

Concept of hash key for the enhancement of secure mining in horizontally distributed database

Raunak Rathi, A.V.Deorankar, Roshni M. Sherkar¹
M.tech student, GCOEA, raunakrathi.rathi@gmail.com,9021904006
²*HOD, GCOEA, avdeorankar@gmail.com, 9405321946*
³*M.tech student, GCOEA, rinku.sherkar@gmail.com*

Abstract: Fast growing area today is data mining which deals with large data collection to extract knowledge, but such collections are divided into several parties. In such case privacy plays an key role, to maintain the privacy of data some measures for prevention may not allow the parties from directly sharing the data and its information. In this paper we used the protocol from paper [1], and proposed a hash function in this part which will increases the efficiency of the protocol. The current protocol which we use in this paper is based on the protocol of two novel secure multi-party algorithm. The hash function will be automatically generated and as the hash is automatic generated the groups will be formed and the privacy will also be maintained. Using the hash key technique with respect to the protocol it will be beneficial in terms of communication cost.

Keywords – secure1, mining2,

1. INTRODUCTION

In order to perform privacy preserving data mining. Furthermore, the problem has been discussed in multiple communities such as the database community, the statistical disclosure control community and the cryptography community. Data mining and KDD(Knowledge discovery in database) are two different kind of research area which examine the auto extraction of earlier unidentified pattern from huge amount of data. To find the solution of secure mining has become more essential in upcoming years due to the rising capability to save personal data about users and the rising complexity of data mining algorithm to influence this information. A number of technique as such classification, kanonymity, association rule mining, clustering had been recommended in upcoming years in order to performed secure data mining. Besides, the difficulty has been discussed in several community such as the database community, the statistical disclosure control community and the cryptography community. Data mining technique has been evolved successfully to extract knowledge in such to maintain a variety of domains weather, national security, forecasting, medical diagnosis, and marketing. Although it is confront to mine such kind of data without violating the data owner’s privacy. For example, how to mine an employee private data is an ongoing problem in multinational company’s application. As such data mining become more enveloping, secure concern are rising.

Marketable concern are also concerned with the secure subject. Most organization gather knowledge about individual for their own exact needs. Very regularly, however, different units within an organization themselves may find it necessary to share information. In such cases, each organization or unit must be sure that the security of an individual is not dishonored or that responsive commercial information is not discovered. Consider, for example, a government, or more properly, one of it’s secured branch involved in rising a system for formative, from passengers whose luggage had been checked, those who must be subjected to additional privacy measures. The data signify the necessity for additional assessment derives from a wide variety of source such as and generally government statistics, police proceedings, airports, banks, passenger information records that generally include personal information (such as name and phone no.); demographic data (such as age and sex); flight information(such as duration, departure and destination); and expenditure data (such as transfer, bank transactions and purchasing). In most countries, this information is regarded as private and to avoid deliberately or involuntarily exposing private information about an individual, it is against the law to make such information liberally available.

While several means of preserving entity information has been enhanced, there are ways for circumventing these methods. In our example, in order to sustain privacy, passenger information records can be unidentified before the records are shared with anyone who is not allowable directly to access the related

data. This can be accomplished by removing from the dataset distinct fields, such as name and phone no. However, even if this information is deleted from there are still other kinds of information, private or behavioral (e.g. number of accounts zip code, date of birth, no. of children, gender, number of calls) that, when connect with other accessible datasets, could potentially identify subjects. To evade these types of violation, we need some other data mining algorithm for secure preserving.

2. Existing work

Data mining is a beneficial technique used to extract data/knowledge from large collection of data, but the collection of data is in distributed form many times. In such case privacy plays an important role to maintain the privacy of data or some part of knowledge about the data. The problem here we will discuss from paper[1]. In horizontal distributed data plays an important factor to look for is the distributed database. Here several players that access homogenous databases, i.e., the database that share the same schema but hold different information. The paper[1] support at least S and confidence C , for some given minimum support size of S and confidence C , that hold in united database, while reducing the information release about the secure (or private) database accessed by such players.

The paper deals with the problem of secure multi-party calculation. If a trusted third party would be present, then the players could devote to such party and such party would evaluate and send them such resulting output. If such third party would not be present, it is need to develop a protocol that player can use on own in order to get their required output Y . If no player learn from such view, these protocol is consider perfectly secure more than that the third party would learn the ideal settings where the calculation is carried out by the trusted third party. The protocol that we used here calculates a parameterized family of functions, which we can say as a threshold function, in which the two excessive cases match up to the problem of calculating the union and insertion of private subsets. Those can be said as general purpose protocol that can be used in other part as well. One more problem regarding secure multi-party calculation is the set of addition problem; namely, the problem in which Bobs holds a private subsets of several ground set, and Alice hold an element in the ground set, and they desire to decide whether Alice's element is within Bob's subsets, exclusive of revealing to either of them Knowledge about the other party's input beyond the above describe addition. Here the existing work is an alternative protocol for the secure calculation of the union private subsets. The protocol get better when we use hash key function which will we elaborate in proposed work part.

3. Hash key

A hash function is the function so as to be used to plot digital data of random size to digital data of permanent size, with small difference in input data producing very large difference in output data. The ideals returned by a hash function are called hash ideals, simply hash, or hashes codes. One sensible make use of is a data structure called a hash table, extensively used in computer software for fast data hunt for. Hash functions speed up table or database search for by detecting duplicated report in a big file. An pattern is finding similar stretch in DNA sequence. They are also helpful in cryptography. A cryptographic hash function allow one to with no trouble verify that some input data match a stored hash value, but makes it hard to rebuild the data from the hash alone. This standard is used by the PGP algorithm for data justification and by a lot of password examination system.

Hash functions are linked to (and often confused with) ciphers, error-correcting codes, randomization function fingerprints checkdigits and checksums. Although these concepts partly cover to some extent, each has its own uses and necessities and is considered and optimized differently. The Hash Keeper database maintain by the American National Drug Intelligence Center, for instance, is more aptly described as a catalog of file fingerprints than of hash values.

4. Proposed Work

An rising number of databases have become web easily reached from end to end HTML form-based search interfaces. The data units return from the fundamental database are typically encoded into the outcome pages dynamically for human browse. For the programmed data unit to be machine procedure able, which is necessary for many application such as deep web data collected works and internet link shopping, they require to be extract out and assigned meaningful labels. In this paper, we present an automatic explanation approach that primary aligns the data units on a consequence page into dissimilar groups such that the data in the similar group have the similar semantic. Then, for every group we explain it from dissimilar aspect and combined the different annotations to forecast a final explanation label for it. An explanation wrapper for the look for site is automatically construct and can be used to explain new end result pages from the similar web database.

5. Advantages

As a rising subject, data mining is playing an increasingly important role in the decision support

activity of every walk of life. Get Efficient Item set result based on the customer request.

6. Conclusion

The paper we presented here decrease the computation time from the protocol we referred. The hash function will increase the speed. The automatic allocation of a hash can be achieved and this will be used for the future scope of the methodology.

REFERENCES

- [1] Tamir tassa, “Secure Mining of Association Rules inHorizontally Distributed Databases”, IEEE transactions on knowledge and data engineering, 2013.
- [2] LiWu Chang and Ira S. Moskowitz, *Parsimonious downgrading and decision trees applied to the inference problem*, In Proceedings of the 1998 New Security Paradigms Workshop (1998), 82–89.
- [3] Mike J. Atallah, Elisa Bertino, Ahmed K. Elmagarmid, Mohamed Ibrahim, and Vassilios S. Verykios, *Disclosure Limitation of Sensitive Rules*, In Proceedings of the IEEE Knowledge and Data Engineering Workshop (1999), 45–52.
- [4] LiWu Chang and Ira S. Moskowitz, *An integrated framework for database inference and privacy protection*, Data and Applications Security (2000), 161–172, Kluwer, IFIP WG 11.3, The Netherlands.
- [5] Nabil Adam and John C. Wortmann, *Security-Control Methods for Statistical Databases: A Comparison Study*, ACM Computing Surveys **21** (1989), no. 4, 515–556.
- [6] Rakesh Agrawal and Ramakrishnan Srikant, *Privacy-preserving data mining*, In Proceedings of the ACM SIGMOD Conference on Management of Data (2000), 439–450.
- [7] David W. Cheung, Jiawei Han, Vincent T. Ng, Ada W. Fu, and Yongjian Fu, *A fast distributed algorithm for mining association rules*, In Proceedings of the 1996 International Conference on Parallel and Distributed Information Systems (1996).
- [8] Chris Clifton, Murat Kantarcioglu, Xiadong Lin, and Michael Y. Zhu, *Tools for privacy preserving distributed data mining*, SIGKDDExplorations **4** (2002), no. 2.
- [9] Chris Clifton and Donald Marks, *Security and privacy implications of data mining*, In Proceedings of the ACM SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery (1996), 15–19.
- [10] Elena Dasseni, Vassilios S. Verykios, Ahmed K. Elmagarmid, and Elisa Bertino, *Hiding Association Rules byusing Confidence and Support*, In Proceedings of the 4th Information Hiding Workshop (2001), 369–383.
- [11] Wenliang Du and Mikhail J. Atallah, *Secure multiproblem computation problems and their applications: A review and open problems*, Tech.
- [12] Wenliang Du and Zhijun Zhan, *Building decision tree classifier on private data*, In Proceedings of the IEEE ICDM Workshop on Privacy, Security and Data Mining (2002).
- [13] Alexandre Ev.mievski, Ramakrishnan Srikant, Rakesh Agrawal, and Johannes Gehrke, *Privacy preserving mining of association rules*, In Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2002).
- [14] Ioannis Ioannidis, Ananth Grama, and Mikhail Atallah, *A secure protocol for computing dot products in clustered and distributed environments*, In Proceedings of the International Conference on Parallel Processing (2002).
- [15] Murat Kantarcioglu and Chris Clifton, *Privacy preserving distributed mining of association rules onhorizontally partitioned data*, In Proceedings of the ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery (2002), 24–31.