Neurocomputing 384 (2020) 21–45

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

A review of privacy-preserving techniques for deep learning

Amine Boulemtafes[a,b,*], Abdelouahid Derhab[c], Yacine Challal[a,d]

[a] Division Sécurité Informatique, Centre de Recherche sur l'Information Scientifique et Technique, Algiers, Algeria
[b] Département Informatique, Faculté des Sciences exactes, Université de Bejaia, 06000 Bejaia, Algeria
[c] Center of Excellence in Information Assurance, King Saud University, Riyadh, Saudi Arabia
[d] Laboratoire de Méthodes de Conception des Systèmes, Ecole Nationale Supérieure d'Informatique, Algiers, Algeria

ARTICLE INFO

Article history:
Received 28 July 2019
Revised 10 November 2019
Accepted 11 November 2019
Available online 4 December 2019

Communicated by Dr. Oneto Luca

Keywords:
Deep learning
Deep neural network
Privacy
Privacy preserving
Sensitive data
Taxonomy

ABSTRACT

Deep learning is one of the advanced approaches of machine learning, and has attracted a growing attention in the recent years. It is used nowadays in different domains and applications such as pattern recognition, medical prediction, and speech recognition. Differently from traditional learning algorithms, deep learning can overcome the dependency on hand-designed features. Deep learning experience is particularly improved by leveraging powerful infrastructures such as clouds and adopting collaborative learning for model training. However, this comes at the expense of privacy, especially when sensitive data are processed during the training and the prediction phases, as well as when training model is shared. In this paper, we provide a review of the existing privacy-preserving deep learning techniques, and propose a novel multi-level taxonomy, which categorizes the current state-of-the-art privacy-preserving deep learning techniques on the basis of privacy-preserving tasks at the top level, and key technological concepts at the base level. This survey further summarizes evaluation results of the reviewed solutions with respect to defined performance metrics. In addition, it derives a set of learned lessons from each privacy-preserving task. Finally, it highlights open research challenges and provides some recommendations as future research directions.

1. Introduction

Deep learning, one of the most advanced approaches of machine learning, has attracted a lot of attention in research as it provides the ability to overcome the dependency on hand-designed features that is faced by traditional learning algorithms. Deep learning or usually deep neural networks (DNNs) typically comprises two phases: a training step to optimize the accuracy of the model and an inference phase where model is used for analysis as classification or prediction (see the next section for more details). Nowadays, deep learning is being used in different domains including big data analytics and different applications such as pattern recognition, speech recognition, computer vision, natural language processing, intrusion detection, and medical predictions [1].

Deep learning experience is particularly improved by leveraging powerful infrastructures such as clouds and adopting collaborative learning for model training. As user devices are limited in terms of resources, the solution is to offload resource-demanding operations to an external infrastructure with high-power computation and massive storage such as a cloud. On the other hand, collaborative learning is applied on large and diversified datasets that are originated from different sources, e.g., medical organizations or patients, which results in achieving better learning accuracy. However, privacy concerns, which are related to sensitive data for both model training and its use for inference, are raised. Such concerns include identification of individuals, unauthorized commercial sharing of confidential information, illegitimate use of private data, and the disclosure of sensitive data or inferred private information like disease risks from health records. Additionally, other privacy concerns related to sharing a deep learning model need to be considered. In fact, it has been shown that if training private data are not well protected, they are subject to leakage through model parameters or predictions [2–8].

To tackle the above mentioned concerns, various approaches have been proposed. This survey aims to present a state-of-the-art of recent deep learning techniques and approaches addressing potential privacy concerns, and particularly related to input data which is the focus of this work, along with potential interesting directions and learned lessons.

In the literature, there are two related surveys [9,10] that deal with privacy-preserving in deep learning. However, these

* Corresponding author at: Division Sécurité Informatique, Centre de Recherche sur l'Information Scientifique et Technique, Algiers, Algeria.
E-mail addresses: aboulemtafes@cerist.dz (A. Boulemtafes), abderhab@ksu.edu.sa (A. Derhab), y_challal@esi.dz (Y. Challal).

https://doi.org/10.1016/j.neucom.2019.11.041
0925-2312/