Data Privacy-Preservation: A Method of Machine Learning

Sandesh Achar

Staff Engineer, Intuit Inc., Mountain View, California, **USA**

ABSTRACT

The privacy-preservation field in cyber security tends to affiliate with the protection measure related to the use of data and its sharing via third parties for activities such as data analysis. The paper's main objective for this research article will be to use machine learning models that tend to aid as a privacy-preservation technique (PPT). The augmentation of machine learning as a technique for privacy preservation has been able to address the challenges facing the current field of cyber security concerning data protection and security. The paper summarizes the methods such as "federated learning" to address the current issue in the network security field relating to data protection. The rise of augmentation of machine learning in privacy preservation is due to the development of cloud-based applications that are usually prone to data protection issues. Thus, the result of machine learning was necessary to counteract data insecurity. However, the use of machine learning in privacy preservation has remained proficient; there still needs to be a literature gap between the theory and the application of machine learning.

Keywords: Cloud, Cyber Security, Machine Learning, Privacy Preservation

11/25/2018 **Source of Support:** Nil, **No Conflict of Interest:** Declared

INTRODUCTION

The use of machine learning as a privacy-preservation technique is an efficient approach to solving the crisis related to data insecurity. This is seconded by the fact that machine learning and Artificial Intelligence require data that is usually raw and critical for the intelligent algorithms to function. An example showing the criticality of data towards machine learning is where the algorithms are typically used to create and develop predictive models for decision-making (Xia et al., 2012). This is usually accompanied by collecting and utilizing data where the two are comparable. Concerning data privacy, data collection becomes an issue due to privacy-related concerns such as reputational harm, breaching of data, and financial depreciation (Adusumalli, 2016).

Augmentation of machine learning into privacy preservation is not only to utilize the act of data protection but also to bridge the gap between the acceptance of benefits outsourced from using machine learning and data privacy as e benefit by itself.

The combination of privacy preservation and machine learning has brought about the rise and development of Privacy-Preserving Machine Learning (PPML). The algorithm's primary and general working description is to allow the use of a large domain consisting of multiple inputs to enhance privacy-enhancing strategies. This, in return, allows the machine learning models to cooperate with the private data that is usually input and saves the exposure of sensitive and personal data in its original form (Chen et al., 2017).

Though the use of machine learning as an avenue of privacy preservation seems efficient and fundamental, the technique plays the role of a necessary evil in the cyber security society. This can be seen from the fact that the machine learning algorithms function by using and training data (Ray, 2016). Thus, cyber adversaries and attackers may employ the use of intrusion techniques so that they can access data in the machine learning algorithm that is used to train the models. Furthermore, cyber attackers are usually drawn to this data as machine learning models require enormous sets of data so that they can perform efficiently. Thus, whenever data leakage occurs, cybercriminals can get access to sensitive and valuable data, which they may tend to sell or demand ransom. Thus, the use of machine learning poses a vulnerable method to be used in data protection.

Stakes involved in deploying machine learning for privacy-preserving

Privacy of data in training

The assurance that cyber adversaries cannot employ their malicious skills in accessing the data is always doubted. The only possible counteract measure to be put in place for this is data encryption (Truong & Dustdar, 2015). However, cyber adversaries will not see it challenging to reverse engineer the data from its ciphered form to original plain text, thus reconstructing the training data and getting access to valuable information.

Data input and output privacy

Data protection will be an assurance whenever an individual or client enters data (Cassel, 2012). The issue here is to allow that the input of data is only seen by a client or individual and not even by the model developer. In addition, the assurance that the client will only see data that will be seen as output remains an issue.

Privacy of the model.

A third-party adversary may tend to steal the model crucial to both company and user roles (Dehury & Sahoo, 2016). If other third-party hostile companies can imitate a particular model, then there will be needless to develop new artificial intelligence models dealing with data protection. This will negatively affect a company as the model will become outdated due to their little to no incentive to develop new models. This will create a vulnerability where cyber attackers can penetrate through intrusion and steal valuable training data.

MATERIALS AND METHODS

The methodological approach was based on evaluating the different techniques and tools, such as python libraries, that could be used in privacy preservation. For example, the various PPML techniques were identified and compared to highlight the common and significant characteristics. In addition, the assessment of clouds such as AWS or Microsoft Azure was also researched as the development of machine learning in privacy preservation is mainly deployed on cloud platforms. Finally, data were collected online, and article journals from the internet that had past and previous records of machine learning use on privacy preservation.

RESULTS

PPML techniques

The techniques involved in PPML allow data security by ensuring that the stealing of data cannot occur. Therefore, the following tactics and methods will be used to counteract various assaults.

Differential privacy

For this, as a method of PPML, relevant information regarding a particular data set is usually allowed for one to provide. This comes with an aftermath where no personal information can be released about it regardless of whether a malicious cyber adversary has access to the data. Furthermore, the information entered cannot be linked to the trained data in the machine learning model; thus, the impact of the outcome does not depend on the individual's record.

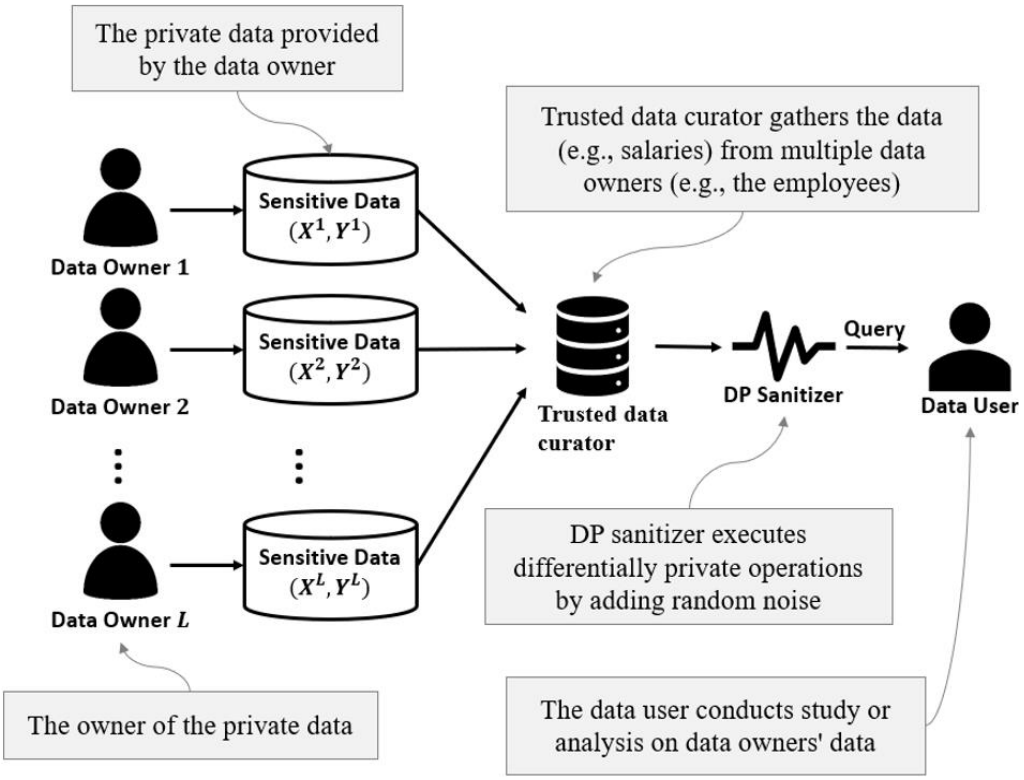

Figure 1. Working methodology of Differential Privacy

Differential privacy working using The Laplace distribution

The Laplace mechanism functions through a symmetric alignment where its distribution is exponential (Iyawa et al., 2017; Michalski et al., 1983). Thus, the instrument can be significantly used for privacy guarantee. This will be so as the tails rationing is constant when comparing two Laplace distributions.

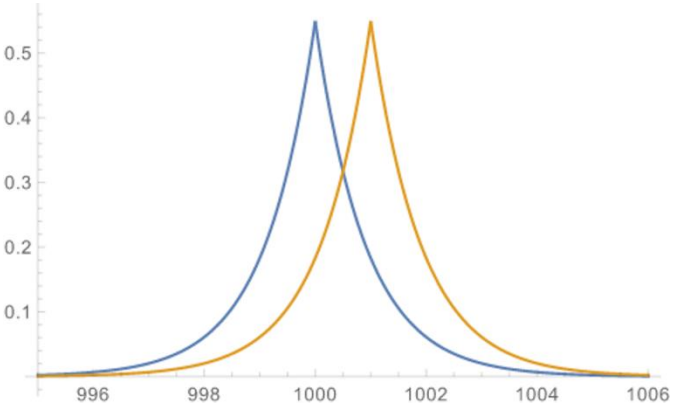

Figure 2. Graph showing an average Laplace Distribution

Privacy-preserving via Ensemble techniques

The definition of ensemble techniques portrays the risks that should be taken by a company or central authority so that privacy is achieved through machine learning (Achar, 2017). The type of technique that will be required will depend on factors such as the data's relevance and sensitivity, the cyber adversaries' abilities, and their resources. So that preservation of privacy is maintained, the proposal of a combination of several and multiple privacy-preserving techniques needs to be put in place so that they can act as a mitigating strategy to curb data breaches. For example, they use different machine learning algorithms to create varying models.

Federated learning.

Federated learning allows the decentralization of the relevant machine-learning processes. This, in return, lowers the amount of exposed information and data from data sets that may have a contributing factor. In addition, the dangers associated with the comprise of data and information are significantly reduced, and the rate of identity privacy breaches is minimized.

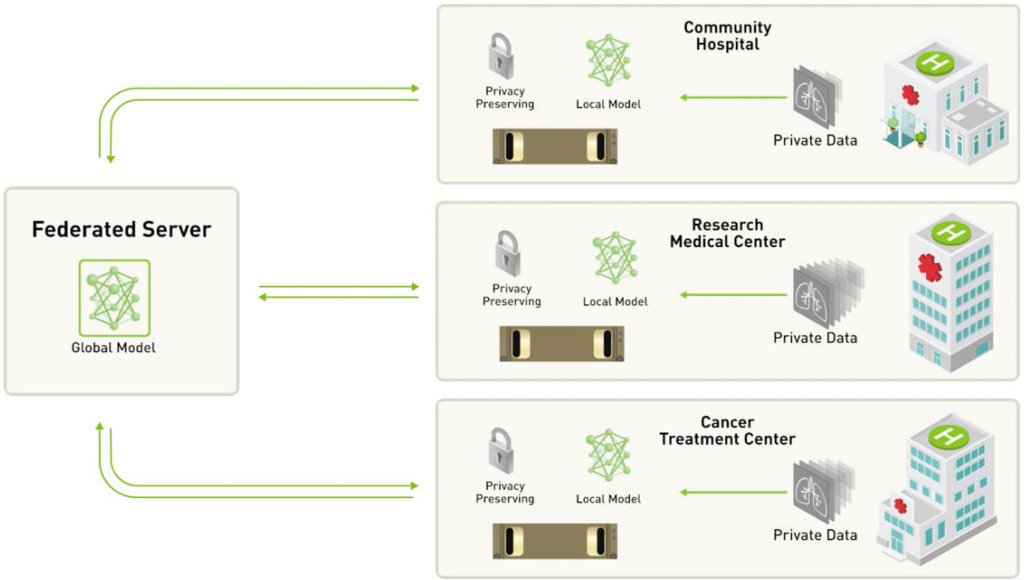

Figure 3. Working methodology of a Federated Learning server algorithm

The working methodology of a federated machine learning model is based on a central model that is possessed by a central authority (Achar, 2016). The enclosed and owned model can have further training and new data sets that may be private from data contributors. Each data contributor can locally train the model. After updating the model's parameter using the central model M., the above procedure is usually repeated and done repetitively until the model has attained efficient training.

PPML TOOLS

PySyft

This is a machine learning toolbox whose raw source code is purely encoded using python and is usually open source. The toolbox offers safety and privacy services while being an Open-Minded program. PySyft is used in the development of artificial intelligent frameworks that tend to respect people's privacy. Examples of privacy techniques that this specific machine learning library supports include Different Privacy and federal learning (Borgia, 2014). Machine learning, deep learning frameworks such as PyTorch and TensorFlow, and APIs such as Kera's are also supported and extended by PySyft.

Machine learning privacy meter

The ML privacy meter function is descriptive enough in data privacy. The library is a python package that asses threats targeting privacy in machine learning models. It does this by using Google's TensorFlow deep learning package.

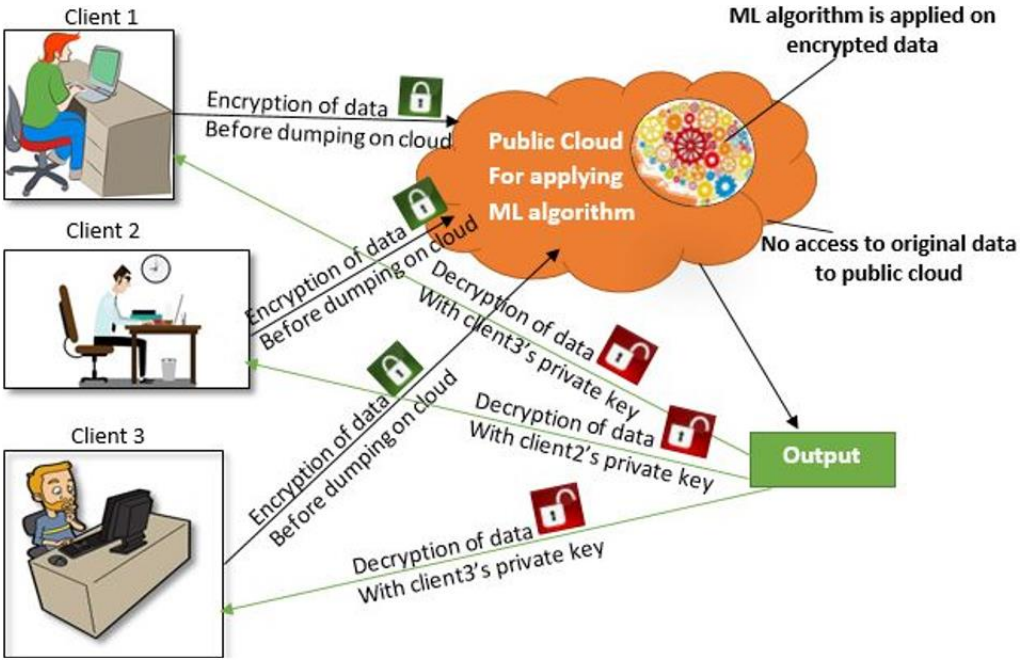

Figure 4. How Machine Learning operates with encryption of data

The tool's operation is based on the development of membership inference attacks, of which the risks associated with privacy are usually calculated later. The calculation of the privacy risk is dependent on the factor of the selected adversary model (Cai et al., 2016). The scores of threats can be used to deploy a hypothesis and resolve a conclusion

concerning the accuracy measure of such attacks. Finally, the model has the visualization capability where the results of privacy risks are usually displayed (Pasupuleti, 2015).

TensorFlow Privacy

Just by the name, the library's function is oriented toward training and the generation of machine learning models that are differentiable and private. The framework of the model is built using google TensorFlow. Its privacy-preserving model is built using differential private SDG. This is usually used to compare machine learning models about privacy provision of utility loss during model decision-making (Fadziso et al., 2018).

DISCUSSION

The PPML methods, such as differential privacy and federated learning, have the same goal. They all address the challenge related to data privacy that aims to bridge and narrow the gap between receiving benefits from machine learning and, at the same time, ensuring data privacy. Though the use of the models may pose a vulnerability by introducing training data to the online platform, thus availing it for cyber assaults, the service and deployment of machine learning in privacy preservation have bearded fruit to the cyber security field (Achar, 2015). In addition, the choice of algorithm model to use depends on the data's relevance and the data used to train the model. Finally, quality data improves the model's efficiency as it will be introduced using a larger dataset. The development of the algorithms is usually similar to that of neural networks. However, since the models use libraries and frameworks such as TensorFlow and PyTorch, the algorithm is developed using the basic neural network structure with related weights and bias configurations. The first part will consist of the input part, where data will be entered, and the last part will be the data output layer. In between will consists of several hidden layers that will perform calculations of related weights and biases for a single network node.

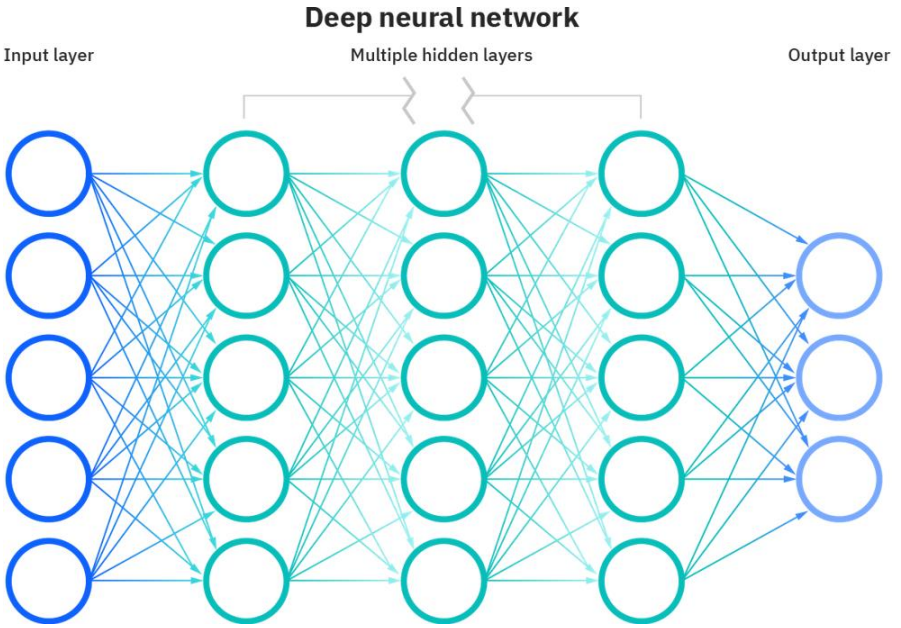

Figure 5. Deep Neural Network Framework

If Z is considered the output, then a neural network model will be viewed as a linear equation with parameter variables whose summation of products of weights and bias add up to result in a single output at each node.

$Z = Bias + W_1X_1 + W_2X_2 + \ldots + W_nX_n$

For the model to function well, the training data must be vast and trained repetitively through several epochs to minimize the cost function that will increase the accuracy of the machine learning models.

CONCLUSIONS

Including machine learning in privacy preservation will help ensure data protection and aid in using machine learning techniques to be affiliated with a central authority. In addition, data will be safe if only the AI models are trained frequently and updated frequently to better and more efficient models that will be difficult for cyber adversaries to penetrate.

REFERENCES

Achar, S. (2015). Requirement of Cloud Analytics and Distributed Cloud Computing: An Initial Overview. *International Journal of Reciprocal Symmetry and Physical Sciences*, *2*, 12–18. Retrieved from https://upright.pub/index.php/ijrsps/article/view/70

Achar, S. (2016). Software as a Service (SaaS) as Cloud Computing: Security and Risk vs. Technological Complexity. *Engineering International*, *4*(2), 79-88. https://doi.org/10.18034/ei.v4i2.633

Achar, S. (2017). Asthma Patients' Cloud-Based Health Tracking and Monitoring System in Designed Flashpoint. *Malaysian Journal of Medical and Biological Research*, *4*(2), 159-166. https://doi.org/10.18034/mjmbr.v4i2.648

Adusumalli, H. P. (2016). How Big Data is Driving Digital Transformation?. *ABC Journal of Advanced Research*, *5*(2), 131-138. https://doi.org/10.18034/abcjar.v5i2.616

Borgia, E. (2014). The Internet of Things vision: Key features, applications and open issues. *Comput. Commun., 54*, 1–31.

Cai, Y., Dai, D., & Hua, S. (2016). *Using machine learning algorithms to improve the prediction accuracy in disease identification: An empirical example*. Athens: The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp). Retrieved from https://search.proquest.com/docview/1806429009?accountid=35493

Cassel, C. K. J. J. (2012). Retail clinics and drugstore medicine. 307(20), 2151-2152.

Chen, M., Hao, Y., Hwang, K., Wang, L., & Wang, L. J. I. A. (2017). Disease prediction by machine learning over big data from healthcare communities. 5, 8869-8879.

Dehury, C. K., Sahoo, P. K. (2016). Design and implementation of a novel service management framework for IoT devices in cloud. *J. Syst. Softw., 119*, 149–161.

Fadziso, T., Adusumalli, H. P., & Pasupuleti, M. B. (2018). Cloud of Things and Interworking IoT Platform: Strategy and Execution Overviews. *Asian Journal of Applied Science and Engineering*, *7*, 85–92. https://upright.pub/index.php/ajase/article/view/63

Iyawa, G. E., Herselman, M., & Botha, A. (2017). *A scoping review of digital health innovation ecosystems in developed and developing countries*. Piscataway: The Institute of Electrical and Electronics Engineers, Inc. (IEEE). Retrieved from https://search.proquest.com/docview/1962316664?accountid=35493

Michalski, R.S., Carbonell, J.G., Mitchell, T.M. (1983). Machine Learning: An Artificial Intelligence Approach. Springer, https://www.springer.com/gp/book/9783662124079

Pasupuleti, M. B. (2015). Data Science: The Sexiest Job in this Century. *International Journal of Reciprocal Symmetry and Physical Sciences*, 2, 8–11. https://upright.pub/index.php/ijrsps/article/view/56

Ray, P. P. (2016). A survey of IoT cloud platforms. *Future Comput. Inform. J., 1*, 35–46.

Truong, H. L., Dustdar, S. (2015). Principles for engineering IoT cloud systems. *IEEE Cloud Comput., 2*, 68–76.

Xia, F., Yang, L.T., Wang, L., Vinel, A. (2012). Internet of things. *Int. J. Commun. Syst., 25*, 1101–1102.

--0--