
Modelado de Filtrado Colaborativo

Alondra Berzunza



Productos

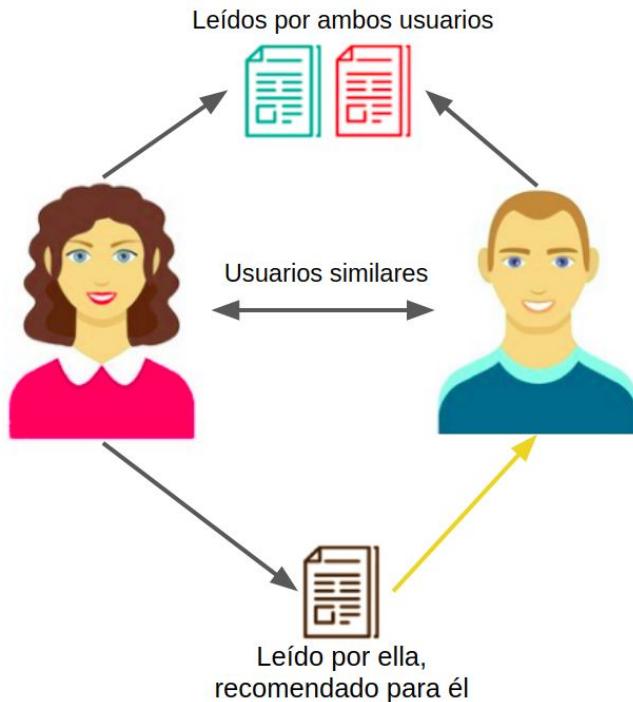
Seleccionar j producto con x_j características
(palabras clave, categoría)



Usuario i con
características x_i
(historial de navegación,
geolocalización, sitios favoritos).

(i,j): **¿Qué producto se le debe recomendar?**
¿El que tiene mayor número de “me gusta”?
¿Los que ha comprado más otro usuario similar a i ?
¿Basado en sus últimas compras?

Filtrado Colaborativo



La similitud entre usuarios puede ser usada para inferir las calificaciones no observadas.

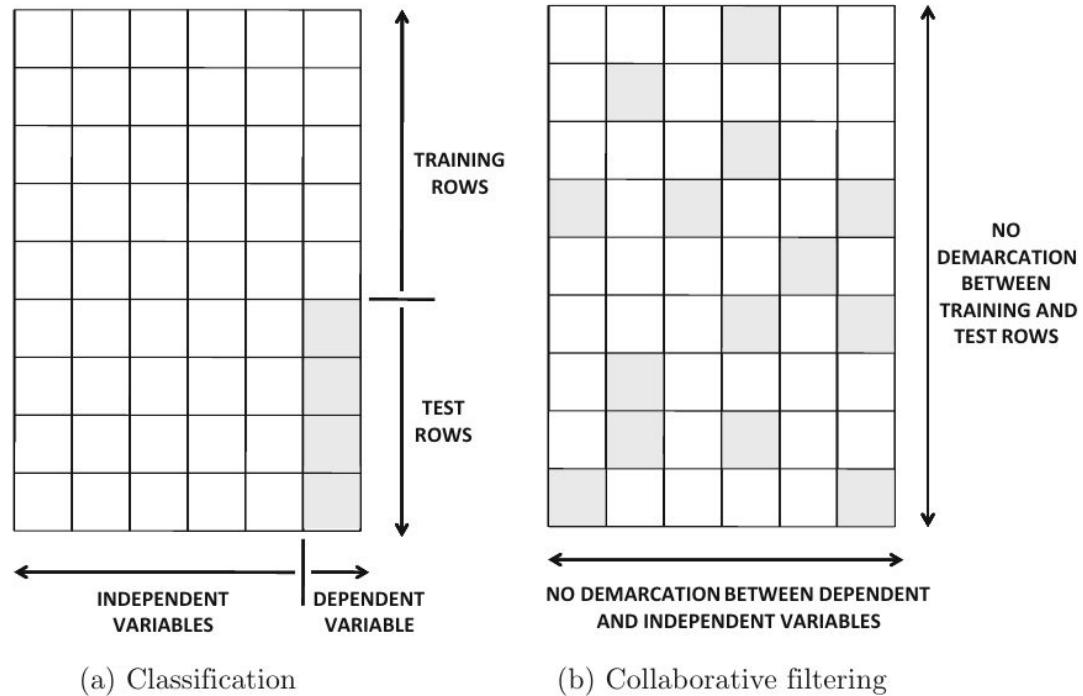
Modelo de Filtrado Colaborativo

- Emplean las calificaciones obtenidas por los usuarios para la generación de las recomendaciones.
- El principal reto de este modelo es la dispersión de las matrices (valores observados y no observados).

Producto	Usuario 1	Usuario 2	...	Usuario n
1	4	?	?	1
2	0	3	?	?
...	?	1	?	?
n	2	?	?	3

Filtrado Colaborativo

Puede ser visto como una generalización del problema de clasificación y regresión.



Clasificación vs Filtrado

Clasificación	Filtrado
Separación entre variables independientes y dependientes (clase).	Cada columna puede ser independiente o dependiente (dependerá sobre lo que se quiere predecir).
Clara separación entre datos de entrenamiento y de prueba.	En el mejor de los casos, los valores observados (serán los de entrenamiento) y los no observados (serán los de prueba).
Las columnas representan características y las filas instancias.	Debido a la naturaleza de los datos, se generan los modelos basados en productos o en usuarios.

Clasificación vs Filtrado

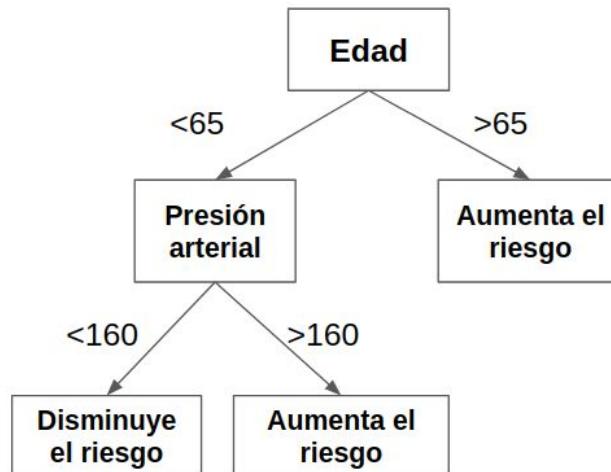
Debido a la generalidad del filtrado colaborativo, diferentes métodos de aprendizaje de máquina son usados en la generación de recomendaciones:

Métodos basados en ensambles (árboles de decisión y regresión):

- Naive Bayes.
 - Redes neuronales.
 - Modelos de factor latente.
-

Árboles de Decisión y Regresión

Son una partición jerárquica del espacio de datos a partir de un conjunto de criterios de decisión en las variables independientes.



Retos en la Construcción de Árboles de Decisión

Los principales desafíos en la extensión de los árboles de decisión en el filtrado colaborativo son:

- Los datos no están separados en forma de variables independientes y clase.
- La matriz de puntuaciones es mayormente dispersa y con datos ausentes.
- Dado que no es claro, las variables independientes y dependientes, ¿qué variable debe predecir el árbol?

¿Qué variable debe predecir el árbol?

- Se construyen árboles de decisión separados para predecir el rating por cada producto.
 - Supongamos, una matriz de $m \times n$ (m usuarios y n productos), el rating del producto que se quiera predecir será la variable dependiente y el resto independiente.
 - El número de árboles a construir será igual al número de productos de la matriz.
-

Ausencia de datos en los árboles de Clasificación

Sin embargo, uno de los principales retos de los árboles de decisión en el filtrado colaborativo es: **la ausencia de datos**.

Para atender este problema:

1. Se usan métodos de reducción de dimensionalidad.
 2. Posteriormente, se calcula la matriz de covarianza usando los vectores propios.
 3. La representación reducida se usa para la construcción de los árboles, como si fuera un problema de clasificación.
-

Filtrado Colaborativo Basado en Reglas

Se aborda la generación de recomendaciones basado en reglas de asociación.

Item ⇒	<i>Bread</i>	<i>Butter</i>	<i>Milk</i>	<i>Fish</i>	<i>Beef</i>	<i>Ham</i>
Customer ↓						
Jack	1	1	1	0	0	0
Mary	0	1	1	0	1	0
Jane	1	1	0	0	0	0
Sayani	1	1	1	1	1	1
John	0	0	0	1	0	1
Tom	0	0	0	1	1	1
Peter	0	1	0	1	1	0

$$\{\text{Butter}, \text{Milk}\} \Rightarrow \{\text{Bread}\}$$

Filtrado Colaborativo Basado en Reglas

Aprovechando las reglas de asociación en el filtrado colaborativo.

Cuando el número de productos es pequeño, cada valor de la combinación producto - calificación puede ser tratado como un pseudo-producto.

Ejemplo:

{Producto = leche, calificación = 'me gusta'} \Rightarrow {Producto = 'pan', calificación = 'me gusta'}

Las reglas de asociación son construidas en términos de pseudo-productos.

Filtrado Colaborativo Basado en Reglas

Ejemplo de reglas basadas en pseudo-productos:

- $\{ \text{Producto} = \text{mantequilla}, \text{calificación} = \text{'me gusta'} \} \text{ AND } \{ \text{Producto} = \text{leche}, \text{calificación} = \text{'me gusta'} \} \Rightarrow \{ \text{Producto} = \text{'pan'}, \text{calificación} = \text{'me gusta'} \}$

Ejemplo de reglas basadas en pseudo-usuarios:

- $\{ \text{Usuario} = \text{Pepe}, \text{calificación} = \text{'me gusta'} \} \Rightarrow \{ \text{Usuario} = \text{'Paco'}, \text{calificación} = \text{'no me gusta'} \} \text{ AND } \{ \text{Usuario} = \text{'Lola'}, \text{calificación} = \text{'me gusta'} \}$