

장르별 선호도 기반 협업 필터링 방식 개발

김도영

성균관대학교 소프트웨어대학

Genre Preference based Collaborative Filtering

Kim Do Young

Sungkyunkwan University

지도 교수: 김유성 교수님

연구실명: 컴퓨터 네트워크 연구실

개요

최근 전자상거래에서 대부분의 개인화 된 추천 시스템들은 사용자의 취향에 맞는 상품을 추천해 매출을 증대시키기 위해 협업 필터링(Collaborative Filtering) 기술을 적용하고 있다. 사용자 기반 협업 필터링은 특정 사용자의 선호도와 가장 유사한 선호도를 가지는 사용자 그룹의 선호도를 바탕으로 그 사용자의 특정 상품에 대한 선호도를 예측하는 기법이다.

기존의 아이템 기반 협업 필터링 방식은 특정 분야에 대한 사용자의 개인적인 선호도를 반영하지 않으며, 평가를 내린 항목 전체의 평균을 이용해 사용자의 선호도를 예측해 세부적인 예측은 불가능하다는 문제가 있다.

본 작품에서는 유저를 선호도에 따라 군집화 한 뒤 유저의 선호도를 반영한 평가값을 예측에 반영해 기존의 협업 필터링 방식을 개선했다.

시스템 구성

영화 아이템에 대한 총 100,000개의 평가치로 구성되어 있다. 각 사용자의 평가치는 최소 20편 이상의 영화 아이템에 대해 존재하며, 평가치의 범위는 1점부터 5점 사이이다. 각 영화는 최소 1개부터 5개의 장르에 속해 있으며 장르의 구성은 인터넷 영화 데이터베이스 사이트인 IMDb (Internet Movie Database) 의 기준에 따라 18개로 구분되어 있다.

유저의 선호도를 평가에 반영하기 위해 피어슨 상관계수 계산 방식을

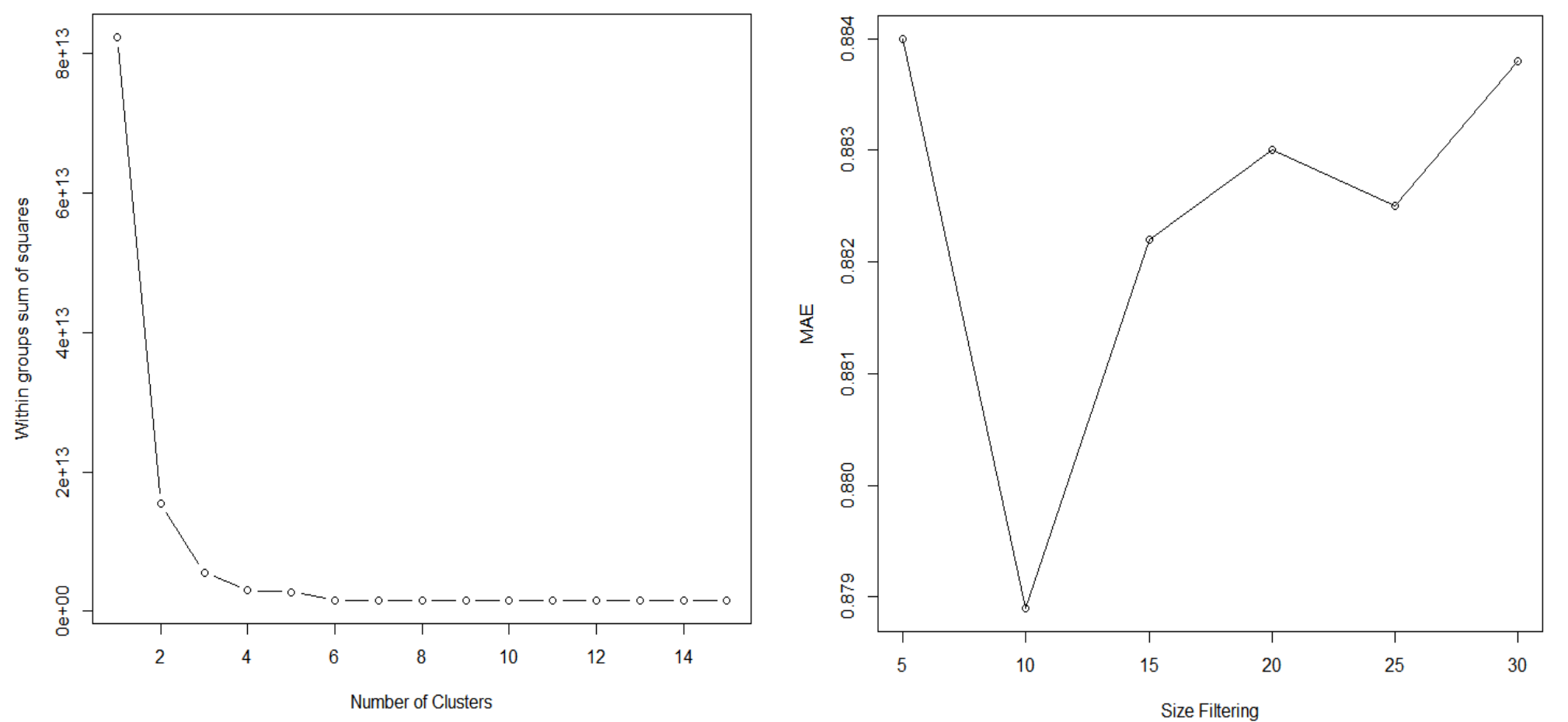
$$P_{a,u} = \frac{\sum_{i=1}^m (r_{a,i} - r'_{a,g}) * (r_{u,i} - r'_{u,g})}{\sqrt{\sum_{i=1}^m (r_{a,i} - r'_{a,g})^2 * \sum_{i=1}^m (r_{u,i} - r'_{u,g})^2}}$$

와 같이 변경하였고, 평가값 계산에도 유저의 특정 장르에 대한 평가치의 평균을 대입해 사용하였다.

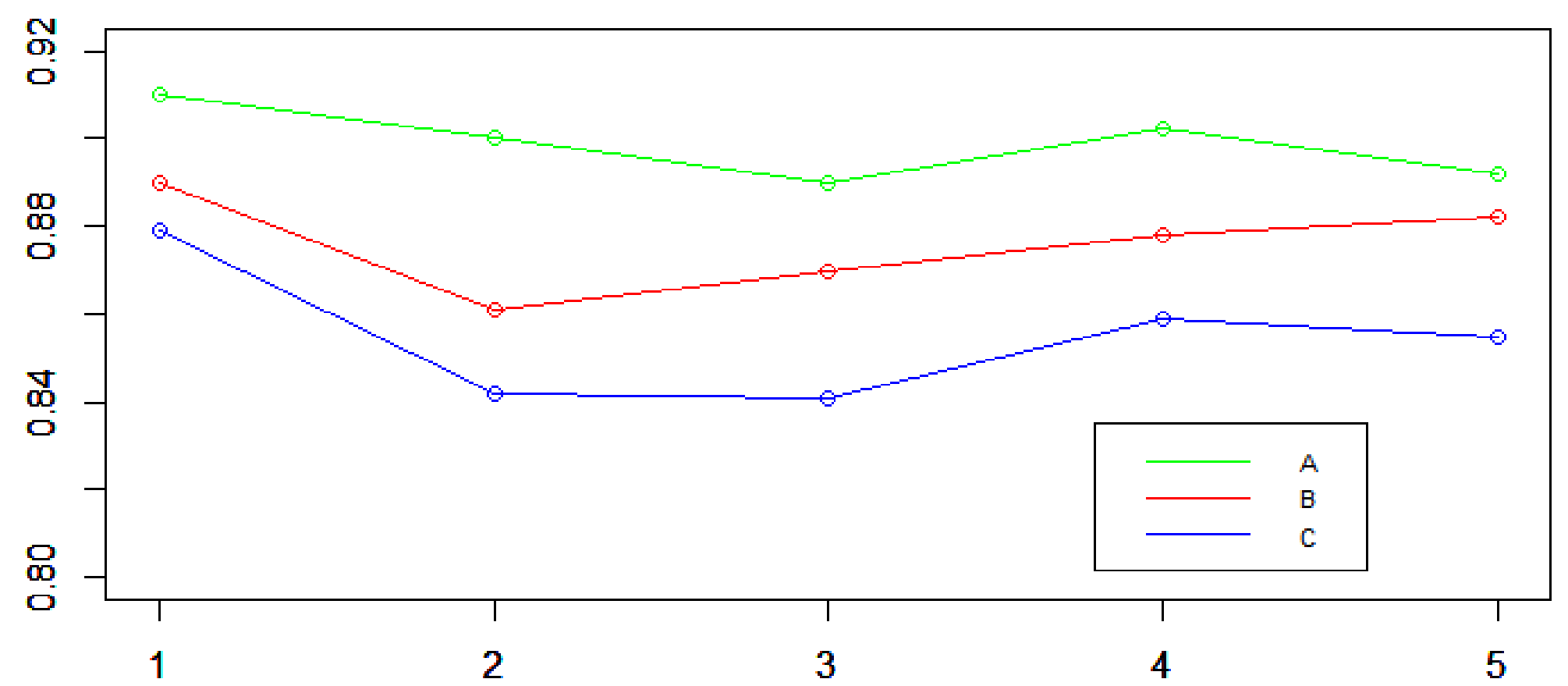
특정 장르에 대한 평가의 수가 충분하지 않을 경우 평가치 평균에 대한 신뢰도가 낮아지므로 특정 장르에 대한 평가치의 수가 최소 임계값 이하인 경우 특정 장르에 대한 평가치 평균에 대한 조정이 필요하다. 이 경우에 대비하여 Size Filtering값을 적용하여 장르에 대한 평가값이 Size Filtering 이상/이하인 경우와 평가치가 0인 경우로 분류해 장르별 평가치 평균을 계산하였다.

Size Filtering 이상	해당 장르의 평점평균 사용
Size Filtering 이하	(해당 장르의 평점평균+ 사용자 전체 평점평균)/2
0인 경우	전체 평점평균 사용 (기존의 방식과 동일)

결과



위 두 그래프는 데이터의 Sum of Squares 분석 그래프와 Size Filtering에 따른 MAE의 변화를 나타낸다. Sum of Squares 분석에 의해 K-means clustering을 위한 최적의 K값은 3임을 알 수 있고, MAE가 가장 낮은 10이 Size Filtering값으로 적절함을 확인할 수 있다. 따라서 Size Filtering값을 10으로 고정하여 기존의 협업 필터링과의 비교를 진행하였다.



위 그래프는 5번의 test를 통해 전통적인 방식과 타 논문에서 개선한 협업 필터링 방식, 새롭게 개발한 협업 필터링 방식의 MAE를 비교한 것이다. 그래프에서 볼 수 있듯이, 모든 테스트에서 기존의 방식보다 MAE값이 낮아졌고, 동시에 기존에 존재하던 방식들에 비해 성능이 개선되었음을 확인할 수 있다.

결론

MAE 평가를 통해 사용자의 장르별 선호도를 고려한 방식이 기존의 협업 필터링 방식에 비해 개선되었음을 확인할 수 있었다. 특정 장르에 대한 유저의 기존 평가가 없는 경우 성능이 전통적인 협업 필터링 방식 수준으로 낮아지지만, 유저가 10개 이상의 평가를 내린 항목들의 경우 기존 방식에 비해 유의미한 개선이 존재한다는 점에서 효과적인 개선 방식이라고 생각한다.