Transportation Research Part C 36 (2013) 261–267

Contents lists available at ScienceDirect

Transportation Research Part C

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

A process for trip purpose imputation from Global Positioning System data

Li Shen *, Peter R. Stopher [1]

Institute of Transport and Logistics Studies, The University of Sydney, NSW 2006, Australia

ARTICLE INFO

Article history:
Received 21 October 2012
Received in revised form 9 September 2013
Accepted 9 September 2013

Keywords:
GPS survey
Trip purpose imputation
Tour-based information

ABSTRACT

Global Positioning System (GPS) devices have been applied in travel data collection for much of the past decade to improve data quality. These devices can record positions, time, and travel speed. However, trip purposes currently cannot be recorded automatically by the devices, therefore, the accuracy of purpose imputation becomes important to improve the quality of travel survey data. This paper proposes an improved process which introduces some additional information (e.g., activity duration, tour information) for trip purpose imputation. An application of this approach is reported in this paper based on a Global Positioning System survey in the Greater Cincinnati region to show the improvement in the accuracy of trip purpose detection over more conventional methods.

1. Introduction

The purpose of this study is to examine the effects of tour-based information and additional activity information on trip purpose imputation from GPS travel data. Global Positioning System (GPS) devices have been applied in travel data collection for much of the past decade to improve data quality. Although some types of travel data (e.g., travel time and location) can be derived directly from GPS devices, travel modes and trip purposes currently cannot be recorded automatically. Modes and purposes are, therefore, usually imputed based on the data that are recorded by GPS devices and other external data (e.g., land use). The accuracy of this detection is as important as that of travel time and location to improve the quality of the travel survey data. While mode detection is more accurate than trip purpose detection currently, because it is more closely related to speeds and routes that can be identified simply based on recorded GPS data, the process of detecting trip purpose still has much room for improvement.

There are only a few papers that have looked into the area of trip purpose imputation. The traditional process of trip purpose imputation is based on either land-use information (Wolf et al., 2001, 2004) or a combination of land use and personal information (e.g., home address, possession of vehicles) (Stopher et al., 2008; Bohte and Maat, 2009; Moiseeva et al., 2010). The former examines the land use of different locations (residential places, shops, stations, etc.) and provides possible purpose for each trip accordingly. The latter adds individual characteristics to improve the accuracy of the detection, especially for return home trips and work trips. However, neither method has achieved good accuracy in imputation.

Based on the traditional process, some additional information about an activity, i.e., activity duration and the time when the activity occurs, is analysed in this paper for the detection. The 2009 National Household Travel Survey (NHTS) in the US (U.S. Department of Transportation, Federal Highway Administration, 2009) is used as a basic data source for analysing the distribution of the additional travel information mentioned above and tour information of people's daily travel. A case study

* Corresponding author. Tel.: +61 2 9114 1884.
E-mail addresses: li.shen@sydney.edu.au (L. Shen), peter.stopher@sydney.edu.au (P.R. Stopher).
[1] Tel.: +61 2 9114 1808.

0968-090X/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.trc.2013.09.004

in the Greater Cincinnati region is described in this paper. It was designed as a household travel survey that commenced in August 2009 and finished in August 2010. Every member in the household over the age of 12 was asked to carry a passive GPS device for three days. After the collection, a prompted recall (PR) web survey was also conducted, in which respondents were assisted to recall their actual travel by receiving GPS-generated maps of where and when they travelled. The new approach described in this paper is applied to the case study to process the data for detecting trip purpose and show the improvement in the accuracy of trip purpose detection over more conventional methods.

Although some research has adopted probabilistic methods to impute purpose from GPS data (Griffin and Huang, 2005; McGowen and McNally, 2007), the approach taken in this research remains a deterministic approach, developing additional rules for classifying purposes. The reason for this stems partly from the fact that early GPS work provided data with much less accuracy than is currently possible to achieve, and therefore has not provided an adequate pool of information that could be used in probabilistic approaches, and partly because a reliable source of 'ground truth' about travel is not yet available (Bohte and Maat, 2009).

Section 2 of this paper analyses the data from NHTS and describes the new process of imputing trip purpose. Section 3 shows the improvement of the new process based on a case study from a GPS-only survey conducted in Ohio, US. Section 4 concludes the research results and suggests future work for this research.

2. Approach and data analysis

The NHTS data is used as source data to obtain the basic information which can be applied to the case study (i.e., the Greater Cincinnati region GPS-only survey). The basic travel information includes the distribution of the activity duration, the distribution of the time when the activity occurs, and tour information.

2.1. Distribution of activity duration

People undertake different activities normally for different durations. Typically, there are some basic rules for some activities in terms of duration, e.g., working may take 4–8 h per day and education may take 3–6 h per day. Since the NHTS data are used and they include residents over the age of five, while the data in the Greater Cincinnati region only included residents over 12, the first step was to exclude children from age five to twelve from the dataset. Also, certain adult proxy reporting data were removed due to a lack of accuracy. This study focuses on the activities of work, education, shopping, being at home and others. Therefore, all the trips in the NHTS are categorised into those groups. In this analysis, in-home activities, which mainly occur during nights, are not counted, because information about these activities could not help examine the effects of activity duration on trip purpose imputation. Fig. 2.1 shows the proportion of each activity. "Shopping" and "Other" are the most probable activities to occur during the day and education makes up the smallest percentage of the activities.

Fig. 2.2 shows the distributions of different activity durations. It illustrates that work and education are more likely to occur when the duration is longer than 4 h. Shopping mostly takes less than 4 h. Working dominates the activities when duration is longer than 8 h. Therefore, a rule is created to test the effect of activity duration on purpose imputation, i.e., if the duration is longer than 4 h and the purpose detected from GPS data is not work or education, this purpose should be suspected as being possibly wrong and the purpose may need to be redefined.

2.2. Distribution of the time when activities occur

Similar to the activity duration, there are also some basic rules for the time when an activity occurs. In this section, an activity (i.e., shopping, work, education, home, other) refers to a "travel to" purpose. Working trips (i.e., go to work) are more likely to occur from 8 to 9 am and finish at 5–6 pm. NHTS data are still used to analyse the basic distribution of the time

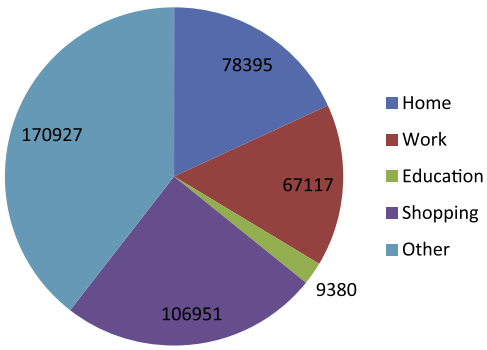

Fig. 2.1. Proportion of each activity.