Journal of The Korea Institute of Information Security & Cryptology
VOL.32, NO.2, Apr. 2022
ISSN 1598-3986(Print)
ISSN 2288-2715(Online)
https://doi.org/10.13089/JKIISC.2022.32.2.405

금융데이터의 성능 비교를 통한 연합학습 기법의 효용성 분석*

장 진 혁,[1†] 안 윤 수,[2] 최 대 선[3‡]
[1,2,3]숭실대학교 (연구원, 대학원생, 교수)

Utility Analysis of Federated Learning Techniques through Comparison of Financial Data Performance*

Jinhyeok Jang[1†], Yoonsoo An[2], Daeseon Choi[3‡]
[1,2,3]Soongsil University (Researcher, Graduate student, Professor)

요 약

AI기술은 데이터 기반의 기계학습을 이용하여 삶의 질을 높여주고 있다. 기계학습을 이용시, 분산된 데이터를 전송해 한곳에 모으는 작업은 프라이버시 침해가 발생할 위험성이 있어 비식별화 과정을 거친다. 비식별화 데이터는 정보의 손상, 누락이 있어 기계학습과정의 성능을 저하시키며 전처리과정을 복잡하게한다. 이에 구글이 2017년에 데이터의 비식별화와 데이터를 한 서버로 모으는 과정없이 학습하는 방법인 연합학습을 발표했다. 본 논문은 실제 금융데이터를 이용하여, K익명성, 차분프라이버시 재현데이터의 비식별과정을 거친 데이터의 학습 성능과 연합학습의 성능간의 차이를 비교하여 효용성을 분석하였으며, 이를 통해 연합학습의 우수성을 보여주고자 한다. 실험결과 원본데이터 학습의 정확도는 91% K-익명성을 거친 데이터학습은 k=2일 때 정확도 79%, k=5일 때 76%, k=7일 때 62%, 차분프라이버시를 사용한 데이터학습은 ϵ=2일 때 정확도 52%, ϵ=1일 때 50%, ϵ=0.1일 때 36% 재현데이터는 정확도 82%가 나왔으며 연합학습의 정확도는 86%로 두번째로 높은 성능을 보여 주었다.

ABSTRACT

Current AI technology is improving the quality of life by using machine learning based on data. When using machine learning, transmitting distributed data and collecting it in one place goes through a de-identification process because there is a risk of privacy infringement. De-identification data causes information damage and omission, which degrades the performance of the machine learning process and complicates the preprocessing process. Accordingly, Google announced joint learning in 2016, a method of de-identifying data and learning without the process of collecting data into one server. This paper analyzed the effectiveness by comparing the difference between the learning performance of data that went through the de-identification process of K anonymity and differential privacy reproduction data using actual financial data. As a result of the experiment, the accuracy of original data learning was 79% for k=2, 76% for k=5, 52% for k=7, 50% for ϵ=1, and 82% for ϵ=0.1, and 86% for Federated learning.

Keywords: Federated learning, credit data, K-anonmity, Differential-privacy, synthesis data

Received(02. 09. 2022), Modified(03. 18. 2022), Accepted(03. 18. 2022)

* 본 논문은 2022년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.2021-0-00511, 엣지 AI 보안을 위한 Robust AI 및 분산 공격탐지기술 개발)

* 본 논문은 2022년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. 2021R1A4A1029650)

† 주저자, jjh4002@ssu.ac.kr
‡ 교신저자, sunchoi@ssu.ac.kr(Corresponding author)

I. 서 론

4차산업혁명의 발전으로 음성인식, 이미지 분류, 객체 탐지 등의 기술로 인해 스마트폰, 의료기술 등 사람의 편의성이 확대되고 있다. 이러한 기술처럼 인공지능의 한 분야인 머신러닝은 빅데이터라는 키워드로 현재까지 이슈화되면서 사람들에게 좀 더 이점을 제공하면서 미래사회에 없어서는 안 될 기술이 되고 있다. 하지만 인공지능의 재료인 데이터로부터 개인정보 노출로 인해 피해를 보는 경우도 있다[1,2].이로인해 데이터 비식별화 처리, 차분프라이버시 재현데이터 등의 방법으로 많은 연구가 진행되고 있다[3, 4,5]. 이러한 연구들은 분산된 데이터를 하나의 서버로 모아 학습을 위해 프라이버시가 강한 데이터를 바탕으로 상대방이 알아보지 못하도록 데이터를 변형시키는 기술이다. 하지만 데이터를 변형시키더라도 데이터의 배경을 알고 있는 공격자(배경지식공격), 공개되어있는 정보들을 결합해 개인을 식별하는 연결공격 등의 공격에 취약하므로 데이터프라이버시에 대한 연구가 진행 중에 있다. 이러한 프라이버시 침해에 대한 또 다른 방안으로 구글이 2017년에 처음으로 연합학습(federated learning)을 제안했다. 그 이후로도 알고리즘, 시스템구조 개선과 같이 연합학습에 대한 연구가 많아지고 있으며 MNIST data를 통한 튜토리얼, 오픈소스 제공을 하여 누구나 연합학습을 활용할 수 있다[6].

본 연구는 실무에서 사용하고 있는 기계학습을 바탕으로 데이터를 한 서버로 모아 학습하는 데에서 발생하는 프라이버시 침해를 해결하기 위해, 지금까지 나온 모든 비식별화 기술을 거친 데이터를 동일한 모델로 학습을 진행한 경우와 연합학습을 진행한 경우의 성능 비교를 통해 연합학습의 우수성을 검증할 예정이다. 또한, 현재까지 적용되지 않았던 금융 데이터를 연합학습에 적용한 사례를 보여 그 우수성을 증명할 것이다. 마지막으로, 연합학습의 파라미터별 성능 추세를 보고자 한다. 기존의 익명화 기법에도 다양한 파라미터에 따라 프라이버시와 데이터 유용성이 변동한다. 본 연구의 기여한 바는 다음과 같다.

- 실제 금융데이터를 사용하여 익명화처리 후 여러 비식별모델과 연합학습의 성능 비교
- 연합학습의 파라미터별 성능측정
- 비식별 모델과 연합학습의 성능 차이를 수치적으로 구체화

본 논문의 구성은 다음과 같다. 2장에서는 관련연구로 비식별화에 와 연합학습이 무엇인지와 학습 과정과 관련 연구를 설명한다. 3장에서 실제 신용데이터를 바탕으로 비식별 데이터와 연합학습의 평가 및 결과에 관해 설명하고 4장에서 실험에 대한 고찰과 결론을 맺는다.

II. 관련연구

최근 데이터 3법 개정안이 통과됨으로 인해 민감정보를 가명 처리한 데이터만을 활용할 때 편리하게 사용할 수 있게 되었다. 하지만 2006년 미국에서 발생한 넷플릭스 사례처럼 인터넷에 돌고 있는 정보들을 모으고 분석하다 보면 사생활이라든지, 내가 알리고 싶지 않은 정보들이 노출된다. 따라서 데이터의 민감한 정보를 식별할 수 없도록 마스킹 처리를 한다든지, 삭제하는 등의 비식별기술에 대해 현재까지 연구되고 있다[7,8].

2.1 익명화 가명화

비식별기술은 데이터의 칼럼에 따라 4개의 그룹으로 나누고 그에 따라 보호 기술을 적용한다. 개인을 식별할 수 있는 속성인 식별자(Identifiers)로 주민등록번호, 이름, ID, 이메일, 계좌번호 등이 있으며 비식별 조치 시 무조건 삭제되어야 한다. 준 식별자(Quasi-Identifiers)는 다른 데이터와의 결합을 통해 추론하여 개인을 식별할 수 있는 속성들이며 우편번호, 연간소득, 혈액형, 몸무게 등이 이에 속한다. 민감정보(Sensitive Attributes)는 개인의 사생활을 담긴 속성으로 계좌 수, 잔액, 투자금액, 병명 등이 이에 해당하고 사생활과 거리가 먼 속성을 Insensitive Attributes로 그룹을 나누고 얼마나 할 것인가에 따라 혹은 어느 방법을 사용할 것인가에 따라 비식별기술들이 매우 다양하다. Table 1은 비식별 조치 방법을 보여 준다.

2.2 프라이버시 보호모델

현재 실무에서는 앞서 언급한 비식별 처리기법만을 사용하진 않는다. 비식별처리를 기반으로 프라이버시모델을 적용하더라도 연결공격, 배경지식공격 등의 공격으로부터 안전하지 않은 부분이 생기기 때문이다. 그에 따라 데이터를 더 안전하게 지키고자 하

Tabel 1. Method de-identify

Processing Technique	Example	Skill
Pseudonymisation	Mike, 45years old, lives in Seoul, Hankook University →Jack, age of 40s , lives in Seoul, Universal University	Heuristic Pseudonymization Encryption Swapping
Aggregation	Mike: 180cm Jack: 180cm, Lily : 160cm, Jenny : 150cm →Students' height sum:670cm, heigt mean: 165cm	Aggregation Micro Aggregation Rounding Rearrangement
Data reduction	Resident Registration number : 901206-1536765 →Born in the 90s, male Date information related to an individual is processed on an annul basis (ex. Acceptance day)	Reducing Identifier Reducing Partial Identifier Reducing Record Reducing Every Identifier
Data Suppression	Mike Miller, 45years old → Miller (Family name), age of 30s~40s	Hiding Random Rounding Controlled Rounding
Data Masking	Mike Miller, 35years old, live in Seoul, Hankook University → ***** Miller,, 3* years old, lives in Se*** , ******* University	Adding Random Noise Black Impute

는 목적의 연구가 많이 진행되고 있다[9,10,11].

2.2.1 k 익명성

k-익명성은 주어진 데이터를 바탕으로 같은 레코드 값이 적어도 K개 이상 존재하도록 하여 다른 데이터로부터 정보를 연결할 수 없도록 하는 방법이며 Latanya Sweeney가 제안하였다[12]. Latanya Sweeney는 K-익명성을 다음과 같이 정의했다. $RT(A_1,A_2,..A_n)$는 데이터 테이블, QI_{RT}는 데이터 테이블의 준 식별자라고 할 때 RT가 K-익명성을 만족하면 $RT[QI_{RT}]$의 레코드값들은 K개의 $RT[QI_{RT}]$가 존재해야 한다는 의미 있다. K-익명성을 적용하면 Fig 1에서 보이는 거 같이 데이터 테이블이 변하게 되는데 같은 레코드를 갖는 데이터가 4개로 나눠진 것을 볼 수 있다. 이는 k가 4일 때를 말하며 K에 증가에 따라 데이터의 식별성이 떨어진다. 이러한 K-익명성은 프라이버시 보호를 위한 기본이 되는 모델로 주로 공개된 데이터에 대한 연결 공격(Linkage Attack)을 방어하기 위한 모델이다. 이 외에 통계적 배경 지식을 이용한 추론공격과 그 공격을 막기위한 방안도 제시되고 있다[13].

Fig 2는 연결 공격에 의료정보데이터와 금융데이터를 연계하여 환자의 신원에 대한 프라이버시 침해를 보여주는 연결 공격의 예이다[14].

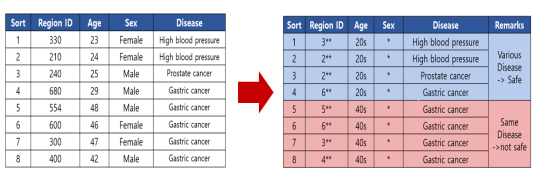

Fig. 1. Medical data with K-anonymity

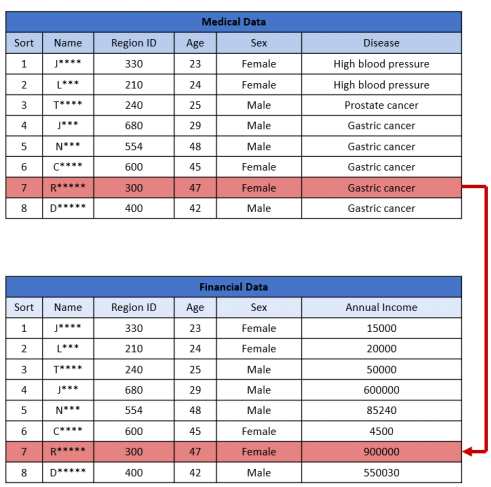

Fig. 2. Linkage Attack

2.2.2 차분 프라이버시

차분프라이버시는 K-익명성의 취약한 부분을 보안하기 위해 Cynthia Dwork가 제안한 수학적 모형이며 한사람에 대한 정보가 데이터 안에 포함되어 있거나 포함하지 않는 두 데이터에 각각 질의(query)를 했을 때, 응답 값을 통해 특정 개인을 식별할 수 없도록 하는 개인정보의 추론을 방지하는 수학적 모델이다[15]. 차분프라이버시는 응답 값에 노이즈를 추가함으로 응답 값의 분포가 일정한 수준 이하로 차이를 갖도록 한다[16]. Fig3은 대화형 차분프라이버시에 대한 그림이며 파라미터로 e(엡실론) 값 을 줄 수 있는데 e값이 작을수록 질의한 응답 값의 차이가 크게 벌어져 데이터 안에 특정 개인의 정보를 추정하기 어렵게 만들어 프라이버시를 보호한다[17]. 실제로 lending club 데이터를 가지고 ARX Tool을 사용하여 e=1일 때 차분프라이버시 기술을 적용한 데이터를 시각화한 자료이며 총 데이터 105,012개 중 61339개가 샘플링되었다. 이 방식을 실무에서 적용하게 된다면 꽤 많은 비율의 데이터가 샘플링 되기 때문에 데이터를 온전히 사용할 수 없게 되어 해당 데이터를 학습시킨 모델의 정확도가 낮아지는 문제가 발생한다.

2.2.3 재현데이터

재현데이터(Synthetic Data)는 사전적 의미로 합성자료, 인위적인 자료이다. 구체적으로 어느 실험을 할 때 시나리오를 위해 가상 자료(재현자료)를 만들어 실험결과를 검증하는데 이용하는 자료로 최근 데이터를 비식별화하는 기술로 연구되고 있다[18]. 재현데이터를 보고 원본데이터를 재식별할 수 없으며 재현데이터의 경우엔 각 칼럼의 평균, 중간값, 최대, 최솟값이 어느 정도 유지된다. 다양한 통계적, 수학적 지식을 필요로 하는 재현자료는 통계적 수치가 매우 유사하기 때문에 연구가치가 매우 높다. Fig 4는 금융데이터를 재현데이터로 바꾸었을 때 통계적으로 얼마나 달라지는 가를 보여 준다[18].

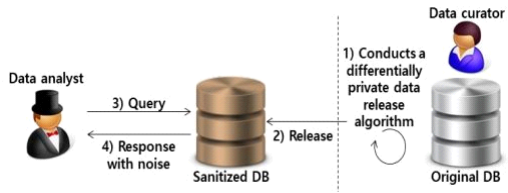

Fig. 3. Interactive differencial privacy environment

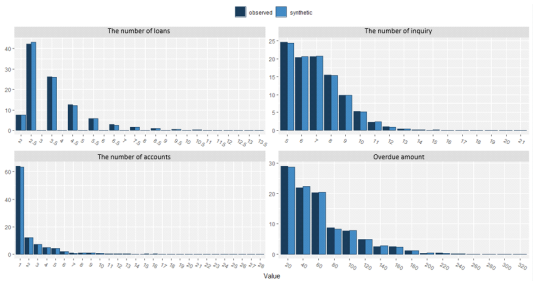

Fig. 4. Statistical numerical comparison between original data and synthetic data

하지만 생성한 재현데이터를 그대로 공개하면서 쓰게되면 인위적으로 재현된 데이터라고 할지라도 원본 데이터가 가진 수리 통계적 특성이 비슷하기 때문에,연결공격, 추론공격들로 [19]인해 결국 프라이버시 침해를 받을 수 있다. 현재는 이러한 공격을 예방하고자 차분프라이버시를 적용하여 재현데이터를 생성하는 방법들 또한 제공된다[20,21].

2.3 연합학습

연합학습은 여러 조직이 분산된 데이터를 공유하지 않고 기계학습을 할 수 있다[6]. 이때 얻어지는 파라미터 값, 가중치 값만을 이용하여 서버의 모델을 학습하는 원리이다. 이러한 연합학습은 각 조직의 데이터를 수집하여 하나의 서버로 모아 대량의 데이터를 가지고 학습을 진행하는 일반적인 학습과는 다르다. 데이터를 공유하지 않는다는 점에서 프라이버시를 보호할 수 있다는 점 과 더욱 민감한 데이터를 가지고 학습을 진행하여 좀 더 새로운 연구를 시도해 볼 수 있다는 점에서 가장 큰 장점을 지니고 있다. 하지만 칼럼을 맞추어야하는 머신러닝 기술의 특성상 서로 다른 칼럼에 대해서 어떤식으로 처리 해야 할지에 대해서는 아직 연구단계이다[22, 23].

2.3.1 연합학습과정

연합학습이란 디바이스 혹은 클라이언트에 저장된 데이터를 상호교환 없이 머신러닝, 딥러닝 모델을 훈련하는 데 사용한다[6]. 구글은 하나의 서버에서 각 디바이스에 응답을 청한 후 모델을 전송하는 형식

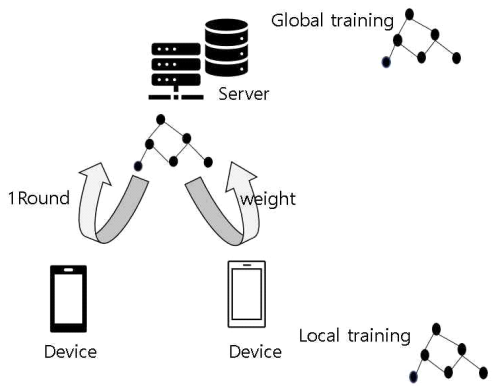

Fig. 5. process of Federated learning

이다[22]. 국내에서는 이를 도입해 각 기관이 보유한 데이터를 연합하여 사용하는 것으로 연구되고 있다. Fig 5는 연합학습의 그림을 보여준다. 하나의 서버로부터 훈련에 참여하는 조직 혹은 디바이스 마다 모델을 전송해주고 디바이스들은 각각의 데이터들을 가지고 로컬학습을 한다. 로컬학습을 진행하면서 업데이트된 값들을 서버로 전달을 해주고 디바이스별로 받았던 업데이트 값들을 평균을 통하여 서버의 모델(글로벌 모델)을 업데이트한다. 그림 8은 연합학습의 흐름을 보여준다. 이때 각 디바이스마다 진행되는 로컬학습은 미니배치 학습을 통해 빠르고 높은 성능을 올릴 수 있다. 구체적으로 디바이스로 부터 서버로 업데이트 값을 1회 전송할 시 1 round, 1 Communicate Cost라고 부른다. 디바이스에서 1회 업데이트 전송마다 글로벌 training을 진행하고 글로벌 모델을 업데이트시킨다. 즉 연합학습은 디바이스와 서버 간의 의사소통이라고 생각하면 될 것이다. 파라미터로 몇 번 의사소통을 진행하는지에 대한 Round, 디바이스 개수인 Client가 있으며 이러한 파라미터에 따라 결과값이 달라진다. Table 2는 연합학습의 Round에 따른 성능을 나타내며 표와 같이 Round가 90에서 91로 증가할 때 성능이 떨어짐을 볼 수 있으며 Round가 높아야만 좋은 것은 아니란 것을 말해준다. 따라서 연합학습의 데이터 분류 성능 부분에서 보여지는것과 같이. 기존의 머신러닝 파라미터를 조절하는 것처럼 연합학습도 파라미터를 조절해서 성능을 조절할 수 있다[24, 25, 26]. 실제 연합학습은 분산된 클라이언트가 가진 데이터를 각 클라이언트의 디바이스에서 모델을 학습시키는 과정을 거친다. 이는 각 클라이언트의 디바이스마다 컴퓨팅 능력이 다르기 때문에 완전히 실제 환경과 일치하는 연합학습을 구현하는데에 어려움이 존재한다. 본 논문은 위 실험에서 사용한 한대의 컴퓨터상으로 구현되었기 때문에, 모든 클라이언트의 디바이스가 동일한 컴퓨팅 능력을 가지고 있는 상황을 가정하게 되므로 실제 환경에서의 연합학습과는 정확도나, 시간 등의 비용에 있어서 다소 차이가 있을 수 있다.

Table 2. Federated learning performance by round

round	accuracy	time
89	0.9778	1048s
90	0.9783	1059s
91	0.9780	1071s

2.3.2 SGD vs AVG

현재 연합학습의 업데이트 과정은 두 가지로 나뉜다. 디바이스에서 받은 업데이트 값을 서버는 이를 합산하여 디바이스 개수만큼 나누며 서버의 모델을 업데이트시키는 알고리즘으로 Federated Stochastic Gradient Descent (FedSGD), 미니배치를 사용하는 Federated Averaging (FedAvg)가 연구되었다[27, 28]. FedSGD는 식 (1), (2)과같이 계산하고 FedAvg는 식 (3), (4)와 같이 계산한다. [27]에 의하면 클라이언트별 골고루 분산된 IID, 클라언트에 복잡하게 섞은 데이터 환경인 Non-IID에서 Fed AVG가 적은 의사소통으로도 99%에 도달할 수 있으며 성능 또한 준수한 것으로 결론이 되었다.

client n개 기준, w: 서버의 가중치, 업데이트 값, g: 디바이스의 가중치, 업데이트 값

$$w_{\neq w} = w_{Old} - Learningrate * g \quad (1)$$

$$g = (g_1 + g_2 + g_3 + ... + g_n)/n \quad (2)$$

$$w_{\neq w} = w_{old} \quad (3)$$

$$w_{\neq w} = (w_1 + w_2 + w_3 + ... + w_n)/n \quad (4)$$

III. 금융데이터 익명화 케이스 스터디

본 실험의 목적은 일반학습과 익명화처리를 한 학습, 연합학습에 대한 비교로 실무에서 활용하기 위함

이다. 예를들어 AI로 신용등급을 매길 때 모델을에 필요한 데이터가 많을수록 정확하다. 그 많은 양의 데이터를 얻기 위해서는 카드사끼리의 데이터를 공유해야 한다. 공유하는 과정에서 비식별처리를 하지만 데이터가 뭉개지거나 각각의 칼럼들이 달라 활용이 어렵다. 연합학습을 활용하면 카드사의 많은 양의 비식별 처리한 데이터, 일반적인 데이터를 전처리 후 MLP(Multi Layer Perception)에 학습을 진행시키고 연합학습모델 또한 MLP를 사용하여 조건(parameter, layer 층, optimizer)을 맞추고 성능을 비교하면서 연합학습이 정말로 쓸 만한지에 대해 평가하고자 한다. 본 논문은 각 카드사의 금융 데이터가 동일한 칼럼을 가진 형식이라는 전제하에 중간 관리자의 역할을 하는 서버를 두고 각 카드사를 클라이언트 라고 가정을 한 후, 클라이언트에서 학습된 모델의 파라미터를 서버에서 업데이트하는 방식으로 실험을 진행했다.

3.1 실험 방법

실무에서 사용하는 일반 데이터를 비식별화한 후 학습과 연합학습의 성능을 비교하기 위해 실제 금융데이터를 이용하여 신용등급을 예측하는 딥러닝 기반 모델을 생성하였다. 각 데이터가 입력되었을 때 신용등급이 몇인지 판단하는 multi-class classification이다. 본 실험의 경우, 민감정보가 드러난 데이터를 구하기 어려운 특성상 어느 정도 비식별화된 데이터를 바탕으로 여러 전처리를 겪고 데이터를 학습하여 모델을 구성하였다. 또한 연합학습은 글로벌 모델 학습용 서버를 여러 사용하는 케이스와 하나의 서버를 사용하는 케이스 두가지중 선택 가능한 성질을 가지지만 본 논문은 구글에서 발표한 방식과 동일하게, 하나의 서버에서 가상의 클라이언트를 생성하여 연합학습을 하는 방식을 구현하여 실험을 진행했다.

3.2 데이터셋과 전처리

본 실험의 데이터는 미국의 P2P 회사의 2020년 상반기 Lending Club 데이터이다[29]. 본 데이터는 105,012명의 신용평가 결과가 A~D등급으로 구분되어 있고 신용에 관한 정보인 주택 소유 여부, 대출금, 연간소득 등으로 나누어져 있으며 전처리 시 신용과 관련 있는 칼럼 24개를 선별했고 각각의 feature는 Table 3에 제시했다. 본 논문에서 사용한 lending club 데이터, 금융데이터 혹은 공공기관 등 인터넷으로 구할 수 있는 데이터들은 어느 정도 비식별화 과정을 거친 데이터이다. 따라서 전처리하기 전의 원본데이터를 수정했으며 기계학습 특성상 범주형 데이터는 원핫 인코딩을 진행하거나 라벨 인코더를 진행한다.

본 데이터는 범주형 데이터를 최대한 수치형 데이

Table 3. Feature and explanation

feature	content
loan_amnt	amount of the loan
funded_amnt	The total amount committed to the loan at that time.
funded_amnt_inv	The total amount committed by the investor for the loan.
term	The number of payment
installment	Monthly payments have to be paid by the borrower when the loan begins.
grade	Credit grade
emp_length	Employment period
home_ownership	Whether the borrower own a house or not
annual_inc	Annial Income
zip_code	Postal code
open_acc	Credit limit
total_acc	Total credit limit
total_pymnt_inv	Payment for a portion of the total amount raised
tot_cur_bal	Total current balance of all acount
open_acc_6m	The number of transaction held over the past 6 months
total_bal_il	Current balance of all installment account
max_bal_bc	Maximum current balance for all account
avg_cur_bal	The average of current balance
mort_acc	The number of mortgage account
num_bc_sats	The number of bank accounts
tot_hi_cred_lim	The highest credit limit
total_il_high_credit_limit	The limit of total installment /Credit limit
revol_bal	Total credit turnover balance
total_pymnt	The amount paid for the total amount

Table 4. Original data preprocessing

feature	Result	preprossed data
zip_code	"780xx", "890"	"780"⟶ 780 "890"⟶ 890
home_ ownership	OWN, RENT,MORTGAGE, ANY	OWN⟶ [1,0,0,0] RENT⟶ [0,1,1,0] MORTAGE⟶ [0,0,1,0] ANY⟶ [0,0,0,1]
emp_length	n/a, < 1 year, 1 year,..., 9year, 10+years	n/a⟶ 0 <1year⟶ 0.5 1year⟶ 1 10+years⟶ 10
term	36months, 60months	36months⟶ 36 60months⟶ 60

터로 바꾸었다. 준 식별자인 우편번호(zip_code), 주택 소유 여부(home_ownership), 고용 기간(emp_length), 지불 횟수(term) 등이 이에 해당하며 원본데이터의 전처리 과정은 Table 4에서 보여준다. 비식별 데이터는 원본데이터로부터 ARX Tools를 사용해 k-익명성의 k=2일 때, k=5일 때, k=7일 때, 차분프라이버시는 e=0.1일 때, e=1일 때, e=2일 때, 재현데이터는 R 라이브러리의 synthpop을 이용하여 비식별 데이터를 얻었다. 이러한 데이터를 아래와 같이 전처리를 진행 후 학습시켰다. ARX Tool에 데이터를 불러와 준 식별자, 민감정보 등을 설정 후 프라이버시 모델을 설정하면 결과값을 얻을 수 있다. Fig 6는 k=2로 설정했을 대 비식별 데이터를 보여준다. 준 식별자로 우편번호, 주택 소유 여부, 고용 기간, 지불 횟수, 연간소득을 설정했으며 보호 강도가 정해진 선에서 원하는 feature의 비식별화 level을 선택할 수 있다. 연간소득은 level이 최고단계인 3단계가 적용되어 "*"처리한 부분을 볼 수 있다. 프라이버시 보호 모델마다 각각 변하는 항목들이 달랐으며 Table3의 전처리를 적용해 학습을 진행했다. Fig 7은 Arx의 프라이버시 보호 모델의 level 1단계 상태를 이용하여 비식별 처리된 데이터의 전처리 후를 보여준다.

emp_lengt	zip_code	home_owr	term	annual_in	loan_amnt	funded_ar	funded_ar	installmen	grade	open_acc	revol_bal
10+ years	0xxxx	MORTGA	3********	*	6500	6500	6500	223.11	C	16	16026
10+ years	0xxxx	MORTGA	3********	*	32000	32000	32000	996.29	A	18	54154
10+ years	0xxxx	MORTGA	3********	*	10000	10000	10000	352.17	C	12	9764
10+ years	0xxxx	MORTGA	3********	*	40000	40000	40000	1268.46	A	26	59517
10+ years	0xxxx	MORTGA	3********	*	5000	5000	5000	155.67	A	22	18788
10+ years	0xxxx	MORTGA	3********	*	6900	6900	6900	243	C	23	73706
10+ years	0xxxx	MORTGA	3********	*	24000	24000	24000	735.14	A	23	28882
10+ years	0xxxx	MORTGA	3********	*	15000	15000	15000	528.25	C	32	22693
10+ years	0xxxx	MORTGA	3********	*	17125	17125	17125	566.43	B	22	3252
10+ years	0xxxx	MORTGA	3********	*	12000	12000	12000	389.07	B	20	10672
10+ years	0xxxx	MORTGA	3********	*	7250	7250	7250	223.93	A	12	4387
10+ years	0xxxx	MORTGA	3********	*	20000	20000	20000	695.66	C	8	11515
10+ years	0xxxx	MORTGA	3********	*	25000	25000	25000	778.35	A	20	61239
10+ years	0xxxx	MORTGA	3********	*	3000	3000	3000	91.9	A	18	8742
10+ years	0xxxx	MORTGA	3********	*	30000	30000	30000	926.59	A	12	41039

Fig. 6. ARX, k=2 de-identification data.

3.3 실험결과

신용등급을 예측하기 위한 모델을 생성하기 위해 Multi-class 인증모델을 생성한다. 그 후 동일한 모델로 제시된 24개의 feature들을 통해 원본 데이터 학습, 비식별 데이터학습, 원본 데이터를 통한 연합학습을 진행하며 성능을 비교한다. 예측에 사용된 머신러닝 알고리즘은 MLP(Multi-layer Perceptron)을 사용했다. 최대한 비슷한 조건에서 원본, 비식별 데이터 학습, 연합학습을 비교하기 위해 조건을 맞추었으며 연합학습의 경우 client 10을 기준으로 100 라운드까지 진행하는 것을 제외한 모든 조건을 맞추었다. 모델에 대한 정확도는 전체 데이터 중 예측 값이 실제 값과 같은 비율을 말하며 식 (5)와 같이 계산한다.

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (5)$$

feature	Before → After
zip_code	"78xxx"--> 78 "89xxx"-->89 "*****"--> 1
home_ownership	(OWN)--> [1,0,0,0] (RENT, MORTGAGE)--> [0,1,1,0] (ANY)--> [0,0,0,1] "*"--> [1,1,1,1]
emp_length	(n/a, < 1 year, 1 year)--> 0.5 (2year,3year,4year)--> 3 (5year,6year,7year)--> 6 (8year, 9year)--> 9 10+years--> 10 "*"--> 1
term	3xxxxxx--> 3 6xxxxxx--> 6 "*******"--> 1
annual_inc	<20000--> 10000 (20000,40000]--> 30000 80000--> 80000 "*"--> 1

Fig. 7. preprocessing of de-identification data

3.3.1 성능평가

동일한 전처리와 모델을 사용하여 정확도를 구하였다. 그 결과 일반학습(ORI)의 경우 약 91%, K-익명성의 k=2일 때 약 80%, k=5일 때 약 76%, k=7일 때 약 62%가 나왔으며 차분프라이버시는 =2일 때 약 52%, =1일 때 50%, =0.1일 때 36%이며 재현데이터의 경우(SYN)는 약 82%가 나왔다. 원본데이터와 비식별 데이터의 학습 과정은 Fig 8에서 볼 수 있다. 연합학습은 약 86%가 나왔으며 원본데이터를 학습한 결과의 다음으로 가장 높은 정확도를 보였다. k-익명성의 k가 증가함에 따라 비식별 강도가 높아지고 차분프라이버시의 epsilon이 증가할 때마다 비식별 강도가 높아져 학습결과가 그만큼 많이 떨어진 것을 볼 수 있다. 모든 실험의 정확도는 Table 5에서 볼 수 있다. ORI는 익명화 처리를 하지 않은 데이터(원본)를 학습 후 테스트 결과이며 K-anom은 K-익명성의 결과를 나타내고 DP는 차분프라이버시, SYN은 재현데이터 FD는 연합학습을 나타낸다. Table 5와 같이 DP(=0.5)〈 DP(=1) 〈DP(=2) 〈k=7〈 k=5 〈 k=2 〈 재현데이터〈 federated learning 〈 original learning를 보였다.

Table 5. MLP Test Accuracy Result

ORI	De-identified Data			FD
	K-anom	DP	SYN	
0.9168	k=2, 0.7968	ϵ=2, 0.529	0.8248	0.8623
	k=5, 0.7606	ϵ=1, 0.5014		
	k=7, 0.6254	ϵ=0.1, 0.3622		

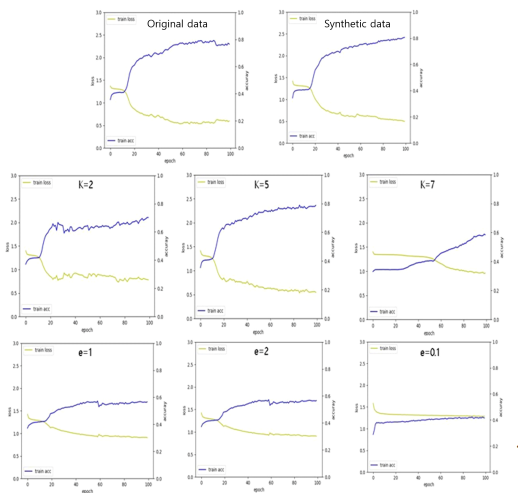

Fig. 8. Learning process: loss & Accuracy

3.3.2 연합학습의 파라미터

연합학습은 3장에서 언급했던 것처럼 디바이스 개수와 의사소통 횟수를 조절할 수 있다. 연합학습 또한 여러 머신러닝 모델처럼 파라미터를 변경하여 최적의 성능을 올릴 수 있다. 따라서 본 챕터에서는 Round 100을 기준으로 기존의 10개 디바이스의 개수를 5개로 줄여보고 20개로 늘려 성능 비교를 할 것이며 디바이스 개수에 따른 학습시간을 체크할 것이다. 디바이스 10개를 기준으로 Round의 수를 50번, 200으로 바꾸어 비교할 것이다. Table 6에서 볼 수 있듯 실험결과 Round 100을 기준으로 디바이스가 5개 일 때 0.9612가 나왔으며 20개일 경우 0.5701이다. 디바이스에 따른 학습시간은 Round를 무한히 설정해 놓고 90%에 도달했을 때 로컬학습 + 글로벌 학습시간을 계산하였고 Fig 9에서 볼 수 있는 바와 같이 디바이스가 적을수록 더 빨리 도달함을 볼 수 있다. 디바이스 10개를 기준으로 Round가 50일 때 0.567, Round가 200일 때 0.9619가 나왔다. 본 파라미터별 성능평가 결과로 Round가 높을수록 디바이스가 적을수록 더 높은 성능을 보여

Table 6. Performance by parameter

client(round=100)	5	10	20
acc	0.9612	0.8623	0.5701
round(Client=10)	50	100	200
acc	0.5676	0.8623	0.9619

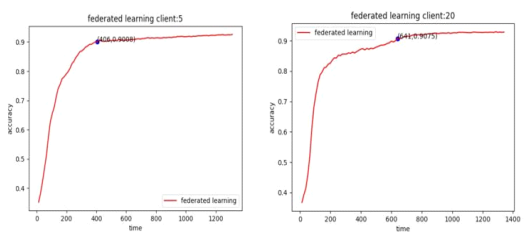

Fig. 9. Learning time depending on the device

주었다. 즉, 의사소통을 많이 하거나 디바이스가 적을수록 본 데이터에서는 높은 성능을 얻을 수 있게 되었다. 이러한 성능의 결과는 이미지 데이터, 텍스트 데이터, 수치 데이터 등등 데이터별 다르기 때문에 머신러닝의 그리드 서치처럼 최적의 파라미터를 찾아 사용해야 한다.

IV. 고 찰

연합학습이 이슈화되면서 국내에도 조금씩 연구중에 있다[30]. 특히 모든 클라이언트의 데이터들이 같은 칼럼을 가졌다는 전제하에 모델을 학습하는 것이 아닌 각 데이터들이 다른 칼럼으로 구성되어 있다는 가정을 할 시, 예를 들어 A카드사가 가지고 있는 데이터의 칼럼이 성별, 우편번호, 연간소득이라면 B카드사의 경우엔 성별, 우편번호, 주택 소유 여부 등인 상황에서 다른 칼럼을가진 데이터들을 어떻게 하면 하나로 학습시킬 수 있을지에 대해 포커스를 두고 있다. 이러한 문제를 연합학습으로도 해결하는 방안으로 제시가 되고 있다. 본 논문은 같은 칼럼을 대상으로 실험을 진행하였지만 아직 나오지 않는 칼럼이 다른 데이터를 이용하여 비식별화 데이터와의 성능 비교실험도 필요하다[28]. 기계학습을 사용할때 원본데이터의 범주형변수를 수치형으로 어떻게 처리하느냐에따라 원본데이터의 학습 정확도는 달라질수있으며 수치형으로 바뀐 원본데이터를 그대로 연합학습에 사용하기때문에 이에따라 연합학습의 학습 정확도 역시 달라질 수 있다.

또한 정확도 이외에 전처리 비용, 프라이버시 보호가 얼마나 되는지에 대한 비교가 필요하다. 프라이버시 보호모델에 적용된 익명화 데이터와 연합학습을 프라이버시 보호 측면에서 공통선상으로 비교함에 있어서 비교할만한 평가지표의 부재로 다른 측면의 비용이나 리스크에 대해서도 고려해야하기 때문이다.

V. 결 론

본 논문에서는 연합학습에 대한 원리와 실제 금융데이터를 사용하여 비식별처리기술을 적용한 데이터에 대해 성능을 비교했다. 연합학습의 결과가 원본데이터에 학습한 결과를 완전히 따라가지는 못하지만 다른 비식별과 비교해서 많이 떨어지지 않는다. 당연한 결과이지만 실무에서는 비식별화를 거친 데이터를 학습시켜 낮은 정확도의 모델을 얻느냐, 각 클라이언트의 디바이스에서 원본 데이터를 학습시켜 파라미터만을 이용해 적당한 정확도의 모델을 얻느냐의 선택상황을 가정해보았기 때문에 본 논문을 바탕으로 연합학습에 대해 연구가 계속해서 개발된다면 앞으로의 기계학습 연구에 많은 기여를 할 수 있을 것이다. 연합학습은 같은 칼럼이 있다는 전제하에 실제 실무에서 발생하는 데이터를 공유하지 않고 민감한 정보를 이용할 수 있다는 점이다. 연구가 지속한다면 금융데이터에서 다른 칼럼들을 이용하여 더 좋은 신용모델을 만들 수 있을 것이고 신용예측정확도가 높아짐에 따라 다양한 상품을 개발하는데 사용될 수 있을 것이다. 이를 통해 갚을 능력이 되는 사회초년생, 프리랜서, 금융정보 부족군에 해당하는 신용 평점 중위등급에 대한 문제도 해결 할 수 있을 것이다. 또한 완전해진 신용모델로 고객이 직접 은행을 방문하여 신용평가 받지 않을 수 있게 되어 비대면 서비스의 확대를 통해 카드사의 일을 덜어줄 수 있을 것이다. 데이터를 모으지 않고 파라미터값만을 통해 모델을 업데이트하는 연합학습의 패러다임은 프라이버시 측면으로 보면 안전하다고 할 수 있지만, 단점으로 연합학습을 구현하는 데 있어서 학습시간이 많이 든다. 실무에서 성능도 중요하지만, 시간효율도 고려해야 할 만큼 중요한 요소이다. 실제로 연합학습이 도입된다면 서비스 시간도 체크 해야 하며 디바이스 마다 응답시간이 다르기 때문에 오래 걸릴 것이다. 시간 효율성을 생각하면 활용가치가 떨어질 수 있지만, 실무에서 대량의 데이터를 모아야 하는 상황에서는 필요하고, 약간의 오버헤드를 감수하는 게 나을 것이다.

향후 연구계획으로는 프라이버시 관점에서 동일한 조건의 비교와 10개 기관의 데이터를 모으는 경우를 가정하여 각 데이터를 익명화 처리를 거친 후 하나로 모아 실험할 예정이다. 또한 칼럼이 다른 데이터인 Non-IID를 바탕으로 보안, 성능 면 뿐만 아니라 활용 가치를 높이는 연구를 할 예정이다[24,28].

References

[1] Hyejung Moon, Hyun Suk Cho, “Risk based policy at big data era: Case study of privacy invasion”, Informatization Policy, vol 19, no 4, pp. 63-82, 2012.

[2] Kangsoo Jung, Seog Park, Daeseon Choi, “Analysis of Privacy Violation

Possibility of Partially Anonymized Big Data", Journal of The Korea Institute of Information Security & Cryptology', vol 28, no 3, pp. 669-673, Jun. 2018.

[3] Jusung Kang, Jinyoung Kang, Okyeon Yi, Dowon Hong, "A study on the algorithms to achieve the data privacy based on some anonymity measures", Journal of the Korea Institute of Information Security & Cryptology, vol 21, no 5, pp. 149-160, Oct. 2011.

[4] Taewhan Kim, Seog Park , "Differentially Private Synthetic Data Generation Methods for Online Community Data", the Korean Information Science Society Conference, pp. 209-211, Jun. 2018.

[5] Neha Patki, Roy Wedge, Kalyan Veeramachaneni, "The Synthetic data vault", 2016 IEEE International Conference on Data Science and Advanced Analytics, pp. 399-410, Oct. 2016.

[6] Google AI, Federated learning [intern et], Available: https://federated.withgoogle.com, 2020.07.02.

[7] Seungwhoun Kim, Sunghae Jun, "Data De-identification using Autoencoder", Journal of Korean Institute of Intelligent Systems vol 30, no 3, pp. 228-235, Jun. 2020.

[8] Heuiju Chun, Hyun Jee Yi, Kyupil Yeon et. al., "Data Quality Measurement on a De-identified Data Set Based on Statistical Modeling", JOURNAL OF THE KOREA CONTENTS ASSOCIATION, vol 19, no 5, pp. 553-561, May. 2019.

[9] Surim Lee, Woongtae Jang, Jaeyoung Bae et.al., "Raising Risk and Suggesting Solution about Personal Information De-identification in Big-Data Environment", the Korea Information Processing Society Conference, Vol 23, no 2, pp. 297-300, Nov. 2016.

[10] Kyoungsung Min, Dohnchung Yon, "An Anonymization Method for Privacy Protection in Data Streams", Journal of KIISE : Databases, vol 41, no 1, pp. 8-20 Feb. 2014.

[11] DongHyun Kang, HyunSeok Oh, WooSeok Yong et.al, , "A Study on the Preservation of Similarity of privated Data", the Korea Information Processing Society Conference, pp.285 - 288, Nov. 2017

[12] LATANYA SWEENEY, "k-anonymity: a model for protecting privacy," International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, vol 10, no 5, pp. 557-570, 2012.

[13] Youngha Ryu, Kangsoo Jung, Seog Park, "Anonymization Technique Preserving Privacy against Inference Attack using Statistical Background Knowledge" Journal of KIISE : Computing Practices and Letters, Vol 17, no 3, pp. 195-199, Mar. 2011.

[14] Chikwang Hwang, Jongwon Choe, ChoongSeon Hong, "A Study on Service-based Secure Anonymization for Data Utility Enhancement", Journal of KIISE, vol 42, no 5, pp. 681-689, May. 2015.

[15] Cynthia Dwork, Aaron Roth, "The Algorithmic Foundations of Differential Privacy", Foundations and Trends in Theoretical Computer Science, vol 9, no (3-4), pp. 211-407, 2014.

[16] Cynthia Dwork, Frank McSherry, Kobbi Nissim et. al. , "Calibrating Noise to Sensitivity in Private Data Analysis", In Theory of Cryptography

Conference (TCC), Spr. 2006.
[17] Hyunil Kim, Cheolhee Park, Dowon Hong et.al, "A Study on a Differentially Private Model for Financial Data", Journal of the Korea Institute of Information Security & Cryptology, vol 27, no 6, pp. 1519-1534, Dec. 2017.
[18] Joungyoun Kim, Minjeong Park, "Multiple imputation and synthetic data", The Korean Journal of applied Statistics, vol 32 no 1, pp.83-97, 2019.
[19] Narayanan, A, and Vitaly S. "Robustde-anonymization of larg sparsedatasets." 2008 IEEE Symposium onSecurity and Privacy, pp.111-125, May. 2008.
[20] Bowen, C. M., and Liu, F. "Comparativestudy of differentially private datasynthesis methods." arXiv preprintarXiv: 1602.01063, Feb. 2016.
[21] Junyoung Kang, Sooyong Jeong, Dowon Hong, Changho Seo, "A Study on Synthetic Data Generation Based Safe Differentially Private GAN", Journal of The Korea Institute of Information Security & Cryptology, Vol.30, no 5, pp 945-956, Oct. 2020.
[22] Qiang Yang, Yang Liu, Tianjian Chen et. al.,"Federated Machine Learning: Concept and Applications", ACM Transactions on Intelligent Systems and Technology, vol 10, no 2, pp. 12:1-19, Jan. 2019.
[23] Yue Zhao, Meng Li, Liangzhen Lai, Naveen Suda, et. al., "Federated Learning with Non-IID Data", arXiv, Jun. 2018
[24] Tian Li, Anit Kumar Sahu, Ameet Talwalkar et. al.,"Federated Learning: Challenges, methods, and future directions", IEEE SIGNAL PROCESSING MAGAZINE, vol 37 no 3, pp. 50-60, May. 2020.
[25] Xin Yao, Tianchi Huang et. al., "Federated Learning with Additional Mechanisms on Clients to Reduce Communication Costs", arXiv, Sep. 2019
[26] Keith Bonawitz, Hubert Eichner, Wolfgang Grieskamp et. al.,"TOWARDS FEDERATED LEARNING AT SCALE: SYSTEM DESIGN", Proceedings of the 2nd SysML Conference, Mar. 2019.
[27] H. Brendan, McMahan Eider, Moore Daniel Ramage et. al., "Communication-Efficient Learning of Deep Networks from Decentralized Data", Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS), vol 54, pp. 1273-1282 , Feb. 2017.
[28] Jakub Konecny, H. Brendan McMahan, Felix X. Yu et. al., "FEDERATED LEARNING: STRATEGIES FOR IMPROVING COMMUNICATION EFFICIENCY", arXiv, Oct . 2017.
[29] LendingClub Statistics, [internet], Available: https://www.lendingclub.com/info/statistics.action , 2020.07.03.
[30] Andrew Hard, Kanishka Rao, Rajiv Mathews, "FEDERATED LEARNING FOR MOBILE KEYBOARD PREDICTION", arXiv, Feb. 2019.

〈저자소개〉

장 진 혁 (Jinhyeok Jang) 학생회원
2020년 2월: 공주대학교 응용수학과 학사
2020년 2월~2020년 8월: 공주대학교 융합과학과 석사과정
2020년 2월~2022년 2월: 숭실대학교 융합소프트웨어학과 석사
2022년 3월~현재: 숭실대학교 산학협력단 연구원
〈관심분야〉 정보보호, 금융보안, 인증, 네트워크보안, 딥러닝

안 윤 수 (Yoonsoo An) 학생회원
2019년 2월: 공주대학교 응용수학과 학사
2021년 9월~현재: 숭실대학교 융합소프트웨어학과 석사과정
〈관심분야〉 정보보호, 금융보안, 딥러닝

최 대 선 (Daeseon Choi) 종신회원
1995년 2월: 동국대학교 컴퓨터공학과 학사
1997년 2월: 포항공과대학교 컴퓨터공학과 석사
2009년 1월: 한국과학기술원 전산학과 박사
1997년 1월~1999년 6월: 현대정보기술 선임
1999년 7월~2015년 8월: 한국전자통신연구원 인증기술연구실 실장/책임연구원
2015년 9월~2020년 8월: 공주대학교 의료정보학과 부교수
2020년 9월~현재: 숭실대학교 소프트웨어학부 교수
2016년~현재: 정보보호학회 이사
〈관심분야〉 인증, 개인정보보호, 이상거래탐지, 의료정보보안, 머신러닝