A SURVEY ON CLOUD DATA SHARING AND TRAINING MECHANISM FOR DATA CATALOGING

R.NallaKumar M.E,(PhD), Assistant Professor, Department of Computer Science,
Anna University Regional Centre Coimbatore.
Dr.N.Sengottaiyan, Professor, Department of Computer Science,
Indira Institute of Engineering and Technology Chennai.
P.Preethi, P.G Scholar, Department of Computer Science,
Anna University Regional Centre, Coimbatore.

Abstract

Cloud computing enables highly scalable services to consume over the Internet. Cloud services are process based on the user request. In cloud environment users' data are usually processed remotely in unknown machines that users do not operate. User data control is reduced on data sharing under remote machines. Anomalous and normal transactions are identified using classification techniques. Neural network techniques are used for the classification process. Back-Propagation Neural network (BPN) is an effective method for learning neural networks. There are three layers in neural network, Input, middle and output layer. Collaborative BPN network learning is applied over arbitrarily partitioned data. A trusted authority (TA), the participating parties (data owner) and the cloud servers entities are involved in the privacy preserved mining process. TA is only responsible for generating and issuing encryption/decryption keys for all the other parties. Participating party is the data owner uploads the encrypted data for the learning process. Cloud server is used to compute the learning process under cloud resource environment. Each participant first encrypts their private data with the system public key and then uploads the ciphertexts to the cloud. Cloud servers execute most of the operations in the learning process over the ciphertexts. Cloud server returns the encrypted results to the participants. The participants decrypt the results by using update weights of the BPN network. Boneh, Goh and Nissim (BGN) doubly homomorphic encryption algorithm is used to secure the private data values. Data splitting mechanism is used to protect the intermediate data during the learning process. Random sharing algorithm is applied to randomly split the data without decrypting the actual value. Secure scalar product and addition operations are used in the encryption and decryption process.

The collaborative learning process is handled without the Trusted Authority (TA). Key generation and issue operations are carried out in a distributed manner. Cloud server is enhanced to verify the user and data level details. Privacy preserved BPN learning process is tuned with cloud resource allocation process.

1. Introduction

Cloud computing is the recent trending which was used to provide the services over the internet as well as the actual cloud infrastructure. Cloud provides the services based on the broadband and wireless internetworking, falling storage costs, and continuous improvements in Internet computing software. Cloud-service clients will be able to request more capacity at peak demand, reduce costs, experiment with new services, and remove unneeded capacity, whereas service providers will increase utilization and also allow the larger investments in software and hardware. It also includes the computing infrastructures and services include virtualization, service-oriented software, grid computing technologies, management of huge facilities, and power efficiency. It provides services in the form of infrastructure-as-a-service (IaaS), platform-as-a-service (PaaS), or soft-

ISSN: 2278 – 7798

ware-as-a-service (SaaS) and sell value-added services to users. Data mining is the process of analysising the data into larger amout of data, in cloud computing, the user request services has to analyze the different volume of the dataset and find the particular data, and also provide the service to the user.

2. Related Work

Several privacy preserving BPN network learning schemes have been proposed recently. Schlitter [9] it introduces a privacy preserving BPN network learning which work with two or more parties is without disclosing the particular private data set. But the solution is proposed only for horizontal partitioned data. In BPN cannot protect the intermediate results, which may also have sensitive data, at the time of the learning process. Chen and Zhong [6] propose a privacy preserving BPN network learning algorithm for two-party scenarios. This algorithm provides the secure production between the intermidate results. However, it supports only vertically partitioned data. To solve this problem, Bansal et. Al [4]. Enhanced this scheme and proposed a solution for arbitrarily partitioned data. Nevertheless, this enhanced scheme was proposed for the two-party scenario. Directly increase them to the multiparty setting will introduce a computation/communication complexity quadratic in the number of participants. In practical implementation, such a complexity represents a high cost on each party considering the already expensive operations on the underlying groups such as elliptic curves. For two-party scenario though it supports arbitrarily partitioned data set. None of existing schemes have solved all these challenges at the same time. There still low efficient and scalable solutions that supports collaborative BPN network learning with privacy preservation in the multiparty setting and allows arbitrarily partitioned data sets.

3. Privacy Preserved Data Mining in Clouds

Back-Propagation [8] is an effective method in the neural network which used in much application; it has to be propagating their weights in each step for further processing. The accuracy of the result will be affect the large amount of data set and it has to be used in the learning process. It has compared with only the local data and it improve the learning accuracy by incorporating more data sets into the learning process [11]: the parties are worked their data sets and also they involved in the other parties data set. The cloud computing infrastructure is used to achieve the joint learning process which was achieved by using the internet [12].

The advantage of the particular neural network processing need to process the privacy data set with high secure prevention. The security between the data set's are used to classify the different data's for the particular parties information [2], It also deals the legal concerns according to the privacy rules such as Health Insurance Probability and Accountability Act (HIPAA) [1].The Internet-wide collaborative learning, it is imperative to provide a solution that allows the participants, who have low mutual trust, to conduct neural network learning jointly without disclosing their respective private data sets. The solution will be efficient and scalable enough to support a random number of participants, each having the random partitioned data sets. Secure multiparty computation (SMC) can be used to solve problems . Due to the high circuit size it has the high computational complexity. To provide practical solutions for privacy preserving Back-Propagation neural (BPN) network learning [10], three main challenges need to be follow: 1) Product the BPN result and intermediate result of the private dataset. 2) The solution and the computational cost should be affordable 3) for collaborative training, the training data sets may be owned by different parties and partitioned in arbitrary ways rather than a single way of partition.

In this work, address this problem by incorporating the computing power of the cloud [7]. The main idea of our scheme can be summarized as follows: Each parties should encrypts their private data by using the private key after that the result will be upload by using the public key the intermediate and final result will be managed in secure by using the Back Propagation Neural network [3]. To support these operations over ciphertexts, we adopt the Boneh, Goh, and Nissim (BGN) [5] doubly homomorphic encryption algorithm and tailor it to split the decryption capability among multiple participants for collusion-resistance decryption. The decryption will be done in the smaller way and also it has to be increased by using the proposed scheme, To protect the intermediate data during the learning process, we introduce a novel random sharing algorithm to randomly split the data without decrypting the actual value. It provides the efficiently to process

the particular data and accuracy of the key generation.

4. Problem Statement

Collaborative BPN network learning is applied over randomly partitioned data. A trusted authority (TA), the participating parties (data owner) and the cloud servers entities are involved

in the privacy preserved mining process. Encryption and decryption key are generated and issues to all parties by TA. Parties are behaving like the data owner. Data owner only uploads the encrypted data for the learning process. In the cloud environment Cloud server compute the learning process. Each participant first encrypts their private data using the system public key and then uploads the ciphertexts to the cloud. A cloud server executes most of the operations in the learning process over the ciphertexts. Cloud server returns the encrypted results to the participants. The participants are decrypting the results with which they modify their respective weights for the BPN network. Boneh, Goh and Nissim (BGN) doubly homomorphic encryption algorithm is used to secure the private data values. Data splitting mechanism is used to protect the intermediate data during the learning process. Random sharing algorithm is applied to randomly split the data without decrypting the actual value. Secure scalar product and addition operations are used in the encryption and decryption process. The following problems are identified from the existing system.

- Centralized key distribution model
- Malicious party attacks are not handled
- Noisy data upload is not controlled
- Resource allocation and data distribution is not optimized

5. Privacy Preserving Multiparty Neural Network Learning

In this paper, several parties jointly work for training the BPN without considering the private data. The input data sets owned by the parties which can be randomly partitioned. The computational and communicational costs on each party will be efficient and scalable. If we are use the 3 layer neural network we can easily convert into the multilayer network. The learning data set for the neural network, which has N samples, is randomly partitioned into $Z(Z \geq 2)$ subsets. Each party Ps holds $x_1^m, x_2^m, ..., x_a^m$ and has

$x_{11}^m + x_{12}^m + ... + x_{1Z}^m = x_1^m$

……

$x_{a1}^m + x_{a12}^m + ... + x_{a1Z}^m = x_{a.1}^m$

Each attribute in sample $\langle x_1^m, x_2^m, ..., x_a^m \rangle, 1 \leq m \leq N,$ is possessed by only one party—if Ps possesses $x_k^m, 1 \leq k \leq a,$ then $x_{ks}^m = x_k^m$ otherwise $x_{ks}^m = 0$. In this paper, w_{jk}^h denotes the weight used to connect

the input layer node and the hidden layer node; w_{ij}^o denotes the weight used to connect the hidden layer node and the output layer node , where the weight between each layer should be one. For collaborative learning, all the parties are to jointly execute the operations defined in the Feed Forward stage and the Back-Propagation stage as shown in Algorithm 1. During each learning stage, except for the final learned network, neither the input data of each party nor the intermediate results generated can be revealed to anybody other than TA.

Algorithm 1: Back-Propagation Neural Network Learning Algorithm

To reach the goals, the main idea of our proposed scheme is to implement a privacy preserving equivalence for each step of the original BPN network learning algorithm which was described in Algorithm 1. It is different from the original BPN network learning algorithm, our proposed scheme lets each party encrypt their input data set and upload the encrypted data to the cloud, allowing the cloud servers to perform most of the operations, i.e., additions and scalar products. To support these operations over ciphertexts, we adopt and tailor the BGN "doubly homomorphic" encryption for data encryption. but BGN algorithm just supports one step multiplication over cipher text, the intermediate results are first securely decrypted and then encrypted to support consecutive multiplication operations. For privacy perspective, the decrypted results known to each party but cannot be the actual intermediate values, for example, the values of the hidden layer. For this purpose, we design a secret sharing algorithm that allows the parties to decrypt only the random shares of the intermediate values. The collaborative parties only known to the intermediate values. Data privacy is monitor continuously. The overall algorithm is described in Algorithm 2, which is the privacy preserving equivalence of Algorithm 1. To support the operations, we propose three other cloud-based algorithms for secure scalar product and addition, secure random sharing, and sigmoid

ISSN: 2278 – 7798

function approximation process. After the entire process of the privacy preserving learning, all the parties jointly establish a neural network representing the whole data set without disclosing any private data to each other.

Algorithm 2: Privacy Preserving Multi-Party BPN Network Learning Algorithm

In this section, we introduce our cloud-based privacy preserving multiparty BPN network learning algorithm randomly partitioned data. As we described in Algorithm 2, The parties are generated and assign the weights in random w^{h}_{jks} s and w^{o}_{ijs} to each P_s and it has the max number of learning iteration max, the learning rate, error threshold and target value t_i of each output layer node at the learning. In the Feed Forward Stage, all the parties are agreed with the sigmoid based function as the activation function which is provide the accuracy requirement and also get the random shares h_{js} for value of hidden layer node and o_{is} for value of output layer node. After the Feed Forward Stage, the parties are worked together and check whether they have any threshold values. If they didn't have any error values, they proceed to the Back-Propagation Stage, which aims at changing the weights so as to achieve correct weights in the neural network. For the weights of each output layer node w^{o}_{ij} each Ps obtains random shares of the changes in weights, denoted as Δw^{o}_{ijs} for Δw^{o}_{ij}. To compute the changes in the weight Δw^{o}_{ij} of each hidden layer node (cf. (2)), Then, each P_s obtains the random shares of each item. Finally, Ps updates its own weights with its shares and the learning rate η.

6. Privacy Preserved Data Cataloging in Clouds

The collaborative learning process is handled without the Trusted Authority (TA). Key generation and issue operations are performed in a distributed manner. Cloud server is used to verify the user and data level details. Privacy preserved BPN learning process is used to allocate resource in the cloud. The cloud data analysis process is used to design the resources training process. Key values are generated and share using the key aggregation process is used to generate and share the key values. Training is performed by using the cloud server with privacy. The system is divided into six major elements. They are cloud server, trusted authority, data provider, upload process, training process and data classification.

The cloud server module is providing resources to the clients. Trusted authority module manages key distribution process. Data provider is designed to share the data in the cloud. Data encryption and upload process are managed under upload process module. Neural network learning process is used the training process module. Data classification module is designed to classify the client data values.

6.1. Cloud Server

The cloud server manages the user and resource details in the cloud server. The cloud server collects resources from different resource providers and also maintains the user authentication process. Resource scheduling is used to allocate computational resources to the training process.

6.2. Trusted Authority

Trusted Authority (TA) application is used for key management process. Public key and private key values are generated in the trusted authority. Key values are issued to the data providers. Cloud server verifies the user account details.

6.3. Data Provider

Shared data values are maintained by the Data provider. Before processing those data's Noise removal process is applied on the data values. Several data providers are participated in the data classification process. Data providers are referred as data owners.

6.4. Upload Process

Data providers upload the Shared data values into the cloud server. The sensitive attributes are secured by using the Encryption process, here Boneh, Goh and Nissim (BGN) doubly homomorphic algorithm is used for the encryption process. There are two types of key generation models are available which was used by the data provider. They are Trusted Authority (TA) based key model and Distributed key model. Trusted Authority generates the key values and also issues the key to the data provider. Aggregation based key generation mechanism is used in distributed key model. And finally Labeled transaction data values are collected and updated in the cloud server.

6.5. Training Process

There are several algorithm used for the training process which are, Resource scheduling process is used to train the scheme. Back Propagation Neural

network (BPN) algorithm is also used for the training process, which have associated with weight for all nodes And Random sharing algorithm is used in the data splitting process which is used for secure the intermediate data values. Training process results are transmitting to the data provider.

6.6. Data Classification

From the cloud server the trained data's are collected and also decrypts those data with help of data provider. Data encryption/decryption processes are achieved by using secure scalar product and addition mechanism. Test data values are compared to the trained data values for the class assignment process.

7. Conclusion

Multi party based collaborative leaning method is used for privacy preserved Back Propagation Neural network (BPN). Data privacy is achieved with encrypted data learning process using cloud resources. Privacy preserved BPN learning process is enhanced without using the Trusted Authority for key management process. The system also handles the malicious party attacks in the learning process. Classification accuracy will be increased by using Collaborative learning model. The system reduces the computational and communication cost for privacy preserved data classification. Data privacy is improved and also Key generation and load is minimized in the aggregation based cryptographic model.

REFERENCES

[1] "The Health Insurance Portability and Accountability Act of Privacy and Security Rules," http://www.hhs.gov/ocr/privacy, 2013.

[2] "National Standards to Protect the Privacy of Personal Health Information," http://www.hhs.gov/ocr/hipaa/finalreg.html, 2013.

[3] Ron C. Chiang and H. Howie Huang, "TRACON- Interference-Aware Scheduling for Data-Intensive Applications in Virtualized Environments" IEEE Transactions On Parallel And Distributed Systems, Vol. 25, No. 5, May 2014

[4] A. Bansal, T. Chen, and S. Zhong, "Privacy Preserving Back-Propagation Neural Network Learning over Arbitrarily Partitioned Data," Neural Computing Applications, vol. 20, no. 1, Feb. 2011.

[5] D. Boneh, E.-J. Goh, and K. Nissim, "Evaluating 2-DNF Formulas on Ciphertexts," Proc. Second Int'l Conf. Theory of Cryptography (TCC '05), 2005.

[6] T. Chen and S. Zhong, "Privacy-Preserving Backpropagation Neural Network Learning," IEEE Trans. Neural Network, vol. 20, no. 10, Oct. 2009.

[7] Qin Liu, Chiu C. Tan, Jie Wu and Guojun Wang, "Towards Differential Query Services in Cost-Efficient Clouds" IEEE Transactions On Parallel And Distributed Systems, Vol. 25, No. 6, June 2014

[8] D.E. Rumelhart and R.J. Williams, "Learning Internal Representations by Error Propagation," Parallel Distributed Processing: Explorations in the Microstructure of Cognition, 1986.

[9] N. Schlitter, "A Protocol for Privacy Preserving Neural Network Learning on Horizontal Partitioned Data," Proc. Privacy Statistics in Databases, 2008.

[10] Huiqi Xu, and Keke Chen, "Building Confidential and Efficient Query Services in the Cloud with RASP Data Perturbation" IEEE Transactions On Knowledge And Data Engineering, Vol. 26, No. 2, February 2014

[11] K. Flouri and P. Tsakalides, "Training a SVM Based Classifier in Distributed Sensor Networks," Proc. 14th European Signal Processing Conf., 2006.

[12] Grossman and Y. Gu, "Data Mining Using High Performance Data Clouds: Experimental Studies Using Sector and Sphere," Proc. 14th ACM Int'l Conf. Knowledge Discovery and Data Mining, 2008.

ISSN: 2278 – 7798