

# 시간적 변화를 고려한 사용자 프로파일과 컨텍스트 정보를 적용한 협력적 필터링

## Collaborative Filtering using User Profiles Considering Temporal Variation and Context Information

이세일<sup>1</sup>, 이상용<sup>2</sup>

<sup>1</sup> 충남 공주시 공주대학교 컴퓨터공학과

E-mail: lsilhr@kongju.ac.kr

<sup>2</sup> 충남 공주시 공주대학교 컴퓨터공학부(교신저자)

E-mail: sylee@kongju.ac.kr

### 요 약

유비쿼터스 환경의 추천 시스템에서는 협력적 필터링을 위하여 컨텍스트 정보를 사용하고 있으나, 컨텍스트 정보의 부족으로 인하여 추천 결과가 정확하지 않는 경우가 발생하게 된다. 이러한 문제를 해결하기 위하여 컨텍스트 정보와 더불어 사용자 프로파일의 서비스 이력 정보를 사용하였으나, 사용자 프로파일의 서비스 이력 정보는 시간이 지남에 따라 사용자의 기호가 변하거나 유행에 영향을 받을 수 있는 문제점이 있다. 또한 컨텍스트 정보와 사용자 프로파일의 서비스 이력 정보는 상황에 따라 적절히 연동하지 못하여 부정확한 예측을 할 수가 있다.

본 논문에서는 시간의 경과에 따라 사용자의 기호나 유행이 변하는 경우, 사용자 프로파일의 서비스 이력 정보들을 일률적인 값으로 적용하는 것이 아니라 시간에 따라 가중치를 달리 적용하는 방법을 사용하였다. 그리고 컨텍스트 정보와 사용자 프로파일의 서비스 이력 정보가 상황에 따라 적절히 연동하지 못하는 문제는 협력적 필터링하여 나온 결과에 컨텍스트 정보와 사용자 프로파일의 서비스 이력 정보의 가중치를 달리 적용하여 통합함으로써 예측성을 높일 수 있었다.

**Key Words** : Ubiquitous, Collaborative Filtering, Context, User Profile

### 1. 서 론

유비쿼터스 환경에서는 각종 센서 네트워크들을 이용하게 되면서 컨텍스트 정보가 빠르게 지속적으로 증가하게 된다[1]. 이와 같은 상황에서 수시로 변화하는 컨텍스트 정보들을 이용하기 위하여 컨텍스트의 통합과 분류과정이 필요하다. 모바일 기기는 필터링된 컨텍스트 정보들을 다양한 방법으로 사용자에게 서비스를 제공하고 있다. 그러나 많은 컨텍스트 정보들이 존재하더라도 서비스 추천에 적합한 컨텍스트 정보가 부족한 경우가 발생하게 된다. 이러한 문제를 해결하기 위해서는 실시간으로 얻어지는 컨텍스트 정보와 사용자의 이력 정보가 저장되어 있는 프로파일 정보를 이용하여 협력적 필터링 함으로써 사용자에게 적절한 서비스를 해줄 수 있었다[2]. 그러나 사용자 프로파일의 서비스 이력 정보는 시간이 지남에 따라 사

용자의 기호나 유행에 많은 영향을 받을 수 있다. 또한 컨텍스트 정보와 사용자 프로파일의 서비스 이력 정보가 상황에 따라 적절히 연동하지 못하는 문제점을 가지고 있다.

본 논문에서는 사용자의 서비스 이력 정보가 시간의 경과에 따라 사용자의 기호나 유행이 변하는 경우 사용자 프로파일의 서비스 이력 정보들을 적용할 때 가중치를 달리 하였다. 그리고 컨텍스트 정보와 사용자 프로파일의 서비스 이력 정보가 상황에 따라 적절히 연동하지 못하는 문제는 사용자 프로파일의 서비스 이력 정보와 실시간 컨텍스트 정보를 협력적 필터링하여 나온 결과에 가중치를 달리 적용하여 통합함으로써 예측성을 높일 수 있었다.

본 논문의 구성은 2장에서 관련 연구, 3장에서는 시간적 변화를 고려한 서비스 추천 시스템, 4장에서는 실험 및 평가한 내용을 기술하였고 5장에서는 결론에 대하여 기술한다.

## 2. 관련 연구

### 2.1 협력적 필터링

협력적 필터링은 방대한 양의 정보 중에서 특정 항목에 대해 서비스를 원하는 사용자의 평가 값을 알고자 할 때, 서비스를 원하는 사용자와 유사성이 높은 경향을 보이는 이웃을 찾아 선호도가 높은 항목을 추천하는 방법이다[3]. 협력적 필터링 과정 중 유사성을 측정하는 방법으로는 피어슨 상관 계수(Pearson Correlation Coefficient), 보안 코사인 유사도(Adjust Cosine Similarity) 방법 등이 있다. 또한 유사성이 높은 이웃을 선정하는 방법에는 사용자간의 유사도 값이 정해진 기준 이상인 값들의 이웃들만 선정하여 예측하는 Thresholding 방법[4]과 서비스를 원하는 사용자와 유사성이 높은 n명을 이웃으로 사용하여 예측하는 Best n-neighborhood 방법[5]이 있다.

### 2.2 컨텍스트

컨텍스트(Context)는 사용자가 처한 환경에 대한 정보를 말한다. 컨텍스트는 활용분야에 따라 사용자의 상황, 물리적 환경 상황, 컴퓨팅 시스템 상황, 사용자-컴퓨터 상호 작용 이력, 가용 자원 상황, 기타 미분류 상황으로 분류한다[6]. 또한 이러한 컨텍스트를 얻어내는 과정을 컨텍스트 인식 기술(Context-Awareness Technique)이라고 한다. 컨텍스트 인식을 지원하는 구조는 Context-Toolkit[7], SOCAM[8] 등이 있다.

### 2.3 사용자 프로파일

사용자 프로파일은 사용자에게 직접 질문하여 정보를 얻어내는 방법이다. 사용자들이 관심 있는 부분이나 실제로 접하고 있는 내용을 참고하여 적합한 서비스 결과를 제공할 수 있다. 사용자 프로파일의 종류 중 정적 프로파일(Static Profiles) 방법은 이름, 성별, 주소, 전화번호, 나이 직업 등과 같이 직접 사용자로부터 얻어낼 수 있는 방법이 있다. 그리고 동적 프로파일(Dynamic Profiles) 방법은 추천 시간, 이동경로, 서비스 정보 등이 있으며 사용자의 이동 패턴이나 서비스된 정보 등으로 얻어낼 수 있다[9].

## 3. 시간적 변화를 고려한 서비스 추천 시스템

센서로부터 수집된 실시간 컨텍스트 정보들은 컨텍스트 인식 및 통합 모듈에서 필요한 컨텍스트 정보들만 분류되어 다음 단계로 이동하

게 된다. 시간적으로 변하는 사용자의 취향이나 유행들을 고려하여 가중치를 달리 결정하였다. 상대적 정량화 단계에서는 컨텍스트 정보들이 연산에 사용할 수 있도록 사용자들 기준으로 정량화한 것을 이용하여 협력적 필터링함으로써 사용자에게 적합한 서비스를 추천하게 된다.

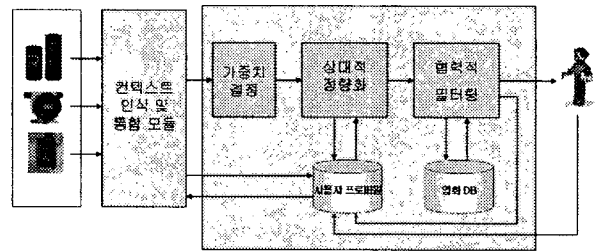


그림 1. 시스템 구성도

그림 1은 컨텍스트 정보와 사용자 프로파일의 서비스 이력 정보를 이용하여 사용자에게 적합한 서비스를 추천하는 과정이며 순서는 다음과 같다.

- ① 실시간 컨텍스트 정보는 컨텍스트 인식 및 모듈 단계에서 필요한 컨텍스트 정보만 분류된다.
- ② 사용자 프로파일의 서비스 이력 정보는 시간에 따라 가중치를 달리 적용한다.
- ③ 상대적 정량화 단계에서는 컨텍스트 정보와 사용자 프로파일의 서비스 이력 정보를 연산에 사용할 수 있도록 사용자들 기준으로 상대적 정량화하여 수치화한다.
- ④ 협력적 필터링 과정에서는 실시간 컨텍스트 정보와 사용자 프로파일의 서비스 이력 정보를 각각 협력적 필터링하여 나온 결과에 가중치를 달리 적용하여 통합함으로써 사용자에게 알맞은 서비스를 한다.

### 3.1 컨텍스트 인식 및 통합

부족한 컨텍스트를 해결하기 위해서는 다른 사용자들의 실시간 컨텍스트 정보가 있어야 한다. 컨텍스트 인식 및 통합 과정에서는 다양한 장치로부터 실시간으로 획득한 사용자 컨텍스트 정보를 사용하기 위하여 인식 및 분류 작업을 한다.

컨텍스트 인식 단계는 외부의 센서를 통하여 컨텍스트 정보를 획득하게 되며 또한 사용자 프로파일 DB를 읽어 사용자를 파악하고 전화번호의 주소록의 그룹명을 확인하여 나와 주변 인물의 관계도 파악할 수 있다.

컨텍스트 통합 단계에서는 불필요한 컨텍스트 정보를 필터링하게 된다. 또한 컨텍스트 정

보의 부족한 부분을 해결하기 위하여 Pure P2P 방법을 이용하여 다른 사용자들의 컨텍스트 정보를 수집한다. 본 논문은 영화관에서 사용자의 동행인이나 관람시간 등을 고려하여 영화를 추천한다. 이에 필요한 컨텍스트 정보는 표 1과 같다.

표 1. 서비스에 필요한 컨텍스트 정보

종 류	내 용
ID	사용자 전화번호
나이	나이(상대적 분류)
동행인	친구, 애인, 등으로 분류
직업	직업구분
성별	남/여
장르	영화장르
제목	영화제목
장소	서비스 장소
시간	서비스 시간(영화 상영 시간)

### 3.2 시간적 변화를 고려한 사용자 프로파일

사용자 프로파일의 정보들 중 서비스 이력 정보들을 활용하기 위해서 서비스에 사용되었던 컨텍스트 정보와 서비스 품목을 저장하게 된다.

사용자의 피드백된 정보를 활용하기 위해서는 서비스 되었던 상황과 서비스된 제목을 알아야 한다. 그러나 예전에 서비스된 상황과 서비스되는 현재의 시점이 모두 일치하더라도 사용자의 취향은 변할 수 있으므로 시간적 변화를 고려하여 가중치를 달리 적용한다. 예를 들면 표 2와 같이 적용할 수 있다.

표 2. 시간적 변화를 고려한 가중치 적용

시 간	가중치 비율
1년 이상	10%
1년 이내	20%
9개월 이내	30%
6개월 이내	40%
3개월 이내	50%

시간의 경과에 따라 사용자의 기호가 변하는 문제점을 해결하기 위해 사용자 프로파일의 서비스 이력 정보들을 시간에 따라 달리 평가 값을 적용하였다.

그리고 컨텍스트 정보와 사용자 프로파일의 서비스 이력 정보는 협력적 필터링에 사용할 수 있도록 상대적 정량화[2] 과정을 거쳐 수치화 된다. 수치화된 정보는 협력적 필터링에 사

용하기에 알맞은 정보가 된다.

### 3.3 협력적 필터링을 이용한 서비스 추천

사용자에게 적합한 서비스를 위해 수치화된 컨텍스트 정보와 사용자 프로파일의 서비스 정보를 이용하여 협력적 필터링에 적용한다. 유사도를 구하는 공식은 피어슨 상관 계수 (Pearson Correlation Coefficient)를 사용하며 식 (1)과 같다.

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

$W_{x,y}$ 는 컨텍스트 정보 x와 컨텍스트 정보 y의 유사도 값이다. 이웃의 수를 결정하는 방법으로는 Best n-neighborhood 방법을 이용하여 유사성이 높은 상위 50%의 이웃만 선정하였다. 그 이웃을 사용하여 예측하는데 그 예측값은 식 (2)와 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 컨텍스트 b에 대한 선호도를 예측한 값이다.

컨텍스트 정보를 이용하여 구한 값은 c이고 사용자 프로파일의 서비스 이력 정보를 이용하여 구한 값을 d로 정하여 식 (3)과 같은 값을 구한다. 식 (3)의 k는 최근에 본 영화의 종류들 중 30%를 선택하여 기간의 평균을 구하고 그것의 가중치를 백분율로 결정한 값이다.

$$f = (c*(1-k)+d*k) \quad (3)$$

식 (3)처럼 가중치를 적용하고 통합함으로써 실시간 컨텍스트 정보와 사용자 프로파일의 서비스 이력 정보가 상황에 따라 적절히 연동하지 못하는 문제를 해결할 수 있었다.

## 4. 실험 및 평가

본 논문은 펜티엄 IV, 2.8Ghz, 512MB의 환경에서 MS Visual Studio 2005, C#, .NET Framework 2.0을 이용하여 설계하고 실험하였다.

설문조사 방법으로 총 427명(남자 421명과 여자 186명)을 조사하여 관련성을 조사하였다. 서비스에 사용된 컨텍스트 정보는 나이, 동행인, 시간, 성별 등이다. 나이는 서비스가 필요한 사용자와 다른 사용자들을 비교하여 분리하

는 상대적인 방법을 사용하였고, 동행인의 경우에는 휴대폰 전화부에 등록된 그룹의 전화번호로 관련성을 조사하였다. 또한 사용자가 영화 보는 시간에 따라 취향이 변하는 것을 확인할 수 있었고, 성별에 따라 영화를 보는 취향의 차이보다는 동행인이 누구이고 어느 시간 때에 영화를 관람하느냐에 따라 많이 달라지는 것을 알 수 있었다. 정량화 단계 이후 사용되는 선호도 값은 1, 2, 3, 4, 5 단계로 표현하였다.

본 논문의 평가 방법으로는 MAE(Mean Absolute Error)를 이용하였다. 식 (4)는 실제 사용자와 예측 사용자와의 예측 값의 차이를 나타낸다.  $r_i$ 는 실제 선호도이고  $v_i$ 는 예측 선호도이며 N은 총 예측 회수이다.

$$|E| = \frac{\sum |v_i - r_i|}{N} \quad (4)$$

본 논문에서 제한한 시스템을 시험한 결과 사용자 프로파일과 컨텍스트 정보를 통합하지 않고 적용한 방법(CF\_UC), 프로파일과 컨텍스트 정보를 통합하고 적용한 결과(CF\_UCI) 그리고 시간적 변화를 고려한 사용자 프로파일과 컨텍스트 정보를 통합하여 적용한 결과(CF\_TUCI)를 가지고 각각 실험하였다. 실험 결과 CF\_TUCI 방법이 다른 두 방법보다 우수함을 알 수 있었다.

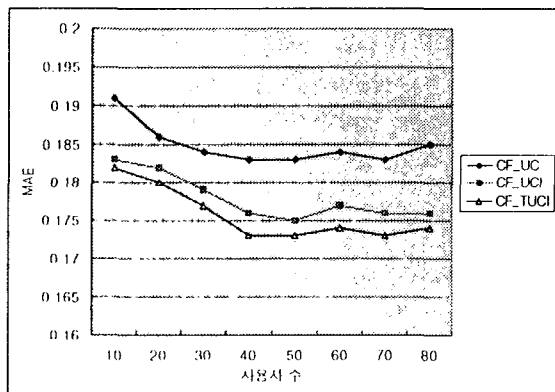


그림 2. 사용자 수에 따른 성능 비교

## 5. 결론

유비쿼터스 환경에서는 각종 센서 네트워크들을 이용하게 되면서 컨텍스트 정보의 양이 많아지고는 있으나, 서비스를 추천하기 위한 상황에서 적합한 컨텍스트 정보가 부족할 경우에는 컨텍스트 정보와 사용자 프로파일의 서비스 이력 정보를 결합하여 사용하였다. 그러나 사용자 프로파일의 서비스 이력 정보는 사용자의 기호나 유행이 시간에 따라 변화를 가져오는 문제점이 있다. 이러한 문제점은 사용자 프

로파일의 서비스 이력 정보를 시간적 변화에 따라 가중치를 달리 적용하여 해결하였다. 또한 컨텍스트 정보와 사용자 프로파일의 서비스 이력 정보가 상황에 따라 적절히 연동하지 못하는 문제는 컨텍스트 정보와 사용자 프로파일의 서비스 이력 정보를 각각 협력적 필터링하여 나온 결과에 가중치를 달리 적용하고 통합함으로써 개선할 수 있었다.

본 논문에서 제한한 시스템은 사용자 프로파일과 컨텍스트 정보(CF\_UC)를 통합하지 않는 방법보다 평균적으로 4.9% 우수하였다.

## 참 고 문 헌

- [1] J. M. Hellerstein, W. Hong and S. R. Madden, The Sensor Spectrum: Technology, Trends, and Requirements. ACM SIGMOD Record, Vol.32, No.4, pp.22-27, 2003
- [2] 이세일, 이상용, "사용자 프로파일 정보와 컨텍스트 정보를 이용한 협력적 필터링", 한국 퍼지 및 지능 시스템학회 논문지, vol.16, No.1, pp.336-339, 2006
- [3] 이기현, 고병진, 조근식, "연관 규칙과 협력적 여과 방식을 이용한 추천 시스템", 한국 지능정보 시스템학회 논문지, 제8권, 제2호, pp. 91-103, 2002
- [4] P. Resnick, et. al., "GroupLens: An Open Architecture for Collaborative Filtering of Netnews", Proc. of ACM CSCW'94 Conference on Computer Supported Cooperative Work, pp.175-186, 1994
- [5] J. S. Breese and D. Heckerman and C. Kadie, "Empirical Analysis of predictive Algorithms for Collaborative Filtering", Proc. of the 14th Conference on Uncertainty in Artificial Intelligence, 1998
- [6] 윤효근, "모바일 P2P 환경에서 실시간 컨텍스트 인식 및 서비스를 위한 다중 에이전트 시스템", 박사학위 논문, 2006
- [7] A.K.Dey and G.D.Abowd, "Towards a Better Understanding of Context and Context-Awareness", Gvu Technical Report GIT-Gvu -99-22. Submitted to the 1st International Symposium on Handheld and Ubiquitous Computing (HUC '99), June 1999
- [8] T. Gu, H. K. Pung, D. Q. Zhang. Towards an OSGi-Based Infrastructure for Context-Aware Applications in Smart Homes. IEEE Pervasive Computing, Vol. 3, Issue 4, 2004.
- [9] 김종원, "사용자 프로파일에 기초한 유즈넷 뉴스그룹 자동 결정 방법", 한국퍼지 및 지능 시스템학회 논문지, vol.14, No.2, pp.142-149, 2004