

컨텍스트 기반 협력적 필터링을 이용한 추천 시스템

A Recommendation System using Context-based Collaborative Filtering

이세일 · 이상용*
Se-Il Lee and Sang-Yong Lee*

공주대학교 컴퓨터공학부

요 약

협력적 필터링은 잠재적인 항목을 추천할 수 있어서 추천시스템에 가장 많이 사용되고 있다. 그러나 협력적 필터링은 평가 항목이 적을 경우, 평가자의 상황이나 기분에 따라 유사도나 선호도에 큰 영향을 끼칠 수 있다. 또한 사용자의 현재 상황을 전혀 고려하지 않고 과거에 평가한 항목만으로 유사도를 계산하여 추천하여 추천의 정확도가 떨어지게 된다. 본 논문에서는 위와 같은 문제점을 해결하기 위해, 먼저 협력적 필터링 과정을 수행하기 전 사용자가 평가한 모든 값을 비교하지 않고 평균 이상인 사용자들만을 비교하여 유사도를 계산함으로써 추천의 정확성을 높였다. 또한 끊임없이 변화하는 유비쿼터스 컴퓨팅 환경에서 사용자의 평가 항목만으로 서비스 정보를 추천하는 것이 적합하지 않기 때문에, 사용자의 실시간 컨텍스트 정보를 이용하여 비슷한 사용자들에게 높은 가중치를 적용하여 유사도를 구하는 방법을 사용하였다. 이러한 방법을 사용한 결과, 추천의 정확도가 평균적으로 16.2% 향상되었다.

키워드 : 협력적 필터링, 추천 시스템, 컨텍스트 정보, 선호도

Abstract

Collaborative filtering is used the most for recommendation systems because it can recommend potential items. However, when there are not many items to be evaluated, collaborative filtering can be subject to the influence of similarity or preference depending on the situation or the whim of the evaluator. In addition, by recommending items only on the basis of similarity with items that have been evaluated previously without relation to the present situation of the user, the recommendations become less accurate.

In this paper, in order to solve the above problems, before starting the collaborative filtering procedure, we calculated similarity not by comparing all the values evaluated by users but rather by comparing only those users who were above the average in order to improve the accuracy of the recommendations. In addition, in the ceaselessly changing ubiquitous computing environment, it is not proper to recommend service information based only on the items evaluated by users. Therefore, we used methods of calculating similarity wherein the users' real time context information was used and a high weight was assigned to similar users. Such methods improved the recommendation accuracy by 16.2% on average.

Key Words : Collaborative Filtering, Recommendation System, Context Information, Preference

1. 서 론

컴퓨터와 인터넷의 결합은 전자상거래의 발전에 초석이 되었다. 전자상거래 방법은 많은 기업들의 온라인 시장에 참여로 다양한 상품들이 출현하게 되었다. 또한 전자상거래 이용자들은 많은 상품들 중 개인의 취향이나 요구 사항에 적합한 상품을 선택할 수 있는 환경을 원하였다. 기업들은 이러한 요구 사항을 받아들이기 위해 추천 시스템을 개발하게 되었다.

추천 시스템은 사용자의 취향이나 요구 사항에 따라 적합한 상품을 추천하는 방법이다. 추천 시스템의 종류로는

협력적 필터링 방법과 내용 기반 필터링 방법 등이 사용되고 있지만, 그 중 협력적 필터링 방법을 가장 많이 사용하고 있다. 협력적 필터링 방법은 많은 항목과 사용자들의 평가 자료를 이용하여 비슷한 선호도를 가지는 항목을 추천하는 방법이다.

협력적 필터링 방법은 잠재적인 항목을 추천할 수 있어 가장 많이 사용되고 있지만 다음과 같은 단점을 가지고 있다. 첫 번째 문제점은 사용자들의 평가 자료가 현재 상황이나 기분에 따라 똑같은 것을 평가하더라도 평가 자료가 달라 질 수 있기 때문에 적은 자료만으로 평가하는 것은 부정확성을 초래한다. 두 번째 문제점은 협력적 필터링 방법이 사용자의 선호도만을 사용하고 있으므로 선호도 이외에는 전혀 다른 사항을 고려하지 않는다는 점이다. 이것은 다른 사항을 고려하지 않으므로 간단하게 평가 한다고 생각할 수 있지만, 유비쿼터스 환경에서 사용자의 처한 상황을 전혀

접수일자 : 2010년 9월 27일

완료일자 : 2011년 4월 15일

* 교신저자

고려하지 않고 평가 자료만으로 항목을 추천한다는 것은 서비스 추천에 영향을 준다.

이러한 문제점들을 해결하기 하기 위해 본 논문에서는 협력적 필터링 과정을 수행하기 전 정확성을 높이기 위해 사용자들이 평가한 모든 값을 비교하지 않고 평균 이상인 사용자만 비교하여 유사도를 계산하는 방법으로 안정성을 높였다. 또한 끊임없이 변화하는 유비쿼터스 컴퓨팅 환경에서 사용자의 평가 자료만으로 사용자에게 서비스 정보를 추천하는 것은 적합하지 않다. 현재 사용자의 정보나 상황정보를 비교하여 비슷한 사용자들에게는 높은 가중치를 적용하는 방법으로 유사도를 구한 후 추천하는 방법을 사용하였다.

본 논문의 구성은 2장에서 관련연구, 3장에서는 컨텍스트 기반 추천 시스템, 4장에서는 실험 및 평가한 내용을 기술하였다. 그리고 마지막으로 5장에서는 결론에 대하여 언급한다.

2. 관련 연구

2.1 추천 시스템

추천시스템(Recommender Systems)은 사용자의 입력 정보나 구매 이력, 웹 사이트의 히스토리 정보를 이용하여 사용자에게 적합한 항목을 추천해주는 시스템이다. 또한 추천으로 끝나는 것이 아니라 사용자의 구매를 유도 하거나 교차판매 증대에 도움을 주고 있다. Amazon, CDNOW, 옥션, Yes24 등과 같은 기업들도 추천 시스템을 사용하여 사용자들에게 편의를 제공하고 있다. 추천 시스템에 많이 이용되고 있는 기법으로는 협력적 필터링(Collaborative Filtering)과 내용 기반 필터링(Content-based Filtering) 등이 있다[1].

내용 기반 필터링은 과거에 사용자가 선호했던 항목과 비슷한 항목을 찾아내어 사용자에게 추천 해 주는 방식이다. 장점으로는 항목만을 사용하기 때문에 단순하다는 점이다. 단점은 사용자가 이전에 경험한 상품만을 추천해야 한다는 점이다[2].

협력적 필터링(Collaborative Filtering)은 추천 시스템 기법 중 가장 많이 이용되고 있다. 이 기법은 다수의 사용자와 다수의 항목을 이용하여 이웃들과 유사성이 높게 보이는 이웃을 찾아 선호도가 높은 항목을 추천하는 방법이다[3].

예를 들어 5명의 사용자와 5개의 항목에 대하여 평가 자료를 1점부터 5점까지 부여한다면, 표 1과 같은 테이블이 완성될 것이다.

표 1. 평가 항목
Table 1. Evaluation Item

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

사용자5의 빈 항목인 항목5를 서비스 받기 위해서는 다른 사용자들과 유사성을 비교한다. 비교 결과 사용자5의 빈 부분은 가장 유사도가 높은 사용자1을 참조하는 것이 좋다는 것을 알 수 있다.

협력적 필터링의 장점은 데이터 양이 충분한 경우 예측력이 높다. 또한, 구매한 상품이 전혀 다르더라도 추천이 가능하다는 점이다. 단점으로는 평가한 자료가 없는 초기에는 평가할 자료가 부족한 초기 평가 문제(Early Rate Problem)가 있다[4]. 많은 사용자와 많은 항목들이 방대하게 퍼져 있는 자료의 희소성 문제(Sparsity Problem)와 방대한 자료들로 인한 연산에 부담을 주는 확장성 문제(Scalability Problem)가 있다[5]. 그리고 사용자의 프로파일 정보나 실시간 컨텍스트 정보를 활용하지 못하는 문제점을 가지고 있다. 초기 평가 문제를 해결하기 위해 내용 기반 필터링 기법 등과 결합하여 해결하고, 희소성 문제를 해결하기 위해 clustering을 이용하였다. 확장 문제는 베이지안 기법 등 다양한 방법으로 개선하고 있다. 그러나 실시간 컨텍스트 정보를 사용하지 않을 경우 현재의 상황을 고려하지 않고 추천하면 정확성이 낮아질 수 있다[6-8].

2.2 컨텍스트

사용자가 처한 환경에서 사용자의 현재 위치, 사물들, 주변 사람과 그들이 일어나는 변화를 Schilit와 Theimer은 컨텍스트라고 정의 했고, Dey는 정서, 장소, 날짜와 시간, 객체, 위치, 사용자의 환경 안에 있는 사람과 사물 등으로 어떤 존재의 상황을 특징 질 수 있는 정보라고 정의 하였다. 그 외 다수의 연구자들은 조금씩 다르게 표현하고 있다. 컨텍스트는 활용분야에 따라 다양하며 사용자의 ID, 시간, 위치, 감정 등이 있다. 컨텍스트를 분야별로 구분하면 컴퓨팅 시스템 상황, 사용자의 상황, 물리적 환경 상황, 컴퓨터와 사용자 상호 이력 상황 그리고 기타 상황 등으로 나눌 수 있다. 이 또한 연구자별로 다르게 표현하고 있다[9].

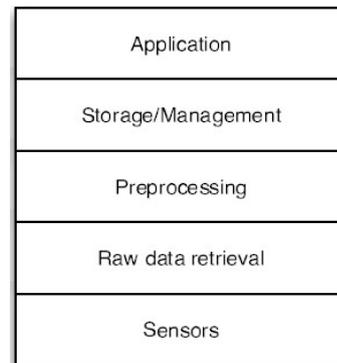


그림 1. 컨텍스트 인식을 위한 계층 개념 체계
Fig. 1. Layered Conceptual Framework for Context-Aware Systems

이러한 주변 환경으로부터 상황정보를 얻어내는 과정을 컨텍스트-인식 기술이라 한다. 컨텍스트-인식 기술을 이용한 시스템은 사용자 주변의 센서를 통해 사용자의 컨텍스트를 파악하고, 인식된 컨텍스트 정보에 따라 적절한 서비스를 제공하는 시스템과 컨텍스트에 맞추어 시스템의 실행 조건이나 주변 환경 등을 스스로 변경하는 시스템으로 나뉜다[10]. 이러한 시스템들은 그림 1과 같은 공통 구조의 계층 구조를 가지고 있다. Sensor Layer에서는 주변의 센서 장

치로부터 온도, 소리, 위치, 행동 등의 컨텍스트 정보를 얻는다. Raw Data Retrieval Layer에서는 외부의 센서들로부터 얻은 정보들을 가공하는 역할을 담당하며, Preprocessing Layer에서는 raw 레벨의 정보들을 애플리케이션에 사용할 수 있도록 가공하는 역할을 담당하고 있다. Store/Management Layer에서는 데이터들을 조직화하고 클라이언트에게 제공하는 역할을 담당하며, 클라이언트는 Application Layer에서 실현된다[11].

3. 컨텍스트 기반 추천 시스템

협력적 필터링에서는 사용자가 서비스 받은 항목을 평가하여 Evaluation 데이터베이스에 저장한다. 이 자료들은 사용자에게 서비스 정보를 추천할 경우 다른 항목들과 연관되어 사용된다. 본 논문에서는 협력적 필터링의 전처리 단계에서 사용자 평가 자료의 처리문제와 실시간 컨텍스트 정보를 이용한 협력적 필터링의 유사도 연산에 가중치 적용을 사용하여 추천하는 방법을 사용한다. 이 시스템은 7개의 처리 모듈과 5개의 데이터베이스로 구성되어 있다.

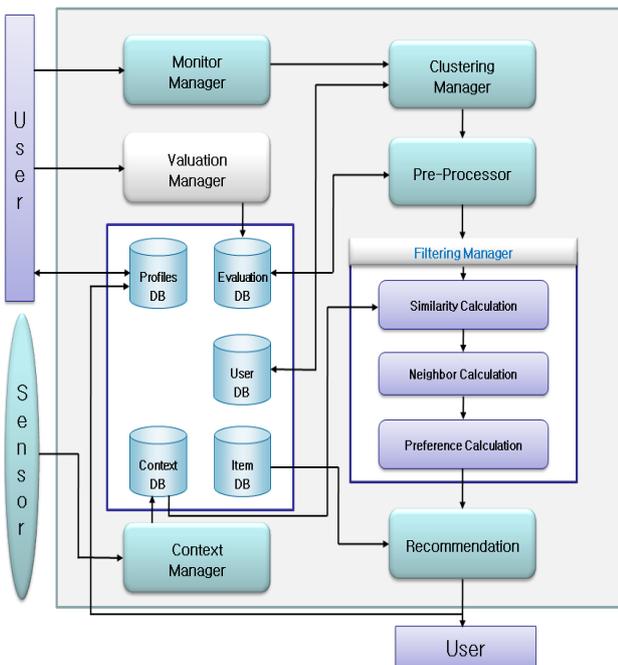


그림 2. 시스템 구조
Fig. 2. System Structure

이러한 시스템을 도식하면 그림 2와 같고, 과정을 나열하면 다음과 같다.

- ① 추천된 서비스를 마친 사용자는 Valuation Manager의 도움을 받아 사용자 만족도에 따라 평가 점수를 입력하여 데이터베이스에 저장한다. 이 과정은 자신과 다른 사람들의 정확한 추천 서비스를 위해 신중히 입력해야 한다.
- ② Monitor Manager는 사용자로부터 서비스 요청을 받을 경우, 이 정보를 Clustering Manager에 전달하는 역할을 한다.

- ③ Clustering Manager에서는 사용자의 프로파일 정보를 이용하여 분류하는 상대적 분류방법을 사용한다.
- ④ Pre-Processor에서는 사용자의 평가 자료가 적은 경우 연산과정의 문제점이 발생하므로 평균 이상인 자료들만 검색하여 협력적 필터링의 연산에 사용할 수 있도록 추출한다.
- ⑤ Filtering Manager에서는 상관계수를 이용하여 유사도를 구한다. 유사도 과정은 사용자의 현재 상황을 파악하여 컨텍스트 정보와 일치하는 사용자에게 가중치를 적용하는 방법을 사용한다.
- ⑥ Recommendation에서는 Filtering Manager 단계에서 계산된 선호도를 이용하여 사용자에게 알맞은 서비스를 한다. 또한 서비스된 정보는 프로파일 데이터베이스에 저장한다.

3.1 사용자 평가 및 모니터링 단계

Valuation Manager는 사용자로부터 평가 자료를 받아 평가 데이터베이스 저장할 수 있도록 관리하는 모듈이다. 사용자 평가 과정은 자신의 실시간 컨텍스트 정보와 함께 평가 자료를 입력받은 Dataset을 사용한다. 사용자는 서비스 받은 후 관심도에 따라 1부터 5까지 평가 자료를 입력한다. 1은 서비스를 추천 했지만 관심이 없어 실행을 하지 않았고, 2는 추천 서비스를 실행 했지만 관심정도가 낮아 관심이 없는 상태이다. 3은 좋은 상태이며, 4는 우수하고 5는 매우 우수한 경우를 말한다. 사용자들의 평가 자료는 평가 데이터베이스에 저장된다.

Monitor Manager는 사용자와 시스템 간에 정보를 주고 받는 역할을 담당하고 있다. 사용자의 서비스 요청이 들어 오면 Monitor Manager는 다음 단계의 모듈로 정보를 이동하는 역할도 하고 있다. 사용자로부터 얻어진 정보는 사용자의 개인 프로파일 정보, 현재 사용자와 동료의 관계 정보 그리고 실시간 컨텍스트 정보를 얻는다.

3.2 상대적 분류 방법과 전처리 단계

Clustering Manager에서는 사용자를 분류하는 모듈이다. 사용자를 분류하는 방법 중 나이를 기준으로 분류하는 방법은 대부분 10대, 20대, 30대 순으로 분류하거나 18부터 27, 28부터 37, 38부터 47 등으로 분류하는 방법을 사용하는 것이 보편적이 방법이다. 이 방법은 한정된 틀을 만들어 놓고 그 안에 사용자를 소속시켜 평가하는 방법이다. 예를 들어 사용자의 나이가 19세라고 가정하여 10대의 분류 그룹에 포함되었다고 생각하자. 이 경우 사용자는 10대의 성향만 가지고 있는 것이 아니라 20대 초반의 성향과 10대 후반의 성향을 모두 가질 수 있지만 10대로 편향되어 10대의 성향만 갖게 되는 단점을 가지게 된다. 이러한 분류 방법보다는 사용자를 기준으로 재분류하는 방식을 사용한다. 즉, 사용자가 19세이고 친구의 나이가 20세인 경우 10대, 20대, 30대 등으로 분류하는 절대적 분류방법은 두 사람의 분류 방법이 달라 다른 그룹에 편성될 것이다[표 1]. 그러나 사용자를 기준으로 분류하는 방법은 같은 그룹에 편입되어 비슷한 나이는 같은 그룹에 속하는 장점이 있다[표 2]. 분류 방법은 사용자를 기준으로 재분류하며, 사용자는 그룹의 중앙에 위치하게 된다. 같은 그룹에 속해있는 사용자는 성향이 비슷하므로 높은 가중치를 적용하고, 사용자와 멀리 있는 그룹 사용자는 가중치를 낮게 적용한다. 또한 남/여도 분류한다.

표 2. 절대적 분류

Table 2. Absolute Classification

색인	나이분류
1	~19
2	20~29
3	30~39
4	40~49
5	50~59
6	60~

표 3. 상대적 분류 방법

Table 3. Relative Classification

색인	나이분류
1	~14
2	15~19~24
3	25~28~34
4	35~39~44
5	45~49~54
6	55~

Pre-Processor는 사용자 평가 자료가 적어 평균계산에 영향을 미치는 자료를 찾아내는 모듈이다. 사용자의 평가는 상황이나 기분에 따라 평가 자료가 달라질 수 있다. 개인의 평가 자료가 많으면 전체적인 연산에 적은 영향을 미치지만, 항목 수가 적으면 평균값에 영향이 있어 협력적 필터링의 유사도와 선호도 계산에 영향을 준다. 표 4의 자료를 보며 예를 들어 본다. User1의 평균은 3이지만, item6을 평가하는 주변 환경의 영향을 받아 3을 입력하였다면 평균은 2로 숫자값이 많이 달라진 것을 알 수 있다. 그러나 user3의 경우 item2의 평가 자료를 3으로 변경하더라도 평균은 2.5에서 2.166으로 user1보다는 덜 영향을 받는다. 이러한 문제를 해결하기 위해 사용자의 평가 자료가 적은 자료들은 연산에서 제거한다. 제거하는 방법은 사용자의 평가 자료가 전체 사용자의 평가 자료의 평균보다 작을 경우 제거하는 방법을 사용한다.

표 4. 사용자 평가 자료

Table 4. User Evaluation Data

	item1	item2	item3	item4	item5	item6
user1	0	1	0	0	0	5
user2	0	1	2	3	0	1
user3	2	5	2	3	2	1
user4	5	1	0	0	4	0
user5	1	0	0	0	3	1

3.3 필터링 및 서비스 추천 단계

Filtering Manager 단계에서는 분류된 사용자, 평가 자료 그리고 실시간 컨텍스트 정보를 이용하여 사용자에게 서비스를 추천한다. 이 모듈에서는 협력적 필터링 방법을 이용하여 사용한다. 이 방법은 사용자의 평가 자료를 비교하여 선호도가 비슷한 사람들의 평가 항목을 가지고 예측하는 방법

이다. 기존의 협력적 필터링 방법은 실시간 컨텍스트 정보를 전혀 고려하지 않았지만 본 논문에서는 협력적 필터링 과정에서 컨텍스트 정보를 비교하여 유사한 사용자에게 가중치를 높게 적용하여 유사도를 계산하는 방법에 사용한다.

Context DB에서 저장된 컨텍스트 정보는 사용자의 위치, 동행인 그리고 서비스 받은 시간 등을 사용하며, 서비스 내용은 영화를 추천하는 것이다. 사용자가 서비스 받을 위치 정보와 다른 사용자들의 위치 정보를 비교하여 일치하는 지역에 있는 사용자만을 비교한다. 동행인에 따라 영화 추천 정보는 전혀 달리 선택될 수 있기 때문에 동행인에 대하여 가족, 동료, 애인, 친구, 싱글로 분리하고 영화 장르는 액션, 로맨스, 드라마, 코미디로 한정하여 연관성을 비교하였다. 또한 시간은 영화 보는 시간 때에 따라 오전, 오후, 저녁, 심야로 나누어 비교하였다. 컨텍스트 정보의 가중치를 이용한 유사도 값을 구하는 식은 식 1과 같다. sWeight는 유사성이 가장 높은 점수는 1이며, 유사성이 가장 낮은 점수는 0.5로 제한하였다. 유사성이 전혀 존재하지 않다고 0의 값을 넣게 되면 다른 부분의 유사성이 높아도 유사도 값은 0이 되기 때문이다.

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

- $W_{x,y}$: 사용자 x와 사용자 y의 유사도 가중치
- \bar{r}_x : 사용자 x의 서비스 정보에 대한 평가 평균값
- \bar{r}_y : 사용자 y의 서비스 정보에 대한 평가 평균값
- $r_{x,a}$: 사용자 x가 서비스 정보 a에 대하여 평가한 값
- $r_{y,a}$: 사용자 y가 서비스 정보 a에 대하여 평가한 값
- n : 비교할 서비스 정보의 개수
- sWeight : 컨텍스트 정보의 유사도 가중치

유사도 식을 구하고 이웃의 수를 결정하는 방법으로는 Best n-neighborhood 방법을 사용한다. 이 방법은 유사성이 높은 50%의 이웃만 선정하여 이웃을 채택한다. 그러나 유사도 값이 0.5보다 작거나 음의 상관을 가지고 있을 경우는 모두 다음 계산과정에서 제거 하였다.

이웃 선정 이후 예측값을 찾기 위하여 선호도 값을 구하며, 식 2와 같다.

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

- $P_{x,b}$: 사용자 x와 서비스 항목 b에 대한 선호도 예측값
- $W_{x,y}$: 사용자 x와 사용자 y의 유사도 가중치
- $r_{y,b}$: 사용자 y가 서비스 항목 b에 대한 평가 값
- \bar{r}_x : 사용자 x의 평균값
- \bar{r}_y : 사용자 y의 평균값
- n : 결정된 이웃의 수

위와 같은 방법을 사용하면 현재의 상황을 고려한 가장 유사성이 높은 서비스 정보를 추천하게 된다. Recommendation에서는 Item DB에 있는 서비스 정보를 사용자에게 추천하게 된다. 또한 서비스된 정보는 프로필

데이터베이스에 저장되어 다음에 사용할 수 있는 정보로 저장된다.

4. 실험 및 평가

실험에 사용된 자료는 미네소타 대학의 Computer Science and Engineering 전공의 GroupLens Research Group에서 영화를 보고 느낀 감정을 평가하여 데이터베이스에 저장한 자료가 MoviesLens dataset이다. MovieLens dataset은 100k dataset이다. 100k dataset은 943명의 사용자, 1682편의 영화 그리고 10만개의 평가 값으로 구성되어 있다. 본 논문에서는 모든 사용자의 나이가 나와 있는 100k dataset을 사용하였다.

MovieLens dataset은 사용자 평가 데이터베이스, 사용자 데이터베이스 그리고 영화 데이터베이스로 구성되어 있다. MoviesLens dataset의 데이터베이스는 사용자 ID, 항목 ID, 평가 자료 그리고 사용자 평가 시간으로 구성되었으며, 컨텍스트 정보를 사용한 데이터베이스는 사용자 ID, 항목 ID, 평가 자료, 평가 시간, 나이, 성별, 위치, 동행인 그리고 서비스 시간 등을 사용하였다.

표 5. 사용자 평가 DB
Table 5. User Valuation Database

user id	item id	rating	timestamp
196	242	3	881250949
186	302	3	891717742
22	377	1	878887116
244	51	2	880606923
166	346	1	886397596

User id와 item id는 각각의 데이터베이스로 구성되어 있어 각각의 아이디만으로 사용자와 아이템을 구분한다. Rating은 사용자가 평가한 평가 자료이며, 1부터 5까지 값을 입력할 수 있다. 예를 들어 1은 보지 않음, 2는 관심 없음, 3은 좋음, 4는 우수함, 5는 매우 우수함의 척도로 평가한다. Timestamp는 unix time으로 1970년 1월 1일 기준으로 하루를 86400초로 계산한다. 예를 들어 “881250949”라는 숫자의 time stamp를 날짜로 계산하는 알고리즘을 이용하여 변환하면 대한민국 표준시로 “1997년 9월 23일 7시 2분 38초”로 표현된다.

또한 사용자의 컨텍스트 정보와 평가 정보를 설문조사 형식으로 데이터베이스에 저장하였다. 설문조사 인원은 남자 241명과 여자 186명으로 총 427명을 조사하여 관련성을 조사하였다. 서비스 추천에 사용된 컨텍스트 정보는 사용자 ID, 나이, 동행인, 성별, 장르, 제목, 장소 그리고 시간 등을 사용한다. 나이는 상대적인 분류방법을 이용하여 분류하였고, 동행인의 경우에는 누구와 동행했는가에 따라 영화의 장르가 결정되는 것을 알 수 있었다. 동행인은 가족, 동료, 애인, 친구, 싱글로 분리하고 영화 장르는 액션, 로맨스, 드라마, 코미디로 한정하여 연관성을 비교하였다. 또한 시간은 영화 보는 시간 때에 따라 오전, 오후, 저녁, 심야로 나누어 비교하였다. 사용된 선호도 값은 1, 2, 3, 4, 5 단계로 표현하였으며, 가장 선호하지 않는 값은 1을 부여했으며, 선호도가 높으면 5를 부여하였다.

논문의 예측성을 평가하기 위하여 MAE(Mean Absolute Error)를 사용하였다. v_i 는 예측 선호도이고 r_i 는 실제 선호도이다. N 은 총 예측 회수이다. MAE는 전체 예측회수에 대하여 발생한 평균 예측의 오차를 의미하며, 오차가 작을 수록 선호도가 높은 것을 의미한다.

$$|E| = \frac{\sum |v_i - r_i|}{N} \tag{3}$$

본 연구에서 제안한 시스템은 아래와 같은 방법으로 실험 평가 하였다. x축의 사용자수는 평가된 사용자들을 말하며, Y축은 평가식 MAE이다. 첫 번째 방법은 사용자 평가 자료에 순수 협력적 필터링을 적용하였다. 두 번째 방법은 필터링된 사용자 평가값을 가지고 평가한 협력적 필터링 방법을 사용하였다. 세 번째 방법은 필터링된 사용자 평가값에 사용자의 상황에 따른 실시간 컨텍스트 정보를 이용하여 협력적 필터링 하였다. 그 결과, 제안한 방법이 순수 협력적 필터링만 사용한 방법1 보다 정확도의 측면에서 평균적으로 16.2% 정도 우수함을 알 수 있었다.

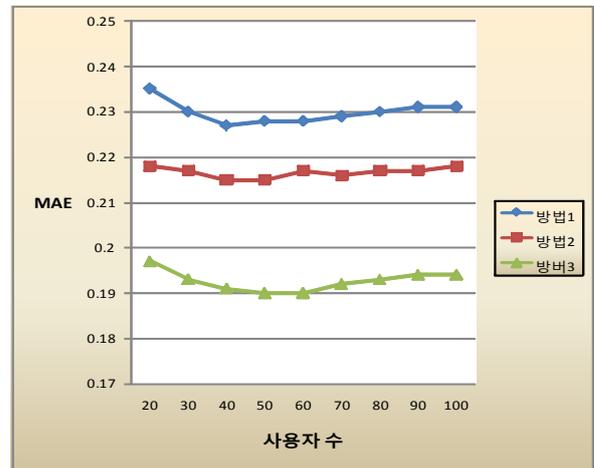


그림 3. 사용자 수에 따른 성능 평가
Fig. 3. Performance Valuation to the Number of Users

5. 결 론

전자상거래의 발전은 추천시스템의 기술을 더욱 발전시키는 계기가 되었다. 추천시스템의 추천 방법은 다양한 방법이 있지만 그 중 가장 많이 사용되고 있는 방법은 협력적 필터링이다. 그러나 협력적 필터링은 몇 가지 문제점을 가지고 있다. 첫째는 사용자가 평가한 모든 자료를 사용하게 되면 평가 값이 적은 경우 상황에 따라 평가 값이 유동적이면 평균에 영향을 미치게 되므로 적절한 제한이 필요하다. 이러한 문제점을 해결하기 위해 모든 자료를 평가하는 것이 아니라 평균 이상인 자료들을 평가한 사용자만 사용하였다. 두 번째는 현재의 상황이나 사용자의 상태를 전혀 고려하지 않고 사용자의 평가 자료만을 가지고 평가한다는 것이다. 이러한 문제점을 해결하기 위해 실시간으로 얻어진 컨텍스트 정보를 협력적 필터링에 적용하였다. 그 결과 현재 사용자의 상황을 파악하여 사용자에게 서비스를 할 수 있기 때문에 평가 자료만을 가지고 평가한 자료보다 더 우수함을 알 수 있었다.

참 고 문 헌

[1] 이형동, 김형주, “협업 필터링 추천시스템에서 취향 공간을 이용한 평가 예측 기법,” *정보과학회논문지*, 제34권, 5호, 2007.

[2] N. Good, J. B. Schafer, J. A. Konstan, A. Borchers, B. Sarwar, J. Herlocker, and J. Riedl, “Combining Collaborative Filtering with Personal Agents for Better Recommendations,” *In Proceedings of National Conference on Artificial Intelligence*, pp. 439-446, 1999.

[3] 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.

[4] 정경용, 류경중, 강운구, 이정현, “내용 기반 여과와 협력적 여과의 병합을 통한 추천 시스템에서 조화 평균 가중치.” *정보과학회논문지*, 제30권, 3호, pp. 239-250, 2003. 4.

[5] 김병만, 이경, 박창석, 김시관, 김주연, “사용자 프로파일 정보를 고려한 협력 필터링,” *한국정보과학회 가을 학술발표논문집*, 제29권, 2호, pp. 286-289, 2002.

[6] Luis M. de Campos, Juan M. Fernandez-Luna, Juan F. Huete, “A Collaborative Recommender System Based on Probabilistic,” *Inference from Fuzzy Observations Fuzzy Sets and Systems*, 159, pp. 1554-1576, 2008.

[7] A. Merve Acilar, Ahmet Arslan, “A Collaborative Filtering Method Based on Artificial Immune Network,” *Expert Systems with Applications*, 36, pp. 8324-8332, 2009.

[8] J. Ben Schafer, Dan Frankowski, Jon Herlocker and Shilad Sen, “Collaborative Filtering Recommendation Systems”, *The Adaptive Web*, 2007

[9] D. Salber, A.K. Dey and G.D. Abowd, “The Context Toolkit: Aiding the Development of Context-Aware Application”, *In the Workshop on Software Engineering for Wearable and Pervasive Computing (Limerick Ireland)*, June, 2000.

[10] 윤효근, 이상용, “협력적 필터링 기법을 이용한 P2P 모바일 에이전트 기반 사용자 컨텍스트 인식 및 서비스 처리 구조,” *한국지능시스템학회논문지*, 제15권, 1호, 2005.

[11] Matthias B. and Scharam D. “A Survey on Context-Aware Systems,” *Distributed Systems Group, Technical University of Vienna*, November 30, 2004.

저 자 소 개



이세일(Se-Il Lee)

1993년 : 대전공대 전자계산학과 졸업 (공학사)

2001년 : 청운대학교 전산전자정보공학과 (공학석사)

2004 ~ 현재 : 공주대학교 컴퓨터공학과 (박사과정)

관심분야 : 유비쿼터스 컴퓨팅, 상황 인식, 협력적 필터링, 게임 알고리즘

E-mail : lsilhr@kongju.ac.kr



이상용(Sang-Yong Lee)

1984년 : 중앙대학교 전자계산학과(공학사)

1988년 : 일본동경대학대학원 총합이공학 연구과(공학석사)

1988년 ~ 1989년 : 일본 NEC 중앙연구소 연구원

1993년 : 중앙대학교 일반대학원 전자계산학과 (공학박사)

1996년 ~ 1997년 : University of Central Florida 방문 교수

1993년 ~ 현재 : 공주대학교 컴퓨터공학부 교수

관심분야 : 인공지능, 유비쿼터스 컴퓨팅, 상황 인식 및 예측

E-mail : sylee@kongju.ac.kr