Limiting Privacy Breaches in Privacy Preserving Data Mining

Alexandre Evfimievski, Johannes Gehrke, Ramakrishnan Srikant

PODS 2003

Presented by *Yanjie Ze*

June 22, 2022

Outline

Section: Background

Notions

- There are N clients $C_1, \cdots, C_N$ connected to one server.
- Each client C_i has private information x_i.
- C_i sends to the server a modified version y_i of x_i.
- The server collects the modified information and recover the statistical properties.

Notions

- $x_i \in V_X$, finite set
- x_i follows the distribution p_X. The client allows the server to learn it.
- Randomization operator $y = R(x), y \in V_Y$.

$$p[x \to y] := \mathbf{P}[R(x) = y]\,. \tag{1}$$

The Problem

By receiving y_i from C_i, the server learns something about x_i.
The problem is:

- Can we measure how much can be disclosed by y_i about x_i?
- Can we find randomization operators that keep the disclosure limited?

Section: Privacy Breaches

Privacy Breaches: Notions

Let C_i be any client, x_i be its private information. We define a random variable X:

$$\mathbf{P}[X = x] := p_X(x)\,. \tag{2}$$

Now, for the randomized value $y_i = R(x_i)$, we define a random variable Y,

$$\mathbf{P}[Y = y] := \sum_{x \in V_X} \mathbf{P}[X = x] \cdot p[x \to y]\,. \tag{3}$$

Random variables X and Y are dependent, their joint distribution is

$$\mathbf{P}[X = x, Y = y] = p_X(x) \cdot p[x \to y]\,. \tag{4}$$

Privacy Breaches: Notions

Given y_i, the server can better evaluate the probablities of possible values for C_i's private information, by Bayes formula,

$$\mathbf{P}[X = x \mid Y = y_i] := \frac{\mathbf{P}[X = x] \cdot p[x \to y_i]}{\mathbf{P}[Y = y_i]} . \qquad (5)$$

Given a *property* $Q(x)$, where $Q : V_X \to \{\text{true}, \text{false}\}$:

$$\mathbf{P}[Q(X) \mid Y = y_i] = \sum_{Q(x), x \in V_X} \mathbf{P}[X = x \mid Y = y_i] . \qquad (6)$$

Informally, a privacy breach is a situation when the large gap occurs between $\mathbf{P}[Q(x)]$ and $\mathbf{P}[Q(X) \mid Y = y_i]$.

Privacy Breaches: Definition

Definition (Privacy Breach)

We say that there is a ρ_1-to-ρ_2 privacy breach with respect to property $Q(x)$ if for some $y \in V_Y$

$$\mathbf{P}[Q(X)] \leqslant \rho_1 \text{ and } \mathbf{P}[Q(X) \mid Y = y] \geqslant \rho_2.$$

Here $0 < \rho_1 < \rho_2 < 1$ and $\mathbf{P}[Y = y] > 0$.

Privacy Breaches: Example

Suppose that private information x is a number between 0 and 1000. This number is chosen as a random variable X such that:

$$\mathbf{P}[X = 0] = 0.01$$
$$\mathbf{P}[X = k] = 0.00099, \quad k = 1 \dots 1000$$

We give three possible $R(x)$ for randomization.

- Let $R_1(x)$ be x with 20% probability, other number with 80% probability.
- Let $R_2(x)$ be $x + \xi (\mod 1001)$, where ξ is chosen uniformly at random in $\{-100, \cdots, 100\}$.
- Let $R_3(x)$ be $R_2(x)$ with 50% probability, and a uniformly random numuber otherwise.

Privacy Breaches: Example

We give two properties and compute prior and posterior probabilities,

- $Q_1(X) \equiv$ "$X = 0$"
- $Q_2(X) \equiv$ "$X \notin \{200, ..., 800\}$"

Given:	$X = 0$	$X \notin \{200, \ldots, 800\}$
nothing	1%	$\approx 40.5\%$
$R_1(X) = 0$	$\approx 71.6\%$	$\approx 83.0\%$
$R_2(X) = 0$	$\approx 4.8\%$	100%
$R_3(X) = 0$	$\approx 2.9\%$	$\approx 70.8\%$

Figure: Prior and posterior (given $R(x) = 0$) probabilities.

Privacy Breaches: Example

We can conclude that there are two important subclasses of privacy breaches

- very unlikely $\rightarrow$ more likely
- uncertain $\rightarrow$ very certain

Privacy Breaches: Two Subclasses

Definition (Two Subclasses of Privacy Breaches)

We say that there is a straight or upward ρ_1-to- ρ_2 privacy breach with respect to Q_1 if for some $y \in V_Y$

$$\mathbf{P}\left[Q_1(X)\right] \leqslant \rho_1, \quad \mathbf{P}\left[Q_1(X) \mid R(X) = y\right] \geqslant \rho_2.$$

We say that there is an inverse or downward ρ_2-to- ρ_1 privacy breach with respect to Q_2 if for some $y \in V_Y$

$$\mathbf{P}\left[Q_2(X)\right] \geqslant \rho_2, \quad \mathbf{P}\left[Q_2(X) \mid R(X) = y\right] \leqslant \rho_1.$$

Using property $Q_2' = \neg Q_2$, we could write this as

$$\mathbf{P}\left[Q_2'(X)\right] \leqslant 1 - \rho_2, \quad \mathbf{P}\left[Q_2'(X) \mid R(X) = y\right] \geqslant 1 - \rho_1$$

We also say that the breach is caused by the value $y \in V_Y$ from the inequalities; we assume that $\mathbf{P}[R(X) = y] > 0$.

Section: Amplification

Motivation

Two problems if we directly use the definition to check privacy breaches:

- There are $2^{|V_X|}$ possible properties, far too many to check them all.
- We can not use the definition if we do not know p_X, the prior distribution.

γ-Amplification

Definition (γ-Amplification)

A randomization operator $R(x)$ is at most γ-amplifying for $y \in V_Y$ if

$$\forall x_1, x_2 \in V_X : \frac{p\left[x_1 \to y\right]}{p\left[x_2 \to y\right]} \leqslant \gamma$$

here $\gamma \geqslant 1$ and $\exists x : p[x \to y] > 0$. Operator $R(x)$ is at most γ-amplifying if it is at most γ-amplifying for all suitable $y \in V_Y$.

Amplification Condition

Theorem (Amplification Condition)

Let R be a randomization operator, let $y \in V_Y$ be a randomized value such that $\exists x : p[x \to y] > 0$, and let $0 < \rho_1 < \rho_2 < 1$ be two probabilities from Definition 2. Suppose that R is at most γ-amplifying for y. Revealing " $R(X) = y$ " will cause neither upward ρ_1-to- ρ_2 privacy breach nor downward ρ_{2-} to-ρ_1 privacy breach with respect to any property if the following condition is satisfied:

$$\frac{\rho_2}{\rho_1} \cdot \frac{1-\rho_1}{1-\rho_2} > \gamma$$

Proof of Amplification Condition (1)

We prove by contradiction. We assume for property $Q(x)$ we have a ρ_1-to-ρ_2 privacy breach.
Since

$$P[Q(X)] \leq \rho_1 < 1$$

$$P[Q(X)|Y = y] \geq \rho_2 > 0\,,$$

the following definitions make sense:

$$x_1 \in \{x \in V_X \mid Q(x) \text{ and } p[x \to y] = \max_{Q(x')} p\left[x' \to y\right]\}$$

$$x_2 \in \{x \in V_X \mid \neg Q(x) \text{ and } p[x \to y] = \min_{\neg Q(x')} p[x' \to y]\}$$

In words,
x_1: satisfy Q and most likely to be into y
x_2: not satisfy Q and least likely to be into y

Proof of Amplification Condition (2)

By conditional probability

$$\mathbf{P}[Q(X) \mid Y = y] = \sum_{Q(x)} \mathbf{P}[X = x \mid Y = y] =$$

$$= \sum_{Q(x)} \frac{\mathbf{P}[X = x] \cdot p[x \to y]}{\mathbf{P}[Y = y]}$$

$$\leqslant \frac{p\left[x_1 \to y\right]}{\mathbf{P}[Y = y]} \cdot \sum_{Q(x)} \mathbf{P}[X = x] = p\left[x_1 \to y\right] \cdot \frac{\mathbf{P}[Q(X)]}{\mathbf{P}[Y = y]}$$

and in the same way,

$$\mathbf{P}[\neg Q(X) \mid Y = y] = \sum_{\neg Q(x)} \mathbf{P}[X = x \mid Y = y] =$$

$$= \sum_{\neg Q(x)} \frac{\mathbf{P}[X = x] \cdot p[x \to y]}{\mathbf{P}[Y = y]}$$

$$\geqslant \frac{p\left[x_2 \to y\right]}{\mathbf{P}[Y = y]} \cdot \sum_{\neg Q(x)} \mathbf{P}[X = x] = p\left[x_2 \to y\right] \cdot \frac{\mathbf{P}[\neg Q(X)]}{\mathbf{P}[Y = y]}$$

Proof of Amplification (3)

We know that $P[Q(X)|Y = y] \geqslant \rho_2 \geqslant 0$, and thus we divide the two inequalities we get,

$$\frac{\mathbf{P}[\neg Q(X) \mid Y = y]}{\mathbf{P}[Q(X) \mid Y = y]} \geqslant \frac{p\left[x_2 \to y\right]}{p\left[x_1 \to y\right]} \cdot \frac{\mathbf{P}[\neg Q(X)]}{\mathbf{P}[Q(X)]}$$

Remember that $R(x)$ is at most γ-amplifying for y, thus

$$\frac{\mathbf{P}[\neg Q(X) \mid Y = y]}{\mathbf{P}[Q(X) \mid Y = y]} \geqslant \frac{1}{\gamma} \cdot \frac{\mathbf{P}[\neg Q(X)]}{\mathbf{P}[Q(X)]}$$

Proof of Amplification Condition (4)

Due to the privacy breach, we have

$$\frac{1-\rho_2}{\rho_2} \geqslant \frac{1-\mathbf{P}[Q(X) \mid Y=y]}{\mathbf{P}[Q(X) \mid Y=y]}; \quad \frac{1-\mathbf{P}[Q(X)]}{\mathbf{P}[Q(X)]} \geqslant \frac{1-\rho_1}{\rho_1}$$

Thus, combine the ineuqality we gain,

$$\frac{1-\rho_2}{\rho_2} \geqslant \frac{1}{\gamma}\frac{1-\rho_1}{\rho_1}$$

which contradicts to our condition:

$$\frac{\rho_2}{\rho_1} \cdot \frac{1-\rho_1}{1-\rho_2} > \gamma\,.$$

Proof of Amplification Condition (5)

To prove the statement for downward ρ_2-to-ρ_1 breaches, we first represent them as upward ρ_1'-to- ρ_2' breaches with $\rho_1' = 1 - \rho_2$ and $\rho_2' = 1 - \rho_1$, and then note that condition stays satisfied:

$$\frac{\rho_2'}{\rho_1'} \cdot \frac{1 - \rho_1'}{1 - \rho_2'} = \frac{1 - \rho_1}{1 - \rho_2} \cdot \frac{\rho_2}{\rho_1} > \gamma$$

Section: Worst Case Information

Motivation

Review amplification approach:

- be independent on the prior distribution
- depend only on the randomization operator itself

We discuss other ways to restrict disclosure, other privacy measures that depend both on the prior distribution of private data and on the operator.

Mutual Information and its Failure

Mutual Information

Mutual information is defined as

$$
\begin{aligned}
I(X;Y) :&= KL\left(p_{X,Y} \| p_X p_Y\right) = \\
&= \mathop{\mathbf{E}}_{y \sim Y} KL\left(p_{X|Y=y} \| p_X\right)
\end{aligned}
$$

where $KL\left(p_1 \| p_2\right)$ is Kullback-Leibler distance between the distributions $p_1(x)$ and $p_2(x)$ of two random variables:

$$
\begin{aligned}
KL\left(p_1 \| p_2\right) &:= \mathop{\mathbf{E}}_{x \sim p_1} \log \frac{p_1(x)}{p_2(x)} \\
p_{X,Y}(x,y) &:= \mathbf{P}[X = x, Y = y] \\
p_{X|Y=y}(x) &:= \mathbf{P}[X = x \mid Y = y]
\end{aligned}
$$

It is assumed that the larger $I(X;Y)$ is, the less privacy is preserved. Unfortunately, there are situations where privacy is obviously not preserved, but mutual information does not show any sign of trouble. Here is an example.

Failure of Mutual Information (1)

Let our private data be just one bit: $V_X = \{0, 1\}$. Assume that both 0 and 1 are equally likely: $\mathbf{P}[X = 0] = \mathbf{P}[X = 1] = 1/2$.

Failure of Mutual Information (2)

Now consider two randomizations, $Y_1 = R_1(X)$ and $Y_2 = R_2(X)$. The first randomization, given $x \in V_X$, outputs x with probability 60% and outputs $1 - x$ with probability 40% :

$$\mathbf{P}\left[Y_1 = x \mid X = x\right] = 0.6,$$
$$\mathbf{P}\left[Y_1 = 1 - x \mid X = x\right] = 0.4$$

The second randomization R_2 can output 0,1 , or "empty record" e. Whatever its input x is, it outputs e with probability 99.99%, otherwise it outputs x with probability 0.0099% and $1 - x$ with probability 0.0001% :

$$\mathbf{P}\left[Y_2 = e \mid X = x\right] = 0.9999,$$
$$\mathbf{P}\left[Y_2 = x \mid X = x\right] = 0.000099 = 99 \cdot 10^{-6}$$
$$\mathbf{P}\left[Y_2 = 1 - x \mid X = x\right] = 0.000001 = 1 \cdot 10^{-6}$$

Failure of Mutual Information (3)

Intuitively, R_2 is a very poor randomizer since if we see, say, $Y_2 = 1$, then we know with very high probability that $X = 1$:

$$\begin{aligned} \mathbf{P}\left[X = 1 \mid Y_2 = 1\right] = \mathbf{P}\left[X = 0 \mid Y_2 = 0\right] = \\ = \frac{99 \cdot 10^{-6} \cdot 0.5}{99 \cdot 10^{-6} \cdot 0.5 + 1 \cdot 10^{-6} \cdot 0.5} = 0.99 \end{aligned}$$

For Y_1, this probability is only 0.6, which is much more reasonable.

Failure of Mutual Information (4)

What does mutual information indicate, however?

Failure of Mutual Information (5)

Let us compute $KL\left(p_{X|Y_i=y}\|p_X\right)$ for $i=1,2$ and $y=0,1,e$:

$$y=0,1:\log\frac{\mathbf{P}\left[X=y\mid Y_1=y\right]}{\mathbf{P}[X=y]}=\log\frac{0.6}{0.5}\approx 0.2630$$

$$\log\frac{\mathbf{P}\left[X=1-y\mid Y_1=y\right]}{\mathbf{P}[X=1-y]}=\log\frac{0.4}{0.5}\approx -0.3219$$

$$KL\left(p_{X|Y_1=y}\|p_X\right)\approx 0.6\cdot 0.2630-0.4\cdot 0.3219\approx 0.02905$$

$$y=0,1:\log\frac{\mathbf{P}\left[X=y\mid Y_2=y\right]}{\mathbf{P}[X=y]}=\log\frac{0.99}{0.5}\approx 0.9855$$

$$\log\frac{\mathbf{P}\left[X=1-y\mid Y_2=y\right]}{\mathbf{P}[X=1-y]}=\log\frac{0.01}{0.5}\approx -5.6439$$

$$KL\left(p_{X|Y_2=y}\|p_X\right)\approx 0.99\cdot 0.9855-0.01\cdot 5.6439\approx 0.91921;$$

$$y=e,x=0,1:\log\frac{\mathbf{P}\left[X=x\mid Y_2=e\right]}{\mathbf{P}[X=x]}=\log\frac{0.5}{0.5}=0$$

$$KL\left(p_{X|Y_2=e}\|p_X\right)=0$$

Failure of Mutual Information (6)

Now we can compute and compare mutual informations. For Y_1, both of $KL\left(p_{X|Y_1=y} \| p_X\right)$ for $y = 0, 1$ are the same, so the average is

$$I\left(X; Y_1\right) \approx 0.02905$$

For Y_2, the average is

$$I\left(X; Y_2\right) \approx 0.9999 \cdot 0 + 0.0001 \cdot 0.91921 \ll I\left(X; Y_1\right)$$

Thus, counter to intuition, mutual information says that R_2 is more privacy-preserving than R_1.

Worst-Case Information

Problem of Mutual Information

Mutual information averages all Kullback-Leibler distances; however, by looking at these distances without taking the average, some breaches become visible.

Worst-Case Information

Definition (Worst-Case Information)

Let X and Y be discrete random variables. We define worst-case information as follows:

$$I_w(X;Y) := \max_y KL\left(p_{X|Y=y} \| p_X\right).$$

Instead of the logarithm, we can use a different numerical function $f(t)$ as long as $tf(t)$ is a convex function on the interval $t > 0$:

Definition (General Worst Case Information)

Let X and Y be discrete random variables, and let $f(t)$ be a numerical function such that $tf(t)$ is convex on $t > 0$. We define worst-case information with respect to f as follows:
$I_w^f(X;Y) := \max_y KL^f\left(p_{X|Y=y} \| p_X\right)$, where

$$KL^f\left(p_1 \| p_2\right) := \mathop{\mathbf{E}}_{x \sim p_1} f\left(p_1(x)/p_2(x)\right).$$

Bound on Privacy Breaches

Bound on Upward Privacy Breaches

Now we are going to show that knowing worst-case information gives a bound on upward privacy breaches.

Theorem (Bound on Upward Privacy Breaches)

Suppose that revealing $R(X) = y$ for some y causes an upward ρ_1-to- ρ_2 privacy breach with respect to property $Q(X)$. *Then*

$$\rho_2 \cdot f\left(\frac{\rho_2}{\rho_1}\right) + (1-\rho_2) \cdot f\left(\frac{1-\rho_2}{1-\rho_1}\right) \leqslant I_w^f(X; R(X))$$

Proof of Bound

Let us denote $Y = R(X)$, and

$$P_1 = \mathbf{P}[Q(X)], \quad P_2 = \mathbf{P}[Q(X) \mid Y = y]\,.$$

By defitinion we have

$$P_1 \leqslant \rho_1 < \rho_2 \leqslant P_2\,.$$

We define the following notions,

$$q_1 = \rho_2 + \alpha\left(1 - P_2\right), \quad q_2 = \rho_1 - \alpha P_1$$
$$\alpha = \frac{\rho_2 - \rho_1}{P_2 - P_1}$$

Therefore,

$$0 \leqslant \rho_2 \leqslant q_1 \leqslant 1 - \left(P_2 - \rho_2\right) \leqslant 1$$
$$0 \leqslant \rho_1 - P_1 \leqslant q_2 \leqslant \rho_1 \leqslant 1$$

So, q_1 and q_2 can serve as probabilities.

Proof of Bound

Define a Boolean random variable Z that depends on X:

- If $Q(X)$, then Z says "true" with probability q_1
- If $\neg Q(X)$, then then Z says "true" with probability q_2

Proof of Bound

Now compute the prior and posterior probabilities of Z :

$$\begin{aligned}
&\mathbf{P}[Z] = q_1 \cdot P_1 + q_2 \cdot (1 - P_1) = \\
&= P_1 (\rho_2 + \alpha (1 - P_2)) + (1 - P_1) (\rho_1 - \alpha P_1) \\
&= \rho_1 + P_1 (\rho_2 - \rho_1) - \alpha P_1 (P_2 - P_1) \\
&= \rho_1 + P_1 (\rho_2 - \rho_1) - P_1 (\rho_2 - \rho_1) = \rho_1
\end{aligned}$$

analogously,

$$\begin{aligned}
&\mathbf{P}[Z \mid Y = y] = q_1 \cdot P_2 + q_2 \cdot (1 - P_2) = \\
&= P_2 (\rho_2 + \alpha (1 - P_2)) + (1 - P_2) (\rho_1 - \alpha P_1) \\
&= \rho_1 + P_2 (\rho_2 - \rho_1) + \alpha (1 - P_2) (P_2 - P_1) \\
&= \rho_1 + P_2 (\rho_2 - \rho_1) + (1 - P_2) (\rho_2 - \rho_1) = \rho_2.
\end{aligned}$$

Proof of Bound

Corollary (1)

Let X, Y, and Z be discrete random variables such that Z is independent from Y given X, and let $tf(t)$ be convex on $t > 0$.

$$I_w^f(Z; Y) \leqslant I_w^f(X; Y)$$

Proof of Bound

Of course, Z is independent from Y given X, so Corollary 1 is applicable:

$$KL^f\left(p_{Z|Y=y} \| p_Z\right) \leqslant I_w^f(Z;Y) \leqslant I_w^f(X;Y).$$

It remains to check that this inequality is exactly what we are proving. Indeed, denote $I = I_w^f(X;Y)$ and "open up" the definition of KL^f :

$$\begin{aligned} \mathbf{P}[Z \mid Y = y] \cdot f\left(\frac{\mathbf{P}[Z \mid Y = y]}{\mathbf{P}[Z]}\right) + \\ + \mathbf{P}[\neg Z \mid Y = y] \cdot f\left(\frac{\mathbf{P}[\neg Z \mid Y = y]}{\mathbf{P}[\neg Z]}\right) \leqslant I. \end{aligned}$$

Proof of Bound

By $\mathbf{P}[Z] = \rho_1, \mathbf{P}[Z|Y = y] = \rho_2$, we thus have

$$\rho_2 \cdot f\left(\frac{\rho_2}{\rho_1}\right) + (1 - \rho_2) \cdot f\left(\frac{1-\rho_2}{1-\rho_1}\right) \leqslant I_w^f(X; R(X)).$$

The theorem is proved.

Proof of Corollary

Corollary (1)

Let X, Y, and Z be discrete random variables such that Z is independent from Y given X, and let $tf(t)$ be convex on $t > 0$.

$$I_w^f(Z;Y) \leqslant I_w^f(X;Y)$$

Proof.

This is equivalent to prove

$$KL^f\left(p_{Z|Y=y} \| p_Z\right) \leqslant KL^f\left(p_{X|Y=y} \| p_X\right)$$

Proof of Corollary

Lemma

If function $tf(t)$ is convex (or strictly convex) on the interval $t > 0$, then so is function $f(1/t)$.

Let us prove by the definition of KL^f. We shall use Jensen's inequality $\mathbf{E}g(\tau) \geqslant g(\mathbf{E}\tau)$ with respect to function $g(t) = f(1/t)$, which is convex on $t > 0$ by Lemma .

$$KL^f\left(p_{X\,|\,Y=y} \parallel p_X\right) = \mathop{\mathbf{E}}_{x\sim X\,|\,Y=y} f\left(\frac{\mathbf{P}\left[X=x \mid Y=y\right]}{\mathbf{P}\left[X=x\right]}\right)$$

$$= \mathop{\mathbf{E}}_{z\sim Z\,|\,Y=y}\ \mathop{\mathbf{E}}_{x\sim X\,|\,\substack{Z=z\\ Y=y}} f\left(1\Big/\frac{\mathbf{P}\left[X=x\right]}{\mathbf{P}\left[X=x \mid Y=y\right]}\right) \geqslant$$

$$\geqslant \mathop{\mathbf{E}}_{z\sim Z\,|\,Y=y} f\left(1\Bigg/\left(\mathop{\mathbf{E}}_{x\sim X\,|\,\substack{Z=z\\ Y=y}} \frac{\mathbf{P}\left[X=x\right]}{\mathbf{P}\left[X=x \mid Y=y\right]}\right)\right);$$

Proof of Corollary

Using the independence of Z from Y given X,

$$
\begin{aligned}
&\mathop{\mathbf{E}}_{x\sim X \mid \substack{Z=z \\ Y=y}} \frac{\mathbf{P}[X=x]}{\mathbf{P}[X=x \mid Y=y]} = \\
&= \mathop{\mathbf{E}}_{x\sim X} \frac{\mathbf{P}[X=x \mid Z=z, Y=y]}{\mathbf{P}[X=x \mid Y=y]} \\
&= \mathop{\mathbf{E}}_{x\sim X} \frac{\mathbf{P}[Z=z \mid X=x, Y=y]}{\mathbf{P}[Z=z \mid Y=y]} \\
&= \mathop{\mathbf{E}}_{x\sim X} \frac{\mathbf{P}[Z=z \mid X=x]}{\mathbf{P}[Z=z \mid Y=y]} = \frac{\mathbf{P}[Z=z]}{\mathbf{P}[Z=z \mid Y=y]}.
\end{aligned}
$$

The first equality is by expectation, the second is by unrolling conditional probability, the third is by independence.
Thus, we could conclude the proof.

Thanks!