Transportation Research Part C 77 (2017) 96–112

Contents lists available at ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

Real-time trip purpose prediction using online location-based search and discovery services

Alireza Ermagun [a], Yingling Fan [a,*], Julian Wolfson [b], Gediminas Adomavicius [c], Kirti Das [a]

[a] Humphrey School of Public Affairs, University of Minnesota, 301 19th Ave S #307, Minneapolis, MN 55455, United States
[b] Division of Biostatistics, University of Minnesota, 420 Delaware St. SE, Minneapolis, MN 55455, United States
[c] Department of Information and Decision Sciences, University of Minnesota, 321 19th Avenue South, Minneapolis, MN 55455, United States

ARTICLE INFO

Article history:
Received 13 August 2016
Received in revised form 23 January 2017
Accepted 24 January 2017

Keywords:
Trip purpose prediction
Online location-based search
Google Places
Nested logit model
Random forest model
Smartphone

ABSTRACT

The use of smartphone technology is increasingly considered a state-of-the-art practice in travel data collection. Researchers have investigated various methods to automatically predict trip characteristics based upon locational and other smartphone sensing data. Of the trip characteristics being studied, trip purpose prediction has received relatively less attention. This research develops trip purpose prediction models based upon online location-based search and discovery services (specifically, Google Places API) and a limited set of trip data that are usually available upon the completion of the trip. The models have the potential to be integrated with smartphone technology to produce real-time trip purpose prediction. We use a recent, large-scale travel behavior survey that is augmented by downloaded Google Places information on each trip destination to develop and validate the models. Two statistical and machine learning prediction approaches are used, including nested logit and random forest methods. Both sets of models show that Google Places information is a useful predictor of trip purpose in situations where activity- and person-related information is uncollectable, missing, or unreliable. Even when activity- and person-related information is available, incorporating Google Places information provides incremental improvements in trip purpose prediction.

1. Introduction

Researchers have traditionally used the Global Positioning System (GPS) technology to identify trip origins/destinations, tracking travel routes, estimating travel time, and classifying vehicle types (Wagner, 1997; Casas and Arce, 1999; Yalamanchili et al., 1999; Draijer et al., 2000; Zhan et al., 2013; Sun and Ban, 2013). Recently, GPS data has been used in conjunction with other sensing data to predict additional trip characteristics (Gong et al., 2014; Wan and Lin, 2013). The smartphone technology has made it possible to generate combined GPS and sensing data in a single device. Smartphones often come with several built-in micro-electro-mechanical-system sensors such as accelerometers, magnetometers, gyroscopes, and location detection services through GPS and network positioning (including WiFi and/or cell networks). The use of multiple sensors combined with GPS has the advantage of providing continuous trip-related data such as location, velocity, orientation, and acceleration that help to predict two crucial trip characteristics—travel mode and trip purpose—with reasonable predictive accuracy (Nitsche et al., 2014; Gong et al., 2014; Moiseeva and Timmermans, 2010). Travel mode refers

* Corresponding author.
E-mail address: yingling@umn.edu (Y. Fan).

http://dx.doi.org/10.1016/j.trc.2017.01.020

to the mode of travel used for a trip, such as car, bus, walk, bicycle, and train. Trip purpose refers to the purpose or activity for a trip, such as work, eating out, shopping, education, personal business, and home.

Between travel mode and trip purpose predictions, the former has been studied more extensively, and more effective ways have been developed to automatically differentiate between travel modes (Gong et al., 2014, 2012; Zhang et al., 2011; Nitsche et al., 2014; Wu et al., 2011). In contrast, trip purpose prediction has received considerably less attention (Montini et al., 2014; Gong et al., 2014). The primary challenge with trip purpose prediction has been the need for multi-factorial approaches, which require additional data sources, like external land use or points of interest (POI) data, in addition to GPS and sensing data (Oliveira et al., 2014). Further, with reliance on external land use or POI data, existing research on trip purpose prediction has focused on developing post-processing models or algorithms that predict trip purposes after a fixed amount of data (e.g., a day or a week's worth) and all relevant offline information are collected. To date, we are not aware of any research that focuses on developing models or algorithms that predict trip purpose concurrently, or "on-the-go," upon the completion of the trip.

Our study addresses this knowledge gap by developing effective trip purpose prediction models that can be integrated with modern smartphone technology. Unlike previous studies that have relied on local land use or POI data, this study is the first to test the utility of nearby places data from online location-based search and discovery services, such as the Google Places Application Programming Interface (API), in predicting trip purpose. Besides being real-time, Google Places data is standardized globally. Our models are usable across many geographies. In addition, this research has practical implications for travel data collection. Implementing on-the-go trip purpose prediction models in travel data collection efforts is likely to reduce respondent burden when travel surveys are conducted using smartphone applications. Rather than having users fill out cumbersome travel diaries or recall surveys, such prediction models could automatically identify trip purposes so that the participant will only be required to confirm or edit (if needed) the system-provided automated prediction. Real-time trip purpose prediction also opens up the possibilities of delivering real-time, customized messages and interventions based upon users' detected trip purposes.

2. Existing studies on trip purpose prediction

Unlike mode detection, trip purpose or activity type detection has not received significant attention in the past. In the existing literature, the level of detail while looking at trip purpose varies significantly, ranging from very general (e.g., indoor vs. outdoor) to very specific (e.g. work, education, and shopping). The level of detail and the classification of activity type play a key role in determining the complexity of models used for prediction. Gong et al. (2014) classify the various methods in the field into three broad categories:

- Rule-based methods that match locational information and the respondent's personal information with a series of predefined heuristic rules to identify trip purposes.
- Statistical methods that use logistic models to calculate the probability of each trip purpose based upon locational information and the respondent's personal information.
- Machine learning and neural networks methods that build sophisticated, computation-intensive classification, and pattern recognition models using existing data of trip purpose, locational information, and the respondent's personal information.

Lee et al. (2016) comprehensively reviewed the studies using emerging data collection technologies for travel mode and trip purpose prediction. Our review of existing studies (Table 1) focuses on studies that included geographic information for trip purpose prediction. As shown in Table 1, the more recent studies tend to use probabilistic methods or machine learning. Although studies using machine learning methods generally produced higher levels of predictive accuracy than those using probabilistic and rule-based methods, using accuracy numbers from different studies may produce misleading results due to the differences in sample sizes, in trip purpose categories (e.g., indoor/outdoor vs. work/recreation), and in the quality of land use and POI data across the studies. Interestingly, we identified only two studies that used more than one method and, thus, were able to compare model performances (Oliveira et al., 2014; Wu et al., 2011). Oliveira et al. (2014) used a two-level nested logit model (probabilistic) and a decision tree model (machine learning) to differentiate between 12 trip purposes. In their study, the decision tree model outperformed the nested logit model by 5% in terms of accuracy (nested logit model accuracy was 60% and decision tree model accuracy was 65%). They also found that it was much faster to generate functioning models using decision trees. Wu et al. (2011) tested two models, a rule-based model and a random forest decision tree model to differentiate between indoor, outdoor static (i.e., when an individual is relatively stationary while outdoors), outdoor walking, and in-vehicle travel activities. They found no striking differences in the performance of the two models, both of which were successful in predicting indoor activities (>96% accuracy) and in-vehicle travel (>88% accuracy) and were only moderately successful in identifying outdoor static and walking points.

As shown in Table 1, input variables used in the existing studies to predict trip purpose typically include information on the destinations' surrounding environment derived from geographically referenced land use or POI data. Input variables also include related activity characteristics (e.g., duration of the activity at the destination, previous activity, activity start/end time) and personal socio-demographics (e.g., respondent's, age, occupation, income, and household structure). For example,