

사용자 평점 데이터의 가중치 재조정을 통한 최상위 콘텐츠 추천 성능의 향상

이은경[○], 한기준, 이문용

한국과학기술원 지식서비스공학과

leklek0816@kaist.ac.kr, Keejun.han@kaist.ac.kr, munyi@kaist.ac.kr

Improving Top-K Contents Recommendation Performance by Reweighting User Rating Data

Eunkyung Lee[○], Keejun Han, Mun Y. Yi

Department of Knowledge Service Engineering, KAIST

요 약

기존의 추천 시스템은 사용자로부터 주어진 데이터를 최대한 활용하는데 중점을 두고 많은 연구가 진행되어 왔다. 본 연구에서는 주어진 데이터를 선택적으로 선별하여 일부 데이터의 영향력을 감소 혹은 증가 시키는 방법을 제안한다. 이와 관련하여 사용자가 남긴 데이터의 가중치를 재조정하는 추천 알고리즘에 대한 프레임워크를 제시한다. 제안하는 기법의 효용성을 검증하기 위해, 실제 사용자 영화 평점 데이터를 사용하여 성능 평가를 실험하였고 그 결과 제안하는 기법이 기존의 일반적인 추천 방식에 비해 향상된 성능을 보임을 확인하였다.

1. 서론 및 관련 연구

최근 콘텐츠 산업 전반에 걸쳐 사용자의 니즈를 자동으로 발견하고 해당 사용자에게 적합한 콘텐츠를 추천해주는 추천시스템이 활발하게 사용되고 있다. 특히, 사용자들이 남긴 평점 정보 기반의 추천 기법들은 아마존, 넷플릭스 등의 콘텐츠 제공 서비스들에서도 널리 사용되고 있는 성공적인 추천 전략이다.

사용자가 남긴 정보를 추천에 활용하는 방법에는 크게 사용자 기반 협업 필터링(User-based Collaborative Filtering) 기법과 아이템 기반 협업 필터링(Item-based Collaborative filtering) 기법이 있다. 그 중 본 연구의 기반이 되는 사용자 기반 협업 필터링 기법의 경우, 비슷한 관심사를 가진 사용자를 같은 그룹으로 분류하고 그룹 내 다른 사용자의 데이터를 기반으로 해당 사용자의 관심사를 예측하여 추천하는 방법이다. 해당 기법은 한 그룹 내의 사용자들이 같은 경향으로 평점을 매겼다면 그러한 경향을 지속할 것이라는 것과 사용자의 선호도는 급격하게 변하지 않을 것이라는 점을 전제로 한 추천 기법이다.

최근에 활발히 서비스되고 있는 콘텐츠 추천 서비스들이 활용하는 데이터는 주로 1(매우 비선호)에서 5(매우 선호)의 5가지 범주로 된 사용자 평점 스케일을 사용하여 나타낸다. 방대한 콘텐츠에 비하여 해당 콘텐츠에 대한 평점 데이터는 충분하지 않은데서 기인한 콜드 스타트(Cold Start) 문제를 해결하기 위한 연구가 콘텐츠 추천 분야에서 중점적으로 연구되어 왔다. 대표적으로, 보유하고 있는 평점 데이터에 기반한 기계학습을 활용하거나 [2], 시간 정보에 중점을 두고 평점을 추측하는 기법 [3], 평점 이외에도 위치, 날씨 등의 사용자의 상황 정보를 활용하는 기법도 소개되었다 [2].

*이 논문은 2014년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2011-0024560)

그러나, 실제 콘텐츠 추천을 사용하는 사용자 행태 분석에서 콘텐츠 추천 서비스 사용자들은 일반적인 상품 추천 서비스들에 비하여 최상위로 추천되는 콘텐츠들을 보다 더 집중적으로 선택하는 경향이 일반적인 상품 추천에 비하여 훨씬 두드러지게 나타난다 [4]. 이는 현재까지의 콘텐츠 추천 연구가 콜드 스타트 문제를 해결하여 전반적으로 균일한 추천 성능을 보이는 데에 초점을 맞추어 온 것에 비하여, 보다 더 최상위에 정확한 추천 콘텐츠를 제공하는 것이 콘텐츠 추천을 활용하는 사용자들의 만족도 향상에 실제적으로 기여할 수 있음을 의미한다.

상기에 소개된 다양하고 복잡화된 추천 알고리즘이 추천결과와 균일한 향상에 초점을 맞추었으나, 알고리즘에 필요한 물리적, 시간적 요구사항과 사용자의 만족감이 항상 비례하지는 않았다 [5, 6]. 또한 현재까지의 연구들이 개인에 따라 평점을 주는 경향성이 다르다는 것을 인지하고 진행되었지만, 1에서 5까지의 평점이 사용자의 선호도를 표현하는데 있어서 모두 동일한 가중치를 가진다는 가정하에 진행되어 왔다.

이에 본 연구는 사용자가 해당 콘텐츠에 대한 평점을 부여할 때 각 평점에 대해 서로 다른 가중치를 부여하고 평점 간의 스케일을 조정하여 최상위에 위치한 콘텐츠들의 추천 정확도를 극대화하는 방법을 제안한다. 이는 5가지 범주로 된 평점 시스템에서 사용자의 변하지 않는 관심과 비관심 분야를 나타내는데 있어서 평점 5와 1의 경우 2, 3, 4에 비해 상대적으로 영향력이 클 수 밖에 없다는 실제 사용자 평점 행태 분석에서 기인한다 [7]. 이는 실제로 평점 스케일에서 2, 3, 4의 가중치는 1과 5보다 낮게 산정되어야 함을 의미한다.

따라서, 본 논문에서 제안하는 ReRAT(Reweight Strategy for Rating Data)을 통해 구현된 추천 시스템이 기여하는 바는 다음과 같다.

- 제안하는 기법은 사용자가 남긴 평점에 있어서 명확하고 강한 의사 표현(1:매우 비선호, 5:매우 선호)에 보다 많은 가중치를 부여하는 것만으로도 개인의 관심사에 맞는 추천이 가능하다
- 제안하는 기법은 추천 알고리즘이 작동하기 전 사용자 데이터의 구조나 형식을 변화시키지 않기 때문에 기존의 어떠한 형태의 사용자 기반 협업 필터링 기법에도 적용이 가능하다

2. 평점 가중치 재조정 추천 시스템: ReRat

추천 시스템의 추천 기법은 주로 두 단계를 거쳐 이루어진다. (1) 사용자 기반 협업 필터링 기법의 경우, 유사한 사용자를 찾기 위해 유사도(Similarity)를 계산한 뒤, (2) 유사도를 기반으로 유사 사용자의 평점을 이용하여 해당 사용자의 평점을 예측한다. 예측 평점을 높은 순으로 정렬하여 추천 아이템을 결정한다.

서론에서 기술한 바와 같이 1, 2, 3, 4, 5 평점의 가중치를 다르게 주는데 있어서 강한 선호를 나타내는 평점 5와 강한 비선호를 나타내는 1에 가중치를 주고 2, 3, 4에 대해서는 영향력을 약화시켜 적용하였다. 사용자 간의 유사도를 구축할 때 피어슨 상관관계계수(Pearson Correlation Coefficient)와 코사인 유사도(Cosine Similarity)가 주로 사용되는데, 사용자 기반 협업 필터링 기법의 경우 피어슨 상관관계계수가 더 높은 정확도를 나타낸다 [1]. 이에 근거하여 본 논문에서는 피어슨 상관관계계수를 이용하였다. 1에서 5까지의 각 평점에 대한 가중치는 수식 (1)과 같다. r 은 사용자로부터 부여 받은 평점을 의미한다. 수식 (2)는 가중치를 적용한 평점에 대하여 사용자 a 와 b 사이의 피어슨 상관관계계수를 나타낸다. 수식 (3)의 r' 은 가중치를 적용한 사용자 평점을 나타낸다. \bar{r}'_a 는 사용자 a 에 의한 평점의 평균, \bar{r}'_b 는 사용자 b 에 의한 평점의 평균을 의미하며, P 는 사용자 a 와 b 모두가 평점을 준 아이템의 집합을 나타낸다.

$$w = \begin{cases} 1, & r = 1, 5 \\ 0, & r = 2, 3, 4 \end{cases} \quad (1)$$

$$\text{sim}(a, b) = \frac{\sum_{p \in P} (r'_{a,p} - \bar{r}'_a)(r'_{b,p} - \bar{r}'_b)}{\sqrt{\sum_{p \in P} (r'_{a,p} - \bar{r}'_a)^2} \sqrt{\sum_{p \in P} (r'_{b,p} - \bar{r}'_b)^2}} \quad (2)$$

$$r' = w * r \quad (3)$$

그림 1은 가중치를 다르게 적용한 평점으로 구축한 사용자 간 유사도 매트릭스와 가중치를 부여하지 않고 모든 평점을 동등하게 사용한 유사도 매트릭스의 일부이다. 매트릭스 상에서 uk ($k = 1, 2, 3, \dots$) 는 서로 다른 사용자를 나타낸다. 본 논문에서 제안하는 매트릭스(ReRat)와 가중치를 부여하지 않은 매트릭스(Non-ReRAT)를 비교하면 제안하는 기법의 유사도값이 상대적으로 적게 나타나는데, 이는 가중치를

적용하여 상대적으로 영향력이 적은 정보를 제거한 결과이다.

수식 (4)는 아이템 i 에 대한 사용자 a 의 예측 평점을 나타낸다. N 은 사용자 a 와 비슷한 선호도를 나타내는 사용자 그룹을 의미하여, $\text{sim}(a, b)$ 는 사용자 a 와 b 간의 유사도를 나타낸다. \bar{r}_a 와 \bar{r}_b 는 각각 가중치를 적용하기 전 사용자 a 와 b 에 의해 주어진 실제 평점의 평균을 나타낸다.

$$\text{pred}(a, i) = \bar{r}_a + \frac{\text{sim}(a, b) * (r_{b,i} - \bar{r}_b)}{\sum_{b \in N} \text{sim}(a, b)} \quad (4)$$

	u1	u2	u3	u4	u5		u1	u2	u3	u4	u5
u1	0	0.1186	0.2618	0.2899	0.1964	u1	0	0.6668	0.7422	0.7529	0.7511
u2	0.1186	0	0.2023	0.2009	0.0642	u2	0.6668	0	0.7360	0.7366	0.7229
u3	0.2618	0.2023	0	0.3287	0.3790	u3	0.7422	0.7360	0	0.8126	0.8084
u4	0.2899	0.2009	0.3287	0	0.2930	u4	0.7529	0.7366	0.8126	0	0.8230
u5	0.1964	0.0642	0.3790	0.2930	0	u5	0.7511	0.7229	0.8084	0.8230	0

그림 1. 가중치를 적용한 유사도 매트릭스(왼쪽)와 가중치를 적용하지 않은 유사도 매트릭스(오른쪽)

3. 실험

3.1 실험 데이터

제안하는 추천 시스템의 성능 평가를 위하여 우리는 MovieLens¹ 데이터 셋을 평가 데이터로 활용하였다. MovieLens의 영화 평점 데이터의 경우, 실제 사용자 평점 데이터를 대용량으로 제공하고 있어 평점 기반의 추천 시스템의 성능을 평가하는 데에 활발하게 사용되고 있다 [7].

우리는 MovieLens의 데이터 중 사용자가 평점 및 댓글을 많이 남긴 상위 1,000개의 영화를 선정하였다. 1,000개의 영화는 최대 34,000에서 최소 2,500명의 서로 다른 사용자로부터 평점을 받았다. 사용자의 경우, 영화에 대한 평가를 활발하게 남긴 사용자 중 임의로 100명의 사용자를 선정하여 실험에 활용하였다. 100명의 사용자는 1,500에서 2,000개의 영화에 대해서 평점을 주었다. 이로 하여 100명의 사용자가 1,000개의 영화에 대해 남긴 평점을 학습 데이터로 활용하였고, 학습 데이터 영화의 10 퍼센트에 해당하는 100개의 영화를 새롭게 임의로 선정하여 평가 데이터로 사용하였다.

3.2 평가 절차

실험에 대한 평가는 다음의 순서로 진행되었다. 사용자 간의 유사도 수치를 이용하여 평가 데이터 100개의 영화에 대하여 평점을 예측하였다. 평가 데이터 중 실제 사용자로부터 주어진 평점이 존재 하는 경우, 이를 정답으로 가정하여 예측 평점을 평가 하였다.

¹ <http://grouplens.org/datasets/movielens/>

이와 같은 방법은 추천시스템의 성능을 평가하는 관련 연구에서 널리 사용되는 방식이다 [8]. 본 논문에서는 평가 방법으로 Mean Absolute Error (MAE) 척도 및 Normalized Discounted Cumulative Gain (NDCG) 척도를 평가 지표로 사용하였다. MAE의 경우, 모든 영화에 대하여 정답셋의 평점과 예측 평점 사이의 오류의 정도를 측정한다. 추천 시스템의 정확도가 높을수록 MAE는 작은 수치를 가진다. 반면 NDCG의 경우, 상위 k 안에 드는 아이템의 추천 결과 내 순위 및 추천 정확도를 정답셋과 비교한다. NDCG 값이 1에 가까울수록 성능이 뛰어나는 것을 나타낸다. 또한 제안하는 기법(ReRAT)을 가중치를 적용하지 않는 추천 시스템 방법(Non-ReRAT)과 비교하여 성능을 평가하였다.

3.3 실험 결과

표 1은 모든 정답셋의 평점을 예측한 MAE 결과이며, 이는 제안하는 기법이 기존 방식에 비해 보다 더 향상된 성능을 보이며, 통계 분석을 통하여 제안하는 기법의 성능 향상이 기존에 비하여 유의한 성능 향상임(P 단측검증 결과: 0.0007, 유의수준 0.05 검증)을 보여준다.

표1 가중치 적용 유무에 따른 성능 비교

	ReRAT	Non-ReRAT
MAE	0.648	0.651
자유도	99	
t 통계량	3.278	
P(T<t) 단측검정	0.0007	

또한 그림 2는 상위 k 아이템에 대한 예측 평점 및 순위를 평가하여 나타낸 평균 NDCG 결과이며, 상위 $k = 3, 5, 10$ 세 가지 경우로 나눠 NDCG를 비교하였다. 두 기법 모두 비교하는 아이템의 개수가 늘어날수록 성능은 감소 하였으나 상위 3개 아이템의 성능 평가의 경우, 가중치를 적용한 제안하는 방법이 기본 추천 알고리즘에 비해 성능이 뛰어남을 보인다. 이는 사용자가 기존의 상품 추천에 비하여 사용자가 최상위로 추천된 콘텐츠에 보다 더 편중되는 콘텐츠 추천에서 우리의 기법이 기존 기법보다 보다 더 효과적인 추천 기법으로 활용될 수 있음을 의미한다.

최종적으로, 두 실험 결과에서 도출할 수 있는 분석 결과는 다음과 같다.

- 사용자가 남긴 데이터를 최대한으로 활용한 기존의 많은 연구들에 비해, 주어진 데이터 안에서 중요도를 판단하여 일부 평점의 영향력을 줄일 때에 현격하게 정확도가 향상된다
- 제안한 방법은 특히 최상위 아이템을 예측하는데 높은 성능 향상을 보였다. 따라서 평점의 가중치를 재조정 할 경우, 최상위 아이템을 더 정확하게 예측한다.

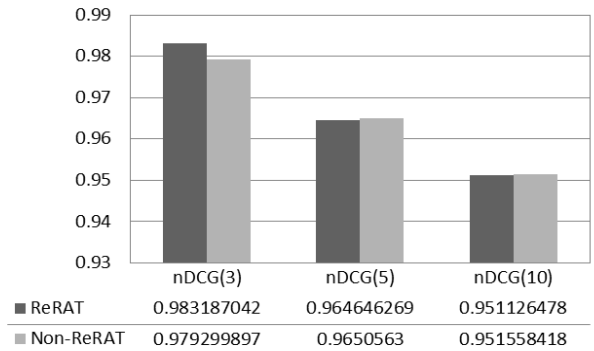


그림 2 가중치 적용 유무에 따른 평균 NDCG 값 비교

4. 결론

본 논문에서는 추천 알고리즘의 정확도를 높이기 위해 평점의 가중치를 다르게 적용하는 프레임워크를 제안하였다. 실험 결과, 제안하는 기법은 기존의 기법과 비교하여 더 나은 성능을 보였다. 가중치를 재조정하여 유사도의 수치가 감소한 것은 데이터를 잃는 현상으로 보이지만, 감소한 절대적 수치 안에 포함된 의미는 더욱 정제되어 있음을 의미한다. 향후 연구에서는 보다 더 가중치를 다변화하는 기법을 통해 제안하는 기법의 성능을 평가하는 것이 필요하다.

참고 문헌

[1] D. Jannach, M. Zanker, A. Felfernig, and G. Friedrich, "Recommender System: An Introduction", Cambridge University Press, 2011.

[2] G. Adomavicius, A. Tuzhilin, "Context-Aware Recommender System", Recommender System handbook Springer, 2011.

[3] Y. Koren, "Collaborative filtering with Temporal Dynamics", In *Proc. of CACM*, 2010.

[4] H. W. Kim, K. Han, M. Y. Yi, J. Cho, and J. Hong, "MovieMine: Personalized Movie Content Search by Utilizing User Comments," *IEEE TCE*, 58(4), pp. 1416-1424, 2012.

[5] C. N. Ziegler, S. M. McNee, J. A. Konstan, and G. Lausen, "Improving Recommendation lists through Topic Diversification", In *Proc. of WWW*, pp. 22-32, 2005.

[6] S. M. McNee, J. Riedl, and J. A. Konstan, "Accurate is not always good: How Accuracy Metrics have hurt Recommender Systems", In *Proc. of CHI*, 2006.

[7] E. Lim, V. Nguyen, N. Jindal, B. Liu, and H. W. Lauw, "Detecting Product Review Spammers using Rating Behaviors", In *Proc. of CIKM*, 2010.

[8] P. Cremonesi, Y. Koren, and R. Turrin, "Performance of Recommender Algorithms on Top-N Recommendation Tasks", In *Proc. of RecSys*, 2010.