Uber Movement:
Speeds Calculation Methodology

Abstract

Uber Movement Speeds provides aggregated speeds by street segments at hourly granularity. Street segments can be mapped to OpenStreetMap (OSM) Way IDs or the SharedStreets Referencing System. This solution is particularly powerful in delivering accurate data where it is most scarce: smaller streets and arterials with limited commercial traffic and where fixed sensor infrastructure is a costly and inefficient investment.

This document provides a comprehensive description of how Movement Speeds data available on movement.uber.com is calculated. We describe how we use anonymized and aggregated location data from the Uber Driver app to measure the speeds on street segments. We believe that the transparency about how we calculate the speeds data is required for the data to be well understood and thus, correctly and effectively used by our users.

Overview

Figure 1 illustrates the steps used in calculation of Movement speed data:

1. Input data (GPS location and Map Data) ingestion
2. Map matching
3. Calculation of traversals
4. Speed aggregation

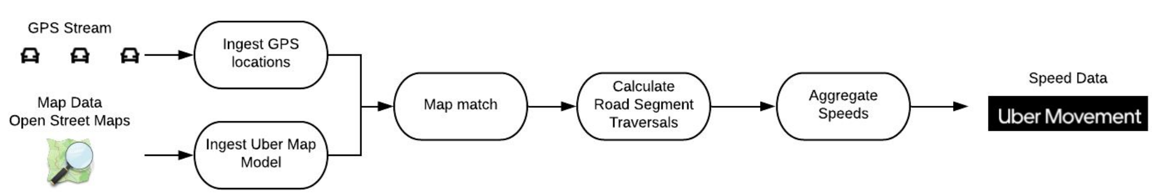

Figure 1. The Overview of Speed Data Calculation Methodology

We describe each step in the following sections.

Input Data

In order to calculate speeds we use two data inputs: (a) GPS locations of vehicles over time, and (b) map data that represents the street network on which vehicles travel.

GPS Data

The Uber Driver-partner app records a location entry every 1 or 2 seconds that include latitude, longitude, speed, course, and a timestamp (date / time) of the GPS location ping. These GPS location pings are ingested in real-time (every 4 seconds) to power multiple Uber business products (e.g. turn-by-turn navigation for driver-partners, fare calculation, matching driver-partners with riders, as well as user experience elements, such as displaying the position of the car in the Uber Rider app). The GPS location data is also stored for offline processing, and when aggregated, can be used to derive average, median, and percentile speed data on any given street segment where there is sufficient data. Please note that the number and quality of GPS location data impacts the quality of speed data that we are able to derive on a given street segment.

Map Data

We rely on a 3rd party map data provider, Open Street Map (OSM), which is commonly used by the research community and transportation industry. Uber ingests snapshots of OSM data regularly and transforms them to an internal map data format consisting of street segments (edges) connected via intersections (nodes). Eventually, each map feature (segments, intersections) has a unique Movement identifier (ID) that allows us to reliably point to it. To facilitate the usage and interoperability of Movement Speeds Data with other data sets, we provide a mapping between Movement Segment IDs to OSM Way IDs as well as a set of tools that can be downloaded on the Movement Download page.

Movement strives to use relatively recent versions of OSM as data are ingested (list of OSM builds can be found on planet.openstreetmap.org/).

Movement presents speed data over a longer historical period of time (ultimately over multiple years). The challenge here is that the map data has a high probability to change over such a long time. The map data changes can happen for a variety of reasons:

- Real world changes - Cities carry out infrastructure projects that add, change, or remove roads, intersections, etc
- Improved coverage/accuracy - OSM contributors add roads that were previously unmapped, fix intersections, update traffic direction (one-way vs two-way), etc.
- Errors - In open source maps such as OSM, erroneous edits are sometimes made and then reverted at a later date.

Movement uses the most recent map data version to show the historical speeds. This means that a particular street segment that existed historically, but not anymore, may not be shown on the Movement map. To learn how we approached the challenge of changing map data see Map Conflation and Speed Rollup.

Map Matching

Having ingested GPS location and street map data, the next step is to assign (or map match) each location data point (latitude, longitude) to a location on a street segment (latitude1, longitude1). Due to limited GPS position accuracy, GPS points often won't align exactly with street segments. We use a map matching algorithm based on a hidden markov chain (HMM) model to assign a raw GPS location ping to the street segment with highest probability. Figure 2 illustrates how this matching works. To learn more about the algorithm, see our talk about map matching.

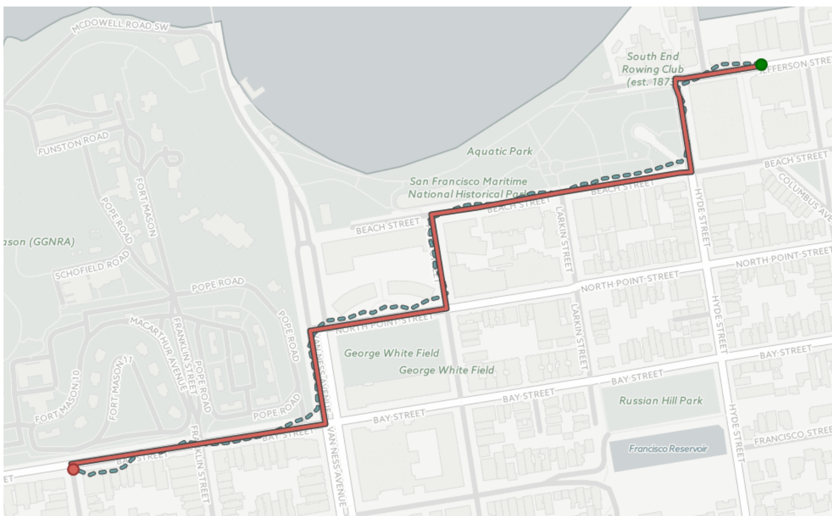

Figure 2. Visualization of map matching. The dotted line presents raw GPS location points, and the red line presents a polyline created from map matched points.

Traversal Speed Calculation

The map matched location data is then used to calculate traversal speed per segment. We use anonymized and truncated trip traces (e.g. we do not include road segments where a driver stopped for pickup or dropoff. Please read more about how we protect the privacy of Uber trips in the Privacy section).

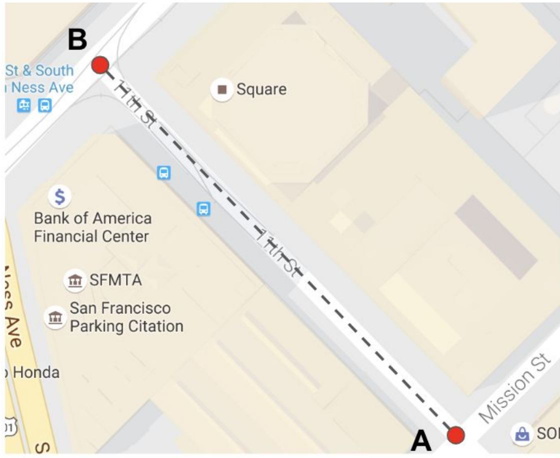

Figure 3. Illustration of traversal calculation. Red points A and B are map matched locations at the endpoints of the indicated segment of 11th St.

Figure 3 illustrates the calculation of travel time on an example street segment. First, the time difference is calculated between when the driver enters the street segment (it's usually a map matched location at the intersection) and when the driver leaves the segment (red points A and B, respectively). Speed is consequently calculated as the distance of the segment (provided in the map data) divided by the time. Please note that if there is a red traffic light at the exit of the street segment (or a stop sign with a queue of cars), the time spent waiting at the exit will be included into traversal time. In this step we apply filters to ensure high quality traversal data. For example, we filter traversals on very short segments, or those with anomalous time stamps (which can indicate problems with upstream location data or map matching algorithm).

Speed aggregation

In this step, the traversals for each segment are initially aggregated into time windows -- for example, the average speed on segment 1234 on 3/19/2018 between 3:00 pm and 3:05 pm was 35 (k)mph. This allows us to re-aggregate the data into time+day blocks that are useful for Movement users. For instance, in Movement, we aggregate speeds per hour, time of day (e.g. AM Peak 7AM-10AM, Midday 10AM-4PM), and quarter (e.g. Jan 1st, 2018 - March 31st 2018). Please note that in order to safeguard rider privacy, we do not publish the speeds on segments where the number of traversals is below a sufficient threshold.

Map Conflation and Speed Rollup

As explained in the Map Data section, the road segment identifiers (IDs) may change over time due to a variety of reasons (adding, removing roads, or simply OSM editing mistakes). Figure 4 illustrates an example when a newly added road segment split the existing segment into two segments with new IDs.

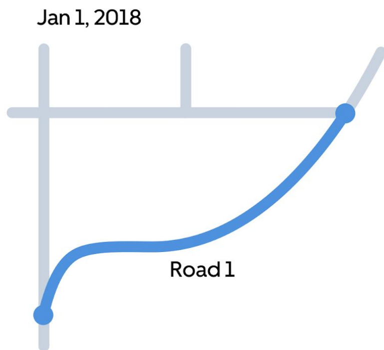

Figure 4. **Road 2** is added to the road network, splitting **Road 1** into **Road 3** and **Road 4**.

The changing IDs can cause confusion for our users when they select a historical time range and don't see speed values for a road segment that happens to have a newer ID. To address this problem we periodically perform map data conflation, described below, which allows us to roll-up speed values to the most recent OSM map data version.

Map data conflation links the IDs of road segments in older map data versions to the ID in the most recent (target) map data version. We consider the following types of operations that can happen to a road segment:

- **Addition** of newly created, or previously unmapped road segments
- **Deletion** of roads that have been removed (or were incorrectly mapped)
- **Alteration** of road, including changes in geometry/length or reclassification.
- **Splitting** of existing roads into multiple segments due to new intersecting road
- **Combination** of segments due to removal of intersecting roads
- **Twinning** of roads where a two-way road is split into two one-way roads separated by a divider.

We detect these changes programmatically by matching metadata like road names, OSM IDs, and doing fuzzy geospatial matching between map versions using Hootenanny's techniques. The result of the map data conflation is a lineage graph, as illustrated in Figure 5.

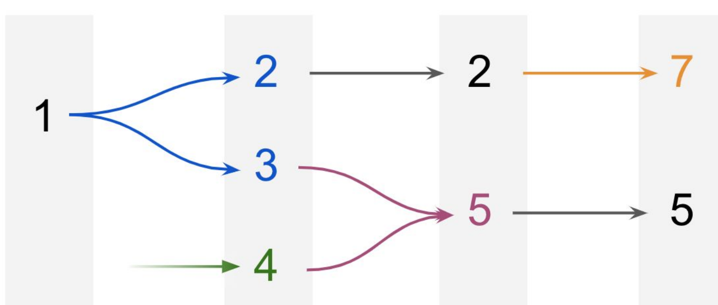

Figure 5. Example lineage graph representing the changes to a road segment with ID 1 over several different map data versions denoted by the grey bars: splitting (blue), combination (purple), addition (green), alteration (yellow).

In some cases, new speed values will be assigned to road segments that were modified during the rollup process. To preserve the ability to estimate speed percentiles and standard deviations, we transform the entire speed distribution on each road segment.

Table 1 describes how the rollup speed value is calculated by segment modification type:

Addition	Segment didn't exist previously so no mapping is required.
Deletion	Segment no longer exists. No action needed
Alteration	No action required since IDs are unchanged
Splitting	Segment A splits into segments $B_1 \ldots B_n$. Copy the speed values from segment A to segments $B_1 \ldots B_n$
Combination	Segments $A_1 \ldots A_n$ get combined into one segment B. Speed on the segment B is calculated as the average of the speed values of $A_1 \ldots A_n$ weighted by each segment's length.
Twinning	We already have different speeds for each direction of a two-way segment, so directly copy them to the two new road segments after twinning (similar for un-twinning)

Table 1. Speed rollup action per segment modification

Other Challenges

Alternating directions of road segments. Some road segments are technically one way but can switch direction throughout the day (e.g. I-5 express lanes in Seattle). In these cases, we "twin" the road in our frontend and show the segment as two road segments with different directions. Speed data only exists for each of these virtual segments during times when traffic is flowing in the relevant direction.

Privacy

Safeguarding the privacy of Uber users is always the number one priority. To ensure privacy we apply the following techniques:

- Anonymization: We do not use or publish personal data, identifiers or other personally identifiable data.
- Aggregation: We only use aggregated data with sufficient sample size. Segments at our lowest hourly granularity that have less than a certain number of traversals are not included in our UI or data downloads. While this may decrease our coverage, it is necessary to prevent any chance of revealing individual trip information.

- Non-identifiable: We recognize that the trip data before the aggregation is sensitive and could be used in attempt to re-identify individuals in even anonymized datasets. In order to further increase the privacy, we truncate Uber trip traces so they do not include the areas close to pickup and dropoff locations.

Interested in how Uber app privacy works? Uber's Privacy Policy privacy.uber.com/policy describes what information Uber collects, how it is used and shared, and your choices regarding this information as a user of Uber app.

More Questions? If you have any questions or concerns about the privacy of Movement data, please email movement-feedback@uber.com and someone from our team will be in touch with you.

Safety

There is a possibility that the aggregated speed data on a particular segment exceeds the speed limit, and at first glance may raise questions about road safety. However, it is important to note that what we're showing is the result of a calculation of speed based on distance over time (not instantaneous speed), which means that more information and context is needed to infer safety conditions. Additionally, worth noting is that, Movement data does not have access to details of the flow of traffic, roadway design or how the posted limit was set. And while we hope that researchers and cities use this information as part of larger speed studies, our primary goal with this tool is to shed light on how people are moving around cities and to contribute to early discussions around roadway engineering, policy, and design.

Additionally, as explained in the Privacy section, Movement data cannot be used to identify individual trips. We require a sufficient number of observations to calculate the aggregation, which means that our data shows overall averages and trends but are not necessarily indicative of a particular driving behavior.

At Uber, we remain committed to safety. We agree with the experts who recognize unsafe speeding as a key road safety risk factor, and again, hope the data presented here can be a part of larger efforts directed at making cities and roads safer for all travelers. For more information about our approach to road safety, please visit Uber's commitment to safety website.