

# DATA CREATOR CAMP

2024 데이터 크리에이터 캠프

대학부 실습영상

## 6강. 추천시스템 개요



과학기술정보통신부

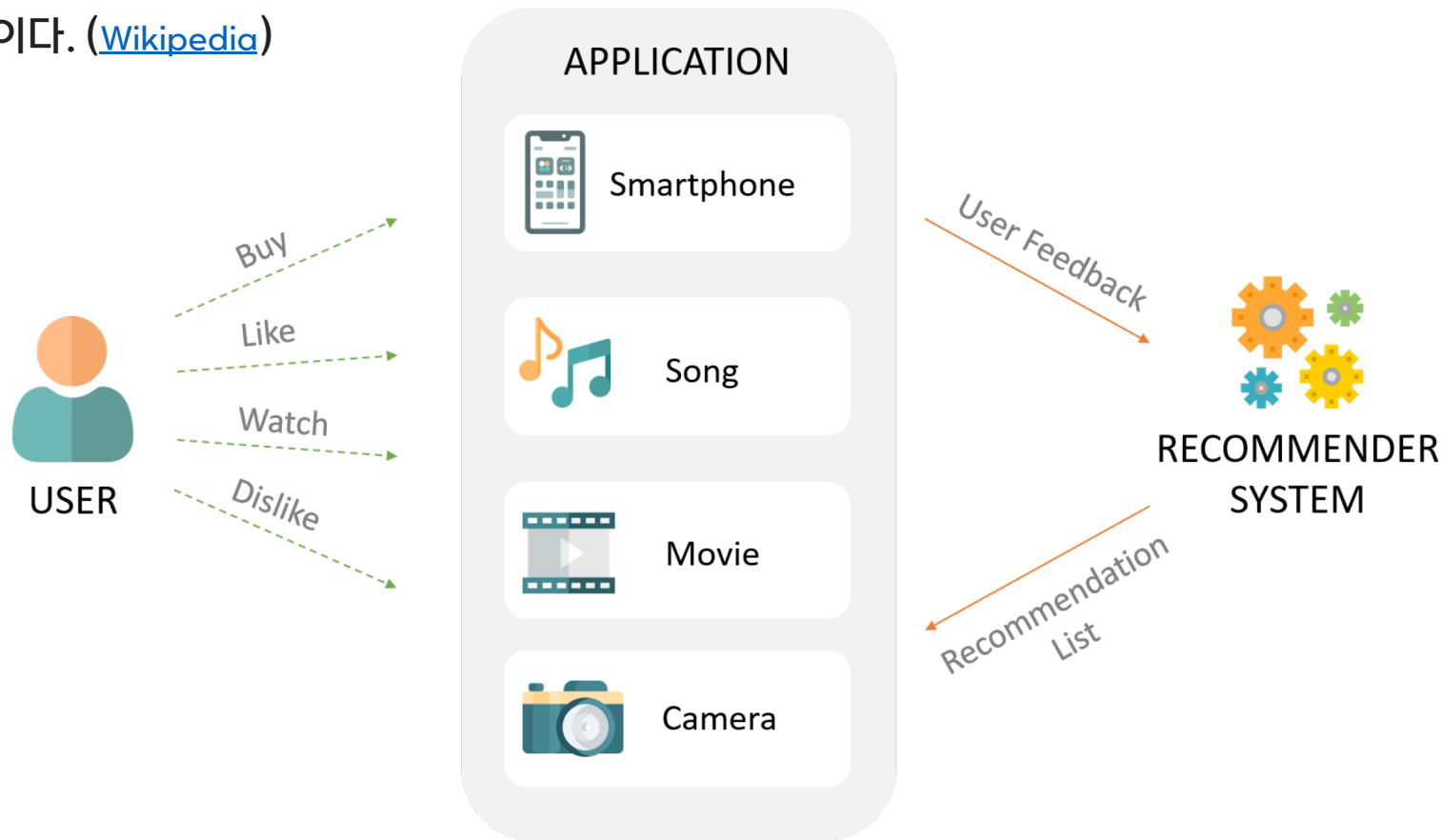
**NIA** 지능정보원  
한국지능정보사회진흥원

## 목차

- 추천시스템이란?
- 추천시스템 분류
- 추천시스템 평가

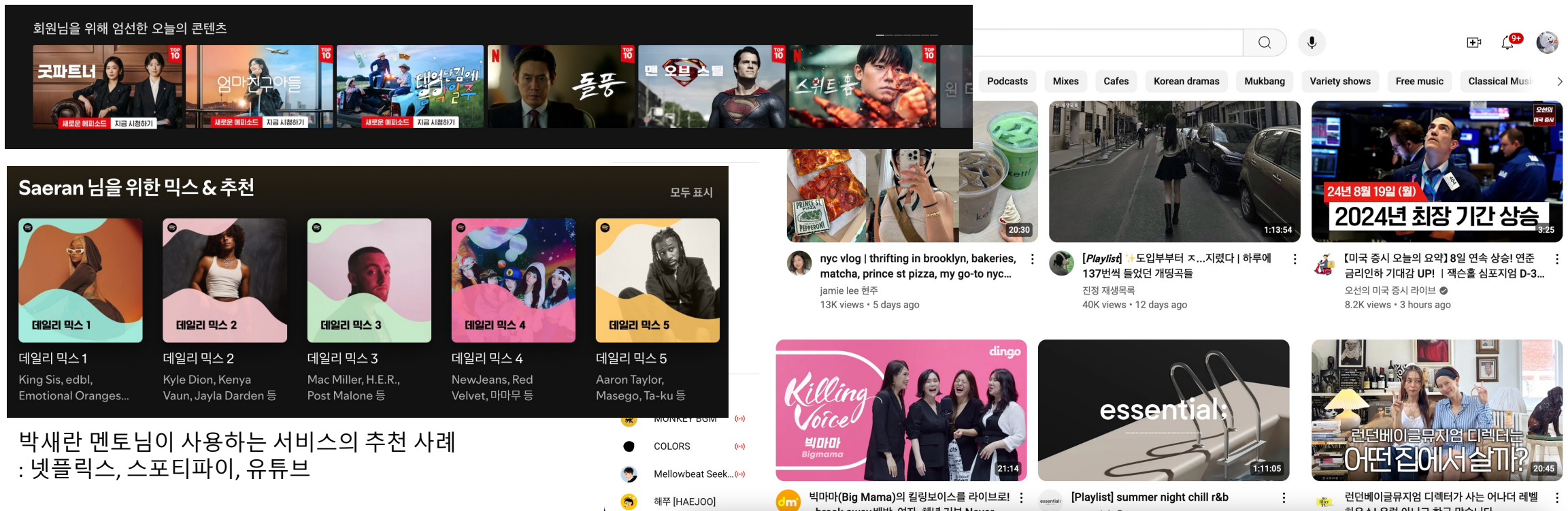
# 1. 추천시스템이란?

추천 시스템은 [정보 필터링](#) (IF) 기술의 일종으로, 특정 사용자가 관심을 가질만한 정보 (영화, 음악, 책, 뉴스, 이미지, 웹 페이지 등)를 추천하는 것이다. ([Wikipedia](#))



# 1. 추천시스템이란?

- 추천시스템은 사용자의 제품을 찾는 시간을 줄여주고, 의사결정의 질을 높여주며, 의사결정에 대한 신념을 높여주고 정신적인 노력을 줄여줌으로써 사용자의 성과에 긍정적인 영향을 미치는 것으로 보고하고 있다. (Hostler et al., 2005)
- 추천 시스템을 사용하는 경우 기업 입장에서는 개인화된 경험을 제공하여 유저의 만족도를 높여 매출 증대를 기대할 수 있다.[1]



박새란 멘토님이 사용하는 서비스의 추천 사례  
: 넷플릭스, 스포티파이, 유튜브

# 1. 추천시스템이란?

## ❖ 추천시스템의 목적

- Relevance : 사용자의 관심사에 관련된 아이템을 추천해주어야 함
- Novelty : 사용자가 과거에 경험해보지 못한 아이템을 추천해주어야 함
- Serendipity : 추천되는 아이템은 단순히 경험해보지 못한 것뿐(novelty) 아니라 놀랍고 신선해야 함
- Diversity : 추천되는 아이템들의 유사성을 줄여 추천의 성공 확률을 높임



## 2. 추천시스템 분류

- ❖ Collaborative Filtering Models
  - Memory-based methods
    - User-based collaborative filtering
    - Item-based collaborative filtering
  - Model-based methods
- ❖ Content-Based Recommender Systems
- ❖ Knowledge-Based Recommender Systems
- ❖ Demographic Recommender Systems
- ❖ Hybrid and Ensemble-Based Recommender Systems

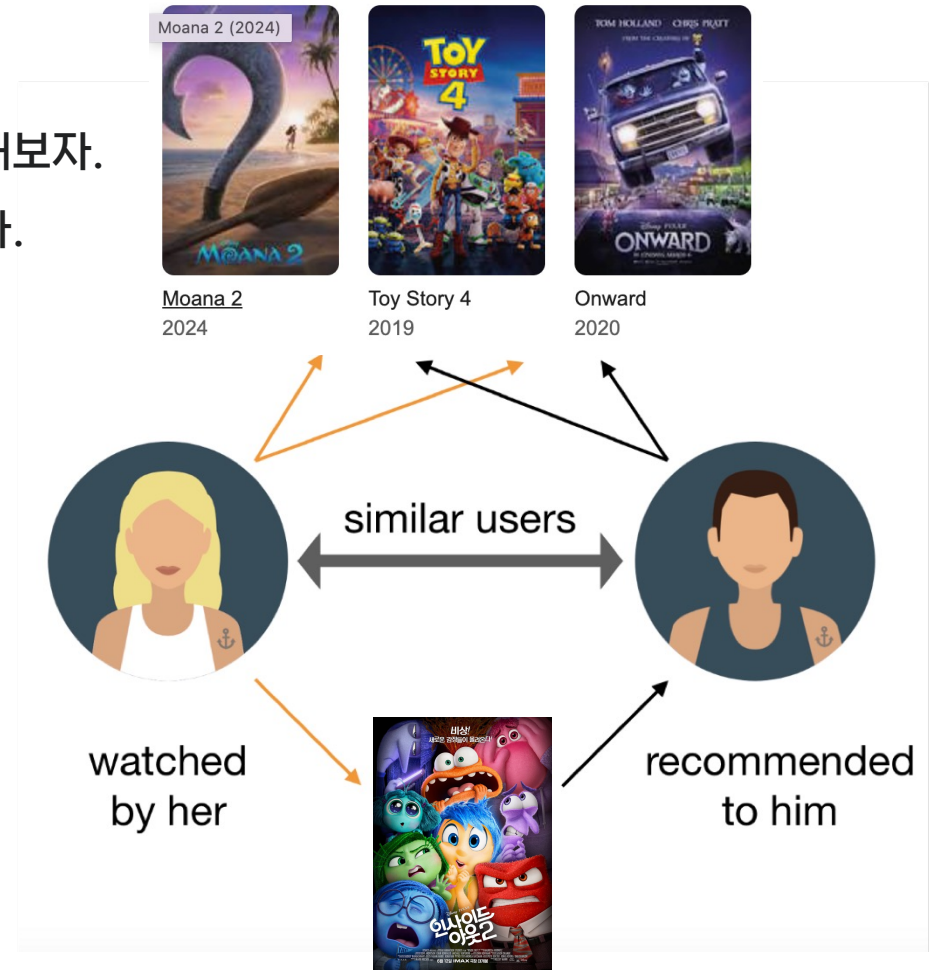
## 2. 추천시스템 분류

- ❖ Collaborative Filtering Models
  - Memory-based methods
    - User-based collaborative filtering
    - Item-based collaborative filtering
  - Model-based methods
- ❖ Content-Based Recommender Systems
- ❖ Knowledge-Based Recommender Systems
- ❖ Demographic Recommender Systems
- ❖ Hybrid and Ensemble-Based Recommender Systems

## 2. 추천시스템 분류

### ❖ Collaborative Filtering(CF)

- 형석이는 "인사이드아웃2"를 좋아할까 ?
  - 형석이와 비슷한 취향을 가진 민정이의 시청기록을 확인해보자.
  - 민정이가 봤던 "인사이드아웃2"도 좋아할 것으로 보인다.
- **특정 사용자**가 **평가하지 않은 아이템**의 평가를 예측할 때,  
해당 아이템에 대한 **다른 사용자**들의 평가를 이용
  - 평가 = rating, likes, buying 등





## 2. 추천시스템 분류

### ❖ Collaborative Filtering(CF)의 종류

#### ▪ Memory-based method

- Neighborhood Models로 사용자가 이전에 평가한 데이터를 사용하여, 유사한 user나 유사한 item을 찾아서 추천함.
- 행렬(matrix) 형태로 저장된 평가 데이터를 사용함.

#### ▪ Model-based method

- 과거의 평가 데이터를 학습하여 예측할 수 있는 모델(예: 결정 트리, 베이지안 모델, 잠재 요인 모델)을 학습함.
- 학습된 모델은 새로운 데이터가 주어졌을 때, 학습한 내용을 바탕으로 예측을 수행함.

## 2. 추천시스템 분류

❖ 협업필터링 (CF) : Memory-based method의 종류

- User-based collaborative filtering

- 이전에 시청한 영화들 중 비슷하게 평가한 사용자를 통해서 타겟 사용자의 타겟 아이템에 대한 점수를 예측함
- 사용자간의 유사성 평가는 row를 기준으로 구함

	7번방의 선물	검은사제들	극한직업	엑시트	파묘	국제시장
사용자1	5	1	3	2	2	5
사용자2	5	2	2	2	2	?
사용자3	1	5	3	3	5	1
사용자4	2	?	2	1	4	1

## 2. 추천시스템 분류

❖ 협업필터링(CF) : Memory-based method의 종류

- User-based collaborative filtering

- 이전에 시청한 영화들 중 비슷하게 평가한 사용자를 통해서 타겟 사용자의 타겟 아이템에 대한 점수를 예측함
- 사용자간의 유사성 평가는 row를 기준으로 구함

	7번방의 선물	검은사제들	극한직업	엑시트	파묘	국제시장
사용자1	5	1	3	2	2	5
사용자2	5	2	2	2	2	?
사용자3	1	5	3	3	5	1
사용자4	2	?	2	1	4	1

## 2. 추천시스템 분류

❖ 협업필터링 (CF) : Memory-based method의 종류

- User-based collaborative filtering

- 이전에 시청한 영화들 중 비슷하게 평가한 사용자를 통해서 타겟 사용자의 타겟 아이템에 대한 점수를 예측함
- 사용자간의 유사성 평가는 row를 기준으로 구함

	7번방의 선물	검은사제들	극한직업	엑시트	파묘	국제시장
사용자1	5	1	3	2	2	5
사용자2	5	2	2	2	2	?
사용자3	1	5	4	4	5	1
사용자4	2	?	4	5	4	1

## 2. 추천시스템 분류

❖ 협업필터링 (CF) : Memory-based method의 종류

- User-based collaborative filtering

- 이전에 시청한 영화들 중 비슷하게 평가한 사용자를 통해서 타겟 사용자의 타겟 아이템에 대한 점수를 예측함
- 사용자간의 유사성 평가는 row를 기준으로 구함

감동적인 한국 영화를 좋아하며, 무서운 영화에 대해서 흥미가 없는 사용자인 것으로 추정  
=> <국제시장>을 좋아할 것이다.

	7번방의 선물	검은사제들	극한직업	엑시트	파묘	국제시장	sim
사용자1	5	1	3	2	2	5	0.98
사용자2	5	2	2	2	2	5	-----
사용자3	1	5	4	4	5	1	0.66
사용자4	2	?	4	5	4	1	0.75



## 2. 추천시스템 분류

❖ 협업필터링 (CF) : Memory-based method의 종류

- User-based collaborative filtering

- 이전에 시청한 영화들 중 비슷하게 평가한 사용자를 통해서 타겟 사용자의 타겟 아이템에 대한 점수를 예측함
- 사용자간의 유사성 평가는 row를 기준으로 구함

오컬트 한국 영화를 좋아하며, 감동적인 영화에는 흥미가 없는 사용자인 것으로 추정된다.  
=> <검은사제들>을 좋아할 것이다.

	7번방의 선물	검은사제들	극한직업	엑시트	파묘	국제시장	sim
사용자1	5	1	3	2	2	5	0.79
사용자2	5	2	2	2	2	?	0.75
사용자3	1	5	4	4	5	1	0.97
사용자4	2	5	4	5	4	1	

## 2. 추천시스템 분류

❖ 협업필터링 (CF) : Memory-based method의 종류

- Item-based collaborative filtering

- 타겟 사용자가 이전에 시청한 영화들과 **비슷한 영화**를 찾아, 타겟 사용자의 평가를 바탕으로 타겟 아이템에 대한 점수를 예측함.
- 아이템간 유사성 평가는 column를 기준으로 구함

	7번방의 선물	검은사제들	극한직업	엑시트	파묘	국제시장
사용자1	5	1	3	2	2	5
사용자2	5	2	2	2	2	?
사용자3	1	5	4	4	5	1
사용자4	2	?	4	5	4	1

## 2. 추천시스템 분류

❖ 협업필터링 (CF) : Memory-based method의 종류

- Item-based collaborative filtering

- 타겟 사용자가 이전에 시청한 영화들과 **비슷한 영화**를 찾아, 타겟 사용자의 평가를 바탕으로 타겟 아이템에 대한 점수를 예측함.
- 아이템간 유사성 평가는 column를 기준으로 구함

	7번방의 선물	검은사제들	극한직업	엑시트	파묘	국제시장
사용자1	5	1	3	2	2	5
사용자2	5	2	2	2	2	?
사용자3	1	5	4	4	5	1
사용자4	2	4	4	5	4	1
	0.56		0.92	0.92	0.97	0.49

## 2. 추천시스템 분류

❖ 협업필터링 (CF) : Memory-based method의 종류

- Item-based collaborative filtering

- 타겟 사용자가 이전에 시청한 영화들과 **비슷한 영화**를 찾아, 타겟 사용자의 평가를 바탕으로 타겟 아이템에 대한 점수를 예측함.
- 아이템간 유사성 평가는 column를 기준으로 구함

	7번방의 선물	검은사제들	극한직업	엑시트	파묘	국제시장
사용자1	5	1	3	2	2	5
사용자2	5	2	2	2	2	5
사용자3	1	5	4	4	5	1
사용자4	2	?	4	5	4	1
	0.93	0.49	0.72	0.59	0.59	

## 2. 추천시스템 분류

### ❖ Collaborative Filtering(CF) 장단점

- 상호작용의(rating, 구매빈도..) 데이터만 있으면 적용할 수 있음. 따라서 도메인에 의존되지 않고 사용할 수 있음
- 일반적으로 content-based보다 좋은 성능을 가짐.
- cold start 문제가 존재함. 새로운 item이나 user들에 대한 feedback(상호작용 데이터)이 부족하기 때문임.



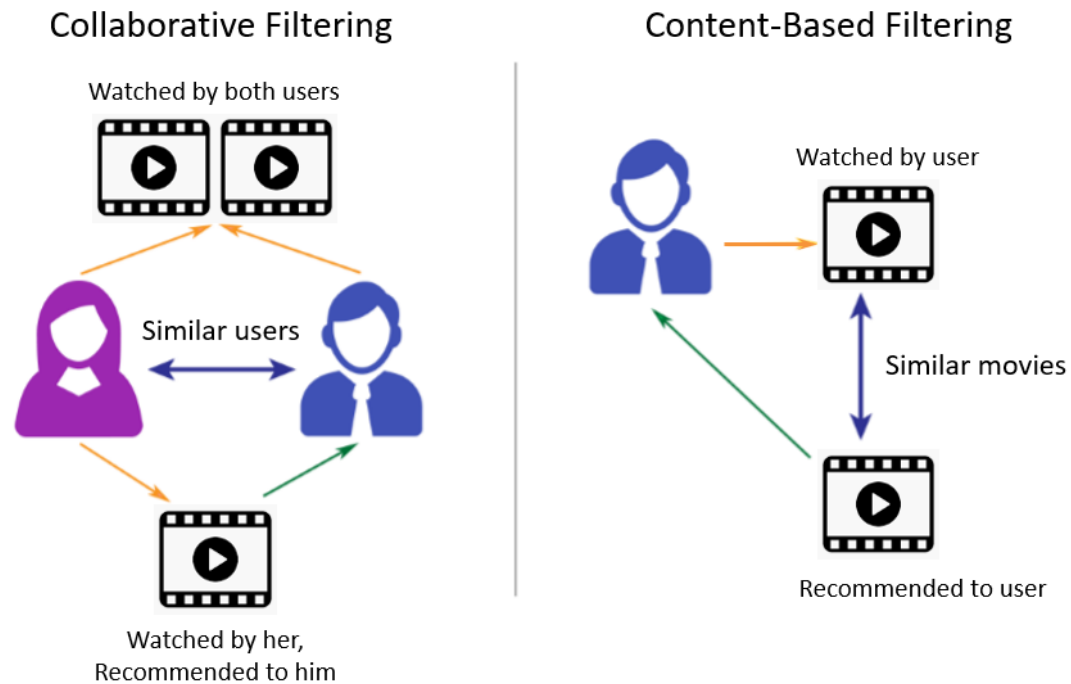
## 2. 추천시스템 분류

- ❖ Collaborative Filtering Models
  - Memory-based methods
    - User-based collaborative filtering
    - Item-based collaborative filtering
  - Model-based methods
- ❖ Content-Based Recommender Systems
- ❖ Knowledge-Based Recommender Systems
- ❖ Demographic Recommender Systems
- ❖ Hybrid and Ensemble-Based Recommender Systems

## 2. 추천시스템 분류

### ❖ Content-based Recommender

- 핵심 아이디어 : 사용자는 과거에 자신이 선택한 아이템들과 비슷한 아이템을 좋아할 가능성이 높다.
- 콘텐츠 기반 추천시스템은 특정 아이템의 구체적인 특징들을 기반으로 추천한다.
- CB 모델에서는 다른 사용자들의 피드백은 사용되지 않으며, 오로지 추천 대상이 되는 "타겟 사용자"에만 초점을 둔다



## 2. 추천시스템 분류

### ❖ Content-based Recommender

#### ▪ 방법

1. 유저 프로필 파악 : 이전 시청 기록을 기반으로 유저의 취향 파악

- 매드 맥스: 분노의 도로와 다크 나이트를 좋아했다면, 이 영화들의 장르(액션, 스릴러), 감독(조지 밀러, 크리스토퍼 놀란), 출연 배우 등의 정보가 사용자 프로필에 반영됨.

⇒ 액션 장르와 스릴러 장르를 좋아하며, 크리스토퍼 놀란이나 톰 하디가 출연한 영화를 선호한다고 예측

2. 유사 아이템 선택

- 인셉션: 크리스토퍼 놀란 감독, 액션/스릴러 장르, 출연 배우 중 톰 하디 포함
- 인터스텔라: 크리스토퍼 놀란 감독, 톰 하디 출연



## 2. 추천시스템 분류

### ❖ Content-based Recommender

#### ▪ 알고리즘 예시

1. TF-IDF (Term Frequency-Inverse Document Frequency)를 사용하여 아이템에 대한 속성 벡터 구하기
  - 아이템에 대한 설명 데이터(text)가 존재한다면, 텍스트 데이터를 벡터로 변환함.
  - 텍스트 데이터를 벡터로 변환시에 TF-IDF 방법을 사용함.
2. 코사인 유사도를 통해서 가장 유사한 아이템 찾기
  - 각 아이템에 대한 벡터를 사용하여 코사인 유사도를 구하고 가장 높은 유사도를 보인 아이템을 추천함.

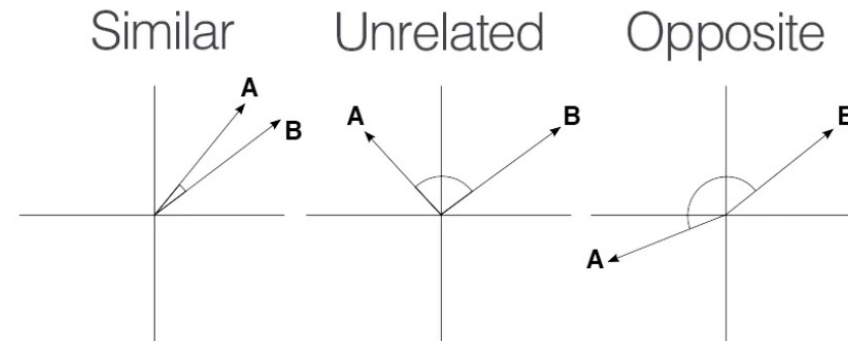
Term Frequency X Inverse Document Frequency

$$w_{x,y} = tf_{x,y} \times \log\left(\frac{N}{df_x}\right)$$

Text1: Basic Linux Commands for Data Science

Text2: Essential DVC Commands for Data Science

	basic	commands	data	dvc	essential	for	linux	science
Text 1	0.5	0.35	0.35	0.0	0.0	0.35	0.5	0.35
Text 2	0.0	0.35	0.35	0.5	0.5	0.35	0.0	0.35



## 2. 추천시스템 분류

### ❖ Content-based Recommender 장단점

- cold-start problem에서 사용가능한 방법
  - 새로운 아이템에 대한 평가 점수가 없더라도 아이템간의 특징만을 비교하여 추천이 가능함.
- 사용자의 과거 구매 이력을 바탕으로 제품을 추천하기 때문에 추천 상품의 다양성 보장이 어려움.
  - 무작위 요소를 추가하여 새로운 아이템을 추천할 수 있음. (유전자 알고리즘 활용)
- 추천하고자 하는 사용자의 독립적인 정보만을 필요함.
  - 다른 사용자의 정보가 부족한 경우에도 유용하게 사용할 수 있는 추천 방법임.
  - 과거 구매 이력이 부족한 경우 추천 성능 보장이 어려움.



## 2. 추천시스템 분류

- ❖ Collaborative Filtering Models
  - Memory-based methods
    - User-based collaborative filtering
    - Item-based collaborative filtering
  - Model-based methods
- ❖ Content-Based Recommender Systems
- ❖ Knowledge-Based Recommender Systems
- ❖ Demographic Recommender Systems
- ❖ Hybrid and Ensemble-Based Recommender Systems

## 2. 추천시스템 분류

### ❖ Knowledge-based Recommender

- 지식 기반 추천 시스템은 평가 데이터를 사용하지 않고, 고객의 요구 사항과 항목 설명 간의 유사성을 바탕으로 추천을 수행함.
- cold-start problem에서 사용가능한 방법임.

## 2. 추천시스템 분류

### ❖ Knowledge-based Recommender 특징

#### 1. 명시적인 사용자 요구사항

- 사용자가 제공하는 구체적인 요구사항을 바탕으로 추천함.
- 자동차를 구입하려고 할 때, "세단", "하이브리드", "5천만 원 이하" 등의 명확한 요구사항을 입력하면, 이 조건에 맞는 자동차를 추천함.

#### 2. 규칙 기반 추천

- "연비가 높은 차"를 선호한다고 했을 경우, 해당 규칙에 부합하는 아이템들을 추천할 수 있습니다.

#### 3. 사용자 행동 이력이 필요하지 않음

- KB 추천 시스템은 협업 필터링이나 콘텐츠 기반 필터링처럼 사용자 행동 데이터를 필요로 하지 않음.

#### 4. 복잡한 제품과 서비스에 적합

- 부동산, 자동차, 금융 상품처럼 다양한 특성과 옵션이 결합된 제품에 대해 매우 효과적임.
- 구매 빈도가 낮고, 사용자마다 요구 조건이 명확하기 때문에 KB 추천 시스템이 잘 맞음.

## 2. 추천시스템 분류

Table 1.2: The conceptual goals of various recommender systems

Approach	Conceptual Goal	Input
Collaborative	Give me recommendations based on a collaborative approach that leverages the ratings and actions of my peers/myself.	User ratings + community ratings
Content-based	Give me recommendations based on the content (attributes) I have favored in my past ratings and actions.	User ratings + item attributes
Knowledge-based	Give me recommendations based on my explicit specification of the kind of content (attributes) I want.	User specification + item attributes + domain knowledge

## 2. 추천시스템 분류

❖ 세가지 방법의 장단점

	장점	단점
Collaborative Filtering	<ul style="list-style-type: none"> <li>- 다양한 추천 가능</li> <li>- 순수 사용자의 행동데이터만 필요함. 메타 데이터 불필요함.</li> <li>- 충분한 사용자 데이터가 확보되면 매우 정확하고 개인화된 추천이 가능함</li> </ul>	<ul style="list-style-type: none"> <li>- Cold start 문제 발생</li> <li>- 데이터 희소성 문제 : 유사도 계산에 어려움이 존재함</li> <li>- 계산 복잡도가 높음</li> </ul>
Content-based Recommendation	<ul style="list-style-type: none"> <li>- Cold Start 문제 완화</li> <li>- 사용자 개인화 : 사용자의 과거 선호도를 분석해 유사한 특성을 가진 아이템을 추천</li> </ul>	<ul style="list-style-type: none"> <li>- 다양한 추천이 어려움</li> <li>- 아이템 메타데이터에 의존적 : 메타데이터가 부족한 경우 부정확함</li> </ul>
Knowledge-based Recommendation	<ul style="list-style-type: none"> <li>- 명확한 통제 및 신뢰성</li> <li>- Cold Start 문제 없음</li> <li>- 복잡한 제품에 적합</li> </ul>	<ul style="list-style-type: none"> <li>- 규칙 설정의 어려움</li> <li>- 사용자가 예상하지 못한 새로운 제품 추천이 어려움</li> <li>- 도메인 지식이 변하는 경우 유지보수에 비용이 많이 듦</li> </ul>



## 2. 추천시스템 분류

### ❖ 추천시스템에서 사용하는 데이터

- 사용자 데이터
  - 사용자 프로필 정보: 사용자의 기본 정보(이름, 나이, 성별, 위치 등)
- 아이템 데이터
  - 아이템 특성 : 영화라면 장르, 감독, 출연 배우, 제품이라면 브랜드, 가격, 카테고리 등 아이템의 속성 정보  
=> 콘텐츠 기반 추천
  - 아이템 메타데이터: 아이템의 설명, 태그, 출시 날짜, 이미지, 사용자 리뷰 등 추가적인 메타데이터  
=> 콘텐츠 기반 추천시 아이템간의 유사도 평가시 더 풍부한 정보 제공 가능
- 사용자-아이템 상호작용 데이터
  - 명시적 피드백(Explicit Feedback): 사용자가 명시적으로 남긴 데이터(ex. 영화에 대한 별점 평가, 리뷰, 좋아요)  
=> 협업 필터링 사용
  - 암묵적 피드백(Implicit Feedback): 사용자의 행동을 통해 수집된 데이터 (ex. 상품 구매, 영화 시청 시간, 웹페이지 조회 수). 사용자의 관심사를 추론가능 => 협업 필터링 사용

### 3. 추천시스템 평가 방법

#### ❖ 사용자 스터디

- 피험자를 모집하여 추천 시스템과 상호작용하도록 요청하고, 이 과정에서 피드백을 수집함.
  - 예) 사용자에게 제품 사이트에서 추천을 받고 그 품질에 대한 피드백을 제공하도록 하거나, 노래를 듣고 평가를 제공하도록 할 수 있음.
- 대규모 사용자 모집이 어렵고 비용이 많이 들며, 모집된 사용자들이 일반 사용자를 대표하지 않을 가능성도 있음.

### 3. 추천시스템 평가 방법

#### ❖ 온라인 평가

- 실 사용자들이 실제 상업 시스템에서 추천 시스템과 상호작용하는 데이터를 활용
- 모집 과정에서 발생하는 편향을 줄일 수 있으며, A/B 테스트를 통해 두 알고리즘의 성능을 비교
- 대규모 사용자 기반이 필요하기 때문에 초기 단계에서는 사용하기 어려움. 주로 상업 시스템 내에서만 활용되기 때문에 연구자들이 다양한 상황에서 테스트하기에는 제한적임.

### 3. 추천시스템 평가 방법

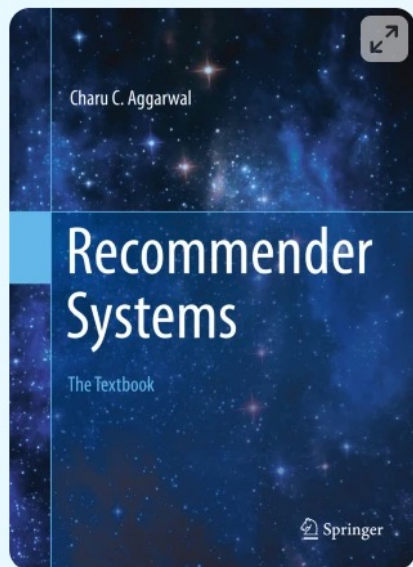
#### ❖ 오프라인 평가

- 우리가 기존에 모델 학습 후에 진행하는 평가하는 것을 오프라인 평가라고 함.
- 주어진 벤치마크 데이터셋에 대해 추천 알고리즘이 얼마나 효과적인지 평가하는 것.
- 오프라인 평가의 단점은 실제 사용자 반응을 측정하지 못하며, 데이터가 시간에 따라 변할 수 있어 미래에 적합하지 않을 수 있음.

# Mini Quiz

- ❖ 협업 필터링(Collaborative filtering)은 사용자의 평가와 동료의 평가를 활용하여 추천을 생성하는 방식이다.  
○ / X
- ❖ 콘텐츠 기반(Content-based) 추천 시스템은 사용자가 과거에 평가하고 행동한 항목의 속성을 기반으로 추천을 제공한다.  
○ / X
- ❖ 지식 기반(Knowledge-based) 추천 시스템은 사용자의 명시적인 요구 사항과 도메인 지식을 활용하여 추천을 제공한다.  
○ / X
- ❖ 콘텐츠 기반 추천 시스템에서는 주로 사용자 평가와 커뮤니티 평가를 입력으로 사용한다.  
○ / X
- ❖ 협업 필터링 방식에서는 아이템 속성(Item attributes)을 입력으로 사용하지 않는다.  
○ / X
- ❖ 지식 기반 추천 시스템에서는 사용자 요구 사항, 아이템 속성, 도메인 지식이 모두 추천의 입력 요소로 활용된다.  
○ / X

# 참고자료



## Recommender Systems

### The Textbook

Textbook | © 2016

✓ Access provided by Korea University

Download book PDF ↓

Download book EPUB ↓



이 문서의  
외부 유출 및  
공유를 금합니다.

본 콘텐츠는 한국지능정보사회진흥원(NIA)의 동의 없이 무단 사용할 수 없으며,  
상업적 목적으로 이용을 금합니다.

# DATA CREATOR CAMP

2024 데이터 크리에이터 캠프

# 감사합니다.



과학기술정보통신부

NIA

지능정보원  
한국지능정보사회진흥원