

DIPLOMADO EN **BIG DATA** PARA LA TOMA DE DECISIONES

Introducción al Modelamiento Estadístico

Parte 4: Sistemas de Recomendación

4.1.- INTRODUCCIÓN

Introducción

- Los sistemas de recomendación cambiaron la forma en que los sitios web, antiguamente estáticos, se comunican con sus usuarios.
- En lugar de proporcionar una experiencia estática en la que los usuarios, por ejemplo, buscan y, potencialmente, compran productos, los sistemas de recomendación aumentan la interacción para proporcionar una experiencia más rica.
- Los sistemas de recomendación generan de forma automática recomendaciones para los usuarios individuales, sobre la base del comportamiento anterior del usuario y del comportamiento de otros usuarios.

Introducción

- Algunos ejemplos de sistemas de recomendación utilizados en sitios web:
 - **LinkedIn:** Crea recomendaciones para personas que tal vez conozcas, puestos de trabajo que te pueden gustar, grupos que te pudiese gustar seguir, o empresas en las que puedas estar interesado. LinkedIn utiliza un método de **filtrado colaborativo**.

Introducción

●

- **Amazon:** Utiliza recomendación basada en el contenido. Cuando una persona selecciona un elemento a comprar, Amazon recomienda otros artículos que otros usuarios han adquirido, en base al elemento original (construye una matriz de ítems que son probables de ser comprados). Amazon patentó este comportamiento, denominado **filtrado colaborativo punto-a-punto**.

Introducción

●

- **Netflix:** En 2006, Netflix llevó a cabo un concurso para mejorar su sistema de recomendación, Cinematch. En 2009, tres equipos se combinaron para construir un conjunto de 107 algoritmos de recomendación, que dieron lugar a una única predicción. Este conjunto demostró ser la clave para mejorar la exactitud de predicción, y el equipo combinado ganó el premio.

4.2.- MÉTODOS BÁSICOS

4.2.1.- FILTRADO COLABORATIVO

Filtrado Colaborativo

- En un método de filtrado colaborativo, también conocido como método de filtrado social, la información se “filtra” mediante el uso de las recomendaciones de otras personas.
- Se basa en la idea de que las personas que estuvieron de acuerdo en su evaluación de determinados artículos en el pasado, es probable que estén de acuerdo también en el futuro.

Filtrado Colaborativo: Ejemplo 1

- Supongamos que una persona que quiere ver una película y le pide recomendaciones a sus amigos.
- Considere la clasificación de cinco personas a cinco películas dada en la siguiente tabla.

Película	Persona				
	Antonio	Carlos	Alejandro	Carolina	Claudia
The Piano	-	-	+		+
Pulp Fiction	-	+	+	-	+
Clueless	+		-	+	-
Cliffhanger	-	-	+	-	+
Fargo	-	+	+	-	?

Filtrado Colaborativo: Ejemplo 1

- En esta tabla, un “+” indica que a la persona le gustó esa película y un “-” que no.
- Para predecir si a Claudia le gustará la película “Fargo”, la clasificación de Claudia se compara con la de los demás.
- En este ejemplo, la clasificación de Claudia y Alejandro son idénticas para todas las películas y, ya que a Alejandro le gustó Fargo, uno podría predecir que a Claudia le gustará la película también.

Filtrado Colaborativo: Ejemplo 1

- En lugar de utilizar sólo a la persona más similar, la predicción se basa normalmente en el promedio ponderado de las recomendaciones de varias personas.
- El peso dado a cada persona esta determinado por el grado de correlación entre esa persona y el sujeto al que se le quiere predecir el gusto.
- Como medida de correlación usualmente se utiliza al coeficiente de correlación de Pearson.

Filtrado Colaborativo

- Una alternativa también utilizada, denominada como enfoque item-al-item, es una inversión del método del vecino más cercano.
- En lugar de medir la similitud entre las clasificaciones de las personas, éstas son utilizadas para medir la correlación entre los ítemes.
- El coeficiente de correlación de Pearson se utiliza nuevamente como medida de asociación.
- En el Ejemplo 1, las clasificaciones de “Fargo” y “Pulp Fiction” tienen correlación perfecta. Dado que a Claudia le gustó “Pulp Fiction”, entonces la predicción es que le guste “Fargo” también.

Filtrado Colaborativo: Ejemplo 2

- Supongamos ahora que usted está construyendo un sitio web para recomendar blogs.
- Mediante el uso de la información de muchos usuarios que se suscriben y leen blogs, se puede agrupar a los usuarios en función de sus preferencias.
- Por ejemplo, se puede agrupar a los usuarios que leen varios del mismo tipo de blogs.
- A partir de esta información, se identifican los blogs más populares que son leídos por ese grupo.

Filtrado Colaborativo: Ejemplo 2

- Entonces, para un usuario en particular en el grupo se recomienda el blog más popular que él o ella no haya leído o se haya suscrito.
- En la siguiente figura se muestra la cantidad de “ingresos” de cuatro usuarios a diferentes tipos de blogs.

Blogs	Marc	Megan	Elise	Jill
Linux	13	3	11	-
OpenSource	10	-	-	3
Cloud Computing	6	1	9	-
Java Technology	-	6	-	9
Agile	-	7	1	8

Articles read per user

Cluster	1	2	1	2
---------	---	---	---	---

Filtrado Colaborativo: Ejemplo 2

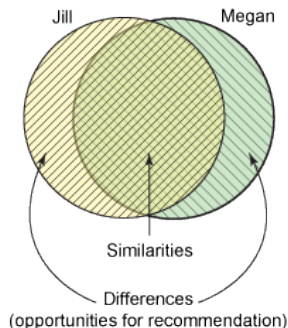
- Al agrupar los usuarios en función de sus hábitos de lectura (por ejemplo, mediante el uso del método del vecino más cercano), se identifican dos grupos de usuarios.
- Marc y Elise leen principalmente artículos sobre Linux y Cloud Computing, y conforman el grupo 1.
- Megan y Jill leen principalmente artículos sobre Java y Agi, conformando el grupo 2.
- Ahora se puede identificar algunas diferencias dentro de cada grupo y hacer recomendaciones significativas.

Filtrado Colaborativo: Ejemplo 2

- En el grupo 1, Marc leer 10 artículos de blogs de código abierto, mientras que Elise no ha leído ninguno. Por otra parte, Elise ha leído un blog de Agile, y Marc no ha leído ninguno.
- No se pueden hacer recomendaciones para Marc porque la diferencia (distancia) entre él y Elise con respecto a la lectura de blogs de Agile es muy pequeña.
- En el grupo 2, Jill ha leído tres blogs de código abierto, y Elise ninguno. Elise ha leído 11 blogs de Linux, y Jill ninguno.
- Se puede recomendar leer un blog de Linux a Jill y un blog de código abierto para Megan.

Filtrado Colaborativo

- Otra forma de ver estas relaciones se basa en sus similitudes y diferencias, como se ilustra en el diagrama de Venn de la siguiente figura:



- Las similitudes definen (en base al algoritmo particular utilizado) los usuarios de un grupo que tienen intereses similares.

Filtrado Colaborativo

- Las diferencias dentro del grupo son las oportunidades que se pueden utilizar para la recomendación, obtenidas, por ejemplo, a través de un filtro de popularidad.

Filtrado Colaborativo: Como un Problema de Clasificación

- El filtrado colaborativo es realmente un problema de clasificación.
- Consideremos el Ejemplo 1, nuevamente.
- El método predictivo tiene que determinar la clase de la película “Fargo” para Claudia.
- El clasificador puede ser entrenado en base a las 4 películas que Claudia ha clasificado.
- “The Piano”, “Pulp Fiction” y “Cliffhanger” son clasificadas como positivas, mientras que “Clueless” es clasificada como negativa.

4.2.2.- FILTRADO BASADO EN EL CONTENIDO

Filtrado Basado en el Contenido

- El filtrado basado en contenido genera la recomendación sobre la base del comportamiento de un usuario.
- Por ejemplo, este enfoque podría utilizar la información histórica de navegación, tales como los blogs, que el usuario ha leído y las características de los blogs.
- Si un usuario lee habitualmente artículos sobre Linux o agregar comentarios en los blogs sobre la ingeniería de software, el filtrado basado en contenido puede utilizar esta historia para identificar y recomendar un contenido similar (artículos sobre Linux u otros blogs sobre la ingeniería de software).
- Este contenido se puede definir de forma manual o en forma automática sobre la base de otros métodos de similitud.

Filtrado Basado en el Contenido

- Consideremos la siguiente figura (nuevamente):

Blogs	Marc	Megan	Elise	Jill
Linux	13	3	11	-
OpenSource	10	-	-	3
Cloud Computing	6	1	9	-
Java Technology	-	6	-	9
Agile	-	7	1	8
Articles read per user				
Cluster	1	2	1	2

- Centremos nuestra atención en Elise.
- En base a ranking del tipo de blog más frecuentemente leído por ella (i.e., sobre la base de sus hábitos de lectura actuales), y en base a una caracterización de contenidos similares, se le podría recomendar algún blog sobre código abierto.

Filtrado Basado en el Contenido

- Este enfoque, que se ilustra en la siguiente figura



- Es importante destacar que este enfoque se basa exclusivamente en el contenido que un único usuario accede, no en el comportamiento de otros usuarios en el sistema.

4.2.3.- MÉTODOS HÍBRIDOS

Métodos Híbridos

- Enfoques híbridos que combinan el filtrado colaborativo y el basado en contenidos han permitido aumentar la eficiencia (y complejidad) de los sistemas de recomendación.
- La incorporación de los resultados de filtrado colaborativo y basado en el contenido puede generar una recomendación más precisa.
- El enfoque híbrido también podría ser utilizado para resolver recomendaciones poco precisas que pueden surgir del filtrado colaborativo cuando parte con pocos datos.
- En éste último caso, se puede ponderar de mayor forma el resultado de un filtrado basado en el contenido inicialmente, para luego reducir su peso en favor de un filtrado colaborativo.

4.3.- LECTURAS SUGERIDAS

Lecturas Sugeridas

- Ricci, F., Rokach, L., Shapira, B. (2011), Introduction to Recommender Systems Handbook, Recommender Systems Handbook, Springer, 2011, pp. 1-35.