

IBM Research Report

Dynamic Update of Public Transport Schedules in Cities Lacking Traffic Instrumentation

Pramod Anantharam, Biplav Srivastava, Raj Gupta

IBM Research Division

IBM Research

New Delhi - 110070. India.

IBM Research Division

Almaden - Austin - Beijing - Delhi - Haifa - T.J. Watson - Tokyo - Zurich

LIMITED DISTRIBUTION NOTICE: This report has been submitted for publication outside of IBM and will probably be copyrighted is accepted for publication. It has been issued as a Research Report for early dissemination of its contents. In view of the transfer of copyright to the outside publisher, its distribution outside of IBM prior to publication should be limited to peer communications and specific requests. After outside publication, requests should be filled only by reprints or legally obtained copies of the article (e.g., payment of royalties). Copies may be requested from IBM T.J. Watson Research Center, Publications, P.O. Box 218, Yorktown Heights, NY 10598 USA (email: reports@us.ibm.com). Some reports are available on the internet at <http://domino.watson.ibm.com/library/CyberDig.nsf/home>

Dynamic Update of Public Transport Schedules in Cities Lacking Traffic Instrumentation

Pramod Anantharam^{*1}, Biplav Srivastava², and Raj Gupta²

¹ Ohio Center of Excellence in Knowledge-enabled Computing (Kno.e.sis),
Wright State University, Dayton, USA

pramod@knoesis.org,

<http://knoesis.org>

² IBM Research, India

{sbiplav, ragupt12}@in.ibm.com

Abstract. A common obstacle for citizens in switching to public transportation is the lack of information about available choices when they need to travel. Although schedules of individual modes like bus or metro may be available as paper pamphlets, or digital files on websites, they do not give an integrated view of the complete services possible when a citizen actually wants to travel. Furthermore, the situation on the roads evolve and this demands timeliness of public transportation information. We want to tackle this problem in the context of cities of developing countries like India, which lacks basic instrumentation to track road conditions or vehicle location.

Our solution extends a public transportation recommender working only with static schedule information to utilize SMS messages about road conditions sent by city authorities. Our solution consists of: (a) extracting events from traffic alert messages, (b) reasoning about traffic delays from extracted events by qualitatively deciding what stops (locations) will be affected, and quantitatively estimating the lower bound on the probability of having a delay at those locations in the city, and (c) utilizing the delay estimates for route recommendation. We use publicly available traffic related SMS messages from Delhi, India, to evaluate our approach and show its promise. Our solution provides dynamic updates for transport network in cities with low investment and quick time to realization.

1 Introduction

Cities around the world want to increase the adoption of public transportation modes such as metros and buses to reduce congestion and consequent pollution. However, a common problem for citizens in adopting public transportation is the lack of information about available choices.

Although schedules of individual modes may be available as paper pamphlets or digital files (e.g. pdf files, web pages), they are usually unavailable when needed, do not give an integrated view of the services possible in the city, and are not amenable to direct analysis. Furthermore, the situation on the roads evolve and this demands timeliness of public transport information. Decision support tools that can help commuters make

* Work done during an internship at IBM Research

journey decisions (journey planners) [1] under such situations play a crucial role in promoting public transportation.

There have been many journey planners proposed in the literature and a few are available operationally. However, they rely on instrumentation on vehicles (e.g., GPS on buses, sensors on smart phones) to provide their recommendations. But, most of the cities around the world do not have such instrumentation in place - the time and investment needed to have sensors is long drawn while their citizens just have low-cost phones. They want journey planners which can work with current data and improve as better sensing becomes available.

Recently, a journey planner was proposed that used readily available schedule data from individual operators and processed them to establish the multi-modal service baseline [2]. However, since it worked only with static data, the recommendations could get stale from day to day or even time to time.

To get dynamic updates, we rely on a trend seen in many parts of the world where cities have an authorized, city-initiated, notification service in place to alert commuters about road conditions. For example, 10³ cities in India including New Delhi [3], provides textual updates through city authorities as Short Message Service (SMS) to the subscribers. The city may themselves get the information from any existing sensor, their field employees, or community; all the information sent is verified. Hence, this becomes a reliable data source whose language and content are consistent. Assuming such a service is in place, we propose a journey planner that can (a) detect and classify events from textual notifications, (b) determine its impact on the multi-modal public transportation network, and (c) use it to rank travel choices least impacted. We develop such a system and show its feasibility in New Delhi, India, where no other similar capability exists. Our contribution include:

- feasibility study of using SMS updates for extracting traffic related events,
- estimation of the traffic disruption caused by event types at various locations in a city,
- estimation of (lower bound on) delays on public route networks based on traffic related events, and
- implementation of a live prototype for providing adaptive route recommendation for the city of Delhi.

In the rest of the paper, we first preview our work and then provide background followed by our approach. Next, we discuss an evaluation on real world data, put our approach in the context of related work and conclude.

2 Preliminaries

This section describes the real world data set used, event representation and its semantics along with some concrete examples.

³ New Delhi (Delhi), Chennai, Bangalore, Pune, Mumbai, Chandigarh, Gurgaon, Nagpur, Kolkatta, Hyderabad.

SIno.	SMS messages
1	Traffic movement is slow from Dhaula kuan to Moolchand due to break down of a DTC low floor bus at the foot of Raj Nagar flyover.msg@9.55am,200712.
2	Traffic is moving in one lane only on Burari road due to MCD work in front of Delhi Jal Board office.msg@10.46am,230612.
3	Due to construction work by DJB at Subhash nagar chowk (Tilak nagar towards Subhash nagar) half of the road is covered by DJB therefore Traffic will remain heavy.msg at 08:56 p.m.,07/07/2012.
4	Traffic is moving slow at G.T.K Road from Mukarba chowk to Azadpur due to work by DJB.Kindly avoid this road.Use Mukundpur to Azadpur road.msg@11.42am,180612.
5	Water logging on Lala Lajpat Rai Marg at North Foot & South foot of Defence Colony flyover, North foot of Moolchand flyover, left slip road from Ring Road to Lala Lajpat Rai Marg, opposite PS Defence colony Ring Road & opposite South Ex-2, Ring Road.msg7.26,130712.
6	The door of one of the compartment full of the stone of Maalgari at going towards Iron Bridge has open and stone have spilled on the road.msg@05.55pm260612

Table 1. A sample of traffic updates sent by Delhi Traffic Police in July 2012.

Get public transit directions

Examples Supports multiple cities

Delhi/AIIMS/Rajiv Chowk Bangalore/Majestic/Marathalli

City:

Search From:

Search To:

Start:

Mode:

City Update:

No. of choices:

No. of hops:

Search for multi-modal commute options between two locations on public network

City updates can be separately analyzed

100 solutions found for source **Moti Bagh** and destination **Dhaulta Kuan**
Please click on the links to explore

Group Hop 0

Solution 1 [# Hops = 0; Modes = Bus ; Update used]

Source: Moti Bagh
Destination : Dhaulta Kuan
Route Id : 711
SMS :
Traffic movement is slow from Dhaulta kuan to Moolchand due to break down of a DTC low floor bus at the foot of Raj Nagar flyover .msg@9.55am, 200712 .

SMS Time: 2012-08-01 09:55:06
Mode : Bus

Solution 2 [# Hops = 0; Modes = Bus ; Update used]

Solution 3 [# Hops = 0; Modes = Bus ; Update used]

Solution and city updates which may cause delay to them

Fig. 1. Snapshot of the journey recommender with dynamic updates.

2.1 Datasets and Notation

New Delhi is the capital city of India as well as part of the bigger provincial state called Delhi. It has many public transportation services including a metro service run by Delhi Metro Rail Corporation (DMRC), a rail service by Indian Railways, bus services by the main public operator Delhi Transport Corporation (DTC) as well as public-private and private bus services by many other organizations. In the absence of any effective multi-modal transportation authority, the schedules of existing services are not available in a uniform format.

We will use the route dataset created by authors in [2] from schedules of bus services by DTC and metro by DMRC. Borrowing GTFS[4] terminology, a public transportation route is expressed as a 5-tuple, $\langle S, R, T, ST, F \rangle$: Stop (S), Route (R), Trip (T), Stop Times (ST), and Frequency (F). Figure 2 shows a sample.

```

Agency
agency_id, agency_name, agency_url,
agency_timezone, agency_phone, agency_lang
DTC, Delhi Transport Corporation, http://www.dtc.com,
GMT + 530, 01123232433, en

Stop
stop_id, stop_code, stop_name
345, , Dhaula Kuan

Route
route_id, route_short_name, route_long_name, route_desc, route_type
R102, 711, 711, 711, 3
R103, Line 1, Line 1, Line 1, 2

Trip
route_id, service_id, trip_id, direction_id
R102, A, 0A_R102, 0
R102, A, 1A_R102, 1

Stop Times
trip_id, arrival_time, departure_time, stop_id, stop_sequence
0A_R102, 0:0:00, 0:1:00, 345, 0
0A_R102, 0:6:00, 0:7:00, 412, 1

Frequency
trip_id, start_time, end_time, headway_secs
0A_R102, 7:25:00, 7:50:00, 5

```

Fig. 2. A sample of multi-modal services made available in New Delhi (India) in GTFS format by authors.

Along with routes, we also need to know geographic details, e.g., of roads, on which transportation services are available. We use OpenStreetMap[5] that provides free geographic data and maps for most of the cities worldwide. We use it to associate transit stops (bus stops, metro stations) with locations on the city map.

Table 1 has SMS messages conveying road conditions, traffic movement, obstructions, and weather from SMSGupShup. A sample route recommendation incorporating dynamic updates is shown in Figure 1.

2.2 Event Representation

Each event e^i is represented using a 6-tuple model $\prec type(e^i_{type}), description(e^i_{description}), location(e^i_{loc_{start}}, e^i_{loc_{end}}, e^i_{loc_{on}}), time(e^i_{time}) \succ$. The event type is an abstraction over collections of events following a common structure extracted from alerts. For example, break down of a Heavy Transport Vehicle (HTV) or break down of a car can be categorized as event type *BreakDown*. Event description can be details of the event (e.g. original message in case of traffic alerts may be event description). Event location has some nuances such as start location, end location, and on location. Event time plays an important role in assessing impact on public transport schedule. From the running example (the first message in Table 1), $\prec e^i_{type} = BreakDown, e^i_{loc_{start}} = Dhaulakuan, e^i_{loc_{end}} = Moolchand, e^i_{loc_{on}} = RajNagarFlyover, e^i_{time} = 20 July, 2012, 9 : 55am \succ$

3 Solution Components

3.1 Event Extraction

We extract events from update messages by specifying common patterns observed in the data we collected for two years for eleven cities in India. The current techniques uses regular expressions which provides a precision of up to 96% and recall of up to 92%. As described earlier, the event tuple has three parts - event type, location and time. Event type denotes the nature and intensity of the incident. Location provides the impact point of the incident on a route network. Location is normally the name of the road, region, etc. SMS alerts and events may not have one to one mapping. Further, each SMS alert may be reporting multiple events or there may be multiple alerts for the same event (e.g. start and end of an event); although, this is not a major issue in Delhi since it is from Delhi traffic police (authoritative and aggregate source of traffic alerts). Our approach can be extended to handle such issues. Time indicates the moment when the incident was reported.

3.2 Reasoning Over Traffic Events

There are two logical steps for finding stops that are impacted by a traffic alert. The first step is to find “what stops are affected?” and the second step is to answer the question “by how much?”

Qualitative reasoning involves what stops are affected by a traffic event. The event locations extracted in the event extraction step $(e^i_{loc_{start}}, e^i_{loc_{end}}, e^i_{loc_{on}})$ are matched with a database of stop names for a city. The stop name matching is not limited to exact name match but we use lat-long information obtained using OSM (Open Street Maps).

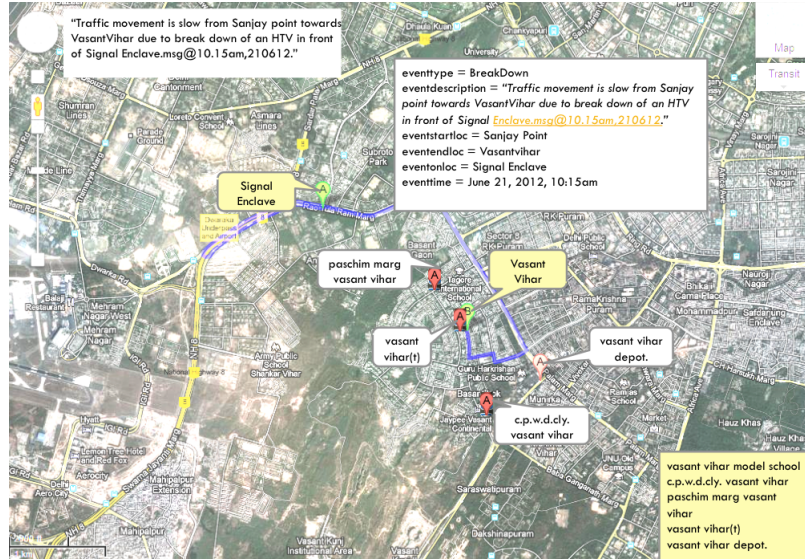


Fig. 3. Illustration of notification, event extraction and its geographical extent.

Quantitative reasoning involves finding the degree of impact on each stop due to a traffic event. We estimate the lower bound on the probability of delay (\hat{P}_{delay}) at a location based on event types reported at that location. This estimation is the lower bound on the delay probability (actual road conditions may be worse leading to higher than estimated delay) due to: (1) incomplete reporting of events, and (2) unknown interactions between different events.

Estimating Delay Probability: To answer queries on delay probabilities, we will consider questions such as: Will event type influence the delay? Will event types uniformly affect delays at different locations? Do locations have prior disposition for event types? We will reconsider these questions in the evaluation section.

An event has certain characteristic properties about how they impact traffic on a road. For example, breakdown of a car on a road may require a call to a mechanic to repair it or tow the car to the workshop. Calling the mechanic and getting the car repaired may lead to a delay of around 4 hours. In this instance, we were able to estimate the delay but it may not be feasible to do such a detailed analysis for all possible events.

This creates the need for a model which can learn the inherent event properties given partial domain knowledge (random variables and connection between them). Our objective is to determine the probability of having a commute delay for a link. For doing this, we need to (a) deal with incomplete, imprecise, and heterogeneous observations, (b) allow specification of domain knowledge for reasoning, (c) use data from the domain to validate and parameterize the domain knowledge, and (d) incorporate historical knowledge in the reasoning process.

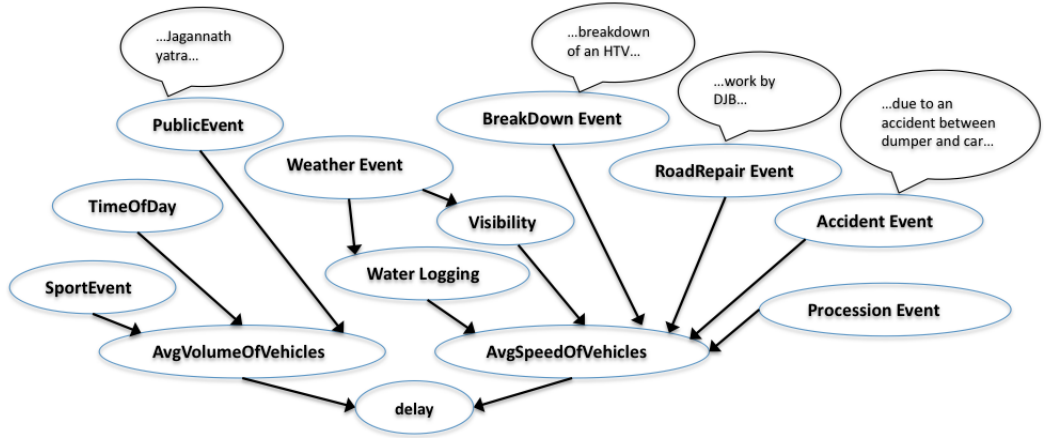


Fig. 4. Bayesian Network for the domain of traffic along with sample instances of event types

We formulate the problem of finding the delay probability associated with a stop (location) using a Bayesian network [6], which combines domain knowledge of traffic and historical observations to reason over current observations of traffic.

Figure 4 is an example Bayesian network for the traffic domain along with some SMS observations from SMSGupShup (an SMS subscription service for traffic alerts). Each node in the network represents a random variable related to the traffic domain and each link imposes a structure among these random variables. All the random variables ending with “event” are the event types we have considered in the analysis.

We have sparse observations about the events in the form of messages, therefore the predicted probability of delays may not be accurate. Due to sparse observations, the priors plays an important role in determining the accuracy of the reasoning process till we accumulate more observations. One way of initializing priors is by using historical traffic related events across different locations in a city. For instance, if the traffic accidents are found in the cities 20% of the time, then we say $P(\text{Accident}) = 20\%$. The priors are initialized and personalized for each location in a city. We initialize the traffic event priors based on traffic alerts collected for two years from Delhi.

Assumption: We assume that each event has an independent effect on the delay. For instance, the event type *Accident* and *Procession* will have an independent effect on the delay. This is a reasonable assumption partly due to the unavailability of data. However, given data about multiple events and delays, we can still use the same framework to capture the collective effect of events on the delay probability estimation.

To assess the probability of having a commute delay at a stop, we associate a random variable $delay_{s_i}$ with all the stops in a route network represented by a set of delays, $D = \{delay_{s_i}, \forall s_i \in S\}$. $delay_{s_i}$ at stop s_i is influenced by various traffic events at node s_i , $E_{s_i} = \{e^1, e^2, \dots, e^n\}$ where e^j is an event type. The probability of having a delay at a stop depends on the (1) events at the stop, and (2) prior probability of having a delay at the stop. The probability of having a delay at a stop given events observed at the

stop in terms of the likelihood, and prior is given by $P(\text{delay}_{s_i} | E_{s_i}) = P(E_{s_i} | \text{delay}_{s_i}) P(\text{delay}_{s_i}) / P(E_{s_i})$.

We illustrate parameterization of the Bayesian network for a place named Dhaula Kuan for clarity. We need to estimate the lower bound on the probability of having a delay at Dhaula Kuan. We noticed that accidents and breakdowns are the events that have major influence on delay at Dhaula Kuan using the traffic alerts for Delhi collected over two years. The conditional probability table is constructed using these observations as shown in Table 2.

3.3 Computation of Priors for Events

From two years of traffic alerts collected at various locations in Delhi, we compute the prior probability of traffic related events. For instance, there may be many reasons for delays at a node. These events occur at different probabilities at different nodes. One node may be prone to a particular type of event compared to other nodes. We observed that (a) breakdowns caused many delays compared to any other events which is orthogonal compared to developed countries, (b) each location has affinity for certain types of events, and (c) not all locations will have events.

3.4 Propagate impact

The delay impact at a node will affect delay at its neighboring node since the nodes are interconnected to each other by links. For computational efficiency we limit the impact propagation from a link to the previous link only i.e., if traveling from s_1 to s_2 , the delay at s_1 is influenced only by the delay at s_2 . This is a realistic assumption in the traffic domain. In other words, we assume that *delay* at a node s_i depends on node s_{i+1} and not on any other nodes in the route network.

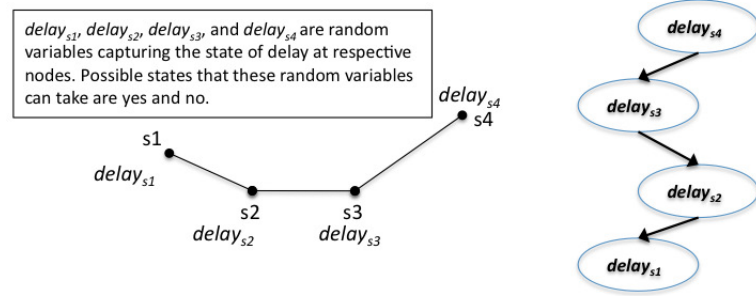


Fig. 5. Impact propagation in terms of delays

Figure 5 shows an example road network and the Bayesian Network modeling the impact propagation in terms of the delays at each node. As it is evident from the network, the random variable delay_{s_1} depends only on delay_{s_2} and is independent of all the other nodes in the network given delay_{s_2} .

3.5 Illustration of Impact Assessment

We will assess the impact of the first message in Table 1 and compute the probability of delay.

Qualify impact In Figure 3, a complete example from SMS notification to event extraction is shown, followed by automatically mapping its locations to their closest stop names for transport network in GTFS. In the running example, the locations are: $e_{loc_{start}}^i = Dhaula\ kuan$, $e_{loc_{end}}^i = Moolchand$, $e_{loc_{on}}^i = Raj\ Nagar\ flyover$.

Accident	BreakDown	delay
0	0	0
1	0	1
0	1	1
0	1	0
1	1	1
0	1	1
0	0	0

	Accident	BreakDown		delay	
		Yes	No	Yes	No
delay	Yes	1	1	0.667	0
	No	0	0	0.333	1

Table 2. Top and bottom table shows the Traffic Observations and the Conditional Probability Table (CPT) respectively for the Bayesian Network constructed for Dhaula Kuan location which is represented as a node in our route network.

Quantify impact The impact assessment of an event from textual observation on the schedule is done using the parameterized background knowledge. We will compute the probability of delay at Dhaula kuan given the example message.

$$P(\text{delay} \mid \text{BreakDown} = \text{"Yes"}, \text{Accident} = \text{"No"}) = \frac{P(\text{BreakDown} = \text{"Yes"}, \text{Accident} = \text{"No"} \mid \text{delay})P(\text{delay})}{P(\text{BreakDown} = \text{"Yes"}, \text{Accident} = \text{"No"})}$$

From Table 2 we consider observations for which we have a breakdown and no accident for computing all the probabilities. There are four instances for which there is a delay, out of which, there are two instances with a breakdown but no accident. Hence, the first probability in the numerator is (1/2). There are seven total instances of observations out of which four of them have delays. Hence the second probability in the numerator is (4/7). Out of seven total instances, there are three instances for which there is a breakdown but no Accident. Hence the denominator in the above equation is (3/7). $P(\text{delay} \mid \text{BreakDown} = \text{"Yes"}, \text{Accident} = \text{"No"}) = (1/2) (4/7)/(3/7) = 0.667$. The same process is used to build the CPT shown in Table 2. This shows that given a breakdown event at *Dhaulta Kuan*, we are 66.7% confident that there will be a traffic delay. We acknowledge that the illustration above did not fully utilize the power of Bayesian network. The proposed approach can be easily extended for adding multiple modalities of observations as they become available.

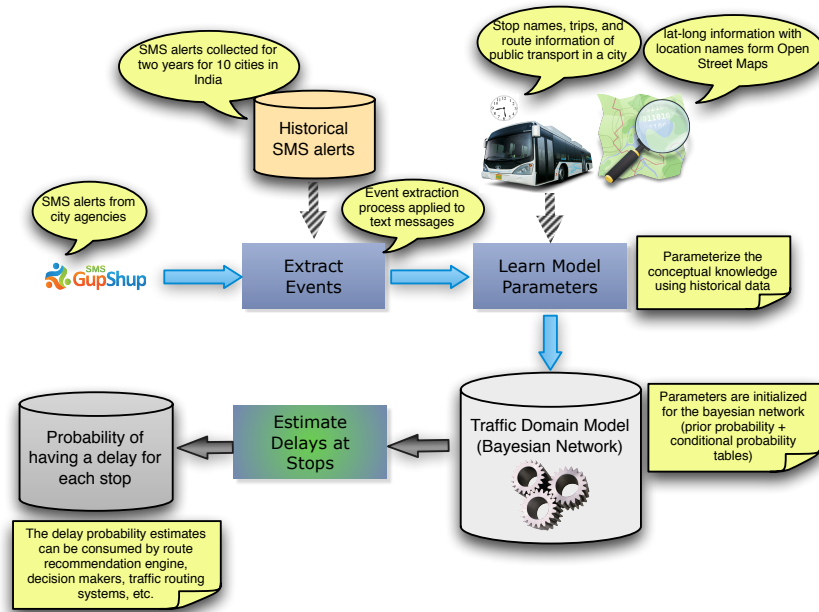


Fig. 6. System architecture with event extraction and reasoning components

4 System Architecture

Figure 6 depicts the overall system architecture of the multimodal dynamic update system. We use the SMSGupShup API⁴ to get all the SMS alerts on an hourly basis (setup as a cron job). For each message, we extract the event related metadata - location (from, to, on), time, and event (type). Event metadata such as location name and time are used to find the affected stops using stop names and trip information. Trip information is time stamped route for a public transport vehicle and it has a direction. Event extraction is run over historical traffic alerts for two years, and for ten cities in India. The event related priors are computed for all the cities and the city (Delhi) specific priors are computed in the learn parameters step. Once the Bayesian network is parameterized, it can be used to reason over new traffic related observations for estimating the probability of having a delay at a node.

5 Evaluation

There has been an increasing body of research on predicting traffic delays and all of them are evaluated in a highly instrumented setting [7–9]. Most of them utilize fine grained sensor observations (usually minute by minute update) to build models and

⁴ <http://api.smsgupshup.com/api>

predict delays. Developing countries like India have far less investment on such an infrastructure and evaluation of traffic delay prediction system is not trivial. The situation is further challenging since there are no formal report of traffic related incidents in the city of Delhi. There has been some recent efforts by citizens in sharing traffic related incidents [10]. We validate our delay predictions by comparing them to a authoritative source of traffic related knowledge [10]. We describe the characteristics of the SMS updates from SMSGupShup and present our evaluation on the data collected for two years for the city of Delhi.

5.1 Dataset

Delhi Traffic Police (DTP) deploys units on roads of Delhi for monitoring traffic conditions. These units report traffic incidents as they occur (and confirmed) in the city. Thus, the number of alerts sent out on a particular day depends on the number of incidents in the city. We collected around 9,000 SMS alerts (on which we evaluate our approach) for the city of Delhi for two years which is on an average, 12 alerts a day. We found that the average alerts per day is not necessarily how it is accumulated in practice, e.g. on a rainy day in Delhi, a location may have more than 12 alerts while no alerts at all on other days.

5.2 Reasoning Over Traffic Events

The evaluation presented in Table 3 answers some of the questions posed regarding events in the quantitative reasoning section. Police and long time residents share prominent traffic jam hot spots as: Dhaula Kuan, AIIMS, Maharani Bagh, Badarpur, ITO and Pragati Maidan [11]. The results are in conformance with these observations. Each location may have multiple independent events and each event may have independent probabilities with which they occur at different locations. Given a location, the delay probability varies with event types e.g. at Lajpat Nagar, BreakDown event causes greater delays compared to Accident. BreakDown affects delays at Lajpat Nagar and Dhaula Kuan with different intensities which can be explained by probabilities $P(\text{Weather}_{Dhau\text{la}Kuan} | \text{delay}) = 0.0588$ and $P(\text{Weather}_{LajpatNagar} | \text{delay}) = 0.2352$. This indicates that given evidence of a delay, the cause being bad weather is 23% at Lajpat Nagar. Similarly, given delay as evidence, bad weather being the cause is 5% at Dhaula Kuan.

6 Discussion and Related Work

6.1 Discussion

SMS as a mode of traffic alerts The SMS alerts are a feasible way of reporting about traffic conditions. In fact, it is known to be the preferred way of conveying real time road traffic conditions [12]. Many cities have realized this and they provide traffic updates and road conditions through SMS alerts. With the data collected for two years

Location	$\hat{P}(\text{delay} \mid \text{BreakDown})$	$\hat{P}(\text{delay} \mid \text{Weather})$	$\hat{P}(\text{delay} \mid \text{Procession})$	$\hat{P}(\text{delay} \mid \text{RepairWork})$	$\hat{P}(\text{delay} \mid \text{Accident})$
Lajpat Nagar	0.0199	0.016	0.0039	0.0017	0.007
Dhaura Kuan	0.0774	0.028	0.0748	0.0160	0.031
Munirka	0.006	0.008	0.0118	0.0035	0.0077
Ashram	0.0553	0.032	0.0039	0.0035	0.0233
ITO	0.0110	0.012	0.0039	0.0071	0.0038
Chirag Delhi	0.0243	0.032	0.0196	0.0089	0.0077
Sansad Marg	0.0066	0.012	0.0275	0.0053	0.0116
Moti Bagh	0.0221	0.012	0.0118	0.0142	0.0116
Noida	0.0309	0.018	0.0177	0.0053	0.0038
Faridabad	0.0044	0.004	0.0039	0.0017	0.0038

Table 3. Delay probability estimation for ten locations in Delhi computed using the priors from the traffic alerts from ten cities.

and the reasoning carried out with new incoming traffic alerts, we showed the feasibility of computing delay probabilities. The delay probability can be consumed by city administration for better planning and allocation of resources.

The event extraction from SMS alerts needs a careful consideration since it directly influences the inference process. Event extraction is used for processing both historical and current observations. Using recurring patterns for determining location and event types resulted in good results.

Sample space for computing probabilities Probabilities presented in Table 3 were computed considering each day as one unit in the sample space. For example, if there were accidents on 4 days out of 30 days at a location, then we consider 30 days as our sample space. This is reasonable in the domain of traffic, since we can accumulate knowledge of the prior probability of accidents at this location over time. We found that some locations are prone to particular types of events more than others thereby influencing the prior probabilities of events associated with it.

Relative ordering of delay probabilities for decision making The relative ordering of delay probabilities is important rather than the absolute value of the delay probability for decision makers. For instance, given that there is going to be a heavy rain in Delhi, which location is affected the most? The location named *Chirag Delhi* is the worst affected location by weather related events. Similarly, given that there is a procession in the city of Delhi, *Sansad Marg* is the worst affected location followed by *Munirka* (from Table 3). Procession, repair work, accidents, and break downs affect *Dhaura Kuan* the most, and out of all the locations presented here, it is the location worst hit by traffic delays. This information is valuable for the deployment of resources such as traffic personnel, city cops, traffic diversion, public transport schedule updates, etc.

Consumption of delay probabilities The route recommendation system IRLTransit uses the delay probabilities in ordering the route recommended to the commuters. This

ordering is based on the relative ordering of delays. For those nodes without any probability of delay estimates due to missing or absence of events at the location, we take the delay probability estimate as zero. Figure 1 shows the route recommendation by considering the dynamic updates on traffic events.

Data sparsity We could estimate the lower bound on the probability of delays given an event at a location. The interaction between events, i.e., two or more events co-located in space and time is unknown from the given observations. This will be useful in estimating delay probabilities in the cities having multiple events at a single location.

6.2 Related Work

There is a rich literature for journey planning, transit analysis and traffic events. Journey planners have started becoming available for prominent cities, e.g., San Francisco’s 511 service[13]. A state-of-the-art algorithm for it is described in [14]. However, such systems assume that real-time updates will be available from vehicles. Google Transit provides a service for travelling in New Delhi but it is not adaptive and the details of how it gets data is not known since none of the city agencies provide their data in its format.

There are studies on public transit dealing with congestion and its impact on vehicle arrival[15]. But such work do not consider the impact of traffic events on bus delays. Considering literature which has dealt with traffic events in a city, [16] studies the impact of an event on traffic patterns using simulation. The impact analysis is done over streets using cluster analysis but does not extend to public transportation. At a general level, [17] looks at traffic modeling and prediction using heterogeneous information sources but assumes a data-rich setting. Our approach is novel in building an adaptive journey planner leveraging multi-modal textual updates (observations) for estimating delay probability and its impact on public transport schedule. We exploit prior knowledge of the domain to deal with data sparsity issues in cities with no instrumentation, i.e., no machine sensors to monitor the physical world.

7 Conclusion

We proposed a solution framework for processing dynamic traffic updates and estimation of the delay probability for each location. We addressed many challenges such as absence of machine sensors for monitoring traffic conditions, extraction of location and event observations from unstructured text, representation and parameterization of a Bayesian network using prior knowledge, dealt with sparse observations, and defined impact propagation. We evaluated our approach using real world dynamic updates on traffic events to recommend solutions leading to timely schedule information of public transport vehicles.

As a future work, one can extend the processing of dynamic updates to social streams such as Twitter. The probability of delays can be extended to have index over time which may be useful for forecasting traffic. The domain knowledge of traffic is assumed to be given and only the parameters of the Bayesian network are learned from

data. We would like to extend our system to extract this background knowledge by processing multi-modal sensor observations. This extension will be useful in domains where there may be missing prior knowledge of random variables and connections between them.

Acknowledgment

The authors would like to thank Srikanth Tamilselvam for his help in implementing the system.

References

1. Esztergár-Kiss, D., Csiszár, C.: Analysis of multimodal journey planners using a multi-criteria evaluation method. (October 2012)
2. Gupta, R., Srivastava, B., Tamilselvam, S.: Making public transportation schedule information consumable for improved decision making. In: Proc. IEEE ITSC. (2012)
3. DTP: Delhi traffic police updates. In: <http://gupshup.me/groups/DTP?page=39>. (2012)
4. Google: General transit feed specification. at http://code.google.com/transit/spec/transit_feed_specification.html, 2012.
5. OSM: Open street map. In: www.openstreetmap.org, en.wikipedia.org/wiki/OpenStreetMap, 2012.
6. Koller, D., Friedman, N.: Probabilistic Graphical Models: Principles and Techniques. MIT Press (2009)
7. Tang, L.A., Yu, X., Kim, S., Han, J., Peng, W.C., Sun, Y., Gonzalez, H., Seith, S.: Multidimensional analysis of atypical events in cyber-physical data. In: Data Engineering (ICDE), 2012 IEEE 28th International Conference on, IEEE (2012) 1025–1036
8. Pan, B., Demiryurek, U., Shahabi, C.: Utilizing real-world transportation data for accurate traffic prediction. In: Data Mining (ICDM), 2012 IEEE 12th International Conference on, IEEE (2012) 595–604
9. Chawla, S., Zheng, Y., Hu, J.: Inferring the root cause in road traffic anomalies. In: Data Mining (ICDM), 2012 IEEE 12th International Conference on, IEEE (2012) 141–150
10. : Traffic fatalities in delhi. http://www.accidentin.com/world/india/fatal_accidents_in_delhi.htm (May 30, 2013)
11. : Delhi - city of organised traffic jams. <http://blogs.timesofindia.indiatimes.com/Diary-of-a-Divorced-Delhi-Male/entry/delhi-city-of-organic-traffic-jams> (Accessed on May 30, 2013)
12. Caulfield, B., O'Mahony, M.M.: Factors influencing preferences for real-time public transport information. In: European Transport Conference. (2007) 1–11
13. 511: San francisco 511 service. In: At <http://511.org>, 2012.
14. Barth, M., Boriboonsomsin, K.: Algorithm for finding optimal paths in a public transit network with real-time data. In: Jour. TRB, No. 2256, TRB. (2011)
15. Chen, M., Liu, X., Xia, J.: Dynamic prediction method with schedule recovery impact for bus arrival time. In: Transportation Research Record: Jour. TRB. (2005)
16. Huang, B., Zhao, T., Zhang, Y.: Traffic incident impact analysis with random matrix theory and cluster analysis. In: IEEE Conf. EMMS. (2010)
17. Horvitz, E., Apacible, J., Sarin, R., Liao, L.: Prediction, expectation, and surprise: Methods, designs, and study of a deployed traffic forecasting service. In: In Twenty-First Conference on Uncertainty in Artificial Intelligence. (2005)