SECURE PRIVACY PRESERVING DEEP LEARNING AGAINST GAN ATTACKS

A Thesis by

Aseem Prashar

Master of Science, Wichita State University, 2020

Bachelor of Engineering, BITS-Pilani, 2013

Submitted to the Department of Electrical Engineering and Computer Science
and the faculty of the Graduate School of
Wichita State University
in partial fulfillment of
the requirements for the degree of
Master of Science

May 2020

SECURE PRIVACY PRESERVING DEEP LEARNING AGAINST GAN ATTACKS

The following faculty members have examined the final copy of this thesis for form and content, and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science with a major in Computer Science.

Sergio Salinas, Committee Chair

Akmal Mirsadikov, Committee Member

Remi Chou, Committee Member

Ajita Rattani, Committee Member

DEDICATION

I dedicate this thesis to my family, friends, and colleagues.

Somewhere, something incredible is waiting to be known.

ACKNOWLEDGEMENTS

I would like to thank my adviser, Sergio Salinas, for his thoughtful input, guidance and support in all stages of this project. I also extend my gratitude to members of my committee, Ajita Rattani, Remi Chou and Akmal Mirsadikov for their valuable time and consideration.

ABSTRACT

Deep learning is a class of machine learning algorithms that use a cascade of multiple layers of non-linear processing units for feature extraction and transformation. Artificial neural network based deep learning is becoming increasingly popular in a variety of fields. Deep learning benefits from larger input data sets and can be revolutionary to organizations that have access to sizeable raw data. In the recent years, researchers have proposed decentralized collaborative learning architectures that allow multiple participants to share their data to train deep learning models. However, privacy and confidentiality concerns limit the application of this approach, preventing certain organizations such as medical institutions to fully benefit from collaborative deep learning. To overcome this challenge, deep learning models that only share abstracted data for collaborative learning have recently been proposed. This approach helps users keep their actual datasets private whilst contributing to and befitting from collaborative learning. However, some researches have outlined threats that can use can take advantage of abstracted data to recreate original data and violate user privacy.

In this paper, we propose a collaborative deep learning approach that allows an organization improve their deep-learning model while preserving its privacy from such attacks.

Specifically, we design our approach to protect organizational data against attacks that involve a malicious participant that can learn meaningful information from the abstracted dataset. Our proposed system protects the organization's privacy by limiting the exposure of private data from users to foreign entities.

Our solution does not involve computationally expensive cryptographic processes and relies on limiting the exposure of private dataset of participants. Our approach is flexible and can be adapted to work with different neural network architectures. We demonstrate the efficacy of approach by calculating the resulting accuracy on benchmark datasets.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

ν	Neutrino
γ	Photon/Gamma
c	Speed of Light

CHAPTER I

INTRODUCTION

In the past few decades, deep learning has generated a lot of interest in the research and academic community due to its great ability to automatically classify large amounts of data. This has led to breakthroughs in many fields ranging from autonomous driving, and natural language processing to genetic research [1, 2, 3, 4].

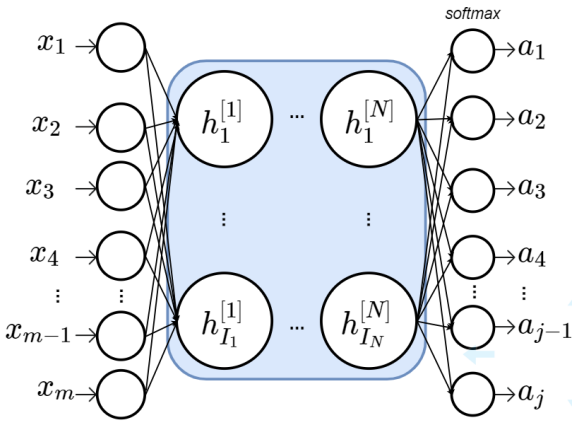

Figure 1: A neural network with m inputs, j outputs, N hidden layers, and I neurons per layer.

CHAPTER II

RELATED WORK

Deep neural networks have outperformed traditional machine learning approaches for many tasks, and are the tool of choice in many fields. Specifically, Deep learning has been successfully used for facial recognition [?, ?, ?], image classification, [?, ?, ?], and speech recognition [?, ?, ?], where it is expected to achieve better performance than humans in the near future. However, directly applying these techniques in fields that deal with private data is challenging. The reason is that they need to centrally collect data at a third-party which organizations may not trust [?]. This is particularly challenging in medical and financial applications where the privacy of the users is governed by federal legislation and is protected by law.

CHAPTER III

DEEP NEURAL NETWORK

Deep neural networks are a type of machine learning that has recently shown high accuracy in data classification tasks. Traditional machine learning requires manual feature selection, which can be time consuming and inaccurate. In contrast, deep learning learns the most relevant features in the data on its own. In other words, the deep neural networks can be trained with raw data without the burden of preprocessing it. Since deep neural networks have more hidden layers compared to traditional neural networks, their accuracy is proportional to the amount of data used for training, i.e., the larger the training data set the more accurate that the deep neural networks become. These advantages make deep neural networks a very effective technique to perform data classification tasks. In this section, we describe the architecture of deep neural networks and their training methods.

3.1 Architecture

In this work, we consider multilayer perception (MLP), which is one of the most common deep neural network architectures. An MLP is formed by multiple layers where each layer consists of many nodes. Each node takes as input a weighted average of the previous layer's node outputs, and the output of an special node called the bias. The nodes use a non-linear activation function to the compute their output. Together, the weights used in the weighted average and the biases from the special neurons are called the parameters of the deep neural network.

Figure **??** shows the structure of a typical classification MLP with m input nodes and j outputs nodes. The neural network has N hidden layers and each layer has I neurons. Intuitively, this MLP takes a data sample represented as a vector of length m on its input layer, and outputs the probability that it belongs to the jth category on the jth output neuron.

$$a_k^i = f(W_k a_{k-1}) \tag{1}$$

3.2 Training

Before a neural network can be used to perform inference, e.g., classify images, it needs to be trained to learn the highly non-linear relationships between the inputs and the correct outputs. Training finds the parameters of the deep neural network, i.e., its the weights and biases, that result in the inferences with the highest accuracy.

3.2.1 Gradient Descent Algorithm

The GD algorithm finds the parameter updates in two steps: error forward propagation and back propagation.

3.2.2 Stochastic Gradient Descent

Although the GD algorithm is effective at finding the parameters of DNNs, all training samples in the dataset need to be processed before a single update is made to the parameters. That is, the algorithm processes the complete training data at each iteration. This is computationally intensive and time consuming.

CHAPTER IV

PROBLEM FORMULATION

In this section, we describe our considered collaborative deep learning model, and the threat model.

4.1 System Model

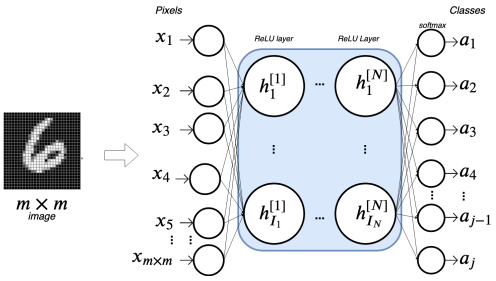

Figure 2: Neural network depicting an image with $m \times m$ pixels fed as input, j outputs N hidden layers and with I neurons in layer.

Table 1: Injection Trials

Trial	Method	Result
1	Hypodermic Injection	Dizziness
2	Oral Ingestion (2.46)	Vomiting
3	Bio-injection suit	Full size control

CHAPTER III

CONCLUSION

With the injection of pym particles I am able to enter the quantum realm and speak with insects. Future work will include exploration of of enlarging myself and developing crime fighting alter egos (see Appendix).

BIBLIOGRAPHY

References

[1] T. Young, D. Hazarika, S. Poria, and E. Cambria, "Recent trends in deep learning based natural language processing," *ieee Computational intelligenCe magazine*, vol. 13, no. 3, pp. 55–75, 2018.

[2] M. Al-Qizwini, I. Barjasteh, H. Al-Qassab, and H. Radha, "Deep learning algorithm for autonomous driving using googlenet," in *2017 IEEE Intelligent Vehicles Symposium (IV)*, pp. 89–96, IEEE, 2017.

[3] B. Huval, T. Wang, S. Tandon, J. Kiske, W. Song, J. Pazhayampallil, M. Andriluka, P. Rajpurkar, T. Migimatsu, R. Cheng-Yue, *et al.*, "An empirical evaluation of deep learning on highway driving," *arXiv preprint arXiv:1504.01716*, 2015.

[4] P. Danaee, R. Ghaeini, and D. A. Hendrix, "A deep learning approach for cancer detection and relevant gene identification," in *PACIFIC SYMPOSIUM ON BIOCOMPUTING 2017*, pp. 219–229, World Scientific, 2017.

[5] B. Banner, "Gamma radiation and other super physics," *Mavel Physics Applied*, 1948.

APPENDIXES

APPENDIX A

Possible Alter Egos

Possible alter egos to facilitate crime fighting: Ant-Man, Giant Man, Quantum-Man, The Wasp, The Whisper.

APPENDIX B

My Awesome Suit

Here's a look at my awesome suit.