

Collaborative Filtering

협업 필터링

마차차 이지연

Collaborative Filtering

협업 필터링이란?

협업 필터링이란 특정 집단에서 발생하는 유사한 사용행동을 파악하여, 비슷한 성향의 사람들에게 아이템을 추천하는 기술이다. 협업 필터링은 성향이 비슷하면, 선호하는 것도 비슷할 것이라는 가정을 전제로 한다.

Collaborative Filtering

협업 필터링이란?

예를 들어, 사용자 A가 온라인 몰에서 선크림과 튜브, 그리고 수영복 함께 구매하고, 또 다른 사용자 B는 선크림과 튜브를 구매했다고 가정해 보자. 알고리즘은 구매 목록이 겹치는 이 두 사용자가 유사하다고 판단해, 사용자 B에게 수영복을 추천한다.

Collaborative Filtering

Memory-Based Approach (유사한 사용자나 아이템을 사용)

특징 : 최적화 방법이나, 매개변수를 학습하지 않음. 단순한 산술 연산만 사용

방법 : Cosine Similarity나 Pearson Correlation을 사용함

장점 : 결과의 설명력이 좋음, 도메인에 의존적이지 않음

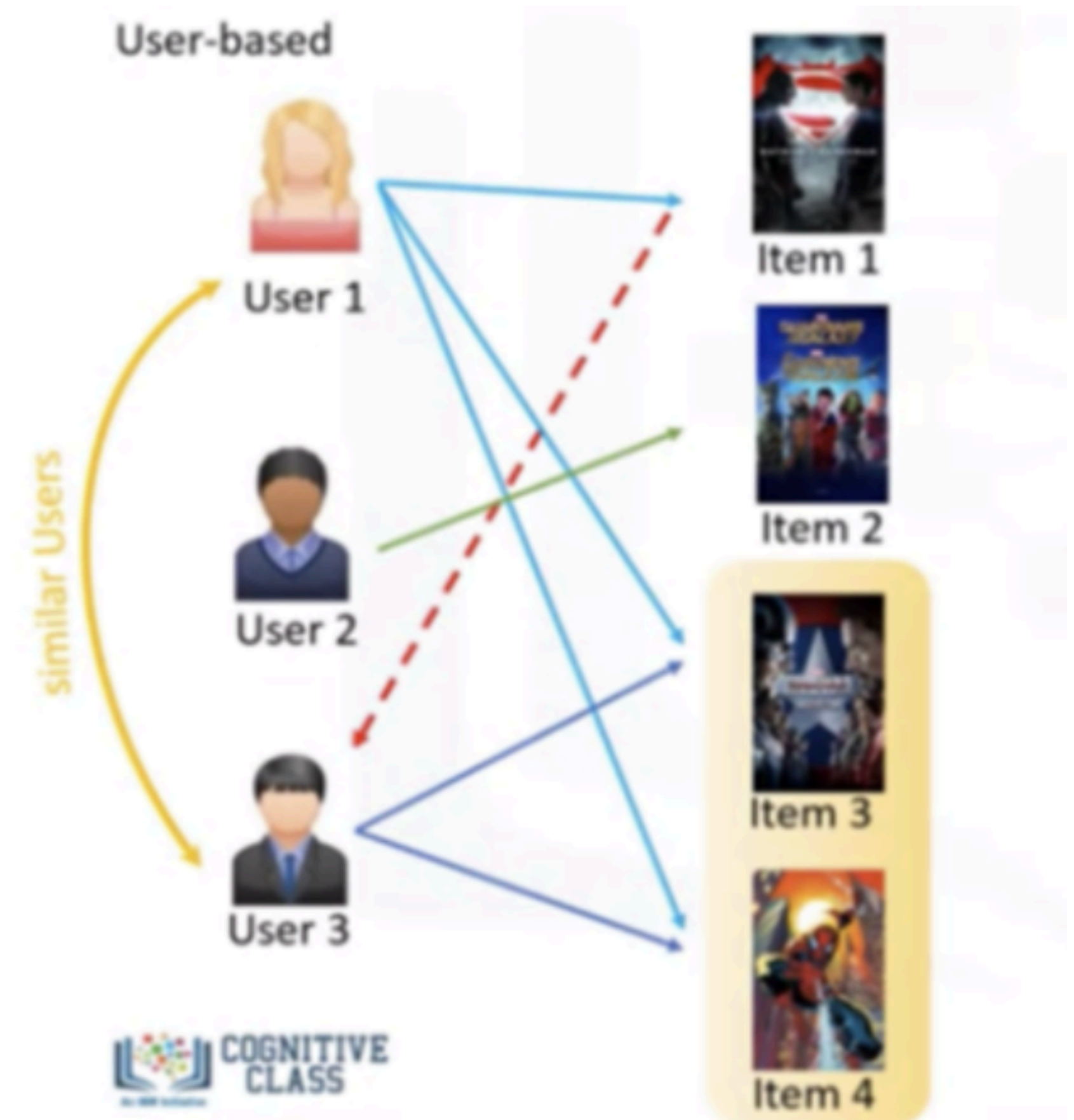
단점 : 확장가능성이 낮음 (∴ 데이터가 너무 많아지면, 속도가 저하됨)

Memory-Based Approach

User-based Filtering

: 특정 사용자(User)를 선택

- ex) 1. 평점 유사도를 기반으로 나와 유사한 사용자(Users)들을 찾음
2. 유사한 사용자가 좋아한 Item을 추천
→ 해석 : 당신과 비슷한 사용자 "A"은,
"B"영화도 좋아했습니다

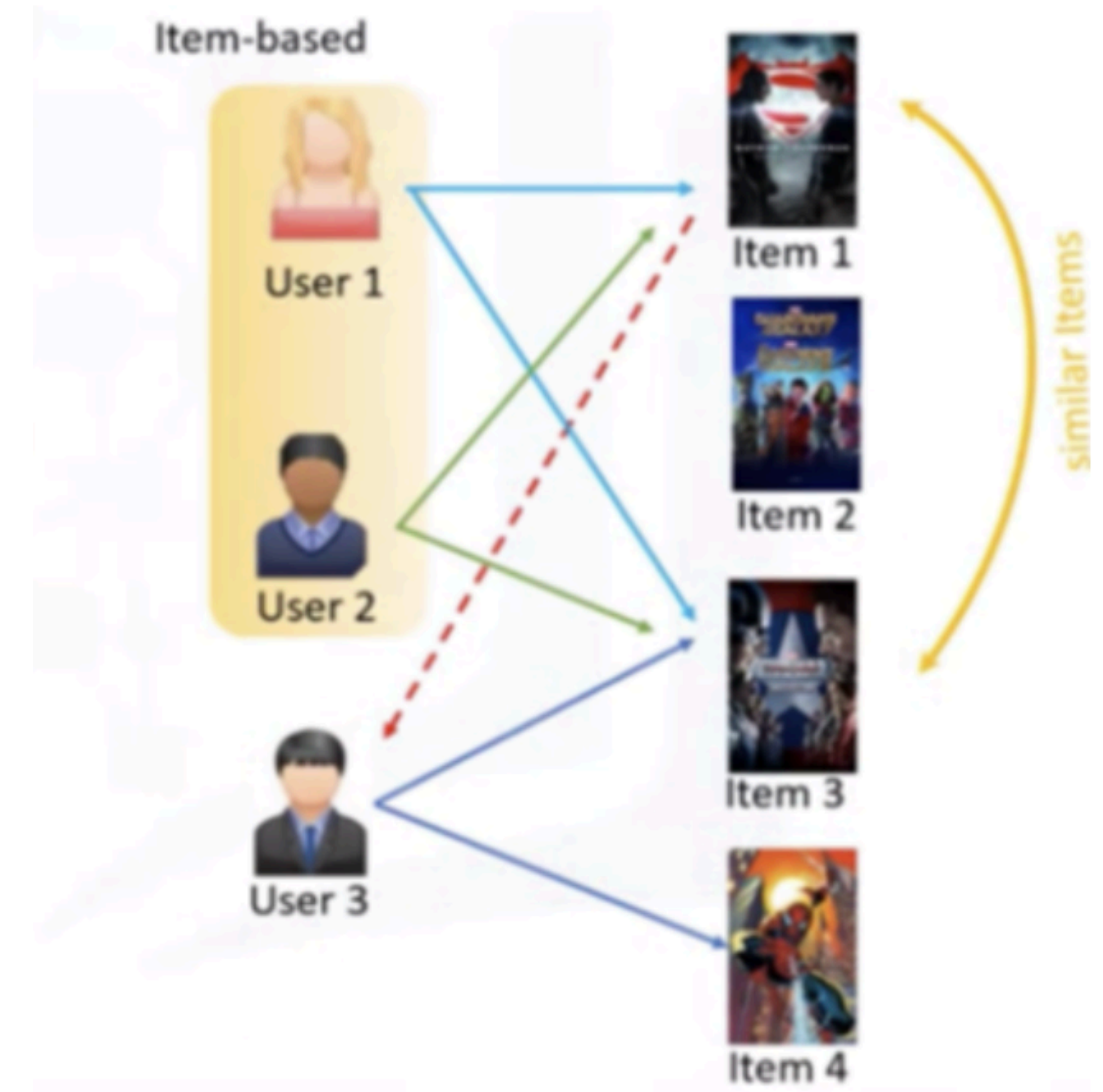


Memory-Based Approach

Item-based Filtering

: 특정 아이템(Item)을 선택

- ex) 1. 특정 Item을 좋아하는 사용자들을 찾음
2. 그 사용자들이 공통적으로 좋아했던 다른 Item을 찾음
→ 해석 : 이 아이템을 좋아하는 사용자는,
“B”영화도 좋아했습니다



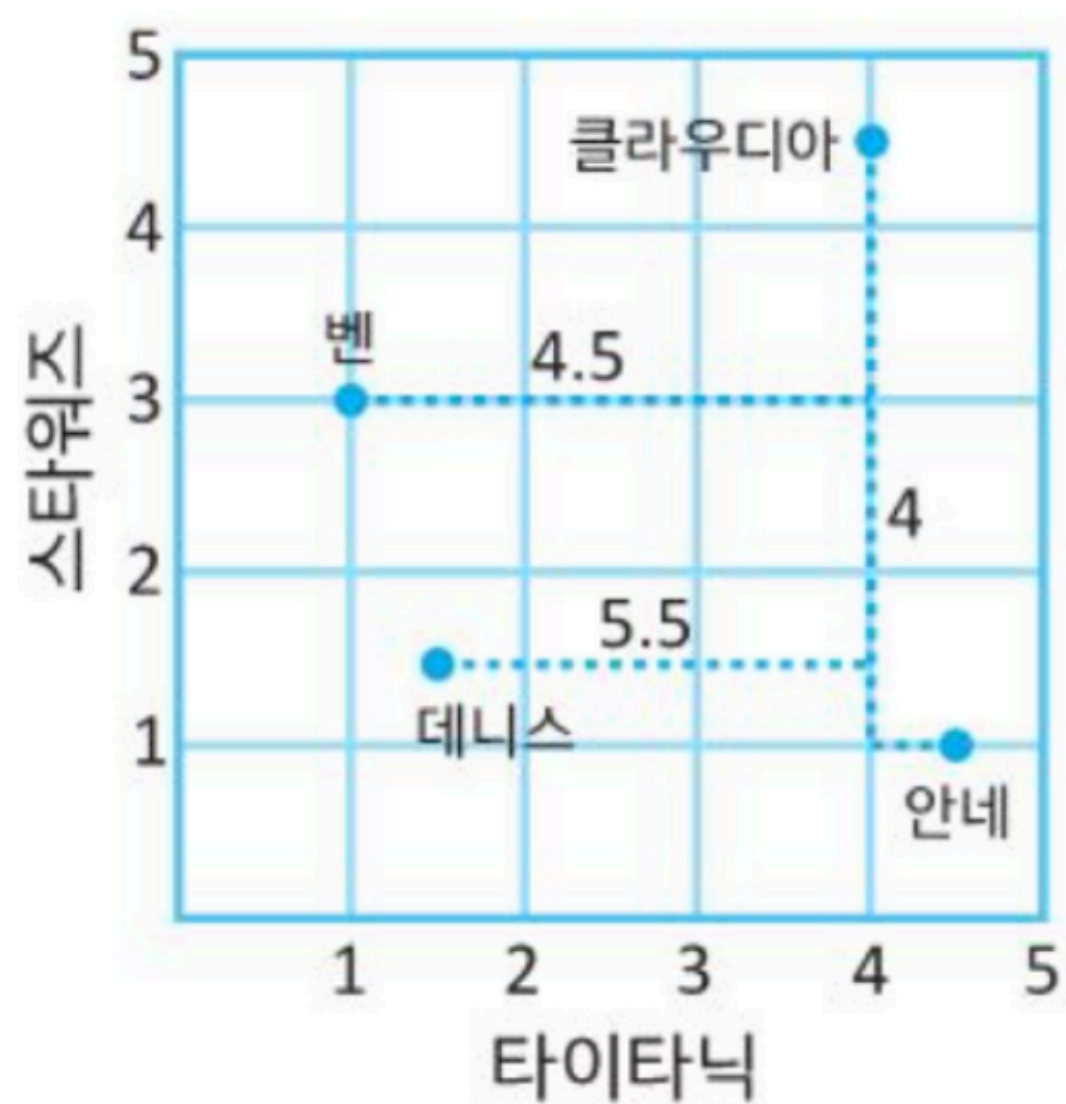
Memory-Based Approach

유사도 측정 방법

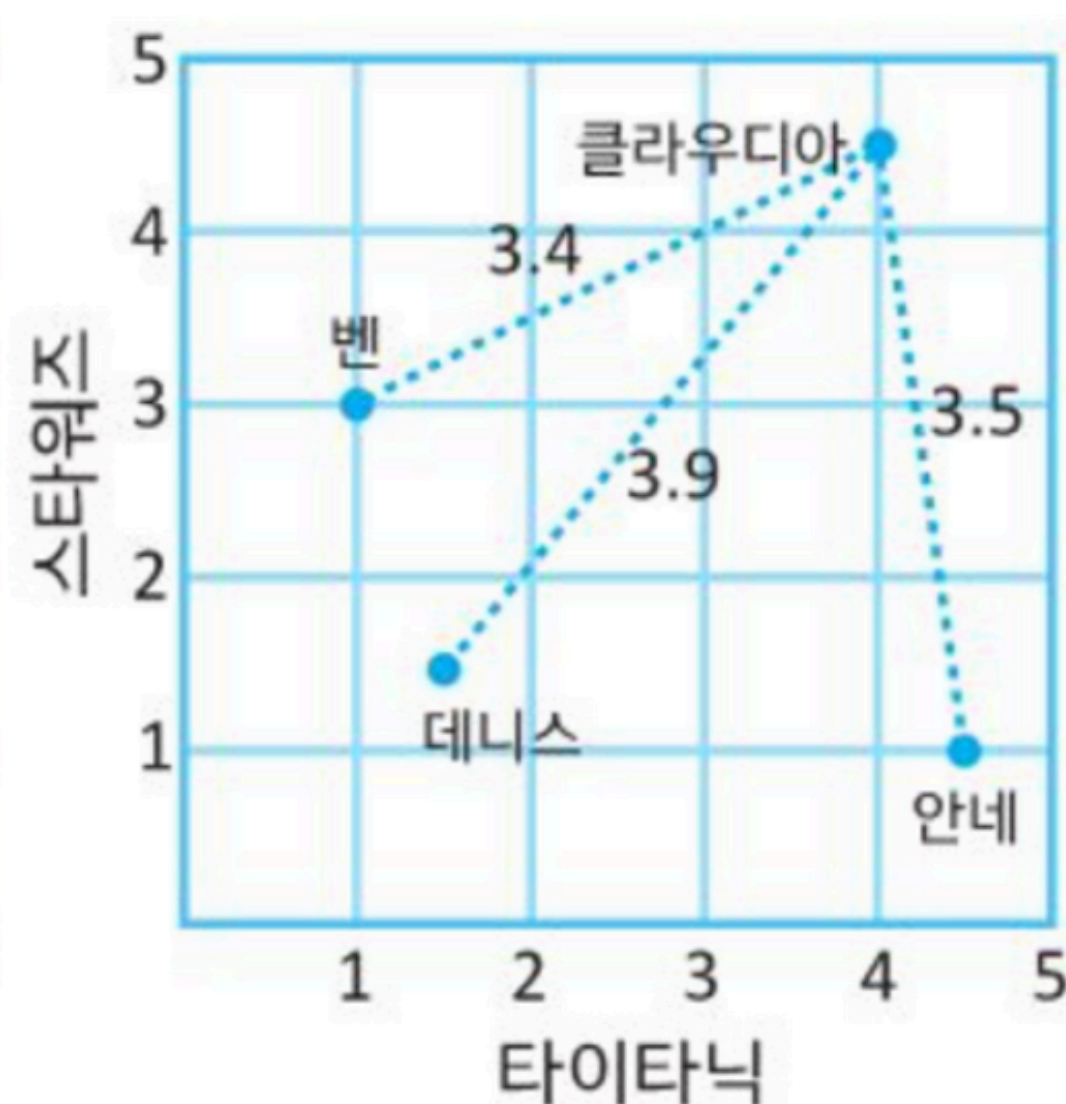
- 유사도(거리)를 측정하여 사용함
- 거리 측정 방법은 다양함
 - 사용자(행) - 아이템(열) 행렬을 사용함

Memory-Based Approach

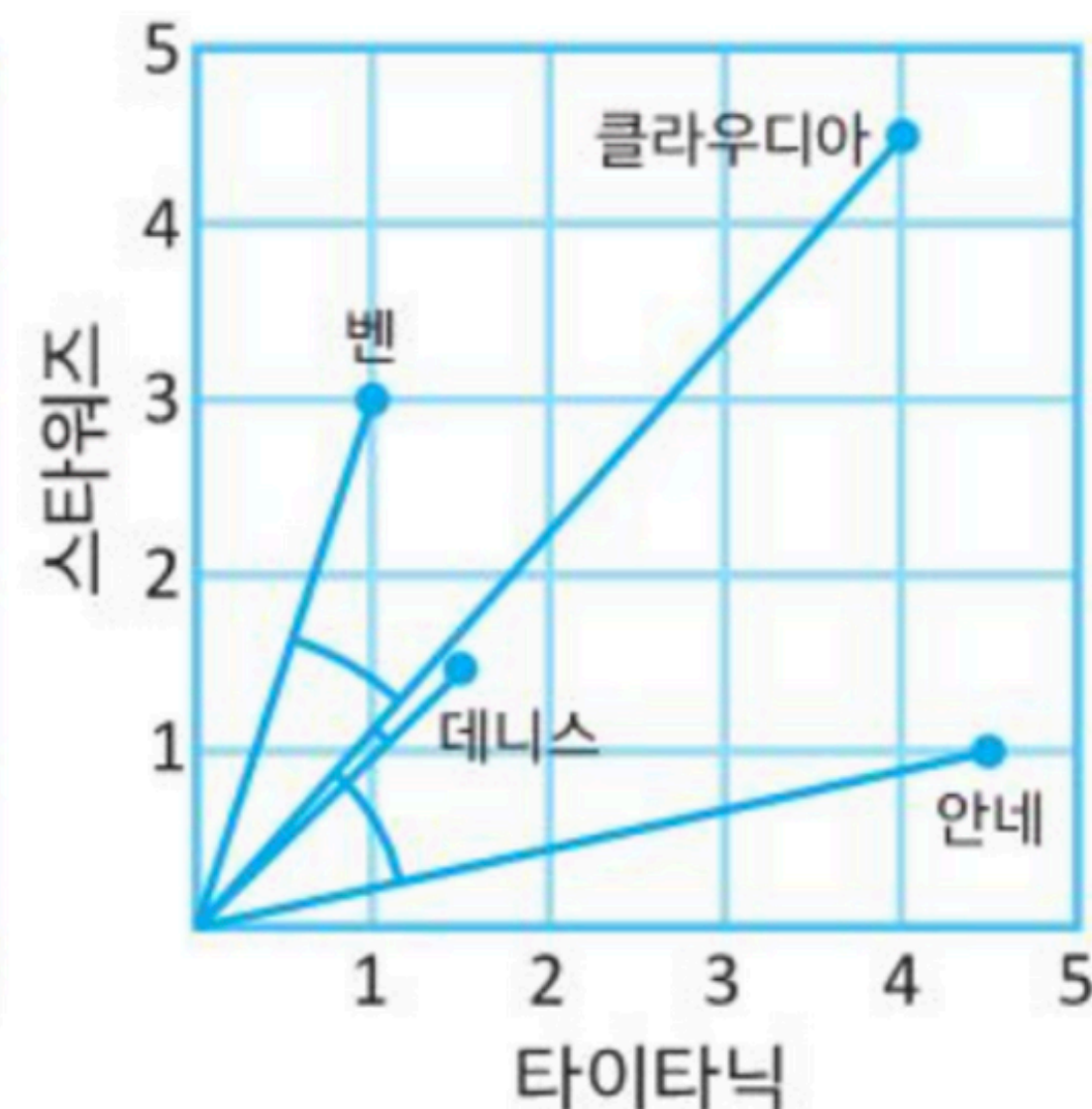
	타이타닉	스타워즈
안네	4.5	1
벤	1	3
클라우디아	4	4.5
데이스	1.5	1.5



<Manhattan Distance>



<Euclidean Distance>



<Cosine Similarity>

Memory-Based Approach

- Cosine Similarity

$$sim(u, u') = cos(\theta) = \frac{R_u \cdot R_{u'}}{\|R_u\| \|R_{u'}\|} = \frac{\sum_{i=1}^n R_{ui} \times R_{u'i}}{\sqrt{\sum_{i=1}^n (R_{ui})^2} \times \sqrt{\sum_{i=1}^n (R_{u'i})^2}}$$

- 사용자 u 와 사용자 u' 가 같은 방향성을 보고 있는지를 평가함
- 일반적으로 사용되는 유사도

- Pearson Correlation Similarity

$$sim(u, u') = \frac{\sum_{i=1}^n (R_{ui} - \bar{R}_u)(R_{u'i} - \bar{R}_{u'})}{\sqrt{\sum_{i=1}^n (R_{ui} - \bar{R}_u)^2} \times \sqrt{\sum_{i=1}^n (R_{u'i} - \bar{R}_{u'})^2}}$$

- 평균적인 경향성에서 얼마나 차이가 나는지를 기반으로 함
- Centered Cosine Similarity 라고도 함
- 평균을 빼므로, 사용자 u 와 사용자 u' 가 함께 사용한 아이템이 꽤나 있어야 함