

추천 시스템

2022.04



1. 개요

❖ 온라인 스토어의 필수 요소

너무 많은 상품으로 가득찬 온라인 스토어



한정된 시간, 어떤 상품을 골라야 할지 선택의 압박



VS

추천엔진은 사용자가 무엇을
원하는지 빠르게 찾아내어
사용자의 온라인 쇼핑 이용
즐거움을 배가시킨다.

〈 추천 시스템으로 사용자의 선택 부담을 해결 〉

1. 개요

❖ 온라인 스토어의 필수 요소



1. 개요

❖ 추천 시스템의 유형

- 콘텐츠 기반 필터링(Contents-based filtering)
- 협업 필터링(Collaborative filtering)
 - 최근접 이웃(Nearest Neighbors) 협업 필터링
 - 잠재 요인(Latent Factor) 협업 필터링
 - ← 행렬 분해(Matrix Factorization) 기법(Netflix 추천 시스템 경연대회)

2. 콘텐츠 기반 필터링

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



장르: SF, 드라마, 미스터리
감독: 드니 빌뇌브
출연: 에이미 아담스, 제레미 러너
키워드: 외계인 침공, 예술성, 스릴러 요소



장르: SF, 액션, 스릴러
감독: 드니 빌뇌브
출연: 라이언 고슬링, 해리슨 포드
키워드: 리들리 스콧 감독의 전작을 리메이크



장르: SF, 액션, 스릴러
감독: 리들리 스콧
출연: 노미 라마스, 마이클 패스벤더
키워드: 에일리언 프리퀼, 액션과 스릴러의 조화



추천



사용자 선호 프로필

선호 장르: SF, 액션, 스릴러
선호 배우: 에이미 아담스, 마이클 패스벤더 등
선호 감독: 리들리 스콧, 드니 빌뇌브

2. 콘텐츠 기반 필터링

- 콘텐츠 기반 필터링은 이전의 행동과 명시적 피드백을 통해 좋아하는 것과 유사한 항목을 추천
 - ex) 내가 지금까지 시청한 영화 목록과 다른 사용자의 시청 목록을 비교해 나와 비슷한 취향의 사용자가 시청한 영화를 추천
- 유사도를 기반으로 추천(코사인 유사도, 유클리드 거리)
- 장점
 - 많은 수의 사용자를 대상으로 쉽게 확장 가능
 - 사용자가 관심을 갖지 않던 상품 추천 가능
- 단점
 - 입력 특성을 직접 설계해야 하기 때문에 많은 도메인 지식이 필요
 - 사용자의 기존 관심사항을 기반으로만 추천 가능

3. 협업 필터링

❖ 개요

- 사용자와 항목의 유사성을 동시에 고려해 추천
- 기존에 내 관심사가 아닌 항목이라도 추천 가능
- 자동으로 임베딩 학습 가능
- 종류
 - KNN(K-Nearest Neighbor), SVD(Singular Value Decomposition), NMF(Non-negative Matrix Factorization), SVD++
- 장점
 - 자동으로 임베딩을 학습하기 때문에 도메인 지식이 필요 없다.
 - 기존의 관심사가 아니더라도 추천 가능
- 단점
 - 학습 과정에 나오지 않은 항목은 임베딩을 만들 수 없음
 - 추가 특성을 사용하기 어려움

3. 협업 필터링

❖ 최근접 이웃(KNN: K-Nearest Neighbor) 협업 필터링

사용자가 평가하지 않은 아이템을 평가한
아이템에 기반하여 예측 평가하는 알고리즘

	Item 1	Item 2	Item 3	Item 4
User 1	3		3	✓
User 2	4	2		3
User 3		1	2	2

로우 레벨 형태의 사용자 - 아이템 평점 데이터

User ID	Item ID	Rating
User 1	Item 1	3
User 1	Item 3	3
User 2	Item 1	4
User 2	Item 2	1
User 3	Item 4	5

변환

사용자 로우, 아이템 칼럼으로 구성된
사용자 -아이템 평점 데이터

	Item 1	Item 2	Item 3	Item 4
User 1	3		3	
User 2	4	1		
User 3				5

- 사용자 기반(User-User): 당신과 비슷한 고객들이 다음 상품도 구매했습니다.
- 아이템 기반(Item-Item): 이 상품을 선택한 다른 고객들은 다음 상품도 구매했습니다.

3. 협업 필터링

❖ 최근접 이웃(KNN: K-Nearest Neighbor) 협업 필터링

		다크 나이트	인터스텔라	엣지 오브 투모로우	프로메테우스	스타워즈 라스트제다이
상호간 유사도 높음	사용자 A	5	4	4		
	사용자 B	5	3	4	5	3
	사용자 C	4	3	3	2	5

사용자 A는 사용자 C 보다 사용자 B와 영화 평점 측면에서 유사도가 높음. 따라서 사용자 A에게는 사용자 B가 재미있게 본 '프로메테우스'를 추천

		사용자 A	사용자 B	사용자 C	사용자 D	사용자 E
상호간 유사도 높음	다크 나이트	5	4	5	5	5
	프로메테우스	5	4	4		5
	스타워즈 라스트제다이	4	3	3		4

여러 사용자들의 평점을 기준으로 볼 때 '다크 나이트'와 가장 유사한 영화는 '프로메테우스'

3. 협업 필터링

❖ 잠재 요인 협업 필터링

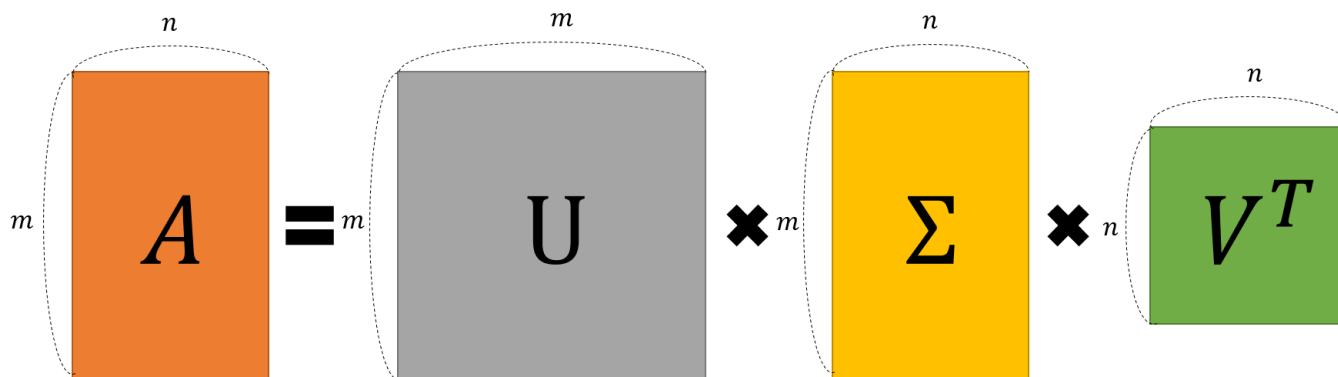


〈 백만 달러의 상금이 걸린 넷플릭스 추천 엔진 경연 대회 우승팀 사진 〉

3. 협업 필터링

❖ 잠재 요인

- 특이값 분해(SVD: Singular Value Decomposition)



- 대체 방법: SGD(Stochastic Gradient Descent)

	아이템1	아이템2	아이템3	아이템4	아이템5		요인1	요인2	요인3		아이템1	아이템2	아이템3	아이템4	아이템5		
사용자1	4.0			2.0		=	사용자1	0.96	0.47	-0.76	*	요인1	1.62	-0.79	1.04	1.07	1.43
사용자2		5.0		3.0	1.0		사용자2	-0.03	0.84	-2.47		요인2	1.51	0.45	-0.06	0.12	-0.21
사용자3			3.0	4.0	4.0		사용자3	2.38	0.11	-1.20		요인3	-2.22	-1.85	0.43	1.18	-0.50
사용자4	5.0	2.0	1.0	2.0			사용자4	0.59	1.10	-1.06							

3. 협업 필터링

❖ 잠재 요인

- 행렬 분해를 통한 잠재요인 협업 필터링

