An Efficient Distributive Framework for Preserving Data Privacy through Block Chain

D Swapna, A Madhuri, T Sri Lakshmi, S Phani Praveen

ABSTRACT--- Deep Learning Models has gained much attention to perform various artificial intelligence tasks. The accuracy of the models relies on the availability of data. Privacy and auditability has become the major concern for data providers. First issue is the centralised server which may become malicious causing break in privacy. Second is no incentives are given for data providers and trainers. Block chain is the most emerging innovation as of late. Decentralised connectivity of block chains gives another approach to interface information without the overheads of security, trust and controls. To address the above issues we propose an algorithm where clients send the model to the block chain for training where the honest trainers are incentivized for training, sharing weights. The weights are averaged; parameters are updated by a smart contract that resides on block chain which guarantees privacy and audit ability

Keywords — Block chain, Network,Privacy

I. INTRODUCTION

Bitcoin is a crypto currency [1] which was introduced by Satoshi Nakomoto in the year 2008 to facilitate distributed method of fund transfer. It provides security for transactions through public key cryptography. Funds are transferred from one person to another only if the owner has the access to private key. In order to create a distributed system without a third party requires consensus to validate the transaction. The technology underlying under Bitcoin is block chain.

Block chain is a collection of digital records which are distributed, unchangeable, shared and stored as a block. Cryptographic Mechanisms are used to protect these blocks from data tampering. Network of users who adhere to the rules and regulations will be maintaining the peer to peer network of block chain and they are responsible for accepting new blocks. Each block in the network contains a timestamp. Every block is connected to the previous block i.e., every block contains the hash of previous block. By its nature data stored in the block chain is immutable. So data once stored in the block chain cannot be changed. By this property Blockchain can be used to make transactions in the network which are verifiable, efficient and untampered.After block is added to the block chain it cannot be updated or modified without altering the linked blocks and modification validation by the group of members in block chain.There are two types of block chain:

Public Block chain is [1] also called as permission less block chain i.e., Any Internet user can read and write to the block chain. Members in the network are unknown to each other. Every user can make transactions and read transactions of other user. Proof of work Consensus algorithm is used to validate the transactions and it can be done by any user called miner. Private Blockchain protocols[1] also called as permission Blockchain .Participants are known and trusted. In this write permissions are allowed only to registered participants. Validation of transactions that is consensus can be done only by registered participants .Therefore private block chain networks are considered as Centralised. Smart contract is termed by Nick Szabo in 1997. Systems today run in a centralised way. Parties who wish to transact with each other will do so via a centralised system. All the parties trust the central system. Smart contracts are just like contracts in the real world the only difference is they are completely digital. Smart contract is a computer program stored inside a block chain which includes contractual agreements that are verified automatically. The main aim of developing smart contracts is to remove the dependency on third party. Smart contract allows transacting directly with untrusted parties. All the nodes in the block chain have a copy of it. It gets executed when an authorised event triggers it. All transactions are stored in a sequential order on the block chain which facilitates future accessibility. Steps required to obtain solution for a deep learning algorithm are information acquisition, training, improvement or optimization that require a certain trust between parties. For example, if an association shares a set of data with a group of researchers, there is a verifiable trust connection between them that, if damaged, can influence the ultimate result of the task. Generally, there is no verifiable record or responsibility of a deep learning algorithm without trusting a centralised authority. Wouldn't it be pleasant on the off chance that we could decentralize the lifecycle of deep learning applications in a way that clients and research hers could work together in a trust less yet secure way.The first challenge is that every data researcher faces is the absence of information. Regardless of how great your model is, no information == no good times.

Revised Manuscript Received on 30 July 2019.
* Correspondence Author

D Swapna*, Assistant Professor, Dept of CSE, PVPSIT, Vijayawada, AP, India

A Madhuri, Assistant Professor, Dept of CSE, PVPSIT, Vijayawada,AP, India.

T Sri Lakshmi, Assistant Professor, Dept of CSE, PVPSIT, Vijayawada, AP, India.

S Phani Praveen, Assistant Professor, Dept of CSE, PVPSIT, Vijayawada, AP, India.

Retrieval Number: B1052078219/19©BEIESP
DOI: 10.35940/ijrte.B1052078219
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

This issue isn't so enormous for huge organizations in fact that their stuff is intended to secure an unbounded stream of information, consider Facebook, Google, and other average cases. Else, they will have enough assets, to gain information. To put it plainly, vast firms will dependably have advantage than littler firms who additionally try to use AI.

Second challenge is simply consider how much time we are effectively and inactively creating while at the same time utilizing our gadgets, there is simply so much information we are giving without end. This is to be sure a discussion about protection, yet who thinks about security? What number of us has really perused the Terms and Conditions before tolerating them? Indeed, we don't generally think that much about it. Notwithstanding, imagine a scenario in which you could procure cash by giving the information. All the more particularly, imagine a scenario in which you can without much of a stretch offer your information in an arrangement that they can't really read or get it. As information is the backbone of current digital society, many are yet required completely to grasp the requirement for proper acquisition and processing of information [2, 3]. Among the key worries in the production and utilization of information are privacy issues. This is significantly more critical in social insurance, where a high level of individual wellbeing information created could be viewed as private.

Gathering personal data became easy with today's computing technologies. Personal information is being constantly analysed by algorithms and making momentous decisions on individuals. The widespread use of algorithms is increasing the risk by violating privacy of individuals. Mistreating algorithms by individuals can become a risk. For an instance people may have benefits in the form of incentives for misreporting their data to algorithms in a strategic environment. In this paper, We propose a block chain based model which has the ability to train a deep learning model by preserving privacy, providing confidentiality for trainers data and clients model and offering incentives for trainers. Blockchain securely aggregates weights received from untrusted trainers ,miners will perform local training and parameter updation based on averaged weights who are also incentivised for participating in the process. Confidentiality and privacy are ensured by using cryptographic techniques. The rest of the paper is organised as follows. In Section 2 we discuss the motivating factors of this article.In Section 3 we introduced the background of this work.In Section 4 We presented the architecture of our proposed block chain based deep learning model trainer.In section 5 we have written a conclusion of our work.

II. MOTIVATION

Motivating examples include: 1. Assume N groups of research around the globe have collected numerous scientific datasets, such as sky survey data, and scientist wish to carry out learning over the combination of these diverse data sets exclusively by avoiding much communication.

3. Presume N hospitals with diverse distributions of patients want to build a classifier to recognize a frequent misdiagnosis. Now, in accumulation to the aim of achieving high accuracy, low communication, and privacy for patients, the hospitals may want to protect their own privacy in some formal way as well

III. BACKGROUND

Blockchain was first innovation has emerged a flood of interests both in the exploration network and industry [1]. It turns into a developing innovation as a decentralized, permanent, sharing and time-arrange record. Exchanges are put away into squares containing timestamps and references (i.e., the hash of a past square) which are kept up as a chain. In Blockchain, exchanges are made by pseudonymous members and intensely gathered to fabricate another square by an element called specialist. The laborer who constructs another and legitimate square can pick up measure of remunerations so the chain is consistently extended by aggressive specialists. That shows the motivating force system in the Blockchain setting. Also, ace creating Blockchain innovations presenting savvy contract bolster Turingcomplete programmability, for example, Ethereum and Hyperledger. Then again, a progression of deals with exchange protection are prominent by applying cryptographic apparatuses into Blockchain, for example, Zerocash, Zerocoin and Hawk . In this manner, Blockchain innovation's motivation highlight and its ace creating advances rouse us to understand our situation issues, for example, the nonappearance of impetus work what's more, cooperation decency.

IV. RELATED WORK

Models of Machine Learning being trained on data available from marketplaces that are based on block chain have the ability to create the most potent artificial intelligences. They merge two powerful primitives: Confidential Machine Learning which enables training on sensitive data without edifying it, and incentives based on block chain, which permits systems to catch the attention of best data and models to build them smarter. The consequence is open market places where anyone can trade their data and maintain privacy of their data, whereas developers can apply incentives to gain the finest data for their algorithms Richard developed Numerai [2]in 2015 based on this idea.Numerai encrypts data available in the market and sends data to data scientists who are working on a competent stock market model. A metamodel is developed by combing all the models.Numerai sells the metamodel and incentivizes data scientists based on the performance of their model. Secure computation methods permit to train models on data without edifying the data. Secure Computation being performed in 3 ways, zero-knowledge proofs, Homomorphic encryption, Secure Multi-party Computation. At present Secure multi party computation is being used widely as Homomorphic encryption is slow for complex models. It has not been clear on applying zero knowledge proofs to machine learning. This federated learning structure, in any case, can't ensure the protection of training information; still the training information is partitioned and put away independently.

For instance, a few researchers demonstrate that the middle gradients can be utilized to deduce vital data concerning training information [3],[4].

Shokri et. al [6] connected differential privacy by adding commotions in the gradients to transfer, accomplishing an exchange off between information security and training exactness. Hitaj et. al [5] called attention to that Shokri's work neglected to secure information protection and exhibited that an inquisitive parameter server can learn private information through GAN learning (Generative Adversarial Network). Phong et. al [7][8] projected to utilize Homomorphic encryption strategy to shield training information security from inquisitive parameter server. The downside of their plan is that they accepted the community oriented members are straightforward however not inquisitive; thus their plan may failure in situation where a few members are interested[9].

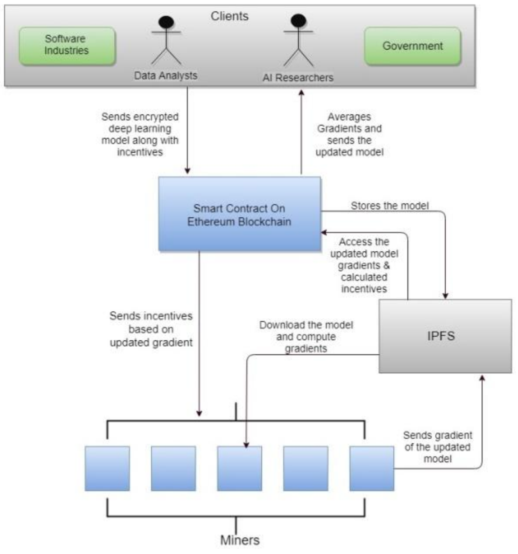
Differential privacy and Federated Learning in private machine learning are a step ahead in data privacy of Machine Learning, But this techniques does not permit users to examine their privacy and data is kept siloed. Danku Contracts by Algorthamia Research incentivizes a model which satisfies certain threshold.To train Deep Learning models open mind is creating a multiparty network above unity that can execute on any device.

V. RESULTS & DISCUSSIONS

Our idea is to build a layer of Deep Learning which will be hosted by participants in the network and is trained on data contributed by the participants of the network. If an individual such as a startup, researcher, data scientist and a government organisation are concerned in training a deep learning model funded by crypto currencies. At present most of the Deep Learning is occurring in locations that are centralised by corporate giants like Google, Facebook etc.Though Apple and Google are training decentralised models on their own mobile platforms, there is lack of availability of open networks.

Our approach for model training is motivated by distributed method of learning, which appeals to be good due to privacy concerns. The data remains in the nodes of the participants of the network and a model is shared. The primary advantage of distributed learning is privacy compared to training on centralised data.

Clients request a randomly initialised deep learning model by sending incentives. Client encrypts the model using pailier cryptosystem. Smart contract initialises the model according to specification and stores the model on Inter Planetary File System. Miners read that a new model is available and make an internal decision as to whether they should participate. A subset of miners downloads the model from IPFS and computes a gradient. Miners submit their gradient to IPFS and receive an address. Miners submit the IPFS address of their gradient to Smart Contract. Smart Contracts decide how each gradient should be weighted. According to weights of the gradient smart contract split the incentives among miners. Using Distributed Averaging Algorithm it averages the gradients and sends the updated model to the client.


Clients
Software Industries
Data Analysts
AI Researchers
Government
Sends encrypted deep learning model along with incentives
Averages Gradients and sends the updated model
Smart Contract On Ethereum Blockchain
Stores the model
Access the updated model gradients & calculated incentives
IPFS
Sends incentives based on updated gradient
Download the model and compute gradients
Sends gradient of the updated model
Miners

To incentivize the miner smart contract evaluates its gradients with the corresponding id, resulting weights and errors. Miners are incentivised only if the best error of the model did not reach the target error.

Total error=model.initialerror-model.targeterror

Solved error=model.besterror-newmodelerror

Incentive=bounty*solvederror/total error

Stochastic Gradient Descent has been used for optimizing the problem, where calculation of gradient is done for a miner who is randomly selected for a single round of communication. For this approach we select M-fraction of miners in each round and gradient computation is done of the overall data loss by these miners. Thus M has a control over the global batch size.

In implementation of Distributed Stochastic Gradient Descent with M=1 and each Miner k compute $gk = \Delta wk$ that is the gradients average on its local data on the present model. The smart contract running on the block chain aggregates the weights and updates as

$$w_{\text{SWA}} \leftarrow \frac{w_{\text{SWA}} \cdot n_{\text{models}} + w}{n_{\text{models}} + 1,},$$

That is every miner locally applies a single step of gradient descent on the present model using their local data and the smart contract running on the block chain retrieves the updated models from IPFS and calculates the average weight of the resulting models

VI. CONCLUSION

The combination of private machine learning with blockchain incentives can create the strongest machine intelligences in a wide variety of applications.

There are significant technical challenges which feel solvable over time. Their long term potential is enormous and a welcome shift away from the current grip large internet companies have on data. They are also a bit scary— they bootstrap themselves into existence, self-reinforce, consume private data, and become almost impossible to shut down, making me wonder if creating them is summoning a more powerful Moloch than ever before. In any case, they are another example of how cryptocurrencies will slowly, then suddenly make their way into every industry.

REFERENCES:

1. Nakamoto S. Bitcoin: A peer-to-peer electronic cash system. Consulted. 2008; 1(2012):28.
2. https://medium.com/numerai
3. M. Abadi, A. Chu, I. Goodfellow, H. Brendan McMahan, I. Mironov, K. Talwar, and L. Zhang. Deep Learning with Differential Privacy. ArXiv e-prints, 2016.
4. Jakub Koneˇcn´y_, H. Brendan McMahan, Felix X. Yu, Ananda Theertha Suresh & Dave Bacon FEDERATED LEARNING: STRATEGIES FOR IMPROVING COMMUNICATION EFFICIENCY ArXiv e-prints, 2017.
5. B Hitaj, G Ateniese, F Pérez-Cruz, "Deep models under the GAN:information leakage from collaborative deep learning," Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security. ACM, 2017, pp. 603-618.
6. R Shokri, V Shmatikov, "Privacy-preserving deep learning," Proceedings of the 22nd ACM SIGSAC conference on computer and communications security. ACM, 2015, pp. 1310-1321.
7. LT Phong, Y Aono, T Hayashi, L Wang, S Moriai, "Privacy-Preserving Deep Learning via Additively Homomorphic Encryption," IEEE Transactions on Information Forensics and Security, 2018, 13(5), pp.1333-1345
8. E. Bagdasaryan, A. Veit, Y. Hua, D. Estrin, and V. Shmatikov, "How to backdoor federated learning," arXiv preprint arXiv:1807.00459,2018.
9. An Optimized Rendering Solution for Ranking Heterogeneous VM Instances SP Praveen, KT Rao - Intelligent Engineering Informatics, 2018