IJDCST @Jan-2015, Issue- V-3, I-2, SW-04
ISSN-2320-7884 (Online)
ISSN-2321-0257 (Print)

Multi Party Privacy Preserving Data Analysis Using Tuple Space Search Methods

Sowmya Budala[1], CH.Ratna Jyothi[2]

[1] Dept. of CSE, PVPSIT,Vijayawada,AP,India.

[2]Assistant Professor, PVPSIT,Vijayawada,AP,India.

ABSTRACT: Privacy and security, particularly maintaining confidentiality of data, have become a challenging issue with advances in information and communication technology. It explores which functionalities can be implemented in a way that participating parties have the incentive to provide their true private inputs upon engaging in the corresponding SMC protocols. Proposed tools from theoretical computer science in general and non cooperative computation (NCC) in particular could be used to analyze incentive issues in distributed data analysis framework. Randomly generating the associations between column values of a bucket significantly lowers data utility. So we propose to replace random grouping with more effective tuple grouping algorithms such as tuple Space Search algorithm based on hashing techniques. The efficiency of tuple grouping algorithms enables its application to handle slicing problems that were previously prohibitive due to high-dimensional data handling and sensitive attribute disclosures.

Index Terms: SMC Protocol, NCC, Tuple space search.

INTRODUCTION

PRIVACY and security, especially keeping up classifiedness of information, have turned into a testing issue with advances in data and correspondence engineering. The capacity to impart and offer information has numerous profits, and the thought of an omniscient information source conveys extraordinary quality to research and building precise information examination models. The organizations send the consortium their deals information, and key assembling expenses and times. At that point, the consortium investigates the information and measurably abridges them in a report of industry patterns, which is made accessible again to consortium parts. For this situation, it is in light of a legitimate concern for organizations to learn genuine industry patterns while uncovering their private information as meager as would be prudent. Despite the fact that SMC conventions can keep the disclosure of the private information, they don't promise that organizations send their actual deals information and other obliged data.

To counteract abuse of information, there is a late surge in laws ordering assurance of secret information, for example, the European Community protection gauges [2], U.s. health awareness laws [7], and California Sb1386. Be that as it may, this security accompanies a genuine cost through both included security consumption and punishments and expenses connected with divulgence. Secure multiparty computation (SMC) [3], [9], [10] has as of late risen as a response to this issue. Casually, if a convention meets the SMC definitions, the partaking gatherings learn just the last come about and whatever can be deduced from the last come about and their inputs. A basic case is Yao's mogul issue [9]: two tycoons, Alice and Bob, need to realize who is wealthier without disclosing their genuine riches to one another. Perceiving this, the examination group has created many SMC protocols, for applications as

differing as estimating [4], choice tree investigation [8] and barters [15] among others.

Our Contributions

In this paper, we dissect what sorts of conveyed functionalities could be actualized in a motivation perfect style. As it were, we investigate which functionalities can be actualized in a manner that taking part gatherings have the motivating force to give their actual private inputs after participating in the comparing SMC conventions. We indicate how instruments from hypothetical software engineering by and large and non-cooperative computation (NCC) [1] specifically could be utilized to break down motivating force issues in appropriated information investigation schema. This is huge on the grounds that include alteration can't be averted before the execution of any SMC-based convention.

NCC	Non-Cooperative Computation
DNCC	Deterministic NCC
PPDA	Privacy Preserving Data Analysis
SMC	Secure Multi-Party Computation
TTP	Trusted Third Party

Table 1: **Notations and Terminologies**

RELATED WORK AND BACKGROUND

We start with an outline of protection protecting dispersed information examination. At that point, we quickly talk about the idea of non-cooperative reckoning. gives regular documentations and wordings utilized widely for whatever remains of this paper. Furthermore, the terms secure and security saving are compatible from there on.

Privacy-Preserving Data Analysis

Numerous privacy-preserving information investigation conventions have been composed utilizing cryptographic techniques. Information are for the most part thought to be either vertically or evenly parceled. On account of evenly apportioned information, diverse locales gather the same set of data about distinctive substances. For instance, diverse charge card organizations may gather Visa exchanges of distinctive people. Security saving dispersed conventions have been created for evenly parceled information for some diverse information mining undertakings, for example, building choice trees, [13], mining affiliation leads, [11], and creating k-means bunches [12] and k-nn classifiers.

3.1 On Honesty in Sovereign Information Sharing.

A key inhibitor in the handy sending of sovereign data imparting has been the failure of the innovation to handle the changing of info by the members. We connected amusement theoretic ideas to the issue and characterized a multi-gathering diversion to model the circumstances. The examination of the amusement formally affirmed the instinct that the length of the members have some advantage from bamboozling, legitimate conduct can't be a harmony of the diversion. We tended to useful issues, for example, what ought to be the recurrence of checking and the punishment sum and how the evaluating gadget can be actualized as a safe system gadget that attains the coveted conclusion without getting to private information of the members.

Fast Algorithms for Mining Association Rules.

We consider the issue of finding affiliation leads between things in a substantial database of offers exchanges. We exhibit two new calculations for taking care of this issue that are in a broad sense not the same as the known calculations. Observational assessment demonstrates that these calculations

outflank the known calculations by variables going from three for little issues to more than a request of size for huge issues. We additionally indicate how the best gimmicks of the two proposed [14] calculations can be consolidated into a half breed calculation, called Apriori hybrid [14]. Scale-up investigations demonstrate that Apriori hybrid scales straightly with the quantity of exchanges. Apriorihybrid likewise has fantastic scale-up properties as for the exchange size and the quantity of things in the database.

Privacy-Preserving Decision Trees over Vertically Partitioned Data.

Protection and security concerns can avoid imparting of information, wrecking information mining tasks. Disseminated learning revelation, if done accurately, can lighten this issue. We present a summed up protection saving variation of the ID3 calculation for vertically parceled information dispersed in excess of two or more gatherings. Alongside a confirmation of security, we talk about what would be important to make the conventions totally secure. We additionally give test results, giving a first exhibit of the pragmatic unpredictability of secure multiparty reckoning based information mining.

Privacy-preserving Distributed Mining of Association Rules on Horizontally Partitioned Data.

Data Mininig can remove paramount learning from vast information accumulations – yet at times these accumulations are part among different gatherings. Security concerns may keep the parties from straightforwardly imparting the information, and a few sorts of data about the information. This paper locations secure mining of affiliation manages over on a level plane apportioned information. The routines fuse cryptographic methods to minimize the data imparted, while adding minimal overhead to the mining track.

EXISTING SYSTEM

Information are by and large thought to be either vertically or evenly apportioned. On account of on a level plane apportioned information, diverse locales gather the same set of data about distinctive substances. Traditionally a privacy preserving distributed protocols can be applicable for accessing and representing data in both horizontally and vertically in processing results efficiently. For distributed data processing on horizontally partitioned data in horizontally using some association and decision trees and for vertical partitioned data representations with k-means (Generalization, bucketization) cluster. These processes assume and achieves each parceling gathering utilize its actual information amid the circulated information mining convention execution to prevent lying about inputs using input-consistency checks.

PROBLEM STATEMENT

Numerous privacy-preserving data protocols conventions have been planned utilizing cryptographic methods. Data are for the most part thought to be either vertically or on a level plane apportioned. The same set of data about diverse elements. Case in point, distinctive Visa organizations may gather Master card exchanges of diverse people. Protection safeguarding appropriated conventions have been created for on a level plane apportioned information for some distinctive information mining errands, for example, building choice trees, mining affiliation administers, and creating k-means bunches and k-nn classifiers.

PROPOSED SYSTEM

Proposed Technique analyzes functionalities for implemented in an incentive compatible fashion. It explores which functionalities can be executed in a

manner that taking an interest gatherings have the motivator to give their actual private inputs after participating in the relating SMC conventions. Proposed tools from theoretical computer science in general and non cooperative computation (NCC) in particular could be used to analyze incentive issues in distributed data analysis framework. It is in our best interests to make the last step of the PPDA task incentive-compatible whenever possible. Improving multi party computation is the basic drawback in present developed privacy preserving technology.

Privacy-Preserving Decision Tree Classification

The point of a choice tree is to give grouping criteria in view of the qualities of an information set. At every hub of a choice tree, the information are "part" into a few subsets of information in view of the paradigm of data increase. The data increase of a part is characterized as the normal distinction in entropy between the first information set, and every information set structured by the part.

Algorithm

The Tuple Space Search calculation is a hash-based calculation. A tuple is characterized as an issue of k lengths, where k is the quantity of fields in a channel. Case in point, in a 5-field channel set, the tuple [7, 12, 8, 0, 16] methods the length of the source IP location prefix is 7, the length of the end IP location prefix is 12, the length of the convention prefix is 8 (a definite convention esteem), the length of the source port prefix is 0 (trump card or "couldn't care less"), and the length of the objective port prefix is 16 (a precise port worth). We can segment the channels in a channel set to the distinctive tuple bunches. Since the filtes in a same tuple gathering have the same tuple determination, they are shared restrictive and none of them covers with others in this tuple bunch. Presently we can perform the parcel order over all the tuples to discover the best matched channel. In the event that different tuple gatherings report matches, we resolve the best matched channel by looking at their needs. The channels in a tuple can be effortlessly composed into a hash table, where we utilize the tuple detail to concentrate the correct number of bits from each one field as the hash key. Accept there is no hash crash in the hash tables. One memory access can figure out whether there is a matched channel in a hash table so the lookup execution is just controlled by the quantity of tuple particulars. In the event that the hash tables are appropriately actualized, this calculation gives a fantastic stockpiling execution, which is direct to the channel set size.

Ordinarily, a channel has its port fields determined as extents, which can't be mapped to the tuple definition specifically. The paper propose to encode the reaches focused around the settling levels. We resolve this by changing over the extent particular to prefixes. In spite of the fact that this will bring about a stretched channel set and more tuples, it evades the need to disentangle the port extent ID, which is frequently executed by an immediate lookup table with 64k passages.

Data Analysis Tasks	DNCC		Non-DNCC	
	Vertical	Horizontal	Vertical	Horizontal
Comparison		√		
Mean & Variance			√	
Mean & Variance (assuming N is public)				√
Mean & Variance (assuming N is private)		√		
$f(v_1, v_2, \ldots, v_n) = f_1(v_1) \oplus f_2(v_2) \oplus \cdots \oplus f_n(v_n)$				√
$f(S_1, \ldots, S_n) = S_1 \cup \cdots \cup S_n$				√
$f(S_1, \ldots, S_n) = S_1 \cap \cdots \cap S_n$				√
$f(v_1, v_2, \ldots, v_n) = f_1(v_1) + f_2(v_2) + \cdots + f_n(v_n)$				√
$f((Y_1, C_1), \ldots, (Y_n, C_n)) = \frac{\sum Y_i}{\sum C_i}$		√		
Dot product of vectors containing real values				√
Dot product of non-zero binary vectors		√		
The predicate $f(Y_1, \ldots, Y_n)$: $(\sum_i^n (Y_i) \geq 0)$		√		
Association rule mining (finding frequent itemsets)	√	√		
Naive Bayes classification	√	√		
Decision tree classification	√	√		
Jaccard coefficient		√		
Dice coefficient		√		
Cosine similarity		√		
Secure similar document detection		√		

Fig : Classification of Common Data Mining Tasks

PPDA task can have many variations, and one common variation is to place a filter at the last step of the task to make the PPDA protocols more secure, it

is always possible to make the entire PPDA task satisfying the DNCC model. Therefore, when designing a PPDA protocol, it is in our best interests to make the last step of the PPDA task incentive-compatible whenever possible. As a part of future research direction, we will investigate incentive issues in other data analysis tasks, and extend the proposed theorems under the probabilistic NCC model.

Tuple pruning

The perception is for any given bundle, the quantity of one of a kind prefixes matched on a specific field is normally little. So on the off chance that we could perform the longest prefix match (LPM) first on some field and evaluate the lengths of the matched prefixes, then just a subset of tuple gatherings need to be sought. On the off chance that Lpms are performed on more extra fields, we can channel out more tuples.

Implementation

We execute the tuple pruning and tuple space seeks calculation. Note that this is just for the execution assessment, so we really did not execute every tuple amass as a hash table. Rather, in the event that we have to inquiry in a tuple bunch, we basically perform a straight hunt on the channels and we just consider this one memory access. Keeping in mind the end goal to abstain from getting to different bitmaps amid the pursuit of a prefix tree, we perform the bitwise OR operation on all the bitmaps along the looking way in the preprocess. In this way, amid the lookup, we just need to recover the bitmap at the longest matched prefix. The prefix tree is executed as the parallel tree with every hub utilizing 4 bytes.

CONCLUSION:

In this paper, we have explored what sorts of PPDA tasks is incentive compatible under the NCC model. Improving multi party computation is the basic drawback in present developed privacy preserving technology; we exhibit a hearty secure system for processing capacities that are spoken to as multivariate polynomials where gatherings hold diverse variables as private inputs, Randomly generating the associations between column values of a bucket significantly lowers data utility. So we propose to replace random grouping with more effective tuple grouping algorithms such as Tuple Space Search algorithm based on hashing techniques. The efficiency of tuple grouping algorithms enables its application to handle cutting issues that were at one time restrictive because of high-dimensional data handling and sensitive attribute disclosures.

REFFERENCES:

[1] Y. Shoham and M. Tennenholtz, "Non-Cooperative Computation: Boolean Functions with Correctness and Exclusivity," Theoretical Computer Science, vol. 343, nos. 1/2, pp. 97-113, 2005.

[2] "Directive 95/46/EC of the European Parliament and of the Council of 24 Oct. 1995 on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data," Official J. European Communities, vol. 281, pp. 31-50, Oct. 1995.

[3] O. Goldreich, S. Micali, and A. Wigderson, "How to Play Any Mental Game - A Completeness Theorem for Protocols with Honest Majority," Proc. 19th ACM Symp. the Theory of Computing, pp. 218-229, 1987.

[4] M.J. Atallah, M. Bykova, J. Li, and M. Karahan, "Private Collaborative Forecasting and Benchmarking," Proc. Second ACM Workshop Privacy in the Electronic Soc. (WPES), Oct. 2004.

[6] Doe News, www.doe.gov, Feb. 2005.

[7] "Standard for Privacy of Individually Identifiable Health Information," Fed. Register, vol. 67, no. 157, pp. 53181-53273, Aug. 2002.

[8] Y. Lindell and B. Pinkas, "Privacy Preserving Data Mining," J. Cryptology, vol. 15, no. 3, pp. 177-206, 2002.

[9] A.C. Yao, "Protocols for Secure Computation," Proc. 23rd IEEE Symp. Foundations of Computer Science, pp. 160-164, 1982.

[10] A.C. Yao, "How to Generate and Exchange Secrets," Proc. 27th IEEE Symp. Foundations of Computer Science, pp. 162-167, 1986.

[11] M. Kantarco_glu and C. Clifton, "Privacy-Preserving Distributed Mining of Association Rules on Horizontally Partitioned Data," IEEE Trans. Knowledge and Data Eng., vol. 16, no. 9, pp. 1026-1037, Sept. 2004

[12] X. Lin, C. Clifton, and M. Zhu, "Privacy Preserving Clustering with Distributed EM Mixture Modeling," Knowledge and Information Systems, vol. 8, no. 1, pp. 68-81, July 2005.

[13] Y. Lindell and B. Pinkas, "Privacy Preserving Data Mining," Proc. Int'l Conf. Advances in Cryptology (CRYPTO '00), pp. 36-54, Aug. 2000.

[14]. Rakesh Agrawal,Ramakrishnan Srikanth. Fast Algorithms for Mining Association Rules.

[15] M. Naor, B. Pinkas, and R. Sumner, "Privacy Preserving Auctions and Mechanism Design," Proc. First ACM Conf. Electronic Commerce, 1999.