DSE Plans for 2016

Pilot Evaluation

March 17-18, 2016
XDATA
Closed
Domain: Multiple

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

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

Full-Scale Evaluation
Multiple Tracks
Open to Everyone
Domain: Multiple
2016 Timeline

Evaluation Plan (final version)
- April 2nd

Test Data Available
- July 15th

Results Available
- August 31st

Core Data Available
- June 15th

Submissions deadline(s)
- August 15th
- August 22nd

Workshop
- TBA in October
Pilot Evaluation

- **Data**
  - Available in Amazon S3 Storage

- **Tasks**
  - Described in the Evaluation Plan

- **Submissions**
  - Upload result files at datascience.nist.gov
  - [beta] Submit systems to be run and benchmarked in a NIST private cloud
Data

Available through S3 Storage

• Traffic Detectors
• Reported traffic events
• Traffic camera videos

Available through Public URL

• OpenStreetMaps
• NOAA Weather Data
• US Census Data

(Optional) Any other public dataset
Tasks

Cleaning (Phase 1)
- Find and correct “dirty” traffic sensor data

Alignment (Phase 1 & Phase 2)
- Match traffic events in video to the reported event

Prediction (Phase 1 & Phase 2)
- Predict how many accidents happened in a location

Forecasting (Phase 1 & Phase 2)
- Predict the flow of traffic
Submission Flow

Phase 1
- Core Data incl.
- "Raw" traffic sensor data

Phase 2
- Core Data incl.
- "Raw" traffic sensor data
- "True" traffic sensor data
Submission Flow

Phase 1

Raw \[\xrightarrow{\text{Cleaning}}\] Cleaned

\[\xrightarrow{\text{Alignment}}\] \[\xrightarrow{\text{Prediction}}\] \[\xrightarrow{\text{Forecasting}}\]

Results \[\xrightarrow{\text{Submission}}\]

Phase 2

Truth \[\xrightarrow{\text{Alignment}}\] \[\xrightarrow{\text{Prediction}}\] \[\xrightarrow{\text{Forecasting}}\]

Results \[\xrightarrow{\text{Submission}}\]
Infrastructure

- Agree to evaluation and data terms
- Review tasks and evaluation plan
- Submit submission files
- View results

- Access core and test data
- Mount the S3 Storage to your EC2 instance(s)
- Download to your local cluster
What’s different from the Pre-Pilot

- Open Evaluation
- More Data (longer time span)
- Enforcing the 2-Phase process
- Metrics
  - **Cleaning**: A new alternate metric to not discourage more aggressive correction systems
  - **Alignment**: Higher cost to discourage false alarms
- Better, more intuitive submission formats
Questions & Feedback
Your Feedback

- Will you participate in the pilot?
- Which tasks will you undertake or *not* and why?
- Does the infrastructure described here meet what you would need to participate in the evaluation?
- Is the proposed timeline a good fit for your team? Where could we shift time around to make it better?
- Do you have proposed changes to task descriptions or metrics? Ideas for other tasks?