

장르 상관관계를 사용한 영화추천 알고리즘

Movie Recommendation Algorithm using Genre Correlation

저자 (Authors)	황태규, 김성권 Tae-Gyu Hwang, Sung Kwon Kim
출처 (Source)	정보과학회 컴퓨팅의 실제 논문지 26(9) , 2020.9, 429-434 (6 pages) KIISE Transactions on Computing Practices 26(9) , 2020.9, 429-434 (6 pages)
발행처 (Publisher)	한국정보과학회 The Korean Institute of Information Scientists and Engineers
URL	http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE09910160
APA Style	황태규, 김성권 (2020). 장르 상관관계를 사용한 영화추천 알고리즘. 정보과학회 컴퓨팅의 실제 논문지, 26(9), 429-434.
이용정보 (Accessed)	원광대학교 61.245.226.*** 2021/07/12 10:21 (KST)

저작권 안내

DBpia에서 제공되는 모든 저작물의 저작권은 원저작자에게 있으며, 누리미디어는 각 저작물의 내용을 보증하거나 책임을 지지 않습니다. 그리고 DBpia에서 제공되는 저작물은 DBpia와 구독계약을 체결한 기관소속 이용자 혹은 해당 저작물의 개별 구매자가 비영리적으로만 이용할 수 있습니다. 그러므로 이에 위반하여 DBpia에서 제공되는 저작물을 복제, 전송 등의 방법으로 무단 이용하는 경우 관련 법령에 따라 민, 형사상의 책임을 질 수 있습니다.

Copyright Information

Copyright of all literary works provided by DBpia belongs to the copyright holder(s) and Nurimedia does not guarantee contents of the literary work or assume responsibility for the same. In addition, the literary works provided by DBpia may only be used by the users affiliated to the institutions which executed a subscription agreement with DBpia or the individual purchasers of the literary work(s) for non-commercial purposes. Therefore, any person who illegally uses the literary works provided by DBpia by means of reproduction or transmission shall assume civil and criminal responsibility according to applicable laws and regulations.

장르 상관관계를 사용한 영화추천 알고리즘 (Movie Recommendation Algorithm using Genre Correlation)

황 태 규 [†] 김 성 권 ^{††}
(Tae-Gyu Hwang) (Sung Kwon Kim)

요 약 영화장르는 주제, 줄거리, 분위기 등이 요약된 특성이고, 같은 장르의 영화들은 비슷한 특성을 가지며, 영화 제작자나 전문가에 의해 영화장르가 분류된다. 사용자가 새로운 영화를 선택할 때, 영화장르로부터 유추된 내용을 바탕으로 의사결정하기 때문에, 영화장르는 선호도를 대표할 수 있는 중요한 단서가 된다. 기존의 방법은 영화장르 분석을 통해 추천 정확성을 향상시켰지만, 협업 필터링 기반의 평점예측으로 인해 복잡도가 크고, 영화분류에 사용된 매개변수 값에 따른 성능편차가 크기 때문에 모델의 최적화가 어려우며, 분류된 영화들의 조합으로 추천목록을 작성하기 때문에 인위적인 추천결과를 보였다. 본 논문에서는 제시된 문제해결을 위한 알고리즘을 제안하며, 실험을 통해 제시된 문제들이 해결됨을 확인할 수 있었다.

키워드: 편향 기반 분석, 장르 상관관계, 평점 예측, 장르 기반 추천, 추천 알고리즘

Abstract Genres categorize movies and help summarize their themes, plots and moods. As such, movies of the same genre should have similar characteristics when they are classified by movie-makers or domain experts. When you choose a new movie, the genre becomes an important clue to match your preferences to something you haven't seen before, basically it helps us make decisions based on analogy to previous movies we have seen in that genre. Although previous methods have improved recommendation accuracy through analysis of movie genre, recommending movies is a high complexity problem due to the collaborative filtering-based rating prediction used. Also, this model is difficult to optimize because of the large variation in performance according to the parameter values used in the movie classification, at the same time the recommended results are artificial because the recommendation list is written using a combination of classified movies. In this paper, we introduce new methods for solving the problems presented, and through experiments we show that our approach successfully solves these problems.

Keywords: bias-based analysis, genre correlation, rating prediction, genre-based recommendation, recommendation algorithm

- 이 논문은 2019년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No. NRF-2019R1F1A1059952)
- 이 논문은 2019 한국소프트웨어종합학술대회에서 '장르 상관관계를 사용한 영화추천 알고리즘'의 제목으로 발표된 논문을 확장한 것임

[†] 학생회원 : 중앙대학교 컴퓨터공학과 학생
tghwang@alg.cse.cau.ac.kr

^{††} 종신회원 : 중앙대학교 소프트웨어학부 교수(Chung-Ang Univ.)
skkim@cau.ac.kr
(Corresponding author)

논문접수 : 2020년 4월 9일
(Received 9 April 2020)
심사완료 : 2020년 6월 23일
(Accepted 23 June 2020)

Copyright©2020 한국정보과학회 : 개인 목적이나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.
정보과학회 컴퓨팅의 실제 논문지 제26권 제9호(2020. 9)

1. 서론

영화추천시스템은 선호패턴 분석을 통해 사용자가 좋아할만한 영화들을 검색 및 여과하는 기법으로써, 평점에 따른 선호패턴 분석으로 선호도를 예측하는 것이 핵심이다. 협업필터링(CF, Collaborative Filtering)은 대표적인 추천기법의 일종으로, 기억기반(Memory-based)[1], 모델기반(Model-based)[2], 콘텐츠기반(Content-based)[3], 상황기반(Context-aware)[4], 혼합모델(Hybrid)로 구분된다. 협업필터링은 학술적, 상업적 성공을 이뤘으나, 정확성은 행렬분해기법을 기점으로 콘텐츠 영역에 따른 소폭조정 수준의 최적화를 보였기 때문에, 성능향상의 한계를 보였다. 이는 기존기법의 진부함과 데이터 및 관련정보의 부족을 원인으로 볼 수 있다.

컴퓨팅기술의 발전으로 신경망 기법은 최근 다분야에 적용되어 기존기법의 한계들을 풀어냈고, 추천시스템에 적용되어 기존성능의 한계를 극복해 주목받고 있다[5,6].

평점은 사용자의 의사결정으로 대표되지만 의사결정에 관여한 과정이 생략되므로, 선호도 이해를 위한 단서들이 부족하다. 기존기법은 선호패턴의 유사도를 반영해 의사결정 결과를 맞췄지만, 의사결정과정에 대한 이해부족으로 추천 성능의 한계가 있다. 메타데이터를 사용한 기법들은 이를 재현해 선호도를 이해하기 위한 접근이지만, 공개된 데이터의 정보를 활용하기엔 한계가 있어, 데이터 수집환경을 구성해 소규모로 데이터를 수집하는 것이 현실이다. 성능한계를 풀기 위해서는 기존과 다른 관점의 분석을 반영해 식 (1)과 같이 접근해야 한다. 식 (1)의 p 는 관점, ω_p 는 관점 p 의 가중치, $f_p(x)$ 는 관점 p 에 따른 분석함수, ϵ 는 관점이 반영되지 않은 오차이다.

$$Decision = \sum_p^{Perspective} \omega_p * f_p(x) + \epsilon \quad (1)$$

장르는 영화, 음악, 영상, 책 등의 콘텐츠 관련 항목에서 찾아볼 수 있고, 콘텐츠에 따른 특성차가 있으나 역할이 동일하기 때문에, 콘텐츠서비스의 핵심 범주로 자리하고 있다. 영화장르는 영화의 주제, 줄거리, 분위기, 기타 등 영화의 내용이 단순 요약되어 사용자에게 전달되는 정보로써, 영화제작자나 관련 전문가에 의해 일관성 있게 분류된다. 영화장르는 사용자가 영화를 선택할 때, 영화의 특성을 유추하기 위한 단서로 사용되기 때문에 의사결정의 근거가 된다. 따라서 선호의 특성이 내포된 주제, 줄거리, 분위기, 기타 등의 요소를 요약하는 영화장르는 사용자의 선호를 이해하기 위한 핵심단서가 될 수 있다.

사용자가 특정 주제, 줄거리 등을 선호한다면 이와 관련된 2~3개 정도의 영화장르에 속한 영화들을 시청할 확률이 높아, 사용자의 선호가 상위 몇 개의 영화장르에 평점빈도가 편향된(집중된) 패턴을 보일 것이다. 제안된

방법은 이러한 관점을 반영하는 것이 핵심이다.

기존의 장르분석기법은 평점에 따른 영화장르 간의 상관관계를 구하고, 이를 기준으로 영화들을 분류 및 추천한 결과, 협업 필터링의 정확성을 향상시켰다[7]. 그러나 영화분류에 사용된 매개변수에 따른 정확성의 편차가 크고, 분류된 영화들의 조합으로 추천목록을 작성하기 때문에, 추천영화들의 다양성이 떨어지는 결과를 보였다. 또한, 장르 상관관계의 계산 복잡도는 데이터 규모만큼 선형적이고, 협업 필터링을 기반으로 평점을 예측하기 때문에 복잡도의 한계가 있다[8].

편향기반 분석모델은 협업 필터링보다 적은 복잡도임에도 높은 정확성을 보여 다수 사용자가 편향된 선호도를 가졌음을 확인할 수 있다[8]. 그러나 사용자와 항목 수준의 편향만 분석했기 때문에, 평점에측 차원이 낮아 추천의 다양성이 떨어졌다.

[9]는 영화장르 선호도를 구분해 선호패턴 분석 및 예측에 직접 반영해, 협업 필터링의 평점에측 성능을 향상시켰다. 그러나 기억기반으로 선호패턴을 분석하기 때문에 복잡도의 한계가 있다.

본 논문에서는 영화장르 상관관계를 사용한 영화추천 알고리즘을 제안한다. 제안된 방법은 기존의 장르분석기법을 기반으로 장르 상관관계를 계산하지만, 편향기반 분석의 개념을 통해 계산과정을 간소화했고, 장르 상관관계를 평점에측에 직접 반영하도록 설계했다. 영화데이터를 대상으로 실험한 결과, 기존의 장르분석기법보다 성능이 향상됨을 확인할 수 있었다.

2. 제안된 방법

제안된 방법은 기존의 장르분석기법[7]의 복잡도를 줄이기 위해 편향기반 분석의 개념을 사용해 설계했다. 평점에측은 편향분석모델에 장르 점수를 반영해 계산하고, 장르점수는 장르 편향에 장르 상관관계를 가중치로 사용해 계산한다. 2.1에서는 장르 상관관계, 2.2에서는 장르 편향, 2.3과 2.4에서는 장르점수 및 평점에측에 대해 소개한다.

2.1 장르 상관관계

장르 상관관계는 목표 장르 g_t 와 다른 장르 g_o 간의 상관관계 값이고, ω_{GP} 와 ω_{GW} 를 사용해 식 (2)와 같이 계산하며, $\omega_{GC}(g_t, g_o)$ 와 같이 표기했다.

$$\omega_{GC}(g_t, g_o) = (1 - \alpha) * \omega_{GP}(g_t, g_o) + \alpha * \omega_{GW}(g_t, g_o) \quad (2)$$

α 는 0~1의 값으로 ω_{GP} 와 ω_{GW} 의 비율조절을 위해 사용했고, 식 (3)과 같이 계산한다. 두 장르의 공통발생 빈도가 클수록 연관성이 높음을 반영할 목적으로, 두 장르 g_t 와 g_o 에 공통으로 속한 영화가 많을수록 ω_{GW} 의 비율을 높이도록 계산된다.

$$\alpha = \frac{\log |I(g_t) \cap I(g_o)|}{\log \max(|I(g_t)|, |I(g_o)|)} \quad (3)$$

$I(g_t)$ 는 장르 g_t 에 속한 영화들의 집합, $|I(g_t)|$ 는 g_t 에 속한 영화들의 수, $\max(|I(g_t)|, |I(g_o)|)$ 는 두 값 중에서 최댓값, $|I(g_t) \cap I(g_o)|$ 는 두 장르에 공통으로 속하는 영화들의 수를 뜻한다.

2.1.1 장르 확률

장르 확률은 장르 상관관계의 비대칭성을 반영하기 위해 사용한다. 상관관계의 비대칭성을 반영하는 이유는 액션을 좋아하는 사용자가 어드벤처를 선호하는 정도와 어드벤처를 좋아하는 사용자가 액션을 선호하는 정도의 차이를 반영하기 위함이다. 장르 확률은 식 (4)와 같이 조건부확률로 계산하고 ω_{GP} 와 같이 표기했다.

$$\omega_{GP}(g_t, g_o) = \frac{\log |I(g_t) \cap I(g_o)|}{\log |I(g_t)|} \quad (4)$$

2.1.2 장르 가중치

장르 가중치는 피어슨 상관관계를 기반으로 설계했고, 두 장르간의 상관관계를 반영할 목적으로 사용한다. 장르 가중치는 ω_{GW} 와 같이 표기했고, 식 (5)와 같이 계산하며, 식 (5)의 $gs_i(g_x)$ 는 식 (6)을 통해 계산한다. 영화는 1개 이상의 영화장르 조합으로 구성되지만, 각 장르들의 기여도나 대표 장르를 모를 경우, 상관관계 계산의 공평성을 부여해야 한다. 식 (7)은 장르의 기여도를 모를 때, 영화 i 에 속한 장르들이 장르 가중치 계산에 공평한 기회를 주기위한 계산이고, pnt_i 와 같이 표기한다. 식 (6)을 통한 상관관계 계산은 데이터 크기가 클수록 기존의 방법보다 빠르게 계산되고, 계산 값의 신뢰성이 향상된다.

$$\omega_{GW}(g_t, g_o) = \frac{\left\{ \sum_{i \in I(g_t) \cap I(g_o)} pnt_i * gs_i(g_t) * gs_i(g_o) \right\} / \sqrt{\sum_{i \in I(g_t) \cap I(g_o)} pnt_i^2 * gs_i(g_t)^2 \sum_{i \in I(g_t) \cap I(g_o)} pnt_i^2 * gs_i(g_o)^2}} \quad (5)$$

$$gs_i(g_x) = \bar{R}(i) - \bar{R}(g_x) \quad (6)$$

$$pnt_i = 2/|G(i)| \quad (7)$$

2.2 장르 편향

사용자들은 선호하는 영화장르가 다르고, 장르에 대한 선호정도의 차이가 있다. 장르편향은 선호정도의 차이를 뜻하고, 편향기반 분석개념을 사용해 식 (8)과 같이 계산하며, $b(\mu, u, g_t)$ 와 같이 표기했다.

$$b(\mu, u, g_t) = \frac{\log |R(u, g_t)|}{\log \max_{x \in G} |R(u, g_x)|} * (\bar{R}(u, g_t) - \mu) \quad (8)$$

$b(\mu, u, g_t)$ 는 μ 를 기준으로 계산된 사용자 u 의 장르 g_t 에 대한 편향, $R(u, g_x)$ 는 사용자 u 가 장르 g_x 에 평가한 평점들의 집합, $\bar{R}(u, g_t)$ 는 사용자 u 가 장르 g_t 에 평

가한 평점들의 평균, μ 는 학습데이터의 모든 평점의 평균이다. 식 (8)의 분모는 사용자 u 가 가장 많이 평점을 부여한 장르 g_x 에 대한 평점의 수로 정규화하고, 평점 빈도가증치는 평점빈도 수의 비율을 로그형태로 반영한다.

2.3 평점 예측

평점 예측은 식 (9)과 같이 편향분석모델을 기반으로 장르에 따른 예측 값을 반영하도록 설계했고, $P(\mu, u, i)$ 와 같이 표기했다. 장르에 따른 예측 값은 사용자 u 가 가장 많이 평가한 장르를 기준으로 영화 i 에 속한 장르간의 관계를 반영해 계산한다.

$$P(\mu, u, i) = \mu + b(\mu, u) + b(\mu, i) + \frac{\sum_{g_o \in G(i)} \omega_{GC}(g_{top}(u), g_o) * b(\mu, u, g_o)}{|G(i)|} \quad (9)$$

$b(\mu, u)$ 는 식 (10)과 같이 μ 를 기준으로 계산된 사용자 u 의 편향, $b(\mu, i)$ 는 식 (11)과 같이 μ 를 기준으로 계산된 영화 i 의 편향, $g_{top}(u)$ 는 식 (12)와 같이 사용자 u 가 가장 많이 평가한 영화장르를 뜻한다. 식 (10)에서 사용된 $\max_{v \in Users} |R(v)|$ 는 평점을 가장 많이 부여한 사용자 v 의 평점의 개수, 식 (11)의 분모에 사용된 $\max_{j \in Items} |R(j)|$ 는 평점을 가장 많이 부여한 영화 j 의 평점의 수를 뜻한다.

$$b(\mu, u) = \frac{\log |R(u)|}{\log \max_{v \in Users} |R(v)|} * (\bar{R}(u) - \mu) \quad (10)$$

$$b(\mu, i) = \frac{\log |R(i)|}{\log \max_{j \in Items} |R(j)|} * (\bar{R}(i) - \mu) \quad (11)$$

$$g_{top}(u) = \arg \max_{g \in G} |R(u, g)| \quad (12)$$

2.4 혼합모델

제안된 기법은 정확성 측면에서 고평가 받는 행렬분해 기법과 편향분석에 최적화된 편향분석모델을 적용해 혼합모델을 설계했다. 두 기법은 장르를 분석하지 않고, 제안된 모델은 선호패턴 분석이 단순하므로, 각 모델의 단점보완을 위해 식 (13)과 같이 평점예측 하도록 설계했다.

$$P(s, u, i) = s + b(s, u) + b(s, i) + q_i^T p_u + \frac{\sum_{g_o \in G(i)} \omega_{GC}(g_{top}(u), g_o) b(s, u, g_o)}{|G(i)|} \quad (13)$$

s 는 편향계산에 적합한 기준 값이고, 편향분석에 적합한 기준점을 탐색하여 고정된다[10]. $b(s, u)$ 는 s 를 기준으로 계산된 사용자 u 의 편향, $b(s, i)$ 는 s 를 기준으로 계산된 영화 i 의 편향, $q_i^T p_u$ 는 인수분해행렬의 곱으로 예측된 점수다.

$$\min_{s, b(s, *), q_u, p_u} \sum_{(u, i) \in \tau} (R(u, i) - s - b(s, u) - b(s, i) - q_i^T p_u)^2 + \lambda(b(s, u)^2 + b(s, i)^2 + \|q_i\|^2 + \|p_u\|^2) \quad (14)$$

혼합모델의 목적함수는 행렬분해모델의 목적함수를 기반으로 식 (14)와 같이 설계했다. τ 는 학습데이터 집합, λ 는 과적합 회피를 위한 일반화 변수, γ 는 학습률 설정을 위한 변수다.

$$\epsilon(s, u, i) = R(u, i) - P(s, u, i) \quad (15)$$

$$b(s, u) = b(s, u) + \gamma^*(\epsilon(s, u, i) - \lambda * b(s, u)) \quad (16)$$

$$b(s, i) = b(s, i) + \gamma^*(\epsilon(s, u, i) - \lambda * b(s, i)) \quad (17)$$

$$q_i = q_i + \gamma^*(\epsilon(s, u, i) \cdot p_u - \lambda \cdot q_i) \quad (18)$$

$$p_u = p_u + \gamma^*(\epsilon(s, u, i) \cdot q_i - \lambda \cdot p_u) \quad (19)$$

혼합모델의 학습은 식 (15)와 같이 실제 값과 예측 값의 오차를 계산해, 식 (16)~(19)의 선형적 증감에 사용해 최적화한다. 편향기반 모델의 최적화는 식 (16)과 (17)과 같고, 행렬분해모델의 최적화는 식 (18)과 (19)와 같다.

3. 실험 및 평가

3.1 실험용 데이터

실험에 사용된 원본 데이터 셋은 표 1과 같고, 평점된 시간을 기준으로 오름차순 정렬해 5회 교차검증 환경을 구성했으며, 실험결과의 평균값을 성능비교에 사용했다.

표 1 데이터 셋

Table 1 Dataset description

Dataset (ML, MovieLens)	ML100K	ML1M	ML10M
No. of users	943	6,040	71,567
No. of items	1,682	3,900	10,681
No. of ratings	100,000	1,000,209	10,000,054
No. of genres	19		
Environment	5-fold cross-validation		

3.2 평가척도

평가척도는 정확성을 RMSE, 복잡도는 runtime로 했다. RMSE(Root mean square error)는 식 (20)과 같이 계산되며, 값이 작을수록 평점예측이 정확함을 뜻한다. Runtime은 모든 사용자에게 모든 영화들의 평점이 예측될 때까지 걸린 시간이고, 값이 작을수록 복잡도가 적음을 뜻하며, 그림 2에서 초단위로 표기해 성능을 비교했다.

$$RMSE = \sqrt{\frac{\sum_{(u,i) \in TestSet} (R(u,i) - P(u,i))^2}{|TestSet|}} \quad (20)$$

3.3 성능평가

3.3.1 장르 상관관계

기존의 장르상관관계는 두 장르에 공통으로 속한 영화들의 평점들을 내적해 계산했지만, 제안된 방법은 장르의 편향을 내적해 계산복잡도를 줄였다. 계산을 간소화했지만 정확성을 얼마나 잃었는지 확인하기 위해 코

표 2 두 장르 상관관계 간의 유사도 비교

Table 2 Comparison of similarities between two genre correlation

Method	Dataset	Previous work [7]		
		ML100K	ML1M	ML10M
Our Method	ML100K	0.9614	0.9291	0.9286
	ML1M	-	0.9796	0.9645
	ML10M	-	-	0.9817

사인 유사도로 두 장르상관관계를 비교했다. 표 2는 두 장르상관관계간의 유사도이고, 코사인 유사도는 대칭이므로 반대 유사도는 비웠다. 추가로, 규모가 다른 데이터 셋과의 유사성을 비교했다.

표 2와 같이 장르 상관관계의 유사도는 92% 이상이고, 데이터 규모에 비례해 유사도가 증가했다. 데이터 규모 측면에서 유사도를 관찰했을 때, ML100K는 더 큰 규모의 데이터 셋의 상관관계보다 멀리 떨어짐을 보였다. 이것은 장르 상관관계를 계산하는데 필요한 정보들이 데이터 규모를 축소하는 과정에서 누락됨으로 볼 수 있다.

3.3.2 성능비교

기존모델[7]은 UBCF-MC, 편향분석모델[8]은 BBP, 장르 선호도 모델[9]은 UBCF-GP로 명시하여 그림 1, 2와 같이 성능을 비교했다.

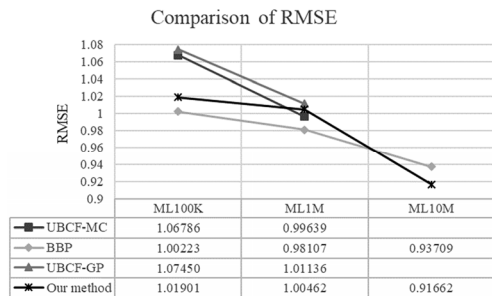


그림 1 기존연구와 정확성 비교

Fig. 1 Comparison of accuracy with previous studies

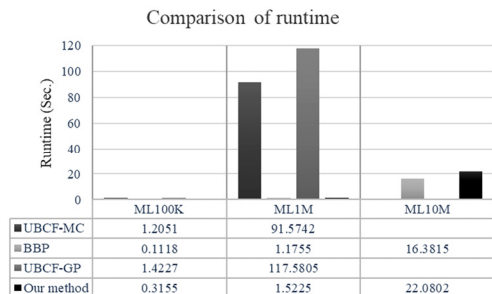


그림 2 기존연구와 복잡도 비교

Fig. 2 Comparison of complexity with previous studies

UBCF-MC와 UBCF-GP는 ML10M에서 공간복잡도의 한계로 성능평가가 불가능했고, 그림 2에서 BBP를 제외할 경우 제안된 방법의 runtime이 가장 좋았다.

그림 1을 통해 제안된 방법과 UBCF-MC의 성능을 비교할 경우, ML100K에서 더 좋고 ML1M에서 근접했고, 그림 2를 함께 고려하면 수행시간이 적음을 근거로 효율적임을 주장할 수 있다. 이를 통해 제안된 모델의 적합성과 사용자들의 선호패턴이 특정 영화장르에 편향되었다는 가설의 합당함을 주장할 수 있다. 반면, 제안된 방법은 평점예측의 기준인 BBP와의 정확성 비교 시, ML10M에서만 좋은 결과를 보였다. BBP의 경우 편향 계산에 사용된 기준점 최적화를 수행했지만, 제안된 방법은 식 (9)와 같이 기준점 최적화가 적용되지 않아 성능이 떨어졌다.

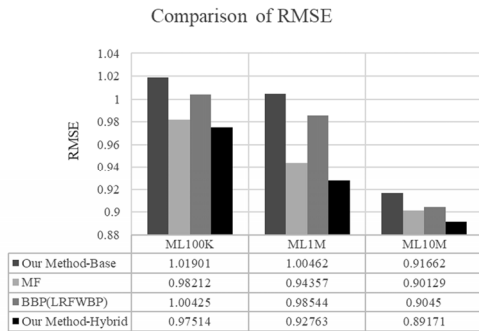


그림 3 최적화를 통한 정확성 비교

Fig. 3 Comparison of accuracy after optimization

표 3 모델 환경설정

Table 3 Model environment setting

Method	Dataset	Model environments
Our Method Base	ML100K	N/A
	ML1M	
	ML10M	
MF	ML100K	$f = 5, iter = 100, \lambda = 0.1, \gamma = 0.005$
	ML1M	$f = 10, iter = 75, \lambda = 0.1, \gamma = 0.005$
	ML10M	$f = 10, iter = 125, \lambda = 0.05, \gamma = 0.005$
BBP (LRFWBP)	ML100K	$n = 19$
	ML1M	$n = 35$
	ML10M	$n = 43$
Our Method Hybrid	ML100K	$f = 5, iter = 100, \lambda = 0.1, \gamma = 0.005, n = 19$
	ML1M	$f = 10, iter = 75, \lambda = 0.1, \gamma = 0.005, n = 35$
	ML10M	$f = 10, iter = 125, \lambda = 0.05, \gamma = 0.005, n = 43$

3.3.3 최적화

그림 3은 최적화를 통한 정확성비교다. BBP[8]는 학습 성능개선[10]을 통해 성능을 향상했다. 최적화 비교에는 개선된 BBP를 사용해 BBP(LRFWBP)로 명시했다. 제안된 기법은 기본모델을 Our Method-Base, 혼합모델을 Our Method-Hybrid로 구분해 명시했고, 행렬 분해모델은 MF로 명시했다. 표 3은 최적화된 변수들의 설정 값이다.

4. 결론

제안된 방법은 실험을 통해 기존모델의 복잡도와 최적화 문제를 해결했고, 정확성이 개선됨을 보였다. 혼합모델은 행렬분해모델의 정확성 개선을 근거로 기존에 반영하지 못했던 관점을 반영했고, 영화장르에서 사용자의 선호를 이해하기 위한 단서가 존재함을 확인했다.

편향을 통해 장르 상관관계 계산을 간소화한 결과, 표 2와 같이 최소 92%이상의 유사도를 보였고, 유사도가 데이터 규모에 비례함을 보였다. 이를 근거로, 영화장르에 편향된 패턴에서 사용자들의 선호패턴이 관측됐고, 데이터 규모와 편향의 신뢰성이 비례할 것으로 볼 수 있다.

편향은 많은 분야에서 축의 이동수준으로 중요성이 적고, 왜곡된 견해로 해석되지만, 선호 및 의사결정의 원인 중 하나로써, 예측에 반영해야 될 사실 중 하나다. 편향분석만으로 극한의 성능을 낼 수 없지만, 저비용으로 기존기법과 결합해 성능을 높여, 관련분야의 보조로 활용될 것으로 기대된다. 향후연구를 통해, 다차원편향 분석을 통한 데이터묘사, 다른 콘텐츠의 편향패턴분석, 편향기반 교차추천으로 편향분석 모델을 확장하고자 한다.

References

- [1] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-based Collaborative Filtering Recommendation Algorithms," *Proc. of the 10th international conference on World Wide Web*, pp. 285-295, 2001.
- [2] Y. Koren, R. Bell, and C. Volinsky, "Matrix Factorization Techniques for Recommender Systems," *Computer*, Vol. 42, Issue 8, pp. 42-49, 2009.
- [3] M. J. Pazzani and D. Billsus, "Content-Based Recommendation Systems," *The Adaptive Web*, pp. 325-341, Springer, Berlin, 2007.
- [4] G. Adomavicius and A. Tuzhilin, "Context-aware recommender systems," *Recommender systems handbook*, pp. 217-253, Springer, Boston, 2011.
- [5] X. He, L. Liao, H. Zhang, L. Nie, X. Hu, and T. S. Chua, "Neural collaborative filtering," *Proc. of the 26th international conference on world wide web*, pp. 173-182, 2017.
- [6] S. Li, J. Kawale, and Y. Fu, "Deep collaborative filtering via marginalized denoising auto-encoder,"

- Proc. of the 24th ACM International on Conference on Information and Knowledge Management*, pp. 811-820, 2015.
- [7] T. G. Hwang, C. S. Park, J. H. Hong, and SK Kim, "An algorithm for movie classification and recommendation using genre correlation," *Multimedia Tools and Applications*, Vol. 75, No. 20, pp. 12843-12858, Oct. 2016.
- [8] T. G. Hwang and S. K. Kim, "Bias-Based Predictor to Improve the Recommendation Performance of the Rating Frequency Weight-based Baseline Predictor," *Journal of KIISE*, Vol. 44, No. 5, pp. 486-495, May. 2017. (in Korean)
- [9] T. H. Park, S. C. Kim, and S. K. Kim, "Improvement of Collaborative Filtering Using Genre Preference," *Proc. of the 37th KIISE Fall Conference*, pp. 264-268, 2010. (in Korean)
- [10] T. G. Hwang and S. K. Kim, "A method of improving the learning accuracy from Bias-Based Predictor," *Proc. of the KIISE Korea Computer Congress 2017*, pp. 796-798, 2017. (in Korean)



황 태 규

2013년 대전대학교 컴퓨터공학과 졸업(학사). 2013년~현재 중앙대학교 컴퓨터공학과 석·박사통합과정. 관심분야는 데이터마이닝, 추천시스템



김 성 권

서울대학교 전자통계학과 졸업(학사). 한국과학기술원 전산학과 졸업(석사). University of Washington, Computer Science & Engineering 졸업(박사). 1996년~현재 중앙대학교 컴퓨터공학부 교수. 관심분야는 암호응용 및 정보보호, 계산기하

학 및 응용, 추천시스템