ECE209AS - Winter 2022
UCLA

“Hariram, Veeramani”
Review for Paper Titled:

“CryptoNets: Applying Neural Networks to Encrypted Data with High Throughput and Accuracy”

1- Summary

This work addresses the Data privacy and security problems faced by Cloud based Machine Learning solution providers operating on highly sensitive critical Data especially in fields such as Finance, Medicine focusing on Inference stage. It proposes CryptoNets that allows Neural Networks to be applied upon encrypted data thereby allowing the user to send their data in encrypted form to cloud service maintaining Data Confidentiality and Privacy and the encrypted operations can be sent back to the owners of the secret key who only can decrypt them. The main components of CryptoNets are the Leveled Homomorphic Encryption which helps to compute Polynomial functions of a maximal degree on the Encrypted Data when provided with the knowledge of the complexity of the Arithmetic circuits in advance and important modifications to Make Neural Networks compatible and efficient with LHE such as the Non-linear Activation Function replaced by Polynomial Activation Functions and the max pooling replaced by scaled mean pooling and collapsing consecutive layers that use only linear transformations into a single layer for time efficient computation

2- Novelty/Contribution

It proposes novel ideas to perform computationally efficient and low noise Addition and Multiplication operations on Plaintext constant. It considers the tradeoff involved in Encoding schemes such as converting real numbers to Fixed precision numbers and then either using the binary representation to convert them into a polynomial given by all their coefficients or by mapping it into a constant polynomial. For the Encoding problem with Large Integer coefficients, it proposes a new technique incorporating Chinese Remainder Theorem (CRT) which allows one to encode exponentially large numbers with linear time and space complexity in the number of primes needed to decompose the polynomial coefficients thereby avoiding the usage of very large ‘t’ field dimensions needed to represent them and the huge noise that comes along with large ‘t’. This CRT decomposition of single polynomial into multiple polynomials is further leveraged during Parallel computation by making the process Single Instruction Multiple data (SIMD) friendly. Also, the converse of the CRT decomposition is used to join the processed data and eventually arrive at a result. Combination of all these powerful techniques results in high performance, efficiency and reduced latency in adopting encrypted Neural Network based solutions and make this work practically feasible.

3- Evaluation

Evaluation of CryptoNets is carried on using the MNIST dataset. The Latency and Throughput metrics of the CryptoNet topology is evaluated at different stages such as the Encoding- Encryption stage, Network Application and Decryption-Decoding stage and evaluates the Messages sizes involved in the Cryptonet transaction from the Owner to Cloud and from the Cloud to Owner. The Encrypted Data is one to three orders of magnitude larger than the unencrypted data which further depends on the natural representation of the data in its raw form. The Latency stage result evaluations remain aligned and consistent with this magnitude of orders difference in message sizes during the corresponding transactions.

4- Unique Strength/Weakness

In addition to the converting normal neural networks to Encrypted Neural Networks, this work thoroughly analyzes the large overhead involved in providing the Encryption and provides novel techniques to mitigate the overhead complexity during different computations and sets the stage for further optimization developments in similar lines.