IOSR Journal of Engineering (IOSRJEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
Vol. 09, Issue 11, November. 2019, Series -II, PP 01-09

An Effective Multi Level Approaches for Detecting and Analyzing Suspects from Dynamic Public Transit Records

Sameera.P[1], Ms.G.Maheswari[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 29 October 2019; Accepted 15 November 2019

Abstract: Detecting Suspects from large scale transit records are the ultimate aim of the proposed system. Handling large scale transit records are possible by applying data mining techniques, because, the transit records are too dynamic and frequently increasing. From this dynamic data, a new level of application is developed to detect the suspect/anomaly. This can detect the various anomaly behaviors of passenger data. This helps in various applications like pickpocket suspects, buglers from the large scale transit records. To achieve this, a new technique is proposed. The followings are the contribution of the system. Firstly, a number of important features that may be extracted from transit data records and are potentially useful for distinguishing thieves from regular passengers. In existing, the automatic fare collection records are used. Secondly, a multi-level approach is used to handle the class imbalance issue in passenger classification. The proposed work is a "Multi Model Batch" multi – Model Batch approach which includes a new classifier named as "Multi Model Batch-SVM" and the new anomaly detection technique "Time Series anomaly detection" Dense local outlier factor, which is improved from the existing LOF. The proposed system studies about the accuracy oriented issue in the suspect detection and learning phase of dynamic transit records. The proposed system aims to efficiently mine the Suspects of uncertain dynamic objects using multi level approaches. The system finally compared with the existing algorithms with different parameters such as time, precision, f-measure. The results show that accuracy of proposed system has been improved and the detection time is reduced with huge dataset.

Keywords: classification, Detecting Suspects, class imbalance, anomaly detection, Multi Model Batch-SVM.

I. INTRODUCTION

Enormous information gathered via mechanized toll accumulation frameworks give chances to considering both individual voyaging practices and aggregate versatility designs in the urban territory. Existing investigations on the Automated Fare Collection information have basically centered on cognizing travelers' development designs. In this paper, be that as it may, we inventively utilized such information for recognizing hoodlums in the open travel frameworks. To be sure, halting pickpockets in the open travel frameworks has been basic for improving traveler fulfillment and open wellbeing. In any case, it is trying to tell cheats from standard travelers by and by. To this end, we built up a suspect identification and observation framework, which can recognize pickpocket presumes dependent on their day by day travel records. In particular, we previously separated various highlights from each traveler's day by day exercises in the travel frameworks. At that point, we adopted a two-advance strategy that adventures the qualities of unaided anomaly location and administered characterization models to distinguish cheats, who show irregular voyaging practices. Trial results showed the adequacy of our technique. We likewise built up a model framework with an easy to use interface for the security work force. Pick pocketing generally has been one of the significant law requirements issues displayed by enormous social occasions of individuals for game occasions, fairs, bazaars, shows and such. It is sensible to accept that the pickpocket issue is winding up increasingly intense all things considered social affairs increment in number, size and recurrence because of upgrades in transportation, expanded populace and development of urban focuses. Shockingly, in any case, investigation into the issues of distinguishing and indicting pickpockets which was made of police and other law authorization authorities of sixty-seven chose urban areas: demonstrated a general conviction that the wrongdoing of pick pocketing is well leveled out. The way wherein this has been practiced is a fascinating case of the viability of utilizing rules characterizing a general wrongdoing which cuts over the standard criminal characterizations so as to stamp out a specific offense; however it likewise displays the topic of the attractive quality of such strategy.

II. PROBLEM DEFINITION

In this Paper [1], the creator depicts the capacity of a metro station territory is crucial for city organizers to think about when building up a setting mindful Transit-Oriented Development arrangement around the station zone. Through, the elements of metro station regions are hard to construe utilizing the static land use distribution and other customary survey datasets. These outcomes conceivably will give us more profound inevitable into how individuals act around the stations by metro frameworks, which will at last advantage the urban arranging and arrangement advancement.

In this Paper [2], the creator portrays ideal making arrangements for open travel framework is one of the keys serving to bring a reasonable execute and a decent personal satisfaction in urban zones. This strategy utilized for transport steering rearrangement which accomplish to change the same number of individuals from private transportation to open transportation as likely given spending limitations on the transport course changes.

In this Paper [3], the maker delineates Understanding urban capacities and their associations with human exercises has extraordinary ramifications for savvy and reasonable urban advancement. In this examination, we present a novel way to deal with revealing urban capacities by accumulating human exercises construed from cell phone situating and internet based life information. The outcomes demonstrate that the proposed methodology can catch citywide elements of both human exercises and urban capacities. It likewise recommends that albeit numerous urban zones have been formally marked with a solitary land-use type, they may give various capacities after some time contingent upon the sorts and scope of human exercises. The investigation shows that consolidating various information on human exercises could yield an improved comprehension of urban capacities, which would profit momentary urban basic leadership and long haul urban approach making.

In this Paper [4], the maker portrays Smart card information assembled via robotized passage accumulation (AFC) frameworks are significant assets for considering urban portability. In this paper, we propose two ways to deal with bunch savvy card information, which can be utilized to remove versatility designs in an open transportation framework

In this Paper [5], the maker delineates Commuting mirrors the long haul travel conduct of individuals and altogether impacts urban traffic blockage and outflow. Ongoing advances in information accessibility give new chances to comprehend driving examples productively and successfully. This investigation builds up a progression of information mining strategies to recognize the spatiotemporal driving examples of Beijing open travel riders. Utilizing one-month travel savvy card information, we measure spatiotemporal consistency of individual suburbanites, including living arrangement, working environment, and takeoff time.

In this Paper [6], the maker delineates The wide utilization of unavoidable registering innovation has took into account the development of huge information on spatial conduct and consequently gives a chance to investigate dynamic urban space. In this paper, an eigende composition technique is proposed to catch the normal examples of travelers' variety after some time among all metro stations just as to investigate the spatial heterogeneity of the dynamic space around the metro stations dependent on the basic examples with low dimensional structures.

In this Paper [7], the creator portrays considerable information accumulated via mechanized admission gathering (AFC) techniques give chances for examining both individual voyaging exercises and joined portability designs in the urban region. Past investigations on the Automated Fare Collection information have essentially centered on perceiving travelers' development designs. In this methodology, in this way, we imaginatively utilized such information for perceiving hoodlums in the open transportation frameworks.

In this Paper [8], the creator portrays Recent years have seen the achievement of double hashing methods in estimated closest neighbor search. In apply; a few hash tables are typically fabricated utilizing hashing to envelop progressively favored result by the hit basins of each table. Subsequently, little work contemplates the consolidated strategies to building a few instructive hash tables utilizing any kind of hashing methods. Meanwhile, for some table pursuits, it additionally absences of a typical inquiry versatile and fine-grained position plan that can improve the double quantization misfortune endured in the standard hashing strategies.

In this Paper [9], the maker portrays although blended use is a developing methodology that has been generally acknowledged in urban getting ready for advancing neighborhood energy, there is no accord on the best way to quantitatively quantify the blend and the impacts of blended use on neighborhood liveliness. Shannon entropy, the most ordinarily utilized decent variety estimation in evaluating blended use, has been seen as lacking in estimating the multifaceted, multidimensional attributes of blended use.

In this Paper [10], the maker delineates This paper tries to comprehend outrageous open travel riders in Beijing utilizing both customary family unit studies and rising new information sources.SCD are utilized to recognize the spatiotemporal examples of these four outrageous travel practices. Likewise, family studies are utilized to enhance the financial foundation and probably profile extraordinary explorers. While the exploration

discoveries are helpful to manage urban administration and arranging in Beijing, our strategy and systems can be reached out to comprehend travel designs somewhere else.

In this Paper [11], the maker delineates Investigating human versatility examples can support analysts and offices comprehend the main thrusts of human development, with potential advantages for urban arranging and traffic the board. In this way, this examination researches urban human intermingling and dissimilarity designs and their associations with the urban utilitarian condition, which is useful for urban approach advancement, urban arranging and traffic the executives. View Full-Text.

In this Paper [12], the maker delineates Transit situated advancement (TOD), which is commonly comprehended as the arrangement of higher-thickness, blended use, enhancement rich, and walk able improvement around fast travel stations, has been advocated as one of the best answers for boosting the potential rate of profitability for existing and future quick travel framework ventures. However, obviously not all usage of TOD is the equivalent in each station catchment region over a travel organizes. This heterogeneity in station territory settings presents huge unpredictability for organizers and policymakers keen on understanding existing TOD conditions, a region's TOD potential, and the applicable strategy and arranging intercessions required to accomplish arranging objectives. It likewise makes inconveniences for scientists keen on partner station settings with different TOD results.

III. PROPOSED SYSTEM

Chapter 3 presents a discussion about the proposed study and the contributions of the proposed work. The system proposes a new effective Framework called as Multi Model Batch SVM with Trajectory Pattern Model **(MMBS--TPM)** this finds and filter the suspected pickpockets and regular passenger from the given s large datasets.

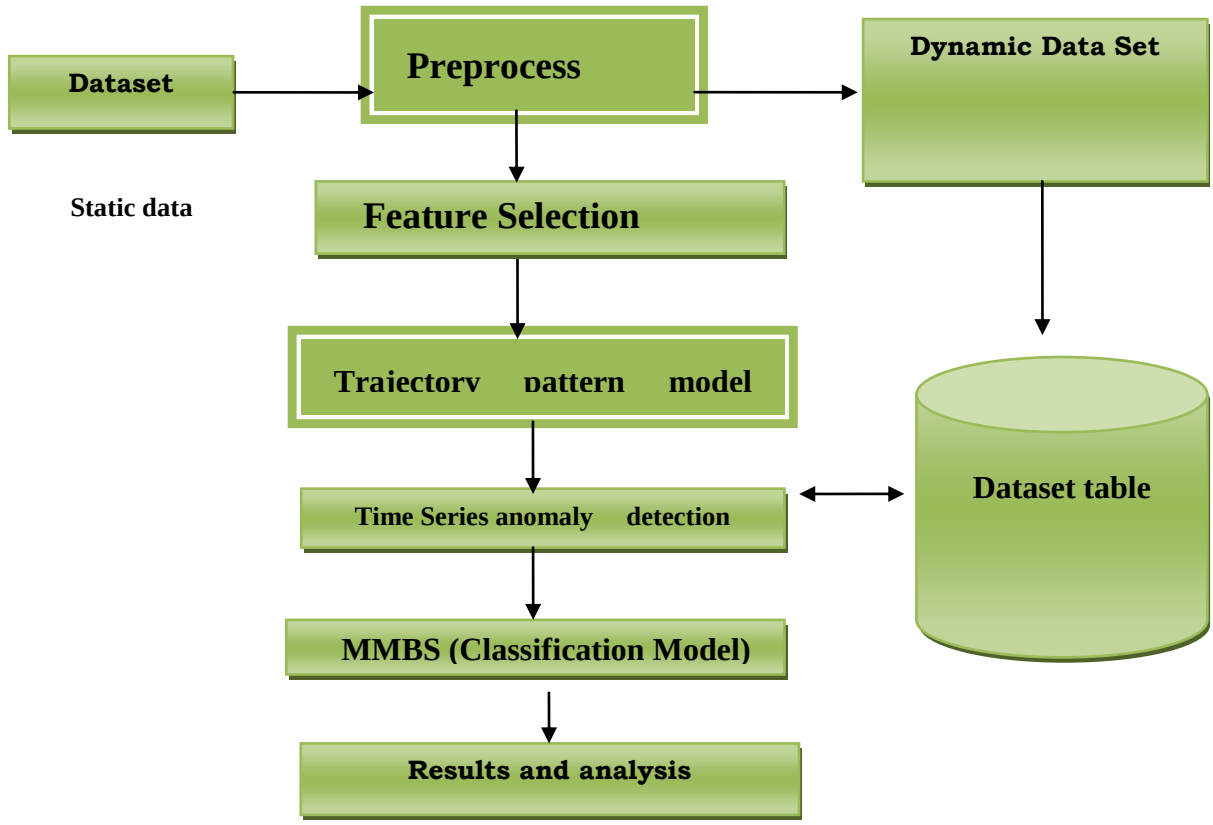


Figure 3.1 Proposed system architecture

The proposed system completely focuses on effectively stopping pickpockets in the public transit system and successfully improving passenger satisfaction. Trajectory pattern model to present individuals' behaviors and take out some effective features for effective pickpocket suspects classification.

The contribution of the Proposed Work

The followings are the contributions work of the proposed system.

- The existing Model is difficult in handling the dynamic Large-Scale Public Transit Records and failed to classify accurate pickpocket suspects.
- The proposed framework Multi Model Batch SVM with Trajectory Pattern Model (MMBS--TPM) this finds and filter the suspected pickpockets and regular passenger from the given s large datasets.
- The proposed system use Time Series anomaly detection. This completely filter regular passenger from Large-Scale Public Transit Records.
- The proposed system uses effective feature selection algorithm this helps to extract only more relevant feature from given Transit Records.
- Compared to the existing method or other popular semi supervised clustering algorithms, the required computational costs and memory requirements are significantly reduced and thus the proposed method is especially preferable in large scale problems.
- Propose apply MMBS algorithm has been used with multi class active learning for fast detection. This reduces the need of training dataset.

IV. METHODOLOGIES

The chapter completely presents the discusses about the algorithm and technique and the steps involved in the Proposed System. Here with list of following methodology explains in this chapter step by step.

- Data set preparation and upload
- preprocessing
- Feature Selection
- Trajectory pattern model
- Time Series anomaly detection
- Multi Model Batch SVM

1. Data set collection

The first step is extraction and crawling of data from Large-Scale Public Transit Records site. After collecting this data complete data is be stored in database with the help of SQL queries.

2. Data pre-processing

Preprocessing is the process of elimination, which eliminates duplicate and incomplete data's from the dataset before processing. It does not include redundant records in the train set as well, so the classifiers will not be biased towards more frequent records. Here are no duplicate records in the proposed test sets; therefore, the performance of the learners is not biased by the methods which have better detection rates on the frequent records.

3. Feature selection

In machine learning, Feature selection (FS) is one of the useful for data analyzing process; it proves which features are significant for prediction, and how these features are associated. FS involves subset generation, subset assessment, terminating the search while achieve the redundant one. While in feature selection not all the features where relevant. Feature selections are labeled as (i) *Relevant:* A relevant feature is one which is related to the minimum cardinality for achieving high predictive data. (ii) Irrelevant: Irrelevant features does not having any control on the output here the values are generated at random level for each data. (iii) *Redundant:* unwanted features occurred in the data. Selecting an original subset feature to the relevant one is not an easy one

4. Trajectory pattern model

Trajectory pattern model computes the probability distribution of an object's location at each time. Specially, this takes a certain number of recent positions in a trajectory, and infers a Gaussian distribution (i.e. a mean value μ and a standard deviation σ) at a present time. This method can be executed in an online manner; whenever a new position is brooked to the system, the estimator computes and populates the corresponding probability distribution. The structure also supports any user-given estimator to infer evolving distributions based on domain-specific knowledge. For this reason data packet may loss at any period because it evolve the distribution regarding specific knowledge of domain. Trajectory Database controls the corresponding probability distributions derived from an evolving density estimator with low accuracy and failed to report the object moving.

Algorithm: Trajectory pattern model

1: $\epsilon_\rho \leftarrow ZScoreLookUp(\rho)$

2: $results \leftarrow$ Ø3: ***if*** *node* is a non – leaf node **then**

4: **for** each node entry e ∈ node **do**

5: **if** $e.t^{-1} < t_q \leq e.t^{-1}$ **then** – temporal filtering

6: $RangeQuery(q, r_q, t_q e.cp)$

7: ***else***

8: ***for*** *each node entry* $e \in node$ ***do***

9: ***if*** $t_q = e.t^{-1}$ **then** – get a single record

10: $rec \leftarrow TrajHash[e.traj\_id, e.loc\_id]$

11: ***else*** – *get two consecutive records*

12: $rec_1 \leftarrow TrajHash[e.traj\_id, e.loc\_id - 1]$

13: $rec_2 \leftarrow TrajHash[e.traj\_id, e.loc\_id]$

14: $rec \leftarrow GetMidPoint(rec_1, rec_2)$

15: ***if*** $||rec.\mu, q|| < r_q + \epsilon_\rho . rec.\sigma$ ***then*** – Spatial filtering

16: ***if*** $PresenceProb(rec, q, r_q, \rho) > \rho$ ***then***

17: *add rec to result*

5. Time Series anomaly detection

This algorithm effectively filters the regular passenger from Large-Scale Public Transit Records. Improve the efficiency of regular passenger filtering by maintaining a large number of base filtering models.

6. **Multi Model Batch SVM**

Multi Model Batch Support vector machines based solution has been applied in the proposed detection process. Multi Model Batch SVM can map the input to a high-dimensional feature space by using non-linear mapping and then the linear hyper plane can be found. After one class classification round, the proposed system will make the next searching round by the changed hyper plane.

This proposed algorithm contains two steps.

Step 1: one is to select individual target through dynamic random, move the other support vector to individual target and the optimal objects from the last iteration, this call it as overall long step search;

Step 2: the second step refers from the detective theory which follows the above step1 in pattern search. Let the optimal support objects have short step partial elaborate search in the neighborhood, in order to find the Optimal class.

Multi Model Batch SVM steps:

Step 1: collect training data samples and test samples.

Step 2: According to gathered data, constructs training sample set and test sample set.

Step 3: Set up parameters, initializes the initial support vector object position, every position corresponding a set of attributes ($a1, a2 \ldots an$) in MCBSVM model, builds up SVM prediction model by parameters and samples.

Step3: From the parameters calculate every class threshold value, and then analyze the hyperplane value.

Step4: Randomly select P objects from initial cluster, find out the optimal object position $bestX$ based on the hyperplane. Set it up as individual target obj_X.

Step5: The non-optimal objects in the initial cluster moving to target class position and make the overall search.

Step6: The optimal object make overall search according to its neighborhood.

Step7: Update every objects class

4. Experiments and Results

Present experimental results employ our proposed framework. First, explain the experimental environments and provide complete implementation details. After demonstrate the effectiveness of our framework by comparing it to several Existing methods and approaches. The data for our research study were collected from multiple sources. These include transit records, geographical information, and theft incident reports.

4.1 Dataset

The experiment uses the synthetic data sets for experiments. Transit Records dataset composed from a public transit system that include buses and and subways. This data set include following attributes are bus/subway route number, boarding station and along with time, exiting station and time.

Smart Card ID	Route Number	Boarding Station	Boarding Time	Exiting Station	Exiting Time
4322	Route 52	a	07:15	b	07:40
4322	Route 26	c	07:46	d	08:23
4322	Route 11	d	17:35	e	18:08
4322	Route 11	e	19:32	f	19:54
4322	Route 16	f	19:58	a	20:15

(c) Transit Records

Geographical Information Dataset

The system can have n number of Geographical Information tuples for experiments.

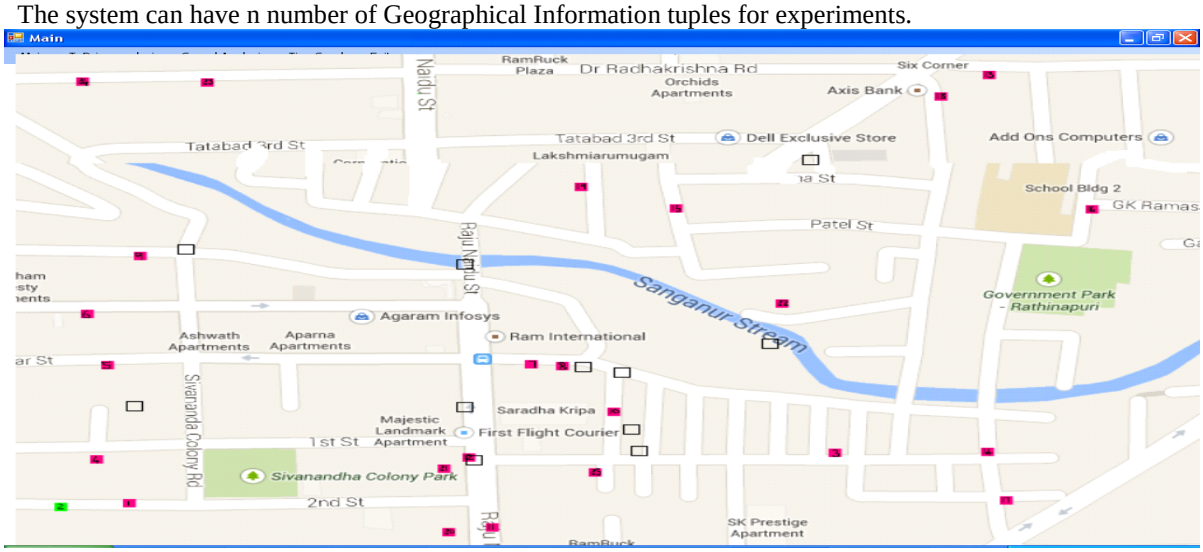


Fig: GPS dataset

V. RESULT AND ANALYSIS

In this chapter, this completely evaluates the Pickpocket Suspects detection accuracy and efficiency of the proposed system, in terms of time, accuracy, performance. This also evaluates the progressiveness of the methods under different dynamic dataset. In this chapter, this evaluates the efficiency of the algorithms, in terms of time consumption against dimensionality d, number of sub space creation m, and indexing threshold q under two distributions of objects' spatial locations. This also evaluates the progressiveness of the methods under different location distributions.

This section evaluates the proposed route network with dynamic trajectory data framework in terms of both indexing overhead and storage performance. We applied Route Net on sample road networks, namely, dynamic route map and the final set of experiments.

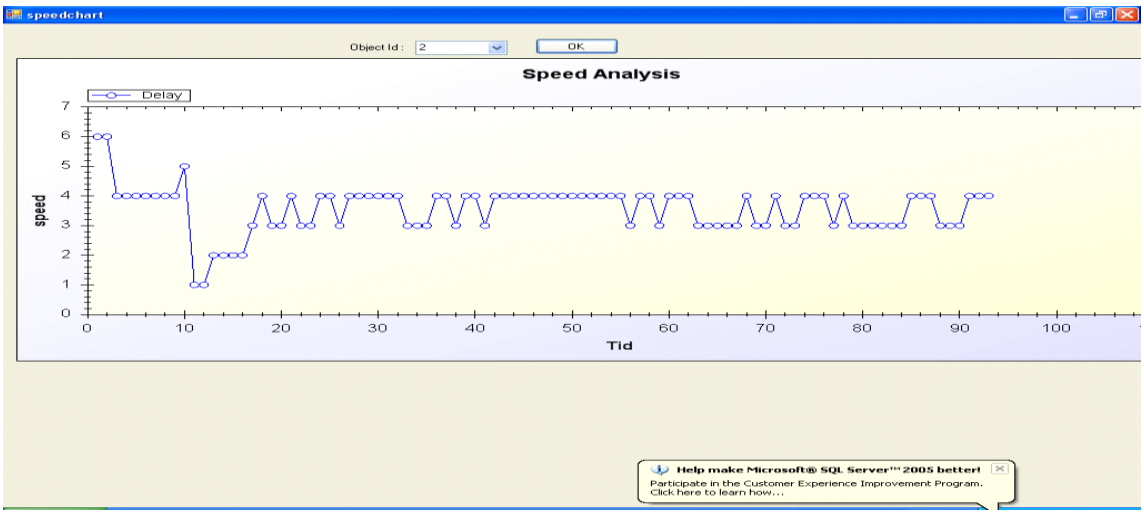
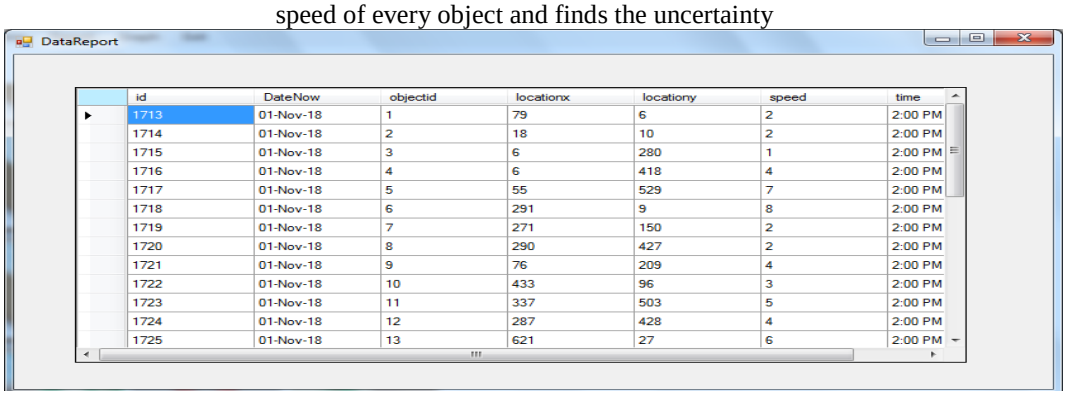


Fig: Object uncertainty

The above figure represents the object mobility at every transaction. The above figure calculates the speed of every object and finds the uncertainty



id	DateNow	objectid	locationx	locationy	speed	time
1713	01-Nov-18	1	79	6	2	2:00 PM
1714	01-Nov-18	2	18	10	2	2:00 PM
1715	01-Nov-18	3	6	280	1	2:00 PM
1716	01-Nov-18	4	6	418	4	2:00 PM
1717	01-Nov-18	5	55	529	7	2:00 PM
1718	01-Nov-18	6	291	9	8	2:00 PM
1719	01-Nov-18	7	271	150	2	2:00 PM
1720	01-Nov-18	8	290	427	2	2:00 PM
1721	01-Nov-18	9	76	209	4	2:00 PM
1722	01-Nov-18	10	433	96	3	2:00 PM
1723	01-Nov-18	11	337	503	5	2:00 PM
1724	01-Nov-18	12	287	428	4	2:00 PM
1725	01-Nov-18	13	621	27	6	2:00 PM

The system has used the Visual Studio.Net framework. And C#.Net has been used for developing the front end and SQL Server for the back end. The reason for using C#.Net is its flexibility. This can add or remove any features without editing the whole code. This separated the standalone functions like port matching and IP address matching in separate functions which are reused again and again. For the back end, this needed a freely distributed and powerful database so SQL Server was a good choice.

Attribute	value
Number of objects	25
Average sampling interval	1 sec
Number of tuples	40000 and above
Number of trajectories	5289

Table 5.1: summary of the datasets

Results and discussion

4.3 Performance Evaluation

For the experiment result, An Intel I5 2.2 GHz processor with 8 Gb RAM was used to measure the execution time and detection accuracy.

Tabulation for Execution Time

Dataset	Execution time (ms)	
	Existing model	**Proposed MMBS--TPM**
100	300	124
300	650	308
500	900	750

Tabulation for Search Time

Dataset	Average Search Time (ms)	
	Existing approach	**Proposed MMBS--TPM**
100	10	7
200	12	8
300	14	12
450	25	16

Measurement of Execution Time to search Suspects

VI. CONCLUSION

Our proposed work re-modulates the existing trajectory models to deal with the uncertainty management data. The system successfully proposed a new and effective approach **MMBS--TPM** to Detecting Pickpocket

Suspects effectively and achieve with maximum accuracy. The optimized algorithm has been expanded with the new optimal classification algorithms, which can handle large category dataset more rapidly, accurately and effectively, and keep the good scalability at the same time. Finally we evaluate our proposed framework with different metrics it gives best result in terms of different dataset. Experimental study on real-world datasets demonstrates the effectiveness and efficiency of our algorithms. In order to check the effectiveness of the proposed model can be extended with some other tree concepts in future.

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]. Tu, W.; Cao, J.; Yue, Y.; Shaw, S.-L.; Zhou, M.; Wang, Z.; Chang, X.; Xu, Y.; Li, Q. "Coupling mobile phone and social media data: A new approach to understanding urban functions and diurnal patterns". International. J. Geogr.Information Science, 2017.

[4]. El Mahrsi, M.K.; Côme, E.; Oukhellou, L.; Verleysen, "M. Clustering Smart Card Data for Urban Mobility Analysis. IEEE Trans", Intell. Transp. Syst. 2017.

[5]. Ma, X.; Liu, C.; Wen, H.; Wang, Y.; Wu, Y.J. "Understanding commuting patterns using transit smart card data". J. Transp. Geogr. 2017.

[6]. Gong, Y.; Lin, Y.; Duan, Z. "Exploring the spatiotemporal structure of dynamic urban space using metro smart card records". Comput. Environ. Urban Syst. 2017.

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

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

[9]. Yue, Y.; Zhuang, Y.; Yeh, A.G.O.; Xie, J.-Y.; Ma, C.-L.; Li, Q.-Q. "Measurements of POI-based mixed use and their relationships with neighborhood vibrancy", Int. J. Geogr. Inf. Sci. 2016.

[10]. Long, Y.; Liu, X.; Zhou, J.; Chai, Y. "Early birds, night owls, and tireless/recurring itinerants: An exploratory analysis of extreme transit behaviors in Beijing, China", Habitat Int. 2016,

[11]. Yang, X.; Fang, Z.; Xu, Y.; Shaw, S.-L.; Zhao, Z.; Yin, L.; Zhang, T.; Lin, Y. "Understanding Spatiotemporal Patterns of Human Convergence and Divergence Using Mobile Phone Location Data". ISPRS Int. J. Geo-Inf. 2016.

[12]. Higgins, C.D.; Kanaroglou, P.S. "A latent class method for classifying and evaluating the performance of station area transit-oriented development in the Toronto region". J. Transp. Geogr. 2016.

Sameera.P." An Effective Multi Level Approaches for Detecting and Analyzing Suspects from Dynamic Public Transit Records." IOSR Journal of Engineering (IOSRJEN), vol. 09, no. 11, 2019, pp. 01-09