ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

IJARMATE
International Journal of Advanced Research in Management, Architecture, Technology and Engineering (IJARMATE) Vol. 2,Special Issue 6, March 2016

ANALYSIS OF SEGMENTING METRO TRIPS USING SMART CARD DATA

S.NAGARESHMI
PG Student
Department of ECE
Rajas Engineering College
Vadakkangulam.
Tamil Nadu. India
reshmis94@gmail.com

C.ANNA PALAGAN
ASSISTANT PROFESSOR
Department of ECE
Rajas Engineering College
Vadakkangulam.
Tamil Nadu. India
Mobile : +919442708891
canna_palagan7467@rediffmail.com

Abstract—This paper investigates an important problem: how to extract spatio-temporal segmentation information of trips inside a metro system. More specifically, for a given trip, we want to answer several key questions: how long it takes for a passenger to walk from the station gantry to the station platform? How much time he/she waits for the next train? How long he/she spends on the train? How long it takes to transfer from one line to another? etc. This segmentation information is important for many application scenarios such as travel time prediction, travel planning, and transportation scheduling. To the best of our knowledge, we are the first to give a practical solution to this important problem. By analyzing the tap-in/tap-out events pattern, our intuition is to pinpoint some special passengers whose transaction data can be very helpful for segmentation. A novel methodology is proposed to extract spatio-temporal segmentation information: first for non-transfer trips by deriving the boarding time between the gantry and the platform, then for with-transfer trips by deriving the transfer time. Evaluation studies are based on a large scale real-system data of Shenzhen metro system, which is one of the largest metro systems in China and serves millions of passengers daily. On-site investigations validate that our algorithm is accurate and the average estimation error is only around 15%.

Keywords:Spatio-temporal,metro trips,smart card,data mining,RFID,segmentation.

I. INTRODUCTION

Contactless smart card systems have gained universal prevalence in modern metros. In addition to its original goal of ticketing,the large amount of transaction data collected by the smart card system can be utilized for many operational and management purposes. This segmentation information is important for many application scenarios such as travel time prediction, travel planning, and transportation scheduling. However, in reality we only assume that only each trip's tap-in and tap-out time can be directly obtained; all other temporal endpoints of segments are unknown. This makes the research very challenging. To the best of our knowledge, we are the first to give a practical solution to this important problem. By analyzing the tap-in/tap-out events pattern, our intuition is to pinpoint some special passengers whose transaction data can be very helpful for segmentation. A novel methodology is proposed to extract spatio-temporal segmentation information: first for non-transfer trips by deriving the boarding time between the gantry and the platform, then for with-transfer trips by deriving the transfer time.We investigate an important problem: how to extract spatio-temporal segmentation information of metro trips by only utilizing the tap-in and tap-out information. To the best of our knowledge, we are the first to provide a practical solution to this important problem. We first propose a set of novel algorithms to identify Border-Walkers by analyzing the tap-in/tap-out events pattern. The transaction data collected by the smart card system canbe utilized for many operational and management purposes. For metro systems, usually both the tap-in and tap-out records for each trip are saved in the database. Each record contains at least the time, the station and the card's ID of the transaction.

This paper compares different models and simulation approaches for an energetic simulation of an automatic subway. For this purpose, several models are carried out from a dynamic model, which is validated by comparison with experimental measurements. Furthermore, two different simulationapproaches are compared, i.e., backward and forward approaches. A simplified model is obtained and allows the reduction of the simulation time by 96 compared with the dynamic model by keeping an accuracy value of more than 99%. Nowadays, metro systems have become the most preferred public transit services in many cities. Compared with other services, metro has the benefits of high efficiency, large volume, and fast speed. In Shenzhen one of the four largest cities in China the number of metro passengers reaches 2.5 million daily, which is around one-third of the total public traffic.

In this paper, we analyze and understand the spatiotemporal travel patterns of individual passengers in Shenzhen, China. The prevalence of smart card fare collection provides a unique opportunity for this study. By the end of 2013, the number of public transit smart card holders in Shenzhen has reached 10 million; these smart cards can be used for both bus and metro systems. The total traction force is thus produced on each wheel to impose the velocity of the subway.Each transaction, in addition to revenue collection, may provide information about how many people use the transport system, where, when, on what route, and even more. Most often in a transit network the data describe for each boarding, the exact time, some precision on location and some extra information about the card itself type of card, type of ticket fare, period of validity, geographic area of validity, and so on. The total traction force is thus produced on each wheel to impose the velocity of the subway. A dynamic model, which reflects the actual operation of the vehicle, is obtained in this part. This model will be taken as a reference for the simplified models. Some assumptions are made, i.e., no slipping phenomena and no curve. Then this paper proposes to take advantage both of data mining methods and public transport planning models in order to describe the regularity in user’s behaviour on a transit network.

II. UNDERSTANDING TEMPORAL AND SPATIAL TRAVEL PATTERNS

Metro systems have become the most preferred public transit services in many cities. It is important to understand individual passengers' spatio-temporal travel patterns inside metro. More specifically, for a specific passenger: what is the temporal access pattern? What is the spatio access pattern? is there any relationship between the temporal and spatio patterns? is this passenger's patterns normal or special? Answer all these questions can help us understanding the major reasons of why thispassenger takes metro. In this paper, analysing patio temporal travel patterns of individual passengers in Shenzhen, China. A systematic approach is proposed to extract temporal,spatial and anomaly features related to metro passengers. We analyze one month smart card datacollected from Shenzhen. Combined with bus transaction data, we give an in-depth analysis and explanations for different groups. These experiments first show that the behaviours of regular transit users evolve with time both in terms of transit stops frequented and time of boarding. Hence, variability of behaviours also changes for various user types.

The contribution of this paper includes the following points given below:

a) A systematic approach is proposed to extract temporal, spatial and anomaly features related to metro passengers. We analyze one month smart card data collected from Shenzhen. Combined with bus transaction data, we give an in-depth analysis and explanations for different groups.

b) Analyse the advantages and disadvantages of smart card fare collection systems compared with traditional approaches. propose an algorithm to infer passengers’ get-off site from smart card data.

A. ANALYZING PASSENGER DENSITY FOR PUBLIC BUS

Besides the major goal of carrying passengersaround, providing a comfortable travel experience for passengers is also an important business consideration. The crowdedness inside a bus can directly affect the number of people choosing thebus. Traditional approaches to obtain passenger density rely on field investigations, which are both non-scalable and incomplete. The wide adoptions of smart card fare collection systems and GPS tracing systems in public transportation provide new opportunities. In this paper, we associate these two independent datasets to derive the passenger density, and evaluate the effectiveness of schedulingchoices. To our best knowledge, this is the first paper which utilizes smart card data and GPS data to calculate the passenger density of bus service.

Several researches have been conducted using probe vehicle technique that is used to collect traffic data in real-time in order to determine the nature of traffic. Most of the probe vehicle techniques are used for determining the link travel time. Where they make use of GPS technology. But the major problem with GPS is that the accuracy of a typical GPS receiver is about 10 meters. This makes it difficult to pin-point a crossing for the purpose of congestion measurement. Secondly, it has been noticed that GPS sends erroneous velocity data even when the vehicle is stationary. Road traffic congestion detection and management has been a challenge and several countries have come up with new concepts and ideas to detect congestion. Today, number of vehicles is increasing at an alarming rate and improvement of road infrastructure, which has its limitation, has not been able to ease the problem. Over the decade, many innovative concepts and technologies have been developed in relation to collecting real time traffic data and use it for gaining knowledge regarding the various aspects of traffic flow.

B. PASSENGER FLOW ASSIGNMENT IN METRO NETWORKS

An integrated Bayesian statistical inference framework to characterize passenger flow assignment model in a complex metro network. In doing so, we combine network cost attribute estimation and passenger route choice modelling using Bayesian inference. We build the posterior density by taking the likelihood of observing passenger travel times provided by smart card data and our prior knowledge about the studied metro network. Given the high-dimensional nature of parameters in this framework, we apply the variable-at-a-time Metropolis sampling algorithm to estimate the mean and Bayesian confidence interval for each parameter in turn. As a numerical example, this integrated approach is applied on the metro network in Singapore. Our result shows that link travel time exhibits a considerable coefficient of variation about 0.17, suggesting that travel time reliability is of high importance to metro operation. The estimation of route choice parameters conforms to previous survey-based studies, showing that the disutility of transfer time is about twice of that of in-vehicle travel time in Singapore metro system.Compared with other transport modes, metro systems have dedicated and exclusive rail-based infrastructures, making it possible to provide superior service with higher speeds and larger capacity. Due to their superiority, metro systems not only attract but also suffer from high passenger demand – especially during rush hours when passenger demand exceeds its designed capacity for not only trains, but also platforms.

III. PROPOSED SYSTEM

The aim of this paper to provide a practical solution for the problems utilizing the tap in and tap-out information in a railway station. The input given is just the number of users entering the station. Initially RFID readers are used to identify the persons with different tags. As they enter the station, readers are also placed in platforms and trains. Here the numbers of users entering the station and reaching the trains are identified. The person moving from the entrance to the platform is calculated as L1.Platform to train is calculated as L2.origin to destination is calculated as L3. If transferred to another train, moving from one train to other is calculated as L4.moving from train to exit is calculated as L5.In peak hours. There is no way to find the number of persons travelling in a particular train. So to detect the edges, We use spatio temporal technique that identify the density of crowd. The values are passed to the control room using GSM technique.

A social phenomenon is called “The Familiar Stranger”. For example, Alice and Bob have different Origin-Destination pairs; they live in different districts and work in different districts. However, they happen to transfer at the same station at approximately the same time every workday; they are familiar with each other, and they have feelings about each other, while none of them dare to make the first contact.“Metro Acquaintance” is an ongoing project with the objective of establishing such a special social network. Directly from historical records, by travel segmentation, we calculate the meet probability, when and where, that each pair of passengers have met each other, either in the same platform, or in the same train. By enter their card ID in the web entry, people can establish a metro social network; the connection recommendation of the social system is mostly based on the meet probability deduced by the travel segmentation algorithms., The Spot on approach estimates the distance between transponder, exploits the density of the transponders, and uses multiple RSS measurements to improve the localizing resolution. Whereas the LANDMARC system uses so called reference tags to get more reference points within the system. The RSS information is evaluated between every tag-to-reader connection However, both systems, Spot on and LANDMARC, use active RFID tags leading to higher costs per tag manufacturing and service.

The block diagram of the proposed system to experiment the spatio temporal segmentation of passengers in a railway station using metro trips is shown in figure 1.

1.RFID CARD

A significant advantage of RFID devices over the others mentioned above is that the RFID device does not need to be positioned precisely relative to the scanner. We're all familiar with the difficulty that store checkout clerks sometimes have in making sure that a barcode can be read. And obviously, credit cards and ATM cards must be swiped through a special reader.In contrast, RFID devices will work within a few feet up to 20 feet for high-frequency devices of the scanner. For example, you could just put all of your groceries or purchases in a bag, and set the bag on the scanner. It would be able to query all of the RFID devices and total your purchase immediately. Read a more detailed article on RFID compared to barcodes. RFID technology has been available for more than fifty years. It has only been recently that the ability to manufacture the RFID devices has fallen to the point where they can be used as a "throw away" inventory or control device. Alien Technologies recently sold 500 million RFID tags to Gillette at a cost of about ten cents per tag. One reason that it has taken so long for RFID to come into common use is the lack of standards in the industry.

ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

International Journal of Advanced Research in Management, Architecture, Technology and Engineering (IJARMATE) Vol. 2,Special Issue 6, March 2016

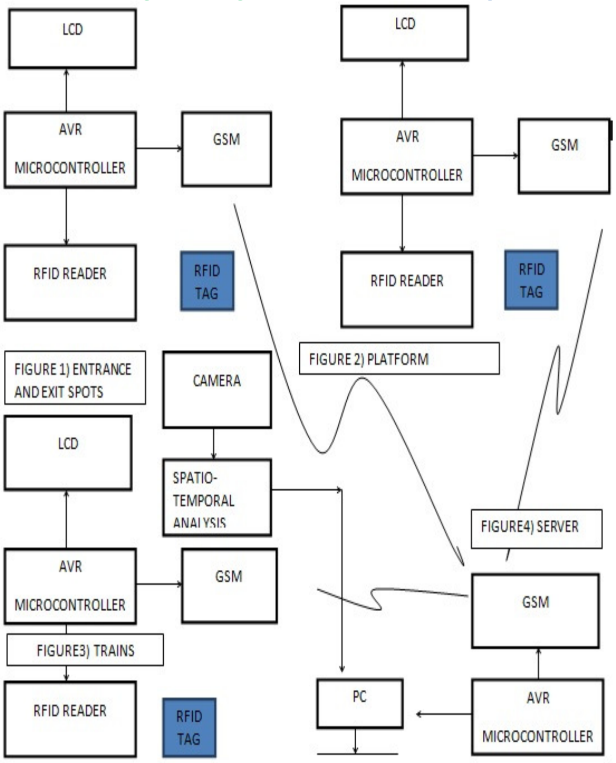

Figure 1. Metro trips operation

Most companies invested in RFID technology only use the tags to track items within their control; many of the benefits of RFID come when items are tracked from company to company or from country to country.The Common Problems with RFID are some common problems with RFID are reader collision and tag collision. Reader collision occurs when the signals from two or more readers overlap. The tag is unable to respond to simultaneous queries. Systems must be carefully set up to avoid this problem.

2. PROCESSING SYSTEM

Our spatio-temporal segmentation algorithm processes large amount of data, and correspondingly, requires intensive Sorting and grouping operations.The implemented system has three layers: the data layer, the model layer, and the application layer.

3. FLOWCHART

Using these special passengers' data, we propose a novel methodology to extract spatio-temporal segmentation information both for non-transfer trips by deriving the boarding time between gantry and platform, and for with transfer trips by deriving the transfer time. We did find abnormal values, which can be several hours. The classic DBSCAN algorithm is again adopted to cluster the points. After excluding the exceptions, the minimum value in the cluster is chosen as the optimal candidate for boarding time or transfer time.By travel segmentation, we can already accurately estimate the historic spatial-temporal density of every operating train from the records.

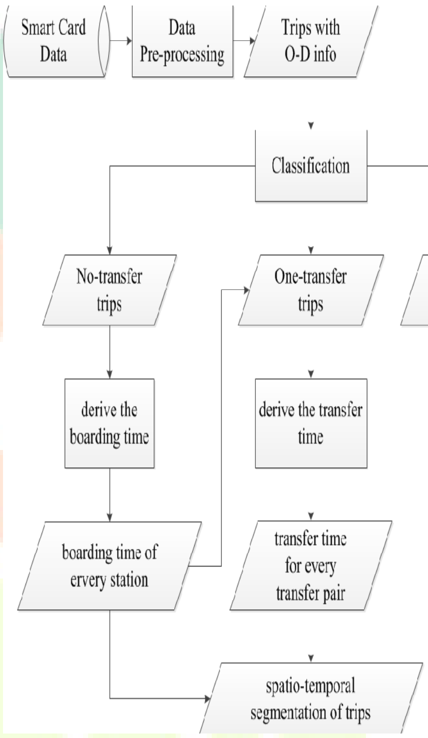

Figure 2. Flowchart of using metro trips

3.1SMARTCARD DATA

Every trip contains one tap-in event and one tap-out event; this step joins them together out ofthe transaction data by matching card ID and time. Also, the redundancy should be removed and the inconsistence shouldbe solved in smart card data.The smartcard data contains details about the passengers travel history and travel duration in the station and train travel.

3.2CLASSIFICATION

This step is to divide the trips from step 1 into three categories: non-transfer trips, one-transfer trips, multi transfer trips. As mentioned above, only non-transfer and one transfer trips are used for boarding transfer time extraction.

3.3DATASET

Our dataset is the metro transaction data from Shenzhen, China. There have been over 10 million public transit smart cards issued, and these smart cards can be used for both the bus and metro systems.

TableII: Transaction record Format

Field	Value
CardID	Anonymous unique card id
TrmnlID	For metro, it represents the stop id
Transact Time	Transaction date and time
Transact Type	Transaction Type

3.4DATA PRE PROCESSING

Every passenger can have many trips, hence the first step is to find all trips belong to a specific passenger. Each metro trip contains one boarding and one alighting records. From all transaction records, each pair of two records is joined together by matching smart card id and time. The second step of pre processing is to filter out passengers that rarely take metro. The distribution of passenger'sversus the total number of active days of taking metro trips. It is clear that about 80 percent of passengers have active days of taking metro less than 7 within 21 weekdays. However, the top 20 present passengers possess 68 present of total transactions, as the distribution of transactions.For passengers seldom travel, there is not enough information to reveal the temporal or spatial characteristic of them; in this paper we only analyze a passenger of his/her card's number of active days is more than 6.

3.5TEMPORAL FEATURES EXTRACTION

Temporal analysis uses n features to express temporal characteristics of a passenger. There are two requirements. First, the chosen n features should convey temporal informationas much as possible. Second, the n value should be as small as possible to improve the scalability of the analysis.There are two observations. First, travels are regular in certain time periods: most days there are trips in two periods, i.e., 08:00_10:59 and 18:00_20:59. Second, there are variations in each period: although there are always trips in the 08:00_10:59 periods, the start time and duration still have variations. Central idea of our temporal feature extraction is to divide time into sequential and overlapped slots.

3.5 SPATIAL AND ANOMALY FEATURES EXTRACTION

From the spatial perspective, we also want to extract m features, or more specifically, m OD Origination-Destination pairs. We first order the frequency of OD pairs in a descending order. The selection rules for parameter m is similar to temporal parameter n: the certain threshold denoting the minimum coverage of information required for spatial features are set to 85%. We conclude that m can be set to 4 in this paper. There are two kinds of obvious abnormal trips. One anomaly is that for a specific trip, the time used from station A to station B is much more than other trips with the same OD pair.

IV. FUTURE WORK

Crowd is a unique group of individual or something involves community or society. The phenomena of the crowd are very familiar in a variety of research discipline such as sociology, civil and physic. Nowadays, it becomes the most active-oriented research and trendy topic in computer vision. Traditionally, three processing steps involve in crowd analysis, and these include pre-processing, object detection and event/behaviour recognition.Meanwhile, the common process for analysis in video sequence of crowd information extraction consists of Pre-Processing, Object Tracking, and Event/Behaviour Recognition. In terms of behaviour detection, the crowd density estimation, crowd motion detection, crowd tracking and crowd behaviour recognition are adopted. In this paper, we give the general framework and taxonomy of pattern in detecting abnormal behaviour in a crowd scene.

V. CONCLUSION

We have to investigate an important problem: how to extract spatio-temporal segmentation information of metro trips by only utilizing the tap-in and tap-out information. To the best of our knowledge, we are the first to provide a practical solution to this important problem. We first propose a set of novel algorithms to identify Border-Walkers by analyzing the tap-in/tap-out events pattern; based on that, we further propose a novel methodology to extract spatiotemporal segmentation information, first for non-transfer trips by deriving the boarding time between gantry and platform, then for with-transfer trips by deriving the transfer time. We study our approach in the case of Shenzhen metro system and perform a large scale on-site investigation. The on-site measured results validate that our algorithm is accurate and the average estimation error is only around 15%.

REFERENCES

1] J. Zhao, C. Tian, F. Zhang, C. Xu, and S. Feng, "Understanding temporal and spatial travel patterns of individual passengers by mining smart card data," in Intelligent Transportation Systems (ITSC), 2014 IEEE 17thInternational Conference on. IEEE, 2014, pp. 2991–2997.

2)B. Agard, C. Morency, and M. Tr´epanier, "Mining smart card data from an urban transit network." 2009.

ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

International Journal of Advanced Research in Management, Architecture, Technology and Engineering (IJARMATE) Vol. 2,Special Issue 6, March 2016

3)M.-P. Pelletier, M. Tr´epanier, and C. Morency, “Smart card data use in public transit: A literature review,” Transportation Research Part C:Emerging Technologies, vol. 19, no. 4, pp. 557–568, 2011.

4)L. Liu, A. Hou, A. Biderman, C. Ratti, and J. Chen, “Understanding individual and collective mobility patterns from smart card records: A case study in shenzhen,” in Intelligent Transportation Systems, 2009.ITSC’09. 12th International IEEE Conference on. IEEE, 2009, pp. 1–6.

5)Y. Gong, Y. Liu, Y. Lin, J. Yang, Z. Duan, and G. Li, “Exploring spatiotemporal characteristics of intra-urban trips using metro smartcard records,” in Geoinformatics (GEOINFORMATICS), 2012 20th InternationalConference on. IEEE, 2012, pp. 1–7.

6)L. Sun, D.-H. Lee, A. Erath, and X. Huang, “Using smart card data to extract passenger’s spatio-temporal density and train’s trajectory of mrtsystem,” in Proceedings of the ACM SIGKDD International Workshopon Urban Computing. ACM, 2012, pp. 142–148.

7) Huang, L. Wang, C. Tian, F. Zhang, and C. Xu, “Mining freight truck’s trip patterns from gps data,” in Intelligent Transportation Systems(ITSC), 2014 IEEE 17th International Conference on. IEEE, 2014, pp. 1988–19

8)B. Li, “Markov models for bayesian analysis about transit route origin– destination matrices,” Transportation Research Part B: Methodological,vol. 43, no. 3, pp. 301–310, 2009.