Data-driven time-dependent freight volume estimation

Chrysovalantis Anastasiou, Luciano Nocera, Cyrus Shahabi
Integrated Media Systems Center
Viterbi School of Engineering
University of Southern California
Motivation

- **Urban planning**
  - Introduce new lanes/roads where trucks over-congest the network
  - Reinforce or more frequently maintain roads that are more likely to be damaged by trucks

- **Air quality**
  - Effect of trucks on air pollution in areas they frequently pass by or drive to
Ideal Scenario -- GPS Tracking

Infeasible due to many reasons, including privacy...
Reality -- Discrete Sensor Observations

Maybe large truck at time $t_2$.

Maybe small truck at time $t_2$.

Maybe medium truck at time $t_3$.

Maybe small truck at time $t_1$.

No idea what is going on at time $t_{\text{any}}$.

One large truck at time $t_2$.

One small truck at time $t_0$. 
Outline

● Motivation
● Problem Statement
● Data Sources
● Algorithms
  ○ Baseline
  ○ Naive / FlowPath
  ○ Reachability-based
● Experiments
Problem Statement

Given a region of interest $R$, its road network $G$, and a sensor-based dataset $\Theta$, estimate the volume of truck movements per time unit (e.g., 1 hour).

Research goal:
To accurately estimate the time-dependent flow of trucks in a road network.
**Sensor Data**

- **RFID sensors** *(very accurate)*
  - Typically at port exits; truck “check-out”
  - `<location, timestamp, truck type> + <truck id>`
  - Refers to a specific truck id
- **Weigh-in-motion (WIM) sensors** *(very accurate)*
  - Sparse but provide checkpoints
  - `<location, timestamp, truck type> + <truck id>`
- **TAMS [1] sensors** *(accurate)*
  - Sparse and probabilistic
  - `<location, timestamp, truck type prob.>`
- **CCTV cameras** *(variable accuracy)*
  - Sparse and probabilistic
  - `<location, timestamp, truck type prob.>`
- **Inductive Loop Detectors**
  - ADMS [2]

---


Sensor Observation Examples

Sensor Observations

<table>
<thead>
<tr>
<th>Sensor</th>
<th>Timestamp</th>
<th>Truck Class Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Small</td>
</tr>
<tr>
<td>$s_1$</td>
<td>12:48pm</td>
<td>0.04</td>
</tr>
<tr>
<td>$s_2$</td>
<td>12:59pm</td>
<td>0.07</td>
</tr>
<tr>
<td>$s_3$</td>
<td>1:15pm</td>
<td>0.05</td>
</tr>
</tbody>
</table>
Truck Flow Estimation

- **Input**
  - $G = (V, E)$ the road network
  - $S$ the set of sensors
  - $\Theta$ the set of sensor observations

- **Output**
  - $T$ the set of truck flow time-series; one per edge/road segment
Baseline Algorithm

- Counts the number of trucks on the sensor’s road segment

Legend
- Intersection
- Road Segment
- Sensor
- Flow Segment

Very sparse estimation → Low recall
Naive Flow Path Expansion

- Counts the number of trucks on the sensor’s road segment.
- Expands backwards and forwards as long as intersection does not affect flow count.
Naive Flow Path Expansion

- Counts the number of trucks on the sensor’s road segment.
- Expands backwards and forwards as long as intersection does not affect flow count.

Legend:
- **Intersection**
- **Road Segment**
- **Sensor**
- **Flow Segment**

Somewhat denser estimation → Better recall
Reachability-based Estimation

- Counts the number of trucks on the sensor’s road segment.
- Expands backwards and forwards as long as intersection does not affect flow count.
- Propagates flow if observation in next sensor is reachable from previous.
  - Requires time-dependent traffic data
Reachability-based Estimation

- Counts the number of trucks on the sensor’s road segment.
- Expands backwards and forwards as long as intersection does not affect flow count.
- Propagates flow if observation in next sensor is reachable from previous.
  - Requires time-dependent traffic data

Legend
- Intersection
- Road Segment
- Sensor
- Flow Segment

<table>
<thead>
<tr>
<th>Time</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>12:13pm</td>
<td>+2</td>
</tr>
<tr>
<td>12:14pm</td>
<td>+2</td>
</tr>
<tr>
<td>~24 minutes</td>
<td></td>
</tr>
<tr>
<td>12:38pm</td>
<td>+2</td>
</tr>
<tr>
<td>12:52pm</td>
<td>+2</td>
</tr>
</tbody>
</table>

I-NUF’22
9th International Urban Freight Conference
 Reachability-based Estimation

- Counts the number of trucks on the sensor’s road segment.
- Expands backwards and forwards as long as intersection does not affect flow count.
- Propagates flow if observation in next sensor is reachable from previous.
  - Requires time-dependent traffic data

**Legend**
- Intersection
- Road Segment
- Sensor
- Flow Segment

Denser estimation → Higher recall
Risk of false positives → Lower precision
Experimental Setup

● Datasets
  ○ SYNT(S, T): Synthetic datasets with S sensors and T trucks
    ■ S = {100, 150, 200, 250, 300}
    ■ T = {250, 500, 750, 1000, 5000}
    ■ “Simulates” truck trajectories and generate sensor observations

● Algorithms
  ○ **Baseline**: Only estimates at edges where data is sensed.
  ○ **FlowPath**: Extrapolates flow based on logic.
  ○ **Reachability-based**

● Metrics
  ○ **Precision**: Percentage of graph edges in estimation that exist in ground truth
  ○ **Recall**: Percentage of graph edges in ground truth that are in estimation
  ○ **MAE**: Mean Absolute Error of flow estimation
  ○ **MAPE**: Mean Absolute Percentage Error

\[
\text{Precision} = \frac{|\text{GroundTruth} \cap \text{Estimation}|}{|\text{Estimation}|} \quad \text{Recall} = \frac{|\text{GroundTruth} \cap \text{Estimation}|}{|\text{GroundTruth}|}
\]
Experimental Results

- 300 sensors, 1000 trucks

Ground Truth

Edges: 888

Baseline

Edges: 94 (TP/94, FP/0)
Precision: 100%
Recall: 10%

MAE: 7.464
MSE: 170.651
RMSE: 13.063
MAPE: 89.22%

Naive Flow Path

Edges: 334 (TP/327, FP/7)
Precision: 98%
Recall: 37%

MAE: 4.858
MSE: 104.845
RMSE: 10.239
MAPE: 63.27%

Reachability-based

Edges: 702 (TP/666, FP/36)
Precision: 95%
Recall: 75%

MAE: 3.854
MSE: 72.999
RMSE: 8.544
MAPE: 40.08%

7x

2x
**Experimental Results**

- **200 sensors, 1000 trucks**

**Ground Truth**

- Edges: 888

**Baseline**

- Edges: 63 (TP/63, FP/0)
- Precision: 100%
- Recall: 7%
- MAE: 7.807
- MSE: 177.910
- RMSE: 13.338
- MAPE: 92.79%

**Naive Flow Path**

- Edges: 226 (TP/219, FP/7)
- Precision: 97%
- Recall: 25%
- MAE: 6.065
- MSE: 130.300
- RMSE: 11.415
- MAPE: 75.28%

**Reachability-based**

- Edges: 601 (TP/564, FP/37)
- Precision: 94%
- Recall: 63%
- MAE: 4.000
- MSE: 90.450
- RMSE: 9.510
- MAPE: 51.20%

- 9x
- 1.5x
Experimental Results
Summary & Future Work

- Critical for planners and decision makers to understand freight flow
- Estimating the volume of trucks from sensor data is feasible
- Reachability-based approach yields more accurate results
  - 9x higher precision compared to the baseline
  - 2x improvement in MAE

Future work

- Improve computational efficiency and accuracy of algorithm
- Validate approach with real-world data
- Infrastructure optimization
  - where should the next sensor be installed in order to improve accuracy?