Sistemas de Recomendación

Álvaro Riascos
Mónica Ribero

Universidad de Los Andes

12 de abril de 2016

Contenido

Introducción

Motivación

- Demasiados contenidos, demasiada oferta.
- Demasiados perfiles de usuarios
- ¿Cómo sugerirle al consumidor?
- ¿Cómo ayudarlo a encontrar productos de su interés ?

- Los Sistemas de Recomendación ordenan la cantidad masiva de datos para identificar potenciales preferencias de los usuarios.
- Identifican recomendaciones basándose en información del uusario, sus compras y el comportamiento de otros usuarios.

Ejemplos

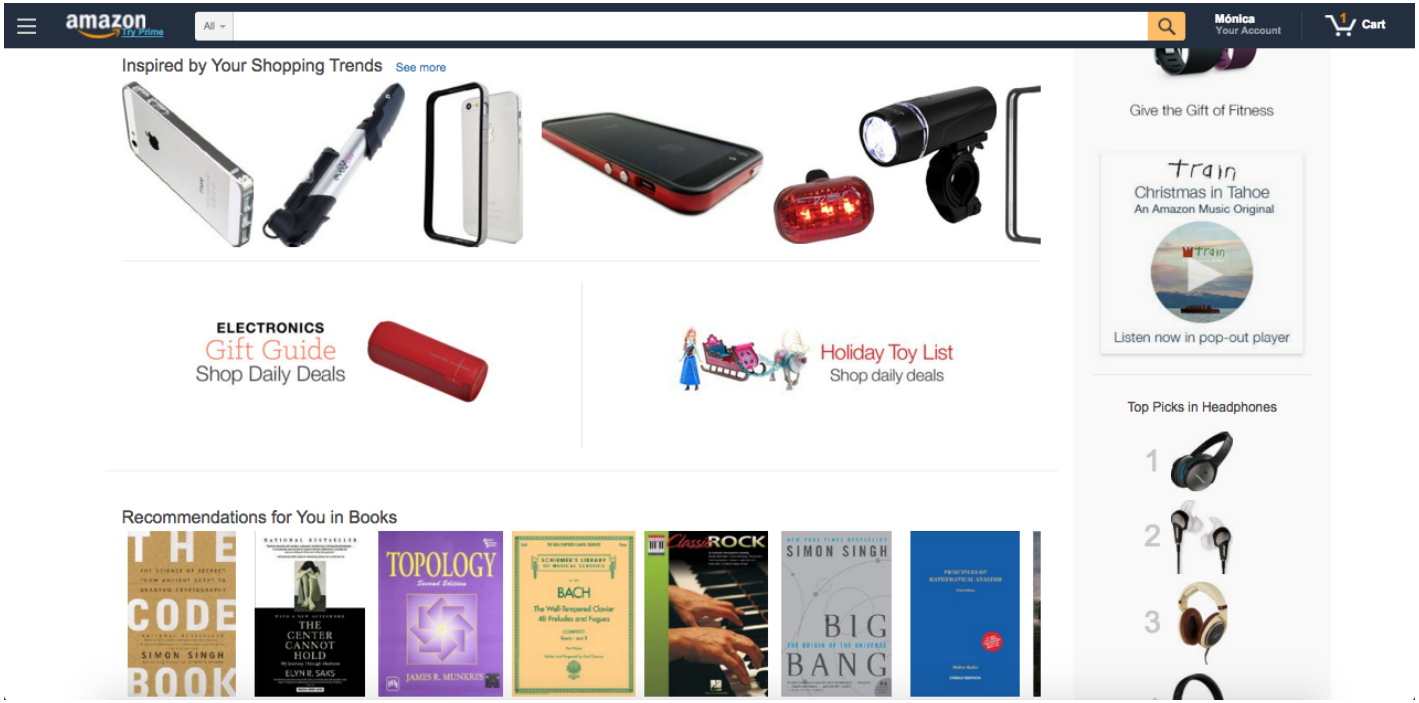


Figura: Página Principal de Amazon

Otros Ejemplos

- Netflix
- YouTube
- Amazon
- Yelp (Base de datos pública)
- Spotify
- The Economist

Definición

Definición

Dado un conjunto de usuarios U y un conjunto de productos P, un sistema de recomendación es un modelo que a partir de datos históricos de cada usuario $u \in U$ predice un producto $p \in P$ que sea de su agrado.
El modelo puede tener diferentes outputs:

- Predicción de una calificación numérica del usuario al producto (Netflix)
- Predicción si al usuario le gusta o no
- Top N productos

Datos

Datos Históricos

Para realizar recomendaciones se puede utilizar diferentes fuentes de datos

1. Perfil del usuario
2. Calificaciones del usuario a los productos que ya ha consumido (Netflix)
3. Productos "vistos" (En barras de búsqueda)
4. Productos "atractivos" (spotify y "likes")
5. Número de visitas a un lugar (Recomendación de restaurantes)

Datos Históricos – Calificaciones

Datos Históricos – Calificaciones

- Se conocen calificaciones de cada usuario a ALGUNOS productos.
- Dado $i \in U$ y $j \in P$, sea r_{ij} la calificación que el usuario i da al producto j
- Objetivo: Predecir r_{ik} para los productos k que el usuario aún no conoce

Datos: Matriz *sparse* de calificaciones

$$R = \begin{array}{c} \\ \textit{usuarios} \end{array} \begin{array}{c} \textit{productos} \\ \begin{bmatrix} r_{11} & & r_{1j} & & r_{1n} \\ & r_{22} & & & \\ \vdots & & & \ddots & \vdots \\ r_{m1} & & & & r_{mn} \end{bmatrix} \end{array}$$

donde r_{ij} = calificación que el usuario i da al producto j.

Cuadro: Entrenamiento

usuario	producto	calificación
1	6	4
1	4	5
2	6	2
3	2	3
3	6	1
3	3	4

Cuadro: Test

usuario	producto	calificación
1	1	
1	2	
1	3	
1	5	
2	1	
2	2	

Modelos

Promedios

Promedios

- Tomar el promedio de las calificaciones
- Dado un usuario $u \in U$, un producto $j \in P$, U_j el conjunto de usuarios que ya calificaron a j y P_i el conjunto de productos calificados por el usuario i estimamos

$$\hat{r}_{uj} = \bar{r} \tag{1}$$

$$\hat{r}_{uj} = \frac{1}{|U_j|} \sum_{i \in U_j} r_{ij} \tag{2}$$

$$\hat{r}_{uj} = \frac{1}{|P_i|} \sum_{j \in P_u} r_{uj} \tag{3}$$

Promedios

- Tomar el promedio de las calificaciones
- Dado un usuario $u \in U$, un producto $j \in P$, U_j el conjunto de usuarios que ya calificaron a j y P_i el conjunto de productos calificados por el usuario i estimamos

$$\hat{r}_{uj} = \bar{r} \tag{1}$$

$$\hat{r}_{uj} = \frac{1}{|U_j|} \sum_{i \in U_j} r_{ij} \tag{2}$$

$$\hat{r}_{uj} = \frac{1}{|P_i|} \sum_{j \in P_u} r_{uj} \tag{3}$$

- Solo captura la mayoría

“Basic Offset”

"Basic Offset"

- Idea: Tener en cuenta el sesgo del usuario y del producto

$$\hat{r}_{uj} = \bar{r} + \alpha_u + \beta_j$$

- $\alpha_u = \frac{1}{|P_u|} \sum_{p \in P_u} r_{pu} - \bar{r}$
- $\beta_j = \frac{1}{|U_j|} \sum_{i \in U_j} r_{ij} - \bar{r}$

Basic Offsets — Ejemplo

Queremos estimar r_{ij} para el usuario i y el producto j en una escala de 1 a 5.

Basic Offsets — Ejemplo

Queremos estimar r_{ij} para el usuario i y el producto j en una escala de 1 a 5.
Suponga:

- Promedio $\mu = 3$

Basic Offsets — Ejemplo

Queremos estimar r_{ij} para el usuario i y el producto j en una escala de 1 a 5.
Suponga:

- Promedio $\mu = 3$
- Usuario i califica en promedio con 4

Basic Offsets — Ejemplo

Queremos estimar r_{ij} para el usuario i y el producto j en una escala de 1 a 5.
Suponga:

- Promedio $\mu = 3$
- Usuario i califica en promedio con 4
- Producto j tiene un promedio de calificaciones de 1

Basic Offsets — Ejemplo

Queremos estimar r_{ij} para el usuario i y el producto j en una escala de 1 a 5.
Suponga:

- Promedio $\mu = 3$
- Usuario i califica en promedio con 4
- Producto j tiene un promedio de calificaciones de 1

$\hat{r_{ij}} = 3 + (4 - 3) + (1 - 3) = 2$

Content Filtering

- Crear perfiles de usuarios y productos que los caracterizan (datos demográficos, género, actores, etc...)
- Utilizar "features" para crear grupos de usuarios y productos

Content Filtering

- Crear perfiles de usuarios y productos que los caracterizan (datos demográficos, género, actores, etc...)
- Utilizar "features" para crear grupos de usuarios y productos
- Utilizar técnicas de segmentación, asociación y clasificación para encontrar patrones en los contenidos (de usuarios y productos) y recomendar productos.
  - k-medias
  - Regresión
  - árboles

Content Filtering

- Crear perfiles de usuarios y productos que los caracterizan (datos demográficos, género, actores, etc...)
- Utilizar "features" para crear grupos de usuarios y productos
- Utilizar técnicas de segmentación, asociación y clasificación para encontrar patrones en los contenidos (de usuarios y productos) y recomendar productos.
  - k-medias
  - Regresión
  - árboles
- Algoritmo utilizado por Pandora

Content Filtering

- Crear perfiles de usuarios y productos que los caracterizan (datos demográficos, género, actores, etc...)
- Utilizar "features" para crear grupos de usuarios y productos
- Utilizar técnicas de segmentación, asociación y clasificación para encontrar patrones en los contenidos (de usuarios y productos) y recomendar productos.
  - k-medias
  - Regresión
  - árboles
- Algoritmo utilizado por Pandora
- Información difícil de recolectar

Collaborative Filtering

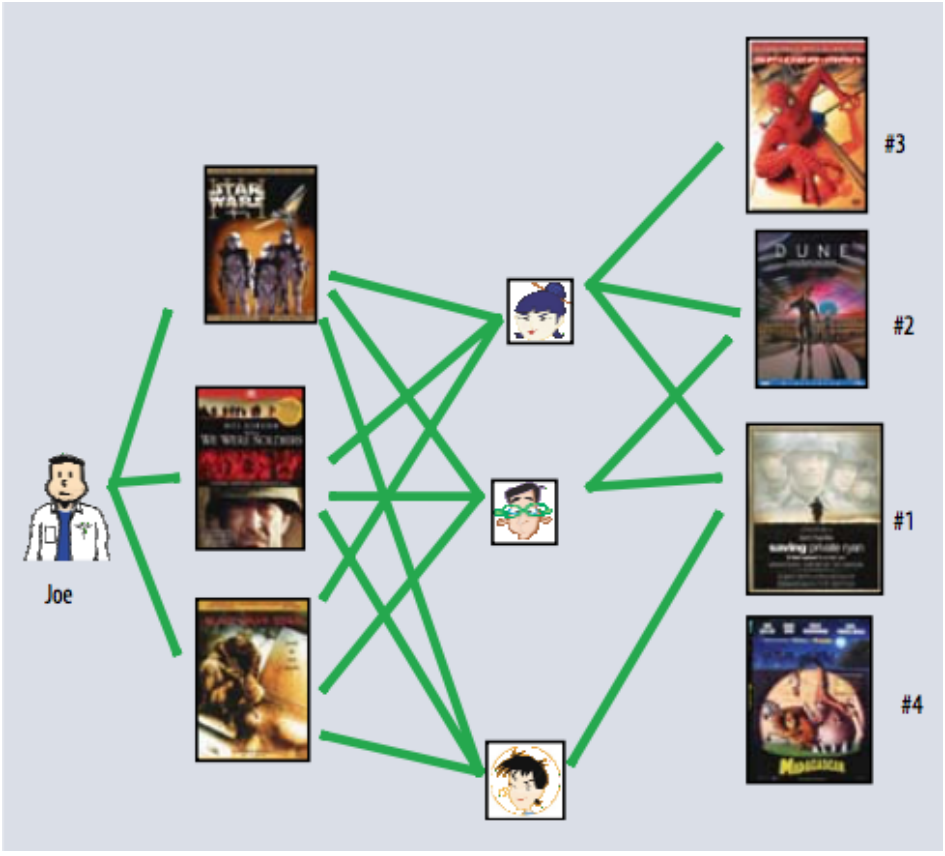
- Basarse únicamente en comportamientos pasados (transacciones, visitas, ...)
- Analizar relaciones entre usuarios e interdependencias entre productos para identificar nuevas asociaciones usuario-producto

Collaborative Filtering

- Basarse únicamente en comportamientos pasados (transacciones, visitas, ...)
- Analizar relaciones entre usuarios e interdependencias entre productos para identificar nuevas asociaciones usuario-producto
- Se utiliza la "colabaración"de múltiples usuarios y se filtra por los que exhiben preferencias o comportamientos similares.
- Problema de *cold start*

Métodos de Vecindarios

Trazar relaciones ente usuarios y productos.



Modelos de Factores latentes

Modelos de Factores latentes

- Capturar razones latentes de las calificaciones

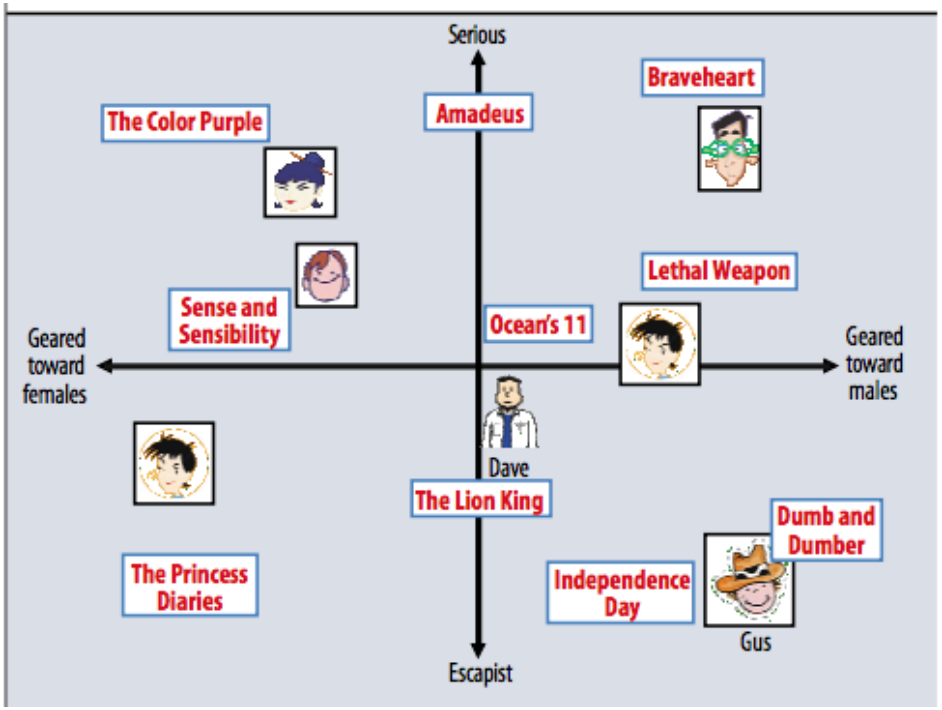
Modelos de Factores latentes

- Capturar razones latentes de las calificaciones
- Metodología: Caracterizar a cada usuario u y cada producto p en K factores (no interpretables) inferidos de patrones de calificación. K es escogido por el modelador.

Modelos de Factores latentes

- Capturar razones latentes de las calificaciones
- Metodología: Caracterizar a cada usuario u y cada producto p en K factores (no interpretables) inferidos de patrones de calificación. K es escogido por el modelador.
- Para los usuarios cada factor significa qué tanto les interesa ese factor particular. Para los productos, el desempeño o contenido del producto de cada factor

Ejemplo Factores Latentes



Factorización de Matrices

Datos: Matriz *sparse* de calificaciones

$$R = \begin{array}{c} \\ \textit{usuarios} \end{array} \begin{array}{c} \textit{productos} \\ \begin{bmatrix} r_{11} & & r_{1j} & & r_{1n} \\ & r_{22} & & & \\ \vdots & & & \ddots & \vdots \\ r_{m1} & & & & r_{mn} \end{bmatrix} \end{array}$$

donde r_{ij} = calificación que el usuario i da al producto j.

Factorización de Matrices

- Para cada usuario $u \in U$ y cada producto $p \in P$ encontrar los vectores x_u y y_p en $\mathbb{R}^K$
- x_u: preferencias del usuario por cada factor
- y_p: desempeño del producto en cada factor
- $r_{up} = x_u^t y_p$

Factorización de Matrices – Ejemplo Productos

$$\vec{y_p} \in \mathbb{R}^k$$

$$\vec{y_p} = \overset{k \text{ factores}}{\begin{bmatrix} servicio \\ calidad \\ precio \\ \vdots \end{bmatrix}}$$

Factorización de Matrices – Ejemplo Productos

$$\vec{y_p} \in \mathbb{R}^k$$

$$\vec{y_p} = \overset{k \text{ factores}}{\begin{bmatrix} servicio \\ calidad \\ precio \\ \vdots \end{bmatrix}} = \begin{bmatrix} 2 \\ 5 \\ 1 \\ 2 \end{bmatrix}$$

Factorización de Matrices – Ejemplo Usuarios

$$\vec{x_u} \in \mathbb{R}^k$$

$$\vec{x_u} = \begin{bmatrix} \text{servicio} \\ \text{calidad} \\ \text{precio} \\ \vdots \end{bmatrix} = \begin{bmatrix} 1 \\ 5 \\ 1 \\ 5 \end{bmatrix}$$

Predicción

Nuevamente,

$$r_{up} = x_u^t y_p$$

$$R = \begin{bmatrix} - & x_{u1} & - \\ & \vdots & \\ - & x_{un} & - \end{bmatrix} \begin{bmatrix} | & & | \\ y_{p1} & \cdots & y_{pm} \\ | & & | \end{bmatrix}$$

Factorización de Matrices

- $r_{up} = x_u^t y_p$
- Encontrar las matrices X y Y tales que

$$X^t Y = R$$

Factorización de Matrices

- $r_{up} = x_u^t y_p$
- Encontrar las matrices X y Y tales que

$$X^t Y = R$$

- Se pueden utilizar métodos de descomposición por valores singulares
- Generalmente falta demasiada información. Se puede estimar entradas faltantes pero esto es muy costoso por el significativo incremento de datos

Factorización de Matrices

- $r_{up} = x_u^t y_p$
- Encontrar las matrices X y Y tales que

$$X^t Y = R$$

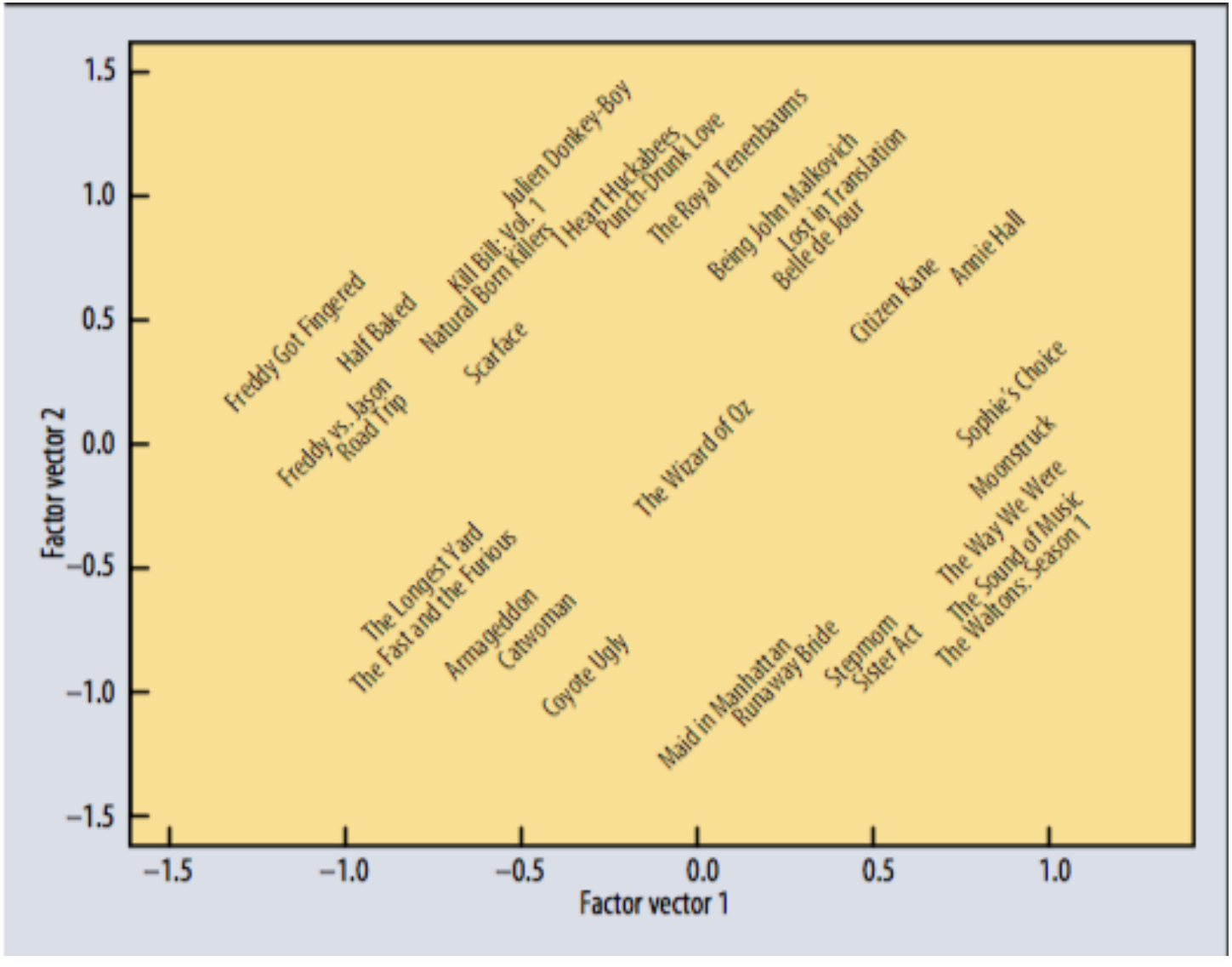
- Se pueden utilizar métodos de descomposición por valores singulares
- Generalmente falta demasiada información. Se puede estimar entradas faltantes pero esto es muy costoso por el significativo incremento de datos
- Utilizar solo las entradas conocidas puede llevar a "overfitting"

Factorización de Matrices – Modelos de regularización

Minimizar el error regularizado sobre el conjunto de calificaciones conocidas

$$min_{x,y} \sum_{(i,j)\in\kappa} (r_{ij} - x_i^t y_j)^2 + \lambda(||x_i||^2 + ||y_j||^2) \quad (4)$$

- κ es el conjunto de calificaciones conocidas (conjunto de entrenamiento).
- λ controla la regularización y es determinado por crossvalidación


Julien Donkey-Boy
Kill Bill: Vol. 1
I Heart Huckabees
Punch-Drunk Love
The Royal Tenenbaums
Being John Malkovich
Lost in Translation
Belle de Jour
Natural Born Killers
Freddy Got Fingered
Half Baked
Scarface
Citizen Kane
Annie Hall
Freddy vs. Jason
Road Trip
The Wizard of Oz
Sophie's Choice
Moonstruck
The Way We Were
The Sound of Music
The Waltons: Season 1
The Longest Yard
The Fast and the Furious
Armageddon
Catwoman
Coyote Ugly
Maid in Manhattan
Runaway Bride
Stepmom
Sister Act
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
Factor vector 2
−1.5
−1.0
−0.5
0.0
0.5
1.0
Factor vector 1

Optimización

Optimización

Descenso estocástico de gradiente

Añadir Información al Modelo

Sesgo

- Capturar características del usario o del producto que no tienen que ver con su interacción

Sesgo

- Capturar características del usario o del producto que no tienen que ver con su interacción
- Sesgo r_{ij} : $b_{ij} = \mu + \alpha_i + \beta_j$
- μ es el promedio general
- α_i y β_j son las desviaciones del usuario i y producto j respectivamente del promedio general

Sesgo

- Capturar características del usario o del producto que no tienen que ver con su interacción
- Sesgo r_{ij} : $b_{ij} = \mu + \alpha_i + \beta_j$
- μ es el promedio general
- α_i y β_j son las desviaciones del usuario i y producto j respectivamente del promedio general

$$min_{x,y,b} \sum_{(i,j)\in\kappa} (r_{ij} - \mu - \alpha_i - \beta_j - x_i^t y_j)^2 + \lambda(||x_i||^2 + ||y_j||^2 + \alpha_i^2 + \beta_j^2)$$

Incluir comentarios

Incluir comentarios

Julian McAuley y Jure Leskovec - “Hidden Factors as Topics (HFT)”

Incluir comentarios

Julian McAuley y Jure Leskovec - "Hidden Factors as Topics (HFT)"

- Incorporar Texto
- Descubrir atributos del usuario y los productos
- combina:
  - Modelo Factores Latentes
  - "Latent Dirichlet Allocation"

Más información...

- Diversidad Geográfica

Más información...

- Diversidad Geográfica
- Diversidad de productos

Otras dificultades

- Detección de “outliers”
- Detección de anomalías
- Privacidad: Caso de Macy’s