SMiLe

Secure Machine Learning Applications with Homomorphically Encrypted Data

Programm / Ausschreibung	IKT der Zukunft, IKT der Zukunft, IKT der Zukunft - 8. Ausschreibung (2019)	**Status**	abgeschlossen
Projektstart	01.04.2021	**Projektende**	31.03.2024
Zeitraum	2021 - 2024	**Projektlaufzeit**	36 Monate
Keywords	homomorphe Verschlüsselung; maschinelles Lernen; Sicherheit; MitarbeiterInnensegmentierung; vorausschauende Wartung		

Projektbeschreibung

Motivation: Der Erfolg von Organisationen hängt zunehmend von ihrer Fähigkeit ab, Daten zur Verbesserung ihrer Effizienz und zur Schaffung innovativer Angebote zu nutzen. In Organisationen werden immer mehr Daten erfasst und gespeichert, ein beachtlicher Teil ihres Potenzials bleibt aber ungenutzt. Vor allem für Anwendungen, die auf Verfahren des maschinellen Lernens setzten, wäre eine Zusammenführung von Daten über Organisationseinheiten und Organisationsgrenzen hinweg wichtig. Besonders sensible Daten werden jedoch kaum geteilt. Dafür gibt es gute Gründe: (1) Gesetze wie die DSGVO verlangen von Organisationen das Ergreifen von Sicherheitsmaßnahmen und schränken die Nutzung und Weitergabe von Daten grundsätzlich ein. (2) Organisationen haben ein starkes Interesse daran geistiges Eigentum zu schützen und sehen es durch das Teilen von Daten gefährdet. Da auf homomorph verschlüsselten Daten zwar beliebige Berechnungen möglich sind, die Daten aber gleichzeitig vor unberechtigtem Zugriff geschützt sind, erscheinen sie im Hinblick auf die Nutzung von sensiblen Daten vielversprechend. Das Fehlen von Know-how und geeigneten Softwarelösungen hat dazu geführt, dass es bisher kaum praktische Anwendungen gibt, die auf maschinelles Lernen auf homomorph verschlüsselten Daten setzen.
Ziele: SMiLe untersucht unter welchen Voraussetzungen Lösungen, bei denen homomorphe Verschlüsselung zum Einsatz kommt, dazu geeignet sind das Potenzial von sensiblen Daten für maschinelles Lernen nutzbar zu machen. Die Weiterentwicklung von relevanten Softwarekomponenten steht genauso im Mittelpunkt des Projekts wie das Zugänglichmachen von erforderlichem Know-how. Das Potenzial des Verfahrens für maschinelles Lernen wird anhand von zwei Anwendungsfällen bewertet, die sich mit Mitarbeitersegmentierung bzw. vorausschauender Wartung befassen. Im Rahmen des Projekts werden sowohl technische als auch soziale, rechtliche und wirtschaftliche Fragen behandelt. Lösungsansätze werden nicht nur hinsichtlich ihrer Analysefähigkeiten, Leistung und Skalierbarkeit, sondern auch im Hinblick auf ihre Kosteneffizienz, Transparenz und Benutzerfreundlichkeit bewertet. Die Vor- und Nachteile des maschinellen Lernens auf homomorph verschlüsselten Daten werden mit alternativen Ansätzen verglichen, die auf synthetischen Daten, Transfer Learning, Secure Multi-Party Computation oder Differential Privacy basieren.
Ergebnisse: SMiLe trägt zur Etablierung eines kooperativ-kreativen Ökosystems bei, in dem verschiedene Akteure vertrauensvoll, symbiotisch und eigenverantwortlich interagieren und bisher nicht vorstellbare Lösungen realisieren, bei denen nicht nur Datenschutz und Sicherheit gewährleistet sind, sondern auch bisher ungenutzte Potenziale von Daten ausgeschöpft werden können. Im Rahmen des Projekts werden bestehende Plattformen erweitert, um Lösungen

bereitzustellen, welche die Erstellung und Nutzung von Modellen für maschinelles Lernen unter Verwendung homomorph verschlüsselter Daten ermöglichen. Bestehende Softwarekomponenten werden weiterentwickelt und in prototypische Lösungen integriert, welche die spezifischen Anforderungen der Anwendungsfälle erfüllen. Die Erkenntnisse werden in Form von praktischen Richtlinien dokumentiert. Die Richtlinien sollen potenziellen Anwendern helfen zu beurteilen, ob maschinelles Lernen auf homomorph verschlüsselten Daten zu ihrem Anwendungsszenario passt.

Abstract

Motivation: Organisations succeed or fail based on their ability to leverage data to improve operational efficiencies and create innovative offerings. Data sources such as organisational databases and software applications are increasingly complemented by sources such as sensors embedded in physical devices and social media. However, much of the potential of large and integrated datasets – especially for machine learning applications – remains unused. If data is considered sensitive, there is a strong reluctance to share that data, both within and between organisations. There are good reasons for this: (1) Legislation such as the GDPR requires organisations to take appropriate security measures and restrict data sharing. (2) Organisations have a strong interest in protecting their intellectual property, and data sharing increases the risk of exposing it. Homomorphic encryption schemes, which allow arbitrary calculations on ciphertexts, promise to eliminate some of the main reasons for the prevailing reluctance to share data. While there have been major recent breakthroughs in fundamental research, due to a lack of both knowledge and software implementations, real-world applications using homomorphic encryption have yet to emerge.
Goals: SMiLe aims to help exploit the potential of sensitive data for machine learning applications through solutions that take advantage of homomorphic encryption. The acquisition of relevant knowledge by organisations is as much the focus of the project as the further development of software implementations both in terms of data platforms and services for machine learning inference and training. The potential of homomorphic encryption for secure machine learning in distributed application scenarios is assessed based on two contrasting use cases dealing with workforce segmentation and predictive maintenance, respectively. Within the scope of the project, technical as well as social, legal and economic questions are addressed. Different solutions are evaluated not only for their analytics capabilities, performance and scalability, but also for their cost efficiency, transparency and ease of use. The advantages and disadvantages of machine learning on homomorphically encrypted data are compared with alternative approaches based on synthetic data, transfer learning, secure multi-party computation and differential privacy.
Results: SMiLe contributes to the establishment of a co-creative ecosystem in which different actors interact in trustful, symbiotic and empowered relationships, and realise previously unattainable solutions, which are not only secure by design but also allow the exploitation of still untapped data potential. The project extends existing data platforms to provide environments that enable the creation and use of machine learning models using homomorphically encrypted data. Basic software implementations are further developed and integrated into prototypical solutions that meet the specific requirements defined for the two use cases. The resulting solutions as well as the generalised lessons learned are documented comprehensively in the form of practical guidelines. It is expected that the guidelines, which are complemented by interviews with project representatives and a video that explains the main concepts in an approachable way, will help potential users assess whether the approach fits their application scenario and, if so, to enter the implementation process in an informed and empowered way.

Projektkoordinator

- Fraunhofer Austria Research GmbH

Projektpartner

- CORE smartwork GmbH
- Software Competence Center Hagenberg GmbH
- Fill Gesellschaft m.b.H.
- VRVis GmbH
- Tributech Solutions GmbH
- MCI Internationale Hochschule GmbH