

개인화 영화 추천 시스템 성능 평가와 개선에 관한 연구

김세준*, 정운해*, 박두순**

*순천향대학교 컴퓨터학과, **순천향대학교 컴퓨터소프트웨어공학과

e-mail : kalkal47@hotmail.com

A Study on the Performance Evaluation and Improvement of Personalized Movie Recommendation System

Se-jun Kim*, Woon-hae Jeong*, Doo-soon Park**

*Dept of Computer Science, Sooncheonhyang Univ.

**Dept of Computer Software Engineering, Sooncheonhyang Univ.

요 약

협업필터링은 추천 시스템 중에서 가장 일반적으로 사용되는 추천 시스템이다. 영화 추천 시스템에서도 이 방법을 가장 많이 사용한다. 추천 시스템에서 가장 많이 사용되고 있는 방법이지만 이 기법만을 적용할 경우 희박성, 확장성 그리고 투명성 등의 문제점을 가진다. 이러한 문제점들을 개선해 보려는 노력들이 많이 연구되어 왔다. 본 논문에서는 개인들의 특징인 개인 성향과 협업 필터링을 기반으로 한 영화 추천 시스템을 제시하고 기존의 영화추천 시스템과 성능 평가한다.

1. 서론

최근 데이터마이닝(Data mining) 추천 기법들이 많이 연구되고 이를 이용하여 여러 분야에서 사용되고 있다. 서비스 제공자의 입장에서 개인화 서비스를 제공하여 획일화된 서비스가 아닌 개개인의 개성과 취향을 존중하여 개인에게 초점을 두어 서비스를 제공함으로써 고객의 구매를 유도하는 시스템이 각광 받고 있다. 대표적인 예로는 Amazon이나 CD Now 등 쇼핑몰에서 많이 사용되고 있다.

데이터마이닝을 이용한 추천 시스템들 중 영화추천 시스템에 가장 일반적으로 사용되는 기법은 협업필터링 기법(Collaborative filtering)이다. 협업 필터링 기법은 많은 사용자들로부터 얻은 프로파일정보에 따라 목표 고객과 가장 비슷한 구매 선호도를 가진 고객을 찾고 그 고객의 구매 제품 중 목표고객이 구매할 가능성이 가장 큰 제품들을 자동적으로 추천해 주는 방법이다. 협업 필터링을 이용한 추천 시스템에서 사용자들은 항목들에 대해 선호도 평가 점수를 부여한다[1].

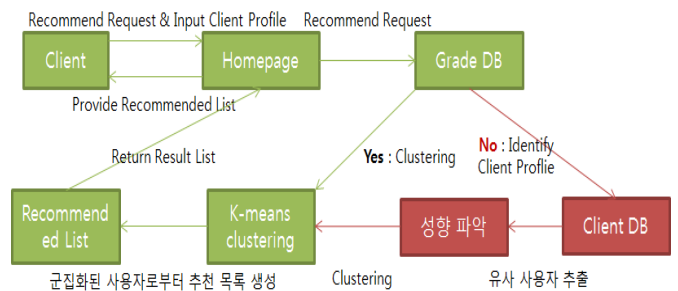
그렇지만 기존의 통상적인 협업필터링 기법은 목표 고객의 유사 선호고객을 결정하는데 있어서 목표고객이 평가한 정보가 없으면 추천을 하는데 정확도가 떨어지게 되어 희박성의 문제가 제기된다[2].

본 논문에서는 희박성 문제를 해결하기 위해 고객들이 가지고 있는 개인들의 특징인 개인 성향을 가지고 비슷한

그룹으로 분류한 뒤, 가장 비슷한 성향의 고객의 영화 평점을 찾아 영화를 추천할 수 있는 시스템을 제안한다. 또한 기존의 방법들 중 허재성[2], 이선영[3], 장슬기[4]가 제시한 방법들과 성능 평가한다.

2. 개인성향과 협업 필터링을 기반으로 한 추천절차

본 연구는 기존 협업필터링 기법에 개인의 성향 중 개인의 특징을 결정할 수 있는 선호장르, 성격, 나이를 가지고 고객과 가장 비슷한 성향의 고객을 추출하여 개인에게 가장 적합한 영화를 추천하기 위한 추천 시스템을 개발한다. 본 연구에서 제안하는 추천시스템은 (그림 1)과 같다.



(그림 1) 영화 추천 시스템 구성도

본 논문에서는 먼저 고객이 서버에 영화추천을 요구하면 평점이 기준치 이상이면 기존의 방법으로 최근접 이웃을 설정하여 영화를 추천한다. 하지만 평점이 기준치 이하이면 고객의 성향 중 선호장르, 성격, 나이 등을 고려하여 다른 사용자와 비교하여 비슷한 유사 성향의 사용자를 추출하여 영화를 추천한다. 여기서 고객들의 평가치를 이용하여 고객들 간의 유사도를 계산한 후 비슷한 성향을 가진 K-means clustering 기법으로 사용자들을 군집화 한다. 고객과 군집화 된 고객들이 본 영화 들 중 평점을 이용해 추천 목록을 구성한다.

3. 영화 추천 시스템의 구현

본 논문에서 구현된 회원 가입 창은 (그림 2)와 같다.

The screenshot shows a web form for user registration. It includes a '회원정보입력' (User Information Input) section with fields for ID (with a 'ID 중복검사' button), password, and email. Below this is a '추가입력' (Additional Input) section with dropdown menus for birth year (1967), month (1), day (3), gender (male), and age (select '성격을 선택해 주세요'). There are '확인' (Confirm) and '취소' (Cancel) buttons at the bottom right.

(그림 2) 회원 가입 창

(그림 2)에서는 사용자의 개인 성향을 파악하기 위해 선호장르, 성격, 나이를 추가하여 입력받아 저장한다.

(그림 3)은 개인 성향을 통해 영화 목록을 추천받아 보여주는 화면이다.

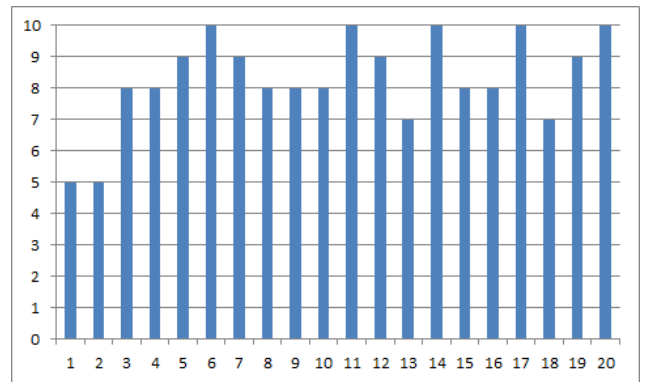
The screenshot displays a list of five recommended movies. Each entry includes a movie poster, the title, year, and a short synopsis. The movies listed are: 1. Casablanca (1942), 2. Close Shave, A (1995), 3. Wrong Trousers, The (1993), 4. Rear Window (1954), and 5. Wallace & Gromit: The Best of Aardman Animation (1996). Each entry has a '자세히 보기' (View details) link.

(그림 3) 개인 성향과 협업필터링 기반 영화 추천 결과

4. 영화 추천 시스템의 성능 평가

기존의 개인 성향과 협업 필터링 기반 영화추천 시스템 [2,3,4] 들은 장르, 성격, 나이, 성별, 혈액형, 지역 등 개인 성향을 가지고 영화 추천을 하였다. 일반적으로 개인 성향이 많으면 많을수록 개인화에 더 적합할 것이라는 판단 하에 혈액형부터 시작하여 나이, 성별, 장르, 성격 지역 등의 예측변수를 사용해왔다.

본 논문에서는 6개의 개인 성향 중에 개인화에 가장 영향을 주는 요인을 분석하여 그 요인을 가지고 구현하였다. 6개의 개인 성향을 가지고 취할 수 있는 모든 방법인 $2^6 = 64$ 개의 방법들을 실제 데이터인 MovieLense[5] 데이터를 가지고 영화 추천을 해서 가장 개인화에 영향을 주는 요인을 발견하였다. 64개의 방법 중 영향을 주는 20개의 방법에 대한 분석표는 (그림 4)와 같다.



(그림 4) 20개 방법에 대한 분석 그래프

각 방법을 설명하면 1번 방법은 혈액형, 2번 방법은 지역, 3번 방법은 성별·선호장르, 4번 방법은 선호장르·혈액형, 5번 방법은 성별·나이·지역, 6번 방법은 나이·선호장르·성격, 7번 방법은 성별·선호장르·성격, 8번 방법은 성별·혈액형·성격, 9번 방법은 성별·성격·지역, 10번 방법은 혈액형·성격·지역, 11번 방법은 나이·성별·선호장르·지역, 12번 방법은 나이·성별·선호장르·지역, 13번 방법은 나이·성별·성격·지역, 14번 방법은 나이·선호장르·성격·지역, 15번 방법은 성별·선호장르·성격·지역, 16번 방법은 성별·혈액형·성격·지역, 17번 방법은 나이·성별·혈액형·성격·지역, 18번 방법은 나이·성별·혈액형·성격·지역, 19번 방법은 나이·선호장르·혈액형·성격·지역, 20번 방법은 나이·성별·선호장르·혈액형·성격·지역 등을 예측 변수로 둔 결과이다.

단, 여기서 각 방법에서 추천된 10개의 영화에 대해서 20개의 방법 중에서 5번 이상 나타나면 개수를 썼다.

그림 4의 결과로부터 개인 성향에 가장 영향을 미치는 요인은 선호장르, 성격, 나이 라는 것을 찾아냈다.

<표 1>, <표 2>, 는 기존의 방법과 본 논문에서 제시한 방법을 가지고 영화 추천한 결과이다.

<표 1> 혈액형만 고려 나이,성별,선호장르,성격을 고려한 추천 목록

혈액형	나이,성별,선호장르,성격
'Prefontaine (1997)'	'Casablanca (1942)'
'Star Kid (1997)'	""Close Shave, A (1995)""
""Great Day in Harlem, A (1994)""	""Wrong Trousers, The (1993)""
'Pather Panchali (1955)'	'Rear Window (1954)'
'Some Mother's Son (1996)'	'Wallace & Gromit: The Best of Aardman Animation (1996)'
""Close Shave, A (1995)""	""Third Man, The (1949)""
'Everest (1998)'	""Manchurian Candidate, The (1962)""
""Wrong Trousers, The (1993)""	'Dr. Strangelove or: How I Learned to Stop Worrying and Love the Bomb (1963)'
'Casablanca (1942)'	'Citizen Kane (1941)'
'Wallace & Gromit: The Best of Aardman Animation (1996)'	'One Flew Over the Cuckoo's Nest (1975)'

<표 2> 나이,성별,선호장르,성격을 고려, 나이,선호장르,성격을 고려한 추천 목록

나이,성별,선호장르,혈액형,성격,지역	나이,선호장르,성격
'Casablanca (1942)'	'Casablanca (1942)'
""Close Shave, A (1995)""	""Close Shave, A (1995)""
""Wrong Trousers, The (1993)""	""Wrong Trousers, The (1993)""
'Rear Window (1954)'	'Rear Window (1954)'
'Wallace & Gromit: The Best of Aardman Animation (1996)'	'Wallace & Gromit: The Best of Aardman Animation (1996)'
""Third Man, The (1949)""	""Third Man, The (1949)""
""Manchurian Candidate, The (1962)""	""Manchurian Candidate, The (1962)""
'Dr. Strangelove or: How I Learned to Stop Worrying and Love the Bomb (1963)'	'Dr. Strangelove or: How I Learned to Stop Worrying and Love the Bomb (1963)'
'Citizen Kane (1941)'	'Citizen Kane (1941)'
'One Flew Over the Cuckoo's Nest (1975)'	'One Flew Over the Cuckoo's Nest (1975)'

<표 1>, <표2>의 결과로부터 개인 성향 중 혈액형 성별, 지역이 개인화에 많은 영향을 주지 못함을 알 수 있다.

결국은 6개의 개인 성향 중 많으면 많을수록 개인화에 맞지만 가장 많은 영향을 주는 것은 나이, 장르, 성격이라는 것을 제안한다.

5. 결론

본 논문에서는 개인화 추천을 위해 많은 예측 변수를 받아야 하는 기존 논문에서 개인의 불필요한 많은 정보를 입력받지 않더라도 그보다 적은 3가지 요소를 가지고도 충분한 성능 향상을 통해 개인에서 맞춤 추천을 할 수 있다.

본 시스템의 특징은 사용자가 웹 브라우저를 통해 세 가지의 정보를 입력함으로써 개인에게 맞는 영화 추천 시스템을 통해 영화를 추천 받을 수 있다는 점이다. 또한 사용자가 늘어나고 데이터가 쌓임으로 인해 더욱더 정확한 영화 정보를 추천하는 시스템으로 확장 될 수 있다.

참고문헌

- [1] 이형동, 김형주 “협업 필터링 추천시스템에서의 취향 공간을 위한 평가 예측 기법” 한국정보과학회, 정보과학회 논문지 : 데이터베이스 34(5), 2007.10, 389-395
- [2] 허재성, 박두순, 정영식 “개인 성향과 협업필터링을 이용한 맞춤형 영화추천 시스템”, 한국멀티미디어학회, 추계학술발표대회 논문집 제 12권 2호, 2009.11
- [3] 이선영, 박두순, 홍민, 이화민, 정영식 “혈액형과 협업 필터링을 이용한 맞춤형 영화추천 시스템”, 한국정보처리학회, 추계학술발표대회 논문집 제 16권 2호, 2009.11
- [4] 장슬기, 박두순, 정영식 “개인 성향과 협업필터링 기반 영화추천 시스템 성능 향상” 한국정보처리학회, 춘계학술발표대회 논문집 제 17권 1호, 2010.4
- [5] <http://www.grouplens.org/node/73>