

Sistema de recomendación - IA 2022

Reinaldo Barrera Travieso - C511

Juan José Lopez Martinez

1 Intoducción

Vemos el uso de sistemas de recomendación a nuestro alrededor. Estos sistemas están personalizando nuestra experiencia web, diciéndonos qué comprar (Amazon) , qué películas ver (Netflix) , con quién ser amigos (Facebook) , qué canciones escuchar (Spotify), etc. Estos sistemas de recomendación aprovechan nuestras compras, patrones de observación, escucha y predecir lo que nos podría gustar en el futuro en función de nuestros patrones de comportamiento. Los modelos más básicos para los sistemas de recomendaciones son modelos de filtrado colaborativo que se basan en la suposición de que a las personas les gustan las cosas similares a otras cosas que les gustan, y las cosas que les gustan a otras personas con gustos similares.

Tipos de técnicas de filtrado colaborativo

- Basado en la memoria
- Basado en modelos
- * Factorización de matrices
- * Agrupación
- * Aprendizaje Profundo

El enfoque de modelos se desarrollan utilizando diferentes mineía de datos para aprendizaje automático predecir la calificación de los elementos no calificados por parte de los usuarios. Hay muchos algoritmos CF basados en modelos. Redes bayesianas, agrupamiento, semánticos latentes como la descomposición de valores singulares , análisis semántico probabiístico latente , factor multiplicativo múltiple, asignación latente de Dirichlet y en el proceso de decisión de Markov modelos basados.

A través de este enfoque, reducción de la dimensionalidad se utilizan principalmente como técnica complementaria para mejorar la solidez y la precisión del enfoque basado en la memoria. En este sentido, métodos como la descomposición de valores singulares , el análisis de componentes principales , conocidos como modelos de factores latentes, comprimen la matriz de elementos de usuario en una representación de baja dimensión en términos de factores latentes. Una ventaja de usar este enfoque es que en lugar de tener una matriz de alta dimensión que contenga una gran cantidad de valores faltantes, estaremos tratando con una matriz mucho más pequeña en un

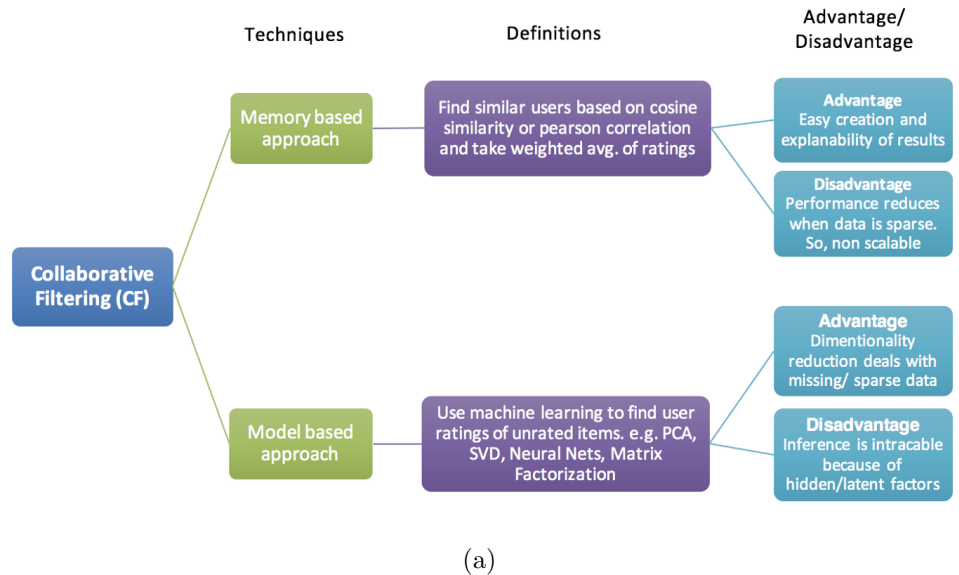


Figure 1: Tipos de enfoques de filtrado colaborativo. Referencia: Wikipedia

espacio de menor dimensión. Se podría utilizar una presentación reducida para los algoritmos de vecindad basados en elementos o basados en usuarios que se presentan en la sección anterior. Hay varias ventajas con este paradigma. Maneja la escasez de la matriz original mejor que las basadas en memoria. Además, comparar la similitud en la matriz resultante es mucho más escalable, especialmente cuando se trata de grandes conjuntos de datos dispersos.

La factorización matricial es una clase de filtrado colaborativo algoritmos los sistemas de recomendación . Los algoritmos de factorización de matrices funcionan descomponiendo la matriz en el producto de dos matrices rectangulares de menor dimensionalidad. Esta familia de métodos se hizo ampliamente conocida durante el premios de Netflix debido a su eficacia, según lo informado por Simon Funk en su publicación de blog de 2006, donde compartió sus hallazgos con la comunidad de investigadores. Los resultados de la predicción se pueden mejorar asignando diferentes ponderaciones de regularización a los factores latentes en función de la popularidad de los elementos y la actividad de los usuarios.

2 Evaluacion del sistema. Métricas

Los sistemas de recomendación son importantes en varios dominios de aplicaciones, como el comercio electrónico, las finanzas, la atención médica, etc. Es importante idear métricas de evaluación para medir qué tan bien funciona un sistema de recomendación.

2.1 Precisión

2.2 Recobrado

2.3 NDCG

NDCG significa ganancia acumulada descontada normalizada, donde la ganancia es solo la puntuación de relevancia para cada elemento recomendado.

La ganancia acumulada en K es la suma de las ganancias de los primeros K elementos recomendados.

$$CG_K = \sum_{i=1}^K G_i$$

La ganancia acumulada descontada pondera cada puntuación de relevancia en función de su posición. Las recomendaciones en la parte superior tienen un peso más alto, mientras que la relevancia de las que están en la parte inferior tienen un peso más bajo.

$$DCG_K = \sum_{i=1}^K \frac{G_i}{\log_2(i+1)}$$

Por tanto la ganancia acumulada descontada normalizada es la DCG con un factor de normalización en el denominador. El denominador es la puntuación DCG ideal cuando recomendamos primero los elementos más relevantes.

$$NDCG_K = \frac{DCG_K}{IDCG_K}$$
$$IDCG_K = \sum_{i=1}^{K^{ideal}} \frac{G_i^{ideal}}{\log_2(i+1)}$$

3 Resultados

Model	Movie				Profuct			
	P	R	NDCG	Score	P	R	NDCG	Score
B-Lineal	0.117	0.069	0.017	-	-	-	-	-
MF-Lineal	0.028	0.017	0.041	2.643	-	-	-	-
MF-L Bias	0.036	0.034	0.057	0.840	-	-	-	-
MF-NonNeg	0.040	0.021	0.056	0.984	-	-	-	-
Deep-MF	0.017	0.015	0.027	0.882	-	-	-	-
Deep-Neural-CF	0.043	0.020	0.062	0.990	-	-	-	-