

Capítulo 2. Sistemas de Recomendación



| | | | | | | | | | | | |
|------------------------|-------------------------|------------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-----------------------------|
| índice | figuras | tablas | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | referencias |
|------------------------|-------------------------|------------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-----------------------------|

Los sistemas de recomendación surgen a partir de la necesidad de poder proveer a los usuarios de información relevante y personalizada. Es decir, como usuario, a uno le gustaría que el periódico de cada mañana consistiera solamente de la clase de artículos que se acostumbra a leer o simplemente que son de nuestro interés. O bien, que a partir de mis hábitos e historial de lectura, se me pudieran recomendar qué libros leer. Ejemplos como éstos, han movido a investigadores de distintas áreas de la computación a desarrollar nuevas herramientas que permitan la construcción de estos sistemas.

Este capítulo realiza una descripción general sobre los sistemas de recomendación, los diferentes mecanismos y algoritmos empleados para la generación de recomendaciones, así como el trabajo relacionado realizado hasta este momento.

2.1 Descripción y contexto

2.2 Modelado de las necesidades de información del usuario

2.3 Mecanismos de recomendación

2.4 Mecanismos de retroalimentación

2.5 Trabajo relacionado

2.1 Descripción y contexto

Wang [1998] define a un sistema de recomendación como “aquél sistema que tiene como principal tarea seleccionar ciertos objetos, de acuerdo a los requerimientos del usuario, dado que estos objetos están almacenados y caracterizados, en base a sus atributos”. Tales objetos pueden ser libros, discos compactos de música, programas de televisión, artículos de periódicos, restaurantes, películas, entre otros.

Evidentemente, en un sistema rudimentario el usuario podría fácilmente consultar la información de alguna base de datos, dado un criterio de búsqueda (*query*), y pareciera que el problema ha sido resuelto. Sin embargo, la experiencia muestra que los usuarios se enfrentan ahora a un nuevo problema: aprender cómo expresar sus necesidades de información en términos de un criterio de búsqueda, que en la mayoría de los casos no es del todo preciso [Belkin 2000; Dharap 1998; Wang 1998]. Podemos percatarnos, entonces, que se deben de generar alternativas a este mecanismo de búsqueda de información. Y es precisamente un sistema de recomendación el que “oculta” ese criterio de búsqueda explícito, sin eliminarlo, y se trabaja ahora en el modelado del usuario, que al final se traducirá en ese criterio de búsqueda, que es generado automáticamente, sin intervención del usuario final.

Esta posición tal vez difiera sutilmente con la asumida por Balabanovic [1998a], dado que argumenta que en la tarea de recomendar el criterio de búsqueda es del todo prescindido. Pero, veremos en los capítulos subsecuentes que esto ha de depender del contexto de la aplicación y de las características propias del conjunto de recomendación, es decir, los objetos a ser recomendados.

2.2 Modelado de las necesidades de información del usuario

Un reto que se presenta indudablemente al momento de crear un sistema de recomendación, es el de cómo modelar al usuario con sus necesidades de información, que no han de permanecer estáticas, y que por tanto habrá que simular ese muy posible dinamismo. Balabanovic [1998a] sugiere que más que modelar al usuario, se tiende a modelar los objetos en los que éste mostró, muestra o posiblemente mostrará interés. Sin embargo, existen otros enfoques muy interesantes sobre este tema, como lo es la teoría de los estereotipos [Balabanovic 1998a; Brajnik et al. 1987; Rich 1979], que no sólo definen las necesidades de información del usuario de una u otra forma, sino que además tomaría en cuenta datos como su nivel de educación, religión, edad, entre otros, llevando con esto a generar recomendaciones que posiblemente sean más efectivas. La investigación se encuentra abierta, mientras tanto la propuesta de modelar las necesidades de información del usuario es plausible y sobre todo efectiva, y por consiguiente muy empleada. En efecto, los trabajos relacionados que en su momento se describirán, no son la excepción. La presente tesis retoma el mismo concepto y lo aplica al contexto de bibliotecas digitales.

Como se ha mencionado, el contexto para el cual el sistema de recomendación es creado, influye totalmente en el modelado del usuario. En un sistema de recomendación de películas, por ejemplo, tal vez sólo sea necesario mantener un registro de las películas que a cierto usuario le han agradado –puesto que así lo ha manifestado al sistema– y recomendarle en un futuro películas del mismo estilo. Mientras que en una biblioteca, donde un usuario requiere información sobre algún tema potencialmente nuevo para él, el sistema no puede solicitarle evaluar libros sobre tal tema, pues preferiría solicitar la recomendación de viva voz de algún experto en el área; pero el sistema sí podría llevar un historial de qué usuarios ya han evaluado libros que hablen del tema en cuestión, y recomendarlos al usuario inexperto hasta entonces.

El siguiente apartado muestra los dos principales mecanismos para la generación de recomendaciones, así como la forma en que se modelan los usuarios.

2.3 Mecanismos de recomendación

En la actualidad encontramos principalmente dos mecanismos bien definidos para llevar a cabo la tarea de generar recomendaciones: *el basado en contenido y el colaborativo* .

2.3.1 Recomendación basada en contenido

Balabanovic y Shoham [1997] describen a un sistema de recomendación basado en contenido, como “aquél sistema en el que las recomendaciones son realizadas basándose solamente en

Capítulo 2. Sistemas de Recomendación

un perfil creado a partir del análisis del contenido de los objetos, que el usuario ha evaluado en el pasado.” Básicamente, un sistema de recomendación que use un mecanismo basado en contenido para generar recomendaciones, debe ser capaz de modelar y representar qué objetos han sido de utilidad o agrado del usuario, para que de esta forma, este último reciba subsecuentemente una serie de objetos similares. Por ejemplo, un sistema de recomendación de discos compactos de música, puede recomendar un álbum nuevo al usuario, ya que en ocasiones pasadas, éste había evaluado positivamente dos álbumes de artistas con similar tipo de música. El ejemplo puede resultar muy obvio, o incluso ingenuo, pero nos da la pauta que debe seguir el mecanismo que estamos analizando.

En este momento, surgen varios cuestionamientos: ¿Cómo representar los objetos a recomendar? ¿Qué métricas nos ayudaran a saber el grado de similitud entre dos o más objetos? ¿Cómo analizar automáticamente el contenido de tales objetos? ¿Cómo mantener un perfil o historial del usuario con respecto a esos objetos? De todas estas interrogantes podemos destacar dos términos importantes: representación y aprendizaje.

Diversos trabajos de investigación han tratado de responder estos cuestionamientos. Uno de los enfoques más empleados es el conocido como *Espacios Vectoriales* [Salton y McGill 1983], que tiene sus raíces en el área de *Recuperación de Información* (*Information Retrieval*) y que representa el carácter temático de los objetos, como un vector de términos. Otro enfoque hace uso de diferentes técnicas del área del *Aprendizaje Automatizado* (*Machine Learning*) como son las *Redes Bayesianas* [Sahami 1998] y derivados de ésta como lo son la *Clasificación* (*Classification*) y la *Agrupación* (*Clustering*). Los dos enfoques han mostrado ser efectivos [Balabanovic 1998a]. En general, el modelo de espacios vectoriales es preferido por su simplicidad.

2.3.1.1 Espacios Vectoriales

El modelo de espacios vectoriales [Sahami 1998; Balabanovic 1998a; Salton y McGill 1983] ha sido empleado repetidas ocasiones para la representación de documentos en el campo de la recuperación de información. A pesar de que este modelo sólo llega a representar el carácter temático de los documentos [Balabanovic 1998a], su aplicación ha resultado exitosa en la mayoría de las ocasiones.

Fundamentalmente, cada documento se representa como un vector buliano o numérico, en un espacio de dimensión p , donde cada dimensión corresponde a un *término* diferente en el conjunto o colección de documentos a caracterizar [Sahami 1998; Balabanovic 1998a]. Un *término* se refiere a una secuencia alfanumérica delimitada por espacios en blanco o signos de puntuación [Sahami 1998], que no es otra cosa que una palabra. De manera más formal, tenemos:

Sea $D = \{d_1, d_2, d_n\}$ el conjunto o colección de documentos.

Sea $\{t_1, t_2, t_p\}$ el conjunto de palabras o diccionario de toda la colección de documentos.

Cada documento \mathbf{d} es representado como el vector de dimensión p , $\mathbf{d} = [d_1, d_2, \dots, d_p]$ donde d_i es el peso de la palabra t_i en el documento \mathbf{d} .

Podemos percatarnos que el peso de cada término para cierto documento \mathbf{d} puede variar. Este peso esta determinado, en la mayoría de los casos, por la frecuencia con la que el término

Capítulo 2. Sistemas de Recomendación

aparece en el documento. En la literatura encontramos distintas funciones que llevan a cabo tal tarea. La más difundida, y tal vez empleada, es la propuesta por Salton y Buckley [1988], mejor conocida como "TFIDF": $TFIDF(t_i, d) = TF(t_i, d) \cdot IDF(t_i)$. $TFIDF(t_i, d)$ indica el peso asignado al término t_i en el documento d . Donde $TF(t_i)$ representa la frecuencia del término o palabra t_i en el documento d ; $IDF(t_i)$, la frecuencia inversa del término t_i en el documento d .

Donde tf_i es el número de ocasiones en que t_i aparece en el documento d ; df_i , es el número de documentos de la colección de documentos que contienen el término t_i ; f_{max} , es la máxima frecuencia de un término por sobre todos los demás términos.

La asignación de pesos, sin embargo, puede ser menos sofisticada y mucho más simple de lo que parece. Alternativamente, podemos emplear tan solo valores binarios, y así tener vectores en los que los pesos vendrían determinados por la ausencia o presencia del término t_i en el documento d [Sahami 1998]. Esto es, la función $f(t_i, d)$, que determina el peso del t_i para el documento d , está definido por el número de ocurrencias λ , en que t_i se encuentra en d :

$$f(t_i, d) = \begin{cases} 1 & \text{si } \lambda > 0 \\ 0 & \text{si } \lambda = 0 \end{cases}$$

Ahora bien, para encontrar la similitud entre dos documentos d_1 y d_2 , calculamos el coseno del ángulo entre los dos vectores de los documentos. Esto puede ser expresado de la siguiente forma: $SIM(d_1, d_2) = d_1 \cdot d_2$

2.3.2 Recomendación colaborativa

Un sistema de recomendación colaborativo puede definirse como "aquél sistema en el que las recomendaciones son hechas basándose solamente en los términos de similitud entre los usuarios" [Balabanovic y Shoham 1997], es decir, no se realiza ningún tipo de análisis de los objetos a recomendar.

Algunos autores se refieren también a la recomendación colaborativa como *filtrado colaborativo* (*collaborative filtering*) [Herlocker et al. 2000; Sarwar 2000; Grasso et al. 1999; Good 1999; Billsus y Pazzani 1998; Konstan et al. 1997; Terveen et al. 1997; Avery y Zeckhauser 1997; Shardanand y Maes 1995]. La idea principal de este mecanismo se basa en que la tarea de realizar las recomendaciones, se lleva a cabo por los mismos usuarios [Terveen et al. 1997]. Se puede afirmar que este mecanismo logra adaptar fielmente el proceso natural de recomendar realizado por los seres humanos, logrando automatizarlo [Shardanand y Maes 1995].

Pero, ¿cómo lograr, entonces, automatizar ese proceso natural de recomendar? Para responder a este cuestionamiento, la mayoría de los investigadores parten de la suposición de que existen tendencias o patrones en los gustos o preferencias de las personas, esto es, que no se encuentran distribuidas de manera aleatoria [Shardanand y Maes 1995]. En otras palabras, nuestra meta es encontrar patrones entre las preferencias de las personas, puesto que se ha observado que esto suele presentarse con mucha frecuencia. De esta forma, se realizarían comparaciones de preferencias o gustos de los usuarios en un campo de aplicación determinado, con el propósito de hallar grupos de personas con características similares.

Capítulo 2. Sistemas de Recomendación

Shardanand y Maes [1995] resumen el proceso que sigue la recomendación colaborativa o el filtrado colaborativo en tres pasos:

1. El sistema mantiene un *perfil de usuario*, que contiene tanto evaluaciones positivas como negativas sobre objetos que han sido de interés para el usuario.
2. Se compara este perfil con los perfiles de los usuarios restantes, para determinar el grado de similitud que existe entre ellos.
3. Habiendo encontrado los perfiles con más similitud, se recomiendan objetos que el usuario no haya evaluado.

Para entender mejor este mecanismo apliquémoslo, a manera de ejemplo, a un sistema de recomendación de películas. Entonces, supongamos que A y B son usuarios de este sistema. El usuario A ha evaluado positivamente a “Los Ángeles al desnudo”, “El silencio de los inocentes”, “Expedientes X: la película”, e hizo lo mismo con “Loco por Mari”. Por su parte, el usuario B evaluó de igual forma todas las películas que el usuario A, excepto por “Loco por Mari”. El sistema debe ser capaz de notar que A y B tiene gustos similares (o al menos en cierta medida), y recomendar al usuario B “Loco por Mari”, puesto que este último no la ha visto aún y al usuario A le agradó.

En cierta forma, en el ejemplo anterior notamos el porqué de la recomendación de la película “Loco por Mari”. Usando el sentido común, agrupamos las tres primeras películas en la categoría de drama–suspense (suponiendo que en verdad entran en esta categoría), y la última en comedia. Los usuarios A y B gustan de películas de drama–suspense, pero a A parece gustarle también cierto tipo de comedia, y tal vez B no ha considerado probar dicha categoría; todo parece indicar que podría gustarle. Ahora bien, ¿cómo conocer, de manera automática, el grado de similitud –en el caso anterior gustos por películas– entre dos usuarios? El siguiente apartado introduce unas de las métricas que más se emplean para llevar a cabo esta tarea.

2.3.2.1 Coeficiente de Pearson

Una de las métricas más empleadas para determinar la similitud entre dos usuarios a e i , es el coeficiente de *Pearson* [Good et al. 1999; Fu et al. 2000; Breese et al. 1998; Konstan et al. 1997; Shardanand y Maes 1995]:

$$w(a, i) = \frac{\sum_j (v_{a,j} - \bar{v}_a)(v_{i,j} - \bar{v}_i)}{\sqrt{\sum_j (v_{a,j} - \bar{v}_a)^2 \sum_j (v_{i,j} - \bar{v}_i)^2}}; \text{ donde}$$
$$\bar{v}_i = \frac{1}{|I_i|} \sum_{j \in I_i} v_{i,j}$$

Donde $v_{a,j}$ representa el voto o calificación del objeto j por parte del usuario a ; $v_{i,j}$, el voto o calificación del usuario i al objeto j ; entiéndase \bar{v}_i , como el voto o calificación promedio del usuario i , por sobre los objetos que haya evaluado, siendo I_i este conjunto de objetos.

2.4 Mecanismos de retroalimentación

Cualquier sistema de recomendación serio, debe contar con un mecanismo de retroalimentación, que ayude a mejorar el nivel de calidad de las recomendaciones emitidas al usuario. Informalmente definamos una taxonomía para estos mecanismos, determinadas

Capítulo 2. Sistemas de Recomendación

por la carga cognitiva en el usuario. Tenemos, entonces, dos tipos de retroalimentación: *la retroalimentación implícita y la retroalimentación explícita*.

La retroalimentación implícita, se refiere a que el usuario no percibe directamente que el sistema está siendo evaluado mediante acciones ejercidas por él. Cuando el usuario emplea más tiempo en la lectura de un artículo en un sistema de recomendación para un periódico virtual, indica en cierta forma que le es de interés, y aunque no lo haya evaluado directamente, ya habrá retroalimentado al sistema. Publicaciones que ahondan sobre esto son [Balavanovic 1998a; Balavanovic 1998b; Das y ter Host 1998; Oard y Kim 1998].

La retroalimentación explícita, por otra parte, requiere que el usuario evalúe el sistema mediante la asignación de una calificación a los objetos recomendados, ya sea individualmente (un libro, una película, un artículo) o en conjunto (serie de libros sobre un tema). [Balavanovic 1998a; Konstan et al. 1997; Terveen et al. 1997] son publicaciones que profundizan en este tópico.

Es válido, cuando se diseña un sistema de recomendación, contemplar los tipos de mecanismos de retroalimentación; y recomendable, emplearlos. Se debe tomar mucho en cuenta que el usuario tiende a no retroalimentar los sistemas de recomendación explícitamente, por diferentes razones (desidia, desinterés, etc.), no importando la mínima complejidad que esto requiera. Es por esto, que los mecanismos de retroalimentación implícita toman un papel trascendente.

2.5 Trabajo relacionado

Los campos para los cuales se decidió crear sistemas de recomendación son muy diversos: el entretenimiento, la investigación científica, el comercio, la industria, entre otros. A continuación se describen de manera concisa algunos proyectos que por su aportación, se han de tomar en cuenta para futuros desarrollos en esta área de investigación aplicada, como lo es caso de la presente tesis.

2.5.1 *Sistemas de Recomendación de libros*

LIBRA [Mooney 2000; Mooney et al. 1998] es un sistema de recomendación de libros implementado en la Universidad de Texas como prototipo. La idea es explorar mecanismos que permitan explotar información semi-estructurada perteneciente a colecciones de libros. La información es extraída de la compañía en línea *Amazon*, que proporciona descripciones de libros solicitados. Los algoritmos empleados para el aprendizaje requieren evaluar ejemplos previamente. Este esquema de aprendizaje, como se explicará, no parece del todo convincente en nuestro contexto, pues compartimos la filosofía del mínimo esfuerzo para el usuario.



Capítulo 2. Sistemas de Recomendación

Figura 2.1 Página inicial de LIBRA



Figura 2.2 LIBRA recomendando libros sobre literatura

Mooney [2000] propone el uso de sistemas de recomendación como servicios en las bibliotecas digitales, puesto que tendrán un fuerte impacto sobre los usuarios, ya de esta forma se les proporcionaría acceso a información personalizada. Esta consideración concuerda definitivamente con las proposiciones hechas por esta tesis. A pesar de esto, LIBRA no forma parte, hasta este momento, de los servicios ofrecidos por alguna biblioteca digital, pero se encuentra a modo de prueba en el sitio <http://www.cs.utexas.edu/users/libra/home.html>. La figura 2.1 muestra la página inicial de LIBRA, en donde el usuario tiene las opciones de: seleccionar un genero en el cual se está interesado; evaluar libros de los cuales ya se tiene conocimiento en particular; revisar sus evaluaciones hechas; y finalmente el de recibir recomendaciones. Al recibir recomendaciones de libros (figura 2.2), el usuario tiene la oportunidad de evaluarlos, asignándoles a los libros diferentes calificaciones, de una escala posible del 1 al 10.

2.5.2 Sistemas de Recomendación enfocados a Internet

Existen diferentes sistemas de recomendación enfocados al material disponible en Internet. Resaltan sistemas que recomiendan páginas electrónicas como Fab [Balavanovic 1998a; 1998b] y *Science Forum* [Bienkowski 1998], y de contenido general en Web como el trabajo descrito en [Imudom y Neuman 1998]. Todos estos sistemas describen interesantes procedimientos, algo-ritmos y consideraciones a tomar en cuenta en el proceso de creación de un sistema de esta índole.

Fab

La presente tesis toma gran parte de la investigación presentada en la disertación doctoral de Balavanovic [1998a] y la lleva a la práctica. Fab es un sistema de recomendación de páginas electrónicas que alude a la concepción de un periódico virtual, que se adecua a los gustos de lectura propios de un usuario –este sistema tiene como trabajo precedente a LIRA (figura 2.3) implementado en 1994. En este estudio se detallan todos los pasos que se siguieron para la creación de Fab, desde la recolección de información hasta la evaluación de los elementos recomendados. Pero, sin temor a errar, la más grande aportación hecha por esta disertación, es haber propuesto combinar los dos mecanismos existentes para la generación de recomendaciones: el basado en contenido y el colaborativo, para tratar de eliminar los problemas y des-ventajas que acarrearía el usar uno solo. En los capítulos 3, 4 y 5, se explica más a fondo sobre esto.

Capítulo 2. Sistemas de Recomendación



Figura 2.3 Interfaz de LIRA (a) y Fab (b), ambas mostrando recomendaciones sobre páginas electrónicas [Balavanovic 1998a]

Science Forum

Science Forum [Bienkowski 1998] es un sitio en Web, que cuenta con un subsistema de recomendación de fuentes de información en la red, con la finalidad de ayudar a académicos, desde preescolar hasta el grado 12, en su actividad de enseñanza.

Esquema del Instituto de Ciencias de la Información de la Universidad de California del Sur

Imudon y Neuman [1998] más que implementar un sistema de recomendación de material en Web, proponen un esquema general basado en un mecanismo colaborativo, para recomendar cualquier tipo de material, aludiendo a ciertas limitantes de sistemas preexistentes. Algunas limitantes, mencionan, se refieren a la carencia de un modelo general para explotar los sistemas comercialmente, la credibilidad de las recomendaciones y de los evaluadores del sistema.

La noción de “experto” introducida por su investigación, es bastante atractiva, y es una de las propuestas para resolver el problema de credibilidad antes mencionado –algo muy relacionado con el concepto de *líder de opinión*, a comentar en el siguiente apartado. El objetivo, entonces, es identificar a aquellos usuarios que poseen un grado más amplio de conocimiento, y hacer que éstos proporcionen además de evaluaciones, comentarios o anotaciones. Esto es de gran ayuda para el resto de la comunidad de usuarios, pues les ofrece un grado más alto de certidumbre respecto a la efectividad de las recomendaciones en un mecanismo colaborativo.

2.5.3 *Sistemas de Recomendación enfocados al entretenimiento*

Algunos investigadores y desarrolladores han llevado la idea potencial de la recomendación a la industria del entretenimiento. Son destacables dos sistemas de recomendación de programas de televisión: *TV Advisor* [Das y ter Horst 1998] y el publicado por Baudisch [1998]; DIVA [Nguyen y Haddawy 1998], un sistema de recomendación interactivo de videos.

TV Advisor

TV Advisor [Das y ter Horst 1998] es un prototipo de un sistema de recomendación de programas de televisión, desarrollado en el laboratorio de investigación de Phillips, en Eindhoven, Holanda. Este sistema se basa en un el perfil de gustos y disgustos del usuario en el contexto de programas de televisión.

El sistema se compone de dos componentes o subsistemas principales: el de recuperación de información referente a la programación de programas de televisión de las principales

Capítulo 2. Sistemas de Recomendación

televisoras holandesas, y el de recomendación de programas.

Algunos aspectos considerados en la investigación señalan puntos importantes a considerar en la construcción de un sistema de recomendación: el mecanismo de retroalimentación, los tipos de usuario y la escala de calificación de los objetos recomendados.

Sistema de Recomendación en GMD

En el Centro Nacional de Investigación Alemán para la Tecnología de la Información (GMD), se lleva a cabo también investigación sobre sistemas de recomendación de programas de televisión. El artículo publicado por unos de sus investigadores, Baudisch [1998], menciona interrogantes que la mayoría de los creadores de sistemas de recomendación se hacen con frecuencia. Uno de ellos, y que se ha comentado, es el de qué tanto el usuario está dispuesto a evaluar las recomendaciones. Ciertamente, para este tipo de sistemas orientados a la televisión es un problema, dado que la actividad en sí es netamente pasiva como se sabe, y debe requerirse por parte del usuario un esfuerzo casi nulo. Esto nos lleva de nueva cuenta a reafirmar que la idea de que no existe una arquitectura preestablecida y suficientemente definida para cualquier sistema de recomendación.

De interés para esta tesis, y que sugiere Baudisch [1998], es la introducción del término *líderes de opinión* (*opinion leaders*). Dicho término se refiere a que la opinión de ciertos usuarios pudieran tener más peso que la del resto. Estos usuarios *líderes* tendrían que interactuar más tiempo con el sistema, y por tanto, lo evaluarían con mucho más conciencia. Como veremos en los siguientes capítulos, se adaptará esta interesante propuesta al trabajo presente.

DIVA

DIVA [Nguyen y Haddawy 1998], es un sistema interactivo de videos desarrollado en el laboratorio de Sistemas de Decisión e Inteligencia Artificial de la Universidad de Wisconsin–Milwaukee. La finalidad es ofrecer al usuario una lista de películas basándose en su perfil, y evitar que busque por su cuenta a través de máquinas de búsqueda o catálogos en línea con miles de títulos. El reto para este grupo, como suele suceder, es el de cómo representar las preferencias del usuario, y al mismo tiempo tener flexibilidad con respecto a los cambios en éstas. A diferencia de otros sistemas, DIVA emplea algoritmos de aprendizaje que requieren de una serie de ejemplos previos para comenzar a generar recomendaciones.

2.5.4 Sistemas de Recomendación comerciales

Al momento de escritura de esta tesis, se encuentran disponibles varios sistemas de recomendación que son explotados comercialmente. Estos sistemas son: *MovieLens*, *Amazon* y *CD–NOW*. Sin embargo, su funcionamiento no es del todo descrito en la literatura, por razones de negocios.

MovieLens

MovieLens (<http://www.movielens.com>) es un sistema de recomendación de películas en Web. Toda su tecnología empleada proviene del grupo *GroupLens* [Good et al. 2000; Herlocker et al. 2000; Konstan et al. 1997], la cual se encuentra a la venta. Las figuras 2.4 y 2.5 muestran a este sistema en acción. Al igual que la mayoría de los sistemas de

Capítulo 2. Sistemas de Recomendación

recomendación, se solicita al usuario evaluar ítems, en este caso películas, para poder recomendar aquellas que se asemejen al gusto mostrado del usuario.



Figura 2.4 MovieLens solicitando al usuario evaluar algunas películas al inicio



Figura 2.5 MovieLens mostrando algunas recomendaciones de películas

Amazon

Amazon (<http://www.amazon.com>) es una de las tiendas departamentales Web más grandes en Web, y explota en buena forma los beneficios de un sistema de recomendación. La intención de *Amazon* es recomendar a los compradores virtuales posibles productos de su interés, pues se tiene registro de las compras realizadas en el pasado. Es muy posible que empleen minería de datos (*data mining*) o almacenamiento de datos (*data warehousing*) para identificar similitudes y patrones entre los compradores por región, sexo, raza, edad, etc., para que de esta manera se pueda generar recomendaciones sobre posibles compras. Este enfoque novedoso para sistemas de recomendación se puede también encontrar en [Chen 1998]. La figura 2.6 muestra a *Amazon* dando recomendaciones sobre libros acerca de los Expedientes X, ya que anteriormente se compraron artículos relacionados con esta serie de televisión.

CD-NOW

CD-NOW (<http://www.cdnw.com>) es una de las tiendas más grandes de música en Web. Al visitar la tienda, siendo o no comprador, es posible recibir recomendaciones sobre qué discos comprar, de acuerdo a nuestro perfil, que contiene nuestras preferencias sobre música. La figura 2.7 muestra a *CD-NOW* recomendando álbumes de Natalie Imbruglia, ya que su tipo de música concuerda con la especificada en el perfil del usuario.



Figura 2.6 Amazon mostrando recomendaciones sobre libros

Capítulo 2. Sistemas de Recomendación



Figura 2.7 CD-NOW recomendando discos compactos



Ramírez Vargas, M. 2001. Services de recomendación en la biblioteca digital de la UDLA-P. Tesis Licenciatura. Ingeniería en Sistemas Computacionales. Departamento de Ingeniería en Sistemas Computacionales, Escuela de Ingeniería, Universidad de las Américas-Puebla. Mayo. Derechos Reservados © 2001, Universidad de las Américas-Puebla.

