정 보 보 호 학 회 지
제27권 제3호, 2017. 6

프라이버시 보존 분류 방법 동향 분석

김 평*, 문 수 빈**, 조 은 지**, 이 윤 호***

요 약

기계 학습(machine-learning) 분야의 분류 알고리즘(classification algorithms)은 의료 진단, 유전자 정보 해석, 스팸 탐지, 얼굴 인식 및 신용 평가와 같은 다양한 응용 서비스에서 사용되고 있다. 이와 같은 응용 서비스에서의 분류 알고리즘은 사용자의 민감한 정보를 포함하는 데이터를 이용하여 학습을 수행하는 경우가 많으며, 분류 결과도 사용자의 프라이버시와 연관된 경우가 많다. 따라서 학습에 필요한 데이터의 소유자, 응용 서비스 사용자, 그리고 서비스 제공자가 서로 다른 보안 도메인에 존재할 경우, 프라이버시 보호 문제가 발생할 수 있다. 본 논문에서는 이러한 문제를 해결하면서도 분류 서비스를 제공할 수 있도록 도와주는 프라이버시 보존 분류 프로토콜(privacy-preserving classification protocol: PPCP) 에 대해 소개한다. 구체적으로 PPCP의 프라이버시 보호 요구사항을 분석하고, 기존의 연구들이 프라이버시 보호를 위해 사용하는 암호학적 기본 도구(cryptographic primitive)들에 대해 소개한다. 최종적으로 그러한 암호학적 기본 도구를 사용하여 설계된 프라이버시 보존 분류 프로토콜에 대한 기존 연구들을 소개하고 분석한다.

Ⅰ. 서 론

폭발적인 계산 능력의 발전은 그것을 이용한 최신 기계 학습 방법들의 발명과 함께 기계 학습의 적용 분야를 매우 광범위하게 만들었다. 예를 들어 스팸 분류, 의료 진단, 신용 평가 등과 같은 분야에서 기계 학습은 학습을 위해 제공된 데이터를 분석하여, 연관된 분야의 의사 결정 수행에 중요한 도움을 줄 수 있다.

그러나 이와 같은 기계 학습의 이점에도 불구하고, 많은 경우, 기계 학습에 사용되는 데이터의 소유자, 기계 학습 서비스 제공자, 그리고 서비스 이용자가 서로 다른 보안 영역에 있는 경우가 많으며 이러한 이유로 보안상의 문제가 발생될 수 있다. 예를 들어, 학습을 위해 필요한 데이터들이 다양한 사용자로부터 생성되며, 이들 각각이 데이터 제공자의 프라이버시와 관련이 있을 수 있다. 또한 제공 서비스의 결과가 사용자의 프라이버시와 연관이 있는 경우도 존재한다. 위의 두 경우의 대표적인 예가 의료 진단 및 신용 평가 서비스이다. 학습에 사용되는 환자의 의료 정보 및 개인의 신용 정보는 해당 정보의 소유자의 프라이버시와 연관이 있으며, 이러한 분야에서 기계 학습 서비스가 제공할 수 있는 사용자의 질병 정보 및 사용자의 신용 정보 분류 결과는 타인에게 노출되면 안되는 주요 프라이버시 보호 정보이다.

위와는 별도로, 기계 학습 서비스를 제공하기 위해 사용되는 정보가 금전적 가치를 갖고 있어, 해당 정보의 소유자가 서비스 제공자가 해당 정보의 무제한적인 사용에 제한을 주고 싶은 경우도 존재한다.

이러한 상황에서 발생할 수 있는 문제점들을 해결할 수 있는 방안으로 프라이버시 보존 기계 학습 방법은 그 중요성이 강조되고 있다.

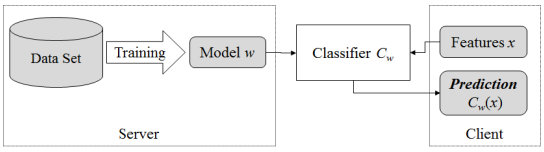
본 논문에서는 프라이버시 보존 기계 학습에서의 한 종류인 프라이버시 보존 분류(Privacy-preserving Classification) 기법에 초점을 둔다. 데이터 분류에서는 이미 분류 결과가 기록되어 있는 다양한 샘플 데이터, 즉 학습 데이터를 바탕으로 분류 알고리즘에 사용되는 모델의 수립을 위한 학습 단계와 학습된 모델을 이용하여, 분류를 원하는 주체가 제공한 데이터들을 이용하여

이 논문은 2017년도 정부 (미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(2016R1A4A1011761, 2016R1C1B2011022), 또한 미래창조과학부의 2017년 고용계약형 SW석사과정 지원 사업을 지원받아 수행한 결과임.
* 서울과학기술대학교 ITM 전공 (firimir@gmail.com)
** 서울과학기술대학교 SW분석설계학과 (subanggu@gmail.com,ejjo321@gmail.com)
*** 서울과학기술대학교 ITM 전공, 교신저자(younholee@seoultech.ac.kr)

분류 알고리즘을 수행, 해당 데이터에 대한 분류 결과를 얻는 예측 단계로 나누어진다. 만약 학습 단계의 결과인 모델을 w라 정의하며, 사용자 입력에 해당하는 분류를 원하는 데이터를 x, 분류 알고리즘(classifier)을 C_w, 분류 결과를 $C_w(x)$라고 정의한다면, 아래의 그림 1과 같은 관계가 될 것이다. 그림 1은 분류 서비스를 제공하는 클라이언트-서버 모델의 개요이다. 그림 1에서 음영이 들어간 상자는 분류 과정에서 프라이버시 보호를 위해 상호간에 노출되어서는 안 되는 정보를 의미한다. 가령, 의료 진단 서비스를 받기 위한 사용자 입력 x와 진단 결과 $C_w(x)$의 경우 사용자의 의료 정보와 발생 가능한 질병에 대한 민감한 정보이기 때문에 서버에게 노출되어서는 안 된다. 한편, 사용자의 선호도 평가를 통해 사용자에게 맞춤 정보를 제공하기 위한 서비스에서 w는 기업의 영업 비밀에 해당되는 정보가 되며 이를 사용자에게 공개하는 것은 기업의 영업 비밀 노출을 야기하여 기업에 대한 금전적인 손해를 입힐 수 있다.

분류 시스템을 이용하여 다양한 민감 정보를 다루는 응용 서비스에 대한 분석은 이전의 연구들에서 이루어졌으며 [1]-[3], 프라이버시 보존 기능을 제공하는 기계 학습 방법들에 대하여 조사한 연구[4] 또한 존재한다. 기존의 프라이버시 보호 기반 기계 학습 연구들은 모델을 추출하는 학습 단계에서 사용되는 데이터의 프라이버시 보호에 초점을 두고 이루어져 왔으며, 사용자 입력 및 예측 결과에 대한 프라이버시의 고려에 미비점이 존재해왔다. 분류 기술에서 프라이버시 문제의 구체적인 정의는 R. Bost 등의 연구[5]에서 처음 이루어졌으며, 본 논문은 이와 같은 프라이버시 요건을 기준으로 분류 알고리즘과 이를 위한 암호기법적인 기반 기술에 대해 조사 분석한다.

향후 논문의 구성은 다음과 같다. II절에서는 프라이버시 보존 분류 알고리즘을 위해 사용될 수 있는 다양한 암호기법적 도구에 대해 다룬다. III절에서는 프라이버시 보존 기능을 제공하는 분류 방법들에 대해 다룬다. IV절에서는 II, III절의 내용을 바탕으로 기존의 프라이버시 보존 기능 제공 분류 방법들을 비교 분석하고 V절에서 결론을 맺는다.



〔그림 1〕 프라이버시 보존 분류 시스템 개요

Ⅱ. 암호기법적 기본 도구

본 절에서는 프라이버시 보호 기반 분류 기술을 구현하기 위해 사용되는 암호기법적인 도구에 대해 정리한다. 데이터 프라이버시를 보호를 위해 데이터를 암호화했을 때, 분류 알고리즘을 수행하기 위해 필요한 암호문들에 대해 연산 방법들을 제공하는 기술들이 이에 해당한다.

2.1. 안전한 양자간 연산(Secure Two-Party Computation)

안전한 양자간 연산의 목표는 상대방과 입력 값을 공유하지 않고 두 당사자가 입력 내용에 대한 임의의 함수를 공동으로 계산할 수 있도록 하는 것이다. 프라이버시 보존 분류 알고리즘 또한 서버는 모델 w를, 또한 클라이언트는 x를 공개하지 않고 분류 알고리즘 C_w을 수행한다는 측면에서 안전한 양자 간 연산에 포함된다고 볼 수 있다.

안전한 양자간 연산에 관한 연구들[6]-[8]에서는 수행을 원하는 함수 f 를 2개의 입력을 받는 게이트를 기반으로 하는 가블드 회로(Garbled Circuit) 형태로 구현한다. 이와 같은 회로의 수행 결과는 2.5 세부절에서 설명할 OT(Oblivious Transfer)를 사용하여 양자간에 공유된다.

안전한 양자간 연산은 프라이버시 보존 분류 알고리즘의 가장 직관적인 해결 방안이 될 수 있지만 가블드 회로를 구성하고 이에 대한 실행 값을 구하는 과정에서 매우 많은 시스템 자원이 요구되기 때문에 비효율적이다. 이에 대한 내용은 3절에서 다시 설명한다.

2.2. (비완전)동형암호와 안전한 양자간 비교 연산 프로토콜

동형암호는 암호문에 대해 특정 유형의 연산을 수행할 수 있도록 한다. 연산의 수행 결과 역시 암호문으로 주어지며, 결과를 담고 있는 암호문은 복호화를 했을 때, 평문에 해당 연산을 수행한 결과와 동일한 값을 갖는다.

이와 같은 동형암호의 특징은 프라이버시 보존 분류 알고리즘의 구현에 있어서 매우 유용하다. 대표적인 동형암호는 가산 연산(+)을 지원하는 Paillier 암호(E_P)[9]와 XOR 연산(⊕)을 지원하는 Goldwa- sser-Micali 암호(E_{GM})[10]이다. (1), (2)는 각 암호가 보존하는 연산에 대한 설명이다.

$$E_P(a)+E_P(b)=E_P(a+b) \quad (1)$$

$$E_{GM}(a)\oplus E_{GM}(b)=E_{GM}(a\oplus b) \quad (2)$$

이러한 비완전동형암호를 사용하여 양자간 안전한 비교연산을 제공해주는 프로토콜로써 DGK 프로토콜이 존재한다[11]. DGK 프로토콜에서는 가산 연산을 지원하는 DGK 암호와 Paillier 암호를 사용된다. 그 중, DGK 암호는 작은 평문 공간에 대해 효율적인 암호이며, 비트 연산으로 정의되는 비교 연산의 효율적인 수행을 위해 사용된다. DGK 암호에서의 비교연산에서는 다음의 사실을 사용한다. l-비트의 a=a1a2...al 와 b=b1b2...bl를 비교할 때 (a1과 b1이 최상위 비트), $z_i=a_i-b_i+1+3\sum_{j<i}(a_j\oplus b_j)=0$을 만족하는 i 가 존재한다면 a<b 가 되고, b<a의 경우에는 $z_i=a_i-b_i-1+3\sum_{j<i}(a_j\oplus b_j)=0$을 만족하는 i가 존재한다. DGK 암호에서는 이러한 암호화된 a, b 가주어졌을 때, 암호화된 z_i값들을 계산할 수 있다.

DGK 프로토콜은 다음과 같이 동작한다. 참여자 A는 Paillier 방법으로 암호화된 (a, b)를 가지고 있고 참여자 B는 두 가지 (비완전)동형암호의 비밀키를 가지고 있는 경우에 프로토콜이 실행되며 프로토콜의 실행 결과 A는 어떤 값이 큰지 여부에 대한 정보를 갖고 B는 값에 대한 어떠한 정보도 갖지 않는다. 프로토콜의 실제 수행은 다음과 같다.

1) A는 비교 과정에서 *a*, *b*의 노출을 방지하기 위해 임의의 *r*을 선택하여 *a'*=*r*, *b'*=*a*-*b*+*r*을 계산하고 *b'*를 B에게 전송한다.
2) B는 *b'*를 복호화하고 DGK 암호를 이용하여 개별 비트에 대해 암호화 후, A에게 반환한다.
3) A는 *a'*를 DGK 암호로 개별 비트를 암호화하고, B에게 비교 결과를 숨기기 위해 *a'*<*b'* 또는 *b'*<*a'* 중에서 하나를 선택하여 그 조건에 맞추어 계산된 암호화된 z_i 전부 $(i=1,..,l)$를 B에게 전송한다.
4) B는 A에게 받은 연산 값을 복호화하고 0이 있을 경우 1을 암호화해서 반환하고 아닐 경우 0을 암호화해서 반환한다.
5) A는 B에게 전송받은 결과를 통해 최종적으로 암호화된 (a, b)의 비교 결과를 구할 수 있다.

2.3. 완전동형암호(Fully Homomorphic Encryption)

비완전동형암호가 한 종류의 연산 또는 제한된 형태의 연산에 대해서 암호문 간이더라도 가능하게 지원한다면 완전동형암호는 컴퓨터에서 가능한 모든 연산을 암호문 상에서 수행할 수 있도록 해준다. 완전동형암호는 2009년 C. Gentry가 처음 제안하였다[12]. Gentry가 제안한 완전동형암호(E_{FHE})의 기본 원리는 (3), (4)와 같은 암호문의 XOR 연산과 AND 연산(∘)을 제한 없이 지원하는 것이다.

$$E_{FHE}(a)\circ E_{FHE}(b)=E_{FHE}(a\circ b) \quad (3)$$

$$E_{FHE}(a)\oplus E_{FHE}(b)=E_{FHE}(a\oplus b) \quad (4)$$

XOR 연산과 AND 연산은 튜링 완전성(Turing-Completeness)을 만족하기 때문에, 이 연산들을 제한 없이 사용해 범용계산성(Universal Computability)을 달성할 수 있다.

그러나 이와 같은 범용계산능력을 갖춘 완전동형암호의 장점에도 불구하고, 성능적인 측면의 문제는 완전동형암호의 활용에 제약이 되고 있다. 완전동형암호는 매우 복잡한 대수적인 연산을 포함하며 적절한 안전도를 달성하기 위해 같은 기반문제를 갖는 일반적인 암호에 비해 큰 수들을 사용하는 이유로 연산 오버헤드가 크다. 그리고 현존하는 주요 완전동형암호 구현은 LWE (Learning With Errors) 문제에 기반한는 관계로 암호문은 노이즈를 포함하는 형태이다. 따라서 암호문 간의 연산이 반복될 경우 암호문 내부의 노이즈가 증가한다. 연산의 정확성을 위해 암호문 내부의 노이즈가 일정 수준에 도달하면 노이즈를 낮춰주기 위해 높은 연산 오버헤드를 요구하는 부트스트래핑(Bootstrapping) 연산을 수행해야 한다. 이와 같은 문제로 인해, 완전동형암호를 사용하는 응용은 이러한 완전동형암호 알고리즘의 특징들을 고려한 성능최적화가 필요하다.

2.4. 편의성 기반 암호기법(Commodity-based Cryptography)과 곱셈 연산

편의성 기반 암호 도구는 프로토콜 실행 이전의 설정 단계에서 상관관계가 있는 데이터(correlated data)를 프로토콜 참여자에게 사전 분배를 통해 안전한 연산을 수행한다[13]. 편의성 기반 암호 도구의 가장 큰 장점은 일반적으로 다른 기술에서는 매우 큰 오버헤드를 가지는 곱셈연산을 효율적으로 달성할 수 있다는 점이다.

편의성 기반 암호 도구에서 프로토콜의 참여자 A, B와 데이터 (a, b)를 가정할 때, 식 (5)를 만족하는 (a_A, b_A)를 A가 가지고 (a_B, b_B)를 B가 갖는 방식으로 값을 공유하며, 가산 연산은 식 (6)과 같이 수행된다.

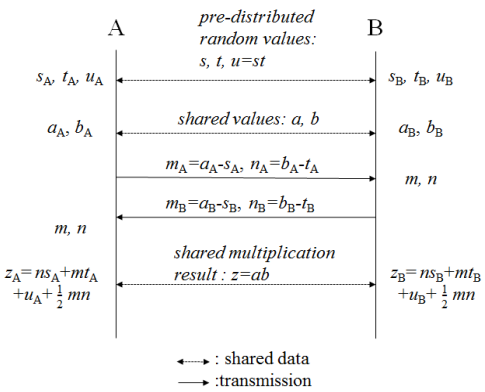
$$a = a_A + a_B, b = b_A + b_B \quad (5)$$

$$(a+b)_A = (a_A + b_A), (a+b)_B = (a_B + b_B) \quad (6)$$

(a, b)의 곱셈 연산은 다음 그림 2와 같이 수행된다.

그림 2에서 s, t, u 값이 설정 단계에서 사전에 공유되는 상관관계가 있는 데이터가 되며, A와 B는 각각 m_A, n_A와 m_B, n_B를 계산하여 공유하는 것만으로 곱셈결과의 공유가 가능하다. 그 과정에서 연산에 필요한 데이터(m, n)을 공유하는 과정과 결과 취합을 위한 추가적인 교신이 필요하다.

이외의 곱셈 연산의 수행은 (7)과 같이 Paillier 암호에서 암호문의 지수 승을 이용하거나 별도의 프로토콜을 사용한다.



〔그림 2〕 편의성 기반 암호기법의 안전한 곱셈 연산 프로토콜

$$E_P(a)^b = E_P(ab) \quad (7)$$

2.5. Oblivious Transfer(OT)

OT 프로토콜은 송신자가 잠재적으로 많은 데이터 중 하나를 수신자에게 전송할 때, 어떤 데이터가 수신자에게 전송되었는지 송신자가 알 수 없도록 하는 프로토콜의 유형이다. 1-out-of-n OT[14]의 경우 송신자가 $(m_1, m_2, ..., m_n)$ 중 하나를 수신자에게 전달하려고 할 때, 수신자는 선택을 위한 $i \in [n]$를 가진다. 프로토콜의 종료 시점에서 수신자는 m_i를 획득할 수 있지만 송신자는 이에 대한 어떠한 정보도 얻을 수 없다.

프라이버시 보존 분류 알고리즘에서 이와 같은 OT와 DGK 프로토콜이 함께 사용될 경우, 수신자는 암호문에 숨겨진 비교 연산 결과에 따라 값의 선택이 가능하다.

Ⅲ. 프라이버시 보존 분류 방법

본 절에서는 이전 절에서 설명한 암호학적 도구를 사용하여 제안된 프라이버시 보존 분류 방법들에 대해서 다룬다.

3.1. 안전한 양자간 연산을 사용한 분류 방법

안전한 양자간 연산을 직접적으로 사용하여 분류 방법을 구현할 경우, 매우 많은 시스템 자원을 요구하며 수행 결과의 정확도에서도 문제점이 발생한다. R. Bost 등의 연구[5]에서는 검증을 위해 많이 사용되는 안전한 양자간 연산 프로토콜인 TASTY[7]와 Failplay[8]를 사용하여 Naïve Bayes 분류 알고리즘을 구현하였다. 성능 평가 결과, 3종류의 값만을 가지는 입력 x에 대하여 256GB 이상의 시스템 메모리가 필요하다는 것을 확인하였다. 또한 안전한 양자간 연산의 성능에 대해 분석한 연구[15]에서는 4GB 이내의 제한된 메모리를 사용할 경우 간단한 가산 연산 및 최소값의 인덱스를 찾는 연산이 실패할 확률이 TASTY에서는 71.4%, Failplay는 14.3%로 매우 높은 값을 갖음을 보였다.

이에 대한 해결 방법으로 안전한 양자간 연산에서 사용되는 가블드 회로의 미세조정을 통해 분류 알고리즘에 최적화 시키는 형태의 연구[16]가 존재한다. 해당 연

구에서는 ECG (Electro-Cardio-Gram)의 해석을 위한 브랜칭 프로그램(Branching Program)을 제안하고 있으며, 이 브랜칭 프로그램은 의사결정트리 분류 알고리즘(Decision Tree Classifier)으로 확장 가능하다.

3.2. (비완전)동형암호를 사용한 분류 알고리즘

비완전 동형암호를 이용할 경우, 동형암호는 다른 암호학적 도구들과 복수 교신 형식의 프로토콜(Interactive Protocol) 형식으로 분류 알고리즘을 구성할 때 사용할 수 있다.

R. Bost 등의 연구[5]는 분류 알고리즘에서 많이 사용되는 연산을 안전한 프로토콜의 구성 블록 형태로 설계하고, 이를 사용하여 프라이버시 보존 분류프로토콜을 구현할 수 있음을 Naïve Bayes 분류, Hyper-plane 분류, 의사결정트리 분류 알고리즘에 대해 보였다. 비교 연산은 DGK 프로토콜을 개선한 연구[20]의 방식을 차용하였으며, 곱셈 연산은 (7)의 방법, 그리고 의사결정트리 분류 알고리즘에서 트리 평가(evaluation)는 완전동형암호를 사용하였다.

CDSS (Clinical Decision Support System)분야에서는 X. Liu 등은 사용자 중심의 의료 진단을 위해 Naïve Bayes 분류 알고리즘을 제안하였다[18]. 해당 연구는 비교 연산의 경우, DGK 프로토콜과 비슷한 방법을 사용하지만, 비트 비교의 효율성을 위한 별도의 동형암호를 사용하지 않는다. 그리고 곱셈 연산을 수행하기 위해 Y. Elmehdwi 등의 연구[19]에서 제안된 프로토콜을 사용하였다.

D. Wu et al. 의 연구[20]는 가산 연산을 지원하는 동형암호와 OT의 방법론을 사용하는 의사결정트리 및 난수 숲(Random Forest) 분류 알고리즘의 실현 방법을 제안하였다. 본 방법에서는 DGK 프로토콜에서의 비트 단위 비교를 위한 계산을 OT에서 트리 평가에 대한 값들 중 1개를 선택하기 위한 값으로 사용한다.

3.3. 완전동형암호를 사용한 분류 알고리즘

완전동형암호는 이론적으로 암호문에 대한 모든 연산의 구현이 가능하기 때문에 분류 알고리즘 또한 암호문 간의 연산만을 통해 구현 가능하다. 이와 같은 특성은 클라이언트-서버 모델에서 1 라운드(전체 2회 메시지 교신)을 수행하는 형식의 프로토콜(Non-interactive Protocol)이 가능케 한다. 그러나 완전동형암호의 성능적인 문제로 인해 제한된 형태의 구현만이 이루어지고 있다.

T. Graepel 등의 연구[21] 에서는 부트스트래핑이 필요하지 않는 범위 안에서만 완전동형암호를 활용[1]하여 실질적으로 제한된 연산만이 가능한 상황에서 선형평균(Linear means), Fisher’s 선형판별(Fisher's linear discriminant) 알고리즘을 구현하였다. 본 연구에서는 모델에 대한 학습 단계도 포함하여 다루고 있으며, 연산 과정에서 사용자 입력에 대한 어떠한 값도 서버에 노출되지 않고, 1 라운드의 통신으로 분류가 가능한 특징이 있다. 그러나 가용 연산 능력의 제한으로 인해 분류 결과 값을 결정하기 위한 최종 비교 연산을 복호화 후 사용자에게 직접 하도록 요구하고 있으며, 이를 위해 사용자에게 모델을 사용한 연산 값의 범위와 대응되는 분류 결과에 대한 정보가 제공된다. 이는 모델 정보의 일부가 사용자에게 노출됨을 의미한다.

J. Bos 등의 연구에서는 완전동형암호로 암호화된 환자의 데이터를 기반으로하는 Cox 비례위험회귀(Cox Proportional Hazards Regression)와 선형 회귀(Linear Regression) 분류 알고리즘을 다룬다[22]. 본 방법에서는 환자를 포함한 모든 사람이 모델 정보를 알 수 있으며, T. Graepel 등의 연구[21]와 유사하게 진단 결과 또한 단순한 분류 결과 보다 더 많은 정보를 노출한다.

A. Khedr 등의 연구는 완전동형암호를 사용한 Bayes 스팸 필터와 의사결정트리 분류 알고리즘의 구현 방안에 대해 다루고 있다[23]. 스팸필터의 경우 단어 검사에 초점을 맞추고 있으며 의사결정트리 분류 알고리즘은 R. Bost 등의 연구[5]에서 완전동형암호를 사용하는 트리 평가 방법과 동일하다. 그리고 아쉽게도 구현된 분류 알고리즘의 성능 평가가 연구에 포함되지 않았다.

W. Lu 등의 연구[24]는 통계분석을 위한 연산들을 데이터가 완전동형암호문인 상황에서도 계산할 수 있도록 하는 완전동형암호 기반 연산 회로들 제안하고 있다. 통계 분석 기법 중 선형 회귀 분석 알고리즘을 분류에 사용할 경우 이 역시, J. Bos 등의 연구[22]의 선형 회귀 분류 알고리즘과 동일한 문제가 발생한다. 단, 제안된 연산 회로를 통해 보완이 가능하지만, 이에 대해 다

1) SWHE (Somewhat Homomorphic Encryption) 이라고 부른다.

루고 있지 않다. 그리고 분류 알고리즘에서 가장 중요한 연산 중 하나인 비교 연산의 구성이 최대값이 2^{20} 미만의 작은 사이즈의 데이터에 적합하며, 최대값이 2^{32} 정도인 데이터를 사용할 경우 급격하게 성능이 저하되는 단점이 있다.

3.4. 편의성 기반 암호기법을 사용한 분류 알고리즘

B. David 등은 편의성 기반 암호기법을 사용하여 Naïve Bayes 분류 알고리즘과 Hyper-plane 분류 알고리즘을 구현하였다[25]. 해당 연구에서는 DGK 프로토콜을 편의성 기반 암호기법으로 구현하였다. DGK 프로토콜에서는 2가지 비완전동형암호가 요구되지만, 편의성 기반 암호기법은 프로토콜에서 참여자들의 모든 데이터는 식 (5)와 같이 안전하게 공유된 형태이기 때문에 별도의 암호학적인 도구가 필요하지 않다. 이와는 별도로, 후속 연구에서는 프로토콜의 최적화 및 기능의 추가를 통해 의사결정트리 분류 알고리즘, SVM(Support Vector Machine), 로지스틱 회귀(Logistic Regression) 분류 알고리즘을 제안하였다[29].

Ⅳ. 비교 연구

본 절에서는 III절에서 기술한 다양한 프라이버시 보존 분류 방법들을 정리하고 비교한다. 표 1은 사용하는 프라이버시 보존 분류 방법과 사용하는 암호학적 기본 도구에 대한 비교이다.

표 2는 III절에서 다루는 프라이버시 분류 방법들에서 노출될 수 있는 정보에 대한 내용이다. 표 2에서 의사결정트리 분류 알고리즘의 경우, D. Wu 등의 연구[20]는 모델의 깊이(depth)를 공개하고 있다. 그리고 복수 교신 형식의 프로토콜 구성의 분류 알고리즘은 클라이언트와 서버가 프로토콜 실행 중에 온라인 상태를 유지해야하므로 수행 시간 측정을 통해 트리의 깊이를 측정하는 부채널공격이 가능해진다. 의사결정트리 분류 알고리즘의 복잡도가 트리의 깊이에 의존하기 때문에

〔표 1〕 프라이버시 보존 분류 알고리즘 비교

	Classification Algorithm	Used cryptographic primitives					1-round Protocol
		STC	HE	FHE	CbC	OT	
M. Barni et al.[16]	Decision tree classifier	√				√	X
R. Bost et al.[5]	Hyper-plane classifier		√				X
	Naïve Bayes classifier		√				
	Decision tree classifier		√	√			
X. Liu et al.[18]	Naïve Bayes classifier		√				X
D. Wu et al.[20]	Decision tree classifier		√			√	O
	Random forest classifier		√			√	X
T. Graepel et al.[21]	Linear means classifier			√			O
	Fisher's linear discriminant						
J. Bos et al.[22]	Cox proportional hazards regression			√			O
	Linear regression classifier						
W. Lu et al.[24]	Linear regression classifier			√			O
B. David et al.[25]	Naïve Bayes classifier				√	√	X
	Hyper-plane classifier						
M. Cock et al.[26]	Decision tree classifier				√		X
	Support vector machine						
	Logistic regression classifier						

STC : Secure Two-party Computatation, HE : Homomorphic Encryption, FHE : Fully Homomorphic Encryption, CbC : Commodity based Cryptography, OT : Obliveious Transfer)

〔표 2〕 프라이버시 보존 분류 알고리즘의 안전성 비교

	Leaked Information
M. Barni et al.[16]	-
R. Bost et al.[5]	The possibility of the side channel attack for the depth of decision tree
X. Liu et al.[28]	-
D. Wu et al.[20]	The depth of decision tree
T. Graepel et al.[21]	More information about calculated value using model than the final sign
J. Bos et al.[22]	More information about calculated value using model than the final sign
W. Lu et al.[23]	More information about calculated value using model than the final sign
B. David et al.[24]	-
M. Cock et al.[26]	The possibility of the side channel attack for the depth of decision tree

트리의 깊이가 짧을 경우는 모델에 대한 프라이버시의 침해 가능성이 증가한다. 완전동형암호를 사용하는 분류 알고리즘의 경우는 비교 연산의 수행을 하지 않기 때문에 3.3에서 설명한 모델에 대한 추가적인 정보 노출이 발생한다. W. Lu 등의 연구[24]는 비교 연산을 제안하고 있지만, 이를 사용하는 분류 알고리즘에 대해서는 다루고 있지 않다.

Ⅴ. 결 론

분류 알고리즘은 중요한 기계 학습 기술 중 하나이며, 분류 알고리즘에 대한 프라이버시 보호는 실제 사용자 데이터뿐만 아니라 분류의 기준이 되는 모델의 보호를 수행한다는 측면에서 향후 분류 방법을 포함하는 기계 학습 방법들의 사용 활성화를 위해 중요한 역할을 수행할 것으로 예상된다. 본 논문은 프라이버시 보존 분류 방법을 실현하기 위해 활용 가능한 암호학적 기본 도구들과 이를 직접 적용한 프라이버시 보존 분류 방법들에 대한 조사를 수행하고, 다양한 측면에서 비교 분석을 수행하였다. 본 논문의 내용은 향후 본 분야를 공부하고자 하는 사람들에게 많은 도움이 되기를 희망한다.

참 고 문 헌

[1] J. Wiens, J. Guttag, E. Horvitz, "Learning evolving patient risk processes for c. diff colonization," *ICML Workshop on Machine Learning from Clinical Data,* 2012.

[2] A. Singh, J.Guttag, "A comparison of non-symmetric entropy-based classificat- ion trees and support vector machine for cardiovascular risk stratification," *Annual International Conference of the IEEE Engineering in Medicine and Biology Society,* pp. 79-82, 2011.

[3] A. Singh, G. Nadkarni, J. Guttag, E. Bottinger, "Leveraging hierarchy in medical codes for predictive modeling," P*roceedings of the 5th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics*. ACM, pp. 96-103, 2014.

[4] C. Aggarwal, S. Philip, "A general survey of privacy-preserving data mining models and algorithms," *Privacy-preserving data mining*, Springer US, pp. 11-52, 2008.

[5] R. Bost, R. Popa, S. Tu, S. Goldwasser, "Machine Learning Classification over Encrypted Data," *NDSS*, 2015.

[6] A. Yao, "Protocols for secure computations," *Annual Symposium on Foundations of Computer Science*, IEEE, pp. 160-164, 1982.

[7] W. Henecka, A. Sadeghi, T. Schneider, I. Wehrenberg, "TASTY: tool for automating secure two-party computations," *Proceedings of the 17th ACM conference on Computer and communications security*, ACM, pp. 451-462, 2010.

[8] D Malkhi, N Nisan, B Pinkas, Y Sella, "Fairplay-Secure Two-Party Computation System." *USENIX Security Symposium,* 4, 2004.

[9] P. Paillier, "Public-key cryptosystems based on

composite degree residuosity classes," *International Conference on the Theory and Applications of Cryptographic Techniques*. Springer Berlin Heidelberg, pp. 223-238, 1999.

[10] S. Goldwasser, S. Micali. "Probabilistic encryption & how to play mental poker keeping secret all partial information," *Proceedings of the fourteenth annual ACM symposium on Theory of computing,* ACM, pp. 365-377, 1982.

[11] I. Damgard, M. Geisler, M. Kroigard. "A correction to'efficient and secure comparison for on-line auctions'," *International Journal of Applied Cryptography,* 1(4), pp. 323-324, 2009.

[12] C. Gentry, "Fully homomorphic encryption using ideal lattices," *STOC*. Vol. 9. pp. 169-178, 2009.

[13] D. Beaver, "Commodity-based cryptography," *Proceedings of the twenty-ninth annual ACM symposium on Theory of computing*, ACM, pp. 446-455, 1997.

[14] G. Asharov, Y. Lindell, T. Schneider, M. Zohner, "More efficient oblivious transfer and extensions for faster secure computation," *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security,* ACM, pp. 535-548, 2013.

[15] J. Ziegeldorf, J. Metzke, M. Henze, K. Wehrle, "Choose wisely: a comparison of secure two-party computation frameworks," *Security and Privacy Workshops*, IEEE, pp. 198-205, 2015.

[16] M. Barni, P. Failla, R. Lazzeretti, A. Paus, A. Sadeghi, T. Schneider, V. Kolesnikov, "Efficient privacy-preserving classification of ECG signals," *First IEEE International Workshop on Information Forensics and Security*, IEEE, pp. 91-95, 2009.

[17] T. Veugen, "Improving the DGK comparison protocol," *International Workshop on Information Forensics and Security,* IEEE, pp. 49-54, 2012.

[18] X. Liu, R. Lu, J. Ma, L. Chen, B. Qin, "Privacy-preserving patient-centric clinical decision support system on naive Bayesian classification," *IEEE journal of biomedical and health informatics* 20(2), pp. 655-668, 2016.

[19] Y. Elmehdwi, B. Samanthula, W. Jiang, "K-nearest neighbor classification over semantically secure encrypted relational data," *IEEE transactions on Knowledge and data engineering,* 27(5), pp. 1261-1273, 2015.

[20] D. Wu, T. Feng, M. Naehrig, K. Lauter, "Privately evaluating decision trees and random forests," *Proceedings on Privacy Enhancing Technologies,* 4, pp. 335-355, 2016.

[21] T. Graepel, K. Lauter, M. Naehrig, "ML confidential: Machine learning on encrypted data," *International Conference on Information Security and Cryptology,* Springer Berlin Heidelberg, pp. 1-21, 2012.

[22] J. Bos, K. Lauter, M. Naehrig, "Private predictive analysis on encrypted medical data," *Journal of biomedical informatics,* 50, pp. 234-243, 2014.

[23] A. Khedr, G. Gulak, V. Vaikuntanathan, "SHIELD: scalable homomorphic implementation of encrypted data-classifiers," IEEE Transactions on Computers, 65(9), pp. 2848-2858, 2016.

[24] W. Lu, S. Kawasaki, J. Sakuma, "Using Fully Homomorphic Encryption for Statistical Analysis of Categorical, Ordinal and Numerical Data," 2017.

[25] B. David, R. Dowsley, R. Katti, A. Nascimento, "Efficient unconditionally secure comparison and privacy preserving machine learning classification protocols," *International Conference on Provable Security*. Springer International Publishing, pp. 354-367, 2015.

[26] M. Cock, R. Dowsley, C. Horst, R. Katti, A. Nascimento, W. Poon, S. Truex, "Efficient and Private Scoring of Decision Trees, Support Vector Machines and Logistic Regression Models based on Pre-Computation," *IEEE*

Transactions on Dependable and Secure Computing, DOI: 10.1109/TDSC.2017.2679189, 2017.

〈저 자 소 개〉

김 평(Pyung Kim)
정회원

2007년 2월 : KAIST 전산학과 학사
2009년 8월 : KAIST 전산학과 석사
2016년 8월 : KAIST 전산학과 박사
2016년 9월 ~ 현재 : 서울과학기술대 ITM 전공 박사후 연구원
관심분야 : 응용암호, 정보보호

문 수 빈(Su-Bin Moon)
정회원

2016년 2월 : 서울과학기술대학교 ITM 전공 졸업
2016년 3월~현재 : 서울과학기술대학교 SW분석설계학과 석사과정
관심분야 : 데이터보안,시스템보안

조 은 지(Eun-Ji Jo)
정회원

2016년 2월 : 부산대학교 정보컴퓨터공학부 졸업
2016년 3월~현재 : 서울과학기술대학교 SW분석설계학과 석사과정
관심분야 : 네트워크, 보안

이 윤 호(Younho Lee)
종신회원

2006년 8월 : KAIST 전산학과 박사
2007년 10월~2009년 2월 : Georgia-Tech Information Security Center 방문 박사후과정
2013년 9월~현재 : 서울과학기술대학교 ITM 전공 부교수
관심분야 : 응용 암호, 데이터 보안