프라이버시 보호를 위한 동형암호의 필요성

서진범 · 조영복*

대전대학교

The Need for Homomorphic Encryption to Protection Privacy

Jin-Beom Seo · Young-Bok Cho*

Daejeon University

E-mail : tjwlsqja2013@naver.com / ybcho@dju.kr

요 약

2020년 데이터 3법의 개정으로 의료데이터의 개인정보를 가명·익명 처리하여 통계작성, 연구, 공익적 기록 보존 등의 목적으로 사용가능 하도록 하고 있다. 그러나 비식별화 한 데이터를 유전정보, 신용정보 등을 이용하여 재식별이 가능하며, 재식별 정보를 통해 개인 건강정보는 민감정보로 프라이버시 침해에 악용될 여지가 있다. 본 논문에서는 민감정보로 분리되는 개인정보의 프라이버시 보호를 위한 동형암호의 필요성을 도출한다.

ABSTRACT

According to the revision of the Data 3 Act in 2020, personal information of medical data can be processed anonymously for statistical purposes, research, and public interest record keeping. However, unidentified data can be re-identified using genetic information, credit information, etc., and personal health information can be abused as sensitive information. In this paper, we derive the need for homomorphic encryption to protect the privacy of personal information separated by sensitive information.

키워드

딥러닝, 동형암호, 의료데이터, 개인정보

Ⅰ. 서 론

4차 산업 혁명으로 빅데이터, 딥러닝 등의 기술이 주목받고 있으며, 빅데이터 활용에 대한 관심이 증가하고 있다. 또한 데이터 활용을 활성화하기 위해 2020년에 「데이터 3법」을 개정하였으며, 「개인정보보호법」에서는 의료데이터의 개인정보사용 시 안정성을 확보할 수 있는 가명·익명 처리화 기술을 적용한 경우 개인의 동의 없이 통계작성, 연구, 공익적 기록 보존 등의 목적으로 사용가능하도록 개정되었다. 이를 통해 의료데이터를 활용한 의료 산업의 혁신이 진행될 것이라 예상되었으나 가명·익명화된 데이터의 활용 분야가 연구, 기록 보존 등의 영역으로 제한적이다. 또한 「의료법」, 「국민건강보험법」에서 의료정보가 별도로 관리되고 있으며, 가명·익명화된 데이터를 유전 정보, 개인 건강정보등을 통해 재식별화 가능한 문제가 존재한다. 딥러닝의 경우 블랙박스모형으로 외부 공격자가 파라미터값을 얻어내지 않는 이상 데이터를 추론하기 어렵지만 의료데이터의 사용에 문제가 발생하며, 현재 이슈가 되는 연합학습의 경우 화이트박스 모형이며, 클라이언트가 데이터를 안전하게 보관하여 공격을 방지할 수 있으나 파라미터 전송, 공유 등의 행위에서 외부 공격자가 파라미터를 관측하고 데이터를 추론 가능한 문제가 발생한다[1].

본 논문에서는 이러한 문제를 해결하기 위한 방안들을 소개하며, 이를 프라이버시 보호를 위한 동형암호의 필요성을 도출하고자 한다.

Ⅱ. 관련 연구

2-1 개인정보와 민감정보

‘개인정보’는 「개인정보 보호법」 제 2조 정의에 의하면 ‘살아 있는 개인에 관한 정보로서 개인

* corresponding author

을 알아볼 수 있는 정보를 말한다.'로 정의되어 있으며, '민감정보'는 「개인정보 보호법」 제23조 제1항 '개인정보 '사상이나 신념에 관한 정보, 건강 등에 관한 정보, 유전자 검사 등의 결과로 얻은 유전 정보, 신체적, 행동적 특징에 관한 정보 등'로 정의되며, 개인의 사생활을 침해할 유려가 있어 법으로 정한 '개인정보'을 '민감정보'라고 한다. '개인정보'는 「개인정보 보호법」의 개정으로 안전한 가명, 익명처리를 통해 연구, 공익적 기록 보존등의 목적으로 정보를 처리할 수 있으나, 의료정보와 같은 '민감정보'는 데이터 유형 별 적절한 가명, 익명처리 방법과 절차를 통해 활용할 수 있되, 정신, 유전, 희귀 질환등 재식별시 개인 인권에 중대한 피해가 발생할 수 있는 정보에 대해 개인의 동의를 받아야 데이터를 사용 가능한 차이가 있다.

2-2 동형 암호화

동형 암호화는 평문에 대한 연산을 수행한 결과와 암호문에 연산을 수행한 결과와 같은 값을 가지는 암호화 방식이다. 또한 암호문에 연산을 수행한 결과는 암호문을 복호화 가능한 키 소유자만 볼수 있어 기밀성을 보장한다. 동형암호는 연산의 유형, 회수등에 따라 부분 동형 암호, 준 동형암호, 완전동형암호로 나뉠 수 있다. 동형 암호의 개념도는 수식 1과 같다.

$$E(M_1) + E(M_2) = E(M_1 + M_2) \quad (1)$$
$$E(M_1) * E(M_2) = E(M_1 * M_2)$$

2-3 동형암호 기반 딥러닝

Yu 등은 부분 동형암호인 Paillier암호와 CNN을 활용하여 스마트 lem에서의 에너지 도난감지 시스템을 연구하였으며, Paillier암호를 통해 에너지 소비 데이터를 암호화하여 사용자 개별의 데이터노출을 방지하여 개인정보 보호와 데이터 기밀성을 보장하였다. Zhang 등은 산업 IoT 환경의 개인정보 보호 딥러닝 시스템을 연구하였다. 연구는 연합학습을 기반으로 로컬 참여자들의 데이터를 통해 학습을 진행하며, Paillier암호를 사용하여 개인정보를 보호하였다. Zhang 등은 클라우드 환경의 의료진단 시스템을 연구하였다.논문은 클라우드시스템에서 존재하는 환자의 개인정보 유출문제를 해결하기 위해 OU암호 와 SVM을 이용하여 의료진단 시스템을 제안하였다[2].

Ⅲ. 결 론

최근 의료분야에서 인공지능 기술이 주목받고 있으며, 인공지능 학습은 빅데이터를 활용하여 학습이 이루어 지고 있다. 빅데이터의 방대한 데이터를 수집하고 인공지능 학습에 사용함에 있어 다양한 문제점들이 존재하며, 의료 데이터의 경우 「데이터 3법」의 개정으로 연구, 공익적 기록 보존 등의 목적으로 사용가능해졌다. 그러나 가명·익명화된 데이터를 유전, 개인건강정보등을 통해 재 식별화 가능한 문제가 존재하여 의료 데이터를 수집하기 어려운 상황이 발생한다. 이러한 문제를 해결하기 위해 다양한 기술들을 사용하고 있다. 이러한 기술 중 연합학습을 사용한 연구가 진행되고 있다. 연합학습은 중앙서버에서 수집, 학습이 이루어지는 형태가 아닌 클라이언트가 가지고 있는 데이터를 클라이언트측에서 학습하고, 학습 결과를 중앙서버로 전송하고, 각각의 클라이언트는 중앙서버로 전송된 파라미터를 사용하여 학습 결과를 높일 수 있으며, 자신에게 특화된 모델을 설계 할 수 있다. 그러나 중앙서버에 전송되는 파라미터 값을 통해 데이터를 추론할 수 있는 문제가 존재한다. 개인정보를 보호하기 위해 동형암호기반 암호화 데이터를 학습에 사용하기 위한 연구가 진행되고 있다. 동형 암호의 경우 암호문에 연산을 수행한 결과와 평문에 연산을 수행한 결과가 같은 암호화 기법이며, 이를 통해 개인정보를 보호할 수 있다.

본 논문에서는 의료 데이터를 사용함에 있어 개인정보 보호를 위해 사용한 연합학습, 동형암호를 조사하였다. 기존 방식의 학습에서는 의료데이터의 개인 정보보호를 보장 할 수 없으며, 연합학습을 이용하더라도 파라미터값을 통한 데이터의 추론이 가능하다. 동형암호을 사용하여 데이터를 암호화하고 학습을 하는 형태에서는 암호화된 데이터의 연산량이 증가하는 단점이 있지만, 지속적으로 발전해가는 하드웨어의 성능과 Azure 클라우드와 같이 클라우드로 제공되고 있는 서버를 통해 이러한 단점을 보완할 수 있을것이라 예상된다. 이를 통해 의료 데이터를 이용한 인공지능 학습, 의료산업 인프라 확장을 위해서는 동형암호를 통한 암호화 데이터가 필요함을 예상 한다.

Acknowledgement

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2018R1C1B 5083789).

References

[1] J. B. Seo, J. B. Cho, "Image Algorithm using Deep Learning", *The Korea Institute of Information and Communication Engineering*, Korea, Vol. 24, No. 1, pp. 40-42 Jul, 2020

[2] J. H. Oh, M. H. Hyun, J. Y. Lee, Y. H. Park, "Research on Security of Artificial Intelligence

Learning Data based on Homomorphic Encryption", *Proceedings of Symposium of the Korean Institute of communications and Information Sciences*, pp. 622-623, Jun, 2021