A Study for Highway Travel Time Prediction Using Markov Chain

Jaehwan Yang  
Master Candidate,  
Dept. of civil and environment engineering  
Seoul National University  
Daehak-dong, Gwanak-gu, Seoul  
Republic of Korea  
Fax: +82-02-873-2684  
E-mail: iyananda@naver.com

Moonjin Song  
Master Candidate,  
Dept. of civil and environment engineering  
Seoul National University  
Daehak-dong, Gwanak-gu, Seoul  
Republic of Korea  
Fax: +82-02-873-2684  
E-mail: mycolordna@hanmail.net

Seungyoung Kho  
Professor  
Dept. of civil and environment engineering  
Seoul National University  
Daehak-dong, Gwanak-gu, Seoul  
Republic of Korea  
Fax: +82-02-873-2684  
E-mail: sykho@snu.ac.kr

Seungpil Kang  
Professor  
Dept. of civil and environment engineering  
Seoul National University  
Daehak-dong, Gwanak-gu, Seoul  
Republic of Korea  
Fax: +82-02-873-2684  
E-mail: spkang@snu.ac.kr

**Abstract:** Providing travel time information is not only basic part of ATIS (Advanced Traveler Information System) but also very important factor for user's route choice. However, although there are various method for collecting travel time data, time-lag have to be came up when data are collected. And it bring about low accuracy of information, especially on peak hour where information providing is more important.

In this study, to estimate stochastic amount of time lag, using Markov Chain that show discrete probabilistic transition. Difference Ratio of TCS between actual travel time were used as transition probability of Chain Process. Through this process, gap between actual travel time and provided travel time became reduced. To verify this model, actual travel time data between Seoul IC and Suwon IC was applied into the model. Consequently, predicted data were verified to correspond with actual travel time by comparing with two-tailed t-test at 95% level of significant.

**Key Words:** Markov Chain, TCS, Travel Time Prediction, Time-lag

**1. INTRODUCTION**

In these days, providing information about travel time is essential requisite of ATIS (Advanced Traveler Information System) that helps travelers choose a best route. At the same time, it enables the administrator to improve efficiency of highway operation.

To provide the information about the travel time, its data are collected by various traffic data collecting device in highway and arterial road. The most representative method is to use Toll Collection System (TCS).

TCS gives information about travel time from departure toll to arrival toll. It is evaluated as the most reliable travel time information which does not require extra installation cost among many methods. But since the data are collected at the time of arrival, there appears time-lag, which means it shows the time that is later than the actual time as the travel time. For this reason, many people have made efforts to eliminate time-lag by developing varied models such as Kalman Filter Model, Pattern Matching and so on. This study was conducted to improve accuracy of travel time prediction. To achieve this, a new model combined with
Markov Chain is proposed and make afford to minimize time-lag. Markov Chain Model is a highly applicable model in the actual field because it has simple data calculation process.

2. PREVIOUS STUDIES & THEORETICAL STUDIES

2.1 Markov Chain
Suppose that things constantly have been changed by iterative process between the past and present, and the transition probability from the present state(time t-1) to future state(time t) only can be calculated based on the characteristic of the present state(time t-1). This is stochastic process called Markov Chain which is independent from all the other states except present state(time t-1). (Ahn, 2006)

In other words, Markov Chain Model is a useful tool of analysis when studying a process of change in particular system. It is named after A.A. Markov (1856-1922), a Russian mathematician who formulated new theory in which every result is only determined by effect of former event. (Kim, 2005)

And stationarity, a important hypothesis about Markov Chain, implies that the distribution of transition probability is uniform throughout any time(time t).

I and m are random integer (l,m≥0)

\[ p^{(l+m)} = p^{(0)} \cdot p^{(m)} \]  

(1)

This is Transition Probability defined as probability to switch state i (time t-1, chosen at random) on state j(time t).

\[ P_t(X_t = S_j|X_{t-1} = S_i) \]

\[ = P_t(X_t = S_j|X_{t-1} = S_i) \]  

(2)

\[ P_t(X_t = S_j|X_{t-1} = S_i) = P_{ij} \]  

(3)

Where, \( X_j \) : stochastic process to switch state i chosen at random(time t)
\( S_j \) : state j
\( X_{t-1} = S_i \) : state i at time t-1
\( X_t = S_j \) : state j at time t
\( P_{ij} \) : Transition Probability from state i to state j
i,j : 1,2,3,···,r

And combination of matrix and vector can show this process. The transition matrix is presented as following.

\[ \sum_{j=1}^{r} P_{ij}(t-1,t) = 1 \]  

(5)

\[ P_{ij}(t-1,t) \geq 0 \]  

(6)

\[ P_{ij}(t-1,t) = \begin{pmatrix} P_{i1}(t-1,t) & \cdots & P_{ir}(t-1,t) \\ \vdots & \ddots & \vdots \\ P_{1j}(t-1,t) & \cdots & P_{rr}(t-1,t) \end{pmatrix} \]  

(7)

---

2.2 TCS (Toll Collection System)
Every tolls collect travel time information of about 3,200 thousands cars, the average daily number of the cars driving the highway. These are TCS time data which include travel time from departure toll to arrival toll. As mentioned, travel time is the time required from departure toll to arrival toll and it is actual travel time reflecting traffic congestions of time-space during the driving time.
This information is evaluated as the most reliable travel time not requiring extra installation cost among many methods. Besides, this calculation of travel time can reflect traffic conditions experienced by driver in actual highway. But even if every car departs at the same time, some cars may have to stay at service area for a long time or stop on shoulder due to a car trouble. As a result, some may arrive later. And vice versa, some may arrive earlier. For example, they can switch the lane frequently or run in bus-only way. In these cases, TCS cannot said to reflect properties of many variables.
Besides, this information shows time-lag because data are collected at arrival time. In other words, data collection is conducted when vehicles starting earlier than actual travel time reach the arrival toll. Results of the comparisons between actual travel time and TCS time show that difference between the two data (TCS, actual travel time) is evident. As you see in <figure 1>, there is time-lag on May 8th, 2007 from Yangjae IC to Suwon IC.

![Figure 1 difference between TCS time and actual travel time](image)

2.3 Previous Studies
Followings are the papers about Markov Chain and various subjects in the paper have analyzed.
Lee (2009) studied on bus delay time prediction using Markov Chain, not using other variables such as traffic volume, on the geometric design elements of roadway. It is possible to use headway in order to calculate the delay time, and this way can overcome the limit of previous studies. Kurauchi et al. (2004) suggest model for capacity constrained transit assignment with common lines. To solve this problem, the concept of hyperpath is brought in assignment and Markov Chain is used for the calculation of choice probability.
Papers related to model for travel time prediction have calculated the travel time reducing errors using ARIMA model, Kalman Filter and Pattern Matching and so on.
Myeong (2010) performed the study based on the ideas that travel time data are collected from past data using detector and TCS. So they show certain differences with the actual travel time. To overcome its weakness, travel time is estimated by Pattern Matching, which finds past data that is similar to the present traffic situation and calculate the travel time under the present conditions. Lint et al. (2005) studied for method to get accurate freeway travel time prediction.
under missing data. To solve this problem, state-space neural networks with 3 steps, Hidden layer, Context layer, Output layer, were used. It was applied to predict actual travel time from Rotterdam to Hague, the change of the result input data missing ratio is observed. Followings are the papers that use Markov Chain for the travel time prediction. Kim(2007) divide congestion conditions into 10 stages where transition probability in each stage make transition matrix and travel time is predicted. For that, Markov-Fuzzy combined with Fuzzy inference and Markov Chain is used. In Yeon et al(2008), the travel time of freeway no.202 on Philadelphia was forecasted with Markov Chain that has congestion status switching probability as its transition matrix. Likewise, the previous studies for the travel time predicting have considered many variables or complicated processes that can provide high accurate estimation. However, in many cases these studies also have difficulties of complexity. Thus, this study will try to consider not any other variables, but just transition matrix direct calculated from the travel time. Moreover, using Markov Chain that can perform rapid estimation, it will be tried that form a model that can forecast the travel time of highway in practice.

3. MODEL FORMULATION

3.1 Base of The Model
The purpose of this paper is to predict the actual travel time at departure point by using the transition probability of difference of TCS data collected in toll and the actual travel time. To do this, we made a model using Markov Chain, which use the absolute value of the ratio of 'difference between present TCS time and the actual travel time' to 'present TCS time'(From now on, it would be called as the Difference Ratio.) as its state of Markov Chain. So the transition matrix should have transition probabilities between each state as its elements. Because difference between TCS time and the actual travel time has almost same probability to be negative or positive, we use absolute value to prevent offset.

TCS time for the latest arrived vehicle is real travel time which is possible to collect at the moment. When there isn't any considerable change in traffic pattern, TCS time could be used as current travel time. In detail, TCS time data have time-lag because it collected at the arrival point. But TCS time can be used as the actual travel time when congestion doesn't exist. So in this study, it is to be classified into two class, the peak time and non-peak time. And Markov Chain would be applied only to peak times where the TCS time cannot present the actual travel time.

The trend of TCS time is used to decide whether the revising value is negative or positive. Generally, when TCS time value has upward trend, the actual travel time is bigger than TCS, and when it is in the contrary condition, TCS is bigger. This phenomenon is used to decide the sign of revising values.

3.2 Definitions and Assumptions
For more effective prediction, some definitions were built to designate the time sections which don't have to be analyzed.

First, the 'Normal State' was defined as 'the time section which has same TCS value more than 30 minutes'.

Second, the 'Peak Hour' was defined as 'the time section which has bigger TCS value more than 30 minutes'.

average travel time value of the entire section'.
In addition, the prediction model used in this paper has basic premises of Markov Chain and
three additional assumptions as following.
First, the transition matrix is same at every phase. This assumption could differ from reality.
But this assumption of 'Stationarity' is the basic premise of Markov Chain.
Second, a condition is remains the same for 30 minutes, so a unit section of the time is 30
minutes.
Third, if a Markov Process is started after the previous process was terminated, revising ratio
should be applied from the first, regardless of the last process.

3.3 Model Formulation

3.3.1 Restriction of Target section
The sections analyzed in this model are restricted to the case of the non-normal state and peak
hour. Since the other cases do not need to be revised, TCS time data can be used intactly in
these cases. Moreover, if when a Markov Process has been started and it became non-peak
hour, the Markov Process would be postponed till non-peak hour ends in 30 minutes. Since
then, if non-peak hour continue over 30 minutes, the Markov Process would be finished. Or
stop postponement and continue the process unceasingly.

3.3.2 Formation of the Transition Matrix
To predict the actual travel time with Markov Chain, the transition matrix should be
estimated first, so the Difference Ratio should be divided into some classes.
In this paper, TCS time data and actual time data from Yangjae IC\(^3\) - Suwon IC\(^4\), from May
2nd 2007 to May 7th 2007, which have interval of a time section of 5 minutes has been used.
To make number of elements of each classes similar, classes are divided into three classes -
less than 5%, 5% and more and less than 10%, 10% and more.
The elements of the transition matrix are set of probability of present class shift to the other
class in 30 minutes'.

\[
R_{ij}(t, t + 30) = \begin{pmatrix}
P_{21} & P_{22} & P_{23} \\
P_{31} & P_{32} & P_{33} \\
P_{41} & P_{42} & P_{43}
\end{pmatrix}
\]

(8)

Where,  \( P_{ij} \) : probability that class i shift to class j
\( t \) : analyzed time section(minutes)

For instance, if there are 100 data in which phase is in status2 (5%~10%) and 40 of these
shift to status1 (less than 5%), \( P_{21} = 40/100 = 0.4 \) would be. Following is the transition matrix
calculated like this method as table.

---

\(^3\) This interchange is placed Seoul, Korea on Gyeongbu Highway.
\(^4\) This interchange is placed Gyeonggi-do, Suwon on Gyeongbu Highway.
### Table 1 The Transition Matrix for Model

<table>
<thead>
<tr>
<th>Revising Ratio</th>
<th>Less than 5%</th>
<th>5%-10%</th>
<th>10% and more</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less than 5%</td>
<td>0.53271</td>
<td>0.33645</td>
<td>0.13084</td>
</tr>
<tr>
<td>5%-10%</td>
<td>0.49038</td>
<td>0.34615</td>
<td>0.16346</td>
</tr>
<tr>
<td>10% and more</td>
<td>0.31313</td>
<td>0.27273</td>
<td>0.41414</td>
</tr>
</tbody>
</table>

3.3.3 Estimate Revising Time with the Transition Matrix

By the transition matrix and assumption of stationary, Difference Ratio after n times of unit interval can be gotten. With this Ratio, revising value that will be added to or subtracted from TCS time could be calculated. These values are satisfied with following equations.

\[
RT = TCS(1 \pm RR) \quad (9)
\]

\[
RR_n = P_0 \times P(30) \times \cdots \times P(30(n - 1)) \times RV \quad (10)
\]

\[
= P_0 \times P(30)^n \times RV \quad (11)
\]

Where, RT: revised travel time

\(R_0\): Initial revising ratio matrix

\(RR_n\): Revising ratio of n\textsuperscript{th} phase

\(RV\): Class representative value matrix

This method is used by Lee(2009) and transformed in this paper. The revising ratio of n\textsuperscript{th} phase can be represented by the multiple of the initial revising ratio matrix, the transition matrix, and the class representative value matrix. The average value of every starting point of Markov Chain of the entire sections can be made into the initial revising ratio matrix. The Average value of revising ratio at the starting point is about 22%, so the initial revising matrix can be made as \{0 0 1\}

### Table 2 Initial revising ratio matrix

<table>
<thead>
<tr>
<th>Average value of revising ratio at starting point</th>
<th>Initial revising ratio matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less than 5%</td>
<td>{1 0 0}</td>
</tr>
<tr>
<td>5%-10%</td>
<td>{0 1 0}</td>
</tr>
<tr>
<td>10% and more</td>
<td>{0 0 1}</td>
</tr>
</tbody>
</table>

The class representative value matrix is needed to estimate the revising ratio. But in the case of the class of 10% and more, the average value was used because this class has open section, which means it doesn't have a representative value. Through this process, the matrix is estimated as \{2.5% 7.5% 20%\}

4. APPLICATION & EVALUATION

4.1 Application

Based on the transition matrix, the initial revising ratio matrix, and the class representative value matrix, the TCS time value was revised, and the followings are the results of the Markov Chain Process.
4.2 Evaluation of the Model

In this paper, t-test is used for significance test between revised value and the actual travel time. A null hypothesis for t-test is set like following.

Null Hypothesis: $H_0: \mu_0 = \mu_1$

$\mu_0$: Mean value of the Actual Travel Time

$\mu_1$: Mean value of the Revised Value of The Model

Under this null hypothesis, the revised value is compared with the actual travel time by a two-tailed t-test at 5% level of significance. Just as came out in <Table 5>, t-value in every case is less than critical region(1.96). So null hypothesis cannot be rejected. Therefore, it cannot be proof that there is significant difference between revised value proposed in the model and the actual travel time.

<table>
<thead>
<tr>
<th>Date</th>
<th>t-value</th>
<th>significance probability</th>
<th>Result of Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>20070507</td>
<td>1.042</td>
<td>0.298</td>
<td>cannot be rejected</td>
</tr>
<tr>
<td>20070508</td>
<td>0.140</td>
<td>0.989</td>
<td>cannot be rejected</td>
</tr>
</tbody>
</table>
5. RESULT AND PROSPECT

In this study, the Markov Chain Process, which can estimate the actual travel time using TCS time as efficiently as possible in a simple process, was used for revising TCS time. This process cannot reflect the factors of real traffic condition that can influence on travel time in detail. However, this model can be considered superior to the others in estimating the travel time value in a highly simple process like Markov Chain.

Although, the model proposed in this paper has certain limits. First, if it is faced with a sudden normal state while Markov Chain Process is being applied, the model could release over-estimated or under-estimated value at high probability. Another limit is that the model use not only Markov Chain but also optional assumptions like the tendency of TCS. If traffic condition factors like traffic density or volume can be included in advance research to overcome their limits, the model proposed in this paper will predict travel time more accurately.

REFERENCES