Tackling Congestion Using Connected Car Data

Project Description

**Project goal:** Activate GM's connected car data in order to showcase the dataset and interesting insights generated from it. Leverage the dataset to systematically support GM's brand promise of zero congestion.

**Problem Statement:** Develop the first GM definition and congestion and use GM vehicle telemetry data to predict future congestion.

**Data**

- GM high speed vehicle telemetry (HSVT) data
- Open source map data
- Open source weather data

- Logs data every 3 seconds along all GM connected car trips
- A trip is defined as ignition on to ignition off
- Includes anonymized IDs, timestamp, latitude, longitude, speed and direction
- Over 180B rows: Required specialized tools, indexing, and pipelines

**Modeling**

**Modeling Scope:** We modelled 7 freeways in the Detroit metropolitan area, training and testing on 2.6B rows of data from January and February 2019

**Methodology:** There are different types of congestion, each of which have different causes, effects and symptoms. In order to predict congestion we separated three types of congestion:

1. **Periodic:** Expected congestion that always occurs at certain times. Ex: rush hour
2. **Irregular:** Unusual congestion but predictable in advance. Ex: construction or snow storm
3. **Event-based:** Chaos events that cannot be predicted in advance Ex: fender bender

In order to capture all three types of congestion, we build a three level model. Each level builds the model’s strength by predicting the difference between the previous levels prediction and the true observed congestion score. The first level can be predicted weeks in advance, the second layer one hour in advance and the third layer ten minutes in advance.

**Level 1: Periodic**

- Model: Mean
  - The first layer of the model captured periodic congestion through historical mean congestion by time of day and location.

**Level 2: Irregular**

- Model: Ridge regression
  - The second layer of the model captures irregular congestion patterns using features engineered to capture their causes and symptoms: snow storms, construction, or unusually high or low numbers of cars on the road. Features include:
    - Recent congestion trends
    - Historical road characteristics
    - Geographic
    - Weather
    - Temporal
    - Non-linear combinations

**Level 3: Event-based**

- Model: Vector Autoregression
  - The third layer models captures how waves of congestion move through the system using multi-dimensional auto-correlation and historical data. VAR uses congestion from the past 30 minutes on the roadway to predict forward 10min

**Impact**

**Activating Dataset:**
GM had not used the HSVT table for tackling congestion. We built and reproducible processes and pipelines

**Congestion Insight:**
We built GM internal congestion modeling expertise. GM could use the expertise to create a real-time congestion app or inform city planners.

**Interest creation:**
GM cannot currently leverage data for live intervention. We met with decision makers and generated interest in investing in needed technologies.

**Deliverables**

- Systematic process for calculating congestion
- Predictive model of congestion
- Visualizations of connected car data and congestion implementation
- Insights into data quality

**Next Steps**

- Make HSVT table update in real time
- Automatically route HSVT trips to roads
- Calculate congestion score live as data is gathered