Transportation Research Part C 58 (2015) 193–207

Contents lists available at ScienceDirect

Transportation Research Part C

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

A modified Density-Based Scanning Algorithm with Noise for spatial travel pattern analysis from Smart Card AFC data

Le-Minh Kieu *, Ashish Bhaskar, Edward Chung

Smart Transport Research Centre, School of Civil Engineering and Build Environment, Science and Engineering Faculty, Queensland University of Technology, Brisbane, Australia

A R T I C L E   I N F O

Article history:
Received 12 July 2014
Received in revised form 6 March 2015
Accepted 18 March 2015
Available online 6 April 2015

Keywords:
Spatial travel pattern
Public transport
Smart Card
AFC
DBSCAN

A B S T R A C T

Smart Card Automated Fare Collection (AFC) data has been extensively exploited to understand passenger behavior, passenger segment, trip purpose and improve transit planning through spatial travel pattern analysis. The literature has been evolving from simple to more sophisticated methods such as from aggregated to individual travel pattern analysis, and from stop-to-stop to flexible stop aggregation. However, the issue of high computing complexity has limited these methods in practical applications. This paper proposes a new algorithm named Weighted Stop Density Based Scanning Algorithm with Noise (WS-DBSCAN) based on the classical Density Based Scanning Algorithm with Noise (DBSCAN) algorithm to detect and update the daily changes in travel pattern. WS-DBSCAN converts the classical quadratic computation complexity DBSCAN to a problem of sub-quadratic complexity. The numerical experiment using the real AFC data in South East Queensland, Australia shows that the algorithm costs only 0.45% in computation time compared to the classical DBSCAN, but provides the same clustering results.

1. Introduction

Smart Card Automated Fare Collection (AFC) system has been increasingly popular in public transport, providing a massive quantity of continuous and dynamic data on passenger boarding and alighting locations. This information provides a tremendous opportunity to analyze spatial travel patterns of the transit users-defined in terms of the regular boarding and alighting stops of transit passengers. An emerging number of studies have extensively explored multiday travel pattern (Chu and Chapleau, 2010; Kieu et al., in press; Ma et al., 2013) to understand individual travel behaviors (Ma et al., 2013), passenger segmentation (Kieu et al., in press), trip purpose (Lee and Hickman, 2014) and potential in transit planning (Utsunomiya et al., 2006). A detailed review of existing advances in AFC analysis could be found in Pelletier et al. (2011).

Spatial travel pattern is defined in this paper as regular origin–destination (OD) that transit passenger usually travels between. The literature of travel pattern analysis using AFC data has been evolving from simple to more sophisticated methods such as from aggregated to individual travel pattern analysis, and from stop-to-stop to flexible stop aggregation. Although considerable research has recently been devoted to capturing the individuality and travel behaviors of transit passengers, rather less attention has been paid to the practical computing constraints which limit those methods in real-world applications. So far, existing method have been confined to first-time analysis of spatial travel pattern – i.e. finding the travel pattern from AFC data without any prior knowledge of passenger travel pattern, leaving the question of updating this

* Corresponding author.
E-mail addresses: leminh.kieu@student.qut.edu.au (L.-M. Kieu), ashish.bhaskar@qut.edu.au (A. Bhaskar), edward.chung@qut.edu.au (E. Chung).

http://dx.doi.org/10.1016/j.trc.2015.03.033

information from the existing travel pattern knowledge unanswered. To make the best use of the individual travel pattern in customized service provision and operational strategies, travel pattern has to be updated daily to observe the changes in passenger behaviors. After the last service of the day, transit operator collects all Smart Card transactions of the day and updates travel pattern of each individual passenger before the first service of the next day. However, existing methods have been developed with increasing complexity and degree of detail, while the number of Smart Cards and transit journeys are also growing rapidly. A full spatial travel pattern analysis would be an absurd task to perform within this short time gap of a few minutes to hours. Consequently, there is a need for a time-effective algorithm to observe and record the evolution of travel pattern.

This paper proposes a new algorithm named Weighted Stop Density Based Scanning Algorithm with Noise (WS-DBSCAN) based on the classical Density Based Scanning Algorithm with Noise (DBSCAN) aiming to rapidly detect and update individual spatial travel patterns while maintains high degree of detail in travel pattern analysis, which enables transit operators to use this information on a daily basis. This research focuses on spatial travel pattern analysis only, because spatial pattern analysis as a two-dimensional problem has much more complexity than temporal travel pattern. We also focus on developing a new algorithm to detect and update individual travel pattern, without actual analysis to mine spatial and temporal travel patterns from transit passengers.

The paper first reviews the existing studies on travel pattern analysis using AFC data. After the description of the data processing method, the paper describes the DBSCAN implementation in travel pattern analysis. The WS-DBSCAN algorithm along with an example of its implementation is then presented. The numerical experiment shows the effectiveness of the proposed method compared to the classical DBSCAN algorithm. A discussion and future research directions finally concludes the paper.

2. Literature review

The use of AFC data enables us to continuously analyze the multiday travel patterns of a much larger population than the traditional travel survey method. The existing studies in the literature have explored travel pattern by different level of aggregation on passenger and stop level.

2.1. Existing passenger aggregation approaches for travel pattern analysis

An emerging number of publications have recently analyzed transit passenger travel pattern by different level of aggregation from whole aggregated dataset to each individual Smart Card user. Utsunomiya et al. (2006) is an example of aggregated dataset analysis. The authors described the data possessing and analysis methods to mine meaningful information from AFC data. Jang (2010) demonstrated the use of AFC data in travel time and transfer locations analysis. The method facilitates the comparison between different transit modes and the identification of passenger transfer choices. Hasan et al. (2013) exploited AFC data to observe both spatial and temporal passenger travel pattern. The authors modeled two important passenger decisions: (a) which place to visit (by assuming a fixed probability of visit to each regular place) and (b) how long to stay (by a hazard based duration modeling). The whole dataset analysis explores general travel patterns from transit passengers.

Some other authors emphasized the similarity of travel pattern by subgroup. Their analyses are based on the aggregation of several similar characteristics of the transit trip and passenger. Morency et al. (2007) aggregated the Smart Card users into five classes according to the card type and the privilege of route usage. The travel profile of each card type could be well observed by investigating the indicators of spatial and temporal travel pattern. Chu et al. (2009) proposed a new framework to mine spatial–temporal distribution of transit demand by different aggregation level such as stop, route, link, node and card type. Lee and Hickman (2014) developed a heuristic rules algorithm and a classification decision tree to group Smart Card users into multiple classes and infer their trip purposes.

Although aggregated travel pattern analysis provides insights into the travel pattern of general user, it fails to capture the individuality of travel behavior. The typologies passenger groups are also predefined which might not reflect the similarity of passengers between the same class, and the difference between classes.

Several studies have recently enriched the travel pattern comprehension by individually analyzing each Smart Card user. Chu and Chapleau (2010) described a disaggregated travel pattern analysis framework for multi-day AFC data. "Anchor points" or repeated travel locations are mined from each Smart Card user and then assigned to known spatial coordinates. Ma et al. (2013) and Kieu et al. (in press) used the classical DBSCAN algorithm, originally proposed in Ester et al. (1996), to mine spatial and temporal travel patterns from AFC data.

2.2. Existing stop aggregation approaches for travel pattern analysis

Travel pattern analysis often spatially breaks down to stop-to-stop repeated journeys. However, the limitation of this method has been identified by several authors (Lee and Hickman, 2013). A transit stop is usually linked with only a single direction or route, while transit passengers normally have several route choices options within their origin destination locations. Any stops within the immediate vicinity that provide the same access should be considered in the same travel pattern,