AutoGAN-based Dimension Reduction for Privacy Preservation

Hung Nguyen, *University Of South Florida,* Di Zhuang, *University Of South Florida,*
Peiyuan Wu, *Member, IEEE,* Morris Chang, *Member, IEEE*

More elaboratively, in the proposed framework, the face images are locally collected which are nonlinearly compressed to achieve DR, and then sent to an authentication center

***Abstract*—Along with the explosion of information and the existence of smart cities, exploiting big data and protecting individual sensitive information for whom data belongs to are emerging objectives. Several methods have been introduced to protect individuals privacy while aiming to maximize the utility of the data which respect to a target in the proposed scenario. In this paper, we introduce a theoretical tool to quantify dimension reduction -based privacy and a private dimension reduction framework for privacy preservation that can work well with machine learning algorithms while protecting privacy. In the experiments, we implement our method on different popular face image datasets. We show that our method meets the requirement of the dual-target.**

***Index Terms*—Generative Adversarial Nets, Auto-encoder, neural-network, privacy, machine learning, dimension reduction, reconstruction.**

介紹 Machine learning as a service (MLaaS) 較有內容

I. INTRODUCTION

Artificial intelligence, data mining and machine learning (ML) are common terms used recently. Despite distinct points, those techniques all aim at utilizing known data to build up models for applications such as prediction or estimation. There are various types of user's data being collected in a smart-city such as patients records, salary information, biological characteristics, Internet access history, personal images, etc. Those types of data could be widely used in daily recommendation systems, business data analysis, or a disease prediction system. However, collecting and using such data might raise privacy issues for individuals who contribute their sensitive data. For instance, in an on-line access control system, there are a number of users who would like to access multiple recourses in a data system secured by a face authentication system. The utility of face features can be effectively used in machine learning tasks for authenticati~~ng~~ on purpose. However, leaking [redacted] the face images might lead to a severe security problem such as exposing individual's behaviors.

on-line access system that collects data via mobile devices raises privacy issues

Several methods have been developing to protect data. However, those methods usually come from security perspective. Thus, applying them directly to machine learning might experience several challenges such as high computational cost and time consumption. In order to tackle those problems, we developed a [redacted] privacy-preserving dimension reduction (DR) system which not only ~~help processing machine learning tasks~~ but also guarantee a ~~certain level of~~ privacy. ~~It is unlike other DR methods in data mining which only~~ [redacted] considers the utility of the data, ~~our method considers both data's utility and privacy.~~ The method can be used in several practical applications. For example, the ~~proposed framework in section VI can~~ (be specific ?) be applied directly to the online access control system mentioned above. ~~Our framework provides a face image private transformation and also a classifier model which can be used to grant permission for authenticated users. The user's face images are locally collected and transformed to DR form, then sent to an authentication center.~~ The authentication server, then performs authenticates tasks with the dimensionality-reduced data ~~for users who achieve positive output from the classifier. With this setup,~~ We assume ~~can accept a~~ semi-honest authentication server which does not deviate from [redacted] authenticating protocols ~~procedures, but~~ while curious about specific user's behaviors. A strong adversary who obtains the image training dataset and the transformation model can hardly ~~and intends to determine a specific user's behaviors by~~ reconstruct~~ing~~ the original face images ~~will face challenges.~~ In addition, the method provides a light-weight face feature form which can lower storage and communicational cost.

The DR mechanism is designed in a way that even

Put to contribution

comment on diff. to DP

utility and privacy definitions unclear.

In order to analytically support privacy guarantees, we first introduce a theoretical tool $\epsilon - DR$ for distance-based privacy evaluation. Secondly, we propose a non-linear dimension reduction framework for privacy preservation ~~based on state-of-the-art deep learning methods~~ motivated by Generative Adversarial Nets and Auto-encoder Nets [1] that [redacted] satisfies the $\epsilon - DR$ privacy. Finally, we perform several experiments on a benchmark dataset for practically proving our method. The experiments illustrate that our method can achieve a high accuracy up to 100% with very low number of reduced dimension and can preserve the privacy.

One experiment is too weak

Mention utility definition

The remainder of the paper is ~~constructed~~ organized as follows. Section II reviews the state-of-the-art PPML methods. Section III reviews a background knowledge of deep learning methods that are utilized in our work. Section IV describes the problem. Section V introduces a definition of $\epsilon - DR$ privacy. Section VI introduces our framework as mentioned above. Section VII presents the experiments and a discussion about issues might be occurred. Finally, the conclusion and future work are mentioned in section VIII.

be specific

Contribution bullets

II. LITERATURE REVIEW

We categorize current PPML methods into two main approaches as follows:

Cryptographic approach: This approach usually applies to the scenarios ~~such that~~ where the data owners do not ~~want~~ wish to expose their plain text sensitive data ~~but want~~ while asks for ML services from a third-party ~~to perform particular machine learning tasks~~. The most common tool used in this approach is fully homomorphic encryption that supports multiplication and addition operations over encrypted data, which enable the ability to perform a more complex function.

Should mention Differential privacy, homomorphic encryption, Compressive privacy, local differential privacy. (Consult Mohammad) and their pros and cons

Our method belongs to compressive privacy. Mention about GAP, and how that is similar/different from our work.

In Hesamifard's work [4], the fully homomorphic encryption is applied to deep neural networks, where the nonlinear activation function is approximated by polynomials. However, polynomials is difficult to _ (unclear to me)

However, the high cost of ~~the~~ multiplicative homomorphic operations ~~makes~~ it ~~hard to apply~~ renders difficult to be applicable on machine learning tasks. Instead, additive homomorphic encryption schemes are widely used in PPML. Thus, the limitation narrows the ability to apply on particular ML techniques or scenarios. (be specific) Such methods in [2], [3] are applicable to simple machine learning algorithms such as decision tree and naive bayes. In [4], Hesamifard aims to perform deep neural network on encrypted data by utilizing property of fully homomorphic encryption. The paper introduced a solution for applying neural network over encrypted data by giving polynomial approximation of the activation function which is usually hard to describe using fully homomorphic encryption properties. Although these encryption-based techniques can protect the privacy in particular scenarios, computational cost is a significant concern. ~~A number of other popular proposed scenarios in PPML are based on~~ In secure multi-party computing (SMC) ~~in which~~ multiple parties ~~would like to~~ collaborate ~~to compute particular machine learning~~ to compute functions without ~~leaking directly from the data of others~~ leaking revealing input to other parties. A widely used tool in SMC is garbled circuit, [5] a cryptographic protocol carefully designed for two-party computation in which they can jointly evaluate a function over their sensitive data without the trust of each other. ~~The idea of garbled circuit was first introduced by Yao in his paper~~ [5]. In [6], Mohammad ~~was~~ introduced a ~~securely multi-party computing~~ SMC protocol for Principle Component Analysis (PCA), which is a hybrid system utilizing additive homomorphic and garbled circuit ~~in order to compute Principle Components Analysis (PCA) of a jointed dataset among multiple parties without learning anything from one to the other~~. In Secret sharing techniques ~~has been used in SMC~~ [7], ~~The idea of this technique is distributing~~ a secret s is distributed over multiple pieces n also called shares, where the secret can only be recovered by a sufficient amount of t shares. ~~and only sufficient number of shares t can be used to recover the secret, but any set of at most t-1 shares reveals no information about the secret. [8] has reviewed~~ A good review of secret sharing-based techniques and encryption-based techniques for PPML is given in [8]. ~~and made a comparison between the two. However, besides computational cost, the paper also showed that~~ Furthermore, as to be elaborated in this manuscript, the high communication cost ~~is still~~ also poses a big concern ~~of~~ for both techniques.

Non-Cryptographic approach: ~~The most widely definition used in this approach is~~ Differential Privacy (DP) ~~introduced in 2006 by C. Dwork~~ [9]. ~~Basically, techniques in this approach aim to achieve differential privacy by~~ adding noise ~~drawn from a certain distribution in order~~ to ~~guarantee individual data confidentiality which can~~ prevent ~~individual~~ membership inference attack. For better understanding the mathematics definition, we can look at an extreme case where the system is designed with $\epsilon = 0$ which implies the best differential privacy protection. D is a salary database of labors in a company and M is the mechanism resulting in the average salary in the company. D is the database when there is a new labor joining the company, thus there is one different element between D and D. If the mechanism satisfies $\epsilon = 0$, the outcome of M is always in S despite the difference between D and D. Hence, it is hard to infer any information about the salary of the individual who joined the company base on the difference between D and D. One can design a system with a certain value of ϵ by adding noise to the data directly. However, this method usually leads to low accuracy results. Indeed, adding noise to output and parameters of the mechanism, are widely used. [10], [11] proposed methods to guarantee -differential privacy by adding noise to outcome of the weights $w^* = w + \eta$, where η drawn from Laplacian distribution and adding noise to the objective function of logistic regression or linear regression models. [12], [13] satisfied differential privacy by adding noise to the objective function while training a deep neural network which using stochastic gradient descent as the optimization algorithm. However, the limitation of those methods is they are designed for specific mechanisms and not working well for other algorithms.

Can't understand logic

unclear to me

is distributed over

Please study how Mohammad introduce DP

There are also existing works proposed differential privacy dimension reduction (DPDR). Dimension reduction is an important process before going through machine learning algorithm. Hence, one can guarantee -differential privacy by perturbing dimension reduction outcome. Principal component analysis (PCA) is a popular method in dimension reduction, in which its output is a set of eigenvectors. The original data will then be represented by its projection on those vectors, which keep the largest variance of the data. One can reduce the data dimension by eliminating insignificant eigenvectors which contain less variance, and apply noise on the outcome to achieve differential privacy [14]. DPDR could be used to protect privacy in a number of scenarios and works with few machine learning algorithms; however, record-level differential privacy, are not effective with image dataset as shown in [15].

Similar to this work, there is another work that utilizes well-known DR techniques for privacy preservation. In [16], after projecting the image data into lower dimension using principal component analysis and linear discriminant analysis techniques, the author claims that they could achieve privacy base on the reconstruction error between the original data and the reconstructed data using a certain number of principal components. However, the work has not considered the case that the adversary can obtain the dataset. As this work is using linear determinant DR techniques, the adversary could utilize the training dataset to compute all principal components and obtain the reconstruction of original data. In contrast, our framework provides a non-linear DR transformation implemented by a convolutional neural network which is hard to be reconstructed. Additionally, our model has considered a strong reconstruction function during training the model which is implemented by an convolutional Auto-encoder.

III. Preliminaries

In order to lead the DR system to a better privacy preserving area, we extend the structure of Generative Adversarial Network (GAN) [17] and utilized the structure of deep Auto-encoder as a reconstruction method to solve the problem. In this section, we briefly review Auto-encoder and GAN.

A. Auto-encoder

Auto-encoder is an artificial neural network framework which aims at learning a lower dimension representation of an unsupervised data or learning generative models of data. Auto-encoder can be used in dimension reduction and denoising