

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 993 453**

51 Int. Cl.:

G06F 21/62 (2013.01)

G06N 20/00 (2009.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **12.04.2022** **E 22167899 (8)**

97 Fecha y número de publicación de la concesión europea: **19.06.2024** **EP 4261725**

54 Título: **Método y dispositivos para proporcionar datos de conformidad con una restricción de acceso**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
30.12.2024

73 Titular/es:
HELSING GMBH (100.00%)
Mühldorfstr. 8
81671 Munich, DE

72 Inventor/es:
ADLER, ANTONIA;
DUBINSKA, MAJA;
SCHNEIDER, MANUEL;
GENDRE, MAXIME y
TURNER, THEO

74 Agente/Representante:
LEHMANN NOVO, María Isabel

ES 2 993 453 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

DESCRIPCIÓN

Método y dispositivos para proporcionar datos de conformidad con una restricción de acceso

5 ANTECEDENTES DE LA INVENCION

Los modelos de aprendizaje automático pueden crearse, por ejemplo, desarrollarse, practicarse y/o evaluarse, en plataformas de procesamiento de datos multiusuario a las que acceden distintos usuarios con diferentes autorizaciones de acceso. De este modo, por ejemplo, un desarrollador de un modelo ML puede no tener acceso de lectura a un conjunto de datos de aprendizaje del modelo ML o a otros datos asociados con la creación del modelo ML. Sin embargo, el desarrollador puede seguir viendo el conjunto de datos de aprendizaje. Los enfoques existentes o bien no permiten que un usuario cree el modelo ML sin proporcionarle los datos necesarios o bien solamente proporcionan procesos engorrosos a este respecto. El mantenimiento de la seguridad de los datos, por ejemplo, de las restricciones de acceso, durante la creación del modelo ML, sin embargo, puede ser de gran importancia cuando se trata de datos sensibles, en particular en el ámbito militar o de defensa.

DAVID F. FERRAILOLO ET AL: "Norma NIST propuesta para el control de acceso basado en funciones", ACM TRANSACTIONS ON INFORMATION AND SYSTEM SECURITY, ACM, NEW YORK, NY, US, vol. 4, nº 3, 1 de agosto de 2001 (01-08-2001), páginas 224-274, XP058098436, ISSN: 1094-9224, DOI: 10.1145/501978.501980 propone una norma para el control de acceso basado en funciones (RBAC).

LEE KEON MYUNG ET AL: "Plataforma autónoma de aprendizaje automático", INTERNATIONAL JOURNAL OF INFORMATION MANAGEMENT, ELSEVIER SCIENCE LTD, GB, vol. 49, 11 de julio de 2019 (11-07-2019), págs. 491-501, XP085857475, ISSN: 0268-4012, DOI: 10.1016/J.IJINFOMGT.2019.07.003 propone una plataforma autónoma de aprendizaje automático que proporciona los factores de decisión que deben tomarse durante el desarrollo de aplicaciones de aprendizaje automático. En la plataforma autónoma de aprendizaje automático propuesta, los procesos de aprendizaje automático se automatizan basándose en la especificación de niveles autonómicos. Esta plataforma autónoma de aprendizaje automático puede utilizarse para obtener un resultado de aprendizaje de alta calidad minimizando las intervenciones de los expertos y reduciendo el número de selecciones de diseño que requieren el conocimiento y la intuición de los expertos. También se demuestra que la plataforma autónoma de aprendizaje automático propuesta es adecuada para las poblaciones inteligentes, que suelen requerir cantidades considerables de información sensible desde el punto de vista de la seguridad.

LUKAS TUGGENER ET AL: "Aprendizaje automático de máquinas en la práctica: Estado de la técnica y resultados recientes", ARXIV.ORG, CORNELL UNIVERSITY LIBRARY, 201 OLIN LIBRARY CORNELL UNIVERSITY ITHACA, NY 14853, 19 de julio de 2019 (19-07-2019), XP081444512, DOI: 10.1109/SDS.2019.00-11 ofrece una visión general del estado de la técnica en AutoML con un enfoque en la aplicabilidad práctica en un contexto empresarial, y proporciona resultados de referencia recientes sobre los algoritmos AutoML más importantes.

OLIVER EBERLE ET AL: "Construir e interpretar modelos de similitud profunda", ARXIV.ORG, CORNELL UNIVERSITY LIBRARY, 201 OLIN LIBRARY CORNELL UNIVERSITY ITHACA, NY 14853, 11 de marzo de 2020 (11-03-2020), XP081619585, propone hacer interpretables las similitudes aumentándolas con una explicación en términos de características de entrada. Muchos algoritmos de aprendizaje, tales como las máquinas kernel, los vecinos más próximos, la agrupación o la detección de anomalías, se basan en el concepto de "distancia" o "similitud". Antes de utilizar las similitudes para practicar un modelo real de aprendizaje automático, nos gustaría verificar que están vinculadas a patrones significativos en los datos.

El objeto de la invención es proporcionar un método que permita acceder de forma segura a los datos de aprendizaje de los modelos ML o a los datos asociados con los modelos ML.

SUMARIO

La presente invención, que se define por las reivindicaciones adjuntas, proporciona una solución puesta en práctica por ordenador para proporcionar datos de forma eficaz y segura de conformidad con una restricción de acceso. En particular, se permite la creación de modelos de aprendizaje automático, ML, manteniendo la seguridad de los datos, por ejemplo, cumpliendo las restricciones de acceso a los datos de desarrollo y/o de aprendizaje para el modelo ML.

Según una de muchas formas de realización, se proporciona un método puesto en práctica por ordenador para proporcionar datos de conformidad con una restricción de acceso, comprendiendo el método: determinar las características de los primeros datos asociados con los primeros datos, estando los primeros datos sujetos a la restricción de acceso; determinar las características de los segundos datos asociadas con los segundos datos; determinar si una similitud de las características de los primeros y segundos datos cumple un umbral predeterminado; y proporcionar, en particular emitir o indicar, los segundos datos si la similitud cumple el umbral predeterminado.

Los primeros datos o conjuntos de datos pueden utilizarse para practicar, evaluar y/o desplegar un modelo ML. Los primeros datos pueden comprender datos sensibles, en particular datos militares. Los segundos datos pueden

proporcionarse a un usuario, por ejemplo, un desarrollador, del modelo ML correspondiente sin tener acceso (de lectura) a los primeros datos. Los segundos datos pueden no estar sujetos a la restricción de acceso. La restricción de acceso puede estar asociada o aplicarse al usuario. Al proporcionar al usuario o desarrollador los segundos datos que cumplen el umbral de similitud, se permite al desarrollador crear de manera adecuada el modelo ML manteniendo la seguridad de los datos.

Las características de los datos pueden indicar, incluir o especificar una forma abstracta de los datos. Dicho de otro modo: Las características de los datos pueden ser indicativas de al menos una de las siguientes características: estructura, forma, tipo, tipo de contenido, sintaxis y semántica de los datos asociados, en particular a un nivel general, de manera que no se dé a conocer el contenido real de los datos asociados. Alternativa o adicionalmente, las características de los datos pueden ser indicativas de la fuente de los datos asociados, o dicho de otro modo, de cómo se adquirieron los datos asociados.

La características de los primeros y segundos datos pueden ser similares si la forma abstracta de los datos asociados se corresponde entre sí al menos en parte. Dicho de otro modo: Una similitud de características de datos puede comprender que al menos dos características de datos sean compatibles y/o al menos en parte iguales entre sí. Dicho de otra manera: Determinar si la similitud de las características de los primeros y segundos datos cumple el umbral de similitud, incluyendo la determinación de si las características de los primeros datos corresponden, al menos en parte, a las características de los segundos datos.

El umbral de similitud puede estar predeterminado, en particular de manera que los segundos datos cumplan ciertos requisitos para una tarea específica, por ejemplo, el desarrollo del modelo. De ese modo, un desarrollador puede efectuar su función en el modelo ML basándose en los segundos datos. Como los segundos datos cumplen el umbral de similitud, el modelo ML así desarrollado puede utilizarse, por ejemplo, practicarse, evaluarse y/o desplegarse, con los primeros datos.

Según una forma de realización, las características de los primeros y/o segundos datos se determinan basándose en los primeros y segundos metadatos almacenados previamente y asociados con los primeros y segundos datos, respectivamente.

Dicho de otro modo: Las características de los datos están comprendidas, indicadas o especificadas por los primeros y segundos metadatos almacenados. Las características de los datos y/o los metadatos almacenados pueden comprender etiquetas de datos incluidas en, o asociadas con, los respectivos datos asociados. En un ejemplo, los metadatos almacenados pueden estar incluidos en los respectivos datos asociados o ser accesibles por separado. De ese modo, las características de los datos, en particular la similitud de las características de los datos, pueden determinarse sin acceder, en particular realizando la lectura, de los primeros y segundos datos, mejorando de este modo la seguridad de los datos.

Según una forma de realización, el método comprende, además: antes de la determinación de los segundos datos, añadir los segundos datos a una pluralidad de datos almacenados; en respuesta a la adición de los segundos datos, determinar y/o almacenar automáticamente segundos metadatos indicativos de las características de los segundos datos; y determinar las características de los segundos datos basándose en los segundos metadatos.

De forma adicional o alternativa, un nuevo tercer, cuarto, etc. dato o conjunto de datos puede añadirse a la pluralidad de datos almacenados de manera similar. De este modo, se crea y mantiene una base de metadatos o características de los datos que corresponde o está asociada a la pluralidad de datos almacenados y en función de la cual se puede determinar, a petición, la similitud entre un conjunto de datos de referencia y al menos uno de entre la pluralidad de conjuntos de datos de forma rápida y eficaz, manteniendo al mismo tiempo la seguridad de los datos.

Según una forma de realización, la determinación automática de los segundos metadatos se realiza mediante un modelo de aprendizaje automático. De este modo, se mejora aún más la eficacia y la seguridad de los datos del método.

El método comprende, además: recibir una solicitud de un usuario para acceder a los primeros datos, en donde la restricción de acceso se aplica al usuario; proporcionar automáticamente, en particular identificar y proporcionar, los segundos datos al usuario en respuesta a la recepción de la solicitud del usuario.

De este modo, el usuario puede indicar específicamente qué datos o tipo de datos necesita o a qué datos debe asemejarse el conjunto de datos proporcionado, mejorando de este modo la precisión del método. A continuación, el método identifica los segundos datos y determina si se cumple el umbral de similitud. Dicho de otro modo, el método identifica conjuntos de datos similares al conjunto de datos solicitado. Si es así, los segundos datos, es decir, los datos similares, se proporcionan al usuario.

Según una forma de realización, las características de los primeros y segundos datos comprenden características sintácticas y/o características semánticas de los conjuntos de primeros y segundos datos, respectivamente.

De este modo, las características de los datos pueden indicar el (tipo de) contenido y/o la estructura del conjunto de datos respectivo, por ejemplo, el formato de almacenamiento y el sistema de datos del conjunto de datos. Además, o de manera alternativa, las características de los datos pueden indicar el tipo de datos y/o la fuente u origen de los datos del conjunto de datos respectivo. Por ejemplo, las características de los datos pueden indicar qué tipos de objetos están representados por el conjunto de datos respectivo, por quién fueron recogidos los datos y/o cómo están organizados o almacenados los datos.

Según una forma de realización, el método comprende, además: realizar un primer proceso sobre los terceros datos utilizando los segundos datos; y después de realizar el primer proceso, realizar un segundo proceso sobre los terceros datos utilizando los primeros datos.

Los terceros datos pueden ser, comprender o representar un modelo ML. Los procesos primero y segundo pueden comprender un desarrollo, un aprendizaje, una evaluación y/o un despliegue de los terceros datos. En un ejemplo, el primer proceso comprende el desarrollo o aprendizaje de los terceros datos y el segundo proceso comprende el aprendizaje, evaluación o despliegue de los terceros datos. Es decir, los datos segundos y/o primeros pueden ser un predecesor de los terceros datos, depender de ellos o constituir de otro modo la base para los terceros datos. En particular, el segundo proceso puede realizarse sin permitir el acceso (a un usuario, por ejemplo, un desarrollador) a los primeros datos. De ese modo, pueden desarrollarse o crearse terceros datos configurados para ser utilizados con los primeros datos sin tener que dar a conocer los primeros datos.

Según una forma de realización, el método comprende, además: después de realizar el segundo proceso, restringir el acceso a los terceros datos y/o permitir el acceso a, en particular proporcionar, los metadatos de los terceros datos.

Como los terceros datos procesados pueden dar a conocer al menos una parte de los primeros datos, en particular una parte sensible de los primeros datos, restringir el acceso a los terceros datos procesados mejora aún más la seguridad de los datos y, más en particular, garantiza la propagación de la restricción de acceso.

Según otra forma de realización, se proporciona un aparato de procesamiento de datos que comprende medios para llevar a cabo el método descrito con anterioridad.

Según otra forma de realización, se proporciona un programa informático de datos que comprende instrucciones que, cuando el programa es ejecutado por un ordenador, hacen que el ordenador lleve a cabo el método descrito con anterioridad.

Según otra forma de realización, se proporciona un medio legible por ordenador de datos que comprende instrucciones que, cuando son ejecutadas por un ordenador, hacen que el ordenador lleve a cabo el método descrito con anterioridad.

BREVE DESCRIPCIÓN DE LOS DIBUJOS

La presente invención puede comprenderse mejor a partir de la lectura de la siguiente descripción de formas de realización no limitantes, con referencia a los dibujos adjuntos, en donde:

la Figura 1 muestra un diagrama de flujo de un método puesto en práctica por ordenador para proporcionar datos de conformidad con una restricción de acceso;

la Figura 2 muestra un diagrama de flujo de un método puesto en práctica por ordenador para determinar las características de los datos que puede realizarse antes del método para proporcionar datos de conformidad con una restricción de acceso;

la Figura 3 muestra un diagrama de flujo de un método puesto en práctica por ordenador para realizar un proceso y/o restringir el acceso a un modelo ML que puede realizarse con posterioridad al método para proporcionar datos de conformidad con una restricción de acceso; y

la Figura 4 muestra un aparato de procesamiento de datos configurado para llevar a cabo cualquiera de las etapas del método descrito con referencia a las Figuras 1 a 3.

DESCRIPCIÓN DETALLADA DE LAS FORMAS DE REALIZACIÓN PREFERIDAS

La Figura 1 muestra un diagrama de flujo de un método puesto en práctica por ordenador 100 para proporcionar datos de conformidad con una restricción de acceso según una o más formas de realización preferidas.

En la etapa 110, se recibe una solicitud de usuario para acceder a unos primeros datos. La solicitud del usuario puede recibirse en una plataforma de procesamiento de datos multiusuario. Los primeros datos están sujetos a una restricción de acceso. Por ejemplo, el usuario que solicita los primeros datos puede tener restringido el acceso a los primeros datos en la plataforma. Más concretamente, el usuario puede no tener acceso de lectura y/o escritura a los primeros

datos en la plataforma.

El usuario puede ser un desarrollador de un modelo de aprendizaje automático, ML. El modelo ML puede ser desarrollado para ser practicado con, desplegado en, o utilizado de otro modo con, los primeros datos. La etapa 110 puede incluir, alternativa o adicionalmente, la etapa de recibir una solicitud del usuario a datos similares a los primeros datos, por ejemplo, cuando el usuario es consciente de su restricción de acceso. Dicho de otro modo, en la etapa 110 puede recibirse una solicitud de segundos datos que puedan utilizarse para desarrollar un modelo ML, en donde la solicitud indica que el modelo ML debe utilizarse con los primeros datos.

En la etapa 120, se determinan las características de primeros datos de los primeros datos. Las características de los primeros datos están asociadas con los primeros datos. Las características de los primeros datos pueden comprender información específica relativa a al menos una de entre la naturaleza, el tipo, la sintaxis, la semántica, la forma de origen o la estructura de los primeros datos. Dicho de otro modo, las características de los primeros datos pueden indicar la estructura y/o el tipo de contenido de los primeros datos, en particular sin dar a conocer el contenido específico de los primeros datos.

Las características de los primeros datos pueden estar incluidas en los primeros datos. En otro ejemplo, las características de los primeros datos pueden almacenarse de manera separada con respecto a los primeros datos. Además, o de manera alternativa, las características de los primeros datos pueden incluirse en los primeros metadatos asociados con los primeros datos. Las características de los primeros datos pueden incluir etiquetas de datos. Las etiquetas de datos pueden etiquetar los primeros datos, en particular cada uno o una parte de entre una pluralidad de datos o conjuntos de datos incluidos en los primeros datos.

En la etapa 130, se identifican los segundos datos. Los segundos datos no están sujetos a la restricción de acceso. Dicho de otro modo, los segundos datos, a diferencia de los primeros, no están clasificados. Dicho de otro modo, el usuario del que se recibe la solicitud de usuario en la etapa 110 tiene acceso, en particular acceso de lectura y/o escritura, a los segundos datos. Los segundos datos pueden identificarse basándose en una pluralidad de datos o conjuntos de datos. La pluralidad de datos puede estar incluida en o puede estar representada por una interfaz de programación de aplicaciones de datos, API de datos. La API de datos puede comprender, además, los primeros datos.

Además, en la etapa 130, se determinan las características de segundos datos con respecto a los segundos datos. Las características de los segundos datos se asocian a los segundos datos del mismo modo o de forma similar a como las características de los primeros datos se asocian a los primeros datos. La descripción anterior de las características de los primeros datos se aplica en consecuencia a las características de los segundos datos. Las características de los primeros y/o segundos datos pueden incluirse en la interfaz API de datos.

En la etapa 140, se determina una similitud de las características de los primeros y segundos datos. La etapa 140 puede comprender la forma de realización de una comparación de las características de los primeros y segundos datos. Las características de los primeros y segundos datos pueden considerarse similares si al menos una parte de las características de los primeros y segundos datos se corresponden entre sí, se solapan o son similares entre sí de algún otro modo. El tipo de similitud puede estar predefinido. Dicho de otro modo, se puede predefinir en qué condiciones las diferentes características de los datos deben considerarse similares. Dicho de otro modo, la similitud puede determinarse en la etapa 140 mediante un algoritmo determinista. Además, o de manera alternativa, la determinación en la etapa 140 puede ser realizada por un modelo ML practicado.

En la etapa 150, se determina si la similitud determinada en la etapa 140 cumple un umbral, también aquí denominado como un umbral de similitud. La similitud puede alcanzar el umbral si las características de los primeros y segundos datos son, al menos en parte, similares entre sí, por ejemplo, si al menos una parte predeterminada de las características de los primeros y segundos datos se corresponden entre sí, o son similares entre sí de otro modo. Dicho de otro modo, la similitud puede alcanzar el umbral si la estructura, la forma y/o el tipo de contenido de los primeros y segundos datos (indicados por las características de los primeros y segundos datos) se corresponden entre sí, en particular, se corresponden totalmente entre sí o se corresponden entre sí en una extensión predeterminada.

Si, en la etapa 150, se determina que la similitud no alcanza el umbral, el método vuelve al etapa 130, en donde se determina otros segundos datos y se determinan otras características de segundos datos de los otros segundos datos.

Si, en la etapa 150, se determina que la similitud cumple el umbral, se proporcionan los segundos datos en la etapa 160. Por ejemplo, los segundos datos se proporcionan a o para el usuario, o se emiten o indican de otro modo al usuario, en particular al usuario del que se recibe la solicitud de usuario en la etapa 110. De ese modo, se proporcionan segundos datos que son similares a los primeros datos solicitados, de tal modo que la tarea que debe realizarse utilizando los primeros datos, por ejemplo, un proceso que debe realizarse utilizando los primeros datos, estando la tarea o el proceso preferentemente especificados en la solicitud, puede realizarse utilizando los segundos datos. La tarea a realizar puede comprender un proceso de desarrollo de un modelo ML, tal como se describe con más detalle a continuación con referencia a la Figura 3.

La Figura 2 muestra un diagrama de flujo de un método 200 para determinar automáticamente las características de los segundos datos. El método 200 puede realizarse antes del método 100. Sin embargo, en general, las etapas del método de cualquiera de los métodos descritos con referencia a las Figuras 1 a 3 pueden realizarse en un orden diferente.

En la etapa 210, los segundos datos se añaden a una pluralidad de datos almacenados, o conjuntos de datos. Tal como se ha descrito con anterioridad, la pluralidad de datos almacenados puede denominarse como la interfaz API de datos.

En respuesta a la adición de los segundos datos a la pluralidad de datos almacenados, en la etapa 220 se determinan automáticamente las características de los segundos datos. Dicho de otro modo, cada vez que se añaden nuevos datos o un nuevo conjunto de datos a la pluralidad de datos almacenados, se determinan automáticamente las características de datos de dichos nuevos datos. Dichas características de datos asociadas con los nuevos datos pueden añadirse a la interfaz API de datos, almacenarse por separado y/o almacenarse dentro de los nuevos datos.

La determinación (automática) de las características de los segundos datos, tal como se realiza en la etapa 220, puede comprender la determinación de uno o más identificadores de datos asociados con los segundos datos. Dicho de otro modo, determinar las características de los segundos datos puede comprender analizar y/o etiquetar los segundos datos. Los identificadores o etiquetas de datos pueden almacenarse como segundos metadatos asociados con los segundos datos. Tal como se ha mencionado con anterioridad, los segundos metadatos pueden estar incluidos en los segundos datos o pueden almacenarse por separado, en particular como parte de la interfaz API de datos. La determinación automática de las características de los segundos datos en la etapa 220 puede ser realizada por un modelo ML. De este modo, se construye una base de datos o una interfaz API de datos que comprende una pluralidad de datos y las correspondientes características de datos asociadas con los mismos. A partir de dicha base de datos, puede identificarse de forma automatizada y eficiente unos segundos datos adecuados cuyas características de datos sean similares a las características de los primeros datos de forma que se cumpla el umbral de similitud.

La Figura 3 muestra un diagrama de flujo de un método 300 para realizar procesos en un modelo ML y definir restricciones de acceso al modelo ML. El método 300 puede realizarse con posterioridad al método 100 y/o al método 200 descritos con referencia a las Figuras 1 y 2.

En la etapa 310, se realiza un primer proceso sobre un modelo ML utilizando los segundos datos que se proporcionan en la etapa 160 del método 100. El primer proceso puede comprender un proceso de desarrollo del modelo ML. Dicho de otro modo, los segundos datos se utilizan para el desarrollo o la creación del modelo ML, en donde el modelo ML desarrollado se utilizará posteriormente con los primeros datos, tal como se indica en la solicitud del usuario recibida en la etapa 110 del método 100.

En la etapa 320, se realiza un segundo proceso en el modelo ML utilizando los primeros datos. El segundo proceso puede comprender un proceso de aprendizaje del modelo ML. Dicho de otro modo, el modelo ML puede ser practicado con los primeros datos después de que el modelo ML haya sido desarrollado o practicado basándose en los segundos datos similares a los primeros datos. En una forma de realización, el segundo proceso puede basarse en, o utilizar, los primeros y los segundos datos. El primer y/o el segundo proceso pueden realizarse (en las etapas 310 y 320) en respuesta a una entrada del usuario, en particular en respuesta a una entrada recibida del usuario a partir de la cual se recibió la solicitud del usuario en la etapa 110 del método 100. Es decir, el usuario puede desencadenar la forma de realización tanto del primer proceso como del segundo. Como el usuario del que se recibe la solicitud de usuario en la etapa 110 puede tener restringido el acceso a los primeros datos, el proceso de aprendizaje realizado en la etapa 320 puede denominarse proceso de aprendizaje a ciegas.

En la etapa 330, se restringe el acceso al modelo ML practicado. Dicho de otro modo, se determina una restricción de acceso o autorización de acceso, que está en consonancia con la restricción de acceso a los primeros datos. Dicho de otro modo, las mismas restricciones de acceso que se aplican a los datos de aprendizaje, es decir, los primeros datos, se aplican al modelo ML practicado. Expresado de otra manera, puesto que el modelo ML practicado depende de los primeros datos, las restricciones de acceso a los primeros datos se aplican al modelo ML practicado. De ese modo, se puede mantener una política coherente de restricción de datos basada en la procedencia o ascendencia de los datos.

En la etapa 340, que puede realizarse, además, o en lugar de la etapa 330, puede habilitarse el acceso a los metadatos del modelo ML practicado, por ejemplo, pueden proporcionarse los metadatos, en particular la salida o indicarse. En concreto, los metadatos pueden proporcionarse al usuario del que se recibe la solicitud de usuario en la etapa 110 del método 100. Los metadatos del modelo ML practicado pueden comprender estadísticas de aprendizaje del modelo ML practicado u otra información asociada con el modelo ML y utilizada o utilizable para desarrollar el modelo ML. Los metadatos del modelo ML pueden no dar a conocer el contenido específico de los primeros datos. Por lo tanto, al realizar el proceso de aprendizaje en el modelo ML en la etapa 320 y proporcionar, por ejemplo, la salida, solamente los metadatos del modelo ML, se realiza un proceso de aprendizaje a ciegas en el modelo ML sin permitir el acceso a, o proporcionar de otra manera, los primeros datos, manteniendo de este modo la restricción de acceso, es decir, los requisitos de seguridad predeterminados.

La etapa 330 y/o la etapa 340 pueden comprender la habilitación o el suministro de un acceso de descubrimiento al modelo ML, en particular para el usuario del que se recibe la solicitud de usuario en la etapa 110 del método 100. Dicho de otro modo, el usuario puede ver o descubrir el modelo ML practicado y/o los primeros datos, sin tener acceso de lectura o de escritura al modelo ML y/o a los primeros datos. Expresado de otra manera, el usuario, por ejemplo, el desarrollador del modelo ML, puede conocer el modelo ML practicado y/o los primeros datos sin poder acceder a los datos sensibles incluidos en los primeros datos y/o en el modelo ML.

La Figura 4 muestra un aparato de tratamiento de datos 400 configurado para llevar a cabo cualquiera de las etapas del método descrito con referencia a las Figuras 1 a 3. Para ello, el aparato de tratamiento de datos 400 puede comprender un procesador 410 y una memoria o medio legible por ordenador 420.

La memoria 420 puede comprender la pluralidad descrita con anterioridad de datos o conjuntos de datos almacenados y/o los metadatos descritos con anterioridad asociados con la pluralidad de datos almacenados. Dicho de otro modo, la interfaz API de datos descrita con anterioridad está comprendida por la memoria 420. Alternativa o adicionalmente, al menos uno de los modelos ML descritos con anterioridad está incluido en, es decir, almacenado en, la memoria 420.

Un programa informático 425 está almacenado en la memoria 420. El programa informático 425 puede comprender instrucciones que, cuando el programa es ejecutado por un ordenador, en particular por el aparato de procesamiento de datos 400, hacen que el ordenador o el aparato de procesamiento de datos 400 lleve a cabo cualquiera de las etapas del método descritas con referencia a las Figuras 1 a 3.

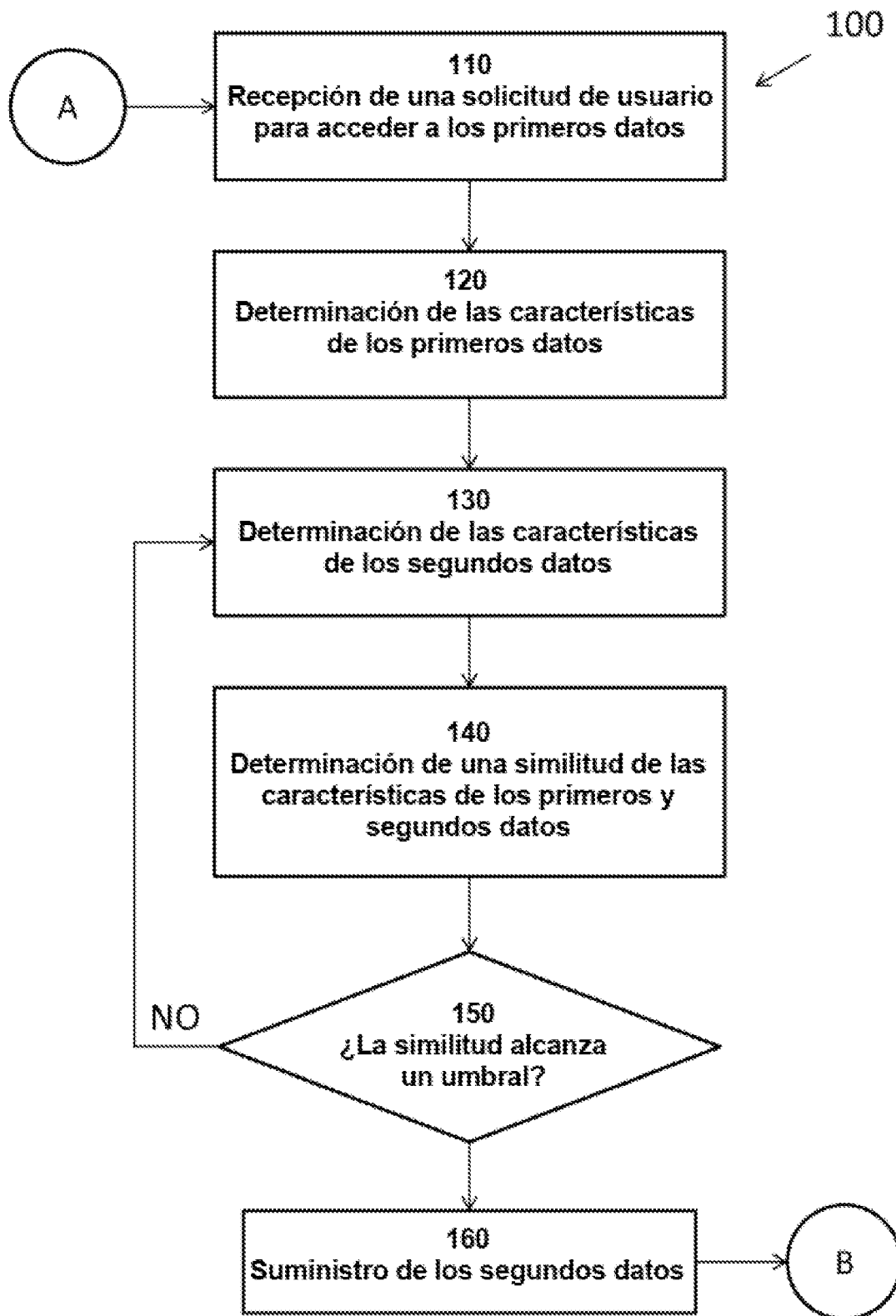
El aparato de procesamiento de datos 400 puede comprender, además, una interfaz 430. La interfaz 430 puede conectar otros componentes del aparato de procesamiento de datos 400, por ejemplo, el procesador 410 y la memoria 420, y/o proporcionar conexión a otros componentes que estén acoplados de forma comunicativa al aparato de procesamiento de datos 400. La interfaz 430 puede ser, además, una interfaz (de usuario) para proporcionar o permitir de otro modo el acceso a los datos determinados en cualquiera de las etapas del método descritas con referencia a las Figuras 1 a 3, en particular en la etapa 160 del método 100 y, más en concreto, al usuario de la solicitud de usuario que fue recibida en la etapa 110 del método 100.

REIVINDICACIONES

1. Un método puesto en práctica por ordenador para proporcionar datos de conformidad con una restricción de acceso, comprendiendo el método:
 - 5 recibir (110) una solicitud de un usuario para acceder a unos primeros datos, estando los primeros datos sujetos a la restricción de acceso, en donde la restricción de acceso se aplica al usuario;
 - 10 determinar (120) unas las características de los primeros datos asociadas con los primeros datos;
 - determinar (130) las características de los segundos datos asociadas con los segundos datos que no están sujetos a la restricción de acceso; determinar (150) si una similitud de las características de los primeros y segundos datos cumple un umbral predeterminado; y
 - 15 proporcionar automáticamente (160), en particular emitiendo o indicando, los segundos datos al usuario en respuesta a la recepción de la solicitud del usuario si la similitud cumple el umbral predeterminado.
2. El método según la reivindicación 1, en donde la primera y/o segunda características de los datos se determinan basándose en los primeros y segundos metadatos previamente almacenados asociados con los primeros y segundos datos, respectivamente.
- 20 3. El método según cualquiera de las reivindicaciones anteriores, que comprende, además:
 - 25 antes de la determinación de los segundos datos, añadir (210) los segundos datos a una pluralidad de datos almacenados;
 - en respuesta a la adición de los segundos datos, determinar y/o almacenar automáticamente (220) segundos metadatos indicativos de las características de los segundos datos; y
 - 30 determinar (130) las características de los segundos datos basándose en los segundos metadatos.
4. El método según la reivindicación 3, en donde la determinación automática de los segundos metadatos se realiza mediante un modelo de aprendizaje automático.
- 35 5. El método según cualquiera de las reivindicaciones anteriores, en donde las características de los primeros y segundos datos comprenden características sintácticas y/o características semánticas de los conjuntos de primeros y segundos datos, respectivamente.
6. El método según cualquiera de las reivindicaciones precedentes, que comprende, además:
 - 40 realizar (310) un primer proceso sobre unos terceros datos utilizando los segundos datos; y
 - después de realizar el primer proceso, realizar (320) un segundo proceso sobre los terceros datos utilizando los primeros datos.
 - 45 7. El método según la reivindicación 6, que comprende, además:
 - con posterioridad a la realización del segundo proceso, restringir (330) el acceso a los terceros datos y/o permitir (340) el acceso a, en particular proporcionar, los metadatos de los terceros datos.
 - 50 8. El método según cualquiera de las reivindicaciones precedentes, que comprende, además:
 - desarrollar, crear o practicar (310) un modelo de aprendizaje automático, ML, utilizando los segundos datos;
 - 55 en respuesta a una entrada del usuario, practicar (320) el modelo ML utilizando los primeros datos después de que el modelo ML haya sido desarrollado, creado o practicado utilizando los segundos datos.
 9. El método según la reivindicación 8, que comprende, además:
 - 60 con posterioridad al aprendizaje del modelo ML utilizando los primeros datos, aplicar (330) la restricción de acceso al modelo ML.
 10. Un aparato de procesamiento de datos (400) que comprende medios para llevar a cabo el método según cualquiera de las reivindicaciones anteriores.
 - 65 11. Un programa informático (425) que comprende instrucciones que, cuando el programa es ejecutado por un

ordenador, hacen que el ordenador lleve a cabo el método según cualquiera de las reivindicaciones 1 a 9.

12. Un soporte legible por ordenador (420) que comprenda instrucciones que, al ser ejecutadas por un ordenador, hacen que este último lleve a cabo el método según cualquiera de las reivindicaciones 1 a 9.



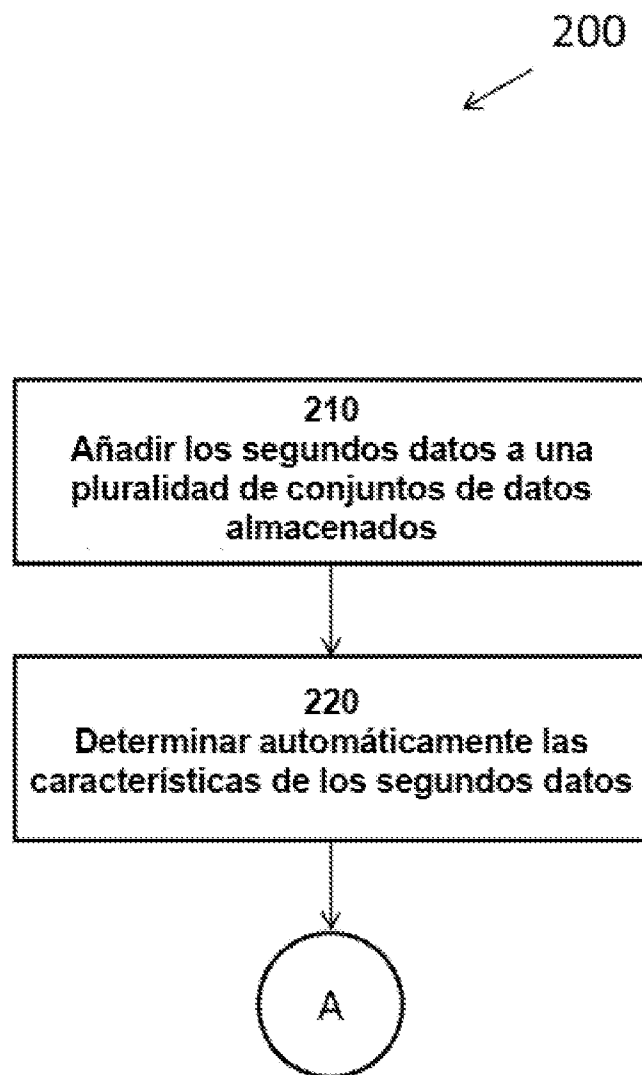


Fig. 2

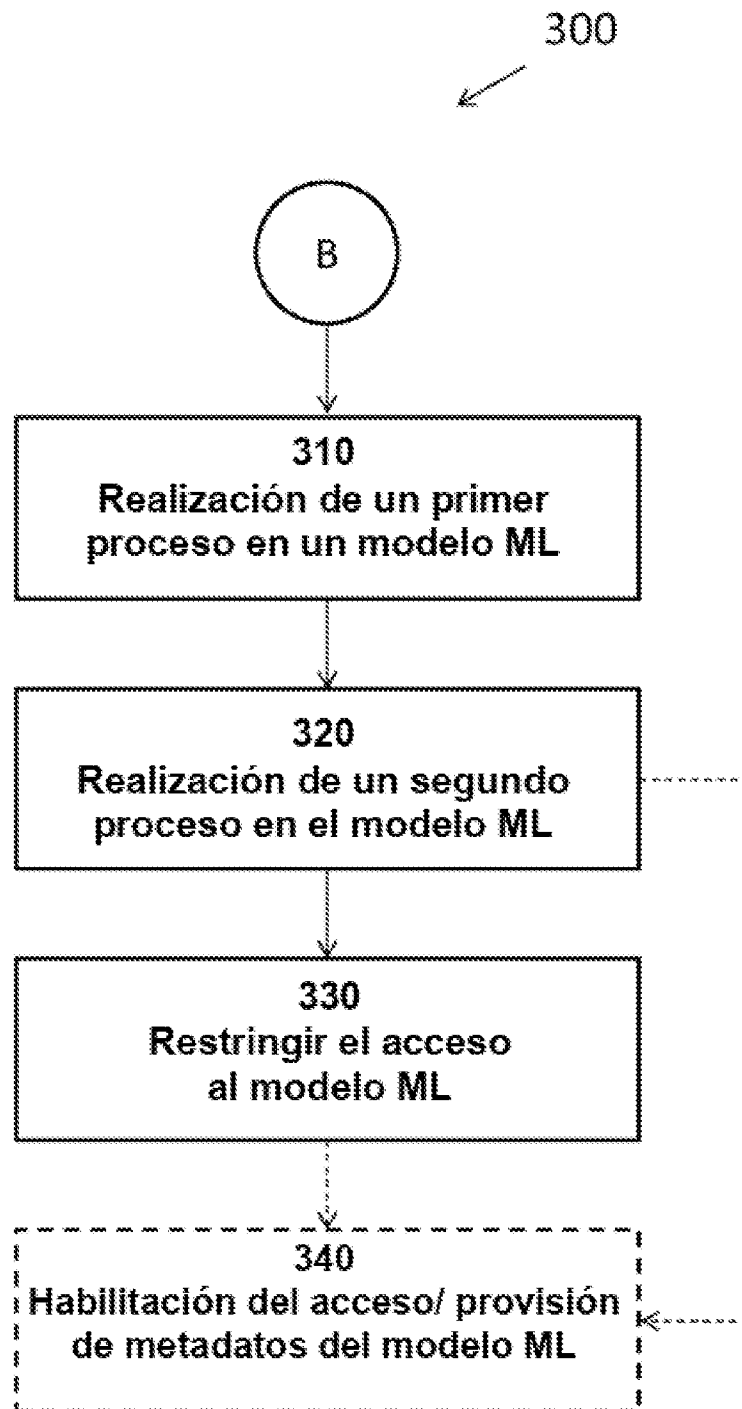


Fig. 3

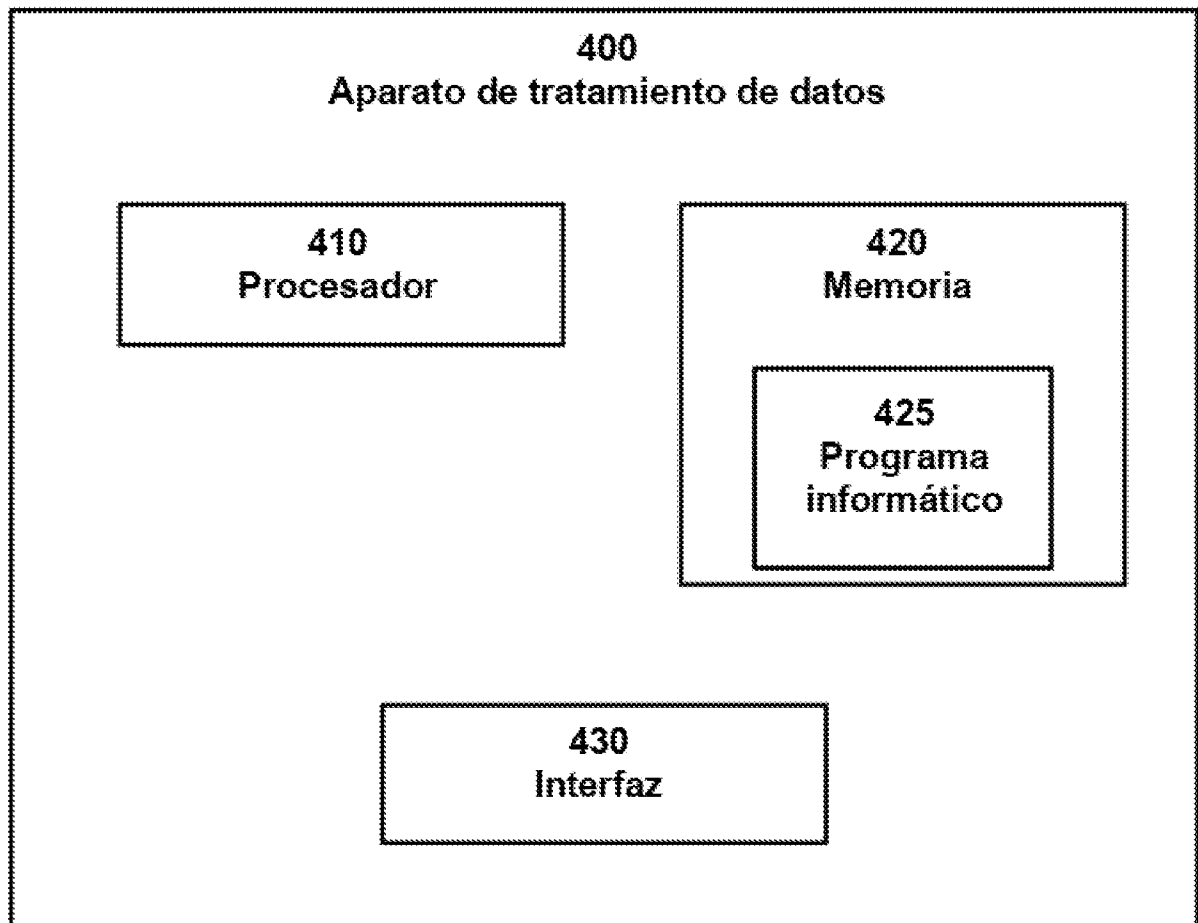


Fig. 4