

사용자 프로파일 정보와 실시간 컨텍스트 정보를 이용한 협력적 필터링

Collaborative Filtering using User Profiles Information and Real-Time Context Information

이세일¹, 이상용²

¹ 충남 공주시 공주대학교 컴퓨터공학과
E-mail: lsilhr@kongju.ac.kr
² 충남 공주시 공주대학교 컴퓨터공학부
E-mail: sylee@kongju.ac.kr

요 약

추천시스템에서 가장 많이 사용하고 있는 협력적 필터링 방법을 모바일 기기 등에서 사용하려면 추천 정보와 사용자들의 평가 정보가 부족하여 추천의 질이 떨어지게 된다. 이러한 문제를 해결하기 위해 실시간으로 얻어진 컨텍스트 정보를 정량화하여 협력적 필터링에 적용함으로써 보다 나은 추천 결과를 얻을 수 있었다. 그럼에도 불구하고 평가를 하기 위한 컨텍스트 정보가 충분하지 못한 경우 부정확한 결과를 가져올 수 있다. 또한 사용자 정보 평가 과정 중 정량화 단계의 분류 과정을 단순히 하게 되면 서비스 받는 사용자가 정확한 그룹에 분류되어 정확도가 결여되는 문제가 발생한다.

본 논문에서는 실시간으로 얻을 수 있는 컨텍스트 정보가 부족한 경우, 내용 기반 필터링에서 많이 사용하고 있는 사용자 프로파일 정보를 실시간 컨텍스트 정보와 결합한다. 그리고 정량화 단계를 개선하여 협력적 필터링함으로써 기존의 방법보다 향상된 결과를 얻을 수 있다.

Key Words : Collaborative Filtering, Context, User Profile, Quantification, Recommendation System

1. 서 론

매우 많은 정보들 중 인터넷이나 모바일 기기를 이용하여 사용자에게 알맞게 서비스하는 시스템을 추천 시스템이라 한다. 이 추천 시스템은 여러 가지 방법이 존재하지만 그 중 가장 많이 이용하는 방법이 협력적 필터링이다. 그러나 모바일 기기에서 협력적 필터링을 이용하여 추천하려면 축적된 정보 부족과 사용자의 평가 정보를 얻기에 힘들다는 문제점을 가지고 있다. 이러한 문제점을 해결할 수 있는 방법으로 사용자의 실시간 컨텍스트 정보를 정량화하여 협력적 필터링에 적용함으로써 보다 나은 결과를 얻을 수 있었다[1]. 그럼에도 불구하고 다음과 같은 문제점이 발생한다. 첫째, 평가를 위한 컨텍스트 정보가 부족하여 서비스를 원하는 사용자에게 정확한 정보를 추천하기가 어렵다. 둘째, 정량화 단계의 분류 과정이 단순하다. 정량화 단계는 시간, 나이, 동행인, 남녀 등

이 있다[1]. 그 중 나이의 분류 과정을 10대, 20대처럼 단순하게 분류하게 되면 서비스 받는 사용자와의 나이차가 적음에도 가까운 나이 그룹에 분류되는 것이 아니라 나이차가 있는 부정확한 그룹에 분류되게 될 수도 있다. 예를 들어 29살의 서비스 사용자는 실제 30대 초반이나 20대 후반의 사용자들과 관련성이 있지만 20대의 나이 그룹에 분류되어 진다.

본 논문에서는 실시간으로 얻을 수 있는 컨텍스트 정보가 부족할 경우 사용자 프로파일과 컨텍스트 정보를 결합한다. 그리고 정량화 단계 중 나이의 분류 과정을 단순하게 분류하는 절대적인 방법이 아니라 서비스 받는 사용자를 기준으로 분류하는 상대적인 분류 방법을 사용하여 협력적 필터링에 적용함으로써 향상된 결과를 얻을 수 있었다.

본 논문의 구성은 2장에서 관련연구, 3장에서는 사용자 프로파일과 실시간 컨텍스트 정보를 이용한 추천 시스템, 4장에서는 실험 및 평

가한 내용을 기술하였다. 그리고 마지막으로 5 장에서는 결론에 대하여 언급한다.

2. 관련 연구

2.1 협력적 필터링

협력적 필터링은 방대한 양의 정보 중에서 각 사용자들이 평가하는 평가치를 근거로 하여 다른 사용자간의 유사성을 예측한다. 유사성을 예측하는 방법으로는 피어슨 상관 계수(Pearson Correlation Coefficient), 보안 코사인 유사도(Adjust Cosine Similarity) 방법 등이 있으나 가장 많이 사용하는 방법은 피어슨 상관 계수이다[2].

예를 들어 6명의 사용자가 5개의 항목에 평가를 하였다면 표 1과 같은 테이블이 만들어 질 수 있다.

표 1. 평가 항목

항목 \ 사용자	항목1	항목2	항목3	항목4	항목5
사용자1	3	5	2	1	4
사용자2		2	4	5	2
사용자3	2	3	5	4	
사용자4	5	5	3		3
사용자5	?	5	1	2	4
사용자6	1	3	1	5	3

사용자5의 항목1을 추천 받기 위해서는 5명의 사용자 중 가장 관련성이 있는 사용자1의 값을 참조하는 것이 좋다는 것을 알 수 있다.

2.2 사용자 프로파일

사용자 프로파일은 사용자로부터 설문조사나 질문하는 방법으로 정보를 얻을 수 있다. 이 프로파일은 사용자들이 관심 있는 부분이나 실제로 접하고 있는 내용을 근거로 사용자에게 적합한 검색 결과를 제공할 수 있다. 이와 같은 방법으로 얻게 되는 사용자 프로파일을 정적 프로파일(Static Profiles)이라고 하며 성별, 나이, 남녀, 주소 등이 있다. 또한 사용자의 현재 상황이나 피드백(Feedback)된 서비스 내용 등으로 얻을 수 있는 동적 프로파일(Dynamic Profiles)이 있다[3].

2.3 컨텍스트

컨텍스트(Context) 정의는 사용자가 처한 환경에서 사용자의 현재 위치, 행동 및 작업 등 사용자에게 대한 정보와 그 정보들의 지속적인 변화를 말한다. 컨텍스트는 활용분야에 따라 다양하며 사용자의 ID, 시간, 위치, 감정 등이 있다[4]. 이러한 컨텍스트를 얻어내는 과정을 컨텍스트 인식 기술(Context-Awareness Technique)이라고 한다.

2.4 정량화

실시간으로 얻어진 컨텍스트 정보는 사용자의 평가 정보를 얻을 수 없으므로 협력적 필터링과 같은 추천 시스템 방법에는 사용하기 힘들다. 그래서 이러한 컨텍스트 정보들을 숫자로 표시한 것을 정량화라 한다[1].

3. 사용자 프로파일과 실시간 컨텍스트 정보를 이용한 추천 시스템

다른 사용자로부터 모집된 실시간 컨텍스트 정보가 부족할 경우에는 사용자 프로파일 정보와 컨텍스트 정보를 통합하고 정량화 과정을 수행한다. 정량화 과정은 사용자가 평가한 정보와 같이 숫자로 만들기 위해서 필요한 단계이다. 그리고 정량화된 정보들을 이용하여 협력적 필터링함으로써 서비스 사용자에게 추천하며, 도식화하면 그림 1과 같다.

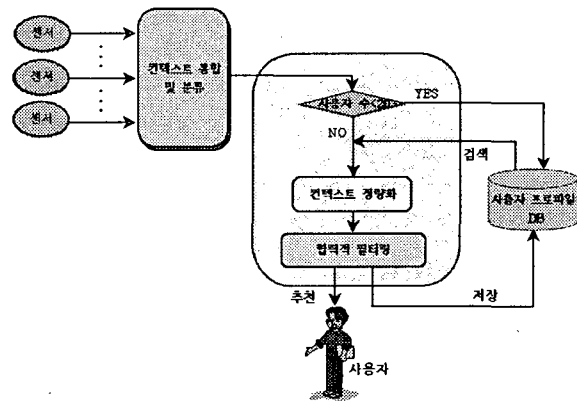


그림 1. 흐름도

다음은 실시간으로 모집된 컨텍스트 정보와 사용자 프로파일의 정보를 이용하여 사용자에게 추천하는 과정이다.

- 1) 센서들로부터 컨텍스트 정보를 모집하여 통합하고 분류한다.
- 2) 컨텍스트 정보의 개수가 20보다 작으면 아래 과정을 수행하며, 그렇지 않으면 3) 과정을 수행한다.

- 사용자 프로파일의 정보를 사용자 프로파일 DB로부터 검색한다.

- 3) 컨텍스트 정보를 정량화한다(컨텍스트 정보가 20미만이면 프로파일 정보와 통합하여 정량화함).

- 4) 협력적 필터링을 적용하여 서비스 받는 사용자에게 추천되며, 같은 정보는 사용자 프로파일에 피드백되어 저장된다.

3.1 컨텍스트 인식

서비스 받는 사용자가 모바일을 이용하여 원하는 서비스를 받으려면 다른 사용자들의 실시간 컨텍스트 정보를 필요로 한다. 컨텍스트 통합 및 분류 에이전트(Integration and Classification Agent)는 센서로부터 모집된 실시간 컨텍스트 정보들을 통합하고 서비스에 필요한 컨텍스트 정보만을 분류하면 표 2와 같다.

표 2. 서비스에 필요한 컨텍스트 정보

종류	내용
ID	사용자 구분
나이	서비스 사용자에게 의한 상대적인 분류
성별	남/여
장르	영화의 장르
제목	영화제목(추천)
장소	서비스 장소
시간	서비스 시간
동행인	의도(애인, 친구, 동료, ...)

또한 지역으로부터 모집된 사용자 수가 20보다 작은 컨텍스트 정보를 얻을 경우 서비스의 정확도가 떨어지므로 사용자 프로파일을 검색한다. 검색된 프로파일 정보는 실시간으로 모집된 컨텍스트 정보와 통합되어 정량화 단계로 이동한다. 검색된 사용자 프로파일의 정보는 사용자에게 서비스된 피드백된 정보들이다.

3.2 개선된 정량화

통합된 정보들 중 나이, 성별, 동행인, 서비스 시간을 이용하여 정량화하며, 사용되는 범위값은 1부터 5사이의 값을 갖는다. 특히 정량화 단계에서 사용하는 나이 분류는 절대적인 방법이 아니라 서비스 받는 사용자의 나이와 비교할 수 있는 상대적인 방법을 사용하였다.

일반적으로 사용되는 절대적인 나이 분류 방법은 표 3과 같다.

표 3. 절대적인 나이별 분류 방법

색인	나이분류	인원
1	~10	19
2	20~29	52
3	30~39	45
4	40~49	26
5	50~	8

이러한 방법은 서비스 사용자의 나이가 29살이면 30대 초반이나 20대 후반의 나이 그룹들의 사람들과 공감대가 형성될 수 있으나 무조건 20대에 편입되어 20대에 영향을 받는다. 이

러한 문제점을 해결하기 위해 서비스 받는 사용자의 나이를 기준으로 분류하는 상대적인 방법을 사용하면 좀 더 관련 있는 나이 그룹에 속하게 된다. 예를 들어 서비스 요구자의 나이가 29이면 표 4와 같이 상대적으로 분류된 것을 알 수 있다.

표 4. 상대적인 분류 방법

색인	나이분류	인원
1	~24	43
2	25~29~34	63
3	35~39~44	27
4	45~49~54	18
5	55~	6

서비스 받는 사용자 나이 그룹과 비교하여 가까울수록 높은 점수를 부여하며, 나이가 많이 날수록 낮은 점수를 부여한다. 그 이유는 나이에 따라 영화의 관심도가 다르기 때문이다. 또한, 성별, 동행인, 서비스 시간도 연관 관계에 따라 차등의 점수를 부여한다[1].

3.3 협력적 필터링에 적용

협력적 필터링은 사용자의 평가치를 기준으로 비슷한 선호도를 가지는 다른 고객들의 평가 항목이 어느 정도 존재하여야 예측할 수 있었다. 그러나 컨텍스트를 이용한 협력적 필터링 방법에서는 서비스된 동류의 항목이 하나만 있어도 다른 컨텍스트 정보들과 함께 이용하여 서비스 요구자에게 추천 서비스를 할 수 있다. 컨텍스트 정보를 이용한 유사도식은 아래 식 1과 같다.

$$w_{x,y} = \frac{\sum_{a=1}^n (r_{x,a} - \bar{r}_x)(r_{y,a} - \bar{r}_y)}{\sqrt{\sum_{a=1}^n (r_{x,a} - \bar{r}_x)^2} \sqrt{\sum_{a=1}^n (r_{y,a} - \bar{r}_y)^2}} \quad (1)$$

a는 사용자 x와 사용자 y가 평가한 컨텍스트 정보, \bar{r}_x 는 사용자 x의 컨텍스트 전체 정보에 대한 평가 평균값이다. $r_{x,a}$ 는 사용자 x가 컨텍스트 정보 a에 대하여 평가한 값이고, n은 컨텍스트 정보의 총 개수이다.

유사도 식과 이웃의 수를 사용하여 값을 예측하는데 그 예측값은 식 2와 같다.

$$P_{x,b} = \bar{r}_x + \frac{\sum_{y=1}^n w(x,y)(r_{y,b} - \bar{r}_y)}{\sum_{y=1}^n w(x,y)} \quad (2)$$

$P_{x,b}$ 는 사용자 x와 서비스 항목 b에 대한 선

호도를 예측한 값이고, \bar{r}_x 는 사용자 x 의 선호도 평균값이다. $w(x,y)$ 는 사용자 x 와 사용자 y 의 유사도 가중치이고, $r_{y,b}$ 는 사용자 y 가 서비스 항목 b 에 대하여 평가한 값이다. \bar{r}_y 는 사용자 y 의 선호도 평균값이고, n 은 결정된 이웃의 수이다.

정량화된 값에 위 식을 적용하면 서비스 요구자와 가장 유사성이 높은 사용자를 참조하여 서비스 정보를 추천하게 된다. 따라서 서비스 요구자는 질적으로 향상된 서비스를 제공 받을 수 있다.

4. 실험 및 평가

본 논문은 펜티엄 IV, 2.8Ghz, 512MB의 환경에서 C#과 J2ME, WIPI를 이용하여 설계하고 실험하였다.

서비스에 필요한 컨텍스트 정보는 나이, 동행인, 시간, 성별 등이다. 나이는 관련성이 많아서 정량화하기가 쉽지만 동행인, 시간 그리고 성별에 대한 관련성을 정량화하는 것은 쉽지 않기 때문에 427명에게 설문조사하여 관련성을 조사하였으며, 설문조사한 인원은 남자 241명과 여자 186명이다.

본 논문의 평가 식으로는 예측의 정확성을 평가하기 위하여 MAE(Mean Absolute Error)를 사용하였다. v_i 는 예측 선호도이며, r_i 는 실제 선호도이고, N 은 총 예측 회수이다.

$$|E| = \frac{\sum |v_i - r_i|}{N} \quad (3)$$

컨텍스트 정보를 이용한 협력적 필터링에 사용자 프로파일을 적용(CFP)한 결과는 기존의 GroupLen방식(CF)보다 월등하게 오차가 적음을 알 수 있었고, 컨텍스트 정보를 기반으로 SOM에 적용(CFS)[1]한 결과보다도 평균 0.0097정도 오차가 적음을 그림 2에서 알 수 있었다.

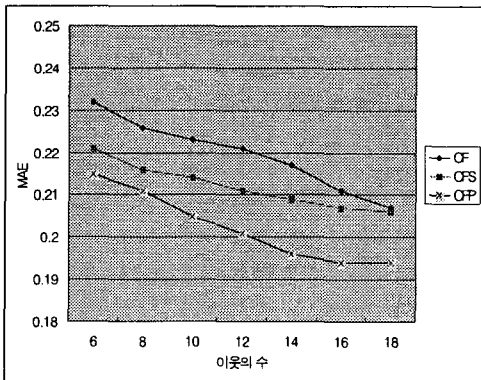


그림 2. 사용자 프로파일을 적용한 결과

정량화 단계의 나이 분류 작업을 상대적으로 분류(CFS_Age)한 방법은 CF보다 0.019 정도 오차가 적음을 알 수 있었고, CFS보다 평균 0.0016 정도 오차가 적음을 그림 3으로 알 수 있다.

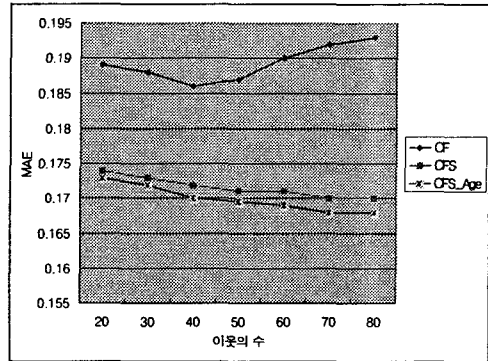


그림 3. 정량화 단계를 보완한 결과

5. 결론

모바일 기기에서 협력적 필터링을 적용할 경우 축적된 자료 부족과 평가 정보가 부족하였다. 이러한 문제점을 해결하기 위하여 실시간으로 모집된 컨텍스트 정보를 정량화하여 협력적 필터링함으로써 양질의 서비스를 얻을 수 있었다. 그러나 컨텍스트 자료가 부족할 경우 양질의 서비스가 어려운 문제점은 피드백된 정보들로 구성된 사용자 프로파일과 통합하여 해결하였다. 또한 정량화 과정의 단순화 문제는 나이 분류 과정을 서비스 받는 사용자의 나이를 기준으로 분류함으로써 정확하지 않은 그룹에 편입되는 것을 개선할 수 있었다.

제안된 방법을 실험한 결과 CFP는 CFS보다 4.6% 향상되었고, CFS_Age도 CFS보다 1.1%로 우수함을 알 수 있었다.

참고 문헌

- [1] 이세일, 이상용, "Pure P2P 환경에서 컨텍스트 정보를 이용한 실시간 서비스 추천 시스템", 한국퍼지 및 지능 시스템학회 논문지, vol.15, No.7, pp.887-892, 2005
- [2] N.Good, B. Schafer, J.Konstan, A. Borchers, B.Sarwar, J. Riedl, "Combining Collaborative filtering with personal Agents for Better Recommendation", AAAI/IAAI, pp.439-446, 1999
- [3] 김종원, "사용자 프로파일에 기초한 유튜브 뉴스 그룹 자동 결정 방법", 한국퍼지 및 지능 시스템학회 논문지, vol.14, No.2, pp.142-149, 2004
- [4] A.K.Dey, "Context-Aware Computing: The CyberDesk Project.", Proc. of the AAAI 1998 Spring Symposium on Intelligent Environments(AAAI Technical Report SS-98-02), pp.51-54, Mar 1998.