Probabilistic Tour Modeling using Mobility Traces

December 16, 2016

1 Introduction

Our goal is to provide analytics that enable users and planners to understand travel patterns. The hope is that such analytics will support convenient and environmentally-friendly travel plans. We decompose the analytics goals into understanding individual patterns and aggregating these patterns to develop models of a community. These tasks are referred to in the transportation modelingas disaggregate tour modeling [1] for individual trips and aggregate tour modeling across many tirps.

We propose to use mobility data, such as motion and geolocation data from mobile phones, to build these models. At the individual level, our model will identify the locations a user visits regularly and common routes used between these locations. In addition to user models, we also propose to learn aggregate models that identify patterns at a community level. Individual and aggregate travel analytics can support a number of varied applications, including recommending routes with lower environmental impact or improving infrastructure on frequently-used routes.

2 Data Requirements

Our model assumes input is given as a series of *segments*. Each segment consists of several features:

- origin geolocation
- starting time
- destination geolocation
- ending time
- inferred travel mode (motorized vehicle, bicycle, or walking)

3 Model

Our modeling approach will include three components:

1. identifying *anchor locations* a user frequently visits
2. finding *frequent trips* between anchor locations
3. aggregating frequent trips across a community to find *local travel patterns*

Initially, we will develop independent solutions to each of these three tasks. Based on insights from these baseline models, we will build a collective model that jointly performs the three tasks.

We will use probabilistic soft logic (PSL) to construct each of these three models. PSL models are specified using rules in first-order logic syntax. We introduce some representative rules below as an initial exploration of the modeling problem, and discuss the rationale behind these rules. Ultimately, models will also include more sophisticated dependencies drawn from this initial modeling exploration.

3.1 Identifying Anchor Locations

Each user's anchor locations are associated with a time of day and travel mode, and the model attempts to deduplicate anchor locations using temporal information.

$$\textsc{Segment}(\texttt{S}) \land \textsc{StartLoc}(\texttt{S},\texttt{L}) \land \textsc{StartTime}(\texttt{S},\texttt{T}) \rightarrow \textsc{AnchorTime}(\texttt{L},\texttt{T}) \quad (1)$$
$$\textsc{Segment}(\texttt{S}) \land \textsc{EndLoc}(\texttt{S},\texttt{L}) \land \textsc{EndTime}(\texttt{S},\texttt{T}) \rightarrow \textsc{AnchorTime}(\texttt{L},\texttt{T}) \quad (2)$$
$$\textsc{Segment}(\texttt{S}) \land \textsc{StartLoc}(\texttt{S},\texttt{L}) \land \textsc{Mode}(\texttt{S},\texttt{M}) \rightarrow \textsc{AnchorMode}(\texttt{L},\texttt{M}) \quad (3)$$
$$\textsc{Segment}(\texttt{S}) \land \textsc{EndLoc}(\texttt{S},\texttt{L}) \land \textsc{Mode}(\texttt{S},\texttt{M}) \rightarrow \textsc{AnchorMode}(\texttt{L},\texttt{M}) \quad (4)$$
$$\textsc{AnchorMode}(\texttt{L},\texttt{M}) \rightarrow \textsc{Anchor}(\texttt{L}) \quad (5)$$
$$\textsc{AnchorTime}(\texttt{L},\texttt{T}) \rightarrow \textsc{Anchor}(\texttt{L}) \quad (6)$$
$$\textsc{AnchorTime}(\texttt{L1},\texttt{T}) \land \textsc{AnchorTime}(\texttt{L2},\texttt{T}) \land \texttt{L1} \neq \texttt{L2} \rightarrow \neg\textsc{Anchor}(\texttt{L2}) \quad (7)$$
$$\textsc{Anchor}(\texttt{L1}) \land \textsc{Near}(\texttt{L1},\texttt{L2}) \land \texttt{L1} \neq \texttt{L2} \rightarrow \neg\textsc{Anchor}(\texttt{L2}) \quad (8)$$

Discussion: Rules 1 and 2 associate starting and ending locations with a time. These times can be coarse-grained (morning, afternoon, evening, night) or fine-grained (15-minute intervals). Rules 3 and 4 associated each location with the travel mode used at that location. Rules 5 and 6 promote locations with strong time and mode propensity to anchor locations. Rule 8 encourages sparsity by preventing many anchor locations from having the same time propensity.

3.2 Finding Frequent Trips

The second component of the modeling approach is determining frequent trips. Trips are defined to involve travel between anchor locations, and can be associated with time and mode, similarly to anchor locations. The model can also attempt to impute missing trips that fill a gap in a user's travel patterns.

$$\begin{aligned}
&\textsc{Segment}(\texttt{S}) \land \textsc{Anchor}(\texttt{L1}) \land \textsc{Anchor}(\texttt{L1}) && (9)\\
&\land \textsc{StartLoc}(\texttt{S},\texttt{L1}) \land \textsc{EndLoc}(\texttt{S},\texttt{L2}) \rightarrow \textsc{FreqTrip}(\texttt{L1},\texttt{L2}) &&\\
&\textsc{Segment}(\texttt{S1}) \land \textsc{Segment}(\texttt{S2}) \land \textsc{Anchor}(\texttt{L1}) \land \textsc{Anchor}(\texttt{L1}) && (10)\\
&\quad \land \textsc{StartLoc}(\texttt{S1},\texttt{L1}) \land \textsc{EndLoc}(\texttt{S2},\texttt{L2}) \land \textsc{StartDay}(\texttt{S1},\texttt{D}) &&\\
&\quad \land \textsc{EndDay}(\texttt{S2},\texttt{D}) \land \textsc{StartTime}(\texttt{S1},\texttt{T1}) \land \textsc{EndTime}(\texttt{S2},\texttt{T2}) &&\\
&\quad \land \textsc{Before}(\texttt{T1},\texttt{T2}) \land \textsc{SimTime}(\texttt{T1},\texttt{T2}) \rightarrow \textsc{FreqTrip}(\texttt{L1},\texttt{L2}) &&\\
&\textsc{FreqTrip}(\texttt{L1},\texttt{L2}) \land \textsc{FreqTrip}(\texttt{L3},\texttt{L1}) \rightarrow \textsc{FreqTrip}(\texttt{L2},\texttt{L3}) && (11)\\
&\textsc{FreqTrip}(\texttt{L1},\texttt{L2}) \land \textsc{FreqTrip}(\texttt{L3},\texttt{L4}) \land \textsc{FreqTrip}(\texttt{L4},\texttt{L1}) \rightarrow \textsc{FreqTrip}(\texttt{L2},\texttt{L3}) && (12)\\
&\textsc{FreqTrip}(\texttt{L1},\texttt{L2}) \land \textsc{StartLoc}(\texttt{S1},\texttt{L1}) \land \textsc{EndLoc}(\texttt{S2},\texttt{L2}) && (13)\\
&\quad \land \textsc{StartTime}(\texttt{S1},\texttt{T1}) \land \textsc{EndTime}(\texttt{S2},\texttt{T2}) \rightarrow \textsc{FreqTripTime}(\texttt{L1},\texttt{L2},\texttt{T1},\texttt{T2}) &&\\
&\textsc{FreqTrip}(\texttt{L1},\texttt{L2}) \land \textsc{StartLoc}(\texttt{S1},\texttt{L1}) \land \textsc{EndLoc}(\texttt{S2},\texttt{L2}) && (14)\\
&\quad \land \textsc{Mode}(\texttt{S1},\texttt{M}) \land \textsc{EndMode}(\texttt{S2},\texttt{M}) \rightarrow \textsc{FreqTripMode}(\texttt{L1},\texttt{L2},\texttt{M}) &&\\
& && (15)
\end{aligned}$$

Discussion: Rule 9 identifies the most basic trip using the anchor locations of frequently occurring segments. Rule **??** is slightly more complex, using anchor locations form temporally adjacent segments on the same day. Rules 11 and 12 impute missing trips in tours of increasing length. Rule

13 determines the travel mode associated with a trip, while Rule 14 associates trips with the time of day.

3.3 Finding Local Travel Patterns

Aggregating trips across users will likely be a more difficult task. Here we present some initial rules for performing simple aggregates. Depending on the downstream applications and additional data sources, a richer set of aggregate modeling can be implemented.

$$\begin{aligned}
&\textsc{FreqTripTime}(\texttt{U},\texttt{L1},\texttt{L2},\texttt{T1},\texttt{T2}) && && (16)\\
&\quad \wedge \textsc{LocArea}(\texttt{L1},\texttt{A}) \wedge \textsc{LocArea}(\texttt{L2},\texttt{A}) && \rightarrow \textsc{FreqTripAreaTime}(\texttt{A},\texttt{L1},\texttt{L2},\texttt{T1},\texttt{T2}) &&\\
&\textsc{FreqTripMode}(\texttt{U},\texttt{L1},\texttt{L2},\texttt{M}) \wedge \textsc{LocArea}(\texttt{L1},\texttt{A}) && && (17)\\
&\quad \wedge \textsc{LocArea}(\texttt{L2},\texttt{C}) && \rightarrow \textsc{FreqTripAreaMode}(\texttt{A},\texttt{L1},\texttt{L2},\texttt{M}) &&
\end{aligned}$$

Discussion: Rules 16 and 17 determine common trips across users that share the same time propensity and mode of travel, respectively. One example of a more complex analytic task would be to determine the anchor locations common to a geographical region and identify trips with similar endpoints that travel through different intermediate anchor locations.

4 Discussion

References

[1] John Gliebe, Ofir Cohen, and John Hunt. Dynamic choice model of urban commercial activity patterns of vehicles and people. *Transportation Research Record*, 2003:17–26, 2007.