국소 차등 개인정보 보호를 위한 유계 계단 메커니즘에 관한 연구

김준영, 박재민, 이시현
한국과학기술원 전기 및 전자공학부

jun0kim@kaist.ac.kr, jaemin.park@kaist.ac.kr, sihyeon@kaist.ac.kr

A Study on Bounded Staircase Mechanism for Local Differential Privacy

Jun-Yeong Kim, Jae-Min Park, Si-Hyeon Lee
School of Electrical Engineering
KAIST

요 약

인공지능과 사물인터넷, 빅데이터 등 정보통신기술이 널리 활용됨에 따라 많은 사용자 정보가 다양한 기업과 기관들에 수집되어 통계, 마케팅 전략 수립 등에 사용되고 있다. 그러나 이에 따라 수집되는 사용자 데이터로부터 개인정보가 유출될 위험성이 존재한다. 본 논문에서는 입력과 출력의 범위가 제한된 상황에서 차등적 정보보호 강도와 실제 값과의 오차 사이의 trade-off 를 개선하는 유계 계단 메커니즘 (bounded staircase mechanism)을 설계하고 분석한다.

Ⅰ. 서 론

Privacy-utility trade-off (PUT)는 사용자의 개인정보 보호를 위한 입력 데이터 변형과 이에 따른 데이터 수집단 (서버)의 통계적 추론 성능 저하 간의 trade-off 를 일컫는다.

개인정보 보호 메커니즘의 출력 범위가 제한된 상황에서 bounded Laplace mechanism [1]이 활용될 수 있다 [2]. 본 논문에서는 bounded Laplace mechanism 과 같이 출력 범위가 제한되어 있는 동시에 그보다 더 좋은 PUT 를 달성할 수 있는 bounded staircase mechanism 을 제시한다.

Ⅱ. 본 론

1. System Model

본 논문에서는 개인정보 보호 제약이 있는 평균 추정 문제를 고려한다. 고려하는 모델에는 n 명의 사용자가 있으며, 사용자 데이터 $t \in \mathbb{R}$ 은 원래 데이터가 가지는 사전 확률분포 $P_T(t)$에 따라 IID 하게 생성된다. 여기서 P_T 는 $[l,u] \subset \mathbb{R}\,(l<u)$ 위에서 정의된 확률 분포라 가정한다. 생성된 사용자 데이터 t는 잡음 추가 개인정보 보호 메커니즘의 잡음 분포 $P_{X|T}$ 를 통해 변형된 출력 $Y = t + X$ 를 생성한다. 이때, 메커니즘 $P_{Y|T}$ 는 P_T 와 동일한 영역 $[l,u]$ 위에서 정의된 확률 분포라 가정한다. 이후 서버는 사용자들로부터 y 를 수집하고, 이를 활용하여 통계적 추론을 수행한다. 통계적 추론이 정확하기 위해서는 메커니즘의 입력 t 와 출력 y 가 유사해야 한다. 따라서 본 논문에서는 메커니즘 입출력 사이의 평균 제곱 오차가 작을 수 있는 메커니즘을 설계한다.

사용자 데이터 보호를 위한 지표인 ϵ − local differential privacy 는 아래와 같다.

Definition 1. ϵ − local differential privacy [3]

개인정보 보호 메커니즘 $P_{Y|T}$와 임의의 입력 값 $t,t' \in \mathcal{T}$, 출력 값 $y \in \mathcal{Y}$, $\epsilon > 0$에 대하여 아래의 식을 만족한다면, 해당 메커니즘은 $\epsilon - LDP$를 만족한다.

$$\mathrm{P}_{Y|T}(y|t) \le e^{\epsilon} \cdot \mathrm{P}_{Y|T}(y|t'), \forall\, t,t',y$$

서버에서의 통계적 추론 정확도에 대한 지표는 아래 정의된 mean squared error (MSE) loss 를 고려한다.

Definition 2. 평균 제곱 오차

잡음 추가 개인정보 보호 메커니즘 $P_{Y|T}$가 P_T와 동일한 영역 $[l,u]$ 위에서 정의된 확률 분포일 때, $Y = T + X$이며, MSE loss 는 아래와 같이 정의한다.

$$\mathbb{E}_T\left[\mathbb{E}_{Y|t}[(Y-t)^2|T=t]\right] = \mathbb{E}_T\left[\mathbb{E}_{X|t}[X^2|T=t]\right]$$

2. Bounded Staircase Mechanism

본 논문은 메커니즘 입력과 출력의 제약이 없는 상황에서 최적의 PUT 를 가지는 계단 메커니즘 [4]을 입력과 출력이 제약이 동일한 범위를 가지는 상황에 맞게 변형하여 제시한다.

Definition 3. 민감도 [4]

임의의 함수 $q : \mathbb{R} \to \mathbb{R}$ 에 대해 민감도는 아래와 같이 정의된다.

$$\Delta(\mathrm{q}) \coloneqq \max_{t_1,t_2 \in \mathbb{R}} |q(t_1) - q(t_2)|$$

Definition 4. 계단 메커니즘 [4]

임의의 함수 q와 민감도 $\Delta(\mathrm{q})$에 대해, $\gamma \in [0,1]$, $\epsilon > 0$가 주어졌을 때 더해지는 잡음의 확률 밀도함수 $f_{\gamma,\epsilon}$는 아래와 같이 정의된다.

$$f_{\gamma,\epsilon}(x) = \begin{cases} a(\gamma,\epsilon) & x \in [0, \gamma\Delta(\mathrm{q})) \\ e^{-\epsilon}a(\gamma,\epsilon) & x \in [\gamma\Delta(\mathrm{q}), \Delta(\mathrm{q})) \\ e^{-k\epsilon}f_{\gamma,\epsilon}\left(x - k\Delta(\mathrm{q})\right) & x \in [k\Delta(\mathrm{q}), (k+1)\Delta(\mathrm{q})) \; for \; k \in \mathbb{N} \\ f_{\gamma,\epsilon}(-x) & x < 0 \end{cases}$$

여기서, $a(\gamma,\epsilon)$은 다음과 같이 정의된다.

$$a(\gamma,\epsilon) = \frac{1 - e^{-\epsilon}}{2\Delta(\mathrm{q})(\gamma + e^{-\epsilon}(1-\gamma))}$$

Definition 5. 유계 계단 메커니즘 (제안 메커니즘)

임의의 함수 q와 민감도 $\Delta(\mathrm{q})$에 대해 $\gamma \in [0,1]$, $\epsilon > 0$, $\hat{\epsilon} > 0$, $t \in \mathbb{R}$이 주어졌을 때 더해지는 잡음의 확률 밀도함수 $\hat{f}_{\gamma,\epsilon,\hat{\epsilon},t}$는 아래와 같이 정의된다.

$$\hat{f}_{\gamma,\epsilon,\hat{\epsilon},t}(x) = \begin{cases} \dfrac{1}{\hat{a}_t(\gamma,\hat{\epsilon})} f_{\gamma,\hat{\epsilon}}(x) & \text{if } l - t \le x \le u - t \\ 0 & \text{otherwise} \end{cases}$$

여기서, $\hat{a}_t(\gamma,\hat{\epsilon})$은 다음과 같이 정의된다.

$$\hat{a}_t(\gamma,\hat{\epsilon}) = \int_{l-t}^{u-t} f_{\gamma,\hat{\epsilon}}(x)\,dx$$

Theorem 6.

P_T와 $P_{Y|T}$가 동일한 영역 $[l,u]$ 위에서 정의된 확률분포이며, $Y = T + X$이고, LDP 상황이므로 $q(t) = t$라는 가정하에 정의 5 의 bounded staircase mechanism 이 ϵ-LDP 를 항상 만족하기 위한 $\hat{\epsilon}$은 아래 표 1 과 같다.

Case	$\hat{\epsilon}$
$0 \le \gamma < \frac{1}{2}$	$\log\left\{\frac{2(1-\gamma)e^{\epsilon}}{-(\gamma e^{\epsilon} - 1 + 2\gamma) + \sqrt{(\gamma e^{\epsilon} - 1 + 2\gamma)^2 + 8\gamma(1-\gamma)e^{\epsilon}}}\right\}$
$\frac{1}{2} \le \gamma < 1$	$\log\left\{\frac{2(1-\gamma)e^{\epsilon}}{-\gamma e^{\epsilon} + \sqrt{(\gamma e^{\epsilon})^2 + 4\gamma(1-\gamma)e^{\epsilon}}}\right\}$
$\gamma = 1$	ϵ

표 1. ϵ-LDP 를 만족하기 위한 $\hat{\epsilon}$

공간의 제약으로 인해 위 정리의 자세한 증명은 생략한다. 간략하게 언급하자면, bounded staircase mechanism 의 worst case 입력 t_1, t_2에 대해 ϵ-LDP 를 만족하도록 하는 $\hat{\epsilon}$ 의 범위를 계산하여 가장 큰 값을 선정했다.

그림 1 은 설정한 시스템 모델에서 제안한 메커니즘과 standard staircase mechanism 확률밀도함수 비교이다. 입력t는 18 이고, l, u, γ, ϵ은 각각 10, 20, 0.3, 1 이며, 정리 6 을 만족하는 $\hat{\epsilon}$은 약 0.7703 의 값을 갖는다. 그림 1 을 통해 제안한 메커니즘은 standard staircase mechanism 과는 다르게 γ, ϵ 뿐만 아니라 입력 t 에 의해서도 확률 밀도 함수가 달라지는 것을 알 수 있다.

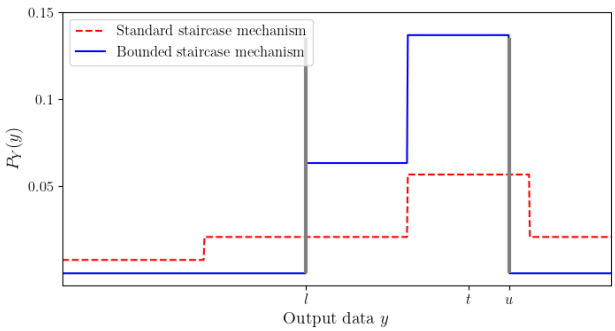

그림 1. 확률밀도함수 비교

3. Numerical Result

설정한 시스템 모델에서 제안한 메커니즘의 PUT 를 확인하고 bounded Laplace mechanism 과의 성능을 비교하기 위해 임의의 데이터를 사용하여 두 메커니즘의 MSE loss 를 실험적으로 확인한다.

유계 메커니즘은 사람의 키와 같이 유의미한 영역을 가지는 데이터에 적용하기에 적절하다 [1]. 본 논문에서는 대한민국 통계청의 2013 년 25~29 세 남성 키 통계 [5]를 토대로 평균 1.758m 표준편차 0.0538 을 가지는 정규분포를 가정하고 100,000 개의 임의의 데이터를 생성했다. 그림 2 는 생성된 데이터의 최소값 1.67m 와 최대값 1.85m 를 l, u로 사용하여 $0.2 \le \epsilon \le 10$에서 MSE loss 를 비교한 결과이다. 그림에서 볼 수 있듯이, 해당 실험 조건 하에서 ϵ 이 $0.2 \le \epsilon \le 10$ 에 포함될 경우 γ 를 $0.16 \le \gamma \le 0.22$ 에서 선정했을 때 bounded Laplace mechanism 보다 제안한 메커니즘이 더 낮은 MSE 를 달성한다는 것을 확인했다.

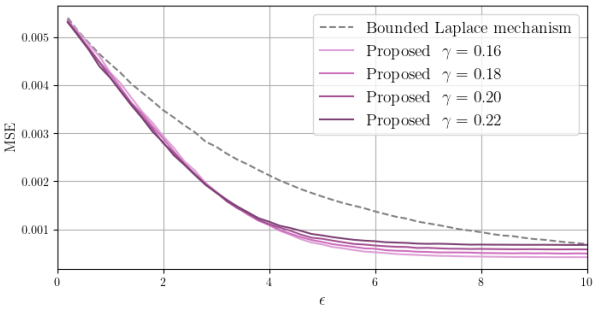

그림 2. 제안한 메커니즘과 bounded Laplace 비교

Ⅲ. 결 론

본 논문에서는 입출력 데이터의 범위가 제한된 상황에서 국소 차등 개인정보 보호와 통계적 추론 성능을 고려한 bounded staircase mechanism 을 제안했으며, 주어진 개인정보 보호 파라미터 ϵ 에 대해 bounded Laplace mechanism 대비 trade-off 가 개선되는 γ영역이 존재하는 것을 실험적으로 보였다.

ACKNOWLEDGMENT

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 지원을 받아 수행된 연구임 [RS-2023-00215700, 신뢰 가능한 메타버스: 블록체인 기반 융합 연구].

참 고 문 헌

[1] N. Holohan, S. Antonatos, S. Braghin, and P. M. Aonghusa, "The bounded Laplace mechanism in differential privacy," arXiv preprint arXiv:1808.10410, 2018.

[2] V. C. Nair, G. M. Gonzalo, and D. Song, "Going incognito in the metaverse: Achieving theoretically optimal privacy-usability tradeoffs in VR," in Proc. 36th Annu. ACM Symp. User Interface Softw. Technol., 2023.

[3] C. Dwork and A. Roth, "The Algorithmic Foundations of Differential Privacy," Now, 2014.

[4] Q. Geng and P. Viswanath, "The optimal noise-adding mechanism in differential privacy," IEEE Trans. Inf. Theory, vol. 62, no. 2, pp. 925-951, Feb. 2016.

[5] 국가통계포털, “2013 년 대한민국 성인 남성 신장,” 2013. [Online]. Available: https://kosis.kr/statHtml/statHtml.do?orgId=113&tblId=DT_113_STBL_1020213