NIST Data Science Evaluation Series Pre-Pilot Overview

Peter Fontana, March 17–18, 2016
DSE - Schedule

2014

XDATA
Closed
Domain: Multiple

2015

Pre-Pilot
Single-Track
Invitation-Only
Domain: Traffic

Pilot
NIST Cloud
Single-Track
Open to Everyone
Domain: Traffic

2016

Full-Scale Evaluation
Multiple Tracks
Open to Everyone
Domain: Multiple

2017
Pre-Pilot Data Sets

- Traffic Lane Detectors
- Traffic Events Listing
- Traffic Camera Video
Other Publicly-Available Sets

- OpenStreet Maps
- U.S. Census Data, Including American Community Survey (ACS)
- NOAA Integrated Surface Hourly Weather Data Sets
- Publicly Available Sets welcome
- **Note:** Cleaning Task restricts allowable data
Traffic Camera Video
Traffic Event Listing

- **Event_ID:** MDOT_CHART_a000a2de48a600680055fa2ec42350
- **Description:** "Incident @ I-495 INNER LOOP PRIOR TO EXIT 33 MD 185 CONNECTICUT AVE [Collision, Personal Injury]"
- **Creation Time:** 5/29/15 13:50
- **Incident Type:** accidentsAndIncidents, injury accident,
- **Location (lat, lon):** 39.00536,-77.077752,5,3
## Traffic Detector Data

<table>
<thead>
<tr>
<th>Lane_id</th>
<th>Measurement start</th>
<th>Speed</th>
<th>Flow</th>
</tr>
</thead>
</table>
Pre-Pilot Data Information

- **Traffic Lane Detectors**: Traffic Detectors on Major routes in the MD-DC-VA area from September 2006–August 2015, 150 GB

- **Traffic Events Listing**: Traffic Events in the MD-DC-VA Area from October 2003–August 2015, 200 MB.

- **Traffic Camera Video**: Traffic Camera Video Feeds for 20 MD Traffic Camera, June 2015–August 2015, 4TB
Pre-Pilot Tasks Overview

- **Cleaning**: Clean traffic lane detector data flow values where a portion of its measurements were manipulated to be incorrect.
- **Alignment**: Match traffic events with the traffic video segments containing those events.
- **Prediction**: Predict the number and types of traffic incidents in a region over a time period.
- **Forecasting**: Provide future traffic flow values.
Pre-Pilot Cleaning Task
Pre-Pilot Cleaning Task

Dirty lane detectors

Open Street Map Data

Cleaning System

Cleaned lane detector measurements

Metric (cost) = Mean Absolute Error in Flow
Pre-Pilot Alignment Task

MDOT_CHART_a000a2de48a600680055fa2ec4235c0a, "Incident @ I-495 INNER LOOP ...

90%

MDOT_CHART_400122fc49a700680055fa2ec4235c0a, "Incident @ I-495 OUTER LOOP ...

3%
**Pre-Pilot Alignment Task**

**Traffic camera video segments**

**Traffic events listing**

**Alignment System**

**Video, event pair \((v,e)\) confidence scores**

**Additional data**

**Metric (cost) = (Miss Rate) + 10*(False Alarm Rate)**

Decision threshold selected to **minimize** metric
Pre-Pilot, Prediction Task

March 1, 2015 to April 1, 2015

10 accidents and incidents
3 obstructions
No other incidents
Pre-Pilot Prediction Task

Geographical bounding boxes and time windows
Traffic events (past)
Additional data

Prediction System

Numbers of incidents of each type for each bounding box, time window pair

Metric (cost) = Mean Root Mean Square Error
Pre-Pilot Forecasting Task
Pre-Pilot Forecasting Task

Traffic zone locations and time intervals
Traffic lane detector data (past)
Additional data

Forecasting System

Forecasted traffic flow values for each measurement in each time interval

Note: here, traffic lane detectors are aggregated traffic into traffic zones.
Metric (cost) = Mean of Mean Absolute Percentage Error
## Pre-Pilot Participants

<table>
<thead>
<tr>
<th>Measure</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Registrants</td>
<td>5</td>
</tr>
<tr>
<td>Number of Participant Sites</td>
<td>1</td>
</tr>
<tr>
<td>Cleaning Task Submissions</td>
<td>3</td>
</tr>
<tr>
<td>Prediction Task Submissions</td>
<td>7</td>
</tr>
</tbody>
</table>

Participants selectively invited (some accepted)

Submissions received only for cleaning and prediction tasks

3 NIST baseline systems for cleaning task

1 NIST baseline system for prediction task
Pre-Pilot Results Now Available

datascience@nist.gov