Extended Cryptotree: Training and Inference on Encrypted Data

Marie-Theres Schier, Penelope Mück, Robert Logiewa

Rheinische Friedrich-Wilhelms-Universität, Bonn, Germany

https://crypto.bit.uni-bonn.de/teaching/20ws/lab/

August 20, 2021

Random Forests are among the most popular machine learning models with applications e.g. in healthcare and finance. Especially the financial and medical sector handle very sensitive data. Privacy-preserving machine learning enables secure outsourcing of machine learning tasks to an untrusted service provider while preserving the privacy of the user's data. The approach we present here uses Homomporhic Encryption, which is an encryption scheme that allows data owners to encrypt their data and let a third party perform computations on it, without knowing what is the underlying data. After obtaining a result on encrypted data, this result can be sent back to the data owner who can decrypt the result.
Training a Random Forest is the process of producing an ensemble of decision trees when given a dataset of labeled examples. The goal of training is to obtain a Random Forest that yields accurate predictions on unseen data by aggregating scores from all trees in the forest to produce a prediction. In the following, we will see how Random Forests can be trained on encrypted data and how we can improve the Random Forests' performance by modeling it efficiently as a deep neural network, called Neural Random Forest. We will see that we can adapt this version of a Random Forest into a Homomorphic Random Forest which is able to do quick inference on encrypted data.

crypto day matters 33 (2021)
Beiträge vom 33. Kryptotag, 17.09.2021, online.
doi:10.18420/cdm-2021-33-32

References

[Aka+19] Adi Akavia et al. “Privacy-Preserving Decision Tree Training and Prediction against Malicious Server.” In: *IACR Cryptol. ePrint Arch.* 2019 (2019), p. 1282.

[Son19] Yongsoo Song. *Introduction to CKKS*. Private AI Bootcamp, Microsoft Research Redmond, USA. 2019. URL: https://yongsoosong.github.io/files/slides/intro_to_CKKS.pdf.

[Huy20] Daniel Huynh. “Cryptotree: fast and accurate predictions on encrypted structured data”. In: *arXiv preprint arXiv:2006.08299* (2020).

[SML21] Marie-Theres Schier, Penelope Mück, and Robert Logiewa. *Extended Cryptotree: Training and Inference on Encrypted Data*. 2021. URL: https://crypto.bit.uni-bonn.de/teaching/20ws/lab/.