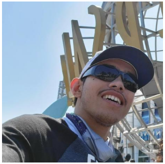
Ensuring Trustworthy Neural Network Training via Blockchain

IEEE CogMI 2023



Edgar Navarro

Kyle Standing

Gaby Dagher

Tim Andersen

Nov 2, 2023

San Diego State University

BYU

Boise State University

Motivation

- **Rise of AI:** The recent rise of Artificial Intelligence is fueling a growing need to ensure the integrity of AI models.
- **Integrity Concerns:** Two aspects of integrity are of particular concern.
  - Ensuring a model has not been compromised or “poisoned".
  - Certifying the completion of training.
- **Complexity and Cost:** The evolution of AI creates more complex and expensive training, leading to difficulty in tracking the integrity of the process.

Challenges

- **Focus:** We focus on a subset of AI: neural networks.
- **Unique Challenges:** Neural networks pose unique challenges in verifying their integrity.
  - Given a trained model, it is impossible to derive details about the training process that the model underwent.
  - The “black-box" nature of neural nets further exacerbates the issue, making it difficult to understand the inner-workings of a model and how it reaches a particular outcome.

Our Proposed Solution

A Blockchain network, tasked with validating neural networks by intelligently retraining select portions of the training process to validate integrity of the resulting models

Summary of Contributions

1. **Blockchain System:** Developed a system for efficient verification of neural network models using blockchain to ensure transparency and provide provenance.
2. **Weight-Analysis Algorithm:** Innovated a weight-analysis algorithm for intelligent distribution of training workload, optimizing computation and validation efficiency.
3. **Implementation and Testing:** Implemented and tested the proposed system within a functioning blockchain network, confirming robustness against adversarial attempts and demonstrating the accuracy and scalability of the algorithm.

Blockchain Technology and Quorum Consensus

- **Blockchain Technology:** A decentralized and distributed digital ledger that securely records transactions across multiple computers. Comprized of a network of nodes who collaborate to generate and validate new bocks.
- **Quorum Consensus:** A consensus process by which a new block is created. A random subset of nodes is chosen. The subset works together to propose, vote, and then broadcast new blocks to the rest of the network.

Blockchain Network for Model Verification

- **System Overview:**
  - *Submitters:* Provide trained models with detailed training information. All of this is packaged into a "transaction"
  - *Verifiers:* Complete the verification process and append approved transactions (models) into a new block to be added to the chain.
- **Verification Process:**
  - 1. Weight-analysis algorithm
  - 2. Re-training

Blockchain System

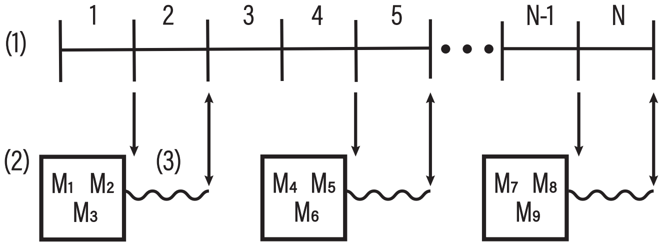


Figure: (1) Represents the model submitted for verification split into its sub intervals (2) Sub groups of the quorum that are selected to retrain specific intervals (3) The process of retraining

Weight-analysis Algorithm

Method	P	R	F1
Abs. Change	0.86	1.0	0.92
L2 Norm	0.80	0.67	0.73
Pct. Change	0.75	1.0	0.86
Cos. Dist.	0.11	0.17	0.13

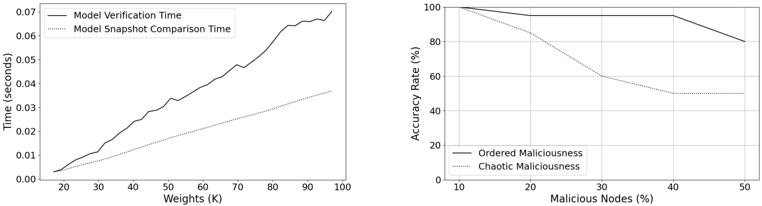
Table: Performance metrics for different methods of analyzing changes in neural network weights.

- **Abrupt Shifts:** Distinguishable from gradual changes in normal training.
- **Approaches:** Four methods were tested - Absolute Change, L2 Norm, Percent Change, Cosine Distance.
- **Best Performance:** The Absolute Change method performed best.

Overview of Experiments

- **Weight-analysis Scalability Experiment:**
  - Evaluate scalability of the detection mechanism against model complexity.
  - Record and average computation times, identify thresholds of efficiency.
- **Robustness Experiment:**
  - Test network robustness against varying degrees and types of malicious nodes.
  - Measure performance based on model integrity verification accuracy.

Experimental Results - Graphical Overview



(a) Scalability Experiment

(b) Robustness Experiment

Figure: Aggregated experimental results showcasing the performance of the poisoning detection mechanism under different conditions.

Future Work

- **Enhanced Detection Methods:** Further refinement or development of new methods for improved detection rates or efficiency.
- **Real-world Applications:** Applying the solution to real-world scenarios for insights into practical performance and potential limitations.
- **Scalability:** Explore performance with larger and more complex models, and develop methods to further optimize computational efficiency.
- **Consensus Mechanism:** Enhance the quorum consensus mechanism for model verification, develop more efficient selection procedures, and incorporate additional checks.

Questions?