

SISTEMAS Y MÉTODOS RELACIONADOS CON ANÁLISIS PREDICTIVOS
QUE USAN LA REPRESENTACIÓN MULTIDIMENSIONAL DE EVENTOS EN
TRAYECTOS DE CLIENTES

5 **REFERENCIA CRUZADA A SOLICITUD RELACIONADA**

Esta solicitud reivindica el beneficio de la solicitud provisional de patente de los EE. UU. núm. 63/433536, titulada “Systems and Methods Relating to
10 Predictive Analytics Using Multidimensional Event Representation in Customer Journeys”, presentada en la Oficina de Patentes y Marcas de los EE. UU. el 19 de diciembre de 2022. Esta solicitud también reivindica el beneficio de la solicitud de patente de los EE. UU. núm.
15 18/545.105, también titulada “Systems and Methods Relating to Predictive Analytics Using Multidimensional Event Representation in Customer Journeys”, presentada concurrentemente en la Oficina de Patentes y Marcas de los EE. UU. el 19 de diciembre de 2023.

20

ANTECEDENTES

La presente invención se relaciona generalmente con los servicios de relaciones con los clientes y la
25 gestión de las relaciones con los clientes a través de centros de contacto y sistemas basados en la nube

asociados, e incluye la prestación de asistencia al cliente a través de centros de contacto y opciones de servicio basadas en Internet. Más particularmente, pero no a manera de limitación, la presente invención se relaciona con la interacción de sitios web, aplicaciones móviles, análisis y centros de contacto mediante el uso de tecnologías tales como análisis predictivo y/o aprendizaje automático, que incluye el uso de representar los trayectos de clientes como una secuencia de eventos multidimensionales para mejorar el rendimiento.

BREVE DESCRIPCIÓN DE LA INVENCION

La presente invención incluye un método implementado por computadora las etapas de: generar, a través de un proceso de datos de entrenamiento, muestras de datos de entrenamiento a partir de muestras de datos de trayecto respectivas, cada una de las muestras de datos de trayecto incluye el trayecto de un cliente representado por datos que describen una secuencia de eventos, para cada uno de los eventos, valores asociados con atributos de evento respectivos de una lista de atributos de evento, y un resultado de trayecto; y entrenar un modelo de aprendizaje automático usando las muestras de datos de entrenamiento generadas. Una entrada del modelo de aprendizaje automático, para cada muestra

de datos de entrenamiento, incluye una secuencia de las incrustaciones vectoriales generadas a través del proceso de datos de entrenamiento de la secuencia de eventos; y una salida del modelo de aprendizaje automático, para

5 cada muestra de datos de entrenamiento, incluye el resultado de trayecto asociado. El proceso de datos de entrenamiento incluye generar una incrustación vectorial para cada uno de los eventos incluidos dentro de una determinada muestra de datos de trayecto que captura el

10 valor para cada uno de los atributos de evento al: dividir los atributos de evento de la lista de atributos de evento en un grupo de cardinalidad baja y un grupo de cardinalidad alta, en donde la división se realiza de acuerdo con si los valores incluidos dentro de las

15 muestras de datos de entrenamiento para un determinado atributo de evento tiene una cardinalidad por encima o por debajo de un umbral de cardinalidad predefinido; para cada uno de los grupos de cardinalidad baja, codificar categóricamente los valores incluidos dentro del grupo de

20 cardinalidad baja de acuerdo con un número total de valores únicos que aparecen en este; para cada uno de los grupos de cardinalidad alta: agrupar los valores incluidos dentro del grupo de cardinalidad alta para crear una pluralidad de grupos de agrupamiento, y

25 codificar categóricamente los valores incluidos dentro del grupo de cardinalidad alta de acuerdo con la

pluralidad de grupos de agrupamientos en los que reside el valor; y tener en cuenta que cada una de las muestras de datos de entrenamiento incluye: la secuencia de las incrustaciones vectoriales generadas a partir de la
5 secuencia de eventos incluidos en una de las muestras de trayecto asociadas, y el resultado de trayecto de la muestra de trayecto asociada.

Estas y otras características de la presente solicitud se harán más evidentes tras la revisión de la
10 siguiente descripción detallada de las modalidades de ejemplo cuando se toman en conjunto con las figuras y las reivindicaciones adjuntas.

BREVE DESCRIPCIÓN DE LAS FIGURAS

15

Una comprensión más completa de la presente invención será más fácilmente evidente a medida que la invención se entienda mejor por referencia a la siguiente descripción detallada cuando se considera en conjunto con
20 las figuras adjuntas, en las que los símbolos de referencia similares indican componentes similares, en donde:

la Figura 1 representa un diagrama de bloques esquemático de un dispositivo informático de acuerdo con
25 modalidades ilustrativas de la presente invención y/o con

los cuales se pueden habilitar o practicar modalidades ilustrativas de la presente invención;

la Figura 2 representa un diagrama de bloques esquemático de una infraestructura de comunicaciones o
5 centro de contacto de acuerdo con modalidades ilustrativas de la presente invención y/o con los cuales se pueden habilitar o practicar modalidades ilustrativas de la presente invención;

la Figura 3 es un diagrama de flujo simplificado
10 que demuestra la funcionalidad de un modelo de aprendizaje automático de acuerdo con modalidades de la presente invención;

la Figura 4 es una representación esquemática de un modelo de aprendizaje automático de acuerdo con la
15 operación ilustrativa de las modalidades de la presente invención; y

la Figura 5 es un método de acuerdo con una modalidad de la presente invención para generar datos de entrenamiento para entrenar un modelo de aprendizaje
20 automático.

DESCRIPCIÓN DETALLADA

Con el fin de promover una comprensión de los principios de la invención, ahora se hará referencia a las modalidades de ejemplo ilustradas en las figuras y se usará un lenguaje específico para describirlas. Sin embargo, será evidente para un experto en la técnica que el material detallado proporcionado en los ejemplos puede no ser necesario para practicar la presente invención. En otros casos, los materiales o métodos bien conocidos no se han descrito en detalle para evitar oscurecer la presente invención. Además, se contempla la modificación adicional de los ejemplos proporcionados o la aplicación de los principios de la invención, como se presenta en la presente, como se les ocurriría normalmente a aquellos expertos en la técnica. Las características, estructuras o funcionalidades particulares pueden combinarse en cualquier combinación y/o subcombinaciones adecuadas en una o más modalidades o ejemplos. Los expertos en la técnica reconocerán que diversas modalidades pueden implementarse por computadora usando muchos tipos diferentes de equipos de procesamiento de datos, con modalidades que se implementan como un aparato, método, o producto de programa informático. Las modalidades ilustrativas, por lo tanto, pueden tomar la forma de una

modalidad de hardware, una modalidad de software, o una combinación de las estas.

Dispositivo informático

5 La presente invención puede implementarse por computadora usando formas diferentes de equipos de procesamiento de datos, por ejemplo, microprocesadores digitales y memoria asociada, que ejecutan programas de software adecuados. Como antecedente, la Figura 1 ilustra
10 un diagrama de bloques esquemático de un dispositivo informático ilustrativo 100 de acuerdo con modalidades de la presente invención y/o con cuales de esas modalidades pueden habilitarse o practicarse.

 El dispositivo informático 100, por ejemplo,
15 puede implementarse a través de un firmware (p. ej., un circuito integrado de aplicación específica), hardware, o una combinación de software, firmware, y hardware. Cada uno de los servidores, controladores, interruptores, puertas de enlace, motores, y/o módulos en las siguientes
20 figuras (que colectivamente pueden denominarse servidores o módulos) pueden implementarse a través de uno o más de los dispositivos informáticos 100. Por ejemplo, los diversos servidores pueden ser un proceso que se ejecuta en uno o más procesadores de uno o más dispositivos
25 informáticos 100, que pueden ejecutar instrucciones de programas informáticos e interactuar con otros módulos

del sistema para realizar las diversas funcionalidades descritas en la presente. A menos que se limite específicamente de cualquier otra manera, la funcionalidad descrita en relación con una pluralidad de dispositivos informáticos puede integrarse en un único dispositivo informático, o las diversas funcionalidades descritas en relación con un único dispositivo informático pueden distribuirse a través de varios dispositivos informáticos. Además, en relación con los sistemas informáticos descritos en las siguientes figuras, tal como, por ejemplo, el centro de contacto 200 de la Figura 2: los diversos servidores y dispositivos informáticos de estos se pueden ubicar en dispositivos informáticos locales 100 (p. ej., en el sitio en la misma ubicación física que los agentes del centro de contacto), dispositivos informáticos remotos 100 (p. ej., fuera del sitio o en un entorno informático en la nube, por ejemplo, en un centro de datos remoto conectado al centro de contacto a través de una red), o alguna combinación de estos. Se puede acceder a y proporcionar la funcionalidad proporcionada por servidores que se ubican en dispositivos informáticos fuera del sitio a través de una red privada virtual (VPN, por sus siglas en inglés), como si tales servidores estuvieran en el sitio, o se puede proporcionar la funcionalidad usando un software como servicio (SaaS, por sus siglas en inglés) al que se

accede a través de Internet usando diversos protocolos, tales como el intercambio de datos a través del lenguaje de marcado extensible (XML, por sus siglas en inglés), JSON, y lo similar.

5 Como se muestra en el ejemplo ilustrado, el dispositivo informático 100 puede incluir una unidad central de procesamiento (CPU, por sus siglas en inglés) o procesador 105, y una memoria principal 110. El dispositivo informático 100 también puede incluir un
10 dispositivo de almacenamiento 115, una interfaz de medios removible 120, una interfaz de red 125, un controlador de E/S 130, y uno más dispositivos de entrada/salida (E/S) 135, que como se representa puede incluir un, dispositivo de visualización 135A, teclado 135B, y dispositivo de
15 señalización 135C. El dispositivo informático 100 puede, además, incluir elementos adicionales, tales como un puerto de memoria 140, un puente 145, puertos de E/S, uno o más dispositivos de entrada/salida adicionales 135D, 135E, 135F, y una memoria caché 150 en comunicación con
20 el procesador 105.

El procesador 105 puede ser cualquier sistema de circuitos lógico que responda y procese instrucciones traídas de la memoria principal 110. Por ejemplo, el procesador 105 puede implementarse por un circuito
25 integrado, p.ej., un microprocesador, microcontrolador, o unidad de procesamiento de gráficos, o en una matriz de

puertas programables en campo o circuito integrado de aplicación específica. Como se ilustra, el procesador 105 puede comunicarse directamente con la memoria caché 150 por medio de un bus secundario o bus trasero. La memoria principal 110 puede ser uno o más chips de memoria capaces de almacenar datos y permitir que los datos almacenados sean accedidos por la unidad central de procesamiento 105. El dispositivo de almacenamiento 115 puede proporcionar almacenamiento para un sistema operativo, que controla tareas de planificación y acceso a recursos del sistema, y otro software. A menos que se limite de cualquier otra manera, el dispositivo informático 100 puede incluir un sistema operativo y un software capaz de realizar la funcionalidad descrita en la presente.

Como se representa en el ejemplo ilustrado, el dispositivo informático 100 puede incluir una amplia variedad de dispositivos de E/S 135, uno o más de los cuales pueden conectarse mediante el controlador de E/S 130. Los dispositivos de entrada, por ejemplo, pueden incluir un teclado 135B y un dispositivo señalador 135C, p. ej., un mouse o un lápiz óptico. Los dispositivos de salida, por ejemplo, pueden incluir dispositivos de visualización de video, altavoces, e impresoras. Más generalmente, los dispositivos de E/S 135 pueden incluir cualquier dispositivo convencional para realizar la funcionalidad descrita en la presente.

A menos que se limite de cualquier otra manera, el dispositivo informático 100 puede ser cualquier estación de trabajo, computadora de escritorio, computadora portátil o notebook, máquina de servidor, 5 máquina virtualizada, teléfono móvil o inteligente, dispositivo portátil de telecomunicaciones, dispositivo de reproducción multimedia o cualquier otro tipo de dispositivo informático, de telecomunicaciones o multimedia, sin limitación, capaz de realizar las 10 operaciones y funcionalidades descritas en la presente. El dispositivo informático 100 puede incluir una pluralidad de tales dispositivos conectados por una red o conectados a otros sistemas y recursos por medio de una red. A menos que se limite de cualquier otra manera, el dispositivo 15 informático 100 puede comunicarse con otros dispositivos informáticos 100 a través de cualquier tipo de red usando cualquier protocolo de comunicación convencional.

Centro de contacto

20 Con referencia ahora a la Figura 2, se muestra una infraestructura de comunicaciones o un sistema de centro de contacto (o simplemente “centro de contacto”) 200 de acuerdo con modalidades ilustrativas de la presente invención y/o con qué modalidades ilustrativas de la 25 presente invención pueden habilitarse o practicarse. Como antecedente, los proveedores de servicio al cliente

generalmente ofrecen muchos tipos de servicios a través de centros de contacto. Tales centros de contacto pueden contar con empleados o agentes de servicio al cliente (o simplemente “agentes”), con los agentes que sirven como
5 interfaz entre una compañía, empresa, agencia gubernamental, u organización (de aquí en adelante denominada indistintamente como una “organización” o “empresa”) y personas, tales como usuarios, individuos, o clientes (de aquí en adelante denominados indistintamente
10 como “individuos” o “clientes”). Por ejemplo, los agentes en un centro de contacto pueden ayudar a los clientes a realizar decisiones de compra, recibir órdenes, o resolver problemas con productos o servicios ya recibidos. Dentro de un centro de contacto, tales interacciones entre los
15 agentes y clientes pueden realizarse a través de una variedad de canales de comunicación, tales como, por ejemplo, a través de voz (p. ej., llamadas telefónicas o llamadas de voz sobre IP o llamadas VoIP), video (p. ej., videoconferencias), texto (p. ej., correos electrónicos y
20 chat de texto), uso compartido de pantalla, navegación conjunta, y lo similar.

Operacionalmente, los centros de contacto generalmente se esfuerzan por proporcionar servicios de calidad a los clientes mientras minimizan los costos. Por
25 ejemplo, una forma de operar de un centro de contacto es atender cada interacción del cliente con un agente vivo.

Si bien este enfoque puede tener una buena puntuación en términos de calidad del servicio, probablemente también sería prohibitivamente costoso debido al alto costo del trabajo del agente. Debido a esto, la mayoría de los centros de contacto utiliza procesos automatizados en lugar de agentes en vivo, tal como sistemas de respuesta de voz interactiva (IVR, por sus siglas en inglés), sistemas de respuesta de medios interactivos (IMR, por sus siglas en inglés), robots de internet o “bots”, módulos de chat automatizado o “chatbots”, y los similar.

Con referencia específicamente a la Figura 2, un proveedor de servicio al cliente puede usar el centro de contacto 200 para proporcionar diversos tipos de servicios a clientes. Por ejemplo, el centro de contacto 200 se puede usar para participar y gestionar las interacciones en las que los procesos automatizados (o bots) o agentes humanos se comunican con los clientes. El centro de contacto 200 puede ser una instalación interna de un negocio o empresa para realizar las funciones de ventas y servicio al cliente en relación con los productos y servicios disponibles a través de la empresa. En otro aspecto, un proveedor de servicios que se contrata para proporcionar servicios de relaciones con los clientes para otra organización puede operar el sistema de centro de contacto 200. Además, el centro de contacto 200 puede desplegarse en equipos especializados para la empresa o proveedor de servicios de

terceros, y/o desplegarse en un entorno informático remoto tal como, por ejemplo, un entorno privado o público de nube con infraestructura para soportar múltiples centros de contacto para múltiples empresas. El centro de contacto 200
5 puede incluir programas o aplicaciones de software, que pueden ejecutarse en instalaciones o remotamente o alguna combinación de estos. Debe apreciarse además que los diversos componentes del centro de contacto 200 pueden estar distribuidos a través de diversas ubicaciones
10 geográficas.

A menos que se limite específicamente de cualquier otra manera, cualquiera de los elementos informáticos de la presente invención puede implementarse en entornos informáticos en la nube o basados en la nube.
15 Como se usa en la presente, la “informática en la nube”, o simplemente la “nube”, se define como un modelo para permitir el acceso de red ubicuo, conveniente, y bajo solicitud a un grupo compartido de recursos informáticos configurables (p. ej., redes, servidores, almacenamiento,
20 aplicaciones, y servicios) que pueden aprovisionarse rápidamente a través de la virtualización y liberarse con un mínimo esfuerzo de gestión o interacción con el proveedor de servicios, y por consiguiente después escalar. La computación en la nube puede componerse de
25 diversas características (p. ej., autoservicio bajo solicitud, amplio acceso a la red, agrupación de recursos,

elasticidad rápida, servicio medido, etc.), modelos de servicio (p. ej., software como servicio (“SaaS”, por sus siglas en inglés), plataforma como servicio (“PaaS”, por sus siglas en inglés), infraestructura como servicio (“IaaS”, por sus siglas en inglés), y modelos de despliegue (p. ej., nube privada, nube comunitaria, nube pública, nube híbrida, etc.). Conocido frecuentemente como una “arquitectura sin servidor”, un modelo de ejecución en la nube generalmente incluye un proveedor de servicios que administra dinámicamente una asignación y aprovisionamiento de servidores remotos para lograr una funcionalidad deseada.

De acuerdo con el ejemplo ilustrado de la Figura 2, los componentes o módulos del centro de contacto 200 pueden incluir: una pluralidad de dispositivos de clientes 205; red de comunicaciones (o simplemente “red”) 210; puerta de enlace de conmutador/medio 212; controlador de llamadas 214; servidor de respuesta de medios interactivos (IMR, por sus siglas en inglés) 216; servidor de enrutamiento 218; dispositivo de almacenamiento 220; servidor de estadísticas 226; pluralidad de dispositivos de agentes 230 que tienen cada uno un contenedor de trabajo 232; servidor multimedia/de redes sociales 234; servidor de gestión de conocimientos 236 acoplado a un sistema de conocimientos 238; servidor de chat 240; servidores web 242; servidor de

interacciones 244; servidor de contacto universal (o “UCS”) 246; servidor de reportes 248; servidor de servicios de medios 249; y un módulo de análisis 250; Se debe entender que cualquiera de los componentes, módulos, o servidores implementados por computadora descritos en relación con la Figura 2 o con cualquiera de las siguientes figuras puede implementarse a través de dispositivos informáticos, tal como el dispositivo informático 100 de la Figura 1. Como se verá, el centro de contacto 200 gestiona generalmente recursos (p. ej., personal, computadoras, equipos de telecomunicaciones, etc.) para permitir el suministro de servicios a través del teléfono, correo electrónico, chat, u otros mecanismos de comunicación. Los diversos componentes, módulos, y/o servidores de la Figura 2 (y otras figuras incluidas en la presente) pueden incluir cada uno o más procesadores que ejecutan instrucciones de programas informáticos y que interactúan con otros componentes del sistema para realizar las diversas funcionalidades descritas en la presente. Además, los términos “interacción” y “comunicación” se usan indistintamente, y generalmente se refieren a cualquier interacción en tiempo real y no en tiempo real que usa cualquier canal de comunicación, que incluyen, pero no se limitan a, llamadas telefónicas (llamadas de PSTN o VoIP), correos electrónicos, mensajes de voz, video, chat, pantalla compartida, mensajes de texto, mensajes de redes

sociales, llamadas de WebRTC, etc. El acceso y control de los componentes del sistema de contacto 200 puede verse afectado a través de las interfaces de usuario (UI, por sus siglas en inglés) que pueden generarse en los dispositivos de cliente 205 y/o los dispositivos de agentes 230.

Los clientes que deseen recibir servicios del centro de contacto 200 pueden iniciar comunicaciones entrantes (p. ej., llamadas telefónicas, correos electrónicos, chats, etc.) hacia el centro de contacto 200 a través de un dispositivo de cliente 205. Si bien la Figura 2 muestra dos dispositivos de cliente de este tipo, debe entenderse que cualquier número puede estar presente. Los dispositivos de cliente 205, por ejemplo, pueden ser un dispositivo de comunicación, tal como un teléfono, teléfono inteligente, computadora, tableta, o computadora portátil. De acuerdo con la funcionalidad descrita en la presente, los clientes generalmente pueden usar los dispositivos de cliente 205 para iniciar, gestionar, y conducir comunicaciones con el centro de contacto 200, tales como llamadas telefónicas, correos electrónicos, chats, mensajes de texto, sesiones de navegación web, y otras transacciones multimedia. Las comunicaciones entrantes y salientes desde y hacia los dispositivos de cliente 205 pueden atravesar la red 210, y la naturaleza de la red depende típicamente del tipo de dispositivo de cliente que se usa y la forma de comunicación. Como ejemplo, la red 210 puede incluir una

red de comunicación de servicios telefónicos, celulares, y/o de datos. La red 210 puede ser una red telefónica conmutada privada o pública (PSTN), red de área local (LAN), red de área amplia (WAN) privada, y/o WAN pública tal como la Internet. Además, la red 210 puede incluir una red de operador inalámbrico que incluye una red de acceso múltiple por división de código, una red de sistema global para comunicaciones móviles (GSM, por sus siglas en inglés), o cualquier red/tecnología inalámbrica convencional en la técnica.

El conmutador/puerta de enlace de medios 212 puede acoplarse a la red 210 para recibir y transmitir llamadas telefónicas entre clientes y el centro de contacto 200. El conmutador/puerta de enlace de medios 212 puede incluir un conmutador telefónico o de comunicación configurado para funcionar como un conmutador central para el enrutamiento de agente dentro del centro. El conmutador puede ser un sistema de conmutación de hardware o implementarse mediante software. Por ejemplo, el conmutador 215 puede incluir un distribuidor de llamadas automático, un intercambio de derivación privado (PBX, por sus siglas en inglés), un conmutador de software basado en IP, y/o cualquier otro conmutador con hardware y software especializados configurados para recibir interacciones de Internet y/o interacciones de red telefónica de un cliente, y enrutar esas interacciones a, por ejemplo, uno

de los dispositivos de agente 230. En general, el conmutador/puerta de enlace de medios 212 establece una conexión de voz entre el cliente y el agente al establecer una conexión entre el dispositivo de cliente 205 y el dispositivo de agente 230. El conmutador/puerta de enlace de medios 212 puede acoplarse al controlador de llamadas 214 que, por ejemplo, sirve como un adaptador o interfaz entre el conmutador y los otros componentes de enrutamiento, monitoreo, y manejo de comunicaciones del centro de contacto 200. El controlador de llamada 214 puede configurarse para procesar llamadas PSTN, llamadas VoIP, etc. El controlador de llamada 214 puede incluir software de integración teléfono-computadora (CTI, por sus siglas en inglés) para interactuar con la puerta de enlace de conmutador/medios y otros componentes. El controlador de llamadas 214 puede extraer datos sobre una interacción entrante, tal como el número telefónico, dirección IP, o dirección de correo electrónico del cliente, y después comunicar estos con otros componentes del centro de contacto al procesar la interacción.

El servidor de medios interactivos de respuesta (IMR) 216 permite la funcionalidad de autoayuda o asistente virtual. Específicamente, el servidor de IMR 216 puede ser similar a un servidor de respuesta de voz interactivo (IVR, por sus siglas en inglés), excepto que el servidor de IMR 216 no está restringido a voz y también puede cubrir una

variedad de canales de medios. En un ejemplo que ilustra voz, el servidor de IMR 216 puede configurarse con un guion de IMR para consultar a los clientes sobre sus necesidades. A través de la interacción continua con el servidor de IMR 216, los clientes pueden recibir servicio sin necesidad de hablar con un agente. El servidor de IMR 216 también puede determinar por qué un cliente entra en contacto con el centro de contacto para enrutar la comunicación al recurso adecuado.

10 El servidor de enrutamiento 218 enruta interacciones entrantes. Por ejemplo, una vez que se determina que una comunicación entrante debe ser atendida por un agente humano, la funcionalidad dentro del servidor de enrutamiento 218 puede seleccionar al agente más
15 adecuado y enrutar la comunicación a este. Este tipo de funcionalidad puede denominarse enrutamiento predictivo. Tal agente puede basarse en qué agente disponible es el más adecuado para atender la comunicación. Más específicamente, la selección del agente adecuado puede estar basada en una
20 estrategia o algoritmo de enrutamiento que es implementado por el servidor de enrutamiento 218. Al hacer esto, el servidor de enrutamiento 218 puede consultar datos que son relevantes para la interacción entrante, por ejemplo, datos relacionados con el cliente particular, agentes
25 disponibles, y el tipo de interacción, que, como se describe más abajo, pueden almacenarse en bases de datos

particulares. Una vez que se selecciona el agente, el servidor de enrutamiento 218 puede interactuar con el controlador de llamadas 214 para enrutar (es decir, conectar) la interacción entrante al dispositivo de agente 230 correspondiente. Como parte de esta conexión, puede proporcionarse información sobre el cliente al agente seleccionado por medio de su dispositivo de agente 230, que puede mejorar el servicio que el agente es capaz de prestar.

Se debe apreciar que el centro de contacto 200 puede incluir uno o más dispositivos de almacenamiento masivo, representados generalmente por el dispositivo de almacenamiento 220, para almacenar datos en una o más bases de datos. Por ejemplo, el dispositivo de almacenamiento 220 puede almacenar datos del cliente que se mantienen en una base de datos de cliente 222. Tales datos del cliente pueden incluir perfiles del cliente, información de contacto, acuerdo de nivel de servicio (SLA, por sus siglas en inglés), e historial de interacción (p. ej., detalles de interacciones previas con un cliente particular, que incluyen la naturaleza de las interacciones previas, datos de disposición, tiempo de espera, tiempo de atención, y acciones tomadas por el centro de contacto para resolver problemas del cliente). Como otro ejemplo, el dispositivo de almacenamiento 220 puede almacenar datos de agentes en una base de datos de

agentes 223. Los datos de agente que mantiene el centro de contacto 200 pueden incluir disponibilidad de agente y perfiles de agentes, programas, habilidades, promedio de tiempo de gestión, etc. Como otro ejemplo, el dispositivo de almacenamiento 220 puede almacenar datos de interacción en una base de datos de interacción 224. Los datos de interacción pueden incluir datos relacionados con numerosas interacciones pasadas entre clientes y centros de contacto. Más generalmente, debe entenderse que, a menos que se especifique de cualquier otra manera, el dispositivo de almacenamiento 220 puede configurarse para incluir bases de datos y/o almacenar datos relacionados con cualquiera de los tipos de información descritos en la presente, con esas bases de datos y/o datos accesibles a los otros módulos o servidores del centro de contacto 200 de maneras que faciliten la funcionalidad descrita en la presente. Por ejemplo, los servidores o módulos del centro de contacto 200 pueden consultar tales bases de datos para recuperar datos almacenados en estas o transmitir datos a estas para su almacenamiento.

El servidor estadístico 226 puede configurarse para registrar y añadir datos relacionados con el rendimiento y aspectos operacionales del centro de contacto 200. Tal información puede estar compilada por el servidor estadístico 226 y está disponible para otros servidores y módulos, tales como el servidor de reportes 248, que

después puede producir reportes que se usan para gestionar aspectos operaciones del centro de contacto y ejecutar acciones automatizadas de acuerdo con la funcionalidad descrita en la presente. Estos datos pueden relacionarse
5 con el estado de los recursos del centro de contacto, p. ej., tiempo de espera promedio, tasa de abandono, ocupación del agente, y otros según lo requiera la funcionalidad descrita en la presente.

Los dispositivos de agentes 230 del centro de
10 contacto 200 pueden ser dispositivos de comunicación configurados para interactuar con los diversos componentes y módulos del centro de contacto 200 para facilitar la funcionalidad descrita en la presente. Un dispositivo de agente 230, por ejemplo, puede incluir un
15 teléfono adaptado para llamadas telefónicas regulares o llamadas de VoIP. Un dispositivo de agente 230 puede incluir además un dispositivo informático configurado para comunicarse con los servidores del centro de contacto 200, realizar el procesamiento de datos asociado
20 con las operaciones, e interactuar con los clientes a través de voz, chat, correo electrónico, y otros mecanismos de comunicación multimedia de acuerdo con la funcionalidad descrita en la presente. Si bien solo se muestran dos dispositivos de agentes de este tipo, puede
25 estar presente cualquier número.

El servidor multimedia/de medios sociales 234 puede configurarse para facilitar interacciones de medios (que no sean de voz) con los dispositivos de cliente 205 y/o los servidores 242. Tales interacciones de medios 5 pueden estar relacionadas, por ejemplo, con correo electrónico, mensaje de voz, chat, video, mensajes de texto, web, redes sociales, navegación conjunta, etc. El servidor multimedia/de medios sociales 234 puede tomar la forma de cualquier enrutador IP convencional en la 10 técnica con hardware y software especializados para recibir, procesar, y reenviar eventos y comunicaciones multimedia.

El servidor de gestión de conocimientos 234 puede configurarse para facilitar interacciones entre 15 clientes y el sistema de conocimiento 238. Generalmente, el sistema de conocimiento 238 puede ser un sistema informático capaz de recibir preguntas o consultas y proporcionar respuestas. El sistema de conocimiento 238 puede incluir un sistema informático artificialmente 20 inteligente capaz de responder preguntas que se exponen en el lenguaje natural mediante la recuperación de información de fuentes de información tales como enciclopedias, diccionarios, artículos de periódico, trabajos de literalidad, u otros documentos sometidos al 25 sistema de conocimiento 238 como materiales de referencia, como se conoce en la técnica.

El servidor de chat 240 se puede configurar para realizar, orquestar, y gestionar comunicaciones de chat electrónico con los clientes. Tales comunicaciones de chat pueden ser realizadas por el servidor de chat 240 de tal manera que un cliente se comuniquen con chatbots automatizados, agentes humanos, o ambos. El servidor de chat 240 puede actuar como un servidor de orquestaciones de chat que despacha conversaciones de chat entre chatbots y los agentes humanos disponibles. En tales casos, la lógica de procesamiento del servidor de chat 240 puede estar regida por reglas para aprovechar una distribución inteligente de la carga de trabajo entre los recursos de chat disponibles. El servidor de chat 240 puede además implementar, administrar y facilitar interfaces de usuario (también IU) asociadas con la característica de chat. El servidor de chat 240 puede configurarse para transferir chats dentro de una sola sesión de chat con un cliente particular entre fuentes automatizadas y humanas. El servidor de chat 240 puede acoplarse al servidor de gestión de conocimiento 234 y los sistemas de conocimiento 238 para recibir sugerencias y respuestas a consultas planteadas por clientes durante un chat para que, por ejemplo, puedan proporcionarse enlaces a artículos pertinentes.

Los servidores web 242 proporcionan hosts de sitio para una variedad de sitios de interacción social a

los que se suscriben los clientes, tal como Facebook, Twitter, Instagram, etc. Aunque se ilustran como parte del centro de contacto 200, debe entenderse que los servidores web 242 pueden ser proporcionados por terceros y/o mantenidos de manera remota. Los servidores web 242 también pueden proporcionar páginas web para la empresa u organización que está soportada por el centro de contacto 200. Por ejemplo, los clientes pueden navegar por las páginas web y recibir información sobre los productos y servicios de una empresa en particular. Dentro de tales páginas web empresariales, pueden proporcionarse mecanismos para iniciar una interacción con el centro de contacto 200, por ejemplo, a través de chat con cámara web, voz, o correo electrónico. Un ejemplo de tal mecanismo es un widget, que puede desplegarse en las páginas web o sitios web alojados en los servidores web 242. Como se usa en la presente, un widget se refiere a un componente de interfaz de usuario que realiza una función particular. En algunas implementaciones, un widget incluye un GUI que se superpone a una página web mostrada a un cliente a través de Internet. El widget puede mostrar información, tal como en una ventana o cuadro de texto, o incluir botones u otros controles que permiten al cliente acceder a ciertas funcionalidades, tales como compartir o abrir un archivo o iniciar una comunicación. En algunas implementaciones, un widget incluye un componente de interfaz de usuario que

tiene una porción portátil de código que puede instalarse y ejecutarse dentro de una página web separada sin compilación. Tales widgets pueden incluir interfaces de usuario adicionales y se configuran para acceder a una
5 variedad de recursos locales (p. ej., una información de calendario o de contacto en el dispositivo de cliente) o recursos remotos a través de red (p. ej., mensajería instantánea, correo electrónico, o actualizaciones de redes sociales).

10 El servidor de interacción 244 se configura para gestionar actividades diferibles del centro de contacto y el enrutamiento de estos a agentes humanos para su finalización. Como se usa en la presente, las actividades diferibles incluyen trabajo administrativo que puede
15 realizarse fuera de línea, p. ej., responder a correos electrónicos, asistir a los entrenamientos, y otras actividades que no implican una comunicación en tiempo real con un cliente.

El servidor de contacto universal (UCS, por sus
20 siglas en inglés) 246 puede configurarse para recuperar información almacenada en la base de datos de cliente 222 y/o transmitir información a este para almacenarse en esta. Por ejemplo, el UCS 246 puede utilizarse como parte de la característica de chat para facilitar el
25 mantenimiento de un historial sobre cómo se atendieron los chats con un cliente particular, que después puede

usarse como referencia sobre cómo deben atenderse chats futuros. Más generalmente, el UCS 246 puede configurarse para facilitar el mantenimiento de un historial de preferencias del cliente, tales como canales de medios preferidos y mejores horas para entrar en contacto. Para 5 hacer esto, el UCS 246 puede configurarse para identificar datos pertinentes al historial de interacción para cada cliente tales como datos relacionados con comentarios de agentes, historial de comunicaciones del 10 cliente, y lo similar. Después, cada uno de estos tipos de datos puede almacenarse en la base de datos del cliente 222 o en otros módulos y recuperarse según lo requiera la funcionalidad descrita en la presente.

El servidor de reportes 248 puede configurarse 15 para generar reportes a partir de datos compilados y agregados por el servidor estadístico 1226 u otras fuentes. Tales reportes pueden incluir reportes casi en tiempo real o reportes históricos y referirse al estado de los recursos del centro de contacto y características 20 de rendimiento, tales como, por ejemplo, tiempo de espera promedio, tasa de abandono, y/u ocupación del agente. Los informes se pueden generar automáticamente o en respuesta a una solicitud y usar para administrar el centro de contacto de acuerdo con la funcionalidad descrita en la 25 presente.

El servidor de servicios multimedia 249 proporciona servicios de audio y/o video para soportar características del centro de contacto. De acuerdo con la funcionalidad descrita en la presente, tales

5 características pueden incluir avisos para un sistema de IVR o IMR (p. ej., reproducción de archivos de audio), música en espera, mensaje de voz/grabaciones de una única parte, grabaciones de varias partes (p. ej., de llamadas de audio y/o video), reconocimiento de voz,

10 reconocimiento de multifrecuencia de doble tono (DTMF, por sus siglas en inglés), transcodificación de audio y video, protocolo de transporte seguro en tiempo real (SRTP, por sus siglas en inglés), audio o videoconferencia, análisis de llamadas, detección de

15 palabras clave, etc.

El módulo de análisis 250 se puede configurar para realizar análisis de datos recibidos de una pluralidad de fuentes de datos diferentes, según lo requiera la funcionalidad descrita en la presente. El

20 módulo de análisis 250 también puede generar, actualizar, entrenar y modificar predictores o modelos, tales como el modelo de aprendizaje automático 251 y/o los modelos 253, basados en los datos recopilados. Para lograr esto, el módulo analítico 250 puede tener acceso a los datos

25 almacenados en el dispositivo de almacenamiento 220, que incluyen la base de datos de cliente 222 y la base de

datos de agente 223. El módulo analítico 250 también puede tener acceso a la base de datos de interacción 224, que almacena datos relacionados con interacciones y contenido de interacción (p. ej., audio y transcripciones de las interacciones y eventos detectados en estas), metadatos de interacción (p. ej., identificador de cliente, identificador de agente, medio de interacción, duración de la interacción, hora de inicio y finalización de la interacción, departamento, categorías etiquetadas), y la configuración de la aplicación (p. ej., el trayecto de interacción a través del centro de contacto). El módulo de análisis 250 puede recuperar tales datos del dispositivo de almacenamiento 220 para desarrollar y entrenar algoritmos y modelos. Cabe entender que, aunque el módulo de análisis 250 se representa como parte de un centro de contacto, la funcionalidad descrita en la presente también se puede implementar en sistemas de clientes (o, como también se usa en la presente, en el “lado del cliente” de la interacción) y usar en beneficio de los clientes.

El modelo 251 de aprendizaje automático puede incluir uno o más modelos de aprendizaje automático, que pueden estar basados en redes neuronales. En ciertas modalidades, el modelo de aprendizaje automático 251 se configura como un modelo de aprendizaje profundo, que es un tipo de aprendizaje automático basado en redes neuronales

en el que se usan múltiples capas de procesamiento para extraer progresivamente características de nivel superior a partir de datos. A modo de ejemplo, el modelo 251 de aprendizaje automático puede configurarse para predecir el comportamiento. Tales modelos de comportamiento pueden ser entrenados para predecir el comportamiento de clientes y agentes en una variedad de situaciones, de manera que las interacciones puedan ser adaptadas personalmente a los clientes y manejadas de manera más eficiente por los agentes. Como otro ejemplo, el modelo de aprendizaje automático 251 puede configurarse para predecir aspectos relacionados con la operación y el rendimiento del centro de contacto. En otros casos, por ejemplo, el modelo de aprendizaje automático 251 también puede configurarse para realizar el procesamiento del lenguaje natural y, por ejemplo, proporcionar reconocimiento de intención y lo similar.

El módulo de análisis 250 puede incluir además un sistema de optimización 252. El sistema 252 de optimización puede incluir uno o más modelos 253, que pueden incluir el modelo 251 de aprendizaje automático, y un optimizador 254. El optimizador 254 se puede usar junto con los modelos 253 para minimizar una función de costo sujeta a un conjunto de restricciones, donde la función de costo es una representación matemática de la operación u objetivos deseados. Debido a que los modelos

253 son típicamente no lineales, el optimizador 254 puede ser un optimizador de programación no lineal. Sin embargo, se contempla que el optimizador 254 puede implementarse usando, individualmente o en combinación, una variedad de diferentes tipos de enfoques de optimización, que incluyen, pero no se limitan a, programación lineal, programación cuadrática, programación no lineal de enteros mixtos, programación estocástica, programación global no lineal, algoritmos genéticos, técnicas de partículas/enjambres, y lo similar. El módulo de análisis 250 puede utilizar el sistema de optimización 255 como parte de un proceso de optimización mediante el cual los aspectos del rendimiento y operación del centro de contacto se optimizan o, al menos, mejoran. Esto, por ejemplo, puede incluir aspectos relacionados con la experiencia del cliente, experiencia del agente, enrutamiento de interacción, procesamiento de lenguaje natural, reconocimiento de intención, asignación de recursos del sistema, análisis del sistema, u otra funcionalidad relacionada con procesos automatizados.

Modelos de aprendizaje automático

La Figura 3 ilustra un modelo de aprendizaje automático 300 ilustrativa, que puede incluirse en una o más de las modalidades de la presente invención. El

modelo 300 de aprendizaje automático puede ser un componente, módulo, programa informático, sistema o algoritmo. Como se describe a continuación, algunas modalidades en la presente usan el aprendizaje automático para proporcionar análisis predictivos para la aplicación en un centro de contacto. El modelo de aprendizaje automático 300 se puede usar como modelo para alimentar esas modalidades. El modelo de aprendizaje automático 300 se entrena con muestras de datos de entrenamiento 306, que pueden incluir un objeto de entrada 310 y un valor de salida 312 deseado. Por ejemplo, el objeto 310 de entrada y el valor 312 del objeto deseado pueden ser tensores. Un tensor es una matriz de n dimensiones donde n puede ser cualquiera de 0 (una constante), 1 (una matriz), 2 (una matriz 2D), 3, 4 o más.

El modelo de aprendizaje automático 300 tiene parámetros internos que determinan su límite de decisión y que determinan la salida que produce el modelo de aprendizaje automático 300. Después de cada iteración de entrenamiento, que incluye introducir el objeto de entrada 310 de una muestra de datos de entrenamiento en el modelo de aprendizaje automático 300, la salida 308 real del modelo de aprendizaje automático 300 para el objeto de entrada 310 se compara con el valor de salida 312 deseado. Uno o más parámetros internos 302 del modelo de aprendizaje automático 300 pueden ajustarse de manera que,

al ejecutar el modelo de aprendizaje automático 300 con los nuevos parámetros, la salida 308 producida estará más cerca del valor de salida 312 deseado. Si la salida 308 producida ya era idéntica al valor de salida 312 deseado, entonces los parámetros internos 302 del modelo de aprendizaje automático 300 pueden ajustarse para reforzar y fortalecer aquellos parámetros que causaron la salida correcta y reducir y debilitar los parámetros que tendían a alejarse de la salida correcta.

El resultado del modelo de aprendizaje automático 300 puede ser, por ejemplo, un valor numérico en el caso de regresión o un identificador de una categoría en el caso de clasificador. Un modelo de aprendizaje automático entrenado para realizar regresión puede denominarse modelo de regresión y un modelo de aprendizaje automático entrenado para realizar clasificación puede denominarse clasificador. Los aspectos del objeto de entrada que pueden ser considerados por el modelo de aprendizaje automático 300 al tomar su decisión pueden denominarse características. Después de que se haya entrenado el modelo de aprendizaje automático 300, se puede proporcionar un nuevo objeto de entrada 320 invisible como entrada al modelo 300. El modelo de aprendizaje automático 300 produce entonces una salida que representa un valor objetivo 304 previsto para

el nuevo objeto de entrada 320, basado en sus parámetros internos 302 aprendidos por el entrenamiento.

El modelo de aprendizaje automático 300 puede ser, por ejemplo, una red neuronal, máquina vectorial de soporte (SVM), red bayesiana, regresión logística, clasificación logística, árbol de decisión, clasificador de conjuntos u otro modelo de aprendizaje automático. El modelo de aprendizaje automático 300 puede ser supervisado o no supervisado. En el caso no supervisado, el modelo de aprendizaje automático 300 puede identificar patrones en datos no estructurados 340 sin muestras de datos de entrenamiento 306. Los datos no estructurados 340 son, por ejemplo, datos sin procesar sobre los que se desea realizar procesos de inferencia. Un modelo de aprendizaje automático no supervisado puede generar la salida 342 que incluye datos que identifican estructura o patrones.

La red neuronal puede consistir en una pluralidad de nodos de red neuronal, donde cada nodo incluye valores de entrada, un conjunto de ponderaciones y una función de activación. El nodo de red neuronal puede calcular la función de activación en los valores de entrada para producir un valor de salida. La función de activación puede ser una función no lineal calculada sobre la suma ponderada de los valores de entrada más una constante opcional. En algunas modalidades, la función de

activación es una función logística, sigmoidea o una función tangente hiperbólica. Los nodos de la red neuronal pueden estar conectados entre sí de manera que la salida de un nodo sea la entrada de otro nodo. Además, los nodos de la red neuronal pueden organizarse en capas, cada una de las cuales incluye uno o más nodos. Una capa de entrada puede incluir las entradas a la red neuronal y una capa de salida puede incluir la salida de la red neuronal. Una red neuronal puede entrenarse y actualizar sus parámetros internos, que incluyen los pesos de cada nodo de la red neuronal, usando retropropagación.

En algunas modalidades, se puede usar una red neuronal convolucional (CNN, por sus siglas en inglés). Una red neuronal convolucional es un tipo de red neuronal y modelo de aprendizaje automático. Una red neuronal convolucional puede incluir uno o más filtros convolucionales, también conocidos como kernels, que operan en las salidas de la capa de red neuronal que la precede y producen una salida para ser consumida por la capa de red neuronal posterior a esta. Un filtro convolucional puede tener una ventana en la que opera. La ventana puede ser espacialmente local. Un nodo de la capa precedente puede estar conectado a un nodo en la capa actual si el nodo de la capa precedente está dentro de la ventana. Si no está dentro de la ventana, entonces no está conectado. Una red neuronal convolucional es un tipo

de red neuronal conectada localmente, que es una red neuronal donde los nodos de la red neuronal están conectados a nodos de una capa precedente que se encuentran dentro de un área espacialmente local. Además, una red neuronal convolucional es un tipo de red neuronal escasamente conectada, que es una red neuronal donde la mayoría de los nodos de cada capa oculta están conectados a menos de la mitad de los nodos en la capa posterior. En otras modalidades, se puede usar una red neuronal recurrente (RNN, por sus siglas en inglés). Una red neuronal recurrente es otro tipo de red neuronal y modelo de aprendizaje automático. Una red neuronal recurrente incluye al menos un bucle posterior, donde la salida de al menos un nodo de red neuronal se introduce en un nodo de red neuronal de una capa anterior. La red neuronal recurrente mantiene el estado entre iteraciones, tal como en la forma de tensor. El estado se actualiza en cada iteración, y el tensor de estado se pasa como entrada a la red neuronal recurrente en la nueva iteración. En aún otras modalidades, la red neuronal recurrente es una red neuronal de memoria larga a corto plazo (LSTM, en inglés). En algunas modalidades, la red neuronal recurrente es una red neuronal LSTM bidireccional. Una red neuronal de alimentación hacia adelante es otro tipo de red neuronal y no tiene bucles de retroceso. En algunas modalidades, una red neuronal de alimentación

hacia adelante puede estar densamente conectada, lo que significa que la mayoría de los nodos de la red neuronal en cada capa están conectados a la mayoría de los nodos de la red neuronal en la capa posterior. En algunas 5 modalidades, la red neuronal de alimentación hacia adelante es una red neuronal completamente conectada, donde cada uno de los nodos de la red neuronal está conectado a cada nodo de red neuronal en la capa posterior. Una red neuronal de secuencia gráfica cerrada 10 (GGSNN, por sus siglas en inglés) es un tipo de red neuronal que se puede usar en algunas modalidades. En una GGSNN, los datos de entrada son un gráfico, que comprende nodos y bordes entre los nodos, y la red neuronal genera un gráfico. El gráfico puede ser dirigido o no dirigido. 15 Se realiza una etapa de propagación para calcular representaciones de nodos para cada nodo, donde las representaciones de nodos se pueden basar en características del nodo. Un modelo de salida se asigna desde representaciones de nodos y etiquetas 20 correspondientes a una salida para cada nodo. El modelo de salida se define por nodo y es una función diferenciable que se asigna a una salida. Además, las modalidades pueden incluir redes neuronales de diferentes tipos o del mismo tipo que están enlazadas entre sí en 25 una serie secuencial o paralela de redes neuronales, donde las redes neuronales posteriores aceptan como

entrada la salida de una o más redes neuronales precedentes. La combinación de múltiples redes neuronales puede entrenarse de extremo a extremo usando la retropropagación desde la última red neuronal a través de la primera red neuronal. Como se indicó, el modelo de aprendizaje automático 251 también puede configurarse como un modelo de aprendizaje profundo. El modelo de aprendizaje profundo es un tipo de aprendizaje automático basado en redes neuronales en las que se usan múltiples capas de procesamiento para extraer progresivamente características de nivel superior de los datos. Los modelos de aprendizaje profundo son generalmente más hábiles en el aprendizaje no supervisado.

La Figura 4 ilustra el uso del modelo de aprendizaje automático 300 para realizar inferencias de la entrada 360 que comprende datos relevantes sobre los trayectos de clientes. Por ejemplo, como se analiza a continuación, la entrada 360 puede incluir datos de trayecto del cliente, es decir, una secuencia de incrustaciones vectoriales multidimensionales generadas para representar una secuencia de eventos web que describe el trayecto de un cliente. El modelo de aprendizaje automático 300 realiza entonces inferencias de los datos basado en sus parámetros internos 302 aprendidos a través del entrenamiento. El modelo de aprendizaje automático 300 genera una salida 370 de un

resultado pronosticado. En una modalidad ilustrativa, el modelo de aprendizaje automático 300 se puede configurar de acuerdo con la deseabilidad de algoritmos particulares de aprendizaje automático para lograr la funcionalidad descrita en la presente. Como ejemplo, el modelo de aprendizaje automático 300 puede incluir una o más redes neuronales. Más específicamente, el modelo de aprendizaje automático 300 puede incluir redes neuronales recurrentes (RNN), que generalmente son efectivas para procesar datos secuenciales, tales como texto, audio o datos de series de tiempo. Dichos modelos están diseñados para recordar o “almacenar” información de entradas anteriores, lo que les permite hacer uso del contexto y las dependencias entre pasos de tiempo. Esto los hace útiles para tareas como la traducción de idiomas, el reconocimiento de voz y la previsión de series de tiempo. En algunas modalidades, la RNN puede incluir redes de memoria larga a corto plazo (LSTM) o unidades recurrentes cerradas (GRU). Tanto las LSTM como las GRU están diseñadas para abordar el problema de los “gradientes desaparecidos” en las RNN, que ocurre cuando los gradientes de los pesos en la red se vuelven muy pequeños y la red tiene dificultades para aprender. Las redes LSTM son un tipo de RNN que usa un tipo especial de celda de memoria para almacenar y generar información. Estas celdas de memoria están diseñadas para recordar información durante largos

períodos de tiempo, y lo hacen mediante el uso de un conjunto de “puertas” que controlan el flujo de información dentro y fuera de la celda. Las puertas en una red LSTM están controladas por funciones de activación sigmoide, que dan salida a valores entre 0 y 1. Las puertas permiten que la red almacene u olvide selectivamente información, dependiendo de los valores de las entradas y el estado previo de la celda. Las GRU, por otro lado, son una versión simplificada de las LSTM que usan una única “puerta de actualización” para controlar el flujo de información hacia la celda de memoria, en lugar de las tres puertas que se usan en las LSTM. Esto hace que las GRU sean más fáciles de entrenar y más rápidas de ejecutar que las LSTM, pero es posible que no sean tan efectivas para almacenar y acceder a dependencias de largo plazo. En otras modalidades, el modelo de aprendizaje automático se puede configurar como un modelo de secuencia a secuencia que comprende un primer modelo codificador y un modelo decodificador. El primer codificador puede incluir una RNN, o red neuronal convolucional (CNN), u otro modelo de aprendizaje automático capaz de aceptar la entrada de secuencia. El decodificador puede incluir un RNN, CNN u otro modelo de aprendizaje automático capaz de generar salida de secuencia. El modelo de secuencia a secuencia puede entrenarse en muestras de datos de entrenamiento, en

donde cada muestra de datos de entrenamiento incluye una secuencia de incrustaciones vectoriales que representan una secuencia de eventos de trayectos del cliente y un resultado de los trayectos. Por ejemplo, el modelo de secuencia a secuencia puede entrenarse introduciendo los datos de entrada en el primer modelo de codificador para crear un primer vector de incrustación. El primer vector de incrustación puede introducirse en el modelo descodificador para crear un resultado de salida de un resultado previsto. El resultado de salida puede compararse con el resultado real, y los parámetros del primer codificador y el decodificador pueden ajustarse para reducir la diferencia entre el resultado previsto y el resultado real. Los parámetros pueden ajustarse mediante retropropagación. En una modalidad, el modelo de secuencia a secuencia puede incluir un segundo codificador que toma información adicional relacionada con la prueba fallida/el cambio disruptivo para crear un segundo vector de incrustación. El primer vector de incrustación puede combinarse con el segundo vector de incrustación como entrada al decodificador. Por ejemplo, el primer y segundo vectores de incrustación pueden combinarse usando concatenación, suma, multiplicación u otra función. En otro ejemplo, las características pueden incluir estadísticas calculadas sobre el recuento y el orden de palabras o caracteres en los datos asociados con

características de eventos particulares, por ejemplo, palabras asociadas con la URL de una página. En otras modalidades, el modelo de aprendizaje automático para la salida de un resultado previsto puede ser un modelo no supervisado, tal como un modelo de aprendizaje profundo. El modelo no supervisado no está entrenado y en su lugar identifica sus predicciones basado en la identificación de patrones en los datos. En una modalidad, el modelo de aprendizaje automático no supervisado puede identificar características comunes en el conjunto de datos de entrenamiento.

Representación multidimensional de eventos para análisis predictivo

La derivación de información procesable de los trayectos de clientes ha sido un área crítica de enfoque de investigación en los últimos años. Se entiende que ser capaz de generar una visualización sucinta de tales trayectos es una etapa importante hacia una mejor comprensión de los patrones que se encuentran dentro de los trayectos de clientes. Hacer esto para los eventos que se producen cuando los clientes interactúan con los sitios web, que pueden denominarse “eventos web” o simplemente “eventos”, es sumamente difícil dada la falta de estructura en cuanto a cómo se desarrollan tales interacciones. Como se verá, la presente invención

proporciona una manera de modelar los trayectos de clientes de manera que se identifiquen los eventos hito que se correlacionan con resultados particulares. Los eventos hito pueden usarse entonces para producir visualizaciones instructivas en relación con las secuencias predictivas de tales eventos web encontrados en los trayectos de clientes. Además, a través de un proceso iterativo habilitado por la presente invención, la capacidad predictiva y la precisión de los modelos de trayectos de clientes basados en eventos web pueden mejorarse significativamente, a medida que se identifican y eliminan eventos/parámetros ruidosos. Estos modelos mejorados se pueden usar para proporcionar visualizaciones más precisas, así como para proporcionar recomendaciones efectivas para la siguiente mejor acción.

De manera más general, el trayecto de un cliente, como se usa en la presente, se refiere a una secuencia de eventos que ocurren cuando un cliente interactúa con una empresa o negocio. Esto puede incluir a un cliente que interactúa con sistemas automatizados, por ejemplo, un agente virtual/bot o un sistema IVR. Como se indicó anteriormente, esto puede incluir además cómo un cliente interactúa con un sitio web de una empresa. Los trayectos de clientes se pueden usar como una forma de organizar la información sobre la forma en que los clientes interactúan con las empresas a lo largo del

tiempo. Los trayectos de clientes son series de tiempo discretas y muestreadas de manera desigual de eventos de los clientes que contienen un conjunto heterogéneo de atributos y características. Pueden contener señales inequívocas de compromiso, como comprar un auto nuevo, así como señales de compromiso más ambiguas, como una serie mensual de compras con tarjeta de crédito o contacto con el servicio al cliente. Se ha observado que el uso del marco del trayecto del cliente mejora el rendimiento del aprendizaje automático, ya sea para sistemas de recomendación o para la personalización de la experiencia. Los trayectos de clientes se pueden usar para impulsar las ventas a través de esas recomendaciones y experiencias personalizadas. Con el uso de los trayectos del cliente, la experiencia del cliente puede tratarse como un factor dinámico en lugar de un factor estático. Por lo tanto, por ejemplo, la experiencia del cliente se convierte en una señal o parámetro continuamente administrado que se puede usar para decidir, recomendar y activar actividades de una empresa a sus clientes y posibles clientes. El modelado, por ejemplo, en el estado de programa de una aplicación de aprendizaje automático o de investigación operativa, de los trayectos del cliente se utiliza para determinar un punto de datos, como un nuevo evento o eventos, que minimiza una brecha de conocimiento en el trayecto del cliente. Estos resultados pueden modelarse para determinar

una “próxima mejor acción” que debe tomar una empresa en relación con un cliente que probablemente produzca un resultado deseado, como realizar una venta, o evitar un resultado no deseado, como un cliente que discontinuara el servicio.

Se ha avanzado mucho en relación con el uso efectivo de ciertos tipos de trayectos de clientes. Estos incluyen aquellos tipos de trayectos de clientes que están más estructurados en cuanto a cómo se les permite a los clientes navegar a través de ellos, por ejemplo, los trayectos de clientes que involucran interacciones con bots de agentes, que pueden denominarse flujos de borrado, y aquellos relacionados con las interacciones con los sistemas IVR, que pueden referirse a trayectorias o trayectos IVR. Más allá de eso, sin embargo, sigue habiendo mucha necesidad de un soporte similar en relación con los trayectos de clientes que describen cómo los clientes interactúan con los sitios web. Los avances en esta materia han demostrado ser un desafío difícil. Una razón principal para esto es que, si bien los flujos de bot y las trayectorias IVR tienen una estructura fija de cómo se pueden secuenciar los eventos dentro de cualquier trayecto de cliente dado, la manera en que un cliente interactúa con un sitio web no lo hace. Esta falta de estructura conduce a una multitud de variaciones en las posibles secuencias de eventos (es decir, el ordenamiento de eventos en los

trayectos de clientes). A manera de ejemplo, el uso de estructuras de datos trie (estructura de datos de tipo árbol) han sido productivos en el modelado de flujos de bots y trayectorias IVR con fines predictivos y para
5 generar visualizaciones útiles, pero esta utilidad no se ha extendido a los trayectos de clientes que involucran eventos web. Una de las razones de esto es que las características inherentes de las estructuras de datos trie hacen que el escalado para este tipo de casos de uso sea
10 casi imposible. Con estructuras de datos trie, se crea una nueva rama para cada nuevo prefijo (de secuencias de eventos), lo que a su vez conduce a la creación de una estructura de datos trie muy grande y engorrosa. La estructura de datos trie crece rápidamente en tamaño con
15 incluso ligeras variaciones en los eventos iniciales o la secuencia de eventos, lo que conduce a un aumento significativo en el uso de la memoria y el tiempo de procesamiento que hacen que no sea factible.

Los eventos que ocurren al interactuar con
20 sitios web son generalmente multidimensionales, es decir, los eventos contienen muchos atributos diferentes, y esto se suma a las dificultades para descifrar los trayectos de clientes en esta área. Para manejar casos tan complejos de altas variaciones en las secuencias de
25 eventos, o para dar cuenta de múltiples atributos (es decir, dimensiones) de eventos mientras se minan patrones

con fines de predicción o visualización, una estructura de datos trie simplemente no funciona. En cambio, se ha encontrado que los algoritmos de aprendizaje automático más avanzados (específicamente algoritmos de aprendizaje de secuencias) son más efectivos, como se propone en esta descripción. Tales grandes variaciones en las secuencias vienen con el problema de la visualización ruidosa, en donde identificar qué eventos son importantes para la visualización se vuelve crítico. Las modalidades de la presente invención proporcionan una solución a estos diversos desafíos.

Pasando ahora a modalidades particulares, se apreciará que descubrir patrones o trayectorias comunes en los trayectos de clientes es extremadamente valioso para las empresas, ya que permite que las empresas hagan predicciones que optimicen la experiencia del cliente y los KPI del negocio. No obstante, como se apreciará, los eventos en un trayecto de cliente no tienen igual importancia. (Tenga en cuenta que los eventos de referencia incluyen referencias a eventos o eventos web). Por lo tanto, para identificar patrones comunes en los trayectos de clientes, es vital eliminar eventos sin importancia (ruido) y enfocarse solo en los eventos importantes (eventos hito) que conducen a lograr un resultado y tienen valor predictivo. Estos eventos hito son eventos sobre los cuales se pueden hacer predicciones

precisas sobre las próximas acciones, deseos o necesidades de un cliente. Por supuesto, diferenciar entre estos dos tipos de eventos suele ser difícil, lo que supone desafíos en torno al análisis del trayecto del cliente, particularmente en relación con interacciones complejas que involucran un sitio web.

Cuando los eventos de trayectos del cliente son multidimensionales, es decir, cada evento tiene varios atributos diferentes, las técnicas actuales para descubrir hitos y patrones en los trayectos simplemente no funcionan bien. Esto es particularmente cierto cuando estos atributos de eventos tienen un alto grado de cardinalidad.

Cuando los trayectos del cliente mapean las interacciones con un sitio web de empresa, cada uno de los eventos puede puntuarse en relación con varios tipos de atributos de eventos diferentes. Por ejemplo, en este contexto, los eventos implican generalmente que un cliente se desplaza entre diferentes páginas de un sitio web, realiza búsquedas o interactúa con aspectos de una de las páginas, y los posibles atributos de evento que describen estas acciones incluyen lo que puede ser una larga lista. Por ejemplo, los atributos de eventos pueden incluir características que describen el sitio web, la página que se está visitando, un orden de páginas visitadas, cadenas de búsqueda, sitios web de referencia, así como muchos

otros. Las modalidades preferidas de la presente invención pueden incluir una lista de atributos de eventos que se usa para describir atributos asociados con cada evento, que puede incluir atributos de página web o página, atributos
5 de cliente, atributos de ubicación, atributos de fecha, atributos de sitios web de referencia, atributos de número de eventos, atributos de navegador, atributos de sesión, atributos de dispositivo de usuario, atributos de consulta o búsqueda, así como otros. Estos atributos pueden
10 determinarse en relación con cada evento web que ocurre dentro del trayecto del cliente, por ejemplo, en relación con cada página web visitada, widget con el que interactúa, búsqueda ingresada, etc. que ocurre cuando un cliente interactúa con un sitio web en particular. De acuerdo con
15 una modalidad ilustrativa, una lista de atributos de evento puede incluir cada uno de los siguientes o un subconjunto de los siguientes:

	geolocation_country;
	geolocation_locality;
	geolocation_regionName;
	visit_date;
	referrer_domain;
	session_referrer_url;
	referrer_keywords;
20	ipAddress;
	visitReferrer_domain;
	referrer_name;
	visitReferrer_pathname;
	visitReferrer_medium;
	referrer_pathname;
	total_EventCount;
	session_referrer_queryString;
	session_referrer_hostname;
	referrer_hostname;
	browser_fingerprint;
25	browser_featuresWebRTC;
	browser_viewheight;
	browser_featuresFlash;
	session_shortId;

```

browser_featuresJava;    browser_lang;    device_category;
device_screenheight;    device_osVersion;    page_domain;
page_URL;    totalPageviewCount;    page_lang;    session_type;
session_pageviewCount; customerIdType; loginId; customerId;
5 session_eventCount; device_type; page_breadcrumb; visitId;
session_createdDate;    device_isMobile;
session_referrer_pathname;    session_referrer_fragment;
device_osFamily;    ipOrganization;    mktCampaign_content;
outcomeName;    visitReferrer_name;    browser_family;
10 marketingCampaign_source; referrer_queryString; outcomeId;
referrer_url;    browser_version;    geolocation_longitude;
session_id;    referrer_fragment;    browser_viewWidth;
session_secondsSincePrevious;    searchQuery;
device_fingerprint;    session_secondsSinceFirst;
15 session_durationInSeconds; visitReferrer_url; page_title;
session_referrer_name;    visitReferrer_hostname;
geolocation_timezone;    geolocation_postalCode;
externalContactId;    mktCampaign_clickId;    page_fragment;
geolocation_source;    geolocation_latitude;
20 visitReferrer_keywords; page_keywords; page_queryString;
session_referrer_keywords;    userAgentString;
referrer_medium;    page_hostname;    page_pathname;
organizationId; y visitReferrer_fragment.

```

De acuerdo con modalidades ilustrativas, se

25 generan incrustaciones vectoriales para cada evento (es decir, evento web) que capturan (es decir,

matemáticamente toman en cuenta o reflejan) datos que describen las puntuaciones o valores para el evento dado para cada uno de los atributos de evento. La generación de tales incrustaciones vectoriales puede realizarse de la siguiente manera.

En primer lugar, los atributos de eventos de cardinalidad baja se codifican categóricamente. En tales casos, la cardinalidad baja puede definirse como cardinalidad por debajo de un umbral elegido, por ejemplo, 10. A manera de ejemplo, el país en el que se encuentra el cliente (por ejemplo, “geolocation_country”) puede ser uno de los atributos del evento que tiene una cardinalidad baja. Entonces, por ejemplo, si hay cinco países únicos que aparecen como un valor en este atributo de evento, cada país puede enumerarse como 1, 2, 3, 4, 5, y después normalizarse, formando esto la base para el vector de incrustación.

De acuerdo con modalidades ilustrativas, se puede usar otro proceso para manejar atributos de eventos que tienen una cardinalidad alta. Para los atributos de evento que tienen una cardinalidad alta, los valores únicos pueden agruparse primero. Por ejemplo, los atributos de eventos asociados con las URL de página pueden tener una cardinalidad alta, ya que los nombres de las URL de página son únicos. En modalidades ilustrativas, para realizar la agrupación, las características tendrían que extraerse

primero de las URL de página. Por ejemplo, esto puede incluir la eliminación de caracteres especiales para que quede una cadena de palabras. Entonces, por ejemplo, se podría crear una incrustación vectorial separada para cada URL de página. Esto se puede hacer a través de un modelo de aprendizaje automático entrenado configurado para este propósito, como lo apreciará un experto en la técnica. Las incrustaciones vectoriales se crean a través de un proceso de aprendizaje automático donde se entrena un modelo para convertir tipos particulares de datos en vectores numéricos. Otro modelo de aprendizaje automático (que está entrenado para realizar agrupaciones) se podría usar entonces para agrupar la incrustación vectorial basado en sus similitudes. Por ejemplo, se puede usar una agrupación basada en redes neuronales. Este tipo de agrupaciones usa redes neuronales para aprender la estructura de clústeres de los datos. Ejemplos de este método son los autocodificadores y la agrupación de incrustación profunda. También se pueden usar otros tipos de algoritmos de agrupación, por ejemplo, agrupación basada en centroides (que utiliza la media o mediana de los puntos de un clúster como centro o centroide del clúster), K-means (algoritmo de agrupación basado en centroides más popular); agrupamiento jerárquico (que crea una jerarquía de clústeres, donde cada clúster es un subconjunto del siguiente clúster de nivel superior), agrupamiento basado en densidad (que agrupa

puntos cercanos entre sí en el espacio de características), como DBSCAN, y agrupamiento basado en distribución (modela los datos como una mezcla de distribuciones de probabilidad) como el modelo de mezcla gaussiana (GMM, por sus siglas en inglés) y el agrupamiento espectral (utiliza el propio vectores de una matriz de similitud para agrupar los datos).

Por ejemplo, en una modalidad ilustrativa, las palabras que aparecen en un valor de URL de página dado podrían tratarse como palabras de PNL. Es decir, después de eliminar caracteres especiales, los trabajos restantes se podrían usar para generar una incrustación vectorial, que entonces se usaría para agrupar URL de página similares. El proceso de agrupamiento se puede usar para reducir drásticamente la cardinalidad que aparece en los valores del atributo de evento. Por ejemplo, en el ejemplo de la URL de página, ahora cada una se puede representar en relación con su clúster, que forma las bases de la incrustación vectorial.

En este punto, se puede preguntar por qué no se generan codificaciones categóricas para todos los atributos de eventos y después se agrupan los eventos vectorizados multidimensionales. La razón de esto es que el resultado arrojaría un número de clúster único para cada evento y, por lo tanto, representaría un trayecto como una secuencia unidimensional. Esto limitaría la

efectividad del análisis predictivo asociado. El motivo de por qué no se hace esto también tiene que ver con la explicabilidad. Aunque agrupar los eventos en su conjunto genera secuencias de trayecto unidimensionales, no ofrece
5 ninguna explicación en cuanto a las propiedades de esos clústeres. Por lo tanto, cualquier resultado obtenido del análisis de trayecto y los eventos hito estarán al nivel de “números de clúster” y, por lo tanto, no serán interpretables de manera significativa.

10 Una vez completadas todas las etapas, el trayecto del cliente se puede representar como una secuencia de eventos. En concreto, el trayecto del cliente se representa como la secuencia de incrustaciones vectoriales, tal como se obtuvo anteriormente, para la
15 secuencia de eventos respectivos. Cada evento en el trayecto se transforma en una incrustación vectorial que captura los valores que el evento tenía en cada atributo de evento de la lista de atributos de evento. De acuerdo con modalidades ilustrativas, esta secuencia de
20 incrustaciones vectoriales constituye la entrada para entrenar un modelo de predicción. En cuanto al resultado, puede basarse en si se logró o no un resultado objetivo en cada trayecto del cliente en particular. Por ejemplo, el resultado objetivo (o resultado) podría ser si se logró o
25 no una venta. O, por ejemplo, el resultado podría ser si la interacción con el sitio web resultó en que el cliente

aceptara una sesión de chat ofrecida con un agente en vivo. En modalidades ilustrativas, la salida puede representarse usando una variable binaria, tal como “0” para el resultado que no se logra y un “1” para el
5 resultado que se logra. Como se apreciará, la salida representa la variable objetivo para entrenar el modelo de predicción.

De acuerdo con modalidades ilustrativas, el modelo de predicción se entrena entonces usando las
10 incrustaciones vectoriales y los datos de resultado. Por ejemplo, el modelo puede ser un modelo LSTM bidireccional basado en la atención o un modelo transformador. Como se apreciará, este modelo está entrenado para predecir la salida dada una entrada y se selecciona de acuerdo con su
15 capacidad para manejar incrustaciones multidimensionales de cardinalidad baja, cuya generación se describió anteriormente.

En una modalidad alternativa, cada grupo de entidades de incrustación de eventos multidimensionales
20 se puede codificar categóricamente para generar trayectos unidimensionales. Por ejemplo, un evento $E1 = (a1, a2, a3, a4, a5)$, $E2 = (b1, b2, b3, b4, b5)$, $E3 = (c1, c2, c3, c4, c5)$. Entonces $E1$ se puede codificar categóricamente como 1, $E2$ como 2 y $E3$ como 3. La cardinalidad de esta
25 codificación categórica será igual al producto de cardinalidades de atributos de eventos individuales. La

codificación categórica se puede hacer de la manera (sin aumentar excesivamente la cardinalidad) en que se han tomado las etapas durante el proceso de incrustación para convertir los atributos de eventos de cardinalidad alta en representaciones de cardinalidad baja usando la agrupación.

Como etapa final, una vez entrenado el modelo, las puntuaciones de atención pueden calcularse para algunos de los eventos que aparecen en los trayectos de clientes seleccionados. Tales trayectos de cliente pueden seleccionarse como aquellos casos en los que el modelo capacitado tiene éxito en predecir el resultado. Tales casos pueden seleccionarse del conjunto de datos de entrenamiento y/o incluir nuevos casos en los que la predicción del modelo sea precisa. Las puntuaciones de atención entonces pueden compararse con un umbral de puntuación de atención, identificándose cada uno de los eventos de puntuaciones de atención más altos que el umbral como eventos hito. Cada uno de los eventos que tienen una puntuación de atención inferior al umbral seleccionado puede tratarse como ruido e ignorarse. Esto puede repetirse en muchos de estos casos de prueba para determinar los eventos que aparecen con mayor frecuencia como eventos hito. Los trayectos de clientes para las interacciones con el sitio web pueden entonces representarse como una secuencia de tales eventos hito.

Como se apreciará, tales eventos hito se pueden usar para generar visualizaciones sucintas e informativas de tipos particulares de trayectos de clientes tomados en relación con el sitio web de una empresa. En modalidades
5 ilustrativas, la visualización puede incluir volumen y dirección del tráfico de un evento hito a otro, similar al gráfico dirigido ponderado.

Como etapa adicional, las puntuaciones de atención altas se pueden correlacionar con atributos de
10 eventos particulares presentes en los eventos hito. Esto proporciona una clave para comprender cuáles de los atributos de eventos que componen los vectores multidimensionales de los eventos son más importantes para determinar los resultados. Con esto hecho, se proporciona
15 explicabilidad adicional. Esta información adicional se puede usar para obtener información predictiva adicional, que incluye recomendaciones sobre la “siguiente mejor acción”. Además, a través de un proceso iterativo, la lista de atributos de eventos puede reducirse en longitud,
20 ya que se encuentra que ciertos son predictivos (o importantes para por qué un evento se identifica como un evento hito) y otros solo añaden ruido. Después se pueden entrenar modelos más precisos centrándose en un menor número de atributos de eventos que se descubra que están
25 fuertemente correlacionados con determinados resultados, y

descartar los atributos de eventos que se descubra que son ruido no correlacionado.

Con referencia ahora a la Figura 5, se muestra un método 500 ilustrativo que ilustra una modalidad de la presente invención. El método 500 comienza, en la etapa 505, generando, a través de un proceso de datos de entrenamiento (que se muestra como un inserto a la etapa 505), muestras de datos de entrenamiento a partir de muestras de datos de trayecto respectivas, cada una de las muestras de datos de trayecto incluye el trayecto de un cliente representado por datos que describen una secuencia de eventos, para cada uno de los eventos, valores asociados con atributos de evento respectivos de una lista de atributos de evento, y un resultado de trayecto. En la etapa 510, el método 500 continúa entrenando un modelo de aprendizaje automático usando las muestras de datos de entrenamiento generadas de la etapa anterior. En el entrenamiento, una entrada del modelo de aprendizaje automático, para cada muestra de datos de entrenamiento, incluye una secuencia de las incrustaciones vectoriales generadas a través del proceso de datos de entrenamiento de la secuencia de eventos; y una salida del modelo de aprendizaje automático, para cada muestra de datos de entrenamiento, incluye el resultado de trayecto asociado.

Con respecto al proceso de datos de entrenamiento, como se muestra, el proceso puede comenzar

en la etapa 515 generando una incrustación vectorial para cada uno de los eventos incluidos dentro de una determinada de las muestras de datos de trayecto que captura el valor para cada uno de los atributos de evento. En las etapas 5 520, 525 y 520, se describen las etapas mediante las cuales el proceso de datos de entrenamiento genera las incrustaciones vectoriales. En la etapa 520, el proceso de datos de entrenamiento incluye dividir los atributos de evento de la lista de atributos de evento en un grupo de 10 cardinalidad baja y un grupo de cardinalidad alta, en donde la división se realiza de acuerdo con si los valores incluidos dentro de las muestras de datos de entrenamiento para un atributo de evento dado tienen una cardinalidad por encima o por debajo de un umbral de cardinalidad 15 predefinido. En la etapa 520, el proceso de datos de entrenamiento incluye, para cada uno de los grupos de cardinalidad baja, codificar categóricamente los valores incluidos dentro del grupo de cardinalidad baja de acuerdo con un número total de valores únicos que aparecen en este. 20 En la etapa 520, el proceso de datos de entrenamiento incluye, para cada uno de los grupos de cardinalidad alta, agrupar los valores incluidos dentro del grupo de cardinalidad alta para crear una pluralidad de grupos de agrupamiento; y codificar categóricamente los valores 25 incluidos dentro del grupo de cardinalidad alta de acuerdo con la pluralidad de grupos de agrupamiento en los que

reside el valor. Por lo tanto, cada una de las muestras de datos de entrenamiento incluye una secuencia de las incrustaciones vectoriales generadas a partir de la secuencia de eventos incluidos en una de las muestras de trayecto asociadas; y el resultado de trayecto de una de las muestras de trayecto asociadas.

En modalidades ilustrativas, los eventos son eventos web. Cada evento web puede incluir acciones tomadas por un cliente mientras el cliente interactúa con un sitio web de una empresa, tales como, acciones relacionadas con la selección de páginas web particulares del sitio web para navegar y enviar una consulta de búsqueda.

En modalidades ilustrativas, cuando se describe en relación con una primera muestra de datos de entrenamiento, que es representativa de cómo se usa cada una de las muestras de datos de entrenamiento para entrenar el modelo de aprendizaje automático, la etapa de entrenamiento del modelo de aprendizaje automático puede incluir: proporcionar como entrada al modelo de aprendizaje automático la secuencia de incrustaciones vectoriales generados para la primera muestra de datos de entrenamiento; generar como resultado del modelo de aprendizaje automático un resultado de trayecto previsto; comparar el resultado de trayecto de la primera muestra de datos de entrenamiento con el resultado de trayecto

previsto y, mediante la comparación, determinar una diferencia entre ellos; y ajustar los parámetros del modelo de aprendizaje automático para reducir la diferencia determinada.

5 En modalidades ilustrativas, el resultado de trayecto puede incluir datos que indican si se logra o no se logra una condición binaria. La condición binaria puede relacionarse con una métrica de rendimiento del negocio, por ejemplo, si se realiza una venta o se
10 satisface una métrica de rendimiento.

 En modalidades ilustrativas, los atributos de evento pueden incluir: múltiples atributos que describen la página web que el cliente está navegando, que incluye al menos una dirección URL y palabras clave asociadas con
15 esta; múltiples atributos que describen al cliente, que incluyen al menos un identificador de cliente, una ubicación del cliente y un atributo que describe un dispositivo del cliente; múltiples atributos que describen un sitio web de referencia, que incluyen al
20 menos palabras clave asociadas con el sitio web de referencia; al menos un atributo que describe una consulta de búsqueda enviada por el cliente en el sitio web de referencia o en el sitio web de la empresa; y al
25 menos un atributo que describe un número total de eventos en una sesión de navegación.

En modalidades ilustrativas, se pueden usar uno o más modelos de aprendizaje automático entrenados en la etapa de agrupar los valores incluidos dentro de cada uno de los grupos de cardinalidad alta. Esta etapa puede
5 incluir además: extraer características de los valores incluidos dentro del grupo de cardinalidad alta; generar una incrustación vectorial para las características extraídas para cada uno de los valores; y agrupar de acuerdo con similitudes encontradas en las incrustaciones
10 vectoriales generadas.

En modalidades ilustrativas, el modelo de aprendizaje automático es un modelo transformador. En otras modalidades, el modelo de aprendizaje automático es una red neuronal recurrente de memoria larga a corto plazo
15 bidireccional basada en la atención que calcula las puntuaciones de atención para los eventos respectivos incluidos dentro de un trayecto del cliente al predecir el resultado de un trayecto. En modalidades ilustrativas, el método puede incluir además la etapa de determinar eventos
20 hito mediante: recibir un umbral de puntuación de atención; recibir un subconjunto de muestras de datos de entrenamiento, el subconjunto de muestras de datos de entrenamiento incluye muestras de datos de entrenamiento en las que el modelo de aprendizaje automático entrenado
25 predice con precisión si la condición binaria se logra o no; determinar las puntuaciones de atención para los

eventos incluidos dentro de cada una de las muestras de fecha de entrenamiento del subconjunto de muestras de datos de entrenamiento; para cada uno de los eventos, comparar las puntuaciones de atención con el umbral de puntuación de atención; determinar si cada uno de los eventos es un evento hito basado en si la puntuación de atención del evento excede el umbral de puntuación de atención; e identificar un tipo de trayecto del cliente como una secuencia de los eventos hito determinados.

En modalidades ilustrativas, el método puede incluir además la etapa de generar una representación visual del tipo de trayecto del cliente identificado. La representación visual puede incluir una secuencia de nodos conectados donde cada uno de los nodos está etiquetado como uno de los eventos hito del tipo de trayecto del cliente. En modalidades ilustrativas, la representación visual generada incluye el volumen y la dirección de la etiqueta de tráfico desde uno de los eventos hito hasta otro de los eventos hito. Esto puede ser en forma de un número que etiqueta el conector para el volumen y el conector que se forma como una flecha para la dirección. En modalidades ilustrativas, la etapa de recibir el umbral de puntuación de atención incluye recibir una pluralidad de umbrales de puntuación de atención diferentes para generar una pluralidad de tipos de trayecto del cliente respectivos, la pluralidad de tipos de trayecto del cliente tiene un número

variable de los eventos hito identificados de acuerdo con la pluralidad de umbrales de puntuación de atención diferentes. El método puede incluir la etapa de generar las representaciones visuales de cada uno de la pluralidad de
5 tipos de trayecto del cliente.

En modalidades ilustrativas, el método puede incluir además la etapa de correlacionar puntuaciones de atención altas logradas al determinar los eventos hito con uno o más atributos de eventos particulares presentes
10 en los eventos hito. En modalidades ilustrativas, el método puede incluir además la etapa que produce una o más recomendaciones con respecto a las acciones a tomar con un cliente futuro que interactúa con el sitio web de la empresa cuando se detecta que uno o más atributos de
15 evento particulares están presentes durante la interacción con el futuro cliente. Tales recomendaciones se pueden realizar en tiempo real en respuesta a interacciones en vivo y/o ejecutar automáticamente. En modalidades ilustrativas, el método puede incluir además
20 las etapas de: determinar uno o más atributos de eventos particulares que son ruido basado en una correlación baja en la determinación de los eventos hito; y la salida de una o más recomendaciones con respecto a la eliminación de uno o más atributos de eventos particulares al
25 entrenar una versión revisada del modelo de aprendizaje automático.

Como apreciará un experto en la técnica, las muchas características y configuraciones variables descritas anteriormente en relación con las diversas modalidades ilustrativas pueden aplicarse, además, 5 selectivamente para formar las otras modalidades posibles de la presente invención. En aras de la brevedad y tomando en cuenta las capacidades de un experto en la técnica, no se proporcionan ni se analizan en detalle cada una de las iteraciones posibles, aunque todas las combinaciones y 10 modalidades posibles abarcadas por las diversas reivindicaciones a continuación o de cualquier otro modo están destinadas a ser parte de la presente solicitud. Además, debe ser evidente que lo anterior se relaciona solamente con las modalidades descritas de la presente 15 solicitud y que pueden realizarse varios cambios y modificaciones en la presente sin apartarse del espíritu y alcance de la presente solicitud como se define por las siguientes reivindicaciones y sus equivalentes.