International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 3 Issue 8, August - 2014

An Efficient Travel Time Prediction Method using Cosine Similarity

Subrina Akter[1], Jamshed Alam Patwary[2]
Department of CSE
International Islamic University,
Chittagong,Bangladesh

Shamima Akter[3], Lutfun Nahar[4]
Department of CSE
University of Chittagong
Chittagong, Bangladesh

***Abstract*— Travel Time Prediction has a great significance in the research area of Intelligent Transportation Systems (ITS). It helps travelers to take their decisions regarding travelling. Cosine Similarity (CS), a grouping technique, can easily be applied on historical traffic data to predict accurate travel time. It is a potential tool to reveal cryptic knowledge which is useful for predicting travel time. In our proposed Cosine Similarity (CS) oriented grouping approach, some historical traffic data are classified into two meaningful groups based on travel time and frequency of travel time. Velocity for an explicit road segment and time group is also considered to sort out the data. Based on the use of same set of historical traffic data, an effective comparison is made with the predicting results of different techniques such as Modified K-Means (MKC), Successive Moving Average (SMA), Chain Average (CA), Naïve Bayesian Classification (NBC) and Artificial Neural Network (ANN). The results imply that every method can predict the travel time but Cosine Similarity (CS) takes much less calculation time and memory to do so. Therefore, the method is more efficient than others.**

Keywords*— Intelligent Transportation System (ITS); Cosine Similarity (CS); Modified K-Means Clustering (MKC); Successive Moving Average (SMA); Chain Average (CA); and Naïve Bayesian Classification (NBC); Travel time prediction;

I. INTRODUCTION

The meaning of travelling is to go from one place to another, especially over a long distance.When people want to travel to a distant place, they want to know the condition of the road, using which they find their destination. If they have sufficient information about road condition then they can anticipate the time to reach their target. Therefore, the forecasting of travel time has a foremost concern in the arena of Intelligent Transportation System (ITS). Intelligent Transportation Systems (ITS) are sophisticated applications, aimed to make inventive services available involving various modes of transport and traffic organization and allow different users to be informed about safer, synchronized and smarter use of transport networks. However, ITS is not limited to road transportation. It comprises information technology which deals with the long-distance broadcast of computerized information and all kinds of communications in automobiles: such as between two vehicles, for example, communication between two cars; and between a vehicle and a fixed location, for instance, communication between a car and an infrastructure. Lately, precise assumption or estimation of travel time has been assumed as pivotal for traffic data analysis to many Advanced Travelers Information System (AITS) and ITS applications. In addition, Travel time forecasting is becoming very important with the advancement of ATIS and ITS applications [1]. Moreover, accurate travel time prediction helps travelers to take any sort of decision regarding traveling. And hence, the dependable and accurate travel time prediction on road network plays a significant role in any kind of dynamic route guidance system to meet the user's target [2]. Over and above all, the importance of travel time prediction is imperative to trace out the fastest possible path, that is, the shortest path according to travel time. The criteria, upon which the travel time prediction is based, are vehicle speed (i.e. distance per unit time), traffic flow, and weather condition along with traffic occurrence [3]. Moreover, uncertainty in road network is also a vital issue that is to be addressed in this research area. It is very difficult to predict under uncertainty. Time-dependent characteristics of traffic flow are also very important, because travel time in peak-hour and off-peak hour is not the same [4]. Consequently, it is very important to conduct research in this problem area to deliver reliable travel time information [5, 6, 7, 8, 9].

For travel time prediction, several methods have been applied. These include time series data analysis along with data mining techniques. The aim of data mining is to find unique, remarkable and useful knowledge from large dataset. Common data mining techniques help to determine recurrent occurrences [10] (e.g., common paths chosen by travelers) and to find out inconsistencies [11] (e.g., abnormally busy travel time). Besides these, other data mining techniques can be applied in predicting travel time. For example, classification techniques [4] can be applied to trained historical traffic data and to predict reliable travel time for query data. In the same way, clustering techniques [12] can also be applied to cluster or bunch the same types of data into the same class so that it can predict travel time based on a class of similar data (rather than the whole dataset). Last few years, a number of travel times prediction procedures have been suggested [13, 14, 15]. For instance, NBC [4] classification process was proposed in KES 2008 travel time prediction. SMA and CA [16] were other two algorithms, applied in KES 2009. The conception of these algorithms was based on moving average, and the prediction results were more accurate. In KES 2010, a clustering algorithm termed as MKC [12] was exposed, which gives better results than SMA, CA [16] and NBC [4]. Moreover, a careful examination reveals that there are some problems in the above procedures (e.g., SMA and CA [16], MKC [12]). In this paper, we emphasize on a novel method that can predict travel time reliably and perfectly. Our proposed method is able to recover the limitations of previous methods. According to the experimental results, our method reveals satisfactory result in terms of calculation time and memory requirements. Furthermore, the result is faster and more accurate than other prediction methods like NBC, SMA, CA and MKC.

In this article firstly some relevant research studies in this area are explained. Then a summary of Cosine Similarity (CS) with examples is examined. Next, a brief experimental calculation is demonstrated. And finally, there is a conclusion.

II. BACKGROUND STUDY

Now-a-days travel time prediction has emerged as a dynamic and intense research area. Many researches have been done in travel time prediction to perfectly predict travel time. Until today, several approaches have been explored for effective calculation and prediction of travel times. These approaches exhibited different levels of performance. In this segment, related works about travel time prediction is described in brief. Park et al [17, 18] proposed a model named Artificial Neural Network (ANN). The model is to predict freeway corridor travel time not the link travel time. Kohonen Self Organizing Feature Map (SOFM) is used in a model for classification of traffic pattern. A clustering technique called Fuzzy c-means is also used in the same purpose. A state-space neural network based method is proposed by Lint et al [19, 20] to predict travel time exactly where gaps in traffic data is considered. Kwon et al [21] predict travel time based on linear regression method. Rice et al [22] proposed a method to predict the time required to pass through a given time in upcoming day. A classification method named Support vector regression (SVR) proposed by Wu et al [23] where comparison was made between its result and other related methods along with travel time prediction. They used real highway traffic data in the paper. A switching model consisting of two linear predictors for travel time prediction was proposed by Erick et al [24]. Lee et al [4] proposed an effective travel time prediction method using NBC which was also scalable to road networks with random travel routes. Its basic idea was to give probable velocity level for any road segment. It was done based on historical traffic data. As an extension to NBC, a rule-based Bayesian classification (RBC) algorithm [25] representing the knowledge (in the form of rules) in a more intuitive way was also proposed. By using the concept of moving average another two travel time prediction algorithms—namely, Successive Moving Average (SMA) and Chain Average (CA) [16]—were formulated. Nath et al. [12] developed the modified k-means addressed and travel time can be efficiently predicted by MKC, but a significant problem will arise in the existing system. For example the present system takes much time and memory to predict travel time. In this study, we try to stamp out the clustering (MKC) [12] method where travel time can be predicted by calculating the average of the mean of travel times of each cluster. Travel time prediction is one of the major issues in ATIS. Uncertain situations can be effectively limitations of previous systems with the help of our proposed Cosine Similarity (CS) technique.

III. PROPOSED METHOD

In this segment, a novel technique using Cosine Similarity (CS) is illustrated for forecasting travel time based on historical traffic data. Cosine Similarity (CS) is one of the popular methods for measuring the similarity between two vectors of n-dimensions [26, 27]. It is widely used in text mining and information retrieval system. The similarity measurement can be transformed to a distance measure which can be used in any distance based classifier. The key advantages of Cosine Similarity (CS) are straightforwardness, containing less memory space, reducing computation time and achieving higher predictive performance. The process is to categorize a given dataset into two groups (according to most similar and the most dissimilar). Suppose two vectors of attributes, $p = \{a_1, a_2, ..., a_n\}$ and $q = \{b_1, b_2, ..., b_n\}$ are given. Then the Cosine Similarity (CS), θ is the measure of the angle between the two vectors (p and q) and can be measured by the following equation [27, 28]:

$$similarity(p,q) = \cos(\theta) = \frac{\bar{p} \cdot \bar{q}}{|p| * |q|} \quad (1)$$

The main disadvantage of previous methods is computational complexity. For example, in MKC method there is a repeating procedure in defining the cluster memberships of tuples and also re-estimating the cluster center until no change in clusters [12]. In Cosine Similarity (CS) these two steps are not required and this method requires no repetition of steps.

TABLE I: DESCRIPTION OF TIME GROUP

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	22:01~12:00	8
12:01~14:00	4	00:01~06:00	9
14:01~16:00	5		

TABLEII: SAMPLE HISTORICAL TRAFFIC DATA

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

In addition the experimental results show that our method based on the Cosine Similarity (CS) is better than other methods regarding computational complexity.

In this paper we used the same data set of Nath et al [12] where the whole day time is divided into different groups based on time, which is shown in table 1. We also used the sample of historical traffic data that used in Nath et al paper, so that we can compare the efficiency of our proposed method with other methods. The sample of historical traffic data shown in table 2.

In table 2 there exists seven attributes. Four of them namely vehicle_id, road_id, start_time and end_time are given. On the other hand time_group is derived from table 1 considering start_time_range, travel_time is calculated from the difference between start_time and end_time and velocity is measured by dividing the length of a road segment by travel_time.

The method based on Cosine Similarity (CS) is illustrated in section 3.1 with an example.

3.1 Cosine Similarity Method

The method is illustrated with the help of following steps:

First Step: Frequency for each travel time has to be determined by counting the total number of occurrences of that travel time in different records.

Second Step: Taking the frequency, travel_time and velocity into account the prediction relation is defined. The value of the travel time of each vector of prediction relation must be distinct.

Third Step: The maximum value of the Frequency attribute ($f_{\max}$) has to be measured. If there is any case where two or more vectors contain the same highest value then the highest Travel_time for available highest frequencies must be determined. We consider vector as base or seed vector $P(X_p, Y_p, Z_p)$, where X_p is the maximum frequency, Y_p is the corresponding travel_time connected with X_p and Z_p is the velocity related with travel_ time Y_p

Fourth Step: Reduce the attributes of the vectors by discarding the frequency_attribute as we consider the record of same travel time as a single record.

Fifth Step: Compare each vector $Q_i(Y_i, Z_i)$ of Prediction relation with the base vector $P(Y_p, Z_p)$ by using the following formula:

$$similarity(p, q_i) = \cos(\theta) = \frac{\bar{p} \cdot \bar{q}_i}{|p| * |q_i|} \qquad (2)$$

Choose vector $Q_i(Y_q, Z_q)$ which is most dissimilar than the base vector.

Sixth Step: Define the group memberships of vectors by assigning them to the nearest group representative vector. The grouping is based on similarity measure which is calculated by equation 2.

Seventh Step: After the grouping of vectors, for each groups desired predicted time is calculated distinctly by using the following formula:

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

Where G_i is the predicted travel time found from i-th group, N is the total number of vectors in respective groups, f_i and t_i is the frequency and travel time of i-th vector respectively.

Eighth Step: Finally the approximate travel time T can be predicted by calculating the average of the desired travel time of groups G_1 and G_2.The formula is,

$$T = (G_1 + G_2)/2$$

Do not use abbreviations in the title or heads unless they are unavoidable.

3.2 Clarification of Cosine Similarity with examplenits

For Road_id =1 and Time group=6, the steps of the method of Cosine Similarity (CS) are explained below. Here we use the sample historical traffic data of Table 2.

First Step: Table 2 contain 10 records where Road_id and Time_group are common. At first the frequency of each distinct travel time derived from Table 2. By observing Table 2 it has been found that the frequency of Travel_time 7 is four

(4) because the number of occurrences 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.

Second Step: Prediction relation is demonstrated in Table 3 which contains only those vectors that have distinct travel time. Each vector in relation has three attributes namely Frequency, Travel_time and Velocity.

TABLEIII: PREDICTION RELATION DERIVED FROM TABLE2

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			

Third Step: From the Frequency column of relation Prediction it is showed that the highest frequency is 4 and

only one
vector contain the highest frequency. So, the base vector for group 1 is the vector $P(X_p, Y_p, Z_p)$ = (4, 7, 1.8725).

Fourth Step: Discarding the Frequency column the following table of prediction relation has been found. Now the bas vector

represented as $P(Y_p, Z_p)$ = (7, 1.8725).

Fifth Step: The similarity of each vector $T_i(Y_i, Z_i)$ from the base vector

TABLEIV: REDUCED PREDICTION RELATION DERIVED FROM TABLE 2

Travel_time (min)	Velocity(km\min)	Travel_time(min)	Velocity(km\min)
16	0.8192	11	1.1916
9	1.456	7	1.8725
13	1.0082		

$P(Y_p, Z_p)$ of Group 1 can be calculated by using Eq. 2 represented in Table 5.

TABLE V: COMPARISON OF EACH VECTOR FROM THE BASE VECTOR OF GROUP 1

Travel_time(min)	Velocity(km\min)	Similarity between(7,1.8725)
16	0.8192	**0.9779**
9	1.456	0.99491
13	1.0082	0.98313
11	1.1916	0.98825
7	1.8725	1

Observing Table 5 we find that the most dissimilar value between base vector and the test vectors is 0.9779. So the vector $Q(Y_q, Z_q)$ = (16, 0.8192) must be the seed vector of Group 2.

Sixth Step: The group memberships of vectors is shown in Table 6 and it has been done by assigning them to the nearest group representative vector. The numbers marked as block indicate the most similar comparison between the test vectors and the group representative vectors by using Eq. 2.Final grouping result also shown in Table 7.

TABLEVI: DETERMINING GROUP EMBERSHIPS

Travel_time(min)	Velocity (km\min)	Similarity between(7,1.8725)	Similarity between(16, 0.8192)
16	0.8192	0.9779	**1**
9	1.456	**0.9949**	0.9940
13	1.0082	0.9831	**0.9997**
11	1.1916	0.9882	**0.9984**
7	1.8725	**1**	0.9779

Seventh Step: Desired travel time from Group 1and Group 2 can be calculated by using Eq. 3

TABLEVII: GROUPING RESULT WITH RESPECTIVE MEMBERS

	Travel_time(min)	Velocity(km\min)
Group 1(G1)	7	1.8725
	9	1.456
Group 2(G2)	16	0.8192
	13	1.0082
	11	1.1916

<u>Expected Travel Time from Group 1</u>

Here, N=2

So, G1 = (4*7+1*9)/ (4+1)
= (28 +9) / 5 = 7.4

<u>Expected Travel Time from Group 2</u>

Here, N=3
So, G2 = (1*16+2*13+2*11) / (1+2+2)
= (16+26+22) / 5 = 12.8

By applying round operation we have found the expected travel time from Group 1, G1 =7 min and from Group2, G2 =13 min.

Eighth Step: Finally, For *Road_id=1* and *Time_group= 6* the final approximate travel time
T = ((7+13)/2) min =10 min.T is predicted by using Eq. 4.

IV. RESULT ANALYSIS

To evaluate the performance of different prediction systems, Pusan National University (PNU) generator is used in our proposed system. The real traffic situation in Pusan city of South Korea is simulated by this generator. To gather real traffic wait for building this well-organized PNU generator, Global Positioning System (GPS) is used. Traffic model of Pusan city was extracted from this PNU generated data. The PNU generated data is almost similar to real data. So, we can use the data in the system as real data. The performance of the proposed algorithm is evaluated effectively by dividing the data into training data set and test data set.

The overall error in travel time prediction is measured with the help of Mean Absolute Relative Error (MARE) which is very easy and eminent method. MARE is also used to compare among all the prediction methods. The MARE is calculated by the following equation:

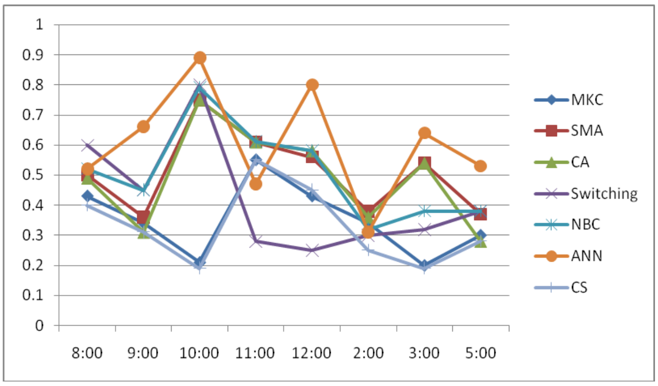
$$MARE = \frac{1}{N}\sum_{t=1}^{N}\frac{\left|x(t) - x^{'}(t)\right|}{x(t)} \qquad (4)$$

Here, the observation value is denoted by , predicted value is denoted by and N is the sample number. In this experiment, proposed method is compared with other methods like MKC, NBC, SMA and CA. Errors in the prediction of all predictors from 8 AM to 6 PM are experimented. The graph shown in Fig.1 demonstrates the comparison of different travel time predictors. Taking everything into account we can claim that the proposed method exhibits better performance than other methods.

V. CONCLUSION

Necessary steps have been taken to make our research technically efficient rendering scalable method for predicting travel time with arbitrary routes in road network. Certainly, our

FIGUREI: MARE OF ALL METHODS DURING DIFFERENT TRAVEL TIME



proposed method based on Cosine Similarity (CS) belongs to the simplest grouping techniques. Unfortunately, the computational complexity of previous methods for predicting travel time is higher than that of Cosine Similarity (CS). Another limitation of previous methods is that they pay no attention to the uncertainty related to data. Given the fact that, in case our output based on Cosine Similarity (CS) is excellent in a greater number of cases, but nevertheless, in uncertain situations the results are hopelessly worse. Therefore, by analyzing both groups we take the predicted travel time so that the algorithm will be able to address the unfavorable or uncertain situation. Yet our method is capable of forecasting in uncertain situation more accurately when compared with other methods. Fortunately, performance analysis portion of this research indicates that our proposed method definitely outperforms other methods in many cases. The shining part of Cosine Similarity (CS) is that the more the historical traffic data set increase the more the predictor becomes more efficient to predict precisely. Hopefully, we shall be able to paint a rosy picture in near future by augmenting our approach in view of not only day time but also week days and finally seasonal patterns. This is very much likely to help us to address uncertain situations more efficiently. Case will be exercised in the relationship between the length of roadways and accuracy of the prediction. We shall leave no stone unturned to improve our algorithm by addressing which data are intimately related with uncertain situation. Hopefully, a brighter future is awaiting us in this noble venture.

REFERENCES

[1] Chen, M., Chien, S.: 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] Wei, C.H., Lee, Y.: Development of Freeway Travel Time Forecasting Models by integrating Different Sources of Traffic Data. IEEE Transactions on Vehicular Technology 56 (2007)

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

[4] Lee, H., Chowdhury, N.K., Chang, J.: A New Travel Time Prediction Method for Intelligent Transportation System. In: Lovrek, I., Howlett, R.J., Jain, L.C. (eds.) KES 2008, Part I. LNCS (LNAI), vol. 5177, pp. 473–483. Springer, Heidelberg (2008)

[5] Kriegel, H.-P., Renz, M., Schubert, M., Z¨ufle, A.: Statistical density prediction in traffic networks. In: SDM 2008, pp. 692–703. SIAM, Philadelphia (2008)

[6] Lee, J.-G., Han, J., Whang, K.-Y.: Trajectory clustering: a partition-and-group framework. In: ACM SIGMOD 2007, pp. 593–604. ACM, New York (2007)

[7] Liu, W., Wang, Z., Feng, J.: Continuous clustering of moving objects in spatial networks. In: Lovrek, I., Howlett, R.J., Jain, L.C. (eds.) KES 2008, Part II. LNCS (LNAI), vol. 5178, pp. 543–550. Springer, Heidelberg (2008)

[8] Nakata, T., Takeuchi, J.: Mining traffic data from probe-car system for travel time prediction. In: ACM KDD 2004, pp. 817–822. ACM, New York (2004)

[9] Takamiya, M., Yamamoto, K., Watanabe, T.: Probabilistic estimation of travel behaviors using zone characteristics. In: Vel´asquez, J.D., R´ios, S.A., Howlett, R.J., Jain, L.C. (eds.) KES 2009, Part II. LNCS (LNAI), vol. 5712, pp. 615–622.Springer, Heidelberg (2009)

[10] Leung, C.K.-S., Brajczuk, D.A.: uCFS$_2$: an enhanced system that mines uncertain data for constrained frequent sets. In: IDEAS 2010, pp. 32–37. ACM, New York (2010)

[11] Leung, C.K.-S., Mateo, M.A.F., Nadler, A.J.: CAMEL: an intelligent computational model for agro-meteorological data. In: ICMLC 2007, vol. 4, pp. 1960–1965. IEEE, Piscataway (2007)

[12] Nath, R.P.D., Lee, H.-J., Chowdhury, N.K., Chang, J.-W.: Modified k-means clustering for travel time prediction based on historical traffic data. In: Setchi, R., Jordanov, I., Howlett, R.J., Jain, L.C. (eds.) KES

2010, Part I. LNCS (LNAI), vol. 6276, pp. 511–521. Springer, Heidelberg (2010)

[13] Id´e, T., Kato, S.: Travel-time prediction using Gaussian process regression: a trajectory-based approach. In: SDM 2009, pp. 1183–1194. SIAM, Philadelphia (2009)

[14] Lee, W.-H., Tseng, S.-S., Tsai, S.-H.: A knowledge based real-time travel time prediction system for urban network. Expert Systems with Applications 36(3), Part 1, 4239–4247 (2009)

[15] Simroth, A., Z¨ahle, H.: Travel time prediction using floating car data applied to logistics planning. IEEE Trans. Intell. Transp. Syst. 12(1), 243–253 (2011)

[16] Chowdhury, N.K., Nath, R.P.D., Lee, H., Chang, J.: Development of an effective travel time prediction method using modified moving average approach. In: Vel´asquez, J.D., R´ios, S.A., Howlett, R.J., Jain, L.C. (eds.) KES 2009, Part I. LNCS (LNAI), vol. 5711, pp. 130–138. Springer, Heidelberg (2009)

[17] Park, D., Rilett, L.: Forecasting multiple-period freeway link travel times using modular neural networks. J. of Transportation Research Record 1617, 163–170 (1998)

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

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

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

[21] Kwon, J., Coifman, B., Bickel, P.J.: Day-to-day travel time trends and travel time from from loop detector data. J. of Transportation Research Record, No. 1717, TRB, National Research Council, Washington, D.C., 120–129 (2000)

[22] Rice, J., Van Zwet, E.: A simple and effective method for predicting travel times on freeways. IEEE Trans. Intelligent Transport Systems 5(3), 200–207 (2004)

[23] Chun-Hsin, W., Chia-Chen, W., Da-Chun, S., Ming-Hua, C., Jan-Ming, H.: Travel Time Prediction with Support Vector Regression. In: IEEE Intelligent Transportation Systems Conference (2003)

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

[25] Chang, J., Chowdhury, N.K., Lee, H.: New travel time prediction algorithms for intelligent transportation systems. J. Intell. Fuzzy Syst. 21(1–2), 5–7 (2010)

[26] A. Karnik, et al., Detecting Obfuscated Viruses Using Cosine Similarity Analysis, in Modelling & Simulation, AMS '07. First Asia International Conference on, 2007, 165-170.

[27] N. Swe Swe, Mining contents in Web page using cosine similarity, in Computer Research and Development (ICCRD), 2011 3rd International Conference on, 2011, 472-475.

[28] Y. Soe-Tsyr and S. Jerry, Ontology-based structured cosine similarity in document summarization: with applications to mobile audio-based knowledge management, Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on, 35 (2005), 1028-1040.