Travel Behaviour and Society 13 (2018) 11–25

Contents lists available at ScienceDirect

Travel Behaviour and Society

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

A model for measuring activity similarity between public transit passengers using smart card data

Hamed Faroqi[a,*], Mahmoud Mesbah[a,b], Jiwon Kim[a], Ahmad Tavassoli[a]

[a] School of Civil Engineering, The University of Queensland, Australia
[b] Department of Civil and Environmental Engineering, Amirkabir University of Technology, Iran

ARTICLE INFO

Keywords:
Time-geography
Data mining
Travel behaviour
Planning
Spatiotemporal

ABSTRACT

An activity is characterized by its location, time and type. Smart card data include the location and time of boarding and/or alighting transactions within the public transit system. This data can be used to study the spatiotemporal range of the activity as it usually happens between an alighting and the next boarding transaction. This kind of activity can also be inferred from the start time and duration of the activity, and the available land use in the vicinity. This paper proposes a model which considers the three main characteristics of the activity to measure similarities between passengers' activities. The model consists of two parallel steps—one for the spatiotemporal aspects and the other for the activity type. The first one uses the concept of Space Time Prism (STP) to measure the spatiotemporal similarity of two activities in a three-dimensional continuous space. The latter models the activity type using a probabilistic decision tree to measure the activity type similarity. The final activity similarity value is the product of the activity type and the spatiotemporal similarity values. The model is implemented for four-day smart card data in Brisbane, Queensland [Australia]. In order to confirm the results of the model, the passengers are clustered and discussed based on the measured activity similarity. The results show more than 81 per cent of the passengers have partial or complete activity similarity with their fellow passengers.

1. Introduction

Understanding activity patterns and interactions between activities of the public transit passengers can improve both public transit network and urban planning structures. A number of studies have studied these issues to detect and examine activities of passengers in the public transit network (Devillaine et al., 2012; Kusakabe and Asakura, 2014; Lee and Hickman, 2014; Nassir et al., 2015; Faroqi and Sadeghi-Niaraki, 2016; Alsger, 2017; Zhang et al., 2018). Passengers travel by public transit to perform an activity, such as work, shopping, a recreational activity, or attending school. The activity occurs at a place and during a certain time. A specific activity type is related to the availability of land use at the dimensions of the time and place. For instance, it is impossible to shop at a shopping centre after the closing hours or shop at a school. Therefore, the passengers' activities—as the main reason for using public transit—should be examined considering the start time, duration, location, and land use.

Smart card datasets include boarding and/or alighting transactions of public transit. The datasets provide required data to study the activities of public transit passengers. The datasets are usually big—they include data for hundreds of thousands of passengers in a metropolitan area (Chen et al., 2016; Faroqi et al., 2017). The public transit smart card datasets provide a great opportunity to develop transport models and new applications, and study travel behaviour and travel demands. A transaction includes the time and location of a boarding and/or an alighting for one passenger who is identified by a specific ID in the dataset (Kieu et al., 2015). The activity can occur between an alighting and the next boarding transaction. A passenger usually performs the activity somewhere and during the time between one alighting and the next boarding transaction in the public transit network. This means the smart card datasets can determine spatial and temporal spans of the activity.

Time-geography presents a framework to study constraints on movements in both spatial and temporal dimensions. Constraints can include the object's maximum walking pace or opening hours of a shopping centre (Long and Nelson, 2013). In addition, the space-time cone and prism are two approaches which measure probability for locations of moving objects passing the time. The Space Time Prism (STP) quantifies movement possibilities amid two known locations that should be determined in time as well (Chen et al., 2013). Moreover, the

* Corresponding author.
E-mail addresses: h.faroqi@uq.edu.au (H. Faroqi), mmesbah@aut.ac.ir, mahmoud.mesbah@uq.edu.au (M. Mesbah), jiwon.kim@uq.edu.au (J. Kim), a.tavassoli@uq.edu.au (A. Tavassoli).

https://doi.org/10.1016/j.tbs.2018.05.004
Received 23 November 2017; Received in revised form 8 May 2018; Accepted 24 May 2018

boarding and alighting transactions from the smart card dataset include both time and location. Likewise, the activity can occur between two consecutive transactions. Therefore, the STP method can model the spatiotemporal possibilities and probabilities for the public transit passengers' activities.

Interactions between public transit passengers can be modelled as the interactions between their activities. The activities of two passengers can be considered similar if they have similar activities in all spatial, temporal and activity type aspects. The STP can model the spatiotemporal aspects of such activity. The STP can measure intersections between activities of different passengers in a three-dimensional spatiotemporal continuous space. Likewise, a probabilistic decision tree approach can model the activity type. A decision tree can infer the activity type and measure the probability that two activity types are similar. Consequently, the interactions between two passengers can be studied by their activities implementing the STP and decision tree approaches.

This paper presents a novel model for exploring the activity similarity among public transit passengers—it considers both the activity type and spatiotemporal elements of the activity. The proposed model uses the STP method to measure the spatiotemporal similarity and the decision tree approach to measure the activity type similarity. The final activity similarity value is the product of the activity type and the spatiotemporal similarity values. The model is implemented for four-day data within Brisbane's public transit network. In order to confirm the results of the model, the passengers are clustered based on the measured activity similarity using an Agglomerative Hierarchical clustering algorithm.

In simple words, in this paper, the activities of passengers are extracted from the smart card dataset that includes boarding and alighting transactions of the passengers. Passengers' behaviour during their activity are focused considering the three main elements of each activity (time, location, and type). The STP method is used to measure the location and time similarity between the activities. Also, the decision tree method is used to infer the activity types according to the available land use, start time, and duration of the activity. Consequently, the most important scientific contributions can be summarised as follows.

1. The model considers the three main elements (time, location, and type) of the activity together in one framework.
2. The model develops a passenger-based perspective which concentrates on the passengers' behaviour between two consecutive trips rather than the riding period on the public transit network.

The remainder of this paper is structured as follows. Firstly, existing literature—considering transport planning activity-based models, smart card data and the STP method—is reviewed. Following this, the methodology—including the STP and activity type similarity steps—is explained. From then on in, the case study and its results are presented and analysed. Next, a Discussion section is dedicated to discuss the important advantages and potential applications of the model. Then, a Validation section that validates the results of the model using a Household Travel Survey (HTS) dataset. The last section of this paper includes conclusions for the proposed model, and plans for expanding the model development and potential applications.

2. Literature review

2.1. Activity similarity

Conventionally, travel demand models take single trips of individuals and model them regarding their generation, distribution, model split and assignment to the public transport network. In contrast, activity-based models are tour-based rather than trip-based and they extract the travel demand from the need to perform activities (Algers et al., 2005). For instance, ALBATROSS, as one of the known activity-based models, explains and anticipates which activities individuals conduct, and where, when and for how long (Arentze and Timmermans, 2000). In addition, measuring the similarity between activities is one of the interesting fields in this regard that attracts researchers' attention. Kwan et al. (2014) explained the necessity and complexity of defining a method that allows researchers to determine how close or similar one's activity is to that of another. They defined the 'activity similarity problem' as a multi-dimensional sequence alignment and they formulated it to a multi-objective optimisation problem. They implemented the proposed method on 50 travel diaries from 30 households during two days. Shen and Cheng (2016) considered the equal roles for space and time in the explanation of people's activities. They determined the spatiotemporal region of interests for 100 police officers from five different police stations in London through finding places with the frequent visit of multiple users in a limited time span. They evolved the concept of "where, when and how long you stay is who you are" into "what [type of)] places, when and how long you stay is who you are".

2.2. Smart card analysis

Smart card technology is developing all around the world. It is utilised in public transport systems, as well as other domains. These cards usually gather fare information from passengers and the more valuable information is stored in a dataset. This information helps to study travel behaviour of passengers or for an organisation to evaluate the performance of the network Pelletier et al., 2011; Faroqi et al., 2018). Agard et al. (2006) mined smart card data in order to discover trip habits of public transit users. They answered the question "is data mining techniques can be used to study user behaviour from observations of smart card data". From their results, they found a high capability of smart card data could be used to study passengers' behaviour. Chapleau et al. (2008) proposed a data frame to enrich smart card datasets using GIS. They also described the Point of Interest (POI) concept as a place where people are inclined to perform their activities. These researchers also designed a method of trip purpose (activity type) inference as a function of fare type, time of day, activity duration, activity location and frequency of the activity.

Finding where and when an activity occurs is another challenge with smart card datasets. Devillaine et al. (2012) considered two criteria for detecting an activity. First, if two consecutive trip legs are on the same routes, then there should be an activity between them. Second, if the time gap between an alighting and a next boarding transaction is more than 30 min, then there should be an activity. They also defined rules on fare type, duration and start time of activity in order to infer the purpose of the person's trip. Nassir et al. (2015) defined a criterion called 'off-optimality'—this is a time deviation between the observed route and the quickest possible path in a time-dependent transit network. This is used to detect short activities between trip legs extracted from smart card data. Lee and Hickman (2014) developed previous works by attaching land use information with temporal and user information in relation to activity type inference processes using a rule-based decision tree. Kusakabe and Asakura (2014) integrated smart card data with HTS using a data fusion technique. They estimated absent behavioural attributes, such as trip purpose through applying naïve classifier method. Langlois et al. (2016) developed a method considering smart cards data to cluster users sharing similar multi-week activity sequences. They focused on the sequence of the activities and examined heterogeneity among the passengers. More recently, Alsger (2017) developed a probabilistic decision tree model to infer the trip's purpose. They evaluated their model using HTS data finding a high level of accuracy.

2.3. Time geography

Peuquet (1994) proposed an integrated method for illustrating spatiotemporal datasets in geographic information systems (GIS).