

# ANÁLISIS DE DATOS MASIVOS

## RECOMENDACIÓN BASADA EN CONTENIDO

---

Blanca Vázquez

4 de noviembre de 2024

## RECOMENDACIÓN BASADA EN CONTENIDO (1)

- Sistemas basados en los atributos de los productos/elementos.
- Combina calificaciones con el contenido
- Muy útil cuando no hay suficientes calificaciones de algún producto/elemento.
- Desventajas
  - Las recomendaciones tienden a no ser tan diversas
  - No es efectiva para nuevos 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