(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 615 104**

(51) Int. CI.:

G06Q 40/02 (2012.01)
G06Q 50/32 (2012.01)
G06F 21/31 (2013.01)
H04L 12/58 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(86) Fecha de presentación y número de la solicitud internacional: **09.01.2013 PCT/US2013/020714**

(87) Fecha y número de publicación internacional: **18.07.2013 WO13106354**

(96) Fecha de presentación y número de la solicitud europea: **09.01.2013 E 13735613 (5)**

(97) Fecha y número de publicación de la concesión europea: **09.11.2016 EP 2803031**

(54) Título: **Clasificación de cuentas de usuario basada en el aprendizaje de máquinas basado en direcciones de correo electrónico y en otra información de las cuentas**

(30) Prioridad:

12.01.2012 US 201213349306

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
05.06.2017

(73) Titular/es:

MICROSOFT TECHNOLOGY LICENSING, LLC (100.0%)
One Microsoft Way
Redmond, WA 98052, US

(72) Inventor/es:

ZHU, BIN BENJAMIN y
XUE, FEI

(74) Agente/Representante:

DE ELZABURU MÁRQUEZ, Alberto

DESCRIPCIÓN

Clasificación de cuentas de usuario basada en el aprendizaje de máquinas basado en direcciones de correo electrónico y en otra información de las cuentas

ANTECEDENTES

Las transacciones habitualmente necesitan cuentas en línea. Cada cuenta en línea contiene información relativa al usuario de la cuenta, tal como dirección de correo electrónico, nombre, dirección del domicilio, números de teléfono, etc. Para evitar ser descubiertos y aumentar la tasa de éxito, los usuarios malintencionados habitualmente utilizan una cuenta para uso ilegal durante un corto periodo de tiempo y, a continuación, cambian a otra cuenta, por lo que necesitan un gran número de cuentas. Las cuentas utilizadas para transacciones ilegales pueden ser cuentas manipuladas de otros usuarios legítimos, o pueden estar creadas por los usuarios malintencionados. Se pueden utilizar programas automáticos (por ejemplo, ordenadores zombie) para generar una gran cantidad de direcciones de correo electrónico para registrar tales cuentas fraudulentas. Por ejemplo, tales cuentas fraudulentas pueden estar asociadas con alguna tarjeta de crédito robada o ilegal para realiza transacciones fraudulentas en línea. La información asociada con una cuenta, tal como la dirección de correo electrónico asociada, se revisa manualmente para la identificación de cuentas maliciosas, lo que resulta costoso en dinero, trabajo y tiempo.

El documento EP 2 112 627 da a conocer que una base de datos de información de contactos, que incluye registros tales como los almacenados en una agenda personal, se aplica para evaluar la reputación de un usuario y para realizar detección de fraude o spam. Se pueden utilizar una serie de factores diferentes seleccionados por valor de predicción de reputación / fraude en un modelo estadístico, para evaluar la reputación de un individuo sobre la base de un identificador, tal como una dirección de correo electrónico. Los factores pueden incluir información útil en la predicción de la reputación de un individuo, tal como en cuántos libros de direcciones aparece la dirección de correo electrónico u otra información, si se han enviado correos electrónicos previamente a esa dirección de correo electrónico, si algunos de tales correos electrónicos han sido devueltos como no entregables, etc.

El documento US 2006/200487 describe un sistema en el que una Entidad de Registro mantiene una base de datos con datos de reputación a la que una Autoridad de Certificación puede acceder. La Entidad de Registro puede actualizar los datos de reputación sobre la base de una variedad de eventos relativos al nombre del dominio Los datos de reputación pueden ser rastreados por el nombre del dominio, las URL, el comprador o registrador del nombre del dominio, y/o direcciones de correo electrónico asociadas con el nombre del dominio. Los datos de reputación pueden incluir varias categorías, tal como prácticas del correo electrónico, contenido del sitio web, políticas y prácticas de privacidad, actividades fraudulentas, quejas relacionadas con el nombre del dominio, reputación global, etc.

COMPENDIO

Este Compendio se proporciona para introducir una selección de conceptos de una forma simplificada, que se describen con más detalle a continuación en la Descripción detallada. Este Compendio no pretende identificar características clave o características esenciales del asunto central reivindicado, ni pretende ser utilizado como ayuda en la determinación del alcance del asunto central reivindicado. El término “técnicas”, por ejemplo, puede hacer referencia a un dispositivo o dispositivos, a un sistema o sistemas, a un método o métodos y/o a instrucciones legibles por ordenador, tal como permite el contexto anterior y la presente descripción.

La invención proporciona un método implementado por un ordenador, un sistema y uno o más medios de almacenamiento legibles por ordenador, según se reivindica a continuación en esta memoria.

La presente explicación proporciona técnicas para identificar si una cuenta entrante es maliciosa al menos parcialmente sobre la base de cierta información, que incluye una dirección de correo electrónico y/u otra información adicional asociada con la cuenta. Las técnicas extraen características de la información asociada con la cuenta, y aplican un modelo de clasificación para generar una puntuación basada en las características extraídas para indicar la probabilidad de que la cuenta entrante sea maliciosa. Tal puntación puede estar determinada, al menos parcialmente, por un grado de la capacidad de ser memorizada, de la dirección de correo electrónico. La capacidad de ser memorizada se refiere a la facilidad para que una persona memorice la dirección de correo electrónico. El grado de capacidad de ser memorizada se puede determinar utilizando una o más características que distinguen una dirección de correo electrónico generada por un humano de una dirección de correo electrónico generada por una máquina, para generar un grado de capacidad de ser memorizada de la dirección de correo electrónico. Además, el dominio de la dirección de correo electrónico puede determinar asimismo parcialmente la puntuación, y esta puede estar basada en otra información adicional relativa a la cuenta, tal como la correlación de la dirección de correo electrónico con otra información de la cuenta tal como nombre, teléfono y domicilio, para ayudar a la determinación de si la cuenta es maliciosa. Por ejemplo, la puntuación se puede determinar asimismo parcialmente mediante información distinta de una dirección de correo electrónico que está asociada con una cuenta, tal como si el nombre, el domicilio, etc., de la cuenta coincide con los de una tarjeta de crédito, validez de un domicilio, cuánto tiempo hace que se creó la cuenta, y otros.

Las presentes técnicas aplican al aprendizaje de máquinas para aprender una o más características extraídas de los datos de aprendizaje que distinguen una cuenta maliciosa de una cuenta benigna, obtener el modelo de clasificación basado en las características, y utilizar el modelo de clasificación para determinar de manera automática la puntuación que indica la probabilidad de que la cuenta sea maliciosa. El modelo de clasificación se puede mejorar de manera continua y entrenar incrementalmente con datos nuevos. Por ejemplo, los datos contrastados de la cuenta entrante acerca de si la cuenta entrante es maliciosa o benigna, que se pueden obtener a partir del resultado de una o más transacciones en línea, se pueden introducir para entrenar al modelo de clasificación para ajustar el modelo de clasificación de acuerdo con ellos. Por ejemplo, los datos contrastados de la cuenta entrante se pueden comparar con un resultado determinado por el modelo de clasificación, y dicho resultado de la comparación se utiliza para mejorar el modelo de clasificación.

BREVE DESCRIPCIÓN DE LOS DIBUJOS

La descripción detallada se describe con referencia a las figuras que se acompañan. En las figuras, el dígito o dígitos de la izquierda de un número de referencia identifica la figura en la que el número de referencia aparece primero. En todos los dibujos se utilizan los mismos números para referenciar características y componentes iguales.

La figura 1 ilustra un escenario de ejemplo de determinación de una puntuación de una cuenta sobre la base de la dirección de correo electrónico de la cuenta y de otra información adicional.

La figura 2 ilustra un diagrama de flujo de un método de ejemplo para determinar una puntuación de una cuenta.

La figura 3 ilustra un diagrama de flujo de un método de ejemplo para aprendizaje de máquinas para identificar características y generar un modelo de clasificación.

DESCRIPCIÓN DETALLADA

Las presentes técnicas determinan una puntuación de una cuenta asociada con una dirección de correo electrónico y con otra información de la cuenta, sobre la base, al menos parcialmente, del grado de capacidad de ser memorizada de la dirección de correo electrónico. La puntuación determina la probabilidad de que la cuenta sea maliciosa o benigna. Además, las técnicas pueden utilizar asimismo otras características asociadas con la cuenta y/o con la dirección de correo electrónico, tal como el dominio de la dirección de correo electrónico, la correlación del número de teléfono y el nombre de usuario asociado con la cuenta y/o con la dirección de correo electrónico, u otras características asociadas con la cuenta, para determinar una puntuación de la cuenta.

La dirección de correo electrónico asociada con una cuenta benigna está generada habitualmente por un humano, mientras que la dirección de correo electrónico asociada con una cuenta maliciosa puede estar generada habitualmente por una máquina. Las direcciones de correo electrónico generadas por humanos están seleccionadas típicamente para una memorización fácil durante un periodo de tiempo. Existen típicamente ciertas características, tales como secuencias y números con un cierto significado, secuencias pronunciables, simetría en las direcciones de correo electrónico generadas por humanos para ayudar a la memorización. Sus longitudes son también típicamente cortas. Las direcciones de correo electrónico generadas por máquinas, actuales, típicamente carecen de estas características. En algunas realizaciones, algunas direcciones de correo electrónico pueden pasar la prueba de capacidad de ser memorizadas y ser consideradas como generadas por un humano, pero, se pueden considerar aún asociadas con cuentas maliciosas si se determina que sus dominios u otra información adicional asociada con la dirección de correo electrónico y/o con la cuenta están asociados con cuentas maliciosas. En otras realizaciones, se pueden utilizar también datos de la cuenta distintos de una dirección de correo electrónico, tal como el nombre asociado con la cuenta y su correlación con el nombre de una tarjeta de crédito utilizada en una transacción en línea, la validez de un domicilio asociado con la cuenta, etc., para determinar la probabilidad de que la cuenta sea o no maliciosa.

Las técnicas de aprendizaje de máquinas se pueden utilizar para extraer características a partir de datos de aprendizaje, para distinguir las cuentas maliciosas de las cuentas benignas, y para generar un modelo de clasificación para determinar una puntuación basada en la probabilidad de que la cuenta sea maliciosa. En una realización, pueden existir dos etapas para obtener el modelo de clasificación. Una primera etapa se supervisa aprendiendo dónde ser utilizan los datos de aprendizaje para aprender un modelo de clasificación sobre la base de la identificación de un conjunto de características extraídas de los datos de aprendizaje que distinguen las cuentas maliciosas de las cuentas benignas al menos sobre la base de la capacidad de ser memorizadas de las direcciones de correo electrónico incluidas en los datos de aprendizaje. Una segunda etapa es utilizar el modelo de clasificación para determinar si una cuenta entrante es maliciosa. Estas dos etapas se pueden operar secuencialmente o de una manera intercalada, en la que se recogen datos adicionales para entrenar incrementalmente al modelo de clasificación. Por ejemplo, se comparan datos contrastados de una cuenta con el resultado de la determinación automática mediante el modelo de clasificación, y se utilizan a continuación para entrenar al modelo de clasificación. Dichos datos contrastados de la cuenta se pueden determinar mediante el resultado de una o más transacciones en línea, para encontrar que la cuenta es maliciosa o benigna. Por ejemplo, una tarjeta de crédito asociada con la cuenta es rechazada por el banco. Dicho dato contrastado de la cuenta puede ser determinado también por un revisor tal como una persona.

La figura 1 ilustra un escenario 100 de ejemplo de determinación de la probabilidad de que una cuenta 102 sea maliciosa. La cuenta 102 se asocia con una dirección de correo electrónico 104 y con información adicional 106. Un usuario 108 utiliza un dispositivo de cliente 110 para enviar una solicitud 112 de servicios a un sistema informático 116, tal como pago en línea o registro de la cuenta (no mostrado) en una red 114. La solicitud 112 se asocia con la cuenta 102 y puede incluir información de la cuenta 102, tal como la dirección de correo electrónico 104 e información adicional 106. La cuenta 102 y su dirección de correo electrónico 104 asociada se puede utilizar como identificación de la solicitud 112 o del usuario 108. Antes de que los servicios estén disponibles para el usuario 108, un sistema informático 116 determina la probabilidad de que la cuenta 102 sea maliciosa, sobre la base, al menos parcialmente, del grado de capacidad de ser memorizada de la dirección de correo electrónico 104. La capacidad de ser memorizada se refiere a la facilidad de memorizar la dirección de correo electrónico 104 por parte de una persona. El sistema informático 116 puede enviar asimismo un mensaje 118 al dispositivo del cliente 110 para informar del resultado de la determinación. Si la cuenta 102 se considera benigna, los servicios serán proporcionados al usuario 108. De lo contrario, los servicios se mantendrán retenidos y la solicitud 112 se rechaza.

El sistema informático 116 puede utilizar asimismo características adicionales asociadas con la dirección de correo electrónico 104, tal como el dominio de la dirección de correo electrónico 104 y las características asociadas con la información adicional 106 de la cuenta 102, tal como el número de teléfono, el nombre del usuario, el domicilio, etc., para determinar la probabilidad de que la cuenta 102 sea maliciosa. Tal información adicional 106 puede estar asociada con la cuenta 102 y está incluida en la solicitud 112, tal como se muestra en la figura 1. De manera alternativa, la información adicional 106 se puede buscar, obtener y comprobar mediante el sistema informático 116 en una o más bases de datos locales o remotas o de la internet. Por ejemplo, el sistema informático 116 puede utilizar la dirección de correo electrónico 104 como clave para buscar la información adicional 106.

El sistema informático 116 incluye uno o más procesadores 120, y la memoria 122. Cualquier número de módulos de programa, aplicaciones o componentes se puede almacenar en la memoria, incluyendo a modo de ejemplo un sistema operativo, una o más aplicaciones, otros módulos de programa, datos de programa, instrucciones ejecutables por un ordenador. En el ejemplo de la figura 1, existe una pluralidad de módulos almacenados en la memoria 122. Los módulos pueden incluir un módulo de recepción 124, un módulo de aprendizaje 126, un módulo de clasificación 128, un módulo de determinación 130 y un módulo de respuesta 132.

El módulo de recepción 124 recibe la solicitud 112 asociada con la cuenta 102 que contiene la dirección de correo electrónico 104 y/o la información adicional 106, si existe, que están asociadas con la cuenta 102.

El módulo de aprendizaje 126 utiliza uno o más datos etiquetados como datos de aprendizaje para aprender un modelo de clasificación 128 sobre la base de una o más características que distinguen una cuenta maliciosa de una cuenta benigna. Por ejemplo, los datos etiquetados pueden incluir una pluralidad de cuentas que ya están etiquetadas como cuentas maliciosas o benignas a partir de una o más fuentes de datos tales como bases de datos. El módulo de aprendizaje 126 analiza las direcciones de correo electrónico y la información adicional asociada con las cuentas a partir de los datos etiquetados, extrae una o más características que distinguen una cuenta maliciosa de una cuenta benigna a partir de la dirección de correo electrónico y/o de la información adicional. Tales características pueden incluir una o más características que determinan la capacidad de ser memorizadas de las direcciones de correo electrónico.

En algunas realizaciones, las una o más características pueden estar basadas en los dominios de las direcciones de correo electrónico. En otras realizaciones, las una o más características pueden estar basadas en la correlación de la dirección de correo electrónico con el nombre, el domicilio, el número de teléfono, etc. en la información adicional. En otras realizaciones adicionales, las una o más características pueden estar basadas en parte en la información adicional, por ejemplo, la validez del nombre, el domicilio o el número de teléfono, o en las correlaciones entre el nombre de la cuenta y el nombre de una tarjeta de crédito utilizada en una transacción en línea. El módulo de aprendizaje 126 aprende el modelo de clasificación 128 sobre la base de las características obtenidas utilizando uno o más métodos de aprendizaje de máquinas. Por ejemplo, los uno o más métodos de aprendizaje de máquinas incluyen una máquina de vector de soporte (SVM – Support Vector Machine, en inglés). El módulo de aprendizaje 126 puede enseñar de manera continua al modelo de clasificación 128 sobre la base de los nuevos datos.

El modelo de clasificación 128 incluye una o más características que distinguen una cuenta maliciosa de una cuenta benigna. El modelo de clasificación 128 puede incluir asimismo formatos de las características, el valor de umbral correspondiente a cada característica que determina si las cuentas asociadas con ciertas direcciones de correo electrónico y/o información adicional son benignas o maliciosas, y el peso de cada característica o una función de las características determinando que una cuenta respectiva es maliciosa.

El módulo de determinación 130 utiliza el modelo de clasificación 128 para determinar una puntuación para indicar la probabilidad de que la cuenta 102 sea maliciosa. Por ejemplo, el módulo de determinación 130 puede analizar sintácticamente y extraer características de la dirección de correo electrónico 104 y de la información adicional 106 de acuerdo con los formatos preestablecidos requeridos por el modelo de clasificación 128, aplican el modelo de clasificación 128 a las características de la cuenta 102, y determinan una puntuación para indicar la probabilidad de que la cuenta 102 sea maliciosa. Existen varios métodos para la determinación. Por ejemplo, el sistema informático

116 puede calcular una puntuación para indicar la probabilidad de que la cuenta 102 sea maliciosa sobre la base de varios valores de características extraídos de la dirección de correo electrónico 104 y/o de la información adicional 106. Si la puntuación es mayor que un umbral preestablecido, la cuenta 102 se considera una cuenta maliciosa, y el sistema informático 116 rechaza el servicio solicitado por el usuario 108. Si la puntuación es menor del mismo o de un umbral diferente preestablecido, la cuenta 102 se considera una cuenta benigna, y el sistema informático 116 proporciona el servicio solicitado. De manera alternativa, una o más de las características pueden ser establecidas como una prueba decisiva. Si los valores de la característica de la cuenta 102 para tales características están en un rango o rangos preestablecidos o no alcanzan uno o más de los umbrales de la prueba decisiva, se determina que la cuenta 102 es maliciosa independientemente de sus valores de características para otras características. Por ejemplo, si el dominio de la dirección de correo electrónico 104 o la información adicional 116 asociada con la clasificación 102 está enumerado en una o más listas negras (no mostradas), la cuenta 102 se considera maliciosa.

El módulo de respuesta 132 devuelve el mensaje 118 que incluye el resultado del módulo de determinación 130 para el dispositivo del cliente 110.

El sistema informático 116 puede tener configuraciones y módulos diferentes en varias realizaciones. En otra realización (no mostrada en la figura 1), el sistema informático 116 puede no necesitar analizar los datos etiquetados y puede no tener el módulo de aprendizaje 126. El modelo de clasificación 128 puede estar preconfigurado con las características ya prealmacenadas en el sistema informático 116.

La memoria 122 puede incluir una memoria volátil, una memoria no volátil, una memoria extraíble, una memoria no extraíble y/o una combinación de cualquiera de las anteriores. Generalmente, la memoria 122 contiene instrucciones ejecutables por un ordenador que son accesibles y ejecutables por los uno o más procesadores 120. La memoria 122 es un ejemplo de medios legibles por ordenador. Los medios legibles por ordenador incluyen al menos dos tipos de medios legibles por ordenador, a saber, medios legibles por ordenador y medios de comunicaciones.

Los medios de almacenamiento en un ordenador incluyen medios no volátiles, extraíbles y no extraíbles implementados en cualquier método o tecnología para el almacenamiento de información tal como instrucciones legibles por ordenador, estructuras de datos, módulos de programa u otros datos. Los medios de almacenamiento en un ordenador incluyen, pero no están limitados a, una memoria de cambio de fase (PRAM – Phase change RAM, en inglés), una memoria de acceso aleatorio estática (SRAM – Static Random Access Memory, en inglés), una memoria de acceso aleatorio dinámica (DRAM – Dynamic Random Access Memory, en inglés), otros tipos de memoria de acceso aleatorio (RAM – Random Access Memory, en inglés), una memoria de solo lectura (ROM – Read Only Memory, en inglés), una memoria de solo lectura programable borrable eléctricamente (EEPROM – Electrically Erasable Programmable Read Only Memory, en inglés), una memoria rápida u otra tecnología de memoria, una memoria de solo lectura de disco compacto (CD-ROM – Compact Disk Read Only Memory, en inglés), discos versátiles digitales (SVS – Digital Versatile Disks, en inglés) u otro almacenamiento óptico, casetes magnéticos, cinta magnética, almacenamiento en disco magnético u otros dispositivos de almacenamiento, o cualquier otro medio no de transmisión que se pueda utilizar para almacenar información para acceso por medio de un dispositivo informático.

En contraste, los medios de comunicación pueden realizar instrucciones legibles por ordenador, estructuras de datos, módulos de programa u otros datos en una señal de datos modulada, tal como una onda portadora u otro mecanismo de transmisión. Tal como se define en esta memoria, los medios de almacenamiento no incluyen medios de comunicación.

El sistema informático 116 puede ser el mismo que o independiente del sistema informático que proporciona el servicio en línea. En algunas realizaciones, el sistema informático 116 puede basarse también en la solicitud 112 a otro sistema informático (no mostrado) para determinar la probabilidad de que la cuenta 102 sea maliciosa.

El sistema informático 116 es solo un ejemplo y no pretende sugerir ninguna limitación al alcance de utilización o funcionalidad de las arquitecturas del ordenador o de la red. En aras de una descripción conveniente, el sistema anterior está dividido por funcionalidades en varios módulos que se describen de manera separada. Cuando se implementa el sistema descrito, las funciones de varios módulos se pueden implementar en uno o más casos de software y/o software.

El dispositivo del cliente 110 o el sistema informático 116 se pueden utilizar en un entorno o en una configuración de sistemas informáticos universales o especializados. Ejemplos del dispositivo del cliente 110 o del sistema informático 116 pueden incluir un ordenador personal, un ordenador servidor, un dispositivo de mano o un dispositivo portátil, un dispositivo de tableta, un sistema de múltiples procesadores, un sistema basado en microprocesadores, un descodificador, un dispositivo electrónico de abonado programable, un PC de red y un entorno informático distribuido que incluye cualquier sistema o dispositivo anterior.

En el entorno informático distribuido, una tarea es ejecutada por dispositivos de procesamiento remotos que están conectados a través de una red de comunicación. En el entorno informático distribuido, los módulos pueden estar situados en medios de almacenamiento en un ordenador (que incluyen dispositivos de almacenamiento de datos) de

ordenadores locales y remotos. Por ejemplo, el sistema informático 116 puede utilizar algunos o todos los módulos anteriores, tales como el módulo de recepción 124, el módulo de aprendizaje 126, el módulo de clasificación 128, el módulo de determinación 130 y el módulo de respuesta 132.

A continuación, se explican con detalle métodos de ejemplo para realizar las técnicas descritas en esta memoria. Estos métodos de ejemplo se pueden describir en el contexto general de las instrucciones ejecutables por un ordenador. En general, las instrucciones ejecutables por un ordenador pueden incluir rutinas, programas, objetos, componentes, estructuras de datos, procedimientos, módulos, funciones y otros, que realizan funciones particulares o implementan tipos de datos abstractos particulares. Los métodos se pueden poner en práctica también en un entorno informático distribuido en el que las funciones son realizadas por dispositivos de procesamiento remotos que están conectados a través de una red de comunicación o una nube de comunicación. En un entorno informático distribuido, las instrucciones ejecutables por un ordenador pueden estar situadas en memorias tanto locales como remotas.

Por conveniencia, los métodos se describen a continuación en el contexto del sistema informático 116 y en el entorno de la figura 1. No obstante, los métodos no están limitados a la implementación en este entorno.

Los métodos de ejemplo se ilustran como una colección de bloques en un gráfico de flujo lógico que representa una secuencia de operaciones que pueden ser implementadas en hardware, software, firmware o una combinación de los mismos. A menos que se indique expresamente otra cosa, el orden en el que se describen los métodos no pretende considerarse como una limitación, y se puede combinar cualquier número de bloques de métodos en cualquier orden para implementar los métodos, o alternar los métodos. Adicionalmente, se pueden omitir operaciones individuales de los métodos sin separarse del espíritu y alcance del asunto central descrito en esta memoria. En el contexto del software, los bloques representan instrucciones informáticas que, cuando son ejecutadas por uno o más procesadores, realizan las operaciones citadas.

La figura 2 ilustra un diagrama de flujo de un método de ejemplo para determinar si la cuenta 102 es maliciosa, sobre la base de una o más características del correo electrónico 104 y/o de la información adicional 106 asociada con la cuenta 102.

En 202, el sistema informático 116 recibe la cuenta 102 e información asociada con la cuenta 102 que incluye su dirección de correo electrónico 104 y/o información adicional 106. En 204, el sistema informático 116 extrae características de la información, tal como el correo electrónico 104 e información adicional 106 asociada con la cuenta 102. En 206, el sistema informático 116 determina un nivel de confianza de la cuenta 102 sobre la base de los valores de las características extraídas.

Por ejemplo, el sistema informático 116 puede analizar sintácticamente la dirección de correo electrónico 104 de acuerdo con una o más características que determinan el grado de capacidad de ser memorizada de la dirección de correo electrónico 104, y obtener los valores de las características correspondientes de la dirección de correo electrónico 104. Las una o más características pueden comprender características relativas a secuencias con significado, características relativas a secuencias pronunciables, características relativas a longitudes de números y sus posiciones en la dirección de correo electrónico, características relativas a un patrón que incluye simetría o anti simetría de secuencias vecinas, o caracteres separados uniformemente en la dirección de correo electrónico. El sistema informático 116 puede calcular un grado de capacidad de ser memorizada de la dirección de correo electrónico 104 sobre la base de los correspondientes valores de características. El sistema informático 116 puede convertir asimismo uno o más caracteres o números para encontrar las secuencias de caracteres con significado de acuerdo con un preajuste de las reglas. Por ejemplo, el número “2” en la dirección de correo electrónico 104 se puede tratar como los caracteres “to”. Algunas características y reglas de conversión de ejemplo se enumeran con detalle a continuación.

Como ejemplo adicional, el sistema informático 116 puede asimismo analizar sintácticamente la dirección de correo electrónico 104 de acuerdo con una o más de las características que se refieren al dominio de la dirección de correo electrónico 104, a la correlación de la dirección de correo electrónico 104 con los datos en la información adicional 106 tal como nombre, domicilio, número de teléfono. El sistema informático 116 puede analizar sintácticamente la dirección de correo electrónico 106 de acuerdo con una o más características que se refieren a la validez del nombre, domicilio, número de teléfono, y comprobar la correlación entre la información adicional 106 de la cuenta 102 con información correspondiente de una tarjeta de crédito utilizada en una transacción en línea. Por ejemplo, en el caso de que se determine que la tarjeta de crédito es una tarjeta de crédito no válida o maliciosa tal como que la tarjeta de crédito sea rechazada por el banco y una o más de la información adicional 106 coincida con la información correspondiente de la tarjeta de crédito, la clasificación 102 puede ser determinada como maliciosa.

Las características pueden estar prealmacenadas en el sistema informático 116 o pueden ser aprendidas por el sistema informático 116 a través de los datos etiquetados procedentes de una o más fuentes por lotes o de una manera progresiva. Por ejemplo, el sistema informático 116 puede analizar cuentas etiquetadas y sus direcciones de correo electrónico y/o información adicional asociadas procedentes de una o más fuentes. Cada una de las cuentas etiquetadas indica que una cuenta etiquetada respectiva es una cuenta maliciosa o una cuenta benigna. El sistema

informático 116 puede aprender una o más características que distinguen una cuenta maliciosa de una cuenta benigna utilizando uno o más métodos de aprendizaje de máquinas tal como SVM. Las una o más características que distinguen una cuenta maliciosa de una cuenta benigna se pueden extraer de una dirección de correo electrónico asociada con la cuenta sobre la base, al menos parcialmente, de la memorización de la dirección de correo electrónico. Las unas o más características se pueden extraer de una dirección de correo electrónico sobre la base de otra información de la dirección de correo electrónico, tal como el dominio de la dirección de correo electrónico, la correlación de la dirección de correo electrónico con los datos en la información adicional 106. Adicionalmente, las una o más características se pueden extraer de la información adicional 106. El sistema informático 116 obtiene el modelo de clasificación 128 sobre la base de las características obtenidas; y utiliza el modelo de clasificación 128 para calcular una puntuación para indicar la probabilidad de que la cuenta sea maliciosa.

En 206, el sistema informático 116 puede determinar un nivel de confianza de la cuenta 101 sobre la base, al menos parcialmente, del grado de capacidad de ser memorizada determinado, de la dirección de correo electrónico 104 asociada con la cuenta 102. En una realización, el sistema informático 116 puede determinar una puntuación sobre la base del grado de capacidad de ser memorizada calculado de la dirección de correo electrónico 104. Si la puntuación calculada es mayor que un umbral preestablecido, el sistema informático 116 determina que la cuenta 102 es maliciosa. Si la puntuación calculada es menor que el umbral preestablecido, el sistema informático 116 determina que la cuenta 102 es benigna.

En otra realización, además del grado de capacidad de ser memorizada de la dirección de correo electrónico 104 asociada con la cuenta 102, el sistema informático 116 puede determinar asimismo la probabilidad de que la cuenta 102 sea maliciosa considerando otras características del correo electrónico 104. Por ejemplo, el sistema informático 116 puede obtener el dominio de la dirección de correo electrónico 104, la correlación de la dirección de correo electrónico 104 con la información adicional 106 de la cuenta 102 tal como el nombre del usuario, el número de teléfono y el domicilio. En otra realización más, el sistema informático 116 puede determinar asimismo la probabilidad de que la cuenta 102 sea maliciosa considerando una o más características de la información adicional 106 de la cuenta 102. Por ejemplo, el sistema informático 116 puede determinar una distancia de que un nombre asociado con la cuenta 102 con respecto al nombre real, una distancia del domicilio de una dirección válida, la correlación del teléfono y el domicilio, la correlación del nombre, el domicilio y el teléfono asociados con la cuenta con una tarjeta de crédito que fue / es utilizada en una o más transacciones en línea asociadas con la cuenta 102. Tal información adicional 106 puede ser proporcionada por el usuario 108 y enviada al sistema informático 116 junto con la dirección de correo electrónico 104 previamente, tal como durante el registro de la cuenta, o en la solicitud 112. De manera alternativa, el sistema informático 116 puede utilizar algunas técnicas de búsqueda para utilizar la cuenta 102 y/o la dirección de correo electrónico 104 como clave para buscar en línea o de una o más bases de datos para encontrar la información adicional 106 asociada con la cuenta 102. El sistema informático 116 puede recibir o utilizar algunas técnicas de búsqueda para encontrar información utilizada para extraer las características de la dirección de correo electrónico 104 y/o la información adicional 106.

El sistema informático 116 puede calcular un nivel de confianza del dominio de la dirección de correo electrónico 104 y la información adicional 106 respectivamente.

Por ejemplo, el sistema informático 116 puede almacenar cuatro listas que incluyen una lista blanca de dominios, una lista negra de dominios, una lista benigna de dominios y una lista maliciosa de dominios para determinar y grabar un nivel de confianza de los dominios. Estas listas pueden estar predeterminadas o basarse en estadísticas o en el cálculo de los datos de aprendizaje que incluyen los datos etiquetados y los nuevos datos etiquetados entrantes continuos. En alguna realización, el sistema informático 116 puede almacenar solamente una o más de las cuatro listas.

El sistema informático 116 puede contar el número de veces que un dominio respectivo está asociado con cuentas etiquetadas como benignas y/o en cuentas etiquetadas como maliciosas en los datos de aprendizaje y calcula un nivel de confianza del respectivo dominio sobre la base de tales contenidos. De manera alternativa, el sistema informático 116 puede calcular un nivel de confianza de que el respectivo dominio esté asociado con cuentas benignas y un nivel de confianza de que el respectivo dominio esté asociado con cuentas maliciosas respectivamente.

La lista blanca de dominios enumera los dominios que se asumen asociados con cuentas benignas. Por ejemplo, la lista blanca de dominios puede incluir dominios de empresas, escuelas, o instituciones de confianza que existen en la lista blanca. Como ejemplo adicional, la lista blanca de dominios puede incluir asimismo dominios cuyos niveles de confianza se calcula que están por encima de un umbral tal como 100% o 99%. En otras palabras, existe al menos 99% o 100% de probabilidad de que tales dominios sean de confianza y estén asociados con cuentas benignas.

La lista negra de dominios enumera los dominios que se suponen asociados con cuentas maliciosas. Por ejemplo, la lista negra de dominios puede incluir dominios de atacantes que existen en la lista negra. Como ejemplo adicional, la lista negra de dominios puede incluir asimismo dominios cuyos niveles de confianza se calcula que están por debajo

de un umbral, tal como 0% o 1%. En otras palabras, existe solo 0% o 1% de probabilidad de que tales dominios sean de confianza, y por lo tanto un 100% o un 99% de probabilidad de que estén asociados con cuentas maliciosas.

Si el nivel de confianza del dominio respectivo basado en los datos de aprendizaje actuales es insuficiente para enumerar los respectivos dominios en la lista negra o la lista blanca, el dominio respectivo puede estar enumerado en la lista benigna de dominios y/o en la lista maliciosa de dominios. La lista benigna de dominios puede incluir el dominio respectivo y una cuenta de distintas direcciones de correo electrónico de dicho dominio, asociadas con las cuentas etiquetadas como benignas analizando los datos de aprendizaje. La lista benigna puede incluir asimismo un nivel de confianza de que el dominio respectivo esté asociado con cuentas etiquetadas como benignas mediante cálculo basado en dichas cuentas. La lista benigna de dominios puede incluir el dominio respectivo y una cuenta de direcciones de correo electrónico distintas de tal dominio, asociadas con cuentas etiquetadas como maliciosas analizando los datos de aprendizaje. La lista benigna puede incluir asimismo un nivel de confianza de que el respectivo dominio esté asociado con cuentas maliciosas mediante el cálculo basado en tales cuentas.

En una realización, la lista benigna de dominios y la lista maliciosa de dominios no son exclusivas. Un dominio puede estar enumerado tanto en la lista benigna de dominios como en la lista maliciosa de dominios. El nivel de confianza o probabilidad de que un dominio respectivo esté asociado con una cuenta benigna, y el nivel de confianza o probabilidad de que el dominio esté asociado con una cuenta maliciosa, puede ser añadido para igualar a 1 o 100%. Por ejemplo, después de que se ha reunido una cantidad suficiente de datos etiquetados para el respectivo dominio, la fórmula para calcular el nivel de confianza de que el dominio respectivo esté asociado con cuentas benignas puede ser el número de veces que el respectivo dominio es asociado con cuentas etiquetadas como benignas dividido por el número total de veces que el respectivo dominio es asociado con todas las cuentas (incluidas tanto las cuentas etiquetadas como benignas como las cuentas etiquetadas como maliciosas) en los datos de aprendizaje. Después de que se ha reunido una cantidad suficiente de datos etiquetados para el respectivo dominio, la fórmula para calcular el nivel de confianza de que el dominio respectivo esté asociado con cuentas maliciosas puede ser el número de veces que el respectivo dominio es asociado con cuentas etiquetadas como maliciosas dividido por el número total de veces que el respectivo dominio es asociado con todas las cuentas (incluidas tanto las cuentas etiquetadas como benignas como las cuentas etiquetadas como maliciosas) en los datos de aprendizaje. Por ejemplo, tal suficiencia puede estar basada en que un umbral de datos etiquetados han sido reunidos y analizados para el dominio respectivo.

Con un cambio del nivel de confianza del respectivo dominio, tal dominio puede estar enumerado en diferentes listas durante un periodo de tiempo.

El sistema informático 116 puede comparar el dominio y la dirección de correo electrónico 104 con los de la lista blanca y/o la lista negra para determinar adicionalmente si la cuenta 102 asociada con la dirección de correo electrónico 104 es maliciosa o no. Por ejemplo, incluso si el grado de capacidad de ser memorizada de la dirección de correo electrónico 104 calculado es mayor que el umbral, es decir, se determina que la dirección de correo electrónico es fácil de recordar por un humano, el sistema informático 116 puede determinar que la cuenta 102 a la que está asociada la dirección de correo electrónico 104 es maliciosa si se encuentra que alguno de los dominios de la dirección de correo electrónico 104 está en la lista negra. El sistema informático 116 puede comprobar asimismo la lista benigna de dominios y/o la lista maliciosa de dominios para obtener un nivel de confianza del dominio.

El sistema informático 116 puede almacenar listas adicionales de información adicional con el fin de evaluar un nivel de confianza de un dominio de una dirección de correo electrónico. Por ejemplo, de manera similar a las cuatro listas de dominios, el sistema informático 116 puede almacenar también cuatro listas, incluidas una lista blanca de información adicional, una lista negra de información adicional, una lista benigna de información adicional y una lista maliciosa de información adicional, para determinar y registrar un nivel de confianza de la información adicional. Dado que la información adicional puede incluir una pluralidad de elementos asociados con la cuenta, tal como el nombre, el número de teléfono, la dirección física, la dirección de IP de la fuente de la solicitud, etc., el sistema informático 116 puede almacenar dichas cuatro listas para cada una de la información adicional, tal como una lista blanca de números de teléfono, una lista negra de números de teléfono, una lista benigna de números de teléfono y una lista maliciosa de números de teléfono. En algunas realizaciones, el sistema informático 116 puede almacenar solo una o más de las cuatro listas. El sistema informático 116 puede almacenar asimismo solo las listas para algunos de los elementos de la información adicional, tal como nombre y número de teléfono.

De manera similar a las cuatro listas descritas anteriormente con respecto al dominio, la lista blanca de listas de información adicional enumera la información adicional que se supone asociada con cuentas benignas. La lista negra de dominios enumera la información adicional que se supone asociada con cuentas maliciosas. Si el nivel de confianza de la respectiva información adicional basado en los datos de aprendizaje actuales es insuficiente para enumerar el dominio respectivo bien en la lista negra o en la lista blanca, el respectivo dominio puede estar enumerado en la lista benigna de información adicional y/o en la lista maliciosa de información adicional. La lista benigna de información adicional puede incluir la información adicional respectiva y una cuenta de tal información adicional asociada con las cuentas etiquetadas como benignas analizando los datos de aprendizaje. La lista benigna puede incluir asimismo un nivel de confianza de que la información adicional respectiva esté asociada con cuentas

benignas mediante cálculo basado en dichas cuentas. La lista maliciosa de información adicional puede incluir la información adicional respectiva y una cuenta de tal información adicional asociada con cuentas maliciosas analizando los datos de aprendizaje. La lista maliciosa puede incluir asimismo un nivel de confianza de que la información adicional respectiva esté asociada con cuentas maliciosas mediante cálculo basado en dichas cuentas.

Dichas lista blanca, lista negra, lista benigna y/o lista maliciosa de dominios y/o la respectiva información adicional se pueden obtener de una o más fuentes o se puede introducir manualmente. Dicha lista maliciosa o benigna se puede obtener a partir del análisis sintáctico y del análisis de los datos de aprendizaje. Alternativa o adicionalmente, dicha lista maliciosa o lista benigna se puede obtener de una fuente separada, tal como una base de datos que de manera independiente reúne y guarda dominios maliciosos o benignos y/o información adicional asociada con las cuentas.

La figura 3 ilustra un diagrama de flujo de un método 300 de ejemplo de técnicas de aprendizaje de máquinas que identifican características y generan el modelo de clasificación 128 para determinar un nivel de confianza de la cuenta 101. En una realización, pueden existir dos etapas para obtener el modelo de clasificación.

En 302, el sistema informático 116 utiliza uno o más datos etiquetados para aprender o construir el modelo de clasificación 128 basado en una o más características que distinguen una cuenta maliciosa de una cuenta benigna. Los uno o más datos etiquetados actúan como datos de aprendizaje que incluyen una pluralidad de cuentas etiquetadas asociadas con direcciones de correo electrónico y posiblemente información adicional procedente de una o más fuentes, tal como una base de datos de cuentas maliciosas y una base de datos de cuentas benignas. Cada una de las cuentas en los datos etiquetados está etiquetada como maliciosa o benigna.

En 304, el sistema informático 116 utiliza el modelo de clasificación para determinar un nivel de confianza de la cuenta 102. En algunas realizaciones, si el nivel de confianza calculado es mayor que un umbral preestablecido, se determina que la cuenta 102 es benigna. Por otro lado, si el nivel de confianza calculado de la cuenta es menor que un umbral preestablecido, que puede ser el mismo o diferente del umbral preestablecido anterior, se determina que la cuenta es maliciosa.

Las operaciones en 302 y 304 se pueden realizar secuencialmente. Por ejemplo, en un aprendizaje por lotes de los datos etiquetados, el modelo de clasificación 128 se obtiene y utiliza para clasificar la dirección de correo electrónico 104.

De manera alternativa, las operaciones en 302 y 304 se pueden operar de una manera intercalada y el modelo de clasificación 128 se mejora con un resultado del aprendizaje en línea. Además de los datos etiquetados, se utilizan datos adicionales tales como una nueva cuenta entrante etiquetada como la cuenta 102 para enseñar incrementalmente al modelo de clasificación 128, mientras que el modelo de clasificación 128 se utiliza en la clasificación de las cuentas en línea. Por ejemplo, se pueden obtener los datos contrastados de la cuenta 102, y compararlos con el resultado determinado mediante el modelo de clasificación 128. Si los datos contrastados están de acuerdo con el resultado, el modelo de clasificación 128 y sus características incluidas se confirman. Si los datos contrastados no están de acuerdo con el resultado, el modelo de clasificación 128 no se confirma. Las nuevas cuentas etiquetadas y/o los datos contrastados devueltos de las cuentas determinadas previamente se pueden utilizar para enseñar incrementalmente al modelo de clasificación 128 y, por ello, el modelo de clasificación 128 mejora de manera continua con nuevos datos de aprendizaje. Por ejemplo, los datos contrastados de la cuenta 102 pueden ser un resultado revisado manualmente por un humano para determinar si la clasificación 102 es maliciosa o benigna.

El sistema informático 116 puede utilizar una o más técnicas de aprendizaje de máquinas para enseñar al modelo de clasificación. Por ejemplo, las técnicas de aprendizaje de máquinas pueden incluir una máquina de vector de soporte (SVM). Se puede aplicar tanto la SVM lineal como no lineal, tal como la Función de Base Radial polinomial o gausiana.

La figura 4 ilustra un diagrama de flujo de un método 400 de ejemplo de técnicas de aprendizaje de máquinas que enseñan continuadamente al modelo de clasificación 128.

En 402, se reciben múltiples cuentas etiquetadas de una o más fuentes. Cada una de las cuentas etiquetadas indica si la respectiva cuenta etiquetada es una cuenta maliciosa o una cuenta benigna. Cada cuenta puede estar asociada con una dirección de correo electrónico y/o información adicional.

En 404, una o más características que se utilizan para distinguir cuentas maliciosas de cuentas benignas se extraen de la pluralidad de cuentas etiquetadas. Por ejemplo, el sistema informático 116 puede extraer una o más características de las direcciones de correo electrónico y/o la información adicional asociada con las cuentas. El sistema informático 116 puede utilizar las una o más características extraídas y los resultados etiquetados de las cuentas para enseñar al modelo de clasificación 128 utilizando uno o más métodos de aprendizaje de máquinas. Por ejemplo, el sistema informático 116 puede utilizar SVM para producir o construir el modelo de clasificación 128. Algunas de las características que el sistema informático 116 extrae de las cuentas pueden estar relacionadas con la

capacidad de ser memorizada de una dirección de correo electrónico. Las características adicionales pueden estar basadas en los dominios de las direcciones de correo electrónico y/o en la correlación correspondiente de una dirección de correo electrónico con la información adicional asociada con la misma cuenta. Además, las características que el sistema informático 116 extrae de las cuentas pueden estar basadas en la información adicional. Por ejemplo, algunas características pueden estar basadas en la validez del nombre, domicilio, número de teléfono, correlación de la información de la cuenta con la de una tarjeta de crédito que ha sido o es utilizada en una o más transacciones en línea. Además, el sistema informático 116 puede asimismo modificar el valor de umbral de cada característica en la determinación del nivel de confianza de la cuenta.

En 406, el modelo de clasificación 128 se aprende o construye utilizando una o más técnicas de aprendizaje de máquinas basadas en las características extraídas y en los resultados de etiquetado de las cuentas. Por ejemplo, se puede utilizar SVM en 406 para producir el modelo de clasificación 128.

En 408, el modelo de clasificación 128 se utiliza para determinar un nivel de confianza de la cuenta 102. En algunas realizaciones, el nivel de confianza calculado se compara con uno o más umbrales para determinar que la cuenta 102 es maliciosa, benigna o incierta.

En 410, el resultado determinado por el modelo de clasificación 128 se compara con los datos contrastados para mejorar el modelo de clasificación 128. Por ejemplo, los datos contrastados de las cuentas se pueden comparar con los resultados determinados por el modelo de clasificación 128 para enseñar incrementalmente al modelo de clasificación 128, de tal manera que el modelo de clasificación 128 se mejora. Si los datos contrastados están de acuerdo con el resultado, el modelo de clasificación 128 y sus características incluidas se confirma. Si los datos contrastados no están de acuerdo con el resultado, el modelo de clasificación 128 no se confirma.

En 412, se reciben nuevos datos etiquetados, y el modelo de clasificación 128 se adapta incrementalmente con nuevas cuentas etiquetadas. Los nuevos datos etiquetados pueden ser nuevos datos etiquetados entrantes o datos etiquetados previamente con cambios en la etiqueta, de tal manera que una cuenta previamente etiquetada como maliciosa se etiqueta como benigna y viceversa. Por ejemplo, se pueden aplicar nuevas cuentas etiquetadas para enseñar al modelo de clasificación 128 incrementalmente para adaptarse a o rastrear cambios de las cuentas. El sistema informático 116 puede continuar recibiendo nuevos datos etiquetados entrantes para adaptar el modelo de clasificación 128.

Lo que sigue enumera algunas características de ejemplo que se refieren a la capacidad de ser memorizada de una dirección de correo electrónico 104. Una o más de estas características se pueden extraer de la dirección de correo electrónico 104 asociada con el sistema informático 116 de la cuenta 102. La contribución de cada una de las características de ejemplo siguientes al nivel de confianza calculado final de la cuenta puede ser determinada por el modelo de clasificación 128. Por ejemplo, las características y sus respectivas ponderaciones en la determinación del nivel de confianza pueden ser las mismas o diferentes durante un periodo de tiempo y se pueden añadir, borrar o modificar a partir del aprendizaje, incluyendo mejorar o adaptar el modelo de clasificación 128 tal como se ha descrito anteriormente.

m_EmailAccountLength. Esta característica representa el número de caracteres en la cuenta de la dirección de correo electrónico antes del símbolo “@” en una dirección de correo electrónico. Tras el símbolo “@”, se encuentra el dominio de la dirección de correo electrónico. Para el ejemplo de la dirección de correo electrónico xuefei0917@gmail.com, su m_EmailAccountLength es 10. El dominio es gmail.com. Para el ejemplo de la dirección de correo electrónico fkajklfa8971jfjlajfqiu@gmail.com, su m_EmailAccountLength es 22, que es difícil de recordar y podría ser asociada con una cuenta maliciosa. Por ejemplo, en el caso de que el número sea alto, la dirección de correo electrónico puede ser creada por una máquina y, por ello, la cuenta puede ser sospechosa. La contribución de esta característica al nivel de confianza calculado final de la cuenta está determinada por el modelo de clasificación 128.

m_NumberMemorableLength. Esta característica representa el número de números memorizables totales, incluidos fecha, caracteres repetidos y simetría, tales como 19870919, 123456, 135531. Para el ejemplo de la dirección de correo electrónico zsf59823@gmail.com, su m_NumberMemorableLength es 0, porque “58923” parece no memorizable. En un ejemplo, los números tales como fechas, simetría, repetidos se pueden considerar como memorizables en su conjunto (es decir, 100% memorizables). Para el ejemplo de la dirección de correo electrónico zsf123321@gmail.com, su m_NumberMemorableLength es 6, lo que describe “123321”. Se debe observar que “123321” es simétrico. En otro ejemplo, los números con simetría se pueden considerar parcialmente memorizables. Para el ejemplo de una secuencia de números “378873”, los últimos 3 números se pueden obtener a partir de los tres primeros números y, de este modo, se supone que no se necesita ningún esfuerzo para recordar los últimos 3 números. Pero una persona puede aún necesitar recordar los tres primeros números, dado que parece que no tienen significado (y por ello es necesario un cierto esfuerzo para recordarlos). Por lo tanto, esta secuencia de números se considera como 3 dígitos memorizables y 3 dígitos no memorizables, es decir, 50% memorizable. Para el mismo ejemplo anterior, si el domicilio en la información adicional 106 contiene 378, por ejemplo, 378 forma parte del código postal del domicilio, entonces estos tres números se consideran también memorizables, y la dirección de correo electrónico contiene, en este caso, 6 dígitos memorizables. En el caso de que el número o la tasa de

números memorizables sea bajo, el correo electrónico puede tender a ser generado por una máquina, y la cuenta asociada puede ser sospechosa.

m_NumberofLetterStrings. Esta característica representa el número de secuencias separadas que incluyen letras. Para el ejemplo de la dirección de correo electrónico xuefei0917@gmail.com, su m_NumberofLetterStrings es 1. Para el ejemplo de dirección de correo electrónico xf0917zww@gmail.com, su m_NumberofLetterStrings es 2, es decir, “xf” y zwww”. Para el ejemplo de la dirección de correo electrónico f7fa18foa@gmail.com, su m_NumberofLetterStrings es 3, es decir, “f”, “fa” y “foa”. Por ejemplo, en caso de que el número sea alto, la cuenta asociada con la dirección de correo electrónico puede ser sospechosa.

m_NumofMemorable. Esta característica representa el número de sub-secuencias memorizables que incluyen letras tales como nombre, y otros. Para el ejemplo xuefei0917@gmail.com, su m_NumofMemorable es 2, como “xue” y “fei” son las dos memorizables como deletreos chinos. Para el ejemplo de nicholas@gmail.com, su m_NumofMemorable es 1, dado que incluye un nombre “Nicholas” en inglés. Para el ejemplo de dirección de correo electrónico gkjhgfhja@163.com, su m_NumofMemorable es 0, puesto que no se detectan palabras en la dirección de correo electrónico. Por ejemplo, en el caso de que el número sea alto, la cuenta asociada con la dirección de correo electrónico puede ser sospechosa.

m_LengthofMemorable. La característica representa el número total de caracteres en las sub-secuencias memorizables que incluyen letras. Para el ejemplo de la dirección de correo electrónico xuefei0917@gmail.com, su m_LengthofMemorable es 6, puesto que el número total de caracteres en las dos secuencias de letras memorizables “xue” y “fei” es 6. Para el ejemplo de la dirección de correo electrónico nicholas.zxy@gmail.com, su m_LengthofMemorable es 8, puesto que el número de caracteres en la secuencia de letras memorizable “Nicholas” es 8. Por ejemplo, en el caso de que el número sea bajo, la cuenta asociada con la dirección de correo electrónico puede ser sospechosa.

m_MemorableRate. Esta característica representa la tasa memorizable que se mide como el número total de caracteres en las secuencias de letras memorizables dividido por el número total de caracteres en la cuenta de correo electrónico antes del símbolo @. Para el ejemplo de la dirección de correo electrónico xuefei0917@gmail.com, su m_MemorableRate es 1.0, porque toda la secuencia de letras es memorizable. Para el ejemplo de la dirección de correo electrónico nicholas.zxy@gmail.com, su m_MemorableRate es 0,727 (es decir, 8/11), dado que el número total de caracteres en la secuencia de letras memorizable es 8, mientras que el número total de caracteres es 11. Para el ejemplo de la dirección de correo electrónico gkjhgfhja@163.com, su m_MemorableRate es 0, puesto que no hay ninguna secuencia de letras memorizable en la dirección de correo electrónico. Por ejemplo, en el caso de que el número sea bajo, la cuenta asociada con la dirección de correo electrónico puede ser sospechosa.

m_FrontMemorableConfidenceLevel. Esta característica representa el nivel de confianza en el rango de 0 a 1, para la cuenta de correo electrónico que empieza con una secuencia memorizable. Los métodos de cálculo detallado pueden variar. Para el ejemplo, de la dirección de correo electrónico xuefei0917@gmail.com su m_FrontMemorableConfidenceLevel es alto, tal como 0,9, porque “xue” se considera una secuencia memorizable en la posición frontal. Para el ejemplo de la dirección de correo electrónico lijffs@gmail.com, su m_FrontMemorableConfidenceLevel es relativamente incierto, tal como 0,5, porque la subsecuencia que empieza con “li” se considera con una probabilidad de 0,5 de ser una secuencia memorizable. Por ejemplo, en el caso de que el número sea bajo, la cuenta asociada con la dirección de correo electrónico puede ser sospechosa.

m_EndMemorableConfidenceLevel. Esta característica representa el nivel de confianza en el rango de 0 a 1, para la cuenta de correo electrónico que finaliza con una secuencia memorizable. Las características pueden no considerar ningún número al final de la dirección de correo electrónico. Para el ejemplo de la dirección de correo electrónico xuefei0917@gmail.com, su m_EndMemorableConfidenceLevel es alto, tal como 0,9, porque “fei” se considera una secuencia razonable en la posición final de letras.

m_MaxLengthofMemorable. Esta característica representa el número de caracteres en la subsecuencia de letras memorizable más larga. Para el ejemplo del correo electrónico de la dirección de correo electrónico nicholas21eo2ben@gmail.com, su m_MaxLengthofMemorable es 8, puesto que la secuencia de letras memorizable más larga es “nicholas”, que tiene 8 caracteres.

m_DistanceBetweenMemorable. Esta característica representa la distancia más grande entre dos subsecuencias memorizables. Para el ejemplo de la dirección de correo electrónico xuefei0917@gmail.com, su m_DistanceBetweenMemorable es 0, puesto que no hay nada entre las dos subsecuencias memorizables “xue” y “fei”. Para el ejemplo de jobghjfsdfhtown@gmail.com, su m_DistanceBetweenMemorable es 8, de la palabra “job” a “town”. Por ejemplo, en el caso de que el número sea alto, la cuenta asociada con la dirección de correo electrónico puede ser sospechosa.

m_MaxNonmemorableLength. Esta característica representa el número de caracteres en la secuencia de letras no memorizable más larga. Esta difiere de la característica anterior por considerar todas las secuencias no

memorizables, mientras que la característica anterior m_DistanceBetweenMemorable considera solo las secuencias no memorizables entre secuencias memorizables. Para el ejemplo de la dirección de correo electrónico xuefei0917@gmail.com, su m_MaxNonmemorableLength es 0, puesto que las dos secuencias de letras “xue” y “fei” son memorizables. Para el ejemplo de xuefeihhfg0917@gmail.com, su m_MaxNonmemorableLength es 4, por la secuencia de letras sin significado “hhfg”. Para el ejemplo de la dirección de correo electrónico ghfiafsdk@gmail.com, su m_MaxNonmemorableLength es 9, por “ghfiafsdk”. Por ejemplo, en el caso de que el número sea alto, la cuenta asociada con la dirección de correo electrónico puede ser sospechosa.

m_BreakPoints. Esta característica representa el número total de secuencias no memorizables compuestas por letras en la cuenta de correo electrónico. Por ejemplo, para xuefei0917@gmail.com, el m_BreakPoints es 0, porque “xue” y “fei” son memorizables, mientras que para kjxuebbbhfei98j@gmail.com, el m_BreakPoints es 3, porque hay 3 subsecuencias no memorizables, “kj”, “bbbh” y “j”. Un valor de esta característica alto puede indicar que la cuenta asociada con el correo electrónico probablemente es maliciosa.

m_NumberofNonmemorable. Esta característica representa el número de secuencias no memorizables que incluyen letras o números en la dirección de correo electrónico. Esta característica es diferente de m_BreakPoints, puesto que esta última no considera las subsecuencias no memorizables compuestas por números. Para el ejemplo de la dirección de correo electrónico jobs472fhs@gmail.com, su m_NumberofNonmemorable es 2, por “472” y “fhs”, mientras que su m_BreakPoints es 1, por “fhs”. Para el ejemplo de la dirección de correo electrónico gjh783ffsj04571fua@gmail.com, su m_NumberofNonmemorable es 5, por “gjh”, “783”, “ffsj”, “04571” y “fua”, mientras que su m_BreakPoints es 0, porque no tiene ninguna secuencia memorizable.

n-gram average y n-gram max. Esta característica representa el valor medio de n-gram en la dirección de correo electrónico. Un n-gram es una secuencia de n caracteres contiguos. Un modelo de n-gram es un modelo de lenguaje probabilístico, que es la probabilidad condicional del ultimo carácter xi, dados los (n-1) caracteres anteriores xi-(n-1), xi-(n-2), xi-1. El modelo de n-gram se puede obtener para una pluralidad de palabras, por ejemplo, un diccionario. La secuencia de un correo electrónico antes del símbolo “@” se puede dividir en una pluralidad de n caracteres continuos. A cada subsecuencia de n caracteres continuos se le asigna una probabilidad o peso mediante el modelo n-gram. La media de estas probabilidades de n-gram es la media de n-gram, mientras que el valor máximo de estas probabilidades de n-gram es el max de n-gram. En algunas realizaciones, el n-gram utilizado para determinar el nivel de confianza de una cuenta puede ser al menos uno de un 2-gram, 3-gram, 4-gram y un 5-gram. Por ejemplo, la secuencia “trean” en una dirección de correo electrónico trean@gmail.com no es una palabra, pero tiene un buen comportamiento en una decisión de n-gram, de tal manera que tiene una mejor media de n-gram, y un max de n-gram que los caracteres aleatorios tales como “ghfjs”, “gkjiu” y “tyttt”, tal como para n = 2, 3, 4 o 5.

m_LengthofNumberStrings. Esta característica representa la longitud de subsecuencias que incluyen números en la dirección de correo electrónico. Para el ejemplo de la dirección de correo electrónico xuefei0917@gmail.com, su m_LengthofNumberStrings es 4, por la subsecuencia “0917”.

m_NumberofNumberStrings. Esta característica representa el número de subsecuencias separadas de números en la dirección de correo electrónico. Para el ejemplo de la dirección de correo electrónico xuefei0917@gmail.com, su m_NumberofNumberStrings es 1, por la subsecuencia “0917”. Para el ejemplo de la dirección de correo electrónico 48hfh519jhfa7888@gmail.com, su m_NumberofNumberStrings es 3, por las subsecuencias “48”, “519” y “7888”.

m_TotalMemorableRate. Esta característica representa la tasa total memorizable, que es la suma de la longitud de subsecuencias memorizables de letras y la longitud del número de subsecuencias memorizables, dividida por la longitud total de la dirección de correo electrónico. Para el ejemplo de la dirección de correo electrónico xuefei0917@gmail.com, su m_TotalMemorableRate es 1, porque la longitud de secuencias memorizables de letras es 6, por “xuefei”, y la longitud de secuencias memorizables de números es 4, por “0917”, que se interpreta como indicativo de una fecha. La longitud total de la dirección de correo electrónico es 10. Así que, la tasa memorizable total es 1, por 10/10. Para el ejemplo de la dirección de correo electrónico benjamin2786ghhf@gmail.com, su m_TotalMemorableRate es 0,5, porque la longitud de subsecuencias memorizables de letras es 8, por “benjamin”, y la longitud del número de subsecuencias memorizables es 0. La longitud total de la dirección de correo electrónico es 16. Así que la tasa memorizable total es 8/16.

m_NameConfidenceLevel. Esta característica representa el nivel de confianza de que la dirección de correo electrónico contenga un nombre. El nivel de confianza puede ser un valor entre [0, 1]. En general, una dirección de correo electrónico generada por un humano puede contener un nombre para indicar un usuario de la dirección de correo electrónico. Para el ejemplo de la dirección de correo electrónico nicholas312@gmail.com, su m_NameConfidenceLevel es alto (aproximadamente 1), porque tiene la subsecuencia “nicholas”, que es un nombre con una longitud de 8 letras. Esta característica puede considerar también si la primera letra del nombre es mayúscula, tal como “N” aquí y una posición del nombre, y puede considerar si el nombre en la información adicional 106 o sus variaciones aparece en la dirección de correo electrónico y qué probabilidad existe de que ese nombre sea un nombre válido. Si el nombre se sitúa al inicio o al final de la dirección de correo electrónico, el valor sería mayor. Para el ejemplo de la dirección de correo electrónico joshfguufsd@gmail.com, su m_NameConfidenceLevel

no podría ser tan alto como el de nicholas312@gmail.com, dado que tiene una probabilidad mayor de un deletreo erróneo del nombre en la dirección de correo electrónico.

m_DomainNameReliability. Esta característica representa la puntuación de fiabilidad (tal como entre 0 y 1) del dominio de la dirección de correo electrónico. A un dominio en la lista blanca (tal como Microsoft.com, ibm.com) se le asigna una alta puntuación de fiabilidad, tal como 1, y a un dominio en la lista negra (tal como “m5n.com”) se le asigna una baja puntuación de fiabilidad, tal como 0. Para un dominio que no está en ninguna lista, la puntuación de fiabilidad se calcula como sigue: existen dos listas adicionales de dominios guardadas, por ejemplo, mediante los datos etiquetados y la determinación de las nuevas direcciones de correo electrónico entrantes. Una lista de dominios benignos enumera una pluralidad de dominios y una cuenta C_B de direcciones de correo electrónico distintas para cada dominio respectivo que está asociado con cuentas benignas. Una lista de dominios maliciosos enumera una pluralidad de dominios y una cuenta C_M de direcciones de correo electrónico distintas para cada dominio respectivo que está asociado con cuentas maliciosas. La puntuación de fiabilidad se puede obtener mediante la fórmula $(C_B + C) / (C_B + C_M + C)$, donde C es un parámetro para que la puntuación sea aproximadamente 0,5 para C_B y C_M bajos. En otras palabras, cuando las muestras de los C_B y C_M no son suficientemente grandes la puntuación de fiabilidad indica neutral. En otro ejemplo, la puntuación de fiabilidad puede estar en un rango menor de (0, 1) tal como (0,3, 0,7) para la puntuación e fiabilidad determinada por la lista de dominios benignas y la lista de dominios maliciosos para que sus puntuaciones de fiabilidad sean diferentes de los dominios de la lista blanca o la lista negra.

Cuando se consideran las características, el sistema informático 116 puede utilizar asimismo una o más reglas de tratamiento de ciertos números o letras como otros números o letras de acuerdo con convenios y abreviaturas. La siguiente es una lista de algunas reglas de conversión de ejemplo.

Por ejemplo, el número “0” se puede tratar como la letra “o” y viceversa. El número “2” se puede tratar como los caracteres “two”. El número “4” se puede tratar como los caracteres “for”. El número “1” se puede tratar como “i” o “1”. El número 3 se puede tratar como la letra “e”. El número “5” se puede tratar como la letra “s”. La letra “y” se puede tratar como “i”. El número “7” se puede tratar como los caracteres “seven”, La secuencia “2k” se puede tratar como “2000”. La letra “c” se puede tratar como los caracteres “see”. La secuencia “im” se puede tratar como “iam”. La secuencia “ezy” se puede tratar como “ez” o “easy”. La secuencia “biz” se puede tratar como “busy” o “business”. La letra “f” o “f4” se puede tratar como “for”. La letra “x” se puede tratar como el símbolo de multiplicación “*”. La secuencia “2b” se puede tratar como “tobe”. La secuencia “nite” se puede tratar como “night”. La secuencia “b4” se puede tratar como “before”. La letra “u” se puede tratar como “you”. El número “8” se puede tratar como “ate”.

Como ejemplo adicional, la frase “xxxst” se puede tratar como calle “xxx”, y la frase “xxxav” se puede tratar como avenida “xxx”. La “xxx” aquí funciona como marcador de posición, y se refiere a cualquier carácter. Los estándares “007” de una famosa película se pueden tratar como un término con significado. La secuencia “xx.xx”, “xx-xx” y “xx_xx” se pueden tratar como números con significado relacionados por “.”, “-“, o “_”.

Conclusión

Aunque el asunto central se ha descrito en un lenguaje específico para las características estructurales y/o actos metodológicos, se debe entender que el asunto central definido en las reivindicaciones adjuntas no está necesariamente limitado a las características o actos específicos descritos. Por el contrario, las características y actos específicos se describen como formas de implementación de ejemplo de las reivindicaciones.

REIVINDICACIONES

1. Un método implementado por un ordenador, que comprende:

recibir (202; 402) información asociada con una cuenta, incluyendo la información una dirección de correo electrónico:
extraer (204; 404) una o más características de la información asociada con la cuenta, en la que al menos una de las una o más características se basa en la capacidad de ser memorizada de la dirección de correo electrónico, estando referida la capacidad de ser memorizada a un patrón que incluye simetría, anti-simetría o caracteres separados uniformemente en la dirección de correo electrónico; y
determinar (206; 408) un nivel de confianza basado al menos parcialmente en las características extraídas.

2. El método de acuerdo con la reivindicación 1, que comprende, además:

determinar que la cuenta es benigna si el nivel de confianza determinado es mayor que un primer umbral predeterminado; y/o
determinar que la cuenta es maliciosa si el nivel de confianza determinado es menor que el primer umbral predeterminado o un segundo umbral predeterminado que es diferente del primer umbral predeterminado.

3. El método de acuerdo con la reivindicación 1, en el que la al menos una de las una o más características comprende una o más características relativas a secuencias con significado en la dirección de correo electrónico.

4. El método de acuerdo con la reivindicación 1, en el que la al menos una de las una o más características comprende una o más características relativas a secuencias impronunciables en la dirección de correo electrónico.

5. El método de acuerdo con la reivindicación 1, en el que al menos una de las una o más características está basada en un dominio de la dirección de correo electrónico.

6. El método de acuerdo con la reivindicación 1, en el que:

la información comprende además información adicional asociada con la cuenta y/o la dirección de correo electrónico, incluyendo la información adicional un nombre, un número de teléfono, una dirección de IP de una fuente de la solicitud y/o un domicilio asociado con la cuenta y/o la dirección de correo electrónico; y
al menos una de las una o más características está basada en la información adicional de la cuenta.

7. El método de acuerdo con la reivindicación 1, en el que determinar (206; 408) el nivel de confianza de la cuenta comprende:

analizar una pluralidad de cuentas etiquetadas procedentes de una o más fuentes, indicando cada una de la pluralidad de cuentas etiquetadas que una cuenta etiquetada respectiva es maliciosa o 5;
determinar una o más características extraídas de la pluralidad de cuentas que distinguen una cuenta etiquetada respectiva que es maliciosa y una cuenta etiquetada respectiva que es benigna;
aplicar uno o más métodos de aprendizaje de máquinas para construir un modelo de clasificación basado en las una o más características obtenidas; y
utilizar el modelo de clasificación para calcular una puntuación del nivel de confianza de la cuenta.

8. Un sistema (116) que comprende:

una memoria (122) que almacena uno o más módulos;
uno o más procesadores acoplados operablemente a la memoria para ejecutar los uno o más módulos, incluyendo los uno o más módulos:

un módulo de recepción (124) para recibir información asociada con una cuenta, incluyendo la información una dirección de correo electrónico; y
un módulo de determinación (130) para extraer una o más características de la información asociada con la cuenta, en el que al menos una de las una o más características está basada en la capacidad de ser memorizada de la dirección de correo electrónico,
estando la capacidad de ser memorizada relacionada con un patrón que incluye simetría, anti-simetría, o caracteres separados uniformemente en la dirección de correo electrónico;

en el que el módulo de determinación (130) utiliza un modelo de clasificación (128) para determinar un nivel de confianza de la cuenta basado, al menos parcialmente, en las características extraídas.

9. Uno o más medios de almacenamiento legibles por ordenador que almacenan instrucciones ejecutables por un ordenador para hacer que el sistema lleve a cabo el método de cualquiera de las reivindicaciones 1 – 7.

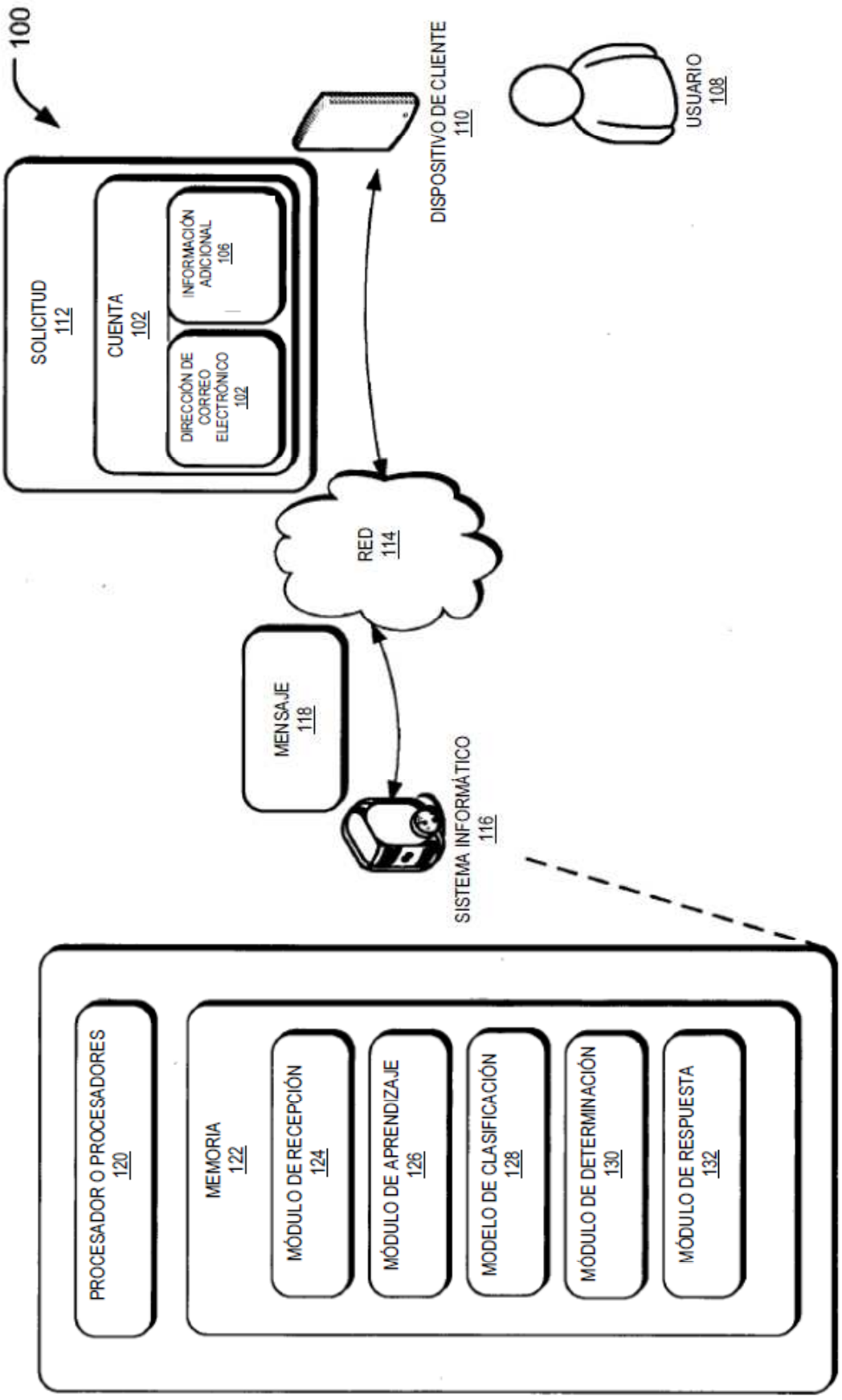
100
SOLICITUD
112
CUENTA
102
DIRECCIÓN DE CORREO ELECTRÓNICO
102
INFORMACIÓN ADICIONAL
106
DISPOSITIVO DE CLIENTE
110
USUARIO
108
RED
114
MENSAJE
118
SISTEMA INFORMÁTICO
116
PROCESADOR O PROCESADORES
120
MEMORIA
122
MÓDULO DE RECEPCIÓN
124
MÓDULO DE APRENDIZAJE
126
MODELO DE CLASIFICACIÓN
128
MÓDULO DE DETERMINACIÓN
130
MÓDULO DE RESPUESTA
132

FIG. 1

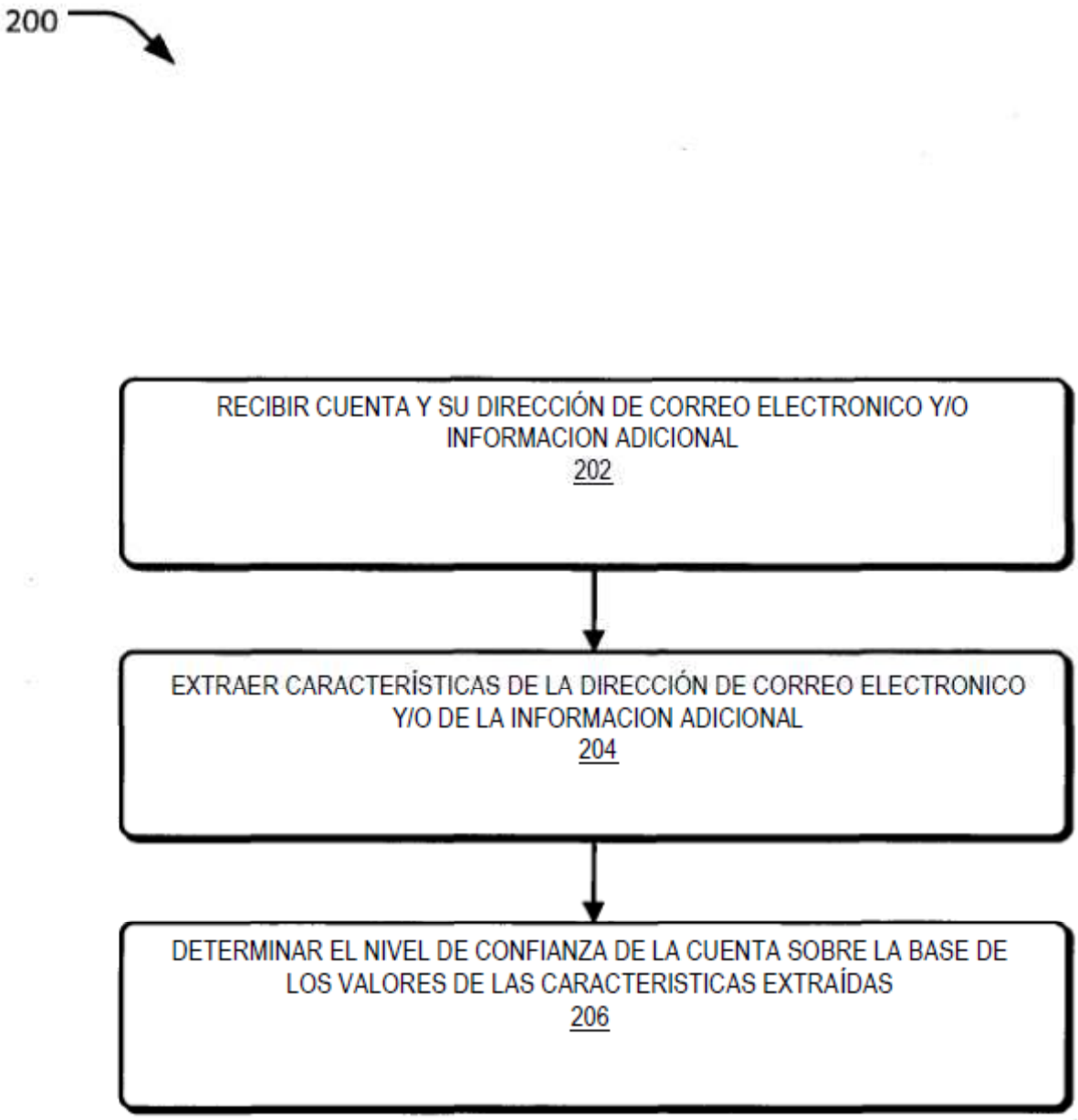
200
RECIBIR CUENTA Y SU DIRECCIÓN DE CORREO ELECTRONICO Y/O INFORMACION ADICIONAL
202
EXTRAER CARACTERÍSTICAS DE LA DIRECCIÓN DE CORREO ELECTRONICO Y/O DE LA INFORMACION ADICIONAL
204
DETERMINAR EL NIVEL DE CONFIANZA DE LA CUENTA SOBRE LA BASE DE LOS VALORES DE LAS CARACTERISTICAS EXTRAÍDAS
206

FIG. 2

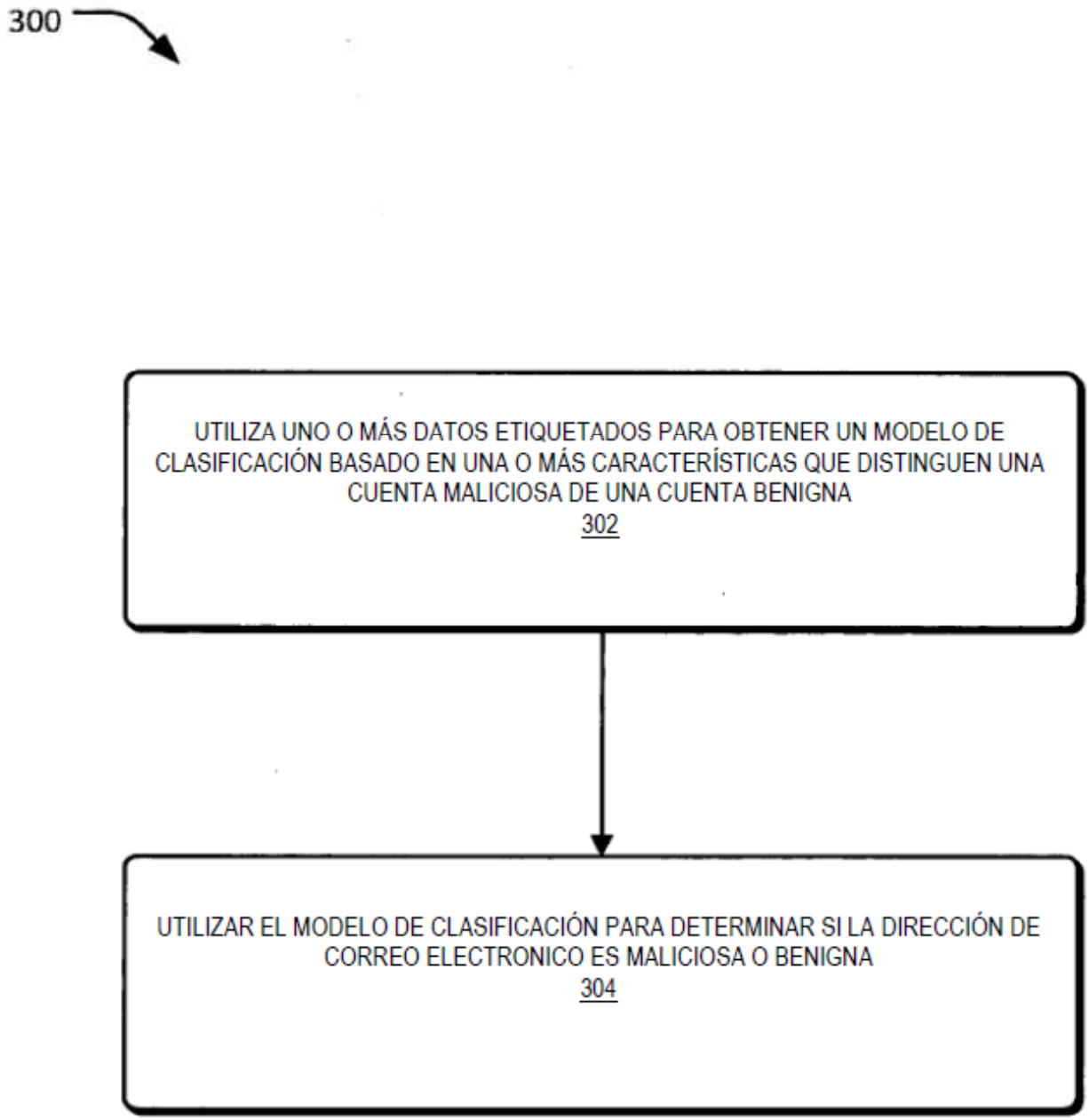
300
UTILIZA UNO O MÁS DATOS ETIQUETADOS PARA OBTENER UN MODELO DE CLASIFICACIÓN BASADO EN UNA O MÁS CARACTERÍSTICAS QUE DISTINGUEN UNA CUENTA MALICIOSA DE UNA CUENTA BENIGNA
302
UTILIZAR EL MODELO DE CLASIFICACIÓN PARA DETERMINAR SI LA DIRECCIÓN DE CORREO ELECTRONICO ES MALICIOSA O BENIGNA
304

FIG. 3

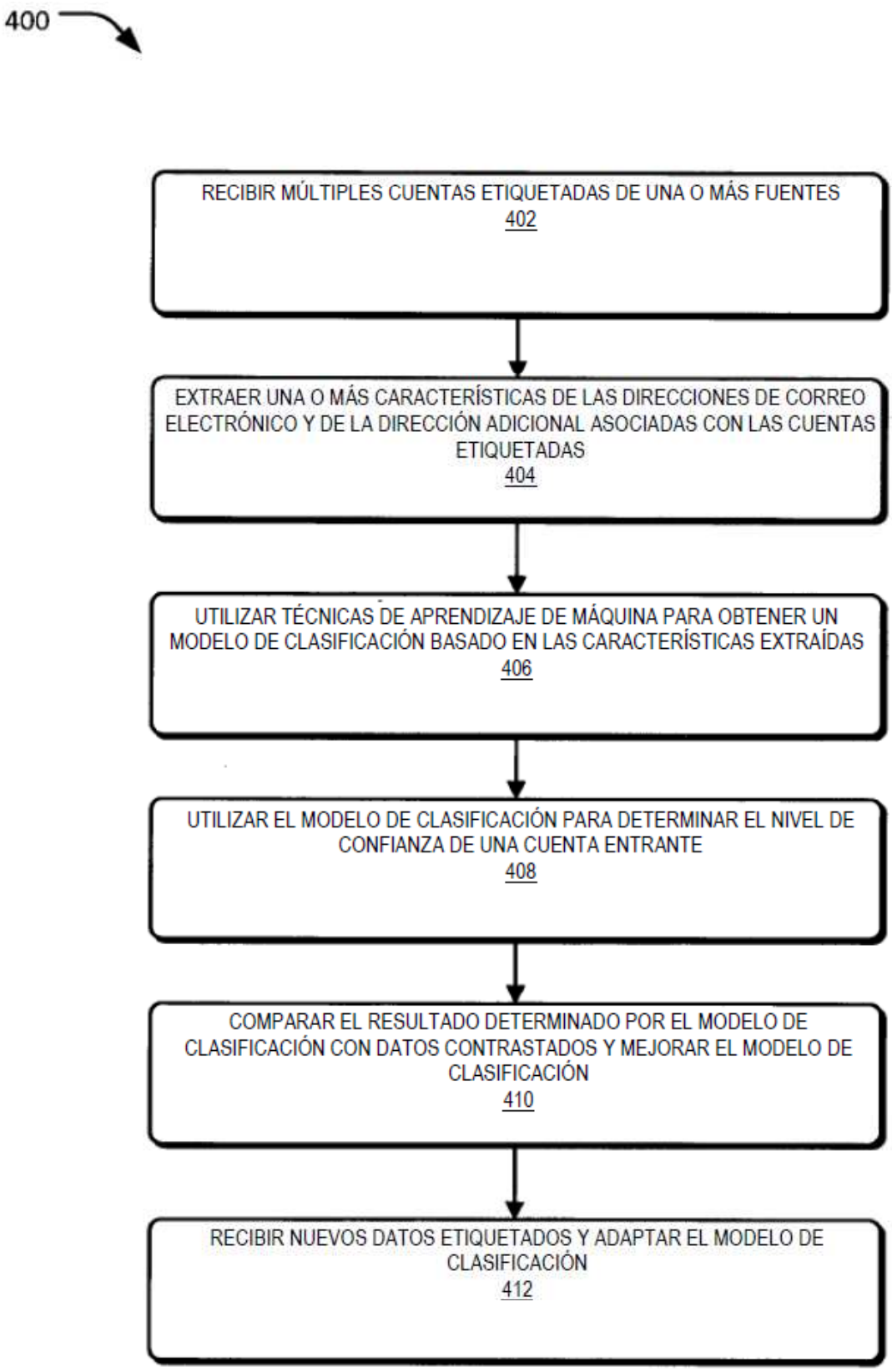
400
RECIBIR MÚLTIPLES CUENTAS ETIQUETADAS DE UNA O MÁS FUENTES
402
EXTRAER UNA O MÁS CARACTERÍSTICAS DE LAS DIRECCIONES DE CORREO ELECTRÓNICO Y DE LA DIRECCIÓN ADICIONAL ASOCIADAS CON LAS CUENTAS ETIQUETADAS
404
UTILIZAR TÉCNICAS DE APRENDIZAJE DE MÁQUINA PARA OBTENER UN MODELO DE CLASIFICACIÓN BASADO EN LAS CARACTERÍSTICAS EXTRAÍDAS
406
UTILIZAR EL MODELO DE CLASIFICACIÓN PARA DETERMINAR EL NIVEL DE CONFIANZA DE UNA CUENTA ENTRANTE
408
COMPARAR EL RESULTADO DETERMINADO POR EL MODELO DE CLASIFICACIÓN CON DATOS CONTRASTADOS Y MEJORAR EL MODELO DE CLASIFICACIÓN
410
RECIBIR NUEVOS DATOS ETIQUETADOS Y ADAPTAR EL MODELO DE CLASIFICACIÓN
412

FIG. 4