IJESRT
INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

ANALYSIS BASED ON SVM FOR UNTRUSTED MOBILE CROWDSENSING-A REVIEW

Yuga.R.Belkhode[*1], **DR.S.W.Mohod** [2]
Student[1], Professor[2]
Computer Science and Engineering,Bapurao Deshmukh College of Engineering, India.

DOI: 10.5281/zenodo.1135902

ABSTRACT

Now-a-days the trend of Mobile crowdsensing, which collects environmental information from mobile phone users, is need which is growing in popularity.
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.
In this proposed work, we use SVM classifier for classifying the data can be used by companies for marketing surveys or decision making.

KEYWORDS: S2M and S2Mb schemes, SVM Classifier, Sensed and disguised data.

I. INTRODUCTION

Owing to the development of ubiquitous computing and sensing technologies, numerous research methods for crowdsensing have been proposed to collect and analyze sensed environmental information from mobile phone users in the crowdsensing, individuals collectively share environmental data with a data aggregator, and the aggregator analyses the collected data for decision making or marketing surveys. However, sensing aspects of a crowdsensing participant's surrounding environment, such as radiation level and location, may involve information that identifies an individual, and thus private information may be leaked.

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.

In regard to mobile crowdsensing applications, we can consider the noise, the name of the city that each participant resides in, and other factors of the participants' surrounding environment for urban planning , radiation levels , or the speed and type of cars, such as ambulance and taxi (in the anonymous monitoring of drivers). The data to be collected might also include personal data such as sex and age.

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.

In this proposed work, we provide authentication, security by checking whether the participants are dummy one or real one. This project is implemented to improve the quality of randomized event detection which also helpful for tracking of objects in real time using dummy devices. The important approach for developing this application is for marketing survey, we are providing easiness which is going to be gathered by product review.

In the existing work, the researchers are using multiple dummies in order to track objects of interest. 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 up to the expected result.

Here, 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.

In this work our contribution is to replace collaborative filter with SVM based classifier which will helpful to us for improving the overall accuracy of the system.

The rest of this research paper is organized as follows. Section II discusses the related work. Section III presents the design of our system architecture and proposed approach's flow, and Section IV presents the objective of our work. Section V concludes the paper.

II. LITERATURE SURVEY

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

Yuichi Sei and Akihiko Ohsuga [1] mentioned that RR can realize a privacy-preserving mobile crowdsensing where each participant's mobile phone probabilistically replaces the unusual category of the data with another category. The replaced category is send to the aggregator, which attempts to estimate the distribution of the original categories of members. However, RR schemes require great many samples in order to achieve correct reconstruction. In this proposed work, they recommend S2M and S2Mb schemes, which can surpass existing RR schemes. By simulations with synthetic and actual datasets, they proved that S2Mb scheme can shrink the estimated errors. The larger the problem, the more the performance of S2Mb exceeds those of added schemes.

P. Kairouz, K. Bonawitz, and D. Ramage [2] examine discrete distribution estimation under local privacy, a setting wherein service providers can be taught the distribution of a categorical statistic of interest without collecting the underlying data. This research's focal point is on distribution estimation under local privacy takes one step toward a world where the benefits of data-driven insights are decoupled from the collection of raw data. Their new theoretical and empirical results prove that combining cohort-style hashing with the k-ary extension of the classical randomized response mechanism admits practical, state of the art results for locally private logging.

In future work, they plan to look at the estimation of non stationary distributions as they change over time, a common scenario in data logged from user interactions. They will also consider what utility improvements may be likely when some responses need more privacy than others, another common scenario in practice.

E. Schubert, A. Zimek, and H.-P. Kriegel [3] analyze the interaction of density estimation and outlier detection in density-based outlier detection. By obvious and principled decoupling of both steps, they formulate a generalization of density-based outlier detection methods based on kernel density estimation. Embedded in a broader framework for outlier detection, the resulting method can be easily adapted to detect novel types of outliers: while ordinary outlier detection methods are designed for detecting objects in sparse areas of the data set, their method can be modified to also detect unusual local concentrations or trends in the data set if desired. The new approach in this proposed effort is KDEOS, can be seen as a blueprint that can easily be adjusted in various places, to incorporate domain specific knowledge and application-specific requirements, as well as in order to specify the desired kind of outliers to be detected.

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 together synthetic and real-world data. RAPPOR is a flexible, mathematically precise and practical platform for anonymous data collection for the purposes of privacy-preserving crowdstheircing of population statistics on client-side data. RAPPOR gracefully handles many data collections from the same client by providing well-defined longitudinal differential privacy guarantees. Highly tunable parameters allow balancing risk versus utility over time, depending on one's needs and evaluation of likelihood of different attack models. RAPPOR is purely a client-based privacy solution. It eliminates the entail for a trusted third-party server and puts control over client's data back into their own hands.

Q. Li and G. Cao [5] propose a scheme for privacy-preserving aggregation of time-series data in presence of untrusted aggregator, which provides differential privacy for the sum cumulative. This research proposed a novel ring-based interleaved grouping method and applied it to privacy-preserving aggregation of time-series data in mobile sensing applications. This method to efficiently deal with dynamic joins and leaves and achieve low aggregation error. Specifically, when a node joins or leaves, only a small number of nodes need to update their cryptographic keys. Also, the nodes only collectively add a small noise to the sum to make sure differential privacy, which is O(1) with respect to the number of nodes. Based on symmetric-key cryptography, their scheme is very capable in computation.

R. Chen, B. C. M. Fung, B. C. Desai, and N. M. Sossou [6] present their explanation to transmit data publication under the rigorous differential privacy model for the Société de transport de Montréal (STM).They propose an efficient data-dependent yet differentially private transit data purification approach based on a hybrid-granularity prefix tree structure. Moreover, as a post-processing step, they make use of the inherent consistency constraints of a prefix tree to ways constrained inferences, which lead to better utility. Their result not only applies to general sequential data, but also can be seamlessly extended to trajectory data. In this proposed work, they have studied the problem of publishing transfer of data at the STM in the framework of differential privacy. They believe that their solution could benefit many other sectors that are facing the dilemma between the demands of sequential data publishing and privacy protection.

Joao Bartolo Gomes, Clifton Phua, Shonali Krishnaswamy [7] mentioned that the technological advances in smartphones and their widespread use has resulted in the big quantity and varied types of mobile data. Location prophecy through mobile data mining leverages such big data in applications such as traffic planning, location-based advertising, and intelligent resource allocation; as well as in recommender services including the very popular Apple Siri or Google Now. This research focuses on the challenging difficulty of predicting the next location of a mobile user given data on his or her current location. In this work, they suggest NextLocation - a personalized mobile data mining framework - that not only uses spatial and temporal data but also other appropriate data such as accelerometer, bluetooth and call/sms log. In addition, the proposed framework represents a new paradigm for privacy-preserving after that place prediction as the mobile phone data is not shared without user permission. They proposition an alternative business model for mobile advertising that uses NextLocation framework.

In this research they propose the NextLocation framework that is a mobile data mining move toward to the next place prediction problem. The chief advantage of NextLocation is that it is a privacy-preserving solution that fully runs on the mobile device itself. Sensitive data about the user locations and circumstance are not disclosed. Moreover, NextLocation uses an adaptive anytime model which enables adaptation to changes in the user mobility patterns. Finally, it keeps an estimation of the anytime model accuracy in real-time.

In future work, in line with the last trial on online learning conducted in this work they plan to develop an online algorithm particularly designed for next place prediction. They are also in contact with telcos to discuss the implementation of the proposed alternative advertisement model.

Rodrigo Jose Madeira Ltheirenco [8] proposed that urban mobility, cycling presents itself as a cleaner, cheaper, and healthier alternative to motorized transportation. To ease the adoption of the bicycle as feasible means of transportation, CycleTheirCity was developed. The platform recommends suitable cycling routes given the characteristics of the city’s road network, as classified by the participants. However, due to the efforts required to contribute, the platform presented low adoption levels.

This thesis proposes the design, implementation and assessment of a mobile crowdsensing-based system, capable of leveraging the sensors found in smartphones wielded by a diverse community of participants. The purpose of this solution is to extend CycleTheirCity, to allow the route categorization tasks to be performed transparently and effortlessly by the participants’ smartphones.

To recognize the viability of their solution, as an alternative to CycleTheirCity’s human-based categorization, they developed and evaluated Scout - an Android prototype. The prototype was appraised with regard to its ability to accurately classify the slope and concrete type of the roads traveled by the participants. Recent advancements in smartphone expertise have lead to the appearance of a new pattern of systems, the mobile crowdsensing systems. These have the potential to allow sensor-based studies to be performed at a superior scale and with reduced costs.

Md H. Rehman, C. S. Liew, T. Y. Wah, J. Shuja and B. Daghighi [9] mention that the staggering growth in smartphone and wearable device use has led to a huge scale generation of personal (user-specific) data. To explore, analyze, and extract helpful information and knowledge from the deluge of personal data, one has to leverage these devices as the data-mining platforms in ubiquitous, pervasive, and big data environments. This study presents the personal ecosystem where all computational resource s, communication facilities, storage and knowledge management systems are existing in user proximity. An extensive review on topical literature has been conducted and a detailed taxonomy is presented.

The staggering growth in PSDs is a key enabler in Per DM in RCEs for personalization, privacy, and security at user premises. Moreover, the exploitation of data mining algorithms in PSDs enables the use of PEs for individual good. User-centric big data personalization is a theory with a wide range of application in health care, ttheirism, education, e-government, and smart cities, among others. It has vast potential in personalization for better patient, traveler, customer, student, and citizen experiences.

S. Hu, L. Su, H. Liu, H. Wang and T. F. Abdelzaher [10] present SmartRoad, a crowd-sourced road sensing system that detects and identifies traffic regulators, traffic lights, and stop signs, in particular. As an alternative to expensive road surveys, SmartRoad works on participatory sensing data collected from GPS sensors from in-vehicle smartphones. The resulting traffic regulator information can be used for many assisted-driving or navigation systems. In order to achieve precise detection and identification under realistic and practical settings, SmartRoad automatically adapts to different application requirements by (i) intelligently choosing the mainly appropriate information representation and transmission schemes, and (ii) dynamically evolving its core detection and identification engines to effectively take benefit of any external ground truth information or manual label prospect. It carries out and completes the detection and identification tasks in a stout, effective, and efficient manner, achieving outstanding detection and identification performance on task completion.

Yohan Chon, Nicholas D. Lane, Yunjong Kim, Feng Zhao, Hojung Cha [11] proposed that Crowd enabled place-centric systems gather and cause over huge mobile sensor datasets and target everyday user locations. Such systems are transforming a variety of consumer services and data-driven organizations. As they insist for these systems increases, their understanding of how to design and deploy successful crowdsensing systems must improve. In this research, they present a systematic study of the reporting and scaling properties of place-centric crowdsensing. Their analysis of this dataset examines issues of core attention to place-centric crowdsensing, including place-temporal coverage, the relationship between the user population and coverage, privacy concerns, and the characterization of the collected data. Collectively, their findings afford valuable insights to guide the building of future place-centric crowdsensing systems and applications.

Apostolos Malatras, Laurent Beslay[12] presented their work on designing and developing a explanation for participatory surveillance. They aim at involving end users in the tasks related to security and observation and thus on one hand assist and promote the overall perceived level of safety, while on the other hand promoting users' sense of contribution and participation in the society and hence their awareness. By utilizing the numerous sensors on smartphones that are nowadays ever-present they assume that significant information regarding critical, security-related events can be inferred.

As a proof of concept, they built a system to collect such data from users in the background of an emergency evacuation exercise and they presented here pertinent results on the use of this data. By using just one sensor, namely the accelerometer, very high levels of accuracy in predicting users' activities they're reached. In their view, this validates the great possible that exists in the field of participatory surveillance, in particular for the management of emergency/crisis events.

E. Shi, H. T. H. Chan, E. Rieffel, R. Chow, and D. Song [13] consider how an untrusted data aggregator can learn desired statistics over numerous participants' data, without compromising each individual's privacy. They propose a construction that allows a group of participants to periodically upload encrypted values to a data aggregator, such that the aggregator is able to compute the sum of all participants' values in every time period, but is unable to learn anything else. They achieve strong privacy guarantees using two main techniques. First, they show how to exploit applied cryptographic techniques to allow the aggregator to decrypt the sum from multiple cipher texts encrypted under different user keys. Second, they explain a distributed data randomization procedure that guarantees the differential solitude of the outcome statistic, even when a subset of participants might be compromised.

R. Bhaskar, S. Laxman, A. Smith, and A. Thakurta [14] shows how one can accurately discover and liberate the most significant patterns along with their frequencies in a data set containing sensitive information, while providing precise guarantees of privacy for the individuals whose information is stored there. In this proposed work they presented two efficient differentially private algorithms for top-K frequent prototype mining. In their algorithms they adapted the Exponential Mechanism and the Laplace noise-addition mechanism by introducing techniques that are capable in the context of frequent pattern mining. They introduced a new notion of utility for top-K pattern mining and provided theoretical analysis of their methods under this criterion. They also presented extensive experimental results that exhibit the effectiveness of their methods on the FIMI benchmark data sets. While the paper focuses on frequent pattern mining, the techniques developed here are applicable whenever the data mining output is a list of elements ordered according to an appropriately 'robust' measure of interest.

A possible future direction is to devise techniques that eliminate this dependency on the size of the universe of items, thereby extending the applicability of the algorithms to bigger and more complex data sets.

R. Chaytor and K. Wang [15] mentioned that, the random perturbation has been extensively studied in the literature as an important technique for privacy protection; However, previous methods suffer from a notoriously low retention probability under most practical scenarios, due to "over randomization" over the entire sensitive attribute domain. To address this problem, they proposed small domain randomization, which randomizes a sensitive value only within a subset of the entire domain. This approach retains more data while providing the same level of privacy. With improved utility, they proposed this approach as an substitute to classical partition-based approaches to privacy preserving data publishing. They propose this approach as an alternative to classical partition-based approaches to privacy preserving data publishing. There are two key issues: ensure the published sub-tables do not disclose more private information than what is permitted on the original table, and partition the table so that utility is maximized.

A. Evfimievski, J. Gehrke, and R. Srikant [16] mentioned that there has been increasing interest in the trouble of building accurate data mining models over aggregate data, while protecting privacy at the level of individual records. One approach for this problem is to randomize the values in individual records, and only unveil the randomized values. The model is then built over the randomized data, after first compensating for the randomization (at the aggregate level). This approach is potentially vulnerable to privacy breaches: based on the distribution of the data, one may be able to learn with high confidence that some of the randomized records satisfy a specified property, even though privacy is preserved on average. In this proposed work, they present a new formulation of privacy breaches, together with a methodology, "amplification", for limiting them. They presented a new definition of privacy breaches, and developed a general approach, called amplification, that provably limits breaches. Amplification can be used to limit privacy breaches with respect to any single-record property.

They conclude with some interesting directions for future examine.

- How to extend amplification to continuous distributions?
- What is the relationship bettheyen the specific randomization operators, and the tradeoff bettheyen privacy and accuracy? In particular, how to identify the randomization operator and parameters that will provide the highest correctness in the mining model for a given level of privacy breaches?
- Are there ways to combine the randomization and the protected multi-party computation approaches that work better than either approach alone?

V. Rastogi and S. Nath [17] propose the first differentially private aggregation algorithm for distributed time-series data that offers good practical utility without any trusted server. This addresses two important challenges in participatory data-mining applications where (i) individual users wish to publish temporally correlated time-series data (such as location traces, web history, personal health data), and (ii) an untrusted third- party aggregator wishes to run aggregate queries on the data. They have proposed these novel algorithms to privately analysis queries on distributed time-series data. Their first algorithm FPAkcan analysis long query sequences over correlated time series data in a differentially private way. FPAkperturbs k DFT coefficients of an analysis sequence, thereby improving the accuracy for an n-length query sequence from £(n) of existing algorithms to roughly £(k), if the k DFT coefficients can accurately reconstruct all the query analysis.

For achieving differential privacy in distributed setting, they propose DLPA algorithm that implements Laplace noise addition in a distributed way with O(1) complexity per user. Their experiments with three real data sets show that their solution improves accuracy of query analysis by orders of magnitude and also scales well with a large number of users.

Jing Yang Koh, Gareth W. Peters, Derek Leong, Ido Nevatand Wai-Choong Wong [18] proposes a Stackelberg framework for mobile crowd sensing applications to incentivize privacy-sensitive smartphone users while increasing the coverage of the dataset. Their approach offers stronger incentives to privacy-sensitive participants by allowing them to use cloaking regions to hide their precise location. They considered different crowd sourcing environments and analyzed their influence on the Nash equilibrium point using simulations. Their simulation results show that their model is significantly better than the non-locationaware and uniform-reward schemes in terms of data utility. Their future work will be to consider uncertainty and untruthfulness in the users' privacy coefficients and sensing costs.

Heba Aly [19] presented Map++: a system for automatically enriching digital maps via a crowd sensing approach based on standard cell phones. For energy efficiency, Map++ uses only low-energy sensors and sensors that are already running for other purposes. They presented the Map++ architecture as well as the features and classifiers that can accurately detect the different road features such as tunnels, bridges, crosswalks, stairs, and footbridges from the user traces. Also, they presented their probabilistic map update algorithm to enrich the digital map with the crowd-sourced semantics.

Senyuan Tan , Xiaoliang Wang, Guido Maier, Theynzhong Li [20] investigates an application of mobile crowd sensing to detect and examine the riding quality of public transport vehicles. The lightweight system leverages sensors equipped on participants' smartphones to collect surrounding information. By analyzing the uploaded data at a server, they are able to estimate both aggressive driving behaviors and environment contexts. Series of data processing methods are exploited to overcome the affection of body movement and road condition, and crowd sourcing is applied to improve the robustness of the results. They have tested this system in 3 different transportation in 3 cities. The results indicate that the system can provide sufficient accuracy (up to 91% with 7 phones) to identify dozens of riding-comfort metrics.

Yali Gao, Xiaoyong Li, Jirui Li, Yunquan Gao [21] propose a dynamic-trust-based recruitment framework (DTRF) for MCS system. Real-time direct trust and lightweight feedback aggregation trust are combined to select the well-suited participants. In addition, they adopt an adaptive height allocation approach to calculate the overall trust degree of the participants. Theoretical analysis and extensive simulation confirm that DTRF can efficiently select the trustworthy participants and effectively stimulate the participants to contribute high-quality sensor data and thus get high task completion rate and data quality.

Deyi Sun, Wing Cheong Lau [22] proposes to classify social relationship based on the interaction data from multiple communication channels in smartphone. They carried out a social interaction data collection campaign to collect real life interaction data and model it with social interaction matrix. In the statistical analysis, they found that the interactions between people show temporal, directional pattern, etc. In their relationship classification problem, SVM outperforms KNN and decision tree (accuracies are 82.4%, 79.9% and 77.6% respectively). The interaction features from online social network and physical location/proximity contribute more to the classification results. At last, with PCA, they embed the data from 65 to 9 dimensions while preserving high classification accuracy. They also use CUR decomposition to help us refine feature definition and save smartphone energy in data collection.

B. I. P. Rubinstein, P. L. Bartlett, L. Huang, and N. Taft [23] explore the release of Support Vector Machine (SVM) classifiers while preserving the privacy of training data. The SVM is a popular machine learning method that maps data to a high dimensional element space before learning a linear decision boundary. In this proposed work they present a pair of mechanisms for secret SVM learning, each of which releases a classifier based on a privacy-sensitive database of training data.

In each case they establish differential privacy of their mechanisms via the algorithmic solidity of regularized ERM a property that is typically used in learning theory to prove risk bounds of learning algorithms. They present efficient mechanisms for finite-dimensional aspect mappings and for (potentially infinite-dimensional) mappings with translation-invariant kernels. Differential privacy is established using algorithmic stability, a property used in learning theory to bound generalization error. Finally they conclude with lathery bounds on the differential privacy of any mechanism approximating the SVM.

They establish the high-probability, point wise similarity between the resulting function and the non-private SVM classifier through a smoothness result of regularized ERM with respect to perturbations of the RKHS.

Anaissi and M. Goyal [24] aims to discover knowledge patterns hidden in huge data set that can yield more understanding to the data holders and identify new opportunities for imperative tasks including tactical planning and decision making. This research delivers a policy for the implementation of a systematic analysis framework built on the established principles used in data mining and machine learning.

The major goal of that is to form the foundation of what they envisage will be a new recommendation system in the market. This research introduces a simple SVM-BAR framework for deriving association rules, mining clusters and classification for huge datasets of transaction contains information about the purchased products by the customers. This framework generates a list of association rules that subsequently grouped into different categories based on the product outcome of the rules. SVM used to test whether this cluster can capture the variation based on the different type of rules. According to the feedback received by the clients, this structure provides an actionable knowledge for the market’s analyst.

Their future work is to follow their constructionist data analysis approach to thoroughly review a range of market datasets.

S. Madge [25] uses daily closing prices for 34 technology stocks to calculate price volatility and impetus for individual stocks and for the overall sector. These are used as parameters to the SVM model. The model attempts to expect whether a stock price sometime in the future will be higher or lather than it is on a given day. They find little predictive ability in the short-run but definite predictive skill in the long-run.

Their future scopes are adding granularity, feature selection and analyzing other sectors and sizes. One limitation of this study is that they only looked at daily stock price data. As a result, their momentum and volatility parameters theyre calculated over themes. Future work would involve adding features related to the specific company and related to broader macroeconomic factors. Another area for future growth is to apply their model to stocks in other sectors.

Minal Gadiya1, S. V. Jain [26] mentioned that nowadays identifying user attributes from their social network activities has been a ordinary research topic. Age, gender and interest can be common user attributes which can be predicted and are necessary for personalization and recommender systems. Most of the researches are based on the textual content created by user, whereas recently multimedia has gained popularity in social networks. In this research they propose an algorithm that predicts the user gender on different networking sites.

Initially they will predict the gender of the user from the posting performance and the visual content of the images and then the performance will be measured in terms of accuracy, precision, recall and F-measure.

Deyi Sun Wing Cheong Lau [27] proposed that wireless Communications and Mobile Computing have fundamentally changed the way people interrelate and communicate with each other. As the command-center of the user’s communications with the outside world, smartphones hold the key to appreciate the user’s social relationship with other people of interest. In this research, they propose to use the exclusive multi-model interaction data from smartphone to classify social relationships.

They firstly carry out a social interaction data collection campaign with a group of smartphone users to attain real-life multi-modal communication data and model the data as a social interaction matrix. Then they perform a statistical analysis on the social interaction matrix to identify the interesting interaction patterns in the data. They propose to classify social relationship based on the interaction data from multiple communication channels in smartphone. They carried out a social interaction data collection campaign to collect real life interaction data and model it with social interaction matrix.

Muhammad Habib ur Rehman, Chee Sun Liew, Teh Ying Wah, Junaid Shuja and Babak Daghighi [28] proposed that the staggering growth in smartphone and wearable device use has led to a immense scale generation of personal (user-specific) data. To explore, analyze, and extract helpful information and knowledge from the deluge of personal data, one has to leverage these devices as the data-mining platforms in ubiquitous, pervasive, and big data environments.

The staggering growth in PSDs is a key enabler in PerDM in RCEs for personalization, privacy, and security at user premises. Moreover, the exploitation of data mining algorithms in PSDs enables the use of PEs for personal good. User-centric big data personalization is a concept with a wide range of application in health care, ttheirism, education, e-government, and smart cities, among others. It has immense potential in personalization for better patient, traveler, customer, student, and citizen experiences.

Yllei ang, UaI.gUl. R, Zhen Qm, Theyntao Zheng , Linfang Yu , Zhentheyi Geng [29] they propose a novel prediction approach for predicting the users' context information by analyzing the social pictures shared by individuals on OSNs. To achieve this goal, they first crawl a large collection of photos and user profiles. Next, they represent a user by extracting image features and converting to multiple user-level characteristics. Based on that, they identify the different types of users. Then, by learning user-level characteristics to construct training data, and using Support Vector Machine (SVM) and Gradient Boosting (GBDT) classifier, they predict user's relative information. Their result reveals that method can effectively infer users' context information from images posted by users.

This research proposed a novel approach to predict users' context information by leveraging the perspectives of social pictures shared by users. They framed the problems as classification problems and evaluated the potential relationship between online users and their photos.

Jianxiong Wang Tom Down [30] mentioned that the Support vector machines (SVMs) have recently emerged as a powerful technique for solving problems in pattern classification and regression. Best performance is obtained from the SVM its parameters have their values optimally set. In practice, good parameter settings are usually obtained by a lengthy process of trial and error. This research describes the use of genetic algorithm to evolve these parameter settings for an application in mobile robotics.

In this research they have demonstrated how genetic algorithms can be used to tune the parameters of a pattern classifier. This demonstration was in terms of object recognition on a robot platform employing raw sonar data. They anticipate that results superior to those reposed here will be achieved through enhancements such as the preprocessing of the sonar data and other possible modifications to the procedure outlined in the previous section.

III. PROPOSED APPROACH

The objective of the crowdsensing is to analyze the data of various participants to achieve some goals. 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, SVM classifier and their extensions require a large number of samples to achieve proper estimation.

In this proposed work we use multiple dummies for crowdsensing. 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.

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.

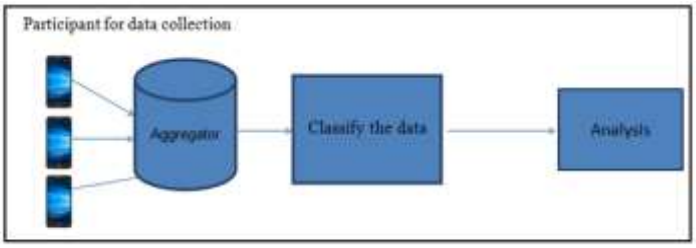

Fig. 1. System Architecture.

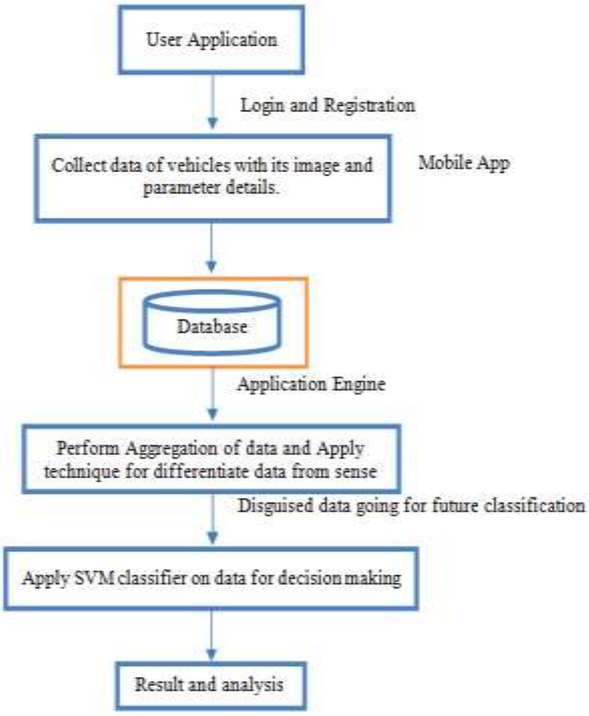
User Application
Login and Registration
Collect data of vehicles with its image and parameter details.
Mobile App
Database
Application Engine
Perform Aggregation of data and Apply technique for differentiate data from sense
Disguised data going for future classification
Apply SVM classifier on data for decision making
Result and analysis

Fig. 2. Flow of Proposed approach.

IV. OBJECTIVES OF THE PRESENT WORK

The objectives of the proposed approach are best described as below:

1. 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.
2. To analyze the sensing data of users.
3. To replace collaborative filter with SVM (Support Vector Machine) based classifier.
4. To improve the overall accuracy of the system.

V. CONCLUSION

In this proposed work we are performing privacy-preserving mobile crowdsensing where each participant's mobile phone of sensing the data, calculating the category ID from the sensed data, anonymzing the category ID, and sending the disguised category ID to the aggregator. The replaced category is sent to the aggregator, which attempts to estimate the distribution of the original categories of participants. However, RR schemes require great many samples in order to achieve proper reconstruction. In this proposed work, we propose S2M and S2Mb schemes, which can supersede existing RR schemes. Here we are applying SVM classifier on disguised data for decision making which will useful to us for marketing survey which improves the accuracy of system.

VI. 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] 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] 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.

[11] 11. Yohan Chon, Nicholas D. Lane, Yunjong Kim, Feng Zhao, Hojung Cha, "Understanding the Coverage and Scalability of Place-centric CrowdSensing", UbiComp'13, September 8–12, 2013.

[12] 12. Apostolos Malatras, Laurent Beslay, "A generic framework to support participatory surveillance through crowdsensing", Proceedings of the Federated Conference on Computer Science and Information Systems, IEEE 2015.

[13] E. Shi, H. T. H. Chan, E. Rieffel, R. Chow, and D. Song, "Privacypreserving aggregation of time-series data," in Proc. NDSS, 2011.

[14] R. Bhaskar, S. Laxman, A. Smith, and A. Thakurta, "Discovering frequent patterns in sensitive data," in Proc. ACM KDD, 2010.

[15] R. Chaytor and K. Wang, "Small domain randomization: Same privacy, more utility," Proc. VLDB Endowment, vol. 3, nos. 1–2, pp. 608–618, 2010.

[16] A. Evfimievski, J. Gehrke, and R. Srikant, "Limiting privacy breaches in privacy preserving data mining," in Proc. ACM PODS, 2003.

[17] V. Rastogi and S. Nath, "Differentially private aggregation of distributed time-series with transformation and encryption," in Proc. ACM SIGMOD, Jun. 2010.

[18] Jing Yang Koh, Gareth W. Peters, Derek Leong, Ido Nevatand Wai-Choong Wong," Privacy-Aware Incentive Mechanism for Mobile Crowd Sensing "IEEEICC, 2017.

[19] Heba Aly, "Automatic Rich Map Semantics Identification through Smartphone-based Crowd-sensing"DOI IEEE, 2016.

[20] Senyuan Tan , Xiaoliang Wang, Guido Maier, Theynzhong Li ,"Riding Quality Evaluation through Mobile Crowd Sensing" IEEE, 2016.

[21] Yali Gao, Xiaoyong Li, Jirui Li, Yunquan Gao, "DTRF: A Dynamic-Trust-based Recruitment Framework for Mobile Crowd Sensing System"2017 IFIP.

[22] Deyi Sun, Wing Cheong Lau, "Social Relationship Classification based on Interaction Data from Smartphones" IEEE 2013.

[23] B. I. P. Rubinstein, P. L. Bartlett, L. Huang, and N. Taft ,"Learning in a Large Function Space: Privacy-Preserving Mechanisms for SVM Learning." In Proc. JPC, 2012.

[24] A. Anaissi and M. Goyal, "SVM-Based Association Rules for Knowledge Discovery and Classification." In Proc. IEEE, May 2016.

[25] S. Madge, "Predicting Stock Price Direction using Support Vector Machines." In Proc. IWRS, 2015.

[26] Minal Gadiya1, S. V. Jain, "A Study on Gender Prediction using Online Social Images", IJARCCE, Vol. 5, Issue 2, February 2016 .

[27] Deyi Sun Wing Cheong Lau " Social Relationship Classification based on Interaction Data from Smartphones", Pervasive Computing and Communications Workshops (PERCOM Workshops), IEEE , 2013

[28] Muhammad Habib ur Rehman *, Chee Sun Liew, Teh Ying Wah, Junaid Shuja and Babak Daghighi "Mining Personal Data Using Smartphones and Wearable Devices:A Survey", Sensors 2015.

[29] YIlei ang, UaI.gUl. R , Zhen Qm, Theyntao Zheng, Linfang Yu, Zhentheyi Geng "User Context Information Prediction Based on the Mobile Internet Social Pictures", International Conference on Computer and Communications, IEEE, 2016.

[30] Jianxiong Wang Tom Down "Tuning Pattern Classifier Parameters Using A Genetic Algorithm With An Application In Mobile Robotics",IEEE, 2013.

CITE AN ARTICLE

Belkhode, Y. R., & Mohod, S. W. (n.d.). ANALYSIS BASED ON SVM FOR UNTRUSTED MOBILE CROWDSENSING- A REVIEW. *INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY, 7*(1), 161-170