

비정형 빅데이터 분석의 응용과 실습

Week-09. Recommender System

서중원 2020.11.14

추천시스템

내가 이걸 좋아한다면?

- 추천 시스템은 정보 필터링 기술의 일종으로,
특정 사용자가 관심을 가질만한 정보를 추천하는 것
 - 예: 온라인 쇼핑몰 상품 추천, 넷플릭스 콘텐츠 추천
- How?
 - User Based Collaborative Filtering
 - Item Based Collaborative Filtering
 - Content based filtering

User Based Collaborative Filtering

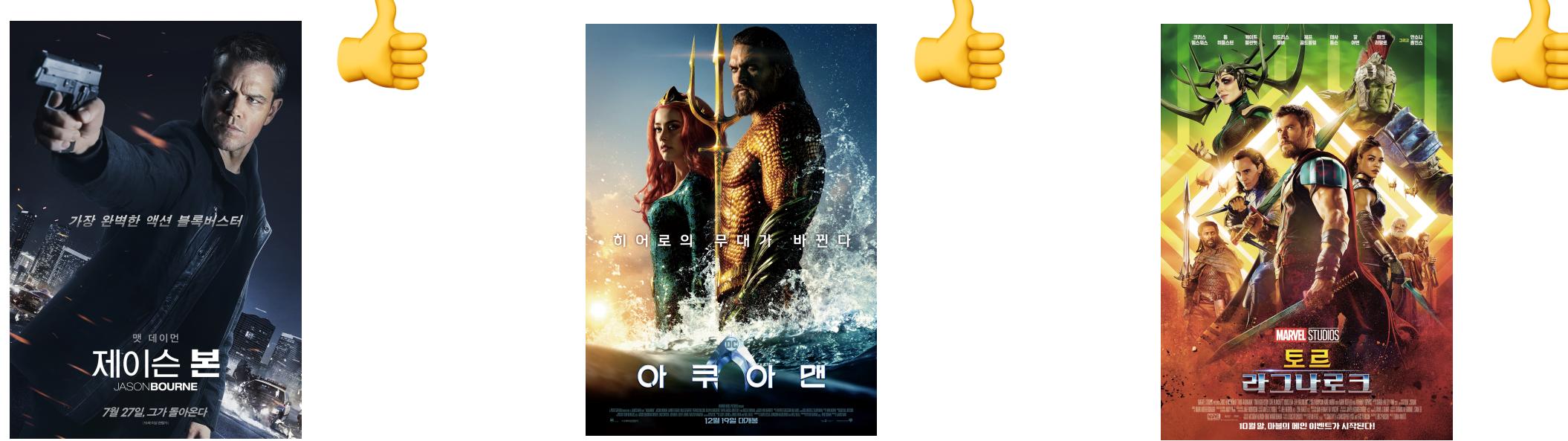
사용자의 행동을 기반으로 쌓은 데이터를 활용

- 소셜 미디어의 “좋아요” 기능
- 쇼핑몰의 장바구니 기능
- 구글 광고 클릭 기능
- 이러한 데이터를 기반으로 사용자들을 Grouping 가능

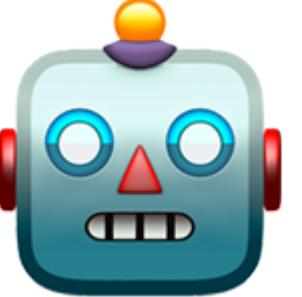
User Based Collaborative Filtering

Example: 영화 추천

User1



User2



User3



User Based Collaborative Filtering

Example: User1과 User3는 비슷한 취향이다!



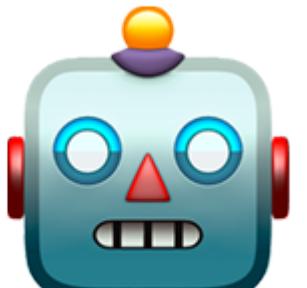
User Based Collaborative Filtering

비슷하다는 것을 어떻게 계산할 수 있을까?

	제이슨본	아쿠아맨	토르	스파이더맨	건축학개론	암살
	1	1	1	0	0	0
	0	0	0	1	1	1
	1	1	0	0	0	0

User Based Collaborative Filtering

한 유저들을 각각의 Vector로 보고 영화는 축으로 => Cosine Similarity

	제이슨본	아쿠아맨	도르	스파이더맨	건축학개론	암살	유사도
	1	1	1	0	0	0	0.81
	0	0	0	1	1	1	0
	1	1	0	0	0	0	1.0

Cosine Similarity

$$\text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}}$$

Item Based Collaborative Filtering

비슷한 영화를 추천해주는 건 어떨까?

- 영화가 비슷하다는 것은?
- Collaborative Filtering 기반에서는, 사용자의 액션에 의해 영화를 Grouping
- 예를들어
 - User1은 어벤져스를 보고, 아바타를 본상태
 - User2는 어벤져스만 본 상태
 - 어벤져스와 아바타는 비슷한 부류의 영화군!

Item Based Collaborative Filtering

비슷하다는 것을 어떻게 계산할 수 있을까? Transpose

	제이슨본	아쿠아맨	토르	스파이더맨	건축학개론	암살
	1	1	1	0	0	0
	0	0	0	1	1	1
	1	1	0	0	0	0
유사도	0.5	0.5	1	0	0	0

이전에 아무도 선택하지 않은 영화는
어떻게 추천 할 수 있을까?

Content Based Filtering

“진짜” 비슷한 영화를 추천해주는 건 어떨까?

- 영화가 비슷하다는 것은?
- Content Based Filtering 기반에서는, 영화의 Feature들을 기반으로 유사도를 계산
- 예를들어, 감독, 장르, 주연배우, 언어 등
 - 더나아가 자연어 처리를 활용한다면 줄거리
 - 마찬가지로 유사도는 코사인 유사도

오늘의 실습

localhost

YONFLIX

Home Link Disabled

로그아웃

◀ 싫어요

▶ 좋아요



닥터 스트레인지

불의의 사고로 절망에 빠진 천재 외과의사 '닥터 스트레인지(베네딕트 컴버배치)'. 마지막 희망을 걸고 찾아 간 곳에서 '에인션트 원(틸다 스윈튼)'을 만나 세상을 구원할 강력한 능력을 얻게 되면서, 모든 것을 초월한 최강의 히어로로 거듭나는데...

추천영화

영화 추천 시스템

내가 좋아하는 영화를 고르면, 비슷한 영화를 추천해줘!

- 우선 사용자가 처음 보여지는 10개의 영화중 본인이 좋아하는 영화를 선택
- 좋아한 영화들을
 - User Based Collaborative Filtering
 - 또는 Content Based Filtering
- 이번 실습에서는 줄거리를 이용한 TF-IDF를 활용한 Content Based Filtering

E.O.D