

개선된 추천시스템을 이용한 전자상거래시스템 설계 및 구현

김 영 설[†] · 김 병 천^{††} · 윤 병 주^{†††}

요 약

인터넷 환경의 급속한 발전과 함께 이를 이용한 전자상거래가 빠르게 증가하고 있다. 증가하는 전자상거래 환경에서 고객에게 필요한 제품을 신속히 제공하고, 제품판매를 증가시킬 수 있는 새로운 전자상거래 시스템의 필요성이 점차 커지고 있다. 이러한 필요성에 의해서 최근에 추천시스템에 대한 많은 연구가 이루어지고 있다. 하지만 지금까지의 추천시스템은 고객의 구매데이터가 증가하면 고객에게 추천을 제공하는데 많은 시간이 소요되어 실시간 추천이 어렵다는 큰 단점을 가졌다. 따라서, 이 논문은 전자상거래 시스템의 경쟁력을 높이는 방안으로 협동적 필터링을 이용한 추천시스템을 연구하고, 성능을 개선하기 위해서 추천에 사용되는 데이터를 제품의 대표상표를 이용하여 줄임으로서 추천소요시간을 단축하여 실시간 추천이 가능한 개선된 추천시스템을 제안하고 실험하였다. 또한 개선된 추천시스템을 Enterprise JavaBeans로 구현함으로써 분산환경에서 사용할 수 있는 전자상거래시스템을 설계하여 경쟁력 있는 전자상거래 시스템 환경을 제공하고자 한다.

Design and Implementation of e-Commerce Applications using Improved Recommender Systems

Young Seol Kim[†] · Byung-Cheon Kim^{††} · Byung-Joo Yoon^{††}

ABSTRACT

With the fast development of Internet environment, e-Commerce is rapidly increasing. In the expanding e-Commerce environment, the need for a new e-Commerce systems that will deliver products to the customer rapidly and increase sales is growing bigger. Recently, these requirements brought many researches on recommender systems. However, until now, those recommender systems have a limit because it takes too much time for recommender systems to give customers the recommendations in real time, if the number of purchase data of customers is large. So this paper concerns on the recommender systems using collaborative filtering as one of the solutions to increase the competitive power. We proposed and experimented the more improved recommender systems which could decrease the data size to shorten the recommending time by using the representative category of the product which customers want to buy. Also, we design and implement a recommender system using Enterprise JavaBeans.

키워드 : 전자상거래(e-Commerce), 추천시스템(Recommender Systems), 분산객체(Distributed Object), EJB(Enterprise JavaBeans)

1. 서 론

인터넷 환경의 급속한 발달과 함께 이를 이용한 전자상거래가 빠르게 증가하고 있다. 증가하는 전자상거래 환경에서 고객에게 필요한 제품을 신속히 제공하고, 제품판매를 증가시킬 수 있는 새로운 전자상거래 시스템의 필요성이 점차 커지고 있다.

그러나 기존의 전자상거래 시스템은 고객이 요구하는 제품만을 제공하는 단순한 형태여서 고객의 요구사항을 충분히

충족시키지 못하였다. 따라서 전자상거래시스템들 간의 경쟁력이 강조되는 현재에는 고객들에게 양질의 서비스를 제공하고, 이를 이익 창출로 연결시키는데 도움이 되는 보다 특성화된 기능의 다양한 해결방안이 요구되고 있다.

이러한 요구에 부응하여 최근 전자상거래 시스템 연구에서는 추천시스템에 대한 연구가 활발히 진행되고 있다. 하지만 지금까지 발표된 대다수의 추천시스템은 고객의 구매 데이터가 증가하면 고객에게 추천을 제공하는데 많은 시간이 소요되어 실시간 추천이 어렵다는 큰 단점을 가졌다.

따라서, 본 논문에서는 전자상거래 시스템의 경쟁력을 높이기 위하여 기존의 추천시스템들을 조사·분석한 후, 추천시스템의 추천 소요 시간을 단축시킬 수 있는 개선된 추천시

[†] 준 회원 : 명지대학교 대학원 컴퓨터공학과

^{††} 종신회원 : 한경대학교 컴퓨터공학과 교수

^{†††} 종신회원 : 명지대학교 컴퓨터공학과 교수

논문접수 : 2001년 9월 19일, 심사완료 : 2001년 12월 17일

시스템을 제안한다. 또한 제안한 개선된 추천시스템에 기반하는 전자상거래 시스템을 설계하여 이를 Enterprise Java-Beans로 구현함으로써, 분산환경에서 사용할 수 있는 전자상거래 시스템의 모형을 제시한다.

2. 관련 연구

2.1 정보과잉과 그 해법

인터넷에서 야기되는 정보과잉(Information overloading) 현상을 해결하기 위하여 정보검색, 정보필터링, 협동적 필터링 등에 관한 연구가 활발히 진행되고 있다[3].

정보검색(Information retrieval)은 사용자가 찾고자 하는 정보에 대해서만 응답하기 때문에 사용자가 원하는 것은 많지만 사용자에게는 유용한 정보를 발견하거나 제공하지 못하며, 정보필터링(Information filtering)은 대상 정보를 이미 분류해 놓거나 그 속성을 알고 있어야만 사용 가능하다는 제약이 있으므로 사용자가 찾고 싶은 내용을 정확히 알지 못한 경우에는 대응하지 못하는 문제점이 있다[1].

협동적 필터링(Collaborative filtering)은 다른 사용자의 의견을 참고하여 대상이 되는 특정 사용자에게 적합한 정보를 제공하는 방법이다. 즉, 사용 가능한 아이템에 대한 사용자들의 의견을 데이터베이스로 구축하여 특정 사용자의 특정 아이템에 대한 평가 값을 예측할 때 데이터베이스에서 특정 사용자와 유사한 사용자(유사그룹)들을 발견하여 유사 사용자의 의견을 참조하여 평가 값을 예측한다[1].

예를 들어, 3명의 사용자가 4개의 영화에 대해 1~5점 사이의 점수로 자신의 취향을 평가(Rating)했다고 가정하자. 평가는 특정 사용자가 특정 아이템에 대해서 의견을 표시하는 가장 단순한 방법이다.

<표 1> 영화에 대한 평가 예

평가자 영화	타이타닉	쉬리	매트릭스	리브레터
철수	5	5	3	2
영희	3	3	5	5
순이	4	4	3	?

<표 1>에서 순이는 영희보다는 철수에 더 가까운 취향을 가지고 있다는 것을 추측할 수 있다. 따라서 순이가 아직 평가하지 않은 '리브레터'에 대한 예측은 영희보다는 철수를 참조하는 것이 더 낫다고 볼 수 있는데, 이와 같은 가정과 방법에 기초한 것이 바로 협동적 필터링이다.

협동적 필터링이 정보검색이나 정보필터링보다 뛰어난 점은 아이템에 대한 사전분류나 파악에 의존하기보다는 다른 사용자들의 의견만을 참조한다는 점, 사용자가 예측하

지 않았지만 취향에 맞는 새로운 정보를 발견할 가능성이 높다는 점, 시스템에서 자동화될 가능성이 매우 높다는 점이다.

2.2 추천시스템의 분류

일반적으로 추천시스템의 추천방법은 다음과 같다[5, 7].

- ① 단순검색(Raw retrieval)은 고객이 찾고자 하는 내용에 대한 검색 결과를 제공한다.
- ② 수동선택(Manual selection)은 고객이 추천을 요구하면 해당 분야의 전문가가 고객이 좋아할 제품을 예측하여 제공한다.
- ③ 통계적 요약(Statistical summaries)은 고객이 추천을 요구하면 고객이 속하는 공동체에서 가장 인기있는 제품을 추천한다.
- ④ 속성기반(Attributed-based)은 아이템의 속성에 대한 고객의 선호도에 따라 추천을 수행한다.
- ⑤ 아이템간 관계(Item-to-item correlation)는 상품을 작은 집합으로 구분하여 고객이 특정 집합에 포함된 상품을 구입하면 같은 집합에 포함된 상품들을 고객에게 추천하는 방식이다.
- ⑥ 사용자간 관계(User-to-user correlation)는 추천을 요구하는 고객과 제품을 구매한 다른 고객과의 관계에 기초하여 고객에게 제품을 추천한다. 이 기술은 개인화된 추천을 제공하기 위하여 유사 사용자(그룹)의 의견을 사용하므로 협동적 필터링의 한 유형으로 볼 수 있다.

2.3 기존의 협동적 필터링

협동적 필터링 방법은 우수한 성능을 보이는 추천시스템의 핵심기술로서 TAPESTRY로부터 유래되었다[1]. TAPESTRY의 문제점을 해결하면서 상당한 수준의 성능향상을 이룬 대표적인 자동화된 협동적 필터링 시스템은 GroupLens이다[2].

GroupLens는 두 가지 평가방법을 사용한다. 첫 번째는 사용자의 직접적인 평가에 의한 방법이며, 두 번째는 사용자의 직접적인 평가가 없는 경우 다른 사용자의 프로파일을 기반으로 하는 상관관계(Correlation)에 의해 문서에 대한 평가를 예측하는 방법이다.

GroupLens에서는 Tapestry에서 문제가 되었던 사용자가 읽은 문서에 대한 평가를 하지 않는 경우는 같은 관심을 가지는 사람들의 평가를 기반으로 해결하려는 방법을 시도하였다.

<표 2>와 같이 6개의 문서에 대한 4명의 사용자의 평가가 있다고 가정하자. 빈칸으로 되어 있는 것은 아직 사용자

가 그 문서를 읽지 않은 것을 의미한다. 그리고 물음표가 되어 있는 것은 문서를 읽었지만 문서에 대한 평가를 하지 않은 것을 의미한다.

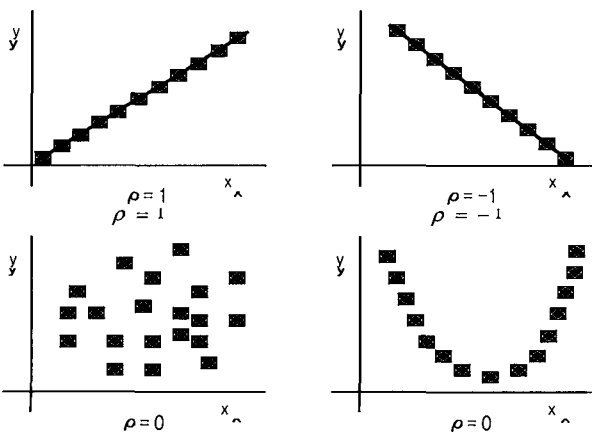
〈표 2〉 사용자의 문서에 대한 평가

문서번호	평가자			
	Ken	Lee	Meg	Nan
1	1	4	2	2
2	5	2	4	4
3			3	
4	2	5		5
5	4	1		1
6	?	2	5	

GroupLens는 Ken의 6번째 문서에 대한 평가를 예측하기 위하여 상관계수(correlation coefficients)를 사용하는데, 이 값은 -1과 1사이의 값을 가진다. Ken과 Lee사이의 상관계수는 다음과 같이 구해진다.

$$\begin{aligned}
 r_{KL} &= \frac{Cov(K, L)}{\delta_K \delta_L} \\
 &= \frac{\sum_i (K_i - \bar{K})(L_i - \bar{L})}{\sqrt{\sum_i (K_i - \bar{K})^2} \sqrt{\sum_i (L_i - \bar{L})^2}} \quad (2.1) \\
 &= \frac{-2-2-2-2}{\sqrt{10} \sqrt{10}} = -0.8
 \end{aligned}$$

식 (2.1)은 Ken과 Lee사이의 상관계수를 구하는 식으로 \bar{K} 과 \bar{L} 는 각각 Ken과 Lee의 문서에 대한 평가 평균값이다. 상관계수가 1이면 perfect positive relationship이라고 하며 -1이면 perfect negative relationship이라고 한다. 만약 0이면 상관관계가 없는 경우이다. 이것을 (그림 1)과 같은 그래프로 나타낼 수 있다.



(그림 1) 상관계수 그래프

위와 같은 방법으로 다른 사람들과의 상관관계를 구해보면 Meg과의 상관계수는 1이며, Nan과의 상관계수는 0이 된다. Ken과 다른 사람과의 상관관계가 모두 구해졌으면 Ken의 6 번째 문서에 대한 평가를 예측하게 된다. 평가의 예측에는 상관계수를 포함한 평균값을 이용한다.

$$\begin{aligned}
 K_{6pred} &= \bar{K} + \frac{\sum_{j \in \text{raters}} (J_6 - \bar{J}) r_{Kj}}{\sum_j |r_{Kj}|} \\
 &= 3 + \frac{2r_{KM} - r_{KL}}{|r_{KM}| + |r_{KL}|} = 3 + \frac{2 - (-0.8)}{|1| + |-0.8|} \\
 &= 4.56 \quad (2.2)
 \end{aligned}$$

Nan과의 상관계수는 0이므로 식 (2.2)에서 Nan과의 관계 값은 제외된다. 따라서 Ken의 6 번째 문서에 대한 평가 값은 4.56으로 예측된다. 이와 같은 방법을 통해서 평가를 하지 않은 문서에 대한 예측을 수행하게 된다.

이와 같이 한 사용자와 다른 사용자의 상관관계에 기초한 평가 예측을 일반화하면 식 (2.3)과 같다.

$$p_{a,i} = \bar{r}_a + \frac{\sum_{u=1}^n (r_{u,i} - \bar{r}_u) * w_{a,u}}{\sum_{u=1}^n w_{a,u}} \quad (2.3)$$

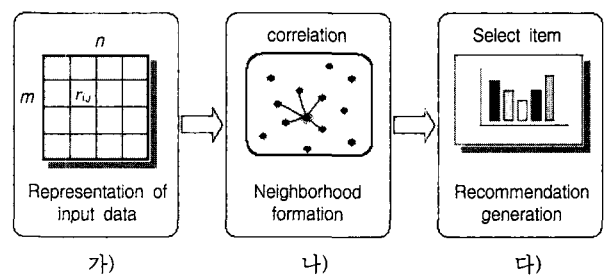
식 (2.3)의 $p_{a,i}$ 는 사용자 a의 아이템 i에 대한 예측 값, \bar{r}_a 는 사용자 a의 평가 평균값, n은 사용자 수, r_{ui} 는 사용자 u와 i번째 아이템 평가 값, \bar{r}_u 는 사용자 u의 평가 평균값, $w_{a,u}$ 는 사용자 a와 사용자 u사이의 유사도(similarity weight)를 나타내며, 식 (2.4)와 같이 피어슨 상관계수로 정의된다.

$$w_{a,u} = \frac{\sum_{i=1}^m (r_{a,i} - \bar{r}_a) * (r_{u,i} - \bar{r}_u)}{\sigma_a * \sigma_u} \quad (2.4)$$

3. 개선된 추천시스템의 제안

3.1 GroupLens의 문제점

(그림 2)는 협동적 필터링을 이용한 추천과정으로 가는 m 명의 고객이 n 개 제품에 대해 평가한 평가 데이터의 집합으



(그림 2) 협동적 필터링의 추천 과정

로 추천시스템의 입력 데이터가 된다. 이것은 보통 $m \times n$ 고객-제품 Matrix로 표현되며 r_{ij} 은 i 번째 고객이 j 번째 제품을 평가했다는 것을 의미하게 된다.

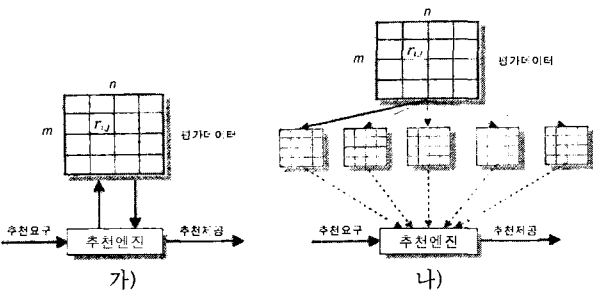
가)의 $m \times n$ 고객-제품 Matrix를 추천시스템의 입력 데이터로 사용하여 나)과정에서 고객들 사이의 상관계수를 구하여 이를 기초로 유사고객을 발견하고 다)과정에서 유사고객의 의견을 참고하여 고객에게 제품을 추천한다.

이와 같은 협동적 필터링 추천 알고리즘은 고객사이의 상관계수를 구하기 위해 많은 시간을 요구하여 추천 요구를 실시간으로 처리하기 힘든 단점을 가진다. 이 때문에 입력 데이터를 전처리 과정을 통해서 줄이는 작업이 요구된다.

3.2 개선된 추천시스템의 제안

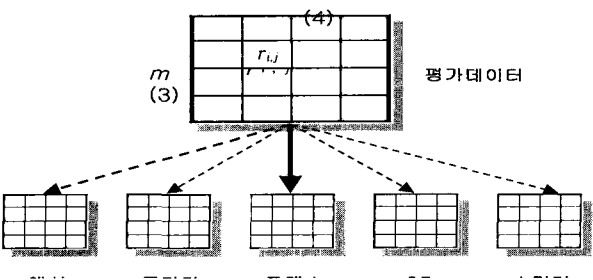
이 논문에서는 추천시스템의 추천 수행시간을 단축시키기 위하여 추천시스템의 입력데이터로 사용되는 평가데이터를 전처리 과정을 통해서 장르별 평가데이터로 분리하여 이 데이터를 추천시스템의 입력데이터로 사용하여 추천 수행시간을 단축할 수 있는 개선된 추천시스템을 제안하였다.

기존의 GroupLens는 (그림 3)의 가)와 같이 추천을 수행할 때 평가 데이터 전체를 추천시스템의 입력 데이터로 사용하여 추천 소요시간이 길어져 실시간 추천이 어려웠으나 제안 방법은 나)와 같이 평가 데이터를 전처리 과정을 통해서 장르별 평가 데이터로 분리하여 사용함으로써 추천시스템의 입력 데이터를 감소시켜 추천 수행시간을 단축할 수 있는 형태를 가진다.



(그림 3) 추천방법의 비교

3.2.1 장르별 데이터 분리

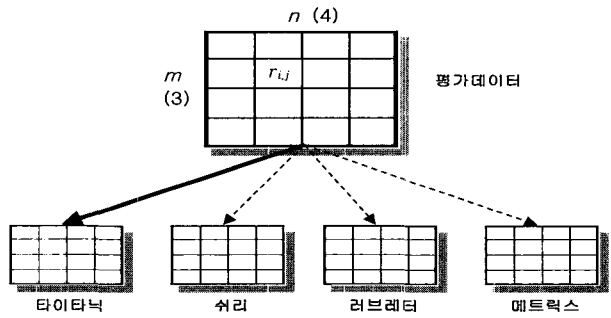


(그림 4) 장르별 평가데이터

제안된 추천시스템의 입력 데이터로 사용될 장르별 평가 데이터를 구하기 위하여 (그림 4)와 같이 평가 데이터로부터 장르별 평가 데이터를 분리한다.

3.2.2 대표장르 결정

영화의 대표장르를 결정하기 위하여 (그림 5)와 같이 전체 평가 데이터로부터 특정 영화를 평가한 데이터를 분리한다.



(그림 5) 영화별 평가데이터

영화에 대한 평가 데이터 분리작업이 끝나면, <표 3>과 같이 해당 영화가 속하는 장르 평가 데이터를 추천시스템의 입력 데이터로 사용하여 사용자 간 유사도에 기초하여 각 사용자의 영화에 대한 평가 값을 예측한다.

<표 3> 장르별 예측에러

사용자	영 화	장 르	실제값	예측값	에 러
철수	타이타닉	드라마	5	4.3064	0.6936
영희	타이타닉	드라마	3	3.6072	0.6072
순이	타이타닉	드라마	4	3.4124	0.5876
철수	타이타닉	로맨스	5	4.7082	0.2918
영희	타이타닉	로맨스	3	2.9927	0.0073
순이	타이타닉	로맨스	4	3.3756	0.6244

이 때 <표 4>와 같이 예측결과와 평균에러가 작은 장르를 이 영화에 대한 대표장르로 결정한다.

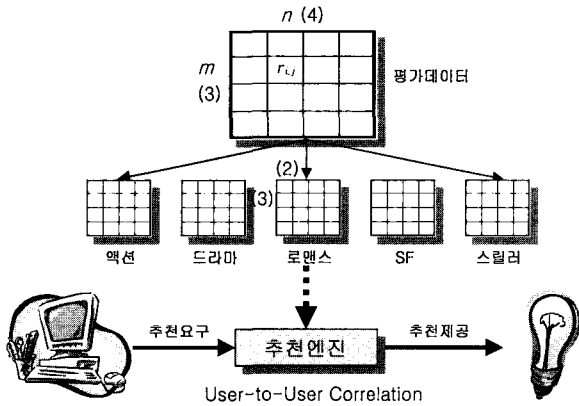
<표 4> 장르별 평균에러

영 화	장 르	평균에러	비 고
타이타닉	드라마	0.6234	
타이타닉	로맨스	0.3078	대표장르

3.2.3 평가값 예측

사용자 a 의 특정 영화에 대한 좋아하는 정도를 예측하기 위해서 이 영화의 대표장르(예, 로맨스) 평가데이터를 사용자 간 상관계수 계산의 입력데이터로 사용하여 사용자 a 의 영화에 대한 평가 값을 예측한다. 이 예측값에 의해 사용자 a 가 이 영화를 얼마나 좋아하는지를 예측하게 된다.

(그림 6)과 같이 대표장르 평가데이터를 추천시스템의 입력 데이터로 사용함으로써 추천에 필요한 입력 데이터의 양을 최소화하여 추천 수행 시간을 단축할 수 있으며 예측시간의 단축으로 인하여 실시간 추천이 가능하게 된다.



(그림 6) 대표장르를 이용한 추천

3.3 실험 및 토의

본 실험은 GroupLens의 예측방법과 이 논문에서 제안한 장르별 평가데이터를 이용하는 방법의 평가값 예측의 표준편차(Mean Absolute Error), 최소에러(Best Error), 최대에러(Worst Error) 그리고 전체 영화 중 예측가능한 영화의 비율을 나타내는 예측율(Coverage)을 비교하여 제안한 방법의 타당성을 살펴보고자 한다.

3.3.1 실험환경 및 DataSet

실험은 펜티엄II-350MHz 3대에서 10일동안 이루어졌으며, 사용 언어는 Java이고 개발틀은 JBuilder4.0을 사용하였다. 오리지날 DataSet은 텍스트 형태였으나 이 데이터를 Microsoft SQL 7.0으로 변환하여 사용하였다.

실험에 사용된 Data Set은 GroupLens Research Project [9]에서 제공한 MovieLens Data Set으로 그 구성은 다음과 같다[2]. 1) 943명의 사용자가 1682개 영화에 대해서 1-5점 사이의 점수로 평가한 100,000개 데이터이다. 2) DataSet은 최소 20개 이상의 영화에 대해서 평가한 사용자 데이터만으로 구성된다. 3) 영화는 19개 장르에 중복해서 포함될 수 있다. 실험을 위하여 총 100,000개의 DataSet 중에서 80,000개를 뽑아 Training Data로 사용하고 20,000개를 Test Data로 사용하였다.

3.3.2 실험

943명의 사용자가 1650개 영화에 대해서 평가한 80,000개 Training Data를 분석하여 <표 5>와 같이 19개 장르별 평가데이터를 분리한다. 1650개 영화에 대해서 제안한 방법의 대표장르 결정방법을 이용하여 1650개 영화에 대한 대표장르를

구한 후 GroupLens 실험에서 평가값 예측에 사용된 Test Data의 「사용자-영화」 평가쌍에 대해서 장르별 평가데이터를 입력데이터로 사용하여 평가값을 예측한다.

<표 6>는 기존의 GroupLens에 의한 실험으로 80,000개의 Training Data 전체를 추천시스템의 입력데이터로 사용하여 Test Data의 「사용자-영화」 평가값을 예측한 결과이다.

<표 5> 장르별 평가데이터

장르	사용자	영화	평가쌍	회소성
action	929	247	20566	91.04%
adventure	880	135	11131	90.63%
animation	619	42	2893	88.88%
children	763	122	5779	93.83%
comedy	935	500	23896	94.89%
crime	883	106	6446	93.12%
documentary	300	46	570	95.87%
drama	942	707	31989	95.20%
...

<표 6> 대표장르 사용한 경우 성능

표준편차	최소에러	최대에러	예측율	예측시간(ms)
0.7725	0.0082	2.6277	100%	1148

한편 <표 7>은 제안방법에 의해 장르별 평가데이터를 추천시스템의 입력데이터로 사용하여 Test Data의 「사용자-영화」 평가값을 예측한 결과이다.

<표 7> 대표장르 사용한 경우 성능

표준편차	최소에러	최대에러	예측율	예측시간(ms)
0.7748	0.0074	2.9762	99%	297

3.3.3 성능측정

GroupLens와 이 논문에서 제안한 장르별 평가데이터를 사용한 방법을 실제 DataSet을 통해서 실험하였다. 실험결과 추천시스템의 추천 정확성 나타내는 표준편차는 GroupLens와 제안한 방법이 큰 차이를 보이지 않았다. 하지만 평가값 예측에 소요되는 시간을 크게 단축할 수 있었다.

<표 8> 예측성능 비교

	GroupLens	제안방법
입력데이터	80,000	18,300
예측율	100%	99%
표준편차	0.7425	0.7748
최소에러	0.0082	0.0074
최대에러	2.6277	2.9762
예측시간	1148 ms	297 ms

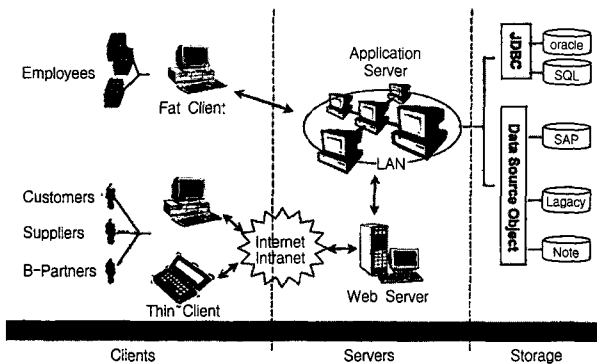
<표 8>은 GroupLens와 제안한 방법의 예측성능 비교표이다. GroupLens가 943명이 1650개 영화에 대해 평가한 80,000개의 Training Data 전체를 사용하는데 비하여 제안한 방법에서는 평균 894명이 341개 영화에 대해 평가한 18,300개의 평가 데이터를 사용하여 추천시스템의 입력 데이터 양을 줄일 수 있었다. 두 방법 모두 예측의 정확성을 나타내는 평가값 예측의 표준편차와 최소에러, 최소에러는 큰 차이가 없었다. 하지만 제안한 방법의 평가값 예측시간은 기존 GroupLens가 약 1148초 소요되는데 비하여 크게 감소한 약 297초가 소요되었다.

결국 제안한 방법은 대표장르를 이용하여 평가 데이터를 전처리 과정을 통하여 추천시스템의 입력데이터로 사용되는 데이터의 양을 감소시켜 기존방법의 표준편차, 최대에러, 최소에러, 예측을 같은 추천 성능의 희생없이 평가값 예측에 소요되는 시간을 단축할 수 있었다. 이에 따라 평가데이터의 양에 관계없이 빠른 추천이 가능하여 실시간 추천의 수행을 가능하게 하였다.

4. 전자상거래시스템 설계 및 구현

폭발적으로 증가하는 전자상거래 환경에 효율적으로 대응하기 위한 방법을 모색하고자 본 논문에서 제안한 추천시스템을 사용하고 Java기반의 EJB(Enterprise JavaBeans)를 활용하여 분산환경에서 운영될 수 있는 전자상거래 모형을 설계하고 구현하였다.

구현을 위한 개발 환경은 다음과 같다. Java 컴파일러는 JDK1.3을 사용하고 EJB 개발 Tool은 JBuilder4.0을 사용하였다. 추천시스템의 주요기능을 구현한 Component의 운영환경인 EJB Container는 Borland/Inprise사의 Inprise Application Server 4.1(IAS4.1)을 사용하였다.



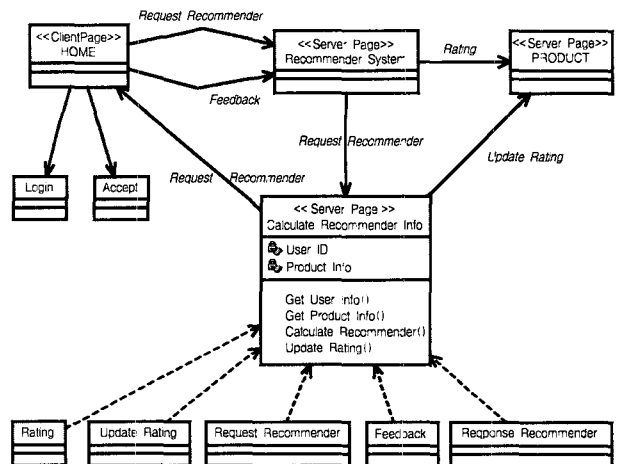
(그림 7) 분산환경에서의 추천시스템

(그림 7)은 분산환경에 응용될 수 있는 추천시스템을 EJB를 이용하여 구현한 전자상거래 시스템의 구성이다. 추천 엔진

의 주요 기능은 EJB 컴포넌트로 구현되어 application server에서 작동한다. 이러한 컴포넌트는 application server로부터 트랜잭션, 보안, 인증 등의 서비스를 제공받게 된다. 컴포넌트는 필요에 따라 JDBC를 통하여 Oracle 또는 MS-SQL과 같은 DBMS에 접근하거나 Data Source Object를 통해서 Legacy System과 통합될 수 있다.

이처럼 분산환경에서 구현된 추천시스템은 클라이언트의 웹브라우저를 통해서 호출되며, 클라이언트의 호출을 받는 웹서버는 application server의 EJB 컴포넌트를 실행하여 그 결과를 클라이언트에 되돌려주는 형태가 된다. 이 과정에서 application server는 fail-over처리나 load-balancing을 자동으로 처리하므로 분산환경에서 발생할 수 있는 오류를 방지해준다.

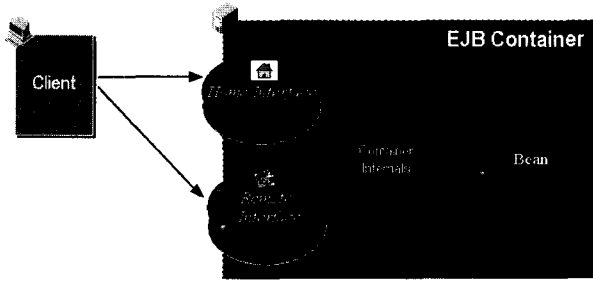
추천시스템의 기본구조를 바탕으로 전자상거래시스템을 기능별 단위로 구분해보면 (그림 8)와 같은 클래스 다이어그램으로 나타난다. 각각은 추천시스템의 모듈을 구현한 Component로 Session Bean 또는 Entity Bean으로 구현된다.



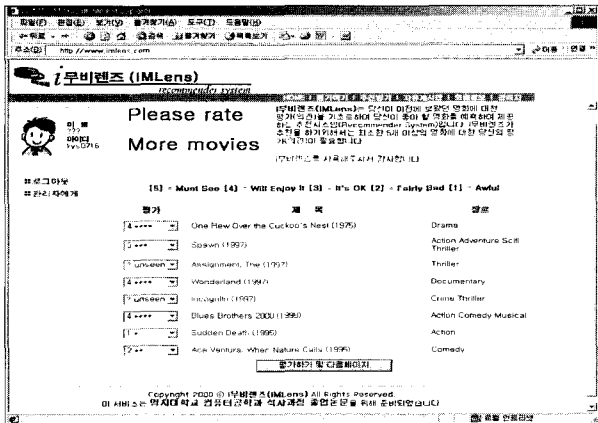
(그림 8) 클래스 다이어그램

추천엔진은 EJB로 구현되어 (그림 9)처럼 EJB Container에 등록된다. 클라이언트 프로그램에서 EJB에 구현된 Business Logic을 호출하기 위해서 JNDI(Java Naming and Directory Interface)를 통해서 Bean의 Home Interface를 얻은 후 Remote Interface에 의해서 EJB의 Business Logic를 호출하게 된다.

한편 클라이언트 프로그램은 JSP(Java Server Pages)로 구현한다. JSP는 Server-side-Applet로 불리기도 하는데 자바코드가 웹서버에서 실행되어 그 수행결과를 HTML 형태로 웹브라우저로 돌려주게 된다. 기존의 CGI방식이 접속사용자마다 프로세스(process)를 생성하여 웹서버의 부담을 가중시키는데 비하여 JSP는 쓰레드(Thread)를 사용함으로써 웹서버의 부담을 줄일 수 있다.



(그림 9) 클라이언트와 서버의 통신



(그림 10) 고객의 선호도 표시

(그림 10)은 제안한 추천시스템을 구현한 i무비렌즈에 접속하여 영화에 대한 고객의 선호도를 입력받는 화면이다. 기본적으로 10개 영화에 대해 평가하면 i무비렌즈로부터 추천을 제공받게 된다. 이 때 사용자는 좀더 정확한 추천을 제공받기 위해서 다양한 장르와 많은 영화에 대한 평가가 요구된다.

5. 결 론

기존의 추천시스템은 고객의 구매 데이터가 증가하면 할수록 추천시스템의 추천 소요시간은 계속 증가하는 것이 큰 단점으로 지적되었다. 본 논문에서는 GroupLens의 알고리즘을 기초로 추천 소요 시간을 단축할 수 있는 개선된 추천시스템을 제안하였다.

실험 결과, 제안한 방법은 고객의 평가데이터를 전처리 과정을 통해서 추천시스템의 입력데이터로 사용되는 데이터의 양을 감소시켜 예측성능을 나타내는 표준편차, 최대에러, 최소에러, 예측율의 희생없이 평가값 예측에 소요되는 시간을 크게 단축할 수 있었다. 또한 개선된 추천시스템을 Java기반의 EJB를 사용하여 분산환경에서 사용할 수 있는 전자상거래시스템 모형을 설계하고 구현하였다. 향후 추천의 정확성을 높이는 연구와, 기존 전자상거래 시스템과 통

합 가능한 새로운 형태의 추천시스템 구조에 대한 연구가 요구된다.

참 고 문 헌

- [1] D. Goldberg, D. Nichols, B. M. Oki, and D. Terry., Using Collaborative Filtering to Weaves and Information TAP-ESTRY, CACM, Vol.35, No.12, pp.61-70, 1992.
- [2] Resnick P., Iacovou N., Sushak M., Bergstrom P., and Riedl J., GroupLens : An open architecture for collaborative filtering of Netnews. Proceedings of the 1994 Computer Supported Collaborative Work Conference, 1994.
- [3] Sarwar, B., Konstan, J., Borchers, A., Herlocker, J., Miller, B., and Riedl, J., Using Filtering Agents to Improve Prediction Quality in the GroupLens Research Collaborative Filtering System. Proceedings of the 1998 Conference on Computer Supported Cooperative Work, Nov., 1998.
- [4] Herlocker, J., Konstan, J., Borchers, A., Riedl, J., An Algorithmic Framework for Performing Collaborative Filtering. Proceedings of the 1999 Conference on Research and Development in Information Retrieval, Aug., 1999.
- [5] Schafer, J.B., Konstan, J., and Riedl, J., Recommender Systems in E-Commerce. Proceedings of the ACM Conference on Electronic Commerce, November, 3-5, 1999.
- [6] Sarwar, B. M., Karypis, G., Konstan, J. A., and Riedl, J., Analysis of Recommender Algorithms for E-Commerce. ACM E-Commerce 2000 Conference, Oct., 17-20, 2000.
- [7] Schafer, J. B., Konstan, J., and Riedl, J., Electronic Commerce Recommender Applications. *Journal of Data Mining and Knowledge Discovery*, January, 2001.
- [8] 김영설, 장수현, 윤병주, 「협동적 필터링 을 이용한 전자상거래에서의 추천시스템」, 2000년 정보처리학회 추계학술발표논문집, 제7권 제2호, pp.289-292, 2000.
- [9] <http://www.grouplens.org>.
- [10] <http://www.imdb.com>.



김 영 설

e-mail : kys0716@mju.ac.kr

1999년 명지대학교 컴퓨터공학과 졸업
(공학사)

2001년 명지대학교 대학원 컴퓨터공학과 졸업(공학석사)

2001년~현재 명지대학교 대학원 컴퓨터공학과 박사과정

관심분야 : 전자상거래, 추천시스템, 분산객체, CRM, DRM/DOI /INDECS



김 병 천

e-mail : bckim@hnu.hankyong.ac.kr

1988년 한남대학교 전자계산학과 졸업
(공학사)

1990년 숭실대학교 대학원 전자계산학과
졸업 (공학석사)

1999년 명지대학교 대학원 컴퓨터공학과 졸업 (공학박사)

1991년~현재 한경대학교 컴퓨터공학과 부교수

관심분야 : Machine Learning, Data Mining, Artificial Neural
Networks



윤 병 주

e-mail : yoonbj@mju.ac.kr

1975년 경북대학교 수학과(학사)

1982년 한국과학기술원 전산학과(석사)

1994년 Florida State University 전산학과
(박사)

1982년~현재 명지대학교 컴퓨터공학과 교수

관심분야 : Machine Learning, Data Mining, Hybrid Intelligent
Systems