Transportation Research Part C 90 (2018) 97–113

Contents lists available at ScienceDirect

Transportation Research Part C

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

Large-scale transit market segmentation with spatial-behavioural features

Le Minh Kieu*, Yuming Ou, Chen Cai

Data61, CSIRO, Sydney, Australia

ARTICLE INFO

Keywords:
Customer segmentation
Travel pattern
Smart card
AFC

ABSTRACT

Transit market segmentation enables transit providers to comprehend the commonalities and heterogeneities among different groups of passengers, so that they can cater for individual transit riders' mobility needs. The problem has recently been attracting a great interest with the proliferation of automated data collection systems such as Smart Card Automated Fare Collection (AFC), which allow researchers to observe individual travel behaviours over a long time period. However, there is a need for an integrated market segmentation method that incorporating both spatial and behavioural features of individual transit passengers. This algorithm also needs to be efficient for large-scale implementation. This paper proposes a new algorithm named Spatial Affinity Propagation (SAP) based on the classical Affinity Propagation algorithm (AP) to enable large-scale spatial transit market segmentation with spatial-behavioural features. SAP segments transit passengers using spatial geodetic coordinates, where passengers from the same segment are located within immediate walking distance; and using behavioural features mined from AFC data. The comparison with AP and popular algorithms in literature shows that SAP provides nearly as good clustering performance as AP while being 52% more efficient in computation time. This efficient framework would enable transit operators to leverage the availability of AFC data to understand the commonalities and heterogeneities among different groups of passengers.

1. Introduction

Maintaining service quality and customer satisfaction is challenging for transit operators due to the heterogeneity in passengers spatial locations and travel patterns. Passenger behaviours and needs vary across different market segments from transit commuters to infrequent passengers, or from adults to students and senior passengers (Kieu et al., 2015). Complex urban structures of spatially diverse educational, recreational and occupational locations further nurture the diversity of these segments and complicate service provisions. Ridership competition from private transport and new shared-mobility transport such as carsharing and ridesourcing (e.g. Uber or Lyft) also forces transit operators to understand more about their customers and cater individual needs.

Market segmentation is a popular procedure in economics to classify a market of customers into segments sharing similar interests, needs or locations. Market segmentation is essential for transit operators to understand the commonalities and heterogeneities among different groups of passengers. It aims to define specific subsets of passengers sharing similar characteristics in demographics, psychograpic, geographic or behavioural so that transit operators can break down the mobility requirements of everyone and align their services to specific needs. Understanding the individual needs will enable operators to implement: (1) targeted surveys to understand specific groups of passengers, (2) incentives and personalised transit service to reward loyal passengers, and (3) on-

* Corresponding author.
E-mail addresses: leminh.kieu@data61.csiro.au (L.M. Kieu), yuming.ou@data61.csiro.au (Y. Ou), chen.cai@data61.csiro.au (C. Cai).

https://doi.org/10.1016/j.trc.2018.03.003
Received 8 August 2017; Received in revised form 28 February 2018; Accepted 2 March 2018
0968-090X/

demand services for under-served areas or during incidents.

The most basic form of transit market segmentation is through age and social situations, which are usually Adult, Senior, Child, and Student. While this level of market segmentation is useful for ticketing purposes, it provides unbalanced segments where customers do not share similar interests, needs or locations. Research on transit market segmentation started early with a study by Tybout et al. (1978), but has only recently been attracting a great interests with the proliferation of automated data collection systems such as Smart Card Automated Fare Collection (AFC). Smart Card AFC data captures a rich information that can potentially reveal a more comprehensive understanding of passenger travel patterns and mobility needs. Literature of data-driven studies using Smart Card data has evolved from simple data enriching studies (Alfred Chu and Chapleau, 2008; Bagchi and White, 2005), to mining individual temporal-spatial travel patterns (Kieu et al., 2015; Kusakabe and Asakura, 2014; Ma et al., 2013) and recently to improvements of transit operation, such as predicting passenger flow (Kieu et al., 2017; Li et al., 2017; Ma et al., 2014) or vehicle arrival time estimation (Min et al., 2016; Zhou et al., 2017).

Literature offers a number of approaches to transit market segmentation using Smart Card AFC data. Agard et al. (2006) adopted the Hierarchical Ascending Clustering to segment transit passengers using only temporal travel patterns. Lathia et al. (2013) applied Agglomerative Hierarchical Clustering to segment passengers using temporal travel profiles, aggregated in five daily time periods. Kieu et al. (2015) adopted a bi-level Density-based Scanning Algorithm with Noise (DBSCAN) (Ester et al., 1996) to mine individual travel patterns and then proposed *a priori*segmentation method to segment transit passengers. Legara and Monterola (2017) introduced the concept of eigentravel matrices to capture passenger travel characteristics and developed a classification technique with promising accuracy. Langlois et al. (2016) inferred passenger travel patterns through a longitudinal representation of multi-week activities. Passenger travel areas were clustered using Agglomerative Hierarchical Clustering, and then longitudinal travel patterns were clustered by principal component analysis. Briand et al. (2017) clustered transit passengers of Gatineau City, Canada using a Gaussian mixture model with Classification Expectation Maximisation. The proposed model is capable of cluster passengers based on continuous temporal travel activities. A model-based mixture model using Expectation Maximisation was also proposed in El Mahrsi et al. (2017), where the authors proposed two approaches to cluster transit passengers from a station-oriented and a passenger-focused standpoints using their temporal travel patterns.

Understanding the geographic market segments in the transit industry is indeed essential because transit service provision is spatially limited. Transit passengers usually walk to stops, thus their rational travel choices often limit within a walking distance. Transit operators who can leverage such spatial understanding about passenger segments will be able to provide better services. For instance, travel information can be given to passenger segments at areas influenced by an incident. The impact of future transit management plans can be foreseen given the passenger segments on the impacted areas. Spatial transit market segmentation is also helpful for passenger choice modelling, such as modal and route choices, because passengers of the same segment living closely are sharing similar travel behaviours and facing a similar choice set.

However, the consideration of individual geographical characteristics in existing passenger segmentation studies is relatively limited. In the conclusion of their work, Agard et al. (2006) recognised that the inclusion of geospatial trip behaviour would enable better understanding of transit supply and demand. In Lathia et al. (2013), the spatial characteristics of each passenger are considered as the number of visited stations, rather than the spatial proximity of adjacent stations. Langlois et al. (2016) considered spatial proximities when clustering the user-specific areas using a predefined threshold distance value among stops/stations. There remains a need for integration of spatial and behavioural features in an integrated spatial-behavioural transit market segmentation.

In this paper, we define the concept of spatial-behavioural passenger segments as the spatially-limited clusters of passengers who have similar behaviours. Identifying these spatial-behavioural segments is challenging. First, a new distance metric will be required to incorporate spatial and behavioural features in segmentation, because they are in different units of measures. Second, since transit service provision is spatially limited, it is important to predefine a maximum spatial size value Δ for each passenger segment. Segment's spatial size measures the maximum distance from any two passengers belong to the same segment. If Δ is large, passengers might be less spatially relevant to each other, but there might be more chances to find passengers of similar travel behaviours. For instance, let us say that a transit provider wants to find spatial segments of regular passengers in an area to provide a coach service. To maximise the utility of each vehicle and limit the boarding time, the transit provider may stop the coach only at the centre of each spatial-behavioural passenger segment, and ask the passengers to walk to those predefined stops. In this example if the spatial size is large, the walk would be more tiresome, though may be more regular passengers can be grouped together to maximise vehicle's utility. If the spatial size is small, the average walking distance would be convenient, but there might also be less passengers of similar behaviours in each spatial-behavioural passenger segment. Too small Δ may even lead to spatial singletons, those are segments at a single location, which is unfavourable from an efficiency point of view. It is challenging to define Δ because the data-driven passenger segmentation results depend on the spatial and behavioural distributions of passengers' characteristics. Choosing a value of Δ may require inputs from an expert with domain-knowledge, since an increase in Δ may lead to an increase in both similarity among passengers of the same segment, and an increase in passengers' walking distance. Multiple implementations of the transit market segmentation algorithm with different values of Δ may be required.

The search for a value of Δ leads us to another problem in large-scale transit market segmentation: scalability. Existing literature in transit market segmentation have been often developed using a random sub-sample of limited size, such as in Lathia et al. (2013), especially when the number of market segments are usually not known, so that a hierarchical method is required (Lathia et al., 2013; Langlois et al., 2016; Agard et al., 2006) or multiple runs with different number of segments are required (Briand et al., 2017; El Mahrsi et al., 2017). Kieu et al. (2015) is an exception where over a million of Smart Cards were considered, but the segmentation method is rather simplistic with a set of *a priori* heuristic rules. Table 1 shows the number of Smart cards, number of trips considered and method adopted in existing transit market segmentation studies.