City Notifications as a Data Source for Traffic Management

Pramod Anantharam[1]
Kno.e.sis, Wright State University,
Dayton, OH.
Email: pramod@knoesis.org

Biplav Srivastava
IBM Research - India
Email: sbiplav@in.ibm.com

ABSTRACT

A common problem for cities of developing countries like India in managing traffic is the lack of basic automated instrumentation to track road conditions or vehicle locations. Still, to help their citizens make informed travel decisions based on changing city dynamics; many cities have an authorized, city-initiated, notification service in place to alert subscribing commuters about road conditions. Here, alternative means may be used to create informal textual notifications – e.g., inputs from field personnel, citizen updates, and pre-authorized events from city calendar. In this paper, we show that collections of such notifications, when processed with information extraction techniques, can turn them into a rich source of data for traffic managers. Specifically, we use Short Message Service (SMS) notifications from the city of Delhi, India to show promising insights.

Keywords: Traffic Data, Developing Countries, City Notifications, Information Extraction

1. INTRODUCTION

Intelligent Transportation Systems envisages applying information and communication technologies (ICT) to efficiently manage the traffic moving in a city using its transportation infrastructure. All such systems rely critically on foundational methods to collect and manage traffic data and lead to applications that make commuters travel efficiently (e.g., enable real-time travel routing, adopt public transportation) or streamline city infrastructure (e.g., region based charging, enable safety measures like road speed enforcement, variable speed limits, and collision and congestion avoidance systems).

A common problem for cities of developing countries like India in managing traffic is that they lack basic automated instrumentation to track road conditions or vehicle locations. There is a slew of techniques available varying in accuracy, coverage and cost to install and maintain. Further, these methods are set up in a diverse setting and are complementary to each other. Traffic measurement/data acquisition has received much attention with technologies like inductive-loop since 1960s, video image analysis since 1970s, floating car data since 1990s, data mining on telecommunication data since 2000s, and the currently popular GPS-based devices. However, getting a traffic-sensing infrastructure in place is a time

---

[1] The work was done while doing internship at IBM Research – India.

consuming and expensive process while managing daily traffic is an immediate necessity.

In this context, we propose a promising new traffic data source based on textual notifications – a trend seen in many parts of the world where a city has an authorized, city-initiated, notification service in place to alert subscribing commuters about road conditions. For example, in many cities[2] in India including New Delhi [5], city authorities send textual updates as SMS to registered mobile phone users. The city may themselves get the information from their field employees, community updates, pre-authorized events from city calendar or any existing sensors; but only verified information is sent out as alerts. Hence, this becomes a reliable data source whose language and content are coherent.

Focusing on Short Message Service (SMS) notifications for the city of Delhi, we study the feasibility of using them, when processed with information extraction techniques, as source of data for traffic managers. Specifically, we automatically extract *events*, a rich representation of what is happening, where and when. We analyze group of events for insights about road conditions in a city, as well as across cities. Some examples of insights are: (a) which places are prone to events, how often and for how long, (b) for different event types, what places are most likely to be affected, (c) how does one city compare with another in terms of different events.

In the rest of the paper, we first describe the datasets and then present our approach to extract events. Next, we present the system architecture and provide empirical validation on the feasibility of our approach. We finally conclude with a discussion of generated insights and their applications.

2 CITY NOTIFICATIONS AND SMS
We start by describing SMS and the datasets used in the paper.

2.1 Why focus on SMS
SMS updates are the best way to convey real time traffic to the users [1] due to its prevalence and availability. In developing economies like India, SMS is prevalent from business transactions to personal message exchanges. In India, many cities provide traffic notifications to citizens using SMS. Most of the low-cost phones, which constitute 95% of all phones (unlike smart phones which are only 5%), have SMS services resulting in a vast coverage. SMS updates from authoritative sources are significantly better in quality than social media status updates like tweets.

---

[2] Cities from where we have data are: New Delhi (Delhi), Chennai, Bangalore, Pune, Mumbai, Chandigarh, Gurgaon, Nagpur, Kolkatta, Hyderabad.

2.2 Nature of SMS Updates

Delhi Traffic Police has been sending alerts in the SMS format to its subscribing citizens about road conditions for over 2 years. This has proven very popular since citizens have no other source of traffic information or sensors to rely on. Table 1 shows some sample SMS messages reporting traffic incidents. Such updates are now being given in eleven other cities in India. In Figure 1, we show the 2269 messages distributed over different locations in Delhi for two years (Aug 2010 to Aug 2012). We notice that alerts are highly localized with some places receiving more alerts than others.

Figure 1. Number of traffic updates received for different locations in Delhi during two years (August 2010 – August 2012)

3. SOLUTION APPROACH

We present our approach by describing the model of *events* we adopt - a rich representation of what is happening, where, and when. Then we describe the extraction steps and present the system architecture.

3.1 Event Representation

Each event e^i is represented using a 6-tuple model: $\langle$type (e^i_{type}), description ($e^i_{description}$),

location ($e^i_{locstart}$, e^i_{locend}, e^i_{locon}), time (e^i_{time})⟩. The event type is an abstraction over the events extracted from the alerts (e.g. break down of a Heavy Transport Vehicle (HTV) or break down of a car can be categorized as event type BreakDown). Event description holds 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. For the first SMS from Table 1, the event 6-tuple looks like: ⟨e^i_{type} = BreakDown, $e^i_{locstart}$ = Dhaula Kuan, e^i_{locend} = Moolchand, e^i_{locon} = Raj Nagar Flyover, e^i_{time} = 20 July, 2012, 9:55⟩.

Table 1. Sample SMS Messages from SMSGupShup for the City of Delhi

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 coverd 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

3.2 Event Extraction

We extract events from update messages using common patterns observed in the data collected for two years as shown in Figure 2. As described earlier, the event 6-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, landmark, 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 the messages are from Delhi traffic police (authoritative and aggregate source of traffic alerts). Our approach can be extended to handle such issues.
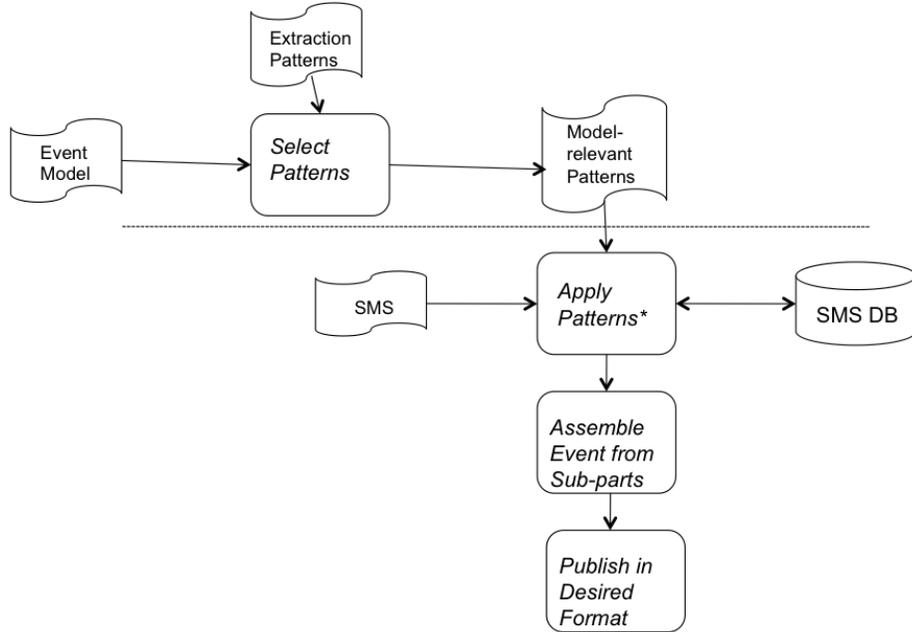
Figure 2. Event processing workflow for processing SMS updates

In the current implementation, we use regular expressions to express patterns for content extraction. Some sample patterns are shown in Table 2.

Table 2. Patterns Used for Content Extraction

Extracted object	Pattern start marker	Pattern stop marker
$e^i_{loc_{start}}$	from	to
$e^i_{loc_{end}}$	to towards	due End Of Line (EOL)
$e^i_{loc_{on}}$	at at at at	due from to near
e^i_{time}	at	,
e^i_{type}	due to due to	at EOL

3.3 Location Extraction

For deriving GIS co-ordinates for each stop, we map the stop names (of public transport vehicles) to OSM location names. This matching is done using Jaro, Levenshtein, MongeElkan, and Needleman-Wunch similarity [2]. A voting is done to pick a confident match (>2 matching algorithms), possible match (==2), and imprecise match (<2). After we enrich the stop names with GIS co-ordinates, we can make semantic queries such as stops "near" a given stop.

5

The locations mentioned in the SMS are processed to get their GIS coordinates. The location GIS coordinates are matched against that of the GIS coordinates of stops to detect those stops that may be potentially affected by an alert.

We mine the alerts to uncover the nature of events and their spatial extent. A domain expert can fix the temporal extent of different events. A sample of extracted events and its components is shown in Figure 3. They are very useful in estimating the nature of delay at different city locations.

Figure 3. Snapshot of event and locations extracted from SMS alerts

3.4 Computation of Priors

We can use the event information for initializing prior probability of events related to traffic at different locations across cities. For instance, if the traffic accidents are found in the cities 20% of the time, then we say P(Accident) = 0.2. After a careful consideration we found that such a general statistics may hold in some domains such as medicine where the disease information is published and the chances of finding a disease in a population can be computed. However, traffic is a very dynamic domain and we need priors to be initialized and personalized to each city. We initialize the traffic event priors based on traffic alerts collected for two years from Delhi. Decision makers can use this prior probability of events for deploying resources for better traffic management.

6

3.5 System Architecture

Figure 4 depicts the overall system architecture of the dynamic update processing system. We use the SMSGupShup [4] to get all the SMS alerts on an hourly basis. For each message we extract the event related metadata - location (from, to, on), time, and event (type). The historical events may be used to estimate the prior probability of different event types at different locations in a city, referred to as model parameters. The GIS coordinates are used to connect stops and locations mentioned in the SMS.
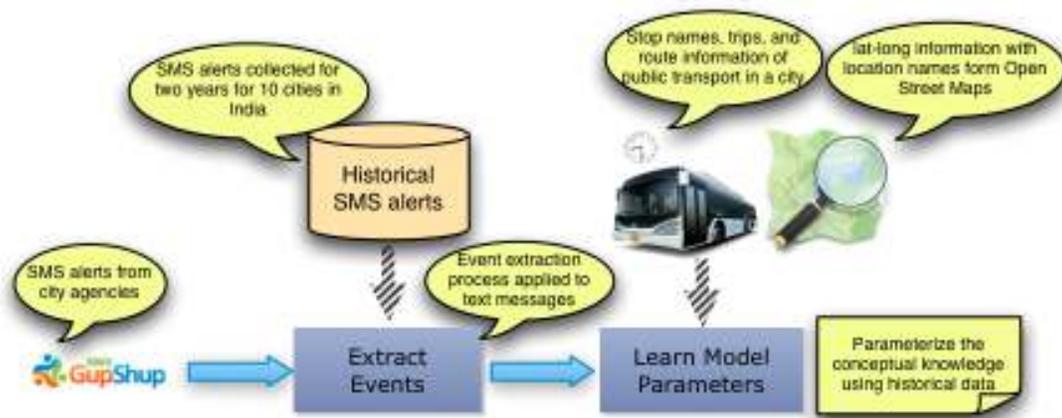

Figure 4. Snapshot of event and locations extracted from SMS alerts

4. EVALUATION

We describe the characteristics of the SMS updates from SMSGupShup and present our evaluation on the data collected for two years in the city of Delhi.

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 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 on clear days.

We want to evaluate the effectiveness of our content extraction of events. Further, we want to investigate if events as a collection can be used to build an understanding of the city's transportation network. We can further correlate the locations with information from public maps (e.g., bus stops, landmarks) to build precise location of events.

4.1 Event Extraction

Traffic related event extraction process is evaluated based on precision and recall measures for location and event type as shown in Table 2. Precision for location extraction is the ratio of correctly identified location (start, end, and on) to the total number of extracted locations. We also present the scores for at least one location being identified from the SMS alert. Recall is the ratio of total number of extracted locations to the actual number of locations in the messages. The definition of precision and recall applies to all location type ($e^i_{locstart}$, e^i_{locend}, e^i_{locon}, $e^i_{locatleastone}$) entries in Table 2. The event mentioned in the messages, follow a structure and remain predictable resulting in high precision and recall. The example message (first message from Table 1) has *break down of a DTC low floor bus* as the event which belongs to a general event category of e^i_{type} = *BreakDown*.

Thus, we can indeed detect events accurately from dynamic updates. As a future work, for improving the extraction process, Named Entity Recognition (NER) [3] can be used to complement the location extraction process.

Table 2. Evaluation of Event Extraction

Location type	Precision	Recall	F-Measure
$e^i_{loc_{start}}$	96.00	86.20	85.60
$e^i_{loc_{end}}$	78.26	79.31	78.78
$e^i_{loc_{on}}$	66.66	13.63	22.63
$e^i_{loc_{atleast_one}}$	83.33	68.18	74.99
e^i_{type}	100	92.30	95.99

4.2 Location Mapping

We now correlate the locations with information from public maps (e.g., bus stops, landmarks) to build precise location of events. We use OpenStreetMap[7] that provides free geographic data and maps for most cities worldwide.

Figure 5 shows a sample mapping between stop names and OpenStreetMap (OSM) locations. The second column is the stop name and the fourth column is the OSM location name to which the stop name is matched. The last two columns are latitude and longitude of the OSM location. The idea of mapping stop names to location names on OSM is to find other stops near this stop using lat-long information. We considered only confident matches and could find a mapping for 1496 stops (to OSM locations) out of 3931 total stops in Delhi.

1695	rithala vill	113	Rithala	28.721	77.107
3839	pratap nager	114	Pratap Nagar	28.667	77.199
3848	pulbangash	115	Pulbangash	28.666	77.207
3803	guru dronacharya	127	Guru Dronacharya	28.482	77.102
3308	najaf garh	131	Arjan Garh	28.481	77.126
1765	petrol pump	147	HP Petrol Pump	28.561	77.222
578	hp depot	147	HP Petrol Pump	28.561	77.222
2198	indian oil	149	Indian Oil Petrol Pump	28.568	77.254
1142	indian oil qtr	149	Indian Oil Petrol Pump	28.568	77.254
1901	talkatora stadium	204	Talkatora Stadium	28.623	77.193

Figure 5. Sample mapping of stop names to OSM names to derive lat-long information

5. DISCUSSION

We have shown that city notifications can be a feasible source for traffic data in developing countries for gaining insights into traffic related events. The event model and extraction technique to populate the model from SMS alerts resulted in promising results for a relatively simple extraction technique. The extracted events provided insights. Some examples are:

(a) Which places are prone to events, how often and for how long? This is feasible with statistical analysis (mining) of historical events.

(b) For different event types, what places are most likely to be affected? This is feasible with statistical analysis (mining) of historical events, and ranking the results based on event types.

(c) How does one city compare with another in terms of different events? This is feasible by comparing historical events from different cities.

However, while the insights are useful, one needs to be careful in interpreting the results since what we have is a sample of the events, as recorded by authorized notifications from the city. There could be more incidents on the road which may go unreported. However, the surety we have from city notifications is that the content is authentic and its quality (in terms of accuracy and grammar) is high.

As one future work, one could extend the approach to consider updates given by citizens on social media sites like Facebook[3] and Twitter[4], as well as correlate with them. Indeed, some of the cities already have official sites on them. Unfortunately, they come with their own sets of challenges. Facebook is primarily a human interaction medium and there are severe restrictions on how data could be exported out for analysis. On Twitter, it is hard to check whether the entity posting the update is authentic and furthermore, the content is of poor

---

[3] http://facebook.com

[4] http://twitter.com

grammatical quality.

In terms of users of extracted events, we have focused on traffic managers. As another future work, one can use the extracted events to estimate the impact of traffic related events on the schedule of public transport vehicles. This can be achieved by estimating the probability of delays using the events form SMS updates and correlating them with existing schedules.

Hence, we argue that city notifications are a promising source for traffic data.

6. CONCLUSION

We proposed a solution framework for processing dynamic updates of traffic related events and assessing their impact on the schedule of public transport vehicles. We evaluated our framework on dynamic updates on traffic from reliable sources. We addressed many challenges such as absence of machine sensors for monitoring traffic conditions, extraction of location, and event observations from unstructured text. We also demonstrated useful insights possible from the events and identified some avenues for future work.

7. ACKNOWLEDGMENT

8. REFERENCES

[1] Caulfield, B. and O'Mahony, M, Factors influencing preferences for real-time public transport information: [in] proceedings of the European Transport Conference 2007, Leeuwenhorst, Netherlands, October 17th-19th 2007, pp[1-11].

[2] "String similarity metrics". Available at http://en.wikipedia.org/wiki/String_metric (Jan 10, 2013)

[3] Xiao Ling and Daniel S. Weld, "Fine-grained entity recognition," in Proc. AAAI, 2012.

[4] "SMSGupShup API". Available at http://api.smsgupshup.com/api (Jan 10, 2013)

[5] "Delhi traffic police updates," Available at http://gupshup.me/groups/DTP?page=39 , (Jan 10, 2013)

[6] John P., Korattyswaroopam N., and Ittyerah N. "The crisis of public transport in India: overwhelming needs but limited resources." Journal of Public Transportation 7 (2004): 95-113.

[7] Open Street Map, At http://www.openstreetmap.org