IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 17, Issue 2, Ver. 1 (Mar – Apr. 2015), PP 75-79
www.iosrjournals.org

Back-Propagation Neural Network Learning with Preserved Privacy using Cloud Computing

[1]Mrs. Valli N, [2]Mr. H.Anwar Basha
[1]PG Scholar S.A. Engineering College Chennai, India
[2]Assistant Professor, CSE S.A.Engineering College Chennai, India

***Abstract*:** *Cloud computing facilitates the owners to share data. Multiple parties join the learning process by conducting joint Back propagation neural network algorithm on the union of their respective data sets. During the learning process none of the party wants to disclose her/his private data to others. The limitations of the existing schemes are learning process for only two parties or the few ways in which the data is arbitrarily partitioned. The proposed solution allows two or more parties, each with an arbitrarily partitioned data set, to jointly conduct the learning. The solution is provided by the magnificent power of cloud computing. In the proposed scheme, each owner encrypts his/her private dataset locally through AES cryptography and uploads the cipher texts into the cloud. The cloud then executes most of the operations over cipher texts via BGN homomorphic algorithm. The cloud is unaware of the original dataset. The Back propagation learning takes place and the owners are benefited through collaborative learning. Thus the scalability of the learning process is improved and the privacy of the data is ensured.*

Keywords: *Privacy preserving, learning, neural network, back-propagation, cloud computing, computation outsource, homomorphic encryption*

I. Introduction

The multiparty learning in Neural Networks through Back Propagation Algorithm has always been a challenging scenario. With cloud computing infrastructure, we can solve this problem.

Cloud computing involves deployment of groups of remote servers and software networks and other resources that allow centralized access to storage and computer services or resources. Clouds can be classified as public, private or hybrid based on their deployment model. Public cloud is accessible by all and private cloud is protected within the boundary of an organization say corporate through firewalls whereas hybrid cloud is a combination of both public and private cloud. With cloud there is a shift from CAPEX model to OPEX model in most of the organizations.

With the increased use of cloud computing, there is a new paradigm emerging – cloud computing security. Cloud computing security is the set of policies, tools and regulations to ensure that data is secure in cloud computing atmosphere. To make sure the data is secure and privacy of the dataset is maintained, we make use of AES (Advanced Encryption Standard) and RSA (Rivest, Shamir and Adleman) algorithms.

The back propagation neural network algorithm is the most widely used algorithm for learning in neural network. It is the workhorse in neural networks. Neural networks are composed of artificial neurons that simulate the human brain. The neural network is trained with numerous examples and it learns from these examples. Back propagation is one such scheme where the supervised learning takes place by examples. It makes use of activation function and makes the network learn by examples by adjusting the weights of the edges in the neural network. The activation function can be a step function or sigmoid function. Usually sigmoid function is used.

II. System Model

The system consists of three major components:

- Trust Agent
- Cloud
- Owners

The trust agent is a government agency that activates the owner's account. The owner registers with the trust agent and uploads the data only when the owners's account is active. The uploaded data is encrypted using AES cryptography. The ciphertext is manipulated by the cloud to perform additive and multiplicative operations using \bgn homomorphic algorithm. The learning takes place using back propagation algorithm. It is then collaboratively shared.

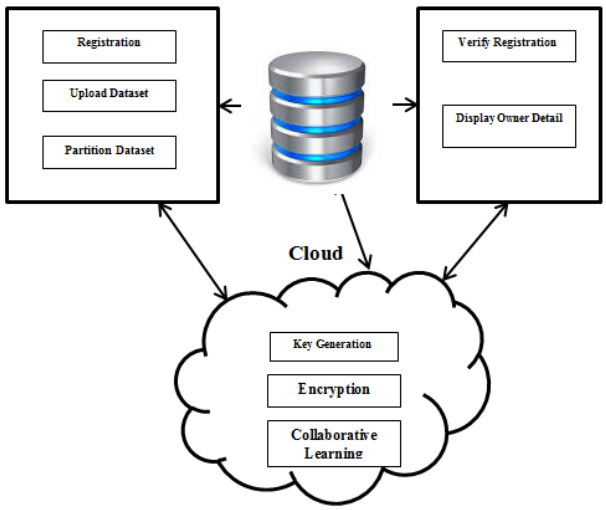

Fig1 shows the overall description of Collaborative Learning

III. Security Model

The Trust Agent is usually a government agency that authenticates the owner by verifying the owner details. In the proposed solution, the cloud generates the keys and distributes to the owners. The security of the system is improved as the cloud does not know the data of the owners.

IV. System Description

A. Authenticating the Owner

It is performed by Trust Agent. Owner's account is activated after the authentication of the owner by the Trust Agent. The owner registers with the trust agent and the trust agent confirms his registration. Only after the confirmation of registration the owner's account is made active.

B. Uploading the data

The owner uploads the data in cloud. The owner can upload as many times as required. Owner owns a private data set and wants to perform collaborative learning. For this he loads the encrypted data in cloud. Each participating party s, denoted as Ps, owns a private data set and wants to perform collaborative BPN network learning with all other participating parties. That is, they will collaboratively conduct learning over the arbitrarily partitioned data set, which is private and cannot be disclosed during the whole learning process. We assume that each participating party stays online with broadband access to the cloud and is equipped with one or several contemporary computers, which can work in parallel if there is more than one.

C. Arbitrarily partitioning data

Combination of horizontal and vertical partitioning is done. The rows and columns are not partitioned in any particular order. It is done arbitrarily. We consider arbitrary partitioning of data between two parties. In arbitrary partitioning of data between two parties, there is no specific order of how the data is divided between two parties. Combined data of two parties can be seen as a database. Suppose D is the data to be partitioned then D_i is the partition of data D such that $D_1 \cap D_2 \cap D_3 \cap \ldots\ldots\ldots\ldots D_n = \emptyset$
and $D1 \cap D_2 \cap D_3 \cap \ldots\ldots\ldots\ldots D_n = D$

D. Encrypting the data

AES (Advanced Encryption System) is used to encrypt the data and the encrypted data is uploaded in cloud. This is the fundamental encryption that takes place. This produces the cipher text upon which BGN algorithm is applied.

E. Homomorphically encrypting data

BGN (Boneh, Goh and Nissim algorithm is used to encrypt the cipher text and the encrypted cipher text is available in cloud. It is called doubly homomorphic encryption. Homomorphic encryption enables operations on plaintexts to be performed on their respective cipher texts without disclosing the plaintexts. Most existing homomorphic encryption schemes only support single operation - either addition or multiplication. It introduced a public-key 'doubly homomorphic encryption scheme (called 'BGN' for short), which simultaneously supports one multiplication and unlimited number of addition operations. Therefore, given

cipher texts C(m1), C(m2), …. C(mi) and C(ˆ m1), C(ˆm2),….C(ˆmi), one can compute C(m1mˆ1+m2mˆ2+ ⋅ ⋅ ⋅ +mimˆi) without knowing the plaintext, where C() is the cipher text of message mi or mˆi, encrypted by the system's public key.

Input: Ciphertext of ϵ
Output: Shares of ϵ: ϵ_s for $P_s, 1 \leq s \leq Z$
begin
  for $s = 1, 2 \cdots, Z$ **do**
    Choose $L_s \xleftarrow{R} (0, u)$
    $C(L_s) = g_1^{L_s} h_1^{r_s q_2}$
    //where u is the upper bound of ϵ
  //Cloud Calculates:
  $C(sumL) = \prod_{s=1}^{Z} C(L_s)$
  case $1.\epsilon > \sum_{s=1}^{Z} L_s$
    $C(\hat{L}) = C(\epsilon) * C(sumL)^{-1}$
  case $2.\epsilon < \sum_{s=1}^{Z} L_s$
    $C(\hat{L}) = C(sumL) * C(\epsilon)^{-1}$
  Decrypt $C(\hat{L})$ with Algorithm 3 and send $\hat{L}$ to P_1
  //Output Shares:
  $\epsilon_1 = L_1 + \hat{L}$ (Case 1) or $\epsilon_1 = L_1 - \hat{L}$ (Case 2)
  for $i = 2, 3 \cdots, Z$ **do**
    $\epsilon_s = L_s$
  end

Fig 3: Scalar Product and Sum

Fig 3[12] illustrates the calculation of scalar product and Sum. The cloud calculates first the product of cipher text and then it calculates the sum of the product obtained in the previous step. The C(m1) is multiplied with C(m^1) and then added with C(m2) multiplied with C(m^2).

F. Back Propagation Learning

Back propagation, otherwise known as "backward propagation of errors", is a common method of learning in artificial neural networks. From a desired output, the network learns from many inputs Back-Propagation neural network learning algorithm is mainly composed of two stages: feed forward and error back − propagation. In Feed Forward the network is trained with many inputs and the output is generated as a sigmoid function of the inputs. The actual output and the expected output is compared and then the difference is calculated as error. The error is then propagated backwards in the network and the internal weights adjusted so that the difference between the actual and expected weight diminishes.
The input and output [10] of the neuron, i, (except for the input layer) in a multilayer perceptron mode, according to the Back Propagation algorithm are:

Input $x_i = \sum w_{ij} o_j + b_i$ (1)
Output $o_i = f(x_i)$ (2)

Where W_{ij} is the weight of the connection from neuron i to node j, b_i is the numerical value and f is the activation function.

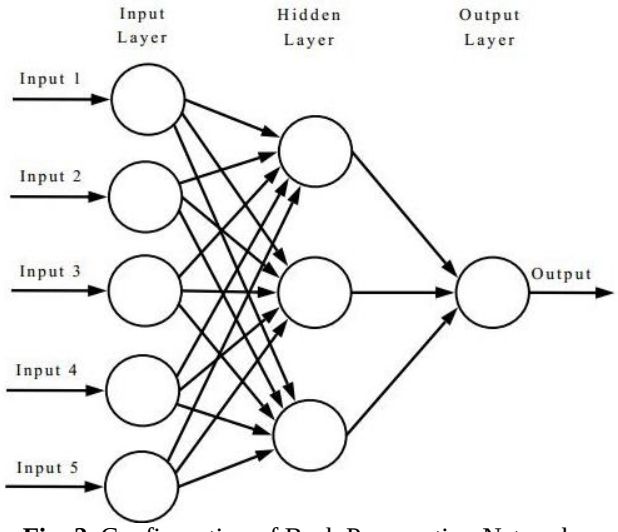

Fig. 2. Configuration of Back Propagation Network

Fig.1[10] shows is a configuration for a three layered Back Propagation network. The inputs are feeded to the input layer. The weights are assigned to the edges connecting these layers and the output is calculated with the help of sigmoid function. The difference between the actual and the desired output is calculated and this difference is observed as error. A fraction of the error is back propagated to the hidden layer and input layer and then the sigmoid function again calculates the actual output. The difference is noted and this process is repeated until the error becomes negligible.
The algorithm can be decomposed in the following four steps:
i) Feed-forward computation
ii) Back propagation to the output layer
iii) Back propagation to the hidden layer
iv) Weight updates
The algorithm is stopped when the value of the error function has become sufficiently small.

The algorithm for back propagation[10] is described below.

[13] Input: N input sample vectors V, 1<i<N with a
dimensions iteration tales place at the rate of n, target value ti, sigmoid function f(x) =1/1+e(-x)

Output: Network with final weights. wjk

Randomly initialize the wjk values
For iteration=1, 2….iteration
Do for sample=1, 2…..N.
//Feed forward stage: for j=1, 2…b do
hj=f ak=1(x(k)*wjk)
For i=1, 2….c do
oi=f ($a(j$=1hj*wjk)
If Error=12 ci=1(ti-0i)*hj
else// it works with Back propagation stage,
$\boldsymbol{\sigma}$wij=-(ti-oi)*hj $\sigma wjk=-hj\ 1-hj\ xk$
ci=1[(ti-oi)*wij)]
wij=wij-µwjk
wjk=whj-µwjk
else //learning finish break.
$\sigma wij=-(ti-oi)hj$
$\sigma wjk=-hj\ 1-hj\ xk$
ci=1[$ti-oi*wjk$)]

G. Collaborative learning

Collaborative learning is a process in which many people gather with their data for learning simultaneously. There are two factors impacting the multiparty learning - the number of participating parties and the size of dataset .As the number of participants increases, the operations for each party to share intermediate results and decrypt the final learning result increases.

V. Performance Evaluation

In this section, the several issues related to performance in back propagation algorithm using cloud is discussed. First the accuracy of the output depends on the number of layers in the network. There should be at least three layers in the back propagation configuration network. The more the number of layers, the accuracy of the output increases at the cost of tradeoff with performance. [11] Choosing number of nodes for each layer will depend on problem Neural Network is trying to solve, types of data network is dealing with, quality of data and some other parameters. Number of input and output nodes depends on training set in hand. If there are too many nodes in hidden layer, number of possible computations that algorithm has to deal with increases. Picking just few nodes in hidden layer can prevent the algorithm of its learning ability. Right balance needs to be picked.

Second the security algorithms namely AES and BGN should be applied to only small datasets. So the datasets are made very relevant and apt.

Third the number of iterations in back propagation learning increases the accuracy of the output. The tradeoff with computation time has to be analyzed.

VI. Conclusion

The cloud does the key generation and distributes the keys to the owner. The owner encrypts the data with the public key of the owner using AES cryptography. The encrypted ciphertext is uploaded in the cloud. The cloud does multiplication and addition operation on the ciphertext and the collaborative learning takes place without disclosing the private data of the owners.

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]. M. Abramowitz and I.A. Stegun, Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematics Tables, DoverBooks on Mathematics. Dover, 1964.
[4]. A. Bansal, T. Chen, and S. Zhong, "Privacy Preserving Back-Propagation Neural Network Learning over Arbitrarily Parti-tioned Data," Neural Computing Applications, vol. 20, no. 1, pp. 143150,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), pp. 325-341, 2005.
[6]. T. Chen and S. Zhong, "Privacy-Preserving Backpropagation Neural Network Learning," IEEE Trans. Neural Network, vol. 20, no. 10, pp. 1554-1564, Oct. 2009.
[7]. L. Cun, B. Boser, J.S. Denker, D. Henderson, R.E. Howard, W.Hubbard, and L.D. Jackel, "Handwritten Digit Recognition with a Back-Propagation Network," Proc. Advances in Neural Information Processing Systems, pp. 396-404, 1990.
[8]. S.D.C. di Vimercati, S. Foresti, S. Jajodia, S. Paraboschi, and P.Samarati, "Over-Encryption: Management of Access Control Evolution on Outsourced Data," Proc. 33rd Int'l Conf. Very Large Data Bases (VLDB '07), pp. 123-134, 2007.
[9]. T. El Gamal, "A Public Key Cryptosystem and a Signature Scheme Based on Discrete Logarithms," Proc. Advances in Cryptology (CRYPTO '85), pp. 10-18, 1985.
[10]. http://www.m-hikari.com/ces/ces2011/ces1-4-2011/mustafaCES1-4-2011.pdf
[11]. http://www.dataminingmasters.com/uploads/studentProjects/NeuralNetworks.pdf
[12]. Jiawei Yuan, Student Member, IEEE, and Schucheng Yu, Member, IEEE Privacy Preserving Back Propagation Neural Network learning made Practical with Cloud Computing, "IEEE transactions on parallel and distributed systems, vol no. 1, January 2014.
[13]. http://www.ijera.com/special_issue/Humming%20Bird_March_2014/Version%20%203/CA0105.pdf