TECH-CHRONICLE
AN INTERNATIONAL E-JOURNAL ON
EMERGING TRENDS IN SCIENCE, TECHNOLOGY AND MANAGEMENT

Survey Paper on Optimizing Differential Data Collection With Multiple Random Dummies Crowd sensing Mobiles Using SVM

Ankita A Mathankar[1], **Mukul Pande**[2]

[1]M.Tech Student,TGPCET,Nagpur,Maharashtra,India
[2]Professor, Information Technology Department,TGPCET,Nagpur,Maharashtra,India

***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. 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. In this , we propose a new anonymized data-collection scheme that can estimate data distributions more accurately. Using simulations with synthetic and real datasets, we prove that our proposed method can reduce the mean squared error and the JS divergence by more than as compared with other existing studies. In this age of computer science each and every thing becomes intelligent and perform task as human. For that purpose there are various tools, techniques and methods are proposed. Support vector machine is a model for statistics and computer science, to perform supervised learning, methods that are used to make analysis of data and recognize patterns. SVM is mostly used for classification and regression analysis. Here we are developing an application for dummies by taking dada from that dummies we are going to analyse the progress.

***Keywords* —** Crowdsensing , SVM, data mining.

I. INTRODUCTION

Knowledge discovery in databases is relatively young and interdisciplinary field of computer science in this domain The actual task is the automatic or semi-automatic analysis of large quantities of data to extract previously unknown interesting patterns such as groups of data records (cluster analysis), unusual records (anomaly detection) and dependencies (association rule mining). This knowledge is formed using discovery of knowledge from the data which is generated by different domain is possible using data mining. Extraction of knowledge from data in a human-understandable structure is the main goal of data mining .Industries, education, business and many other domains required knowledge for growth and with stand in the point where they reached this kind of data is helpful.

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 phoneusers .In the crowdsensing, individuals collectively share environmental data with a data aggregator, and the aggregator analyzes 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.

A lot of studies have been proposed for privacy-preserving data aggregation. However, most of them require an a priori estimate of the fraction of malicious participants. In

crowdsensing, it is difficult for each participant to know how many participants there are.

Randomized response (RR)is a promising method for anonymized data collection. It can protect each participant's data even if the aggregator and $N - 1$ of N participants collude with each other. In RR, a sensed value is categorized as one of the predefined categories. That category is replaced by another category with certain probability, and then the disguised category is sent to the aggregator. Because the participant sends the true data to the server with probability p and the disguised data to the server with probability $1 - p$, the privacy of the participant is protected at a certain level.

Here we are using SVM (Support Vector Machin) SVM is mostly used for classification and regression analysis. The support vector machine has been chosen because it represents a framework both interesting from a machine learning perspective . A SVM is a linear or non-linear classifier, which is a mathematical function that can distinguish two different kinds of objects. These objects fall into classes, this is not to be mistaken for an implementation.

II. LITERATURE SURVEY

2.1 Differential Private Data Collection and Analysis Based on Randomized Multiple Dummies for Untrusted Mobile Crowdsensing Yuichi Sei and Akihiko Ohsuga, Member, IEEE 2017

This paper focuses on 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

2.2 S. Agrawal and J. R. Haritsa, "A framework for high-accuracy privacy preserving mining," in Proc. IEEE ICDE.

This paper introduce To preserve client privacy in the data mining process, a variety of techniques based on random perturbation of data records have been proposed recently. In this paper, we present a generalized matrix-theoretic model of random perturbation, which facilitates a systematic approach to the design of perturbation mechanisms for privacy-preserving mining.

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

This paper introduces 'Interactive Classifier System' a fast learning method that enables a mobile robot to acquire autonomous behaviours from interaction between human and robot. A mobile robot is able to quickly learn rules by directly teaching from an operator.

2.4 V. Rastogi and S. Nath, "Differentially private aggregation of distributed time-series with transformation and encryption," in Proc. ACM SIGMOD.

This paper intends to propose PASTE, the first differentially private aggregation algorithms for distributed time-series data that offer good practical utility without any trusted server. PASTE addresses important challenges in participatory data-mining applications.

2.5 M. M. Groat, B. Edwards, J. Horey, W. He, and S. Forrest, "Application and analysis of multidimensional negative surveys in participatory sensing applications," Pervasive Mobile Comput.

Participatory sensing applications rely on individuals to share personal data to produce aggregated models and knowledge. In this setting, privacy concerns can discourage widespread adoption of new applications. We present a privacy-preserving participatory sensing scheme based on *negative surveys* for both continuous and multivariate categorical data. Without relying on encryption, our algorithms enhance the privacy of sensed data in an energy and computation efficient manner.

2.6 Jinho Kim , Byung-Soo Kim, Silvio Savarese, "Comparing Image Classification Methods: K-Nearest-Neighbor and Support-Vector-Machines".

In this paper for a robot or a computer to perform tasks, it must recognize what it is looking at. Given an image a computer must be able to classify what the image represents. While this is a fairly simple task for humans, it is not an easy task for computers. Computers must go through a series of steps in order to classify a single image. In this paper, we used a general Bag of Words model in order to compare two different classification methods. Both K-Nearest-Neighbor (KNN) and Support-Vector-Machine (SVM) classification are well known and widely used. We were able to observe that the SVM classifier outperformed the KNN classifier. For future work, we hope to use more categories for the objects and to use more sophisticated classifiers.

2.7 J. S. Raikwal, Kanak Saxena, " Performance Evaluation of SVM and K-Nearest Neighbor Algorithm over Medical Data set".

In this paper explain that this age of computer science each and every thing becomes intelligent and perform task as human. For that purpose there are various tools, techniques and methods are proposed. Support vector machine is a model for statistics and computer science, to perform supervised learning, methods that are used to make analysis of data and recognize patterns. SVM is mostly used for classification and regression analysis. And in the same way k-nearest neighbor algorithm is a classification algorithm used to classify data using training examples. In this paper we use SVM and KNN algorithm to classify data and get prediction (find hidden patterns) for target. Here we use medical patients nominal data to classify and discover the data pattern to predict future disease, Uses data mining which is use to classify text analysis in future.

V.CONCLUSIONS

By comparing all the result it is found that Randomize Response have some restriction on producing result [1] and it also have some complexity in it. As the crowdsensing is get performed in our project so it is basically based on cloud computing so no need of security has takesplace[2]. Interactive classifier method isnot as good as SVM to produce result[3]. It also contain security which is not required[4]. It present a privacy-preserving participatory sensing scheme based on negative surveys for both continuous and multivariate categorical data[5]. By comparing both KNN and SVM it is found that the SVM produce result more fast and accurate than KNN.

VI. REFERENCES

[1] Differential Private Data Collection and Analysis Based on Randomized Multiple Dummies for Untrusted Mobile Crowdsensing Yuichi Sei and Akihiko Ohsuga, Member, IEEE 2017

[2] S. Agrawal and J. R. Haritsa, "A framework for high-accuracy privacy preserving mining," in Proc. IEEE ICDE.

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

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

[5] M. M. Groat, B. Edwards, J. Horey, W. He, and S. Forrest, "Application and analysis of multidimensional negative surveys in participatory sensing applications," Pervasive Mobile Comput.

[6] Jinho Kim , Byung-Soo Kim, Silvio Savarese, "Comparing Image Classification Methods: K-Nearest-Neighbor and Support-Vector-Machines" .

[7] J. S. Raikwal, Kanak Saxena, " Performance Evaluation of SVM and K-Nearest Neighbor Algorithm over Medical Data set".