[9장: 추천 시스템]

1. 추천 시스템의 개요와 배경
2. 추천 시스템의 개요

추천 시스템 도입 후 큰 폭의 매출 증가가 발생

추천 시스템을 통해 자신이 좋아하는지 몰랐던 취향을 시스템이 발견하고 그에 맞는 콘텐츠를 추천하면 사용자는 해당 사이트를 더욱 강하게 신뢰하게 되어 더 많은 추천 콘텐츠를 선택하게 됨

결론적으로 더 많은 데이터가 축적되며 더욱 정확하고 다양한 결과를 얻을 수 있ㄴㄴ 선순환 시스템 구축 가능

따라서 추천 시스템의 고도화에 많은 비용 투자하게 됨

1. 온라인 스토어의 필수 요소, 추천 시스템

너무 다양한 상품 이미지와 번잡한 카테고리들은 온라인 쇼핑에 대한 제품 선택의 어려움 등 부정적인 이미지가 강해져 매출 감소로 이어질 수 있음

시간은 한정적이므로 너무 다양한 선택지는 오히려 사용자에게 어떤 상품을 골라야 할지에 대한 압박감을 느끼게 함 🡪 추천 시스템이 해결

온라인 스토어에서 필수 요소가 됨

온라인 스토어는 많은 고객과 상품 관련 데이터를 가지고 있음

데이터를 통해 사용자가 흥미 가질만한 상품 즉각적으로 추천

ex) 사용자가 어떤 상품을 구매했는지, 장바구니에 넣은 상품이나 둘러본 상품, 무엇을 클릭했는지 등등

도출된 데이터를 기반으로 우리가 흔히 아는 ‘이 상품을 선택한 다른 사람들이 좋아하는 상품들’ 과 같은 친숙한 문구로 구매를 유도함

텍스트, 스크린샷이(가) 표시된 사진

자동 생성된 설명

1. 추천 시스템의 유형

크게 콘텐츠 기반 필터링(Content based filtering), 협업 필터링(Collaborative Filtering) 방식으로 나뉨

협업 필터링은 최근접 이웃(Nearest Neighbor) 협업 필터링과 잠재 요인(Latent Factor) 협업 필터링으로 나뉨

초창기: 콘텐츠 기반 필터링, 최근접 이웃 기반 협업 필터링 주로 이용

이후: 잠재 요인 협업 필터링 기반 추천 시스템 적용 (행렬 분해 기법 이용) but 서비스하는 아이템 특성 따라 초창기 그대로 유지하는 사이트도 많음 ex) 아마존 – 아이템 기반의 최근접 이웃 협업 필터링 방식을 추천 엔진으로 이용

최근: 개인화 특성 강화하기 위해 하이브리드 형식으로 콘텐츠 기반과 협업 기반을 적절히 결합해 사용하는 경우도 증가

1. 콘텐츠 기반 필터링 추천 시스템

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

1. 최근접 이웃 협업 필터링 = 메모리 협업 필터링 🡪 거진 아이템 기반

축적된 사용자 행동 양식(User Behavior)만을 기반으로 사용자가 아직 평가하지 않은 아이템을 예측 평가하여 추천하는 것을 협업 필터링이라고 부름

텍스트, 도표, 번호, 폰트이(가) 표시된 사진

자동 생성된 설명

최근접 이웃 방식과 잠재 요인 방식으로 나뉨 🡪 둘 다 사용자-아이템 평점 행렬 데이터에만 의지해 추천 수행함

행: 개별 사용자

열: 개별 아이템

행렬 해당 위치: 평점

다차원 행렬, 희소 행렬 특성

사용자 기반: 당신과 비슷한 고객들이 다음 상품도 구매했습니다

아이템 기반: 이 상품을 선택한 다른 고객들은 다음 상품도 구매했습니다

* 행이랑 열이 반대

최근 특정 사용자와 타 사용자 간의 유사도를 측정해준 뒤 가장 유사도가 높은 TOP-N 사용자를 추출해 그들이 선호하는 아이템을 추천, 코사인 유사도 use

1. 잠재 요인 협업 필터링

사용자-아이템 평점 매트릭스 속에 숨어 있는 잠재 요인을 추출해 추천 예측을 할 수 있게 함

행렬 분해 use

잠재요인은 각자의 목적에 맞게 고정해서 사용하면 됨

1. 행렬 분해의 이해

SVD, NMF가 대표적

factorization(인수분해)



SVD로 분해 불가능할 경우 확률적 경사 하강법이나 ALS 방식 이용하여 SVD 수행

1. 확률적 경사 하강법을 이용한 행렬 분해

회귀의 경사 하강법의 한 종류

P와 Q 행렬로 계산된 예측 R 행렬 값이 실제 R 행렬 값과 가장 최소의 오류를 가질 수 있도록 반복적인 비용 함수 최적화를 통해 P와 Q를 유추해내는 것

폰트, 타이포그래피, 텍스트, 화이트이(가) 표시된 사진

자동 생성된 설명

친필, 폰트, 서예, 텍스트이(가) 표시된 사진

자동 생성된 설명

1. 콘텐츠 기반 필터링 실습 – TMDB 5000 영화 데이터 세트
2. 장르 속성을 이용한 영화 콘텐츠 기반 필터링

칼럼 값의 유사도를 비교한 후 그중 높은 평점을 가지는 컨텐츠 추천

1. 데이터 로딩 및 가공
2. 장르 콘텐츠 유사도 측정

genres를 문자열로 변경한 후 이를 CountVectorizer로 피처 벡터화한 행렬 데이터 값을 코사인 유사도로 비교하는 것

1. 장르 콘텐츠 필터링을 이용한 영화 추천
2. 아이템 기반 최근접 이웃 협업 필터링 실습

데이터 가공 및 변환 🡪 영화 간 유사도 산출 🡪 아이템 기반 최근접 이웃 협업 필터링으로 개인화된 영화 추천

1. 행렬 분해를 이용한 잠재 요인 협업 필터링 실습
2. 파이썬 추천 시스템 패키지 – Surprise
3. Surprise 패키지 소개

파이썬 기반에서 사이킷런과 유사한 API와 프레임워크를 제공

텍스트, 폰트, 영수증, 화이트이(가) 표시된 사진

자동 생성된 설명

1. Surprise를 이용한 추천 시스템 구축
2. Surprise 주요 모듈 소개
3. Dataset

데이터 세트 칼럼 순서: 사용자 아이디, 아이템 아이디, 평점

텍스트, 스크린샷, 영수증, 폰트이(가) 표시된 사진

자동 생성된 설명

텍스트, 스크린샷, 폰트, 번호이(가) 표시된 사진

자동 생성된 설명

1. OS 파일 데이터를 Surprise 데이터 세트로 로딩

* 실습 참고

1. Surprise 추천 알고리즘 클래스

텍스트, 영수증, 폰트, 스크린샷이(가) 표시된 사진

자동 생성된 설명

텍스트, 폰트, 화이트, 대수학이(가) 표시된 사진

자동 생성된 설명

텍스트, 영수증, 폰트, 스크린샷이(가) 표시된 사진

자동 생성된 설명

1. 베이스라인 평점

베이스라인 평점: 개인의 성향을 반영하여 아이템 평가에 편향성 요소를 반영하여 평점을 부과하는 것

텍스트, 영수증, 폰트, 화이트이(가) 표시된 사진

자동 생성된 설명

1. 교차 검증과 하이퍼 파라미터 튜닝

사이킷런과 유사한 cross\_validate()와 GridSearchCV 클래스 제공

1. Surprise를 이용한 개인화 영화 추천 시스템 구축