Regional Transportation Network Travel Time Estimation Based on Transition of Traffic Flow Phase

PengCheng Yuan¹, ZhiCai Juan²

Antai College of Economics & Management, Shanghai Jiao Tong University, Shanghai, 200052, PR China
¹danis_cx@126.com; ²zjuan@sjtu.edu.cn

Abstract- Travel time is widely recognized as an important performance measure for assessing transportation system, it is a meaningful parameter in theory and practice. Previous researches about travel time mainly focus on two areas: first, the research object is the unit of the transportation network, and there is little research on the area transportation network travel time; second, previous researches about travel time mainly focus on travel time forecast which using history traffic data. But, in fact, the confidence of the forecasted travel time can’t be tested because the instantaneous travel time can’t be measured in practice. Not only that, previous researches about the travel time forecasting are under one traffic state (congestion state), they don’t consider the impact of the traffic state variation. The traffic flow state of the area transportation network is mutative, and the travel time of the multi-state area transportation system can be estimated using the random process theory. In this paper, we divide the traffic flow into three states using C-mean fuzzy clustering method and develop the travel time function under every state. We also analyze the state changing of the unit of the area transportation network using semi-Markov random process theory and develop a travel time estimating model to evaluate the area transportation network travel time. We use two types of data to develop and verify the model. The first type of the data is traffic flow data (include traffic flow velocity, speed and occupation). The second type of the data is the travel time, which was got by the plate matching method. Using the two types data of changzhou china of 1/3-10/3(2010), we develop the model. Using the data of changzhou china of 11/3-11/3(2010), we verify the result. Estimates from the model are compared to the field-measured travel time; the paired t-test for the mean difference is conducted too. The results show that at the significance level of α=0.05, there is no significance difference between the estimated travel time and the field mean travel time.

Keywords- Regional Transportation Network; Travel Time; Random Process; Semi-Markov Chain

I. INTRODUCTION

The evaluation on regional traffic network performance has been concerned by traffic managers, traffic planning designers and traffic engineers. While, the evaluation on travel time of regional traffic network shows great significance. Since it can objectively reflect the service level of regional road network, as well as provide travellers’ with visualized and easily perceived regional road network travel parameters. However, previous researches have shown little concentration on the evaluation indexes of regional traffic network travel time. Several of them mainly focus on predicting network unit travel time, in which network travel time prediction is simply treated as the accumulation of the travel time of all units for composing a network. Since, actual traffic flow is dynamic and changes with time, so instantaneous travel time cannot be measured practically. Thereby, it is difficult to test the reliability of the previous travel time data obtained by forecasting. In addition, due to being characterized by the multi-states (i.e. being unblocked, traffic jam, congestion) of road network unit traffic flow, the regional traffic network is provided with more complex states. Thus theoretically, previous researches of travel time basing on merely one traffic state (congestive state) are limited. In practice, although the travel time of network unit in each state can not be measured, the instantaneous travel time of network unit can be induced through traffic flow velocity (T=L/v, T is instantaneous travel time, L refers to road length, and v is instantaneous flow velocity). This study attempts to estimate the average travel time of traffic network in morning peak period (7:00-10:00), evening peak period (17:00-19:00) and Flat peak (13:00-15:00), by investigating how to utilize the instantaneous travel time of network unit.

Travel time is an important index in evaluating the service level of traffic systems and shows great significances for both theoretical and practical researches. According to traffic distribution theory, the most important part in travelers’ travel cost is travel time. In practical fields, travel time is also one of the widely used evaluation parameter. To offer travelers’ varied information i.e. variable message system, many large cities in China such as Beijing and Shanghai, employ travel time as a basic traffic information data. Due to its universality and significance, in recent years, many researchers have attempted to predict it using various methods. Among which, the traffic data provided by ATIS (Advanced Transportation Information System), including GPS data, link traffic flow data (flow velocity, speed and occupation), and floating vehicle data etc., are most widely used. By analysing the correlation of travel time and these data, travel time prediction model can be established (Oh. et al., 2003; Coifman and Ergueta, 2003; Davis, 2010). The using the methods including auto-regressive moving average method (ARMA), the least square method, time series method, artificial neural network method and Kalman filtering method (Shaw, 2002; Zhang and Rice, 2003), and traffic flow prediction method (Davis, 2010). These previous methods focus on the prediction of

MTTER Volume 2, Issue 1 January 2013 PP. 68-77 www.vkingpub.com ©American V-King Scientific Publishing
travel time, which is forecasted using historical traffic data. It is so useful in traffic engineering. Especially, they can provide traveller with references of real-time travel time data. However, in fact, these methods cannot be tested due to the unavailable predictive values (actual traffic flow is dynamic and time-dependent, while real-time predictive travel time can not be measured) theoretically. Besides, the present methods mainly estimate the travel time in one traffic flow state, while the variations of the following traffic states, in particular, the variations of congestive effect brought by the sudden change of recurrent traffic state (Elefteriadou et al., 1995; Persaud et al., 2001), are rarely considered. As what indicated by the researches of Yang et al. (1999), Vanlint and Zuylan (2005), the law of travel time changing with traffic demands are absolutely different in varied traffic states. In free flow state, travel time presents little fluctuation with traffic demand changes, which can be treated as a constant. On contrast, in congestive state, it increases with the addition of traffic demands, and presents large fluctuation. Moreover, the traffic state changing law should be considered in evaluating travel time. So, the researches concerning division and changing laws of traffic states are involved.

The division of the traffic states is subjected to the classification in the condition of lacking of prior knowledge, which is a fuzzy conception that can not be defined by specific figures (Peng, 2012; Gao, et al 2009; Tian, et al 2009; Zhu 2010;). Hence some researchers investigate the discrimination of traffic state using C-mean fuzzy clustering method according to the spatial distribution characteristics of traffic flow data. Where, the traffic state identification method, such as clustering analysis, synergetic principle, time series method, traffic parameter analysis and data probabilistic method, are employed (Helbling et al., 1999; Jin 2011; Peng, et al 2012; Hu, et al, 2012 ). In addition, in practical applications, in order to simplify the calculation process and algorithm, certain one or several traffic flow parameters are utilized as state evaluation index to classify the traffic state, i.e. speed, occupancy, travel time and delay. German physicist Kerner (1996) summarized previous research results, and explored the defects of some fundamental assumptions in original traffic flow theory through a lot of practice and experiment. Furthermore, he proposed a three-phase traffic flow theory based on the characteristics of traffic flow. According to the three-phase traffic flow theory, when the traffic flow is in congestive state, the flow-density numerical distribution in equilibrium is located in a two-dimensional plane and traffic congestive state is divided into two phases: synchronized flow and wide moving congestion. Thus traffic flow is divided into three categories by three-phase traffic flow theory: free flow, synchronized flow and wide moving congestion. Among which, synchronized flow and wide moving congestion belong to congestive state. According to the statistical analysis on the characteristics of city expressway, Guan and He (2008) divided the traffic flow state of city expressway into the following four stable phases: free flow, harmonic flow, synchronized flow and congestion. Where, free flow and harmonic flow phase are in an “unblocking” traffic state, while synchronized flow and congestion phase are in a “congestive” state. In fact, the traffic flow state classification above is also indirectly reflects the basic conclusions by Yang (1999), Vanlint and Zuylan (2005), namely in different congestive states, travel time exhibits significant differences with the traffic demand changes. Additionally, Evans et al.(2001), Yeon et al.(2008) analysed the mutation features of traffic flow of expressway, based on which, they classified traffic flow state, calculated the transition probability of traffic flow state and analyzed the changing law of traffic flow state. Considering the limitations of previous researches, and especially inspired by the researches of Kernner( 1996, 2002), Evans et al. (2001), Yeon et al (2008), this study puts forward an estimation method of average travel time of area transportation network using semi-Markov chain and the changing law of network state. On one hand, the average travel time of area transportation network within a certain time is estimated by this method. So the estimated results can be compared with the field data, and the credibility of the estimated travel time can be acquired; on the other hand, the changing law of network state is fully accounted in the travel time estimation process, so it can reflect the characteristic of dynamic traffic more effectively.

Previous research assumed that vehicle speeds are a random environment and can be considered as a finite-state Markov process. The random environment process includes physical factors and traffic factors. Based on the assumption that the environmental process is a semi-Markov chain, an exact analytical expression is obtained for the Laplace Transformation of the link travel-time cumulative distribution function. Thus, to estimate the regional travel time, the impact of traffic congestion should be considered, and stochastic processes can be employed to analyze changes of regional traffic states over time and space and estimate the expected travel time considering the probability of congestion. The basic thought of this study is explained as follows: firstly, the link traffic state of regional traffic network unit is divided into three categories using C-mean clustering analysis method; considering the similarities and differences of the congestion effect of travel time in various traffic states, the travel time functions in each state are analyzed; using semi-Markov chain stochastic process theory, the stochastic changing law of each state of regional traffic network in a given period are analyzed; combining with the travel time function of the unit in each state, the estimation method of the average travel time in a given period is proposed.

This paper is organized as following: section 2 introduces the stochastic process of semi-Markov chain and C-mean fuzzy clustering method; section 3 is the model building; section 3.1 clusters the unit traffic state using C-mean fuzzy clustering analysis method; section 3.2 calibrates the instantaneous travel time functions under various states; section 3.3 classifies the regional traffic network state and calculates the limiting probability of each state using stochastic process theory; section 3.4 gives the prediction model of regional traffic network travel time; section 4 calibrates the model according to specific calculation samples and tests the estimation results of the
model; section 5 is the conclusion of this research.

II. SEMI-MARKOV CHAIN STOCHASTIC PROCESS AND C-MEAN FUZZY CLUSTERING METHOD

A. Semi-Markov Chain

A stochastic process \( \{X(T), t \in T\} \) is a set of stochastic variables; i.e. for each \( t \in T \), \( X(t) \) is a stochastic variable. Index \( t \) is often interpreted as time, \( X(t) \) is defined as the state of process lies in, set \( T \) refers to the index set of this process. When \( T \) is a denumerable set, stochastic process is noted as a discrete stochastic process. If \( T \) is a real interval, the stochastic process is called a continuous time process. The state space of stochastic process is defined as all the possible values of stochastic variable \( X(t) \).

Let \( \{X_n, n=0, 1, 2 \ldots \} \) be the stochastic process of finite values, if \( X_n = i \), then the state of process in time \( t \) is recorded as \( i \). Assuming that as long as the process is in the state of \( i \), there is a fixed probability \( P_{ij} \) who make the process in the state of \( j \) in the next period, namely, assuming that, for all state \( i_0, i_1, i_2, \ldots, i_n \), \( i, j \) and all \( n \geq 0 \), there

\[
P(X_{i+1}=j | X_0=i, X_1=i_n,...,X_t=i_t, X_{i}=i_0) = P_{ij}
\]

(1)

The stochastic process above is called a discrete time Markov process. Formula (1) can be interpreted as that, given the past state of \( X_0, X_1, X_2, \ldots, X_{n-1} \) and current state \( X_n \), the conditional distribution of future state \( X_{n+1} \) is independent of past state, and merely relies on current state. \( P_{ij} \) represents the transition probability of the process to the next state \( j \), when the process is in state \( i \). Since probability is nonnegative, and the process has to be transferred to a certain state, so we get:

\[
P_{ij} \geq 0, i,j \geq 0; \sum_{j=0}^{\infty} P_{ij} = 1, i=0,1
\]

Noting \( P \) as the matrix of the one-step transition probability \( P_{ij} \), there is

\[
P = \begin{bmatrix}
P_{00} & P_{01} & P_{02} & \cdots & P_{0n} \\
P_{10} & P_{11} & P_{12} & \cdots & P_{1n} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
P_{n0} & P_{n1} & P_{n2} & \cdots & P_{nn}
\end{bmatrix}
\]

Where, \( n \) is the state number. Write

\[
p^n_i = P(X_{n+1} = j | X_0 = i) , n \geq 0 , \quad i,j \geq 0 ,
\]

where \( p^n_i \) is called as the transition probability of the \( n \)-step, expressing the probability of the process in state \( j \) after \( n \) times of transition from state \( i \). Record

\[
\pi_j = \lim_{n \to \infty} p^n_j, \quad j \geq 0
\]

as the limiting probability of the process in state \( j \), thus \( \pi_j \) is the unique solution of formula (2) (Ross 2010)

\[
\pi_j = \sum_{i=0}^{\infty} \pi_i P_{ij}, \quad j \geq 0, \quad \sum_{j=0}^{\infty} \pi_j = 1
\]

(2)

In assumption of that a process can be in arbitrary state of \( N \) states 1,2,...,\( N \) and each state \( i \) maintains a stochastic time with mean value \( \mu_i \), then the process is transmitted to state \( j \) in probability \( P_{ij} \). If time \( T \) stays in each state obeys exponential distribution, this stochastic process is a continuous-time Markov process; otherwise, it is called semi-Markov process. The counting process based on semi-Markov is denoted as a renewal process.

For a semi-Markov process, \( X_n \) is noted as the state after \( n \) transitions, then \( \{X_n, n \geq 0\} \) is a discrete Markov process with \( P_{ij} \) as transition probability, so \( \pi_i \) can be obtained by formula (2). As long as the process has access to state \( i \), the process may stay for a expected time \( \mu_i \), then the limiting probability \( P_i \) of semi-Markov process shall be the weighted mean of \( \pi_i \), that is,

\[
P_i = \frac{\pi_i \mu_i}{\sum_{j=1}^{N} \pi_j \mu_j} \quad i=1, 2, ..., N
\]

(3)

Where, \( \pi_i \) can be obtained by formula (2).

B. C-mean Fuzzy Clustering Method

According to the differential count of C-mean fuzzy clustering, it is considered that the samples in classified sample set can be classified into a certain category by different memberships. Therefore, a certain category can be treated as a fuzzy subset of sample set, and the classification matrix corresponding to each such classification result is a fuzzy classification matrix.

Let \( U=(u_{ik})_{n \times c} \) be fuzzy classification matrix (where, \( n \) represents sample number, \( c \) refers to the classification number, \( u_{ik} \) is the membership of the \( i \) sample belonging to the \( k \) category), set \( X=\{x_1,x_2,\ldots,x_n\} \) be classified sample set, among which, each sample \( x_i \) owns \( c \) characteristic indexes, namely \( x_i=\{x_{i1},x_{i2},\ldots,x_{ic}\} \). Then sample set \( X \) is divided into \( c \) clusters. Assume that \( T \) is the centre vector of \( c \) clusters. To obtain the optimal fuzzy classification, objective function is defined as:

\[
J(U,T) = \sum_{i=1}^{n} \sum_{k=1}^{c} u_{ik}^m \|x_i-t_k\|^2
\]

Where, \( t_k \) is cluster center, \( m \) is weighted index, \( x_{i}-t_k \) represents the Euclidean distance from the \( k \) sequence to the centre of the \( i \) cluster. The objective function refers to the cluster square sum of feature points of each category to the distance centre. By selecting the \( U \) and \( T \) that satisfy the objective function, the objective function can achieve the minimum value. The detailed process of fuzzy C-mean clustering are indicated in the following:

Step1: Set \( c \), \( m \) and initial membership matrix \( U^0 \), and iteration step number \( l=0 \).

Step2: computing clustering centre \( T \).

Step3: modifying \( U \).

Step4: for a given \( \lambda \geq 0 \), the given initial value should be computed iteratively until \( \max \|u_{ik}^l - u_{ik}^{l-1}\| < \lambda \) in actual calculation, then algorithm is terminated, otherwise, \( l=l+1 \), the calculation turns to Step2.
III. THE MODEL

Semi-Markov stochastic process can be used to analyze the transition law between states and build the estimation model of the average travel time of regional traffic network. The main steps for model building are indicated as follows: First step, divide the state of the network unit (link); Second step, calibrate the instantaneous travel time functions of regional network unit in different states; Third step, classify the network state according the network unit state and network structure; Fourth step, compute the duration time of each state, the state transition probability and the limiting probability of the regional traffic network; Fifth step, estimate the regional travel time according to the state transition characteristic and the link’s instantaneous travel time function.

A. The Classification of Network Unit traffic State

In practice, the traffic flow parameters that are easy to be measured include flow rate, velocity and time occupancy. The data of the traffic flow is obtained through the detecting loops lay in the road, which is used to test the average speed and average time occupancy of vehicles in a small time period and transmits these data to the database of monitoring centre. The travel time of link is collected through the vehicle license plate recognition system installed in the entry and exit end of road, which is employed to record the plates of the vehicles passing thorough this link in a certain time period. In the monitoring centre, the travel time samples are acquired through the comparison of the time of vehicle plates entering and exiting this link in this time period. In the study, the traffic data of the express way in changzhou is employed. According to the samples of the three parameters above, network unit traffic flow state can be classified using the C-mean fuzzy clustering method motioned in section 2.2. And in actual operation, the number of the category can be determined based on personal demands. Generally, the more the unit states are, the higher the model accuracy is. However, sharper network system state will be induced as well, resulting in geometrical series increase in the calculation of model that is not beneficial for calculation. Therefore, in this study, the state of network unit is divided into three categories.

Take network unit as a system, the state of which can be regarded as 3 states. Note $x(t)$ as the link traffic state at $t$ time, then, there are 3 possible values for $x(t)=1,2,3$, represent the 3 states of network unit respectively.

The network unit state can be recognized by the canonical discrimination method. At first, the data of 1/3-10/3 2010 are clustered by the method of C-mean fuzzy clustering analysis, and we get the typical samples of every state. Then the states of the unknown samples (1/3 2010) are discriminated using canonical discrimination method, until all states are determined. Thereby, the state classification of network unit and state transition process are acquired. Figure 1 shows the link state transition process.

B. Instantaneous Travel Time

Instantaneous travel time can be obtained by calculating link length and instantaneous flow velocity ($T=L/v$. $T$ is instantaneous travel time, $L$ refers to link length, $v$ represent instantaneous flow velocity). Observing the relationship of instantaneous velocity and traffic flow velocity, there are obvious differences in the function forms of instantaneous travel time changing with traffic flow velocity in various states. In summary, in state 1, the randomness of instantaneous travel time show little variations, and congestion effect is not significant. In state 2 and 3, instantaneous travel time show notable variations with flow changing, and congestion effect is significant. Figure 2 shows the variation curves of the instantaneous travel time with traffic flow velocity in each state. In figure 2, the horizontal axes represent the traffic flow velocity, while the longitudinal axis represents the instantaneous travel time. The instantaneous travel time can be computed by $T=L/v$, where $T$ is instantaneous travel time, $L$ refers to link length, $v$ represent instantaneous flow velocity. The traffic flow velocity data is got by the traffic flow acquisition devices.
Where, in state 1(Figure 2a), instantaneous travel time shows little variation with traffic flow velocity and stronger randomness. State 2 (Figure 2b) presents a similar linear variation with traffic flow velocity, illustrating significant congestion effect. In state 3 (Figure 2c), with traffic flow velocity changes, instantaneous travel time exhibits sharp variations, demonstrating a similar exponential variation. Based on the characteristics above, the instantaneous travel time in state 1 is calibrated by a constant value, while that in state 2 by a linear function and that in state 3 by an exponential function.

C. Regional Transportation Network State

Transportation network is constituted by links, if one link shows \( m \) states, the regional traffic network containing \( n \) links will present \( m^n \) states. Record \( X(t) \) as the state of network at time \( t \), then \( X(t) \) denotes the vector \( x_i(t) \) that comprises \( m^n \) unit states. Where, \( x_i(t) \) represents the state of the \( i \)th network unit. That is

\[
X(t) = \begin{bmatrix}
x_1(t) \\
x_2(t) \\
\vdots \\
x_n(t)
\end{bmatrix}
\]

Where, \( x_i(t) = i, \ i=1,2 \ or \ 3 \). For example, the state of a road network containing two link units can be expressed by the state variables in Table 1.

**TABLE 1 THE DEFINITION OF ROAD NETWORK STATE**

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

D. State Limiting Probability

According to the definition of the state of regional traffic network, the transition process curves of the state can be acquired using traffic data. Figure 3 illustrates the state transition process curves of regional traffic network comprising two links (9 states in total).

Network state transition probability \( P_{ij} \) can be estimated by the transition curve. The process can be explained in the following: ignore the state duration time shown in Figure 4, and then the figure of state renewal process is regarded as a discrete stochastic process, the one-step transition time \( n \) from state \( i \ (i=1,2,3...9) \) to state \( j \ (j= i=1,2,3...9) \) are counted. Then, the statistics of one-step transition probability time \( N \) from state \( i \) to all states are statistically counted, thus \( P_{ij} = n/N \).

Basing on the state duration time shown in Figure 3, the duration time mean \( \mu_i \ (i=1,2,3...9) \) of each state is computed, and the limiting probability \( P_i \ (i=1,2,3...9) \) (namely long-travel time proportion) of each state is calculated by formula (3).

E. Average travel Time Estimation Model

Let \( T_{ikj}(f) \) represent the travel time function of link \( k \) in \( j \) states. Since the time proportion of each state in a certain time has been calculated in section 3.4, the expected travel time in this period can be obtained by multiplying the proportion of the occurrence of each state and the travel time function in these states. The formula can be express as:
\[ \bar{T} = \sum_{j=1}^{n} \sum_{k=1}^{m} P_{j,k}(f) \]  \hspace{1cm} (4)

Where, \( n \) refers to the unit link number comprised by network, \( m \) is the state number of each unit link, \( \bar{T} \) denotes average travel time estimated, \( P_r \) represents the limiting probability of regional network in state \( j \).

IV. SOME RESULTS AND MODEL VALIDATION

In this section, a calculation example is introduced to explicitly indicate the estimation process of average travel time by the model proposed in this paper. Meanwhile, the reliability of the model is verified through example.

A. Data Collection

In this paper we use a real road network of changzhou china for an example for developing and verifying the model. The road network is a part of city center road network of changzhou. For convenience in calculation, the regional road network including 4 link units is selected, which are link1 (north huaiide road), link2 (xiheng road), link3(beida street) and link4(west yanling road). The topology and the data we collect of the network is shown in Figure4.

Since model development and model test require appropriate data, the data acquired is mainly divided into two types; one is used for model development, and another for model validation test. The data for model development is the characteristic data of traffic flow of a specific regional road network in a certain time period, while the model validation data is the travel time data of each link comprised in the network in corresponding time period. The characteristic data of traffic flow is obtained through the detecting loop lay in the road, which is used to test the number of vehicles, average traffic flow velocity and average occupancy. For example, the max vehicle speed of link 1 is 68km/h, the min speed is 19km/h, the mean speed is 34km.h. And then transmit these data to the database of monitoring centre. Link travel time is collected through the vehicle plate recognition system installed in the entry and exit end of link, which is employed to record the plates of the vehicles passing by the link in certain time period. Then in the terminal of the monitoring centre, travel time samples are acquired through the comparison of the vehicle plates entering and exiting the link in this time period. The traffic data of a regional network in the centre of Changzhou city in China is applied in this study. For convenience in calculation, the regional road network including 4 link units is selected. Though the road network structure selected is not complicated, it can fully explain the issue. We rescored the traffic flow data of the 4 links and the plate recognition data. The data collection is shown in Figure 4.

In Figure 4, traffic flow data is obtained by means of the detection loop installed in the link centre. The plates of vehicles are gained by the license plate recognition detector set in the upper and lower of the link when the vehicles go through the links, travel time is acquired by recognize the plates entering and existing these links. In collecting plates, it should be noted that the plates entering into each network unit are collected at the same time at every time period, then record the time they exit. Basing on the way above, with 1 second as a time period, the plates of 60 time periods in 7:00-10:00 are acquired. The general condition of the data collected is shown in Table 2.

![Figure 4](image_url)

**Fig. 4 The data collection of the traffic flow and travel time of the links**

The specific data includes that: (1) the traffic flow data in 5:00-20:00 of 10 days (from 3/1 to 3/10) in 2010 is introduced in the clustering of traffic flow state; (2) the traffic data in 7:00-10:00, 13:00-14:00, 17:00-19:00 in 3/11 in 2010 is used to analyze the changing law of the traffic state in these period of this day and model development; (3) the plate recognition data (fielded travel time data) in 7:00-10:00, 13:00-14:00, 17:00-19:00 in 3/11 in 2010 is utilized for model validation, namely, the credibility of the travel time estimated by the model is validated and analyzed.

B. The Clustering and Discrimination of State

The traffic parameters of the 4 links in the 10 days from 3/1 to 3/10 in 2010 is conducted by C-mean fuzzy clustering using fuzzy C mean clustering method. The clustering number is set as 3, weight \( m=1.5 \), \( \varepsilon=0.1 \), and the initial matrix \( U \) is considered as the stochastic matrix uniformly distributed in interval \([0, 1]\) randomly generated. After 19, 19, 20, 21 iterations, the three state distribution results of the traffic flow of the 4 links treated by C-mean fuzzy clustering are obtained, as shown in figure 5.
According to the clustering results of each link above, samples of each state are extracted. Then the traffic flow states in the time period that will be investigated are distinguished using typical discrimination method. And the process is indicated as follows:

Note the typical sample number of each state as \( n \), thus \( n_1=729 \), \( n_2=189 \), \( n_1+n_2+n_3=1169 \)

\[
\begin{align*}
\bar{x}_1 &= 39.97, & \bar{x}_2 &= 47.12, & \bar{x}_3 &= 49.93 \\
\bar{x}_1 &= 13.19, & \bar{x}_2 &= 80.92, & \bar{x}_3 &= 214.72 \\
\bar{x}_1 &= 23.12, & \bar{x}_2 &= 42.08, & \bar{x}_3 &= 65.71 \\
\end{align*}
\]

The eigenvalues and eigenvectors of \( E^{-1}B \) are computed, and according to the accumulate contribution rate, the discriminate number is determined.

\[
E^{-1}B = \begin{bmatrix}
0.8183 & 0.0341 & 0.2225 \\
1.6432 & 0.0493 & 0.4951 \\
2.0504 & 0.0248 & 0.4398 \\
\end{bmatrix}
\]

The eigenvalues are \( \lambda_1=2.5 \), \( \lambda_2=0.06 \) and the corresponding eigenvectors are

\[
t_1 = \begin{bmatrix}
-0.6643 \\
-0.2807 \\
0.6928 \\
\end{bmatrix}, \quad t_2 = \begin{bmatrix}
-0.1523 \\
0.7469 \\
0.6472 \\
\end{bmatrix}
\]

When \( r=1 \), the cumulative contribution rate \( \lambda_1/\lambda_1+\lambda_2 \) is up to 97.6\%, which is a rather high level. Hence, 1 discriminants are used to perform highly accurate discrimination to the state sample. The centralized typical discriminants are:

\[
y_i = t_1(x_i - \bar{x}) = -0.6643(x_i - 27.16) - 0.2807(x_i - 41.67) + 0.6928(x_i - 36.59)
\]

The group means of the discriminants are:

\[
\bar{y}_{11} = 27.16, \quad \bar{y}_{31} = 1.26, \quad \bar{y}_{31} = -77.8
\]

Arbitrary traffic parameter sample \( x \) can be discriminated by the following formula:

\[
x \in \pi_i , \quad \text{if} \quad (y_i - \bar{y}_i)^2 = \min_{1 \leq j \leq 3} (y_i - \bar{y}_j)^2 \quad (5)
\]

Using the discriminant above, the state transition curves of the collected data of each links can be obtained, as shown in Figure 6.

\[
\tau = \frac{1}{n} \sum_{i=1}^{n} \eta_{ni} = \begin{bmatrix}
67.41 \\
41.67 \\
36.59 \\
\end{bmatrix}
\]

\[
B = \sum_{i=1}^{n} (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x}) = 10^6 \times \begin{bmatrix}
7.6244 & 0.0963 & 1.6432 \\
0.0963 & 0.0078 & 0.0334 \\
1.6432 & 0.0334 & 0.3787 \\
\end{bmatrix}
\]

\[
E = \sum_{i=1}^{n} \sum_{j=1}^{n} (x_{ij} - \bar{x})(x_{ij} - \bar{x}) = 10^6 \times \begin{bmatrix}
3.4706 & -0.1268 & 0.2928 \\
-0.1268 & 0.1684 & 0.0078 \\
0.2928 & 0.1045 & 0.4508 \\
\end{bmatrix}
\]
We use the Least-squares regression method to analyze the relationship between instantaneous travel time and the traffic flow velocity. From the regression results of Table 3, it can be found that the regression result \( t \) value is far greater than 1.96 in all states, indicating an obvious regression effect. Moreover, the values of \( R^2 \) are larger, which is greater than 0.8, explaining that demand shows notable influences to the travel time in the two traffic states. And the reasonability of calibrating travel time by traffic demand in these two traffic states is further confirmed.

**D. The State Limiting Probability**

The state changes of regional network divided in section 4.2 are statistically analyzed. The transition probabilities of each network state are obtained using the method in section 3.4, and the results are listed in Table 4.

**E. The Average Traffic Flow in Each State**

Since link travel time is a function concerning traffic demand, it is necessary to conduct statistics to the average traffic demand of each link state in corresponding time period. The results are presented in Table 5.
F. The Estimation and Test of Average Travel time

After model calibration, this model can be used for travel time estimation and model test. According to the limiting probabilities of each network state obtained in Table (6), the travel time function calibrated in Table (3), as well as the average traffic demand corresponded to the time period, and the state of each network unit (Table 7), the average travel time of each link can be acquired by formula (4). The results is shown in Table 6.

On the basis of the average travel time estimated, the probability distributions (sample size is 300) of the travel time data collected from the plate recognition equipment in corresponding measured time period are compared to test the estimated average travel time. The test is assumed as:

\[ H_0: \bar{T}_{estimated} = \bar{T}_{measured} \quad H_1: \bar{T}_{estimated} \neq \bar{T}_{measured} \]

The testing results are shown in Table 6.

<table>
<thead>
<tr>
<th>Time</th>
<th>field samples mean</th>
<th>Sd</th>
<th>estimated value</th>
<th>Confidence interval</th>
<th>Accept H0?</th>
</tr>
</thead>
<tbody>
<tr>
<td>7:00-10:00</td>
<td>132.8s</td>
<td>34.56</td>
<td>129.3</td>
<td>[125.66,135.31]</td>
<td>yes</td>
</tr>
<tr>
<td>13:00-15:00</td>
<td>51.6s</td>
<td>8.31</td>
<td>53.11</td>
<td>[49.94,52.22]</td>
<td>yes</td>
</tr>
<tr>
<td>17:00-19:00</td>
<td>249.3s</td>
<td>77.21</td>
<td>257.8</td>
<td>[242.00,264.35]</td>
<td>yes</td>
</tr>
</tbody>
</table>

where, in a confidence level of \( \alpha = 0.05 \), the model results are all located in the confidence intervals, representing a higher reliability possessed by the model in the view of the statistic law.

V. CONCLUSIONS

In this study, a model for estimating the average travel time of regional network is proposed using semi-Markov stochastic process theory and C-mean fuzzy clustering method. The process is indicated as follows: firstly, basing on three parameters, namely flow, velocity and occupancy, the historical traffic flow states of network unit are clustered; Using typical discrimination method, the traffic flow state of network unit for investigating in this work are distinguished and classified, based on which, the instantaneous travel time model of network unit is calibrated; Then the regional traffic states of network unit are calibrated according to the state variation of network unit; by employing semi-Markov stochastic process theory, the characteristic of the probability occurred in each regional network state is analyzed, and limiting probabilities of each state are calculated. Combining the limiting probabilities with instantaneous travel time function of network unit, the average travel time model is calibrated. Through the acquisition of the plates of the vehicles passing by certain road network in the centre of Changzhou city, the travel time data in the corresponding period are collected. Then the model results are compares with the field travel time, and the results calculated by the model are analyzed. The result indicates that, in a significant level of \( \alpha = 0.05 \), there is no significant differences between the average travel time estimated and the travel time mean measured. Thus, the higher of reliability of the model proposed is verified. The model constructed in this study can provide traffic planning designers and traffic engineers with a new method for evaluating travel time of road network.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (Grant No. 50978163)

REFERENCES


ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China (Grant No. 50978163)

Pengcheng Yuan. Ph. D candidate in Management Science and Engineering. Since 2010 he has been working in Institute of Transportation Studies in Shanghai Jiao Tong University. His research activity has been focused on Transportation System Planning & Management, Transportation System Simulation and Logistical System Analysis.

Zhicai Juan. Ph. D in Engineering; Professor of management & Engineering Department of Antai Management School in Shanghai Jiao Tong University; PhD tutor (Supervisor); Director of Institute of Transportation Studies of Shanghai Jiao Tong University, He taught at Jilin University from 1984 to 2004 and was the dean of the Transportation College in Jilin University of Technology from 1995~2000. He has authored over 60 publications in these areas including Sixteen papers which can be searched by SCI/EI/ISTP.