

UNIDAD 4: SISTEMAS DE RECOMENDACIÓN

RECOMENDACIÓN BASADA EN CONTENIDO

Blanca Vázquez y Gibran Fuentes-Pineda

Diciembre 2020

- La recomendación basada en contenido se enfoca en las propiedades de los productos o elementos
 - Útil para nuevos elementos
- Componentes generales
 1. Perfiles de productos/elementos
 2. Perfiles de usuarios

- Representación de cada producto usando atributos distintivos
- Por ej. una película
 - Actores
 - Director
 - Año de realización
 - Género
- Otros dominios: videojuegos, música, libros, noticias, etc.

- No es fácil asignar directamente un conjunto de atributos general
- Ejemplos
 - Artículos científicos
 - Noticias
 - Blogs
- Cálculo de similitud de documentos
 - Una estrategia para representar los documentos es usar las n palabras con mayor pesos *tf-idf*
 - Métricas comunes: Jaccard y coseno

PERFILES DE USUARIOS

- Tienen la misma representación que perfiles de productos/elementos
- Agregado de los perfiles de productos/elementos que le gustan al usuario
- Una estrategia es usar el promedio
 - Por ej. si el 30 % de las películas que le gustan a un usuario son de acción, ese atributo tendría el valor 0.3 en el perfil del usuario
- Es posible calcular la similitud o distancia de perfiles de usuarios y productos/elementos

- Otra estrategia es usar los perfiles de productos/elementos y las matrices de utilidad para entrenar clasificadores
- Para cada usuario se entrena un clasificador que predice las calificaciones que le dará a cualquier producto/elemento
- Ejemplos de entrenamiento
 - Entrada: perfil de producto/elemento calificado por el usuario
 - Salida: calificación otorgada

- Modela distribución conjunta de atributos y clases $P(x_1, \dots, x_d, y)$, asumiendo independencia condicional de los atributos dada la clase
- En el clasificador bayesiano ingenuo, la probabilidad conjunta está dada por

$$P(x_1, \dots, x_d, y) = \left(\prod_{j=1}^d P(x_j | y = c) \right) P(y = c)$$

CLASIFICADOR BAYESIANO INGENUO (2)

- Para obtener la probabilidad de cada clase para un nuevo dato $\tilde{\mathbf{x}} = [\tilde{x}_1, \dots, \tilde{x}_d]$ usamos teorema de bayes

$$P(y = c | \tilde{x}_1, \dots, \tilde{x}_d) = \frac{P(\tilde{x}_1, \dots, \tilde{x}_d | y = c) P(y = c)}{P(\tilde{x}_1, \dots, \tilde{x}_d)}$$

- Debido a que

$$\left(\prod_{j=1}^d P(\tilde{x}_j | y = c) \right) P(y = c) \propto P(y = c | \tilde{x}_1, \dots, \tilde{x}_d)$$

- Podemos obtener la clase más probable como:¹:

$$\hat{y} = \arg \max_y \left(\prod_{j=1}^d P(\tilde{x}_j | y = c) \right) P(y = c)$$

¹En algunas aplicaciones se requiere conocer las probabilidades para la toma de decisiones, por lo que es necesario calcular $P(\tilde{x}_1, \dots, \tilde{x}_d)$