Transportation Research Part C 71 (2016) 447–463

Contents lists available at ScienceDirect

Transportation Research Part C

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

Detecting trip purposes from smartphone-based travel surveys with artificial neural networks and particle swarm optimization

Guangnian Xiao [a], Zhicai Juan [a,*], Chunqin Zhang [b]

[a] School of Antai Economics & Management, Shanghai Jiao Tong University, Shanghai 200052, China
[b] School of Civil Engineering and Architecture, Zhejiang Sci-Tech University, Hangzhou 310018, China

ARTICLE INFO

Article history:
Received 17 March 2016
Received in revised form 4 July 2016
Accepted 15 August 2016

Keywords:
Trip purpose
Smartphone
Travel survey
Artificial neural network
Particle swarm optimization

ABSTRACT

Travel surveys based on global positioning system (GPS) data have exponentially increased over the past decades. Trip characteristics, including trip ends, travel modes, and trip purposes need to be detected from GPS data. Compared with other trip characteristics, studies on trip purpose detection are limited. These studies struggle with three types of limitations, namely, data validation, classification approach-related issues, and result comparison under multiple scenarios. Therefore, we attempt to obtain full understanding and improve these three aspects when detecting trip purposes in the current study. First, a smartphone-based travel survey is employed to collect GPS data, and a surveyor-intervened prompted recall survey is used to validate trip characteristics automatically detected from the GPS data. Second, artificial neural networks combined with particle swarm optimization are used to detect trip purposes from the GPS data. Third, four scenarios are constructed by employing two methods for land-use type coding, i.e., polygon-based information and point of interest, and two methods for selecting training dataset, i.e., equal proportion selection and equal number selection. The accuracy of trip purpose detection is then compared under these scenarios. The highest accuracies of 97.22% for the training dataset and 96.53% for the test dataset are achieved under the scenario of polygon-based information and equal proportion selection by comparing the detected and validated trip purposes. Promising results indicate that a smartphone-based travel survey can complement conventional travel surveys.

1. Introduction

Traffic congestion has become an issue that has attracted the attention of people in many cities over the past couple of decades. To address this issue, transportation policymakers worldwide have attempted to develop travel demand models to predict the effect of transportation demand management (TDM) policies (subsidizing transit costs for employees or residents, road pricing tolls during peak hours, and workplace travel plans) on individual travel patterns. Travel surveys (household or personal) are used to collect explicit information on individuals (socio-economic and demographic), their household (size, structure, and relationships), and a diary of their trips during the survey period. The collected data are used to calibrate these travel demand models. As a result, the accuracy and completeness of trip characteristics collected in travel surveys have a critical effect on model results and the effectiveness of TDM policies.

* Corresponding author.
E-mail address: zcjuan@sjtu.edu.cn (Z. Juan).

http://dx.doi.org/10.1016/j.trc.2016.08.008

Travel surveys have evolved in terms of involved technologies over time. Early surveys were conducted based on a paper-and-pencil interview (PAPI) with face-to-face or mail-out/back interviews. PAPI, although easy to implement, requires a high cost. During the 1980s and 1990s, computer-assisted telephone interview (CATI) became popular because of the high penetration rate of telephones and the low calling cost (Badoe and Steuart, 2002). However, CATI poses a heavy burden on respondents because they are required to report the explicit characteristics (trip timing, travel destination, travel mode, and trip purpose) of each trip. Recently, global positioning system (GPS) technologies, which record latitude, longitude, altitude, heading, instantaneous speed, and positioning quality (the number of satellites and the horizontal dilution of precision) according to user-specified time interval, have introduced a new mode of travel survey (Chung and Shalaby, 2005). GPS-based travel surveys have potential advantages over conventional ones in the following aspects: (1) positioning data are passively collected without burdening respondents; (2) complete routes are recorded for most trips, making it possible to recover unreported trips; (3) accurate trip start and end times as well as trip lengths can be automatically detected (Wolf et al., 2001).

While GPS technologies bring new hopes in obtaining accurate and complete positioning data, certain important trip characteristics, including trip ends, travel modes, and trip purposes, cannot be observed directly from passively collected data (Du and Aultman-Hall, 2007). Trip purpose detection has received less attention in data mining of GPS data than trip end and travel mode identification (Shen and Stopher, 2014). Few studies have looked into the area of trip purpose detection. Trip purposes are generally detected based on either land-use type (Wolf et al., 2001; Wolf et al., 2004) or a combination of land-use type and personal information (home address and workplace) (Bohte and Maat, 2009). A summary of studies on trip purpose detection is shown in Table 1. However, three aspects of limitations are observed in existing literature on trip purpose detection, namely, data validation, classification approach-related issues, and result comparison under multiple scenarios.

Detected trip purposes are generally compared with reported trip characteristics to determine the accuracy of employed approaches. Therefore, the completeness of GPS data and the accuracy of reported trip characteristics have a critical effect on the evaluation of classifiers. Smartphones and dedicated GPS devices are generally used to collect GPS data. Respondents are more likely to provide complete GPS data by smartphones than dedicated GPS devices because most of them are accustomed to taking smartphones with them and checking smartphones frequently. In terms of data validation, several existing studies have asked respondents to complete a web-based or paper-based diary during the survey period (Maruyama et al., 2014). However, the reported trip characteristics in this way might not accurately match the recorded GPS data because of under-reporting or erroneous memories. Prompted recall (PR) surveys can effectively improve the consistency between reported trip characteristics and GPS data by providing respondents with a map with GPS trajectories and automatically derived trip ends, travel modes, and trip purposes. Respondents are required to correct these trip characteristics if necessary. However, respondents might have different understandings of "trip" and report different trip characteristics for the same trip. As a result, the trip characteristics collected from the PR survey remain far from the "ground truth" because of the self-report error, similar to those in conventional travel surveys (Shen and Stopher, 2014). Therefore, surveyors are introduced into the process of validating trip characteristics.

Most studies on trip purpose detection have depended on rule-based and/or machine-learning approaches. Bohte and Maat (2009) labeled a trip as the type of a point of interest (POI) when the distance between the trip end and the POI was less than 50 m, an unlabeled trip as shopping when the trip end of the trip was within the shopping center polygon, and unlabeled trips with destination less than 100 m from home/workplace addresses as trips with purposes of home/work.

Table 1
A summary of studies on trip purpose detection.

Author/s	Classification method	Features	Land-use type coding	Best accuracy (%)	Ground truth
Wolf et al. (2001)	Rule-based approach	Land-use type	Polygon-based information	60.9	Paper-based diaries
Stopher et al. (2008)	Rule-based approach	Land-use type, addresses of home, workplace/school, and two most frequently visited grocery stores	n/a	Over 60	PR surveys
Bohte and Maat (2009)	Rule-based approach	Land-use type, home and workplace/school addresses	POI and polygon-based information	43	PR surveys
McGowen and McNally (2007)	Discriminant analysis and classification tree	Land-use type	Polygon-based information	74	Paper-based diaries
Deng and Ji (2010)	Decision tree (C4.5)	Land-use type, individual and trip related characteristics	Polygon-based information	87.6	Web-based diaries
Lu et al. (2013)	Decision tree, support vector machine, metalearner	Land-use type, individual and trip related characteristics	POI and polygon-based information	80.6	PR surveys
Feng and Timmermans (2013)	Bayesian network, Decision tree, Random forest	Land-use type, trip related characteristics	POI	96.8	PR surveys