International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169
Volume: 4 Issue: 4 740 - 742

Classification of Secured Encrypted Relational Data over Cloud by Using SVM

Ms. Pradnya Chavarkar
Department of Computer Engineering
Pillai HOC College of Engineering & Technology
pchavarkar@mes.ac.in

Prof. Rajesh Bhise
Department of Computer Engineering
Pillai HOC College of Engineering & Technology
rbhise@mes.ac.in

Abstract **-** Now a day's data mining is being widely used in several applications such as banking, medicine, scientific research and various other fields where data is huge. Data mining applications are commonly used for classification. In last few years, due to popularity of numerous protection issues, various theoretical and practical answers for the classification issue have been proposed under different security models. Now a day users can access as well as store their data anywhere and anytime by making use of cloud computing.

Cloud servers may not be trusted and therefore data need to be encrypted without losing the originality. In privacy preserving classification technique performance grow slowly with value of dataset record and paillier encryption key size(k).In this regards the Support Vector Machine(SVM) method is proposed to solve the classification problem over encrypted information under semi-honest model. One is able to build classifier whose performance is better than the existing methodology under different parameter settings.

Keywords - Data mining; Data encryption; Cloud computing; Privacy-preserving; Data decryption; SVM classifier.

*****

I. INTRODUCTION

Since a decades ago, to store, recovering and additionally handling the information distributed computing is utilized as a perfect model for dealing with the association's information. Distributed computing utilized as a part of number of associations, trustfully because of its potential relying upon its cost-effectiveness, adaptability, and offload of administrative overhead as a rising preparing perfect model. Reliably, association stores their computational operations notwithstanding their information to the cloud.

Confirmation and security issues in the cloud are dismissing relationship to use those positive circumstances, paying little regard to huge enthusiasm of diversion that the cloud offers. Precisely when information is to a great degree delicate, the information ought to be encoded before outsourcing to the cloud. Then again, when information are blended, free of the basic encryption techniques, performing any information mining assignments winds up being amazingly troublesome while never translating the information. We likewise can characterize information mining as a system of recovering information frame vast arrangement of information or information mining will be mining of learning from information.

In data industry today the endless sum information is available. On the off chance that it is not in changed over in some valuable data it is of no utilization. It is particularly obligatory to study this gigantic information and recover the helpful data from it. Extraction of information is not just the procedure we need to perform; data mining moreover information cleaning, information combination, information change, information mining, design assessment and information presentation are some issue to explain. Once every one of these techniques are over, we would have the ability to use this information in various applications, for example, extortion discovery, market investigation, generation control, science investigation, etc.

II. LITERATURE SURVEY

P. Williams et al. [1], query process is the read and write request for data item. The aim of the client is to find out data item without knowing to server. The system provides two conditions to achieve access pattern. First, no data item is queried twice using same label and second, is access pattern must appear indistinguishable. To find the query result bloom filter is use. Bits of bloom filter are encrypted so it hides the result of user query. This technique also reduces the complexity and storage overhead by using reshuffling protocol. Additional work of this paper is to allow throughput of queries per second on 1 Tbyte+database. This approach provides computational privacy and accuracy at large scale as compared with existing approaches.

De Capitani di Vimercati et al.[2],present the main privacy risk that occur in data outsourcing and cloud computing. Different privacy risks are introduced such as privacy risk in cloud, privacy risk for users, privacy risk for stored data, privacy risk for data access and try to illustrate some solution to solve these risks. Privacy risk in cloud is solved by data dissemination and sharing, anonymous communication, collaborative query result, external data storage, digital interactions. By using these solutions, privacy of the user's risk is preserved. Privacy risk for users is solved by attribute-based access control and users privacy preferences. Privacy risk for stored data is maintained by providing confidentiality and integrity to the stored data and using selective access

methodology. Privacy risk for data access is handled by maintaining the intigrity, query privacy.

R. Agrawal et al [3], develop accurate model for data mining without access of personal information of an individual record. In prior to this method decision tree classifier was build from training data but resulted data was different from original data. In [3] novel reconstruction procedure was proposed to correctly find out the distribution of original data. This technique build classifier whose accuracy is comparable to accuracy of classifiers built with original data. Author provided privacy-preserving method such as value class membership in which values for an attribute were partitioned into a set of disjoint, mutully exclusive classes. Reconstruction of original distribution is done by reconstructing the distribution for each attribute once at beginning using training data.

A. Evfimievski et al. [4] proposed classification model using training data in which sensitive numeric values of users record have been randomized so that true value cannot be find out. Before sending a transaction to server, client takes each item and replaces it by new item with probability P. New item is not originally present in transaction. While doing this, privacy breaches may occur. Privacy preservation and association mining rule is done by finding out problem of privacy breaches, also mathematical equation for randomization algorithm was used to derive formula for support and prediction.

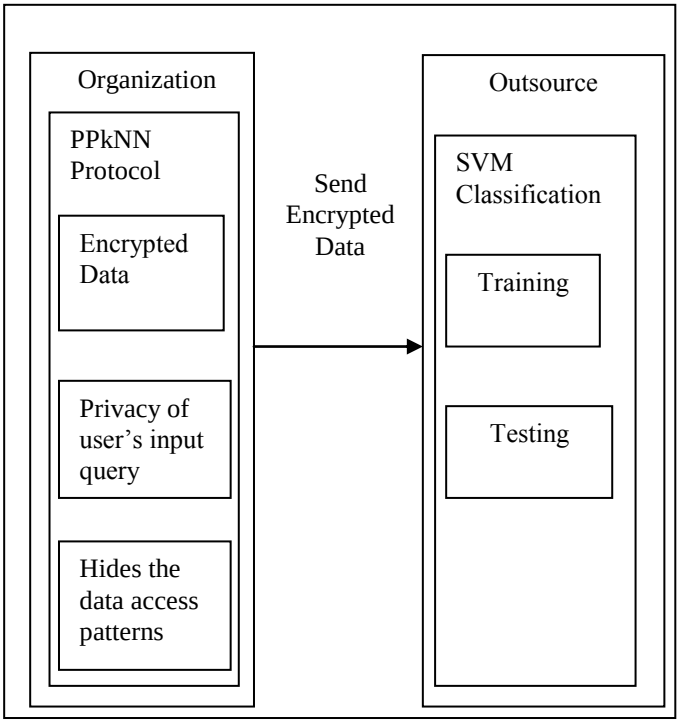
R. J. Bayardo et al. [5], hide the details of personal information by using k-anonymization. Even though data has been hidden such as name, social security number but by linking attack attacker can find out the personal information. In [5] proposed method for determining an optimal k-anonymization of given dataset. Algorithm proposed was not only finding k-anonymization but find out best or optimal solution. Tree-search strategy was used to solve the problem, which involve searching the power-set of a special alphabet of domain value

Y. Elmehdwi et al. [6] introduces two new SkNN protocols for data encryption in cloud. Initial protocol plays role a fundamental solution but leaks data of cloud. Next protocol provides more security. This protocol was able to convey the secrecy of the information, input query as well as avoid disclosing of data access patterns. Second protocol was more costly than the basic protocol. Efficiency of protocols over the various parameters was examined.

B. K. Samanthula et al. [7], proposed novel methods to effectively solve the DMED problem assuming that the encrypted data are outsourced to a cloud. They introduced privacy preserving k-NN classification protocol (PPkNN) which was constructed using protocols such as secure squared Euclidean distance, secure bit-decomposition, secure minimum, secure minimum out of n numbers, Secure Bit-OR (SBOR).Goal of the protocol was to classify user's query record in a privacy preserving manner. Goal was achieved by securely retrieving of k-Nearest Neighbour class and by securely computing majority class. PPkNN protocol protects the confidentiality of the data, users input query and hide the data access pattern. Performance of the protocol was measure under different parameter settings.

III. IMLEMENTAION DETAILS

System Overview



System Architecture

In this system new method is developed to overcome the issue of DMED successfully on the basis of encrypted data which is shared with a cloud. Mainly, our concentration is on the issue of classification .Classification is familiar task as compared to other data mining tasks. This method is used on encrypted information within the cloud computing. SVM (support vector machines) algorithm is utilized in our proposed system for classification as well as to prevent the issue of classification on encrypted data. SVM is used under the supervision and on the learning models. Those models are assigned with learning algorithms and algorithms are analysed as well as identify the patterns utilized for classification and regression analysis. Proposed algorithm saves the memory as well as it consumes less time as compared with existing and all data procedure are executed on the cloud.

Proposed system includes multiple clouds, client, and a data source, where a data source sends a data to cloud to store it in encrypted format. At the time of storing the data the data source gives the encrypted private key to the other cloud. When a client request a data to the cloud where we are using SVM classification. A client or a user sends encrypted query to the cloud, there is a process of training and testing on encrypted data. Both clouds perform training and testing operations on the query. The result is send back to the client in

encrypted form where the client decrypts the result. We also assume that both the clouds are working on semi honest models to work on the proposed system.

Mathematical Model

M = (Q,∑, P, q0, F)

Where

Q is the set of States

∑ is the set of inputs

p State Transition Table

q0 is the initial Stage

F is the final Stage

- Q: { S1, S2, S3 }

Where,

S1: Organizer

S2: PPkN Protocol

S3: Outsource

- { W1, W2, W3 }

Where

W1: input data

W2: Encrypted data

W3: SVM Classifier

- q0 : {S1 }
- F: {S3 }

State Transition Table

	W1	W2	W3
S1	S2		
S2		S3	
S3			S3

SVM Algorithm:

candidateSV = { nearest pair from opposite classes }
While there are disregarding points
do
Find a disregarder candidate
SV = candidateSV ∪ S Disregarder
If any αp < 0 due to addition of c to S
then candidateSV = candidateSV \ p
repeat until all such points are trimmed
end if
end while

Support vector machines (SVMs) are used to observe learning models with related learning algorithms. Algorithms are analysing information to use for classification and reverting analysis. Set of training samples provided and every set is separately related with one of two classifications. A SVM training algorithm builds a model which assign new cases into one classification. This model makes it a non-probabilistic binary linear classifier.

IV. CONCLUSION

In the cloud user data security is a primary concern and there are different privacy-preserving classification approaches are available to provide security. Previous approaches are not appropriate to database as well as it shares encrypted data with third-party server. To overcome this issue we propose a new SVM classification protocol for privacy preservation. Data is classified by SVM protocol stored in cloud. Our protocol mainly ensures the privacy of data, privacy of users query and also makes access patterns invisible. This protocol analyzed using different parameters.

Our aim is to build secure SVM classifier over encrypted information under semi-honest model

REFERENCES

[1] P. Williams, R. Sion, and B. Carbunar, 11 Building castles out of mud: Practical ac-cess pattern privacy and correctness on untrusted storage," in Proc. 15th ACM Conf.Comput. Commun. Security, 2008, pp. 139-148.

[2] De Capitani di Vimercati, S. Foresti, and P. Samarati, "Managing and accessing data in the cloud: Privacy risks and approaches," in Proc. 7th Int. Conf. Risk Security Internet Syst., 2012, pp. 1–9.

[3] R. Agrawal and R. Srikant, "Privacy-preserving data mining," ACM Sigmod Rec., vol. 29, pp. 439–450, 2000.

[4] A. Evfimievski, R. Srikant, R. Agrawal, and J. Gehrke, "Privacy preserving mining of association rules," Inf. Syst., vol. 29, no. 4, pp. 343–364, 2004.

[5] R. J. Bayardo and R. Agrawal, "Data privacy through optimal k- anonymization," in Proc. IEEE 21st Int. Conf. Data Eng., 2005, pp. 217–228.

[6] Y. Elmehdwi, B. K. Samanthula, and W. Jiang, "Secure k-nearest neighbor query overencrypted data in outsourced environments," in Proc. IEEE 30th Int. Conf. Data Eng., 2014, pp. 664-675.

[7] B. K. Samanthula, Y. Elmehdwi, and W. Jiang, "k-nearest neighbor classification over semantically secure encrypted relational data," eprint arXiv:1403.5001, 2015.