International Journal on Future Revolution in Computer Science & Communication Engineering
Volume: 4 Issue: 4
ISSN: 2454-4248
388 – 392

Analysis Based on SVM for Untrusted Mobile Crowd Sensing

[*]Ms. Yuga. R. Belkhode , Dr. S. W. Mohod
*Student, Professor
Computer Science and Engineering, Bapurao Deshmukh College of Engineering, India.
Email id: yuga.belkhode@gmail.com

Abstract: Mobile crowdsensing, which collects environmental information from mobile phone users, is growing in popularity. These data can be used by companies for marketing surveys or decision making. However, collecting sensing data from other users may violate their privacy. Moreover, the data aggregator and/or the participants of crowdsensing may be untrusted entities. Recent studies have proposed randomized response schemes for anonymized data collection. We have Developed vehicle Survey Mobile Application for decision making and predict marketing survey. This kind of data collection can analyze the sensing data of users statistically without precise information about other users' sensing results. In this proposed work, we use SVM classifier for classifying the data can be used by companies for marketing surveys or decision making. In which we worked on Parameter of a city, which will help in analyzing vehicle count as well their availability according to vehicle type, vehicle model etc. The Result analyses will directly affects in predicting the result oriented strategies.

Key Words: *Mobile crowdsensing, privacy, data mining, SVM.*

*****

I. Introduction

Mobile crowdsensing, which collects environmental information from mobile phone users, is growing in popularity. These data can be used by companies for marketing surveys or decision making. However, collecting sensing data from other users may violate their privacy. Moreover, the data aggregator and/or the participants of crowdsensing may be untrusted entities. Recent studies have proposed randomized response schemes for anonymized data collection. This kind of data collection can analyze the sensing data of users statistically without precise information about other users' sensing results. However, traditional randomized response schemes and their extensions require a large number of samples to achieve proper estimation.

Participants of crowdsensing perceive their surrounding environment through their mobile phones, and the mobile phones send the sensed data (e.g. radiation level, location) to the aggregator. We assume that the aggregator reconstructs the true data distribution, that is, it generates an estimated contingency table of the sensed data. For this reason, the aggregator requires categorical attribute values.

Our objective is to collect dataset of various participants with the help of android app to evaluate marketing survey and for providing security; aggregator can assign a certain crowd sensing application ID to one honest participant which is used to analyze the sensing data of users for decision making purpose.

We trained and considered different types of classifiers using a supervised learning approach, which included SVM, which we are using for classify disguised data for decision making and marketing survey.

The rest of this research paper consists of related work followed by objective and motivation of research. Then the proposed system followed by implementation. Finally, concluded the research.

II. Related Work

There is a rich literature for crowdsensing and SVM and here we discuss the most related work.

Yuichi Sei and Akihiko Ohsuga [1] in this paper, they use collaborative flittering. Our contribution is to replace collaborative filter with SVM based classifier which will helpful to us for improving the overall accuracy of the system. P. Kairouz, K. Bonawitz, and D. Ramage [2] examine discrete distribution estimation under local privacy, a setting wherein service providers can learn the distribution of a categorical statistic of interest without collecting the underlying data. E. Schubert, A. Zimek, and H.-P. Kriegel [3] analyzed the interplay of density estimation and outlier detection in density-based outlier detection. By clear and principled decoupling of both steps, they formulate a generalization of density-based outlier detection methods based on kernel density estimation. Ú. Erlingsson, V. Pihur, and A. Korolova [4] describes and motivates RAPPOR, details its differential-privacy and utility guarantees, discusses its practical deployment and properties in the face of different attack models, and, finally, gives results of its application to both synthetic and real-world data. Q. Li and G. Cao [5] mentioned that to provide strong privacy guarantee, existing approaches add noise to each node's data and allow the aggregator to get a noisy sum aggregate. R. Chen, B. C. M. Fung, B. C. Desai, and N. M. Sossou [6] they examine data utility in terms of two popular data analysis tasks conducted at the STM, namely count queries and frequent sequential pattern mining. J. B. Gomes, C. Phua, S. Krishnaswamy [7] mentioned that the technological advances in smartphones and their widespread use has

 ISSN: 2454-4248
388 – 392

resulted in the big volume and varied types of mobile data.

Rodrigo Jose Madeira Ltheirenco [8] proposed for urban mobility, cycling presents itself as a cleaner, cheaper, and healthier alternative to motorized transportation. Md H. Rehman, C. S. Liew, T. Y. Wah, J. Shuja and B. Daghighi [9] this study presents the personal ecosystem where all computational resources, communication facilities, storage and knowledge management systems are available in user proximity. S. Hu, L. Su, H. Liu, H. Wang and T. F. Abdelzaher [10] in this article they present SmartRoad, a crowd-sourced road sensing system that detects and identifies traffic regulators, traffic lights, and stop signs, in particular.

III. Objective and Motivation

The objectives of the proposed approach are: To collect dataset of various participants. Then analyze the sensing data of users and replace collaborative filter with SVM (Support Vector Machine) based classifier.

The motivation for research is: The tracking is being done using collaborative filtering, and is usually efficient in terms of tracking accuracy of objects. The existing system takes more time for tracking and the quality of tracking is not unto the expected result. The result analysis obtained by participant may not useful for aggregator, as it may be sensed or not.

IV. Proposed Work

In this proposed work we use multiple dummies for crowdsensing as shown in Figure.1. These dummies collect information through individual smart phones and send sensed data to the data aggregator. Aggregator collects and analyzed sensed data and reconstructs true data distribution that is it generates estimated contingency table for sensed data. From this table we can analyze issues related to sensed values and take action according to it.

We implemented our node protocol as a smartphone application for Android to verify the feasibility of the protocols. We measured the time it took for a smartphone to anonymized its sensed data and send the disguised data. Because our target is a crowdsensing system, the calculation cost of the randomization algorithm conducted in smartphones should be light.

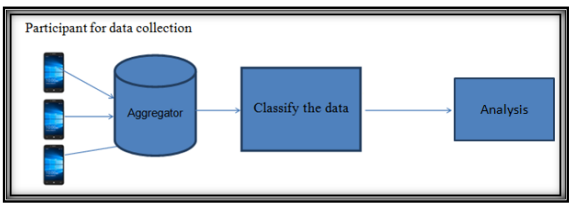


Figure.1 System Architecture.

In the proposed method, the aggregator in crowdsensing systems can be used to estimate data distributions more accurately than other randomization methods. Moreover, the participants do not need to confirm the fraction of malicious participants.

We are doing this by using advanced methods for classification of the sensed input data, and then using a prediction engine in order to check the current and next state of the object. Here after analyzing disguised data from sensed data we applying SVM classifier on disguised data for decision making purposed which will helpful us for marketing survey.

The project flow starts from the implementation of Mobile application which we are developing in Android as shown in Figure. 2. It will be useful in data gathering that is going to be done by Participant which are permitted by Aggregator. The application will send collected data to the aggregator where all the classification of sensed data will do.

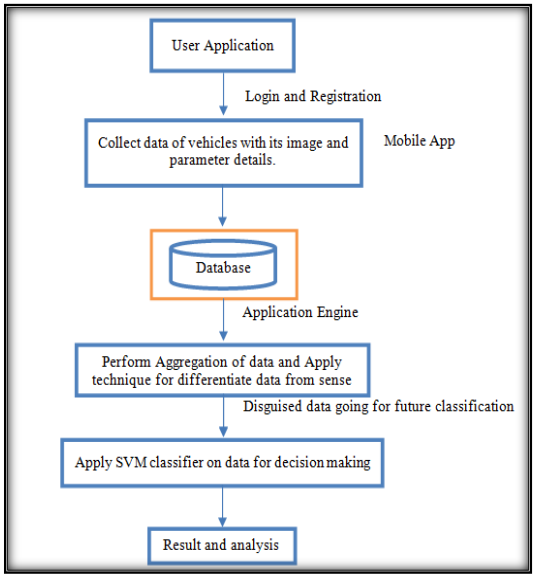


Figure.2 Flow of Proposed approach.

After getting the sensed data, it will generate cluster according to the region i.e., City which is done according to parameter decided for the SVM.

Implementation:

Anonymization Algorithm:

We assume that each participant installs a smartphone application for anonymized data collection provided by the aggregator. If the participant agrees to join mobile crowdsensing, the application's anonymization algorithm of the application starts.

Algorithm 1 shows the node protocol. Algorithm 1 Node protocol for a participant P

Input: P's true category ID t, IDs of categories K, Parameters s and q

Output: Set of anonymized values of a participant P

1: Creates empty set R_P
2: if rand() < q then
4: R_P (ftg [getRandElements(K n ftg; s1)
5: Else R_P (getRandElements(K n ftg; s)
6: end if

 ISSN: 2454-4248
388 – 392

SVM Methodology

Gather all the data of vehicle survey from the participant. Classify the Data with reference to the RTO_ data set where we are comparing the necessary information such as vehicle owner name, vehicle type, vehicle model and number. This classified data will be treated as input to the SVM classifier.

Algorithm:

1. Begin
2. LOOP through each data instance and label pair according to parameter.
3. For each entry in the data instance.
4. If entry is sensed with reference to RTO_dataset then ADD to classifier of that particular city region.
5. If present then show into SVM.
6. Else discard that entry
7. End

The implemented system will execute as follows:
In the first step we will be gathering data of vehicle survey by the means of a mobile application, and the collected data will send to the web application where, the admin can see all the details of registered participants who will be collecting data for the vehicle survey analysis. Also he is able to see the collected data by them.

Then we will be classifying the sensed data from the collected data with the help of a dataset we have maintained already. The sensed data will be useful to perform the SVM so the system can provide the end result which will be useful in marketing survey and decision making.

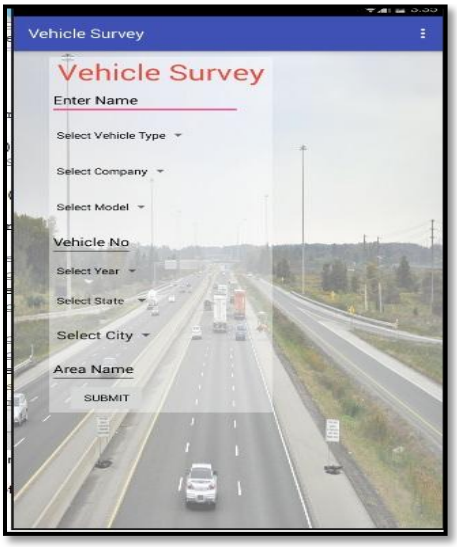
Here we can get the glimpse of the vehicle survey, where we will understand it as per the parameter "City" we had selected.

Scr. No. 1: Login Page

Here, Participant can Login using the Credentials provided by the admin to him respectively such as Participant id and password. Also there is one sign up tab where new user can register them by giving their suitable details.

Scr. No.2: Registration Form

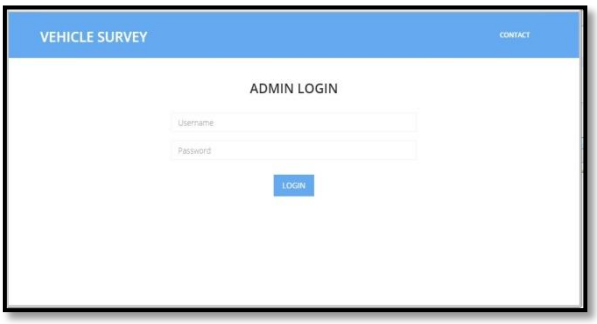
This is the registration form where Participant will be giving all the necessary details such as name, contact, email, and password. Owing to this admin will decide whether to allow this participant to gather the data or not. The details of participant will be sent to admin by this way.



Scr. No.3: Vehicle Survey

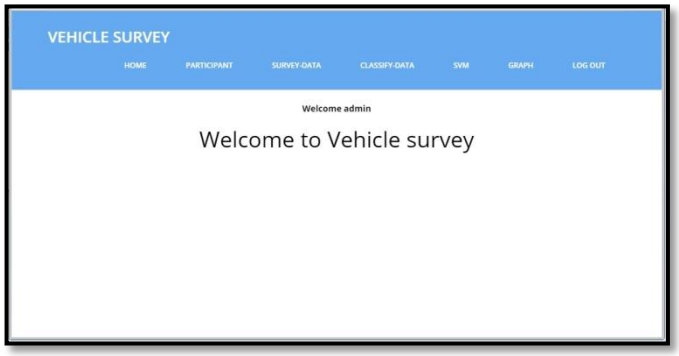
This is the vehicle survey form where registered participant will be taking vehicle surveys, in this we will be needing vehicle owner's name, his vehicle type, vehicle model, vehicle company, vehicle number, purchase year, state and city also the area in which the participant is taking the survey. On the Submit button click it is going to send to the admin by the participant.

 ISSN: 2454-4248
388 – 392



Scr. No.4: Admin Login

This is the form for admin login where admin will use his username and password to sign in to the vehicle survey admin site.



Scr. No. 5: Welcome Page

This is the welcome page where admin can see all the necessary tabs such as participant, survey data, classify data, SVM and result analysis tab.

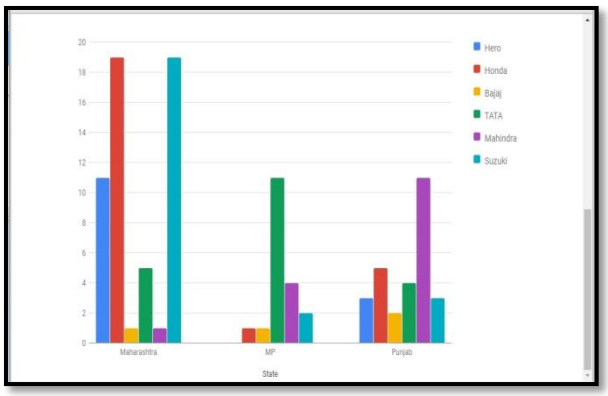
In the participant tab, admin can view the details of participant such his name, id, contact, mail, participant id and password. Also here admin can give all participant id and password by clicking on the "give password" tab. Only those are going to be allowed who will have authorized Participant ID and password.

In the survey data tab, admin can view survey data gathered by the participant such as the participant id, person name, his vehicle type, vehicle model, Vehicle Company, vehicle number, purchase year, state and city.

In Classify Survey tab, we have Classify the data owing to the data present in the RTO_Dataset, the sensed data will be shown here in this tab. As admin already have RTO_Dataset maintained at his site, so here he will cross check and after analyzing the present data with RTO he will come on conclusion such that it is Sensed or not.

In the SVM classifier tab, we have classified data according to cities and made cluster of it. After Classification we came to know that which one is sensed and which one is not, so depend on that output we have created cluster for SVM classifier such that what amount of classified data is present in which particular city.

Scr. No. 6: Graph 1



Scr. No. 7: Graph 2

It is the overall graph analysis according to vehicle type. Where first graph shows Analysis according to State, where admin can see in which state how many true vehicle owner are there, they are using which vehicle.

In our analysis we worked on 2 wheeler and 4 wheeler vehicle data so owing to this we have generated graph like which model have how much frequency for that particular state respectively.

This result may help to make market decision for upcoming production to that vehicle company, so that it will generate more revenue with the help of new strategies.

V. Conclusion

In this project we are performing privacy-preserving mobile crowdsensing where each participant's mobile phone will perform sensing the data, calculating the category ID from the sensed data, anonymzing the category ID, and sending the disguised category ID to the aggregator.

After collecting the data, aggregator will perform classification anonymization algorithm to get sensed data from large amount of data. Here we are applying SVM classifier on sensed data for decision making which will useful to us for marketing survey which improves the accuracy of system.

 ISSN: 2454-4248
388 – 392

References

[1] Yuichi Sei and Akihiko Ohsuga, "Differential Private Data Collection and Analysis Based on Randomized Multiple Dummies for Untrusted Mobile Crowdsensing," in Proc. IEEE Transactions on Information Forensics and Security, Vol. 12, No. 4, April 2017.

[2] P. Kairouz, K. Bonawitz, and D. Ramage, "Discrete distribution estimation under local privacy," in Proc. ICML, 2016.

[3] E. Schubert, A. Zimek, and H.-P. Kriegel, "Generalized outlier detection with flexible kernel density estimates," in Proc. SIAM SDM, 2014.

[4] Ú. Erlingsson, V. Pihur, and A. Korolova, "RAPPOR: Randomized aggregatable privacy-preserving ordinal response," in Proc. ACM CCS, 2014.

[5] Q. Li and G. Cao, "Efficient privacy-preserving stream aggregation in mobile sensing with low aggregation error," in Proc. PETS, 2013.

[6] R. Chen, B. C. M. Fung, B. C. Desai, and N. M. Sossou, "Differentially private transit data publication: A case study on the montreal transportation system," in Proc. ACM KDD, 2012.

[7] Joao Bartolo Gomes, Clifton Phua, Shonali Krishnaswamy, "Where will you go? Mobile Data Mining for Next Place Prediction", Data Warehousing and Knowledge Discovery. DaWaK 2013.

[8] Rodrigo Jos´e Madeira Ltheirenc¸o, "Cycle Their City goes Mobile", 2012.

[9] Md H. Rehman, C. S. Liew, T. Y. Wah, J. Shuja and B. Daghighi "Mining Personal Data Using Smartphones and Wearable Devices: A Survey" in Proc.ISSN, Feb. 2015.

[10] S. Hu, L. Su, H. Liu, H. Wang and T. F. Abdelzaher, "SmartRoad: Smartphone-Based Crowd Sensing for Traffic Regulator Detection and Identification."In Proc. ACM TSN, Vol. 11, No. 4, Article 55, July 2015.