

# 유사 취향 사용자의 시간 상황에 따른 선호 아이템에 가중치를 둔 음악 추천

박성은<sup>o</sup>, 이동주, 강민석, 이상구

서울대학교 컴퓨터공학부

{separk1031, therocks, minsuk, sglee}@europa.snu.ac.kr

## Music Recommender System Weighting Similar Users' Preference in the Temporal Context

Sung Eun Park<sup>o</sup>, Dongjoo Lee, Minsuk Kahng, Sang-goo Lee

School of Computer Science and Engineering, Seoul National University

### 요 약

사용자와 취향이 비슷한 사용자를 찾고, 이 유사 사용자가 선호한 아이템을 추천하는 협력적 필터링 방식은 일반적으로 많이 사용되는 추천 방식이다. 하지만 협력적 필터링 방식은 어떤 상황적 요소도 고려하지 않아 모든 상황에서 동일한 추천 결과를 제시하게 된다. 반면, 상황을 고려한 추천 방식은 다른 상황에서 그 상황에 적합하다고 판단되는 추천 리스트를 보여주는 다양성을 가지지만 개인의 선호를 반영하지 못하는 한계를 가진다. 이에 협력적 필터링 방식과 상황에 따른 추천 방식을 함께 고려하려는 시도가 있다. 본 논문에서는 시간 상황에 따른 음악 추천 시, 전체 상황에서 가장 유사한 사용자를 찾고 이 유사 사용자의 현재 상황에서의 선호 아이템을 추천하는 모델을 제시하고 실험을 통하여 이 모델의 한계와 실용 가능한 상황을 제시한다.

### 1. 서론

추천의 대상이 되는 사용자의 취향을 판단하여 그 사용자와 취향이 비슷한 사용자를 찾고, 이 유사 사용자가 선호한 아이템을 대상 사용자에게 추천하는 협력적 필터링 방식은 그 성능이 인정되어 일반적으로 많이 쓰이는 추천 방식이다. 이 협력적 필터링은 개인의 취향을 기반으로 추천한다는 면에서 개인화된 추천 모델이다. 이 모델은, 그 사용자가 속해 있는 환경을 고려하지 않아, 어떤 시점에서나 같은 데이터를 바탕으로 같은 추천 결과를 제시한다.

반면, 상황인지 추천 기법은 사용자가 속한 상황에 어울리는 음악을 추천하지만, 개인의 특성을 고려하기 보다는 상황에 따른 대중의 경향성을 파악하고 그 경향성에 따라 적합한 아이템을 추천을 하는 방식이다. 따라서 단순한 상황인지 추천 기법에서는 개인의 취향을 고려하지 않는다.

상황인지와 협력적 필터링을 모두 고려하는 연구는 이런 배경으로부터 추천 목록의 다양성과 정확성을 높일 수 있는 방법으로 기대되며 연구가 이루어지고 있다. 상황인지와 협력적 필터링을 모두 고려하여 추천할 때, 개념적으로 유사 사용자를 구하는 대상 데이터에 따라 그 방식은 두 가지가 될 수 있다. 첫째는,

상황에 상관없이 전체적으로 유사한 성향을 보이는 사용자를 먼저 구한 뒤, 이 사용자가 추천하는 시점과 유사한 상황에서 선호한 아이템을 추천하는 방식이다. 둘째는, 추천을 시점의 상황에서, 그 상황에서 유사한 사용자를 찾아서 그 사용자가 그 상황과 같은 상황에서 선호한 아이템을 추천하는 방식이다. 첫 번째 방식은 비슷한 취향의 사용자가 상황에 따라 변화하는 취향도 비슷하다는 가정에서 나오는 모델이라고 할 수 있고, 두 번째 방식은 일반적으로 비슷한 취향을 가진 사용자도 상황에 따라서 변화하는 취향은 다르기 때문에 각각의 상황에서 유사 사용자를 매번 계산해야 한다는 가정을 따르는 것이다.

예를 들어, 사용자 A, B, C에게 음악을 추천하려고 한다고 하자. B는 C보다 일반적으로 A와 더 비슷한 취향을 가지지만, 여름 밤이라는 특정 시점에서는 C가 B보다 더 비슷한 취향의 음악을 듣는다고 가정하자. 이때 첫 번째 방법에서는 B가 여름 밤에 들은 음악을 추천하고, 두 번째 방법에서는 C가 여름 밤에 들은 음악을 추천한다. 이전의 연구에서는 두 번째 방법만을 고려하여 왔다.

이에, 본 논문에서는 첫 번째 방식을 모델로 제시하여 구현하고 실험한다. 본 논문의 목표는 협력적 필터링을 이용한 추천이나 상황정보를 이용한 추천 성능을 높이기 위한 모델을 제시하는 것이 아니고, 다루어지지 않았던 가정을 검증하기 위한 모델을 제시하고 이전의

본 연구는 지식경제부 및 정보통신산업진흥원의 대학 IT연구센터 육성·지원사업(NIPA-2010-C1090-1031-0002)의 연구결과로 수행되었음.

방식[2]과 비교함으로써, 상황을 고려한 협력적 필터링으로 음악을 추천할 때 어떤 가정을 따르는 것이 옳은가에 대해 고민하는 것이다. 이 때, 개인의 선호는 사용자의 음악청취 로그로부터 추론한다.

본 논문의 구성은 다음과 같다. 먼저 2장에서는 협력적 필터링 방식과 상황인지를 이용한 추천에 대한 관련연구를 살펴본다. 3장에서는 협력적 필터링과 상황인지를 함께 고려한 모델을 소개한다. 마지막으로 4장에서는 이 모델의 성능을 실험하고 비교한다.

## 2. 관련 연구

협력적 필터링을 이용한 추천은 추천 사용자의 프로파일과 가장 유사한 사용자를 찾은 후 그 유사 사용자의 아이템을 바탕으로 추천하는 방식으로 [5]를 시작으로 꾸준히 연구되어, 최근까지 활발하게 연구되고 있다[6,7,8].

사용자의 상황을 로그로부터 추출하고 이를 반영하여 추천하는 연구는 [1]에서 시도되었다. [1]는 사용자와 아이템 사이의 연관도를 상황을 기반으로 데이터를 군집한 계층을 통해서 알아내는 방식을 사용하였다. 이 연구에서는 상황을 고려하는 모델을 제시하였을 뿐, 개인의 취향을 반영하지는 않았다. 이 외에도 상황을 고려한 추천은 [9,10]를 비롯하여 활발하게 연구되고 있다.

상황과 개인의 취향을 모두 고려한 추천 모델은 [2]에서 제시된 적이 있었다. 이 모델은 각 상황에서 협력적 필터링을 사용하여 유사 사용자를 찾은 후, 이 유사 사용자의 평가를 반영하는 방식을 사용하였다. 본 논문의 방식은 협력적 필터링을 먼저 적용하여 유사 사용자를 찾고, 추천을 해야 하는 상황에서 이 유사 사용자가 비슷한 상황에서 한 평가만을 반영하여 추천한다는 점에서 [2]의 방식과 다르다. 또한 [3]의 연구에서는 사용자가 같은 아이템을 선호한 상황을 유사 상황으로 정의하고, 유사한 상황에서 사용자가 선호한 아이템을 기반으로 비슷한 상황을 찾고, 추천 상황에서 아이템에 대한 사용자의 평가를 그 사용자가 추천상황과 유사한 상황에서 아이템에 대해 가진 선호를 계산하여 구한 뒤, 이 평가 값으로 협력적인 필터링을 하여, 유사 사용자를 찾고 이들의 평가를 반영하는 방법을 사용하였다.

## 3. 협력적 필터링과 상황인지를 사용한 추천 모델

본 논문에서 협력적 필터링과 상황인지 추천을 위한 사용자의 선호도는 사용자의 음악청취 기록을 기반으로 하여 추론된다. 사용자의 로그는 크게 사용자와 상황정보, 청취된 음악으로 이루어진다. 즉  $n_u$ 개의 사용자 로그 집합  $L$ ,  $n_u$ 명의 사용자 집합  $U$ ,  $n_s$ 개의 음악 집합  $S$ ,  $n_c$ 개의 상황 집합  $C$ 는 다음과 같이 정의될 수 있다.

$$L = \{l_1, l_2, \dots, l_y, \dots, l_{n_l}\}$$

$$U = \{u_1, u_2, \dots, u_j, \dots, u_{n_u}\}$$

$$S = \{s_1, s_2, \dots, s_i, \dots, s_{n_s}\}$$

$$C = \{ctx_1, ctx_2, \dots, ctx_k, \dots, ctx_{n_c}\}$$

이 때,  $(u, s, ctx)$ 으로 구성되는 각 로그  $l$ 은 다음과 같이 표현될 수 있다.

$$\{l \mid l \in L \wedge l.s = s_i \wedge l.u = u_j \wedge l.ctx = ctx_k\}$$

### 3.1 협력적 필터링을 이용한 추천 모델

협력적 필터링을 적용할 때, 각 사용자  $u_j$ 의 음악  $s_i$ 에 대한 평가  $r_{s_i, u_j}$ 는 전체상황을 고려하여 이루어지며 다음과 같다.

$$r_{s_i, u_j} = \left\{ l \mid l \in L \wedge l.s = s_i \wedge l.u = u_j \right\}$$

이 사용자의 음악에 대한 평가를 바탕으로 추론되는 사용자  $u_j$ 의 음악  $s_i$ 에 대한 예상 사용자 평가  $p_{s_i, u_a}$ 는 아래와 같다.

$$p_{s_i, u_a} = \bar{r}_{u_a} + \frac{1}{\alpha} \sum_{j=1}^{topk} (r_{s_i, u_j} - \bar{r}_{u_j}) \cdot \text{sim}(u_a, u_j)$$

이 때, 사용자 간의 유사도  $\text{sim}(u_a, u_j)$ 는 다음과 같이 코사인 유사도를 이용하여 구한다.

$$\text{sim}(u_a, u_j) = \frac{\sum_{i=1}^{n_s} r_{s_i, u_a} \times r_{s_i, u_j}}{\sqrt{\sum_{i=1}^{n_s} r_{s_i, u_a}^2} \sqrt{\sum_{i=1}^{n_s} r_{s_i, u_j}^2}}$$

$\alpha$ 는 곡 아이템 사이에 사용자간의 유사도 값의 차이를 보정하기 위한 값으로 다음과 같이 구한다.

$$\alpha = \sum_{j=1}^{topk} \text{sim}(u_a, u_j)$$

### 3.2 시간 상황 정보를 이용한 추천 모델

시간 상황 정보를 표현하는 방식은 시간을 계절, 월, 요일, 하루 중 때에 해당되는 개념으로 구분하고 각 로그의 해당 정도를 퍼지 멤버십 함수를 통해 계산하는 [1]의 방식을 따른다. [1]의 방식에서는 상황을 “월요일”, “여름”, “아침”과 같은 상황 개념으로 정의한 후 로그가 이에 해당하는 정도를 퍼지함수를 통하여 계산하였다.

상황  $ctx$ 에서 사용자  $u_j$ 의 곡 아이템  $s_i$ 에 대한 선호는 다음과 같이 표현된다.

$$r_{s_i, u_j, ctx} = \left\{ l \mid l \in L \wedge l.s = s_i \wedge l.u = u_j \wedge l.ctx = ctx \right\}$$

### 3.3 협력적 필터링과 상황인지를 사용한 추천 모델

협력적 필터링과 상황인지를 모두 고려한 추천 모델은 다음과 같다.

$$R: U \times S \rightarrow Preferences$$

$$p_{s_i, u_a, ctx} = \bar{r}_{u_a} + \frac{1}{\alpha} \sum_{j=1}^{topk} (r_{s_i, u_j, ctx} - \bar{r}_{u_j}) \cdot \text{sim}(u_a, u_j)$$

$$\alpha = \sum_{j=1}^{topk} \text{sim}(u_a, u_j)$$

이 때, 사용자간의 유사도는 3.1절에서 언급한 것과 같이 코사인 유사도를 사용한다.

## 4. 실험

### 4.1 실험 데이터 및 실험 방식

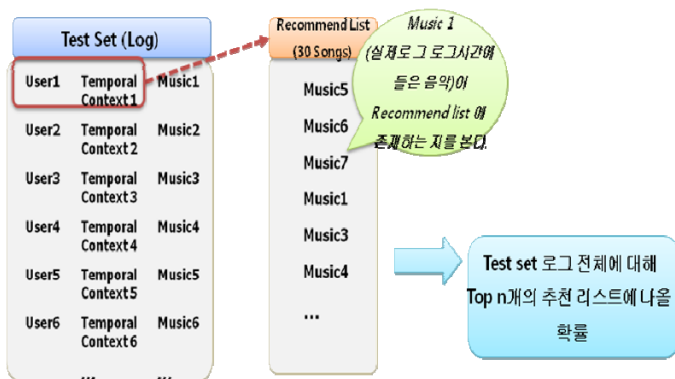
위 모델의 성능을 확인하고 [2]의 모델과 공정하게 비교하기 위하여 [2]과 동일한 데이터를 이용하여 실험하였다. 데이터의 상세정보는 다음과 같다.

표 1. 데이터 셋 상세정보

데이터 셋	SET1	SET2
곡수	10,000	3,000
청취 이력	2,421,362	1,357,137
사용자당 평균 청취 횟수	582.6회	335.1회
사용자당 평균 청취 곡수	207.6곡	113.1곡
곡당 평균 청취 횟수	2.8회	3.0회
곡당 청취자 수	242.1명	452.4명

본 실험에서는 각각의 데이터 셋 중 10%를 평균으로 랜덤으로 추출하고, 나머지 90%를 학습군으로 사용하였다.

실험 방식은 역시 공정한 비교를 위하여 [2]에서 제안한 HR@n을 사용하였다. 이 실험방식은 테스트 셋에 있는 각 로그에 n개의 아이템을 리스트로 추천한 후 이 리스트에 실제로 사용자가 들은 음악이 들어있는 지를 확인하는 방식으로 아래 그림과 같다.



### 4.2 실험 결과 및 분석

그림 1과 2에서 CA(Context Aware)는 현재 상황과 유사한 상황에서 가장 인기있는 음악을 추천한 경우이고, CF(Collaborative Filtering)은 상황에 상관없이 협력적 필터링을 이용하여 추천한 방식이다. GCF\_CA(Global Collaborative Filtering and Context Aware)는 본 논문에서 제안한 모델로 상황에 상관없이 전체 데이터를 대상으로 협력적 필터링을 해서 구해진 유사 사용자의 현재 상황에서의 선호 아이템을 추천하는 방식이다. CTXCF\_CA(CF in Context and Context aware)는 추천 상황에서 매번 유사 사용자를 구하여 이 유사 사용자의 선호 아이템을 추천하는 방식으로 3.1절의 예상 사용자 평가와 같은 방식이지만 사용자 유사도만 다음과 같은 방식으로 계산된다.

$$\text{sim}_{ctx}(u_a, u_j) = \frac{\sum_{i=1}^{ns} r_{s_i, u_a, ctx} \times r_{s_i, u_j, ctx}}{\sqrt{\sum_{i=1}^{ns} r_{s_i, u_a, ctx}^2} \sqrt{\sum_{i=1}^{ns} r_{s_i, u_j, ctx}^2}}$$

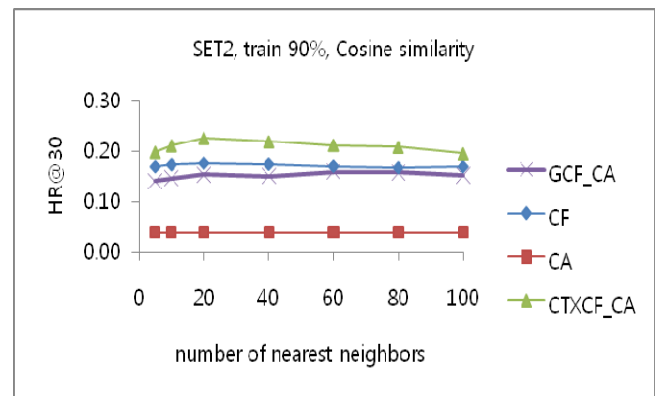


그림 1. 데이터 셋1에서 유사 사용자 수에 따른 성능 평가

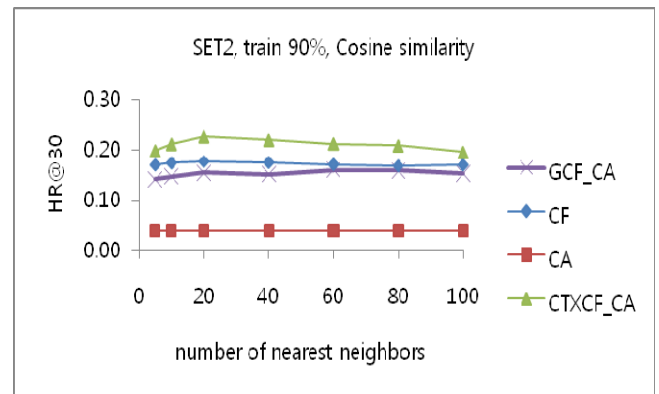


그림 2. 데이터 셋2에서 유사 사용자 수에 따른 성능 평가

GDF\_CA는 그림 1과 2에서 보는 것과 같이 HR@30의 실험 방식에서 유사 사용자의 수와 상관없이 협력적 필터링의 성능과 비슷하거나 더 낮은 성능을 보였다. 유사 사용자를 찾고 그 사용자의 아이템 선호를 고려하는 것을 추천 상황에 한정하는 모델인

CTXCF\_CA가 협력적 필터링보다 좋은 성능을 보이는 것을 감안하면 논문의 서론에서 제기하였던 두 가지 가정 중에, 일반적으로 선호가 비슷한 사용자들 사이에도 상황에 따라서 변화하는 취향이 다를 수 있기 때문에 각각의 상황에서 유사 사용자를 매번 계산해야 한다는 가정이 더 합리적이라고 판단할 수 있다.

반면, GCA\_CF의 성능 저하 요인으로, 추천 대상 사용자와 일반적으로 비슷한 선호를 가진 유사 사용자를 미리 구해놓기 때문에 이 유사 사용자의 아이템에 대한 평가가 추천 시점과 비슷한 상황에 반드시 존재한다는 보장이 없다는 사실을 말할 수 있다. 아래 표는 각 상황에서 일반적 상황에서의 유사 사용자의 청취기록 존재 확률을 데이터 셋2에서 실험한 결과이다.

표 2. 추천 상황에 청취기록을 가지는 일반적으로 유사한 취향의 사용자 수 및 확률

$k$	5	10	20	40	60	80	100
평균 유사 사용자 수(명)	3.9	8.2	16.8	34.3	51.8	69.3	86.7
유사 사용자의 청취기록 소유 확률(%)	78.7	81.9	84.1	85.6	86.3	86.6	86.7

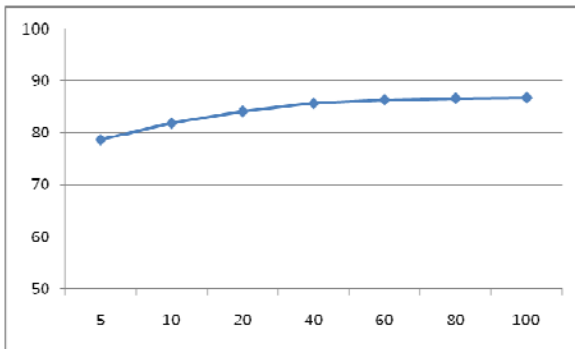


그림 3. 일반적으로 유사한 사용자가 추천 상황에 평가를 가질 확률 변화

표 1과 그림 3에서 보는 것과 같이 일반적인 상황에서 유사한 취향을 가진 사용자가 추천 시점에 로그를 가질 확률은 평균 84%이며, 추천 시 고려하는 유사 사용자 수가 증가함에 따라 조금씩 증가하는 경향을 보이지만 40명을 넘어서면 거의 일정해진다. 이는 16% 성능 저하의 한가지 요인으로 예상할 수 있다.

하지만 이 모델은  $k$ 가 작고, 추천하는 목록의 길이가 짧을 때 협력적 필터링보다 높은 성능을 보였다. 다음 표는  $k=5$ 이고 1개의 음악을 추천했을 때의 성능을 나타낸 것이다. 즉,  $k=5$ 일때 HR@1에서의 성능을 평가한 것이다.

표 3. HR@1에서의 성능평가

	상황인지	협력적 필터링	본 논문의 방식
SET1	0.08	0.11	0.14
SET2	0.03	0.14	0.18

이 결과는 사용자에게 여러 개의 추천 목록을 제시하는 것이 아니라 한 개의 아이템을 추천하는 경우가 많은 곳에서 유리하게 사용될 수 있음을 의미한다.

### 5. 결론

본 논문에서는 상황을 고려한 협력적 필터링으로 음악을 추천할 때 고려되지 않았던 가정에 따른 모델을 고려하고, 실험하여 보았다. 일반적으로는 특정 상황에서 유사한 사용자를 매번 구하여 추천하는 것이 추천 성능을 높일 수 있다는 것을 확인하였다. 또한, 본 모델은 몇 개의 음악만을 추천해야 하는 경우에 협력적 필터링과 상황인지의 발전된 모델로 사용할 수 있다.

### 6. 참고문헌

- [1] Dongmin Shin, Jae-won Lee, Jongheum Yeon, Sang-goo Lee, Context-Aware Recommendation by Aggregating User Context, IEEE Conference on Commerce and Enterprise Computing (CEC), page 423-430, 2009
- [2] 이동주, 이상근, 이상구, 시간 상황 정보를 고려한 협업 필터링을 이용한 음악 추천, 한국컴퓨터종합학술대회 논문집 Vol.36, No.1(C), 2009.
- [3] Chen, A. "Context-Aware Collaborative Filtering System: Predicting the User's Preference in Ubiquitous Computing". LoCA 2005, LNCS 3479, page 244-253, 2005.
- [4] <http://www.last.fm>
- [5] Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., and Riedl, J. "GroupLens: an open architecture for collaborative filtering of netnews." Proc. CSCW, 1994.
- [6] Cheung, K., Tian, L.F., "Learning User Similarity and Rating Style for Collaborative Recommendation.", Information Retrieval 7(3-4), page 395-410, 2004.
- [7] J. Ben Schafer, Dan Frankowski, Jon Herlocker and Shilad Sen, "Collaborative Filtering Recommender Systems", Proc. Conf. Am.Assoc. Artificial Intelligence (AAAI-99), page 439-446, July 1999. 2007.
- [8] Candillier, L., Meyer, F., Fessant, F. "Designing Specific Weighted Similarity Measures to Improve Collaborative Filtering Systems." Proc. ICDM 2008. LNCS (LNAI), vol. 5077, Springer, page 242-255, 2008.
- [9] K. Stefanidis and E. Pitoura. "Fast contextual preference scoring of database tuples." In EDBT '08, page 344-355, 2008.
- [10] L. Baltrunas. "Exploiting contextual information in recommender systems", In RecSys '08, page 295-298, 2008.