A Naïve Clustering Approach in Travel Time Prediction

Rudra Pratap Deb Nath *rudra@kde.cs.tut.ac.jp*
Department of Computer Science & Engineering
Toyohashi University of Technology
Toyohashi, Aichi, 441-8580, Japan

Nihad Karim Chowdhury *umchowdn@cs.umanitoba.ca*
Department of Computer Science
University of Manitoba
Winnipeg, R3T 2N2, Canada

Masaki Aono *aono@kde.cs.tut.ac.jp*
Department of Computer Science & Engineering
Toyohashi University of Technology
Toyohashi, Aichi, 441-8580, Japan

Abstract

Travel time prediction plays an important role in the research domain of Advanced Traveler Information Systems (ATIS). Clustering approach can be acted as one of the powerful tools to discover hidden knowledge that can easily be applied on historical traffic data to predict accurate travel time. In our proposed Naïve Clustering Approach (NCA), we partition a set of historical traffic data into several groups (also known as clusters) based on travel time, frequency of travel time and velocity for a specific road segment, day group and time group. In each cluster, data objects are similar to one another and are sufficiently different from data objects of other groups. To choose centroid of a cluster, we introduce a new method namely, Cumulative Cloning Average (CCA). For experimental evaluation, comparison is also focused to the forecasting results of other four methods namely, Rule Based method, Naïve Bayesian Classification (NBC) method, Successive Moving Average (SMA) and Chain Average (CA) by using same set of historical travel time estimates. The results depict that the travel time for the study period can be predicted by the proposed strategy with the minimum Mean Absolute Relative Errors (MARE) and Mean Absolute Errors (MAE).

Keywords: Travel Time Prediction, Advanced Traveler Information Systems (ATIS), Naïve Clustering Approach (NCA), Cumulative Cloning Average (CCA) ,Successive Moving Average (SMA), Chain Average (CA), Naïve Bayesian Classification (NBC).

1.INTRODUCTION

In the research area of Intelligent Transportation Systems (ITS), travel time prediction is a very important issue and is becoming increasingly important with the advancement of ATIS [1]. Moreover, information, provided by travel time forecasting, helps traveler to decide whether they should change their routes, travel mode, starting time or even cancel their trip [2]. Therefore, the reliable and accurate travel time prediction on road topology plays an indispensible role in any kind of dynamic route guidance systems such as trip planning, vehicular navigation systems, etc. to fulfill the users' whim. Most importantly, the importance of travel time information is also significant to find the shortest path in terms of time. On top of that, accurate travel time estimation can improve the service quality of delivery industries by delivering products on time.

Predicted travel time information provides the capacity for road users to organize travel schedule pre-trip and en-trip. It helps to save transport operation cost and reduce environmental impacts. As congestion increases on urban freeways, more and more journeys are impacted by delays. Unless a traveler routinely traverses a given route, the extent of possible delays are unknown

before departing on a journey and the uncertainty must be addressed by allocating extra time for traveling. ATISs attempt to reduce the uncertainty by providing the current state of the system and sometimes a prediction of future state. In this context, travel time is an important parameter to report to travelers. Generally, prediction of travel time depends on vehicle speed, traffic flow and occupancy that are extremely sensitive to external event like weather condition and traffic incident [3]. Addressing the uncertainty on road network is also a crucial research issue. Additionally, prediction on uncertain situation is very complex, so it is important to reach optimal accuracy. Yet, the structure of the traffic flow of a specific road network fluctuates based on daily, weekly and occasional events. For example, the traffic structure of weekend may differ from that of weekday [17]. So, time-varying feature of traffic flow is one of the major issues to estimate accurate travel time [12].

In this research, we propose a new clustering way that is able to predict travel time accurately and reliably. Here, we attempt to combine the merits of our previous methods namely NBC [12], Rule based method, SMA and CA [13] by eliminating the shortcomings of those methods. Actually, this is the update version of our most recent research [16]. With the same set of historical traffic data, comparison is also made to evaluate our proposed method. Experimental results show the superiority of our proposed method over other prediction methods namely, NBC, Rule based, SMA and CA.

The remaining portions of this paper are organized as follows: Section 2 introduces some related researches in this field. An outline of our proposed NCA with example is demonstrated in section 3, Section 4 presents a concise experimental evaluation. Finally, the conclusion words and guidelines of future research are discussed in section 5.

2.LITERATURE REVIEW AND MOTIVATION

Nowadays, travel time prediction has emerged as an active and intense research area. So, a healthy amount of researchers have paid their concentration on the accurate travel time prediction. Several methodologies have been developed till date to compute and predict travel time with varying degree of success. A wide-ranging literature review on the topic of travel time prediction is presented in this section.

Park et al [5], [6] proposed Artificial Neural Network (ANN) models for forecasting freeway corridor travel time rather than link travel time. One model used a Kohonen Self Organizing Feature Map (SOFM) whereas other utilized a fuzzy c-means clustering technique for traffic pattern classification. Lint et al [7], [8] proposed a state-space neural network based approach to provide robust travel time predictions in the presence of gaps in traffic data. In [14], Kitaoka et al. developed a new computational method that they called the “Three-Range Composite Prediction Method” to realize optional dynamic route guidance and arrival travel time prediction with the TOYOTA G-BOOK telematic service. Kwon et al [9] proposed linear regression method to predict travel time.

A linear predictor consisting of a linear combination of the current times and the historical means of the travel times was proposed by Rice et al [10]. They proposed a method to predict the time that would be needed to traverse a given time in the future. Wu et al [3] applied support vector regression (SVR) for travel time predictions and compared its results to other baseline travel-time prediction methods using real highway traffic data. Most recent research in this field was proposed by Erick et al [11]. They investigated a switching model consisting of two linear predictors for travel time prediction. UI et al. [15] investigated an approach based on pattern matching which had relied on historic data patterns for estimating future travel times.

An efficient method for predicting travel time by using NBC was proposed by Lee et al [12] which had also been scalable to road networks with arbitrary travel routes. The main idea of NBC was that it would give probable velocity level for any road segment based on historical traffic data. It was shown from experiments that NBC could reduce MARE significantly rather than the other predictors. Another effective rule-based method was proposed by Chang et al [17] in where they had considered vehicle’s current road information, day time and week day information to extract

best suited decision rule. In [13], we formulated two completely new methods, namely SMA and CA that were based on moving average. In that research, we eliminated the drawbacks of conventional moving average approach such as unwanted fluctuation in data set. These methods were also scalable to large network with arbitrary travel routes. Moreover, both methods were less expensive in terms of computational time. Consequently, it was revealed that these proposed methods can reduce error significantly, compared with existing methods.

The prediction of travel time has been received an increasing attention in recent years that urges many researchers to motivate themselves in the research of travel time forecasting. Besides, travel time estimation and prediction form an integral part of any ATIS and ITS. In NBC and rule-based methods, a whole day and velocity of vehicle are divided into several groups in an effective and efficient manner. Moreover, in rule-based method, authors also concentrate on week days. But, the calculation of velocity level for a particular route enhances the complexity. Furthermore, it emphasizes only on those data that have high probability i.e. it doesn't take all data in consideration. In rule-based method, road information, day time and week day information are taken into account to carry out rule generation process. Generated rules are used to predict velocity class. As they generate some fixed rules, so it is unable to address uncertain situation. On the other hand, SMA and CA compute all data and are not based on probability theory. Although, SMA and CA provide an almost accurate travel time, those are unable to find uncertain data from the available traffic data. Clustering is one of the powerful leading data mining tools for discovering hidden knowledge that can be applied in the large historical traffic data set. To address the uncertain situation, and predict travel time more accurately, we propose NCA. In this study, our attempt to eliminate the shortcoming of NBC, rule-based, SMA and CA as well as combining their facilities. The key challenges of this research are to increase prediction accuracy and to address uncertain situation. On top of that, proposed method can also be scalable to large network with arbitrary travel routes. To clarify our method, the complete scenario of our method is presented in the next section.

3. PROPOSED NAÏVE CLUSTERING APPROACH

Cluster analysis or clustering is an assignment of separating the set of observations into subset. A cluster is therefore a collection of objects which are similar between themselves and are dissimilar to the objects belonging to other clusters. From available clustering techniques, partitioning and hierarchical clustering ways are popular and effective. In our research, we emphasize on partitioning clustering. For its simplicity and speed, K-means clustering, one of the partitioning clustering techniques is a better candidate to run on large data set. The procedure of K-means follows a simple and easy way to classify a given data set through a certain number of clusters (assume K clusters) fixed a priori. The main concept is to define K centroids, one for each cluster. The main disadvantage of K-means clustering is that it doesn't yield the same result with each run, since the resulting clusters depend on the initial random assignment. In contrast, we formulate our approach in a cunning way so that it eliminates all shortcomings of traditional K-means algorithm. Our algorithm can automatically determine the number of clusters without the intervention of users i.e. no fixed K clusters. Apart from it, we incorporate a technique so that centroids of different clusters maintain a sufficient difference by placing them as much as possible far away from each other. Furthermore, the initial random assignment problem is also handled. In addition, to re-estimate the centroid from a cluster, we introduce a new method, namely Cumulative Cloning Average which is described in section 3.1.

At first, an origin with start time, day and destination is provided by user. A route may consist of several road segments from origin to destination. Initially, we apply our NCA on the data set of the first road segment to calculate the end time of first road segment which in turn becomes the start time of the next road segment. Finally, applying successive repetition approximate travel time from origin to destination can be measured.

3.1. Cumulative Cloning Average (CCA)

To re-estimate a suitable centroid from available data of a cluster, a new method has been proposed. For the better understandability of the reader, CCA method with an appropriate example is presented, here.

Let $t = (t_1, t_2, ...t_n)$ be the data set where n is the number of elements in that set. The value of $\tau[i, j]$ gives the desired result for $t_i, t_{i+1}, ...t_j$ where $1 \le i \le j \le n$. Finally, the value of $\tau[1, n]$ indicates the CCA of the data set, *t*. CCA can be mathematically defined by following formula:

$$\tau[i, j] = \begin{cases} t_i & \text{if } i = j \\ \dfrac{\sum_{k=1}^{i+1} \tau[k,\ (j-i)+k-1]}{(i+1)} & \text{if } i < j \end{cases} \quad (1)$$

3.1.1. CCA with Example

A set of data with five elements is given below i.e. n=5, here. So, let's see how CCA works.

Sample Data $(t_1, t_2, t_3, t_4, t_5)$: 5, 3, 5, 4, 2

Total Sample Data (n):5

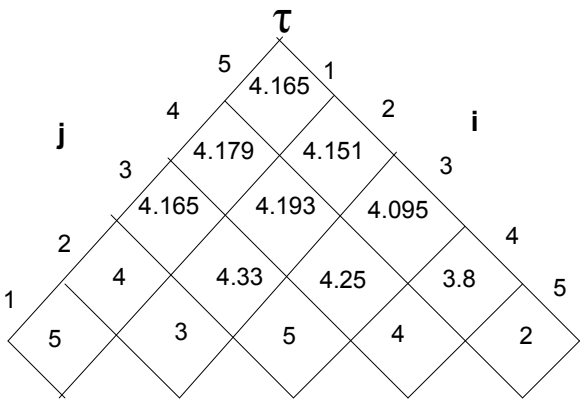

FIGURE 1: τ table for Cumulative Cloning Average

The τ table is used for storing the value of $\tau[i, j]$. Figure 1 illustrates CCA method on a sample data set where n =5. When i=j, then the value of $\tau[i, j] = t_i$. Using equation 1, we can calculate the value of $\tau[2,4]$ as $\frac{\tau[1,2] + \tau[2,3] + \tau[3,4]}{3} = 4.193$. Therefore, CCA of this data set travel would be 4 after applying round-off operation.

3.2. Definition of Time Group and Day group

The road environment of the same road network for running vehicles on the different time periods of a day is different. In NBC, the whole day time is separated into several groups according to the time. In our research, we also use their time grouping table which is illustrated in Table 1 [12].

Start_time_range	Time_group	Start_time_range	Time_group
06:01~10:00	1	16:01~18:00	6
10:01~11:00	2	18:01~22:00	7
11:01~12:00	3	22:01~00:00	8
12:01~14:00	4	00:01~06:00	9
14:01~16:00	5		

TABLE 1: Time group definition

Name of Day_ group	Symbol
Holiday	HD
Day_Before_holiday	BD
Remaining_day	RD

TABLE 2: Definition of Day group

Vehicle_ID	Road_ID	Time_group	Start_time	End_time	Travel_time (min)	Velocity (km/min)	Day_group
1	1	6	16:50	16:57	7	1.8725	RD
2	1	6	17:20	17:31	11	1.1916	RD
3	1	6	17:43	17:56	13	1.0082	RD
4	1	6	16:02	16:11	9	1.456	RD
5	1	6	16:16	16:32	16	0.8192	RD
6	1	6	16:05	16:18	13	1.0082	RD
7	1	6	17:03	17:10	7	1.8725	RD
8	1	6	17:11	17:18	7	1.8725	RD
9	1	6	17:35	17:46	11	1.1916	RD
10	1	6	16:09	16:16	7	1.8725	RD

TABLE 3: Sample historical traffic data

If a vehicle starts from any road segment between 16:01 and 18:00, its *Time_group* will be 6. The traffic flow of road network also depends on holiday, before holiday and remaining day. For our convenience, we group all national holiday and week holiday into holiday group. The previous day of holiday is also crucial in traffic structure. So, we put them in another category and the remaining days are kept in another group. For example, if it is Saturday, the group will be HD. Table 2 exhibits the day group definition. Table 3 illustrates the sample snapshot of historical traffic data for any road segment. Each record of the table contains seven attributes. The value of *Time_group* is calculated from the *Start_time. Travel_time* is the difference from *End_time* to *Start_time.* Dividing length of road segment by *Travel_time, Velocity* is measured.

To calculate approximate travel time for any road segment, we introduce NCA in the following section with appropriate example.

3.3. Procedure of Naïve Clustering Approach

When start time, day and the destination are given, our algorithm extracts the related data from the large data set according to the time_group, day_group, and road segment. Then the following step by step procedure is executed to predict the travel time of that road segment.

PROCEDURE

Step 1: Frequency for each travel time is measured by counting the repetition of that travel time in different records.
Step 2: Define Prediction relation that contains three attributes namely *Frequency*, *Travel_time* and *Velocity*. Each record of Prediction relation must contain distinct travel time.
Step 3: Find the greatest value from the *Frequency* attribute (f_{max}). A tuple $P\ (x_p\ ,\ y_p\ ,\ z_p)$ is chosen as centroid of a cluster, where x_p is the maximum frequency, y_p is the corresponding travel_time associated with x_p and z_p is the velocity associated with travel_time y_p. If two or more tuples contain the greatest value then make those tuples as the centroids, each for one cluster. Hence, we get a set of centroids, P where each centroid has maximum frequency.
Step 4: Compare each tuple T_i (x_i, y_i, z_i) of relation *Prediction* with the selected each centroid P_k (x_p, y_p, z_p) by using the following formula:

$$COST\ (P_k, T_i) = |x_p - x_i| + |y_p - y_i| + |z_p - z_i| \quad (2)$$

Where, subscript k, is the centroid number and can be ranged from 1 to n, depends on the duplication of frequency number. Choose tuple Q_k (x_q, y_q, z_q) as the centroid of another cluster, where *COST* (P_k, Q_k) is maximum. In this way, we also get another set of centroids, Q. Now, to select the final centroids, we perform intersection operation i.e. P ∩ Q. So, the number of tuples or elements in (P ∩ Q) set is the total cluster number.

Step 5: Build clusters where the centroid of each cluster is the distinct element of (P ∩ Q) set.

Step 6: Define the cluster memberships of tuples by assigning them to the nearest cluster representative tuple. The cost is given by Eq.2.

Step 7: Re-estimate the cluster centre by assuming the memberships found above are correct. To re-estimate we use our CCA method which has been illustrated in section 3.1.

Step 8: Step 6 and Step 7 are repeated until no change in clusters

Step 9: After complete preparation of clusters, desired predicted time is calculated separately for each cluster by using the following formula:

$$\tau_r = \frac{\sum_{i=1}^{N} f_i * t_i}{\sum_{i=1}^{N} f_i} \quad (3)$$

Where τ_r is the travel time obtained from r-th cluster, N is the total number of tuple in associated cluster, f_i is the *Frequency* of the i-th tuple, and t_i is the *Travel_time* of the i-th tuple.

Step 10: If the number of elements of (P ∩ Q) is R i.e. |P ∩ Q| =R, then the final predicted approximate travel time, T for the road segment of the specific time group and day group can be defined by following formula:

$$T = \frac{\sum_{i=1}^{R} \tau_i}{R} \quad (4)$$

3.4. Explanation of NCA method with example

Considering the sample historical traffic data of Table 3 that contains data for Road_id =1, Time_group=6 and day_group=RD. Steps of NCA procedure are explained below:

Step 1: There are 10 records in Table 2 where *Road_id* and *Time_group* and day_group are common. First step of NCA reveals to find the frequency of each distinct travel time. If we observe Table 3, then we find that the frequency of *Travel_time* 7 is four (4) because the number of repetition of *Travel_time* 7 in different records is four. Similarly, frequencies of *Travel_time* 16,9,13, and 11 are 1, 1, 2, and 2 respectively.

Step 2: *Prediction* relation is illustrated in Table 4. Each tuple in relation has three attributes namely *Frequency*, *Travel_time* and *Velocity*. The relation also reveals that it contains only those tuples that have distinct travel time.

Frequency	Travel_time(min)	Velocity (km/min)	Frequency	Travel_time(min)	Velocity (km/min)
1	16	0.8192	2	11	1.1916
1	9	1.456	4	7	1.8725
2	13	1.0082			

TABLE 4: *Prediction* relation of Table 2.

Step 3: The Frequency column of relation *Prediction* represents that the maximum value of it is 4. No more than one tuple contain the highest frequency. So, only one member in P set that is the tuple $P\ (x_p, y_p, z_p)$ = (4, 7, 1.8725).
Step 4: Table 5 calculates the cost of each tuple T_i (x_i, y_i, z_i) from the seed of *P Set* by using Eq.2

Frequency	Travel_time (min)	Velocity (km/min)	Distance from (4,7,1.8725)
1	16	0.8192	| 4-1 | + | 7-16 | + | 1.8725 - 0.8192 | = 3 + 9 + 1.0533 = **13.0533**
1	9	1.456	3 + 2 + 0.4165 = 5.4165
2	13	1.0082	2 + 6 + 0.8643 = 8.8643
2	11	1.1916	2 + 4 + 0.6809 = 6.6809
4	7	1.8725	0

TABLE 5: Comparison of each tuple with the centroid of *P set*

The maximum cost **(13.0553)** from centroid (4, 7, 1.8725) is marked as block in the Distance column of Table 5. So, the tuple $Q\ (x_q, y_q, z_q)$ = (1, 16, 0.8192) is selected as the centroid of Q *Set* As Set P has only one element, Set Q also contains only one element. Here, $|P \cap Q| = 2$.
Step 5: Two clusters are built where the centroid of *Cluster1* is the tuple $P\ (x_p, y_p, z_p)$ = (4, 7, 1.8725) and that of *Cluster2* is the tuple $Q\ (x_q, y_q, z_q)$ = (1, 16, 0.8192).
Step 6: Table 6 decides the cluster memberships of tuples by assigning them to the nearest cluster representative tuple. The numbers marked as block indicate the lowest cost comparison to other. Eq.2 is also used to find cost. 1st scenario of both clusters is shown in Table 7.

Freq-uency	Travel _time (min)	Velocity (km/min)	Distance from *Cluster1 centroid* (4,7,1.8725)	Distance from *Cluster2 centroid* (1,16,0.8192)
1	16	0.8192	3 + 9 + 1.0533 = 13.053	**0**
1	9	1.456	3 + 2 + 0.4165 = **5.4165**	0+7+0.6368=7.6368
2	13	1.0082	2 + 6 + 0.8643 = 8.8643	1+3+0.189=**4.189**
2	11	1.1916	2 + 4 + 0.6809 = 6.6809	1+5+0.3724=**6.3724**
4	7	1.8725	**0**	3+9+1.0533=13.0533

TABLE 6: Deciding cluster memberships.

	Frequency	Travel_time(min)	Velocity(km/min)
Cluster1	4	7	1.8725
	1	9	1.456
Cluster2	1	16	0.8192
	2	13	1.0082
	2	11	1.1916

TABLE 7: 1st scenario of both clusters with their members.

Step 7: Re-estimating of new centroid for each cluster. We calculate the new centroid for each cluster by using CCA (Eq. 1) method separately for frequency, travel_time and velocity.

New centroid for *Cluster1 using CCA*
$P_1\ (x_p, y_p, z_p)$ = (2.5, 8, 1.664).

New centroid for *Cluster2 using CCA*
$Q_1\ (x_q, y_q, z_q)$ = (1.55, 13.91, 0.96).

Step 8: Repetition of Step 6 with new centroids of both clusters. Blocking numbers indicate lowest cost comparing to other. Detail description illustrates in Table 8.

Frequency	Travel_time (min)	Velocity (km/min)	Distance from *Cluster1 new centroid* (2.5,8,1.664)	Distance from *Cluster2 new centroid* (1.55,13.91,0.96)
1	16	0.8192	1.5+8+0.84=10.34	0.55+2.09**+**0.14=**2.7**
1	9	1.456	1.5+1+0.208=**2.708**	0.55+4.91+0.49=5.95
2	13	1.0082	0.5+5+0.655=6.155	0.45+0.91+0.048=1**.408**
2	11	1.1916	0.5+3+0.4724=3.9724	0.45+2.91+0.23=**3.589**
4	7	1.8725	1.5+1+0.2085=**2.7085**	2.45+6.91+0.912=10.27

TABLE 8: Deciding cluster memberships with new centroids.

Re-estimating the cluster memberships from Table 8, 2nd scenario of both clusters has been represented in Table 9.

	Frequency	Travel_time(min)	Velocity(km/min)
Cluster1	4	7	1.8725
	1	9	1.456
Cluster2	1	16	0.8192
	2	13	1.0082
	2	11	1.1916

TABLE 9: 2nd scenario of both clusters with new centroids.

After repetition of step 7 we get that the most recent centroids of *Cluster1* P_{new1} (x_p, y_p, z_p) *and Cluster2* Q_{new2} (x_q, y_q, z_q) *are* (2.5, 8, 1.664) and (1.55, 13.9, 0.96) respectively. The most recent centroids of both clusters are similar to the 2nd most recent centroids. So, the need of repetition of step 6 and step 7 again and again are unnecessary. Table 8 shows the final clusters.

Step 9: By using Eq. 3, desired travel time from *Cluster1* and *Cluster2 can* be measured

Expected Travel Time from *Cluster1*

Here, N=2

So, τ_1 = (4*7+1*9)/ (4+1)
= (28 +9) / 5
= 37/5
= 7.4

Expected Travel Time from *Cluster2*

Here, N=3

So, τ_2 = (1*16+2*13+2*11) / (1+2+2)
= (16+26+22) / 5
= 64/5
= 12.8

So, expected travel time from *Cluster1,* τ1 =7 min (applying round operation) and expected travel time from *Cluster2,* τ2 =13 min (applying round operation)

Step 10: The final approximate travel time, T (for *Road_id=1*, Day_group=RD *Time_ group= 6)* is predicted by using Eq. 4 such as the simple arithmetic mean of τ_1 and τ_2. So, the final approximate travel time is T= ((7+13)/2) min =10 min.

4. PERFORMANCE ANALYSIS

4.1. Data Set Description

To measure the performance of different predictors, a real data set is used in our research. The data set generator is based on real traffic situation in Pusan city, South Korea. GPS sensor is used to collect real traffic delay for building this well–organized PNU generator. Traffic pattern of Pusan city was extracted from this data. According to this traffic pattern, generator simulates and generates trajectory data which almost same as real data. User interface of PNU (Punsan National University) is shown in the following figure 2.

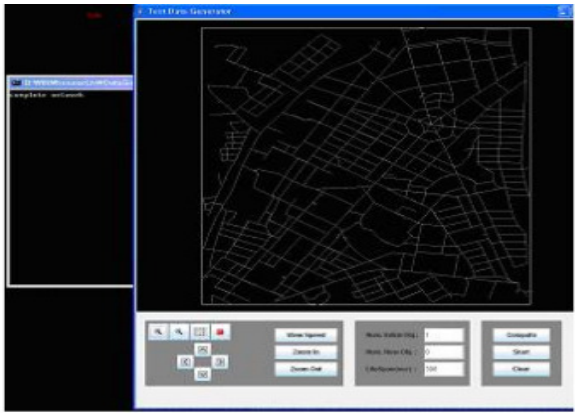

FIGURE 2: User interface of PNU trajectory data generator

By using this generator, 167,669 trajectories are generated. Every trajectory may compose of several road segments. The period of real traffic data covers both week days and weekends, and both peak hours and non-peak hours. This data organization format sufficiently reflects real traffic situations. For computing easily and efficiently and accurate evaluation of performance of the algorithms, data is divided into two categories, namely training data and test data sets. 365 days traffic data are used as training data set and 30 days traffic data are used as testing data set. Data from 365 training days are used for fitting the model. However, 30 days test data are used to measure prediction performance for all methods.

4.2. Comparison of Prediction Accuracy

The prediction error indices, Mean Absolute Relative Error (MARE) and Mean Absolute Error (MAE) are used to compare the accuracy among all prediction methods. MARE is the simplest & well-known method for measuring overall error in travel time prediction. MARE measures the magnitude of the relative error over the desired time range. The MARE is measured by the following formula:

$$MARE = \frac{1}{N}\sum_{t}\frac{|\, x(t) - x^{*}(t)\,|}{x(t)} \tag{5}$$

where $x(t)$ is the observation value; $x^{*}(t)$ is the predicted value and *N* is the number of samples.

On the other hand, the Mean Absolute Error (MAE) is a quantity used to measure how close forecasts or predictions are to the eventual outcomes. The MAE is a common measure of forecast error in time series analysis. This error measurement is defined as:

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|\, x(t) - x*(t)\,| = \frac{1}{n}\sum_{t=1}^{n}|\, e(t)\,| \tag{6}$$

As the name suggests, the mean absolute error is an average of the absolute errors
e(t) = x(t) – x*(t), where x(t) is the prediction and x*(t) is the true value. In equation (6), n is the number of samples.In experimental evaluation, proposed methods are tested against other predictors like NBC, Rule-based, SMA and CA. In this section, mean relative absolute error (MRAE) and mean absolute error (MAE) among all travel time predictors are investigated. Prediction errors of all predictors from 8 AM to 6 PM are examined. There are 11 test cases are evaluated between 8 AM to 6 PM. The line chart shown in figure 3 illustrates relative performance of all travel time predictors according to MARE. From the overall point of view, proposed method performs much better than NBC, SMA, CA and Rule based methods. In case of NCA method, it is shown that eight test cases exhibit errors less than 0.40.

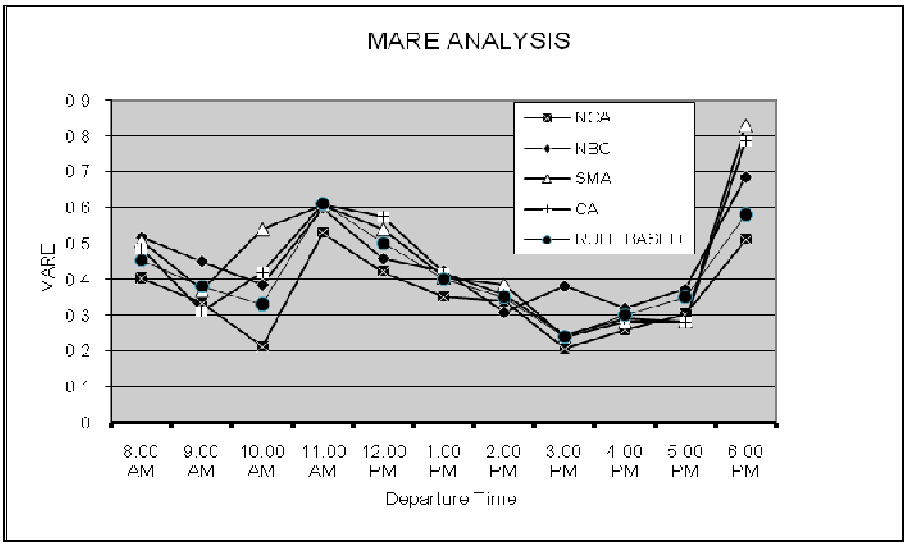

FIGURE 3: MARE of each method during different time interval.

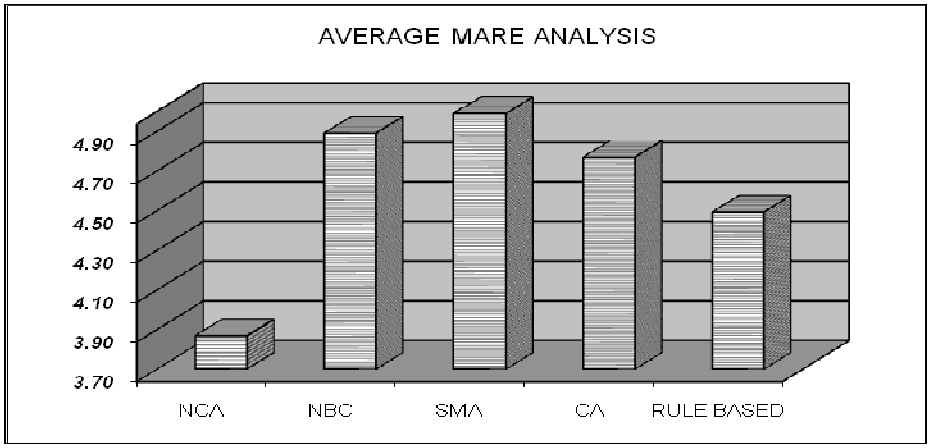

FIGURE 4: Summarized MARE of each prediction method.

Summarized MARE for different methods are shown in figure 4. Summarized MARE of NCA, NBC, SMA, CA and Rule based methods are 3.8692, 4.891, 4.9902, 4.769 and 4.493 respectively. Hence, our method reduces MARE from NBC, SMA, CA and Rule based methods by 20.89 %, 22.4%, 19%, and 14% respectively.

MAE of different methods during different time interval are shown in figure 5. In major cases, our method outperforms other methods in most of the cases. Figure 6 displays that the summarized MAE of NCA, NBC, SMA, CA and Rule based methods are 2.9601, 3.0727, 3.1648, 3.2173 and 3.24 respectively and our method reduces MAE from NBC, SMA, CA and Rule based method by 3.66%, 6.4%, 8% and 8.63% respectively.

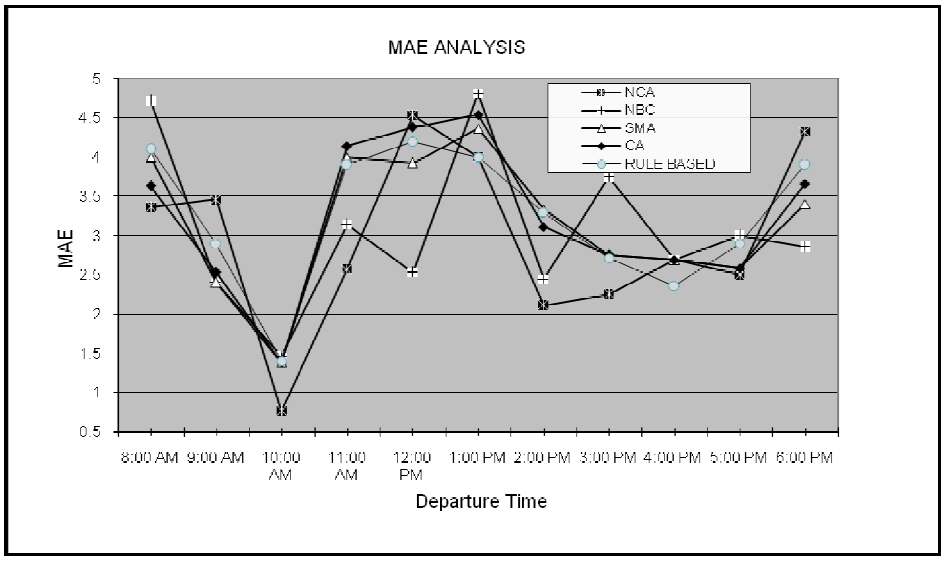

FIGURE 5: MAE of each method during different time interval.

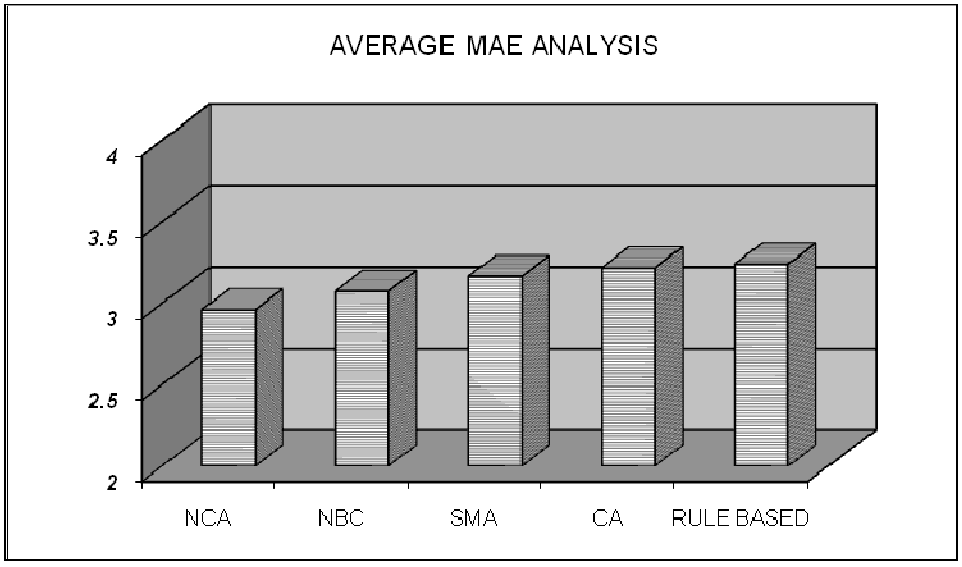

FIGURE 6: Summarized MAE of each method .

5. CONCLUSION

In this research, we focus an effective and efficient method to predict travel time more accurately. From the performance analysis portion, we can easily conclude that our method significantly reduces errors comparing with other methods. We also formulate our method in a cunning way so that we can eliminate so called partitioning problems. The centroids of the clusters are placed in a cunning manner so that they maintain as much as possible far way from each other. The superiority of our method is that the more the historical data set increases, the more the predictor is able to predict accurately. In our future plan, we will extend our NCA approach considering not only time and day but also seasonal event. The relationship between the length of roadways and accuracy of the prediction will also be tried to focus. Most importantly, analysis of our NCA will be extended with respect to real field data.

ACKNOWLEDGEMENTS
We would like to thank Prof. Jae-Woo Chang and Prof. Ki-Joune for providing us the PNU (Pusan National University) trajectory data generator.

6.REFERENCES

[1] M. Chen and S. Chien. "Dynamic freeway travel time prediction using probe vehicle data: Link-based vs. Path-based". J. of Transportation Research Record, TRB Paper No. 01-2887, Washington, D.C. 2001

[2] C. H. Wei and Y. Lee. "Development of Freeway Travel Time Forecasting Models by Integrating Different Sources of Traffic Data". IEEE Transactions on Vehicular Technology. Vol. 56, 2007

[3] W. Chun-Hsin, W. Chia-Chen, S. Da-Chun, C, Ming-Hua and H. Jan-Ming. "Travel Time Prediction with Support Vector Regression". IEEE Intelligent Transportation Systems Conference, 2003

[4] J. Kwon and K. Petty. "A travel time prediction algorithm scalable to freeway networks with many nodes with arbitrary travel routes". Transportation Research Board 84th Annual Meeting, Washington, D.C. 2005

[5] D. Park and L. Rilett. "Forecasting multiple-period freeway link travel times using modular neural networks". J. of Transportation Research Record, vol. 1617, pp.163-170. 1998

[6] D. Park and L. Rilett. "Spectral basis neural networks for real-time travel time forecasting". J. of Transport Engineering, vol. 125(6), pp.515-523, (1999)

[7] J. W. C. V. Lint, S. P. Hoogenoorn and H. J. V. Zuylen. "Towards a Robust Framework for Freeway Travel Time Prediction: Experiments with Simple Imputation and State-Space Neural Networks". Presented at 82 Annual Meeting of the Transportation Research Board, Washington ,D.C., 2003

[8] J. W. C. V. Lint, S. P. Hoogenoorn and H. J. V. Zuylen. "Freeway Travel Time Prediction with State-Space Neural Networks: Modeling State-Space Dynamics with Recurrent Neural Networks". In Transportation Research Record: Journal of the Transportation Research Board, No. 1811, TRB, National Research Council, Washington, D.C., pp. 30-39. 2002

[9] J. Kwon, B. Coifman and P. J. Bickel. "Day-to-day travel time trends and travel time predicti-on from loop detector data". J. of Transportation Research Record, No. 1717, TRB, National Research Council, Washington, D.C., pp. 120-129. 2000

[10] J. Rice and E. Van Zwet. "A simple and effective method for predicting travel times on freeways". In: IEEE Trans. Intelligent Transport Systems, vol. 5, no. 3, pp. 200-207, 2004

[11] J. Schmitt Erick and H. Jula. "On the Limitations of Linear Models in Predicting Travel Times". In: IEEE Intelligent Transportation Systems Conference, 2007

[12] H. Lee, N. K. Chowdhury and J. Chang. "A New Travel Time Prediction Method for Intellige-nt Transportation System". In: International Conference on Knowledge-Based and Intelligent Information & Engineering Systems, LNAI 5177, pp: 473-483, 2008

[13] N. K. Chowdhury, R. P. D. Nath, H. Lee and J. Chang. "Development of an Effective Travel Time Prediction Method using Modified Moving Average Approach". 13th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems. Part 1. LNAI 5711, pp: 130-138 2009

[14] H. Kitaoka, T. Shiga, H. Mori, E. Teramoto and T. Inoguchi. "Development of a Travel Time Prediction Method for the TOYOTA G-BOOK Telematics service". R & D Review of TOYOTA CRDL vol. 41 no.4 ,2006

[15] S. Ul, I. Bajwa and M. Kuwahara, "A Travel Time Prediction Method Based on Pattern Matching Technique". In proceedings of the 21st ARRB and 11th REAAA Conference. Transport. Vermont South, Victoria 3133, ZZ N/A Australia.2003.

[16] R. P. D. Nath, H. Lee, N. K. Chowdhury and J. Chang. "Modified K-means Clustering for Travel Time Prediction Based on Historical Traffic Data". 14th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems. Part 1. LNAI 6276, pp: 511-521, 2010.

[17] J. Chang, N. K. Chowdhury and H. Lee. "New travel time prediction algorithms for intelligent transportation systems". Journal of intelligent and fuzzy systems, vol.21, pp: 5-7, 2010.