SCIS 2022 2022 Symposium on
Cryptography and Information Security
Osaka, Japan & Online, Jan. 18 – 21, 2022
The Institute of Electronics,
Information and Communication Engineers

Data Lineage Management with Unlearning Method for Machine Learning Security and Privacy Issues

Haibo ZHANG * Toru NAKAMURA † Takamasa ISOHARA ‡ Kouichi SAKURAI §

Keywords: Data Lineage, Security, Privacy, Machine Learning, Machine Unlearning

Abstract

Privacy protection has been a concern for researchers for a long time. In today's big data environment, users interact with data on various web platforms, such as sending and receiving emails, browsing news, etc., almost every day. For users, once they have provided their information in an application, it is difficult to remove it from the root. When machine learning is widely used today, most advanced features are obtained based on understanding and training users' data. As a result, users' privacy has been spread in every corner of the application, makings it more accessible for attackers to steal users' private data.

Recently, an increasing number of laws have governed the useability of users' privacy. For example, Article 17 of the General Data Protection Regulation (GDPR), the right to be forgotten, requires machine learning applications to remove a portion of data from a dataset and retrain it if the user makes such a request [1, 2]. This maintains the user's right to use their privacy from a privacy protection perspective [3]. From the security perspective, if an attacker compromises the machine learning model by injecting some pollution data into its dataset, it is also necessary to remove such data from the dataset and retrain it [4]. For example, an attacker can open a backdoor in a machine learning model by injecting malicious data into the dataset used for training. The attacker can steal all the private data in the model.

For solving the above problems, it is necessary to retrain the machine learning model. However, the existing retraining methods cause a large amount of computational power and time consumption. Therefore, researchers propose machine unlearning as a more efficient research method [5].

This paper provides an in-depth analysis of machine learning models' security and privacy concerns. Firstly, we illustrate the privacy and security concerns based on the review of related academic and industrial works. Then, we compare the traditional retraining method with machine unlearning methods to solve security and privacy protection issues. Furthermore, we also discuss the future research direction and possibilities in this field.

* Department of Information Science and Technology, Graduate School of Information Science and Electrical Engineering, Kyushu University (E-mail: haiboz0105@gmail.com)
† KDDI Research Inc. (E-mail: tr-nakamura@kddi-research.jp)
‡ KDDI Research Inc. (E-mail: ta-isohara@kddi-research.jp)
§ Department of Information Science and Technology, Faculty of Information Science and Electrical Engineering, Kyushu University (E-mail: sakurai@inf.kyushu-u.ac.jp)

Acknowledgement

This research is partially supported by the Japan Science and Technology Agency (JST) Strategic International Collaborative Research Program (SICORP).

References

[1] S. Schelter, "Towards efficient machine unlearning via incremental view maintenance."

[2] L. Graves, V. Nagisetty, and V. Ganesh, "Amnesiac machine learning," *arXiv preprint arXiv:2010.10981*, 2020.

[3] M. Al-Rubaie and J. M. Chang, "Privacy-preserving machine learning: Threats and solutions," *IEEE Security & Privacy*, vol. 17, no. 2, pp. 49–58, 2019.

[4] N. Baracaldo, B. Chen, H. Ludwig, and J. A. Safavi, "Mitigating poisoning attacks on machine learning models: A data provenance based approach," in *Proceedings of the 10th ACM Workshop on Artificial Intelligence and Security*, 2017, pp. 103–110.

[5] L. Bourtoule, V. Chandrasekaran, C. A. Choquette-Choo, H. Jia, A. Travers, B. Zhang, D. Lie, and N. Papernot, "Machine unlearning," in *2021 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2021, pp. 141–159.