International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

Survey on K-Nearest Neighbor Categorization over Semantically Protected Encrypted Relational Information

Pranali D. Desai [1], Vinod S. Wadne [2]

[1]Department of Computer Engineering, Imperial College of Engineering And Research, Wagholi, Pune

[2]Professor, Department of Computer Engineering, Imperial College of Engineering And Research, Wagholi, Pune

Abstract: ***Data Mining has wide use in many fields such as financial, medication, medical research and among govt. departments. Classification is one of the widely applied works in data mining applications. For the past several years, due to the increase of various privacy problems, many conceptual and realistic alternatives to the classification issue have been suggested under various protection designs. On the other hand, with the latest reputation of cloud processing, users now have to be able to delegate their data, in encoded form, as well as the information mining task to the cloud. Considering that the information on the cloud is in secured type, current privacy-preserving classification methods are not appropriate. In this paper, we concentrate on fixing the classification issue over encoded data. In specific, we recommend a protected k-NN classifier over secured data in the cloud. The suggested protocol defends the privacy of information, comfort of user's feedback query, and conceals the information access styles. To the best of our information, our task is the first to create a protected k-NN classifier over secured data under the semi-honest model. Also, we empirically evaluate the performance of our suggested protocol utilizing a real-world dataset under various parameter configurations.***

Keywords: Security, k-NN classifier, outsourced databases, encryption

1. Introduction

Lately, the cloud computing model [1] is changing the landscape of the organizations' way of working their information especially in the way they save access and process data. As a growing processing model, cloud processing draws many organizations to think about seriously concerning cloud potential with regards to its cost-efficiency, versatility, and offload of management expense. Most often, organizations assign their computational functions in improvement to their information to the cloud. Regardless of remarkable benefits that the cloud offers, security and comfort issues in the reasoning are avoiding companies to utilize those benefits. When information is extremely delicate, the information need to be encoded before freelancing to the cloud. Nevertheless, when information are secured, regardless of the actual security plan, executing any information mining tasks turns into very complicated without ever decrypting the information. There are other privacy worries, confirmed by the following example.

Example 1: assume an insurance provider contracted its secured clients database and relevant data mining task to a cloud. When a representative from the company needs to figure out the threat stage of a potential new client, the representative can use a classification method to figure out the threat stage of the client. Initial, the representative requires generating a details history q for the client containing certain private details of the client, e.g., credit rating, age, marriage status, etc. Then this history can be sent to the cloud, and the cloud will estimate the class label for q. However, since q contains vulnerable details, to secure the customer's privacy, q should be encoded before delivering it to the cloud.

The above example reveals that data mining over encoded information (denoted by DMED) on a cloud also requires securing a user's history when the history is a part of a data mining procedure. Furthermore, cloud can also obtain helpful and delicate information about the real information products by monitoring the information accessibility styles even if the information are encoded [2], [3]. For that reason, the privacy/security specifications of the DMED issue on a cloud are threefold: (1) comfort of the encoded information, (2) comfort of a user's query history, and (3) concealing information accessibility patterns.

Current work on privacy-preserving data mining (PPDM) (either perturbation or protected multi-party computation (SMC) centered approach) cannot fix the DMED issue. Perturbed information do not have semantic protection, so information perturbation techniques cannot be applied to secure highly delicate information. Also the perturbed information do not generate very precise information mining outcomes. Secure multi-party computations centered strategy represents information are spread and not secured at each taking involving party. In inclusion, many advanced calculations are conducted depending on non-encrypted information. As an outcome, in this paper, we suggested novel methods to successfully resolve the DMED issue supposing that the secured information is contracted to a cloud. Particularly, we concentrate on the category issue considering that it is one of the most common data mining tasks. For the reason that each category strategy has their own benefits, to be tangible, this document focuses on performing the k-nearest neighbor category method over secured information in the cloud processing atmosphere.

2. Literature Survey

In this paper [4], a new realistic procedure for remote data storage space with efficient accessibility pattern comfort and correctness is introduced. A storage space customer can set up this procedure to problem secured read, writes, and inserts to a potentially curious and harmful storage space service agency, without exposing information or accessibility types. The supplier is incapable to set up any connection between subsequent accesses, or even to differentiate between a read and a write. Furthermore, the consumer is presented with strong correctness guarantees for its functions – illegal company behavior does not go unnoticed. We developed first realistic system orders of magnitude quicker than present implementations that can perform over various queries per second on 1 Tbyte+ databases with full computational comfort and correctness.

In paper [6], a completely homomorphic security plan is recommended – i.e., a plan that allows one to assess circuits over secured information without being able to decrypt. Our remedy comes in three actions. Initial, we offer a common outcome that, to build an security plan that allows assessment of irrelevant circuits, it suffices to create an security plan that can assess (slightly enhanced editions of) its own decryption circuit; we contact a plan that can assess its (augmented) decryption circuit boots trappable. Upcoming, we explain a public key security plan using perfect lattices that is almost boots trappable. Lattice-based cryptosystems generally have decryption algorithms with low circuit complexness, often covered with an inner item computation that is in NC1. Also, perfect lattices offer both preservative and multiplicative homeomorphisms (modulo a public-key perfect in a polynomial band that is showed as a lattice), as required to assess common circuits.

In this paper [8], they display how to divide data D into n items in such a way that D is quickly reconstruct able from any k items, but even finish details of k - 1 items shows definitely no details about D. This strategy allows the development of effective key management techniques for cryptographic techniques that can operate safely and effectively even when misfortunes damage 50 percent the items and protection breaches reveal all but one of the staying items.

In paper [9], collecting and handling delicate data is a challenging work. In fact, there is no common formula for building the necessary computer. In this document, they provide a provably protected and efficient general-purpose calculations system to address this issue. Our solution—SHAREMIND—is a virtual machine for privacy-preserving information processing that depends on share computing strategies.

This is a conventional way for safely analyzing features in a multi-party calculations atmosphere. The unique of our remedy is in the choice of the secret sharing plan and the design of the protocol package. We have created many realistic choices to create large-scale discuss handling possible in training. The protocol of SHAREMIND is information-theoretically protected in the honest-but-curious design with three handling members. Although the honest-but-curious design does not accept harmful members, it still provides considerably improved comfort maintenance when compared to conventional centralized databases.

In this paper [10], the problem of privacy preserving data mining is addressed. Particularly, a situation in which two parties having private databases wish to run a data mining algorithm on the partnership of their databases, without exposing any needless details. Performance is inspired by the require to both protected fortunate details and allow its use for research or other reasons. The above issue is a specific instance of protected multi-party calculations and as such, can be fixed using known general protocol. Nevertheless, data mining algorithms are typically complicated and, moreover, the feedback usually includes large details sets. The general protocol in such a case are of no realistic use and for that reason more effective methods are required. We concentrate on the issue of decision tree learning with the popular ID3 algorithm. Our protocol is significantly more effective than general alternatives and requirements both very few units of interaction and affordable data transfer bandwidth.

In paper [11], a structure for mining association rules from dealings made up of particular products where the information has been randomized to protect comfort of personal dealings. While it is possible to restore organization guidelines and protect comfort using a uncomplicated "uniform" randomization, the found guidelines can unfortunately be utilized to discover comfort breaches. Evaluate the characteristics of privacy breaches and recommend a type of randomization providers that are much more efficient than consistent randomization in restricting the breaches. Then obtain formula for an unbiased support estimator and its difference, which allow us to restore item set facilitates from randomized datasets, and display how to integrate these formula into exploration methods. Lastly, we existing trial outcomes that confirm the criteria by implementing it on actual datasets.

In paper [12], the capability of databases to arrange and work together often improves comfort issues. Data warehousing along with data mining, providing data from several resources under a single authority, improves the risk of comfort offenses. Privacy protecting data mining presents a means of dealing with this problem, especially if data mining is done in a way that doesn't reveal information beyond the outcome. This paper provides a technique for independently processing k − nn category from allocated resources without exposing any details about the resources or their data, other than that exposed by the final category outcome.

In paper [13], allocated privacy preserving data mining methods are crucial for mining several databases with a lowest information disclosure. We present a structure along with a general model as well as multi-round algorithms for exploration side to side partitioned databases using a comfort protecting k Nearest Neighbor (kNN) classifier.

In this paper [14], the issue of assisting multidimensional variety queries on secured information is researched. The issue is inspired by protected information freelancing

applications where a client may shop his/her information on a remote server in secured type and want to perform concerns using server's computational abilities. The remedy strategy is to calculate a protected listing tag of the information by implementing bucketization (a general way of information partitioning) which stops the server from studying actual principles but still allows it to check if a history meets the question predicate. Queries are evaluated in an estimated way where the came back set of information may contain some false-positives. This information then needs to be weeded out by the consumer which consists of the computational expense of our plan. In this paper create a bucketization process for responding to multidimensional variety concerns on multidimensional information. For a given bucketization plan we obtain price and disclosure-risk analytics that calculate client's computational expense and disclosure-risk respectively. Given a multidimensional dataset, its bucketization is presented as a marketing issue where the objective is to prevent disclosure while maintaining question price (client's computational overhead) below a certain user specified limit value. We provide tunable information bucketization criteria that allow the information proprietor to control the compromise between disclosure threat and price. We also research the trade off features through a comprehensive set of tests on real and artificial information. Service like Google and Amazon are shifting into the SaaS (Software as a Service) business. They turn their large facilities into a cloud-computing atmosphere and strongly hire companies to run programs on their systems. To implement protection and comfort on such a service model, we require securing the information running on the system. However, conventional protection methods that aim at offering "unbreakable" protection are often not sufficient because they do not assistance the efficiency of applications such as database queries concerns on the secured information. In this paper [15], we talk about the common issue of protected computations on a secured databases and recommend a SCONEDB (Secure Computation ON an Encrypted Database) design, which catches the efficiency and protection specifications. As a research study, we concentrate on the issue of k-nearest neighbor (kNN) calculations on secured databases. We create a new asymmetric scalar-product-preserving encryption (ASPE) that maintains a special type of scalar item. We use APSE to create two protected techniques those assistance kNN calculations on secured data; each of these techniques is proven to avoid realistic strikes of a different qualifications knowledge level, at a different expense cost. Comprehensive efficiency research is performed to assess the expense and the efficiency of the techniques.

3. Conclusion

To secure user privacy, numerous privacy-preserving category methods have been suggested over the past several years. The current methods are not appropriate to contracted database surroundings where the information exists in secured form on a third-party server. This paper suggested a novel privacy-preserving k-NN classification protocol over secured information in the cloud. Our protocol defends the privacy of the information, user's input query, and conceals the information access patterns. We also analyzed the efficiency of our protocol under various parameter configurations. Considering that helping the performance of SMINn is an important first step for helping the performance of our PPkNN protocol, we plan to examine alternative and more efficient solutions to the SMINn issue in our future work. Also, we will examine and increase our research to other category algorithms.

References

[1] P. Mell and T. Grance, "The NIST definition of cloud computing (draft)," NIST Special Publication, vol. 800, p. 145, 2011.
[2] S. 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] P. Williams, R. Sion, and B. Carbunar, "Building castles out of mud: Practical access pattern privacy and correctness on untrusted storage," in Proc. 15th ACM Conf. Compute. Common. Security, 2008, pp. 139–148.
[4] P. Williams, R. Sion, and B. Carbunar, "Building castles out of mud: Practical access pattern privacy and correctness on Untrusted storage," in Proc. 15th ACM Conf. Compute. Common. Security, 2008, pp. 139–148.
[5] C. Gentry, "Fully homomorphic encryption using ideal lattices," in Proc. 41st Annu. ACM Sympos. Theory Compute. 2009, pp. 169– 178.
[6] B. K. Samanthula, Y. Elmehdwi, and W. Jiang, "k-nearest neighbor classification over semantically secure encrypted relational data," e-print arXiv: 1403.5001, 2014.
[7] C. Gentry and S. Halevi, "Implementing gentry's fully-homomorphic encryption scheme," in Proc. 30th Annu. Int. Conf. Theory Appl. Cryptographic Techno. Adv. Cryptol., 2011, pp. 129–148.
[8] A. Shamir, "How to share a secret," Common. ACM, vol. 22, pp. 612–613, 1979.
[9] D. Bogdanov, S. Laur, and J. Williamson, "Sharemind: A framework for fast privacy-preserving computations," in Proc. 13th Eur. Symp. Res. Compute. Security: Compute. Security, 2008, pp. 192–206.
[10] Y. Lindell and B. Pinkas, "Privacy preserving data mining," in Proc. 20th Annu. Int. Cryptol. Conf. Adv. Cryptol., 2000, pp. 36–54.
[11] 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.
[12] M. Kantarcioglu and C. Clifton, "Privately computing a distributed k-nn classifier," in Proc. 8th Eur. Conf. Principles Practice Know. Discovery Databases, 2004, pp. 279–290.
[13] L. Xiong, S. Chitti, and L. Liu, "K nearest neighbor classification across multiple private databases," in Proc. 15th ACM Int. Conf. Inform. Know. Manage. 2006, pp. 840–841.
[14] B. Hore, S. Mehrotra, M. Canim, and M. Kantarcioglu, "Secure multidimensional range queries over outsourced data," VLDB J., vol. 21, no. 3, pp. 333–358, 2012.
[15] W. K. Wong, D. W.-l. Cheung, B. Kao, and N. Mamoulis, "Secure kNN computation on encrypted databases," in Proc. ACM SIGMOD Int. Conf. Manage. Data, 2009, pp. 139–152.