

REIVINDICACIONES

Se reivindica:

- 5 1. Un método implementado por computadora que comprende las etapas de:

generar, a través de un proceso de datos de
entrenamiento, muestras de datos de
entrenamiento a partir de muestras de datos
10 de trayecto respectivas, cada una de las
muestras de datos de trayecto comprende el
trayecto de un cliente representado por
datos que describen:

una secuencia de eventos;
15 para cada uno de los eventos, valores
asociados con los atributos de evento
respectivos de una lista de atributos de
evento; y

un resultado de trayecto;
20 en donde el proceso de datos de
entrenamiento comprende generar una
incrustación vectorial para cada uno de
los eventos incluidos dentro de una
determinada de las muestras de datos de
25 trayecto que captura el valor para cada
uno de los atributos de evento mediante:

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 muestras de datos de entrenamiento para un atributo de evento dado tienen 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 cardinalidad baja de acuerdo con un número total de valores únicos que aparecen en este;

para cada uno de los grupos de alta cardinalidad:

agrupar los valores incluidos dentro del grupo de cardinalidad alta para crear una pluralidad de grupos de agrupamiento; y

y codificar categóricamente los valores incluidos dentro del grupo de

cardinalidad alta de acuerdo con la pluralidad de grupos de agrupamiento en los que reside el valor;

tener en cuenta que 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;

entrenar un modelo de aprendizaje automático usando las muestras de datos de entrenamiento generadas, en donde:

una entrada del modelo de aprendizaje automático, para cada muestra de datos de entrenamiento, comprende la secuencia de las incrustaciones vectoriales; y

una salida del modelo de aprendizaje automático, para cada muestra de datos de entrenamiento, comprende el resultado de trayecto asociado.

2. El método implementado por computadora de la reivindicación 1, en donde los eventos comprenden eventos web, cada evento web comprende una acción tomada por un cliente a medida que el cliente interactúa con un sitio web de una empresa, que incluye acciones relacionadas con la selección de páginas web particulares del sitio web para navegar y enviar una consulta de búsqueda.

3. El método implementado por computadora de la reivindicación 2, en donde 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 comprende:

proporcionar como entrada al modelo de aprendizaje automático la secuencia de incrustaciones vectoriales generadas 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 estos; y

ajustar los parámetros del modelo de aprendizaje automático para reducir la diferencia determinada.

5 4. El método implementado por computadora de la reivindicación 2, en donde el resultado de trayecto comprende datos que indican si se logra o no una condición binaria, la condición binaria se relaciona con una métrica de rendimiento del negocio.

10 5. El método implementado por computadora de la reivindicación 4, en donde la lista de atributos de eventos comprende:

15 múltiples atributos que describen la página web que el cliente está navegando, que incluye al menos una dirección URL y palabras clave asociadas con esta;

20 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 menos palabras clave asociadas con el sitio web de referencia;

25 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 menos un atributo que describe un número total de eventos en una sesión de navegación.

5 6. El método implementado por computadora de la reivindicación 4, en donde, usar uno o más modelos de aprendizaje automático entrenados, la etapa de agrupar los valores incluidos dentro de cada uno de los grupos de cardinalidad alta comprende:

10 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;

15 agrupar de acuerdo con similitudes encontradas en las incrustaciones vectoriales generadas.

7. El método implementado por computadora de la reivindicación 6, en donde el modelo de aprendizaje automático comprende un modelo transformador.

8. El método implementado por computadora de la reivindicación 6, en donde el modelo de aprendizaje automático comprende una red neuronal recurrente de memoria larga a corto plazo bidireccional basada en la atención que
25 calcula las puntuaciones de atención para los eventos

respectivos incluidos dentro de un trayecto del cliente al predecir un resultado de trayecto.

9. El método implementado por computadora de la reivindicación 8, que comprende además la etapa de
5 determinar eventos 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
10 muestras de datos de entrenamiento en las que el modelo de aprendizaje automático entrenado 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
15 las muestras de fecha de entrenamiento del subconjunto de muestras de datos de entrenamiento;
para cada uno de los eventos, comparar las
20 puntuaciones de atención con el umbral de puntuación de atención;
determinar si cada uno de los eventos comprende un evento hito basado en si la puntuación de atención del evento excede el
25 umbral de puntuación de atención; y

identificar un tipo de trayecto del cliente como una secuencia de los eventos hito determinados.

10. El método implementado por computadora de la reivindicación 9, que comprende además la etapa de generar una representación visual del tipo de trayecto del cliente identificado, la representación visual comprende una secuencia de nodos conectados donde cada uno de los nodos se etiqueta como uno de los eventos hito del tipo de trayecto del cliente.

11. El método implementado por computadora de la reivindicación 10, en donde la representación visual generada incluye el volumen y la dirección del etiquetado del tráfico desde uno de los eventos hito hasta otro de los eventos hito.

12. El método implementado por computadora de la reivindicación 11, en donde 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.

13. El método implementado por computadora de la reivindicación 12, que comprende además la etapa de generar

las representaciones visuales de cada uno de la pluralidad de tipos de trayecto del cliente.

14. El método implementado por computadora de la reivindicación 9, que comprende además la etapa de:

5 correlacionar puntuaciones de atención altas logradas al determinar los eventos hito con uno o más atributos de eventos particulares presentes en los eventos hito.

15. El método implementado por computadora de la reivindicación 14, que comprende además la etapa de:

10 producir 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
15 o más atributos de evento particulares están presentes durante la interacción con el futuro cliente.

16. El método implementado por computadora de la reivindicación 15, que comprende además la etapa de:

20 determinar uno o más atributos de eventos particulares que comprenden ruido basados en una correlación baja en la determinación de los eventos hito;
obtener una o más recomendaciones con
25 respecto a la eliminación de uno o más atributos de eventos particulares al entrenar

una versión revisada del modelo de aprendizaje automático.

17. Un sistema que comprende:

un procesador; y

5 una memoria, que almacena instrucciones que, cuando son ejecutadas por el procesador, hacen que el procesador realice las etapas de:

10 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 comprende el trayecto de un cliente representado por datos que describen:

15 una secuencia de eventos;
para cada uno de los eventos, valores asociados con los atributos de evento respectivos de una lista
20 de atributos de evento; y
un resultado de trayecto;
en donde el proceso de datos de entrenamiento comprende generar una
incrustación vectorial para cada
25 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 mediante:

5 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
10 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;

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
20 valores únicos que aparecen en este;

para cada uno de los grupos de alta
cardinalidad:

agrupar los valores incluidos dentro
25 del grupo de cardinalidad alta para

crear una pluralidad de grupos de agrupamiento; y

y codificar categóricamente los valores incluidos dentro del grupo de cardinalidad alta de acuerdo con la pluralidad de grupos de agrupamiento en los que reside el valor;

tener en cuenta que 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;

entrenar un modelo de aprendizaje automático usando las muestras de datos de entrenamiento generadas, en donde:

una entrada del modelo de aprendizaje automático, para cada muestra de datos de entrenamiento, comprende la secuencia de las incrustaciones vectoriales; y

una salida del modelo de aprendizaje automático, para cada muestra de datos de entrenamiento, comprende el resultado de trayecto asociado.

5 18. El sistema de la reivindicación 17, en donde los eventos comprenden eventos web, cada evento web comprende una acción tomada por un cliente a medida que el cliente interactúa con un sitio web de una empresa, que incluye acciones relacionadas con la selección de
10 páginas web particulares del sitio web para navegar y enviar una consulta de búsqueda.

 19. El sistema de la reivindicación 18, en donde el modelo de aprendizaje automático comprende 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; y

 en donde la memoria almacena instrucciones
20 que, cuando son ejecutadas por el procesador, hacen que el procesador realice las etapas de:

 determinar eventos hito mediante:

 recibir un umbral de puntuación de
25 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 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 comprende 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.

20. El sistema de la reivindicación 19, en donde la memoria almacena instrucciones adicionales que, cuando son ejecutadas por el procesador, hacen que el procesador realice las etapas de:

- 5 generar una representación visual del tipo de trayecto del cliente identificado, la representación visual comprende una secuencia de nodos conectados donde cada uno de los nodos se etiqueta como uno de los
- 10 eventos hito del tipo de trayecto del cliente.