Introduction to NIST Pilot Evaluation

Dihong Gong
Dept. of CISE

Data Overview

Lane detector measurements
Sensor-measured traffic speeds and traffic flow values

Street Maps*
OpenStreetMap data with road maps and location labels

Traffic camera Video
Video feeds of traffic cameras on major highways

Traffic Events
Accidents, construction, roadwork, severe weather, and others

NOAA Weather*
Station sensor data and severe weather alerts

U.S. Census*
Census data with American Community Survey (ACS)

Data Overview

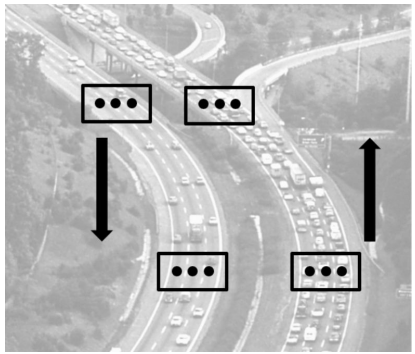
- Traffic Lane Detectors
- Traffic Events Listing
- Traffic Camera Video

Data Overview -- other public data

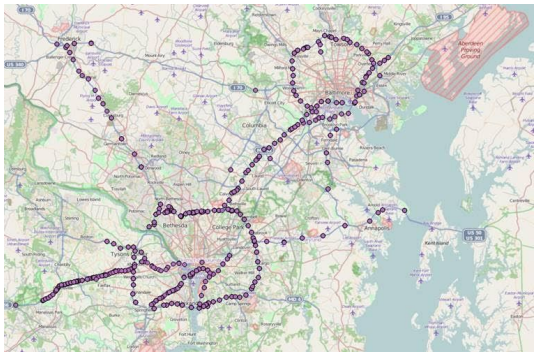
- OpenStreet Maps
- U.S. Census Data, Including American Community Survey (ACS)
- NOAA Integrated Surface Hourly Weather Data Sets
- Publicly Available Sets welcome
- **Note: Cleaning Task restricts allowable data**

Data Overview

- detector_lane_inventory.csv
  a. lane_id: uniquely identify a detector (totally 2,139).
  b. zone_id: identifier of a zone in a road (around 1,000).
  c. road: on which road, e.g. I-66.
  d. location_description: e.g. I-66 NEAR Sudley Rd @ MM 49.02
  e. Geographical coordinate: (latitude, longitude)
  f. There are 11 other less important fields.



lane and zone illustration (courtesy by NIST)



Detector distribution (courtesy by Sreten Cvetojevic)

Data Overview

- lane_measurement/*.tsv (totally 118 files with over 250GB)
  b. lane_id: identifier of a detector that this record is collected from.
  c. measurement_start: timestamp when measurement starts, e.g. 2007-04-09 14:04:12-04
  d. speed: measured average speed (mph) of the last interval, e.g. 70.
  e. flow: number of vehicles passed through the lane detector in the last interval, e.g. 9.
  f. occupancy: the average percent of time a vehicle was in front of detector in the last interval, e.g. 2.
  g. Two other less important fields: trial_id and quality.

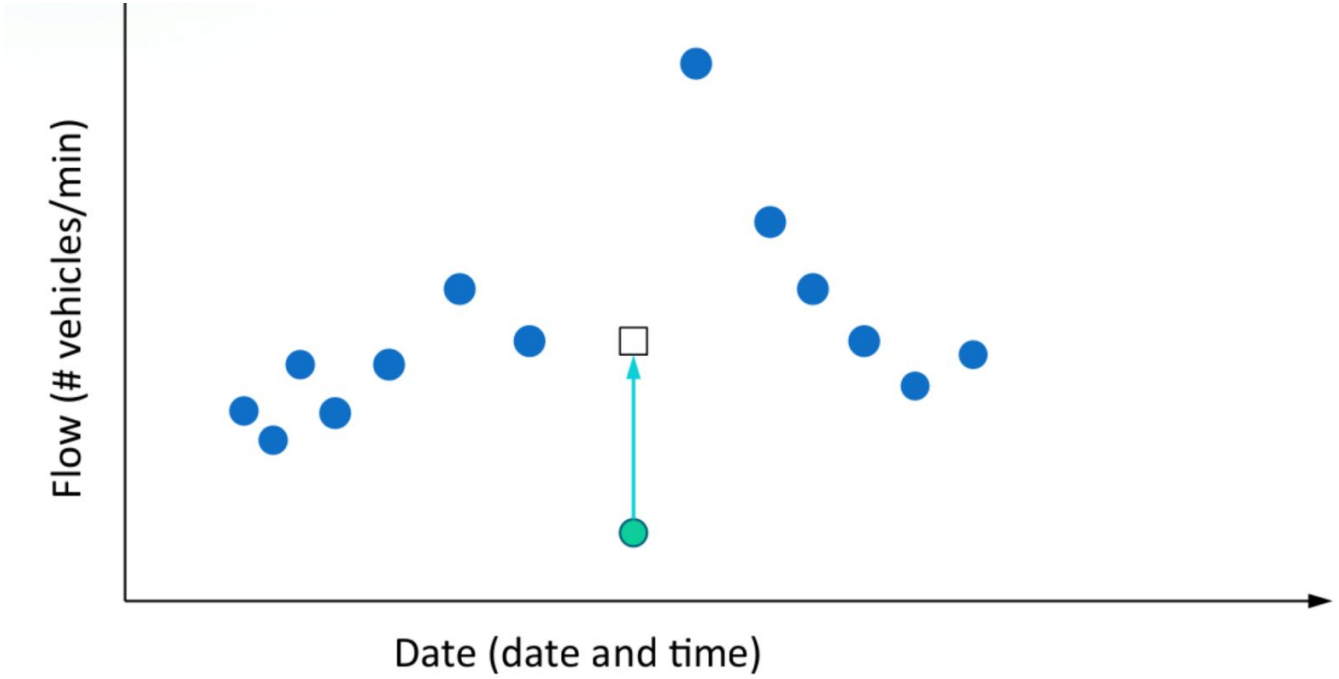
	A	B	C	D	E	F	G
1	trial_id	lane_id	measurement_start	speed	flow	occupancy	quality
2	c_06_09_000000000	12	2006-09-01T00:00:07-04:00	65	0	0	0
3	c_06_09_000000001	13	2006-09-01T00:00:07-04:00	63	3	2	0
4	c_06_09_000000002	14	2006-09-01T00:00:07-04:00	64	-2	1	0
5	c_06_09_000000003	15	2006-09-01T00:00:07-04:00	59	4	3	0
6	c_06_09_000000004	16	2006-09-01T00:00:07-04:00	66	5	1	0
7	c_06_09_000000005	17	2006-09-01T00:00:07-04:00	0	255	4	0
8	c_06_09_000000006	18	2006-09-01T00:00:07-04:00	67	13	7	0
9	c_06_09_000000007	19	2006-09-01T00:00:07-04:00	61	4	1	0
10	c_06_09_000000008	20	2006-09-01T00:00:07-04:00	65	0	0	0

***Note: The measurement values are mixed with noises.**

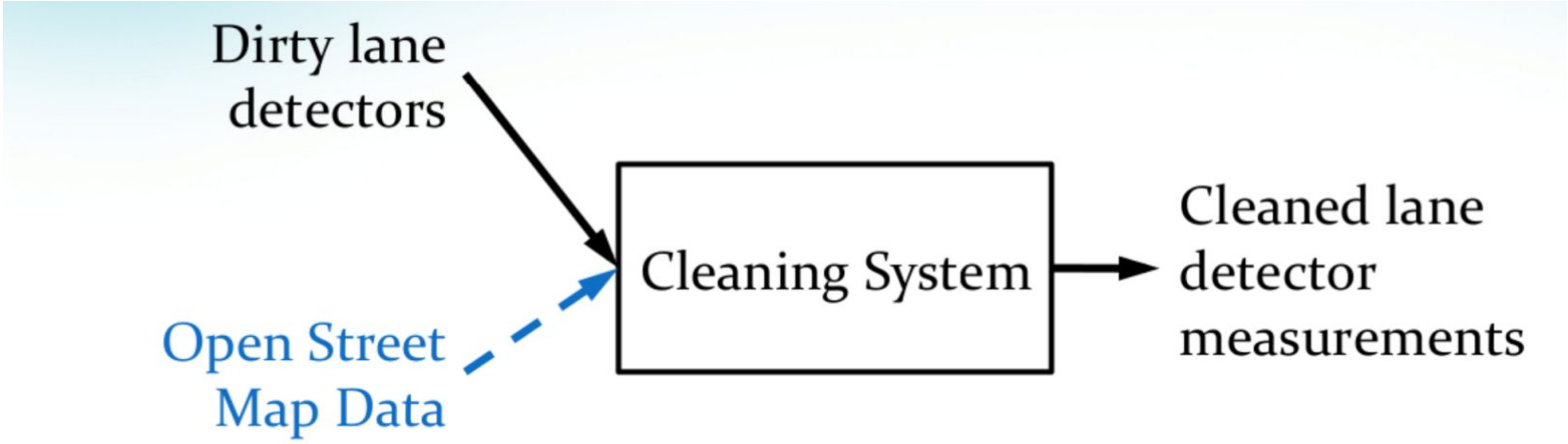
Data Overview

- events_train.csv
  a. event_id: uniquely identify an event, e.g. “MDOT_CHART_4aff02b300110095003f0be8b3035daa”
  b. event_description: a text description about an event, e.g. “Disabled Vehicle Event @ I-495 AT MD 187”
  c. Timestamps: times the event was created, confirmed, and closed (some are missing).
  d. event_type: the type of an event, e.g. “accidentsAndIncidents”.
  e. geographical location: (latitude, longitude)
  f. There are 9 other less important fields.

Cleaning (1)



Cleaning (2)



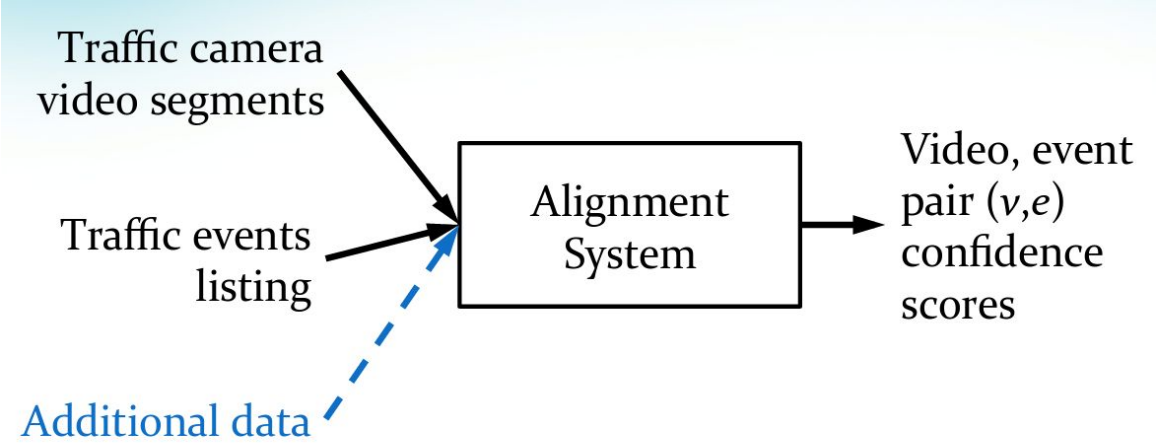
Metric (cost) = Mean Absolute Error in Flow

Alignment (1)

MDOT_CHART_a000a2de48a600680055fa2ec4235c0a,"Incid
ent @ I-495 INNER LOOP … **90%**

MDOT_CHART_400122fc49a700680055fa2ec4235c0a,"Incide
nt @ I-495 OUTER LOOP … **3%**

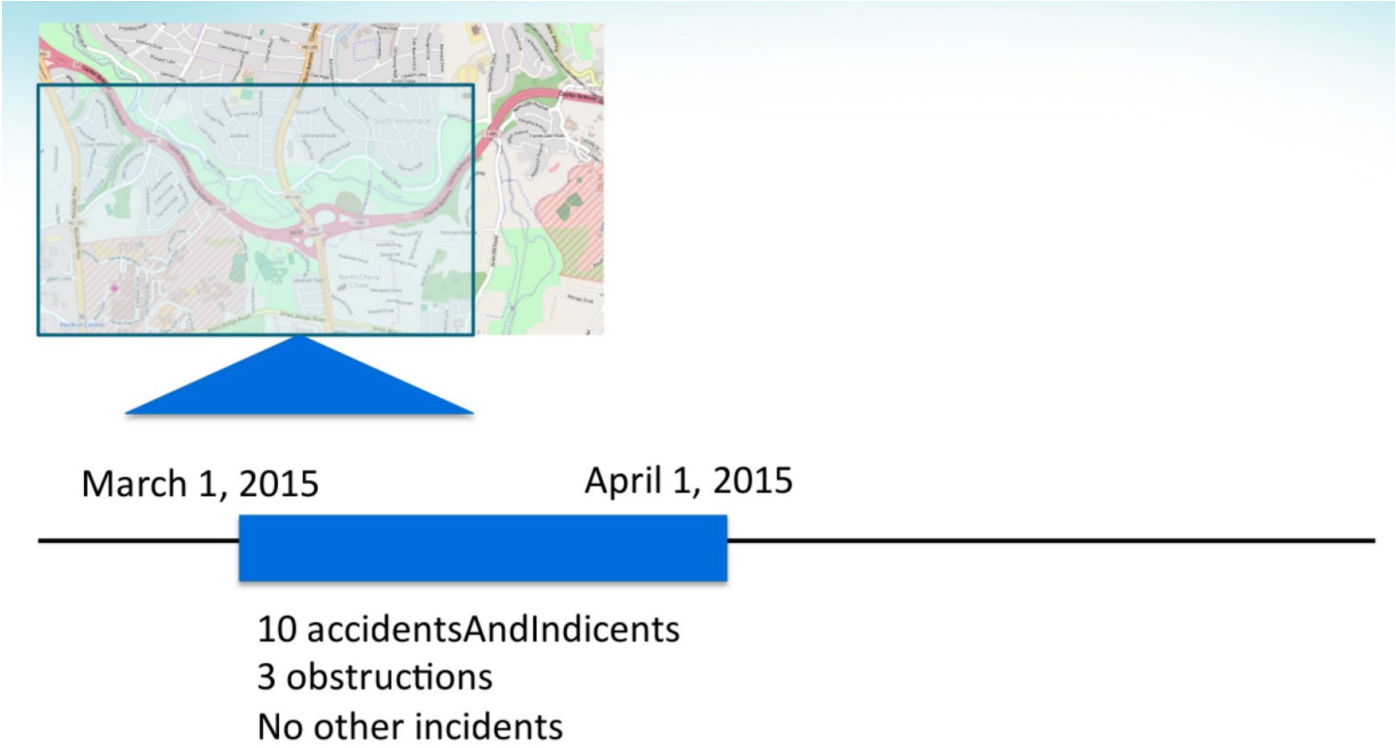
Alignment (2)



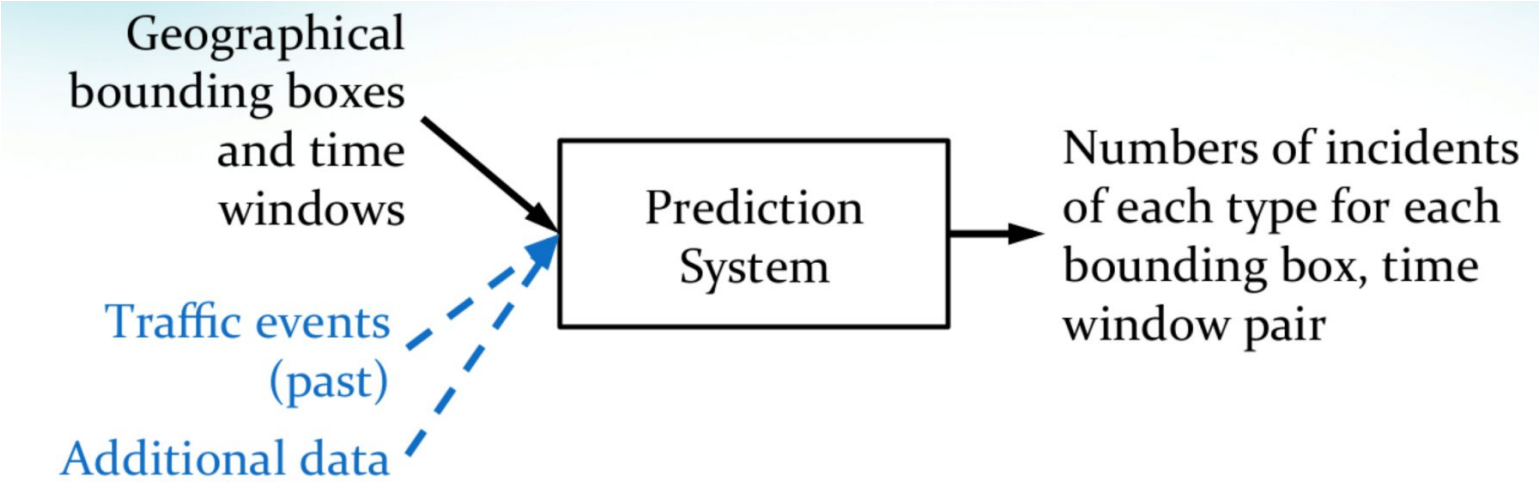
Metric (cost) = (*Miss Rate*) + 10*(*False Alarm Rate*)

Decision threshold selected to **minimize** metric

Prediction (1)

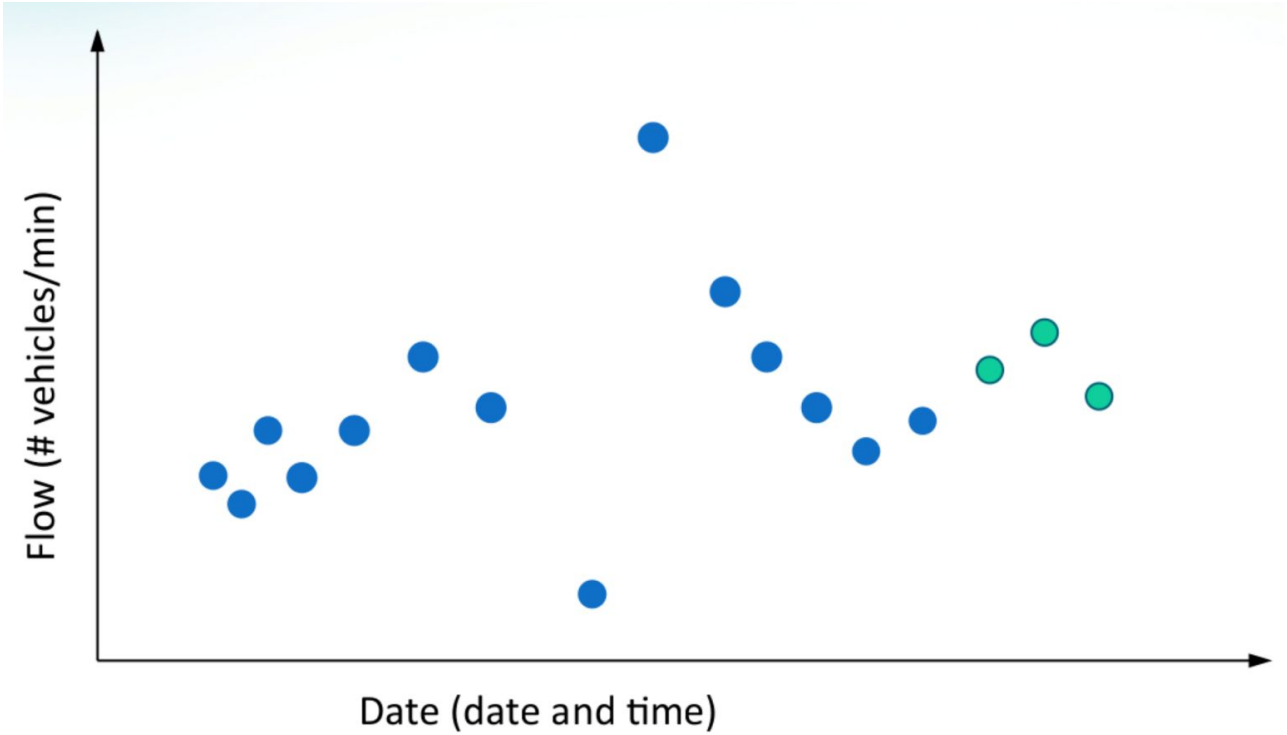


Prediction (2)

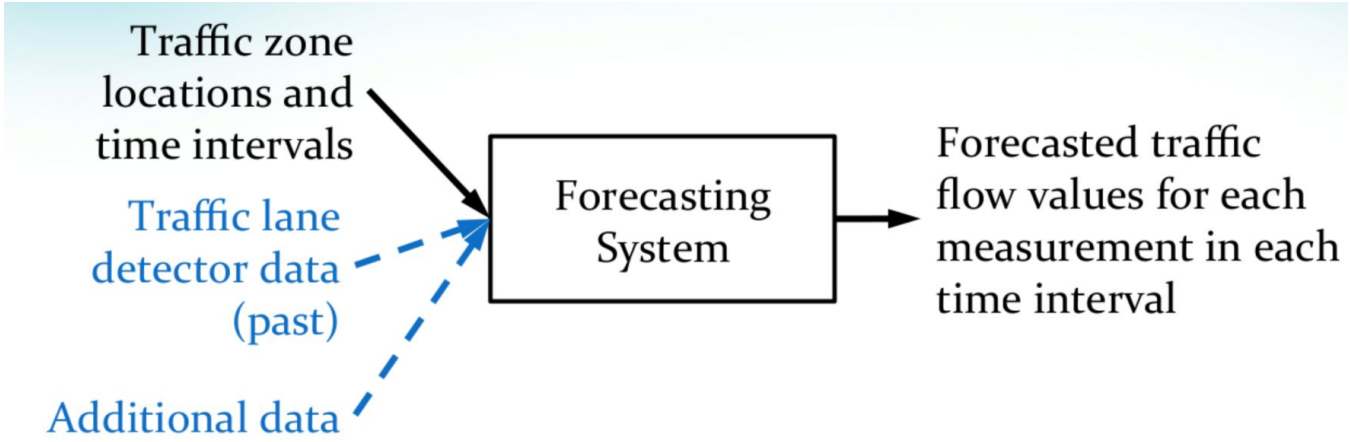


Metric (cost) = Mean Root Mean Square Error

Forecasting (1)



Forecasting (2)



Note: here, traffic lane detectors are aggregated traffic into **traffic zones**.
Metric (cost) = Mean of Mean Absolute Percentage Error

Final Project Tasks Preview

1. **Cleaning Task**
    a. clean traffic lane detector measurements containing incorrect flow values, providing correct traffic flow values for the erroneous traffic flow measurements.
2. Alignment Task
    a. analyze video from camera feeds to detect an event and match it to a separate inventory of traffic events (disabled car, accidents, etc).
3. **Prediction Task**
    a. develop a system that can predict the number and types of traffic events by type for a given (geographical bounding, interval of time) pair.
4. Forecasting Task
    a. leverage past traffic information and current conditions (weather, maps) to forecast vehicle flows on major roads.

* We will only participate in task #1 and #3.