الجمهورية الشعبية الديمقراطية الجزائرية
République Algérienne Démocratique et Populaire
وزارة التعليم العالي و البحث العلمي
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
المدرسة العليا للإعلام الآلي ٠08 ماي 1945٠ بسيدي بلعباس
École Supérieure en Informatique
-08 Mai 1945- Sidi Bel Abbès

Mémoire de Fin d'étude

Pour l'obtention du diplôme d'**ingénieur d'état**
Filière : **Informatique**
Spécialité : **Ingénierie des Systèmes Informatiques (ISI)**

Thème

TenSEAL: A Library for Encrypted Machine Learning Using Homomorphic Encryption

---

Présenté par :

- Ayoub Benaissa
- Bilal Retiat

Soutenu le : **17/09/2020**

Devant le jury composé de :

Mr Abdellatif Rahmoun	Professeur	Président
Mr Mohamed Kamel Faraoun	Professeur	Examinateur
Mr Alaa Eddine Belfedhal	Docteur	Encadreur

Année Universitaire : 2019/2020

Abstract

Machine learning algorithms have achieved remarkable results and are widely applied in a variety of domains. These algorithms often rely on sensitive and private data such as medical or financial records. It is therefore vital to draw further attention regarding privacy threats and corresponding defensive techniques for machine learning.

Research community has proposed a wide range of defensive techniques to preserve data privacy in these systems, one of the promising approach is homomorphic encryption. Thus we revisit existing works that have contributed to reducing the cost of evaluating neural networks on encrypted data, mainly using homomorphic encryption schemes, as well as training neural network on encrypted data.

We describe our implementation of the TenSEAL library and the client-server framework, which can build privacy-preserving machine learning services using homomorphic encryption. Finally, we show empirical results of our libraries' evaluation, and that we can implement a convolutional neural network on the MNIST dataset and achieve 0.5 MB communication cost.

ملخص

حققت خوارزميات التعلم الآلي نتائج باهرة ويتم تطبيقها على نطاق واسع في مجالات متعددة، غالبًا ما تعتمد هذه الخوارزميات على بيانات خاصة و حساسة، مثل: السجلات الطبية أو المالية، لذلك كان من الضروري جذب المزيد من الانتباه فيما يتعلق بتهديدات الخصوصية والتقنيات الدفاعية في مجال التعلم الآلي.

اقترح الباحثون مجموعة واسعة من التقنيات الدفاعية للحفاظ على خصوصية البيانات في هذه الأنظمة، ومن بين أحد أهم الحلول الواعدة هي: التشفير التماثلي، ولذلك أردنا أن نعيد النظر في الأعمال الموجودة التي ساهمت في تقليل تكلفة عملية التصنيف في الشبكات العصبونية على البيانات المشفرة، وخصوصا التي وظفت التشفير التماثلي، بالإضافة إلى عملية تدريب الشبكات العصبونية على البيانات المشفرة.

نحن نصف تنفيذنا لمكتبة TenSEAL وإطار عمل خادم عميل، والذي يمكنه بناء تطبيقات التعلم الآلي التي تحافظ على الخصوصية باستخدام الشفير التماثلي. أخيرًا، نعرض النتائج التجريبية لتقييم مكتباتنا، وأنه يمكننا تنفيذ شبكة عصبونية التفافية على مجموعة بيانات MNIST وتحقيق تكلفة اتصال 0.5 ميجا بايت.

Résumé.

Les algorithmes d'apprentissage automatique ont obtenu des résultats remarquables et sont largement utilisés dans divers domaines. Ces algorithmes dépendent souvent de données privées et sensibles telles que des données médicales ou financières. Il est donc essentiel d'attirer davantage l'attention sur les menaces à la privacy et les techniques de défense correspondante pour l'apprentissage automatique.

La communauté des chercheurs a proposé différentes techniques de défense pour la préservation de la privacy dans ces systèmes, l'une des approches prometteuses étant le chiffrement homomorphe. Nous revisitons donc les travaux existants qui ont contribué à la diminution du cout de calcul lors de l'évaluation des réseaux de neurones sur des données chiffrées, principalement en utilisant le chiffrement homomorphe, ainsi que l'entrainement des réseaux de neurones sur des données chiffrés.

On décrit notre implémentation de la librairie TenSEAL et le framework client-server permettant de construire des services d'apprentissage automatique préservant la privacy, en utilisant le chiffrement homomorphe. Au final, on présente les résultats empiriques après évaluation de notre librairie, et on montre qu'on peut implémenter un réseau neuronal convolutif sur le jeu de données MNIST et atteindre un coût de communication de 0.5 Mo.