



Facultad de
Comunicación y Documentación

UNIVERSIDAD DE GRANADA

GRADO EN INFORMACIÓN Y DOCUMENTACIÓN

TRABAJO FIN DE GRADO

INTRODUCCIÓN A LOS SISTEMAS DE RECOMENDACIONES

Presentado por:

D./D^a. Josefa Aguado Ferrón

Tutor:

Prof. Dr. Enrique Herrera Viedma

Curso académico 2015 / 2016

D./Dña.: Enrique Herrera Viedma, tutor/a del trabajo titulado **Introducción a los Sistemas de Recomendaciones** realizado por el alumno/a **Josefa Aguado Ferrón**, INFORMA que dicho trabajo cumple con los requisitos exigidos por el Reglamento sobre Trabajos Fin del Grado en *Información y Documentación* para su defensa.

Granada, _____ de _____ de _____

Fdo.: _____

Por la presente dejo constancia de ser el/la autor/a del trabajo titulado **Introducción a los Sistemas de Recomendaciones** que presento para la materia Trabajo Fin de Grado del Grado en Información y Documentación, tutorizado por el/la profesor/a Enrique Herrera Viedma durante el curso académico 2015- 2016.

Asumo la originalidad del trabajo y declaro que no he utilizado fuentes (tablas, textos, imágenes, medios audiovisuales, datos y software) sin citar debidamente, quedando la Facultad de Comunicación y Documentación de la Universidad de Granada exenta de toda obligación al respecto.

Autorizo a la Facultad de Comunicación y Documentación a utilizar este material para ser consultado con fines docentes dado que constituyen ejercicios académicos de uso interno.

____ / ____ / ____

Fecha

Firma

AGRADECIMIENTOS

Lejos de un agradecimiento formal y cumplido, en esta ocasión, está más que justificado el ofrecer mi gratitud a todas aquellas personas que con su ayuda me han permitido llevar a cabo esta incierta apuesta inicial de volver a la universidad.

A esta simpática generación de compañeros, en el que las redes sociales forman parte de su ADN y que conjugan el verbo compartir con absoluta normalidad. Especialmente a Esther, que permitió que alcanzara el don de la ubicuidad.

Al profesorado en general, por su flexibilidad y comprensión de mis circunstancias laborales y geográficas.

A mi tutor, Enrique Herrera, por la confianza depositada en mí, dándome la autonomía necesaria para realizar mi trabajo.

A mi marido, Juan Antonio, mi principal apoyo desde el inicio de esta aventura,
y el reanimador en los momentos de mayor desaliento.

INDICE

1. INTRODUCCIÓN.....	1
2. MOTIVACIÓN.....	3
3. METODOLOGÍA.....	3
4. HISTORIA Y SITUACIÓN ACTUAL.....	4
5. LA RECUPERACIÓN DE INFORMACIÓN Y LOS SISTEMAS DE RECOMENDACIÓN.....	8
6. SISTEMAS DE RECOMENDACIONES.....	11
6.1. Definición y usos	11
6.2. Estructura de los SRs	13
7. TIPOLOGIA DE LOS SISTEMAS DE RECOMENDACIÓN.....	15
7.1. SRs BASADOS EN EL CONTENIDO	17
7.1.1. Principales ventajas y limitaciones asociados a estos enfoques	20
7.1.2. Sobreespecialización y Serendipia	21
7.2. SISTEMAS DE RECOMENDACIONES COLABORATIVOS	24
7.2.1. <i>Sistemas colaborativos basados en la memoria</i>	25
7.2.2. <i>Filtrado colaborativo basado en modelos</i>	28
7.2.3. Principales ventajas y limitaciones del filtrado colaborativo	32
7.2.4. Cold start	34
7.3. SISTEMAS DE RECOMENDACIONES HÍBRIDOS	35
7.4. OTRAS TÉCNICAS Y CLASIFICACIONES DE SISTEMAS DE RECOMENDACIÓN	36
7.4.1. SRs basados en el conocimiento	36
7.4.2. SRs semánticos	38
7.4.3. Técnicas basadas en la teoría de los conjuntos difusos	39
7.4.4. SRs demográficos	40
7.4.5. SRs basados en redes sociales	40
8. EVALUACIÓN DE LOS SISTEMAS DE RECOMENDACIÓN	41
9. ALGUNOS EJEMPLOS DE TÉCNICAS Y SISTEMAS DE RECOMENDACIONES.....	44
10. CONCLUSIONES	45

RESUMEN

Los sistemas de recomendación tienen por objeto ayudar al usuario a seleccionar, de entre una variedad de posibles opciones, los elementos más adecuados, basándose en sus preferencias anteriores, o a través de los gustos y preferencias de usuarios similares. Desde la aparición de las primeras técnicas de filtrado colaborativo para afrontar el problema de sobrecarga de información, estas técnicas se han extendido a multitud de dominios, llegando a ocupar un papel prominente en las plataformas de comercio electrónico. Se han desarrollado diversidad de algoritmos para mejorar su precisión y rendimiento, y adoptado numerosos enfoques para solucionar los problemas asociados a estas técnicas. El objetivo de este trabajo es realizar un repaso por la literatura existente , al objeto de resaltar aquellos conceptos e ideas que pueden ayudarnos a comprender la base y funcionamiento de estos sistemas.

ABSTRACT

Recommendation Systems are intended to help users select from among a variety of options, the most appropriate items, based on their past preferences, or through the tastes and preferences of similar users. Since the appearance of the first collaborative filtering techniques to tackle the problem of information overload, these techniques have spread to many domains, taking a prominent role in e-commerce platforms. Great deal of algorithms have been developed to improve accuracy and performance, and many approaches have been adopted to solve the problems associated to these techniques. The aim of this paper is to review the literature in order to highlight those concepts and ideas than can help us to understand the basis and operation of these systems.

1. INTRODUCCIÓN

Nadie podía imaginar que Internet, una tecnología creada para uso militar en plena guerra fría en los años 60, se convertiría en tal soporte de información y comunicación que tan profundamente cambiaría nuestras vidas. Si en un principio el principal problema residía en poder contar con suficiente información para tomar una decisión, hoy contrariamente nos vemos desbordados por un exceso de la misma.

Cuando accedemos a ese gran depósito de conocimiento, la satisfacción de nuestra necesidad dependerá en gran medida de la habilidad que tengamos para expresar la misma, el manejo de herramientas adecuadas y la identificación de las mejores fuentes de información. Sin embargo, muchas veces no sabemos expresar aquello que queremos buscar o incluso ni siquiera conocemos de su existencia. Nos vemos ahogados en una multitud de posibles opciones y donde apenas disponemos de tiempo, y a veces de ganas, de valorar la opción más adecuada. Es este escenario pues, el que motiva el desarrollo de técnicas de recuperación que ayudan a los usuarios a dirigirlos hacia los elementos más adecuados en función de sus preferencias o gustos, e incluso a través de los gustos y preferencias de otros usuarios.

La tarea de recomendar se basa en intentar predecir el interés de un usuario sobre objetos que no conoce fundamentándose en las preferencias del mismo o en las preferencias o gustos de usuarios similares. Consumir contenidos que otras personas conocidas o amigas ya han visto siempre ha sido una práctica muy extendida y al igual que otras, con la llegada de Internet, se automatizan.

Aunque actualmente tenemos la sensación de que los Sistemas de Recomendaciones, en adelante SRs, tienen como fin el objetivo económico de sus anunciantes, desde su nacimiento como herramienta para afrontar la sobrecarga de información, y durante las dos últimas décadas, estas técnicas se han aplicado a multitud de dominios (películas, videos, música, compras, entornos de e-learning, e incluso amigos en las redes) y se han desarrollado diversidad de algoritmos y enfoques para mejorar su precisión y rendimiento, alimentándose de millones de preferencias explícitas e implícitas con el permiso o a veces no, de sus usuarios.

El núcleo de los SRs son sus algoritmos de filtrado, alcanzando desde la simple concordancia de palabras clave hasta la incorporación de una amplia variedad de técnicas de

inteligencia artificial. Aunque existen varios paradigmas de diseño, tradicionalmente los SRs se han agrupado en dos categorías básicas: los *basados en contenido* y los *colaborativos*.

Los sistemas basados en contenido se basan en la elaboración del perfil del usuario a través de las características de los objetos preferidos anteriormente por ellos y la utilización de éstas para recomendarle aquellos que son muy similares en su contenido. Por otra parte, los sistemas colaborativos se basan en la utilización de las valoraciones de otros usuarios para filtrar y recomendar elementos a un usuario basada en la similitud de éste con aquellos.

Mientras que las cuestiones básicas a las que deben responder los sistemas basados en contenido se centran en las técnicas a utilizar para extraer o aprender automáticamente las características o atributos de los ítems, las métricas usadas para determinar la similitud entre cada par de usuarios se convierten en el aspecto más significativo de los sistemas de filtrado colaborativos.

Cada enfoque de recomendación tiene sus ventajas y limitaciones. Mientras el principal problema de los SRs basados en contenido es la sobreespecialización de sus recomendaciones, la escalabilidad, dispersión o arranque en frío, constituyen los principales problemas de los sistemas colaborativos. Para solventar estos problemas se han propuesto diferentes alternativas, desde los sistemas híbridos (la combinación de ambos enfoques para aliviar los problemas que cada uno presenta por separado), hasta técnicas más avanzadas procedentes de la inteligencia artificial para la construcción de modelos de recomendación, como el clasificador bayesiano, redes neuronales artificiales, algoritmos genéticos o los enfoques basados en la teoría de los conjuntos difusos.

Sin embargo, con la explosión de la web colaborativa, el problema de la recomendación se redefine en un panorama dominado por las redes sociales y la generación de contenidos por parte de los usuarios. A ello le añadimos el desarrollo de dispositivos móviles como los smartphones y las tablets que están haciendo que se conviertan en uno de los medios más importantes para el entretenimiento y la adquisición de información. La navegación por la web y el uso de distintas aplicaciones se graban en los registros de nuestros dispositivos y pueden ser utilizados para la extracción de preferencias personales.

2. MOTIVACIÓN

Antes de la popularización del uso de Internet, cuando necesitábamos información para tomar cualquier decisión, la práctica habitual era la de pedir consejo a conocidos, amigos o expertos, y la decisión final, siempre acababa adoptándose sobre la afinidad de nuestros gustos con los de la persona que nos asesoraba. La sobrecarga de posibles opciones que nos ofrece la red, ha llevado al desarrollo de técnicas que automatizan esa práctica habitual, permitiéndonos casi sin apercibirnos de ello, el recibir consejo sin pedirlo.

Cuando una práctica se convierte en habitual, tenemos la sensación de que siempre ha existido, o al menos es lo que percibimos en el uso de estas técnicas por parte de la generación actual. Sin embargo, la edad, me permite evocar el recuerdo de aquel primer contacto, *no consciente*, con estos cotidianos "consejeros virtuales". En esta iniciativa de aprendizaje y actualización, de forma azarosa en la selección del Trabajo Fin de Grado, este tema se presentó ante mí, permitiéndome explorar y entender en parte, por qué aquella pretenciosa aplicación -como la llamé años atrás- se atrevió a anticipar mis preferencias musicales, por el mero hecho de haberle solicitado una sola canción.

Es indiscutible que nos encontramos en una área multidisciplinar, y que son muchos los aspectos relacionados con estos sistemas que podrían tratarse con una menor o mayor complejidad. Es el objetivo del presente trabajo, realizar un modesto recorrido a través de la literatura existente acerca del tema, al objeto de resaltar aquellos conceptos que creo nos pueden ayudar a comprender la base y funcionamiento de estos habituales consejeros virtuales.

3. METODOLOGÍA

La mayoría de las búsquedas se han realizado en ProQuest, Science Direct, y sobre todo, en las plataformas IEEE Xplore Digital Library y ACM Digital Library, dado que es un tema que ocupa gran interés en el campo de las Ciencias de la Computación. Aunque he de resaltar los problemas asociados al idioma y la poca literatura en español, no obstante, he podido contar con el auxilio de alguna que otra tesis presentada en nuestras universidades españolas y que a modo de introducción, en el ofrecimiento de algún algoritmo, técnica o enfoque, han permitido introducirme, valga la redundancia, en este complejo mundo de algoritmos y predicciones.

4. HISTORIA Y SITUACION ACTUAL

Los sistemas de recomendaciones, SRs, pueden considerarse como el resultado de una serie de cambios en las investigaciones sobre sistemas de información.

Durante los 70's gran parte de la investigación se enfocó en las técnicas de tratamiento automático de la información. En este escenario Salton y sus colaboradores desarrollaron el modelo de espacio vectorial como parte del proyecto SMART (Salton, 1971) donde los documentos se representan como vectores en un espacio vectorial permitiendo establecer su similitud mediante el coseno de los ángulos entre los vectores que formaban ese espacio multidimensional. Mediante esta representación se podía medir la similitud entre diferentes objetos (documentos y consultas, documentos y documentos, oraciones y consultas) además de proporcionar un método de ranking de los resultados. Los objetivos de la recuperación se basaban en la búsqueda de información relevante en poco tiempo donde las medidas utilizadas para evaluar estos sistemas eran por excelencia *precisión* y *recall*.

Es en la década de los 80's cuando el uso creciente de algunas aplicaciones como los procesadores de textos o el correo electrónico causaron los primeros problemas de sobrecarga de información (Martin et al., 2011) y por lo tanto, la necesidad de nuevas iniciativas en investigación, en un entorno donde la Inteligencia Artificial ilusionaba con la automatización de la conducta humana. (Maes, 1994; citado en Martin et al., 2011) expone la necesidad del desarrollo de agentes inteligentes para reducir la carga de trabajo y sobrecarga de información utilizando la expresión "asistente personal" para ejemplarizar la colaboración entre el usuario y un agente inteligente capaz de aprender los intereses de los usuarios, sus hábitos y preferencias y poder así hacer frente a la sobrecarga de información y trabajo. La idea era de que cuanto más aprendiese el agente sobre como asistir al usuario aumentaría el conjunto de tareas que se le podrían delegar (filtrado de información, recuperación de información, gestión de correo y citas, selección de libros, películas, música...).

Los objetivos se centraron más en eliminar información no relevante que encontrar información relevante. Los investigadores llegaron a la conclusión de que aprovechando la evaluación colectiva de muchos individuos se podrían establecer criterios de prioridad y clasificación. Esta idea fue la base del primer sistema colaborativo en el Centro de Investigación de Xerox Palo Alto (Tapestry). Goldberg utiliza por primera vez el término filtrado colaborativo para describir como el sistema almacenaba las anotaciones con las

opiniones de los artículos o noticias que los usuarios habían leído para poder ser utilizado por otros usuarios que aún no la habían leído y establecer así, si esa información era relevante o no (Goldberg et al, 1992).

Uno de los primeros grupos de investigación en el desarrollo del filtrado colaborativo fue el proyecto GoupLends (aún activo) de la Universidad de Minnesota. Proporcionaron gran parte de los algoritmos de muchos sistemas de recomendaciones. Fueron los primeros en introducir el filtro colaborativo automático usando un algoritmo de búsqueda de vecinos para proporcionar predicciones en los grupos de noticias de USENET (Galan, 2007).

Paralelamente, el sistema Ringo del Instituto Tecnológico de Massachussets (MIT) hizo lo mismo con los album de música y Belcore Video Recommenders para películas. Estos primeros algoritmos fueron tan efectivos que se convirtieron en el punto de referencia y objeto de comparaciones.

La atención sobre el filtrado de información ha vivido desde entonces una tremenda expansión, la investigación en estos sistemas ha dado lugar a infinidad de algoritmos y técnicas provocando cientos de publicaciones como se iría comprobando en las conferencias anuales de la ACM (Association for Computing Machinery).

El interés por el filtrado colaborativo llevó a la acuñación del término “Sistemas de Recomendación” para describir mejor el objetivo de estos sistemas ayudando a conciliar las diferencias entre los dos enfoques basados en contenido y de colaboración (Resnick y Varian, 1997), ya que por estas fechas estos sistemas no solo se limitaban al filtrado de información y habían aparecido nuevos sistemas en el que no se utilizaban los opiniones de otros usuarios y las aplicaciones se empezaron a aplicar sobre productos u objetos basándose en las características de los mismos. Por otra parte, la recomendación incluía sugerir objetos interesantes además de filtrar únicamente objetos no deseados.

Un gran impulso para el desarrollo de estos sistemas vino también a través de la convocatoria del Premio Netflix, un concurso que premiaba con un millón de dólares para el filtrado colaborativo que mejor aproximase las preferencias de usuarios.

Aunque las raíces de estos sistemas se puedan encontrar en la ciencia cognitiva, teoría de la aproximación o recuperación de información entre otras, surgió como una área independiente de investigación cuando las investigaciones empezaron a centrarse en los

problemas de las recomendaciones que se basaban explícitamente sobre la estructura y predicción de las calificaciones (Adomavicius y Tuzhilin , 2005).

Esta capacidad de selección ha ido evolucionando, desde la simplificación del proceso de toma de decisiones mediante la identificación de alternativas que satisfagan las necesidades específicas hasta proporcionar esta información de una manera altamente personalizada. Concretamente, la aplicación de técnicas de minería de datos en el análisis de preferencias, han sido eficaces en el suministro de información personalizada al usuario (Park et al, 2012). Identificados los inconvenientes asociados a estos primeros sistemas de filtrado, como los límites en el análisis de contenido o *sobre especialización* en el caso de los basados en contenido y el de *arranque en frío, escasez y escalabilidad* en el caso de los filtrados colaborativos, durante las últimas dos décadas han ido surgiendo nuevos enfoques para resolver estos problemas y aplicarlos a situaciones del mundo real. Distintos enfoques, con sus correspondientes fortalezas y debilidades y que investigadores intentan superar mediante la combinación de algunos de ellos en los sistemas híbridos.

Durante la última década, se han abordado numerosos aspectos en relación a estos sistemas. Se ha tomado especial interés en la fortaleza de los sistemas frente a ataques (inyección de perfiles falsos, bien para aumentar la venta de un producto, para hacer daño a un competidor o bien el sabotaje en sí mismo) y el desarrollo de métricas para detectar estos ataques (Mobasher et all., 2007; Zhou et all., 2014). Otro de los mayores problemas planteados ha sido el de la privacidad ya que las personas somos cada vez más reticentes a dar nuestra información personal y exhibir nuestros gustos y preferencias. Aunque algunos sistemas permiten la participación bajo anonimato o pseudónimo, no ocurre lo mismo en los sistemas empleados en el comercio electrónico, donde los clientes están perfectamente identificados. También se ha dedicado especial atención a la necesidad de diversidad, novedad y serendipia en las listas de recomendaciones (Herlocker et all., 2004; Adamopoulos y Tuzhilin, 2015; de Gemmis, 2015; Bedi et all., 2015).

Especial relevancia cobra la necesidad de incorporar al proceso de recomendación la información contextual del escenario de decisión del usuario y la proactividad (no solo la elección del objeto a recomendar sino la decisión sobre el cómo y el cuándo mostrar esa recomendación), (Woerndl et all., 2011; Adomavicius et all., 2005; Paniello y Gorgoglione., 2012; Gallego et all., 2013).

Hasta el nacimiento de la web colaborativa, el problema habitual de las recomendaciones se basaban en la valoración (ratings) de los artículos (ítems) objeto de la recomendación y en la búsqueda del mejor sistema de correspondencia entre los usuario y dichos ítems. La llegada de la bautizada web social ha llevado a la proliferación de sitios web de colaboración en el que el número de elementos que se pueden recomendar aumentan de manera significativa. La relevancia que cobra el contenido generado por los usuarios pone de manifiesto que la localización de las mejores fuentes de información requieran sistemas recomendaciones que incorporen la naturaleza social de una web que va más allá de la primigenia Internet (Pascual et al., 2011).

A la luz de las últimas investigaciones, (Martin et all., 2011) proponen una visión de lo que conformará la nueva generación de sistemas de recomendaciones. Destacan 3 dimensiones prominentes: la *conciencia* contextual, el almacenamiento y programación en la nube y los objetivos a largo plazo del usuario. Las funciones cada vez más avanzadas de los modernos smartphones parecen ser el canal perfecto para conducir esta nueva generación de SR. Estos proporcionan una base cada vez mayor de información (calendarios, correo electrónico, coordenadas del usuario mediante GPS, detección del gesto, medición de la tensión...) y que ayudan a impulsar sugerencias contextualizadas. En este sentido vaticinan el regreso a la metáfora del “asistente personal” de Maes ya comentado anteriormente, para ponerse del lado de los usuarios y ayudarles a hacer frente al problema de sobrecarga de información en una amplitud de ámbitos. (Así lo ejemplarizan con *Siri*, el precursor del asistente personal móvil). Aunque los SRs actuales se centran más en los objetivos a corto plazo (venta cruzada o sugerencia de productos para consumo inmediato) se espera que los nuevos sistemas pongan un mayor énfasis en el manejo de los objetivos a largo plazo.

La cantidad y diversidad de datos que procesan los SRs aumentan enormemente mientras que el costo de procesar y almacenar parece disminuir. El almacenamiento y programación en la nube parece reconfigurar el paisaje, proporcionando una potencia muy inferior en su coste.

En definitiva, un escenario ideal para que investigadores y profesionales aborden nuevos desafíos para los SRs.

5. LA RECUPERACIÓN DE INFORMACIÓN Y LOS SISTEMAS DE RECOMENDACIONES

El crecimiento de forma exponencial de la información a lo largo del tiempo, ha motivado gran cantidad de líneas de investigación en torno a la recuperación de información.

Básicamente, las técnicas de recuperación empleadas en la web proceden de las empleadas en los Sistemas de Recuperación tradicionales. Como ya sabemos, la Recuperación de Información intenta resolver el problema de encontrar y presentar ordenados documentos relevantes que satisfagan las necesidades de información del usuario expresados en forma de consulta. Se encarga por lo tanto, de dar respuesta a necesidades de información puntuales que puedan tener los usuarios. Con la llegada de la web, las características de información en la red hacen que la recuperación de información haya de enfrentarse a diferentes retos motivando el desarrollo de nuevas técnicas de acceso a la información, entre ellas las técnicas de filtrado, para afrontar el grave problema de sobrecarga de información.

(Belkin y Croft, 1992) sostiene que ambos conceptos se asemejan en el hecho de que intentan proporcionar información relevante al usuario pero, los sistemas de filtrado presentan algunas características que los identifican y permiten abordarlos:

- . Estos sistemas están diseñados para datos no estructurados o semiestructurados con significados difíciles de representar.
- . Los sistemas de filtrado trabajan con grandes cantidades de datos
- . Los sistemas de FI eliminan la información irrelevante de flujos de entrada de información o reúnen información relevante de diferentes bases de datos heterogéneas de acuerdo al perfil del usuario, frente a los sistemas de IR que seleccionan los ítems de las bases de datos que coinciden en la consulta.
- . En el filtrado, la descripción de la información se realiza sobre las preferencias individuales o de grupo llamadas perfiles, representando intereses a largo plazo, en tanto que en los sistemas de RI la necesidad de información se expresa en forma de consulta y no necesita saber o tener información sobre los usuarios, carecen de modelos usuario/contexto.

Para los autores, recuperación de información y sistemas de filtrado son las caras de una misma moneda. Usando el filtrado de información se puede depurar la información

seleccionada por los sistemas de recuperación de información, de manera que la información mostrada finalmente a los usuarios se adapte lo mejor posible a sus necesidades.

Este nuevo entorno caracterizado por la sobrecarga de información nos lleva al acceso de cada vez más y más de fuentes de datos. Los usuarios se enfrentan al reto de elegir de entre miles de opciones de información que no pueden manejar directamente si no es a través de herramientas como son los buscadores (Google, Yahoo...). Por otra parte el rendimiento de estos sistemas sufren diversos problemas como:

- *Webcrawling*. Debido al tamaño, estructura y rápido crecimiento de la web, los motores de búsqueda pueden cubrir solo una pequeña parte de la totalidad de la web (Lawrence y Giles, 1998).
- *Cloacking*, constructores de páginas web que insertan en su descripción términos que nada tienen que ver con su contenido.
- *Links forms*, páginas donde puedes insertar enlaces a tus páginas web, conduciendo a una artificial popularidad de la página.
- *Spamming*, los algoritmos utilizados por los motores de búsqueda pueden ser fácilmente manipulados para promover ciertas páginas hacia la parte más alta del conjunto de resultados.

Las limitaciones que presentan estos sistemas como el lenguaje de búsqueda y comprensión de los documentos se acentúan cuando el usuario no tiene claro o no es capaz de formular correctamente su necesidad de información. A esto se añade que ante las mismas consultas tienden a responder con los mismos resultados, sin tener en cuenta el perfil del usuario ni el contexto. Plasmar en una consulta objetiva un concepto subjetivo, como es el gusto o preferencia particular de un usuario, no es nada fácil.

Es en este contexto, donde los SR se ofrecen como herramientas que buscan disminuir el esfuerzo del usuario mediante el aprendizaje de patrones que permitan predecir sus posibles elecciones sobre un conjunto de elementos sobre los que no tiene experiencia.

Es común en la literatura justificar que algunos sistemas de recomendación se deriven de la intersección de las áreas de la RI y la Inteligencia Artificial. Desde la RI la investigación sobre las técnicas de recomendación forman parte de diversas filosofías de trabajo englobadas bajo un mismo concepto, el de *Acceso a la Información* en inglés “*Information Seeking*”,

término que describe cualquier proceso que hace posible filtrar la gran cantidad de información disponible, y que el usuario únicamente acceda a información relevante para él (Porcel, 2005). En los sistemas de RI las necesidades de información se representan a través de una consulta, normalmente una serie de palabras clave, mientras que las necesidades de información en los SR se representan a través del perfil del usuario. Desde la IA el problema de la recomendación se trata como un problema de aprendizaje sobre el conocimiento de los usuarios, ofrecen técnicas para poder categorizar nuevos elementos de información basados en la utilización de información vista anteriormente por el usuario y que “etiquetó” explícita o implícitamente como interesante o no. En definitiva, la de generar un modelo predictivo (Lops et al., 2011).

La evolución de los sistemas de recomendaciones desde la recuperación de información ha llevado a un cambio en las dimensiones en sus modelos de matrices (Perugini et al., 2004): Desde sus inicios, la RI tenía como objetivo la recuperación de información relevante a las consultas a corto plazo. Los sistemas se modelaban como matrices término×documento permitiendo emparejar documentos con consultas mediante la representación de los mismos en vectores de términos ponderados y midiendo su similitud mediante el coseno del ángulo que formaban los vectores en ese espacio multidimensional.

Con el aumento de la información electrónica las investigaciones se centraron en la eliminación de información irrelevante más que en recuperar información pertinente dando lugar a los sistemas de filtrado. Los sistemas de filtrado modelan las características de los documento en perfiles de usuario sustituyéndolos así por los términos.

Estos sistemas conocidos posteriormente por la comunidad como basados en el “contenido” ya que se empezaron a utilizar para recomendar otra serie de objetos como libros y películas, se centraron en modelar las características de dichos objetos y recomendar dichos objetos a través de la consulta de tales características mediante las palabras claves o preferencias suministrada por los usuarios.

Los filtrados colaborativos, basados en la utilidad del trabajo de “los otros”, introdujeron otro cambio en las investigaciones. En lugar de calcular la similitud entre los contenidos, se calcula la similitud entre los usuarios basándose en modelos de perfiles que encapsulan sus características y preferencias. Este cambio sustituye las características con representaciones de personas en el modelo de matriz.

La llegada de la llamada web colaborativa ha llevado a la proliferación de sitios web de colaboración en el que el número de elementos que se pueden recomendar aumentan de manera significativa. El entorno es distinto y las características de los datos difieren considerablemente, el documento asume un significado más amplio en la dimensión de la matriz. La generalización y aceptación de la web colaborativa, hace que el filtrado social sea necesario para reducir la sobrecarga de información, constituyendo ahora el objetivo principal de los actuales sistemas de recomendación. El filtrado social se puede ver como una función que acepta una representación de usuarios y un conjunto universal de objetos como entrada y devuelve un subconjunto de recomendaciones de objetos como salida. Los sistemas de recomendaciones pretenden conectar grupos de individuos con intereses similares y aprovechar esa experiencia colectiva en lugar de centrarse en la búsqueda de información de un individuo específico tan característico en la RI. Este último cambio ha reemplazado los documentos por objetos en la matriz de modelado:

Concept	Modeling matrix
Information retrieval	terms \times documents
Information filtering	features \times documents
Content-based filtering	features \times artifacts
Collaborative filtering	people \times documents
Recommender systems	people \times artifacts

Fig. 1. Evolución en los modelos de matrices de los SR desde la IR (Perugini et al, 2004)

6. SISTEMAS DE RECOMENDACIONES

6.1. DEFINICIÓN Y USOS

La definición de sistema de recomendación ha evolucionado desde que aparecieron los primeros sistemas de filtrado de información. En un artículo seminal de 1997, Resnick y Varian describieron así los sistemas de recomendación:

“En un sistema de recomendación típico, las personas proporcionan información en forma de entradas, que luego el sistema agrega y dirige a destinatarios apropiados. En algunos casos la transformación primaria se encuentra en la agregación: en otros el valor del sistema reside en su capacidad para establecer conexiones acertadas entre los recomendadores y aquellos que buscan recomendaciones.”

Posteriores investigadores extendieron esta definición para incluir sistemas que sugieren elementos de interés, independientemente de cómo se producen las recomendaciones:

“Cualquier sistema que proporcione recomendaciones individualizadas como outputs o guíe al usuario de forma personalizada hacia objetos útiles o interesantes en un amplio rango de opciones posibles”, (Burke 2002).

El filtrado colaborativo encontró el éxito rápidamente en varios dominios sobre todo, en el comercio electrónico. La publicidad era casi la única manera de dar a conocer un producto. A través del análisis del comportamiento de los clientes, explícita o implícitamente, las empresas pueden analizar las tendencias y hacer predicciones para dirigir así sus ofertas y promociones. De hecho hay numerosas razones por las cuales los proveedores de servicios quieren explotar esta tecnología (Ricci et al., 2011):

.Aumentar el número de artículos vendidos. Quizás la más importante función para los SR comerciales, ser capaz de vender un conjunto adicional de artículos en comparación a los que por lo general se venden sin ningún tipo de recomendación. Tratan de aumentar el número de usuarios que aceptan la recomendación y consumen un producto en comparación con el número de visitantes que sólo navegan a través de la información.

. Venta más diversa de artículos. Permitir al usuario seleccionar elementos que podrían ser difíciles de encontrar sin una recomendación precisa. Ya que los proveedores de servicios no pueden permitirse el riesgo de publicitar artículos que no son susceptibles de satisfacer el gusto de un usuario en particular, los SR pueden sugerir artículos impopulares a los usuarios adecuados.

. Aumentar la satisfacción del usuario. La combinación de una buena interfaz de usuario y unas recomendaciones precisas aumenta el uso del sistema y la probabilidad de que se acepten las recomendaciones.

. Aumentar la fidelidad del usuario. Cuanto mayor sea la interacción del usuario con el sitio, el modelo de representación de sus preferencias se hacen más refinadas y las recomendaciones por lo tanto se harán más coincidentes con sus gustos.

. Un mayor entendimiento de lo que quiere el usuario. El conocimiento de las preferencias de los usuarios pueden ser aprovechados para la mejora de acciones y productos.

Otro uso lo podemos ver en el desarrollo de entornos de e-learning. Estos entornos caracterizados por la gran cantidad de información, una fuerte interactividad y una gran cobertura, hacen que la personalización se convierta en una de sus herramientas más importantes. Cada alumno utiliza sus herramientas, métodos y procesos y la orientación en el proceso de aprendizaje se personaliza. El SR utiliza la información sobre los alumnos y las actividades de aprendizaje y recomienda elementos tales como documentos, páginas web,

cursos y otros recursos de aprendizaje que respondan a las características e intereses pedagógicos de los alumnos (Klasnja-Milicevic et all., 2015).

Una amplia difusión la podemos observar también en los servicios de streaming de música. Las sugerencias de nuevas canciones que nos puedan gustar, radios personalizadas y en algunas ocasiones la inclusión de contexto, fidelizan a los usuarios. Hacer que el usuario descubra nueva música hace que se implique más a la hora de evaluar las canciones que escucha, repercutiendo así en la mejora de recomendaciones. Mejora nuestra satisfacción, con lo que habrá más posibilidades de que en el futuro gastemos nuestro dinero en ellos.

Libros, películas, música, hoteles, viajes, programas de TV, compras..., un sin fin de ámbitos en los que los SRs se nos ofrecen como ayuda en la complicada tarea de elegir la mejor opción.

6.2. ESTRUCTURA DE LOS SISTEMAS DE RECOMENDACIÓN

Cada sistema de recomendación se compone de 3 elementos básicos:

- Los artículos que se recomiendan (libros, música, páginas web..)
- Modelado de usuario. Perfil de preferencias de los usuarios. Este perfil se crea después de identificar sus preferencias a través de diversas técnicas.
- El algoritmo de recomendación, también llamado método o técnica de recomendación. Este componente es el mecanismo que genera la recomendación.

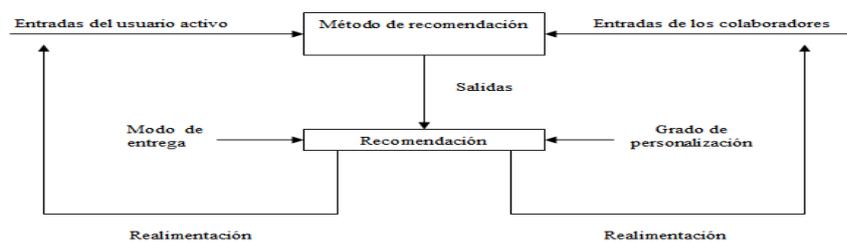


Fig. 2. Esquema del proceso de generación de una recomendación (Herrera et all., 2004)

El objetivo es sugerir nuevos elementos a un usuario basándose en sus elecciones anteriores y en elecciones de gente con similar historial de ratings o valoraciones.

Existen pues dos tipos de entrada o formas de recoger estas valoraciones y que no tienen por qué ser exclusivas:

Entradas del usuario activo. Esta información puede venir dada de dos formas:

De forma *implícita*, donde la información se obtiene a partir de la experiencia y actuaciones anteriores de los usuarios con respecto a los ítems encontrados (número de veces que escucha la canción, tiempo que pasa leyendo una determinada página, enlaces que sigue). Estos datos se infieren a través de procesos de descubrimiento del conocimiento como la minería de datos, monitoreo de navegación, etc (Herlocker et al., 2004). La mayor ventaja de estas entradas radica en su mayor disponibilidad ya que el sistema puede obtener información sobre sus gustos o preferencias sin necesidad de solicitarlas en ningún momento (Lops et al., 2011). En la mayoría de casos el usuario no es consciente de este seguimiento.

De forma *explícita*, donde el usuario sí expresa de manera explícita sus preferencias al sistema. A través de los famosos pulgares hacia arriba o hacia abajo los artículos se clasifican en relevantes o no relevantes (puntuaciones binarias), las calificaciones a través de una escala numérica, o comentarios de texto sobre un ítem en particular que se recogen y se presentan al usuario como medio para facilitar la toma de decisiones. Aunque ofrece la ventaja de la simplicidad, puede no estar adecuado para la captura de los sentimientos del usuario sobre los elementos (Lops et al., 2011).

Método o Técnicas de Recomendación. El modelo de recomendación utilizado debe identificar los elementos útiles para el usuario. Para ello debe “predecir” su utilidad (estima que valoración daría el usuario a cada elemento) y los compara para decidir y extraer a continuación los elementos más recomendables. Estas predicciones se calculan con algoritmos específicos y el uso de diferentes fuentes de conocimiento sobre usuarios e ítems que se desarrollaran más adelante.

La salida del sistema estará constituido por las recomendaciones generadas por el sistema que variarán dependiendo del tipo, cantidad y formato de la información proporcionada al usuario. Las formas más elementales de presentar la salida son:

Sugerencia o lista de sugerencias al usuario de una serie de ítems. El enfoque habitual en este caso requiere que en esta lista solo se incluyan solo los ítems que el usuario activo no ha comprado, visto o puntuado.

Presentar al usuario predicciones del grado de satisfacción que se asignará al ítem concreto. Estas predicciones pueden ser presentadas como personalizadas al usuario o como estimaciones generales del conjunto de colaboradores.

7. TIPOLOGIA DE LOS SISTEMAS DE RECOMENDACIONES

Aunque actualmente existen varios paradigmas de diseño (basados en los datos que se van a utilizar, la información proporcionada por el usuario o las características del dominio (Jannach et al., 2011), tradicionalmente según los enfoques o la forma de predecir los ítems que le puedan interesar a los usuarios, los sistemas de recomendaciones se han agrupado en 3 categorías básicas: los basados en el contenido, los colaborativos y los sistemas de recomendaciones híbridos, (Adomavicius y Tuzhilin, 2005). No obstante, es común en la literatura encontrarnos también clasificaciones que incluyen a los basados en el conocimiento, semánticos, en utilidad y demográficos (Burke, 2002). Otra clasificación más reciente la encontramos en (Lu et al., 2015) donde divide a las técnicas de recomendación en técnicas tradicionales (métodos colaborativos, basados en el contenido, basados en el conocimiento e híbridos) y los métodos avanzados desarrollados recientemente, como los basados en los conjuntos difusos, los basados en la red social, en el conocimiento del contexto, los basados en la confianza y los basados en grupos.

Ya que las técnicas de recomendaciones se pueden considerar como un conjunto de fuentes de conocimiento y enfoques algorítmicos que utilizan dichas fuentes para generar recomendaciones, sería conveniente tener una visión general sobre los distintos tipos de fuentes de conocimiento para entender los distintos enfoques que utilizan estos sistemas y las funciones que pueden realizar.

Según (Felfernig y Burke, 2008), los conocimientos necesarios para generar recomendaciones se pueden extraer de 4 fuentes: del usuario en sí, a partir de otros usuarios del sistema, a partir de datos de los ítems recomendados y del dominio de la recomendación y el conocimiento sobre cómo se usan los ítems recomendados y las necesidades que satisfacen. Para ello establecen la siguiente clasificación:



Fig.3. Fuentes de conocimiento (Felfernig y Burke, 2008)

Social (Colaborativo). El conocimiento social o colaborativo es el conocimiento sobre otros usuarios. Normalmente son perfiles de calificaciones numéricas (ratings), aunque hay otros tipos de conocimiento que se pueden considerar como colaborativos como los datos demográficos que también se utilizan para extraer semejanzas entre usuarios, las etiquetas sociales o los links de las web que pueden tratarse como votos colaborativos en el algoritmo PageRank.

Individual (Usuario). Como no podía ser de otra forma, para hacer recomendaciones personalizadas, necesitamos tener conocimiento individual sobre el usuario. Aunque muchas veces el conocimiento histórico sobre las preferencias del usuario podría bastar para realizar las recomendaciones, en muchas ocasiones el usuario utilizará el sistema con una intención particular y el sistema deberá responderle. Estos requisitos del usuario pueden venir de varias formas:

- . *Query.* Dependiendo de la interfaz de usuario éste puede introducir una query o el interfaz puede mostrar las características o atributos de los requisitos que deben ser especificados.

- . *Restricciones.* En algunos dominios hay muchos tipos de restricciones que debe cumplir la solución recomendada.

- . *Contexto.* Circunstancias exteriores asociadas con la recomendación o la situación del usuario (p.ej. la ubicación del usuario podría ser un factor importante en la recomendación de un restaurante).

Contenido. Aquí se incluye cualquier tipo de conocimiento que no ha sido generado por el usuario. El área de conocimiento de los contenidos es muy amplio. La taxonomía distingue cuatro tipos diferentes de conocimiento del contenido:

- . *Atributos de los ítems.* Obviamente es punto de partida para cualquier tipo de razonamiento sobre los ítems debe ser el conocimiento sobre esos ítems en sí. Este podría ser desde un par de valores de atributos asociados al ítem en una base de datos o atributos estructurados en el caso de ítems complejos

- . *Contexto.* Como señalábamos antes, la localización por ejemplo, puede ser una señal contextual en los SR de móviles.

. *Conocimiento del dominio*. Un SR puede necesitar un conocimiento más amplio sobre los productos que recomendar y los usos que puede tener. Este conocimiento viene dado en una variedad de formas:

- *Conocimiento del medio y del fin*, es quizás el componente más complejo del conocimiento del contenido. Es el conocimiento inferido que permite al sistema razonar sobre como un ítem en particular puede satisfacer las necesidades o requerimientos del usuario
- *Ontologías*. Los atributos de un ítem puede mostrar al recomendador las características que están asociadas a ese ítem pero muchas veces tendrá que saber la relación entre esas características.
- *Limitaciones del dominio*. Los usuarios tienen límites sobre los requisitos que deben cumplir los ítems recomendados y los ítems también pueden tener restricciones que rigen su idoneidad en diferentes circunstancias.

(Burke y Ramezani, 2011), ilustran la conjunción entre las distintas fuentes del conocimiento y las distintas técnicas que las atienden:

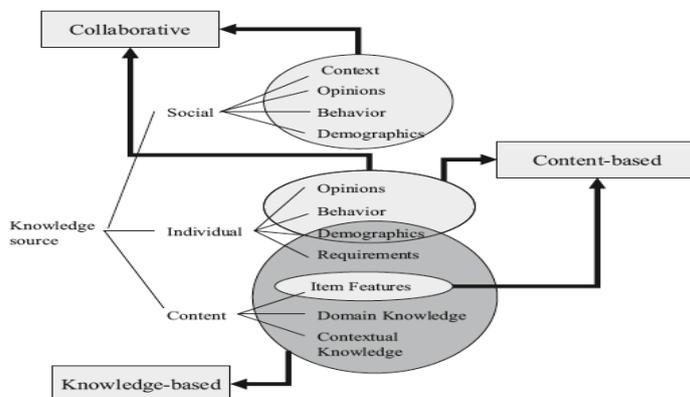


Fig 4. Fuentes de conocimiento y técnicas de recomendaciones (Burke y Ramezani, 2011)

7.1. SISTEMAS DE RECOMENDACIONES BASADOS EN EL CONTENIDO

Estos sistemas se basan en la elaboración del perfil del usuario a través de las características de los objetos preferidos anteriormente por ellos y la utilización de éstas para recomendarle

aquellos que son muy similares en su contenido (Lops et al., 2011; Adomavicius y Tuzhilin, 2005). En este tipo de sistemas, la previsión de recomendaciones personalizadas requiere que el sistema contenga conocimiento sobre las preferencias del usuario, por lo tanto la recomendación se basará en la habilidad del sistema para extraer la descripción de los ítems y un perfil que asigne la importancia de esas características. En este contexto, los sistemas de recomendaciones deben responder a dos cuestiones básicas:

- Cómo puede el sistema adquirir y mejorar continuamente los perfiles de usuarios automáticamente, y
- Qué técnicas se deben usar para extraer o aprender automáticamente las características o atributos de los ítems

El proceso de recomendación se realiza en 3 etapas llevadas a cabo por 3 componentes distintos (Lops et al., 2011):

. **Content Analyzer.** La responsabilidad principal de este componente es la de representar el contenido de los ítems procedentes de las fuentes de información de una manera adecuada para que puedan ser procesados por el siguiente componente. Normalmente toman prestadas técnicas de RI, técnicas de extracción para representar las características en forma de texto (términos, frases, N-gramas), e incluso de forma no textual como información de diseño o etiquetas XML.

. **Profile Learner.** Quizás el componente principal del sistema ya que consiste en el proceso de modelado del usuario, los intereses del usuario se infieren a través de la interacción del usuario con los elementos. Este modulo recoge los datos representativos de las preferencias de los usuarios (los objetos que el usuario utilizó, descargó... o compró en el pasado) y trata de generalizarlos para construir el perfil del usuario. Esta generalización se realiza a través de técnicas de aprendizaje automático capaces de inferir un modelo de intereses del usuario a partir de los objetos que fueron de su gusto o por los que sintió aversión.

. **Filtering Component.** Este modulo aprovecha el perfil de usuario para realizar las recomendaciones comparando dicho perfil y los elementos a recomendar mediante métricas de similitud dando como resultado una lista clasificada de artículos predeciblemente interesantes para el usuario.

La mayoría de los SRs basados en el contenido utilizan los modelos tradicionales de recuperación, como la concordancia de palabras clave o el Modelo de Espacio Vectorial donde los documentos pueden ser codificados como vectores en un espacio multidimensional euclídeo y donde las dimensiones del espacio se corresponden con las palabras clave. La representación de los documentos en este modelo de espacio vectorial plantea dos cuestiones: la ponderación de los términos y la medida de similitud entre los vectores. La técnica más usada para la ponderación de los términos es la llamada TF*IDF *Term Frequency-Inverse Document Frequency* (Salton, 1989), basada en observaciones empíricas de los textos. Esta técnica de ponderación se apoya en que los términos que ocurren con frecuencia en un solo documento (TF) pero rara vez en el resto del corpus (IDF) probablemente sean más relevantes para el tema del documento. Además la normalización del resultado en los vectores impiden que los documentos más largos tengan una mayor oportunidad de ser recuperados. Estos supuestos están bien ejemplificados con la función TF*IDF:

$$|TF-IDF(t_k, d_j) = \underbrace{TF(t_k, d_j)}_{TF} \cdot \underbrace{\log \frac{N}{n_k}}_{IDF}$$

donde N es el número de documentos en el corpus, y n_k el número de documentos en la colección en el que el término t_k ocurre al menos una vez

$$TF(t_k, d_j) = \frac{f_{k,j}}{\max_z f_{z,j}}$$

donde se calcula el máximo de las frecuencias $f_{z,j}$ de todos los términos t_z que ocurren en el documento d_j .

Para normalizar la longitud de los documentos, se suele utilizar la normalización del coseno:

$$w_{k,j} = \frac{TF-IDF(t_k, d_j)}{\sqrt{\sum_{s=1}^{|T|} TF-IDF(t_s, d_j)^2}}$$

En estos sistemas tanto los perfiles de usuarios como los ítems se representan como vectores de términos ponderados. Los perfiles de usuarios se obtienen mediante el análisis del contenido de los ítems vistos y calificados previamente por el usuario utilizando técnicas de análisis de palabras clave procedentes de la RI. El perfil se define también como un vector de

pesos donde cada peso denota la importancia de la palabra clave al usuario. La predicción de interés se calculará mediante alguna medida de similitud que permita determinar su proximidad. Entre las más variadas(Coeficiente de Dice, coseno, Jaccard) la más utilizada es la similitud del Coseno:

$$sim(d_i, d_j) = \frac{\sum_k w_{ki} \cdot w_{kj}}{\sqrt{\sum_k w_{ki}^2} \cdot \sqrt{\sum_k w_{kj}^2}}$$

Al igual que los filtrados colaborativos, los filtrados basados en contenido no sólo se apoyan en métodos heurísticos tradicionales, también se basan en enfoques basados en modelos. No calculan la predicción de utilidad mediante fórmulas heurísticas como el coseno sino que se apoyan en métodos de aprendizaje automático. Desarrollan primero un modelo de los ratings del usuario y tratan el problema como un problema de predicción estadística y calculan el valor esperado para cada ítem en función de los ratings anteriores. El problema del aprendizaje de los perfiles de usuario se lanza como un problema de categorización, cada documento tiene que ser clasificado como interesante o no con respecto a las preferencias del usuario. Para ello se utilizan diversos algoritmos de aprendizaje como pueden ser los clasificadores bayesianos, el clustering o los árboles de decisión entre otras.

7.1.1. Principales ventajas y limitaciones asociadas a estos enfoques

Los SR basados en contenidos presentan ciertas ventajas frente a los filtrados colaborativos entre las que destacamos las siguientes:

- . Puede determinar las mejores recomendaciones para cada usuario individualmente en lugar de estar limitados por estereotipos (Beel et al., 2015), explotan únicamente las calificaciones proporcionadas por el usuario activo, no necesitan las calificaciones de otros usuarios con gustos similares.
- . Transparencia en el sentido de que pueden proporcionar explicaciones de cómo funciona el sistema mediante una lista explícita del contenido o las características que causan que el ítem sea recomendado.
- . No padece en exceso el problema del arranque en frío ya que cuando se añade un nuevo ítem se expresan también sus propiedades.

. No tiene el problema de escasez (*sparsity*) pues el modelado de la información está presente en las características del ítem y no necesitan proveerlas de otros usuarios.

Entre las limitaciones asociadas a este enfoque podemos mencionar:

. Límite en cuanto al número y tipos de características que se pueden asociar a los ítems, ya sea manual o automáticamente. Ciertas representaciones solo capturan ciertos aspectos del contenido y hay otros aspectos que si podrían influir al usuario tales como calidades estéticas o información adicional multimedia. Aunque las técnicas de IR trabajan bien en la extracción automática de características de documentos textuales se hace más difícil en otros dominios como en datos multimedia, imágenes, audios o videos.

.Requiere más costo computacional ya que cada elemento debe analizarse en base a sus características, se debe construir el modelo de usuario y se deben realizar los cálculos de similitud. Si hay muchos usuarios y muchos objetos los cálculos requieren muchos recursos.

. Ignoran la calidad y popularidad de los elementos. Si los documentos están representados por sus palabras clave más importantes, los sistemas basados en contenido no pueden distinguir la calidad entre un buen o mal artículo si ambos utilizan los mismos términos (Shardanand y Maes, 1995; citado en Adomavicius y Tuzhilin, 2005).

.Nuevo usuario. Estos sistemas requieren una cantidad significativa de información sobre los usuarios para entender sus preferencias e inferir recomendaciones precisas (Albadvi y Shahbazi, 2009). Los usuarios no suelen sentirse cómodos proporcionando información personal sobre sus preferencias debido a cuestiones de privacidad, paciencia, voluntad o bien porque no disponen de tiempo para ello. Esta falta de información provoca muchas veces que estos sistemas no puedan proporcionar recomendaciones exitosas derivando en una desconfianza del usuario sobre la fiabilidad de las mismas. Pero sin embargo, de una forma paradógica, si se extraen numerosas características de las preferencias de los usuarios, sería difícil entender la principal (Rajabpour et al., 2014).

Pero sin duda alguna, su gran debilidad es la sobre especialización y baja serendipia, (Adomavicius y Tuzhilin, 2005), ya que no tiene en cuenta la posible arbitrariedad en los gustos e intereses de los usuarios (Lops et al., 2011), y que por ello tratamos en un epígrafe aparte.

7.1.2. SOBRESPECIALIZACIÓN Y SERENDIPIA.

Es enteramente asumido el hecho de que los sistemas basados en contenido sufren de sobreespecialización, ya que solo recomienda ítems que son muy similares a aquellos que ya fueron puntuados o vistos por los usuarios en el pasado. El objetivo de estos sistemas es encontrar ítems que más se asemejen a las preferencia del usuario con el fin de mejorar la precisión, independientemente de la utilidad real de las recomendaciones. El problema no solo se reduce a que estos sistemas no puedan recomendar ítems diferentes de los puntuados anteriormente sino que no deberían recomendar ítems que son demasiado similares a los ya conocidos por los usuarios (Lops et al., 2011).

Interesantes reflexiones de Parisier, y Gori y Witten, referenciadas en (de Gemmis et al., 2015) sobre la opacidad de los sistemas de recomendaciones. Parisier argumenta que los filtrados personalizados crean universos de información individuales para nosotros y que nos alimentamos de información que nos es familiar y confirma nuestras creencias. Resalta la opacidad de estos filtros en el sentido de que privan de encuentros casuales que encienden la chispa de la creatividad, la innovación y el intercambio democrático de ideas. Por otra parte, Gori y Witten utilizan la metáfora de *Dragones guardianes del tesoro* para los motores de búsqueda, justificándolo por el hecho de que el gran tesoro que guardan es el depósito del conocimiento de la sociedad y que todos aceptamos a esos dragones como intermediarios para acceder a dicho tesoro. En este sentido, de Gemmis los asemeja con los SRs ya que dichos sistemas analizan el comportamiento de un usuario y usan esa información para sugerir ítems. El usuario estaría provisto de ítems dentro de su gama actual de intereses y su tendencia a un determinado comportamiento se reforzaría por la creación de un bucle autorreferencial.

La *serendipia* en un sistema de recomendaciones puede verse como la experiencia de recibir un ítem inesperado y fortuito.

Es común en la literatura la necesidad de distinción entre novedad, diversidad y serendipia. (Herlocker et al, 2004), (Adomapiolius y Tuzhilin, 2015) señalan que la novedad se produce cuando el sistema sugiere al usuario un elemento desconocido que éste podría haber descubierto de una forma autónoma. La diversidad representa la diversificación en la lista de recomendaciones, utilizado generalmente para evitar listas homogéneas en la que todos los artículos son muy similares entre sí. La serendipia o recomendaciones casuales ayudan al usuario a encontrar sorprendentemente ítems interesantes que tal vez no hubiese descubierto por sí mismo. La diversidad no implica serendipia ya que la diversidad de ítems

podrían caer dentro del rango de las preferencias del usuario, pero ambas se relacionan en el sentido de que proporcionando una lista diversa se puede facilitar lo inesperado.

Como justifica (Toms, 2000; citado en de Gemmis et al., 2015), hay 3 tipos de búsqueda de información: buscar información sobre un objeto bien definido, buscar información sobre un objeto que no puede ser totalmente descrito pero que lo reconoceríamos a primera vista y, la adquisición de información de manera casual o accidental. Aunque la serendipia es bastante inútil para las dos primeras, es bastante importante para la tercera.

Aunque hay pocos estudios sobre el tema, los investigadores hacen hincapié en la importancia de tener en cuenta factores distintos de la precisión que contribuyan a una mayor calidad en las recomendaciones. Diversificación para maximizar la variedad de ítems en las listas de recomendaciones, definiciones de lo inesperado, métricas específicas para medir tanto lo inesperado como las recomendaciones, algoritmos para generar este tipo de recomendaciones, la integración del contexto y la implicación del estudio y diseño de la interacción entre los usuarios y los SR ocupan las principales líneas de investigación en este tema.

Entre algunas propuestas para inducir serendipia en los sistemas de recomendaciones podemos destacar algunas:

Toms sugiere 4 estrategias, desde la más simple hasta la más compleja:

- Un golpe de suerte o del azar, implementado a través de un nodo generador de información al azar
- Principio *Pasteur*, “*la suerte favorece sólo a la mente preparada*”, es decir, destellos repentinos de visión no ocurren por casualidad sino que son el producto de la preparación. Los SR podrían aplicar este paradigma mediante la aplicación de la información de las preferencias del usuario en varios contextos (p. ej., si el sistema sabe que al usuario le gustan las películas de ciencia ficción podría explotar dicha información cuando está buscando un hotel y sugerirle el Hilton de Las Vegas ya que alberga un simulador de Star Trek).
- Anomalías y excepciones, que pueden ser implementadas utilizando medidas de distancia capaces de identificar elementos diferentes a los que les gustó al usuario en el pasado.

- Razonamiento por analogía, lo que implica un mecanismo de abstracción permitiendo al sistema descubrir la aplicabilidad de un esquema existente a una nueva situación.

(de Gemmis et al., 2015) siguiendo el principio de Pasteur, define una estrategia basada en un proceso intensivo de conocimiento denominado *KI*, Infusión de Conocimiento. Construye automáticamente un fondo de conocimiento legible, la memoria del sistema, y explotada por un algoritmo de razonamiento para encontrar correlaciones significativas ocultas entre los elementos. La hipótesis es que si el proceso de recomendación explota las asociaciones descubiertas en lugar de similitudes de características, se pueden proporcionar al usuario recomendaciones más casuales.

(Yamaba et al., 2013) proponen la representación de ítems a través de las impresiones que tienen los usuarios sobre ellos mediante las folksonomias.

7.2. SISTEMAS DE RECOMENDACIONES COLABORATIVOS

Los sistemas colaborativos, definidos por primera vez por Goldberg, pretenden conectar grupos de individuos con intereses similares y aprovechar esa experiencia colectiva para realizar las recomendaciones. Estos sistemas utilizan la información conocida sobre las preferencias de otros usuarios para realizar la recomendación al usuario que la precise. Definida de una manera más formal (Adomavicius, 2005): la utilidad $u(c,s)$ del ítem s para el usuario c se estima en base a las utilidades $u(c_{c_j},s)$ asignadas al ítem s por todos aquellos usuarios $c_j \in C$ que son similares al usuario c .

Estos sistemas se apoyan pues en las tendencias marcadas por grupos de personas con las mismas preferencias que el usuario activo. Un usuario cuyos gustos o preferencias son similares a los de un grupo determinado pasa a formar parte de ese grupo. Si personas de ese grupo valoran positivamente un nuevo ítem, el sistema de filtrado establece que es muy probable que los demás miembros del grupo al que pertenecen valoren positivamente al nuevo ítem. Ejemplarizando, si a un usuario U le han gustado las películas “Avatar”, “Regresión” y “El corredor del laberinto” y al usuario activo U_a le ha gustado también “Regresión” y “El corredor del laberinto” hay muchas posibilidades de que le pueda gustar también “Avatar” por ser usuarios con gustos muy parecidos. El SR debería recomendar esta película al usuario activo.

Este tipo de sistemas han de realizar 3 tareas básicas para elaborar las recomendaciones demandadas por los usuarios (Rodríguez, M. et al, 2010):

- 1) Análisis y selección de un conjunto de datos. El conjunto de datos debe ser almacenado y optimizado por el sistema
- 2) Agrupación de usuarios. Para elaborar las recomendaciones, los algoritmos colaborativos seleccionan un grupo de usuarios con preferencias similares
- 3) Generar la predicción. Una vez que los usuarios han sido agrupados por intereses el sistema los utiliza para calcular mediante diversas técnicas la predicción para el usuario activo.

La idea básica se puede resumir de la siguiente manera: Dado un usuario activo y un objeto que todavía no ha visto:

- Encontrar el conjunto de usuarios que más se parezcan al usuario activo (le gustan objetos similares y han puntuado el objeto)
- Aplicar este proceso sobre todos los objetos que el usuario no ha puntuado
- Recomendar los que tienen mayor puntuación

Las estrategias o algoritmos de filtrado colaborativo pueden ser agrupados en 2 clases generales: los basados en memoria, *memory-based* (comúnmente conocidos como *K-Neighbourhood* por ser quizás la técnica más empleada, y los basados en modelos, *model based*). Las diferencias entre estos dos enfoques radican en la forma en que procesan la información disponible en la matriz de puntuaciones (Formoso, 2013). Los basados en memoria procesan la matriz cada vez que calculan una recomendación o predicción, usando generalmente medidas de similitud para encontrar usuarios o ítems con un patrón de puntuaciones similares para luego llevar a cabo la recomendación. Los basados en modelos utilizan la información presente en la matriz para entrenar un modelo previamente especificado, el cual intenta reflejar características o relaciones entre ítems y usuarios. Las recomendaciones o predicciones se hacen a partir de dicho modelo sin que sea necesario volver a procesar la matriz continuamente. El modelo se entrena en un proceso offline y solo se vuelve a entrenar cuando los cambios en las puntuaciones presentes en la matriz requieren una actualización del mismo.

7.2.1. SISTEMAS COLABORATIVOS BASADOS EN LA MEMORIA

Las calificaciones de los ítems de los usuarios se almacenan en el sistema y se usan directamente para predecir calificaciones de los nuevos ítems de dos maneras distintas (Desrosiers y Karypis, 2011) conocidas como basadas en usuario o en ítems según utilicen la semejanza entre usuarios o entre ítems respectivamente.

7.2.1.1. Vecino más cercano basado en el usuario

La idea principal de este enfoque es el siguiente: Dada una base de datos de ratings y un usuario activo como entrada, identificar otros usuarios (vecinos más cercanos) que tuvieron preferencias similares al usuario activo en el pasado. Para cada ítem que el usuario activo no ha visto aún, se calcula la predicción sobre los ratings que para ese ítem hicieron los vecinos similares.

(Tomamos como ejemplo para ilustrar esta técnica el proporcionado en (Jannach et all., 2011):

Tomamos como base la matriz de valoraciones:

	Item1	Item2	Item3	Item4	Item5
Alice	5	3	4	4	?
User1	3	1	2	3	3
User2	4	3	4	3	5
User3	3	3	1	5	4
User4	1	5	5	2	1

Como se puede observar en la matriz de ratings formada por el usuario activo (Alice) y otros usuarios, y un ítem (Item 5) aún no visto o calificado por el usuario activo, la tarea del sistema de recomendaciones consistiría en determinar la predicción del usuario activo sobre ese ítem y en el caso de que fuese alta incluir dicho ítem en la lista de recomendaciones.

Para realizar la recomendación se deben llevar a cabo una serie de pasos:

1)- **Calculo de similitud entre usuarios.** En primer lugar se ha de elegir una métrica para determinar la similitud entre una pareja de usuarios (Coseno, MSD, Correlación de Pearson..)

Por ejemplo utilizamos la correlación de *Pearson*, una medida de similitud muy popular en los sistemas colaborativos basados en usuario. La similitud $sim(a,b)$ de los usuarios a y b dada da matriz de ratings R , se define como:

$$sim(a, b) = \frac{\sum_{p \in P} (r_{a,p} - \bar{r}_a)(r_{b,p} - \bar{r}_b)}{\sqrt{\sum_{p \in P} (r_{a,p} - \bar{r}_a)^2} \sqrt{\sum_{p \in P} (r_{b,p} - \bar{r}_b)^2}}$$

Los símbolos \bar{r}_a y \bar{r}_b corresponden a la puntuación media del usuario a y b respectivamente

2)- **Calcular los K-vecinos.** Después de utilizar la métrica de similitud seleccionada se obtienen los K-usuarios mas similares al usuario activo. A estos se les llaman los k-vecinos del usuario activo.

	Usuario 1	Usuario 2	Usuario 3	Usuario 4
Alice	0.85	0.70	0.00	-0.79

Como se observa en la tabla del ejemplo anterior los usuarios 1 y 2 valoraron en el pasado de una forma similar a Alice, por lo que la elección obvia sería elegir a este par de usuarios para predecir las valoraciones de Alice.

3)- **Calcular las predicciones de los ítems.** A partir de los K-vecinos del usuario activo se determinan las posibles valoraciones que el usuario activo haría sobre los ítems que no ha votado, es decir, se predice cómo el usuario valoraría esos ítems. Una posible fórmula para calcular la predicción sería:

$$pred(a, p) = \bar{r}_a + \frac{\sum_{b \in N} sim(a, b) * (r_{b,p} - \bar{r}_b)}{\sum_{b \in N} sim(a, b)}$$

En el ejemplo la predicción de la valoración de Alice para el ítem 5 basándose en las valoraciones de los usuarios 1 y 2 sería:

$$0+1/(0.85+0.7)*(0.85*(3-2.4)+0.70*(5-3.8))=4.87$$

Según estos cálculos se podrían calcular predicciones de valoraciones para el usuario activo sobre todos los ítems que no ha visto o calificado aún e incluiría los que tienen un valor de predicción alto en la lista de recomendaciones.

7.2.1.2. Vecino más cercano basado en ítems

Este enfoque se utiliza sobre todo en grandes webs de comercio electrónico, donde es imposible calcular las predicciones en tiempo real al necesitar explorar tan inmensas

cantidades de potenciales vecinos.

A diferencia del enfoque anterior este sistema se basa en calcular las predicciones usando la similitud entre ítems.

Basándonos en el ejemplo anterior, la predicción de Alice para el ítem 5 se llevaría a cabo de la siguiente manera: Primero compararíamos los vectores de otros ítems y buscaríamos aquellos ítems que tienen ratings similares al ítem 5.

Entre las métricas más utilizadas para hallar la similitud entre los ítems es la medida del coseno, al parecer la que ofrece unos resultados más precisos y que como ya vimos anteriormente, se utiliza sobre todo para medir la similitud entre dos vectores n-dimensionales mediante el ángulo que forman entre ellos. Otra medida muy utilizada también es una variante de la anterior, la llamada medida del coseno ajustado, al tener en cuenta las diferencias entre las escalas de los usuarios. Las puntuaciones de cada usuario se normalizan respecto a su media con el objeto de minimizar el impacto de la variabilidad entre usuarios. Para ello introduce en la fórmula el parámetro R_u que indica la valoración media de ese usuario:

$$sim(a, b) = \frac{\sum_{u \in U} (r_{u,a} - \bar{r}_u)(r_{u,b} - \bar{r}_u)}{\sqrt{\sum_{u \in U} (r_{u,a} - \bar{r}_u)^2} \sqrt{\sum_{u \in U} (r_{u,b} - \bar{r}_u)^2}}$$

El valor de similitud del coseno ajustado para los ítems 5 y 1 sería:

$$\frac{0.6 * 0.6 + 0.2 * 1.2 + (-0.2) * 0.80 + (-1.8) * (-1.8)}{\sqrt{(0.6^2 + 0.2^2 + (-0.2)^2 + (-1.8)^2} * \sqrt{0.6^2 + 1.2^2 + 0.8^2 + (-1.8)^2}} = 0.80$$

Después de determinar la similitud entre los ítems se puede predecir la valoración de Alice sobre el ítem 5 calculando una suma ponderada de las valoraciones de Alice sobre los ítems que son similares al ítem 5. Formalmente se puede predecir la calificación de un usuario u sobre un ítem p :

$$pred(u, p) = \frac{\sum_{i \in ratedItems(u)} sim(i, p) * r_{u,i}}{\sum_{i \in ratedItems(a)} sim(i, p)}$$

7.2.2. FILTRADO COLABORATIVO BASADO EN MODELOS

Como señalábamos anteriormente, los sistemas de filtrado basados en modelos utilizan la información presente en la matriz para entrenar un modelo previamente especificado, el cual intenta reflejar características o relaciones entre ítems y usuarios. Estos modelos se

construyen a partir de técnicas de inteligencia computacional entre las que podríamos incluir modelos probabilísticos basados en técnicas bayesianas, minería de reglas de asociación, redes neuronales artificiales, clustering, algoritmos genéticos...Resumimos a continuación las ideas básicas de algunas de ellas:

7.2.2.1. Modelos probabilísticos

Se basan en la existencia de clases que determinan las características de usuarios e ítems. Utilizan un modelo probabilístico donde la pertenencia a una clase se modela como una variable latente. La clase a la que pertenece un usuario y/o ítem según el modelo, determinará la distribución de sus puntuaciones. Volvemos a tomar como ejemplo el proporcionado anteriormente y utilizar un método probabilístico básico para calcular la predicción sobre Alice en cuanto al ítem 5:

	Item1	Item2	Item3	Item4	Item5
Alice	5	3	4	4	?
User1	3	1	2	3	3
User2	4	3	4	3	5
User3	3	3	1	5	4
User4	1	5	5	2	1

La tarea de predicción se formula como el problema de calcular el valor más probable para calificar al ítem 5, dadas las otras calificaciones de Alice y las de los otros usuarios. Según este método se calculan las probabilidades condicionales para cada posible valor de calificación de Alice dadas las otras calificaciones y seleccionar como predicción la de mayor probabilidad.

Para predecir la probabilidad de calificar con un 1 al ítem 5 se debe calcular la probabilidad condicional $P(\text{Ítem } 5=1/X)$ siendo X las otras puntuaciones de Alice:

$X = (\text{Ítem } 1=1, \text{ ítem } 2=3, \text{ ítem } 3= 3, \text{ ítem } 4=2)$

Para calcular esa probabilidad se utiliza generalmente el *Teorema de Bayes* que permite calcular la probabilidad posterior $P(Y/X)$ a través de la probabilidad condicional de clases $P(X/Y)$, la probabilidad de Y (en el ejemplo, la probabilidad del valor 1 para el ítem 5), y la probabilidad de X. Mas formalmente:

$$P(Y|X) = \frac{P(X|Y) \times P(Y)}{P(X)}$$

Considerando que los atributos son condicionalmente independientes (las calificaciones de los usuarios) se puede calcular la probabilidad posterior para cada valor de Y con el clasificador

bayesiano *naive*, siendo d el número de atributos en cada X :

$$P(Y|X) = \frac{\prod_{i=1}^d P(X_i|Y) \times P(Y)}{P(X)}$$

Como $P(X)$ es un valor constante se puede omitir en los cálculos.

$P(Y)$ se puede estimar para cada calificación basada en la base de datos de calificaciones $P(\text{Item } 5=1) = 2/4$, porque 2 o 4 posiciones para el ítem 5 tenían el valor 1,

$P(\text{item } 5=2) = 0$ y así en adelante

$P(X/\text{item } 5=1) = P(\text{item } 1=1/\text{item } 5=1) * P(\text{item } 2=3/\text{item } 5=1) * P(\text{item } 3=3/\text{item } 5=1) * P(\text{item } 4=2/\text{item } 5=1) = 2/2 * 1/2 * 1/2 * 1/2 = 0.125$

$P(X/\text{item } 5=2) = P(\text{item } 1=1/\text{item } 5=2) * P(\text{item } 2=3/\text{item } 5=2) * P(\text{item } 3=3/\text{item } 5=2) * P(\text{item } 4=2/\text{item } 5=2) = 0/0\dots$

Basandose en estos cálculos, dado $P(\text{item } 5=1) = 2/4$, la probabilidad posterior del valor 1 para el ítem 5 es $P(\text{item } 5=1/Y) = 2/4 * 0.125 = 0.0625$. En la matriz de ratings del ejemplo, $P(\text{item } 5=1)$ es más alto que las otras probabilidades lo que significa que la predicción probable para calificar el ítem 5 para Alice sería 1.

Generalmente este método es complejo y no funciona bien con bases de datos pequeñas con pocas puntuaciones. Los enfoques que más utilizan los modelos probabilísticos se basan en la idea de agrupar usuarios o ítems similares en clusters.

7.2.2.2. Extracción de reglas de asociación, Association rule mining.

Una regla de asociación es una expresión del tipo X implica Y , donde X e Y son conjunto de ítems. A menudo se escriben en la forma $X \Rightarrow Y$ expresando que siempre que los elementos de X (*antecedente* de la regla) están contenidos en una transacción T , es muy probable que los elementos en Y (*consecuencia* de la regla) sean elementos de la misma transacción.

La minería de reglas de asociación es una técnica muy utilizada para identificar reglas como patrones de relación en operaciones de venta. Se basa en la idea de “la cesta de la compra” para detectar que pares de productos se compran juntos. Cuando esa relación se conoce, ese conocimiento puede explotarse para ventas cruzadas.

En un SR colaborativo se traduciría en detectar automáticamente reglas del tipo “Si a un usuario X le gusta el ítem 1 e ítem 2 probablemente le gustará también el ítem 5”. Las

recomendaciones para el usuario activo se realizan evaluando qué reglas detectadas se aplicarán, es decir, comprobar si al usuario le gustó o compró los ítems 1 y 2 y generar una lista ordenada de ítems propuestos en base a estadísticas sobre la coocurrencia de ítems en las operaciones de venta

El objetivo de estos algoritmos es cómo detectar automáticamente a priori tales reglas y calcular una medida de calidad o control que defina el grado de interés de la regla. Las dos medidas standar se conocen como *soporte* y *confianza*. El *soporte* se define como el porcentaje de operaciones que contienen a $X \cup Y$ (probabilidad de coocurrencia de X e Y en una operación) del total de operaciones. La *confianza* se define como el porcentaje de número de transacciones que contienen a $X \cup Y$ sobre el número total de operaciones que contienen a X , la probabilidad condicional de Y dado X .

7.2.2.3. Métodos de reducción de la dimensionalidad

Básicamente se apoya en la proyección de los usuarios e ítems en un espacio reducido latente que captura sus características más importantes.

Cuando las matrices de puntuaciones son grandes y dispersas estas técnicas tienen como objetivo transforman esta matriz en una de dimensiones menores que refleje las características más importantes. Una vez seleccionadas las características más representativas, esta información se representaría con una matriz de reducidas dimensiones en las q se representa el grado de afinidad de un usuario a cada una de las características y el grado de pertenencia de un ítem a los mismos. La idea es que dichas características representan atributos latentes en la matriz original, los cuales permiten expresar el comportamiento de los usuarios. Esto aceleraría el proceso de recomendación, ya que únicamente habría que considerar dichas características en lugar de todas las puntuaciones, al tiempo que minimizaría los problemas relacionados con la dispersión de la matriz o la presencia de datos anómalos.

7.2.2.4. Algoritmos basados en regresión.

Orientados fundamentalmente a la tarea de predicción, se basan en que la puntuación de un ítem puede servir para estimar la de otro (Formoso, 2013). En lugar de sumar directamente las notas de los elementos similares, se utiliza una aproximación basada en la recta de regresión. Partiendo de esa premisa, entrenan un modelo de regresión para par de ítems $i, j \in J$, utilizando los datos disponibles en la matriz de puntuaciones. El modelo que se denota

como $F_{j,i}: R \rightarrow R$ puede verse como un experto capaz de estimar la puntuación de un ítem i a partir del ítem j .

Normalmente el usuario puntúa varios ítems, por lo que las estimaciones obtenidas a partir de cada uno de ellos se combinan para obtener la predicción final, es decir:

$$\hat{r}_{ai} = \sum_{j \in \mathcal{I}(a)}^* f_{ji}(r_{aj})$$

La función de agregación \sum^* es generalmente algún tipo de suma ponderada. Normalmente el uso de estos algoritmos requiere predecir una puntuación para todos los ítems no puntuados por el usuario activo, y luego ordenarlos para devolver los que tengan una puntuación mayor.

Cono este método se intenta compensar el problema que se da al evaluar similitudes mediante medidas del coseno o la correlación y es que vectores con alta similitud pueden encontrarse distantes en sentido euclídeo.

7.2.3. Principales ventajas y limitaciones del filtrado colaborativo

Los enfoques colaborativos superan algunas limitaciones de los basados en contenidos en algunos aspectos:

- . Permite recomendar ítems cuyos contenidos son difíciles de analizar a través de la retroalimentación de otros usuarios.
- . Habilidad para filtrar ítems basados en calidad y no solo en sus características. En el filtrado colaborativo la calidad esta siempre evaluada por personas ya que para la recomendación solo se tienen en cuenta las preferencias. Como ya vimos en los SR basados en contenido la recomendación implica la comparación en el contenido de los ítems y el hecho de que compartan un número de términos no implica que la calidad sea la misma.
- . A diferencia de los basados en contenido, los filtrados colaborativos pueden recomendar ítems con muy diferentes contenidos, siempre y cuando otros usuarios hayan mostrado interés por los mismos. Por otra parte, no se necesita conocer las características de los mismos, lo que permite que estos sistemas puedan ser usados en dominios diferentes. Otras técnicas requieren de cierto conocimiento sobre el dominio, lo que los hace exclusivos de aquel sistema para el que fueron diseñados, o que incluso no puedan ser usados en otros contextos. Para este tipo de sistemas recomendar libros sería lo mismo que recomendar películas u otros tipos de ítems.

. Capacidad para proporcionar recomendaciones casuales. Otros sistemas nunca recomendarían ítems que no fuesen muy diferentes a los valorados positivamente por el usuario. Al basarse enteramente en las preferencias estos sistemas pueden recomendar ítems inesperados y que no guardan relación aparente con los preferidos por el usuario, eso sí, si le hubiese gustado a otro usuario con gustos similares al usuario activo.

. Adaptabilidad. Cuando el número de usuarios y puntuaciones crecen estos sistemas tienen un mejor rendimiento.

A pesar de que los filtrados colaborativos poseen ciertas ventajas que no están presentes en otros tipos de sistemas, estos también poseen una serie de limitaciones:

. El problema de dispersión o escasez de puntuaciones, (*sparsity*). En algunos sistemas de recomendaciones, el número de calificaciones o puntuaciones obtenidas es muy pequeño comparado con el número de puntuaciones que necesita predecir. Eso significará que la matriz de puntuaciones será una matriz muy dispersa en la que desconoceremos el valor de la mayor parte de las celdas, complicando así la búsqueda de similitud entre usuarios. Por otra parte, es posible que dos usuarios sean muy similares pero al no haber puntuado los mismos ítems el sistema no sea capaz de averiguarlo, así como el caso de usuarios con gustos inusuales en comparación con el resto de la población para los que no habrá usuarios particularmente similares dando lugar a recomendaciones pobres. Una forma de superar la escasez de puntuaciones es utilizar información demográfica en el perfil del usuario cuando se calcula de similitud. Es decir, dos usuarios podrían considerarse similares no solo si calificaron los mismos ítems de una forma similar, también si pertenecen al mismo segmento demográfico (género, edad, educación, empleo, formación..., lo que se conoce como técnicas de filtrado demográfico (Pazzani, 1999).

. Escalabilidad. A medida que crece la cantidad de usuarios e ítems también crece la cantidad de cálculos de vecinos más cercanos para determinar los usuarios similares.

. El problema de “*the gray sheep*”. Se refiere a los usuarios cuyas opiniones no están de un modo consistente de acuerdo o no con cualquier grupo de personas y sus perfiles por lo tanto caen en la frontera de dos grupos de usuarios, haciendo difícil determinar para ellos una recomendación adecuada.

. Ataques y manipulación de preferencias. Al basarse por completo en información introducida por los usuarios, el filtrado colaborativo es más susceptible de sufrir ataques de spam sobre todo por parte de usuarios interesados en que se recomiende un determinado ítem

e introducen deliberadamente puntuaciones manipuladas para engañar así al sistema.

De la misma manera que la sobre especialización y falta de serendipidad caracterizan las desventajas de los sistemas basados en contenidos, el problema del arranque en frío o *cold start* lidera las limitaciones asociadas a este tipo de sistemas:

7.2.3.1. Cold start

Este problema se manifiesta con usuarios e ítems recientemente introducidos en el sistema. El usuario tiene que evaluar un número suficiente de ítems para que el sistema pueda aprender sus preferencias y presentarle recomendaciones fiables. Por lo tanto un usuario nuevo que tiene pocas puntuaciones no sería capaz de obtener recomendaciones precisas (Adomavicius y Tuzhilin, 2005). De la misma manera sucede con los ítems nuevos, hasta que el nuevo ítem no sea puntuado por un número substancial de usuarios el sistema no podría recomendarlo. Debido al hecho de que estos sistemas trabajan calculando las similitudes y correlaciones entre el usuario activo y otros usuarios, cuando no existe una información inicial o esta es escasa, la similitud entre los conceptos involucrados es baja o nula y por lo tanto, no se producen recomendaciones. Este problema disminuye la eficiencia general de este tipo de sistemas y más aún la confianza de los usuarios en las recomendaciones. La calidad y eficiencia del sistema dependerá por tanto de la capacidad para predecir con éxito recomendaciones con la mínima cantidad de información sobre usuarios e ítems (Rodríguez et al., 2010).

Aunque como veremos más adelante, los sistemas híbridos combinan varias técnicas de recomendación para aliviar las limitaciones que tanto las técnicas colaborativas como las basadas en contenido tienen por separado, podemos mencionar algunas propuestas aparecidas en la literatura como alternativas para solucionar el problema del arranque en frío (Adomavicius y Tuzhilin, 2005), (Rodríguez et al., 2010):

. (Rashid et al., 2002) propone la solicitud de puntuaciones mediante la presentación de algunos ítems, destacando la importancia en la elección de las preguntas para minimizar el esfuerzo del usuario mientras aprende sus preferencias. Propone varias estrategias para seleccionar dichos ítems como estrategias basadas en la popularidad, selección al azar o en base a la entropía de los ítems.

. (Melville et al., 2002) proponen el uso de los clasificadores bayesianos para estimar las puntuaciones desconocidas a partir de la características de los ítems. Posteriormente el

perfil completo se utiliza en un algoritmo de filtrado colaborativo tradicional.

. (Ahn H.J, 2008) propone una nueva medida de similitud (PIP, Proximity- Impact- Popularity). El factor proximidad, además de basarse en una diferencia aritmética entre dos puntuaciones considera además, si esas calificaciones coinciden o no, penalizando a las que están en desacuerdo duplicando la distancia entre las puntuaciones. El factor impacto considera la fuerza con la que un ítem agrada o desagrada al usuario y por lo tanto da una mayor credibilidad a la similitud. El factor de popularidad da un mayor valor de similitud a aquellos ítems que están muy lejos de la puntuación media de los ítems co-puntuados.

. (Victor et all, 2007), ofrecen la conexión de un nuevo usuario a una red de confianza subyacente entre los usuarios del sistema, estudiando el efecto de guiar al nuevo usuario a través del proceso de conexión y en particular la influencia que tiene esto en la cantidad de recomendaciones generadas, afirmando que es más beneficioso que un recién llegado se conecte a una figura clave identificada en lugar de un usuario al azar.

7.3. SISTEMAS DE RECOMENDACIONES HÍBRIDOS

El objetivo de estos sistemas es solucionar los inconvenientes o limitaciones que presentan los anteriores sistemas, combinando diferentes técnicas para mejorar así la precisión en las recomendaciones. Normalmente se combinan las técnicas del filtrado colaborativo con las basadas en contenido clasificándose en las siguientes categorías (Adomavicius y Tuzhilim, 2005):

Implementando por separado ambos métodos. Existen cuatro maneras diferentes de combinar estos dos sistemas por separado (Nguyen y Santos, 2013):

. Método de hibridación ponderado. Las puntuaciones obtenidas individualmente por cada sistema se combinan entre sí para obtener una única recomendación final utilizando bien una combinación lineal o un esquema de votación.

. Método de hibridación cambiado. El sistema va cambiando entre las técnicas, seleccionando el método que da mejores resultados en una situación dada dependiendo de alguna medida de calidad.

. Método de hibridación mixto, donde los resultados de diferentes subsistemas de recomendaciones se presentan simultáneamente.

. Método de hibridación en cascada. En este método, una de las técnicas se utiliza para producir una clasificación de candidatos en bruto y la otra se centra en aquellos elementos en los que hace falta un refinamiento adicional.

Añadiendo características de las técnicas basadas en contenido a los enfoques colaborativos. Algunos autores proponen la colaboración a través del contenido en la que la similitud entre los usuarios no se calcula sobre las puntuaciones obtenidas sino mas bien en el perfil basado en el contenido de cada usuario. Otros autores como (Melville et al., 2002) citados en (Adomavicius y Tuzhillim, 2005), utilizan un enfoque colaborativo donde el vector de puntuaciones del usuario tradicional se aumenta con puntuaciones adicionales calculadas utilizando un predictor basado en el contenido.

Añadiendo características del filtrado colaborativo a los enfoques basados en contenido. La principal técnica de esta categoría es aplicar la reducción de la dimensionalidad en un grupo de perfiles basados en contenido. La técnica más popular es la introducida por (Soboroff et al., 1999) donde utilizan la indexación semántica latente para crear una opinión colaborativa de los perfiles de usuario donde dichos perfiles se representan como vectores de términos.

Modelo unificado de recomendación. Algunos ejemplos propuestos en la literatura son p. ej., el uso de características colaborativas y de contenido (edad del usuario, género de una película) en un solo clasificador basado en reglas o la utilización de un método probabilístico que combina recomendaciones colaborativas y basadas en contenido.

Podemos citar el ejemplo de filtrado híbrido propuesto en (Albadvi y Shahbazi, 2009)

7.4. OTRAS TÉCNICAS Y CLASIFICACIONES DE SISTEMAS DE RECOMENDACIONES

7.4.1. Sistemas de Recomendaciones basadas en el Conocimiento

(Felfernig y Burke, 2008), los definen como los sistemas que se basan en fuentes de conocimiento distintas de las explotadas por los sistemas basados en el contenido y colaborativos y caracterizados por lo tanto por los dos aspectos del conocimiento que no se encuentran en los otros diseños, es decir, las necesidades del usuario y dominio del

conocimiento.

Normalmente los sistemas que se basan en el filtrado colaborativo se apoyan en las calificaciones de los usuarios como la única fuente de conocimiento para generar ítems para sus usuarios y no suele introducirse conocimiento adicional como características o disponibilidad de los ítems. Por otra parte en los basados en contenido, la fuente de conocimiento se basa en información categorizada, palabras clave que son extraídas automáticamente de la descripción textual de un ítem. Aunque ambos enfoques guían al usuario hacia ítems que puedan serle de interés, hay situaciones en las que dichos enfoques no ofrecen la mejor solución. Lo que distingue a estos sistemas basados en el conocimiento es su “énfasis” en la situación del usuario y cómo los ítems recomendados pueden satisfacer esa necesidad particular (quizás porque las preferencias evolucionan a lo largo del tiempo y quizás, porque no siempre tengamos las mismas necesidades). Estos sistemas permiten abordar esos cambios. Además tiene la ventaja de que no se necesitan ratings para calcular recomendaciones, las recomendaciones se calculan independientemente de las calificaciones individuales de los usuarios.

Estos sistemas manejan 3 tipos de conocimiento (Burke, 2002) :

. *Catalog Knowledge*, conocimiento sobre los ítems que se están recomendando

. *Functional Knowledge*, cómo las características de los productos se relacionan con las necesidades del usuario

. *User Knowledge*, el conocimiento que el sistema ha obtenido sobre las necesidades del usuario

Una característica de las aplicaciones que utilizan estos sistemas es su alta interactividad, razón por la cual se les conoce como sistemas conversacionales. La interacción se modela en forma de un cuadro de dialogo donde los usuarios pueden especificar sus requisitos como forma de respuesta a las preguntas. Permiten también la interacción en lenguaje natural flexibilizando así las consultas (p.ej. los usuarios pueden especificar propiedades sin ser obligado a responder a un elevado número de preguntas, o pueden apoyar consultas que no estén directamente relacionadas con búsqueda de ítems como cuestiones relativas a funcionalidad o preguntas técnicas).

Hay dos tipos de enfoques en estos sistemas, los conocidos como basados en restricciones, *constraint-based* y los basados en casos, *cased-based*. Ambos enfoques son similares en el

proceso de recomendación. Es decir, el usuario debe especificar los requisitos y el sistema debe encontrar una solución, proponer cambios en los requisitos si no encuentra la solución y dar explicaciones de apoyo a los ítems seleccionados. Difieren en la forma en que usan el conocimiento proporcionado. Las recomendaciones basadas en casos utilizan las métricas de similitud para recuperar ítems (dentro de un umbral predefinido) sobre los requisitos especificados por el usuario, con el entendimiento de que esta similitud implica consideraciones sobre el conocimiento de ese dominio específico. Los basados en restricciones, se basan en ir descartando aquellos ítems incompatibles con los requisitos especificados por el usuario hasta que se tenga un número limitado de opciones. Si no hay ningún ítem que realmente se adapte a las necesidades del usuario (el valor de similitud no alcanza un umbral o el conjunto de restricciones es incompatible con el conjunto de restricciones proporcionadas por el usuario), ambos enfoques explotan mecanismos de apoyo para determinar el conjunto mínimo de cambios en el conjunto de requisitos del usuario para poder encontrar una solución.

Estos sistemas son adecuados para búsquedas casuales, cuando no existe información del usuario o ésta es escasa. Los sistemas colaborativos p. ej., necesitan un periodo hasta obtener información sobre el usuario la calidad de las recomendaciones puede ser bastante bajo e incluso pueden llegar a no realizarse durante dicho periodo. Los basados en conocimiento no tienen esa desventaja ya que pueden funcionar bien con una pequeña cantidad de conocimiento sobre el usuario, sin necesidad de información histórica. Tampoco parece sufrir el problema de “*the grey sheep*”.

Sin embargo estos sistemas también presenta algunas desventajas siendo la más destacada la de requerir la adquisición de un conocimiento explícito. Muchas veces este conocimiento no es fácil de obtener o de ser extraído con técnicas automáticas. Por otra parte, cuando la cantidad de ítems es muy grande, el proceso de proveer un ejemplo para expresar las necesidades del usuario puede ser una dura tarea. Además debemos añadir la posibilidad de que el usuario no encuentre un ejemplo que sea apto exactamente a sus necesidades.

Ejemplo de este enfoque lo podemos encontrar en (Zhen et all, 2010)

7.4.2.Sistemas de recomendaciones semánticos

Este enfoque intenta mejorar y enriquecer la representación de la información mediante la aplicación de tecnologías de la web semántica.

Los perfiles basados en palabras clave no pueden capturar el valor semántico de los intereses del usuario ya que se basan en una operación de coincidencia en la cadena de caracteres. Si esa cadena de caracteres se encuentra en el perfil y en el ítem, se considera relevante sin tener en cuenta los problemas derivados del lenguaje natural como pueden ser la sinonimia y polisemia. El principal enfoque para resolver este problema es la integración del análisis semántico. La idea es la adopción de bases de conocimiento como léxicos u ontologías para anotar ítems y representación de perfiles y para obtener una interpretación semántica de las necesidades de información de los usuarios. El análisis semántico permite un aprendizaje más preciso al contener referencias a conceptos definidos en bases de conocimiento externas. El principal motivo de este enfoque es proporcionar un SR que tenga un conocimiento lingüístico y cultural para interpretar el lenguaje natural de los documentos y razonar sobre su contenido (Janach et al., 2011).

7.4.3. Técnicas basadas en la teoría de los conjuntos difusos (fuzzy set theory)

El hecho de que la subjetividad y la incertidumbre sean propiedades típicas de cualquier proceso de acceso a la información, sobre todo si intervienen los usuarios, ha dado lugar a que estas técnicas se hayan convertido en una excelente herramienta para el manejo de la subjetividad y la imprecisión en la definición de los sistemas de acceso a la información (Porcel, 2005). La teoría de conjuntos difusos ofrece una rica variedad de métodos para la gestión de incertidumbre y adecuada sobre todo para el manejo de información imprecisa, la falta de nitidez en situaciones o clases de objetos y, en la gradualidad de perfiles de preferencia.

En relación a los ítems, la incertidumbre se asocia al grado (p. ej., bajo o alto) en el cual los ítems tienen algunas características. En relación con el comportamiento de los usuarios, como podría ser el interés, se asocia con los métodos empleados para medir y representar el interés de estos de una manera tan precisa como sea posible. En relación con la tarea de recomendación, la incertidumbre se asocia a los tipos de relación que existe entre primero, el comportamiento del usuario y las características de los ítems, segundo, entre usuarios en términos de comportamiento y tercero, entre ítems en cuanto a sus características (Zenebe y Norcio, 2008).

La noción de conjunto refleja la tendencia a “encapsular” los objetos en clases o conjuntos cuyos miembros comparten una serie de características o propiedades y que implican por lo tanto la función de pertenencia a ese conjunto o clase. La lógica difusa, que se fundamenta en

el concepto de conjunto difuso, suaviza el requerimiento anterior y admite valores intermedios en esa función de pertenencia. Esto permite una interpretación más realista de la información puesto que la mayoría de las categorías que describen los objetos del mundo real no tienen unos límites claros y bien definidos (Porcel, 2005).

En concreto, se puede aplicar la lógica difusa en dos ámbitos:

- Para modelar las preferencias y opiniones de los usuarios
- Para modelar los operadores de agregación que están relacionados con el problema de combinar diferentes valoraciones (puntuaciones, preferencias o graduaciones) de varias fuentes de información.

Ejemplos de aplicación de estos enfoques lo podemos ver en (Herrera-Viedma et al., 2007) y (Porcel y Herrera-Viedma, 2010)

7.4.4. Sistemas de recomendaciones demográficos

Los SR demográficos tratan de categorizar al usuario en función de atributos personales y hacer recomendaciones basadas en clases demográficas (Pazzani, 1999; Burke, 2002; Safoury y Salah, 2013). Realizan el proceso de recomendación en 3 etapas: entrada de datos, cálculo de similitud y predicción de la recomendación. La entrada de datos la constituirán los datos demográficos del nuevo usuario (usuario activo) u las clasificaciones y datos demográficos del resto de usuarios. Se calcula la similitud de los datos demográficos de los usuarios para obtener un grupo de usuarios similares al usuario activo (vecindario) y finalmente se sugieren al usuario los ítems que han sido puntuadas como positivas por el vecindario.

El beneficio de este enfoque radica en que no necesita de un perfil histórico de datos de usuario como es el caso de los basados en contenido y colaborativos ayudando así a paliar en cierta medida el problema del arranque en frío. Sin embargo, la recogida de datos demográficos necesarios se hace difícil ya que las personas normalmente somos reticentes a la hora de proporcionar información personal, necesitando por tanto investigación y estadística para saber traducir los grupos culturales de la persona a las necesidades informativas (Formoso, 2013).

7.4.5. Sistemas de Recomendaciones basados en redes sociales

Con la explosión de la web 2.0, la generación de contenidos por parte de los usuarios en

blogs, foros de discusión, redes sociales, profesionales..., suponen un nuevo desafío para afrontar toda esta sobrecarga de información. En este sentido, estos sistemas se centran en aprovechar estas nuevas fuentes de conocimiento para mejorar los enfoques en las técnicas de recomendación introduciendo los datos de la web social. Exploran la información existente en las redes sociales y analizan su utilidad para la formulación de recomendaciones.

La integración de las redes sociales pueden mejorar técnicamente el rendimiento de los SR actuales (Che at all., 2013): Primero en términos de exactitud en la precisión y en segundo lugar, con la información de los amigos en las redes sociales ya no es necesario encontrar amigos similares calculando la similitud en las puntuaciones, ya que cuando las personas son amigos, se supone que hay ciertas cosas en común entre ellos.

Cobran también vital importancia las folksonomias o etiquetado social (conjunto de términos en lenguaje natural, *tags*, empleados por los usuarios para describir el contenido de un documento o recurso web). Estas etiquetas se convierten en una importante fuente de información tanto para describir los temas de interés en el perfil del usuario como para describir el contenido y las valoraciones de un determinado ítem, permitiendo a su vez la posibilidad de filtrado para distintos tipos de contenido.

Otro factor que se tiene en cuenta y que constituye un profundo interés en estos enfoques es la influencia social ya que juega un papel muy importante en el comportamiento del consumidor. A través de las redes sociales se pueden tener en cuenta las relaciones sociales explícitas entre los usuarios y se puede descubrir información adicional para predecir las preferencias de los usuarios y mejorar los algoritmos de predicción existentes. Los usuarios comparten sus experiencias tanto positivas como negativas sobre cualquier producto o servicio, ejerciendo una enorme influencia en la formación de opiniones de otros usuarios (Zhou. et all., 2012).

8. EVALUACIÓN DE LOS SISTEMAS DE RECOMENDACIONES

La evaluación de los SR y sus algoritmos es una tarea inherentemente difícil por varias razones: los algoritmos pueden ser mejores o peores según los diferentes conjunto de datos con los que trate, los objetivos de las evaluaciones pueden ser distintos y sobre todo, la elección de medidas para utilizar en la evaluación (Herlocker at all., 2004) .

Como ya se destaca ampliamente en la literatura, la evaluación en los SR no debe centrarse solamente en la precisión de la predicción, sino medir también la utilidad del sistema. Los objetivos del usuario, métricas de confianza que ayuden al usuario a tomar decisiones más eficaces o métricas que midan la posibilidad de introducir novedad y diversidad en las recomendaciones para un usuario, deben ser medidas a tener en cuenta para evaluar la utilidad del sistema y la satisfacción del usuario sobre su rendimiento.

No obstante, creo conveniente hacer un breve repaso por las métricas más utilizadas y comentadas tradicionalmente en la literatura.

Métricas de relevancia:

Procedentes de la IR, tienen como objeto medir el grado en que se satisfacen los gustos del usuario activo al realizar una recomendación. Las ampliamente conocidas *Precision* y *Recall* de Cleverdon, (Jannach et al., 2011):

La **precisión** viene definida por el conjunto de ítems relevantes, número de aciertos ($hits_u$) entre todos los ítems recomendados, ($recset_u$):

$$P_u = \frac{|hits_u|}{|recset_u|}$$

Recall viene definida por el conjunto de ítems relevantes recomendados entre todos los teóricamente relevantes del conjunto, ($testset_u$):

$$R_u = \frac{|hits_u|}{|testset_u|}$$

Unidas a éstas, otra métrica muy utilizada es una combinación de ambas, la **F1**, para presentar resultados evaluativos que sean más comparables universalmente. Dicha métrica realiza un promedio entre *precision* y *recall* con un sesgo en los valores más débiles:

$$F1 = \frac{2 \cdot P \cdot R}{P + R}$$

Otra medida muy popular es la **NDCG**, *Normalized Discounted Cumulative Gain*, que evalúa la capacidad del sistema para mostrar el ítem en las primeras posiciones. Evalúa la relevancia de los ítems teniendo en cuenta su posición en el ranking, ya que se puede colocar más un ítem poco relevante en las primeras posiciones que uno con mayor relevancia que esté más abajo en el ranking. La métrica admite además la distinción entre diferentes grados de

relevancia, a diferencia de la precisión que solo considera relevancia binaria

Métricas de Precisión:

Con la predicción como objeto, no es de extrañar que las estrategias en la evaluación se centrasen en las medidas de precisión en la predicción (Konstan y Riedl, 2012). Las primeras evaluaciones se centraron en evaluar los SR utilizando medidas de error y correlación. Las medidas de error y error cuadrático medio proporcionaron una evaluación de lo que las votaciones predichas coincidían con las puntuaciones reales. La correlación proporcionaba una medida similar pero se centra en la predicción relativa correcta en lugar de predicción absoluta.

MAE, *Mean absolute error*, mide la desviación absoluta media entre las valoraciones predichas y las valoraciones verdaderas de los usuarios (Jannach et al., 2011):

$$MAE = \frac{\sum_{u \in U} \sum_{i \in testset_u} |rec(u, i) - r_{u,i}|}{\sum_{u \in U} |testset_u|}$$

Calcula la desviación media entre los valores calculados de la recomendación $rec(u,i)$ y los valores reales ($r_{u,i}$) para todos los usuarios $u \in U$ y todos los ítems en el conjunto de pruebas ($testset_u$). Algunos autores proponen normalizar dicha medida para el rango de valores puntuados:

$$NMAE = \frac{MAE}{r_{max} - r_{min}}$$

donde r_{max} y r_{min} representan los valores más altos y más bajos de las puntuaciones.

Para (Shani y Gunawadana, 2011), **RMSE**, *Root Mean Square Error*, sea quizás la medida más popular para evaluar la predicción de las puntuaciones predichas:

RMSR. El sistema genera predicciones para $\hat{r}_{u,i}$ para un conjunto de pruebas T de pares de usuario-items (u,i) para los que se conocen las verdaderas puntuaciones r_{ui} . Estas puntuaciones r_{ui} se conocen bien porque proceden de experimentos offline, se obtuvieron a través de un estudio de usuarios o a través de experimentos online. El RMSE entre las votaciones previstas y las puntuaciones reales viene dado por:

$$RMSE = \sqrt{\frac{1}{|T|} \sum_{(u,i) \in T} (\hat{r}_{ui} - r_{ui})^2}$$

Métrica **ROC**, muy utilizada, proporciona una idea de la potencia de diagnóstico. Las curvas ROC dibujan la especificidad (probabilidad de que un elemento malo del conjunto sea rechazado por el filtro) y la sensibilidad (probabilidad de que un elemento bueno al azar sea aceptado). Si un elemento es bueno o malo viene dado por el usuario. Las curvas se dibujan variando el umbral de predicción a partir del cual se acepte el elemento. El área bajo la curva se va incrementando si cuando el filtro es capaz de retener más elementos buenos y menos malos.

9. ALGUNOS EJEMPLOS DE TÉCNICAS Y SISTEMAS DE RECOMENDACIÓN

Comercio electrónico, AMAZON

La recomendación de sus productos a sus clientes es uno de sus activos más importantes. Utiliza el filtrado colaborativo ítem a ítem. Asocia cada producto comprado por el cliente con una lista de productos similares y que se obtiene en función de los elementos que han sido adquiridos en un mismo pedido o almacenados en una “lista de deseos”.

Información Científica y Técnica, SCIENCEDIRECT

La plataforma ScienceDirect de Elsevier, incorpora una técnica de recomendación implícita basada en el contenido. En los detalles de las páginas del documento se incluyen enlaces a artículos recomendados, aunque no se especifica el porqué de esa recomendación y su relevancia. La técnica de recomendación basada en el comportamiento (descarga del artículo) se muestra en “Other users also viewed these articles”.

Musicales, SPOTIFY

Realiza la elección de preferencias en base a los hábitos de escucha y el estado anímico de la persona al otro lado de la pantalla. Los artistas que escuchas se clasifican según la afinidad que tengas con ellos en base al número de escuchas y encuentran similitud entre ellos, como lo podemos observar en las radios personalizadas de nuestros artistas favoritos. Por otra parte, la música es distinta según el momento del día y la actividad que realicemos, ofreciéndonos un simpático “*prepárate para el finde*” un jueves tarde, “*de camino al trabajo*” a las 8 de la mañana o un “*cafe, libros*” para una tranquila tarde de domingo...

Servicios y Turismo, MINUBE

Esta plataforma se utiliza tanto para buscar servicios (vuelos, alojamiento, hostelería..) como

una red social en la que los miembros comparten sus experiencias. Ofrecen guías de distintos destinos, creadas a partir de la selección de las mejores recomendaciones y fotografías compartidas por los usuarios. En la aplicación para móvil, una vez detectada nuestra posición, nos ofrece sitios para comer, dormir o hacer turismo. Estas sugerencias se ordenan según su popularidad o distancia a nuestra localización.

Entretenimiento, FILMAFFINITY

Uno de los mejores ejemplos de filtrado colaborativo. Su pilar se basa en recurrir a la sabiduría popular. A partir de las votaciones de los miembros de la comunidad podemos encontrar a nuestras “almas gemelas” (aquellos con los que tenemos mayor coincidencia en puntuaciones realizadas). La mayor afinidad se logra simplemente votando mayor cantidad de películas.

Bibliotecas digitales, Cyclades

El objetivo de Cyclades es proporcionar un entorno integrado para usuarios y grupos de usuarios que desean utilizar de una manera personalizada los recursos digitales abiertos y compartidos que se encuentran en los repositorios de la OIA, Open Archives Initiative. Los algoritmos de recomendación se basan tanto en la información personalizada como en las opiniones de los usuarios, utilizando el filtrado colaborativo y el basado en contenido tanto en combinación como por separado.

10. CONCLUSIONES

- La sobrecarga de Información que ha traído consigo la popularización y uso de Internet, se ha convertido en el escenario indiscutible para motivar la investigación y desarrollo de estas técnicas de recuperación. Su fin, el de ayudar a los usuarios actuando como consejeros o intermediarios, y aliviar así su esfuerzo en la ardua tarea de elegir entre las multitudinarias posibles opciones que se presentan casi a diario.
- Aunque son muchos los enfoques que se les ha dado a estas técnicas para abordar distintos problemas, podemos decir que el filtrado colaborativo y el filtrado basado en contenidos son la base en las metodologías adoptadas por estos sistemas.
- Los criterios de interés y personalización se convierten en los aspectos fundamentales que separan a los Sistemas de Recomendaciones de otros sistemas de recuperación de información y motores de búsqueda.

- La aceptación unánime de la aplicación de técnicas de minería de datos en el análisis de preferencias, para garantizar la eficacia en el suministro de información personalizada.
- Aunque son diversos los dominios en los que se han aplicado, y otros, en los que pretenderán aplicarse, la mayoría de sistemas de recomendación se ofrecen como uno de los mayores activos y mejores servicios de intermediación entre el usuario y empresas de diversos sectores. Un servicio de asesoramiento "exento" de retribución, pero condicionado a la concesión de lo que la mayoría de nosotros consideramos nuestra privacidad.
- Los más optimistas, auguran un Sistema de Recomendaciones o "asistente personal" adornando nuestras muñecas, capaces de percibir nuestro estado anímico y físico, la realidad que nos rodea y actuar en consecuencia. Al igual que, en el ya citado trabajo de (Martin et al, 2011), me pregunto si alguna vez el sistema de recomendaciones de nuestro banco, nos sugerirá que transfiramos nuestro dinero a otro banco que ofrece un mejor interés, que los servicios de streaming de películas nos sugieran un programa de TV, que televisen en ese momento en lugar de buscar en sus repositorios, o Amazon sugiriéndonos que tomemos prestado el libro que nos interesa de un amigo que lo compró y leyó no hace mucho.

BIBLIOGRAFIA

- . Adamopolius, P. y A. Tuzhilin., (2015) "On Unexpectedness in Recommender Systems: or How to better Expect the Unexpected " . ACM Transactions on Intelligent Systems and Technology. Vol. 5, núm 4. Art. 54
- . Adomavicius, G y A. Tuzhilin, (2005). "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions". Knowledge and Data Engineering, *IEEE Transactions on*. Vol. 17, núm 6. pp. 734-749.
- . Adomavicius, G., Sankaranarayanan, R., Sen, S. y A. Tuzhilin. "Incorporating contextual information in recommender systems using a multidimensional approach". En ACM Transactions on Information Systems. Vol. 23, núm. 1. pp. 103-145.
- . Ahn, H.J., (2008). "A new similarity measure for collaborative filtering to alleviate the new user cool-starting problem". En *Information Sciences*. Vol. 178, núm 1. pp. 734-49.
- . Albadvi, A. y M. Shahbazi., (2009). "A hybrid recommendation technique based on product category attributes". *Expert Systems with Applications*. Vol. 36, núm 9. pp.11480-8.
- . Beel, J., Gipp, B., Langer, S. y C. Breitingner, (2015). Research-paper recommender systems: a literature survey. *International Journal on Digital Libraries*. [Aún no publicado, disponible]
- . Belkin N J. y W. Croft, (1992). " Information Filtering and Information Retrieval: The two sides of the same coin" en *Information Filtering*. Communications of the ACM. Vol.35, n 12, pág. 29-38
- . Burke, R., (2002). "Hybrid recommender systems: Survey and experiments. *User Modeling and User-Adapted Interaction*. Vol. 12, pp. 331-370
- . Burke, R. y M. Ramezani. (2011). "Matching Recommendation Technologies and Domains". En Ricci (ed). *Recommender Systems Handbook*. pp.367-386.
- . Che, S., Owusu, S. y L. Zhou., (2013). "Social Network based Recommendation Systems: A short survey". *IEEE Conference Publications*. 2013 National conference on Social Computing. pp. 882-885.
- . Desrosiers, C. y G. Karypis. "A Comprehensive Survey of Neighborhood-based Recommendation Methods". En Ricci (ed) *Recommender Systems Handbook*. Boston: Springer Science Business Media. pp. 107-144.
- . Felfernig, A. y R. Burke, (2008). "Constraint-based Recommender Systems: Technologies and Research Issues". *ICEC '08 Proceedings of the 10th international conference on Electronic Commerce*. Austria, Innsbruck. Art.3
- . Formoso López, V., (2013). "Técnicas eficientes para la recomendación de productos basadas en filtrado colaborativo". Tesis de Doctorado. Universidad da Coruña. Departamento de Tecnoloxias da Información e as Comunicacions. [En línea]. Disponible en: www.tdx.cal/handle/10803/120245
- . Galan, N., (2007). "Filtrado colaborativo y sistemas de recomendación". [En línea] Disponible en : <http://www.it.uc3m.es/jvillena/irc/practicas/06-07/31.pdf>
- . Gallego, D., Fumero, A. y G. Huecas, (2013). "Proactividad y Contextualización". *El*

profesional de la información. Vol. 22, número 1, pp.19-25

. de Gemmis, M., Lops, P., Semeraro, G. y C. Musto., (2015). "An investigation on the serendipity problem in recommender systems". En *Information Processing and Management*. Vol. 51, pp.695-717.

. Goldberg, D., Nichols, D., Oki, B. y D. Terry., (1992) "Using collaborative filtering to weave an information tapestry". *Communications of the ACM*, special issue on information filtering, Vol. 35, núm 2, pp.61-70.

. Herlocker, J., Konstan, L.G., Terveen, L.G. y J.T Riedl, (2004). "Evaluating Collaborative Filtering Recommender Systems". *ACM Transactions on Information Systems*. Vol 22, número 1, pp.5-53

. Herrera-Viedma, E., Porcel, C. y L. Hidalgo., (2004). "Sistemas de recomendaciones: herramientas para el filtrado de información en Internet". [En línea]. Disponible en www.hipertext.net, número 2, 2004.

. Herrera-Viedma, E., Alonso, S., Chiclana, F. y F. Herrera., (2007). "A Consensus Model for Group Decision Making with Incomplete Fuzzy Preference Relations". *IEEE Transactions on Fuzzy Systems*. Vol. 15, número 5. pp. 863-877.

. Jannach, D., Zanker, M., Felfernig, A. y G. Friedrich., (2011). "*Recommender Systems: An Introduction*". New York: Cambridge University Press.

. Klasnja-Milicevic, A., Ivanovic, M. y A. Nanopoulos, (2015). "Recommender systems in e-learning environments: a survey of the state of the art and possible extensions" En *Artificial Intelligence Review*. Vol. 44, número4. pp. 571-604

. Konstan J.A. y J. Riedl., (2012). "Recommender systems: from algorithms to user experience. *User Modeling and User-Adapted Interaction*. Vol, 1,2., pp.101-123.

. Lops, P., de Geminis, M. y P. Kantor., (2011). "Content-based Recommender Systems. State of the Art and Trends" en Ricci (ed) *Recommender Systems Handbook*. Boston: Springer Science Business Media.

. Lu, J., Wu, D., Mao, M., Wang, G. y G. Zhang., (2015). "Recommender systems application developments: A Survey". *Decision Support Systems*. 74. pp.12-32.

. Maes, P., (1994). "Agents that reduce work and information overload. En *Communications of the ACM*. Vol. 37, número 2.

. Martin, F., Donaldson, J., Ashenfelter, A., Torrens, M. y R. Hangartner., (2011). "The big promise of recommender systems" en *AI Magazine*. Vol. 32, número 3. pp. 19-27.

. Melville, P., Mooney, R.J. y R. Nagarajan. (2002). "Content-boosted collaborative filtering for improved recommendations". *Proceedings of the 18th International Conference on Artificial Intelligence*, Canada. 2002, pp.187-192.

. Mobasher, B., Burke, R., Bhaumik, R. y JJ. Sandvig., (2007). "Attacks and Remedies in Collaborative Recommendation". *IEEE Intelligent Systems*. Vol. 22, número 3. pp. 56-63.

. Nguyen, H y E. Santos, (2013). "Híbrido user model for capturing a user's information seeking Intent. En *Multimedia Services in Intelligent Environments: Advances in Recommender Systems*. Tsirintzis, J.(Ed). Heidelberg: Springer International Publishing.

- . Panniello U. y M. Gorgoglione, (2012). "Incorporating context into recommender systems: an empirical comparison of context-based approaches" . *Electronic Commerce Research*. Vol. 12, núm 1. pp.1-30.
- . Park, D., Kim, H-K., Choi, I-Y. y J.Kim, (2012). " A Literature review and Classification of Recommender System Research ". *Expert Systems and Applications*. Vol. 39
- . Pascual, F., Chaparro, J. y A. Fumero, (2011). "Presente y futuro de los sistemas recomendadores en la Web 2.0". *El Profesional de la Información*. Vol. 20, número 6, pp. 645-651.
- . Pazzani, M., (1999). "A Framework for collaborative, Content-Based and Demographic Filtering". *Artificial Intelligent Review*. Vol, 13, pp. 393-408.
- . Perugini, S., Gonçalves, M. y Edward A. Fox, (2004). "Recommender Systems Research: A Connection-Centric Survey". *Journal of Intelligent Information Systems*. Vol. 23, número 2, pp. 107-143.
- . Porcel Gallego, C., (2005). *Sistemas de acceso a la información basados en información lingüística difusa y técnicas de filtrado*. Tesis de Doctorado. Granada, Escuela Técnica Superior de Ingeniería Informática. Universidad de Granada.
- . Porcel, C. y E. Herrera-Viedma., (2010). "Dealing with incomplete information in a fuzzy linguistic recommender system to disseminate information in university digital libraries". *Knowledge-based Systems*. Vol. 23. pp. 32-39.
- . Rajabpor, N., Bardsiri, V., Mohammadighavam, A. y E. Molaei, (2014). "Application domain of recommender systems: a survey". *International Journal of advanced studies in Computer Science and Engineering*. Vol. 3, número 2, pp. 8-14.
- . Rashid, A.M., , Albert, I., Cosley, D., Lam, S., McNee, S., Konstan, J y J Riedl (2002). "Getting to know you: Learning New User Preferences in Recommender Systems. International Conference on Intelligent User Interfaces. Canada: San Francisco. 2002. pp.127-134
- . Resnick, P. y H. Variant, (1997). "Recommender Systems". *Communications of the ACM*. Vol. 40, Número 3.
- . Ricci, F., Rokach, L. y B. Shapira., (2011). "Introduction to recommender Systems Handbook". En *Recommender Systems Handbook*. pp.1-35. Boston: Springer Science+Business Media.
- . Rodríguez, R (2010),"Using linguistic incomplete preference relations to cold start recommendations", en *Internet Research*, Vol. 20 Iss 3 pp. 296 -315
- . Safoury, L. y A. Salah, (2013). "Exploring user demographic attributes for solving cool-start problem in recommender systems". *Lecture notes on software Engineering*. Vol 1, número3.
- . Salton, G., (1989). "Automatic Text Processing: the transformation, analysis and retrieval of information by computer". Reading, Mass: Addison-Wesley.
- . Salton, G. (ed)., (1971). *The SMART Retrieval Systems: Experiments in Automatic Document Processing*. Prentice-Hall

- . Shani, G. y A. Gunawadana., (2011). "Evaluating Recommendation Systems". en Ricci (ed). *Recommender Systems Handbook*. pp. 257-297. Boston: Springer Science+Business Media
- . Shardanand, U. y P. Maes, (1995). "Social Information Filtering: Algorithms for Automating 'Word of Mouth'". Proceedings of the Conference on Human Factors in Computing Systems. USA, Denver.
- . Trujillo, M., Millan, M. y E. Ortiz., (2007). "A recommender system based on multi-features". En Heidelberg, S.B. (Ed.), *Computational Science and Its Applications – ICCSA 2007*, Springer-Verlag, Berlin, pp. 370-82.
- . Victor, P., De Cock, M., Cornelis, C. y A.M. Teredesai, (2007). "Getting cold start users connected in a recommender system's trust network". *Computational Intelligent in Decision and Control*. pp.877-882.
- . Woerndl, W., Hueber, J., Bader; R. y D. Gallego-Vico. "A model for proactivity in mobile, context-aware recommender systems". En Proceedings of the fifth ACM Conference on Recommender Systems. pp. 273-276.
- . Yamaba, H., Tanoue, M., Takatsuka, K., Okazaki, N. y S. Tomita., (2013). "On a serendipity-oriented recommender system based on folksonomy". En *Artificial life and Robotics*. Vol, 18, número 1, pp.89-94
- . Zenebe, A. y A. Norcio., (2008). "Representation, similarity measures and aggregation methods using fuzzy sets for content-based recommender systems. *Fuzzy Sets and Systems*. 160 (2008). pp. 76-94.
- . Zhen, L., Huang, G.Q. y Z. Jiang., (2010). "An inner-enterprise Knowledge recommender system". *Expert Systems with Applications*. Vol. 37, número 2. pp. 1703-1712.
- . Zhou, W., Koh, Y.S., Wen, J., Alam, S. y G. Dobbie., (2014). "Detection of abnormal profiles on group attacks in recommender systems". *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*. pp. 955-958.
- . Zhou, X., Xu, Y., Li, Y., Josang, A. y C. Cox., (2012). "The state of the art in personalized recommender systems for social networking". *The Artificial Intelligence Review*. 37.2. pp119-132.