

15기 정규 세션

ToBig's 14기 이원도

추천 시스템 기초

Recommendation System

contents

Unit 01 | 추천 시스템 이해

Unit 02 | 콘텐츠 기반 추천과 협업필터링

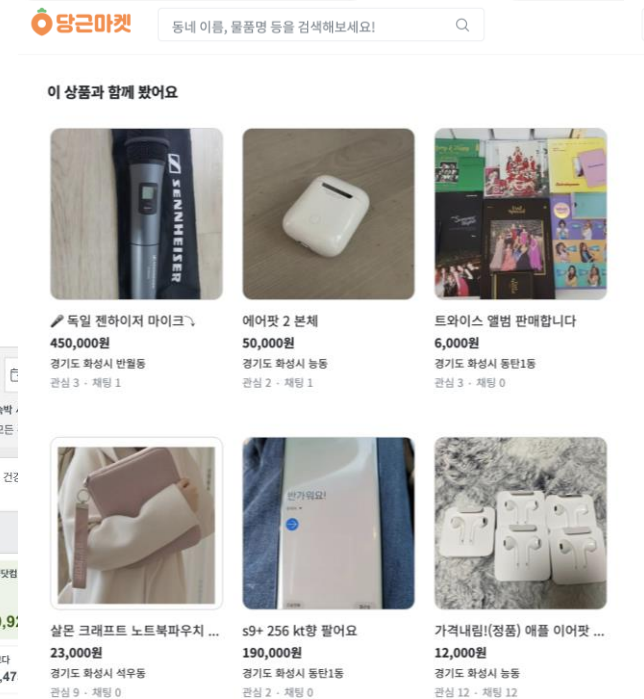
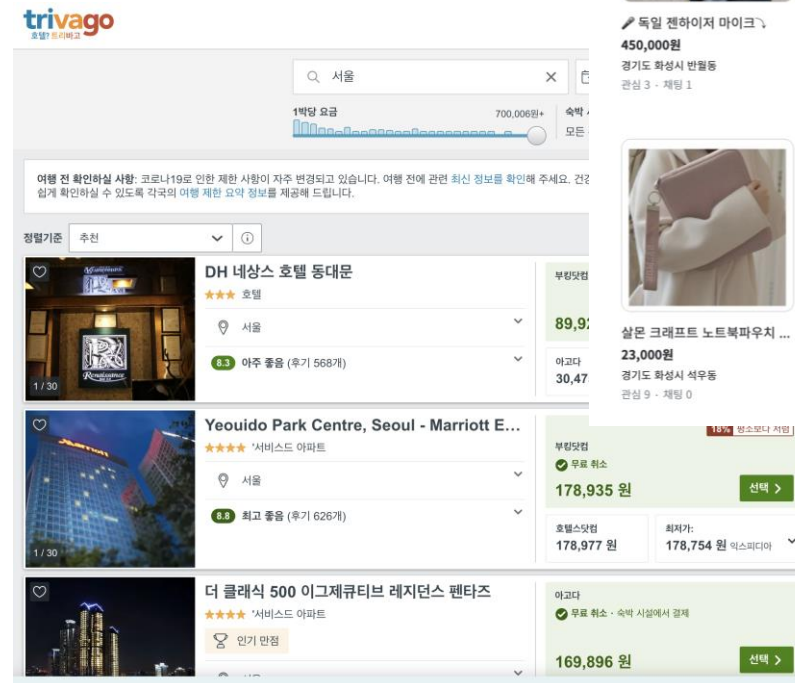
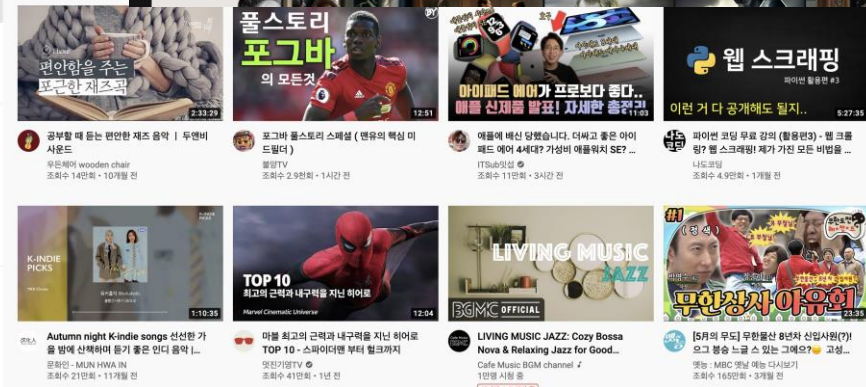
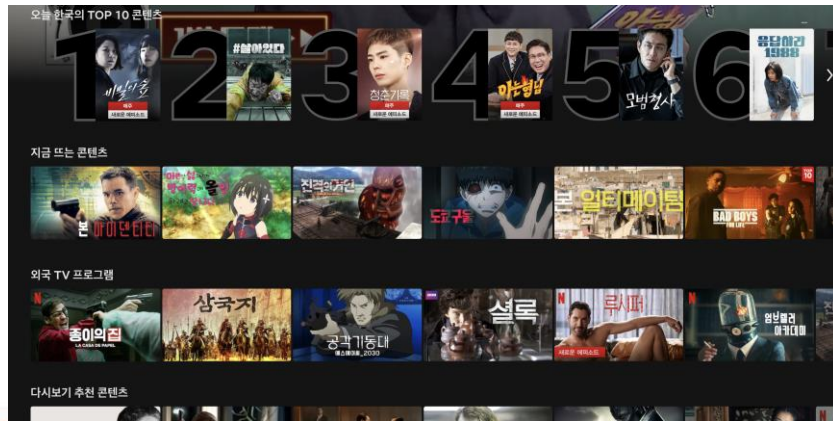
Unit 03 | Matrix Factorization

Unit 04 | 추천 시스템 평가방법

Unit 01 | 추천 시스템 이해

Unit 01 | 추천 시스템 이해

추천 시스템이란?



Unit 01 | 추천 시스템 이해

추천 시스템이란?

- ✓ 정보필터링 기술의 일종으로, 특정 사용자가 관심을 가질만한 정보 (영화, 음악, 책, 뉴스, 이미지, 웹 페이지 등)를 추천하는 것
(출처 : Wiki)
- ✓ 방대한 양의 항목들 중에서 사용자(고객)의 관심에 맞추어 양질의 제품 또는 서비스를 추천해주거나 고객의 구매를 예측하는 시스템
(출처 : Yonsei Univ, buss177)
- ✓ 사용자(user)에게 상품(item)을 제안하는 소프트웨어 도구 이자 기술
(출처 : Recommender Systems Handbook)

Unit 01 | 추천 시스템 이해

추천 시스템이란?

- ✓ 정보필터링 기술의 일종으로, 특정 사용자가 관심을 가질만한 정보 (영화, 음악, 책, 뉴스, 이미지, 웹 페이지 등)를 추천하는 것
(출처 : Wiki)
- ✓ 방대한 양의 항목들 중에서 사용자(고객)의 관심에 맞추어 양질의 제품 또는 서비스를 추천해주거나 고객의 구매를 예측하는 시스템
(출처 : Yonsei Univ, 2001/7/7)
- ✓ 사용자(user)에게 상품(item)을 제안하는 소프트웨어 도구 이자 기술
(출처 : Recommender Systems Handbook)

추천 시스템의 목표: 어떤 사용자에게 어떤 상품을 어떻게 추천할 것인가?

Unit 01 | 추천 시스템 이해

추천 시스템 Trend

연관규칙분석

- Apriori 알고리즘

~2010

Spark ML을 활용한

- FPGrowth
- Matrix Factorization

~2013

~2015

개인화 추천과 Re-Ranking

- Factorization Machine
- Deep Learning-based
- Multi-Armed Bandit(MAB)
- Reinforcement Learning
- GNN

~2017

Netflix Prize 이후 (2006~2009)

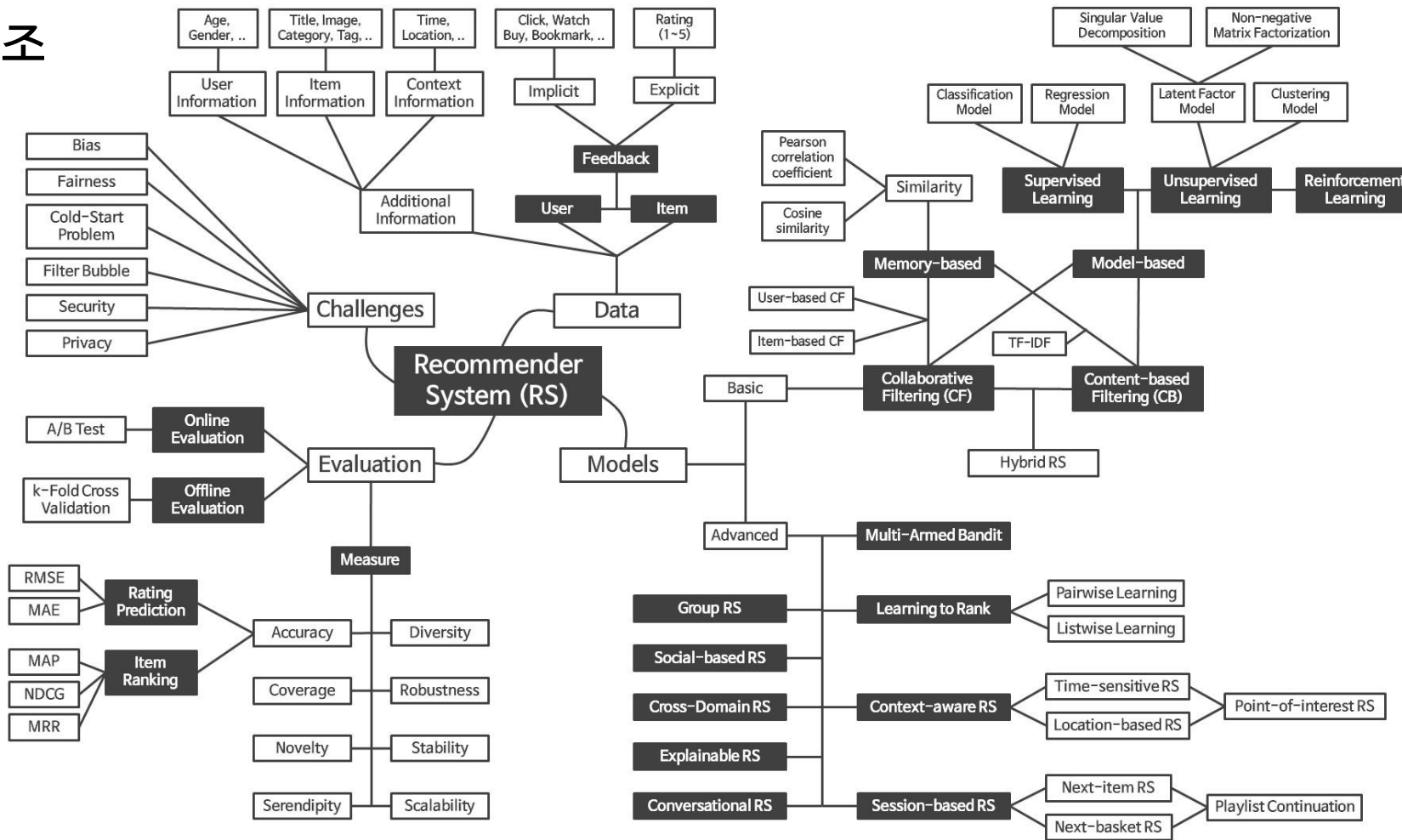
- Content-based
- 협업필터링(CF)
- User-based, Item-based
- SVD, SVD++

딥러닝을 활용한

- Item2Vec/User2Vec + CF
- YouTube-Recommendation
- Google-Wide&Deep Model

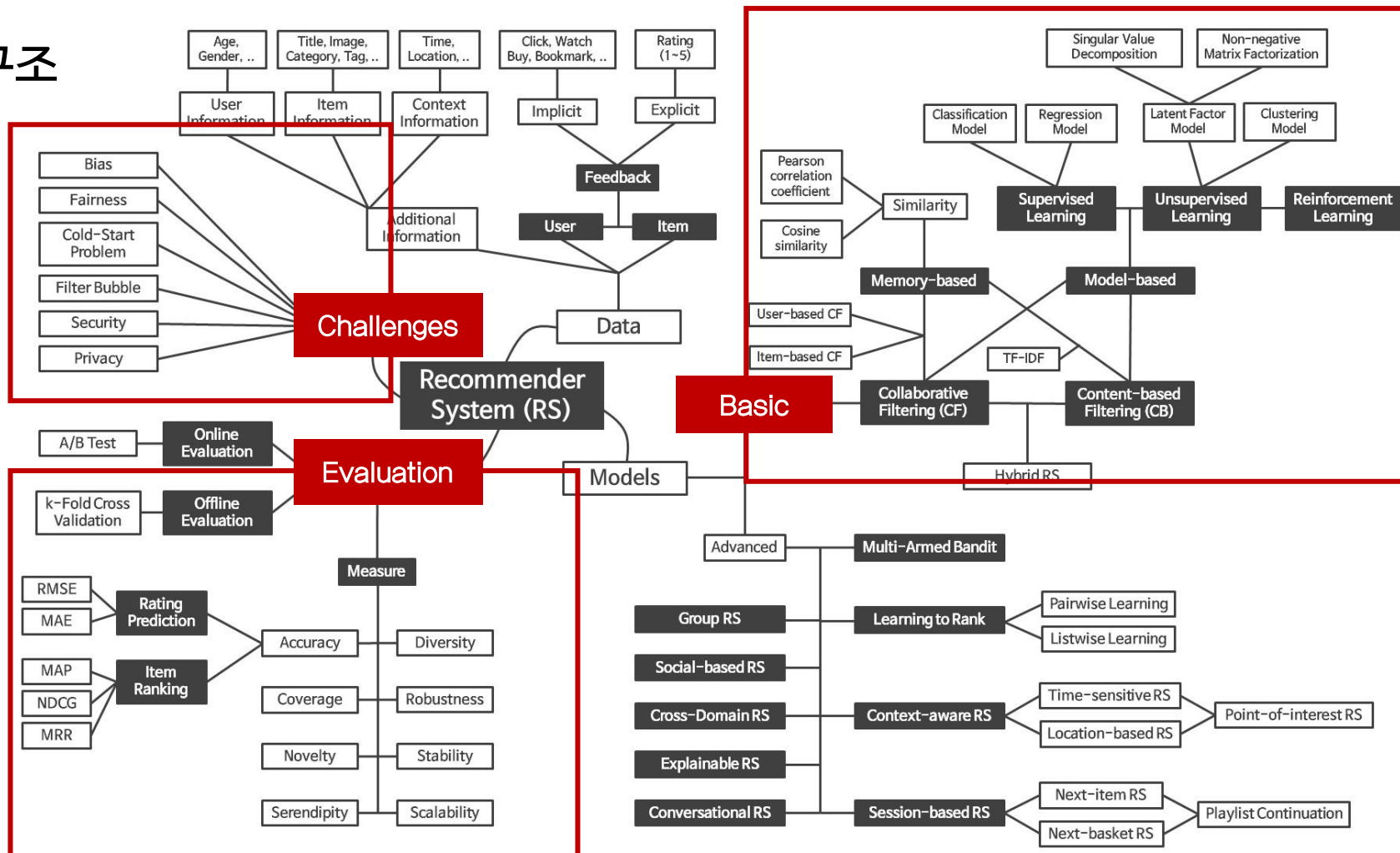
Unit 01 | 추천 시스템 이해

추천 시스템 구조



Unit 01 | 추천 시스템 이해

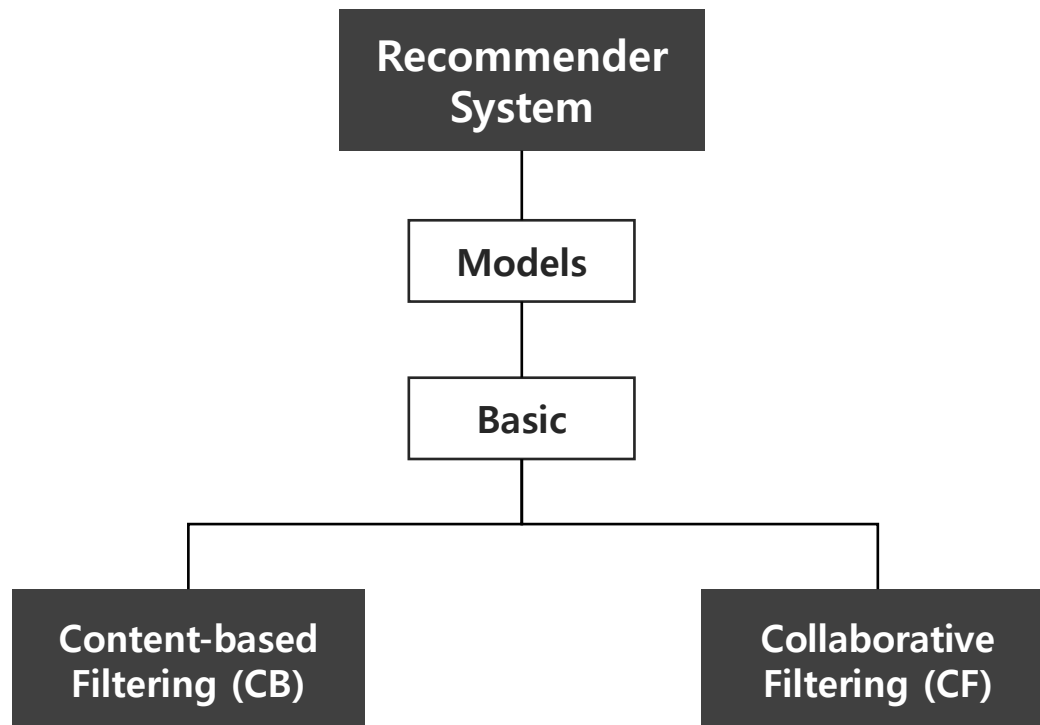
추천 시스템 구조



Unit 02 | 콘텐츠 기반 추천과 협업필터링

Unit 02 | 콘텐츠 기반 추천과 협업필터링

콘텐츠 기반 추천과 협업필터링



Unit 02 | 콘텐츠 기반 추천과 협업필터링

콘텐츠 기반 추천, Content-based filtering

: 아이템의 콘텐츠를 직접 분석하여 아이템과 아이템 간 유사성을 분석하고 이를 토대로 사용자에게 아이템을 추천해주는 방식

Unit 02 | 콘텐츠 기반 추천과 협업필터링

콘텐츠 기반 추천, Content-based filtering

: **아이템**의 **콘텐츠**를 직접 분석하여 아이템과 아이템 간 유사성을 분석하고 이를 토대로 사용자에게 아이템을 추천해주는 방식

고객 또는 사용자가 소비할 수 있는 무언가 또는
서비스

- ✓ 영화
- ✓ 동영상
- ✓ 제품

Unit 02 | 콘텐츠 기반 추천과 협업필터링

콘텐츠 기반 추천, Content-based filtering

: **아이템**의 **콘텐츠**를 직접 분석하여 아이템과 아이템 간 유사성을 분석하고 이를 토대로 사용자에게 아이템을 추천해주는 방식

고객 또는 사용자가 소비할 수 있는 무언가 또는 서비스

- ✓ 영화
- ✓ 동영상
- ✓ 제품

-> 러닝타임, 장르, 배우, 감독, 개봉일, 배급사, 포스터...

-> 카테고리, 업로드 시간, 영상 길이, 조회수, 썸네일, 제목, 설명...

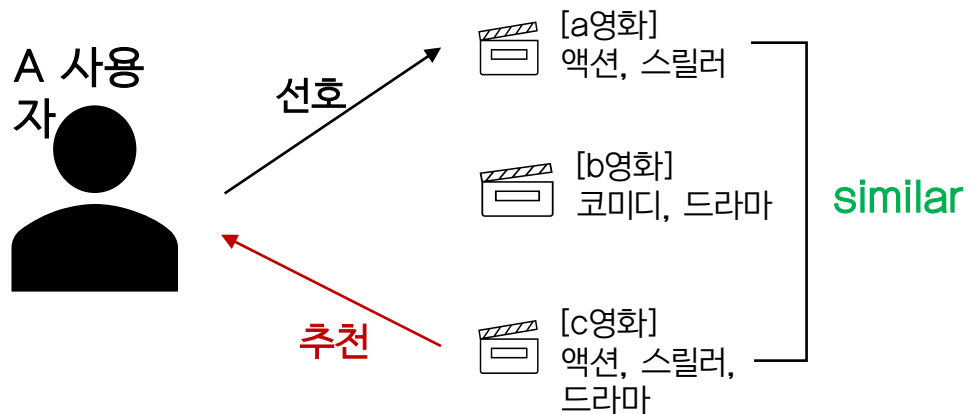
-> 제품명, 가격, 제조사, 판매처...

Unit 02 | 콘텐츠 기반 추천과 협업필터링

콘텐츠 기반 추천, Content-based filtering

: 아이템의 콘텐츠를 직접 분석하여 아이템과 아이템 간 유사성을 분석하고 이를 토대로 **사용자에게 아이템을 추천해주는 방식**

-> 어떤 사용자가 선호했던 아이템과 유사한 아이템을 선호할 가능성이 높다.

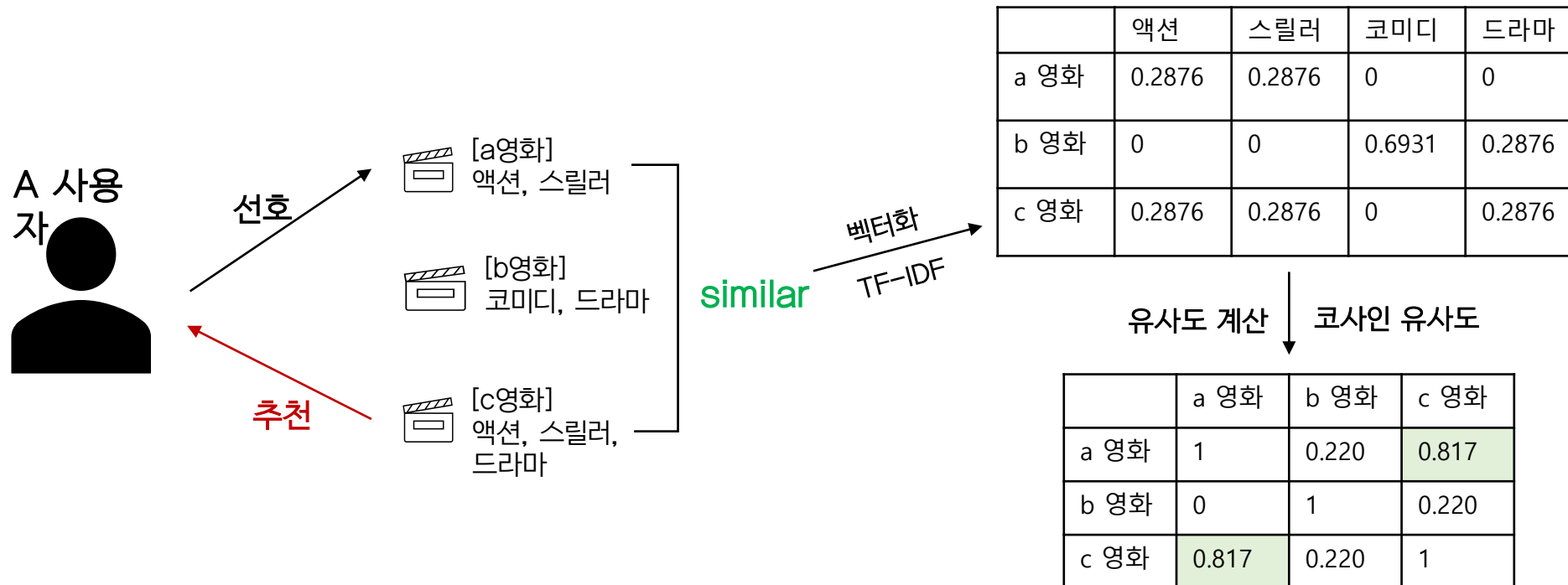


Unit 02 | 콘텐츠 기반 추천과 협업필터링

콘텐츠 기반 추천, Content-based filtering

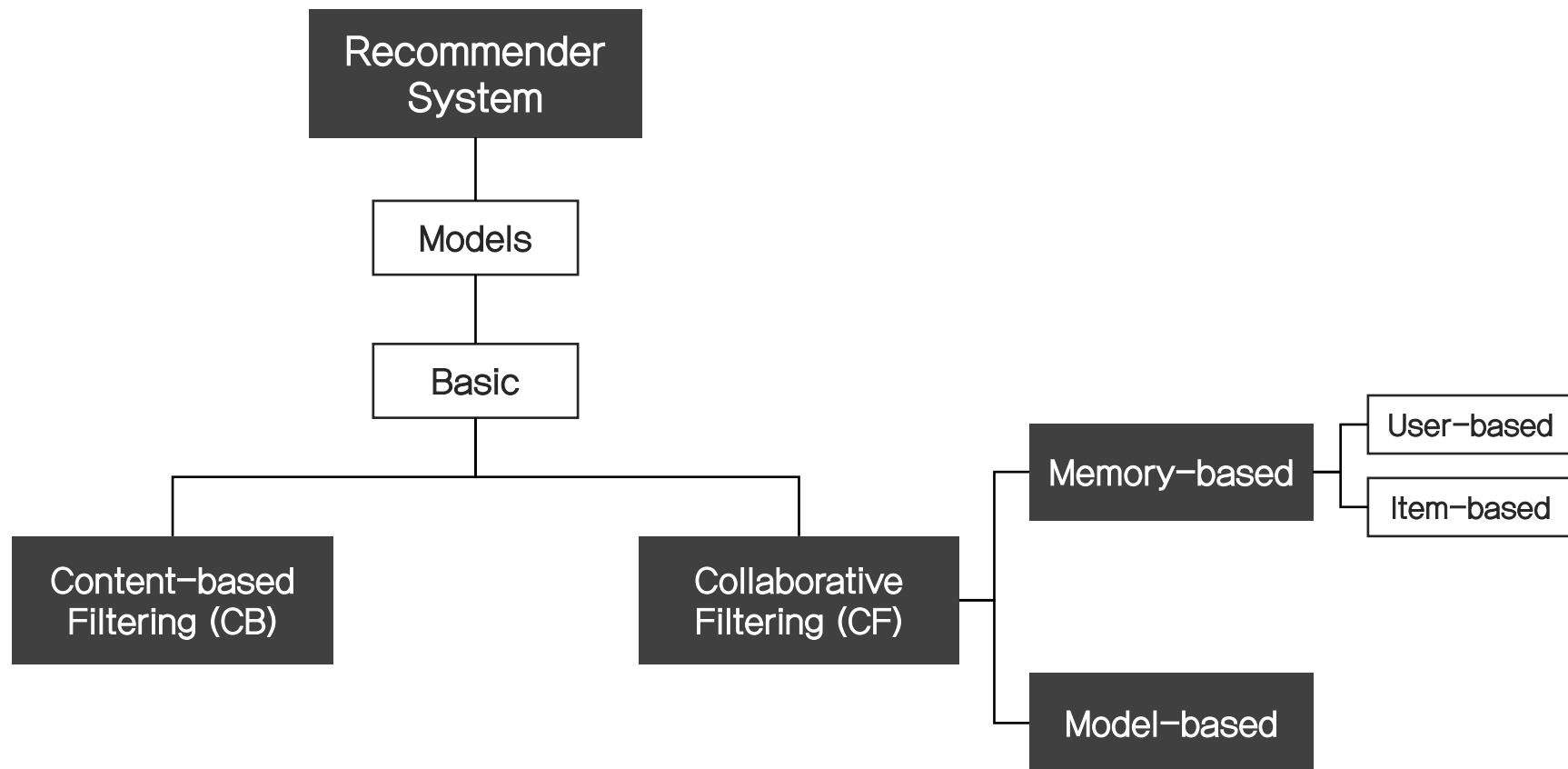
: 아이템의 콘텐츠를 직접 분석하여 **아이템과 아이템 간 유사성을 분석**하고 이를 토대로 사용자에게 아이템을 추천해주는 방식

→ 어떤 사용자가 선호했던 아이템과 유사한 아이템을 선호할 가능성이 높다.



Unit 02 | 콘텐츠 기반 추천과 협업필터링

협업필터링, Collaborative filtering



Unit 02 | 콘텐츠 기반 추천과 협업필터링

협업필터링, Collaborative filtering

- : 사용자와 아이템 간의 상호작용 정보를 통해 사용자 혹은 아이템 간 유사도를 계산하여 유사한 성향의 아이템을 추천해주는 방식
- **User-based**: 아이템에 대해 선호도가 유사한 사용자는 다른 아이템에 대해서도 비슷한 선호도를 보일 가능성이 높다.
 - **Item-based**: 사용자로부터 선호도가 유사한 아이템은 다른 사용자에게 비슷한 선호도를 받을 가능성이 높다.

Unit 02 | 콘텐츠 기반 추천과 협업필터링

협업필터링, Collaborative filtering

: 사용자와 아이템 간의 상호작용 정보를 통해 사용자 혹은 아이템 간 유사도를 계산하여 유사한 성향의 아이템을 추천해주는 방식

- **User-based**: 아이템에 대해 선호도가 유사한 사용자는 다른 아이템에 대해서도 비슷한 선호도를 보일 가능성이 높다.
- **Item-based**: 사용자로부터 선호도가 유사한 아이템은 다른 사용자에게 비슷한 선호도를 받을 가능성이 높다.

✓ Explicit
Feedback

→ 영화 평점, 맛집 별점...

✓ Implicit Feedback

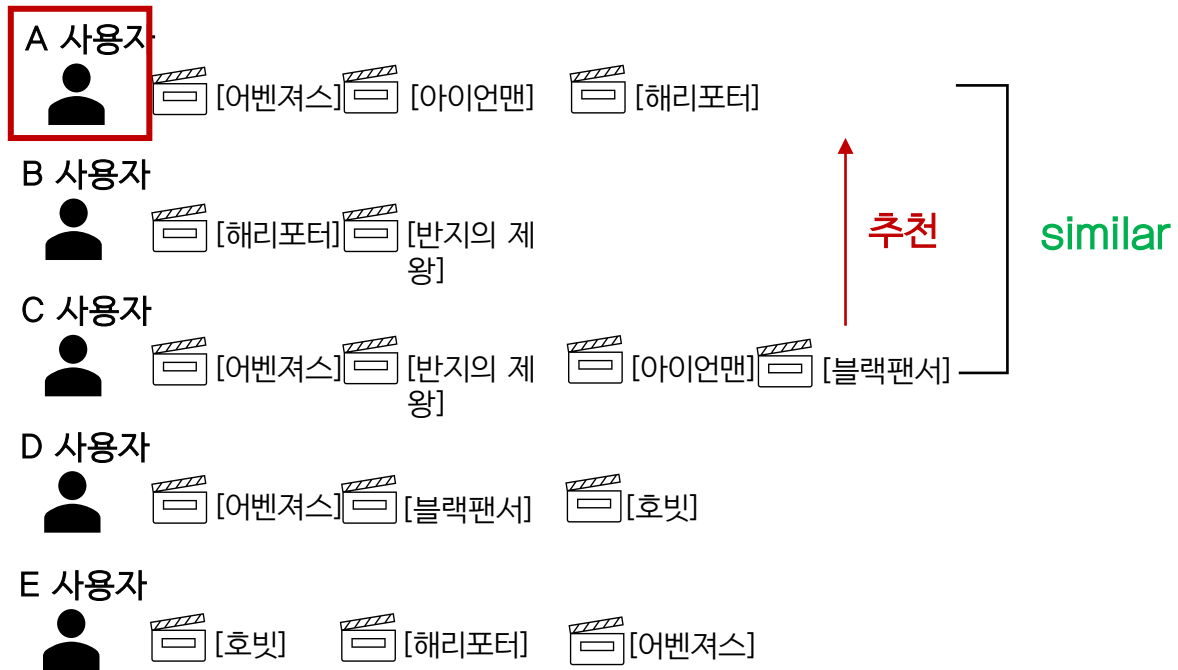
→ 클릭 기록, 머문 시간, 시청 시간, 검색 기록...

Unit 02 | 콘텐츠 기반 추천과 협업필터링

협업필터링, Collaborative filtering

: 사용자와 아이템 간의 상호작용 정보를 통해 사용자 혹은 아이템 간 유사도를 계산하여 유사한 성향의 아이템을 추천해주는 방식

→ **User-based**: 아이템에 대해 선호도가 유사한 사용자는 다른 아이템에 대해서도 비슷한 선호도를 보일 가능성이 높다.



Unit 02 | 콘텐츠 기반 추천과 협업필터링

협업필터링, Collaborative filtering

: 사용자와 아이템 간의 상호작용 정보를 통해 사용자 혹은 아이템 간 유사도를 계산하여 유사한 성향의 아이템을 추천해주는 방식

→ **User-based**: 아이템에 대해 선호도가 유사한 사용자는 다른 아이템에 대해서도 비슷한 선호도를 보일 가능성이 높다.

	어벤져스	아이언맨	반지의 제왕	해리포터	호빗	블랙팬서
A 사용자	5	4		3		
B 사용자	2		5	3		
C 사용자	5	5	2			4
D 사용자	4	5			2	5
E 사용자	2			4	5	

피어슨 상관계수



유사도 계산

Unit 02 | 콘텐츠 기반 추천과 협업필터링

협업필터링, Collaborative filtering

: 사용자와 아이템 간의 상호작용 정보를 통해 사용자 혹은 아이템 간 유사도를 계산하여 유사한 성향의 아이템을 추천해주는 방식

→ **User-based**: 아이템에 대해 선호도가 유사한 사용자는 다른 아이템에 대해서도 비슷한 선호도를 보일 가능성이 높다.

	어벤져스	아이언맨	반지의 제왕	해리포터	호빗	블랙팬서
A 사용자	5	4		3		
B 사용자	2		5	3		
C 사용자	5	5	2			4
D 사용자	4	5			2	5
E 사용자	2			4	5	

피어슨 상관관계수
→
유사도 계산

	A 사용자	B 사용자	C 사용자	D 사용자	E 사용자
A 사용자	1	-0.516	0.707	0.369	-0.834
B 사용자	-0.516	1			
C 사용자	0.707		1		
D 사용자	0.369			1	
E 사용자	-0.834				1

A 사용자와 유사한 사용자

Unit 02 | 콘텐츠 기반 추천과 협업필터링

A 사용자의 사용자 기반 선호도 예측 - 블랙박스

: 사용자와 아이템 간의 상호작용 정보를 통해 사용자 혹은 아이템 간 유사도를 계산하여 유사한 성향의 아이템을 추천해주는 방식

추천 대상 사용자 선호도 평균 비교 사용자의 선호도 평균 추천 대상 사용자와 비교 사용자 간 유사도

$$P_{a,i} = \bar{r}_a + \frac{\sum_{u \in U} (r_{u,i} - \bar{r}_u) \times w_{a,u}}{\sum_{u \in U} |w_{a,u}|}$$

$$4 + \frac{((4 - 4) * 0.707) + ((5 - 4) * 0.369))}{0.707 + 0.369}$$

사용자는 다른 아이템에 대해서도 비슷한 선호도를 보일 가능성이 높다.

A 사용자와 유사한 사용자

	A 사용자	B 사용자	C 사용자	D 사용자	E 사용자
A 사용자	1	-0.516	0.707	0.369	-0.834
B 사용자	-0.516	1			
C 사용자	0.707		1		
D 사용자	0.369			1	
E 사용자	-0.834				1

← 평점 예측

Unit 02 | 콘텐츠 기반 추천과 협업필터링

A 사용자의 사용자 기반 선호도 예측 - 블랙박스

: 사용자와 아이템 간의 상호작용 정보를 통해 사용자 혹은 아이템 간 유사도를 계산하여 유사한 성향의 아이템을 추천해주는 방식

추천 대상 사용자 선호도 평균 비교 사용자의 선호도 평균 추천 대상 사용자와 비교 사용자 간 유사도

$$P_{a,i} = \bar{r}_a + \frac{\sum_{u \in U} (r_{u,i} - \bar{r}_u) \times w_{a,u}}{\sum_{u \in U} |w_{a,u}|}$$

C 사용자 D 사용자

$$4 + \frac{((4 - 4) * 0.707) + ((5 - 4) * 0.369))}{0.707 + 0.369}$$

사용자는 다른 아이템에 대해서도 비슷한 선호도를 보일 가능성이 높다.

A 사용자와 유사한 사용자

	A 사용자	B 사용자	C 사용자	D 사용자	E 사용자
A 사용자	1	-0.516	0.707	0.369	-0.834
B 사용자	-0.516	1			
C 사용자	0.707		1		
D 사용자	0.369			1	
E 사용자	-0.834				1

평점 예측

Unit 02 | 콘텐츠 기반 추천과 협업필터링

협업필터링, Collaborative filtering

: 사용자와 아이템 간의 상호작용 정보를 통해 사용자 혹은 아이템 간 유사도를 계산하여 유사한 성향의 아이템을 추천해주는 방식

→ **User-based**: 아이템에 대해 선호도가 유사한 사용자는 다른 아이템에 대해서도 비슷한 선호도를 보일 가능성이 높다.

A 사용자와 유사한 사용자

	어벤져스	아이언맨	반지의 제왕	해리포터	호빗	블랙팬서
A 사용자	5	4	2	3	2	4.34
B 사용자	2		5	3		
C 사용자	5	5	2			4
D 사용자	4	5			2	5
E 사용자	2			4	5	

← 평점 예측

	A 사용자	B 사용자	C 사용자	D 사용자	E 사용자
A 사용자	1	-0.516	0.707	0.369	-0.834
B 사용자	-0.516	1			
C 사용자	0.707		1		
D 사용자	0.369			1	
E 사용자	-0.834				1

Unit 02 | 콘텐츠 기반 추천과 협업필터링

협업필터링, Collaborative filtering

: 사용자와 아이템 간의 상호작용 정보를 통해 사용자 혹은 아이템 간 유사도를 계산하여 유사한 성향의 아이템을 추천해주는 방식

→ **Item-based**: 사용자로부터 선호도가 유사한 아이템은 다른 사용자에게 비슷한 선호도를 받을 가능성이 높다.



Unit 02 | 콘텐츠 기반 추천과 협업필터링

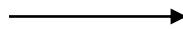
협업필터링, Collaborative filtering

: 사용자와 아이템 간의 상호작용 정보를 통해 사용자 혹은 아이템 간 유사도를 계산하여 유사한 성향의 아이템을 추천해주는 방식

→ **Item-based**: 사용자로부터 선호도가 유사한 아이템은 다른 사용자에게 비슷한 선호도를 받을 가능성이 높다.

	A 사용자	B 사용자	C 사용자	D 사용자	E 사용자
어벤져스	5		4	2	
해리포터	1	5	2		2
아이언맨	4		4	3	5
반지의 제왕	4	4		5	3
트랜스포머	5	2		2	4

보완 코사인 유사도



유사도 계산

Unit 02 | 콘텐츠 기반 추천과 협업필터링

협업필터링, Collaborative filtering

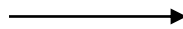
: 사용자와 아이템 간의 상호작용 정보를 통해 사용자 혹은 아이템 간 유사도를 계산하여 유사한 성향의 아이템을 추천해주는 방식

→ **Item-based**: 사용자로부터 선호도가 유사한 아이템은 다른 사용자에게 비슷한 선호도를 받을 가능성이 높다.

	A 사용자	B 사용자	C 사용자		
어벤져스	5		4	2	
해리포터	1	5	2		2
아이언맨	4		4	3	5
반지의 제왕	4	4		5	3
트랜스포머	5	2		2	4
평균	3.8	3.66	3.33	3	3.5

$$w_{i,j} = \frac{\sum_{u \in U} (r_{u,i} - \bar{r}_u)(r_{u,j} - \bar{r}_u)}{\sqrt{\sum_{u \in U} (r_{u,i} - \bar{r}_u)^2} \sqrt{\sum_{u \in U} (r_{u,j} - \bar{r}_u)^2}}$$

보완 코사인 유사도



유사도 계산

Unit 02 | 콘텐츠 기반 추천과 협업필터링

협업필터링, Collaborative filtering

: 사용자와 아이템 간의 상호작용 정보를 통해 사용자 혹은 아이템 간 유사도를 계산하여 유사한 성향의 아이템을 추천해주는 방식

→ **Item-based**: 사용자로부터 선호도가 유사한 아이템은 다른 사용자에게 비슷한 선호도를 받을 가능성이 높다.

	A 사용자	B 사용자	C 사용자	D 사용자	E 사용자
어벤져스	5		4	2	
해리포터	1	5	2		2
아이언맨	4		4	3	5
반지의 제왕	4	4		5	3
트랜스포머	5	2		2	4
평균	3.8	3.66	3.33	3	3.5

보완 코사인 유사도
→
유사도 계산

	어벤져스	해리포터	아이언맨	반지의 제왕	트랜스포머
어벤져스	1				
해리포터		1			
아이언맨			1		
반지의 제왕	-0.560	0.967	-0.22	1	-0.52
트랜스포머	0.99	-0.871	0.39	-0.525	1

Unit 02 | 콘텐츠 기반 추천과 협업필터링

C 사용자의 아이템 기반 선호도 예측 - 트랜스포머

: 사용자와 아이템 간의 상호작용 정보를 통해 사용자 혹은 아이템 간 유사도를 계산하여 유사한 성향의 아이템을 추천해주는 방식

→ Item-based Collaborative Filtering: 유사한 아이템은 다른 사용자에게 비슷한 선호도를 받을 가능성이 높다.

$$P_{a,i} = \frac{\sum_{n \in N} r_{a,n} \times w_{i,n}}{\sum_{n \in N} |w_{i,n}|}$$

예측 아이템과
다른 아이템 간의 유사도

$$3.33 + \frac{((4 \times 0.99) + (2 \times -0.871) + (4 \times 0.39))}{(0.99 + 0.871 + 0.39)}$$

	어벤져스	해리포터	아이언맨	반지의 제왕	트랜스포머
어벤져스	5		4	2	
해리포터		5			2
아이언맨			5	3	5
반지의 제왕	4	4		5	3
트랜스포머	5	2	2	4	4
평균	3.8	3.66	3.33	3	3.5

평점 예측

	어벤져스	해리포터	아이언맨	반지의 제왕	트랜스포머
어벤져스	1				
해리포터		1			
아이언맨			1		
반지의 제왕	-0.560	0.967	-0.22	1	-0.52
트랜스포머	0.99	-0.871	0.39	-0.525	1

Unit 02 | 콘텐츠 기반 추천과 협업필터링

C 사용자의 아이템 기반 선호도 예측 - 트랜스포머

→ Item-based CF: 사용자와 아이템 간의 상호작용 정보를 통해 사용자와 유사한 다른 아이템 간 유사도를 계산하여 유사한 성향의 아이템을 추천해주는 방식

→ Item-based CF: 예측 아이템과 다른 아이템 간의 유사도

$$P_{a,i} = \frac{\sum_{n \in N} r_{a,n} \times w_{i,n}}{\sum_{n \in N} |w_{i,n}|}$$

3.33 + $\frac{((4 \times 0.99) + (2 \times -0.871) + (4 \times 0.39))}{(0.99 + 0.871 + 0.39)}$

	어벤져스	해리포터	아이언맨	C 사용자	E 사용자
어벤져스					
해리포터					
아이언맨					
반지의 제왕	4	4	5	3	5
트랜스포머	5	2	2	2	4
평균	3.8	3.66	3.33	3	3.5

예측 아이템과 다른 사용자에게 비슷한 선호도를 받을 가능성이 높다.

평점 예측

	어벤져스	해리포터	아이언맨	반지의 제왕	트랜스포머
어벤져스	1				
해리포터		1			
아이언맨			1		
반지의 제왕	-0.560	0.967	-0.22	1	-0.52
트랜스포머	0.99	-0.871	0.39	-0.525	1

Unit 02 | 콘텐츠 기반 추천과 협업필터링

협업필터링, Collaborative filtering

: 사용자와 아이템 간의 상호작용 정보를 통해 사용자 혹은 아이템 간 유사도를 계산하여 유사한 성향의 아이템을 추천해주는 방식

→ **Item-based**: 사용자로부터 선호도가 유사한 아이템은 다른 사용자에게 비슷한 선호도를 받을 가능성이 높다.

	A 사용자	B 사용자	C 사용자	D 사용자	E 사용자
어벤져스	5		4	2	
해리포터	1	5	2		2
아이언맨	4		4	3	5
반지의 제왕	4	4	2.23	5	3
트랜스포머	5	2	5	2	4
평균	3.8	3.66	3.33	3	3.5

← 평점 예측

	어벤져스	해리포터	아이언맨	반지의 제왕	트랜스포머
어벤져스	1				
해리포터		1			
아이언맨			1		
반지의 제왕	-0.560	0.967	-0.22	1	-0.52
트랜스포머	0.99	-0.871	0.39	-0.525	1

Unit 02 | 콘텐츠 기반 추천과 협업필터링

콘텐츠 기반 추천, Content-based filtering

장점

- 추천 대상 사용자의 독립적인 정보 혹은 아이템 콘텐츠 데이터만 필요하므로 다른 사용자의 정보가 부족하여도 추천이 가능함
- 아이템 혹은 새로운 아이템에 대한 평가 데이터가 존재하지 않더라도 아이템 간 속성 데이터를 파악하여 추천이 가능함
- 추천에 아이템 속성이 활용되기 때문에 추천에 대한 설명이 명시적으로 가능함

단점

- 과거 구매 데이터, 아이템 선호 데이터가 부족하다면 추천이 불가능할 수 있음
- 사용자가 선호하는 아이템과 유사한 아이템만이 추천되기 때문에 추천 후보 리스트의 다양성이 보장되지 않음
- 아이템 속성 추출을 위한 도메인 지식이 요구됨

Unit 02 | 콘텐츠 기반 추천과 협업필터링

협업필터링, Collaborative filtering

장점

- 사용자와 아이템 간의 선호도 혹은 구매 정보만을 사용하기 때문에 도메인 지식이 많이 요구되지 않음
- 다른 사용자라는 이웃을 활용하기 때문에 아이템 추천의 다양성이 어느정도 보장됨

단점

- 사용자-아이템 행렬이 희소하게 되면 추천의 성능이 현저히 떨어짐
- 새로운 사용자의 경우, 다른 사용자와 유사도를 계산할 데이터가 존재하지 않아 추천이 이뤄지지 않는 **Cold start problem**이 발생함
- 선호도 평가를 받지 않은 아이템은 추천이 이뤄지지 않는 **First rater problem**이 발생함
- 소수의 인기있는 아이템이 전체 추천 후보 리스트에 많은 비율을 차지하게 되는 **Long-tail problem**이 발생함
- 실시간으로 갱신되는 사용자-아이템 행렬을 통해 추천이 유지되어야 하기 때문에 빠른 계산 속도와 메모리가 요구됨

Unit 02 | 콘텐츠 기반 추천과 협업필터링

하이브리드 시스템, Hybrid System

: 각 추천 알고리즘의 장점을 극대화하면서 단점은 보완하고 다양한 정보를 효과적으로 활용할 수 있도록 다양한 추천 모델을 결합하는 방식

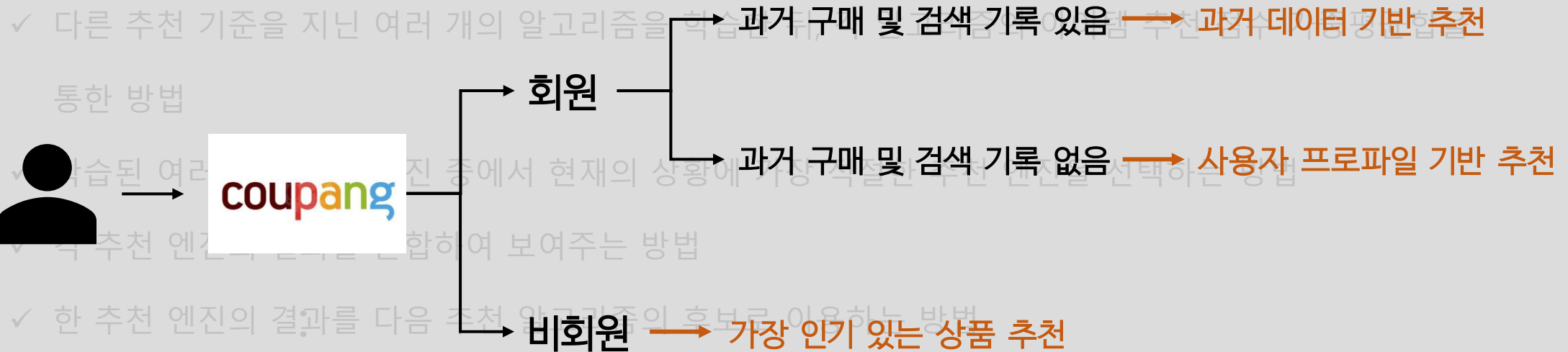
- ✓ 다른 추천 기준을 지닌 여러 개의 알고리즘을 학습한 뒤, 각 알고리즘의 아이템 추천 점수 가중평균합을 통한 방법
- ✓ 학습된 여러 개의 추천 엔진 중에서 현재의 상황에 가장 적절한 추천 엔진을 선택하는 방법
- ✓ 각 추천 엔진의 결과를 혼합하여 보여주는 방법
- ✓ 한 추천 엔진의 결과를 다음 추천 알고리즘의 후보로 이용하는 방법

⋮

Unit 02 | 콘텐츠 기반 추천과 협업필터링

하이브리드 시스템, Hybrid System

: 각 추천 알고리즘의 장점을 극대화하면서 단점은 보완하고 다양한 정보를 효과적으로 활용할 수 있도록 다양한 추천 모델을 결합하는 방식



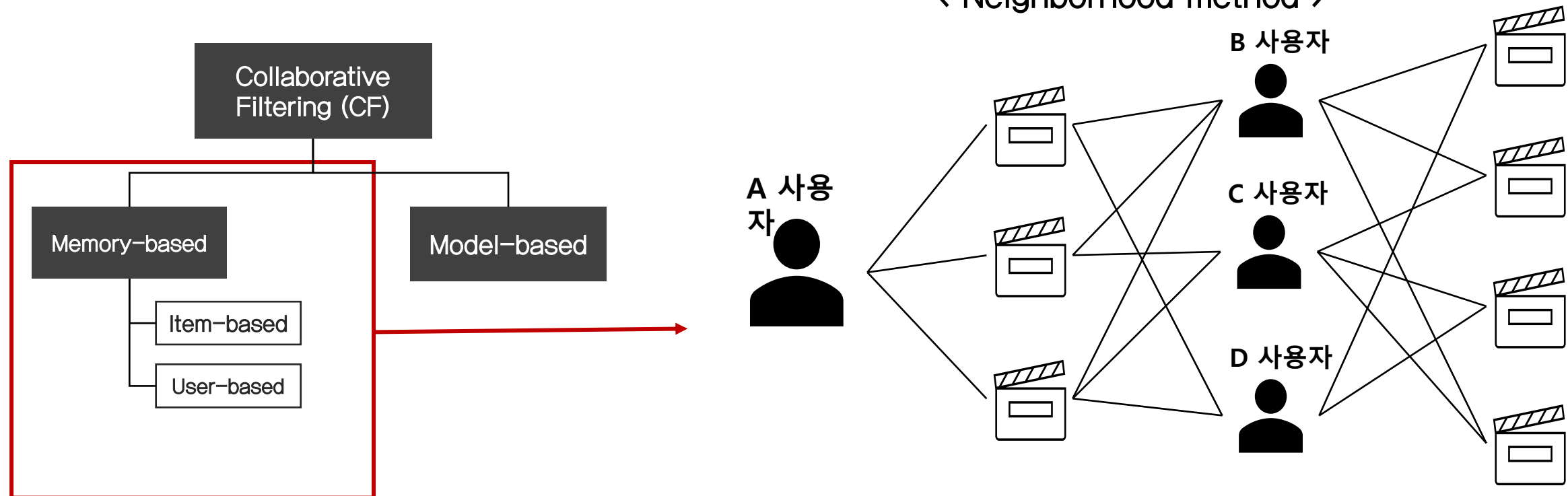
Unit 03 | Matrix Factorization

Unit 03 | Matrix Factorization

Collaborative Filtering – Memory-based

: 평점 혹은 선호도 데이터를 통해 아이템 간의 혹은 사용자 간의 유사도를 계산하고 관계를 정의하여 추천 하는 방식
(Explicit data)

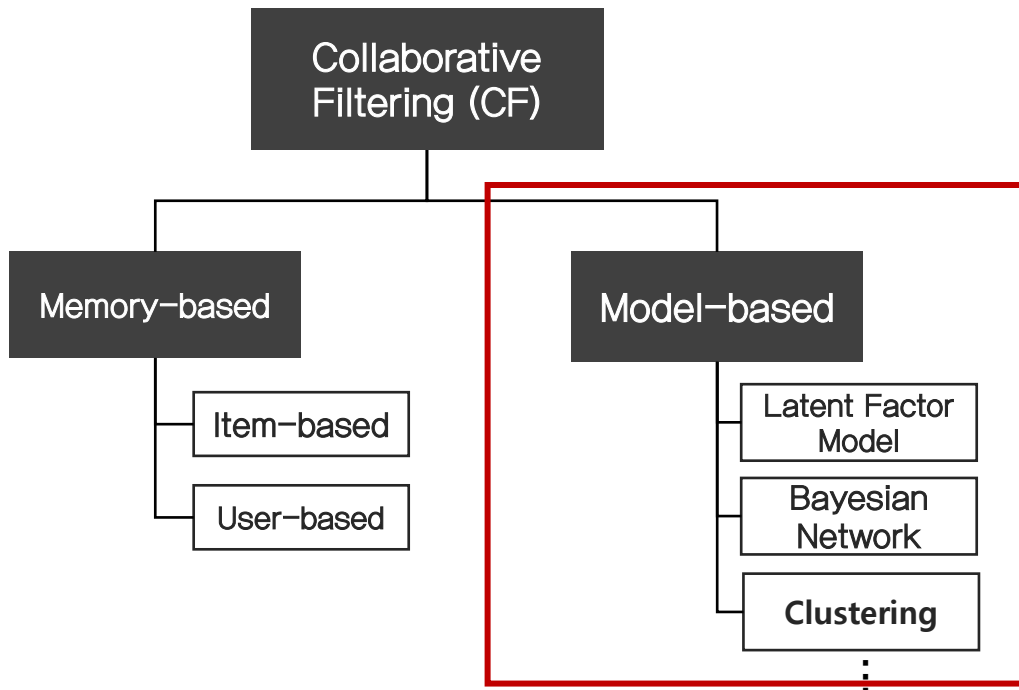
< Neighborhood method >



Unit 03 | Matrix Factorization

Collaborative Filtering – Model-based

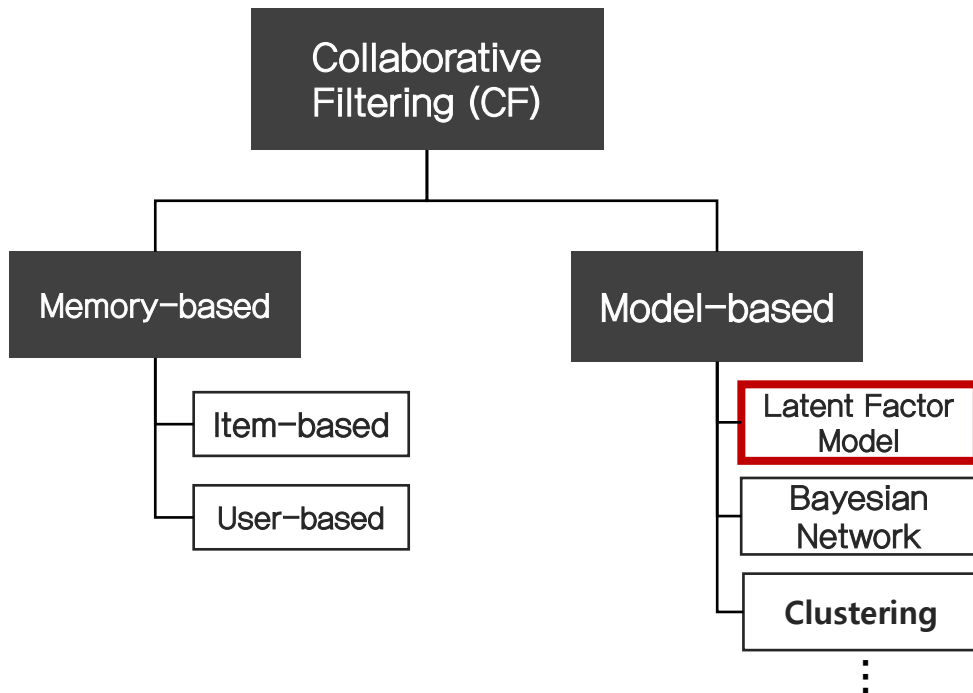
: 사용자-아이템 행렬의 학습데이터를 통해 모델을 구축한 뒤, 사용자에게 아이템을 추천하는 모델 기반 협력필터링 방식



Unit 03 | Matrix Factorization

Collaborative Filtering – Latent Factor Model

: 사용자와 아이템 간의 관계에서 잠재 요인(Latent Factor)을 찾아내고 잠재 요인을 기반으로 추천이 이뤄지는 방식

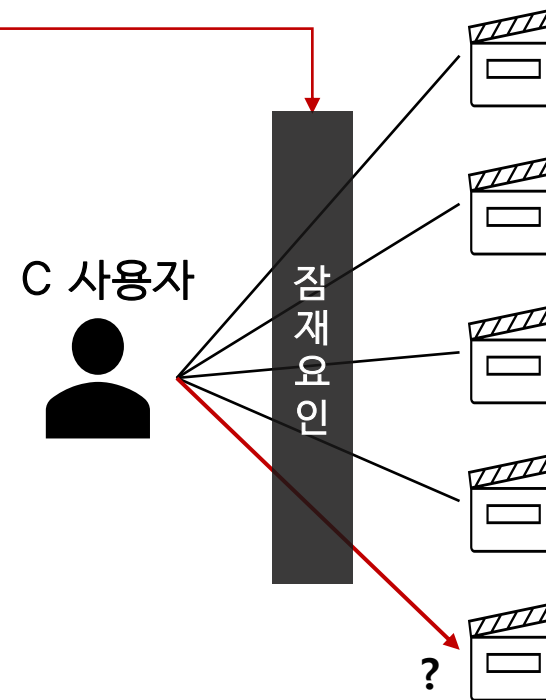


Unit 03 | Matrix Factorization

Collaborative Filtering – Latent Factor Model

: 사용자와 아이템 간의 관계에서 **잠재 요인(Latent Factor)**을 찾아내고 잠재 요인을 기반으로 추천이 이뤄지는 방식

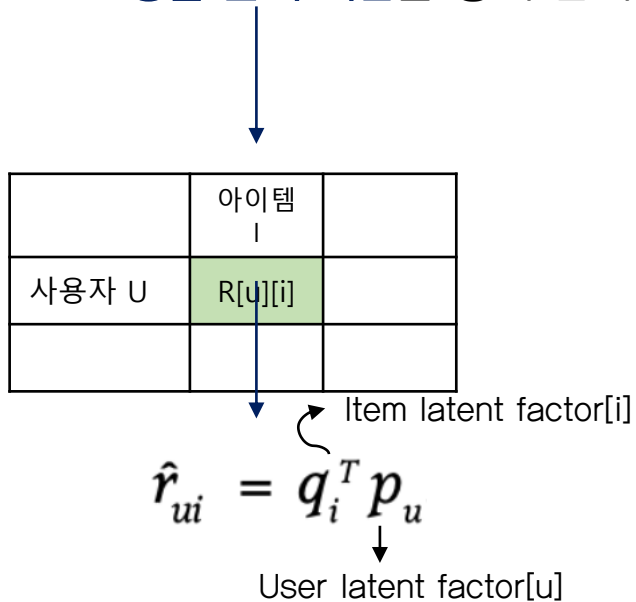
	어벤져스	아이언맨	반지의 제왕	해리포터	호빗	블랙팬서
A 사용자	5	4		3		
B 사용자	2		5	3		
C 사용자	5	5	2	1		4
D 사용자	4	5			2	5
E 사용자	2			4	5	



Unit 03 | Matrix Factorization

Matrix Factorization

: 행렬 분해 기법을 통해 잠재 요인(Latent Factor)을 찾아내고 행렬 복원(Matrix Completion)을 거쳐 추천이 이뤄지는 방식



Unit 03 | Matrix Factorization

Matrix Factorization

: 행렬 분해 기법을 통해 잠재 요인(Latent Factor)을 찾아내고 행렬 복원(Matrix Completion)을 거쳐 추천이 이뤄지는 방식



Unit 03 | Matrix Factorization

Matrix Factorization

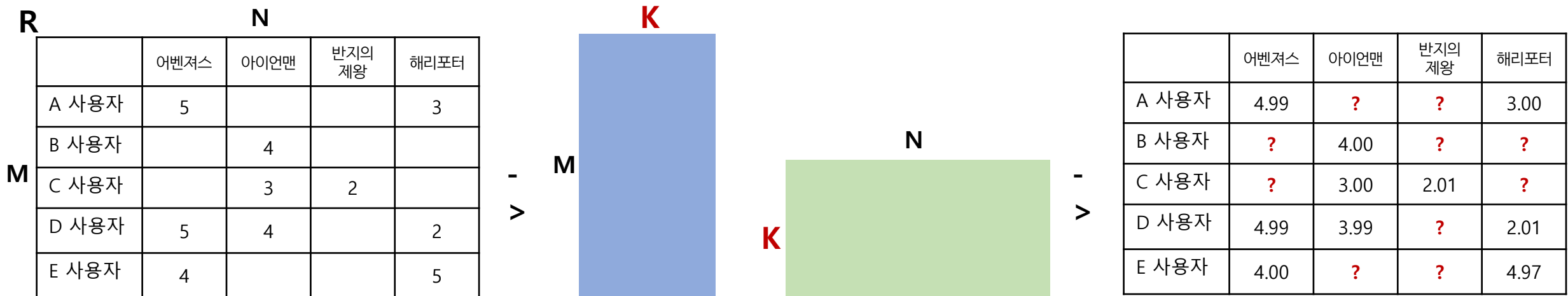
: 행렬 분해 기법을 통해 잠재 요인(Latent Factor)을 찾아내고 행렬 복원(Matrix Completion)을 거쳐 추천이 이뤄지는 방식



Unit 03 | Matrix Factorization

Matrix Factorization

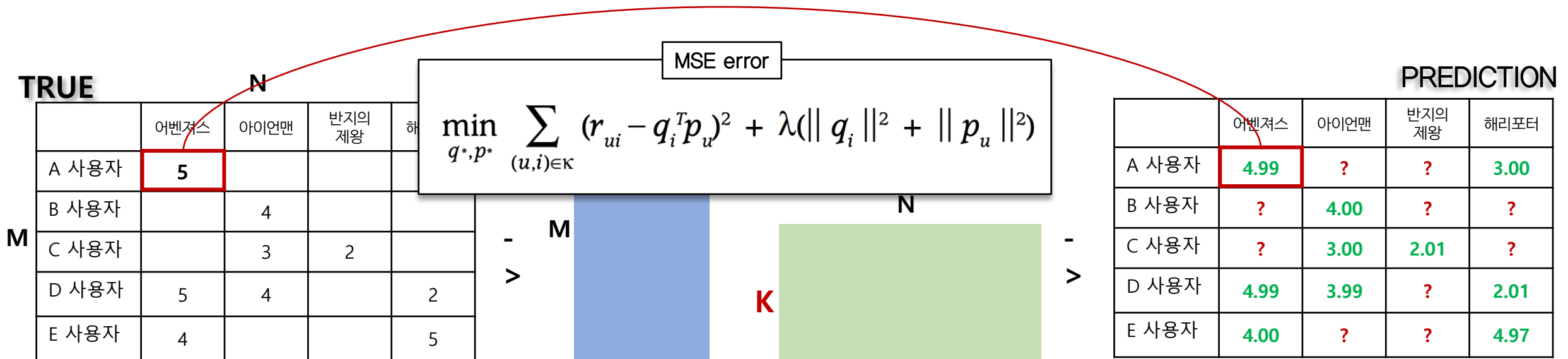
: 행렬 분해 기법을 통해 잠재 요인(Latent Factor)을 찾아내고 행렬 복원(Matrix Completion)을 거쳐 추천이 이뤄지는 방식



Unit 03 | Matrix Factorization

Matrix Factorization

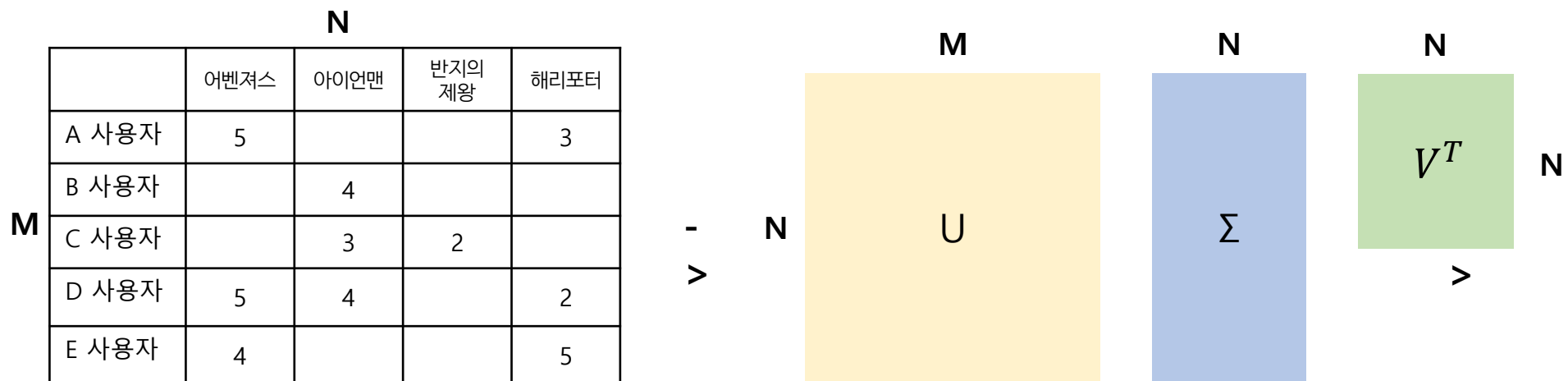
: 행렬 분해 기법을 통해 잠재 요인(Latent Factor)을 찾아내고 행렬 복원(Matrix Completion)을 거쳐 추천이 이뤄지는 방식



Unit 03 | Matrix Factorization

Matrix Factorization – SVD, Singular Value Decomposition

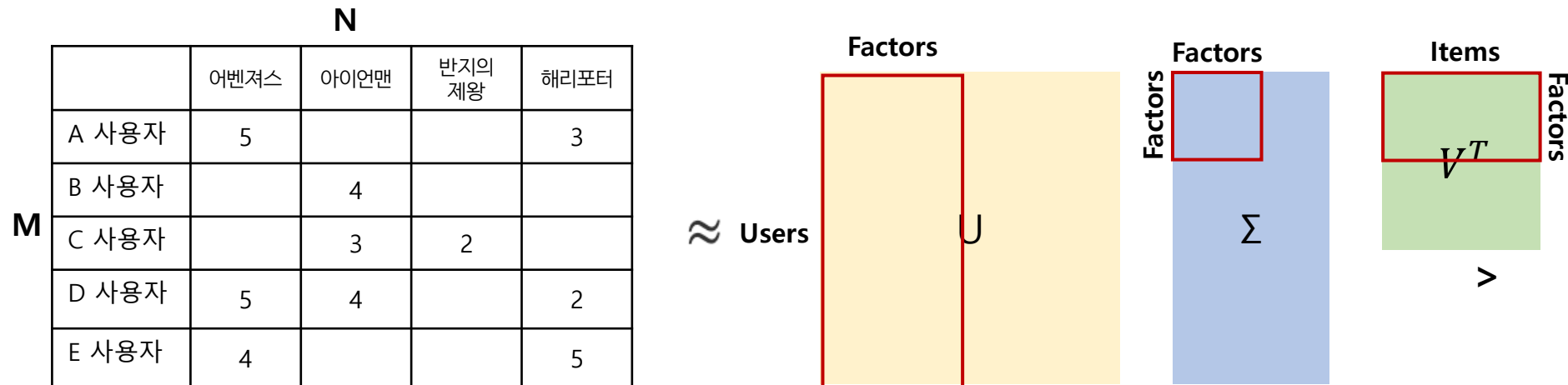
: 특이값의 대각행렬을 기준으로 행렬을 3개의 행렬의 곱으로 나타내는 방법



Unit 03 | Matrix Factorization

Matrix Factorization – SVD, Singular Value Decomposition

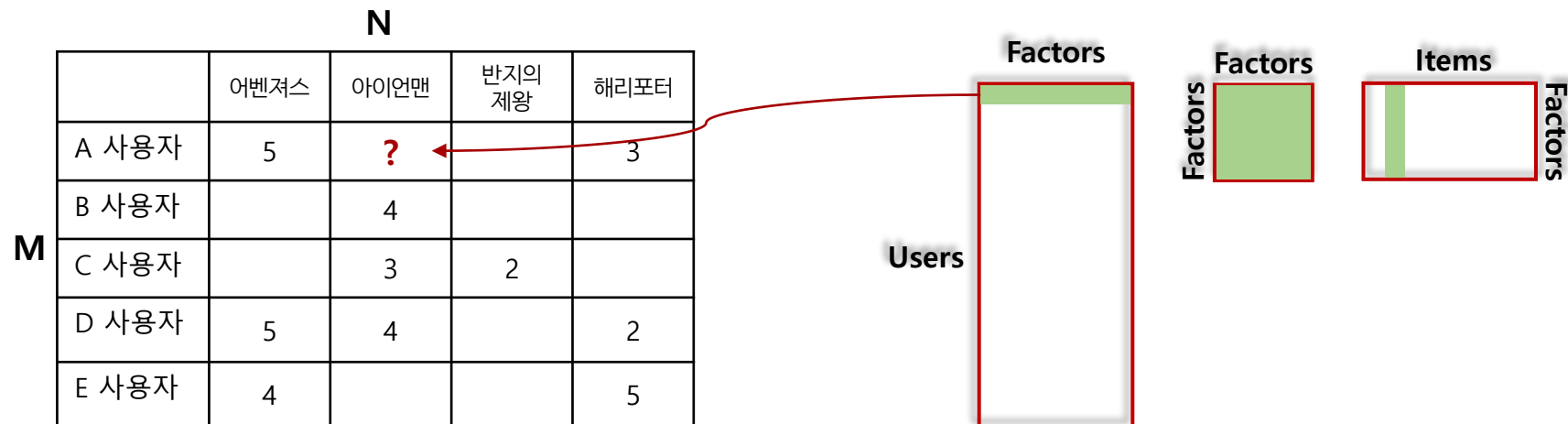
: 특이값의 대각행렬을 기준으로 행렬을 3개의 행렬의 곱으로 나타내는 방법으로 상위 특이값 K개만 사용하는 Truncated SVD



Unit 03 | Matrix Factorization

Matrix Factorization – SVD, Singular Value Decomposition

: 특이값의 대각행렬을 기준으로 행렬을 3개의 행렬의 곱으로 나타내는 방법으로 상위 특이값 K개만 사용하는 Truncated SVD



Unit 03 | Matrix Factorization

Latent Factor Model – Matrix Factorization

장점

- 명시적으로 설명하기 어려운, 정의되지 않은 부분이 잠재 요인으로 표현될 수 있음
- 잠재 요인, latent factor를 찾음으로써 차원이 축소되고 저장 공간을 절약할 수 있음
- 사용자-선호도 행렬과 같은 explicit data뿐만 아니라 implicit data에서도 잠재 요인을 추출하여 활용할 수 있음

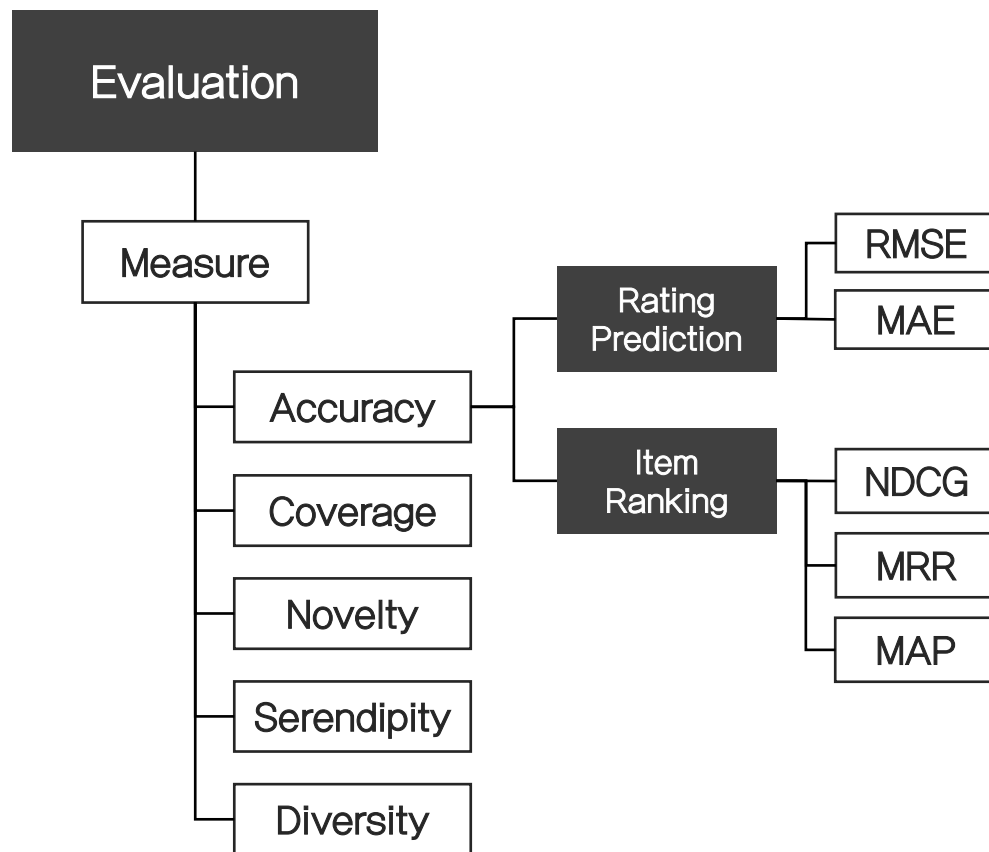
단점

- 잠재 요인에 대한 명백한 해석이 어려움
- 사용자-아이템 선호도 행렬의 결측치 비율이 높다면(Sparse) 좋은 성능을 이끌어내기 어려움
- 사용자 프로필, 아이템 콘텐츠에 대한 데이터는 활용할 수는 없음

Unit 04 | 추천 시스템 평가방법

Unit 04 | 추천 시스템 평가방법

추천 시스템 평가방법



Unit 04 | 추천 시스템 평가방법

점수 예측 추천의 평가방법 – Rating Prediction

: 아이템에 대한 실제 점수(선호도 및 별점)와 예측 값의 차이를 고려한 평가 지표

- MAE(Mean Absolute Error)

$$MSE = \frac{1}{N} \sum (p_{i,j} - r_{i,j})^2$$

- MSE(Mean Squared Error)

$$MSE = \frac{1}{N} \sum (p_{i,j} - r_{i,j})^2$$

- RMSE(Root Mean Squared Error)

$$RMSE = \sqrt{\frac{1}{N} \sum (p_{i,j} - r_{i,j})^2}$$

- MMSE(Masked Mean Squared Error)
- NMAE(Normalized Mean Absolute Error)

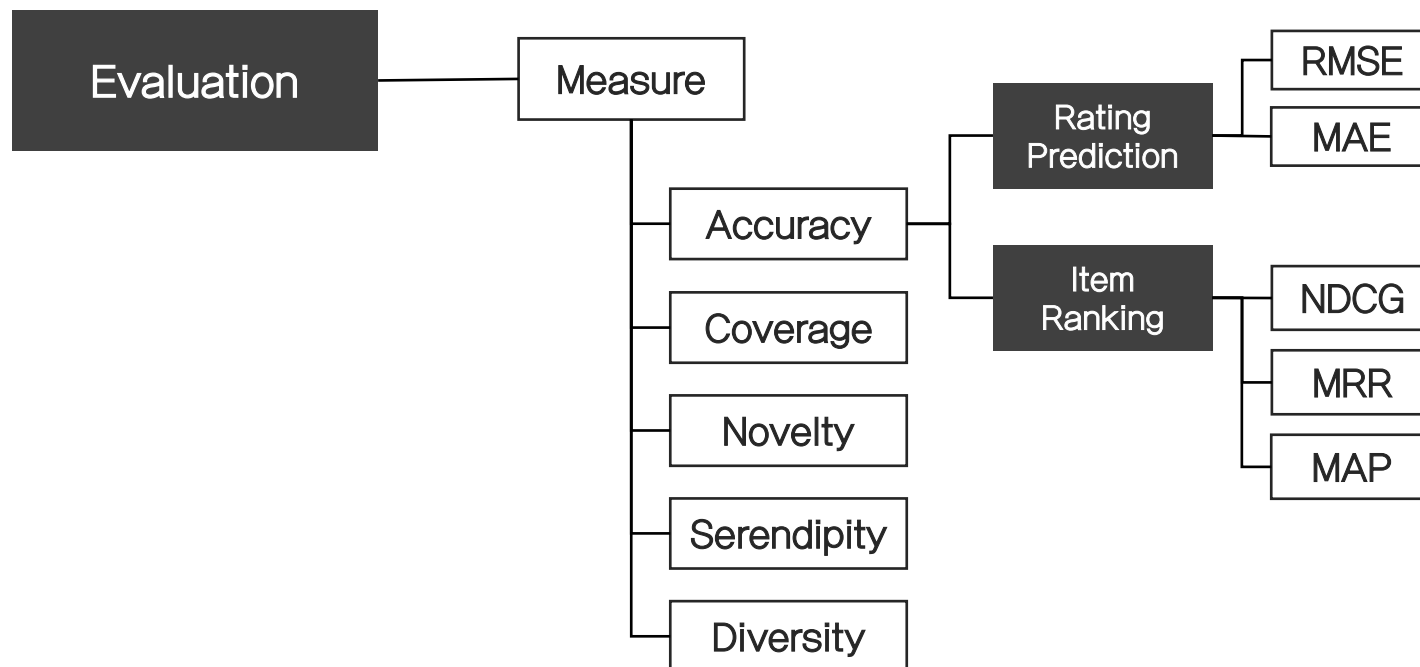
	어벤져스	아이언맨	반지의 제왕	해리포터
A 사용자	5			3
B 사용자		4		
C 사용자		3	2	
D 사용자	5	4		2
E 사용자	4			5

	어벤져스	아이언맨	반지의 제왕	해리포터
A 사용자	4.99	?	?	3.00
B 사용자	?	4.00	?	?
C 사용자	?	3.00	2.01	?
D 사용자	4.99	3.99	?	2.01
E 사용자	4.00	?	?	4.97

Unit 04 | 추천 시스템 평가방법

랭킹 기반 추천의 평가방법 – Item Ranking

: 추천 아이템의 순서와 순위를 고려한 평가 지표



Unit 04 | 추천 시스템 평가방법

랭킹 기반 추천의 평가방법- MRR(Mean Reciprocal Rank)

: 추천된 아이템 랭킹 중 사용자가 소비한 아이템 순위의 역수를 사용한 평가 지표

순위	아이템	예측
1	어벤져스	어벤져스
2	해리포터	트랜스포머
3	트랜스포머	해리포터
4	아이언맨	아이언맨
5	토르	다크나이트
6	다크나이트	호빗
7	호빗	반지의 제왕
8	블랙팬서	아이덴티티
9	반지의 제왕	인터스텔라
10	분노의 질주	블랙팬서

▪ MRR(Mean Reciprocal Rank)

RR(Reciprocal Rank) = 1/3

⋮

$$\text{MRR} = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \frac{1}{\text{rank}_i}$$

Unit 04 | 추천 시스템 평가방법

랭킹 기반 추천의 평가방법- MAP(Mean Average Precision)

: 특정 상위 N개의 추천 아이템의 정밀도(Precision)를 고려한 평가 지표

	Predict Positive	Predict Negative
Actual Positive	True Positive	False Negative
Actual Negative	False Positive	True Negative

$$Precision = \frac{\text{옳게 추천한 아이템 개수}}{\text{추천한 전체 아이템 개수}}$$

$$Recall = \frac{\text{옳게 추천한 아이템 개수}}{\text{고객이 실제로 구매한 아이템 개수}}$$

Unit 04 | 추천 시스템 평가방법

랭킹 기반 추천의 평가방법- MAP(Mean Average Precision)

: 특정 상위 K개의 추천 아이템의 정밀도(Precision)를 고려한 평가 지표

순위	아이템	정답
1	어벤저스	정답
2	해리포터	정답
3	트랜스포머	정답
4	아이언맨	오답
5	토르	정답
6	다크나이트	오답
7	호빗	정답
8	블랙팬서	오답
9	반지의 제왕	오답
10	분노의 질주	정답

P(k=3) = 1

P(k=5) = 0.8

P(k=7) = 0.714

P(k=10) = 0.6

▪ AP(Average Precision)

$$AP@K = \frac{1}{m} \sum_{j=1}^K P(j) \cdot rel(j) \dots \begin{cases} rel(j) = 1 & \text{if } j^{th} \text{ item is relevant,} \\ rel(j) = 0 & \text{if } j^{th} \text{ item is not relevant,} \end{cases}$$

- K : Cutoff 개수
- m : 추천 아이템 중 relevance가 있는 아이템 개수
- j : 해당 추천 아이템의 index
- P(j) : j번째까지의 precision
- rel(j) : j번째의 relevance 여부

Unit 04 | 추천 시스템 평가방법

랭킹 기반 추천의 평가방법- MAP(Mean Average Precision)

: 특정 상위 K개의 추천 아이템의 정밀도(Precision)를 고려한 평가 지표

순위	아이템	정답
1	어벤저스	정답
2	해리포터	정답
3	트랜스포머	정답
4	아이언맨	오답
5	토르	정답
6	다크나이트	오답
7	호빗	정답
8	블랙팬서	오답
9	반지의 제왕	오답
10	분노의 질주	정답

P(k=3) = 1

P(k=5) = 0.8

P(k=7) = 0.714

P(k=10) = 0.6

■ MAP(Mean Average Precision)

$$AP@5 = \frac{1}{4} \cdot \left(\frac{1}{1} + \frac{2}{2} + \frac{3}{3} + 0 + \frac{4}{5} \right) = 0.95$$

$$AP@7 = \frac{1}{5} \cdot \left(\frac{1}{1} + \frac{2}{2} + \frac{3}{3} + 0 + \frac{4}{5} + 0 + \frac{5}{7} \right) = 0.90$$

$$AP@10 = \frac{1}{6} \cdot \left(\frac{1}{1} + \frac{2}{2} + \frac{3}{3} + 0 + \frac{4}{5} + 0 + \frac{5}{7} + 0 + 0 + \frac{6}{10} \right) = 0.85$$

$$MAP@K = \frac{1}{U} \sum_{u=1}^U (AP@K)_u$$

Unit 04 | 추천 시스템 평가방법

랭킹 기반 추천의 평가방법- NDCG(Normalized Discounted Cumulative Gain)

: 아이템 관련성(relevance)을 기반으로 랭킹을 고려한 평가 지표

- CG(Cumulative Gain): 각 아이템의 relevance의 합

$$CG_p = \sum_{i=1}^p rel_i$$

- DCG(Discounted Cumulative Gain): 추천 순위에 따른 relevance의 합

$$DCG_p = \sum_{i=1}^p \frac{rel_i}{\log_2(i+1)}$$

$$DCG_p = \sum_{i=1}^p \frac{2^{rel_i} - 1}{\log_2(i+1)}$$

- NDCG(Normalized Discounted Cumulative Gain): ideal DCG를 통해 정규화된 DCG

$$nDCG_p = \frac{DCG_p}{IDCG_p}$$

Unit 04 | 추천 시스템 평가방법

랭킹 기반 추천의 평가방법- NDCG(Normalized Discounted Cumulative Gain)

: 아이템 관련성(relevance)을 기반으로 랭킹을 고려한 평가 지표

순위	Ideal	추천 01	relevance	추천 02	relevance
1	3	A	3	A	3
2	3	E	1	B	3
3	2	C	2	C	2
4	2	D	2	G	0
5	1	F	1	E	1

▪ DCG(Discounted Cumulative Gain)

- 추천01 $\frac{3}{\log_2(2)} + \frac{1}{\log_2(3)} + \frac{2}{\log_2(4)} + \frac{2}{\log_2(5)} + \frac{1}{\log_2(6)} = 5.879$
- 추천02 $\frac{3}{\log_2(2)} + \frac{3}{\log_2(3)} + \frac{2}{\log_2(4)} + \frac{0}{\log_2(5)} + \frac{1}{\log_2(6)} = 6.280$

▪ IDCG

$$\frac{3}{\log_2(2)} + \frac{3}{\log_2(3)} + \frac{2}{\log_2(4)} + \frac{2}{\log_2(5)} + \frac{1}{\log_2(6)} = 7.141$$

▪ NDCG(Normalized Discounted Cumulative Gain)

- 추천01 $\frac{5.879}{7.141} = 0.823$
- 추천02 $\frac{6.280}{7.141} = 0.879$

Unit 04 | 추천 시스템 평가방법

기타 추천 시스템 평가방법

- **Coverage:** 전체 아이템 중 추천 모델이 추천에 사용한 아이템의 비율
- **Diversity:** 추천 리스트에 포함되어 있는 아이템의 다양성
- **Novelty:** 사용자가 알지 못하는 아이템 추천의 새로움
- **Serendipity:** 사용자 만족을 높일 수 있는 예상치 못한 추천의 우연성

Unit 04 | 추천 시스템 평가방법

기타 추천 시스템 평가방법

- **Coverage:** 전체 아이템 중 추천 모델이 추천에 사용한 아이템의 비율

추천 시스템의 궁극적 목표

- **Diversity:** 추천 리스트에 포함되어 있는 아이템의 다양성

고객 및 사용자로 하여금 **만족도**를 극대화하여 지속적, 장기적으로 서비스를 운영하는 것

- **Serendipity:** 사용자 만족을 높일 수 있는 예상치 못한 추천의 우연성

✓ 정확성
✓ 다양성
✓ 참신성
✓ 우연성

최신 트렌드

- 정보검색이론 및 추천시스템 2020 Trend(SIGIR)

Topics

- Neural IR and Semantic Matching
- Knowledge and Explainability
- Graph-based Analysis
- Knowledge for Personalization
- User Behavior and Experience
- Evaluation
- Bias and Fairness
- Learning to Rank
- Question Answering
- Query and Representation
- Graph-based Recommendation
- Neural Networks and Embedding
- Learning for Recommendation
- Information Access and Filtering
- Neural Collaborative Filtering 1~2
- Context-aware Modeling
- Conversation and Interactive IR
- Text Classification and Transfer Learning
- Multi-modal Retrieval and Ranking
- Sequential Recommendation
- Domain Specific Retrieval Tasks
- Domain Specific Applications 1~2
- Industrial Session 1~6
- Short Paper

최신 트렌드

- 추천시스템 (RecSys 2020 Workshop)

RecSys Workshops

- CARS: Context-Aware Recommender Systems
- ComplexRec: Recommendation in Complex Environments
- FAccTRec: Responsible Recommendation
- fashionXrecsys: Recommender Systems in Fashion and Retail
- HealthRecSys: Health Recommender Systems
- ImpactRS: Impact of Recommender Systems
- IntRS: Interfaces and Human Decision Making for Recommender Systems
- OHARS: Online Misinformation- and Harm-Aware Recommender Systems
- ORSUM: Online Recommender Systems and User Modeling
- PodRecs: Podcast Recommendations
- REVEAL: Bandit and Reinforcement Learning from User Interactions
- RecSys Challenge 2020 Workshop

참고자료

- [추천 시스템 기법 연구동향 및 분석](#), 2015, 손지은 외 4명
- T아카데미, [추천시스템 분석 경진대회 참여하기](#), 김현우
- [협업 필터링 추천 시스템 \(Collaborative Filtering Recommendation System\)](#), scvgoe
- [Matrix Factorization Techniques for Recommender Systems](#), jeongchul
- [Matrix Factorization 설명 및 논문 리뷰](#), Gorio's learning
- [정보 검색\(Information Retrieval\) 평가는 어떻게 하는 것이 좋을까?\(2/2\)](#), lamttic's playgroud
- [Normalized Discounted Cumulative Gain \(nDCG\) 예제와 코드](#), ride-or-die's blog
- [SVD와 PCA, 그리고 잠재의미분석\(LSA\)](#), ratsgo's blog
- [\[선형대수학#4\] 특이값 분해\(Singular Value Decomposition, SVD\)의 활용](#), 다크 프로그래머
- [\[KR\] 추천시스템의 평가 지표 : MAP](#), Blog from swy
- [\[KR\] 추천시스템의 평가 지표 : nDCG](#), Blog from swy
- [추천시스템 Basics](#), 이웅원
- [갈아먹는 추천 알고리즘](#), 갈아먹는 머신러닝

#assignment

#assignment – 추천 시스템 기초 실습

- Data: MovieLens
 - movies.csv: 영화 제목과 영화 장르에 대한 데이터
 - ratings.csv: user(사용자)-movie(아이템) 선호도 데이터
 - tags.csv: user가 해당 movie에 부여한 tag 데이터
- Content-based Filtering 실습: 특정 아이템과 가장 유사한 아이템 TOP10 추천
- Collaborate Filtering 실습: 특정 사용자와 가장 유사한 사용자의 가장 높은 선호도 영화 TOP5 추천
- Matrix Factorization 실습
- 본인만의 추천 리스트 실습: 각 user의 다음 추천 아이템 TOP10 리스트 만들어보기

Q & A

들어주셔서 감사합니다.