MJRET
Open Access
Multidisciplinary Journal of Research in Engineering and Technology
Journal homepage: www.mjret.in
ISSN:2348 - 6953

K-NN CLASSIFICATION OVER SECURE ENCRYPTED RELATIONAL DATA IN OUTSOURCED ENVIRONMENT

Akshay Dabi, Arslan Shaikh, Pranay Bamane,
Vivek Thorat, Prof.Popat Borse.

Computer Engineering. Dr. D.Y Patil School of Engineering
Pune, India

Abstract: *Data Mining has wide applications in numerous zones for example banking, medicine, and exploratory examination and among government offices. Classification is one of the ordinarily utilized assignments as a part of data mining applications. For as long as decade, because of the ascent of different security issues, numerous hypothetical and handy answers for the arrangement issue have been proposed under diverse security models. On the other hand, with the later prevalence of distributed computing, clients now have the chance to outsource their information, in encoded structure, and additionally the information mining assignments to the cloud. Since the information on the cloud is in encoded structure, existing security safeguarding characterization procedures are not persistent. In this paper, we concentrate on taking care of the arrangement issue over encoded information. Specifically, we propose a protected K-NN classifier over encoded information in the cloud. The proposed convention ensures the secrecy of information, security of client's information question, and conceals the information access designs. To the best of our insight, our work is the first to add to a safe K-NN classifier over encoded information under the semi-fair model. Likewise, we observationally investigate the proficiency of our proposed convention utilizing a real world dataset under diverse parameter settings.*

Keywords*: Security, k-NN Classifier, Outsourced Databases, Encryption.*

1. INTRODUCTION

As of late, the cloud computing [1] world view is reforming the associations' method for working their information especially in the way they store, access and process information. As a developing registering world view, distributed computing draws in numerous associations to consider truly with respect to cloud potential as far as its expense proficiency, adaptability, and offload of managerial overhead. Frequently, associations delegate their computational operations not withstanding their information to the cloud. In

spite of enormous points of interest that the cloud offers, protection and security issues in the cloud are anticipating organizations to use those favourable circumstances.

At the point when information is exceedingly delicate, the information should be encoded before outsourcing to the cloud as suggested in [2]. Be that as it may, when information are encoded, independent of the fundamental encryption plan, performing any information mining assignments turns out to be extremely difficult without ever unscrambling the information. There are other security concerns, shown by the accompanying sample.

Data mining over encrypted data (denoted by DMED) [3] on a cloud also needs to protect a client's record when the record is a part of a data mining process. However cloud can also abstract useful and sensitive information about the outsource data items by observing the data access patterns even if the data are encrypted. Therefore, the privacy/security requirements of the DMED problem on a cloud are of three types: (1) privacy of the encrypted data, (2) privacy of a user's query record, and (3) hiding data access patterns

2. OBJECTIVE

- Novel PPkNN protocol: A secure k-NN classifier over semantically secure encrypted data.
- Privacy Preserving Approach.

3. SCOPE

This project may be wide applications in many areas such as banking, medicine, scientific research and among government agencies.

4. PROBLEM DEFINITION

Suppose John owns a database DB of n records t 1 , . . . , t n and m + 1 attributes. Let t p,q denote the q the attribute value of record t p . Initially, Ricky encrypts his database attribute-wise, that is, he computes E pk (t p,q), for 1 ≤ p ≤ n and 1 ≤ q ≤ m + 1, where column (m + 1) contains the class labels. We assume that the underlying encryption scheme is secure [45]. Let the encrypted database be denoted by DB′ . We assume that John outsources DB ′ as well as the further classification process to the outsourced data.

Let Bob be an authorized user who wants to classify his input record r = hr rickypointing39@gmail.com1 , . . . , r m p by applying the k-NN classification method based on DB ′ . We refer to such a process as privacy-preserving k-NN (PPkNN) classification over encrypted data in the cloud. Formally, we define the PPkNN protocol as:

PPkNN(DB ′ , r) → c r

Where c r denotes the class label for r after applying k-NN classification method on DB ′& r.

5. BACKGROUND

As suggested in [4] fully homomorphism system can solve the problem as third party is used for the arbitrary functions but such techniques are costly. By using Shamir's scheme [5] we can develop PPkNN but our work is different in other way. Existing work on PP Data Mining

(either perturbation or secure multi-party computation based approach) cannot solve the DMED problem. Uneasy data do not possess semantic security, so data perturbation techniques cannot be used to encrypt highly delicate data. Also the uneasy data do not produce very accurate data mining results. Secure user computation (SMC) based approach assumes data are distributed and not encrypted at each participating party.

6. MODULES

- PRIVACY-PRESERVING PRIMITIVES
- Secure Minimum (SMIN)
- Secure Minimum out of n Numbers (SMINn)
- Secure Frequency (SF)

7. BASE ALGORITHMS

PRIVACY-PRESERVING PRIMITIVES:-Here we present a set of generic sub-protocols that will be used in constructing our proposed k-NN protocol as given in [3].All of the below protocols are considered under two-client semi-honest setting. In particular, we consider the presence of two semi honest clients P1 and P2 such that the Palliser's secret key sk is known only to P2 whereas ik is public.

• **Secure Minimum (SMIN):-** In this protocol, P1holds private input (u′, v′) and P2 holds sk, where u′ = ([u],Epk(su)) and v′ = ([v],Eik(sv)). Here su (resp., sv) denotes the secret associated with u(resp., v). The goal of SMIN is for P1 and P2 to jointly Here we present a set of generic sub-protocols that will be used in constructing our proposed k-NN protocol .All of the below protocols are considered under two-clients semi-honest setting. In particular, we consider the presence of two semi honest clients P1 and P2 such that the Palliser's secret key sk is known only to P2 whereas ik is public.

• **Secure Minimum out of n Numbers (SMINn):-**In this protocol, we consider P1 with n encrypted vector's ([d1], [dn]) along with their corresponding encrypted secrets and P2 with sk. Here[dp] = hEik(dp,1), . . . ,Eik(dp,l)i where dp,1 anddi,l are the most and least significant bits of integer irrespectively, for 1 ≤ p≤ n. The secretor dp is given by sdi . P1 and P2 jointly compute[min(d1, . . . , dn)]. In addition, they compute Epk(smin(d1,...,dn)). At the end of this protocol ,the output ([min(d1, . . . , dn)],Epk(smin(d1,...,dn)))is known only to P1. During SMINn, no information regarding any of dp's and their secrets is revealed to P1 and P2.

• **Secure Frequency (SF):-** Here P1 with private input (hEik(c1), . . .Eik(cw)p, hEik(c′1), . . . ,Eik(c′k)p)and P2 securely compute the encryption of the frequency of cq , denoted by f(cq), in the listhc′1, . . . , c′kp, for 1 ≤ q≤ w. Here we explicitly assume that cq 's are unique and c′p ∈{c1, . . . , cw}, for 1 ≤ p≤ k. The output Eik(f(c1)), . . . ,Eik(f(cw))p will be known only to P1. During the SF protocol, no data regarding c′p, cq , and f(cq) is revealed to P1 andP2, for 1 ≤ p≤ k and 1 ≤ q≤ w.

8. MATHEMATICAL MODEL

Let S is the Whole System Consist of
S= {Q, PPKNN, D', SRKNN, SCMCK, PPP}.
Where Q is set of query entered by user.

Q={q1, q2, q3,…..qn}.
D' = Encrypted Data set.
PPKNN = process as privacy-preserving k-NN.
SRKNN = Secure Retrieval of k-Nearest Neighbours.
SCMCK = Secure Computation of Majority Class.
PPP = Privacy-Preserving Primitives.

9. PROCEDURE

The proposed PPkNN protocol mainly consists of the following two stages:

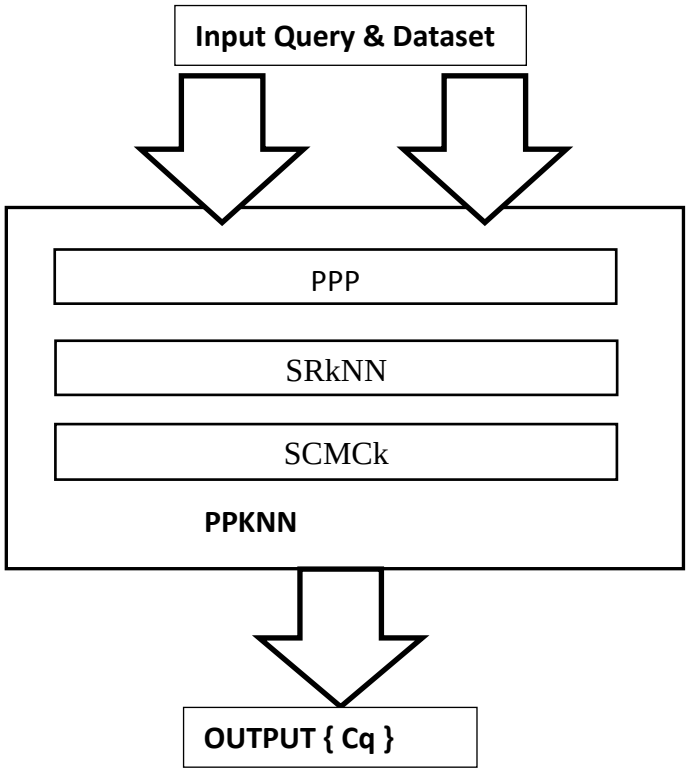
Stage 1: Secure Retrieval of k-Nearest Neighbors (SRkNN):

- In this stage, User initially sends his query q (in encrypted form) to C1.
- After this, C1 and C2 involve in a set of sub-protocols to securely retrieve (in encrypted form) the class
- Labels corresponding to the k-nearest neighbors of the input query q.
- At the end of this step, encrypted class labels of k-nearest neighbors are known only to C1.

Stage 2: Secure Computation of Majority Class (SCMCk):

- C1 and C2 jointly compute the class label with a majority voting among the k-nearest neighbors of q.
- At the end of this step, only User knows the class label corresponding to his input query record q.

Procedural flow diagram:-



10. CONCLUSION

In this paper, we have discussed the PP techniques for data mining, and the usage of the best applicable data perturbation technique with multilevel trust privacy.. Here the different processes are to be done on the relational data.

ACKNOWLEGEMENT

Authors of this paper are thankful to Prof. Popat Borse for his valuable and constructive suggestions. Authors sincerely thank Head of Department (Computer Engineering) Prof. Soumitra S. Das for his reassuring encouragement throughout the preparation of our Paper. Also thanking to all others who have tried hard to make their work easy to accomplish.

REFERENCES

[1] "The NIST Definition of Cloud Computing" by Peter Mell and Tim Grance ,Version 15, 10-7-09.

[2] "Building Castles out of Mud: Practical AccessPattern Privacy and Correctness on Untrusted Storage" by Peter Williams ,Radu Sion ,Bogdan Carbunar.

[3] "k-Nearest Neighbor Classification over Semantically Secure Encrypted Relational Data" by Bharath K. Samanthula, Member, IEEE, Yousef Elmehdwi, and Wei Jiang, Member, IEEE.

[4] "C. Gentry, "Fully homomorphic encryption using ideal lat-tices," in ACM STOC, pp. 169–178, 2009."

[5] Shamir, "How to share a secret," Commun. ACM, vol. 22,pp. 612–613, Nov. 1979.