microwave Detectors](image)

10 fields from the traffic data detected by microwave detectors were selected as basic data and highlighted in bold in Table 3.1.
Table 3.1: Traffic Data Detected by Microwave Detectors

<table>
<thead>
<tr>
<th>NO.</th>
<th>NAME</th>
<th>NO.</th>
<th>NAME</th>
<th>NO.</th>
<th>NAME</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>occur_time</td>
<td>7</td>
<td>speed95</td>
<td>13</td>
<td>direction_id</td>
</tr>
<tr>
<td>2</td>
<td>device_id</td>
<td>8</td>
<td>occupy</td>
<td>14</td>
<td>lane_id</td>
</tr>
<tr>
<td>3</td>
<td>sensor_no</td>
<td>9</td>
<td>vehicle_length</td>
<td>15</td>
<td>road_id</td>
</tr>
<tr>
<td>4</td>
<td>unit_id</td>
<td>10</td>
<td>distance</td>
<td>16</td>
<td>transferred</td>
</tr>
<tr>
<td>5</td>
<td>intensity</td>
<td>11</td>
<td>car_class</td>
<td>17</td>
<td>repair_flag</td>
</tr>
<tr>
<td>6</td>
<td>speed</td>
<td>12</td>
<td>integration_period</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3.2 Data Processing

Microwave detectors can detect the traffic characteristics of each lane. The following equation is used to estimate the average speed at a cross section:

\[
\bar{V} = \frac{\sum \text{speed} \cdot \text{intensity}}{\sum \text{speed}}
\]  

(3.1)

where, \(\bar{V}\) = the average speed of all vehicles across the detection cross section; \(\text{speed}\) = vehicles’ average speed of each lane in the detection period, which is the value of speed in the database; \(\text{intensity}\) = the traffic volume of each lane in the detection period, which is the value of intensity in the database.

According to equation (3.1), we obtained the results of one cross section speed. In Figure 3.2, speed frequency shows a normal distribution.

![Figure 3.2 Frequency Distribution Diagram of Speed](image_url)
IV. METHODOLOGY

The core of the pattern matching method is searching for historical traffic flow patterns that are similar to current traffic pattern. The procedures are described as follows:

Step One: constitute a pattern through determining state vector;
Step Two: calculate travel time and establish a pattern library;
Step Three: search for similar patterns from the pattern library;
Step Four: predict travel time through calculation.

4.1 State Vector Selection

The state vectors are the foundation of pattern matching. We selected traffic volume and speed as the important state vectors.

It is commonly known that traffic flow varies by time and space. Currently in China, truck volume at night is significantly higher than that during daytime. As shown in Figure 4.1, the proportion of trucks between 22:00-6:00 was higher than the proportion by day with a top rate of 43%. During the daytime, the proportion of large vehicles is less than 10%. Therefore, time is selected as the third state vector, and is divided into daytime and nighttime.

Finally, the state vectors are \([T, S, Q]\), where \(T\) is time period, \(S\) is the vehicle mean speed, and \(Q\) is traffic volume. In fact, many other factors, such as the proportion of large vehicles, traffic events, and roadway types, might have an effect on traffic flow characteristics. Based on the same principles, these factors can be added according to the actual need.

![Figure 4.1 Proportion Change Curves of Different Types of Vehicles from Taihu to Zhangjiawan.](image-url)
4.2 Calculation of Travel Time

4.2.1 Calculating mean speed based on two-dimensional linear interpolation

Microwave detectors were installed along the section at 3-4 kilometers apart. While calculating spot speed between detectors with harmonic mean method or single space interpolation method, we found that the large distance caused the large nonlinear prediction error and the low prediction precision. Therefore, the time-space two-dimensional linear interpolation method was adopted to calculate the mean speed. Figure 4.2 shows the principle of calculating mean speed of cross section. The procedures are described as follows:

Figure 4.2 Average Speed Calculation Principles.

Step One: Calculate the mean speeds of two adjacent detectors using the time linear interpolation method. The equations are as follows:

\[
V_{d,t} = \frac{v(d,h) + (v(d,t) + v(x,t))}{2} \tag{4.1}
\]

\[
V_{d+h} = \frac{v(d,h) + (v(d,t) + v(x,t))}{2} \tag{4.2}
\]

where, \( V_{d,t} \) = the mean speed of detection point \( d \) at time \( t \) in the period \( h \); \( V_{d,h} \) = the mean speed of detection point \( d \) in the period \( h \); \( V_{d+h} \) = the mean speed of detection point \( d+h \) at time \( t \) in the period \( h \); \( t_0 \) = beginning of the period \( h \); \( \Delta \) = length of a period.

Step Two: Calculate the mean speed of a cross section between two adjacent detectors by using the space linear interpolation. The equation is expressed as follow:
where, \( V_{x,t} \) = the mean speed of point \( x \) on road at time \( t \); \( x, t \) = the position of vehicle on road at time \( t \) in the period \( h \); \( x_d \) = the location of detection point \( d \) on road; \( l = distance between detection point d and d' \).

4.2.2 Calculating travel time by piecewise method

As shown in Figure 4.3, A and B are the toll stations, \( D_i \) is the microwave detector, the road AB can be divided into several sections such as \( AD_1, D_1 D_2, D_2 D_3, \ldots, D_n B \), and \( L_i \) is the length of each section. The procedures of adopting the piecewise method to calculate travel time are as follows:

\[
V_{x,t} = \frac{1}{l} \int_{l}^{l} \cdot \quad (4.3)
\]

Figure 4.3 Sections of the road AB

Step One: Calculate the speed at the toll station by using the time and space interpolation method based on the mean speeds from the detectors on both sides of the toll station;

Step Two: Calculate \( T_i \) of each road section \( L_i \) between toll station A and B;

1) If \( L_i \leq \frac{1000}{i} \) meters, then \( \bar{v} = \frac{1}{2} \cdot \frac{v}{v} \), \( T_i = \frac{1}{v} \)

where, \( v_i = cross section mean speed; \bar{v} = mean speed of road section \( L_i \). \)

2) If \( L_i > \frac{1000}{i} \) meters, the road can be divided into \( j \) segments

\[
j = \left[ \frac{r}{\bar{v}} \right].
\]

Subsequently, \( T_i \) is the sum of the travel times of each segment, and can be calculated by the piecewise method.

Step Three: Calculate the travel time between toll station A and B by \( t = \sum T_i \).
4.3 Establishment of Pattern Library

The patterns stored in the library, which consist of the state vector \([T, S, Q]\) and decision attribute [t] (travel time). The procedures for building up Pattern library A_B are as follows:

Step One: Choose traffic volume and speed at station A as state vectors. However, the traffic volume cannot be measured directly. Therefore, the traffic volume for the first detector after station A is chosen to replace that at station A.

Step Two: Count ranges of traffic volume and speed, then divide them into different intervals. In order to accelerate the searching speed of the matching process, every interval is coded. When matching is being performed, the search only needs to be in a corresponding speed (S) – volume (Q) range.

Based on the data from September 1, 2013 to October 24, 2013, the intervals of S and Q were analyzed and divided. As shown in Table 4.1, the traffic data accorded with traffic engineering theory.

Step Three: Fill the pattern library with the mean speed, traffic volume and corresponding travel time.

The structure of the pattern library is shown in Figure 4.4. The total library consists of different libraries between two stations. A, B, C, and D present different toll stations. Each library contains speed, traffic volume, speed code \(i\), traffic volume code \(R\), and time code \(Time\_number\).

<table>
<thead>
<tr>
<th>Q</th>
<th>S</th>
<th>0-50</th>
<th>51-60</th>
<th>61-70</th>
<th>71-80</th>
<th>81-90</th>
<th>≥91</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0-35</td>
<td>0</td>
<td>0</td>
<td>35</td>
<td>1584</td>
<td>1439</td>
<td>14</td>
</tr>
<tr>
<td>0</td>
<td>36-70</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>248</td>
<td>1824</td>
<td>235</td>
</tr>
<tr>
<td>1</td>
<td>71-105</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>50</td>
<td>1268</td>
<td>1257</td>
</tr>
<tr>
<td>2</td>
<td>106-140</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>38</td>
<td>1047</td>
<td>3117</td>
</tr>
<tr>
<td>3</td>
<td>141-175</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>249</td>
<td>750</td>
</tr>
<tr>
<td>4</td>
<td>176-210</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>35</td>
<td>163</td>
</tr>
<tr>
<td>5</td>
<td>211-245</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>79</td>
</tr>
<tr>
<td>6</td>
<td>246-280</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>9</td>
<td>32</td>
</tr>
<tr>
<td>7</td>
<td>281-315</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>8</td>
<td>≥316</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>0</td>
</tr>
</tbody>
</table>
4.4 Travel Time Prediction

4.4.1 Searching Strategy

In order to find out the similar patterns from the pattern library, the K-nearest neighbor search strategy is adopted. At first, set a number K. Then, search for K patterns which are the nearest patterns. If the number of patterns is less than K, all of these will be used. If not, only the nearest patterns will be used.

\[ d_{W,P} \] is set as the distance between all patterns \( P = \ldots \) and the match state \( W \), The value of \( d \) can be used to describe the state similarity.

**Definition 1:** The distance between the two state vectors.

According to the Euclidean distance, \( d \) is defined as

\[
    d_{W,P} = \sqrt{w - W + T_p - Q_w} \quad (4.4)
\]

**Definition 2:** search area

The search area of the state vector \( W = Q_w \) is the set in the corresponding sub-library, where \( T_p = \ldots \).
4.4.2 Travel Time Prediction

Based on the nearest neighbor data, the weighted average method is then used. Set K nearest neighbors as $P_1 \sim P_k$ respectively, in which time values are $t_1 \sim t_k$. The distances between $P_i \sim P_k$ and W state are $d_1 \sim d_k$, the prediction time can be calculated as follows:

$$t' = \sum_{i=1}^{k} \beta_i$$ \hspace{1cm} (4.5)

$$\beta = \sum_{j=1}^{k}$$ \hspace{1cm} (4.6)

where, $t'$ = the predicted time, $\beta$ = the weight value.

V. APPLICATION

The Beijing section of the G1 Expressway is closely linked to the capital Ring Roads. The distance between the sections entering Beijing from the Taihu toll station to the Zhangjiawan toll station is 5.11 kilometers, and from the Taihu toll station to the Huoxian toll station is 13.18 kilometers. Due to the distinct time-space characteristics of traffic flow, we chose the two sections to verify the algorithms. We predicted the travel times of the normal and holiday respectively. We then compared the prediction accuracy of the dividing time mode before dividing time mode. The prediction time period is 5 minutes.

5.1 Prediction Performance Indices

In order to evaluate the result of predicted travel time, we adopted the Absolute Percentage Error (APE) and the Mean Absolute Percentage Error (MAPE) to evaluate the accuracy of prediction.

$$APE(t) = \frac{T}{\sum_{R(t)}^{I_{R(t)}}} \times$$ \hspace{1cm} (4.7)

$$MAPE = \frac{1}{L} \sum_{L}^{I_{R(t)}} \times$$ \hspace{1cm} (4.8)
Where, $\text{APE}(t) = \text{the relative error of period } t$; $T_{P(t)} = \text{the predicted travel time of period } t$; $T_{R(t)} = \text{the actual travel time of period } t$; $\text{MAPE} = \text{the average relative error}$; $L = \text{the number of total prediction periods the whole day}$.

5.2 Optimization of K

By selecting different values of K, we can study and analyze the prediction accuracy. Firstly, from Taihu to Zhangjiawan, with the change of K, we determined the fit curve between the relative error and K. Then, we determined the fit curve from Zhangjiawan to Huoxian. As shown in Figure 5.1, for K value, using a larger sample distance is not always better. The best K value is affected by the pattern library. The best prediction result was when K=20.

![Figure 5.1](image1.png)

![Figure 5.1](image2.png)

Figure 5.1 (a) K value-error curves from Taihu to Zhangjiawan, (b) from Zhangjiawan to Huoxian
5.3 Prediction Results and Error Analysis

5.3.1 Normal Traffic Flow

Figure 5.2 (a) shows travel time prediction results before dividing time and after dividing time under a normal traffic flow. We can see that both predicted travel times are close to the actual travel time. Meanwhile, the prediction result of dividing time is better than before dividing time in several periods. Figure 5.2 (b) shows the relative errors of travel time. We can see that the APE of dividing time is smaller. As shown in Table 5.1, the MAPE of dividing time is reduced to 2.4%. And prediction accuracy is applicable to travel time prediction in normal traffic flow.

Figure 5.2 (a) Travel time prediction result curves in normal traffic flow,
(b) APE curves in normal traffic flow
Table 5.1 Travel Time Prediction Performance Indices in Normal Traffic Flow

<table>
<thead>
<tr>
<th>Performance Indices</th>
<th>APE Section of All Periods</th>
<th>MAPE of All Periods</th>
</tr>
</thead>
<tbody>
<tr>
<td>after dividing the time</td>
<td>(-10%,11%)</td>
<td>2.4%</td>
</tr>
<tr>
<td>before dividing the time</td>
<td>(-10%,11%)</td>
<td>2.8%</td>
</tr>
</tbody>
</table>

5.3.2 Holiday Traffic Flow

Figure 5.3(a) shows two kinds of travel time prediction curves with holiday traffic flow. We can see that holiday traffic flow was unstable; and that traffic jams occurred from 10:00 to 12:00. We can also see that the two predicted travel times are relatively close to the actual travel time. The APE curves of travel time prediction are shown in Figure 5.3(b), and travel time prediction performance indices are shown in Table 5.2. Corresponding to the two methods, the MAPEs are reduced from 3.9% to 3.4%. The prediction accuracy of dividing time is applicable to travel time prediction in holiday traffic flow.
FIGURE 5.3 (a) Travel time prediction result curves in holiday traffic flow,  
(b) APE curves in holiday traffic flow

<table>
<thead>
<tr>
<th>Performance Indices</th>
<th>APE Section of All Periods</th>
<th>MAPE of All Periods</th>
</tr>
</thead>
<tbody>
<tr>
<td>after dividing the time</td>
<td>(-10.7%,10.0%)</td>
<td>3.4%</td>
</tr>
<tr>
<td>before dividing the time</td>
<td>(-13.8%,12.1%)</td>
<td>3.9%</td>
</tr>
</tbody>
</table>

VI. CONCLUSIONS

By using the cross section traffic data collected from microwave detectors, this paper presents a 
highway travel time prediction algorithm based on the pattern matching method. Because the 
sparseness of data somewhat constrains the prediction accuracy, the space-time two-dimensional 
linear interpolation method is used to calculate the mean speed. Secondly, traffic volume, speed 
and time are selected as pattern state vectors, and then the pattern libraries are established. 
Thirdly, using the K Nearest Neighbor approach, similar patterns are obtained, and predicted 
travel time is calculated by the Weighted Average method. 
Finally, the case study shows that the pattern matching method for travel time prediction can 
produce sufficient accuracy, and effectively solves the problem of sparse detectors. The APEs of 
all prediction periods are less than 11% and the MAPEs are less than 4% for the normal and 
holiday traffic flow. After dividing periods of time, the prediction accuracy is improved.
VII. CONFLICT OF INTEREST

The authors have declared that no conflict of interest exists.

VIII. ACKNOWLEDGEMENTS

This work is supported by the Fundamental Research Funds for the Central Universities (2013JBM080). The help from the China Scholarship Council is greatly appreciated by the authors.

REFERENCES


276-281.
[18] Xiaohong Peng, Zhi Mo and Riyao Liao. Traffic signal control for urban trunk road based
on wireless sensor network and intelligent algorithm. International Journal on Smart Sensing