mdt.

Ex Datis, Claritas

Tipos de privacidad diferencial: global y local

Privacidad diferencial global (o central) y local: en qué se diferencian, qué ventajas e inconvenientes tiene cada una y cómo elegir entre ellas.

Manuel Delgado Tenorio · 28 de abril de 2019

Publicado originalmente en www.manueldelgado.com/blog/tipos-privacidad-diferencial-global-local



Los sistemas que cumplen con la definición de la privacidad diferencial se dividen en dos grandes grupos, los que aplican privacidad local y los que aplican privacidad global (o central).

Cada uno de estos enfoques cuenta con sus ventajas y desventajas. Conocer esos matices es esencial para elegir el esquema de nuestro sistema. Veamos sus características y diferencias.

Privacidad diferencial global

Decimos que un sistema adopta la privacidad diferencial desde una perspectiva global (algunos la llaman "central") cuando **en él existe un repositorio de datos en bruto**, es decir, sobre los que aún no se ha aplicado el ruido.

El propietario/responsable del repositorio puede usar los datos para sus propósitos analíticos en su forma original. Sin embargo, cuando da acceso a esos datos a terceras partes, aplica el ruido sobre los datos que salen de su sistema, protegiendo la privacidad de quienes proporcionaron originalmente la información.

Privacidad diferencial local

Por su parte, decimos que la privacidad diferencial **es local si se aplica el ruido sobre los datos en origen**, en el mismo punto en que se recopila el dato.

En el enfoque local no existe, por tanto, una versión de los datos "real" y "limpia". Lo que se recopila, centraliza y analiza es siempre la versión "sucia" de los datos, pasada por el algoritmo correspondiente.

Ventajas y desventajas

Cada caso de uso se beneficiará de las peculiaridades de cada uno de los dos enfoques.

Con el esquema global, **el propietario del sistema tiene acceso a una versión real y limpia de los datos.** Esto permite que todo análisis realizado por el propietario del sistema sea preciso, incluso aunque sea sobre un subconjunto pequeño.

De forma simétrica, **en un sistema con privacidad diferencial local, nadie dispone de una versión limpia de los datos.** Ni siquiera el propietario del sistema. Puesto que todo dato tiene añadido su correspondiente ruido, ninguno es fiable por separado. Hacen falta grandes cantidades de datos para que el ruido añadido por el sistema pueda ignorarse y la información sea útil. Por eso, el enfoque local sólo es recomendable cuando prevemos recopilar datos a gran escala.

En el enfoque global, el propietario del sistema mantiene para sí mismo la máxima utilidad de la información incluso en pequeños subconjuntos. Y, a su vez, puede compartir datos con terceros manteniendo la promesa de privacidad realizada a los usuarios. A simple vista, parecería que este enfoque es el que ofrece más ventajas, pero no hay que olvidar sus contrapartidas:

- Existe el riesgo de acceso o uso malintencionado a los datos limpios, lo que exige un mayor esfuerzo en su securización y restringe los modos en los que puede compartirse el sistema con terceros

- Desde la perspectiva de la confianza, un sistema con privacidad local permite hacer una promesa de privacidad mucho más sencilla y robusta. Frases como "*nadie, ni siquiera nosotros, tendremos acceso al detalle de tus datos, que sólo se usarán de forma agregada*" son imbatibles en términos de confianza.

El diablo está en los detalles

En cualquier caso, no hay que minimizar la dificultad de implantar correctamente un sistema que cumpla con la definición de la privacidad diferencial. Elegir, simplemente, un método u otro no resuelve de un plumazo las muchas dificultades a las que se enfrenta toda buena implementación.

Podemos tomar como ejemplo el caso de Apple y su plasmación de la privacidad diferencial a partir de la versión 10 de iOS. Optaron por el enfoque local, el que ofrece mayor confianza. Esto significa que los datos de uso (de las aplicaciones que están bajo este sistema, como el teclado) que se envían desde un terminal iOS hacia los servidores de Apple han sido convenientemente modificados con su ruido numérico correspondiente, lo que dificulta que se nos pueda identificar y relacionar con lo que escribimos.

Su apuesta fomentó la popularidad del concepto y seguro que elevó el listón de exigencia para sus competidores, con lo que todos los ciudadanos salimos ganando. Sin embargo, un sistema que, a primera vista, ofrecía tantas garantías de privacidad **fue duramente criticado por varios investigadores,** como se recoge en este paper.

En la práctica, la privacidad del usuario no estaba garantizada con la robustez que se considera necesaria en ámbitos académicos, particularmente la de aquellos usuarios que hagan un uso más bien limitado de las características en cuestión. Una decisión de diseño (concretamente, que el presupuesto de privacidad no consumido se acumule de un día para otro) tiene consecuencias indeseadas para un tipo concreto de usuarios.

Así que, al afrontar una iniciativa de adopción de la privacidad diferencial en los datos de usuario que recopilamos, **debemos recordar que ni siquiera a Apple le salió bien a la primera.** Es un problema correoso que requiere mucha atención.



Sobre el autor

Manuel Delgado Tenorio

Manuel Delgado Tenorio es asesor independiente de estrategia de datos e inteligencia artificial, profesor e investigador. Cofundó y vendió una *startup* de inteligencia artificial, fue socio de la práctica de Data & Analytics para servicios financieros en EY y responsable de Data Analytics para el sur de Europa en AWS. Es profesor asociado en IE University, donde imparte la primera asignatura de grado sobre Customer Lifetime Value en España, investiga sobre el CLV en su doctorado en la Universidad Complutense de Madrid y es autor del libro "Customer Lifetime Value (CLV): No todos tus clientes valen lo mismo".

manueldelgado.com · LinkedIn · X

Recibe cada artículo en tu correo con la newsletter Ex Datis, Claritas: manueldelgado.com/newsletter/suscribete