Mobility21

A USDOT NATIONAL
UNIVERSITY TRANSPORTATION CENTER

Carnegie Mellon University

Penn
University of Pennsylvania

The Ohio State University

CCAC

Modeling and Enhancing Freight Mobility in the Philadelphia Region

Sean Qian
Carnegie Mellon University

https://orcid.org/0000-0001-8716-8989

Research Assistant: Arnav Choudhry

FINAL RESEARCH REPORT

Contract # 69A3551747111

Modeling and Enhancing Freight Mobility in the Philadelphia Region

PI: Sean Qian (ORCID: 0000-0001-8716-8989)

Research Assistant: Arnav Choudhry

DISCLAIMER

Description of the problem

The Philadelphia region has a large and complex freight transportation network that includes more than 1,000 miles of the National Highway System and 9.8 million vehicle-miles of daily truck travel. The mobility of commercial trucks and the efficiency of freight infrastructure are essential to regional transportation infrastructure planning and economic development. Unfortunately, characteristics of freight demand, such as when and how trucks travel, freight destinations and truck routing behavior, are unclear. This becomes the main hurdle for improving truck mobility. In addition, how roadway construction projects would impact freight mobility is unknown. For instance, during the reconstruction of the I-95 corridor in the region, accommodation of truck demand, including planning of construction projects and detour strategies for truck demand, is of critical importance to freight mobility, and ultimately to the economic prosperity. Accurate modeling of freight transportation, on both demand and infrastructure, is needed to support those activities. Emerging truck data provides a unique opportunity to learn, estimate and predict truck demand, and furthermore to establish a regional freight transportation model to support regional planning and development.

Delaware Valley Regional Planning Commission's (DVRPC) existing freight demand model is limited to modeling light and heavy truck trips within the context of a traditional 4-step model. This model is based on a truck survey conducted in 2000 that determines truck trip generation rates and trip length frequency distributions. Vehicle classification traffic counts on several typical weekdays were collected and used to calibrate the model. While the existing model is able to reasonably predict the average truck demand at the level of traffic analysis zones (TAZs), it is unable to encapsulate truck flow in high spatio-temporal granularity. Given sparse (and sometimes synthesized) data, the model is also unable to predict the mobility of truck demand induced by 'what-if scenarios', such as roadway construction, new freight terminals and land-use change.

Main modeling challenges:

- Extensive amounts of traffic data for both cars and trucks has become available, but how to integrate them into truck demand modeling and understand truck demand is unclear.
- Traffic congestion is endogenous to truck routing and operations, which in return considerably affects traffic congestion. A challenge lies in modeling of the mixture of truck flow and car flow that can be validated by real-world data.
- Modeling trucks in the DVRPC regional network raise significant challenges to account for high spatio-temporal granularity, while ensuring computational tractability.

Objectives: This research project analyzes freight transportation data and enhance the regional freight transportation model to better estimate and forecast truck travel in the Philadelphia region. In particular, this project infers truck activity from GPS traces data (acquired by DVRPC). Identifying truck activity in this case refers to identifying stops made by the truck, identifying hubs of operations (if they exist), and identifying tours and trips made by the truck. In an effort to achieve the above stated goal, two additional sub-activities were undertaken. First, developing device matching/chaining methods to deal with the issue of several device identifiers resetting at midnight. Second, developing clustering methods for stops to remove false positive stops and infer hubs of operations.

Approach

Heuristic methods were used to identify stops from the GPS events, as well as chain devices and find clusters. For a particular GPS event, the method uses the speed to the next GPS event and the speed from the previous GPS event to find whether that particular GPS event is a stop or not. The chaining algorithm looks at various rules based on features such as direction of travel and speed to match two devices. The clustering algorithm used pre-identified microzones, or a type of agglomerative clustering with maximum diameter of cluster.

Methodology

Data Description

The data was provided by INRIX and contained GPS traces for Consumer Vehicles, Field Service/Local Delivery Fleets, and For hire/private trucking fleets across Low, Medium, Heavy weight categories. The data is for a duration of four weeks, for the weeks of 01/21, 04/22, 07/15 and 10/14 in 2018. A subset of the information (this was the information used in the project) contained in each row is as follows:

- CaptureDate: The capture date and time of the waypoint in UTC, ISO-8601 format
- Latitude: The decimal degree latitude coordinates of the waypoint
- Longitude: The decimal degree longitude coordinates of the waypoint
- DeviceId: A device's unique identifier

This information was augmented by adding provider information, weight class information as well as land use information. The frequency of this information is variable.

Stop identification

The first task was to identify the stops from all the GPS events. In order to do this, the "speed from previous GPS event", let's call it *a*, and "speed to next GPS event", let's call it *b*, was calculated for every GPS event. A speed threshold was then fixed, such as 5 mph, and each GPS event was classified as either starting, stopping, moving or stopped, based on the rules below:

1. If a > 5, b > 5, then the GPS event is "moving"
2. If a < 5, b > 5, then the GPS event is "starting"
3. If a > 5, b < 5, then the GPS event is "stopping"
4. If a < 5, b <5, then the GPS event is "stopped"

Among these, we were interested in just the events classified as "stopping" and label them as stops for a device. The GPS events classified as "starting" were used to find the time a device spent at a stop as well as the time a device took to reach an inferred stop, and then dropped. GPS events classified as "moving" or "stopped" were just dropped from the analysis. On the inferred stops, a stop duration threshold was applied to remove false positive stops such as traffic stops. A variety of speed and stop duration thresholds were tested such as 3 mph, 4 mph, 5 mph, 6 mph, 7 mph and 5 minutes, 10 minutes, 15 minutes respectively.

Device matching

Device matching algorithms match each eligible device to one of the candidate devices. An eligible device is a device whose device ID is suspected to have been reset. Candidate devices are devices with device IDs that an eligible device's device ID might have reset to. Given an eligible device, the following steps are undertaken to find the right candidate device:

1. A list of candidate devices is constructed such that the first time of transmission of the candidate device is within 15 minutes after the last time of transmission of the eligible device.
2. The candidate devices are filtered such that the distance between the first point of transmission of the candidate device and last point of transmission of the eligible device is less than 37,000 meters.
3. The candidate devices are filtered such that the direction of travel of the candidate device is within a 90-degree cone of the direction of travel of the eligible device.
4. The candidate devices are filtered such that the weight class of the candidate device is the same as the weight class of the eligible device.
5. The candidate devices are filtered such that the implied speed, which is calculated using the distance calculated from the first point of transmission of the candidate device and last point of transmission of the eligible device divided by the time between these two events, is less than 60 miles per hour.
6. The candidate devices are filtered such that the implied speed is within 10 miles per hour of the inferred speed of the candidate device, calculated by looking at the first few GPS events of the candidate device.

Finally, the candidate device which is found to be the closest to the eligible device after applying all the filters is matched. If there are multiple candidates at the same minimum distance, then there is no matching as we wanted the matching to be conservative. All the numbers in the rules above can be changed and were found after trying out different ranges of values.

Stop clustering

Two methods were tried out for the clustering methodology. In the first method, a predetermined map of *microzones* was used to associate each stop with a “MAZ ID”. Microzones are smaller subdivisions of Traffic Analysis Zones (TAZs). Consecutive stops with the same MAZ ID were then supposed to belong to the same cluster, and aggregated into one stop with a stop duration equal to the sum of stop durations and travel durations for all the stops in that cluster. In the second method, each stop is considered to be an individual cluster at first, and then the clusters are merged if they are within some distance *d* of each other. A variety of values of *d* such as 500 feet, 1000 feet, 1500 feet and 1 mile were considered.

Truck activity identification

Finally, given all the information for stops we were able to identify trips and tours for the devices. A trip is made by a truck/device going from point A to point B. A tour is a collection of trips, where a truck/device returns to the hub. Finding tours is trivial, as it is just going to the next stop. For finding the tour, we need to find the hub. For finding the hub, we employed two strategies. The first strategy was to just just designate the first stop a device starts transmitting information from as the hub. But it suffers from obvious drawbacks such as what if the device ID is reset and the device starts transmitting from the middle of a trip. For that reason, the second strategy we used to find hubs was to find the stop a device visited most frequently and designate that as the hub. In cases that there is no mode for the stops visited by a device, we default to the first strategy.

Findings

#stops	Speed threshold		
stop duration	3 mph	5 mph	7 mph
<5 min	10,185	11,739	12,435
>=5 min	12,641	16,161	20,258
Total	22,826	27,900	32,693

stops %	Speed threshold		
stop duration	3 mph	5 mph	7 mph
<5 min	21%	24%	26%
>=5 min	8%	11%	14%
Total	11%	14%	17%

Table 1: Number of stops on the highway and their proportion with respect to all stops

For stop identification, we found that a higher speed threshold was more tolerant to small speed changes at starting or stopping, like moving within a parking lot. However, a higher speed threshold was also found to be less tolerant to congestion/stop and go situations, and thereby resulted in more false stops on the highway, associated with land use category 4010 (LU4010). Table 1 shows the number of stops at LU4010. Hence, we decided to use a lower threshold of 3 mph with a higher value of stop duration threshold to identify stops.

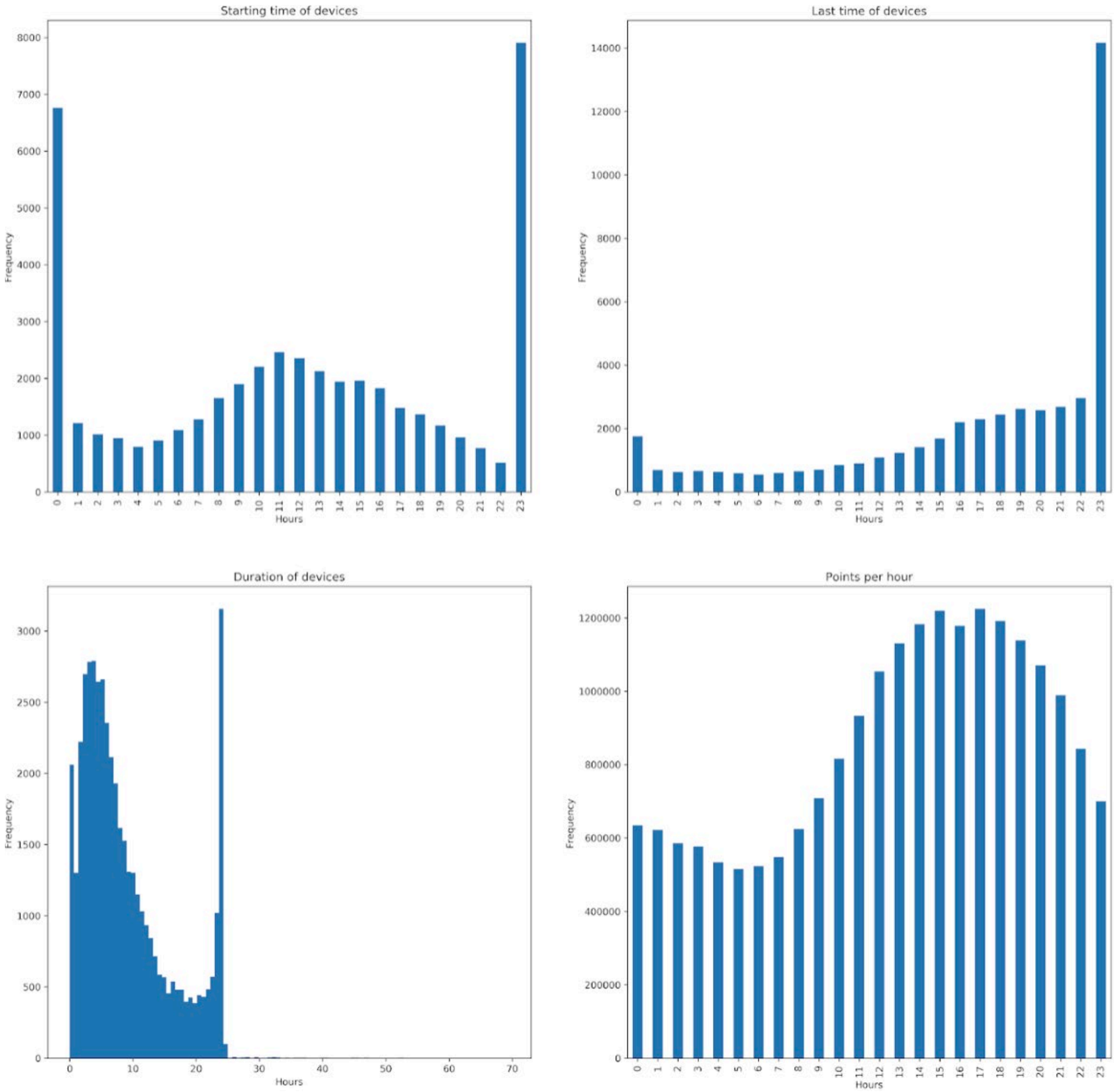

Figure 1: Summary statistics for provider ‘45c48cce2e2d7fbdea1afc51c7c6ad26’

For device matching, INRIX had informed us that they did not know the device resetting rules, but the rules are consistent within a provider. We found that only one provider in fifteen had a device resetting that we could safely deduce. Figure 1 shows the summary statistics of the GPS traces from that provider. There was a clear trend that a lot of devices end transmitting information at midnight and have a transmission duration of 24 hours. Therefore we only performed device matching for the one provider, for devices which end transmission between 11:45:00 PM to 11:59:59 PM using the rules outlined in the “Methodology” section. Table 2 shows a summary of results for data from the week of 01/21.

Total devices	Ineligible devices	Eligible devices	Devices with no Match	Devices with Multiples matches	Devices with One match	Match Rate
46,621	35,747	10,874	3,317	17	7,540	69%

Table 2: summary of results for data from the week of 01/21

For clustering we concluded that the MAZ-based clustering works best since the median MAZ is roughly the size of 850 meters by 850 meters which means we're not grouping together stops too far apart, and also MAZs have a physical interpretation based on actual land usage. Which means that a distance based cluster may cluster two points on either side of the highway as long as they are less than 1000 feet apart, but an MAZ would not extend over the highway and hence not group these stops together, which is the desirable behavior. For the week of 01/21 we found 25,455 unique MAZ associated with stops, which is 10.8% of total MAZs.

Conclusions

This research project analyzes freight transportation data and enhance the regional freight transportation model to better estimate and forecast truck travel in the Philadelphia region. In particular, this project infers truck activity from GPS traces data (acquired by DVRPC). Identifying truck activity in this case refers to identifying stops made by the truck, identifying hubs of operations (if they exist), and identifying tours and trips made by the truck. In an effort to achieve the above stated goal, two additional sub-activities were undertaken. First, developing device matching/chaining methods to deal with the issue of several device identifiers resetting at midnight. Second, developing clustering methods for stops to remove false positive stops and infer hubs of operations. Finally, files containing the device/truck activity information were created. The files can be found here. The associated metadata is posted in the file 'stop_metadata.txt' here.