Journal of The Korea Institute of Information Security & Cryptology ISSN 1598-3986(Print)
VOL.27, NO.6, Dec. 2017 ISSN 2288-2715(Online)
https://doi.org/10.13089/JKIISC.2017.27.6.1519

금융 데이터 상에서의 차분 프라이버시 모델 정립 연구*

김 현 일,† 박 철 희, 홍 도 원,‡ 최 대 선
공주대학교

A Study on a Differentially Private Model for Financial Data*

Hyun-il Kim,† Cheolhee Park, Dowon Hong,‡ Daeseon Choi
Kongju National University

요 약

데이터 비식별화 기법은 데이터 내에 속한 개인 정보에 대한 프라이버시를 만족하면서 동시에 데이터 분석가들에게 유용한 정보를 습득할 수 있게 하는 반드시 필요한 기술 중 하나이다. 그러나 k-익명성과 같은 기존의 비식별화 기법은 공격자의 사전지식(Background knowledge)에 근본적으로 취약한 약점을 지니고 있다. 하지만 차분 프라이버시(Differential privacy)는 기존의 비식별화 기법들과는 다르게 개인 정보에 대한 강력한 안전성을 보장하는 모델로써 최근 들어 이에 대한 연구가 매우 활발히 진행 중에 있다. 본 논문은 이러한 차분 프라이버시가 적용된 기술에 대한 연구 및 분석을 통해 금융 데이터 상에서의 차분 프라이버시 모델을 정립하였으며 이러한 모델들은 금융 데이터 상에서 유용하게 사용될 수 있음을 입증하였다.

ABSTRACT

Data de-identification is the one of the technique that preserves individual data privacy and provides useful information of data to the analyst. However, original de-identification techniques like k-anonymity have vulnerabilities to background knowledge attacks. On the contrary, differential privacy has a lot of researches and studies within several years because it has both strong privacy preserving and useful utility. In this paper, we analyze various models based on differential privacy and formalize a differentially private model on financial data. As a result, we can formalize a differentially private model on financial data and show that it has both security guarantees and good usefulness.

Keywords: De-identification, Differential privacy, Financial data

I. 서 론

빅 데이터 시대의 출현으로 인하여 다수의 기업들은 각자의 목적을 위해 방대한 양의 데이터에 대한 분석을 통해 유용한 정보를 얻어내는 기술을 많이 사용하고 있다. 이 때 각 기업들이 자신이 보유하고 있는 데이터를 공개하여 상호 공유한다면 더욱 더 의미 있는 유용한 정보를 얻어낼 수 있게 된다. 하지만 이러한 데이터는 Sweeny[1]에 의해 준식별자(Quasi-identifier, QID)인 우편번호, 성별, 생일 만으로 미국 인구의 87%가 재식별화가 가능함이 밝혀져 식별자(Identifier, ID)가 제거된 데이터라도 안전하지 않음이 확인되었다. 따라서 이러한 문제점을 보완하기 위해 프라이버시에 위협이 될 수 있는 준식별자의 속성(attribute)들을 알아볼 수 없는 형태로 변환하거나 삭제 등을 수행하는 데이터 비식별화(De-identification)기법이 반드시 수행되어야만 한다.

Received(09. 18. 2017), Modified(10. 24. 2017.), Accepted(10. 24. 2017)
* 본 연구는 2016년도 정부(미래창조과학부/교육부)의 재원으로 한국연구재단의 지원을 받아 수행되었습니다.(2016R1A4A1011761, 2016R1D1A1B03931071).
† 주저자, hyunil89@kongju.ac.kr
‡ 교신저자, dwhong@kongju.ac.kr(Corresponding author)

이에 따라 국내에서는 2016년 6월 데이터에 속해 있는 프라이버시 보호를 위한 개인정보 비식별화 조치 가이드라인을 발표하였다[2]. 이로 인해 모든 기업 및 단체들은 비식별화된 데이터를 상호 공유할 수 있게 되었으며 금융관련 기업 등 여러 기업들은 상호 공유된 데이터에 대한 통계를 통해 마케팅, 위험 분석 등을 유용하게 사용할 수 있게 되었다. 실제 2016년 10월 미래창조과학부로부터 빅데이터 개인정보 비식별화에 대한 실증과제 의뢰를 받은 SK텔레콤은 지난 6개월 간 한화생명, 서울보증보험 등과 함께 빅데이터 개인정보 비식별화 작업을 진행하였으며 SK텔레콤과 한화생명에 동시 가입된 약 218만명을 대상으로 다음과 같은 속성을 가진 비식별화 된 데이터를 결합하였다.

- SK텔레콤 : 통신료 연체금액, 멤버쉽 사용금액, 통신료 미납횟수 등
- 한화생명 : 직업, 신용대출건수, 총신용대출금액, 최근신용등급 등

그 결과, SK텔레콤은 결합된 데이터가 고객의 통신요금 연체발생을 판단할 수 있는 보조정보로 활용될 가치가 있음이 입증되었다고 발표하였다[3]. 이처럼 비식별화 된 데이터에 대한 통계 분석은 여러 측면에서 매우 유용하게 사용될 수 있다.

이와 동시에 학계에서는, 데이터에 대한 유용성 뿐만 아니라 개인 프라이버시를 고려한 안전성 측면에서의 데이터 비식별화 기법에 대해 많은 연구들이 진행되어왔다. 대표적으로 Sweeny[1]에 의해 제안된 k-익명성은 QID로 인한 재식별화를 방지하기 위해 모든 레코드(Record)들을 적어도 k개의 QID 그룹으로 일반화(Generalization)를 수행하는 기법으로 특정 개인의 QID들을 알고 있더라도 적어도 $1/k$ 만큼의 안전성을 보장하는 기법이다. Table 1은 개인의 신용불량 여부를 민감속성으로 갖는 원본 데이터를 나타내며 Table 2는 이에 대해 3-익명성을 만족하는 데이터를 나타낸다. 하지만 Machanavajjhala 등[4]은 데이터에 k-익명성 모델을 단순하게 적용할 경우 동일한 QID 그룹에 대해 민감한 속성이 다양하게 표현되지 않는다면 공격자의 배경지식 공격(Background knowledge attack)에 취약함을 발견하였다. Table 2에서의 QID 그룹 중 하나인 〈Artist, [30-40), M〉에 해당하는 레코드의 수는 4개이므로 1/3 이상의 안전성을 보장하는 것으로 보이지만 해당 QID 그룹의

Table 1. The original data for credit status

Job	Age	Sex	Credit status (sensitive)
Engineer	33	M	○
Engineer	31	M	X
Professor	34	M	X
Professor	49	F	X
Architect	43	F	X
Architect	41	F	X
Writer	39	M	○
Poet	38	M	○
Dancer	36	M	○
Dancer	37	M	○

Table 2. Credit status data with 3-anonymity

Job	Age	Sex	Credit status (sensitive)
Professional	[30~40)	M	○
Professional	[30~40)	M	X
Professional	[30~40)	M	X
Professional	[40~50)	F	X
Professional	[40~50)	F	X
Professional	[40~50)	F	X
Artist	[30~40)	M	○
Artist	[30~40)	M	○
Artist	[30~40)	M	○
Artist	[30~40)	M	○

레코드 모두 신용불량 이력을 가지고 있으므로 공격자는 쉽게 해당 QID 그룹에 속한 피공격자의 민감한 정보를 알 수 있게 된다. 따라서 이러한 공격을 방지하기 위해 동일한 QID 그룹 내의 민감 정보가 최소 l개만큼 다양하게 표현되는 l-다양성 모델이 제안되었다[4]. 하지만, l-다양성 모델은 민감 정보의 자연스러운 분포도로 인해 확률적 추론 공격(Probabilistic Inference Attack)에 대한 문제점이 있다. 의료 데이터에서 민감 속성인 질병 여부를 예로 들면, 자연스럽게 에이즈(HIV)는 독감(Flu)에 비해 매우 희귀하다. 따라서 동일한 QID 그룹에 대해 $1/k$ 만큼의 안전성을 만족할 수 없으며, $1/k$ 만큼의 안전성을 만족하게끔 QID 그룹을 표현한다 해도 너무 많은 일반화 과정을 수행해야 하므로 데이터의 유용성에 매우 큰 문제점이 발생한다. 이후 제안된 (α,k)-익명성[5] 및 t-근접성[6] 등의 여러 익명화 기법들 모두 k-익명성과 l-다양성처럼 QID에 대한 일반화 모델의 변형 기법들이며, 이러한 기법들은 근본적으로 공격자의 사전지식에 취약한 약점을 지니고 있다[7]. 이에 대한 해외 실증

사례로써, 인터넷 미디어 스트리밍 서비스를 제공하는 Netflix는 사용자들의 성향을 분석해 보다 나은 미디어 추천 서비스를 제공하기 위해 기존의 모델을 통해 비식별화 된 데이터를 공개하고 이를 이용해 가장 효율적인 추천 시스템을 만든 팀에게 100만 달러를 수여하는 Netflix prize를 개최하였다. 하지만 이는 동일한 서비스를 제공하는 IMDb사에서 공개한 비식별화 된 데이터베이스와의 연결(linkage)로 인해 재식별화가 가능함이 발견되었다[8]. 또한 메사추세츠 보험 위원회(Massachusetts Group Insurance Commission(GIC))의 비식별화 된 의료기록 데이터베이스 역시 유권자 등록 데이터베이스와의 연결로 인해 재식별화 되었다[9]. 이처럼 기존의 비식별화 기법은 프라이버시 측면에서의 여러 문제점을 지니고 있다.

하지만, 이전의 비식별화 기법들과는 다르게 Dwork 등[11]에 의해 제안된 차분 프라이버시(Differential Privacy) 모델은 개인의 프라이버시를 보존할 수 있는 강력한 안전성을 보장하는 모델이다. 차분 프라이버시란 임의의 인접한 두 입력 데이터 집합에 대해 각각 차분 프라이버시를 만족하는 특정 메커니즘의 출력 값에 대한 확률의 차이가 크지 않다는 것을 의미한다. 즉, 차분 프라이버시를 만족하는 결과 값을 이용해 특정 개인이 속한 데이터집합과 속하지 않은 데이터집합에 대한 구분을 할 수 없음을 나타낸다. 따라서 차분 프라이버시는 공격자의 배경지식과는 무관하게 개인에 대한 강력한 프라이버시를 보장하는 기법이다.[10]

본 논문에서는 기존의 비식별화 기법이 아닌 차분 프라이버시를 이용한 기술에 대한 연구 및 이를 이용해 금융 데이터 상에서의 차분 프라이버시 모델에 대한 방향을 정립한다. 앞서 SK텔레콤의 실증 사례와 같이 차분 프라이버시가 금융 데이터 비식별화 기법에 적용된다면 개인에 대한 정보보호를 강력하게 보장하면서 동시에 유용하게 쓰일 것으로 예상된다. 또한 금융 산업이 로보어드바이저, 금융보안 및 신용평가 등의 서비스에서 인공지능 기술 활용도가 높을 것으로 기대[13]되기 때문에 차분 프라이버시를 만족하는 머신러닝 모델 또한 매우 유용하게 쓰일 것으로 예상된다. 이에 따라 본 논문의 구성은 다음과 같이 이루어진다. 2장에서는 차분 프라이버시 및 본 논문에 대한 기본개념을 서술하며, 3장에서는 현재까지 적용된 차분 프라이버시를 이용한 기술에 대해 분류하고 분석한다. 4장에서는 금융 데이터 상에서의 차분 프라이버시 모델 방향 정립 및 공개된 금융 데이터를 이용해 차분 프라이버시가 적용된 기법에 대한 데이터 유용성을 분석한 후 마지막으로 5장에서 결론을 맺는다.

II. 기본개념

2.1 차분 프라이버시(Differential Privacy)

차분 프라이버시는 이전의 비식별화 모델과는 다르게 공격자가 특정 개인에 대해 다른 데이터베이스로 부터 얻은 사전지식이 있다 해도 해당 데이터베이스에서의 개인 레코드에 대한 프라이버시에는 영향을 미칠 수 없는 강력한 프라이버시 모델이다. 거꾸로 말하면, 차분 프라이버시는 특정 개인이 해당 데이터베이스에 있든 없든 질의의 응답 값을 통해 이전에 공격자가 알고 있는 특정 개인에 대한 정보보다 더 많은 정보를 얻을 수 없다는 것을 의미한다.[12] 이에 대한 차분 프라이버시의 정의는 다음과 같다.

정의 1 (이웃 데이터베이스(neighboring database)).
데이터 전체집합 $\mathbb{N}^X$ 상에서의 두 데이터베이스 $D, D' \in \mathbb{N}^X$에 대해 l_1-노름(l_1-norm) $\|D\|_1$은 해당 데이터베이스의 크기를 말하며 이는 레코드의 수를 의미한다. 이 때 $\|D \triangle D'\|_1$은 두 데이터베이스의 레코드의 수의 차이를 말하며 $\|D \triangle D'\|_1 \leq 1$를 만족할 시 두 데이터베이스 D, D'를 서로 이웃 데이터베이스(neighboring database)라 한다.

정의 2 ((ϵ, δ)-차분 프라이버시). 모든 사건 $S \subseteq Range(M)$에 대해 이웃 데이터베이스인 $D, D' \in \mathbb{N}^X$를 입력 값으로 하는 랜덤성을 갖는 메커니즘 M: $\mathbb{N}^X \rightarrow Range(M)$이 다음과 같은 식을 만족할 때 M은 (ϵ, δ)-차분 프라이버시를 만족한다고 한다.

$$\Pr[M(D) \in S] \leq e^{\epsilon} \times \Pr[M(D') \in S] + \delta$$

이 때 $\delta = 0$일 시에는 ϵ-차분 프라이버시라고 한다. Fig. 1은 각 데이터베이스의 출력 값의 분포에 대해 ϵ-차분 프라이버시 및 (ϵ, δ)-차분 프라이버시 관점에서의 차이 값을 나타낸다.[14] 이 때 안전성 파라미터인 $\epsilon, \delta \geq 0$는 다음과 같은 뜻을 갖는다. 우선 ϵ값은 작아지면 작아질수록 두 응답 값이 사건 S에 속할 확률 값에 대한 차이가 거의 나지 않기 때문에 해당 응답 값으로 두 데이터베이스를 구분하기 매우 어려워짐을 의미한다. 해당 ϵ값은 해당 시스템의 파라미터에 따라

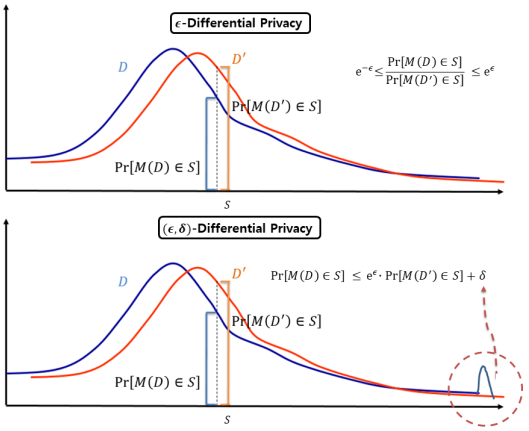

Fig. 1. Differential privacy for ϵ, δ[14]

결정되지만 보통 0.1, 0.5, 1 정도의 값을 주로 사용한다. 다음으로 δ는 그 값에 따라 δ만큼의 차이를 용인한다는 뜻으로써 이 역시 시스템 파라미터에 의해 결정되나 보통 10^{-4}보다 작은 값을 사용한다.[15]

이처럼 차분 프라이버시는 이전의 비식별화 모델들과는 다르게 특정 속성에 대한 삭제 및 일반화를 수행하는 것이 아닌 해당 질의에 대해 특정 개인의 정보가 드러나지 않을 만큼의 랜덤성을 갖는 응답 값을 반환하는 메커니즘을 제공하는 것을 의미한다. 만일 이웃 데이터베이스 D, D'에 대해 공격자가 각각 "신용 불량자의 수는 몇 명인가?"와 같은 특정 질의를 수행하여 데이터베이스에 속한 특정 공격 대상자의 신용 여부를 알려 한다고 가정하자. 이러한 경우 기존의 비식별화 기법이 적용된다 해도 두 응답 값의 차이가 존재하지 않으면 신용불량자가 아님을, 존재한다면 신용불량자임을 쉽게 알 수 있다. 하지만 차분 프라이버시는 입력되는 안전성 파라미터 (ϵ, δ)에 대해 (ϵ, δ)-차분 프라이버시를 만족하는 응답 메커니즘을 제공하여 특정 공격 대상자가 실제로 신용불량자라고 해도 확률적 특성을 이용하여 응답 값의 차이가 존재할 수도, 존재하지 않을 수도 있게끔 조정되어 응답한다. 이로 인해 특정 개인이 데이터베이스에 속한다는 배경 지식을 가지고 있는 공격자일지라도 쉽게 신용불량이력을 알 수 없게 된다.

이러한 차분 프라이버시를 만족하는 메커니즘 M은 기본적으로 특정 질의에 해당하는 응답 값에 노이즈를 추가하는 방식을 사용한다. 이 때 추가되는 노이즈의 크기는 해당 질의 함수의 민감도에 의존하게 되며, 질의 함수에 대한 민감도의 정의는 다음과 같다.[10]

정의 3 (민감도(sensitivity)). 질의 함수 $Q: \mathbb{N}^X \rightarrow \mathbb{R}^d$ 및 이웃 데이터베이스 $D, D' \in \mathbb{N}^X$에 대하여 민감도 $\triangle Q$는 다음과 같이 계산된다.

$$\triangle Q = \max_{D, D'} \| Q(D) - Q(D') \|_1$$

2.2 차분 프라이버시를 만족하는 메커니즘

이번 절에서는 차분 프라이버시를 만족하는 메커니즘 중 대표적인 Laplace 메커니즘 및 Exponential 메커니즘에 대해 정의한다.

2.2.1 Laplace 메커니즘

Dwork 등[11]에 의해 제안된 Laplace 메커니즘은 해당 질의 Q가 "신용 불량 이력이 있는 사람은 몇 명인가?"와 같이 수에 기반한 질의(numeric query)에 대해 가장 많이 사용되는 메커니즘이다. Laplace 메커니즘의 정의는 다음과 같다.

정의 4 (Laplace 메커니즘(Laplace Mechanism)). 어떠한 데이터베이스 D에 대해 민감도가 $\triangle Q$인 질의 $Q: \mathbb{N}^X \rightarrow \mathbb{R}^d$를 수행할 시의 Laplace 메커니즘 $\mathcal{M}_L$은 다음과 같은 식을 만족한다.

$$\mathcal{M}_L(x, Q(\cdot), \epsilon) = Q(x) + (Y_1, \ldots, Y_d)$$

이 때 각각의 $Y_i (1 \leq i \leq d)$는 Laplace 분포 $Lap(\triangle Q/\epsilon)$에서 독립적으로 균등하게 (Independent and identically distributed) 선택되며 이를 수식으로 나타내면 $Y_i \sim Lap(\triangle Q/\epsilon)$이다. 이 때 $X \sim Lap(b)$는 Laplace 분포의 확률밀도함수 $\frac{1}{2b} exp(-\frac{|x|}{b})$를 따른다.

정리 1[11][10]. 정의 4에 의해 정의된 Laplace 메커니즘 $\mathcal{M}_L$은 ϵ-차분 프라이버시를 만족한다.

2.2.2 Exponential 메커니즘

기본적으로 Laplace 분포를 따르는 노이즈 값을 선택하여 원래의 응답 값에 노이즈 추가 후 응답하는

Laplace 메커니즘은 수에 기반 하지 않은 질의(non-numeric query)에는 바람직하지 않다. 해당 질의 Q가 "신용 불량 이력을 가진 사람이 가장 많은 직업군은 무엇인가?"와 같이 최적의 속성을 선택해야 하는 경우를 예로 들 수 있다. Mcsherry 등[16]에 의해 제안된 Exponential 메커니즘은 이러한 수에 기반 하지 않은 최적의 값을 선택해야하는 질의에 대해 가장 많이 사용되는 메커니즘 이다. Exponential 메커니즘의 정의는 다음과 같다.

정의 5 (Exponential 메커니즘(Exponential Mechanism)). 데이터 전체집합 $\mathbb{N}^X$에서의 스코어 함수 $u:(\mathbb{N}^X\times T)\rightarrow R$ 및 이웃 데이터베이스 D와 D'에 대한 스코어 함수의 민감도를 $\triangle u = \max_{\forall t, D, D'}|u(D,t)-u(D',t)|$ 라고 할 때 Exponential 메커니즘 $\mathcal{M}_E(D,T,Q,\epsilon)$은 출력 집합 T에 대해 출력 값 $t\in T$가 선택될 확률이 $\exp(\frac{\epsilon u(D,t)}{2\triangle u})$에 비례한다.

정리 2[16][10]. 정의 5에 의해 정의된 Exponential 메커니즘 $\mathcal{M}_E$는 ϵ-차분 프라이버시를 만족한다.

2.2.3 차분 프라이버시의 합성

상기 $\mathcal{M}_L$ 및 $\mathcal{M}_E$와 같은 차분 프라이버시를 만족하는 메커니즘 들은 보통 특정 모델 설계 시 한번만 사용되는 것이 아닌 여러 번 사용하게 된다. 이 때 다음과 같은 두 가지의 경우를 생각할 수 있다. 첫 번째는 동일한 데이터베이스 상에서 차분 프라이버시를 만족하는 메커니즘을 반복하는 경우이며, 두 번째는 데이터 일부분이 겹치는 여러 개의 서로 다른 데이터베이스 상에서 각각 차분 프라이버시를 만족하는 메커니즘을 반복하는 경우이다. 두 방식 모두 반복적으로 포함되는 개인 데이터가 반드시 존재하므로 이러한 다중 합성(composition) 시의 차분 프라이버시의 구성 또한 필요하다. 따라서, 이번 절에서는 k번의 메커니즘 구성 시 k-폴드 적응 합성(k-fold adaptive composition)에 대해 설명한다.[17][10] 또한 개인 데이터가 전혀 겹치지 않는 서로 다른 데이터베이스 상에서의 차분 프라이버시의 병렬적 합성(parallel composition)[18]에 대해서도 설명한다.

정리 3(k-폴드 적응 합성(1))[17][10]. ϵ-차분 프라이버시를 만족하는 메커니즘을 k번 구성할 시 k-폴드 적응 합성상에서 $k\epsilon$-차분 프라이버시를 만족한다.

정리 4(k-폴드 적응 합성(2))[17][10]. 안전성 파라미터 $\epsilon,\delta'\geq 0$에 대해 ϵ-차분 프라이버시를 만족하는 메커니즘을 k번 합성할 시 k-폴드 적응 합성상에서 다음과 같은 ϵ'에 대해 (ϵ',δ')-차분 프라이버시를 만족한다.

$$\epsilon' = \sqrt{2k\ln(1/\delta')}\,\epsilon + k\epsilon(e^{\epsilon}-1)$$

정리 5[18] 데이터에 대한 교집합이 없는 서로 다른 데이터 D_i에 대해 각각 ϵ-차분 프라이버시를 만족하는 메커니즘은 데이터의 개수와 상관없이 ϵ-차분 프라이버시를 만족한다.

III. 차분 프라이버시가 적용된 기술에 대한 분류

2.1절에서 언급하였듯이 차분 프라이버시는 질의에 대한 응답 값을 다루는 랜덤성을 갖는 메커니즘을 제공하는 모델이다. 이러한 질의 기반의 차분 프라이버시 메커니즘은 보통 이에 대한 응답 값을 반환하는 온라인 질의 시스템 모델에 용이하게 적용된다. 하지만 이 뿐만 아니라 데이터 세트의 통계를 이용한 머신러닝 모델의 분류 작업 및 프라이버시 보존형 데이터 공개(Privacy Preserving Data Publishing, PPDP) 모델에도 적용 시킬 수 있으며[19], 개인 데이터 수집 시 혹은 비정상 행위 탐지 등에도 적용 가능하다. 이번 장에서는 이러한 차분 프라이버시가 적용된 기술 동향에 대해 분석하고 분류하며 이를 기반으로 4장에서 금융 데이터 상에서의 차분 프라이버시 모델을 정립한다.

3.1 프라이버시 보존형 데이터 공개 기술

Fig. 2는 온라인 질의 시스템에 기반한 차분 프라이버시 모델을 나타내며, Fig. 3은 데이터 공개 기술을 사용할 시의 차분 프라이버시 모델을 나타낸다. 이처럼 차분 프라이버시를 만족하는 프라이버시 보존형 데이터 공개 기술은 온라인 질의 시스템 방식이 아닌 데이터 관리자(curator)가 차분 프라이버시를 만족하는 데이터 공개 알고리즘을 사전에 수행해야 한다[7]. 이러한 데이터 공개 알고리즘은 해당 데이터에 대해 "위생처리(sanitized)" 된 데이터베이스를 공개하는

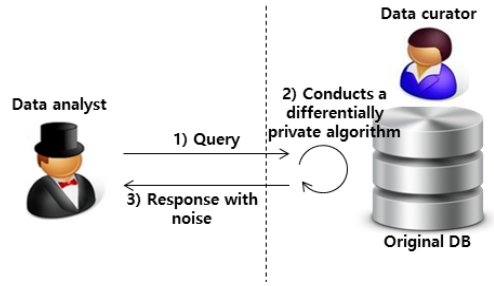

Fig. 2. Differential privacy models with online query systems

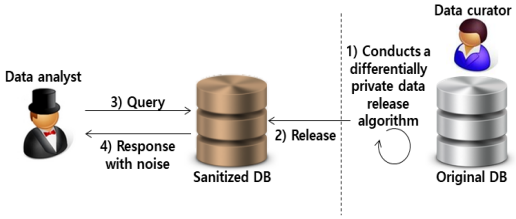

Fig. 3. Differential privacy models with data release

방식이며, 이는 크게 두 가지 모델로 나눌 수 있다.

첫 번째는 데이터에 대한 분할(partition)작업을 통하여 준식별자에 대한 일반화(Generalization) 수행 후 이에 대한 합성 테이블(contingency table)을 공개하는 방식이다. Fig. 4는 Table 1에 대한 합성 테이블 생성 시 사용되는 분류 체계에 대한 예를 나타내며, Table 3는 Fig. 4를 기반으로 최종적으로 공개되는 합성 테이블이다. 이러한 과정은 QID에 대해 일반화 될 속성을 차분 프라이버시하게 선택 후, 최종적으로 일반화가 진행된 QID 그룹에 노이즈가 추가 된 테이블을 공개하게 된다. 이에 대한 대표적인 기법으로, Mohammed 등[7]에 의해 제안된 DiffGen은 준식별자에 대한 분류 체계 트리(taxonomy tree)를 이용하여 일반화를 수행하는

Table 3. A contingency table for Fig.4

Job	Sex	Age	Credit status
Professional	[30 ~ 40)	M	Yes=2, No=2
Professional	[40 ~ 50)	F	Yes=0, No=2
Artist	[30 ~ 40)	M	Yes=5, No=2

TDS(Top-Down Specialization)[20]에 최초로 차분 프라이버시를 적용한 일반화 알고리즘이다. 해당 기법은 TDS에서 사용하는 정보획득량(Information gain)과 최대함수(Max)를 $\mathcal{M}_E$의 스코어 함수로 사용하고 이를 통해 일반화 될 속성들을 선택한 후 분류체계를 생성한다. 이후 각각의 QID 그룹에 신용불량유무에 대해 $\mathcal{M}_L$을 이용한 노이즈를 추가하여 이에 대한 합성 테이블을 공개한다. 또한 미국국립보건원(National Instituted of Health, NIH)의 지원을 받아 수행된 프로젝트인 SHARE[21] 또한 차분 프라이버시를 만족하는 히스토그램 기법인 DPCube[22]를 통해 합성 테이블을 생성하여 공개한다. DPCube는 준식별자의 값을 균일하게 나눈 후, 각 셀마다 $\mathcal{M}_L$을 이용한 노이즈를 추가하여 데이터를 히스토그램으로 나타낸다. 이후 노이즈가 추가 된 값을 이용해 kd-tree[23]를 구성한 후 이를 기반으로 셀을 통합한다. 마지막으로 공개되는 히스토그램의 값은 최종적으로 통합된 셀에서의 카운트에 노이즈가 추가 된 값을 사용한다. 이외에도 데이터 유용성 측면에서 해당 노이즈에 대한 에러를 최소화 하는 방식 및 다차원 데이터를 다루기 위한 기법들이 제안되어 왔으며 현재에도 해당 연구에 대해 계속 진행 중에 있다.[15] 또한, SHARE는 고혈압 측정 데이터와 같은 시계열 데이터(Time series)에 알맞은 차분 프라이버시 기법인 DPTrie[21]를 사용한다. 이는 혈압의 값을 저혈압(L), 정상(N), 고혈압(H)으로 나타내어 그 후에 시간별로

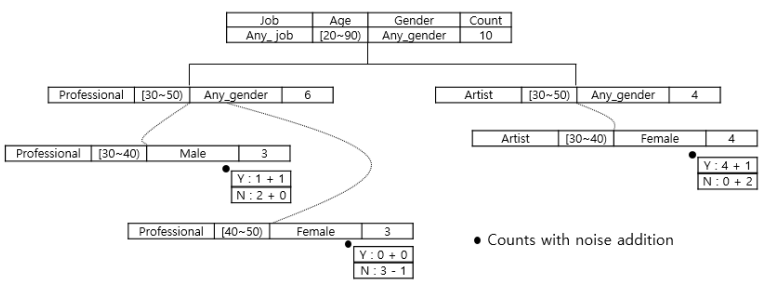

Fig. 4. A taxonomy tree example for Table 1

값을 트리화 하여 분류한다. 이러한 분류 진행 시 카운트 값이 매우 작거나 0인 경우에는 더이상 분할하지 않으며, 최종적으로 생성된 트리의 각 노드에 $\mathcal{M}_L$을 이용한 노이즈를 추가하게 된다.

두 번째는 원본 데이터에 대해 재현 데이터(synthetic data)를 만들어내는 과정이다. 재현 데이터란 랜덤 샘플링을 통해 원본 데이터의 일부분을 선택하여 생성되며, 이는 원본 데이터와 통계적, 확률적 특성이 동일하게끔 데이터를 구성한다.[24] 이에 대한 대표적인 기법으로는 미 통계국(US Census Bureau)이 온라인 상으로 개인 혹은 민간 업체 등이 유용하게 사용할 수 있는 미국 지역 별 인구 통계를 제공하는 OnTheMap[25][26]이 있다. 이는 현재 2002년도부터 2014년까지 나이, 수입, 인종, 학력, 성별을 준식별자로 가진 데이터를 제공하며, Dirichlet 분포를 이용한 차분 프라이버시를 만족하는 재현 데이터 생성 알고리즘[26]을 사용한다. 그리고 Li 등[27] 또한 차분 프라이버시를 만족하는 재현 데이터 생성 알고리즘을 제안하였다. Li 등은 랜덤 샘플링 및 이상치(outlier)에 대한 제거(pruning)를 통해 차분 프라이버시를 만족하는 k-익명성 알고리즘을 제안하였다.

첫 번째 방식인 합성 테이블 생성은 Table 3의 예시에서 카운트 질의처럼 해당 질의의 유형에 대해 두 번째 방식에 비해 상대적으로 높은 유용성을 가진다. 하지만 이는 곧 질의의 종류에 제한된다는 말로, 각각의 질의 유형에 대한 합성 테이블이 필요한 단점을 가지고 있다. 이에 반해 두 번째 방식은 데이터베이스가 가지고 있는 속성 값들을 그대로 공개하는 방식이기 때문에 범용적 분석이 가능한 이점을 가지고 있다. 하지만 원본 데이터의 확률적 특성을 그대로 보존하려면 데이터 열(data row)의 개수가 많아야만 하고 첫 번째 방식에 비해 상대적으로 낮은 유용성을 가지고 있는 단점을 지니고 있다.

3.2 프라이버시 보존형 머신러닝 기술

해당 학습 데이터로부터 의미 있는 패턴을 자동적으로 찾아 새로운 데이터에 대해 분류(classification), 예측(prediction), 군집(clustering) 등을 수행하는 머신러닝은 최근 수많은 분야에서 매우 유용하게 사용되고 있다. 이러한 의미 있는 패턴이라는 것은 해당 데이터가 가지고 있는 확률적 분포의 특성을 통해 여러 속성 값들로 이루어진 데이터 레코드들에 대한 규칙을 분석하는 것과 같다. 따라서 머신러닝 기술은 차분 프라이버시와 깊은 연관성을 가진다. 즉, 머신러닝 또한 개인의 데이터에 의존하기보다는 해당 학습 데이터 세트의 분포로부터 정보를 얻어내기 때문에 데이터 분포를 다루면서 동시에 개인 데이터를 드러내지 않는 차분 프라이버시 모델을 적용하기에 매우 용이하다.[10]

최근 들어 R.Shokri 등[43]과 M.Fredrikson 등[44]에 의해 안전성을 고려하지 않은 머신러닝 알고리즘 구동 시 해당 학습 데이터에 속한 개인의 민감 정보가 누출됨이 발견되었다. 이에 따라 다수의 차분 프라이버시 기반의 머신러닝 알고리즘에 대한 여러 연구가 진행되고 있다[28]. Fig. 5는 차분 프라이버시 기반 머신러닝 모델 생성(학습)을 나타낸다. 모델 설계자는 질의를 통해 어떠한 특성을 만족하는 통계 데이터를 습득하고 이를 기반으로 머신러닝 모델을 생성(학습)할 수 있다. 만일 Table 1의 데이터를 의사결정나무(Decision Tree)를 통해 분류하였다면 Fig. 6과 같이 수행될 수 있다. 하향식 의사결정 흐름도를 가진 의사결정나무는 루트 노드에서부터 분류를 시작한다. 이 때 각 레벨은 두 번의 질의를 통해 분류를 수행하게 된다. 첫 번째 질의는, 해당 레벨에 신용여부가 "예"에 해당하는 레코드의

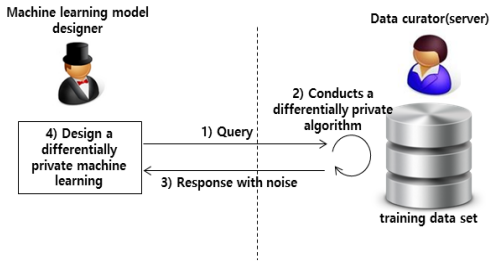

Fig. 5. Differential privacy model with machine learning

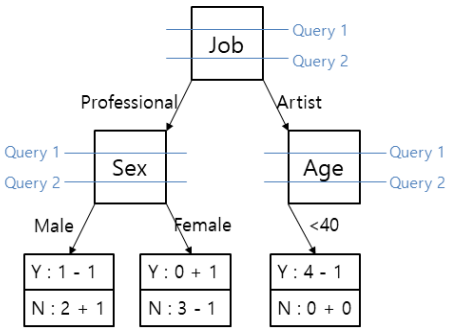

Fig. 6. An example for Table 1 using decision tree algorithm

수와 "아니오"에 해당하는 레코드의 수이다. 그 후 두 번째 질의에서 어떠한 속성을 분할지점으로 선택할 것인지 첫 번째 질의의 레코드의 수를 통해 결정하게 된다. 이러한 방식으로 분류 체계를 생성한 후 최종적으로 분류된 레코드의 수에 $\mathcal{M}_L$을 이용한 노이즈를 추가하게 된다. 기존의 프라이버시를 고려하지 않은 ID3[29]와 C4.5[30]는 정보획득량 및 이를 통한 획득비율량(Gain ratio)을 이용해 분할시킬 속성을 결정하였다. 이후의 차분 프라이버시를 만족하는 의사결정나무 알고리즘들은 다음과 같이 해당하는 질의에 대응하는 분류체계 생성 시 정보획득량 뿐만 아닌 어떠한 연산(operator)을 사용해야 하는지, 카테고리형 속성이 아닌 수를 지닌 속성들에 대한 분할지점이 어떻게 선택되어야 하는지, 또는 트리의 깊이 등의 파라미터가 얼마나 영향을 미치는지 등을 통해 차분 프라이버시 환경에서 안전성을 만족함과 동시에 더 나은 유용성을 보이는가를 목표로 하고 있다[31]. 이러한 차분 프라이버시 측면에서의 고려사항은 의사결정나무 뿐 만 아니라 다수의 머신러닝 알고리즘 상에서도 해당 알고리즘의 특성에 맞는 기술 연구가 여러 방면으로 많이 진행 되고 있다.[28]

3.3 기타 차분 프라이버시 적용 기술

차분 프라이버시는 데이터 공개 시와 머신러닝 모델 구성 시 널리 사용되며 이에 대한 많은 연구가 진행 중에 있다. 이번 절에서는 데이터 공개 및 학습 시의 차분 프라이버시 뿐 만 아니라 데이터 수집, 비정상 행위 탐지 및 질의 기반의 프로그래밍 언어로써 차분 프라이버시를 제공하는 기술 등에 대해 간략하게 정리하고 분석한다.

3.3.1 프라이버시 보존형 데이터 수집 기술

2장에서 다루었던 차분 프라이버시는 기본적으로 데이터 관리자를 신뢰할 수 있는 개체(trust entity)로 가정한 상태에서 진행되는 모델이다. 즉, 데이터에 접근 가능한 신뢰할 수 있는 개체인 데이터 관리자가 존재하며, 이 때 공격자는 차분 프라이버시를 만족하는 알고리즘의 출력 값에만 접근이 가능한 상황이다. 하지만 로컬 차분 프라이버시(Local differential privacy, LDP)[32]라는 개념은 데이터 제공자가 데이터 관리자를 신뢰하지 않을 시를 고려하는 모델이며, 이러한 경우는 프라이버시 관점에서 데이터에 속

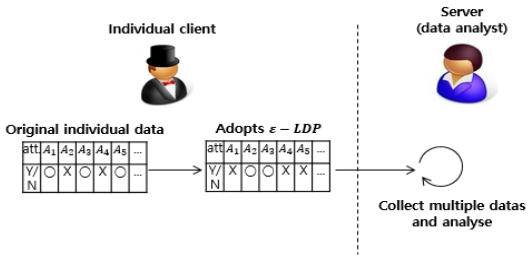

Fig. 7. An example for local differential privacy[32]

한 개인을 단위로 하는 것이 아닌 개인이 보유하고 있는 속성의 실행 유무를 단위로 하게 된다. Fig. 7은 로컬 차분 프라이버시가 적용된 모델의 실행 과정을 나타낸다. 따라서 로컬 차분 프라이버시 모델에서 사용하는 메커니즘은 주로 속성의 유무에 노이즈를 추가하여 응답하게 된다. 이에 대한 대표적인 기술은 Google에서 제공하는 RAPPOR[33]이다. RAPPOR는 chrome 브라우저 사용자들의 특정 웹 사이트 접근 기록에 대해 로컬 차분 프라이버시를 적용하여 안전하게 수집하고 그에 대한 통계를 이용해 악성 소프트웨어 및 웹 페이지의 특성을 파악하여 클라이언트에게 악의적인 사이트를 차단시켜주는 기술을 제공한다.[34] 또한 Apple 역시 개인 데이터 수집 시 차분 프라이버시를 적용하여 iOS 사용자들의 데이터를 이용해 어플리케이션 콘텐츠 순위 결정 등의 서비스를 제공한다.[35] 하지만 Google과 달리 해당 기술에 대한 세부사항은 공개하지 않은 상태이다.

3.3.2 기타 차분 프라이버시 관련 기법

Fan 등[36]에 의해 제안된 기법은 최초로 비정상 행위 탐지에 차분 프라이버시를 적용하였다. 본 기술은 유저로부터 받은 데이터를 군집화하고 $\mathcal{M}_L$을 이용하여 노이즈를 추가한 후 신뢰할 수 없는(untrusted third-party) 서버가 이를 기반으로 비정상 행위 탐지를 수행한다.

또한 차분 프라이버시를 제공하는 질의 기반의 프로그래밍 언어에 대한 연구도 진행 중에 있다. McSherry 등[18]에 의해 제안된 PINQ(Privacy Integrated Queries)는 .NET 기반의 질의 언어인 LINQ(Language-Integrated Query)에 차분 프라이버시를 추가한 기법으로써 해당 언어는 SQL언어와 유사한 동작 구조를 가진다. Reed 등[37]에 의해 제안된 Fuzz와 Gaboardi 등[38]에

의해 제안된 DFuzz 또한 차분 프라이버시를 제공하는 질의 기반의 프로그래밍 언어를 제공한다.

IV. 금융 데이터 상에서의 차분 프라이버시

상기 차분 프라이버시가 적용된 기술들에 대한 분류는 금융 데이터 상에서의 차분 프라이버시 모델 정립의 기반이 되며, 이는 다음과 같은 세 가지 방식으로 크게 나눌 수 있다. 첫 번째로는, 개인 금융 데이터를 수집하는 방식에서의 차분 프라이버시 적용 방안이다. 하지만 현재 대부분의 금융 서비스에서는 사용자가 자신의 정보를 금융 서비스 제공자인 서버에게 전송하게 되면 서버는 특정 개인에게 알맞은 서비스를 제공하게 된다. 따라서 대개 특정 개인의 데이터에 의존하게 되므로 서버를 신뢰할 수 있는 개체로 둘 수 밖에 없으며 이러한 경우에는 로컬 차분 프라이버시 모델이 적합하지 않다. 두 번째로는, 금융사 간 또는 금융사-비금융사 간 데이터 공유 및 결합을 수행 할 시의 차분 프라이버시 적용 방안이다. 이 때 각 업체는 보유하고 있는 데이터를 비식별화하여 데이터 공개를 수행하고, 해당 데이터를 분석하려는 업체는 이를 분석하여 유용하게 사용한다. 따라서 차분 프라이버시를 만족하는 데이터 공개 기법을 적용할 수 있다. 마지막으로, 금융 데이터 상에서의 머신러닝을 이용한 금융 서비스에서의 차분 프라이버시 적용 방안이다. 금융보안원의 보고서에 의하면, 금융 산업은 의료 산업에 이어 인공지능 기술 활용도가 두 번째로 높을 것으로 기대된다. 주로 로보어드바이저, 시장분석, 금융보안, 신용평가 등의 금융서비스에 인공지능 기술이 유용하게 적용 될 것으로 예상하고 있다.[13] 따라서 각각의 금융 서비스에 알맞은 학습 데이터의 프라이버시를 보존하는 차분 프라이버시를 만족하는 머신러닝 모델을 설계할 수 있다.

이번 장에서는 상기 열거한 사항을 기반으로 금융 데이터 상에서의 차분 프라이버시 모델 정립을 목표로 한다. 따라서 기존의 모델을 적용하고 응용하여 공개된 금융 데이터를 이용해 데이터 공유 및 결합을 위한 금융 데이터 공개 및 머신러닝 사용 시의 차분 프라이버시 모델을 정립한다. 또한 공개된 금융 데이터를 이용하여 금융 데이터에 차분 프라이버시가 적용된 경우의 데이터 유용성에 대하여 분석한다. 해당 실험 및 분석은 금융 데이터 공유 모델에 대해 랜덤 샘플링 후 $k-$익명성을 적용하는 버전[27]을 이용하였고 금융 데이터 머신러닝 모델에 대해 ID3[29][39] 버전의 의사결정나무 알고리즘을 이용하였다.

4.1 금융 데이터 공유 및 결합을 위한 차분 프라이버시 모델

서론에서 언급되었던 SK텔레콤의 실증 예는 비식별화 된 데이터가 얼마나 유용하게 쓰일 수 있는지를 직접 보여준 사례라고 할 수 있다. 우선, 비식별화를 위해 속성이 식별자(ID), 준식별자(QID) 및 민감속성(sensitive)로 분류되어야 한다. 이 때 만일 데이터가 식별자 속성을 보유하고 있다면 이는 특정 개인을 정확하게 식별할 수 있으므로 삭제되어야만 하며[40], 이후 차분 프라이버시를 만족하는 데이터 공개 기술을 적용하게 된다. Fig. 8은 3.1절을 기반으로 한 금융 데이터 상에서의 차분 프라이버시 데이터 공개 알고리즘 적용 모델이다. “신용 불량자의 수는 몇 명인가?”와 같은 질의에 보다 정확한 응답이 필요할 시엔 데이터에 대한 합성 테이블을 공개하는 방식을 선택한다. 충분한 데이터 열을 가지고 있는 상태에서 보다 범용적인 분석을 수행할 수 있게 데이터를 공개하려면 랜덤 샘플링을 통해 차분 프라이버시를 만족하는 재현 데이터를 생성하여 공개한다. 두 기법을 통한 모델 설계 시에 큰 차이점은 차분 프라이버시의 파라미터인 ϵ(또는 ϵ,δ)이 어떠한 과정에서 사용되는지에 대한 여부이다. 이는 즉, 차분 프라이버시를 만족하는 메커니즘이 어떠한 과정에서 사용되는 지를 통해 두 모델이 분류된다. 1)합성테이블 생성 후 공개 기법 수행 시엔 차분 프라이버시를 만족하는 분류체계를 생성하여 데이터를 분할하게 되며, 이를 통해 최종적으로 분류 된 데이터에 대해 해당 질의의 응답값에 노이즈 추가 후 합성 테이블을 공개한다. 이와 다르게 2)랜덤 샘플링 기법은 차분 프라이버시를 만족하는 재현 데이터 생성을 기반으로 한다. 따라서 보통 ϵ(또는 ϵ,δ)이 주어지면 이를 기반으로 얼마만큼의 샘플링을 할 것이고, 확률 분포를 유지하기 위

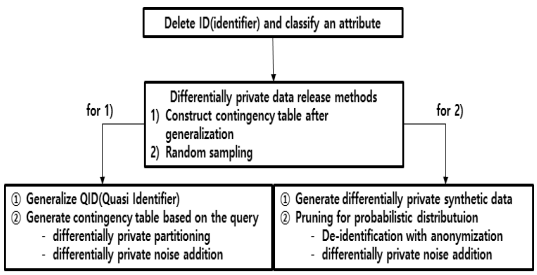

Fig. 8. Classfication for differentially private data release models on financial data

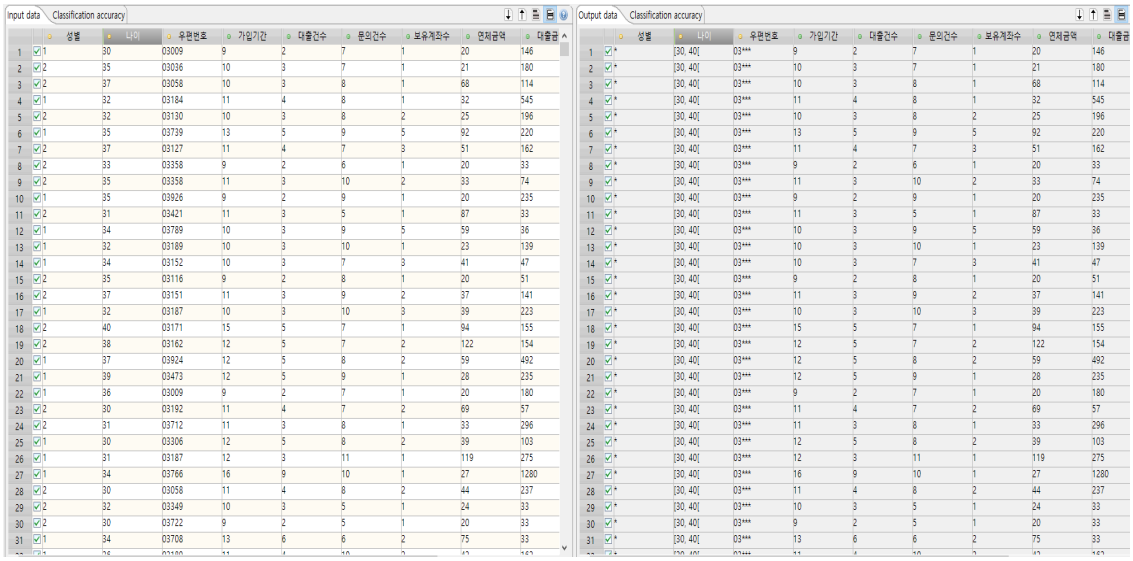

Fig. 9. The original data(left) and a synthetic data using a differentially private synthetic data release method(right)

Table 4. A taxonomity tree for Fig. 9

Attribute	Type	Hierarchy				
Sex (string)	masking	Level1		Level2		
		1(Male) 2(Female)		*		
Zipcode (string)	masking	Level1	Level2	Level3		Level4
		06253	0625*	062**		06***
Age (Integer)	interval	Level1	Level2	Level3	…	Level7
		10, 19, 20, 21, 22, …50, … , 92, 94	[0~29),[30~33), [34~35), … , [67~73), [74~100)	[0~29), [30~37) …, [74~100)	…	[0~100)

해 데이터 열을 얼마만큼 제거 할 것이며, 이를 기준으로 준식별자에 대한 비식별화를 진행하게 된다.

본 절에서는 두 번째 방식의 대표적인 예인 차분 프라이버시를 만족하는 랜덤 샘플링 후 $k-$익명성을 적용하는 기법[27]을 통해 금융 샘플 데이터에 대한 재현 데이터 생성 결과(Fig. 9)를 나타낸다. 해당 기법은 입력 파라미터 ϵ, δ를 기준으로 랜덤 샘플링 확률 γ 및 k를 결정하게 된다. 이 때 QID는 성별, 나이, 우편번호로 설정하며 민감속성은 대출건수, 문의건수, 보유계좌 수 및 예·적금 금액 등으로 이루어져 있다. 해당 분석 시 사용된 QID에 대한 분류체계는 Table 4와 같다. 문자(string)의 성질을 가진 성별 및 우편번호는 마스킹 처리를 하였으며, 수(integer)로 이루어진 나이는 간격(interval)을 두어 분류체계를 생성하였다. 금융 데이터의 특성 상 30세부터 66세까지의 레코드 개수가 데이터의 대부분을 차지하고 있어 일정한 간격이 아닌 약 1000개의 레코드 수를 기준으로 분할하였으며, 상대적으로 레코드 수가 거의 없는 [0~29) 및 [74~100)는 분석 시의 유용성을 위해 더 이상의 일반화 없이 Level4 까지 동일 레벨로 진행된다. 이 때 입력 파라미터인 ϵ, δ는 개인 데이터에 대한 안전성과 분석 시의 유용성 모두를 만족하기 위해 $\epsilon=1, \delta=10^{-5}$를 기준으로 실험을 수행하였다. 이 때 해당 기법은 $\epsilon=1, \delta=10^{-5}$의 안전성을 만족하기 위해 $k=62$로 설정되어진다. 실험 시 사용된 해당 기법은 위와 같은 입력 파라미터와 분류 체계를 이용하여 $(1, 10^{-5})$-차분 프라이버시를 만족하게끔 랜덤 샘플링 될 데이터의 개수를 결정하게 된다. 만일 성별:Level2, 나

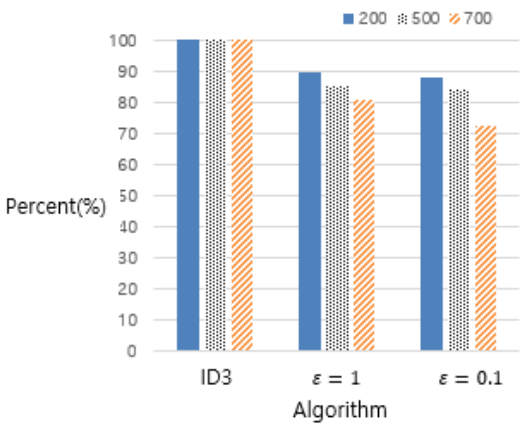

Fig. 10. Utility comparison(%) for non-private ID3[29] and a differentially private ID3[39]

이:Level6 우편번호:Level5로 일반화 레벨이 설정된다면 총 18918개의 데이터 중 12100개(66%)의 데이터가 샘플링 되며, 안전성 측면에서 차분 프라이버시를 만족함과 동시에 유용성 측면에서 민감한 속성들의 속성 값을 그대로 유지하기 때문에(Fig. 9) 데이터 분석가는 매우 범용적인 분석을 수행할 수 있게 된다. 하지만 나이:Level6에 대해 [10, 54), [55, 94)와 같이 간격이 너무 넓게 설정되기 때문에 데이터 분석가 입장에서는 해당 연령별로 데이터 분석을 수행하기 어려운 점이 발생한다. 따라서 일반화 레벨을 낮추어 성별:Level2, 나이 : Level4, 우편번호 : Level4로 설정할 수 있다. 하지만, 이러한 경우 해당 차분 프라이버시를 만족하기 위해 4839개(26%)의 데이터만이 샘플링 되는 것으로 나타났다. 즉, 랜덤 샘플링을 이용한 차분 프라이버시를 만족하는 데이터 공개 기법은 데이터 분석가들이 보다 범용적인 데이터 분석을 원함과 동시에 충분한 데이터 열을 가지고 있을 시 유용하게 사용될 수 있음을 알 수 있다. 하지만, 샘플링 되는 개수를 비율적으로 바라본다면 이는 원본 데이터의 분포를 충분히 만족하지 못할 가능성이 크다. 따라서 이후의 기법들은 주로 이러한 측면에서 원본 데이터의 분포도를 더욱 더 보존하면서 동시에 차분 프라이버시를 만족하기 위한 방향으로 연구 진행 중에 있다.

4.2 금융 데이터 상에서의 차분 프라이버시를 만족하는 머신러닝 모델

금융 산업은 영업 및 마케팅, 투자 관리/트레이딩, 사기 및 부정방지 및 신용 평가/심사 등 여러 서비스를 제공하기 때문에 인공지능 기술 활용도가 매우 높을 것으로 예상된다. 금융보안원의 보고서에 의하면 업무 자동화, 신용 평가, 자산관리 서비스 등 매우 여러 방면에서 머신러닝 알고리즘이 유용하게 사용되고 있으며,[41][13] 이로 인해 금융사 혹은 비금융사들은 각각 자신들이 보유한 고객 데이터를 통해 통계 분석 및 학습을 수행하여 특정 서비스를 위한 머신러닝 기술을 활용하고 있다. 이러한 경우 고객 데이터는 주로 사내에서 활용되나 현재 개인정보 보호법은 당초 개인정보 수집 목적에 상기의 목적이 포함되어 있지 않다면 이를 위한 개별 동의를 받거나 데이터 비식별화 조치를 취하도록 요구하고 있다. 따라서 차분 프라이버시가 적용된 머신러닝 학습 모델이 적용된다면 데이터에 대한 강력한 프라이버시를 보존함과 동시에 기존에 데이터 프라이버시를 고

Algorithm 1. Differentially private ID3 algorithm[39]

1: def DPID3(T, A, C, d, β)
2: Input: Private dataset T, a set of attribute $A=\{A_1, \ldots, A_d\}$, class attribute C, maximal tree depth d and differential privacy budget β
3: Output: Differentially private ID3 taxonomy tree
4: $\varepsilon = \frac{\beta}{2d}$
5: Build_DPID3(T, A, C, d, ε)
6: def Build_DPID3(T, A, C, d, ε)
7: $t = \max_{A_i \in A} |A|$
8: $N_T = \#T + \mathcal{M}_L(\triangle Q/\varepsilon)$
9: if ($A = \varnothing$ or $d = 0$ or $\frac{N_T}{t|C|} < \frac{\sqrt{2}}{\varepsilon}$):
10: for c in C:
11: $T_c = \{\tau \in T | \exists c \in C\}$
12: $N_c = \#T_c + \mathcal{M}_L(\triangle Q/\varepsilon)$
13: return a leaf labeled with $\arg\max_c(N_c)$
14: $\overline{A} = \mathcal{M}_E(T, A, q, \varepsilon)$
15: for i in $\overline{A}$:
16: $T_i = \{\forall \tau \in T | \exists i \in \overline{A}\}$
17: $subtree_i$ = Build_DPID3($T_i, A \setminus \overline{A}$, $c, d-1, \varepsilon$)
18: return a taxonomy tree

려하지 않은 알고리즘과 유사한 효율성을 보일 수 있어 매우 유용하게 사용될 수 있다.

금융 데이터 상에서의 차분 프라이버시 머신러닝 모델 또한 Fig. 5와 같은 구조를 가지게 된다. 설계자는 모델 설계 전 해당 서비스에 알맞은 머신러닝 모델을 선택하게 되며, 온라인 질의를 통해 차분 프라이버시를 만족하는 데이터 학습 이후 이를 기반으로 차분 프라이버시를 만족하는 머신러닝 모델을 설계하게 된다.

Fig. 10은 차분 프라이버시를 만족하는 ID3[39]를 직접 구현한 후 이를 이용하여 UCI 저장소[42]에서 제공하는 독일 금융 데이터(German credit data)를 분류한 결과에 대해 기존의 ID3[29]와의 유용성 비교를 %로 나타낸다. 해당 데이터는 1000개의 데이터 열을 보유하고 있으며 계좌잔액조회, 계좌기간, 신용여부이력 등 총 20개의 속성 값을 가지고 해당 특정 개인에 대해 신용 상태 여부를 클래스로 하여 이를 좋음(Good) 혹은 나쁨(Bad)로 나타낸다. 이 때 학습 데이터 세트의 크기 $t=200,500,700$에 대해 각각의 분류율을 측정하였으며 ID3 알고리즘은 차분 프라이버시를 적용하지 않았을 때, $\varepsilon=1$일 때 및 $\varepsilon=0.1$일 때의 측정 결과를 나타낸다. Algorithm 1은 차분 프라이버시가 적용된 ID3 알고리즘[39]을 나타낸다. 해당 알고리즘은 line 8에서 데이터 개수의 질의 및 line 14에서 분할 지점 선택 질의를 수행한 후, 트리의 말단 노드에서 해당 클래스 별 데이터 개수의 질의 수행 후 최종 클래스를 정하게 된다. 이 때 학습 데이터 세트의 프라이버시 총량, 즉, 학습 데이터 세트의 프라이버시 파라미터 β에 대해 질의의 수 만큼 ε을 분배한다. 이 때 해당 레벨에 대해 각각 두 번의 질의를 수행하므로 정리 3의 $k-$폴드 적응 합성을 만족하게 되면서 동시에 동일 레벨에 대해서는 정리 5인 병렬 합성의 성질을 만족하므로 ε은 $\frac{\beta}{2d}$만큼 분배받게 된다. 해당 알고리즘을 이용한 측정 결과, 많은 속성의 개수 및 적은 데이터 열의 개수로 인해 차분 프라이버시가 적용되지 않은 ID3 알고리즘은 $t=700$을 기준으로 45.51%의 분류 성공률을 보였다. 이에 반해 안전성을 고려하여 설계된 차분 프라이버시를 적용시킨 ID3 알고리즘은 기본적으로 분류되는 데이터 수에 노이즈가 추가되므로 효율성이 낮아지는 것을 알 수 있다. 이 때 $\varepsilon=1$ 일시에는 36.65%의 분류 성공률을 보였으며 $\varepsilon=0.1$ 일시에는 33.05%의 분류 성공률을 보임을 알 수 있다. 하지만 ID3 알고리즘의 분류 성공률인 45.51%에 대비해 각각 80.5% 및 72.6%만큼의 효율성을 보임으로 데이터의 안전성을 강력하게 보장하면서 동시에 기존 ID3와 크게 차이나지 않는 효율적인 분류 성공률을 보이는 것을 알 수 있다. 추가적으로 차분 프라이버시가 적용된 경우 오히려 학습 데이터의 개수가 200개 일 때 가장 높은 것으로 나타났다. 그 이유는 알고리즘 1의 line 14에서 사용되는 스코어 함수인 정보획득량 함수의 민감도 때문이다. 본 알고리즘에서 사용되는 스코어 함수는 다음과 같이 정의된다.

$$q(T,A)=-\sum_{i\in A}\sum_{c\in C}\tau^{A}_{i,c}\bullet\log\frac{\tau^{A}_{i,c}}{\tau^{A}_{i}}$$

따라서 해당 스코어 함수의 민감도는 $\triangle q=\log(T+1)+1/\ln 2$ 이다. 이는 데이터의 개수가 200, 500, 700개 일 때 각각 3.745, 4.142, 4.288의 민감도를 가지게 된다. 이러한 결과는 차분 프라이버시를 만족하는 메커니즘을 사용할 시 해당 함수의 민감도의 값이 클수록 알고리즘의 효율성이 낮아지게 되는 것을 알 수 있다. 따라서 차분 프라이버시를 만족하는 의사결정나무 알고리즘들은 보통 민감도가 데이터의 개수에 비례하는 정보획득량 보다 작은 민감도를 가진 함수를 사용하여 보다 향상된 효율성을 가지게끔 구성한다.[31][39]

결과적으로 차분 프라이버시가 적용된 의사결정나무의 예와 같이 기존의 머신러닝 기술에 차분 프라이버시를 적용한다면 개인 데이터에 대한 프라이버시를 보존하면서 동시에 해당 학습 데이터의 분포를 유지하므로 기존의 프라이버시를 고려하지 않은 알고리즘의 분류도와 크지 않은 차이로 적절한 유용성을 보임을 알 수 있다.

V. 결론 및 향후 전망

본 논문에서는 데이터 비식별화 기법 중 하나인 차분 프라이버시에 대한 연구 및 기술 분류를 통해 금융 데이터 상에서의 차분 프라이버시 모델에 대한 적용 방향을 정립하였다. 차분 프라이버시는 기존의 비식별화 기법과는 달리 공격자의 배경지식과는 무관하게 개인에 대한 강력한 프라이버시를 보존하는 기법이다. 이러한 차분 프라이버시는 데이터 공개 모델 및 머신러닝 모델에서도 데이터의 분

포도에 큰 영향을 미치지 않으면서 동시에 개인 데이터에 대한 프라이버시를 보존하는 특성으로 인해 매우 유용하게 사용될 수 있다. 또한 개인 데이터 수집 등의 기술에도 적용 가능한 장점을 가지고 있다. 본 논문은 이를 바탕으로 금융 데이터 상에서의 차분 프라이버시 적용 방향을 제시하였다. SK 텔레콤의 실증 사례로 보아 비식별화 된 데이터 공개 및 공유는 업체 상호간에 매우 유용하게 사용될 수 있는 것으로 파악되었다. 이에 대해 본 논문은 금융 데이터 상에서 차분 프라이버시를 만족하는 데이터 공개 기법을 크게 두 가지 방식으로 제시하였다. 또한 금융 데이터 기반의 머신러닝 기술 또한 다양한 분야에서 널리 사용 중에 있다. 머신러닝 기술에 적용되는 고객 데이터는 주로 사내에서 활용되지만, 현재 개인정보 보호법에 의해 이에 대한 비식별화 조치가 필요하므로 차분 프라이버시가 적용된 머신러닝 모델이 매우 유용하게 사용될 수 있을 것으로 예상된다.

이 외에도 금융데이터의 여러 특성으로 보아 다음과 같은 분야에 대한 차분 프라이버시 기술 연구가 필요할 것으로 전망된다. 첫 번째는 데이터 수집 환경에서의 로컬 차분 프라이버시 적용 방안이다. 물론 앞서 기술하였듯이 현재는 주로 신뢰할 수 있는 개체가 데이터를 수집한 후 특정 개인에게 알맞은 서비스를 제공하기 때문에 이러한 경우 로컬 차분 프라이버시는 적절하지 않다. 하지만 데이터 수집 시에도 현 개인정보 보호법에 의해 개인정보 제공 동의를 받은 속성에 대해서만 정보 수집 및 활용이 가능하며 이러한 제약조건으로 인해 데이터 분석가 입장에서 이는 데이터를 충분히 활용하지 못할 가능성이 크다. 이와 동시에 금융 고객 입장에서는 신뢰할 만한 개체를 서버로 두고 있다고 해도 자신이 제공 동의한 속성들이 얼마나 공개되고 사용되는지 알 수 없기 때문에 금융 데이터 상에서 로컬 차분 프라이버시를 이용한 개인 데이터 수집을 수행한다면 고객의 개인 데이터의 프라이버시를 보존하면서 동시에 데이터 분석가에게 조금 더 유리한 분석 환경을 제공할 수 있을 것으로 기대된다. 두 번째는 속성 간의 상관관계 분석이다. 4.2절에서 사용된 예제와 같이 금융 데이터는 보통 계좌가입 기간, 신용 이력, 보유한 계좌 수 등 수많은 속성들을 보유하고 있기 때문에 이에 대한 상관관계 분석을 통해 어떠한 속성이 많은 영향을 미치는지가 매우 유용하게 사용될 수 있다. 따라서 이러한 경우에 차분 프라이버시가 적용된다면 개인정보 노출 없이 데이터를 분석할 수 있으므로 이에 대한 관련 기술도 필요한 연구 중 하나일 것으로 예상된다. 세 번째는 시계열 데이터 분석이다. 예를 들어 투자 관련 속성 등은 유동성이 크기 때문에 보통 시계열 데이터로 나타내어진다. 이러한 데이터 분석 시에도 역시 차분 프라이버시 기술을 접목한다면 유용하게 사용할 수 있을 것으로 기대되며 이에 대한 관련 기술 또한 필요한 연구 중 하나일 것으로 예상된다. 마지막으로 비정상 행위 탐지에서도 유용하게 사용될 것으로 예상된다. 3.3절에서 언급하였듯이 비정상 행위 탐지 시의 차분 프라이버시 기술 또한 한 분야로써 활발히 연구 중에 있다. 따라서 현재 금융 서비스 시스템에 알맞게 연구되어 적용된다면 이 또한 유용하게 사용될 수 있을 것으로 기대된다.

References

[1] L.Sweeney, "k-anonymity: A model for protecting privacy," International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, vol. 10, no.5, pp.557-570, 2002.

[2] Office for government Policy Coordination, Prime Minister's Secretariat, Ministry of the Interior and Safety, Korea Communications Commission, Financial Services Commission, Ministry of Science and ICT, Ministry of Health & Welfare, "Guidelines for data de-identification - Guidance on de-identification standard, support and management system,", https://www.privacy.go.kr/inf/gdl/selectBoardArticle.do?nttId=7187&bbsId=BBSMSTR_000000000044&bbsTyCode=BBST01&bbsAttrbCode=BBSA03&authFlag=Y&pageIndex=1&searchCnd=&searchWrd=&replyLc=0&nttSj, June, 2016.

[3] J.Kim, "Presentation of data linkage case of SK Telecom: Creation and distribution demonstration of personal information de-identification data," Seminar on de-identified demonstaration for big data on the fourth industrial revolution, 2017.

[4] A.Machanavajjhala, D.Kifer, J.Gehrke and M.Venkitasubramaniam, "L-diversity: Privacy beyond k-anonymity," ACM

Transactions on Knowledge Discovery from Data (TKDD), vol. 1, no. 1, Article 3, 2007.

〔5〕 C.Wong, J.Li, W.Fu and K.Wang, "(α,k) -anonymity: an enhanced k-anonymity model for privacy-preserving data publishing," Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, pp.754-759, 2006.

〔6〕 N.Li, T.Li, and S.Venkatasubramanian, "t-closeness: Privacy beyond k-anonymity and l-diversity," Data Engineering, 2007. ICDE 2007. IEEE 23rd International Conference on, pp. 106-115, April, 2007.

〔7〕 N.Mohammed, R.Chen, B.Fung and P.S.Yu, "Differentially private data release for data mining," Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, pp.493-501, 2011.

〔8〕 A.Narayanan and V.Shmatikov, "Robust de-anonymization of large sparse datasets," Security and Privacy, IEEE Symposium on, pp. 111-125, May, 2008.

〔9〕 P.Ohm, "Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization," UCLA Law Review, Research Information Network, vol.57, no.6, pp-1701-1777, 2009.

〔10〕 C.Dwork, A.Roth, "The algorithmic foundations of differential privacy," Foundations and Trends® in Theoretical Computer Science, pp.211-407, 2014.

〔11〕 C.Dwork, F.McSherry, L.Nissim and A.Smith, "Calibrating noise to sensitivity in private data analysis, " Third Theory of Cryptography Conference(TCC), vol.3876, pp.265-284, 2006.

〔12〕 C.Park, D.Hong, C.Seo "Differentially private data release method for general use of data," Korea Computer Congress, pp.1036-1038, 2017.

〔13〕 Financial Security Institue, "Present condition on introduction for domestic and foreign financial machine learning techniques," http://www.fsec.or.kr/user/bbs/fsec/42/312/bbsDataView/899.do, 2017.

〔14〕 K.Ligett, "Introduction to differential privacy, randomized response, basic properties," The 7th BIU Winter School on Cryptography, BIU, 2017.

〔15〕 J.Wang, S.Liu and Y.Li, "A review of differential privacy in individual data release," International Journal of Distributed Sensor Networks, vol.11, no.10, 2015.

〔16〕 F.McSherry and K.Talwar, "Mechanism design via differential privacy,", Foundations of Computer Science, pp.94-103, 2007.

〔17〕 C. Dwork, G. N. Rothblum, and S. P. Vadhan, "Boosting and differential privacy," Foundations of Computer Science, pp 51 - 60. 2010.

〔18〕 F.McSherry, "Privacy integrated queries: an extensible platform for privacy-preserving data analysis," Communications of the ACM, vol. 53, no. 9, pp. 89 - 97, 2010.

〔19〕 S.L.Garfinkel, "NISTIR8053: De-identification of personal information," Technical report, National Institute of Standards Technology, 2015.

〔20〕 B.C.Fung, K.Wang, P.S.Yu, "Top-down specialization for information and privacy preservation, " Data Engineering, Proceedings 21st International Conference on IEEE, pp.205-216, 2005.

〔21〕 J.Gardner, L.Xiong, Y.Xiao, J.Gao, A.R.Post, X.Jiang and L.Ohno-Machado, "SHARE: system design and case studies for statistical health information release," Journal of the American Medical Informatics Association, vol.20, no.1, pp.109-116, 2012.

[22] Y.Xiao, L.Xiong, C.Yuan, "Differentially private data release through multi-dimensional partitioning,", Secure Data Management, pp.150-168, 2010.
[23] J.L.Bentley, "Multidimensional binary search trees used for associative searching,", Communications of the ACM, vol.18, no.9, pp.509-517, 1975.
[24] Y.Lim, "Evaluation and future challenges of de-identification techniques," Big data utilization and privacy protection: Information technology solution for object conflicts, Financial Information Society of Korea, Korea Money and Finance Association, Common policy symposium on spring, 2017.
[25] "https://onthemap.ces.census.gov/", OnTheMap.
[26] A.Machanvajjhala, D.Kifer, J.Abowd, J.Gehrke and L.Vilhuber, "Privacy: Theory meets practice on the map," Data Engineering, IEEE 24th International Conference on, pp.277-286, 2008.
[27] N.Li, W.H.Qardaji and D.Su, "Provably private data anonymization:Or, k-anonymity meets differential privacy, " CERIAS Technical Report, 2010.
[28] Z.Ji, Z.Lipton and C.Elkan, "Differential privacy and machine learning: a survey and review," arXiv preprint, 2014.
[29] J.R. Quinlan, "Induction of decision trees," Machine learning, vol.1, no.1, pp.81-106, 1986.
[30] J.R. Quinlan, C4.5: Programs for machine learning, Elsevier, 2014.
[31] S.Fletcher, M.Z.Islam, "Decision tree classfication with differential privacy: A Survey,", arXiv preprint, 2016.
[32] S.P.Kasiviswanathan, H.K.Lee, K.Nissim, S.Raskhodnikova and A.Smith, "What can we learn privately?," SIAM Journal on Computing, vol.40, no.3, pp.793-826, 2011.
[33] U.Erlingsson, V.Pihur and A.Korolova, "RAPPOR: Randomized aggregatable privacy-preserving ordinal response, " Proceedings of the 2014 ACM SIGSAC conference on computer and communications security, pp.1054-1067, 2014.
[34] Google, "Chrome Privacy Whitepaper, "https://www.google.co.kr/intl/ko/chrome/browser/privacy/whitepaper.html
[35] Apple, "guides and sample code," https://developer.apple.com/library/content/releasenotes/General/WhatsNewIniOS/Articles/iOS10.html
[36] L.Fan and L.Xiong, "Differentially private anomaly detection with a case study on epidemic outbreak detection," Data Mining Workshops, IEEE 13th International Conference on, pp.833-840, 2013.
[37] J.Reed and B.C.Pierce, "Distance makes the types grow stronger: a calculus for differential privacy," ACM Sigplan Notices, vol.45, no.9, pp.157-168, 2010.
[38] M.Gaboardi, A.Haeberlen, J.Hsu, A.Narayan and B.C.Pierce, "Linear dependent types for differential privacy," ACM SIGPLAN Notices, vol.48, no.1, pp.357-370, 2013.
[39] A.Friedman and A.Schuster, "Data mining with differential privacy," Proceedings of the 16th ACM SIGKDD International Conference on Konwledge Discovery and Data Mining, pp.493-502, 2010.
[40] J.Gardner and L.Xiong, "HIDE: an integrated system for health information DE-identification," Computer-Based Medical Systems, 2008.
[41] Financial Security Institute, "Survey on machine learning technologies," http://www.fsec.or.kr/user/bbs/fsec/42/312/bbsDataView/355.do?page=7&column=&search=&searchSDate=&searchEDate=&bbsDataCategory=
[42] UCI Repository, "German Credit Data, https://archive.ics.uci.edu/ml/data-

sets/Statlog+%28German+Credit+Data%29

〔43〕 R.Shokri, M.Stronati, C.Song and V.Shmatikov, "Membership inference attacks against machine learning models," Security and Privacy, IEEE Symposium on, pp.3-18, 2017.

〔44〕 M.Fredrikson, S.Jha and T.Ristenpart, "Model inversion attacks that exploit confidence information and basic countermeasures," Proceedings of the 22nd ACM SIGSAC conference on computer and communications security, pp.1322-1333, 2015.

〈저자소개〉

김 현 일 (Hyun-il Kim) 학생회원
2014년 2월: 공주대학교 응용수학과 학사
2014년 3월~2016년 2월: 공주대학교 융합과학과 석사
2016년 2월~현재: 공주대학교 융합과학과 박사과정
〈관심분야〉 암호모듈 구현, 데이터 보호 기술

박 철 희 (Cheolhee Park) 학생회원
2014년 2월: 공주대학교 응용수학과 학사
2014년 8월~2016년 8월: 공주대학교 수학과 석사
2017년 2월~현재: 공주대학교 수학과 박사과정
〈관심분야〉 암호모듈 구현, 데이터 보호 기술

홍 도 원 (Dowon Hong) 종신회원
1994년 2월: 고려대학교 수학과 학사
2000년 2월: 고려대학교 수학과 박사
2000년 4월~2012년 2월: 한국전자통신연구원 팀장, 책임연구원
2012년 3월~현재: 공주대학교 응용수학과 교수
〈관심분야〉 암호기술, 프라이버시 보호기술

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~현재: 공주대학교 의료정보학과 부교수
2016년 현재: 정보보호학회 이사
〈관심분야〉 인증, 개인정보보호, 이상거래탐지, 의료정보보안, 머신러닝