

항목 간 선호도 차이와 사용자 평가 성향을 이용한 영화 추천 방법

김현경[†], 김현진, 박상현[‡]

연세대학교 컴퓨터과학과

[†]교신저자

E-mail: timetobloom@yonsei.ac.kr, {chriskim, sanghyun}@cs.yonsei.ac.kr

Movie Recommendation Method Using User Rating Tendency Based on Rating Difference Between Items

Hyunkyoung Kim[†], Hyunjin Kim, Sanghyun Park[‡]

Department of Computer Science, Yonsei University

요약

스마트기기의 대중화 및 모바일 서비스의 활성화로 다양하고 방대한 양의 멀티미디어 콘텐츠가 공급되면서 사용자들은 수많은 콘텐츠들 중에서 자신이 원하는 콘텐츠를 찾는 데에 많은 어려움을 겪고 있다. 이에 따라 사용자들이 원하는 콘텐츠를 찾을 수 있게 도와주는 추천시스템에 대한 연구가 다양하게 이루어지고 있다. 본 논문에서는 기존의 사용자 기반 추천 방법과 항목 기반 추천 방법의 문제를 보완한 항목 간 선호도 차이를 이용한 추천 방법을 기반으로 사용자의 평점 성향을 반영한 새로운 영화 추천 시스템을 제안한다. 제안하는 방법은 사용자의 평점 결정 성향을 반영하여 새로운 콘텐츠에 대한 사용자의 평가를 보다 정확하게 예측할 것으로 기대된다.

1. 서론

최근 시공간의 제약 없이 멀티미디어 콘텐츠를 재생할 수 있는 스마트기기가 대중화되면서 영상 콘텐츠에 대한 수요가 지속적으로 늘고 있다. 이에 따라 방대한 양의 영상 콘텐츠가 실시간으로 공급되고 있지만 사용자들은 수많은 콘텐츠들 중에서 자신이 원하는 콘텐츠를 찾는 데 어려움을 겪고 있다 [1].

많은 사용자들은 이전에 상영했던 영화 중 자신에게 적합한 영화를 찾기 위해 영화 정보를 제공해주는 TV 프로그램이나 포털 사이트의 지식정보 커뮤니티를 이용하고 있다. 그러나 이러한 방식은 사용자의 개인적인 취향을 완전히 반영할 수 없으며, 추천의 범위가 한정되어 있어 추천의 질이 떨어진다는 단점이 존재하기 때문에 사용자에게 적절한 영화 추천이 이루어지기 어려운 면이 있다.

위와 같은 문제를 해결하기 위해 Netflix¹, IMDb² 등은

자동화된 추천시스템을 제공해왔으며, 우리나라에서도 Watcha³등의 사이트가 등장하면서 영화 추천 시스템이 활성화되고 있다.

영화 추천 시스템에서 사용하는 방식은 크게 Collaborative Filtering과 Contents based recommendation으로 구분된다. Contents based recommendation은 추천의 정확도가 높다는 장점 때문에 Contents based recommendation 방식보다 널리 사용되고 있다.

위의 Contents based recommendation과 Collaborative Filtering 방식을 결합한 혼합 방식에 대한 연구도 이루어지고 있다. 논문 [2]에서는 사용자들의 선호도 정보를 근거로 항목 간 거리를 구하고, 이를 이용해서 새로운 항목에 대한 목표 사용자의 평가 점수를 예측하는 추천 방법을 제안

¹ <https://www.netflix.com>

² <https://www.imdb.com>

³ <https://watcha.net>

* 이 논문은 2015년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(NRF-2015R1A2A1A05001845)

하였다. 그러나 이 방법에서는 항목에 대한 평가 점수를 결정하는 사용자의 성향은 고려하지 않았다는 단점이 있다.

따라서 본 논문에서는 항목 간 선호도 차이를 이용한 영화 추천 방법 [2]에 사용자의 평점 결정 성향을 반영한 추천 방법을 제안한다.

2. 관련 연구

영화 추천 문제를 해결하기 위한 방법은 크게 Collaborative Filtering과 Contents based recommendation 두 가지로 나눌 수 있다.

먼저 Contents based recommendation은 사용자의 선호도가 높은 항목과 특징이 비슷한 새로운 항목을 추천한다 [3]. 이 방식은 데이터가 충분하지 않은 상황에서도 비교적 잘 작동한다는 장점이 있다. 이 방법은 항목의 본질적인 특성을 나타내는 특징들이 비슷하다면 사용자의 선호도 역시 유사할 것이라는 가정을 전제로 한다. 하지만 이를 사용한 항목간 유사도 지표는 사용자의 선호도를 전혀 반영하지 못한다는 단점이 있다.

Collaborative Filtering은 사용자들 간의 유사도를 조사하여 목표 사용자와 유사한 사용자들이 좋게 평가한 항목들을 추천하는 방식으로 [4], 추천의 정확도가 높다는 장점 때문에 Contents based recommendation 방식 보다 널리 사용되고 있다. 그러나 이 방식은 영화에 대한 사용자의 평가 점수를 기반으로 추천이 이루어지기 때문에 사용자의 평점 정보가 부족하면 사용자간 유사도 계산의 신뢰도가 떨어진다는 한계가 있다 [5].

위의 Contents based recommendation과 Collaborative Filtering 방식을 결합하여 두 방식의 장점을 취하는 혼합 방식에 대한 연구도 이루어지고 있다. 논문 [2]에서는 항목 간 선호도 차이를 이용한 영화 추천 방법을 제안하였다. 이 방법에서는 항목 간 선호도 차이를 계산하고, 이를 바탕으로 목표 사용자의 항목 선호도를 예측한다.

3. 제안하는 방법

본 논문에서는 기존 연구 [2]에서 소개한 방법에 더하여 사용자의 평점 결정 성향을 반영한 새로운 평점 예측 방법을 제안하고자 한다.

3.1. 항목 간 선호도 차이를 이용한 영화 추천 방법

기존 연구 [2]에서는 사용자들의 선호도 정보를 근거로 항목 간 거리를 구하고, 이를 이용해서 새로운 항목에 대한 목표 사용자의 평가 점수를 예측한다.

1) 항목 간 선호도의 차이 계산

이 방식에서는 먼저 사용자들이 항목에 대해 평가한 점수를 바탕으로 항목 간 선호도의 차이를 계산한다. 두 항목 i 와 j 에 대한 사용자들의 선호도 차이의 평균 $d_{i,j}$ 는 다음 식에 의해 구해진다.

$$d_{i,j} = \frac{\sum_{a \in U_i \cap U_j} (r_{a,i} - r_{a,j})}{|U_i \cap U_j|} \quad (1)$$

이 식은 주어진 두 항목을 모두 평가한 각 사용자의 두 항목 평가 점수가 얼마나 다른지에 대한 정보를 담고 있다.

2) 선호도 예측

식 (1)에서 계산한 선호도의 차이를 이용하여 새로운 항목 i 에 대한 사용자 u 의 평가 점수 $\hat{r}_{u,i}$ 를 구하는데, 이는 다음 식에 의해 계산된다.

$$\hat{r}_{u,i} = \frac{\sum_{j \in I_u} (r_{u,j} + d_{i,j})}{|I_u|} \quad (2)$$

위의 식에서 항목 j 를 이용해 새로운 항목 i 에 대한 사용자 u 의 평가 점수 예측치를 구할 수 있는데, 이는 사용자 u 가 이전에 평가한 항목 j 에 대한 선호도 $r_{u,j}$ 에 항목 i 와 항목 j 의 선호도의 차이 $d_{i,j}$ 를 더해서 구한다. 이렇게 구한 각 항목 j 를 이용한 예측값들을 단순 평균하여 $\hat{r}_{u,i}$ 를 구할 수 있는데, 이 때 각각의 예측값들에 대한 중요도는 모두 상수 1로 적용한 것이다.

3) 항목의 중요도를 가중치로 계산

위의 경우처럼 각각의 예측값들의 중요도를 1로 적용한 것에 비해서 $d_{i,j}$ 를 구할 때 사용한 데이터의 수 $|U_i \cap U_j|$ 를 항목 j 의 가중치로 적용한 경우 각 항목 j 의 상대적인 중요도를 반영할 수 있다. 가중 평균을 적용한 식은 다음과 같다.

$$\hat{r}_{u,i} = \frac{\sum_{j \in I_u} \{(r_{u,j} + d_{i,j}) \cdot |U_i \cap U_j|\}}{\sum_{j \in I_u} |U_i \cap U_j|}$$

3.2. 사용자의 평점 결정 성향을 이용한 추천 방법

위의 항목 간 선호도 차이를 이용한 추천 방법에 사용자의 평점 결정 성향을 반영하여 사용자 평점 예측의 정확도를 높인다.

1) 사용자의 성향을 반영한 평점 정규화

콘텐츠에 대한 사용자의 평가 점수 결정 방식은 개인의 성향에 따라 조금씩 차이가 있다. 가령 영화 I 를 보고

75/100의 만족감을 느낀 영화 평가 척도가 재미, 시나리오, 캐릭터, 내용 구성, 주제, 영상미 6가지로 동일한 사용자 u_1 과 u_2 가 있을 때, u_1 라는 사용자는 자신의 영화 평가 기준 중 재미, 시나리오, 내용 구성, 영상미의 측면을 좋게 평가 하여 8.0/10의 점수를 주고, u_2 라는 사용자는 앞의 네 가지 측면에서는 만족하였으나 캐릭터와 주제가 만족스럽지 않아 4점을 제하고 6.0/10이라는 점수를 준다고 가정한다. 이렇게 평가 성향이 다른 경우에 생기는 차이를 사용자의 평가 성향에 비추어 정규화 해준다면 또 다른 사용자 u_3 에게 보다 정확한 추천을 해줄 수 있게 된다. 평점 정규화는 사용자들의 평점 데이터가 중앙값 5.5점으로부터 양쪽으로 분포되어 있는 정도를 조정해주는 과정으로, 정규화 과정은 다음과 같이 진행된다.



그림 1. 중앙값을 기준으로 한 평점 정규화

사용자가 각 항목에 대해 1부터 10까지의 점수를 매길 수 있을 때, 중앙값인 5.5점을 중심으로 사용자의 점수 최대값과 최소값이 각각 10점, 1점이 되도록 분포를 조정한다. 예컨대 사용자 u_1 의 평점 최대값이 8.0일 때 u_1 의 점수 7.0을 정규화한다면 중앙값 5.5를 중심으로 오른쪽에 위치한 점수 7을 $5.5 + \left\{ (7 - 5.5) \times \frac{10-5.5}{8-5.5} \right\}$ 를 계산하여 만점이 10점일 때의 점수로 바꾸어주는 것이다. 최소값의 경우도 같은 방식으로 적용된다. 평점의 최소값이 2.0이라면 중앙값 5.5를 넘지 않는 점수 3.0을 $5.5 - \left\{ (5.5 - 3) \times \frac{5.5-1}{5.5-2} \right\}$ 으로 조정해주어 사용자의 평점을 정규화한다.

2) 항목 간 선호도 차이를 이용한 방법에의 적용

논문 [2]에서 제안한 항목 간 선호도 차이를 이용한 추천 방법에 사용자 성향에 따른 평점 정규화를 적용한다. 평점 정규화는 논문 [2]의 항목 간 선호도 차이 식 (1)에 적용할 수 있다.

$$d_{i,j} = \frac{\sum_{a \in U_i \cap U_j} (r_{a,i} - r_{a,j})}{|U_i \cap U_j|}$$

위 식에서 항목 i 와 j 를 모두 평가한 사용자를 각각의 항목 i 와 j 의 평점 차이를 모두 더하는데 이 때 $r_{a,i}$ 와 $r_{a,j}$ 의 값에 각 사용자의 성향에 따른 평점 정규화를 적용시키면 위의 항목 간 선호도 차이를 이용한 목표 사용자의 평점 예측에 보다 객관적인 지표로 사용할 수 있다.

4. 결 론

본 논문에서는 항목 간 선호도 차이 정보를 바탕으로 사용자의 평점 결정 성향을 반영하여 사용자의 선호도를 예측하는 영화 추천 방법을 제안하였다. 사용자들이 영화에 대해 평가 점수를 결정할 때 그 척도가 각 개인마다 다양하므로 평가 점수의 분산 정도가 차이가 나게 된다. 이에 따라 사용자들이 정한 평점을 정규화하여 보다 객관적인 지표로 사용할 필요가 있다.

제안한 방법에서는 사용자들의 평점 결정 성향에 따라 평점을 정규화하였고, 항목 간 선호도 차이 정보에 정규화된 평점을 적용하여 목표 사용자의 평점을 예측하였다.

제안한 방법은 데이터 수가 적은 초기에는 희박성문제에서 자유롭지 못하다는 문제점이 있다. 초기에 사용자 데이터가 충분치 않은 경우 사용자 정보를 이용한 네트워크를 생성하여 평점을 Network Propagation의 방법으로 예측하게 하는 등 향후 부족한 데이터를 채우는 방법에 대한 연구가 진행된다면 희박성 문제를 개선하여 성능을 향상시킬 수 있을 것으로 기대된다.

참고문헌

- [1] George Lekakos, Petros, "A hybrid approach for movie recommendation," *Multimedia Tools and Applications*, Vol. 36, Issue 1-2, pp. 55-70, 2008.
- [2] Se-Chang Oh, Min Choi, "A Movie Recommendation Method Using Rating Difference Between Items," *Journal of Korea Institute of Information and Communication Engineering*, Vol. 17, No. 11, pp. 2602-2608, 2013.
- [3] S. H. Jo, "Weight Recommendation Technique Based on Item Quality to Improve Performance of New User Recommendation on The Web," Ph.D. dissertation, Hannam University Graduation School, 2008.
- [4] S. J. Lee and T. R. Jeon, G. D. Baek, S. S. Kim, "A Movie Rating Prediction System of User Propensity Analysis based on Collaborative Filtering and Fuzzy System," *Journal of Korean institute of intelligent systems*, Vol. 19, No. 2, pp. 242-247, 2009.
- [5] Gediminas Adomavicius, Alexander Tuzhilin, "Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions," *IEEE Transactions on Knowledge and Data Engineering*, Vol. 17, No. 6, 2005.