

DS 11주차 수업 팀과제

추천 시스템

팀 AOA

경영학과 2014109050 박한솔
경영학과 2015126062 신지수

01

평점 예측 알고리즘의 평점 측정 방법

Rapidminer의 performance operator에서는 RMSE, MAE, NMAE를 성능 측정 척도로 사용하였다.

얼마나 사용자의 선호도를 잘 파악해서 평점값을 제대로 예측하는가, 추천된 아이템을 선택할 가능성이 높은지 파악할 수 있다.

$$RMSE = \sqrt{\frac{1}{N} \sum (\text{실제값} - \text{예측값})^2}$$

RMSE는 평균 제곱근 오차로서, RMSE 값이 작을수록 오차가 작다는 것을 의미하므로 성능이 좋다는 것을 의미한다.

$$MAE = \frac{1}{N} \sum |\text{실제값} - \text{예측값}|$$

MAE는 평균 절대 오차로서, 이상치로 의심되는 구역이 많을 경우 사용하는 지표이다.

NMAE

평균 절대 오차인 MAE를 정규화한 값을 의미한다.

02

일반적인 추천시스템의 성능 측정 방법

얼마나 사용자의 선호도를 잘 파악해서 평점값을 제대로 예측하는가, 추천된 아이템을 선택할 가능성이 높은가를 측정한다.

Error 측정 (예측 평점과 실제 평점간의 오차)

흔히 사용되는 지표로는 MAE, RMSE 가 있음.

예측값과 실제값 사이의 상관관계 측정

양 값 사이에 얼마나 유사성이 있는지 알아볼 수 있음

분류 결과에 따른 평가 척도 활용

평점이 좋은(좋다고 예측하는) 결과를 True로, 나쁜(나쁘다고 예측하는) 결과는 False로 설정하고, 예측 결과와 실제 결과를 토대로 2 by 2 matrix를 생성할 수 있다. matrix에서 precision과 recall을 계산하여 평가 척도로 활용할 수 있다.

03

사용자 만족도

추천 시스템은 사용자들과 접점이 매우 많은 실용적인 시스템이기 때문에 단순히 사용자의 선호도를 예측하는 것 이상으로, 사용자들이 실제 추천된 아이템의 만족 여부가 중요한 척도일 것이다.

사용자 만족도 측정 방법은 다음과 같다.

1. CTR : 추천된 상품이 조회 / 구매되는 비율

CTR이 높다면 사용자가 추천된 아이템에 흥미를 느꼈다고 생각할 수 있다.

2. Duration/Retention time : 추천된 아이템을 즐기는 시간과 재방문율을 의미한다.