Privacy-preserving Deep Learning Models

Abstract

The blooming of deep learning techniques has led to impressive achievements in a wide variety of applications. Training the models usually requires massive, representative datasets. Distributing the trained model to adversaries can raise severe privacy issues as the datasets usually contain sensitive information. Addressing this goal, this project aims at studying novel deep learning algorithms to tackle the privacy problems. We plan to implement privacy-preserving deep learning strategy over distributed learning systems or data augmentation.

I. Problem Statement

Recently, deep learning has been applied in various areas, especially classification for complex data such as speech, image and text. The remarkable progress of deep learning techniques highly depends on the availability of large datasets. For example, in the case of health care and medical research, the datasets might be (a) images such as the X-ray images and magnetic resonance imaging (MRI) scans; (b) numerical information, such as vital signs such as heart rate, respiratory rate, and temperature. Obviously, these datasets are highly related to human's biometrics. As pointed out in [1], once the biometrics is stolen, there is no way to get it back. In some scenarios, an ambitious adversary may be able to recover training data from the model. Therefore, privacy guarantee is a significant aspect that has be taken into account in designing deep learning algorithms over data analysis.

To ensure the studied deep learning algorithms fulfill the demands of applications as well as preserve privacy, previous study investigated how the data augmentation, e.g. flipping or cropping the images, preserves privacy [2]. Additional to data augmentation, applying distributed deep learning technique to meet the privacy constraint is studied in [3]. Distributed training, a method of collaborating trained models based on different datasets, provides an efficient strategy for designing privacy-preserved models. It has the advantage that it allows training with large datasets without sharing the raw users' data among institutions.

II. Project Goals

We consider reproducing the results in [2] firstly. Afterwards, if time allows, results in [3] should be reproduced. Finally, based on those previous work, we will investigate and compare the performance of these two methods, e.g. robustness and accuracy.

Reference

[1] B. Schneier, "Inside risks: The uses and abuses of biometrics," Commun. ACM, vol. 42, no. 8, pp. 136, 1999.

[2] Sablayrolles, A., Douze, M., Schmid, C., & Jégou, H. (2018). Déjà Vu: an empirical evaluation of the memorization properties of ConvNets. CoRR, abs/1809.06396.

[3] K. Beaulieu-Jones, Brett & Yuan, William & G. Finlayson, Samuel & Steven Wu, Zhiwei. (2018). Privacy-Preserving Distributed Deep Learning for Clinical Data.