Experienced Travel Time Prediction in Congested Freeway Routes

By

Mehmet Yildirimoglu
Urban Transport Systems Laboratory
School of Architecture, Civil and Environmental Engineering
Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland
Phone: +41(21)693-2484
E-mail: mehmet.yildirimoglu@epfl.ch

and

Nikolas Geroliminis*
Urban Transport Systems Laboratory
School of Architecture, Civil and Environmental Engineering
Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland
Phone: +41(21)693-2481
E-mail: nikolas.geroliminis@epfl.ch

For Presentation Only at the 92nd Annual Meeting
Transportation Research Board
Washington, D.C.
January 2013

Number of words: 5620 words + 6 figures + 2 Tables = 7620 words

*corresponding author
ABSTRACT

Travel time is considered as one of the most important performance measures for roadway systems, and dissemination of travel time information can help travelers to make reliable travel decisions such as route choice or time departure. Since the traffic data collected in real time reflects the past or the current conditions on the roadway, a predictive travel time methodology should be used to obtain the information to be disseminated. However, an important part of the literature either uses instantaneous travel time assumption, and sums the travel time of roadway segments at the starting time of the trip, or uses statistical forecasting algorithms to predict the future travel time. This study benefits from the available traffic flow fundamentals (e.g. shockwave analysis, bottleneck identification), and makes use of both historical and real time traffic information to provide travel time prediction. The experimental results based on the loop detector data on Californian freeways indicate that the proposed method provides promising travel time predictions under varying traffic conditions.
INTRODUCTION

Predictive travel time is valuable information required by drivers and transportation managers to make better travel and control decisions. The provision of travel time information through Advanced Traveler Information Systems (ATIS) enables drivers to make decisions such as route choice and departure time. In addition, besides the fundamental traffic parameters, travel time can be used by transportation agencies to deploy efficient control measures and to prevent potential traffic congestion.

Data required to estimate travel time can be obtained through loop detectors, test vehicles, license plate matching techniques (automatic vehicle identification, AVI) and ITS probe vehicle techniques (1). All of the detection technologies except the one based on loop detectors provide direct measurement of experienced travel time. As freeways are usually equipped with loop detectors that collect flow, speed and occupancy information, travel time estimation in freeways should rely on them. Travel time measurement can be either based on local velocity measurements, or more sophisticated models that attempt to correlate vehicle observations at multiple locations (2). However, the essential problem with travel time information is that it always has to refer to future conditions in the roadway. On the contrary, traffic data collected in real time reflect past or current conditions in the roadway. Therefore, the provision of travel time information always requires prediction of future conditions on the roadway. The approach of instantaneous travel times might create considerable errors when traffic conditions are varying in time and space.

A speed contour plot is presented for a section in freeway I-5S in California in FIGURE 1. A few active bottlenecks can be seen in the site that start at different times and propagate upstream. FIGURE 1 clearly shows that the difference between instantaneous and experienced travel time by plotting a few vehicle trajectories with instantaneous and experienced travel time. Note that these differences can be quite significant especially during the congestion onset and dissipation. This indicates that estimation of travel time should not be solely based on the traffic data collected in real time, but also the future recurrent traffic conditions can be integrated from historical data.

FIGURE 1 - Speed Contour Plot and Trajectories
The need for short-term travel time prediction led to the development of various forecasting algorithms. These methods can be broadly classified into two major categories; parametric methods (e.g., linear regression (3), time series models (4), Kalman filtering (5)) and non-parametric methods (neural network models (6), support vector regression (7), Bayesian models (8), simulation models (9) etc.). Data-driven approaches are consistent under some cases with the transitional physics of traffic flow and they are capable of constructing the underlying behavior of traffic without strong assumptions on its temporal evolution. However, data-driven approaches might not be able to integrate spatiotemporal traffic flow dynamics under all cases, as these are mainly governed by the queue formation and dissipation at point bottlenecks. Abrupt changes of traffic phenomena (e.g., lane changes, capacity drop, merge behavior, oscillations) can affect congestion development and propagation in various ways that require physical than statistical models to be explained. These characteristics of traffic’s transitional behavior and the existence of variant traffic regimes may not be identified by statistically-oriented or data driven approaches and increase their estimation and prediction errors.

This paper presents a predictive travel time methodology based on speed data at fixed loop detectors. However, in contrast to the aforementioned existing methodologies, it benefits from the available traffic flow fundamentals (e.g., shockwave, bottlenecks). The proposed method makes use of both historical and real-time traffic information to provide travel time prediction. Instead of identifying traffic flow patterns using statistical methods, that sometimes might not succeed to capture complex phenomena of traffic flow, we propose to integrate in the methodology, identification of traffic patterns with traffic flow theory fundamentals, for example with shockwave analysis and bottleneck estimation. First, an existing bottleneck identification algorithm is utilized to determine the location and spatial extent of the bottlenecks (10). It uses speed readings at fixed detector locations as an indicator of bottleneck activation. Identified congested locations are used in this study to restore the major traffic events likely to be observed on the roadway and to construct the link between real-time traffic information and historical dataset. Using the shockwave phenomena and identified congested locations in real-time, the impact of a bottleneck can be predicted before it completely develops. Historical information can be very useful to determine the characteristics of the bottlenecks (i.e., spatial extent and duration) and so, predict their impacts. Nevertheless, as we will show later, traffic conditions significantly vary from day to day (even for similar demand conditions) and as a result the size of a bottleneck in the time-space domain and travel speed of vehicles in this domain have high fluctuations. Thus, a simple prediction based on historical average or a partitioning of traffic conditions based on days (weekdays-weekends) or times of day (AM or PM peak) might introduce significant estimation errors.

**METHODOLOGY**

**Bottleneck Identification Algorithm**

Chen et al. (10) developed an algorithm to automatically identify bottleneck locations, their activation and deactivation times, and their spatial extents using loop detector data and focusing on speed measurements. Chen method compares each pair of detectors adjacentlly located and determines the existence of bottleneck when
Yildirimoglu M., Geroliminis N.

- Speed difference between upstream and downstream detectors is above the minimum speed differential, $\Delta v_{\text{min}}$ threshold.
- Speed at upstream detector is below the maximum speed threshold, $v_{\text{max}}$.

Chen et al. (10) chose values of $v_{\text{max}}=40\text{mph}$ and $\Delta v_{\text{min}}=20\text{mph}$ with data aggregated at 5min intervals taken from California freeways. These parameters may need to be adjusted depending on the application.

The congested region affected by a bottleneck can be defined using the speed measurements at upstream locations. This description is slightly modified in our estimation from the original algorithm. A congested region associated with a bottleneck ends at the detector location where two consecutive upstream detectors have more than 40 mph speed, while a single detector with more than 40 mph is sufficient to enclose the congested region in the original algorithm. We note that with this alternation, the methodology provides better identification especially in the offset of congestion. Identification of bottlenecks in an automated way allows us to restore the major traffic events that occur on the roadway and to keep track of traffic conditions in real time.

**Clustering**

Historical traffic patterns can be used for prediction of travel time. To use the historical dataset in a useful and efficient manner, days with similar traffic patterns (i.e. speed profiles) should be identified. Otherwise large variations and temporal bias might be experienced by utilizing very heterogeneous data. Clustering techniques have been already used in transportation field to analyze traffic flow patterns, see for example Weijermars and Van Berkum (11). Since travel times are computed using local velocity measurements in this study, time-dependent speed measurements along the roadway must be used in the clustering step. Without clustering the variance of travel time for a given departure time is significantly larger and this has a direct erroneous effect in the prediction.

**Principal Component Analysis**

Since high number of detectors and time periods in a day lead to a large number of observations, it is not straightforward to define a metric to cluster days with strong similarity for the freeway route under consideration. Principal Component Analysis (PCA) is a well-established technique to reduce the dimensions of the dataset and to compress the data, see for example Nagendra and Khare (12). PCA, using the orthogonal transformation, converts a set of observations with correlated variables into a set of observations with linearly uncorrelated variables, which are called principal components (PC). In other words, it transforms the data into a new space which has most of the information of the original data, but with a lower dimension. It lists the components in such a way that the first principal component has the largest variance. With respect to our problem, a freeway route might have detectors installed every a few hundred meters that provide speed and flow measurements every a few minutes. This creates a huge information of data, that cannot be directly utilized to cluster different days.

Suppose $(m \times n)$ matrix $X$ is the original data set whose rows correspond to observations (e.g. days) and columns correspond to the variables (e.g. time-dependent local velocity measurement). PCA algorithm can be summarized as follows;

- Subtract from each element the mean value of the corresponding column:
\[ X[m,n] \leftarrow X[m,n] - E_n[X] \]
- Compute correlation matrix \( C = XX^T \)
- Compute eigenvalues \( |C - \lambda_i I| = 0, i = 1,2, \ldots n \)
- Compute eigenvectors \( C e^i = \lambda_i e^i \)
- Choose \( p < n \) eigenvectors: \( e^1, \ldots, e^p \) with \( \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \)
- Project data onto new space \( S = P \ast X, P = \begin{bmatrix} e_1^1 & \cdots & e_p^p \\ \vdots & \ddots & \vdots \\ e_1^n & \cdots & e_p^n \end{bmatrix} \)

A certain variance threshold can be used to determine first \( p \) eigenvectors or principal components (PC).

**Gaussian Mixture Modeling**

After reducing the dimensions of the dataset, Gaussian Mixture Model (GMM) is applied to create clusters in the historical dataset. GMM is the combination of multivariate normal density components, and it fits the data using expectation maximization (EM) algorithm. GMM is often used for clustering purposes, and unlike other clustering methods, it is not solely based on the distance between the observations, but it is based on the distribution of data points. GMM is a more appropriate method than \( k \)-means clustering, when cluster have different sizes and correlation within them, which is why it is applied in this study.

Consider \((mxn)\) matrix \( X = \{x_i\}_{i=1}^m \) whose rows correspond to observations (e.g. days) and columns correspond to the principal components. The probability density function (PDF) of \( X \) will be modeled as mixture of \( K \) gaussians.

\[
p(X) = \sum_{k=1}^{K} \alpha_k \ast p(X|\mu^k, \Sigma^k) \quad (1)
\]

PDF of each Gaussian: \( p(X|\mu^k, \Sigma^k) = N(\mu^k, \Sigma^k) \)

Mixing coefficients: \( \Sigma^k \alpha_k = 1 \),

\( \mu^k, \Sigma^k \): mean and covariance matrices of Gaussian \( k \),

Probability that the data is explained by Gaussian \( k \): \( \alpha_k = p(k) = \sum_{i=1}^{M} p(k|x^i) \).

The parameters of GMM are the means, covariance matrices and mixing coefficients;

\[
\Theta = \{\mu^1, \ldots, \mu^K, \Sigma^1, \ldots, \Sigma^K, \alpha^1, \ldots, \alpha^K\} \quad (2)
\]

However, the parameters of GMM cannot be directly estimated because of unobserved latent variables. Expectation-Maximization (EM) is an iterative method to find maximum-likelihood estimates of the parameters, where the model depends on unobserved latent variables. EM attempts to find the optimum of the likelihood of the model given the data;

\[
\max_{\Theta} L(\Theta|X) = \max_{\Theta} p(X|\Theta) \quad (3a)
\]
At each estimation step, new parameters can be estimated as:

\[
\alpha_k^{(t+1)} = \frac{1}{m} \sum_l p(k|x^l, \Theta^{(t)})
\]

\[
\mu_k^{(t+1)} = \frac{\sum_l p(k|x^l, \Theta^{(t)}) x^l}{\sum_l p(k|x^l, \Theta^{(t)})}
\]

\[
\Sigma_k^{(t+1)} = \frac{\sum_l p(k|x^l, \Theta^{(t)}) (x^l - \mu_k^{(t+1)})(x^l - \mu_k^{(t+1)})^T}{\sum_l p(k|x^l, \Theta^{(t)})}
\]

The parameter set \( \Theta \) is updated iteratively until the log likelihood is increased by less than a certain threshold value.

Optimal number of clusters (Gaussians) can be determined by the use of average silhouette width and information criteria such as Akaike Information (AIC) and Bayesian Information (BIC) criteria. In addition to the optimal number of clusters, the stability of the results is crucial to clustering. GMM, whose initialization is random or based on k-means results, should return the same results every time it is repeated to ensure the accuracy and the robustness of the algorithm. By varying the number of components in clustering, we can identify a value that will create robust clusters.

**Stochastic Congestion Maps**

To add a probabilistic flavor in the bottleneck identification algorithm, that will contribute in the accurate prediction of experienced travel times, we introduce a new physical concept of describing spatiotemporal traffic patterns, which is defined as stochastic congestion map. It represents the likelihood of observing congestion at a given space-time point. Stochastic maps can be created for each cluster, and they can be used to predict travel time for a day which belongs to the given cluster. The probability of observing congestion at roadway segment \( i \) at time \( t \) is calculated separately for each cluster \( k \).

\[
p^k(i, t) = \frac{1}{D_k} \sum_{m=1}^{D_k} f(i, t, m)
\]

\[
f(i, t, m) = \begin{cases} 1 & \text{if segment } i \text{ is congested at time } t \text{ on day } m \\ 0 & \text{otherwise} \end{cases}
\]
where \( D_k \) is the number of days in cluster \( k \). Once clusters are created and bottleneck identification algorithm is applied, a stochastic congestion map (FIGURE 2a), which represents the likelihood of observing congestion at a given space-time point, can be created for each cluster and it can be used to predict travel time for a day which belongs to the given cluster.

FIGURE 2 - a. Stochastic Congestion Map, b. Subsets in Congestion Map

To integrate the stochastic congestion maps in an automated framework, a discretization is made. Each cluster is divided into subsets using certain threshold probability values (e.g. from 0.05 to 1), and a congestion map corresponding to a certain likelihood of occurrence is created for each subset. Congestion map associated with the lowest threshold is constructed by the bottleneck points (in the time space domain) whose probability is greater than 0.05 (i.e. occurring more than 5% of the analyzed days), the map associated with second lowest threshold is constructed by the points whose probability is greater than 0.10, and so on so forth (like a cumulative 2D distribution). FIGURE 2b represent the subsets created within the congestion map for two different clusters. Note that darker colors represent higher threshold values, and the higher the value of the threshold is, the smaller the size of the bottleneck is. Also note that the difference between subsets is roughly in the shape of asymmetric rings around a given core. These results are of great importance to our analysis because they show that even the location and duration of bottlenecks are roughly known a priori, a more careful look identifies strong stochastic phenomena that can vary travel times from one day to the other. For example, bottleneck block #5 in FIGURE 2b starts at location with milepost 40, but its extension in time and space varies from day to day.

This study attempts to predict experienced travel times under recurrent traffic conditions. The method might also be able to integrate non-recurrent phenomena once these have been identified. For example, if an accident creates much higher delays than recurrent bottlenecks (as this is expressed by low values of thresholds in the stochastic congestion maps), the travel time prediction can integrate some component of this oversized delay. However, travel time for a specific departure time can vary significantly even within the cluster. To address the travel time variability and to provide more accurate travel time information, an update algorithm is developed in this study that allows us to switch between subsets and congestion maps associated with them.
Before delving into the online update algorithm, one should notify that there are multiple cores in the congestion map around which the rings form. In addition, although there may be a correlation between the size of the blocks for a given day, it is intuitively known that they are not fully dependent. Therefore, the congestion map is divided into blocks as it is represented by the numbers in FIGURE 2b, and the update algorithm is applied separately for each of them to select the threshold value that best represents the real-time traffic conditions on that part of space-time plot.

The idea behind the update algorithm is to provide travel times based on the expected traffic conditions at the very beginning and integrate real-time congestion information to specify the shape of the bottleneck in future time periods. If there is no real-time information about the given block at the departure time, the expected block size (e.g. probability of 0.5) is used to compute the predicted travel time. However, the given departure time in FIGURE 2b allows us to compare real-time traffic information with congestion maps for bottleneck blocks #5 and #6. Hence, the algorithm starts switching between threshold values to find the bottleneck shape that would best represent the real-time bottleneck information obtained till the departure time. Threshold value returned by the update algorithm for each block is used to construct the predicted congestion map, and travel time for the given departure time is computed on this time-dependent congestion map. A graphical representation of the algorithm is provided later.

For a given block, selection of threshold value is done by the following similarity metric;

\[
\begin{align*}
\arg \max_i & \left( f(i,j,k) + g(i,j,k) \right) \\
& f(i,j,k) = TP(c^k_{i,j}, RTI) \\
& g(i,j,k) = TN(c^k_{i,j}, RTI, c^k_{i,j})
\end{align*}
\]

\(i\) is the threshold index, \(j\) is the block index, \(k\) is the cluster index, \(RTI\) real-time information (congested points observed till the departure time), \(c^k_{i,j}\) is the set of points defined by block \(j\), threshold \(i\) and cluster \(k\), \(TP\) (true positive) is the number of rightly classified congested points of \(RTI\) by \(c^k_{i,j}\), \(TN\) (true negative) is the number of rightly classified non-congested points of \(RTI\) by \(c^k_{i,j}\) in the set defined by \(c^k_{i,j}\).

Note that the set of points that the algorithm uses to determine \(TN\), is defined by the maximum size that the given block can get. Since real-time information is available only till departure time, the points up to the departure time are used in the update step. In addition, a moving time window of 2 hours is used to keep track of varying conditions. The algorithm starts with the set of points defined by the lowest threshold value (the maximum size of the block) and up to the departure time, and then switches to a time window of 2 hours after the time difference between the departure time and the start of the block gets larger than 2 hours.
If the same performance value is computed for multiple threshold values, then the algorithm picks the one that is closest to the expected threshold value. This indicates that the algorithm always selects the conditions that are most likely to be observed.

**Speed Profile**

The developed methodology in travel time prediction requires two pieces of future information (i) prediction of major traffic events on the roadway (e.g. bottlenecks) and (ii) prediction of speed profiles. By the methodology described so far, this study attempts to predict major traffic congestion events on the roadway. However, the future speed profile inside and outside the bottleneck time-space domain is still unknown. Our approach will utilize two different speeds, one for congested, $V_c$, and one for uncongested conditions, $V_f$. and will estimate the trajectory of a vehicle which runs in the predicted time-space domain of a congestion map.

Instantaneous speed information (i.e. speed measurements at the departure time) can bring valuable insight into this problem. The assumption that instantaneous speed will remain constant separately for congested and non-congested segments can help us estimate a speed profile.

$$V_c = \frac{1}{N_c} \sum_{i=1}^{N_c} v_{i}^{t_d} \quad \text{and} \quad V_f = \frac{1}{N_f} \sum_{i=1}^{N_f} v_{i}^{t_d}$$ (7)

where $t_d$ is departure time, $v_{i}^{t_d}$ is the speed measurement on roadway segment $i$ at $t_d$, $N_c$ is the number of detectors registered as congested at $t_d$, $N_f$ is the number of detectors registered as uncongested at $t_d$.

However, the speed profile created by Equation (7) does not account for the speed variability along the roadway and it can produce erroneous results, especially when congested speeds are different as threshold values vary. Therefore, the congested speed, $V_c$, is modified for each roadway segment $s$. If a particular roadway section is registered as congested at the departure time, and if it remains congested in the predicted congestion map, its speed at the time of arrival, $v_{i}^{t_a}$, is the average of the congested speed and the speed of the section at the time of departure, i.e. $v_{i}^{t_a} = \frac{1}{2}(V_c + v_{i}^{t_d})$. In this way, local congestion phenomena are integrated in the prediction.

**CASE STUDY**

For the application, the data from California freeway performance measurement system (PeMS) is used. PeMS collects 30-sec loop detector flow and occupancy data throughout the state. Then, it processes them and fills in the missing detector data to compute 5-minute flow, occupancy and speed averages (13).

For this study, a 60 mile section of I-5S in the district of San Diego/Imperial is selected. Considering the detector quality and the effect of recurrent congestion, selected roadway section is quite suitable for this study. 5-minute loop detector data is
downloaded through PeMS website for the year of 2011. The dataset is divided randomly into two parts; Training Set (~80%-289 days) and Testing Set (~20%-76 days).

This section will give details about the application of clustering step and the development of stochastic congestion maps.

**Clustering**

Before clustering, PCA is applied to reduce the dimensions of the dataset. 100 principal components carry 95% of the variance in the original data of 16020 variables (89 detectors * 180 time periods between 6AM-9PM). Therefore, the rest of the operations is carried out with the reduced dataset of 100 variables.

FIGURE 3 presents 100 realizations of clustering for several number of clusters, and the average silhouette width values that result from these realizations. It clearly shows that results are not stable for any cluster number other than three. In addition, mean average silhouette width reaches its maximum value at three clusters. Therefore, considering both the stability of the results and the similarity of observations within the cluster (i.e. average silhouette width), optimal number of clusters is selected to be three. We have noticed that by utilizing a higher number of clusters in the methodological framework, the experienced travel time prediction does not improve.

![FIGURE 3 - Average Silhouette Width vs. Cluster Number](image)

Although AIC and BIC show a decreasing trend for increasing number of clusters, average silhouette width reaches its optimum value at three clusters. The optimal number of clusters indicated by AIC and BIC is also tested. In fact, any number of clusters except three brought unstable results. Therefore, considering both the stability and the average silhouette width results, three clusters is selected to be the most appropriate configuration in this analysis.

FIGURE 4 presents the GMM results based on days of the week and based on clusters. Ellipses in FIGURE 4 represent 50% and 90% of the variance of the clusters along the dimensions of two PC’s. Note that PC values do not have a physical meaning;
they represent the values of the new uncorrelated variables. It clearly shows that the
clusters are mainly dominated by certain features of days of the week. The first cluster
shown in FIGURE 4b is dominated by weekend days, the second week days other than
‘Friday’s, and the third by ‘Friday’s.

![Figure 4 - GMM Results](image)

Table 1 presents the distribution of days along the clusters. Most of the week days
classified in the first cluster are holidays and they are not subject to a significant level of
congestion. The second cluster is mainly composed of week days other than ‘Friday’s.
These days have significant level of congestion. However, the level of congestion is not
as high as it is observed on ‘Friday’s. Therefore, the clustering algorithm creates a
separate cluster mainly for ‘Friday’s. Although the clusters do not totally belong to a
certain day of the week, this information is very useful to identify expected traffic
conditions on the roadway for a particular day. Hence, each cluster is assigned to
dominating days of the week in a predetermined way, and travel time prediction for a
given day is executed within the corresponding cluster and its associated congestion map.

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Days</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Mon 6</td>
</tr>
<tr>
<td>2</td>
<td>Mon 36</td>
</tr>
<tr>
<td>3</td>
<td>Mon 0</td>
</tr>
</tbody>
</table>

**Stochastic Congestion Map**

By combining the results obtained from bottleneck identification and clustering steps,
stochastic congestion maps can now be created. By simply estimating the average
number of congestion observation for a space-time point, such a map can be constructed.
FIGURE 2a presents the stochastic congestion map for the third cluster described above.
Then, it is divided into subsets for different threshold values.

Note that the shape of a block in the congestion maps defined by threshold values
may not be totally proper regarding the traffic flow essentials (e.g. shockwave speeds).
This may be because of the flaws of the bottleneck identification algorithm, or taking the average number of occurrence may ignore the points that would help us have a well-defined shape. However, the purpose of this study is to predict experienced travel times, not to predict the shape of the bottlenecks. The important feature of an accurate methodology is to estimate with some confidence the number of congested points (in the time-space domain) that the predictive trajectory will hit, not the shape of the bottleneck.

RESULTS

Described methodology is tested on the training set (76 days). Since the weekend days are not subject to significant level of congestion, they are not considered in the evaluation step.

Travel time computation in this study is done by constant speed interpolation. For a link between two successive detectors, the speed measurement at downstream or upstream detector, or the average of two measurements can be used to represent the velocity. All constant speed interpolation methods imply instantaneous speed changes, which do not occur in real-time. However, considering the distance between the detectors (about 500m) in our study site, this phenomenon is not expected to largely affect the results. Travel time can also be computed by using linear and quadratic speed interpolation methods, which do not require instantaneous speed changes. Alternatively, one could apply a more detailed traffic flow model (of first or higher order) to estimate speed between the detectors. Nevertheless, we do not expect the accuracy of the results to improve. However, this study only uses constant speed interpolation method, with the average speed measurement of two successive detectors.

Instantaneous travel time is calculated by the use of the speed measurements at the departure time and the constant speed interpolation method, described above. On the other hand, experienced travel time is calculated by traveling a trajectory through the velocity field. The time it takes to travel each segment is calculated, and the speed measurements at the time when the trajectory reaches the next segment is used to compute its travel time. Predicted travel time is calculated in the same manner as experienced travel time. However, instead of velocity field which is unknown at the departure time, predictive trajectory walks over the congestion map and uses an estimated speed profile to compute the time it takes to traverse each segment.

Congested periods in the testing set are identified with an automated algorithm. Note that a period is registered as congested if any of the instantaneous, predictive and experienced trajectory hits a congested point. The implication of doing this is that a period can be registered as congested, although experienced travel times do not indicate delays. In case traffic conditions are less congested than the expected ones of the given cluster, predictive trajectory may hit some congested points, although experienced travel times do not suffer from that. Therefore, our travel time prediction methodology experiences some time lag to adapt to real-time traffic conditions.

For the sake of comparison, historical travel times are also computed for each day of the week. The median value of the experienced travel times at a given departure time on a given day is taken as historical average value.

To measure the effectiveness of the methods, two statistics, namely mean absolute error (MAE) and mean absolute percentage error (MAPE) are utilized;
\( MAE = \frac{1}{n} \sum_{t=1}^{n} | T(t) - \hat{T}(t) | \)  

(8)

\( MAPE = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{T(t) - \hat{T}(t)}{T(t)} \right| \times 100 \)  

(9)

where \( n \) is the number of observations, \( T(t) \) is the experienced travel time, and \( \hat{T}(t) \) is the travel time provided by the methodology.

The weighted average of MAE (for identified 103 congested periods) with respect to congestion duration is 2.10, 3.05 and 5.83 minutes for predictive, instantaneous and historical average approaches respectively, which shows that our methodology outperforms the other two.

FIGURE 5 presents MAE values of instantaneous and predictive travel time for identified 103 congested periods in the testing set. Since historical average performs clearly worse than the other methods, it is not shown in FIGURE 5. In 85 out 103 cases, predictive travel time methodology produces better results than instantaneous travel time assumption.

FIGURE 5 - Instantaneous vs. Predictive Travel time Performance for Congested Periods
FIGURE 6 - Comparison of Models a. 10-Feb-2011, b. 31-Mar-2011, c. 22-Nov-2011
FIGURE 6 presents the travel times provided by predictive methodology, instantaneous approach and historical average method for 3 different days. It clearly shows that historical average is not capable of producing accurate results under congested conditions. In overall, predictive travel time produces better results during both the onset and offset of the congestion. However, the morning peak in FIGURE 6a and the afternoon peaks in FIGURE 6c show that prediction model has a slightly better performance compared with the instantaneous approach during the congestion onset, while it has a significantly better performance during the offset. As we get closer to the end of the congestion, we collect newly available real-time information about the bottleneck and the congested section affected by that. This enables the algorithm to produce better congested area and travel time predictions during the congestion offset.

The most congested days in the testing set can be identified based on the number of congested sections observed. 14 most congested days in the testing set are identified based on this criterion, and the average errors throughout the days are reported in Table 2. For all the days, average performance of predictive methodology is better than both instantaneous and historical average approaches.

<table>
<thead>
<tr>
<th>Date</th>
<th>Day</th>
<th>Prediction MAE</th>
<th>Prediction MAPE</th>
<th>Instantaneous MAE</th>
<th>Instantaneous MAPE</th>
<th>Historical Average MAE</th>
<th>Historical Average MAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>24/06</td>
<td>Fri</td>
<td>2.31</td>
<td>3.34</td>
<td>2.54</td>
<td>3.76</td>
<td>12.08</td>
<td>16.31</td>
</tr>
<tr>
<td>21/07</td>
<td>Thu</td>
<td>1.47</td>
<td>2.46</td>
<td>1.86</td>
<td>3.13</td>
<td>6.49</td>
<td>10.44</td>
</tr>
<tr>
<td>08/07</td>
<td>Fri</td>
<td>1.43</td>
<td>2.29</td>
<td>2.40</td>
<td>3.61</td>
<td>7.85</td>
<td>10.98</td>
</tr>
<tr>
<td>02/12</td>
<td>Fri</td>
<td>1.77</td>
<td>2.60</td>
<td>2.93</td>
<td>4.39</td>
<td>7.52</td>
<td>10.48</td>
</tr>
<tr>
<td>14/07</td>
<td>Thu</td>
<td>2.91</td>
<td>4.08</td>
<td>3.08</td>
<td>4.55</td>
<td>6.98</td>
<td>9.34</td>
</tr>
<tr>
<td>18/01</td>
<td>Tue</td>
<td>2.72</td>
<td>4.09</td>
<td>3.17</td>
<td>4.94</td>
<td>5.38</td>
<td>7.71</td>
</tr>
<tr>
<td>16/11</td>
<td>Wed</td>
<td>1.92</td>
<td>3.13</td>
<td>2.67</td>
<td>4.39</td>
<td>3.87</td>
<td>6.16</td>
</tr>
<tr>
<td>22/11</td>
<td>Tue</td>
<td>2.32</td>
<td>3.51</td>
<td>3.90</td>
<td>5.97</td>
<td>5.67</td>
<td>7.69</td>
</tr>
<tr>
<td>09/12</td>
<td>Fri</td>
<td>1.05</td>
<td>1.78</td>
<td>2.10</td>
<td>3.42</td>
<td>4.76</td>
<td>7.29</td>
</tr>
<tr>
<td>10/02</td>
<td>Thu</td>
<td>1.23</td>
<td>2.12</td>
<td>1.91</td>
<td>3.31</td>
<td>3.72</td>
<td>6.15</td>
</tr>
<tr>
<td>06/12</td>
<td>Tue</td>
<td>1.51</td>
<td>2.54</td>
<td>2.50</td>
<td>4.19</td>
<td>3.47</td>
<td>5.45</td>
</tr>
<tr>
<td>25/08</td>
<td>Thu</td>
<td>0.93</td>
<td>1.65</td>
<td>1.80</td>
<td>3.02</td>
<td>3.27</td>
<td>5.28</td>
</tr>
<tr>
<td>12/12</td>
<td>Mon</td>
<td>3.14</td>
<td>4.54</td>
<td>3.22</td>
<td>4.85</td>
<td>9.53</td>
<td>13.06</td>
</tr>
<tr>
<td>20/04</td>
<td>Wed</td>
<td>1.18</td>
<td>2.04</td>
<td>1.87</td>
<td>3.17</td>
<td>1.45</td>
<td>2.37</td>
</tr>
</tbody>
</table>

CONCLUSION

The disposition of travel time information through ATIS or its use as in ATMS to deploy efficient control measures always requires the prediction of traffic conditions on the freeway. The aim of this paper is to predict travel times by using traffic flow fundamentals, not any statistical procedure.
First, an automated bottleneck identification algorithm is used to detect the major traffic events that occur on the freeway. Then, the historical (or training) dataset is partitioned based on the clusters obtained through GMM. The results obtained from the first two parts are combined to create stochastic congestion maps for each cluster. Next, using the estimated speed profile, the congestion maps associated with threshold values and an update algorithm that compares real-time and historical traffic conditions, this study predicts the experienced travel times.

The experiment results based on the loop detector data of I-5 segment in California/San Diego indicate that the proposed method provides promising travel time predictions under varying traffic conditions. This methodological framework could have a great potential to be applied with trajectory data (e.g. GPS devices or smart phones) instead of loop detectors or with a combination of different sensoring techniques. Existing advances in bottleneck identification with trajectory data can make this approach easily implementable. An estimation and prediction of travel time distributions for different departure times should also be a research priority, as reliability measures can improve the planning of travel trips for various users and provide tools to traffic management for more efficient control.

REFERENCES


