IOSR Journal of Engineering (IOSRJEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
Vol. 09, Issue 10, October. 2019, Series -IV, PP 01-05

A Survey on Detecting Pickpocket Suspects from Large-Scale Public Transit Records-Recent Approaches

Sameera.P[1], Ms.G.Maheshwari[2]
M.Phil Scholar, Department of Computer Science, Sree Narayana Guru College, Coimbatore, Tamil Nadu, India[1]
Assistant Professor, Department of Computer Science, Sree Narayana Guru College, Coimbatore, Tamil Nadu India[2]
Received 16 October 2019; Accepted 31 October 2019

Abstract: Data mining is useful in a wide variety of applications, the most unique and popular applications such as anomaly/suspect detection from the large surveillance system. The abnormality detection is performed on various datasets, which has unique challenges and issues. Outlier detection from the Global Positioning System (GPS) and dynamic trajectory data is much more complicated due to its dynamic update nature. This paper reviews the related works of various suspect, anomaly and outlier detection schemes on mobility datasets. Several works proposed in the literature to detect the abnormality from the traveling behavior which is collected from various wireless media. The unique features among high dimensional dynamic datasets are always developed different types of issues. Detecting and analyzing such dynamic features for finding anomaly is carried out by many types of research in different ways such as cluster oriented, context-aware, dimensionality reduction techniques, graph approaches. After a successful analysis of these approaches, this survey gives an outline to tackle the problems of those techniques.

Index Terms- Outlier Detection, Trajectory Behaviors, Mobility Patterns, Public Safety, Anomaly Detection.

I. INTRODUCTION

The public transportation system passengers have been the major aim for pickpockets. In several cities, thefts occur regularly in public transportation systems, for the reason that passengers tend to pay less attention to their things when they are in a rush or in a crowded environment. Many other big cities in the world are also reported to suffer from the pickpocket problem, which has led to public safety concerns. Certainly, it is challenging to detect theft behaviors committed by cunning thieves who know how to run away without being disclosed. In spite of the considerable cost in manpower and resources, many thieves are still at greater.

It is dangerous to provide a smart observation and tracking tool for the security personnel of the transportation systems. The improving technologies like information technology and data processing have capabilities; through transactional records collected by automated fare collection (AFC) systems have become valuable for understanding passengers' mobility patterns and the urban dynamics. However, most of the existing studies focused on identifying regular, collective mobility patterns, such as commute flows and transit networks. Our study is the first to focus on identifying thieves based on AFC data. Actually, it is probable to identify thieves using AFC records because behavioral differences are coined in the mobility footprints, which can help to separate suspects from regular passengers. Such characteristics, which can make suspects detect, involve traveling for an extended length of time, making unnecessary transfers, and/or wandering on certain routes while making random stops.

Through, finding thieves based on AFC records is not an easy outlier detection problem. We can see a number of trajectories between hot regions A and B. By precise examination, we can see that most passengers move from one place to another place using near-optimal pattern (e.g., shortest time/distance, or a minimal number of transfers).

In the review, to recognize thieves from AFC records, we are faced with a number of inbuilt challenges.

The initial challenge is how to recognize useful attributes to differentiate thieves from normal passengers. These features should not only help us understand the behaviors of pickpockets, but also help us build a suspect detection and tracking system for supporting the security personnel.

The other way, using normal outlier detection methods tends to outcome in a large number of fake positives. Especially, not every trip made by a normal passenger looks normal. Normal commuters may irregularly make trips to visit friends or places of interest, and some of such trips may look suspicious by how much they deviate from regular behaviors.

And other way, a large number of AFC records are being gathered from number of passengers, only a tiny fraction of which are pickpockets. Identifying such a small group of people in such a large-scale dataset is like looking for a needle in the hashtag.

Finally, we also need to effectively transform our knowledge based on model development into a decision support system, so that real-time, personalized deployment recommendations could be made to help to guide security personnel to perform their work more efficiently.

Finally, in this approach, a comprehensive is taken to meet the previous challenges. Specially, we first construct a feature representation for profiling passengers. In addition, we elaborate a two-step framework to divide regular movement patterns from irregular behaviors, and eventually, distinguish thieves from regular passengers. Lastly, we influence real-world datasets from several sources for model training and validation, and implement a prototype system for end users.

We first partition the city area into regions with functional categories. Then, the mobility behaviors of passengers are separated from transit records and incident reports. Furthermore, we construct a person mobility database to store the profile of each passenger. After that, we develop our framework by normal passenger filtering and suspect detection. Finally, the user feedback information, such as newly confirmed thieves, will be entered as ground truth for future model training.

Transit Records

In this study is based on a large-scale transit records dataset collected from a public transit system that includes buses and subways. Passengers utilizing the transportation service are calculated charge by the distance they travel. An enables secure access smart card is issued to each passenger, who has to swipe the card when they board or exit a vehicle. The AFC methods then measure the fare according to the stops of boarding and exiting. As a result, each raw AFC record consists of the smart card ID, the route number, the event (i.e., boarding or exiting), the station, and the time stamp. We transformed the information so that each transportation record consists of one boarding and one exiting event of the same ID. After deleting replicates and tremendously infrequent riders, we are left with over 1.6 billion records that involve nearly 6 million passengers.

With the purpose of this approach describe the data and consequent feature extraction process clearly; here we simplify two concepts, transportation records and trips, using a real example. A) Is the original trajectory on the city map; Part (b) Separates the trajectory into three split trips; and Part (c) describes the corresponding transportation records in our dataset.

Since the number of passengers and the movement of each passenger, the dataset is perpetually increasing. It becomes unfeasible for anyone without dedicated tools to have a complete view and to understand what is going on into the data. Also we believe that data mining tools may be powerful to extract such knowledge available from data acquisition systems. 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 behavior on a transit network. Such knowledge, for important sets of users, may provide important information about every day utilization and periodic evolutions. Also seasonality information may be discovered. The focus of the approach is on trips behaviors (in time). Also an interest is on travel behavior variability.

Data mining tools and applications

Data mining is a group of methods and tools devoted to the discovery of non-trivial, implicit, previously unknown, and potentially useful and understandable patterns from large data sets.

Different classifications of data mining tools are available. If we keep out word, sound and video mining, classical tools may be categorized as classification, estimation, segmentation, and description

- Classification is devoted to allocate labels to data based on preparations constructed on historical data.
- Estimation evaluates missing values of a record as a function of the fields in other records.
- Segmentation (or clustering) creates subsets in a population. Elements are placed in a subset to maximize the homogeneity of the subset and to maximize the heterogeneity between the different subsets.
- Description and realization are used to show the affairs among the information. Regular patterns in the data are noted as association rules (with some measures of regularity). Graphical projections and representation may emphasis particular characteristics.

II. LITERATURE REVIEW

In this Paper [1], the author describes the function of a metro station area is vital for city planners to consider when establishing a context-aware Transit-Oriented Development policy around the station area. Through, the functions of metro station areas are difficult to infer using the static land use allocation and other traditional review datasets. The described approach gather the features involving around the metro station catchment areas according to the patterns of staying behaviors derived from smart card data. We initially define the staying behaviors by the spatial and sequential constraints of the two consecutive alighting and boarding

records from the individual travel profile. After that we cluster and label the whole staying behaviors by considering the activities of duration, frequency, and start time. By predicting the percentage of special types of aggregated activities happening around each metro station, we cluster and explore the functions of the metro station area. Enhancing as a case study, we analyze the conclusion of metro systems and discuss the similarities and differences between the functions and the land use distribution around the station area. The conclusion show that even through there exist some agreements, there is also a gap between the people activities and the land uses around the station area. These results possibly will give us deeper imminent into how people act around the stations by metro systems, which will ultimately benefit the urban planning and policy development.

In this Paper [2], the author describes optimal planning for public transit system is one of the keys serving to bring a sustainable implement and a good quality of life in urban areas. Compared to private transit, public transit helps road space more effectively and gives fewer accidents and emissions. Therefore, in many cities people have a preference to take private transportation further than public transportation owing to the problem of public transportation services. In this approach, we focal point on the recognition and optimization of defective region pairs with difficult bus routing to develop utilization efficiency of public transportation systems, according to people's real insist for public transportation system. To this final, we first produce an incorporated mobility pattern recognize between the location traces of taxicabs and the mobility records in bus transactions. Derived from the mobility patterns, we propose a contained transportation mode option model, with which we can dynamically predict the bus travel demand for several bus routing by taking into account both bus and taxi travel strain. This technique used for bus routing reorganization which achieve to change as many people from private transportation to public transportation as probably given budget constraints on the bus route changes. We also influence the model to recognize region pairs with defective bus routes, which are efficiently, optimized using our method. The techniques in common studies are performed on real-world information together in Beijing which contains 19 million taxi trips and 10 million bus trips.

In this Paper [3], the author describes substantial data gathered by automated fare collection (AFC) methods give chances for studying both personal traveling activities and united mobility patterns in the urban area. Previous studies on the Automated Fare Collection data have primarily focused on recognizing passengers' movement patterns. In this approach, therefore, we innovatively leveraged such data for recognizing thieves in the public transportation systems. Certainly, stopping pickpockets in the public transportation systems has been difficult for increasing passenger happiness and public safety. Therefore, it is demanding to tell thief from routine passengers in practice. To this final, we implement a suspect detection and observation system, which can recognize pickpocket suspects, based on their every day transit records. Particularly, we initially extracted a number of attributes from each passenger's every day activities in the transportation systems. Then, we took two-step methods that exploit the value of unsupervised outlier recognition and supervised categorization models to predict thieves, who exhibit irregular traveling activities. Investigational outcome demonstrated the efficiency of our method. We also implement a prototype method with a user-friendly interface for the security personnel.

In this Paper [4], the author describes Recent years have witnessed the success of binary hashing techniques in approximate nearest neighbor search. In apply; several hash tables are normally built using hashing to wrap more preferred outcome in the hit buckets of each table. Therefore, few work studies the combined methods to constructing several informative hash tables using any type of hashing techniques. In the meantime, for many table searches, it also lacks of a common query-adaptive and fine-grained position scheme that can improve the binary quantization loss suffered in the standard hashing methods. To resolve the above risk, in this approach, we initially regard the table construction as a selection risk in excess of a set of applicant hash functions. With the graph demonstration of the function set, we propose a well organized solution that consecutively applies normalized dominant set to finding the most instructive and independent hash functions for both tables. To extend to decrease the redundancy in the middle of tables, we find out the reciprocal hash tables in a boosting manner, where the hash function graph is efficient with high weights emphasized on the misclassified national pairs of existing hash tables. To process the ranking of the retrieved buckets within a convinced Hamming radius from the query, we offer a query-adaptive bitwise weighting method to allow fine-grained bucket ranking in each hash table, exploiting the discriminative power of its hash functions and their supplement for nearest neighbor search.

In this Paper [5], the author describes we instantly detect interesting phenomena, entitled black holes and volcano's, from an STG. Specifically, a black hole is a sub graph (of an STG) that has the overall inflow greater than the overall outflow by a threshold, while a volcano is a sub graph with the overall outflow greater than the overall inflow by a threshold (detecting volcano's from an STG is proved to be equivalent to the detection of black holes). The online detection of black holes/volcano's can timely reflect anomalous events, such as disasters, catastrophic accidents, and therefore help keep public safety. The methods of black holes/volcano and the interaction between them expose human mobility techniques in a city, thus help to create a better city preparation or develop a system's operation in good organization. Derived from a well-designed

STG catalog, we suggest a two-step black hole prediction algorithm: The initial step finds a group of applicant grid cells to start from; the other step explores an initial edge in an applicant cell to a black hole and prunes other applicant cells after a black hole is predicted. Then, we adapt this prediction algorithm to a nonstop black hole detection situation. We estimate our method based on Beijing taxicab information and the bike trip information in New York, analyzing urban anomalies and human mobility patterns.

In this Paper [6], the author describes to the promising literature analyzing public transportation fare card data for a good accepting of passengers' mobility patterns and path choices. A new heuristic is anticipated to approximation the stop-level source and destinations by finding the traveler behaviors in the observed transactions in a fare card dataset. The major concentration focus in this approach is evaluating the actual passenger trajectories for multi-leg journeys. If the fare card dataset concludes together boarding and alighting sequence of each transaction, the major challenge is the opinion of origins and destinations by distinguishing the transfer interchanges from the activity locations. Construct on usually used criteria for recognizing transfers, this approach proposes a new method to increase the accuracy of short behavior detection to estimate the passengers' true origins and destinations. The set of criteria in this approach is based on the planned concept of "off-optimality" for a more precise recognition of short/hidden behavior within the labeled shares. The measure of off-optimality incorporates different variables of the transit service between the given journey ends (including alternative paths and routes, service headways, walk distances/times, transfer points, etc.) and reflects persons into a minimal quantity to develop the accuracy of opinion. Additionally, the time gap among two transactions, the overall travel period, and the circuit of the path trajectories are additional variables that are used in distinguishing the true transfers from behaviors.

In this Paper [7], the author describes whereas swiping frequency changes are associated with travelers' activity intensity and activity type, it is possible to present the characteristics of the metro station all the way through patterns of aggregated staying behaviors that occurred approximately metro stations. For instance, if a person checks out of the metro station at 8 a.m. and evaluates in at the similar station at 6 p.m., his staying behavior around the station has a high option to be identified as work. Following this reason, we divide the transit records consistent with the staying activity that happened around the station area and inspect the aggregated activity patterns. As an alternative of centering on the trip episodes as in the existing literature, we spotlight the staying duration episodes to unveil the activity patterns hidden behind the card swiping behavior in this paper. The idea of staying behavior is not new, although it is addressed by different names. The author denoted the activities between the trip chains as the reliable Passengers Transit activities and targeted detecting home and work activities;

In this Paper [8], the author describes as identified the typical durations of the activity intervals. Therefore, both works focused more on the duration distribution of the activity chains of single passengers. The physical environment around various stations has not yet been connected to human behavior patterns. In addition, whether there are spatial variations of staying behaviors in different metro stations and how to infer the functions of the catchment area of a metro station from activities recorded in SCD remains unknown. In the direction of address this issue, this approach studies how the combined staying activities in various station areas derived from SCD contribute to characterizing the social functions of the stations.

III. CONCLUSION

From this survey, various activity patterns are studied and analyzed to detect anomaly behaviors. Anomaly detection from the travel log or trajectory data is a more promising process nowadays. This is very challenging due to several reasons such as mobility patterns are unique in every user; the trajectories are dynamic and need frequent updating. Detecting outlier/ anomaly from those dynamic and updatable datasets need more concentration. The techniques should be developed carefully. This paper surveyed various applications related to the public transit records such as identifying pickpocket suspects, daily activity pattern detection, bus route planning, traffic abnormality detection, etc. from this summary; a new technique can be developed and integrated into the recent applications.

REFERENCES

[1]. Yang Zhou , Zhixiang Fang , Qingming Zhan, "Inferring Social Functions Available in the Metro Station Area from Passengers' Staying Activities in Smart Card Data" , International Journal of Geo Information, 2017.

[2]. Yanchi Liu, Chuanren Liu, Nicholas Jing Yuan, "Intelligent bus routing with heterogeneous human mobility patterns", Knowledge and Information Systems, 2017.

[3]. Bowen Du, Chuanren Liu, Wenjun Zhou, "Catch Me If You Can: Detecting Pickpocket Suspects from Large-Scale Transit Records", ACM, 2016.

[4]. Xianglong Liu, Cheng Deng, "Query-Adaptive Reciprocal Hash Tables for Nearest Neighbor Search", IEEE, 2016.

[5]. Liang Hong, Yu Zheng, Duncan Yung, "Detecting Urban Black Holes Based on Human Mobility",ACM, 2015.
[6]. Nassir, N.; Hickman, M.; Ma, Z.L. Activity detection and transfer identification for public transit fare card data, Transportation, 2015.
[7]. Bouman, P.; Kroon, L.; Vervest, P., "Detecting Activity Patterns from Smart Card Data", Benelux Conference on Artificial Intelligence, 2013.
[8]. Chakirov, A.; Erath, A. "Activity identification and primary location modeling based on smart card payment data for public transport Activity Identification and Primary Location Modelling based on Smart Card Payment Data for Public Transport", International Conference on Travel Behaviour Research (IATBR),2012.

Sameera.P." A Survey on Detecting Pickpocket Suspects from Large-Scale Public Transit Records-Recent Approaches." IOSR Journal of Engineering (IOSRJEN), vol. 09, no. 10, 2019, pp. 01-05