Traffic Management Enabled By High Performance Computing
Challenges for a Civil Engineer to Perform HPC
Motivation

Tianhe-2 (Milkyway-2) Supercomputer
Authors Forrest Hoffman (standing) and Bill Hargrove sit "inside" the computer they constructed from commodity PCs.
(Concept of MPI)
Hadoop History

- 2002: Google publishes GFS & MapReduce papers
- 2003: Doug Cutting & Mike Cafarella started working on Nutch
- 2004: Yahoo! hires Cutting, Hadoop spins out of Nutch
- 2005: NY Times converts 4TB of Image archives over 100 EC2s
- 2006: Yahoo! sorts a TB, 3.5mins over 910 nodes
- 2007: Facebook launches Hive: SQL Support for Hadoop
- 2008: Hadoop Summit 2009, 750 attendees
- 2009: Doug Cutting joins Cloudera

Source: Cloudera, Inc.

www.edureka.in/hadoop-admin
Hadoop 1.x Core Components

Hadoop is a system for large scale data processing. It has two main components:

✓ **HDFS – Hadoop Distributed File System (Storage)**
  ✓ Distributed across “nodes”
  ✓ Natively redundant
  ✓ NameNode tracks locations.

✓ **MapReduce (Processing)**
  ✓ Splits a task across processors
  ✓ “near” the data & assembles results
  ✓ Self-Healing, High Bandwidth
  ✓ Clustered storage
  ✓ JobTracker manages the TaskTrackers

✓ **Additional Administration Tools:**
  ✓ Filesystem utilities
  ✓ Job scheduling and monitoring
  ✓ Web UI

www.edureka.in/hadoop-admin
Hadoop 1.x Core Components (Contd.)
**Secondary Name Node**

- Secondary NameNode:
  - Not a hot standby for the NameNode
  - Connects to NameNode every hour*
  - Housekeeping, backup of NameNode metadata
  - Saved metadata can build a failed NameNode

---

[www.edureka.in/hadoop-admin](http://www.edureka.in/hadoop-admin)
Map Reduce
The overall MapReduce word count process

Input
- Deer Bear River
- Car Car River
- Deer Car Bear

Splitting
- Deer Bear River
- Car Car River
- Deer Car Bear

Mapping
- Deer, 1
  - River, 1
- Car, 1
  - Car, 1
  - River, 1
- Deer, 1
  - Car, 1
  - Bear, 1

Shuffling
- Bear, 1
- Bear, 1
- Car, 1
- Car, 1
- Car, 1
- Deer, 1
- Deer, 1
- Deer, 1
- Deer, 1
- River, 1

Reducing
- Bear, 2
- Bear, 1
- Car, 3
- Car, 1
- Car, 1
- Deer, 2
- Deer, 1
- River, 2

Final result
- Bear, 2
- Car, 3
- Deer, 2
- River, 2
Example Problem

• Sentiment Analysis
  • A twitter log is obtained from city of Chicago with each tweet listed in a single line
  • Need to find relative frequency of word Iowa State for a given day
Security breach problem

- IP address, Login Attempt, Time of Day, Status is provided
- Is it possible to detect anomalous activity?
More advance logic can be coded

• Djikstra’s shortest path
• Page Rank

https://courses.cs.washington.edu/courses/cse490h/08au/lectures/algorithms.pdf
Traffic Performance Measure in Hadoop – MapReduce

1. Easy job can be quickly handled in Pig
2. Some complicated job may need your own MapReduce program
Apache Hadoop Ecosystem

Provisioning, Managing and Monitoring Hadoop Clusters

- Ambari
- YARN Map Reduce v2
  - Distributed Processing Framework
- HDFS
  - Hadoop Distributed File System
- Sqoop
  - Data Exchange
- Oozie
  - Workflow
- Pig
  - Scripting
- Mahout
  - Machine Learning
- R Connectors
  - Statistics
- Hive
  - SQL Query
- Flume
  - Log Collector
- Zookeeper
  - Coordination
Traffic Management

- Data Quality Assurance
- Probe Data
- Point Detector Data
- Long/Short Term Decision – Policy
- Mobility
- Work Zone
- Real Time Decision Support
- Traffic Risk Prediction
- Information Dissemination
Traffic Management

Data Quality Assurance

- Probe Data
- Point Detector Data
Reliability of Detectors

Probe Data
• Coverage

• Accuracy

Point Detectors
• Erroneous Detectors
• **Benchmark Dataset:** Wavetronix
• **Probe Dataset:** INRIX
• **Analysis Period:** 1 month
• **Segment - Sensor Pairs:** 100 (Freeways: 60 and Non-Freeway: 40)
• **Association Rules:** Bearing, Proximity
Trend Analysis - Empirical Mode Decomposition

Benchmark

Probe Data

1. Get Corresponding probe and Benchmark Data
2. Extract Short, Medium and Long-term trends
3. Calculate the similarity between trends
4. Similarity Measure $\geq$ Threshold
5. Yes
6. Estimate latency between dataset
Corresponding Trends

- **Long – Term Trends**: Weekly, Monthly, etc
- **Medium – Term Trends**: 1 – 3 Hourly and daily.
- **Short – Term Trends**: 15 – 30 minutes

Red Line – INRIX.
Black Line – Wavetronix.
Latencies: Short-Term Example

**Original Dataset**

**Congestion Detection**

Get Corresponding probe and Benchmark Data → Extract Short, Medium and Long-term trends → Calculate the similarity between trends

Similarity Measure $\geq$ Threshold → Estimate latency between dataset
Latencies: Medium – Term Example

Get Corresponding probe and Benchmark Data → Extract Short, Medium and Long-term trends → Calculate the similarity between trends → Similarity Measure \(\geq\) Threshold → Yes → Estimate latency between dataset
Latencies

Get Corresponding probe and Benchmark Data

Extract Short, Medium and Long-term trends

Calculate the similarity between trends

Similarity Measure >= Threshold

Yes

Estimate latency between dataset
## Congestion Detection Reliability

<table>
<thead>
<tr>
<th>Road Type</th>
<th>Congestion Type</th>
<th>Test</th>
<th>Synchronized Events - $S$ (%)</th>
<th>Lag - $L$ (minutes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Freeways/Expressway</td>
<td>Short - Term</td>
<td>True Positives</td>
<td>73.6</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>False Negatives</td>
<td>16.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>False Positives</td>
<td>9.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Medium - term</td>
<td>True Positives</td>
<td>95</td>
<td>Mean lag = 8.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>False Negatives</td>
<td>4.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>False Positives</td>
<td>1.7</td>
<td></td>
</tr>
<tr>
<td>Non-Freeway/Non-Expressway</td>
<td>Short - Term</td>
<td>True Positives</td>
<td>62.6</td>
<td>Mean lag = 12.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>False Negatives</td>
<td>12.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>False Positives</td>
<td>24.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Medium - term</td>
<td>True Positives</td>
<td>68</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>False Negatives</td>
<td>15.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>False Positives</td>
<td>16.7</td>
<td></td>
</tr>
</tbody>
</table>
Radar based point sensors
<table>
<thead>
<tr>
<th>Category</th>
<th>Type</th>
<th>Definition</th>
<th>Selecting Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Error</td>
<td>Count and Speed Error</td>
<td>Null (or zero) count value associated with positive speed value; Null (or zero) speed value associated with positive count value</td>
<td>Happened &gt;50% time of study period</td>
</tr>
<tr>
<td></td>
<td>Vehicle Class Count Error</td>
<td>The sum of vehicle class counts is not equal to the total count</td>
<td>Happened &gt;90% time of study period</td>
</tr>
<tr>
<td></td>
<td>Negative Speed Error</td>
<td>Speed value is negative</td>
<td>Happened &gt;10% time of study period</td>
</tr>
<tr>
<td></td>
<td>Same Name Error</td>
<td>Different sensors with different coordinates have exactly same name in dataset and inventory</td>
<td>All</td>
</tr>
<tr>
<td>Downtime</td>
<td>Failure</td>
<td>Data with failure status</td>
<td>Happened &gt;50% time of study period</td>
</tr>
<tr>
<td></td>
<td>Off</td>
<td>Data with off status</td>
<td>Happened &gt;50% time of study period</td>
</tr>
<tr>
<td></td>
<td>Missing Data</td>
<td>Data are not reported at the expected timestamp</td>
<td>Happened &gt;50% time of study period</td>
</tr>
<tr>
<td>Anomalous Events</td>
<td>Count Outlier</td>
<td>Count value exceeds 13 veh/20s/ln (equivalent to 2340 veh/h/ln)</td>
<td>Happened &gt;10% time of study period</td>
</tr>
<tr>
<td></td>
<td>Speed Outlier</td>
<td>Speed value exceeds 100 mph</td>
<td>Happened &gt;10% time of study period</td>
</tr>
<tr>
<td></td>
<td>a. Count Difference</td>
<td>Maximum absolute count difference between adjacent sensors (downstream-upstream)</td>
<td>&gt;50</td>
</tr>
<tr>
<td></td>
<td>b. Speed Difference</td>
<td>Maximum absolute speed difference between adjacent sensors (downstream-upstream)</td>
<td>&gt;65 &amp; the corresponding speed &gt;75 or &lt;5 mph</td>
</tr>
<tr>
<td></td>
<td>c. Small Vehicle Difference</td>
<td>Maximum absolute small vehicle count difference between adjacent sensors (downstream-upstream)</td>
<td>&gt;30</td>
</tr>
<tr>
<td></td>
<td>d. Medium Vehicle Difference</td>
<td>Maximum absolute medium vehicle count difference between adjacent sensors (downstream-upstream)</td>
<td>&gt;20</td>
</tr>
<tr>
<td></td>
<td>e. Large Vehicle Difference</td>
<td>Maximum absolute large vehicle count difference between adjacent sensors (downstream-upstream)</td>
<td>&gt;15</td>
</tr>
</tbody>
</table>
| Traffic Flow Theory | Average Effective Vehicle Length | \[
\frac{5280 \times speed \times occupancy\%}{hourly\ volume}\ (ft) >75 \text{ ft or } <10 \text{ ft}
\] |                                           |

*In Performance category, we only use the data during nighttime (10pm-4am) and from adjacent sensors with no ramp in between. Weighted performance score is used to rank the sensors.
# Traffic Flow Theory

<table>
<thead>
<tr>
<th>Category</th>
<th>Type</th>
<th>Definition</th>
<th>Selecting Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traffic Flow Theory</td>
<td>Average Effective Vehicle Length</td>
<td>$\frac{5280 \times \text{speed} \times \text{occupancy%}}{\text{hourly volume}}$ (ft)</td>
<td>&gt;75 ft or &lt;10 ft</td>
</tr>
</tbody>
</table>
Example-Four Sensors on I-35

Average Length Computed by Traffic Flow Theory (Blue) and Vehicle Class Count Data (Red)
Example-Four Sensors on I-35

Original Speed-Occupancy Curve By Volume
Example-January, 2016
Example-January, 2016
Traffic Management

Long/Short Term Decision – Policy
Congested Hours

- Calculate hours of speed less than 45 mph
- Look at each minute of data
  - Congested if speed is less than 45 mph and real time score (30)
- Summarized data by time of day, day of week and month
Traffic Performance Measure in Hadoop – Example 1: filter by speed < 45

Data

<table>
<thead>
<tr>
<th>tmc_code</th>
<th>measurement_tstamp</th>
<th>speed</th>
<th>confidence_score</th>
</tr>
</thead>
<tbody>
<tr>
<td>29243</td>
<td>2/3/2014 20:15</td>
<td>67.51</td>
<td>30</td>
</tr>
<tr>
<td>29244</td>
<td>2/3/2014 20:15</td>
<td>66.13</td>
<td>30</td>
</tr>
<tr>
<td>29245</td>
<td>2/3/2014 20:15</td>
<td>68.32</td>
<td>30</td>
</tr>
<tr>
<td>29246</td>
<td>2/3/2014 20:15</td>
<td>69.93</td>
<td>30</td>
</tr>
<tr>
<td>29247</td>
<td>2/3/2014 20:15</td>
<td>68.89</td>
<td>30</td>
</tr>
<tr>
<td>29248</td>
<td>2/3/2014 20:15</td>
<td>69.25</td>
<td>30</td>
</tr>
<tr>
<td>29249</td>
<td>2/3/2014 20:15</td>
<td>71.37</td>
<td>30</td>
</tr>
<tr>
<td>29250</td>
<td>2/3/2014 20:15</td>
<td>66.39</td>
<td>30</td>
</tr>
<tr>
<td>29251</td>
<td>2/3/2014 20:15</td>
<td>63.59</td>
<td>30</td>
</tr>
<tr>
<td>29252</td>
<td>2/3/2014 20:30</td>
<td>70.53</td>
<td>30</td>
</tr>
<tr>
<td>29253</td>
<td>2/3/2014 20:30</td>
<td>71.57</td>
<td>30</td>
</tr>
<tr>
<td>29254</td>
<td>2/3/2014 20:30</td>
<td>71.61</td>
<td>30</td>
</tr>
<tr>
<td>29255</td>
<td>2/3/2014 20:30</td>
<td>70.83</td>
<td>30</td>
</tr>
<tr>
<td>29256</td>
<td>2/3/2014 20:30</td>
<td>70.86</td>
<td>30</td>
</tr>
<tr>
<td>29257</td>
<td>2/3/2014 20:30</td>
<td>64.91</td>
<td>30</td>
</tr>
<tr>
<td>29258</td>
<td>2/3/2014 20:30</td>
<td>69.35</td>
<td>30</td>
</tr>
<tr>
<td>29259</td>
<td>2/3/2014 20:30</td>
<td>70.75</td>
<td>30</td>
</tr>
<tr>
<td>29260</td>
<td>2/3/2014 20:30</td>
<td>72.16</td>
<td>30</td>
</tr>
<tr>
<td>29261</td>
<td>2/3/2014 20:45</td>
<td>67.2</td>
<td>30</td>
</tr>
<tr>
<td>29262</td>
<td>2/3/2014 20:45</td>
<td>68.62</td>
<td>30</td>
</tr>
<tr>
<td>29263</td>
<td>2/3/2014 20:45</td>
<td>63.92</td>
<td>30</td>
</tr>
<tr>
<td>29264</td>
<td>2/3/2014 20:45</td>
<td>68.13</td>
<td>30</td>
</tr>
</tbody>
</table>

1. A short pig script
2. Put data and .pig file in through WinSCP
Traffic Performance Measure in Hadoop – Example 2: monthly performance measure for each tmc segment

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
</tr>
</thead>
<tbody>
<tr>
<td>tmc</td>
<td>month</td>
<td>avg_speed</td>
<td>speed&lt;45</td>
<td>speed&lt;avg_15</td>
<td>speed&lt;avg_20</td>
<td>speed&lt;avg_25</td>
<td>speed&lt;avg_30</td>
<td>speed&lt;avg_45</td>
<td>Count</td>
</tr>
<tr>
<td>118+04641</td>
<td>201301</td>
<td>60</td>
<td>158</td>
<td>158</td>
<td>105</td>
<td>35</td>
<td>8</td>
<td>3</td>
<td>44283</td>
</tr>
<tr>
<td>118+04641</td>
<td>201302</td>
<td>60</td>
<td>313</td>
<td>313</td>
<td>240</td>
<td>122</td>
<td>26</td>
<td>2</td>
<td>39962</td>
</tr>
<tr>
<td>118+04641</td>
<td>201303</td>
<td>60</td>
<td>40</td>
<td>40</td>
<td>35</td>
<td>24</td>
<td>13</td>
<td>0</td>
<td>44070</td>
</tr>
<tr>
<td>118+04641</td>
<td>201304</td>
<td>60</td>
<td>19</td>
<td>19</td>
<td>17</td>
<td>9</td>
<td>2</td>
<td>2</td>
<td>43057</td>
</tr>
<tr>
<td>118+04641</td>
<td>201305</td>
<td>60</td>
<td>19</td>
<td>19</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>44140</td>
</tr>
<tr>
<td>118+04641</td>
<td>201306</td>
<td>60</td>
<td>23</td>
<td>23</td>
<td>21</td>
<td>18</td>
<td>6</td>
<td>4</td>
<td>42080</td>
</tr>
<tr>
<td>118+04641</td>
<td>201307</td>
<td>60</td>
<td>73</td>
<td>73</td>
<td>67</td>
<td>60</td>
<td>47</td>
<td>29</td>
<td>44507</td>
</tr>
<tr>
<td>118+04641</td>
<td>201308</td>
<td>60</td>
<td>101</td>
<td>101</td>
<td>78</td>
<td>74</td>
<td>62</td>
<td>40</td>
<td>44636</td>
</tr>
<tr>
<td>118+04641</td>
<td>201309</td>
<td>60</td>
<td>44</td>
<td>44</td>
<td>33</td>
<td>32</td>
<td>27</td>
<td>20</td>
<td>42902</td>
</tr>
<tr>
<td>118+04641</td>
<td>201310</td>
<td>60</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>2</td>
<td>2</td>
<td>44402</td>
</tr>
<tr>
<td>118+04641</td>
<td>201311</td>
<td>60</td>
<td>160</td>
<td>160</td>
<td>147</td>
<td>90</td>
<td>52</td>
<td>41</td>
<td>42939</td>
</tr>
<tr>
<td>118+04641</td>
<td>201312</td>
<td>59</td>
<td>475</td>
<td>418</td>
<td>298</td>
<td>234</td>
<td>152</td>
<td>64</td>
<td>44438</td>
</tr>
</tbody>
</table>
Top 10 Most Congested - 2013

Interstate 80 Westbound
The top congested segment in Iowa was located in the Quad Cities on Interstate 80 westbound at the Illinois border. This half mile segment of road had 378.6 hours of congestion in 2013 which was primarily between June and November. Congested peaked with over 90 hours a month during both July and August. Construction on the Mississippi River bridge likely caused the congestion. Congestion primarily occurred between 9 am and 6 pm with a majority during the PM Peak. Friday was the peak of congestion while all other days had consistent congestion hours.

Interstate 29 Northbound
Up to 326.3 hours of congestion was experienced between Exit 141 (Sargent Bluffs) and Exit 147A (Floyd Blvd) on I-29 northbound in Sioux City during 2013. This 5.8 miles of road had the most congestion between the months of August and October and 66.3 hours of congestion in September. The congestion was fairly consistent across the time of day and for all days of the week but was slightly higher between 9 am and 6 pm.
Top 10 Most Congested - 2014

Interstate 29 Southbound (Sioux City)
Congestion on Interstate 29 southbound through Sioux City increased from 326.3 hours to 586.2 hours due to construction beginning in May. After the construction began, the congestion was consistent for the remaining months in 2014. Most congestion occurred during the week with slightly higher amounts on Fridays. Both the AM and PM peaks showed congestion from Exit 1478 (Tyson Event Center) and Exit 144 (US 20).

Interstate 29 Northbound (Sioux City)
Another new addition to the top 10 in 2014 was along Interstate 29 northbound between Exit 141 (Sergeant Bluff) and Exit 147A (Floyd Blvd). The 370.7 hours of congestion was due to the construction that began in May and continued until the end of the year. This stretch of road covered 5.8 miles through the Sioux City area. Very little of the congestion occurred on the weekends and during the week congestion was slightly higher on Fridays. The congestion was also consistently during the PM Peak.
Top 10 Most Congested - 2015

Interstate 29 Northbound (Sioux City)
The most congested roadway in Iowa went from 170.7 hours in 2014 to 665 hours in 2015 through Sioux City between Exit 141 (Sergeant Bluffs) and Exit 147A (Floyd Blvd) on I-29 northbound. This 5.8 miles of congestion was primarily due to construction in the area that restricted traffic to one lane in each direction. Congestion was heaviest on Wednesday thru Friday with very little occurring on weekends. Congestion was consistent beginning in March and mostly occurring in the PM peak.

665.0 hours

Interstate 29 Southbound (Sioux City)
The southbound direction through Sioux City experienced a considerable increase in congested hours due to the construction. This section of I-29 between Exit 147B (Tyson Event Center) and Exit 144 (US 20) experienced 631.2 hours of congestion compared to 585.2 hours in 2014. Most of the congestion occurred between April and November but peaked during the month of August. Congestion was common on weekdays with slightly higher congestion on Mondays. The congestion was occurring both in the AM and PM peak.

631.2 hours
Interstate 35 - 2014
Traffic Management

Long/Short Term Decision – Policy

Office of Traffic Operations Performance Report
Reporting Period: September 6- September 19, 2015

Performance Report

Work Zone
Traffic Management

Real Time Decision Support

Traffic Risk Prediction
Traffic Incident Detection and Incident Risk Prediction → SUPERVISED LEARNING
(Pre-identified events such as crashes)
Detection vs. Prediction

- Both belong to classification problem
- Similar methodology
- Different input window
Input windows

• Detection: Input window covers the detected target.

• Prediction: Input window before the predicted target.
Detection and Prediction as a Classification Problem

• Aspects to work on for improving accuracy:
  
  • Increase the separability between classes:
    • Feature engineering and feature selection
  
  • Improve the training dataset
    • Resampling method vs. big data
  
  • Reduce the model variance and bias
    • Model selection, ensemble learning, etc.
Increase the separability between classes

- Traffic speed vs. incident

- Distribution of speed at the time of incident

- Adding useful features to pull the two classes further away
  - Speed from different locations and different time
  - Standard deviation of speed
  - Speed differentials
  - Geometry, weather, volume, occupancy, etc..
Improve the training dataset

- Balancing the two classes
  - Reduce major class
  - Increase minor class
  - Combinations

- Big data
  - Train with the “population”
  - Appropriate analytic tools required due to the large data size.
Reduce the model variance and bias

• Single models
  • naive Bayes, SVM, logistic regression, perceptron, multilayer perceptron, neuro networks, Bayes net, decision tree, etc..
  • [SVMStruct = svmtrain(Training,Group)]

• Meta models (ensemble learning)
  • Boosting – Multiple model ensemble
  • Bagging – Bootstrap aggregation
  • etc.
Model Performance Measure

- Receiver Operating Characteristic (ROC) curve
- Area Under the Curve (AUC)
Examples of performance from risk prediction models (naive Bayes)
Examples of performance from different risk prediction models

• Comparable results from naive Bayes training on the “population” in H2O (Big Data model):
Help from the Big Data Tech

• Pros:
  • More training data provides higher confidence
  • Possibly handle large roadway network
  • Possibly handle real-time application

• Cons:
  • Still lack of easy-to-use tools
Traffic Management

Real Time Decision Support

Information Dissemination
Evaluating VSL system installed on I-35 SB

- Using existing Wavetronix sensors
- Friction sensors
- Visibility sensors
- RWIS
- Portable DMS
Issues

-Reviewed winter events on December 24\textsuperscript{th} and 28\textsuperscript{th}.

-Determined the current systems within TransSuite were not designed for a dynamic advisory messaging.

-Messages were not continuous, messages were fluctuating, and saw random messages when speeds were not slow
Today’s TAC Meeting

- Present initial algorithm developed for Dynamic Advisory Messaging and feedback

- Present next stages of algorithm development

- Discuss integration of other real time data source (friction, visibility, other weather sources, snow plow)
Overall Schema

Raw Sensor Data — Speed & Occupancy → Smooth Sensor Data via Fourier Filter → K-Means for Data Clustering

Assign Labels → SVM to Train the Function $f$

$f(x) \rightarrow y$

$x$: Streamed New Speed & Occu.
$y$: Advised Speed Limit
Data Smoothing

Initial algorithm:
Fast Fourier Transform

Training: 15 min history

Use speed and occupancy from Wavetronix sensors
Expert Characterization of Messages - Kmeans

K = 2

K = 3

K = 4

Time of Day

Speed
Overall Schema

De Noise

Labeling (Unsupervised Learning)

Training a Classifier (Supervised Learning)
<table>
<thead>
<tr>
<th>VSL Station</th>
<th>Cluster</th>
<th>Short Msg.</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1 on Dec 24&lt;sup&gt;th&lt;/sup&gt;</td>
<td>1/(4)</td>
<td>8/(13)</td>
<td>16/(34)</td>
</tr>
<tr>
<td>#1 on Dec 28&lt;sup&gt;th&lt;/sup&gt;</td>
<td>1/(1)</td>
<td>8/(2)</td>
<td>23/(33)</td>
</tr>
<tr>
<td>#2 on Dec 24&lt;sup&gt;th&lt;/sup&gt;</td>
<td>3/(0)</td>
<td>8/(1)</td>
<td>19/(6)</td>
</tr>
<tr>
<td>#2 on Dec 28&lt;sup&gt;th&lt;/sup&gt;</td>
<td>3/(0)</td>
<td>16/(1)</td>
<td>36/(30)</td>
</tr>
<tr>
<td>#3 on Dec 24&lt;sup&gt;th&lt;/sup&gt;</td>
<td>0/(1)</td>
<td>3/(2)</td>
<td>12/(13)</td>
</tr>
<tr>
<td>#3 on Dec 28&lt;sup&gt;th&lt;/sup&gt;</td>
<td>3/(16)</td>
<td>7/(27)</td>
<td>33/(120)</td>
</tr>
<tr>
<td>#4 on Dec 24&lt;sup&gt;th&lt;/sup&gt;</td>
<td>1/(0)</td>
<td>4/(0)</td>
<td>11/(5)</td>
</tr>
<tr>
<td>#4 on Dec 28&lt;sup&gt;th&lt;/sup&gt;</td>
<td>1/(3)</td>
<td>2/(10)</td>
<td>14/(85)</td>
</tr>
</tbody>
</table>
Proposed Logic vs. Existing Logic

2015/12/24 @ Ankeny

Proposed
Cluster: 0
Short Message: 1
Total Message: 4

Existing
Cluster: 0
Short Message: 0
Total Message: 5

Dry
Trace Moisture
Wet

Ice Warning
Ice Watch

Friction
Time of Day(hour)
Transforming Technologies
Naturalistic Driving—Deep Learning a New Paradigm
Traffic Accident

Normal Congestion

Training Network

Input Training Data

Output

“Traffic Accident”

Input Training Data

Training Network

Output

“Normal Congestion”
Video storyline for high impact event
Accident occurs: 07:23:45
Accident reported: 07:40:00

Secondary Crash
07:26:53 (wait for there to be two semi’s close to each other, it occurs behind them)
Building a searchable video story

- Accident occurs: 07:23:45
- Accident reported: 07:40:00
- Secondary Crash: 07:26:53
- First Responder: 07:35:00
- Police Car: 07:37:34
- The ambulance left shortly after the officer arrived
- The officer and the firemen eventually reattach the trailer to the truck and the driver leaves with the trailer
- By 07:52:34, the last emergency vehicle leaves the scene
- The final time that was reported was 08:01:00, and the actual was 07:52:34
- Total actual time was about 29 minutes, and the reported time was 20:47
### 3+ vehicle collision in I-235 Exit 10A on 1st Feb, 2016

<table>
<thead>
<tr>
<th>Event</th>
<th>Event Start</th>
<th>Event End</th>
</tr>
</thead>
<tbody>
<tr>
<td>TMC reported time</td>
<td>7:36:00 AM</td>
<td>9:43:00 AM</td>
</tr>
<tr>
<td>Seen from video data</td>
<td>7:21:42 AM</td>
<td>9:43:00 AM</td>
</tr>
<tr>
<td>Seen from INRIX data in nearest u/s segment</td>
<td>7:21:00 AM</td>
<td>8:50:00 AM</td>
</tr>
<tr>
<td>Highway Helper Arrival and Departure time</td>
<td>7:30:07 AM</td>
<td>--</td>
</tr>
</tbody>
</table>

**Images:**
- 7:21 AM
- 7:26 AM
<table>
<thead>
<tr>
<th></th>
<th>Event Start</th>
<th>Event End</th>
</tr>
</thead>
<tbody>
<tr>
<td>TMC reported time</td>
<td>7:36:00 AM</td>
<td>9:43:00 AM</td>
</tr>
<tr>
<td>Seen from video data</td>
<td>7:21:42 AM</td>
<td>9:43:00 AM</td>
</tr>
<tr>
<td>Seen from INRIX data in nearest u/s segment</td>
<td>7:21:00 AM</td>
<td>8:50:00 AM</td>
</tr>
<tr>
<td>Highway Helper Arrival and Departure time</td>
<td>7:30:07 AM</td>
<td>--</td>
</tr>
</tbody>
</table>
3+ vehicle collision in I-235 Exit 10B on 1st Feb, 2016 (Contd.)
How to Automate
Thank you!

Anuj Sharma,
Associate Professor
Iowa State University

anujs@iastate.edu