Application of Homomorphic Encryption in Machine Learning

Sagarika Behera

Department of Computer Sc. and Engineering
CMR Institute of Technology
Bengaluru, India
sagarika.b@cmrit.ac.in

Jhansi Rani Prathuri

Department of Computer Sc. and Engineering
CMR Institute of Technology
Bengaluru, India
jhansirani.p@gmail.com

2020 2nd PhD Colloquium on Ethically Driven Innovation and Technology for Society (PhD EDITS) | 978-1-7281-9608-4/20/$31.00 ©2020 IEEE | DOI: 10.1109/PHDEDITS51180.2020.9315305

***Abstract*— The linear regression is a machine learning algorithm used for prediction. But if the input data is in plaintext form then there is a high probability that the sensitive information will get leaked. To overcome this, here we are proposing a method where the input data is encrypted using Homomorphic encryption. The machine learning algorithm can be used on this encrypted data for prediction while maintaining the privacy and secrecy of the sensitive data. The output from this model will be an encrypted result. This encrypted result will be decrypted using a Homomorphic decryption technique to get the plain text. To determine the accuracy of our result, we will compare it with the result obtained after applying the linear regression algorithm on the plain text.**

Keywords*— *Homomorphic encryption, Machine learning, Linear regression

I. INTRODUCTION

Homomorphic Encryption (HE) is a form of encryption technique with an additional computing capability which can perform computation over encrypted data without accessing the secret key. The output of the computation remains encrypted. The popularity of this method is increasing since it will keep the privacy and secrecy of the sensitive data intact. There are different types of homomorphic encryption methods depending on the type of operation it can perform on the encrypted data such as Partially Homomorphic Encryption (PHE), Somewhat Homomorphic Encryption (SHE), Fully Homomorphic Encryption (FHE). The homomorphic encryption method consists of four steps: key generation, encryption, decryption and evaluation.

II. LITERATURE SURVEY

Homomorphic encryption algorithms detailed survey explained in [1]. In paper [2] authors proposed a symmetric verifiable FHE scheme and claimed that it is noise free. A new protocol named as privacy-preserving linear regression (PPLR) is proposed in [3] on horizontally partitioned data. Authors in paper [4] proposed a method where they are combining the contents of different databases and applying linear regression, ridge regression on that combined data. To achieve the security of data, they have applied the Homomorphic encryption technique. They have found that their model is practically possible for medium size databases not for large database.

NOMENCLATURE

TABLE I. SYMBOLS AND DESCRIPTIONS

SYMBOL	**DESCRIPTION**
N_1 & N_2	Two large distinct prime numbers
P_1 & P_2	Two plain text messages, where P_1 & P_2 Є Z_n^*
S_1 & S_2	Two cipher texts, where S_1 & S_2 Є Z_{n2}^*
g	Random integer, where g Є Z_{n2}^*
N_3	$N_3 = N_1 * N_2$, it divides the order of g
r1,r2	Random numbers, where r1,r2 Є Z_n^*
pub_k	Public key
sec_k	Secret key
enc(P_1)& enc(P_2)	Encryption of message P_1 and P_2

III. PROPOSED METHOD

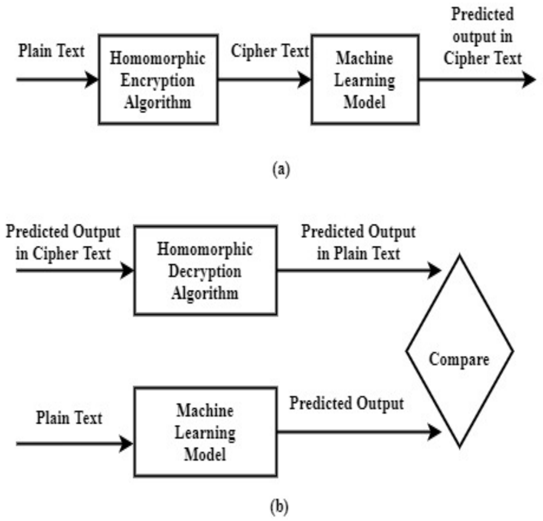
In Fig. 1(a) the plain text is encrypted using a Homomorphic encryption method. The encrypted data are given as input to the machine learning (ML) model. In the ML model, there are two phases. It takes the encrypted data as input in the training process, and gives the output in encrypted form. To predict the outcome, the prediction process takes place on the encrypted data. The linear regression method is used to predict the result. The predicted result is in encrypted form. In Fig. 1(b) a Homomorphic decryption algorithm is applied to this encrypted result to decrypt it.

The decrypted result will be compared with the predicted result which we got after applying a machine learning algorithm on the plain text.

IV. RESULT AND DISCUSSION

Here we have implemented the Pailler cryptosystem. Pailler cryptosystem follows additive Homomorphic encryption property. The result is shown below. Here we have randomly generated two large distinct prime numbers N1 and N2 of size 181 digits. P1 and P2 are two plain text messages which are encrypted. Addition and multiplication operation performed on the encrypted message. The result is decrypted and compared with the original sum and product of two messages. We found that both are the same.

Since Pailler cryptosystem is an asymmetric encryption cryptosystem, it has a public and private key pair (pubk, seck). Where public key for encryption is, pubk = (N3,g) and private key for decryption is, seck = lcm(N1-1,N2-1).



Plain Text
Homomorphic Encryption Algorithm
Cipher Text
Machine Learning Model
Predicted output in Cipher Text
(a)
Predicted Output in Cipher Text
Homomorphic Decryption Algorithm
Predicted Output in Plain Text
Compare
Plain Text
Machine Learning Model
Predicted Output
(b)

Figure 1: Machine learning model on Homomorphic encrypted data (a) Machine learning algorithm applied on cipher text and predicted output is in cipher text form and (b) Comparison of predicted output of Machine learning model on plaintext with predicted output of Machine learning model on ciphertext

The additive Homomorphic property of the Pailler cryptosystem is shown below. Pailler cryptosystem has the property that :

enc (P1).enc (P2) mod N3 = enc(P1+P2) mod N3.

After applying encryption operation on two plain texts P1 and P2, two cipher texts S1 and S2 are generated.

Where S1 and S2 are :

$$S_1 = g^{P_1}.r1^{N_3}\, mod\; {N_3}^2$$

$$S_2 = g^{P_2}.r2^{N_3}\, mod\; {N_3}^2$$

$$S_1.S_2 = g^{P_1}.r1^{N_3}.g^{P_2}.r2^{N_3}\, mod\; {N_3}^2$$

$$= g^{P_1+P_2}(r1r2)^{N_3} mod\; {N_3}^2$$

So it satisfies the property of additive Homomorphic encryption.

N$_1$value: 2473184644907985314134520166626191608169210259481401241520596150820670390120306645374855753137357631521187198060016627511956958123580168216761790126852145033213411043530800277770447

N$_2$value:

2404722932691950547855084204894304557776612669102036300588191719930015811054210912763914162222737624869026118919932515570357304292367943114019399880350855507944873334125057142003187

P1 decrypted value: 30, P2 decrypted value: 60

Original Sum: 90, Decrypted Sum: 90

Original Product: 1800, Decrypted Product: 1800

Here we found that if we will apply addition and multiplication on cipher text and the decrypted result is same as the result we got from the plain text after applying the same operation.

Now we are going to show how the linear regression can be secured using Pailler Homomorphic encryption technique.

The linear equation for Linear Regression is:

$$Y = A.X + b$$

The regression coefficient vector is :

$$\hat{A} = (X^T X)^{-1} X^T Y$$

Slope b is:

$$\hat{b} = \bar{Y} - \hat{A}\,\bar{X}$$

Here we have taken sales data as input. Where X is sales executive id and Y is sales value. We have predicted the sales value given the sales executive id. Using the key pair (pubk, seck), Y and X values are encrypted.

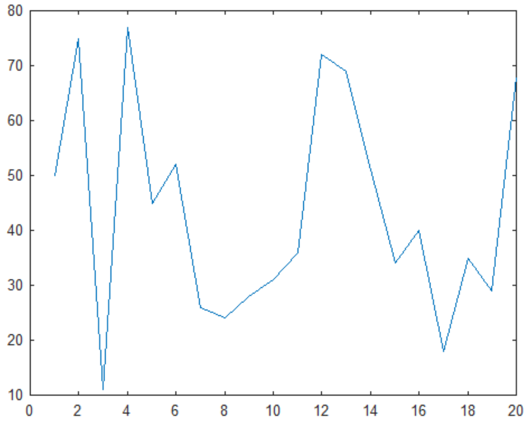


Figure 2: Relationship between sales executive and sales value

V. Conclusion and Future Work

Pailler Homomorphic encryption method is implemented in this paper and result is shown. Using Pailler Homomorphic encryption method the input data for linear regression is encrypted and the output graph is shown in Fig.2. In future work we are planning to find the error.

References

[1] Alkharji, Majedah, Hang Liu, and Washington CUA. "Homomorphic encryption algorithms and schemes for secure computations in the cloud." In *Proceedings of 2016 International Conference on Secure Computing and Technology*. 2016.

[2] El-Yahyaoui, Ahmed, and Mohamed Dafir ECH-CHERIF EL KETTANI. "A verifiable fully homomorphic encryption scheme for cloud computing security." Technologies 7, no. 1 (2019): 21.

[3] Guowei Qiu, Xiaolin Gui, and Yingliang Zhao, "Privacy-Preserving Linear Regression on distributed data by homomorphic encryption and data masking,"IEEE Access,Volume 8,2020.

[4] Hall Rob, Stephen E. Fienberg, and Yuval Nardi, "Secure multiple linear regression based on homomorphic encryption", Journal of Official Statistics, 27(4), 2011,P.669.