Armonizando Privacidad y Utilidad en la Publicación de Datos mediante Privacidad Diferencial Individual y Microagregación

Jordi Soria-Comas
Universitat Oberta de Catalunya
Rambla del Poblenou 156, Barcelona, 08018, Catalunya
jordi.soria@urv.cat

David Sánchez Ruenes, Josep Domingo-Ferrer,
Sergio Martínez, Luis Del Vasto-Terrientes
Universitat Rovira i Virgili
Departament d'Enginyeria Informàtica i Matemàtiques,
CYBERCAT-Centre de recerca en ciberseguretat de Catalunya,
Av. Paisos Catalans 26, Tarragona, 43007, Catalunya
{david.sanchez@urv.cat, josep.domingo@urv.cat,
sergio.martinezl@urv.cat, luismiguel.delvasto@fundacio.urv.cat}

***Resumen*—La ϵ-privacidad diferencial (DP, por sus siglas en inglés) es un modelo de privacidad muy conocido que ofrece garantías sólidas de privacidad. Sin embargo, cuando se aplica a la publicación de datos, DP deteriora significativamente la utilidad analítica de los datos protegidos. Sin embargo, las implementaciones de DP para la publicación de datos han utilizado parámetros de privacidad débiles (altos valores de ϵ), que diluyen las garantías de privacidad de DP. En este trabajo, abordamos este problema utilizando una formulación alternativa de las garantías de privacidad de DP, llamada privacidad diferencial individual ϵ (iDP), que produce una distorsión menor de los datos a la vez que proporciona la misma protección que DP a los sujetos. Mediante ésta y aplicando un paso de preprocesamiento basado en microagregación para reducir la sensibilidad a los cambios de registros que determina la cantidad de ruido requerido para imponer iDP (y DP) conseguimos un mejor equilibrio entre privacidad y utilidad.**

I. Introducción

El análisis de datos se ha convertido en una herramienta esencial en el mundo actual. Sus aplicaciones van desde la mejora de la experiencia de los usuarios hasta el soporte a la toma de decisiones estratégicas, y en general, puede mejorar sustancialmente la vida humana. Sin embargo, al utilizar datos personales para estos fines, se debe preservar la privacidad de los sujetos. Esto es cada vez más crítico debido a regulaciones de privacidad más estrictas, ejemplificadas por el Reglamento General de Protección de Datos (GDPR) europeo. Por tanto, resulta necesario proteger la información personal antes de ser compartida para su análisis. Uno de los enfoques más comunes es el de la publicación de conjuntos de datos anonimizados.

La protección de la privacidad en publicaciones de microdatos es una disciplina con larga historia. Se desarrolló inicialmente en el contexto de las estadísticas bajo el nombre de Control de Divulgación Estadística (SDC, por sus siglas en inglés). En ese momento, el número de controladores de datos era limitado y los datos se recopilaban bajo un fuerte compromiso de privacidad. Esto permitía hacer suposiciones razonables sobre el conocimiento lateral disponible que podrían utilizar intrusos para llevar a cabo ataques de inferencia que condujeran a la revelación. Tales suposiciones fueron muy útiles para ajustar los métodos de SDC con el fin de obtener una protección adecuada contra la revelación. Sin embargo, el panorama ha cambiado radicalmente tras los avances de las TIC.

DP se ha popularizado recientemente en la publicación de conjuntos de datos anonimizados. Sin embargo, en dicho entorno, DP introduce grandes distorsiones en los resultados protegidos, lo que dificulta significativamente su utilidad analítica. Como resultado, DP solo se ha implementado en una medida limitada en aplicaciones del mundo real y, cuando se ha hecho, los requisitos de privacidad (valor ϵ) se han relajado considerablemente para mantener los datos razonablemente útiles. Un ejemplo de esto es el uso reciente de DP por parte de la Oficina del Censo de EE. UU. para proteger la liberación del Censo Decenal 2020.

Con la finalidad de armonizar la utilidad de los datos con las garantías de privacidad similares a DP, hemos propuesto la privacidad diferencial individual ϵ (ϵ-iDP) [6], un modelo de privacidad que puede incurrir en menos pérdida de información que la DP estándar, al tiempo que proporciona a los sujetos la misma protección que DP. El foco de este trabajo es diseñar mecanismos para usar iDP en publicación de datos y, por lo tanto, beneficiarse de su mejorada relación utilidad-privacidad.

II. De la Privacidad Diferencial a la Privacidad Diferencial Individual

La privacidad diferencial ϵ (ϵ-DP)[2] es un modelo de privacidad bien conocido cuyas garantías de privacidad son independientes del conocimiento lateral disponible para los intrusos. Sin embargo, a diferencia de otros modelos de privacidad desarrollados originalmente para publicación de datos (por ejemplo, k-anonymity[3]), DP fue concebido específicamente para entornos interactivos, en los que solo se protegen los resultados de consultas enviadas a una base de datos alojada por un servidor de confianza.

La privacidad diferencial [2] destaca por las sólidas garantías de privacidad que ofrece. Suponiendo que cada registro corresponde a un individuo, la justificación de DP se explica en [9]: "Cualquier divulgación dada será, dentro de un factor multiplicativo, igual de probable tanto si el individuo participa como si no en la base de datos. Como consecuencia, hay un

riesgo nominalmente mayor para el individuo en participar, y solo una ganancia nominal al ocultar o tergiversar sus datos".

La noción de privacidad diferencial se formaliza de acuerdo con la siguiente definición.

Definición 1 (ϵ-Privacidad diferencial)**.** *Una función aleatorizada κ proporciona ϵ-privacidad diferencial si, para todos los conjuntos de datos D_1 y D_2 que difieren en un registro (también conocidos como conjuntos de datos vecinos), y para todos los $S \subset Range(\kappa)$, tenemos*

$$\Pr(\kappa(D_1) \in S) \leq \exp(\epsilon) \Pr(\kappa(D_2) \in S).$$

Los valores seguros para ϵ están entre 0.01, 0.1 [8]; para valores más grandes, las garantías de privacidad de DP tienden a desaparecer porque los resultados protegidos por DP pueden verse sustancialmente afectados por la presencia o ausencia de cada individuo, lo que aumenta el riesgo de divulgación [6].

Aunque varias distribuciones de ruido son posibles, la distribución de Laplace [2] es la más comúnmente empleada.

En un intento de mejorar la precisión de los datos protegidos, se han propuesto varias relajaciones de DP, como (ϵ, δ)-DP [10], DP concentrado [11], DP concentrado en cero [12] o DP de Rényi [13]. Básicamente, estas relajaciones permiten que las garantías de DP se rompan ya sea por una pequeña cantidad o con una pequeña probabilidad. Una relajación alternativa de DP es iDP [6]. Esta última es particularmente interesante porque, a diferencia de las anteriores, preserva las garantías de privacidad que DP ofrece a los sujetos individuales para un ϵ dado.

Consideremos un sujeto individual I que debe decidir si participa en un conjunto de datos o no. Para neutralizar cualquier reticencia de I a divulgar su información privada, a I se le dice que las respuestas a las consultas basadas en el conjunto de datos no permitirán que nadie aprenda nada que no fuera aprendible sin la presencia de I; esta es precisamente la garantía de privacidad intuitiva que ofrece DP. Para alcanzar tales garantías, DP requiere que la respuesta sea indistinguible entre cualquier par de conjuntos de datos vecinos. Si bien tal requisito produce las garantías de privacidad deseadas, es excesivo, porque no se permite que la parte de confianza aproveche su conocimiento del conjunto de datos.

De acuerdo con la discusión anterior, la privacidad diferencial individual (ϵ-iDP) se define de la siguiente manera.

Definición 2 (ϵ-Privacidad diferencial individual [6])**.** *Dado un conjunto de datos D, un mecanismo de respuesta $\kappa(\cdot)$ satisface ϵ-privacidad diferencial individual (o ϵ-iDP) si, para cualquier conjunto de datos vecino D' de D, y cualquier $S \subset Range(\kappa)$, tenemos*

$$\exp(-\epsilon) \Pr(\kappa(D') \in S) \leq \Pr(\kappa(D) \in S) \leq \exp(\epsilon) \Pr(\kappa(D') \in S).$$

Siguiendo la Definición 1, iDP requiere que la probabilidad de cualquier resultado difiera entre conjuntos de datos vecinos en un factor a lo sumo de $\exp(\epsilon)$. Sin embargo, a diferencia de la Definición 1, el papel de los conjuntos de datos D y D' no es intercambiable: D se refiere al conjunto de datos actual y D' a un conjunto de datos vecino de D. La asimetría entre D y D' es relevante, porque la indistinguibilidad se logra solo entre D y sus conjuntos de datos vecinos. Como efecto secundario de esta asimetría, necesitamos hacer cumplir explícitamente un límite superior ($\Pr(\kappa(D) \in S) \leq \exp(\epsilon) \Pr(\kappa(D') \in S)$) y un límite inferior ($\exp(-\epsilon) \Pr(\kappa(D') \in S) \leq \Pr(\kappa(D) \in S)$). Esto no era necesario en la Definición 1, porque el límite superior se podía obtener a partir del límite inferior intercambiando los roles de D y D'.

Esta diferencia tiene una consecuencia práctica importante y beneficiosa. A diferencia de DP, que requiere calibrar el ruido agregado a la sensibilidad global (es decir, al mayor cambio entre cualquier par de conjuntos de datos vecinos), iDP puede lograrse calibrando el ruido a la sensibilidad local (que normalmente es mucho menor).

Definición 3 (Sensibilidad local [14])**.** *La sensibilidad local de una función de consulta f en un conjunto de datos D es*

$$LS_f(D) = \max_{y:d(y,D)=1} \|f(y) - f(D)\|_1,$$

donde $d(y, D)$ significa que el conjunto de datos y difiere de D en un registro.

El siguiente resultado se demuestra en [6].

Proposición 1. *Sea f una función de consulta que toma valores en $\mathbb{R}^k$. El mecanismo $\kappa(x) = f(x) + (N_1, \ldots, N_k)$, donde N_i son ruidos aleatorios Laplace independientes e idénticamente distribuidos $Laplace(0, LS_f(D)/\epsilon)$, proporciona ϵ-iDP.*

III. Conjuntos de datos iDP mediante Microagregación por Clasificación Individual

Aunque la aplicación natural de DP e iDP es el entorno interactivo, ambos pueden utilizarse para generar conjuntos de datos protegidos mediante enmascaramiento de datos. A continuación, discutimos primero cómo esto se puede lograr para DP estándar, y luego adaptamos el enmascaramiento para aprovechar las ventajas de preservación de utilidad de iDP.

III-A. Conjuntos de Datos DP

Se ha propuesto un enfoque alternativo con el fin de reducir la sensibilidad a cambios en los registros (que es el factor que básicamente determina el ruido necesario para alcanzar DP basado en el enmascaramiento de los valores de atributos de los registros en el conjunto de datos original [4], [5] y, por lo tanto, el daño de utilidad incurrido). Este enfoque aplica un paso de microagregación antes del enmascaramiento.

Para que el enmascaramiento sea una opción viable para generar conjuntos de datos DP, la sensibilidad de los registros individuales debe reducirse. Esto se logró en [5], [15] para DP estándar agregando un paso de microagregación antes de la generación del conjunto de datos DP. Aunque la microagregación fue inicialmente propuesta como una técnica de anonimización en sí [7], en este trabajo la utilizaremos como un medio para reducir la sensibilidad de las consultas.

La microagregación procede en dos pasos: 1) Dividir el conjunto de datos en grupos de registros similares con cardinalidad mayor o igual a k (un parámetro dado) y 2) Calcular un registro representativo de cada grupo y reemplazar cada registro en el grupo por una copia del representante.

El Algoritmo 1 formaliza el proceso. El algoritmo recibe el conjunto de datos original, el presupuesto de privacidad

Algorithm 1 Generación de conjuntos de datos ϵ-DP mediante microagregación univariante y adición de ruido Laplaciano

01 **Requerido**:
02 D : conjunto de datos con atributos $A^1, \ldots, A^m$
03 ϵ_a : presupuesto de privacidad asignado a A^a (con $\epsilon = \sum \epsilon_a$)
04 M_a: algoritmo de microagregación univariante sobre el atributo A^a
05 **Salida**:
06 $\bar{D}_\epsilon$: conjunto de datos ϵ-DP

07 **para** $a = 1$ **hasta** m
08 **sea** $(C_j^a, c_j^a)_j$ el conjunto de grupos y centroides producidos aplicando M_a a A^a
09 **para cada** C_j^a
10 **sea** Δc_j^a la sensibilidad global del grupo c_j^a
11 **sea** n_j^a una muestra de una distribución $Laplace(0, \Delta c_j^a/\epsilon_a)$
12 Reemplace cada $c_j^a \in C_j^a$ por $c_j^a + n_j^a$
13 **fin para**
14 **fin para**
15 **devolver** D

asignado a cada atributo y el algoritmo de microagregación asociado con cada atributo. Para cada atributo, ejecutamos la microagregación (línea 08), calculamos la sensibilidad global de los centroides (línea 10), extraemos el ruido de la distribución de Laplace (línea 11) y enmascaramos cada centroide (línea 12). El algoritmo recibe los algoritmos de microagregación como parámetro. Como se discute en [5], una elección sensata es usar microagregación univariante (formar grupos que contengan valores consecutivos del atributo) y calcular el centroide como la media aritmética de los valores en el grupo.

III-B. Conjuntos de Datos iDP

Para hacer cumplir DP, la propuesta descrita en el Algoritmo 1 necesita utilizar la sensibilidad global de los atributos. Desafortunadamente, las sensibilidades globales pueden ser muy grandes, especialmente cuando los atributos tienen dominios que son mucho más grandes que sus rangos de valores reales en el conjunto de datos original. Esto resulta en la adición de mucho ruido, lo que daña severamente la precisión del conjunto de datos generado. Estos problemas son el resultado de que los mecanismos que hacen cumplir DP no pueden aprovechar el conocimiento de D, debido a la formulación estricta de DP (ver Sección II).

A continuación, presentamos una propuesta para generar conjuntos de datos iDP con el fin de resolver estos problemas.

III-B1. Conjuntos de Datos iDP Usando Sensibilidad Local: Al usar iDP, podemos ajustar el ruido a la sensibilidad local de los atributos en lugar de a su sensibilidad global:

- La sensibilidad local de un centroide de atributo es el cambio máximo en el valor del centroide que puede ocurrir al cambiar de D a un conjunto de datos vecino.
- Por otro lado, la sensibilidad global de un centroide es el máximo de la sensibilidad local entre todos los pares de conjuntos de datos vecinos. Por lo tanto, en general, la sensibilidad global es mayor que la sensibilidad local.

El primer mecanismo que proponemos es una adaptación del Algoritmo 1 a iDP: simplemente reemplazamos la sensibilidad global (Δc_j^a) por la sensibilidad local en $\bar{D}$, que se calcula para cada centroide de grupo ($LS_{c_j^a}(\bar{D})$). Dado que cada centroide se calcula en un conjunto disjunto de registros, mediante composición paralela podemos trabajar con cada centroide de forma independiente. Más específicamente, los cambios en el Algoritmo 1 son:

- En la línea 10, calculamos $LS_{c_j^a}(\bar{D})$, la sensibilidad local del centroide c_j^a en $\bar{D}$.
- En la línea 11, extraemos el ruido n_j^a de una distribución $Laplace(0, LS_{c_j^a}(\bar{D})/\epsilon_a)$.

Para calcular $LS_{c_j^a}(\bar{D})$, se requiere fijar la forma en que se calculan los centroides de atributos, como puede ser la media aritmética de los valores de atributo en el grupo.

Proposición 2. *Sea $C^a = \{C_1^a, \ldots, C_p^a\}$ un agrupamiento de registros creado con respecto a los valores del atributo A^a. Sea $c_j^a = \frac{1}{|C_j^a|} \sum_{x \in C_j^a} x^a$, para $j = 1, \ldots, p$, el centroide asociado con el clúster C_j^a, donde x^a es el valor del registro x para el atributo A^a. La sensibilidad local para cada centroide, $LS_{c_j^a}(\bar{D})$, es*

$$\frac{\max\{\max A^a - \min_{x \in C_j^a} x^a,\ \max_{x \in C_j^a} x^a - \min A^a\}}{|C_j^a|}$$

donde $\max A^a$ y $\min A^a$ son, respectivamente, el máximo y el mínimo del dominio de A^a.

Prueba: La sensibilidad local de c_j^a mide el mayor cambio que puede ocurrir en c_j^a como consecuencia de un cambio en uno de los registros en C_j^a. Dado que el centroide se calcula como la media aritmética de los valores del atributo A^a para todos los registros en el clúster, el cambio más grande en c_j^a ocurre cuando el cambio en el registro para dicho atributo es el mayor.

Específicamente, el cambio máximo en un registro $x \in C_j^a$ se alcanza cuando se cambia el valor del registro x^a por uno de los extremos del dominio: ya sea $\max A^a$ o $\min A^a$. Por lo tanto, podemos expresar la sensibilidad local como

$$LS_{c_j^a}(\bar{D}) = \frac{\max_{x \in C_j^a}\{\max A^a - x^a,\ x^a - \min A^a\}}{|C_j^a|},$$

que es equivalente a

$$\frac{\max\{\max A^a - \min_{x \in C_j^a} x^a,\ \max_{x \in C_j^a} x^a - \min A^a\}}{|C_j^a|}.$$

■

III-B2. Conjuntos de Datos iDP Usando Sensibilidad Local Basada en Agrupamientos: Como se discutió en la Sección III, para calcular las sensibilidades global y local necesitamos que los dominios de los atributos estén acotados. Sin embargo, para algunos atributos (por ejemplo, ingresos), puede que no haya un límite natural y necesitemos acotar artificialmente los atributos para aplicar el algoritmo. Acotar el dominio del atributo a los valores máximo y mínimo en el conjunto de datos es problemático, porque esos valores máximo y mínimo corresponden a individuos específicos y DP está diseñado precisamente para ocultar información sobre cualquier individuo. Además, iDP puede tener los mismos

problemas relacionados con valores atípicos, porque utiliza los límites de los atributos para calcular la sensibilidad local (ver Proposición 2).

En esta sección abordamos estos problemas y mejoramos la generación de conjuntos de datos iDP haciendo que la sensibilidad local de un centroide dependa solo de los valores dentro del grupo correspondiente. Esto tiene dos ventajas importantes: (i) evitamos los problemas de los atributos de rangos ilimitados, y (ii) reducimos la sensibilidad local (y por lo tanto la cantidad de ruido requerido) cuando los valores de los atributos en el grupo no abarcan todo el dominio.

Para hacer que la sensibilidad local de un centroide dependa solo de los valores del grupo asociado, aplicamos el siguiente preprocesamiento al conjunto de datos antes de usar el método descrito en la Sección III-B1. Para cada atributo A^a y grupo C^a_j, seleccionamos un individuo entre aquellos con el valor más pequeño para A^a en C^a_j y reemplazamos su valor por el segundo valor más pequeño de A^a en C^a_j; de manera similar, seleccionamos un individuo entre aquellos con el valor más grande para A^a en C^a_j y reemplazamos su valor por el segundo valor más grande de A^a en C^a_j.

Denotemos por D' el conjunto de datos que resulta de aplicar este paso de preprocesamiento a D. El propósito del preprocesamiento es asegurar que la modificación de un solo registro de D mantenga los valores del clúster C^a_j en D' dentro del rango $[x^a_{min^a_j}, x^a_{max^a_j}]$. De hecho, si modificamos un registro a un valor menor que $x^a_{min^a_j}$, respectivamente, mayor que $x^a_{max^a_j}$ (lo que hace que el valor modificado se convierta en el más pequeño, respectivamente, el más grande en C^a_j), el paso de preprocesamiento reemplazará automáticamente el valor modificado por $x^a_{min^a_j}$, respectivamente, $x^a_{max^a_j}$ (es decir, el anterior valor más pequeño, respectivamente, el valor más grande, que ahora es el segundo valor más pequeño, respectivamente, el segundo valor más grande).

Componer el paso de preprocesamiento con la microagregación puede considerarse como un algoritmo de microagregación alternativo, que esperamos que sea sustancialmente menos sensible a los cambios de los registros en D. Para lograr iDP, necesitamos ajustar el ruido a la sensibilidad de este algoritmo de microagregación alternativo. Dicha sensibilidad se calcula en la siguiente proposición.

Proposición 3. *Sea $C^a = \{C^a_1, \ldots, C^a_p\}$ un agrupamiento de registros creado con respecto a los valores del atributo A^a y supongamos que cada clúster contiene al menos tres valores (no necesariamente diferentes). Sea $P^a = \{P^a_1, \ldots, P^a_p\}$ el agrupamiento que resulta de reemplazar los valores más grandes y más pequeños del atributo A^a en cada clúster por el segundo más grande y el segundo más pequeño valor en el clúster (es decir, los clústeres que resultan de aplicar el paso de preprocesamiento previamente descrito a C^a). Sea $p^a_j = \frac{1}{|P^a_j|}\sum_{x \in P^a_j} x^a$, para $j = 1, \ldots, p$, el centroide asociado con el clúster P^a_j, donde x^a es el valor del registro x para el atributo A^a. La sensibilidad local para cada centroide, $LS_{p^a_j}(D)$, es*

$$\text{máx}\{|x_{max^a_j} - x_{min'^a_j}| + |x_{min''^a_j} - x_{min'^a_j}| + |x_{max^a_j} - x_{max'^a_j}|,$$
$$|x_{min^a_j} - x_{max'^a_j}| + |x_{max''^a_j} - x_{max'^a_j}| + |x_{min^a_j} - x_{min'^a_j}|\} \quad (1)$$

dividido por $|P^a_j|$, donde $x_{min''^a_j}$ y $x_{max''^a_j}$ son, respectivamente, el tercer valor más pequeño y el tercer valor más grande en el clúster.

Prueba: Supongamos el cambio máximo posible de un registro. Esto ocurre cuando el valor más pequeño se vuelve mayor que el valor más grande hasta ahora, o cuando el valor más grande se vuelve menor que el valor más pequeño hasta ahora. Comencemos con el primer caso.

Supongamos que $x_{min^a_j}$ cambia a un valor mayor que el valor más grande $x_{max^a_j}$. Inicialmente, el valor preprocesado era $x_{min'^a_j}$ y se convierte en $x_{max^a_j}$ después del cambio (ya que $x_{max^a_j}$ es el segundo valor más grande del clúster en el conjunto de datos modificado). Además, después del preprocesamiento, el anterior segundo valor más pequeño $x_{min'^a_j}$ cambia al tercer valor más pequeño $x_{min''^a_j}$ y $x_{max'^a_j}$ se convierte en $x_{max^a_j}$. Por lo tanto, el cambio en la suma de los valores del clúster preprocesado es

$$|x_{max^a_j} - x_{min'^a_j}| + |x_{min''^a_j} - x_{min'^a_j}| + |x_{max^a_j} - x_{max'^a_j}|. \quad (2)$$

De manera simétrica, si $x_{max^a_j}$ cambia a un valor menor que el valor más pequeño $x_{min^a_j}$, el cambio en la suma de los valores del clúster preprocesado es

$$|x_{min^a_j} - x_{max'^a_j}| + |x_{max''^a_j} - x_{max'^a_j}| + |x_{min^a_j} - x_{min'^a_j}|. \quad (3)$$

Consideremos ahora cambios en el segundo valor más pequeño o segundo valor más grande, que también son relevantes para el rango del clúster preprocesado. Si el segundo registro más pequeño cambia de manera máxima, es decir, se vuelve mayor que el valor más grande hasta ahora, el clúster preprocesado tendrá valores en el rango $[x_{min''^a_j}, x_{max^a_j}]$. Por lo tanto, el cambio en la suma de los valores del clúster preprocesado es el mismo que en la Expresión (2). De manera simétrica, si el segundo valor más grande cambia de manera máxima, el cambio en la suma de los valores es el mismo que en la Expresión (3).

Así, el cambio máximo en la suma de los valores del clúster preprocesado es el máximo de las Expresiones (2) y (3). Para obtener la sensibilidad, debemos dividir por la cardinalidad $|P^a_j|$ del clúster. Esto concluye la prueba. ■

IV. Evaluación Experimental

Esta sección presenta la evaluación empírica de las dos estrategias de microagregación que proponemos para generar conjuntos de datos iDP. Se han utilizado dos conjuntos de datos:

- Censo (census), extraído de la Encuesta de Población Actual de 1995 de la Oficina del Censo de los Estados Unidos [16]. Contiene 1,080 registros con atributos numéricos y ha sido ampliamente utilizado para evaluar métodos de preservación de la privacidad [4], [5], [17]. En nuestros experimentos, se han utilizado los siguientes atributo: AFNLWGT, AGI, EMCONTRB, FEDTAX, STATETAX, TAXINC, POTHVAL, INTVAL y FICA.
- Calidad del Vino (Wine Quality), es un conjunto de datos utilizado comúnmente para tareas de clasificación/regresión encontrado en UCI. Utilizamos el conjunto de datos de vinos blancos, que contiene significativamente más instancias que el conjunto de datos de

vinos tintos (4,898 vs. 1,600). Los atributos son discretos o numéricos, y describen propiedades fisicoquímicas del vino.

Todos los atributos de los dos conjuntos de datos que utilizamos representan magnitudes numéricas no negativas y la mayoría de ellos no tienen límites superiores. Para calcular las sensibilidades (global y local), fijamos los dominios de los atributos como

$$[0, \ldots, (\alpha \times \text{max_valor_de_atributo_en_conjunto_de_datos})],$$

y tomamos $\alpha \in \{1.5, 3.0\}$. Al variar α probamos la influencia del tamaño de los dominios de los atributos en la sensibilidad que se aplica para cada método. Dado que la distribución de Laplace toma valores en el rango $(-\infty, +\infty)$, para mantener la consistencia, limitamos las salidas con ruido agregado a los rangos de dominio definidos anteriormente.

A continuación, presentamos la medición de la pérdida de información incurrida por nuestros métodos respecto a la DP estándar con métricas generales.

IV-A. Evaluación de la Pérdida de Información

La pérdida de información se refiere a la diferencia entre los datos enmascarados y los datos originales, es decir, al daño infligido por el enmascaramiento a la precisión analítica de los datos originales. La Suma de Errores al Cuadrado (SSE, por sus siglas en inglés) es una medida estándar de la pérdida de información [1], y se define como la suma de los cuadrados de las distancias entre los registros originales y enmascarados:

$$SSE = \sum dist(r_i, (r_i)')^2, \tag{4}$$

donde r_i es el i-ésimo registro en el conjunto de datos original y $(r_i)'$ representa su versión enmascarada. La SSE media, calculada dividiendo la SSE por el número de registros n, es generalmente más informativa porque no depende de la cardinalidad del conjunto de datos.

Para calcular la SSE, necesitamos una distancia entre registros:

$$d((x_1, ..., x_m), (y_1, ..., y_m)) =$$
$$= (1/m)\sqrt{(d_1(x_1, y_1)/\sigma_1^2)^2 + \ldots + (d_m(x_m, y_m)/\sigma_m^2)^2}, \tag{5}$$

donde $d_j(\cdot, \cdot)$ es la distancia entre los valores del j-ésimo atributo, σ_j^2 es la varianza muestral del j-ésimo atributo en el conjunto de datos original y m es el número de atributos.

Para la privacidad diferencial, hemos considerado valores de ϵ de 0.01, 0.1 y 1.0, que cubren el rango de valores razonablemente seguros [8]. Para evaluar la influencia del paso de microagregación, y considerando la cardinalidad de los conjuntos de datos *Censo* y *Calidad del Vino*, hemos tomado el parámetro k entre 3 y 100 para el primer conjunto de datos y entre 3 y 400 para el último.

Para contextualizar los resultados obtenidos con nuestros métodos iDP utilizando microagregación y sensibilidad local (*iDP-LS*) e iDP utilizando microagregación y sensibilidad local basada en clústeres (*iDP-CBLS*), los comparamos con las siguientes alternativas:

- Adición de ruido Laplaciano simple para privacidad diferencial ϵ (*DP*) sin microagregación previa, como se describe en la Sección II. Este es el mecanismo ingenuo para producir conjuntos de datos diferencialmente privados. Lo utilizamos como límite superior para la pérdida de información.
- Privacidad diferencial utilizando microagregación univariante (*DP-UM*), como se describe en la Sección III-A. Este mecanismo utiliza el paso de microagregación previo para reducir la sensibilidad, aunque el ruido aplicado a los centroides necesita ajustarse a la sensibilidad global.

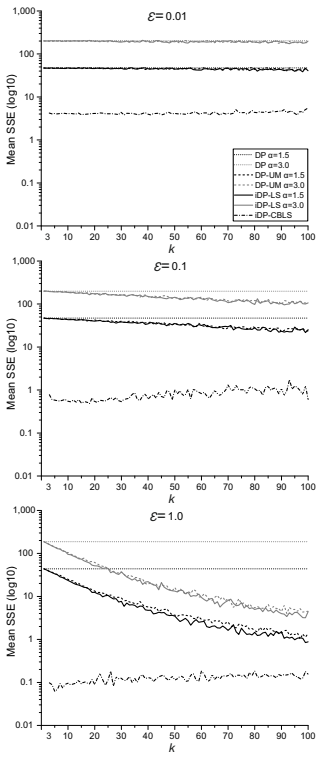


Figura 1. Conjunto de datos *Censo*: SSE promedio para los métodos propuestos (*iDP-LS* e *iDP-CBLS*) y baselines (*DP-UM* y *DP*)

Para cada conjunto de datos, cada método mencionado anteriormente y cada parametrización del método (en términos de valores de α, k y ϵ), realizamos 10 ejecuciones y calculamos la SSE media sobre ellas. El resultado se muestra en las Figuras 1 y 2. Observa que el método de adición de ruido Laplaciano simple (*DP*) se muestra como una línea horizontal, porque no implica microagregación y, por lo tanto, es independiente de k. Además, dado que los valores de SSE son bastante diversos entre los diferentes métodos, se ha utilizado una escala $\log_{10}$ en las ordenadas.

De los resultados obtenidos para los dos conjuntos de datos podemos comentar lo siguiente:

- La adición de ruido Laplaciano simple (*DP*), sin microagregación previa, resulta en la SSE más alta debido a la gran sensibilidad global. De hecho, la SSE apenas disminuye al pasar de $\epsilon = 0.01$ a $\epsilon = 1.0$ y, por lo tanto, podemos considerar que los datos enmascarados son casi aleatorios en todos los casos.
- Tanto *DP-UM* como *iDP-LS* arrojan resultados similares, con SSE disminuyendo a medida que aumenta el nivel de microagregación (k). Esto muestra los beneficios

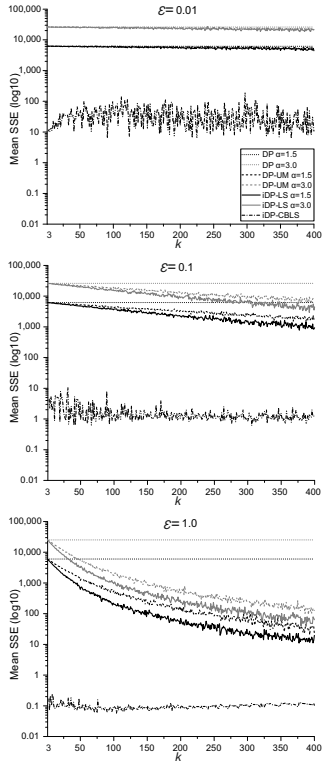


Figura 2. Conjunto de datos *Calidad del vino*: SSE promedio para los métodos propuestos (*iDP-LS* e *iDP-CBLS*) y baselines (*DP-UM* y *DP*)

del paso de microagregación previo para reducir la sensibilidad. *iDP-LS* logra una SSE más baja gracias a su capacidad para utilizar la sensibilidad local en lugar de la sensibilidad global empleada por *DP-UM*. Las diferencias entre los dos métodos son mayores para valores más altos de ϵ, pero permanecen proporcionales al variar los límites de los dominios de los atributos (α). En este sentido, aunque *iDP-LS* utiliza la sensibilidad local, aún depende de los límites de los dominios, como se afirma en la Proposición 2.

- Para los tres métodos anteriores, los valores de SSE disminuyen a medida que los dominios de los atributos se hacen más pequeños (menor α). Esto ilustra que los dominios grandes (o incluso ilimitados) deterioran gravemente la utilidad de los datos enmascarados cuando se utiliza el enfoque directo para generar conjuntos de datos diferencialmente privados.
- Nuestra estrategia más sofisticada (*iDP-CBLS*) es capaz de mejorar la precisión de los métodos anteriores en varios órdenes de magnitud. De hecho, su SSE para $\epsilon = 0.01$ es similar a las de *DP-UM* e *iDP-LS* para $\epsilon = 1.0$ y grandes k, lo que muestra que *iDP-CBLS* puede adaptarse a requisitos de privacidad más estrictos (menor ϵ).

V. Conclusión

Las sólidas garantías de privacidad de DP solo se mantienen para valores pequeños de ϵ, lo que generalmente resulta en una precisión de datos limitada en entornos no interactivos. Para abordar este problema, hemos aprovechado las ventajas de iDP en cuanto a la utilidad de los datos, junto con la reducción de sensibilidad (y ruido) habilitada por el enmascaramiento DP basado en microagregación. Se han propuesto dos estrategias de microagregación, de las cuales la segunda (*iDP-CBLS*) es la que aporta los mayores beneficios en cuanto a utilidad. Por lo tanto, nuestro método es capaz de armonizar pequeños valores de ϵ (digamos $\epsilon < 0.1$) que son necesarios para que DP ofrezca una privacidad real, con la precisión de datos que se necesita para que las publicaciones de datos sean útiles.

Agradecimientos

Esta investigación fue financiada por la Comisión Europea (proyecto H2020-871042 "SoBigData++"), el Gobierno de Catalunya (Premios ICREA Acadèmia a J. Domingo-Ferrer y D. Sánchez y 2021SGR-00115), MCIN/AEI/ 10.13039/501100011033 y "ERDF A way of making Europe" por medio del proyecto PID2021-123637NB-I00 "CURLING", e INCIBE y Unión Europea NextGenerationEU/PRTR (proyecto "HERMES" e INCIBE-Cátedra de Ciberseguridad URV).

Referencias

[1] A. Hundepool, J. Domingo-Ferrer, L. Franconi, S. Giessing, E.S. Nordholt, K. Spicer, P.P. de Wolf, Statistical Disclosure Control, Wiley, 2012.

[2] C. Dwork, F. McSherry, K. Nissim, A.D. Smith, Calibrating noise to sensitivity in private data analysis, in: 3rd Theory of Cryptography Conference, Springer, 2006, pp. 265–284.

[3] P. Samarati, 2001, Protecting respondents' identities in microdata release, IEEE Transactions on Knowledge and Data Engineering 13(6)1010–1027.

[4] J. Soria-Comas, J. Domingo-Ferrer, D. Sánchez, S. Martínez, 2014, Enhancing data utility in differential privacy via microaggregation-based k-anonymity, The VLDB Journal 23(5)771–794.

[5] D. Sánchez, J. Domingo-Ferrer, S. Martínez, J. Soria-Comas, 2016, Utility-preserving differentially private data releases via individual ranking microaggregation, Information Fusion 30(1)1–14.

[6] J. Soria-Comas, J. Domingo-Ferrer, D. Sánchez, D. Megías, 2017, Individual differential privacy: a utility-preserving formulation of differential privacy guarantees, IEEE Transactions on Information Forensics and Security 12(6)1418–1429.

[7] J. Domingo-Ferrer, J.M. Mateo-Sanz, 2002, Practical data-oriented microaggregation for statistical disclosure control, IEEE Transactions on Knowledge and Data Engineering 14(1)189–201.

[8] C. Dwork, 2011, A firm foundation for private data analysis, Communications of the ACM 54(1)86–95.

[9] C. Dwork, Differential privacy, in: Automata, Languages and Programming (ICALP), Springer, 2006, pp. 1–12.

[10] C. Dwork, K. Kenthapadi, F. McSherry, I. Mironov, N. Naor, Our data, ourselves: privacy via distributed noise generation, in: Advances in Cryptology - EUROCRYPT, Springer, 2006, pp. 486–503.

[11] C. Dwork, G.N. Rothblum, Concentrated differential privacy, CoRR abs/1603.01887, 2016, http://arxiv.org/abs/1603.01887

[12] M. Bun, T. Steinke, Concentrated differential privacy: Simplifications, extensions, and lower bounds, in: Theory of Cryptography Conference, Springer, 2016, pp. 635–658.

[13] I. Mironov, Rényi differential privacy, in: IEEE 30th Computer Security Foundations Symposium (CSF), IEEE, 2017, pp. 263–275.

[14] K. Nissim, S. Raskhodnikova, A. Smith, Smooth sensitivity and sampling in private data analysis, in: Proceedings of the 39th Annual ACM Symposium on Theory of Computing, 2007, pp. 75–84.

[15] J. Soria-Comas, J. Domingo-Ferrer, Differentially private data sets based on microaggregation and record perturbation, in: Modeling Decision for Artificial Intelligence (MDAI), Springer, 2017, pp. 119–131.

[16] R. Brand, J. Domingo-Ferrer, J.M. Mateo-Sanz, Reference data sets to test and compare SDC methods for protection of numerical microdata, CASC Project Deliverable, 2002, https://research.cbs.nl/casc/CASCrefmicrodata.pdf/ (accessed 5 June 2024).

[17] M. Laszlo, S. Mukherjee, 2005, Minimum spanning tree partitioning algorithm for microaggregation, IEEE Transactions on Knowledge and Data Engineering 17(7)902–911.