Privacy-Preserving application using Homomorphic Encryption on NLP algorithms

Jose Contreras (jose.angel.contreras.gedler@ut.ee)

Karlos Taaniel Lillemäe (karlos.taaniel.lillemae@ut.ee)

April 4, 2023

Description

Privacy is critical in various applications, particularly in NLP, where sensitive data is often involved. For instance, medical records contain confidential information NLP can analyze to predict a patient's disease. However, the data should not be shared with the NLP algorithm to maintain privacy. Similarly, businesses may predict customer sentiment through NLP while keeping the data confidential. Homomorphic encryption offers a solution by enabling encrypted data to undergo operations without decryption, thereby allowing NLP algorithms to be performed on encrypted data without compromising privacy. Our project will implement a homomorphic encryption scheme to conduct NLP algorithms on encrypted data, ensuring data privacy.

More formally, we can define HE: Let E be an encryption scheme and K be a key space. Then, E is a homomorphic encryption scheme if there exists a function f such that for all $x, y \in \mathbb{Z}$ and all $k \in K$, we have that $E_k(f(x, y)) = f(E_k(x), E_k(y))$.The function f is called the homomorphic operation.

A similar work can be found here: encrypted_sentiment_analysis. For the implementation, we will use the library concrete-numpy[3]. Furthermore, the privacy-preserving models will be implemented using the library Concrete ML[2].

Data

We will use the IMDB dataset from HuggingFace[1] to train the models. The dataset contains 50,000 reviews from IMDB, labeled as positive or negative. The data is split into 25,000 reviews for training and 25,000 reviews for testing.

Evaluation

One method to evaluate the performance of a homomorphic encryption scheme is to measure the accuracy of the predictions over encrypted data and compare it with the accuracy of the predictions over plaintext data. We can also measure the time it takes to perform the predictions. An assumption is that the time it takes to perform the predictions over encrypted data will be higher than the respective time over plain data, and the accuracy will decrease. However, we expect the accuracy will still be high enough to be useful.

References

[1] Andrew L. Maas, Raymond E. Daly, Peter T. Pham, Dan Huang, Andrew Y. Ng, and Christopher Potts. Learning word vectors for sentiment analysis. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 142–150, Portland, Oregon, USA, June 2011. Association for Computational Linguistics.

[2] Arthur Meyre, Benoit Chevallier-Mames, Jordan Frery, Andrei Stoian, Roman Bredehoft, Luis Montero, and Celia Kherfallah. Concrete-ML: a privacy-preserving machine learning library using fully homomorphic encryption for data scientists, 2022-*. `https://github.com/zama-ai/concrete-ml`.

[3] Arthur Meyre, Benoit Chevallier-Mames, Jordan Frery, Andrei Stoian, Roman Bredehoft, Luis Montero, and Celia Kherfallah. Concrete-numpy:a library to turn programs into their homomorphic equivalent., 2022-*. `https://github.com/zama-ai/concrete-numpy`.