www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

CATCH ME IF YOU CAN: DETECTING PICKPOCKET SUSPECTS FROM LARGE-SCALETRANSIT RECORDS

[1]MR K.KOTESWARA CHARI, [2]R.PRASHANTH, [3]O.SRINATH, [4]MD.JIFEAN HUSSAIN

[1](Assistant Professor) ,**CSE.** Teegala Krishna Reddy Engineering College Hyderabad

[234]B.tech scholar ,**CSE.** Teegala Krishna Reddy Engineering College Hyderabad

ABSTRACT

Massive data collected by automated fare collection (AFC) systems provide opportunities for studying both personal traveling behaviors and collective mobility patterns in the urban area. Existing studies on the AFC data have primarily focused on identifying passengers movement patterns. In this paper, however, we creatively leveraged such data for identifying thieves in the public transit systems. In deed, stopping pickpockets in the public transit systems has been critical for improving passenger satisfaction and public safety. However, it is challenging to tell thieves from regular passengers in practice. To this end, we developed a suspect detection and surveillance system, which can identify pick pocket suspects based on their daily transit records. Specifically, we first extracted a number of features from each passengers daily activities in the transit systems. Then, we took a two-step approach that exploits the strengths of unsupervised outlier detection and supervised classification models to identify thieves, who exhibit abnormal traveling behaviors. Experimental results demonstrated the effectiveness of our method. We also developed a prototype system with a user-friendly interface for the security personnel.

1.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. 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.

2. LITERATURE SURVEY

Author: Liang Hong, Yu Zheng and Duncan Yung

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.

Author: Xianglong Liu, Cheng Deng

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. 3 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 peoples 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.

Author: G. R. Newman and M. M. McNally

The author describes that identity theft generally involves three stages: acquisition of the identity information, the thief's use of the information for personal gain to the detriment of the victim of identity theft, and discovery of the identity theft. Evidence indicates that the longer it takes to discover the theft, the greater the loss incurred and the less likely it is that prosecution will be successful. Older persons and those with less education are less likely to discover the identity theft quickly and to report it after discovery. The research also found that access to personal information about potential victims and the anonymity the Internet offers would-be thieves are major facilitators of identity theft. Major topics on identity theft reviewed in this report are the definition of identity theft, the extent and patterns of identity theft, types of identity theft, recording and reporting identity theft, law enforcement issues and response, the cost of identity theft, and issues that need more research. Regarding the latter topic, the researchers recommend more research on the best ways to prevent identity theft crimes. Specifically, research should address practices and operating environments of document-issuing agencies that allow offenders to exploit opportunities to obtain identity documents. Research should also focus on practices and operating environments of document-authenticating agencies that allow offenders access to identity data. Also, the 4 structure and operations of the information systems involved with the operational procedures of the aforementioned agents should be researched. The report reviewed more than 160 literature sources that ranged from traditional journal articles to Web sites and presentations.

Author: Catherine Morency, Martin Trepanier, Bruno Agard

The author describes that various measures regarding the variability of travel behaviours of transit users. The analyses are performed with smart card data collected over a ten months period. The variability in terms of boarding per day and new stops frequented with the days of travel on the transit network is examined. Data mining techniques are then used to classify days of travel according to the similarity of the boarding time periods. In this view, the use of two specific smart cards is examined in more details. 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. 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. Such knowledge, for important sets of users, may provide important information about every day utilization and periodic evolutions. Also seasonality information may be discovered.

www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

The focus of the study is on trips habits (in time)

3.SYSTEM DESIGN

3.1 System Architecture:

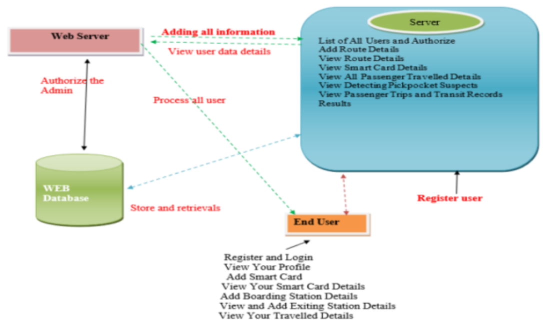


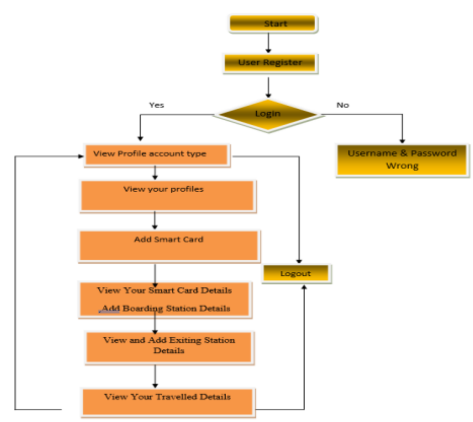
Fig:3.1 System Architecture

EXPLANATION OF ARCHITECTURE:

The architecture of Pick pocket Detection System shows that both the user and admin can login to the server through their login credentials. After logging in successfully user can check their respected travel details using their smart card which is very unique , which includes user ID, name smart card details, Boarding details ,Exiting details and Previous travel details . Admin also should log in through log in credentials like server name and password. Admin can perform several tasks which includes list of all users and 9 Authorize, Add route, View route, view smart card details, View all passengers travelled details, view detecting pickpocket suspects, view passengers trips and transit records results and log out.

3.2 ACTIVITY DIAGRAM

Activity diagrams are graphical representations of workflows of stepwise activities and the actions with support for choice, iteration, and concurrency. In the UML, activity diagrams can be used to describe the business and operational step-bystep workflows of components in a system. An activity diagram shows the overall flow of control. Each rectangular box represents an activity or action that occurs within the system. Arrow indicate the flow of control from one activity to another. The diamond-shaped boxes represent decision points, where the flow of control may take different paths based on conditions. The rounded rectangle represent the start and end points of the process. The diagram illustrates the sequential steps involued in placing an order on an ecommerce Website starting from selecting items to completing the order. Each step must be completed before moving on to the next, ensuring a structed and orderly process.



4.OUTPUT SCREENS

The implementation of the project is done via the following steps by using machine learning algorithms, the process starts by collecting dataset

EXPLANATORY OF KEY FUNCTIONS:

In propose paper to detect person abnormal behavior author employing Two-Step SVM which analyzing person behavior by using travelling frequency and Short Rides. If person has too many travelling frequency with short rides then that person will be pick pocketing suspect. So in this paper first we are applying Two-Step algorithm to identify person behavior and then employing SVM algorithm on person behavior to calculate suspect prediction Precision and FSCORE. As existing algorithm we are using One Class SVM (OCS) algorithm to identify person behavior and then training with SVM, Decision tree and Logistic Regression and then comparing OCS performance with propose Two-Step SMV. To train algorithms author is using Transit passenger record dataset but this exact dataset not available on internet so we are using passenger dataset and in below screen we are showing dataset details.

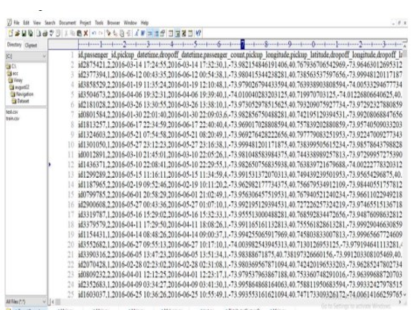


FIG 4.1

In Fig 4.1 dataset screen first row contains dataset column names and remaining rows contains dataset values and this value has passenger traveling latitude and longitude with trip duration.

To implement this project we have designed following modules

1) Upload Transit Records Dataset: using this module we will upload dataset to application

2) Preprocess Dataset: using this module we will select passenger ID from the dropdown box and then application will read all rows from the dataset for selected passenger and then remove missing values and then normalize the values

3) Run One Class SVM: using this module we will apply OCS (one class SVM) to identify passenger behavior and then apply SVM, Decision Tree and Logistic Regression to calculate suspect prediction precision and FSCORE values

4) Run Propose Two-Step SVM (TS-SVM): using this module we will apply TS-SVM (Two Step SVM) to identify passenger behavior and then apply SVM, Decision Tree and Logistic Regression to calculate suspect prediction precision and FSCORE values

5) Accuracy Comparison Graph: using this module we will plot comparison graph between existing OCS and propose TS-SVM algorithms.

SCREENSHOTS

To run project double click on "rat.bat" file to get below screen.

Fig 4.2

In Fig 4.2 click on 'Upload Transit Records Dataset' button to upload dataset and get below screen

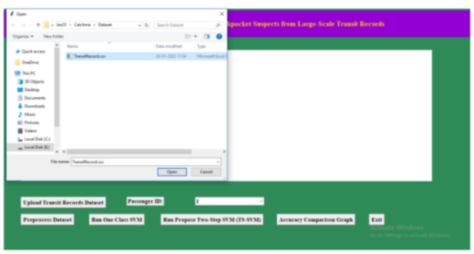


Fig 4.3

In Fig 4.3 selecting and uploading dataset file and then click on 'Open' button to load dataset and get below output

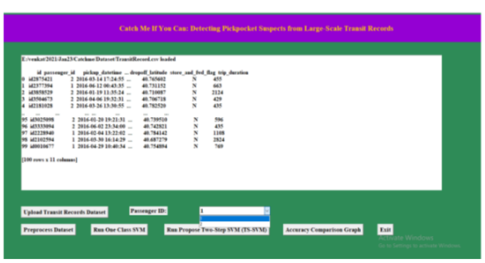


Fig 4.4

In Fig 4.4 dataset loaded and now from drop down box select any passenger ID and then click on 'Preprocess Dataset' button to read all records from selected passenger and then normalize values to get below output

5.CONCLUSION

This paper introduces an innovative system for detecting and tracking pickpocket suspects by analyzing large-scale transit records, addressing a crucial need for enhancing public safety in crowded transportation networks. The system leverages advanced data mining and machine learning techniques to identify suspicious behaviors and enable proactive surveillance in high-risk areas. The approach begins with constructing a feature representation method that profiles passengers based on behavioral and movement patterns derived from transit records.

These features are designed to balance data richness and computational efficiency, making the system scalable for large and complex transit systems. The core framework employs a two-step process to distinguish pickpocket suspects from regular passengers. In the first step, the system filters out ordinary passenger profiles, narrowing the focus to individuals exhibiting potentially suspicious activities.

The second step conducts a detailed analysis of these filtered profiles to identify patterns commonly associated with pickpocketing, such as frequent short trips, irregular travel behaviors, or clustering near high-risk zones. This hierarchical approach enhances accuracy while optimizing computational resources, ensuring practicality in real-world applications. The system was validated using real-world datasets from multiple sources, including transit card records and historical incident reports. These datasets were used to train and test the system, demonstrating high precision and recall in identifying suspects. A prototype implementation further showcased its usability for transit operators and security personnel, enabling real-time monitoring and preventive action in high-risk areas. This study highlights the transformative potential of leveraging large-scale data and machine learning to address urban safety challenges. By integrating practical functionality with ethical considerations like data privacy, the system offers a scalable and effective solution for reducing crime in public transit networks. It represents a significant advancement in urban security, combining innovation with real-world applicability.

6.FUTURE ENHANCEMENT

Future enhancements for pickpocket detection in large-scale transit data can focus on improving accuracy, scalability, and real-world applicability. Below are potential avenues for advancing this system:

1. Integration of Multi-Modal Data: Incorporating additional data sources, such as real-time video surveillance, GPS tracking, and social media alerts, can enrich the system's ability to detect suspicious behaviors. Multi-modal data fusion can provide a more comprehensive picture of passenger activities, improving detection accuracy.

2. Advanced Machine Learning Models: Employing cutting-edge machine learning techniques, such as deep learning and graph neural networks, can enhance the system's ability to analyze complex behavioral patterns. These models can learn from larger datasets and detect subtle deviations in passenger behavior, reducing false positives and negatives.

3. Real-Time Processing and Scalability: Upgrading the system to handle real-time data streams can enable immediate detection and response. Implementing distributed computing frameworks, such as Apache Kafka or Apache Flink, can support largescale data processing without compromising performance.

4. Context-Aware Analysis:

Enhancing the system to consider contextual factors, such as time of day, location, and passenger density, can improve detection precision. For example, the system could prioritize specific behaviors during peak hours or in high-risk zones.

5. Integration with Predictive Policing:

Expanding the system to predict potential hotspots for pickpocketing based on historical data and seasonal trends can help transit authorities allocate resources more effectively and proactively.

6. Enhanced User Interfaces: Developing intuitive dashboards and mobile applications for law enforcement and transit operators can streamline decisionmaking. Features such as automated alerts and visualization of risk zones can improve usability. 7. Ethical and

Privacy Considerations: Future iterations should include robust data anonymization techniques and compliance with data protection regulations to maintain passenger trust and ensure ethical deployment.

8. Continuous Learning Frameworks:

Implementing systems capable of incremental learning can adapt to evolving tactics used by pickpockets, ensuring the system remains effective over time.

By incorporating these advancements, the system can become a more robust, scalable, and ethical solution for ensuring passenger safety in large-scale transit systems.

7.REFERENCE

[1] G. R. Newman and M. M. McNally, "Identity theft literature review," United States Department of Justice, Report 210459, July 2005.

[2] M. Felson and R. V. Clarke, "Opportunity makes the thief: Practical theory for crime prevention," Policing and Reducing Crime Unit: Police Research Series, Report 98, 1998.

[3] T. L. C. da Silva, J. A. F. de Macˆedo, and M. A.Casanova, "Discovering frequent mobility patterns on moving object data," in MobiGIS, 2014, pp. 60–67.

[4] I. Ceapa, C. Smith, and L. Capra, "Avoiding the crowds: understanding tube station congestion patterns from trip data," in UrbComp, 2012, pp. 134–141.

[5] X. Ma, Y.-J. Wu, Y. Wang, F. Chen, and J. Liu, "Mining smart card data for transit riders travel patterns, "Transportation Research Part C, vol. 36, pp. 1–12, 2013.

[6] K. Zheng, Y. Zheng, N. J. Yuan, S. Shang, and X. Zhou, "Online discovery of gathering patterns over trajectories,"IEEE Transactions on Knowledge and Data Engineering, vol. 26, no. 8, pp. 1974–1988, 2014.

[7] Y. Liu, C. Liu, J. Yuan, L. Duan, Y. Fu, H. Xiong, S. Xu, and J. Wu, "Intelligent bus routing with heterogeneous human mobility patterns, "Knowledge and Information Systems, 2016, forthcoming, DOI: 10.1007/s10115-016-0948-6.

[8] B. Du, C. Liu, W. Zhou, Z. Hou, and H. Xiong, "Catch me if you can: Detecting pickpocket suspects from large-scale transit records," in KDD, 2016,pp. 87–96.

[9] C. Morency, M. Tr´epanier, and B. Agard, "Analysing the variability of transit users behaviour with smart card data," in ITSC, 2006, pp. 44–49.

[10] M. J. Sung, "Analysis of travel patterns of the elderly using transit smart card data," in TRB, no. 11-2357,2011.