Traffic Analysis for Holidays and Planned Events

Jing Dai
T.J. Watson Research Center
IBM
jddai@us.ibm.com

Chang-Tien Lu
Department of Computer Science
Virginia Polytechnic Institute and State University
ctlu@vt.edu

Xiang Fei
T.J. Watson Research Center
IBM
xfei@us.ibm.com

Abstract — Holidays and planned special events can affect not only the traffic towards the events, but also the normal commuters using the adjoining roadways. Transportation users who may be affected include the general public, transit (e.g., regional, local, or specialized services), and other service providers (e.g., law enforcement, medical, or fire rescue responders). Accurately predicting the traffic during holidays and planned events can support effective traffic management and trip planning, and contribute to better roadway performance and safety. This paper proposes an innovative approach to model the traffic patterns of holidays and planned events by analyzing routine traffic flows and historical instances of holidays and events. Spatial-temporal information of demand growth and road condition change is applied to estimate the traffic of future holidays and events. Case studies on a practical ITS, HOMES, demonstrated the effectiveness of the proposed approach.

Keywords — traffic prediction, holiday, planned events, spatial-temporal data mining

1 Introduction

Nowadays Intelligent Transportation Systems (ITS) provide the functionalities to monitor transportation performance and to analyze traffic patterns. However, holidays and planned events exhibit very unique traffic behavior comparing to daily routine traffic patterns. Holidays and planned events embrace a multitude of activities across a broad range of geographic areas, including both urban and rural environments. More than ten holidays, during which public transportation systems switch schedules and change rates, occur every year throughout the United States. Planned special events can take the form of major one-time events (e.g., Inauguration Day, Olympic Games, and Super Bowl) or occur on a cyclical basis (e.g., football games, NASCAR races, and the Rose Parade). These activities occur in a variety of cities and rural areas across the country where intense periods of attendee arrival and departure overwhelm the transportation system. Regardless of the size, type, or location of the activities, the accompanying roadways, transit systems, and parking facilities must be capable of handling the increased traffic volume. Throughout the duration of any holiday or planned special event, one of the primary goals of transportation agencies is to reduce travel time for motorists, and minimize the disruption of traffic flow for local motorists. One way to achieve this goal is to accurately model the traffic during holidays and planned events.

One major challenge of modeling the traffic of holidays and planned events is that the sample set is limited. Due to the small amount of the historical occurrences of these holidays and events, it's difficult to obtain these patterns with various activities types and different road and weather conditions. Therefore, utilizing daily routine traffic patterns to empower the process of modeling holidays and events becomes critical towards holiday and events traffic analysis. On the other hand, unlike traffic incidents or natural disasters, for a planned event, information on the location, time, duration, weather, and demand is usually expected to be known. The prior knowledge can be exploited to help transportation agencies to estimate the traffic, plan the operations, coordinate resources, and apply advanced traffic management techniques to mitigate adverse impacts that may result from these events.

This paper presents an innovative approach to analyze traffic for holiday and planned events utilizing the Highway Operation Monitoring and Evaluation System (HOMES) [1]. HOMES is an effective visualization system for observing the summarization of spatiotemporal patterns and trends in traffic data. It is designed for browsing the spatial-temporal dimension hierarchy via integrated roll-up and drill-down operations. Figure 1 illustrates the detector stations monitored by HOMES. Using the outputs of HOMES, the proposed approach integrates the limited holiday and planned event traffic records and the daily routine traffic patterns. Simulated traffic patterns for holidays and planned events can be extracted to support active traffic management and travel planning. Case studies applying this approach on Interstate-66 in northern Virginia area have demonstrated its effectiveness.

The rest of this paper is organized as follows. A brief introduction on ITS and traffic data mining is presented in Section 2. Section 3 explains the challenges of holiday and planned event traffic analysis. Details of the proposed solution are discussed in Section 4. Case studies demonstrate the outputs of the proposed approach in Section 5. Finally Section 6 concludes this paper and discusses future directions.

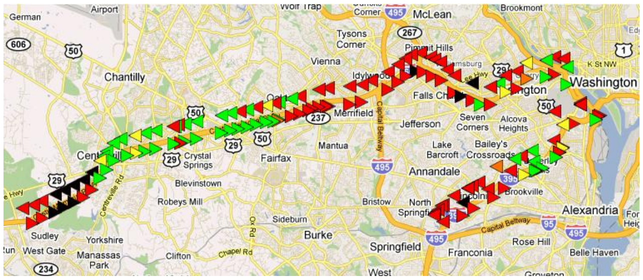


Figure 1. Locations of all stations on I-66 and I-395.

2 Related Work

ITS is defined as the integrated application of advanced sensor, computer, and communication technologies and management strategies to enhance the safety and efficiency of the transportation system. By consistently integrating different system components and technologies, the benefits of increased mobility can be achieved. ITS can be applied in a variety of environments—rural and urban, freeway and arterial—through the use of interconnected traffic signals and area-wide information services. These ITS technologies provide a core communication network, system monitoring, and information processing capabilities that can act as a foundation for the coordinated operation of the transportation system. These key elements make ITS a powerful tool for localities that host planned special events.

Leading practical ITS include PEMS, PORTAL, Smart Trek, and HOMES. PEMS [2] was initially deployed to monitor real-time freeway traffic in California. Various transportation measures are visualized in different presentation formats in PEMS to provide insights to the traffic data in a comprehensive way. PORTAL [3] is a transportation archive system for the Portland-Vancouver metropolitan region. It integrates data sources from traffic sensors, cameras, constructions, incidents, transit system, freight system, and weather stations into a complete transportation information system. Smart Trek [4] contains multiple traffic monitoring and prediction tools that provide real-time traffic performance and transit status of the City of Seattle. To provide real-time monitoring and long-term evaluation of the roadway conditions, an advanced version of AITVS [5], the Highway Operation Monitoring and Evaluation System (HOMES) [1] has been developed as a traffic visualization system that allows for browsing the spatiotemporal dimension hierarchy via integrated data cube operations. The Virginia Department of Transportation (VDOT) currently utilizes HOMES to monitor traffic incidents, analyze roadway behaviors, develop operation strategies, and verify highway designs. In general, HOMES can identify traffic patterns, rules, and anomalies to achieve the following benefits: efficient roadway designs, objective evaluation metrics of traffic policies, improved management of operations and emergent events, and better utilization of the roadway network. A screenshot of the web-based dashboard of HOMES is shown in Figure 2.

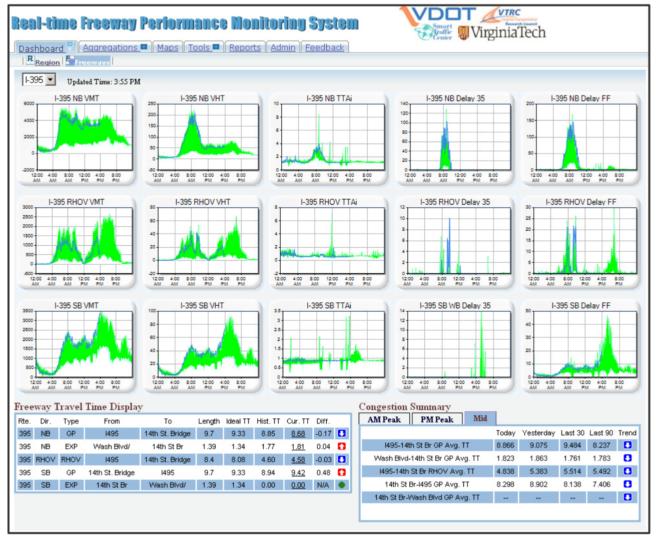


Figure 2. A Screenshot of HOMES User Interface.

Currently, HOMES monitors highways I-66 and I-395 within the Washington Metropolitan Area. It collects streaming information from approximately 850 radar sensors and loop detectors. An important challenge in the development of HOMES is the provision of real-time query processing for both current and historical data under various cube operations. To address this issue, HOMES employs a spatial data warehouse approach as the underlying data management structure [6]. Fast cube-based query response times are achieved by maintaining concurrent sets of aggregated and non-aggregated sensor information. Quick traffic data updates are accomplished by an incremental approach for computing the updated aggregate and non-aggregate representations of the traffic data streams. Based on the traffic data engine of HOMES, incident detection components [7, 8] have been developed for real-time alert.

The traffic contour plot generated by HOMES provides an intuitive overview of freeway performance, and can be used to present daily traffic patterns. An example of contour plot is shown in Figure 3, where the x-axis represents the timestamps, the y-axis represents the mileposts along the freeway, and the color shows the values of traffic measure of the corresponding location and time. Utilizing the spatial and temporal information conveyed in this plot, data mining techniques have been applied to identify incidents in real-time [8]. Both spatial and temporal information are considered to identify the potential incidents. Meanwhile, adaptive learning ability and short detection response time were achieved in this approach. To analyze the high dimensional traffic data, Mahalanobis distance was applied to discover potential incidents according to the traffic pattern. Lifeline style detection and visualization was utilized to provide intuitive user interface.

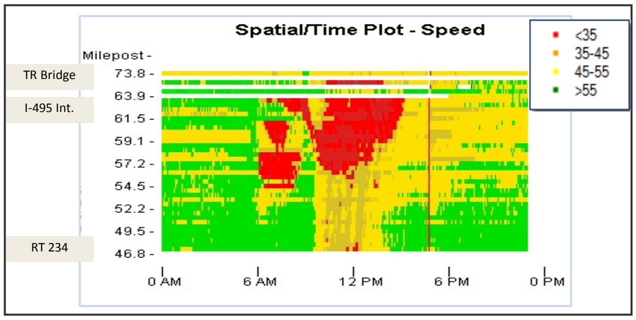


Figure 3. An Example of Traffic Contour Plot.

3 Problem

In this section, we define the targeted problem in this paper. Utilizing the cub-based queries and traffic contour plots generated in HOMES, a complete set of daily routine traffic patterns, PT, for certain date ranges can be obtained by integrating non-incident historical traffic data with weather information. This set of patterns includes all the combinations of traffic on different weekdays with different road conditions. Each item in this set is denoted as PT(days, Road Condition). For example, PT(Mondays in 2009, Clear), PT(Mondays in 2009, Rain), PT(Fridays in Dec. 2008, Snow). Besides these routine patterns, at least one instance of the traffic of the target holiday or planned event has been recorded in the system. Each instance is denoted as INS(holiday/event, Road Condition). For example, the Labor Day 2010 is the target holiday, and there is an instance of Labor Day 2009 with clear road condition, INS(Labor_Day 2009, Clear). In addition, the estimated road condition of the target holiday or planed event needs to be obtained before applying the proposed approach. To appropriately estimate traffic over a certain road segment, the capacity of the road needs to be considered as it determines the upper bound of the possible traffic flow. The terms required to define the problem are listed in Table 1.

Table 1. Terminology.

Terms	Description
Road Condition	Clear, Rain, Snow, Ice, Fog, etc.
Capacity	The maximum traffic flow allowed on given freeway.
PT(days, Road Condition)	Traffic patterns of a given set of days in date range under given Road Condition.
INS(day, Road Condition)	Traffic instance record of a given day under given Road Condition.

The purpose of the proposed approach is to predict the traffic condition of a given holiday or planned event, so that the public agencies can better manage the transportation resources and the travelers can make appropriate plans. Traffic measures, including speed, volume/flow and occupancy, can represent the traffic patterns and instances. Following the convention, traffic flow is used for the traffic prediction. A formal definition of this problem can be described as follows:

Given a complete set of traffic patterns PT, at least one instance of H, INS(H, RC), and capacity of the freeway, estimate the traffic pattern PT(H, RC*), where H is a holiday or planned event, and RC* is the road condition of the target H and RC* != RC.

In this definition, the target holiday or planned event should have at least one occurrence in the data archive. Most holidays occur every year, however, some events are not recurrent at the same location, e.g., Olympic Games. If there is no instance recorded for a target holiday or planned event, the most similar one will be selected by domain experts as replacement. On the other hand, if there are multiple instances recorded for a target holiday or planned event, the generated PT(H, RC*) should take all these instances into consideration.

Generally speaking, to estimate the traffic situation of a given day, the more daily traffic records under similar conditions available, the more accurate prediction can be achieved. There are usually insufficient number of samples of holidays and events for good prediction, because they do not occur frequently. Therefore, a sophisticated approach is needed to solve this problem

4 Proposed Approach

The fundamental idea of the proposed approach is to apply the existing instances of the target holiday or planned event as the baseline, and utilize the daily routine traffic under different road conditions to generate the condition offset and the growth factor accordingly. The condition offset measures the impact of different road conditions on traffic flow. The growth factor reflects the growth of traffic demand from the baseline to the target holiday or planned event. Integrating the baseline with the corresponding offset and growth factor, an appropriate estimation can be obtained. For example, as shown in Figure 4, the existing instance of Thanksgiving Day of 08' under clear road condition is treated as the baseline to estimate Thanksgiving Day of 09'. The condition offset is calculated as the difference between the traffic patterns of typical Thursday under fog and rain condition and of typical Thursday under clear condition, using the 2 months before Thanksgiving Day 09'. Meanwhile, the growth factor is computed to capture the difference between the traffic patterns of the 2 months before Thanksgiving Day 08' and the 2 months before Thanksgiving Day 09'. The pattern of Thanksgiving Day 09' under fog and rain condition can then be estimated by applying both the condition offset and growth factor to the baseline.

In the above example, there are four major steps to generate the final pattern, namely, Baseline Generation,

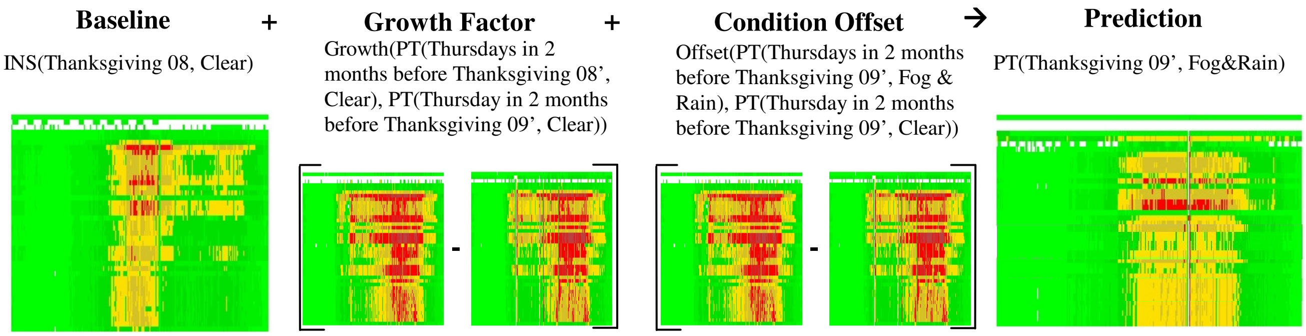


Figure 4. Example of Holiday Traffic Prediction.

Growth Factor Computation, Condition Offset Calculation, and Pattern Estimation. These four steps are described in detail as follows

Baseline Generation

In this step, existing instances of the traffic of the target holiday or planned event under different road conditions will be used to construct the baseline. The baseline may consist of instances/patterns, at most one for each road condition. Taking a holiday or planned event *H* as inputs, the pseudo code to generate the baseline *B* is illustrated in Algorithm 1. This algorithm inspects each available road conditions for *H* in historical records. If there is only one instance for a specific road condition, that instance is added to *B*; if there are multiple instances for a road condition, a procedure named getPattern() will be invoked to create a traffic pattern based on these instances. Various approaches can be used to generate a pattern from multiple instances. In this implementation, the mean of all the instances is chosen as the pattern to demonstrate the idea.

```
Input: Target holiday/planned event H
Output: Baseline B

For each available road condition RC for existing instances
of target holiday/planned event
   If there is only one instance of H, h, under RC
      PT(H', RC) = INS(h, RC); //treat as a pattern
   Else
      PT(H', RC) = getPattern(I | I= INS(h, RC));
   B = B + PT(H', RC);
Return B;
```

Algorithm 1. Baseline Generation.

Growth Factor Computation

The change of the traffic demand from the existing instances to the target holiday or planned event is computed as the growth factor in this step. The growth factor reflects the demand change caused by construction, regional development, season, and economy. Taking the target holiday or planned event *H*, the baseline *B* created in the first step, the forecasted road condition *RC* for *H*, and the number of sample days *S* as inputs, the algorithm to calculate the growth factor *G* is shown in Algorithm 3. The algorithm first determines the corresponding date range by taking *S* days before *H*. Then the traffic pattern during these days under the same road condition as *H* are generated using the getPattern() procedure. Similarly, the sample date ranges are obtained by taking *S* days before each instance in *B*. Each pattern generated using these sample days under road condition *RC* is input into the getDiff() procedure to compute its difference against the traffic pattern of the sample days before *H*. Either absolute or relative differences can be used to calculate the difference between two daily traffic patterns. This implementation applies the relative difference in the implementation. Using the notations in Algorithm 3, the getDiff() procedure can be expressed as *PT(DRH, RC) / PT(DR, RC)*.

```
Input: Target holiday/planned event H, Baseline B, # of
        Sample days S, Forecasted road condition RC
Output: Growth Factor G

DRH = getDateRange(H, S); //get S days before H
PT(DRH, RC) = getPattern(I | I=INS(h, RC), h ∈ DRH);
For each pattern PT(H', RC) in B
   DR = getDateRange(H', S); //get S days before H'
   PT(DR, RC) = getPattern(I | I=INS(h, RC), h ∈ DR);
   GDIF(H, H') = getDiff(PT(DRH, RC), PT(DR, RC));
   G = G + GDIF(H, H');
Return G;
```

Algorithm 2. Growth Factor Computation.

Condition Offset Calculation

The difference between two traffic patterns is calculated as the offset in this step. The offset contains a set of differences between the traffic patterns under the forecasted road condition and other road conditions of the corresponding routine day. Taking the target holiday or planned event *H*, the forecasted road condition *RC* for *H*, the number of sample days *S*, and the baseline *B* created in the first step as inputs, the algorithm to calculate the condition offset *C* is shown in Algorithm 3. The algorithm first determines the corresponding routine day by taking the same weekday as *H* in the past *S* days. Then the traffic patterns of this routine day under different road conditions are generated using the getPattern() procedure. Each pattern with its road condition other than *RC* is input into the getDiff() procedure to derive its difference against the traffic pattern under *RC*. These differences are then output as the condition offsets.

```
Input: Target holiday/planned event H, Forecasted road
        condition RC, Baseline B, # of Sample days S
Output: Condition Offset C

H' = getRoutineDay(H, S); //get same weekdays as H in past
S days
PT(H', RC) = getPattern(I | I=INS(h, RC), h ∈ H');
For each road condition RC' in B and RC'!=RC
    PT(H', RC') = getPattern(I | I=INS(h, RC') , h ∈ H');
    CDIF(RC, RC') = getDiff(PT(H', RC), PT(H', RC'));
    C = C + CDIF(RC, RC');
Return C;
```

Algorithm 3. Condition Offset Calculation.

Pattern Estimation

```
Input: Forecasted road condition RC, Baseline B, Growth
        Factor G, Condition Offset C, Capacity CP
Output: Estimated pattern PT(H, RC)

CAND = ∅;
For each pattern PT(H', RC') in B and RC'!=RC
    Get GDIF(H, H') from G;
    Get CDIF(RC, RC') from C;
    CAND = CAND + applyCapacity(applyCondition
(applyGrowth (PT(H', RC'), GDIF(H, H')), CDIF(RC,
RC')), CP);
PT(H, RC) = getPattern(I | I ∈ CAND);
Return PT(H, RC);
```

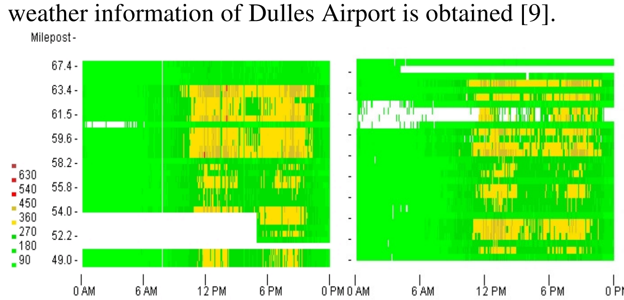
Algorithm 4. Pattern Estimation.

Once both the baseline and difference are obtained, the future traffic of the target holiday or planned event under the forecasted road condition can be estimated. As illustrated in Algorithm 4, taking the forecasted road condition *RC*, the baseline *B*, the growth factor *G*, the offset *C*, and the capacity *CP* as inputs, the estimated pattern can be generated in the following process. For each instance/pattern in *B*, a procedure applyGrowth() is called to apply the corresponding growth fact in *G* on the baseline instance/pattern. Since the relative difference is used in getDiff(), this procedure adds the growth factor to the appropriate baseline by multiplying *GDIF(H, H')*. After the growth factor is applied, a procedure applyCondition() is invoked to apply the corresponding condition offset from *C* using a similar process as applyGrowth() to generate a candidate pattern. The candidate pattern then is refined by setting the maximum traffic volume as the capacity of the road *CP*. Once all the candidate patterns are gathered, the procedure getPattern() is invoked to generate the final estimated pattern.

5 Implementation & Case Study

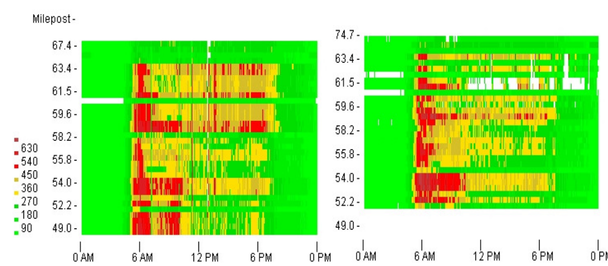
We implemented the proposed holidays and planned events traffic prediction on HOMES. Based on the traffic contour plot function and traffic data engine in HOMES, integrated with historical weather data, the prediction component is able to estimate the daily traffic pattern to accommodate traffic demand growth and road condition change. Data fusion and preparing are performed in this service to clean and fuse the data before store them into the traffic archive. A data processing module, which contains spatial-temporal modeling functions, is used to organize the traffic data, and to represent inherent patterns from the traffic archive.

The real-time traffic data is collected from Virginia Department of Transportation (VDOT) every minute, and is aggregated to every five minutes for analysis. In this implementation, traffic volume is used for prediction, because it is an appropriate measurement for transportation planning and management. Because HOMES monitors the traffic information in northern Virginia area, the historical weather information of Dulles Airport is obtained [9].



a) Thanksgiving Day 07' b) Thanksgiving Day 09'

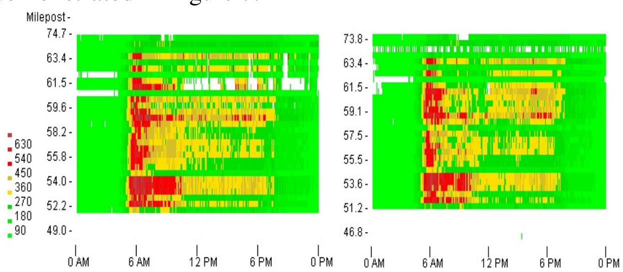
Figure 5. Volume Contour Plots of Thanksgiving Days.



a) Rain 07' b) Rain 09'

Figure 6. Traffic Patterns for Rain 07' vs. Rain 09'.

In this case study, we estimate the daily traffic volume for each link of I-66 East Bound on Thanksgiving Day 2009, when there was fog and rain, by analyzing the historical daily traffic instances. Following the proposed approach, we collect the daily traffic instance of Thanksgiving Day 2007 (rain), weekday daily instances with rain in 2 months before these two Thanksgiving Days, and weekday daily instances with fog and rain in 2 months before Thanksgiving Day 2009. The traffic volume contour plots of the two Thanksgiving Days are illustrated in Figure 5. The traffic patterns generated from getPattern() for raining days before Thanksgiving Day 2007 and before Thanksgiving Day 2009 are shown in Figure 6. The patterns under road condition rain and fog with rain are demonstrated in Figure 7.



a) Rain 09' b) Fog and Rain 09'

Figure 7. Traffic Patterns for Rain 09' vs. Fog and Rain 09'.

The estimated pattern for Thanksgiving Day 2009 was generated by applying the growth factor and condition offset to Thanksgiving Day 2007. Table 2 lists the comparison between the estimated pattern and the real traffic volumes of Thanksgiving Day 2009, as well as the comparisons between the intermediate patterns and the real volumes. As can be observed from the comparisons, about 83% of the estimated traffic volumes were no more than 50 vehicles away from the corresponding real volumes, which is negligible compared to the road capacity. In total about 96% of the estimated volumes were less than 100 vehicles away from the real traffic. The average difference was 28 across all the estimated volumes. On the other hand, either growth factor or condition offset can provide estimation with a certain level of accuracy. Combining these two impacts together, a more accurate daily pattern can be derived.

Table 2. Estimated Patterns vs. Real Thanksgiving Day 09'.

Difference to Real Volumes	**0~50**	**50~100**	**100~150**	**150~200**
w/ Growth Factor	82%	12%	5%	1%
w/ Condition Offset	80%	12%	7%	1%
w/ Growth Factor + Condition Offset	83%	13%	3%	1%

Another case study has been conducted to estimate the traffic for Christmas Day 2009, using Christmas Day 2007 as a base line. Christmas Day 2009 had fog and rain, while Christmas Day 2007 had clear weather. Following the proposed approach, a set of clear days before Christmas Day 2007 and 2009, and the days with fog and rain before Christmas 2009 were analyzed to calculate the corresponding growth factor and condition offset. The comparisons between the real traffic volumes of Christmas Day 2009 and the estimated patterns are shown in Table 3. The final predicted pattern had about 79% of the estimated traffic volumes that were no more than 50 vehicles away from the corresponding real volumes, and 94% no more than 100 vehicles away.

Table 3. Estimated Patterns vs. Real Christmas Day 07'.

Difference to Real Volumes	**0~50**	**50~100**	**100~150**	**150~200**
w/ Growth Factor	77%	16%	6%	1%
w/ Condition Offset	73%	18%	7%	1%
w/ Growth Factor + Condition Offset	79%	15%	5%	1%

6 Conclusion & Future Work

In this paper, we propose an innovative approach to predict traffic patterns for holidays and planned events. It integrates historical traffic instances, weather information, and demand change based on the spatial-temporal data view in HOMES. Computing the offset caused by road condition change and the demand growth, this approach provides accurate estimations using a limited number of historical instances of the target holiday or events. Case studies have been conducted in real system to validate the effectiveness of this approach. Future efforts will be devoted to apply appropriate statistic model for offset computation and pattern generation in this approach. Comprehensive experiments with different types of events (e.g., football games and Inauguration Day) and different traffic measure (e.g., speed, occupancy, or travel time) could be conducted to further evaluate this approach in various application scenarios. This approach can also be integrated with transportation management systems for planning operations such as detour and evacuation for special events.

References

[1] C. T. Lu, A. P. Boedihardjo, J. Dai, and F. Chen, "HOMES: Highway Operation Monitoring and Evaluation System," in *the 16th ACM International Conference on Advances in Geographic Information Systems*, Irvine, CA, 2008, pp. 529-530.

[2] Berkeley, "Performance Measurement System (PeMS)," http://pems.eecs.berkeley.edu, Accessed in Jan., 2010

[3] ODOT, "Portland Oregon Regional Transportation Archive Listing," http://portal.its.pdx.edu/, Accessed in Jan., 2010

[4] U. Washington, "Smart Trek," http://www.its.washington.edu/bbone/, Accessed in Jan, 2010

[5] C. T. Lu, A. P. Boedihardjo, and J. Zheng, "AITVS: Advanced Interactive Traffic Visualization System," in *the 22nd IEEE International Conference on Data Engineering*, Atlanta, GA, 2006, pp. 167-168.

[6] S. Shekhar, C. T. Lu, X. Tan, and S. Chawla, "Map Cube: A Visualization Tool for Spatial Data Warehouses," in *Geographic Data Mining and Knowledge Discovery*, H. J. Miller and J. Han, Eds.: Taylor and Francis, 2001, pp. 74-109.

[7] A.P.Boedihardjo and C. T. Lu, "AOID: Adaptive Online Incident Detection System," in *the 9th IEEE International Conference on Intelligent Transportation Systems*, Toronto, Canada, 2006, pp. 858-863.

[8] Y. Jin, J. Dai, and C. T. Lu, "Spatial-Temporal Data Mining in Traffic Incident Detection," in *SIAM Data Mining, Spatial Data Mining Workshop*, Bethesta, MD, 2006.

[9] WUnderground, "Weather Underground," http://www.wunderground.com/, Accessed in Jan, 2010