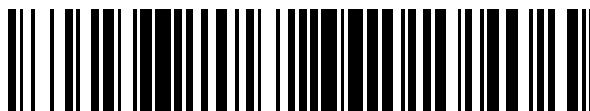


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 668 899**

51 Int. Cl.:

H04N 21/25	(2011.01)
H04N 21/466	(2011.01)
H04N 21/442	(2011.01)
G06Q 30/00	(2012.01)
G06Q 30/02	(2012.01)
H04N 21/236	(2011.01)
H04N 21/2665	(2011.01)
H04N 21/434	(2011.01)
H04N 21/45	(2011.01)
H04N 21/462	(2011.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

- 86 Fecha de presentación y número de la solicitud internacional: **12.04.2012 PCT/IT2012/000103**
- 87 Fecha y número de publicación internacional: **17.10.2013 WO13153562**
- 96 Fecha de presentación y número de la solicitud europea: **12.04.2012 E 12728325 (7)**
- 97 Fecha y número de publicación de la concesión europea: **03.01.2018 EP 2837199**

54 Título: **Recomendaciones de lado de cliente en redes de radiodifusión unidireccional**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
23.05.2018

73 Titular/es:
CONTENTWISE S.R.L. (50.0%)
Via Simone Schiaffino 11
20158 Milano, IT y
POLITECNICO DI MILANO (50.0%)

72 Inventor/es:
CREMONESI, PAOLO y
TURRIN, ROBERTO

74 Agente/Representante:
SUGRAÑES MOLINÉ, Pedro

ES 2 668 899 T3

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

DESCRIPCIÓN

Recomendaciones de lado de cliente en redes de radiodifusión unidireccional

5 **Campo de la invención**

La presente invención se refiere a un sistema y método de recomendación de radiodifusión, en particular a un método de recomendación para una pluralidad de artículos que van a ser comprados por usuarios en redes de radiodifusión unidireccional, tales como TV por satélite (DVB-S) y televisión digital terrestre (DVB-T).

10

Antecedentes de la invención

Tal como es sabido, los sistemas de recomendación ayudan a los usuarios a buscar grandes cantidades de contenido digital y servicios al permitir a los mismos identificar los artículos (por ejemplo, películas, música, libros, noticias, páginas web, ...) que es probable que sean más atractivos o útiles, deduciendo tales recomendaciones en función de elementos diferentes (por ejemplo, popularidad, información demográfica, elecciones y preferencias pasadas individuales o agregadas, valoraciones explícitas sobre una muestra de elementos sugeridos).

15

Cuánto interesa y atrae un artículo a un usuario se denomina relevancia. Por lo tanto, un sistema de recomendación estima la relevancia de un usuario basándose en una colección de artículos. Tal información es especialmente útil en situaciones en donde el número de artículos está creciendo continuamente (por ejemplo, la enorme cantidad de contenidos disponibles en los canales de radiodifusión) y la cantidad de datos relevantes es relativamente baja.

20

Estos sistemas realizan recomendaciones al usuario en función de un modelo del perfil de usuario, es decir, sus gustos, intereses o actitudes. Por lo tanto, por ejemplo, si la colección de artículos consiste en programas de TV y el usuario al que realizar recomendaciones es un bebé, el sistema propondrá una lista de dibujos animados debido a que el perfil del bebé coincide de algún modo con dibujos animados y no coincide con películas de terror o de acción.

25

Este escenario cada vez más amplio de sistemas de recomendación se puede clasificar a lo largo de múltiples dimensiones. Un criterio taxonómico clave es distinguir entre sistemas que ofrecen recomendaciones *no personalizadas* y *personalizadas*. Mientras que los sistemas de recomendación no personalizados (por ejemplo, las películas más populares) no tienen en cuenta el perfil de usuario, los sistemas personalizados realizan sus recomendaciones en función de un "modelo de usuario" (por ejemplo, comportamiento pasado del usuario, intereses de usuario explícitos o implícitos).

30

35

Las dos clases principales de sistemas personalizados son los sistemas *basados en contenido* y *colaborativos*. Los primeros sugieren artículos cuyo contenido es similar al contenido de artículos que el usuario valoró positivamente en el pasado, mientras que los últimos ignoran el contenido y aprovechan las "preferencias colectivas de la multitud".

40

En particular, los algoritmos basados en contenido recomiendan artículos cuyo contenido es similar al contenido de artículos que el usuario ha valorado positivamente en el pasado. La similitud se calcula basándose en las características de contenido explícitas asociadas con los artículos (por ejemplo, una película se puede caracterizar por un género, un director, una lista de actores y un resumen). Debido a que tales sistemas de recomendación tienden a recomendar artículos con las mismas características que las que gustaron a un usuario en el pasado, por lo general los artículos carecen de novedad, lo que quiere decir que el sistema propone una variedad limitada de recomendaciones inesperadas.

45

Por otro lado, los algoritmos colaborativos recomiendan artículos en función de las valoraciones proporcionadas por grupos de usuarios.

50

Tales enfoques generan recomendaciones usando los perfiles de valoración de usuarios diferentes y sugieren artículos que "gustaron" a otros usuarios con gustos similares en el pasado. El grado hasta el cual dos gustos de usuario se consideran similares se basa en la similitud de sus historiales de valoración. En términos generales, el enfoque se puede resumir como "las personas que vieron este programa de TV también vieron...". Las técnicas colaborativas son las más populares y ampliamente implementadas debido a que su integración en los dominios existentes es relativamente sencilla y, en general, su calidad es más alta que la de otras técnicas. Se sabe que los sistemas de recomendación colaborativos proporcionan recomendaciones lo más precisas en comparación con los sistemas de recomendación basados en contenido.

55

60

No obstante, por lo general los sistemas de recomendación colaborativos se diseñan para su aplicación a entornos en dos sentidos, o bidireccionales, en donde una pluralidad de preferencias de los usuarios se pueden detectar y recopilar, procesarse en un lado de extremo posterior y entonces realimentarse de nuevo al dispositivo de usuario en forma de recomendación. Convencionalmente, un sistema de recomendación colaborativo se diseña para un servicio de TV bidireccional (por ejemplo, IPTV, OTT), en donde el canal de retorno - es decir, la comunicación desde el decodificador del usuario al proveedor de servicios - permite que un servicio centralizado identifique y supervise las

65

acciones de los usuarios con el fin de generar recomendaciones personalizadas.

Los sistemas de TV tradicionales solo permiten una comunicación unidireccional (por ejemplo, DVB-S y DVB-T), evitando el uso de un motor de recomendaciones colaborativas centralizado. De hecho, los sistemas de distribución de medios de radiodifusión tradicionales tales como DVB-S y DVB-T se basan en canales de comunicación unidireccional y no permiten usar sistemas de recomendación tradicionales implementados como servicio de extremo posterior, en donde las recomendaciones para un usuario específico son calculadas por un servidor de extremo posterior. De hecho, un sistema de recomendación de radiodifusión no puede interrogar al servidor de extremo posterior. En una arquitectura de ese tipo, el componente a cargo de la generación de recomendación es el STB (descodificador), es decir, el aparato con el que interactúa el usuario y que visualiza ambos medios radiodifundidos por el proveedor de contenido y otra información, entre las cuales hay recomendaciones (es decir, sugerencias acerca de artículos que es probable que sean de interés para el usuario).

Esta configuración hace que el filtrado colaborativo no sea factible debido a que tal enfoque a la recomendación opera sobre colecciones de múltiples perfiles de usuario, información no disponible debido a que cada STB es consciente solo de sus propias preferencias de usuario.

Además, los STB ejecutan aplicaciones de soporte lógico (por ejemplo, un sistema de recomendación) pero se enfrentan a sistemas incrustados con capacidades de memoria y de cálculo limitadas.

Por lo tanto, los métodos de recomendación de radiodifusión convencionales están así limitados a enfoques no personalizados o basados en palabras clave.

No personalizados

La forma más simple de proporcionar recomendaciones es por medio de enfoques no personalizados. Por ejemplo, el proveedor de servicios puede componer una lista editorial de películas mejor valoradas o más vistas y radiodifundir esta lista a cada aparato de TV, que mostrará esta lista a los usuarios. Además, se puede añadir al STB la capacidad de descartar la visualización de películas/programas de TV que ya ha visto el usuario en el pasado, de tal modo que no se recomiende al usuario algo que este ha ya comprado o visto.

Basados en palabras clave

El STB selecciona los artículos a recomendar de acuerdo con una puntuación de recomendación que depende de las palabras clave que representan el contenido de artículo y el perfil de usuario, expresados en términos de palabras clave. Las palabras clave del contenido de artículo se pueden extraer a partir de la descripción del artículo (por ejemplo, la EPG para programas de TV), mientras que las palabras clave que describen el perfil de usuario se pueden obtener a partir de las palabras clave de artículos previamente valorados por el usuario. Cada palabra clave en el perfil de usuario tendrá un peso que depende del tiempo empleado por el usuario viendo programas con esa palabra clave.

Cuanto más alta sea la coincidencia entre las palabras clave del artículo y las palabras clave del perfil de usuario, más altas serán las puntuaciones de recomendación. Los artículos con la puntuación de recomendación más alta serán recomendados al usuario por el STB.

En particular, de acuerdo con la técnica anterior, ya se ha propuesto implementar un sistema de recomendación dentro del mismo descodificador para este tipo de servicios de TV unidireccional. Por lo general, este es un sistema basado en contenido, que se basa solo en descripciones de los artículos y un modelo de usuario construido sobre el historial de comportamiento del mismo usuario del dispositivo. Esta solución adolece de las desventajas típicas de un sistema de recomendación basado en contenido. En particular, la totalidad de los medios de radiodifusión han de contener meta-etiquetas u otros artículos legibles que permiten que el sistema reconozca el contenido de los medios, lo almacene y lo evalúe y, por último, implemente la recomendación. Adicionalmente, tal como ya se ha dicho, este tipo de diseño requiere una capacidad de cálculo elevada del descodificador para implementar localmente el modelo de recomendación y actualizar este con datos de historial.

El documento US2008022307 divulga un ejemplo de un sistema de recomendación basado en contenido de ese tipo para un aparato de TV de canales de radiodifusión. Se divulga un método de recomendación de programas de radiodifusión, que puede seleccionar fácilmente un canal de programas de radiodifusión deseado para un usuario mediante la selección de una pluralidad de canales de programas de radiodifusión clasificados usando un historial de uso de dicho usuario.

Resumiendo, los sistemas de recomendación colaborativos actuales para los servicios de TV requieren la presencia de un canal de comunicación bidireccional: el canal de retorno permite que el proveedor identifique al usuario y adquiera, o bien de forma explícita o bien de forma implícita, sus preferencias para ser usadas por el sistema de recomendación para sugerir un contenido personalizado.

Desafortunadamente, los sistemas de distribución de TV tradicionales solo prevén una comunicación unidireccional (por ejemplo, TV por satélite y televisión digital terrestre), evitando que el sistema de recomendación identifique a los usuarios y recopile y compare activamente sus preferencias. En consecuencia, el STB es el único componente del sistema consciente de la actividad de usuario y, por lo tanto, de las preferencias de usuario. En este entorno de comunicación de radiodifusión, el STB es el único componente que se puede encontrar a cargo del cálculo de la lista de recomendaciones para un usuario de TV. Esto limita en gran medida las capacidades de recomendación, principalmente por dos razones:

(i) el STB del usuario no tiene conocimiento alguno del comportamiento colectivo de otros usuarios; por lo tanto, el STB no puede aplicar técnicas de recomendación colaborativa basándose en el comportamiento social de una comunidad de usuarios; esta es una fuerte limitación debido a que se sabe que las técnicas de recomendación colaborativa son las más eficaces y precisas, debido a que las mismas son capaces de proporcionar recomendaciones novedosas y útiles;

(ii) el STB del usuario es un sistema incrustado con capacidades de cálculo y de memoria limitadas; por esta razón, la vasta mayoría de los algoritmos de recomendación no se pueden ejecutar en el STB debido a que sus requisitos de cálculo y de memoria superan las capacidades del STB. El documento EP1975866 divulga un sistema de recomendación de acuerdo con el preámbulo de la reivindicación 1. El documento "Osman Nurg Osmanli: A SINGULAR VALUE DECOMPOSITION APPROACH FOR RECOMMENDATION SYSTEMS" (30 de julio de 2010, XP05504 74 77, URL: <https://etd.lib.metu.edu.tr/upload/12612129/index.pdf>) da algunos antecedentes para los sistemas de recomendación híbridos.

Sumario de la invención

Por lo tanto, un objeto de la presente invención es solucionar las desventajas anteriores, mediante la provisión de un método y un sistema en un servicio de radiodifusión unidireccional (en donde no está disponible un canal de retorno) que es capaz de basarse en una solución de recomendaciones basándose en un filtrado colaborativo.

El objeto anterior se logra a través del sistema tal como se describe en las reivindicaciones adjuntas.

En particular, la solución de la invención está compuesta por dos módulos, uno que se ejecuta en un ordenador centralizado conectado a Internet y uno que se ejecuta en el dispositivo descodificador (STB) que recibe datos de radiodifusión.

De acuerdo con un primer aspecto de la invención, se proporciona un sistema de recomendación de lado de cliente en una red de radiodifusión unidireccional, que comprende al menos un canal unidireccional que conecta una estación de radiodifusión con un dispositivo descodificador (STB) de un primer conjunto de usuarios, comprendiendo dicho descodificador (STB) una memoria, una unidad de cálculo y de accionamiento para al menos rastrear la actividad de usuario, almacenar un perfil de usuario, generar y visualizar, en un aparato de TV, una clasificación de artículos de medios recomendados, siendo determinada dicha clasificación por dicha unidad de capacidad de cálculo a través de un modelo de recomendación almacenado en dicha memoria, en donde este comprende un módulo doble colaborativo, un primer módulo de extremo posterior que se ejecuta en un servidor auxiliar conectado a una red bidireccional (Internet) y que pertenece a un dominio auxiliar (dominio A), aprovechando este primer módulo de extremo posterior las fuentes de datos externas para recopilar información basada en comunidad sobre al menos parte de dichos artículos de medios, mientras que dicho modelo de recomendación almacenado en dicha memoria del descodificador (STB) es un segundo módulo de extremo frontal, y por que la estación de radiodifusión y dicho dispositivo descodificador pertenecen a un dominio objetivo (dominio T), dicho servidor auxiliar es parte de dicha red bidireccional, que comprende al menos dos canales para recopilar datos de realimentación a partir de un segundo conjunto de usuarios sobre al menos parte de dichos artículos de medios, dicho primer módulo de extremo posterior es procesado por dicho servidor auxiliar a través de técnicas colaborativas que incluyen dichos datos de realimentación recopilados a partir de dicho segundo conjunto de usuarios, dicho primer módulo de extremo posterior es transformado en una representación compacta por dicho servidor y se radiodifunde unidireccionalmente a dicho descodificador (STB) en donde este se implementa como dicho segundo módulo de extremo frontal.

De acuerdo con otro aspecto de la invención, dicho segundo conjunto de usuarios es diferente del primer conjunto de usuarios.

De acuerdo con un aspecto adicional, la unidad de capacidad de cálculo, a través de dicho segundo módulo de extremo frontal, rastrea localmente el comportamiento del usuario de dicho primer conjunto de usuarios y lo compara con el comportamiento de los otros usuarios de dicho segundo conjunto de usuarios, con el fin de proporcionar dicha clasificación de artículos de medios recomendados.

Preferiblemente, dicho primer módulo de extremo posterior se procesa a través de algoritmos de recomendación híbridos mediante la combinación de técnicas tanto colaborativas como basadas en contenido que incluyen artículos adicionales diferentes de dichos artículos de medios.

De acuerdo con otro aspecto, la invención proporciona un método de recomendación de programas de radiodifusión unidireccional que comprende proporcionar un canal unidireccional que conecta una estación de radiodifusión con un dispositivo descodificador (STB) de un primer conjunto de usuarios en un dominio objetivo (dominio T), en donde dicho descodificador (STB) se dispone para rastrear la actividad de usuario, almacenar un perfil de usuario, generar y visualizar, en un aparato de TV, una clasificación de artículos de medios recomendados, siendo determinada dicha clasificación por una unidad de cálculo de dicho descodificador (STB) a través de un modelo de recomendación almacenado en una memoria del mismo, y que comprende adicionalmente las etapas de operar un módulo doble colaborativo, en donde un primer módulo de extremo posterior se ejecuta en un servidor auxiliar que pertenece a una red bidireccional en donde se desarrolla un modelo colaborativo entre un conjunto auxiliar de usuarios y un conjunto auxiliar de artículos, que se solapa al menos parcialmente con dichos artículos de medios, que pertenece a un dominio auxiliar (dominio A), compactando y adaptando dicho modelo colaborativo desarrollado en el dominio auxiliar (dominio A) para radiodifundirse en dicho dominio objetivo (dominio T) que se va a usar como dicho modelo de recomendación almacenado en dicha memoria del descodificador (STB) al aprovechar dicha actividad de usuario y perfil de usuario que pertenece al dominio objetivo (dominio T).

Además, un aspecto de la invención es un método de recomendación de programas de radiodifusión unidireccional que comprende proporcionar un canal unidireccional que conecta una estación de radiodifusión con un dispositivo descodificador (STB) de un primer conjunto de usuarios, en donde dicho descodificador (STB) se dispone para rastrear la actividad de usuario, almacenar un perfil de usuario, generar y visualizar, en un aparato de TV, una clasificación de artículos de medios recomendados, siendo determinada dicha clasificación por una unidad de cálculo de dicho descodificador (STB) a través de un modelo de recomendación almacenado en una memoria del mismo, que comprende la etapa de operar un módulo doble colaborativo, un primer módulo de extremo posterior que se ejecuta en un servidor auxiliar conectado a una red bidireccional (Internet) y que pertenece a un dominio auxiliar (dominio A), aprovechando este primer módulo de extremo posterior las fuentes de datos externas para recopilar información basada en comunidad sobre al menos parte de dichos artículos de medios, y siendo un segundo módulo de extremo frontal dicho modelo de recomendación almacenado en dicha memoria del descodificador (STB) que pertenece a un dominio objetivo (dominio T), en donde dicho servidor auxiliar comprende al menos dos canales para recopilar datos de realimentación a partir de un segundo conjunto de usuarios sobre al menos parte de dichos artículos de medios, dicho primer módulo de extremo posterior es procesado por dicho servidor auxiliar a través de técnicas colaborativas que incluyen dichos datos de realimentación recopilados a partir de dicho segundo conjunto de usuarios, dicho primer módulo de extremo posterior es transformado en una representación compacta por dicho servidor auxiliar y se radiodifunde unidireccionalmente a dicho descodificador (STB) en donde este se implementa como dicho segundo módulo de extremo frontal.

Breve descripción de los dibujos

En cualquier caso, las características y ventajas adicionales del sistema y método de acuerdo con la invención serán más evidentes a partir de la siguiente descripción detallada de una realización preferida de la misma, dada a modo de ejemplo e ilustrada en los dibujos adjuntos, en donde:

la figura 1 es una vista esquemática que muestra la integración de esquemas de recomendaciones colaborativas en un canal de comunicación unidireccional de acuerdo con la invención;

la figura 2 es un diagrama de flujo que muestra una técnica de aumento de características con filtrado unidimensional aplicada en una realización de la invención; y

la figura 3 es un diagrama de flujo que muestra una técnica de Knn de inyección de similitud unidireccional aplicada en otra realización de la invención.

Descripción detallada de las realizaciones preferidas

La invención supera ambas de las limitaciones explicadas en relación con los métodos de la técnica anterior, mediante la adopción de un sistema de recomendación colaborativo de dos módulos original.

El primer módulo del sistema de recomendación se ejecuta en un ordenador centralizado conectado a Internet: este primer módulo aprovecha las fuentes de datos externas para recopilar información basada en comunidad requerida para recomendaciones colaborativas. Esta información externa se transforma en una representación compacta de los gustos de los usuarios (a la que se hace referencia como modelo) y se radiodifunde a los descodificadores STB del sistema. El segundo módulo del sistema de recomendación se ejecuta en el STB de cada usuario. Estos usuarios pueden ser diferentes de los usuarios analizados por el primer módulo. El segundo módulo es un programa liviano que rastrea localmente, en el STB, las actividades del usuario y las compara con el comportamiento de otros usuarios (en términos del modelo procesado e instalado en el mismo STB) con el fin de proporcionar recomendaciones al espectador de TV. El comportamiento de los otros usuarios se incluye en el modelo procesado por el primer módulo del sistema de recomendación.

Con más detalle, de acuerdo con la invención el modelo de recomendación es construido por un componente de extremo posterior que tiene acceso a un entorno bidireccional basado en Internet (en donde un modelo colaborativo se puede procesar fácilmente usando tecnologías conocidas) y se envía regularmente al STB (por ejemplo, diaria o

semanalmente), que lo usa junto con las preferencias de usuario con el fin de seleccionar los artículos a visualizar al usuario.

- 5 Debido a que el modelo es construido por el extremo posterior, no se puede usar un algoritmo de recomendación clásico, sino que se desarrolla un nuevo método de recomendación que se descompone en dos fases: la primera fase está dedicada a la creación de un modelo; la segunda fase proporciona recomendaciones a partir del modelo y el perfil del usuario objetivo. Debido a las características de comunicación unidireccional del escenario de TV, es necesario que la segunda fase sea desacoplada totalmente por la primera fase, con la excepción del modelo.
- 10 El método de recomendación de radiodifusión permite implementar también algoritmos de recomendación híbridos mediante la combinación de recomendaciones tanto colaborativas como basadas en contenido.

En la siguiente descripción detallada, se usa una notación específica que tiene el significado explicado en la tabla 1.

15

Tabla 1

Dominio T	Dominio objetivo con un canal de comunicación unidireccional. El Proveedor T opera en este dominio.
Dominio A	Dominio auxiliar con un canal de comunicación bidireccional. El Proveedor A opera en este dominio.
u, e	Índices de los usuarios. Convencionalmente, u indica un usuario en el dominio T al que se van a realizar recomendaciones, mientras que e indica un usuario en el dominio A.
i, j	Índices de artículos.
b_{ui}	Sesgo de valoración para el usuario u acerca del artículo i .
N	Número de usuarios en la base de clientes del dominio A.
M	Número de artículos disponibles en el catálogo del dominio T (es decir, adecuados para ser recomendados).
\bar{m}	Número de artículos disponibles en el catálogo del dominio A.
r_u	Perfil del usuario u . Un vector de m -dimensiones cuyos valores distintos de cero se corresponden con las valoraciones no sesgadas del usuario u .
R	Matriz de valoraciones de usuario (URM). Colección de valoraciones expresadas por usuarios en el dominio A.
W	Matriz de contenidos de artículo (ICM). Colección de características de contenido de artículos en el dominio T.
Operador *	Este indica la traspuesta de una matriz o un vector (por ejemplo, \mathbf{Q}^* es la traspuesta de la matriz \mathbf{Q}).
Operador *	Este indica el producto interno entre dos matrices o vectores (por ejemplo, $\mathbf{U} \cdot \mathbf{S}$ es el producto interno entre la matriz \mathbf{U} y la matriz \mathbf{S}).
Función norma(a)	Función aplicada al vector a que normaliza el vector mediante su norma euclídea.
Función norma(A)	Función aplicada a la matriz A que normaliza todas las columnas de la matriz mediante sus normas euclídeas respectivas.

20

El sistema de recomendación de radiodifusión está compuesto por un módulo de extremo posterior y uno de extremo frontal. El módulo de extremo posterior se ejecuta en un servidor centralizado y es responsable de distribuir todos los datos requeridos por el extremo frontal para generar recomendaciones. El módulo de extremo frontal se ejecuta en el STB (descodificador) de cada usuario y es responsable de rastrear la actividad de usuario, almacenar el perfil de usuario, generar y visualizar recomendaciones y mostrar el contenido de medios seleccionado por el usuario (al que se hace referencia genéricamente como artículos).

25

Además del flujo de contenido de medios, los datos clave enviados por el componente de extremo posterior al STB es el modelo de recomendación que se va a usar para proporcionar recomendaciones.

30

El componente de extremo posterior se comunica con el extremo frontal a través de una red de radiodifusión unidireccional, en donde la comunicación es unidireccional desde el extremo posterior al extremo frontal, al tiempo que ningún dato puede ser enviado por el STB al extremo posterior.

Se prevé que el STB rastree la actividad de usuario y represente preferencias de usuario en forma de vector disperso. Para cada artículo (por ejemplo, una película), el usuario puede expresar una valoración, es decir, un valor entero en la escala de 1 a 5. Se hace referencia a esta valoración como valoración explícita. Por ejemplo, después de ver una película, se le da al usuario la posibilidad de dar una valoración al hacer clic en un esquema convencional

de valoración de 1 a 5 estrellas. Si el usuario simplemente ve el artículo (por ejemplo, en el caso de un programa de TV) sin expresar una valoración, se guarda una valoración implícita. El valor de la valoración implícita varía de 3 a 5 y este es linealmente proporcional al porcentaje de contenido que ha visto el usuario. Si el usuario ha visto menos de 5 minutos, el STB descarta tal actividad. Para cada visionado mayor que o igual a 5 minutos, la valoración implícita se calcula como: $3 + 2 \cdot t/l$ en donde t es el tiempo (en minutos) empleado por el usuario viendo el artículo, mientras que l es la longitud total (en minutos) del artículo.

Indíquese con r_u el vector de m -dimensiones de valoraciones recopiladas para un usuario genérico u en el dominio objetivo. El elemento r_{ui} del vector de perfiles de usuario representa la valoración (o bien explícita o bien implícita) del usuario u para el artículo i , ajustada apropiadamente tal como sigue:

- para todas las valoraciones no ausentes, r_{ui} es igual a la valoración menos la constante b_{ui} . El valor b_{ui} es el sesgo de valoración y depende de: el sesgo de valoración de los usuarios promedio, la tendencia de valoración de usuario subjetiva y la tendencia de valoración de artículo. El sesgo se puede definir por cada par (u,i) . En la invención de los presentes inventores, se ha establecido b_{ui} igual a 2,5 para todos los pares (u,i) .
- para todas las valoraciones desconocidas, r_{ui} se establece a cero. De hecho, al dedicarse una tarea de las N recomendaciones más altas, no preocupa a los inventores de la presente invención una predicción de valoración exacta, sino que solo están interesados en clasificar correctamente los artículos. La elección de los ceros, que no afecta a la potencia de predicción, da la oportunidad de usar herramientas de soporte lógico optimizadas para matrices dispersas.

Se almacenan vectores dispersos de perfil de usuario en una estructura de datos optimizada para valores dispersos. Para cada artículo valorado, se almacena un par de campos: el ID único de programa de TV (por ejemplo, un dato entero de 32 bits de longitud) y la valoración de usuario no sesgada - o bien implícita o bien explícita - relacionada (por ejemplo, un dato en coma flotante de 32 bits), que se corresponde con un requisito de 8 bytes por artículo visto. Por ejemplo, un usuario que ha visto 500 artículos requiere 4 Kbytes para almacenar sus preferencias.

El método permite desplegar enfoques tanto híbridos como de filtrado colaborativo, en donde filtrado colaborativo y basado en contenido se usan conjuntamente dentro del mismo algoritmo.

Cada vez que el STB ha de calcular recomendaciones para un usuario, el STB:

1. usa el modelo de recomendación más actualizado con el fin de predecir la puntuación s_{ui} para cada artículo i no valorado por el usuario. El modelo ha sido enviado previamente por el módulo de extremo posterior a través del canal de comunicación de radiodifusión.
2. clasifica los artículos en orden decreciente de puntuación.
3. selecciona los primeros N artículos de esta lista clasificada y los visualiza al usuario. Por ejemplo, el STB muestra 10 artículos a los usuarios.

Obsérvese que los artículos ya valorados por el usuario se ignoran y no se recomiendan.

Colaborativo

En un sistema de TV unidireccional, el proveedor de servicios no puede recopilar centralmente las preferencias de los usuarios del sistema de TV. Tales datos representan la información que se requiere para usar sistemas de recomendación colaborativos.

Debido a que esta información no está disponible, el método colaborativo de acuerdo con la invención se basa en transferir el conocimiento requerido de un dominio externo y bidireccional - en donde las interacciones de usuario se pueden recopilar globalmente - al dominio unidireccional en donde los inventores de la presente invención desean que se recomiende a los usuarios usando una técnica colaborativa. Haciendo referencia a la figura 1, se hace referencia al dominio unidireccional como dominio objetivo (dominio T) y se hace referencia al dominio bidireccional como dominio auxiliar (dominio A) o como fuente externa. El proveedor que opera en el dominio objetivo es el proveedor T, mientras que el proveedor que opera en el dominio auxiliar es el proveedor A.

Para cada artículo, se supone que el proveedor T conoce un conjunto de información en relación con el contenido (por ejemplo, el título, el conjunto de actores, el directorio, la trama), mientras que se supone que el proveedor A conoce las preferencias de sus bases de clientes acerca de artículos en el catálogo.

Tal como se muestra en la figura 1, el proceso se descompone en tres etapas:

1. El Proveedor A recopila preferencias de los usuarios en el dominio A acerca del conjunto de artículos comunes al dominio T. También se pueden recopilar, preferiblemente, preferencias acerca de artículos que existen en el dominio A pero que no existen en el dominio T. De hecho, a pesar de que no se recomendarán estos, pueden aprovechar relaciones indirectas entre preferencias de usuario. Un modelo de comunidad se construye usando

tales datos y un algoritmo de recomendación adecuado (conocido de por sí).

2. El modelo de comunidad se envía al proveedor T y se adapta al dominio T.

3. El modelo de comunidad adaptado se radiodifunde a los STB con el fin de posibilitar recomendaciones.

5 Por ejemplo, el dominio A consiste en un portal web a través del cual un número de usuarios pueden dar su valoración a un número de artículos (películas, espectáculos, ...), mientras que el dominio T es un servidor de extremo posterior de una empresa de radiodifusión de TV. No se evita que el mismo distribuidor de TV pueda también gestionar un portal web en el que los usuarios pueden ver programas de TV, leer la descripción y expresar una valoración. Se ha de hacer notar que no hay necesidad alguna de que los usuarios de los dos sistemas (por ejemplo, el aparato de TV tradicional y el portal web de TV actualizado en el dominio A) sean los mismos. Ni siquiera existe la necesidad de identificarlos. En consecuencia, la fuente externa (dominio auxiliar) no ha de ser gestionada necesariamente por el mismo proveedor que el que gestiona el dominio objetivo.

15 Esto abre la oportunidad de implementar un tipo de mercado de modelos de recomendación, en donde los patrones de comportamiento de los usuarios en un dominio (dominio A) se pueden vender/transferir a otro dominio (dominio T) en donde esta información está ausente o es pobre debido al canal unidireccional.

El dominio auxiliar permite que el proveedor externo recopile preferencias de su base de clientes.

20 Las preferencias de los usuarios de dominio auxiliar se almacenan de forma compacta en una matriz de valoraciones de usuario (URM) $n \times \bar{m}$, indicada por \mathbf{R} , es decir, una colección de todas las valoraciones expresadas de forma implícita o explícita por los n usuarios registrados con el proveedor externo acerca de los \bar{m} artículos disponibles en el catálogo del dominio A. Obsérvese que \mathbf{R} almacena las valoraciones de usuario no sesgadas, es decir, a cada valoración del usuario e acerca del artículo i se le resta el sesgo b_{ei} .

25 El Proveedor A ejecuta un algoritmo de recomendación colaborativo sobre la URM recopilada y obtiene un modelo de recomendación de comunidad. Por lo tanto, los únicos datos que se requiere compartir con el proveedor del dominio objetivo es el modelo de recomendación. Debido a que tal modelo contiene solo una representación de patrones de comportamiento de usuario, esto asegura el nivel más alto de privacidad.

30 Con el fin de usar apropiadamente los datos que provienen del dominio auxiliar, preferiblemente estos se proporcionan para alinear los artículos en los dos dominios.

Si se supone que existe un identificador único (de un tercero) común al proveedor A y T, que permite identificar de forma única un artículo con un identificador independiente, los datos que provienen de la fuente externa se pueden alinear directamente con el catálogo en el dominio objetivo. Como un ejemplo no limitante, las películas y los programas de TV se pueden identificar de forma única a través del identificador de IMDb (base de datos de películas de Internet, <http://www.imdb.com/>). De esta forma, tanto el dominio objetivo como el auxiliar son capaces de mapear sus películas y programas de TV con el identificador de IMDb.

En contraposición, si el proveedor A y el proveedor T no comparten un identificador común y único, supóngase que:

- 40
- los catálogos de artículos en los dominios objetivo y auxiliar están compuestos por los mismos conjuntos de m artículos (por ejemplo, m películas de televisión a la carta); no obstante, los índices de artículo en los dos catálogos podrían no corresponderse (por ejemplo, el artículo i en el dominio objetivo puede ser diferente del artículo i en el dominio auxiliar).
- 45
- el modelo colaborativo construido sobre el dominio auxiliar es representado por una matriz $k \times m$, indicada con \mathbf{M} , en donde las m columnas de \mathbf{M} modelan los m artículos en el catálogo.
 - $\text{imdbid}(c,i)$ es una función de mapeo que, dado un índice de artículo i y un catálogo c , devuelve el identificador de IMDb para el artículo i en el catálogo c . El índice de artículo i está en el intervalo $[0, m]$, mientras que c es o bien el dominio 'objetivo' o bien el 'auxiliar'.

50 Dado este conjunto de información, se puede construir un modelo $\bar{\mathbf{M}}$ adecuado para el dominio objetivo partiendo del modelo \mathbf{M} .

- 55
1. Para cada artículo $j \in [1, m]$, se obtiene un identificador de IMDb mediante la llamada a $\text{imdbid}(\text{'objetivo'}, j)$. Indíquese el identificador con i_j .
 2. Para cada artículo $i \in [1, m]$, se obtiene el identificador de IMDb mediante la llamada a $\text{imdb}(\text{'auxiliar'}, i)$. Cuando se halla un artículo i de tal modo que $\text{imdbid}(\text{'objetivo'}, j)$ es igual a $\text{imdbid}(\text{'auxiliar'}, i)$ la i -ésima columna de \mathbf{M} se **asigna** a la j -ésima columna de $\bar{\mathbf{M}}$.

60 Cada vez que se produce un intercambio de datos entre el dominio A y el dominio T, se necesita alinear tales datos usando el mecanismo previo. Por razones de simplicidad, en el resto del análisis se supone que:

- El catálogo del dominio A se corresponde con el catálogo del dominio B y $\bar{m} = m$.
- 65
- Los artículos ya se han alineado, es decir, $\text{imdbid}(\text{'objetivo'}, i) = \text{imdbid}(\text{'auxiliar'}, i)$ para todos los artículos.

SVD unidireccional (OWSVD)

Los modelos de factores latentes - también conocidos informalmente como modelos de SVD - intentan explicar las valoraciones mediante la caracterización de artículos y usuarios con factores que se deducen automáticamente a partir de la realimentación de usuario. Una realización preferida de la invención prevé que una descomposición en valores singulares (SVD), una técnica particular de factorización de matrices usada en diferentes campos, se emplee en la primera etapa del proceso para construir el modelo de comunidad.

La idea clave de OWSVD es factorizar, por medio de SVD, la matriz de valoraciones de artículo de usuario **R** construida con los datos del dominio auxiliar. La matriz de valoraciones de usuario se descompone en el producto de dos matrices de menor rango, una que contiene los así denominados 'factores de usuario', mientras que la otra contiene los así denominados 'factores de artículo'.

Se supone que los usuarios y artículos se pueden representar por medio de *f* características. Por lo tanto, cada usuario *e* se representa con un vector de factores de usuario de *f* dimensiones $\mathbf{p}_e \in \mathbb{R}^f$. De forma similar, cada artículo, *i* se representa con un vector de factores de artículo $\mathbf{q}_i \in \mathbb{R}^f$. Dada la matriz dispersa **R**, es decir, la URM del dominio auxiliar, el proveedor A puede usar paquetes de soporte lógico muy optimizados existentes para realizar una SVD convencional sobre matrices dispersas (por ejemplo, se han llevado a cabo experimentos con SVDLIBC basándose en la biblioteca de SVDPACKC; véase M. W. Berry. "Large-scale sparse singular value computations". The International Journal of Supercomputer Applications, 6(1):13-49, primavera de 1992.).

Por medio de SVD, la matriz de valoraciones de usuario **R** se aproxima en el espacio de características de *f* dimensiones mediante el producto:

$$\mathbf{U} \cdot \mathbf{S} \cdot \mathbf{Q}^*$$

en donde, **U** es una matriz ortonormal $n \times f$, **Q** es una matriz ortonormal $m \times f$, y **S** es una matriz diagonal $f \times f$ que contiene los primeros *f* valores singulares, en orden de magnitud decreciente.

La matriz **Q** es calculada por el proveedor A y representa el modelo que se va a transferir del dominio auxiliar al dominio objetivo. Por lo tanto, se envía al decodificador STB esta matriz **Q** que se usará con el fin de realizar recomendaciones al usuario en tiempo real.

Recuérdese que **Q** se va a alinear con índices de artículo de dominio objetivo al transferirse del dominio A al dominio T.

Los requisitos de almacenamiento del modelo de recomendación **Q** dependen linealmente del número de artículos. Para cada artículo, se requiere almacenar *f* características, por ejemplo, *f* datos en coma flotante de 32 bits por artículo. Por ejemplo, suponiendo un catálogo de programas de TV compuesto por 10.000 artículos, este necesitaría aproximadamente 8 Mbytes para almacenar el modelo de recomendación. Tal modelo de 8 Mbytes es radiodifundido regularmente por el Proveedor T a cada STB con el fin de actualizar el modelo de recomendación.

Una vez que el modelo **Q** ha sido recibido por el STB, este puede ser usado por el módulo de soporte lógico incrustado con el fin de recomendar a los usuarios interactuar con tal aparato.

De hecho, defínase $\mathbf{P} = \mathbf{U} \cdot \mathbf{S}$, de tal modo que la *e*-ésima fila de **P** representa el vector de factores de usuario \mathbf{p}_e , mientras que la *i*-ésima fila de **Q** representa el vector de factores de artículo \mathbf{q}_i . Además, debido a que **U** y **Q** tienen columnas ortonormales, se puede obtener directamente que:

$$\mathbf{P} = \mathbf{U} \cdot \mathbf{S} = \mathbf{R} \cdot \mathbf{Q}$$

Debido a estas propiedades, el STB puede predecir, con el tiempo, la puntuación s_{ui} del usuario *u* para cualquier artículo *i* realizando el siguiente cálculo en dos etapas:

$$\overline{\mathbf{r}}_u = \mathbf{r}_u \cdot \mathbf{Q} \quad \text{(etapa 1)}$$

en donde \mathbf{r}_u es el vector de *m* dimensiones de valoraciones conocidas y $\overline{\mathbf{r}}_u$ representa el modelo del usuario *u* en un espacio de *f*-dimensiones

$$s_{ui} = b_{ui} + \overline{\mathbf{r}}_u \cdot \mathbf{q}_i^* \quad \text{(etapa 2)}$$

Las dos operaciones matemáticas anteriores, ejecutadas en ese orden específico, hacen que el cálculo sea factible

en cada STB, con independencia de las capacidades de cálculo y de memoria.

De hecho, suponiendo que el usuario u ha valorado $|r_u|$ artículos:

- 5 La etapa 1 requiere $f \cdot |r_u|$ multiplicaciones y $f \cdot |r_u|$ sumas.
La etapa 2 requiere f multiplicaciones y f sumas.

10 Obsérvese que la etapa 1 es común a todos los artículos para los que se necesita calcular la puntuación; por lo tanto, con el fin de calcular la puntuación para cualquier otro artículo, solo es necesario ejecutar la etapa 1 una vez y repetir la etapa 2 para cada artículo.

Los artículos con las puntuaciones más altas serán recomendados al usuario.

15 La SVD unidireccional ofrece los beneficios de representar los usuarios como una combinación de características de artículo sin parametrización específica de usuario alguna. Esta también ofrece una optimización conveniente, que no requiere afinar las constantes de aprendizaje. El único parámetro de OWSVD es f , es decir, el número de características usadas para representar usuarios y artículos. Tal parámetro se puede afinar a través de validación cruzada. No obstante, en las aplicaciones típicas, un valor de f en el intervalo [200, 300] conduce, en general, a recomendaciones de buena calidad. El hecho de que la OWSVD requiera el establecimiento de un único parámetro hace que el algoritmo sea particularmente adecuado para transferirse entre dominios diferentes, como en los ajustes del sistema unidireccional de la invención en el que el algoritmo se entrena con usuarios del dominio A, pero usado sobre usuarios del dominio T.

25 *Colaborativa + de contenido híbridas*

Algunos catálogos tienen un gran número de artículos sin valoración alguna. En particular, las aplicaciones de TV se ocupan de un catálogo de artículos disponibles (por ejemplo, programas de TV) que es muy dinámico y el porcentaje de artículos valorados es tan bajo como un 20 %. Además, puede ser el caso que algunos artículos que existen en el dominio objetivo no existan en el dominio auxiliar en donde se va a calcular un modelo colaborativo puro, lo que conduce a una URM (construida con los datos del dominio auxiliar) solo parcialmente completa. En particular, todas las columnas de \mathbf{R} en relación con artículos que existen en el dominio objetivo pero que no existen en el auxiliar estarán sin rellenar.

35 Por otro lado, todos los artículos que existen en el dominio auxiliar pero que no existen en el dominio objetivo se pueden mantener opcionalmente en la URM. De hecho, a pesar de que no se recomendarán estos, pueden aprovechar relaciones indirectas entre valoraciones.

40 En lo sucesivo se proporcionan dos técnicas (colaborativa + de contenido) híbridas adaptadas para sistemas de recomendación de radiodifusión: una primera denominada aumento de características con filtrado unidimensional (OWFFA) y una segunda denominada Knn de inyección de similitud unidireccional (OWSIKNN). La idea principal tras la primera técnica híbrida es usar filtrado basado en contenido para aumentar las valoraciones existentes usadas para el entrenamiento del algoritmo colaborativo. Por otro lado, la segunda técnica híbrida fusiona entre sí las similitudes de artículo a artículo calculadas con filtrado colaborativo y basado en contenido.

45 Las dos técnicas híbridas se han desarrollado con el fin de: (i) funcionar con conjuntos de datos tanto implícitos como explícitos, (ii) conceder recomendaciones buenas incluso cuando solo está disponible información de contenido (por ejemplo, artículos en el dominio objetivo que no existen en el dominio auxiliar), (iii) tener una complejidad en línea comparable con la de las técnicas no híbridas, (iv) no tener parámetro específico de usuario alguno. La totalidad de las propiedades anteriores conceden el uso de tales técnicas en STB livianos que operan en canales unidireccionales.

50 *Aumento de características con filtrado unidimensional (OWFFA)*

OWFFA es un método de aumento de características. A diferencia de otras técnicas híbridas, el enfoque de OWFFA no necesita que se aprenda parámetro específico de usuario alguno.

OWFFA calcula patrones de comportamiento de usuario en función de las valoraciones aumentadas con pseudo-valoraciones obtenidas a partir del filtrado basado en contenido, pero usa los perfiles de usuario originales para predecir valoraciones.

60 La figura 2 muestra un diagrama de flujo que explica el proceso de aprendizaje:

1. En el dominio objetivo, el módulo de filtrado basado en contenido, CBF, está compuesto por un algoritmo basado en contenido que recibe como entrada los datos de contenido disponibles \mathbf{W} , es decir, la matriz de contenidos de artículo ICM. El Proveedor T entrena el algoritmo basado en contenido con \mathbf{W} y obtiene el modelo basado en contenido, que se indica por \mathbf{L} . Tal modelo se comparte con el proveedor A en el dominio auxiliar A.
2. Para cada usuario en el dominio auxiliar, el proveedor A usa \mathbf{L} en el módulo de recomendación de filtrado

basado en contenido, CBFrec, con el fin de predecir todas sus valoraciones desconocidas. Se hace referencia a estas valoraciones adicionales como pseudo-valoraciones.

3. Parte de las pseudo-valoraciones se añade a la matriz de valoraciones de usuario URM original (construida con datos de dominio auxiliar). El componente de filtro se encuentra a cargo de la selección del subconjunto de pseudo-valoraciones que se van a insertar en la matriz **R**. La URM incrementada con las pseudo-valoraciones seleccionadas por el filtro se indica por URM aumentada (a saber, aURM).

4. La URM aumentada (aURM) se usa como entrada para un algoritmo de filtrado colaborativo, implementado en el módulo de filtrado colaborativo CF. El Proveedor A ejecuta el algoritmo en el módulo de CF y obtiene un modelo **Q**, que se comparte con el proveedor T. Tal modelo se enviará al STB por medio del módulo de radiodifusión y se usará para calcular recomendaciones.

Con detalle, el módulo de CBFrec calcula recomendaciones basadas en contenido para los usuarios en el dominio auxiliar. Se supone que, para cada artículo, tal información se representa en un vector de BOW (bolsa de palabras), que desatiende la "estructura semántica" de la descripción del artículo (por ejemplo, el "significado" de los términos y sus relaciones). Todos los vectores de contenidos de artículo se almacenan de forma compacta en una matriz, a la que se hace referencia como ICM (matriz de contenidos de artículo), indicada por **W**, cuya *i*-ésima columna se corresponde con el vector de BOW del artículo *i*. El elemento w_{xi} representa la importancia de la característica de contenido *x* para el artículo *i*. Como ejemplo, w_{xi} se puede establecer a 1 si la característica *x* existe en el artículo *i* y a 0 si tal característica no está presente. Los datos de contenido están directamente disponibles en el dominio objetivo, por ejemplo en forma de EPG (guía de programación electrónica). Se supone que **W** tiene una dimensión $w \times m$, en donde *w* es el número de características de contenido y *m* el número de artículos en el catálogo del dominio objetivo. La matriz **W** se ha de compartir con el proveedor del dominio auxiliar con el fin de enriquecer los datos. Las siguientes etapas son realizadas por el proveedor del dominio auxiliar con el fin de obtener un modelo de recomendación.

La correlación entre las BOW del usuario y el artículo está relacionada con el interés del usuario por el artículo. El módulo de CBF se basa en LSA (análisis de semántica latente), una técnica ampliamente usada en la recuperación de información. LSA usa SVD (descomposición en valores singulares) para representar los vectores de BOW en un espacio de pocas dimensiones, en donde se atenúa el ruido de los datos y se refuerzan las correlaciones entre características de contenido relacionadas.

Dada la ICM como entrada, usando LSA el proveedor del dominio objetivo calcula la matriz **L**, que representa artículos en un espacio de pocas dimensiones. **L** es una matriz $k \times m$ que modela las características latentes de los artículos. El artículo *i*-ésimo se representa en un espacio de características de *k*-dimensiones mediante la *i*-ésima columna de **L**. La matriz **L** se calcula tal como sigue. Dada **W** ($w \times m$), por la totalidad de la descomposición SVD esta se puede aproximar como el producto $\mathbf{B} \cdot \mathbf{G} \cdot \mathbf{V}^*$, en donde **G** es una matriz diagonal $k \times k$ formada por los *k* valores singulares más altos (en orden decreciente de magnitud), **B** es una matriz $w \times k$ de tal modo que $\mathbf{B}^* \mathbf{B} = \mathbf{I}_k$, y **V** es una matriz $m \times k$ de tal modo que $\mathbf{V}^* \mathbf{V} = \mathbf{I}_k$, con \mathbf{I}_k representando la matriz identidad $k \times k$. Por último, se define $\mathbf{L} := \text{norma}(\mathbf{G} \cdot \mathbf{V}^*)$, en donde cada columna del producto $\mathbf{G} \cdot \mathbf{V}^*$ se normaliza mediante la norma euclídea.

La matriz **L** se comparte con el proveedor del dominio auxiliar que calcula la pseudo-valoración de cada uno de los valores de valoración desconocida de sus usuarios. La pseudo-valoración p_{ei} en relación con el usuario *e* acerca del artículo *i* se calcula usando la fórmula:

$$p_{ei} = \text{norma}(\mathbf{r}_e \cdot \mathbf{L}^*) \cdot \mathbf{I}_i$$

en donde \mathbf{I}_i indica la *i*-ésima columna de la matriz **L**.

Una vez que se han calculado pseudo-valoraciones, el filtro selecciona solo las más relevantes para añadirse a la URM. Se implementan dos filtros diferentes: FFAt y FFAg. El primero (FFAt) excluye todas las pseudo-valoraciones cuyo valor es menor que un umbral fijado. Por otro lado, el último usa la medida de impureza de Gini con el fin de añadir pseudo-valoraciones tanto altas como bajas para aumentar la información intrínseca a los perfiles de artículo (FFAg). Se preferirá FFAg en el caso en el que la mayor parte de las valoraciones recibidas por el artículo son explícitas, mientras que se preferirá FFAt en el caso en el que la mayor parte de las valoraciones recibidas por el artículo son implícitas.

Para cada artículo *i*, el valor de impureza de Gini g_i se calcula como una función que depende del número de valoraciones recopiladas para tal artículo - indicado por β . La función se define de tal modo que: (i) cada artículo tiene al menos β_{\min} valoraciones, (ii) las valoraciones de usuario conocidas tienen prioridad sobre las pseudo-valoraciones, y (iii) también los artículos con muchas valoraciones se potencian con un número de pseudo-valoraciones. La fórmula para g_i es:

$$g_i = \beta_{\min} - \beta + (h \cdot \beta) \quad \text{si } \beta \leq \beta_{\min}$$

$$(h \cdot \beta^2) / \beta \quad \text{de lo contrario}$$

en donde h es un parámetro adicional que se puede optimizar a través de validación cruzada (los inventores de la presente invención han experimentado con $h = 0,3$). Además, β_{\min} se establece igual al número promedio de valoraciones por artículo.

- 5 Una vez que se ha construido la URM aumentada, el algoritmo colaborativo SVD unidireccional se puede usar con el fin de calcular un modelo transportable. En particular, el algoritmo SVD unidireccional se usa en donde la matriz de entrada \mathbf{R} es la URM aumentada. El modelo de salida es la matriz \mathbf{Q} que se envía al STB con el fin de usarse en tiempo real para realizar recomendaciones al usuario. Los requisitos de almacenamiento de \mathbf{Q} son los mismos que los analizados para la OWSVD, es decir, para cada artículo, se requiere almacenar f características, por ejemplo, f datos en coma flotante de 32 bits por artículo.

El STB radiodifunde el modelo \mathbf{Q} que permite predecir la puntuación del usuario u para el artículo i con el siguiente cálculo en dos etapas:

$$\overline{r_u} = \mathbf{r}_u \cdot \mathbf{Q} \quad (\text{etapa 1})$$

$$s_{ui} = b_{ui} + \overline{r_u} \cdot \mathbf{q}_i^* \quad (\text{etapa 2})$$

- 15 La complejidad de este cálculo en dos etapas es la misma que la de la OWSVD, de tal modo que este puede ser calculado por cualquier STB. Por último, los artículos con las puntuaciones más altas son seleccionados por el STB y recomendados al usuario.

20 *Knn de inyección de similitud unidireccional (OWSIKNN)*

OWSIKNN construye un modelo usando similitudes de artículo a artículo obtenidas mediante una técnica colaborativa y una basada en contenido. La idea básica es fusionar los dos datos siguientes:

- 25 Las similitudes de artículo a artículo calculadas por medio de filtrado colaborativo entrenado sobre las valoraciones del dominio auxiliar. Tales similitudes se pueden almacenar de forma compacta en una matriz $m \times m$ que se indica por \mathbf{Z}^A , cuyo elemento z_j^A representa la similitud entre el artículo i y el artículo j de acuerdo con las valoraciones disponibles en el dominio auxiliar. Tal matriz es calculada por el proveedor del dominio auxiliar y se comparte con el proveedor del dominio objetivo.

- 30 Las similitudes de artículo a artículo calculadas por medio de filtrado basado en contenido entrenado sobre los datos de contenido recuperados por la EPG (guía de programación electrónica) disponible para el proveedor del dominio objetivo. Tales similitudes se pueden almacenar de forma compacta en una matriz $m \times m$ que se indica por \mathbf{Z}^T , cuyo elemento z_{ij}^T representa la similitud entre el artículo i y el artículo j de acuerdo con la descripción de contenido disponible en el dominio objetivo. Tal matriz es calculada por el proveedor del dominio objetivo.

La figura 3 resume el proceso. Con detalle, la matriz de similitudes \mathbf{Z}^A es calculada por el módulo de CF de acuerdo con las siguientes etapas:

- 40 El Proveedor A usa la técnica de OWSVD con el fin de calcular el modelo colaborativo \mathbf{Q} sobre la URM del dominio auxiliar. El vector \mathbf{q}_i representa de nuevo la i -ésima fila de \mathbf{Q} . Para cada artículo i :

en la i -ésima columna de \mathbf{Z}^A , indicada por \mathbf{z}^A , se almacenan las similitudes de i con todos los artículos $j \in [1, m]$; la columna se calcula como $\mathbf{z}^A = \text{norma}(\mathbf{Q}) \cdot \text{norma}(\mathbf{q}_i^*)$;

- 45 se seleccionan los k valores más altos distintos de cero de \mathbf{z}^A , estableciendo el valor del elemento restante de la i -ésima columna a ceros; esta etapa se requiere con el fin de controlar y limitar el tamaño de \mathbf{Z}^A .

La matriz de similitudes \mathbf{Z}^A se comparte con el proveedor del dominio objetivo.

- 50 Por otro lado, la matriz de similitudes \mathbf{Z}^T es calculada por el proveedor T usando los datos de contenidos de artículo (es decir, la ICM) con un algoritmo basado en contenido implementado en el módulo de CBF. En particular:

usando la misma técnica basada en LSA descrita para OWFFA, la matriz \mathbf{L} se calcula usando la ICM como entrada. Para cada artículo i :

- 55 - en la i -ésima columna de \mathbf{Z}^T , indicada por \mathbf{z}^T , se almacenan las similitudes de i con todos los artículos $j \in [1, m]$; la columna se calcula como $\mathbf{z}^T = \text{norma}(\mathbf{L}^*) \cdot \text{norma}(\mathbf{l}_i)$, con \mathbf{l}_i i -ésima columna de \mathbf{L} ;
- 60 - se seleccionan los k valores más altos distintos de cero de \mathbf{z}^T , estableciendo el valor del elemento restante de la i -ésima columna a ceros; esta etapa se requiere con el fin de controlar y limitar el tamaño de \mathbf{Z}^T .

Una vez que se han calculado las dos matrices de similitudes, el proveedor del dominio objetivo las puede fusionar

en una matriz de similitudes de artículo a artículo única \mathbf{Z} mediante la adopción del siguiente proceso en dos etapas:

- para cada columna de \mathbf{Z}^A , los k elementos más altos se copian en \mathbf{Z} ; por lo tanto, cada columna se rellena con los k valores de similitud más altos que se obtienen a partir del filtrado colaborativo calculado sobre los datos del dominio auxiliar;
- para cada columna de \mathbf{Z}^T , los elementos de arriba se insertan en los elementos vacíos (por ejemplo, de valor cero) correspondientes de \mathbf{Z} de una forma tal que se tenga, para cada artículo, un total de $2k$ valores de similitud distintos de cero.

Por último, la matriz \mathbf{Z} representa el modelo de recomendación que ha de ser enviado por el módulo de radiodifusión al descodificador STB con el fin de generar recomendaciones. Los requisitos de almacenamiento del modelo de recomendación \mathbf{Z} dependen linealmente del número de artículos. Para cada artículo, se necesita almacenar k características, por ejemplo, k datos en coma flotante de 32 bits por artículo. Por ejemplo, suponiendo un catálogo de programas de TV compuesto por 10.000 artículos, se necesitan aproximadamente 8 Mbytes para almacenar el modelo de recomendación. Tal modelo de 8 Mbytes es radiodifundido regularmente por el Proveedor T a cada uno de sus descodificadores STB con el fin de actualizar el modelo de recomendación.

El STB, dado el vector de valoraciones de usuario \mathbf{r}_u con valoraciones conocidas del usuario u al que se van a realizar recomendaciones, puede calcular la puntuación para cualquier artículo no valorado i como:

$$s_{ui} = b_{ui} + \mathbf{r}_u \cdot \mathbf{z}_i^*$$

en donde \mathbf{z}_i indica la i -ésima columna de la matriz \mathbf{Z} .

La complejidad del cálculo de la puntuación depende de: (i) el número de valoraciones de usuario conocidas $|r_u|$ y (ii) el parámetro k . El k óptimo se puede calcular por medio de validación cruzada. No obstante, en las aplicaciones típicas, un valor de k en el intervalo [150, 300] conduce, en general, a recomendaciones de buena calidad. En el peor caso, el cálculo de s_{ui} requiere k_s multiplicaciones y k_s sumas, con k_s igual al mínimo entre $|r_u|$ y $2 \cdot k$. Esto hace también que OWSIKNN sea factible en cualquier STB.

Por último, los artículos con las puntuaciones más altas serán recomendados al usuario por el STB.

Tal como se puede evidenciar a partir de la descripción anteriormente notificada, el sistema y método de acuerdo con la invención permite lograr los resultados deseados.

De hecho, a través de la composición de un sistema de recomendación colaborativo de módulo doble, un modelo colaborativo también se puede aplicar para radiodifundir canales de comunicación unidireccional, con el fin de recomendar eficientemente artículos a los usuarios. El primer módulo del sistema de recomendación se ejecuta en un ordenador centralizado conectado a Internet, módulo que puede aprovechar las fuentes de datos externas para recopilar información basada en comunidad requerida para recomendaciones colaborativas, sin restricción de cálculo alguna. La información externa recopilada por el primer módulo en un dominio auxiliar, se transforma en una representación compacta de los gustos de los usuarios (el modelo) y se radiodifunde fácilmente a los descodificadores STB del sistema. El segundo módulo, que suministra la recomendación actual y final, se puede ejecutar fácilmente en el STB de cada usuario dando una recomendación fiable incluso si los usuarios son diferentes de los usuarios analizados por el primer módulo.

Las ventajas y puntos diferenciadores clave de la invención sobre la técnica anterior son:

1. Las recomendaciones se generan en el STB en función de un modelo de recomendación que es generado por un servicio de extremo posterior y se envía regularmente al STB.
2. Las recomendaciones avanzadas pueden ser generadas por los STB con independencia de su capacidad limitada en términos de cálculo y memoria.
3. El marco de trabajo permite el uso de sistemas de recomendación tanto de 'filtrado basado en contenido' como de 'filtrado colaborativo', no habiéndose explorado este último en las soluciones de la técnica anterior.
4. Las recomendaciones basándose en 'filtrado colaborativo' pueden ser generadas por el STB en función de un modelo de recomendación colaborativo enviado por el servicio de extremo posterior al STB. Tal modelo se puede o bien construir en función de los datos (por ejemplo, valoraciones, visionados de usuarios) recuperados en dominios alternativos que funcionan en canales bidireccionales, o bien incluso intercambiarse entre proveedores de contenido. Esto conduce a un mercado de modelos de recomendación potencial.
5. Los algoritmos de recomendación 'basados en contenido' se basan principalmente en los metadatos disponibles que describen los artículos. En particular, en los ajustes de programas de TV, tales metadatos se extraen habitualmente de la EPG (guía de programación electrónica), en donde por lo general la información disponible no examina el 'género' (por ejemplo, espectáculo de TV, serie de TV, boletín de noticias). Uno metadatos pobres quieren decir unas recomendaciones pobres. Por el contrario, se sabe que las soluciones colaborativas aplicadas de acuerdo con la invención son independientes de la presencia de metadatos, y que son capaces de producir recomendaciones con una calidad general más alta que los enfoques basados en contenido.

Otra ventaja de la invención está conectada con cuestiones de privacidad. Los sistemas de recomendación tradicionales recopilan y usan información de usuario delicada, lo que representa un posible problema de privacidad. El uso de la invención sobre canales de distribución unidireccional cambia significativamente este escenario debido a que las actividades y preferencias de usuario se almacenan en el lado de cliente (en el STB) y nunca se comparten en un servidor centralizado, eliminando todos los problemas de privacidad. Obsérvese que el modelo construido con datos de los usuarios del dominio A, y compartido con el proveedor T y todos los STB conectados, no contiene dato específico de usuario alguno, sino solo información agregada de los patrones de comportamiento de la base de clientes, conservando aún la privacidad.

No obstante, se entiende que la invención no se limita a las realizaciones particulares ilustradas anteriormente, que representan solo un ejemplo no limitante del alcance de la invención, sino que es posible un número de variantes, todas ellas dentro del alcance de un experto en el campo, sin apartarse del alcance de la invención tal como se define mediante las reivindicaciones adjuntas.

REIVINDICACIONES

1. Un sistema de recomendación para una red de radiodifusión unidireccional, que comprende al menos una estación de radiodifusión, un descodificador (STB), un aparato de TV y un servidor auxiliar,
 5 un canal unidireccional que conecta dicha estación de radiodifusión con dicho dispositivo descodificador (STB) de un primer conjunto de usuarios, comprendiendo dicho descodificador (STB) una memoria, una unidad de cálculo y de accionamiento para al menos rastrear la actividad de usuario, almacenar un perfil de usuario, generar y visualizar, en dicho aparato de TV, una clasificación de artículos de medios recomendados, siendo determinada dicha clasificación por dicha unidad de cálculo a través de un modelo de recomendación almacenado en dicha memoria, y que
 10 comprende un módulo de extremo posterior que se ejecuta en dicho servidor auxiliar conectado a una red bidireccional (Internet) y que pertenece a un dominio auxiliar (dominio A), aprovechando dicho módulo de extremo posterior las fuentes de datos externas para recopilar información basada en comunidad, **caracterizado por que** el sistema de recomendación comprende un módulo doble colaborativo que comprende un primer y un segundo módulo, siendo dicho primer módulo dicho módulo de extremo posterior, aprovechando las fuentes de datos
 15 externas para recopilar información basada en comunidad sobre al menos parte de dichos artículos de medios y para desarrollar un modelo colaborativo, mientras que dicho modelo de recomendación almacenado en dicha memoria del descodificador (STB) es dicho segundo módulo en forma de módulo de extremo frontal, y **por que** la estación de radiodifusión y dicho dispositivo descodificador pertenecen a un dominio objetivo (dominio T), dicho servidor auxiliar es parte de dicha red bidireccional, que comprende al menos dos canales para recopilar datos de realimentación a partir de un segundo conjunto de usuarios sobre al menos parte de dichos artículos de medios,
 20 dicho modelo colaborativo en dicho primer módulo de extremo posterior es procesado por dicho servidor auxiliar a través de técnicas colaborativas que incluyen dichos datos de realimentación recopilados a partir de dicho segundo conjunto de usuarios, dicho modelo colaborativo es transformado en una representación compacta por dicho servidor y se radiodifunde unidireccionalmente a dicho descodificador (STB) en donde dicha representación compacta va a ser usada por dicho segundo módulo de extremo frontal,
 25 en donde las preferencias de los usuarios de dicho dominio auxiliar se almacenan de forma compacta en una matriz de valoraciones de usuario R (URM) $n \times m$, que es una colección de todas las valoraciones expresadas de forma implícita o explícita por “n” usuarios registrados con un proveedor externo acerca de los “m” artículos disponibles en un catálogo del dominio auxiliar,
 30 en donde dicha matriz de valoraciones de usuario R (URM) se factoriza por medio de un modelo de descomposición en valores singulares (SVD) para descomponer la matriz de valoraciones de usuario R (URM) en un producto de matrices de menor rango para aproximarse en un espacio de características de f dimensiones mediante el producto:

$$U \bullet S \bullet Q^*$$

35 en donde “U” es una matriz ortonormal $n \times f$, Q^* es la traspuesta de “Q” que es una matriz ortonormal $m \times f$, “S” es una matriz diagonal $f \times f$ que contiene los primeros “f” valores singulares, en orden de magnitud decreciente, y en donde Q es dicha representación compacta del modelo colaborativo que se va a transferir de dicho dominio auxiliar a dicho dominio objetivo.

40 2. El sistema de recomendación de acuerdo con la reivindicación 1, en donde dicho segundo conjunto de usuarios es diferente de dicho primer conjunto de usuarios.

45 3. El sistema de recomendación de acuerdo con la reivindicación 1 o 2, en donde dicha unidad de cálculo, a través de dicho segundo módulo de extremo frontal, rastrea localmente el comportamiento del usuario de dicho primer conjunto de usuarios y lo compara con el comportamiento de los otros usuarios de dicho segundo conjunto de usuarios, con el fin de proporcionar dicha clasificación de artículos de medios recomendados.

50 4. El sistema de recomendación de acuerdo con la reivindicación 1, 2 o 3, en donde dicho primer módulo de extremo posterior se procesa a través de algoritmos de recomendación híbridos mediante la combinación de técnicas tanto colaborativas como basadas en contenido que incluyen artículos adicionales diferentes de dichos artículos de medios.

55 5. Un método de recomendación de programas de radiodifusión unidireccional que comprende proporcionar un canal unidireccional que conecta una estación de radiodifusión con un dispositivo descodificador (STB) de un primer conjunto de usuarios, en donde dicho descodificador (STB) se dispone para rastrear la actividad de usuario, almacenar un perfil de usuario, generar y visualizar, en un aparato de TV, una clasificación de artículos de medios recomendados, siendo determinada dicha clasificación por una unidad de cálculo de dicho descodificador (STB) a través de un modelo de recomendación almacenado en una memoria del mismo, **caracterizado por que** el método
 60 de recomendación comprende la etapa de operar un módulo doble colaborativo que comprende un primer y un segundo módulo, siendo dicho primer módulo dicho módulo de extremo posterior, ejecutándose dicho primer módulo de extremo posterior en un servidor auxiliar conectado a una red bidireccional (Internet) y que pertenece a un dominio auxiliar (dominio A), aprovechando dicho primer módulo de extremo posterior las fuentes de datos externas para recopilar información basada en comunidad
 65 sobre al menos parte de dichos artículos de medios y para desarrollar un modelo colaborativo, y siendo un segundo módulo de extremo frontal el segundo módulo y dicho modelo de recomendación almacenado en

dicha memoria del descodificador (STB) que pertenece a un dominio objetivo (dominio T), y **por que** dicho servidor auxiliar comprende al menos dos canales para recopilar datos de realimentación a partir de un segundo conjunto de usuarios sobre al menos parte de dichos artículos de medios, dicho modelo colaborativo en dicho primer módulo de extremo posterior es procesado por dicho servidor auxiliar a través de técnicas colaborativas que incluyen dichos datos de realimentación recopilados a partir de dicho segundo conjunto de usuarios, dicho modelo colaborativo es transformado en una representación compacta por dicho servidor auxiliar y se radiodifunde unidireccionalmente a dicho descodificador (STB) en donde dicha representación compacta va a ser usada por dicho segundo módulo de extremo frontal, en donde las preferencias de los usuarios de dicho dominio auxiliar se almacenan de forma compacta en una matriz de valoraciones de usuario R (URM) $n \times \bar{m}$, que es una colección de todas las valoraciones expresadas de forma implícita o explícita por "n" usuarios registrados con un proveedor externo acerca de los " \bar{m} " artículos disponibles en un catálogo del dominio auxiliar, y dicha matriz de valoraciones de usuario R (URM) se factoriza por medio de un modelo de descomposición en valores singulares (SVD) para descomponer la matriz de valoraciones de usuario R (URM) en un producto de matrices de menor rango aproximadas en un espacio de características de f dimensiones mediante el producto:

$$U \bullet S \bullet Q^*$$

en donde "U" es una matriz ortonormal $n \times f$, Q^* es la traspuesta de "Q" que es una matriz ortonormal $m \times f$, "S" es una matriz diagonal $f \times f$ que contiene los primeros "f" valores singulares, en orden de magnitud decreciente, y en donde Q es dicha representación compacta del modelo colaborativo que se va a transferir de dicho dominio auxiliar a dicho dominio objetivo.

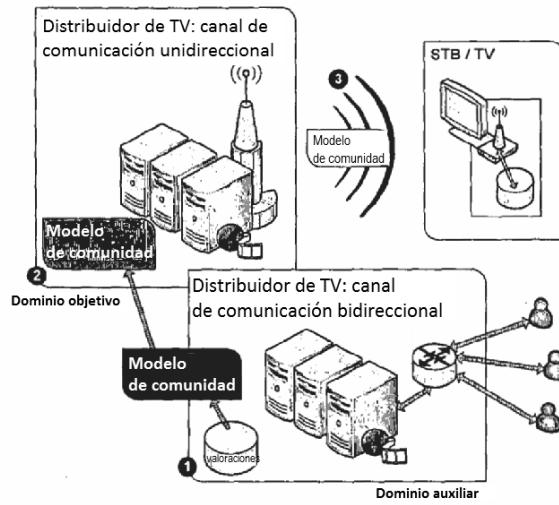


FIG. 1

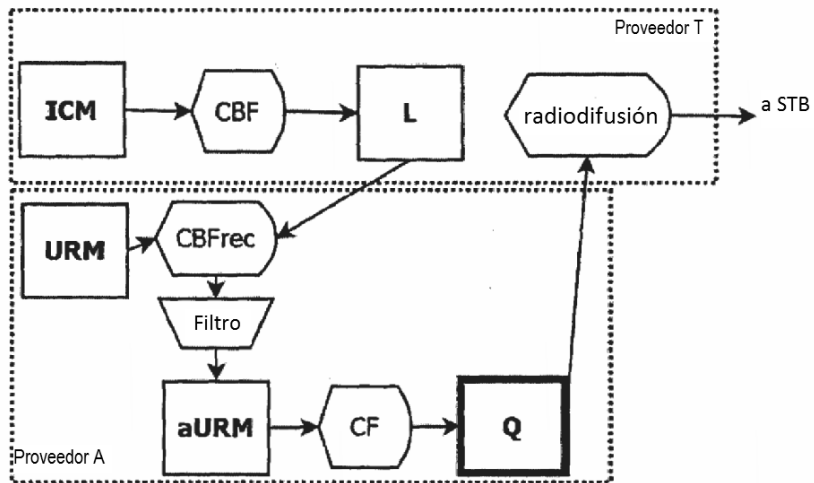


FIG. 2

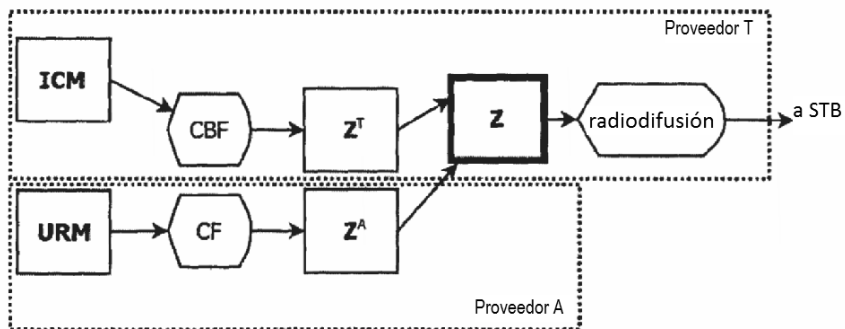


FIG. 3