

# 추천 시스템

# 목차

1-1 추천 시스템이란?

1-2 추천 시스템의 유형

1-3 콘텐츠 기반 필터링 추천 시스템

1-4 최근접 이웃 협업 필터링

1-5 잠재 요인 협업 필터링

## 1-1 추천 시스템이란?

- 정보 필터링을 사용하여 사용자에게 흥미로운 정보 아이템을 제공하는 시스템
- 사용자에게 개인 신상, 관심분야, 선호도 등을 질의하여 사용자의 정보를 획득.
- 획득한 정보를 기반으로 고객의 심리와 선호도 정보에 알맞은 정보 및 상품을 추천하거나 제공하는 방법

## 1-1 추천 시스템이란?

- 영화, 음악, 뉴스, 책, 연구주제, 탐색 질의, 상품 등 검색에 적용 가능
- 아마존, 이베이, 유튜브, 전자 상거래 업체 등 다양하게 활용하고 있음.

# 1-1 추천 시스템이란?

## ▶ 추천 시스템에 사용되는 데이터 예

- 사용자가 어떤 상품을 구매했을까?
- 사용자가 어떤 상품을 둘러보거나 장바구니에 넣었을까?
- 사용자가 평가한 영화 평점은? 제품 평가는?
- 사용자가 좋아요 한 상품은 무엇인가?
- 사용자가 어떤 상품을 클릭했는가?

# 1-2 추천 시스템의 유형

## ▶ 추천 시스템의 종류

- 콘텐츠 기반 필터링(Content based filtering)
- 협업 필터링(Collaborative filtering)
  - A. 최근접 이웃(Nearest Neighbor) 협업 필터링(Collaborative filtering)
  - B. 잠재 요인(Latent Factor) 협업 필터링

\* 추천 시스템 초창기에는 콘텐츠 기반 필터링이나 최근접 이웃 협업 필터링이 주로 사용되었으나, 넷플릭스 추천 시스템 경연 대회에서 행렬 분해(Matrix Factorization) 기법을 이용한 잠재 요인 협업 필터링 방식이 우승하여 많은 온라인 스토어에서 이를 적용하고 있음.

# 1-3 콘텐츠 기반 필터링 추천 시스템

## ▶ 콘텐츠 기반 필터링 방식

- 사용자가 특정한 아이템을 많이 선호하는 경우, 그 아이템과 비슷한 콘텐츠를 가진 다른 아이템을 추천하는 방식

⇒ 사용자가 특정 영화에 높은 평점을 주었다면 그 영화의 장르, 출연 배우, 감독, 영화 키워드 등의 콘텐츠와 유사한 다른 영화를 추천해 주는 방식

# 1-4 최근접 이웃 협업 필터링

## ▶ 최근접 이웃 협업 필터링

- 자기와 취향이 같은 친구들이 무엇을 샀는지 물어보기

## ▶ 종류

A. 사용자 기반 협업 필터링

B. 아이템 기반 협업 필터링



## 1-4 최근접 이웃 협업 필터링

### ▶ 사용자 기반(User-User) 협업 필터링

- 사용자-사용자 간의 유사도를 측정

### ▶ 아이템 기반(Item-Item) 협업 필터링

- 아이템-아이템 간 유사도를 측정

## 1-4 최근접 이웃 협업 필터링

### ▶ 측정 지표

A. 유클리디안

B. 코사인

C. 피어슨

D. 자카드

# 1-4 최근접 이웃 협업 필터링

## ▶ 사용자 기반 협업 필터링

	Item A	Item B	Item C	Item D	Item E
사용자 A	5	5	3	2	5
사용자 B	4	5	4	2	
사용자 C	2	2	3	2	

- A는 아이템 a, b, c, d에 5,5,3,2, e에 5점을 주었다.
- B는 아이템 a, b, c, d에 4,5,4,2 라는 평점을 주었다.
- A, B는 사용자 간 유사도가 높기 때문에 A가 좋아한 아이템 E를 B에게 추천한다.

# 1-4 최근접 이웃 협업 필터링

## ▶ 아이템 기반 협업 필터링

	사용자 A	사용자 B	사용자 C	사용자 D	사용자 E
죽은 시인의 사회	5	5	4	4	5
굿 윌 헌팅	4	5	4	3	
어벤저스	3	3	2	3	

- 사용자 A, B, C, D, E는 죽은 시인의 사회의 영화에 5,5,4,4,5를 평점을 주었습니다.
- 사용자 A, B, C, D는 굿 윌 헌팅의 영화에 4,5,4,3을 주었습니다.
- 죽은 시인의 사회를 좋아한 사용자 E에게 굿 윌 헌팅을 추천해 줍니다.

# 1-4 최근접 이웃 협업 필터링

## ▶ 아이템 기반 협업 필터링

$$\hat{R}_{u,i} = \sum_N (S_{i,N} * R_{u,N}) / \sum_N (|S_{i,N}|)$$

$\hat{R}_{u,i}$  : 사용자 u, 아이템 i의 개인화된 예측 평점 값

$S_{i,N}$  : 아이템 i와 가장 유사도가 높은 Top-N개 아이템의 유사도 벡터

$R_{u,N}$  : 사용자 u의 아이템 i와 가장 유사도가 높은 Top-N개 아이템에 대한 실제 평점 벡터

```
def predict_rating(ratings_arr, item_sim_arr):  
    val = np.array( [np.abs(item_sim_arr).sum(axis=1)] ) # 유사도의 열의 합  
    pred = ratings_arr.dot(item_sim_arr) / val  
    return pred
```

## 1-4 최근접 이웃 협업 필터링

### ▶ 사용자 기반 협업 필터링

A. 특정 사용자와 타 사용자 간의 유사도(Similarity)를 측정한 뒤 가장 유사도가 높은 TON-N 사용자를 추출해 그들이 선호하는 아이템을 추천하는 것이다.

# 1-5 잠재 요인(Latent Factor) 협업 필터링

## ▶ 잠재 요인 협업 필터링

A. 행렬 분해(matrix factorization)을 기반으로 사용.

- 다차원 매트릭스를 저차원 매트릭스로 분해하는 기법 SVD, NMF등이 있다.

B. 사용자-아이템 평점 매트릭스 속에 숨어 있는 잠재 요인을 추출해 추천 예측을 할 수 있도록 하는 기법.

C. 사용자-아이템 평점 행렬 데이터만을 이용해 말 그대로 '잠재 요인'을 끄집어 내는 것을 의미