USE OF TRAFFIC AND CITIZEN TWEETS FOR INCIDENT MANAGEMENT FOR DISTRICT DEPARTMENT OF TRANSPORTATION

Rakesh Nune
Transportation systems Engineer
District Department of Transportation
Washington DC
Rakesh.Nune@dc.gov
Outline

• Introduction
• Methodology
• Extraction
• Ranking and selection
• Storing and accessing
• Display Results
• Validation
• Future Work
Introduction

• Social media has grown exponentially

• Information becomes valuable for TMC operators as more people and news organizations tweet instantaneously about the traffic incidents.
4 car accident, 2 taxis, 2 sedans on E Street ramp coming off bridge. R lane under overpass blocked before E street/fwy split.

Washington, DC

4:13 AM - 30 Apr 2014
Methodology

• The sheer volume of the information can be overwhelming

• Extracting, storing and sorting the right tweet is a challenge.

• Approaches were taken to extract tweet ranging from news organizations to normal citizens tweeting about incident.
Extraction

• A set of keywords were extracted from historical traffic tweets posted by influential users.

• Keywords (query) are sent to a twitter data crawler to extract the tweets limited in Washington DC area.

• The extracted tweets are scored and ranked based on the keywords list.
Extraction

• Four “Trusted Accounts” were selected in our case WTOP Traffic, drgridlock, Dcpolicedep
• Term frequency–inverse document frequency aka tf-idf was used to determine each word importance

\[ tf(t, d) = \frac{f(t, d)}{\max\{f(w, d): w \in d\}} \]
<table>
<thead>
<tr>
<th>#dctrffic</th>
<th>#mdtraffic</th>
<th>#vattrffic</th>
<th>crash</th>
<th>due</th>
</tr>
</thead>
<tbody>
<tr>
<td>delays</td>
<td>st</td>
<td>traffic</td>
<td>right</td>
<td>ave</td>
</tr>
<tr>
<td>lane</td>
<td>lanes</td>
<td>rd</td>
<td>nw</td>
<td>accident</td>
</tr>
<tr>
<td>buses</td>
<td>left</td>
<td>md</td>
<td>nb</td>
<td>bridge</td>
</tr>
<tr>
<td>sb</td>
<td>earlier</td>
<td>street</td>
<td>near</td>
<td>blocked</td>
</tr>
<tr>
<td>loop</td>
<td>congestion</td>
<td>expect</td>
<td>va</td>
<td>dc</td>
</tr>
<tr>
<td>update</td>
<td>road</td>
<td>work</td>
<td>following</td>
<td>closed</td>
</tr>
<tr>
<td>open</td>
<td>vehicle</td>
<td>inner</td>
<td>car</td>
<td>killed</td>
</tr>
<tr>
<td>new</td>
<td>get</td>
<td>minute</td>
<td>directions</td>
<td>close</td>
</tr>
<tr>
<td>schedule</td>
<td>police</td>
<td>beltway</td>
<td>operating</td>
<td>us</td>
</tr>
</tbody>
</table>
Extraction

• Single keywords generate lot of noise

• Apriori algorithm is used to construct association rules
Ranking and selection

• Ranking determines which tweet is more relevant to incident scenario.

• Each keyword is associated a particular weight based on its occurrence in historical set of words.

• Weights were added and final value of the tweet is generated.
Extracted Tweet Volume

Tweet Volume for 02-Nov-2013
Extracted Tweet Volume

Tweet Volume for 31-Oct-2013
MongDB: Storing and accessing

• Scalability is a big issue with Twitter data

• Number of data fields provided vary for the data

• MongoDB, a NoSQL database was selected to meet the requirements
# MongoDB: Storing and accessing

<table>
<thead>
<tr>
<th>Key</th>
<th>Value</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>_id</td>
<td>476575427474243584</td>
<td>Int64</td>
</tr>
<tr>
<td>lang</td>
<td>en</td>
<td>String</td>
</tr>
<tr>
<td>favorited</td>
<td>false</td>
<td>Boolean</td>
</tr>
<tr>
<td>pj.UTCtime</td>
<td>1402459410</td>
<td>Int32</td>
</tr>
<tr>
<td>retweeted_status</td>
<td>false</td>
<td>Boolean</td>
</tr>
<tr>
<td>truncated</td>
<td></td>
<td></td>
</tr>
<tr>
<td>text</td>
<td>West / in-bound lanes on NY Ave @ N-Capitol are close... Wed Jun 11 04:03:30 +0000 2014</td>
<td>String</td>
</tr>
<tr>
<td>created_at</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pj_scoreObj</td>
<td>Array [1]</td>
<td>Array</td>
</tr>
<tr>
<td>retweeted</td>
<td>false</td>
<td>Boolean</td>
</tr>
<tr>
<td>user_mentions</td>
<td>Array [1]</td>
<td>Array</td>
</tr>
<tr>
<td>source</td>
<td></td>
<td></td>
</tr>
<tr>
<td>user</td>
<td></td>
<td></td>
</tr>
<tr>
<td>retweet_count</td>
<td>1</td>
<td>Int32</td>
</tr>
<tr>
<td>id</td>
<td>476575427474243584</td>
<td>Int64</td>
</tr>
</tbody>
</table>
Display Results

• The main front end user interface is implemented by bootstrap framework.
• Data-Driven Display (D3.js) is leveraged to provide visualization results (histogram)
DDOT Social Media Project

The District of Columbia Department of Transportation’s Social Media Project dedicates on utilizing the full use of existing social media data source to help transportation agents visualizing, analyzing and even predicting the transportation events.

Recent Tweets Display
This application filters the tweets based on the score of each tweet. The scores can be calculated by multiple methods such as the term frequency inverse document frequency, the SVM classifier score and the Naive Bayesian classifier score.

Incident Data Supplement
The Incident Data Supplement application is capable for retrieving the incidents' records from DDOT CapTop databases, identifies the records that lack of detailed descriptions, then turn to the social media data sources from supplementation.

Date Filter
The function for Date Filter application is to retrieve crawled social media data documents based on the user specified datetime. This application can verify whether DDOT CapTop incidents database records the tweeted real-time traffic events.
Display Results

Most Recently Posted Tweets

WTOP Traffic @WTOPtraffic
#Quantico I-95 both directions to exit Russell Rd (x148) AND Rt 619 (x150): STOPPED to get into base. Expect delays #vtraffic #dctrffic
8:08 AM - 5 Nov 2014

Metrobus Info @Metrobusinfo
S1, S2, S4, S9: Due to traffic congestion at 16th & U St NW, buses are experiencing up to 20 minute delays in both directions.
8:04 AM - 5 Nov 2014

Jim Russ @JimWBLATraffic
UPDATE: CRASH - INNER LOOP at US 40/Pulaski Hwy - RIGHT lane BLOCKED. Jammed from I 83/Harrisburg Expwy @wbaltoday #dctrffic
8:03 AM - 5 Nov 2014

AA County 1st Alert @AACounty1stAlert
#MDtraffic | Inner Loop 695 before Route 170 | Crash on Right Shoulder | Back up before Route 10
8:01 AM - 5 Nov 2014

Metrobus Info @Metrobusinfo
US, W4: Due to traffic congestion along Benning Rd SE near E St, buses are experiencing up to 30 minute delays in both directions.
8:01 AM - 5 Nov 2014

WTOP Traffic @WTOPtraffic
#Alexandria King St @ Commonwealth Ave: Crash in the intersection, follow police direction #vtraffic #dctrffic
8:00 AM - 5 Nov 2014

Rob Base 92Q @RobBase92Q
I/L 695 before Route 170 RIGHT SHOULDER accident JAMMED before Route 10 #AnneArundelCounty #dctrffic
7:56 AM - 5 Nov 2014

MoCo Shady Grove Rd WB past Mid-County Hwy: Crash CLEARED #dctrffic
7:55 AM - 5 Nov 2014

WTOP Traffic @WTOPtraffic
W4: Due to traffic congestion along Eastern Ave, buses are experiencing up to 30 minute delays in both directions.
7:54 AM - 5 Nov 2014

WTOP Traffic @WTOPtraffic
#PGCo Beltway Inner Loop @ Rt 450: 2 RIGHT lanes BLOCKED due to crash
7:30 AM - 5 Nov 2014

Sarah Caldwell @SarahWBL
Motorists can’t get a break this AM 1/4 A A St NW #Dctrffic LCRA\"ed
7:30 AM - 5 Nov 2014

WTOP Traffic @WTOPtraffic
#MDtraffic Rt 15 SB on Point of Rocks Bridge Should be CLEARED
7:30 AM - 5 Nov 2014

Metrobus Info @Metrobusinfo
MoCo Annex on the north side of Coles Crossing Rd: Traffic CLEARED #vtraffic #dctrffic
7:16 AM - 5 Nov 2014

WTOP Traffic @WTOPtraffic
Update: CRASH - INNER LOOP at US 40/Pulaski Hwy - RIGHT lane BLOCKED. Jammed from I 83/Harrisburg Expwy @wbaltoday #dctrffic
7:12 AM - 5 Nov 2014

Metrobus Info @Metrobusinfo
US, W4: Due to traffic congestion along Benning Rd SE near E St, buses are experiencing up to 30 minute delays in both directions.
7:11 AM - 5 Nov 2014

AA County 1st Alert @AACounty1stAlert
#MDtraffic | Inner Loop 695 before Route 170 | Crash on Right Shoulder | Back up before Route 10
7:10 AM - 5 Nov 2014

Metrobus Info @Metrobusinfo
US, W4: Due to traffic congestion along Benning Rd SE near E St, buses are experiencing up to 30 minute delays in both directions.
7:08 AM - 5 Nov 2014

AA County 1st Alert @AACounty1stAlert
#MDtraffic | Inner Loop 695 before Route 170 | Crash on Right Shoulder | Back up before Route 10
7:06 AM - 5 Nov 2014

Metrobus Info @Metrobusinfo
US, W4: Due to traffic congestion along Benning Rd SE near E St, buses are experiencing up to 30 minute delays in both directions.
7:05 AM - 5 Nov 2014

AA County 1st Alert @AACounty1stAlert
#MDtraffic | Inner Loop 695 before Route 170 | Crash on Right Shoulder | Back up before Route 10
7:04 AM - 5 Nov 2014

Metrobus Info @Metrobusinfo
US, W4: Due to traffic congestion along Benning Rd SE near E St, buses are experiencing up to 30 minute delays in both directions.
7:01 AM - 5 Nov 2014

AA County 1st Alert @AACounty1stAlert
#MDtraffic | Inner Loop 695 before Route 170 | Crash on Right Shoulder | Back up before Route 10
6:59 AM - 5 Nov 2014

Metrobus Info @Metrobusinfo
US, W4: Due to traffic congestion along Benning Rd SE near E St, buses are experiencing up to 30 minute delays in both directions.
6:57 AM - 5 Nov 2014

AA County 1st Alert @AACounty1stAlert
#MDtraffic | Inner Loop 695 before Route 170 | Crash on Right Shoulder | Back up before Route 10
6:55 AM - 5 Nov 2014

Metrobus Info @Metrobusinfo
US, W4: Due to traffic congestion along Benning Rd SE near E St, buses are experiencing up to 30 minute delays in both directions.
6:54 AM - 5 Nov 2014

AA County 1st Alert @AACounty1stAlert
#MDtraffic | Inner Loop 695 before Route 170 | Crash on Right Shoulder | Back up before Route 10
6:52 AM - 5 Nov 2014

Metrobus Info @Metrobusinfo
US, W4: Due to traffic congestion along Benning Rd SE near E St, buses are experiencing up to 30 minute delays in both directions.
6:51 AM - 5 Nov 2014

AA County 1st Alert @AACounty1stAlert
#MDtraffic | Inner Loop 695 before Route 170 | Crash on Right Shoulder | Back up before Route 10
6:50 AM - 5 Nov 2014

Metrobus Info @Metrobusinfo
US, W4: Due to traffic congestion along Benning Rd SE near E St, buses are experiencing up to 30 minute delays in both directions.
6:49 AM - 5 Nov 2014

AA County 1st Alert @AACounty1stAlert
#MDtraffic | Inner Loop 695 before Route 170 | Crash on Right Shoulder | Back up before Route 10
6:47 AM - 5 Nov 2014
<table>
<thead>
<tr>
<th>Tweet Id</th>
<th>Created at</th>
<th>Text</th>
<th>Score</th>
<th>Location</th>
<th>User</th>
<th>Keywords</th>
</tr>
</thead>
<tbody>
<tr>
<td>5298030882569155020</td>
<td>04-Nov-2014 20:11:14</td>
<td>The unit - 200 blk. of Anacostia Ave &amp; 300 blk. of Blaine St NE is BLOCKED due to an unattended package in the park/EGO responding /7066</td>
<td>4.86246532111</td>
<td>NA</td>
<td>bforte22</td>
<td>ave</td>
</tr>
<tr>
<td>529802936420649520</td>
<td>04-Nov-2014 20:10:30</td>
<td>Red Line: Expect residual delays to Glenmont due to an earlier train malfunction at Gallery Place.</td>
<td>6.0366972477</td>
<td>Washington, DC Metro Area</td>
<td>dcmetroalerts</td>
<td>red</td>
</tr>
<tr>
<td>52980280727136348017</td>
<td>04-Nov-2014 20:10:20</td>
<td>The unit - 200 blk. of Anacostia Ave &amp; 300 blk. of Blaine St NE is BLOCKED due to an unattended package in the park/EGO responding /7066</td>
<td>4.86246532111</td>
<td>NA</td>
<td>omarin41</td>
<td>ave</td>
</tr>
<tr>
<td>5298026983121954816</td>
<td>04-Nov-2014 20:09:37</td>
<td>The unit - 200 blk. of Anacostia Ave &amp; 300 blk. of Blaine St NE is BLOCKED due to an unattended package in the park/EGO responding /7066</td>
<td>4.86246532111</td>
<td>Washington, DC</td>
<td>DCPoliceDept</td>
<td>ave</td>
</tr>
<tr>
<td>528600230177121854</td>
<td>04-Nov-2014 20:08:06</td>
<td>RED Red Line: Expect residual delays to Glenmont due to an earlier train malfunction at Gallery Place.</td>
<td>6.22014934623</td>
<td>Washington D.C.</td>
<td>dcmta</td>
<td>red</td>
</tr>
<tr>
<td>52980208854838910</td>
<td>04-Nov-2014 20:07:15</td>
<td>CONSTANTLY FAILING: @MetrorailInfo Red Line: Expect residual delays to Glenmont due to an earlier train malfunction at Gallery Place.</td>
<td>6.0386972477</td>
<td>Washington, D.C.</td>
<td>1stLadySHAE</td>
<td>red</td>
</tr>
<tr>
<td>528601890319773028</td>
<td>04-Nov-2014 20:06:30</td>
<td>Red Line: Expect residual delays to Glenmont due to an earlier train malfunction at Gallery Place.</td>
<td>6.0366972477</td>
<td>Washington, DC</td>
<td>MetrorailInfo</td>
<td>red</td>
</tr>
<tr>
<td>529796546787608676</td>
<td>04-Nov-2014 20:05:30</td>
<td>DULLES TOLL ROAD VA 267 WB Before Exit 9 at VA 281</td>
<td>207.100917431</td>
<td>NA</td>
<td>VA166thm</td>
<td>road</td>
</tr>
</tbody>
</table>
Display Results
Validation

- The results found so far are very accurate and are near real time compared to tweets and incidents.

<table>
<thead>
<tr>
<th>Incident Id</th>
<th>Date</th>
<th>Time</th>
<th>Street Id (A)</th>
<th>Street Id (B)</th>
<th>Description</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>69673</td>
<td>6/18/2014</td>
<td>09:08</td>
<td>branch ave</td>
<td>pennsylvania ave</td>
<td>DC eg 3526</td>
<td>NULL</td>
</tr>
<tr>
<td>69724</td>
<td>6/19/2014</td>
<td>10:50</td>
<td>first st</td>
<td>florida ave</td>
<td>NULL</td>
<td>nothing found ROP 4 responded to incident.</td>
</tr>
<tr>
<td>69734</td>
<td>6/19/2014</td>
<td>16:59</td>
<td>Park Street and W Street</td>
<td>Branch Ave</td>
<td>NULL</td>
<td>NULL</td>
</tr>
<tr>
<td>69735</td>
<td>6/19/2014</td>
<td>17:10</td>
<td>eastern ave</td>
<td>295</td>
<td>NULL</td>
<td>slow traffic condition nb 295 prior to crash in Maryland side.. camera out of order</td>
</tr>
</tbody>
</table>
Validation
Future Work

• Update the keywords weights dynamically.

• Integrate the results to our ATMS and auto fill accidents descriptions

• All the code is available for download for free at http://ddot-toa.github.io/
Questions?