차분 프라이버시 및 지역 차분 프라이버시 개념 및 활용

: 상명대학교 컴퓨터학과 2019104034 박찌진

차례

기존 프라이버시 모델의 한계 : 상명대학교 컴퓨터학과 20190737 박미진

구문적 (syntactic) 프라이버시 모델의 프라이버시 보호 원칙

구문적 (syntactic) 프라이버시 모델? K, l, t 모델

→ "공격자가 특정한 수준 이하의 배경지식을 가지고 있다고 가정할 때, 공격자는 배포된 테이블에서 배경지식 이상의 정보를 얻어서는 안된다."

즉, 공격자가 한 사람의 정보에 대한 배경지식을 가지고 있을 때, 어떠한 일이 발생하면 안된다.

따라서, 원본 데이터를 안전한 데이터로 바꿔야 한다.

구문적 (syntactic) 프라이버시 모델이 사용하는 가정의 한계

→ 또한 테이블의 속성은 식별자, 준식별자, 민감 속성으로 나눌 수 있음

→ 준식별자와 민감 속성을 정확하게 구분하기 어려움

식별자와 준식별자 구분도 마찬가지

→ 공격자는 오직 준식별자 속성값만을 배경지식으로 가지고 있음

→ 공격자는 민감 속성 값도 부분적으로 배경지식으로 가지고 있을 수 있음

한계: 구문적 (syntactic) 프라이버시 모델은 가정을 기반으로 만들어져서 이 가정을 벗어나게 되면, 프라이버시가 위험해진다

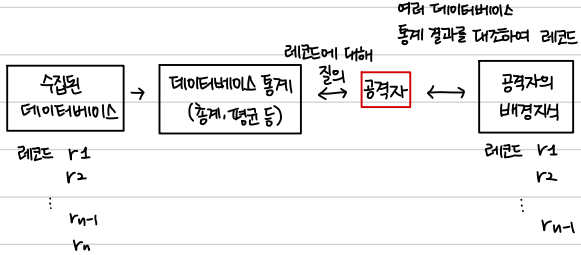
예시)

→ 공격자는 특정한 학생 한 명을 제외한 모든 학생의 몸무게를 배경지식으로 가지고 있음 (n명 중 n-1 명)

→ 평균값(통계적 값)에서 특정한 학생의 몸무게 유추 가능

→ 데이터베이스로부터 생성된 통계에서도 프라이버시 침해 가능

이를 막기 위해 Symantic 프라이버시가 탄생



<데이터베이스통계 배포 및 공격자 모델 예시>

차분 프라이버시 모델의 개념

: 상명대학교 컴퓨터학과 2018104039 박예진

차분 프라이버시 모델 = symantic 프라이버시 모델

Cynthia Dwork 가 2006년에 제안한 프라이버시 모델

차분 프라이버시 모델이 가정하는 상황

→ 원본 데이터베이스는 배포하지 않음

↔ syntactic (k, l, t) 프라이버시 모델 : 원본 데이터를 가공해서 배포

→ 질의를 받아 데이터베이스로부터 결과를 계산하여 반환

user ⟶ query ⟵ answer database

차분 프라이버시 모델의 프라이버시 보호 목표

→ "Release statistical information without compromising the privacy of the individual respondents."

→ 통계 결과로부터 발생하는 프라이버시 침해 방어

차분 프라이버시의 정의 : 상명대학교 컴퓨터학과 그이8이0939 박띠진

인접 데이터베이스 (neighboring database)

→ 임의의 두 데이터베이스 D와 D'에 대하여 두 데이터 베이스가 하나의 레코드 t를 제외한 모든 레코드가 동일할 때, 인접 데이터베이스 관계에 있다고 정의한다

→ D = D' ± t D → n 개의 데이터 D' → n-1개 또는 n+1개의 데이터가 포함

공격자가 인접한 데이터 베이스를 알고 있더라도 특정 한 사람의 프라이버시가 노출되면 안된다

ε-차분 프라이버시 (ε- differential privacy)

→ 데이터베이스로부터 생성한 통계 결과를 무작위로 변조하는 메커니즘 A가 다음식을 만족하면 ε- 차분 프라이버시를 보장

→ $\frac{Pr[A(D)=S]}{Pr[A(D')=S]} \leq e^{\varepsilon}$ $(\varepsilon \geq 0,\ S \in Range(A))$

→ 두 인접 데이터베이스에서 도출한 통계 결과를 변조했을 때, 동일한 결과가 나올 확률의 비율이 크게 차이나지 않아야 한다는 의미

→ ε이 프라이버시 보호 수준을 결정

→ ε = 0 ($e^0 = 1$) 인 경우, 두 인접 데이터베이스에서 동일한 통계 결과가 나올 확률이 같음

→ 공격자는 통계 결과에서 절대 특정 레코드를 추론할 수 없음

한 사람의 정보가 D에 있든 D'에 있든 프라이버시가 노출되지 않고, 안전하다는 의미

→ ε = ∞ ($e^{\infty} > 1$) 인 경우, 두 인접 데이터베이스에서 동일한 통계 결과가 나올 확률의 차이가 커짐

→ 공격자는 통계 결과에서 특정 레코드를 추론하기 쉬움

ε의 값이 클수록 프라이버시 위험이 높아진다.

특징 : 메커니즘 A에 대해 한 사람의 프라이버시가 위험하다, 안전하다 의 정도를 식을 이용해서 정량적, 수학적으로 표현할 수 있다.

라플라스 매커니즘 : 상명대학교 컴퓨터학과 2019104939 박예진

"결과값에 noise를 추가한다."

차분 프라이버시를 보장하는 기초적인 매커니즘

→ 지수 매커니즘 : 질의 결과가 숫자값이 아닌 경우에 사용하는 방법

↔ 라플라스 매커니즘 : 질의 결과가 숫자값인 경우에 사용하는 방법

라플라스 확률 분포에서 추출한 임의의 실수를 통계 결과에 더해서 변조

임의의 실수를 더하므로 주로 통계 결과가 수와 관련된 값일 경우 사용

→ ex) 총계(count), 최소·최대(min / max), 평균(average) 등

데이터베이스D ←— 질의 F —→ user

실제 결과 : F(D) ——→ 변조된 결과 : F(D) + X (← 라플라스 매커니즘)

(X는 라플라스 분포에서 생성한 임의의 실수)

※ 동일한 질문(F)을 여러번 해도 매번 다른 결과값이 나온다.

라플라스 메커니즘 : 상명대학교 컴퓨터학과 2019104934 박찌진

임의의 실수는 어떻게 만들어질까?

라플라스 분포의 확률 밀도 함수 (probability density function)

→ $f(x|\mu, b) = \frac{1}{2b} e^{\left(\frac{|x-\mu|}{b}\right)}$

→ μ : 평균 , $2b^2$: 분산

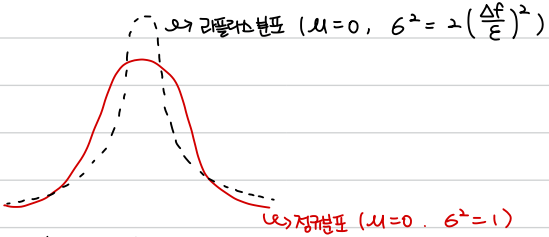
평균이 $\mu=0$ 이고 분산을 $2\left(\frac{\Delta f}{\varepsilon}\right)^2$ 로 하는 라플라스 분포를 사용

→ 평균이 0 이므로 변조된 결과의 기대값 (Expectation value) 은 실제 결과와 동일

→ Δf 는 질의 f 의 전역 민감도 (global sensitivity)

→ $\Delta f = \max_{D, D'} |f(D) - f(D')|$

→ 임의의 레코드 하나가 질의 결과에 미칠 수 있는 최대한의 영향력



<라플라스 분포와 정규 분포>

예시) count를 사용한다면 D는 100명 D' 은 99명의 데이터를 가지고 있다. 이때 $\Delta f = |100-99| = 1$

$\varepsilon=0$ 이면, 라플라스 분포에서 여러가지 값들을 사용한다. → 프라이버시 안전한 편

ε 의 값이 크면 클수록, 라플라스 분포에서 평균값을 빨리 더 많이 사용한다. → 프라이버시 위험한 편

구성 정리 Composition theory : 상명대학교 컴퓨터학과 20190737 박찬진

"각 단계별 ε값과 전체 ε값의 관계가 무엇인지를 나타낸다."

복잡한 알고리즘이 어떠한 차분 프라이버시 보호 기준을 보장하는지 측정하는 것은 어려움

알고리즘의 부분이 차분 프라이버시를 만족하는 것을 증명하면 전체 알고리즘의 프라이버시 보호 수준을 알 수 있음

순차 구성 정리

→ 임의의 알고리즘 A를 구성하는 메커니즘 $A_1, A_2 \cdots A_n$이 각각 $\varepsilon_{i=1,2\cdots n}$ - 차분 프라이버시를 보장하며, 각 메커니즘이 동일한 데이터베이스 적용될 때, A는 $\sum_{i=1}^{n} \varepsilon_i$ - 차분 프라이버시를 보장

결과값 $\xrightarrow{\varepsilon_1}$ 결과값 $\xrightarrow{\varepsilon_2}$ 결과값 $\xrightarrow{\varepsilon_3}$ 최종결과값 최종 ε값 = $\varepsilon_1 + \varepsilon_2 + \varepsilon_3$ 서로 noise 값에 영향을 준다

→ 스무고개와 유사

병렬 구성 정리

→ 임의의 알고리즘 A를 구성하는 메커니즘 $A_1, A_2 \cdots A_n$이 각각 $\varepsilon_{i=1,2\cdots,n}$ - 차분 프라이버시를 보장하며, 각 메커니즘이 서로 다른 데이터베이스 적용될 때, A는 $\max_{i=1,2,\cdots,n} \varepsilon_i$ - 차분 프라이버시를 보장

결과값 → ε_1 → 최종 결과값
→ ε_2 → 최종 결과값
→ ε_3 → 최종 결과값

최종 ε값 = max[ε1, ε2, ε3]

서로 noise값에 영향을 주지 않는다

구성 정리 때문에 일반적으로 ε를 "프라이버시 비용(privacy budget)"이라 부름

→ 전체 프라이버시 보호 수준이 ε일 때, 알고리즘의 각 부분은 ε의 일부를 나눠서 사용해야 함

ε의 값이 크면 클수록 퀄리티가 낮아진다.

퀄리티를 높이기 위해서 적절한 ε의 값을 정해야 한다.

차분 프라이버시를 이용한 프라이버시 보호 데이터 분석 → 대화형 환경 : 상명대학교 컴퓨터학과 2019104394 박찌진

데이터베이스 관리자가 데이터베이스와 데이터 분석가 사이에 존재

관리자는 분석가로부터 질의를 받고 실제 질의 결과를 변조하여 분석가에게 전달

전체 프라이버시 비용 ε 를 설정하고 질의 결과를 반환할 때마다 프라이버시 비용ε을 차감하는 방법을 사용

→ 이유?

→ 순차 구성 정리에 따라 무한정 질의를 받으면 필연적으로 데이터베이스의 모든 정보가 드러남

→ 분석가들끼리 서로 공모가 가능함 (질의 결과를 공유)

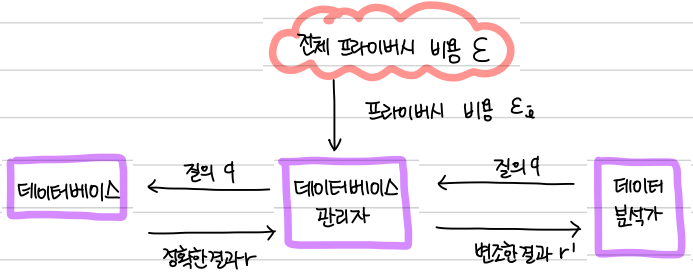
데이터 분석가는 매번 질의를 할 때마다 다른 결과를 얻는다. 계속 질의를 하다보면 정확한 결괏값에 가까워져 유추 가능할 수 있다.

이때, 데이터 베이스 관리자는 질의를 중단한다.

목표

→ 한정된 프라이버시 비용 내에서 많은 질의의 유용성을 보장

대화형 환경은 프라이버시 비용 ε 때문에 비현실적이다.



〈대화형 차분 프라이버시 환경〉

차분 프라이버시를 이용한 프라이버시 보호 데이터 분석 → 비대화형 환경 non-interactive : 상명대학교 컴퓨터학과 201810939 박예진

대화형처럼 질의를 계속 주고받는 것이 아니라 데이터베이스 관리자가 데이터 분석가에게 대량의 데이터를 배포하고, 이를 분석하는 것이다.

데이터베이스 관리자가 데이터 분석가가 사용할 질의에 따라 합성 데이터 (synthetic data)를 생성하여 배포하는 방법

→ 원본 데이터베이스를 바탕으로 히스토그램(histogram)을 생성하여 배포 ex) ppsp

→ 원본 데이터베이스의 레코드를 변조한 익명화 데이터베이스 (anonymized database)를 배포

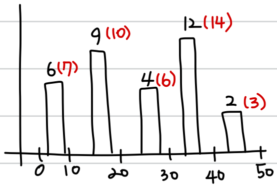
→ 원본 데이터베이스를 바탕으로 비슷한 분포를 가지는 합성 데이터베이스 (synthetic database)를 생성하여 배포

히스토그램 (통계) 배포

가장 많이 사용

→ 데이터베이스의 각 속성을 일정한 구간으로 나누고 특정 구간의 도수에 노이즈를 더해 배포하는 방법

→ 임의의 레코드가 미치는 영향력을 특정 구간으로만 한정할 수 있기 때문에 민감도를 제한할 수 있다는 장점이 있음



히스토그램

프라이버시보호 공간 데이터 통계 배포 → PPSP (Privacy Preserving Statistics Publishing) : 상명대학교 컴퓨터학과 2019109037 박찌진

개인들의 공간/위치 데이터에 대한 통계 정보 배포 → 다양한 빅데이터 활용

예) 유동인구 밀집도 분석, 유동인구 기반 노선 최적화, 유동인구 기반 배차간격 조정

위치 정보는 특정한 개인의 거주지, 근무지, 생활반경 등을 드러내므로 매우 민감한 정보

→ 공격자는 위치 정보를 바탕으로 특정한 개인을 미행, 감시 가능

프라이버시 보호 공간 데이터 통계 배포 → PSD (Private Spatial Decomposition) : 상명대학교 컴퓨터학과 고이화여자 박찬진

PSD : 프라이버스를 고려한 공간분할

→ 특정한 개인의 위치가 드러나지 않도록 생성한 공간 히스토그램

전체 공간을 여러 개의 작은 구역으로 나누고 각 구역에 포함된 사람들의 총계에 라플라스 노이즈를 더하여 생성

예시)

→ 인구분포를 바탕으로 격자 PSD를 생성

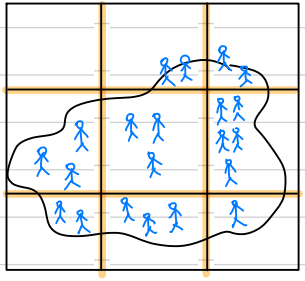
→ 왼쪽의 수는 실제 총계를 의미하며 오른쪽 총계는 라플라스 노이즈를 더한 총계

→ PSD에서 점선의 영역 질의 (range query)의 총계 추정

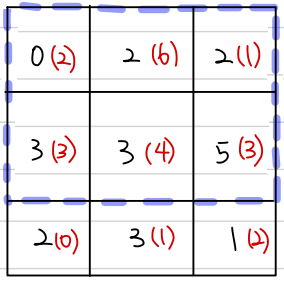
→ (2+6+1+3+4+3) = 19 (실제 총계: 0+2+2+3+3+5 = 15)

영역 질의의 총계는 실제 총계와 오차가 있다.

이러한 오차를 줄이면서 개인의 프라이버시를 보호할 수 있는 방법이 없을까?



〈서울시 인구 분포〉



〈격자 PSD〉

프라이버시 공간 데이터 통계 배포 → PSD (private Spatial Decomposition)

: 상명대학교 컴퓨터학과 2018104394 박찌진

PSD는 임의의 영역 질의에 대해 총계값을 정확하게 추정하는 것을 목표

임의의 영역 질의 q의 유용성은 상대 오차 (relative error)를 통해 계산

→ $\text{RelativeError}(q) = \frac{|RC - VC|}{RC}$

→ RC (Real Count)는 실제 총계, NC (Noised Count)는 PSD를 사용하여 추정한 총계를 의미

→ 앞의 예시의 상대 오차는 $\frac{|15-19|}{15} = \frac{4}{15} = 26.7\%$ 라 할 수 있음

PSD에서 오차가 발생하는 원인

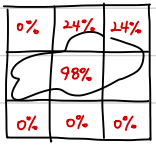
→ 변조 오차 (perturbation error)

→ 각 하위 영역에 더해진 라플라스 노이즈에서 발생

→ 비균일성 오차 (non-uniformity error)

→ 영역 질의와 부분적으로 겹치는 하위 영역에서 발생

→ 하위 영역 내에 사람들이 균등하게 분포하고 있다고 가정하여 추정



〈비균일성 오차〉

PSD 기법의 분류

→ 데이터 비의존적 (data-independent) 기법

데이터를 고려하지 않고, 공간을 구분하는 것

공간의 크기가 일정 하지만 공간 내에 있는 인구 수는 다르거나 같다 → 비균일성 오차 발생

→ 전체 프라이버시 비용 ε을 라플라스 노이즈 생성에 사용

→ 격자 기법 (grid), 쿼드트리 (quadtree) 등

→ 데이터 의존적 (data-dependent) 기법

데이터를 고려하고, 공간을 구분하는 것.

공간의 크기가 일정하지 않지만 공간 내에 있는 인구 수는 모두 동일하다.

→PSD 비균일 공간으로 인한 에러 고려해야 함. 하지만 비균일성 오차가 줄어든다.

→ 전체 프라이버시 비용 ε을 PSD 구조 생성 및 라플라스 노이즈 생성에 나누어 생성

→ kd-트리 (kd-tree), h-트리 (h-tree) 등

지역 차분 프라이버시의 개념 및 정의 LDP: Local Differential Privacy : 상명대학교 컴퓨터학과 2019109039 박찌진

LDP ↔ CDP

지역 차분 프라이버시 (ε-local differential privacy)

→ 데이터 수집 환경에서 사용하는 차분 프라이버시 개념

사용자 —가공 데이터→ 서버

→ 일반적인 차분 프라이버시는 데이터베이스 전체를 알고 있어야 하기 때문에 데이터 수집 환경에 적합하지 않음

→ 인접 데이터베이스 (D, D′) 가 존재하지 않음, 따라서 전역민감도 계산 불가능

∴ 프라이버시 보호를 위해 서버에는 원본 데이터가 아닌 가공된 데이터를 가지고 있기 때문

데이터 제공자가 제공할 수 있는 모든 값의 쌍 v_1, v_2 에 대하여 어떤 임의화 알고리즘 A가 다음의 식을 만족하면 ε-지역 차분 프라이버시 모델을 만족한다고 정의한다.

$$\frac{Pr[A(v_1) \in R]}{Pr[A(v_2) \in R]} \leq e^{\varepsilon} \quad (R \subseteq Range(A))$$

→ ε(ε≥0) 은 프라이버시 보호 수준을 결정하는 매개 변수, R은 메커니즘 A가 도달할 수 있는 모든 결과값의 집합

지역 차분 프라이버시 기반 프라이버시 보호 데이터 수집 : 상명대학교 컴퓨터학과 2019104034 박찌진

확률 기반 응답 (randomized response)

→ 1965년에 Warner 에 의해 제안된 설문조사 방법

→ 민감한 사안에 대한 응답을 수집하기 위해 제안된 방법 ex) 마약여부, 살인여부

→ 확률 기반 응답 수집 방법

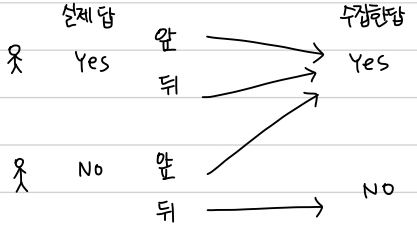
→ 응답자가 응답할 때 피응답자가 알지 못하게 동전을 던짐

→ 뒷면이 나온다면 진실을 말함

→ 앞면이 나온다면 질문에 관계없이 항상 긍정으로 응답함

→ 수집한 긍정 비율 Y' 를 이용하여 실제 긍정비율 Y를 2Y'-1 로 추정 가능

<고전적인 확률기반 응답 모델>



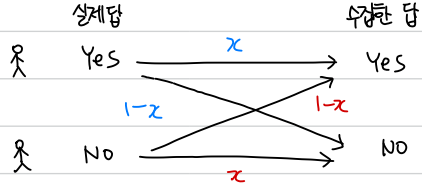
개인은 자신의 프라이버시 보호 걱정을 안할 수 있고, 정보를 수집하는 사람은 정확한 확률통계값을 계산할 수 있다.

→ 응답의 경우의 수가 두 가지인 이진 데이터 확률 기반 응답 모델은 지역 차분 프라이버시를 만족함이 증명됨

→ $\varepsilon = \ln\left(\frac{x}{1-x}\right)$ (x: 진실을 말할 확률)

$\varepsilon=0$ 이면 완벽한 프라이버시를 보장할 수 있다.

<일반화된 이진 데이터 확률 기반 응답 모델>



$\varepsilon=0$, $\frac{Pr[A(V_1)\in S]}{Pr[A(V_2)\in S]} \le e^0 = 1$, if $X=0.5$

$\varepsilon = \ln\left|\frac{0.5}{0.5}\right| = 0$ → 프라이버시 안정

지역 차분 프라이버시 기반 프라이버시 보호 데이터 수집 : 상명대학교 컴퓨터학과 고이화여자 박찬진

이전 데이터 확률 기반 응답의 확장

→ 응답의 경우의 수가 n개인 경우에 대해 확률 기반 응답을 n번 하여 차분 프라이버시를 만족하는 데이터 수집을 할 수 있음

→ 당신은 1번 응답을 선택했습니까? Yes or No ?

→ 당신은 2번 응답을 선택했습니까? Yes or No ?

→ ···

→ 당신은 n번 응답을 선택했습니까? Yes or No ?

→ 예시

→ 당신은 어느 회사의 스마트폰을 사용하십니까?

→ 당신은 삼성의 스마트폰을 사용하십니까? 1 or 0 ?

→ 당신은 애플의 스마트폰을 사용하십니까? 1 or 0 ?

→ 당신은 LG 의 스마트폰을 사용하십니까? 1 or 0 ?

→ 당신은 기타 스마트폰을 사용하십니까? 1 or 0 ?

〈원본 응답〉

0	1	2	···	n
1	0	0	0 0	0

→

〈확률 기반 응답〉

0	1	2	···	n
1	0	1	·····	0

지역 차분 프라이버시 기반 프라이버시 보호 데이터 수집 RAPPOR : 상명대학교 컴퓨터학과 2019104034 박예진

RAPPOR

→ Randomized aggregatable privacy-preserving ordinal response

→ 2014년 Google 에서 발표한 프라이버시 보호 데이터 수집 방법

→ 유한한 경우의 수를 가지는 데이터 수집에 지역 차분 프라이버시를 적용

→ 기존 확률 기반 응답 기법에서 발생하는 문제점 해결

  → 여러 번의 데이터 수집에서도 차분 프라이버시 보장

  → 응답의 경우의 수가 많은 경우에도 성능 보장

〈데이터 수집단계〉 크롬 브라우저 사용자의 시작 페이지

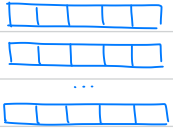
NAVER
Daum
Google

⇒ 해시 함수 h_1 / 해시 함수 h_2 / ... / 해시 함수 h_k

⇒ 블룸 필터 (Bloom filter) [1 | 1 | 0 | ... | 0]

⇒ 변조된 블룸 필터 [0 | 1 | 0 | ... | 1]

〈분포 추정단계〉 모든 사용자들로부터 수집한 블룸 필터



⇒ 라쏘 회귀분석 (Lasso regression)

⇒ 추정분석