개인화 추천 시스템에서 연관 관계 군집에 의한 아이템 기반의 협력적 필터링 기술

(An Item-based Collaborative Filtering Technique by Associative Relation Clustering in Personalized Recommender Systems)

정 경 용 † 김 진 현 †† 정 헌 만 ††† 이 정 현 ††††
(Kyung-Yong Jung) (Jin-Hyun Kim) (Heon-Man Jung) (Jung-Hyun Lee)

요 약 추천 시스템은 예전에는 몇몇 혁신적인 전자상거래(E-commerce) 사이트에서만 사용되어 왔으나 현재는 전자상거래를 새롭게 재구성하는 필수적인 비즈니스 도구가 되어가고 있다. 그리고 협력적 필터링은 이론과 실무 분야 모두에서 가장 성공적으로 평가 받은 추천 기법 중 하나이다. 그러나 개인화 추천 시스템을 구축하기 위해서는 두 가지 문제를 동시에 고려해야 한다. 즉 초기 평가 문제와 희박성 문제이다. 본 논문에서는 연관 관계 군집과 연관 규칙의 향상도를 이용하여 이러한 문제를 해결하고자 한다. 사용자의 평가 데이타를 사용하여 아이템간의 향상도를 산출하고, α-cut에 의한 임계값을 아이템들간의 연관성에 적용한다. 연관 관계 군집의 효율성을 높이기 위해서 기존의 Hypergraph Clique Clustering 알고리즘과 본 연구에서 제안하는 Split Cluster Method를 이용하였다. 군집이 완성되면, 각 군집 내부에서 아이템간의 유사도를 산출하고 빠른 액세스를 위해 인덱스를 데이타베이스에저장한다. 새로운 아이템들의 선호도 예측 시에 생성한 인덱스를 적용시킨다. 성능을 평가하기 위해서 기존의 협력적 필터링 기술과 비교 평가하였다. 그 결과 기존의 협력적 필터링 기술의 문제점을 해결하여 예측의 정확도를 높이는데 효과적임을 확인하였다.

키워드 : 협력적 필터링, 추천시스템, 연관규칙

Abstract While recommender systems were used by a few E-commerce sites former days, they are now becoming serious business tools that are re-shaping the world of E-commerce. And collaborative filtering has been a very successful recommendation technique in both research and practice. But there are two problems in personalized recommender systems, it is First-Rating problem and Sparsity problem. In this paper, we solve these problems using the associative relation clustering and "Lift" of association rules. We produce "Lift" between items using user's rating data. And we apply Threshold by -cut to the association between items. To make an efficiency of associative relation cluster higher, we use not only the existing Hypergraph Clique Clustering algorithm but also the suggested Split Cluster method. If the cluster is completed, we calculate a similarity item in each inner cluster. And the index is saved in the database for the fast access. We apply the creating index to predict the preference for new items. To estimate the performance, the suggested method is compared with existing collaborative filtering techniques. As a result, the proposed method is efficient for improving the accuracy of prediction through solving problems of existing collaborative filtering techniques.

Key words : Collaborative Filtering, Recommender System, Association rule

· 본 연구는 2004년도 인하대학교의 지원에 의하여 연구되었음(INHA-31613)
† 정 회 원 : 가천길대학 뉴미디어과 교수
kyjung@gcgc.ac.kr
†† 비 회 원 : (주)동양시스템즈 사원
adorC@shinbiro.com
††† 비 회 원 : 인하대학교 전자계산공학과
hmjung@inhatc.ac.kr
†††† 종신회원 : 인하대학교 컴퓨터공학부 교수
jhlee@inha.ac.kr
논문접수 : 2003년 5월 7일
심사완료 : 2004년 1월 2일

1. 서 론

추천 시스템은 예전에는 몇몇 혁신적인 전자상거래 사이트에서만 사용되어 왔으나 현재는 전자상거래를 새롭게 재구성하는 필수적인 비즈니스 도구가 되었다. 수많은 대형 웹사이트들은 이미 추천 시스템을 사용하여 고객들의 구매활동을 지원하고 있다. 추천 시스템은 고

객이 제공하는 다양한 유형의 정보로부터 학습을 하며 수많은 상품들 중에서 고객이 가장 좋아할 많은 것들을 추천해준다.

전자상거래의 초기에는 기본적인 플랫폼을 다루는 커머스 서버나 보안 솔루션 등이 주로 부각된 반면, 전자상거래내의 경쟁력이 강조되는 현재에는 고객에게 양질의 서비스를 제공하고, 이를 이익창출로 연결시키는데 도움이 되는 보다 다양한 기능의 솔루션들이 부각되고 있다. 특히 CRM과 ERP라는 용어가 자주 거론되었는데, 이는 새로운 고객의 유치보다도 기존의 고객과의 관계를 증진시켜서 평생고객 가치를 지향하고자 하는 요구에 대한 하나의 대안으로 기대를 모았다. 또한 인터넷 사용 사이트를 개설하기만 하면 이익이 창출된다고 생각하기보다 이제는 인터넷이란 새로운 채널에 맞는 새로운 마케팅 정책을 구사하고자 하는 연구가 진행되고 있다. 예를 들어 판매자는 고객들이 자신의 홈페이지를 찾아 오기만을 기다리는 것이 아니라, 방문객을 빠르게 구매자로 변화시키고, 각각의 고객에서 구매할 기회를 최대화하고, 한번 고객이 되면 그 관계를 오래 지속시킬 수 있는 정책적 활동에 대한 필요성이 점차 느끼고 있다.

이에 따라 마케팅의 특화된 솔루션들에 대한 연구가 최근에 진행되고 있고, 관련된 제품들도 인터넷 시장에 등장하고 있다. 리바이스 홈페이지의 "스타일 찾기" 서비스는 사용자가 사이트에서 주어지는 청바지에 대한 질문에 대해서 기준 이상의 개수의 질문에 선호도를 표시하면 선호도에 따라서 청바지 리스트를 제시하는 서비스를 말하는데, 여기서 협력적 필터링 기술[1-5]이 이용되었다. 본 연구에서는 개인화 추천 시스템에서 연관 관계 군집에 의한 아이템 기반의 협력적 필터링 기술을 제안한다. 여기서 유사한 아이템들 간의 군집을 생성한다. 군집내의 아이템들에 대해서 데이타마이닝의 연관 규칙 기법에서 사용하는 Apriori 알고리즘[5]을 적용하여 연관 관계를 추출한 후 협력적 필터링 기술에 의한 개인화 추천 시스템에 적용하였다.

협력적 필터링 기술을 이용한 개인화 추천 시스템에서 군집을 적용하는 장점은 아이템에 대한 평가 데이타가 적더라도 아이템이 속한 군집에 대한 데이타는 많다는 점이다. 본 연구에서는 연관 관계에 의한 군집을 협력적 필터링 기술에 적용하여, 사용자가 평가한 데이타가 적은 아이템은 다른 아이템들과의 향상성(Lift)이 높다는 것을 발견하였다. 이를 이용하면 기존의 협력력 필터링 기술의 희박성 문제(Sparsity)[3,5]를 해결하였다. 그리고 본 연구에서 제안한 연관 관계에 의한 군집의 기본값을 새로운 아이템을 평가하지 않은 사용자들에게 적용하여 협력적 필터링의 초기 평가 문제(First-Rater problem)[3-5]를 해결하였다. 제안한 방법의 성능을 평가하기 위해서 기존의 협력적 필터링 시스템과 비교 평가하였다.

2. 관련 연구

2.1 협력적 필터링 기술

협력적 필터링 기술은 사용자와 유사한 선호도를 가지는 이웃을 찾아내고 사용자들간의 선호도를 평가한 아이템의 선호도를 예측하기 위해서 사용된다. 대표적인 유사도 가중치의 기준 값으로는 피어슨 상관계수, 벡터 유사도 등이 사용된다[3,4]. 이러한 방법을 응용하여 아이템들간의 유사도 가중치를 계산하는데 적용되는데, 엔트로피를 이용한 유사도 가중치, 코사인 기반의 유사도, 기본 선호도 평가, 역 사용자 빈도 등이 사용된다[3-5].

피어슨 상관 계수는 협력적 필터링 기술에서 대표적으로 쓰이는 사용자 유사도 가중치를 계산할 때의 일반적인 방법으로, 사용자가 평가한 데이타에서 오차와 표준편차를 구함으로써 유사도 가중치를 계산한다. 그러므로, 유사도 가중치는 사용자의 평가 데이타의 양에 많은 영향을 받는다. 평가 데이타가 너무 적으면 유사도 가중치를 계산할 때 특정 사용자에게 편중되므로 예측 값이 부정확한 결과가 나타나고, 평가 데이타가 방대하게 많으면 유사도 가중치의 계산량이 많아진다. 특히, 예측하려는 아이템을 평가한 사용자의 개수가 많아지면, 그 계산량은 제곱 승으로 증가하게 된다.

2.2 연관 규칙

연관 규칙은 "사용자의 행동 패턴은 일정한 규칙을 가진다."는 가정 하에 유용한 규칙을 찾아내는 방법이다. 이 방법은 대체로 장바구니 분석과 같은 사용자의 구매 경향을 파악하려는 곳에서 많이 쓰이는 방법이다. 대표적으로 Apriori, FP-Tree, SETM, DIC, ARHP 알고리즘이 있다[7,8].

연관 규칙의 평가 기준으로는 지지도, 신뢰도, 향상도 그리고, α-related가 있다. 이 중, 지지도와 α-related는 상대적으로 적은 평가 데이타에 대해서 부정확한 값을 나타내고, 신뢰도는 방향성을 가지므로 사용할 수 없다[8,9]. 그러므로, 본 연구에서 연관 관계 군집에 의한 협력적 필터링 방법에 사용될 척도 기준으로 향상도를 사용한다.

2.2.1 지지도(Support)와 신뢰도(Confidence)

연관 규칙은 'A→B'로 표현할 수 있다. 여기서 A와 B는 항목들(Items)의 집합이다. 이는 "A를 포함하고 있는 데이타베이스 내의 트랜잭션은 B도 함께 포함하고 있다."는 의미이다. 예를 들어 슈퍼마켓의 판매 데이타로부터 "맥주를 구입하는 고객들은 기저귀도 함께 구매하는 경향이 있다. 그런데 맥주를 포함하고 있는 트랜잭션의 30%는 기저귀를 함께 포함하고 있고, 전체 트랜잭

션 가운데 2%는 맥주와 기저귀 항목을 모두 포함하고 있다."는 연관 규칙 {맥주}→{기저귀}를 발견할 수 있다. 여기서 맥주와 기저귀는 항목이고, 30%는 연관 규칙의 신뢰도(Confidence 또는 Accuracy)이고 2%는 지지도(Support 또는 Coverage)를 의미한다. 연관 규칙을 발견하기 위해서는 지지도와 신뢰도라는 측정 기준이 있어야 하는데, 지지도와 신뢰도는 다음과 같이 정의한다.

$$\text{지지도} = \frac{\#Tuple\ comtaining\ both\ A\ and\ B}{\#Tuple\ containing\ A}$$

$$\text{신뢰도} = \frac{\#Tuple\ comtaining\ both\ A\ and\ B}{total\ \#\ of\ tuples}$$

지지도는 연관 규칙을 반영하는 트랜젝션이 전체 데이타베이스에서 얼마만큼의 비율을 차지하고 있는지를 나타내는 측정 기준으로서 통계적 중요성을 반영한 것이다. 그리고, 신뢰도는 규칙이 실제로 정확한지를 판단하는 정도로서, 항목들 사이의 강도를 나타내는 측정 기준(결합도)으로 사용된다[7].

2.2.2 향상도(Lift)와 α-related

연관 규칙 'A→B'가 의미 있는 규칙이라면, 전체 트랜젝션의 수에서 항목 B를 포함하고 있는 트랜젝션의 비율보다는 항목 A를 포함하는 트랜젝션에서 항목 B를 포함하고 있는 트랜젝션의 비율이 더 클 것이다. 따라서 항목 A와 B를 포함하고 있는 트랜젝션이 서로 상호관련이 없다면(두 항목이 독립이라면), Pr(B|A)는 Pr(B)와 같게 된다. 이를 상대적으로 표현하기 위해서 규칙 'A→B'의 향상도는 실제의 신뢰도를 독립 가정 하에서의 신뢰도로 나눈 값, 즉

$$\frac{\Pr(B|A)}{\Pr(B)} = \frac{\Pr(A \cap B)}{\Pr(A)\Pr(B)}$$

로 정의되며, 이는 실제의 지지도를 독립가정 하에서의 지지도로 나눈 값(식의 우변)과 동일하다. 이를 해석하면 다음과 같은 비율로 표현된다. 상호 대칭적으로 향상도(A→B) = 향상도(B→A)이다.

$$\text{향상도} = \frac{\#Tuple\ comtaining\ both\ A\ and\ B}{\#Tuple\ containing\ A \times \#Tuple\ containing\ B}$$

따라서, Pr(B|A)의 값은 Pr(B)의 값보다 향상도의 배수만큼 크다. 그러므로, 이 값이 1에 가까우면 독립에 가까운 사건, 1보다 크면 두 항목이 양의 연관 관계, 1보다 작으면 두 항목이 음의 연관 관계로 해석을 한다. 따라서, 의미 있는 연관성 규칙을 추출하려면 향상도 값이 1이상이 되어야 할 것이다[9].

α-related는 Hypergraph Clique Clustering 방법[9,10]에서 항목간의 연관 관계의 척도로서 처음 사용되었으며, 다음과 같이 정의한다.

$$\alpha - related = \frac{\#Tuple\ comtaining\ both\ A\ and\ B}{\#Tuple\ containing\ A\ or\ B}$$

α-related는 지지도에서 트랜젝션 항목이 너무나 많거나 항목간의 트랜젝션이 불균등하게 발생할 경우, 정확한 측정을 하지 못하는 문제를 개선한 것으로, 분모의 항목을 데이타베이스의 모든 트랜젝션에서 A이거나 B가 포함된 항목으로 제한한 것이다. 이렇게 함으로써 다른 트랜젝션의 영향을 줄이는 효과가 있다[10].

3. 연관 규칙 군집에 의한 선호도 예측 시스템

그림 1은 본 연구에서 제안하는 연관 관계 군집에 의한 아이템 기반의 협력적 필터링 기술에 의한 개인화 추천 시스템의 전체적인 구성도이다.

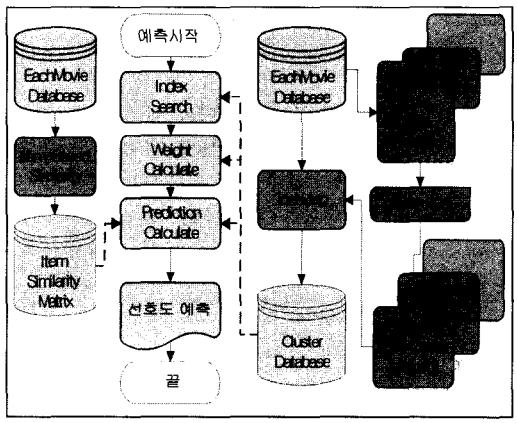


그림 1 연관 관계 군집에 의한 아이템 기반의 협력적 필터링 기술에 의한 개인화 추천 시스템

3.1 연관 규칙을 이용한 아이템의 향상도 측정

Apriori 알고리즘을 이용하여 Eachmovie 데이타에서 사용자가 평가한 아이템들간에 연관 규칙을 찾아낸다. 여기서, 기존의 연구에서는 지지도나 신뢰도를 사용하여 연관 규칙을 찾아냈지만, 본 연구에서는 향상도(Lift)를 사용한다. 향상도는 실제의 신뢰도를 독립 가정 하에서의 신뢰도로 나눈 값으로 식 (1)과 같이 정의한다.

$$L(A,B) = \frac{\Pr(B|A)}{\Pr(B)} = \frac{\Pr(A \cap B)}{\Pr(A)\Pr(B)} \quad (1)$$

식 (1)은 실제의 지지도를 독립 가정 하에서의 지지도로 나눈 값과 동일하다. 또한 상호 대칭적이므로 방향성이 없다. 이 값이 1에 가까우면 아이템 A와 B는 서로 독립에 가깝고, 1보다 작으면 음의 연관 관계, 1보다 크면 양의 연관 관계를 가진다. 그러므로, 아래 3.2절에서 사용할 값은 1이상으로 정한다. Eachmovie 데이타

에 있는 아이템들을 식 (1)을 사용하여 Item Similarity Matrix를 생성한다. 이 노드들은 모든 아이템에 대해 값이 생성되므로 2차원 형태의 메트릭스로 구성할 수 있다. 이 메트릭스는 두 아이템간에 방향성을 가지고 있지 않으므로, 직각 삼각형 구조가 된다. 표 1은 향상도에 의해 구해진 Item Similarity Matrix의 일부이다.

표 1 향상도에 의한Item Similarity Matrix

	Item1	Item2	Item3	Item4	Item5
Item1	2.652	1.072	1.157	1.257	1.057
Item2	-	7.477	1.011	1.211	1.357
Item3	-	-	5.326	1.213	1.257
Item4	-	-	-	4.326	1.754
Item5	-	-	-	-	6.326

Item1: Toy Story, Item2: Jumanji, Item3: Sabrina
Item4: Titanic, Item5: Romeo and Juliet

3.2 α-cut

α-cut은 소속 함수의 [0,1]사이의 값에서 임의의 α ($0 \le \alpha \le 1$)값이 되는 함수 값에 대한 퍼지 상태 변수의 구간을 나타낸다. 이 α-cut은 퍼지 집합의 원소들에 대해 집합에 속할 기준을 정의할 때 사용되는 방법이다. 임의의 X을 원소로 하는 퍼지 집합 A에 대해서 임의의 $\alpha \in [0,1]$값을 가진 α-cut을 적용한 퍼지 집합 A_α는 다음과 같이 정의한다.

$$A\alpha = \{x | A(x) \ge \alpha\}$$

따라서, 퍼지 집합 A_α는 퍼지 집합에 속할 소속정도의 값이 α값 이상으로 이루어진 집합이다. 본 연구에서는 Item Similarity Matrix의 값을 Boolean 값으로 변환시켜서 아이템간의 노드를 만드는 역할을 한다. 또한, 자기 자신과의 관계는 1보다 높으므로 제외시킨다. Item Similarity Matrix에 α-cut을 적용시켜 아이템 군집에 필요한 노드를 생성한다. 여기서, 연관성이 높은지 낮은지에 대한 기준값을 정해야 하는데, 이것이 임계값이다. 향상도에서는 1보다 높으면 양의 연관관계를 가지고, 1보다 낮으면 음의 연관관계를 가진다고 하였으므로 임계값을 1로 설정하여 실험을 진행하였다. 그러나 임계값을 1로 설정하였을 경우, 연관성이 낮은 아이템들간의 군집이 형성이 되고, 많은 아이템들이 하나의 군집에 속하게 되어 상대적으로 군집의 수는 적어지게 된다. 그림 2는 아이템에 대한 사용자와 군집 간의 관계를 임계값 변화에 의해 나타낸 결과이다.

그림 2의 아이템에 대한 사용자와 군집간의 관계에서 임계값을 {3, 6, 9}로 하여 EachMovie 데이타[11]에 적용한 것이다. 여기서 각 "x", "o", "+" 표시는 하나의 아이템을 나타내는 것이다. X축(Person Count)은 그 아이템을 실제로 평가한 사용자의 수이고, Y축(Cluster Count)은 아이템이 속한 군집의 수를 나타낸다. 그림 2에서 실제로 아이템들을 평가한 사용자의 수가 적을수록 많은 클러스터에 아이템이 군집되는 결과를 알 수 있다. 아이템에 대한 사용자와 군집 간의 관계에서 적절한 임계값을 적용하는 것이 중요하다.

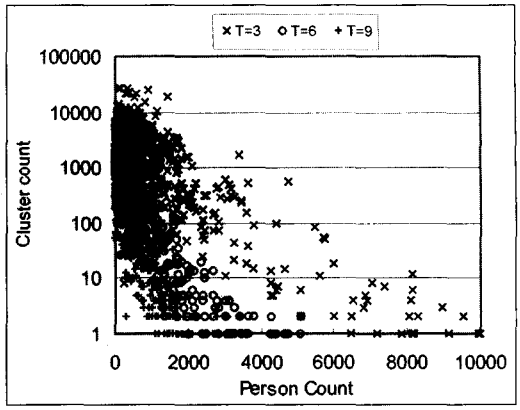


그림 2 아이템에 대한 사용자와 군집 간의 관계

3.3 Split Cluster Method

본 연구에서는 α-cut을 적용하여 만들어진 노드들을 서로 연결하여 연관 관계 그래프를 생성한다. 이 그래프를 이용하여 아이템들간의 연관성을 가진 군집을 생성한다. 기존의 연구에서 제시된 Hypergraph Clique Clustering 방법은 Bottom-up tree 방식의 구조를 사용하는 알고리즘으로 O(n!)이라는 복잡도를 가진다[10]. 그러나 이 알고리즘은 방대한 양의 데이타에서는 성능이 낮은 단점이 있다. 본 연구에서는 엄청난 양의 데이타를 가지고 있는 Eachmovie 데이타에 연관 관계 군집을 하기 위하여 Hypergraph Clique Clustering 알고리즘을 적용하기 어렵다. 그러므로 아이템들의 수를 증가시키면서 군집을 생성하는 기존의 구조에서 변경하여, 하나의 군집에서 작은 군집으로 분리시키는 Split Cluster Method를 본 연구에서는 제안한다. 알고리즘 1은 본 연구에서 제안한 Split Cluster Method에 대한 대략적인 구조를 나타낸 것이다.

이 알고리즘은 삭제된 노드들을 이용하여 전체 하나의 군집에서 단계별로 분리해 나가는 방식을 사용한다. 분리된 군집들은 기존 군집들과 비교하여 포함관계가 성립하면 삭제한다. 이렇게 삭제된 노드들을 모두 이용하여 분할된 최종 군집을 본 연구에서는 사용한다. 알고리즘 1의 복잡도는 $O(n^3)$이다. 그림 3에서는 Split Cluster Method를 이용하여 8개의 노드를 군집하는 과정을 예를 들어 나타내었다. 그림 3에서 각 단계별로 진행될 때마다, 향상도가 낮은 노드들을 이용하여 군집을 분리시

알고리즘 1 Split Cluster Algorithm

```
DC←#Delete Node Classes;
EC←#Equivalence Classes;
while(EC[i] in Each EC) {
pNode[0]←#All Unique Item in EC[i];
while(DC[j] in Each DC) {
  while(pNode[k] in Each pNode) {
    if (pNode[k]⊃DC[j])
      Split pNode[k];
  }
}
while(pNode[k] in Each Splited pNode) {
  while(pNode[l] in Each Unsplited pNode) {
    if (pNode[l]⊃pNode[k])
      Delete pNode[k];
  }
}
pCluster[]←#pNode[] 삽입
Clear pNode[];
}
Assign(pCluster[]);
```

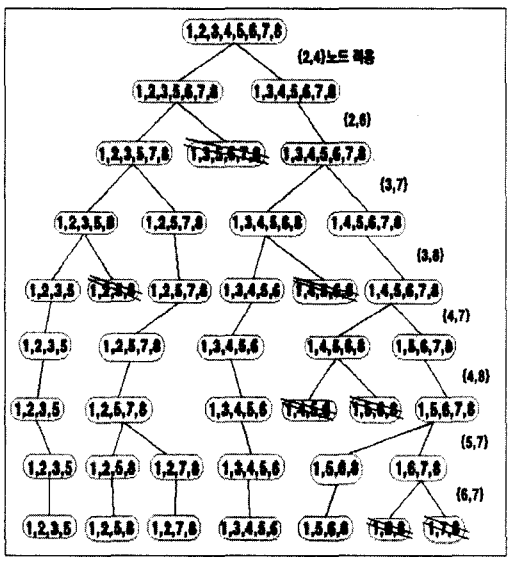


그림 3 Split Cluster Method 적용 예

킨다. 그리고, 분리된 군집들은 분리되지 않은 군집들과 비교하여 포함관계를 조사한다. 만약, 분리된 군집들이 분리되지 않은 군집에 속할 경우 그 군집은 삭제된다. 이렇게하여 향상도가 낮은 노드들을 모두 적용하면 마지막 단계에서 최종 군집이 형성된다. 그림 3의 마지막 단계에서 {{1,2,3,5}, {1,2,5,8}, {1,2,7,8}, {1,3,4,5,6}, {1,5,6,8}}의 다섯 개의 군집이 생성된 것을 볼 수 있다.

3.4 군집 내의 아이템 간의 유사도 계산

3.3절의 Split Cluster Method에 의해 군집이 생성되면, 각 군집 속한 아이템들간의 유사도 가중치를 계산해야 한다. 아이템들간의 유사도 가중치는 협력적 필터링 기술에서 일반적으로 많이 쓰이는 피어슨 상관 계수[3-5]를 사용한다. 기존의 협력적 필터링 방법의 피어슨 상관 계수에 의한 유사도 가중치는 사용자간의 유사도 가중치를 계산할 때 사용하는 방식(사용자 기반의 유사도)이나 본 연구에서는 아이템 간의 유사도 가중치를 계산(아이템 기반의 유사도)하는데 적용한다[12-14,22,23]. 아이템들 간의 유사도 가중치는 식 (2)와 같이 정의한다.

$$sim(x,y)=\frac{\sum_{u\in U}(R_{u,x}-\overline{R_u})(R_{u,y}-\overline{R_u})}{\sqrt{\sum_{u\in U}(R_{u,x}-\overline{R_u})^2\sum_{u\in U}(R_{u,y}-\overline{R_u})^2}} \quad (2)$$

$$where\ -1\le sim(x,y)\le +1$$

식 (2)에서 U는 아이템 x와 아이템 y를 동시에 평가한 사용자들의 집합이다. $R_{u,x}$는 아이템 x에 대해서 사용자들의 집합 U안에 속해 있는 사용자 u의 평가한 선호도이고, $R_{u,y}$는 아이템 y에 대해서 사용자들의 집합 U안에 속해 있는 사용자 u의 평가한 선호도이다. $\overline{R_u}$는 사용자 u의 선호도의 평균값이다.

각 군집에 속한 아이템들 간의 유사도 가중치를 계산한 후, 각 군집을 대표하는 기본값과 가중치를 계산할 수 있다. 기본값은 각 아이템에 대해 사용자들이 평가한 선호도의 평균과 아이템들간의 유사도 가중치에 의해 정의된다. 군집을 대표하는 기본값은 다음 식 (3)과 같이 정의된다.

$$D_a=\frac{\sum_{x,y} sim(x,y)\overline{R_x}\,\overline{R_y}}{\sum_{x,y} 2sim(x,y)} \quad (3)$$

식 (3)에서 D_a는 군집 a를 대표하는 기본값을 나타낸 것이다. 군집 a에 속한 sim(x,y)는 아이템 x와 아이템 y의 유사도 가중치이다. $\overline{R_x}$는 아이템 x의 선호도를 평가한 사람들에 대한 선호도 평균값이고, $\overline{R_y}$는 아이템 x의 선호도를 평가한 사람들에 대한 선호도 평균값이다. 만약 한번도 아이템에 대해서 선호도를 평가하지 않은 사용자에게 아이템이 속한 군집의 기본값(D_a)을 적용하여 협력적 필터링 기술의 초기 평가 문제를 해결하였다.

군집 a에 대한 대표 가중치를 계산하는 식은 식(4)와 같이 정의한다.

$$W_a = \frac{\sum_{x,y}^{N} sim(x, y)}{N} \quad (4)$$

군집 a에 대한 대표 가중치는 군집 내의 아이템들간의 유사도 가중치의 평균으로 계산하게 된다. 군집에 속한 아이템들의 대표 가중치 값은 새로운 사용자가 특정한 아이템에 대해서 선호도를 예측해야 할 경우 적용을 한다. 본 연구에서 제안한 sim(x,y), D_a, W_a는 연관 관계에 의한 군집 a내에 속한 아이템으로 한정한다.

군집 내의 아이템 간의 유사도 계산하는 과정을 그림으로 나타내면 그림 4와 같다.

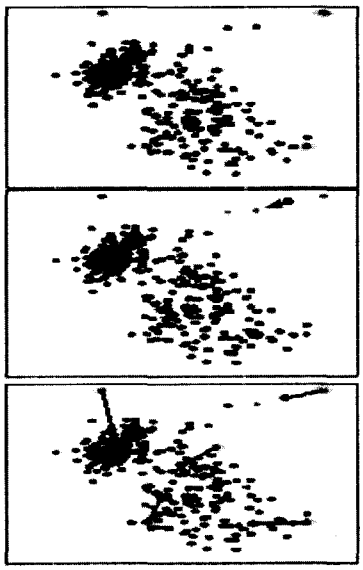


그림 4 군집 내의 아이템 간의 유사도 계산하는 과정

그림 4의 (a)는 군집 a의 아이템들 집합이다. 그리고 그림 4의 (b)는 식 (3)의 D_a에 의한 군집 a를 대표하는 기본값을 나타낸다. 그림 4의 (c)는 식(4)에 의해서 군집 a에 대한 대표 가중치를 계산하는 과정을 나타낸다.

3.5 데이타베이스 저장 및 인덱스 생성

앞에서 제시한 과정을 통하여 군집내의 아이템들간의 유사도 가중치, 군집의 기본값, 군집의 대표 가중치를 계산한다. 그 결과를 데이타베이스에 저장한다. 협력적 필터링 시스템에서 새로운 아이템에 대한 선호도 예측을 실시간으로 처리하게 되면, 많은 계산 시간으로 인해 본 시스템을 이용하는 사용자들은 매우 불편하게 될 뿐만 아니라, 여러 사용자가 동시에 서버에 접속하기 때문에 서버에 많은 부하가 생기게 된다. 본 연구에서는 이 문제점을 해결하기 위해서 계산 시간을 줄이기 위해서 미리 모든 값들을 계산한 후, 데이타베이스에 저장하는 것이다. 그리고 사용자는 평가하기 않은 아이템에 대해서 예측을 해야 할 경우 각 아이템들이 어느 군집에 속해 있는지를 알 수 있는 인덱스를 생성하였다. 이 과정은 3.3절의 Split Cluster Method을 이용하여 군집을 생성할 때 인덱스를 생성하거나 군집 내의 아이템들간의 유사도 가중치를 계산하고 그 결과를 데이타베이스에 저장한 후 인덱스를 생성한다.

3.6 새로운 사용자의 선호도 예측

본 연구에서 예측할 아이템이 속한 군집에 대해서, 아이템들간의 유사도 가중치를 계산하여야 한다. 그러나 아이템들은 군집하면 전체 데이타에 대해서 왜곡이 발생한다[15]. 그래서 이 왜곡을 식 (4)에서 정의한 군집의 대표 가중치를 이용하여 최소화하였다. 군집의 대표 가중치를 이용한 새로운 아이템의 예측을 수식으로 표현하면 다음 식 (5)와 같이 정의한다.

$$P_{i,k} = \overline{v_i} + \frac{\sum_{a}^{A}\sum_{j=1}^{N} sim(i, j)(v_{j,k} - \overline{v_j})}{\sum_{a}^{A} W_a \sum_{j=1}^{N} sim(i, j)} \quad (5)$$

식 (5)의 $P_{i,k}$는 새로운 사용자 i가 아이템 x에 대해서 선호도를 예측하는 식이다. 여기서 $\overline{V_i}$는 사용자 i의 가중치가 부여된 선호도 평균값이다. 군집 a는 아이템 k가 속해있는 군집들을 나타내고, W_a는 군집 a에 대한 평균 가중치를 나타낸다. N은 사용자 a와 다른 사용자들간의 유사도가 0이 아닌 사용자 수이다. sim(i,j)는 사용자 i와 사용자 j의 군집 a에 대한 유사도 가중치이다. $\overline{V_i}$는 사용자 j의 군집 a에 대한 선호도의 평균 값을 나타낸다.

4. 실험 및 성능 평가

본 장에서는 연관 관계 군집을 이용한 협력적 필터링의 알고리즘과 기존의 알고리즘과의 성능 평가를 실험을 통해 알아 본다.

4.1 실험 데이타

실험 데이타로는 컴팩 연구소에서 18개월 동안 협력적 필터링 알고리즘을 연구하기 위해서 영화에 대한 사용자의 선호도를 조사한 EachMovie 데이타[11]를 사용한다. 이 데이타 셋은 총 3개의 텍스트 파일(person.txt, movie.txt, vote.txt)로 구성된다. 세개의 텍스트 파일을 MS SQL Server 2000에서 가져오기("Import")를 하여 데이타베이스에 레코드들을 저장하였다.

이 데이타는 총 72916명의 사용자와 1628종류의 영화에 대해서 0.0에서부터 1.0까지 0.2간격으로 명시적으로 평가한 선호도로 구성되어 있다. 또한 사용자가 실제로 영화를 보았는지의 여부를 알 수 있는 가중치 정보가

존재한다. 영화는 액션, 애니메이션, 외국 예술, 고전, 코미디, 드라마, 가족, 공포, 로맨스, 스릴러의 10개의 장르로 구분되어 있다. 그림 5는 1,612개의 영화에 대한 장르 정보를 담고 있는 데이타베이스의 일부분이며, 그림 6은 영화를 평가한 30,861명 사용자 정보의 일부분이다. 그림 7은 사용자가 영화에 대해서 0.0에서부터 1.0까지

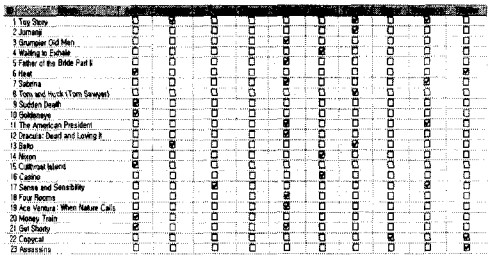


그림 5 영화별 장르 정보 데이타베이스

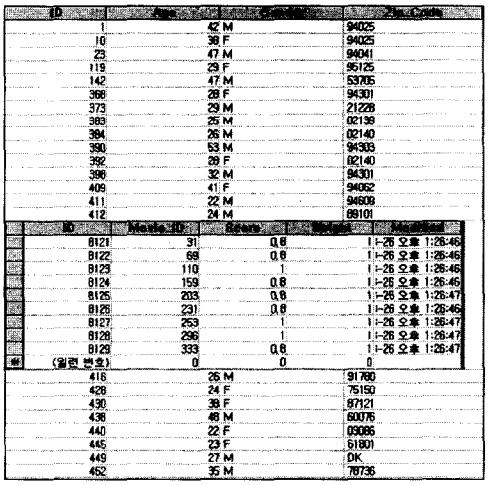


그림 6 평가한 사용자의 정보 데이타

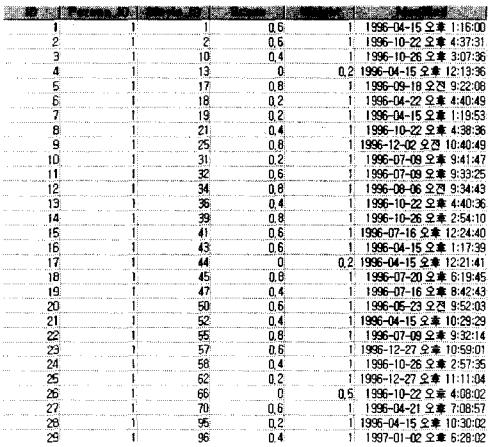


그림 7 영화에 대한 평가 데이타

0.2간격으로 총 6단계로 평가한 데이타의 일부분이다.

4.2 실험 방법 및 결과

본 연구에서 제안한 사용자 선호도 예측 방법은 MS-Visual Studio C++ 6.0으로 구현되었으며, 실제 실험 환경은 PentiumIV, 1.9GHz, 256 MB RAM 환경에서 수행되었다. 실험 방법은 3장에서 제안한 순서대로 하였으며, 연관 관계에 의한 군집에서는 기존의 Hypergraph Clique Clustering(HCC)과 본 연구에서 제안한 Split Cluster Method(SCM)을 실험을 통하여 비교 평가 하였다. 새로운 사용자에 대한 선호도 예측은 기존의 협력적 필터링을 이용한 피어슨 상관 계수(PCC)를 이용한 방식과 본 논문에서 제안한 연관 관계 군집에 의한 협력적 필터링 방식(ARCF) 두 가지를 모두 실험하였다.

본 연구에서는 EachMovie 데이타를 전처리[16]하여 30,861명의 사용자와 1,612종류의 영화에 대해서 실험을 진행하였다. 전처리 작업은 각각의 테이블마다 관계 설정을 하여 무결성 검사를 하였다. Vote 테이블을 기준으로 Movie와 Person 테이블의 무결성 검사를 하였고, Person과 Movie 테이블을 기준으로 Vote 테이블의 무결성 검사를 하였다. 또한, 사용자의 평가 데이타가 1,428,362개이기 때문에 이를 모두 적용하여 실험하기에는 많은 시간이 걸린다. 이를 해결하기 위해서 사용자가 평가한 데이타를 다음과 같이 분할하여 실험을 진행하였다. 즉, Movie 테이블에 있는 영화를 100씩 10개의 군집으로 나누어 실험을 진행하였고, [T10I100D1395k]라는 형식을 사용하는데 이것은 10개의 Transaction, 100개의 아이템, 1395k개의 데이타 로그라는 의미이다[10,15].

표 2 두 알고리즘의 실행 시간[T10I100D1395k]

	Item의 개수				
	30개	40개	50개	60개	70개
HCC	2초	6초	11초	21초	52초
SCM	2초	3초	3초	5초	11초

HCC : Hypergraph Clique Clustering (단위: 초)
SCM : Split Cluster Method

표 2는 Hypergraph Clique Clustering(HCC)와 본 연구에서 제안한 Split Cluster Method(SCM)을 수행시간(단위: 초)에 대해서 실험을 통하여 비교 평가한 결과이다. 아이템의 개수가 증가할수록 Split Cluster Method 알고리즘(SCM)의 성능(수행시간)이 우수하다는 결과를 얻을 수 있다. 그림 7은 Apriori 알고리즘[7,8]으로 생성된 연관 규칙에서 아이템들 간의 "지지도(support)", "신뢰도(confidence)", "향상도(lift)", "α-related"의 값을 나타낸다.

그림 8의 아이템들간의 연관 규칙에서 "지지도", "신뢰도", "향상도", "α-related"는 연관 규칙의 성능을 측정하기 위한 것이다. 본 연구에서는 "향상도"를 사용하여 Item Similairty Matrix을 구성한다. 그림 9는 아이템에 대한 사용자와 군집간의 관계를 나타내기 위해서 Item Similairty Matrix에 임계값이 6인 α-cut을 적용하여 분리해낸 노드들의 집합니다.

그림 10은 그림 9에서 분리해내고 남은 노드들의 집합을 나타낸다. 그림 9의 노드들은 Hypergraph Clique Clustering 알고리즘에 사용하는 노드들이고, 그림 10은 본 연구에서 제안한 Split Cluster Method 알고리즘에 사용하기 위한 노드이다.

그림 11은 연관 관계 군집에 의해 아이템들을 클래스로 나눈 것이고, 그림 12는 연관 관계 군집 안의 아이템들 간의 유사도 가중치를 피어슨 상관 계수를 이용하여 계산한 결과이다.

그림 13에서 Score 필드는 Person 필드의 사용자가 Item 필드의 아이템을 평가한 선호도 값이다. 이는 Eachmovie 데이타의 영화에 대한 평가 데이터(Vote 테이블)의 Score 필드의 값과 동일하다. ARCF 필드와

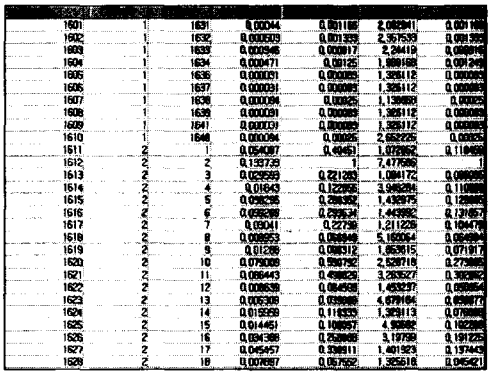


그림 8 아이템들간의 연관 규칙

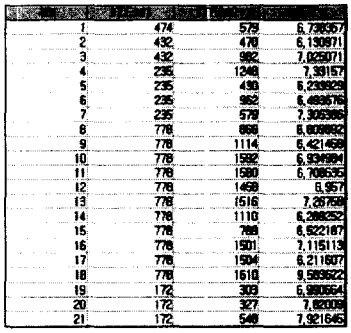


그림 9 Item Equivalence Node

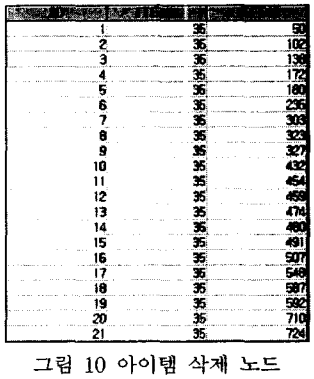


그림 10 아이템 삭제 노드

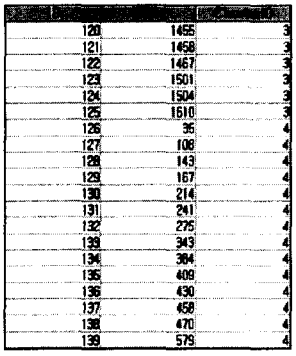


그림 11 아이템들간의 군집

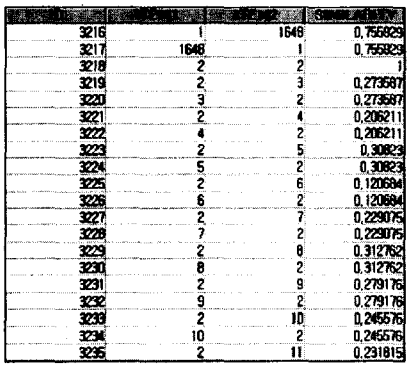


그림 12 아이템들간의 유사도 가중치

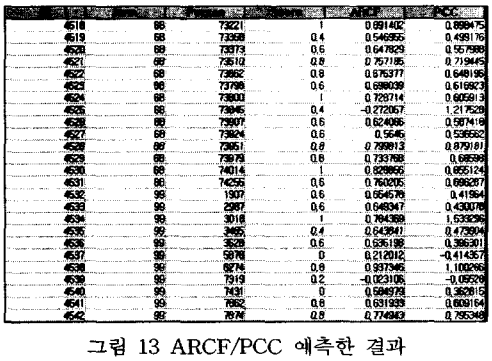


그림 13 ARCF/PCC 예측한 결과

PCC 필드는 본 연구에서 제안한 연관 관계 군집에 의한 협력적 필터링(ARCF) 방법과 기존의 협력적 필터링에 의한 피어슨 상관 계수(PCC)를 이용한 방법으로 예측한 값이다. Score, ARCF, PCC의 필드의 값을 기반으로 하여 4.3절의 분석 및 평가에서 MAE로 제안된 알고리즘의 성능 평가를 하였다.

4.3 분석 및 평가

예측 알고리즘을 평가하는 여러 가지 방법 중에서 예측 값과 실제 값의 차이를 표시하여 정확성 측면에서 성능을 평가하기 위해 MAE(Mean absolute error) 방식[3,4,17-23]을 사용하여 성능 평가하였다. MAE는 예측의 정확성을 판단하는데 가장 많이 쓰이는 방법이다. Error는 실제 선호도 값과 예측된 선호도 값과의 차이로 정의되고 MAE는 Error의 절대값들의 평균을 의미한다. MAE는 절대적으로 알고리즘이 얼마나 정확하게 예측을 했는지를 알 수 있다. 다음은 MAE의 계산식이다.

$$|E| = \frac{\sum_{i=0}^{N} |\varepsilon_i|}{N} \quad (6)$$

식 (6)에서 N은 총 예측 횟수이고, ε_i는 예측 값과 실제 값의 차이를 나타내며, i는 각 예측 단계를 나타낸다.

본 논문에서 제안한 군집의 기본값을 아이템에 대해서 한번도 선호도를 평가하지 않은 사용자에 적용하여 협력적 필터링 시스템의 초기 평가 문제를 해결에 관한 실험이다. 표 3에서 산술평균, 기본 선호도 값, 군집의 기본값에 대한 성능을 비교한 것을 나타낸다. 여기서 기본 선호도 값은 사용자중 어떤 사람도 선호도를 입력하지 않은 새로운 아이템들에 대해서 기본값을 우선적으로 적용함으로써 추천이 가능하도록 할 수 있는 것이다. 대부분의 경우 기본 선호도 값 d는 중립적이거나 다소간 비선호도의 값을 사용하는 경우가 많다. 더 나아가 기본 선호도 값은 암묵적인 선호도 데이타의 경우 웹 페이지의 방문 여부나 어떤 제품의 구매 여부 등과 같이 방문이나 구매를 1로 볼 수 있는 경우 방문하지 않거나 구매하지 않은 경우를 기본 선호도 값을 0으로 설정할 수 있다. 표 3에서 산술 평균은 다른 사용자의 선호도의 산술평균을 나타낸 것이다. 여기서 산술 평균에 관한 연구는 [20]에서 진행하였다.

표 3에서 MAE에 의한 성능을 보면 군집의 기본값을 적용한 것이 가장 성능이 우수함을 보인다. 따라서 본 연구에서는 군집의 기본값(D_a)과 선호도를 평가하지 않은 아이템에 적용함으로 협력적 필터링 시스템에서의 초기 평가 문제를 어느 정도 해결할 수 있었다.

표 3 산술평균/d(기본선호도 값)/군집의 기본값(D_a)

	산술평균	d	D_a
MAE	1.903	1.847	1.686

본 연구에서 식 (6)을 기반으로 제안하는 방식(ARCF)과 협력적 필터링에 의한 피어슨 상관 계수에 의한 방식(PCC)를 실험하여 MAE에 의해 예측의 성능을 평가한 것이다. 표 4는 MAE에 의한 평가에 의한 결과값이다. 그림 14는 표 4를 기반으로 평가한 사용자 수를 증가 시킴에 따른 MAE에 의한 성능 평가를 나타낸다.

한 아이템에 대해 평가한 사용자가 적을 경우는 대체로 상관 관계가 높게 나타나는데, 이것은 그 아이템을 평가한 사용자가 평가한 다른 아이템에 대해 모두 높은 관계를 가지기 때문이다. 이러한 경우 이 아이템이 속한 군집의 크기는 매우 큰 반면, 이 아이템이 속한 군집의 수는 매우 적다. 그래서 군집을 평가한 사용자들이 많아지게 되고, 예측 값의 정확도는 높아진다. 반면, 많은 사용자들이 평가한 아이템에 대해서는 많은 군집에 속하게 되어 평가의 정확도가 낮아진다. 그림 13을 보

표 4 MAE에 의한 성능평가

	P_Count	ARCF	PCC
MAE	2	1.169	1.391
	400	1.038	1.273
	860	0.938	1.122
	1490	0.871	0.958
	2460	0.833	0.926
	2734	0.86	0.92
	3400	0.912	0.94
	4010	0.94	0.952
	5048	1.96	0.975
	6539	1.074	1.152

P_Count : 한 아이템에 평가한 사용자의 수(단위 : 명)
ARCF : 연관 관계 군집에 의한 협력적 필터링
PCC : 피어슨 상관계수에 의한 기존의 협력적 필터링

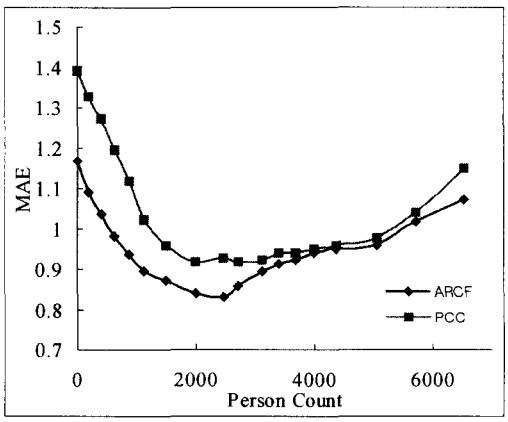


그림 14 ARCF/PCC의 MAE에 의한 성능평가

면, 사용자의 평가 수가 0에서 2,500명인 지점에서 본 연구에서 제안한 알고리즘이 더 좋은 성능을 보임을 알 수 있었다. 그러나 사용자 수가 증가할수록 기존 알고리즘과 비슷한 성능이 보인다. 그러므로, 아이템들이 군집 속에 들어가는 수를 제한하고, 그 수보다 많아질 경우 기존의 피어슨 상관 계수를 이용하여 예측하고, 그 수보다 적을 경우 본 연구에서 제안하는 방법을 사용하면 정확도 면에서 좋은 성능이 나타난다.

본 논문에서 제안한 연관 관계 군집에 의한 아이템 기반의 협력적 필터링 기술은 기존의 협력적 필터링의 단점을 보완하는 역할을 한다. 협력적 필터링에 의한 개인화 추천 시스템은 고객이 시스템에 많은 정보를 제공해야 하며, 고객의 선호에 대한 정보를 축척하기까지 많은 시간이 걸리는 단점을 가지고 있다. 또한 반드시 다른 고객들이 먼저 점수를 부여한 아이템에 대해서만 추천이 가능하다는 제약이 있다[3-5,20,21]. 반면 본 논문에서 제안한 방법은 의미있는 아이템들간의 연관 규칙에 의해 추천을 제공하는 것이다. 즉 기존의 대규모 데이타베이스의 거래 데이타 중 서로 일정 기준점 이상의 지지도와 신뢰도를 만족하는 모든 연관 규칙을 찾아내는 것이다. 따라서 특정 고객의 선호에 대한 정보가 없더라도 단지 아이템간의 연관성에 의해 추천이 가능하다. 그리고 사용자에게 단지 몇 가지 아이템 만을 추천한다고 할 때, 그 사용자가 평가하지 않은 모든 아이템에 대해 예측을 하는 것이 아니라 사용자가 좋아할 만한 아이템 집합을 휴리스틱한 방법에 의해 추출함으로써 데이타베이스 접속횟수와 계산횟수를 줄이는 방법이다. 이런 방식으로 예측을 계산할 아이템 항목의 수를 원천적으로 줄임으로써 추천의 질을 떨어뜨리지 않고도 시스템 반응 속도를 향상시킬 수 있다.

5. 결 론

본 연구에서는 협력적 필터링 시스템에서의 사용자-아이템 행렬의 차원 수를 감소시키기 위하여, 아이템들간의 연관 관계를 이용하여 아이템들을 군집하고, 군집 안에 있는 아이템들에 대해서 평가한 사용자들의 선호도를 기반으로 아이템들을 선호도를 예측하였다. 군집을 적용하여 아이템에 대한 평가 데이타가 적더라도 아이템이 속한 군집에 대한 데이타는 많다는 점을 이용하여 협력적 필터링 시스템의 희박성 문제를 해결하였다. 그리고 본 연구에서 제안한 연관 관계에 의한 군집의 기본값을 새로운 아이템을 평가 하지 않은 사용자에게 적용하여 협력적 필터링 시스템의 초기 평가 문제를 해결하였다. 이는 기존의 연구의 Nave Bayesian 추정치[6]를 이용하는 방법, 기본 선호도 값에 의하여 초기 평가 문제를 해결한 연구보다 성능이 좋은 것을 알 수 있다. 본 연구에서 제안하는 방법을 기존의 협력적 필터링 시스템에 적용하면 선호도 예측의 정확도는 향상되었으나, 평가 데이타가 많아지면 군집의 개수가 엄청나게 늘어나 이를 처리하는 연산 시간도 늘어나게 된다. 연산 시간을 줄이면서 예측의 정확도를 높이는 연구는 향후 연구할 과제이다.

참 고 문 헌

[1] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-based Collaborative Filtering Recommendation Algorithms," In Proc. of the 10th International World Wide Web Conference, Hong Kong, 2001.

[2] J. J. Jung, K. Y. Jung, G. S.Jo, "Ontological Cognitive Map for Sharing Knowledge between Heterogenous Businesses," LNCS 2869, the 18th International Symposium on Computer and Information Sciences, Springer-Verlag, pp.91-98, 2003.

[3] K. Y. Jung, J. K. Ryu, J. H. Lee, "A New Collaborative Filtering Method using Representative Attributes-Neighborhood and Bayesian Estimated Value," Proc. of International Conference on Artificial Intelligence: Las Vegas, USA, June 24-27, 2002.

[4] K. Y. Jung, Y. J. Park, J. H. Lee, "Integrating User Behavior Model and Collaborative Filtering Methods in Recommender Systems," International Conference on Computer and Information Science, Seoul, Korea, August 8-9, 2002.

[5] K. Y. Jung, J. H. Lee, "Prediction of User Preference in Recommendation System using Association User Clustering and Bayesian Estimated Value," LNAI 2557, 15th Australian Joint Conference on Artificial Intelligence(AI'02), Springer-Verlag, pp.284-296, 2002.

[6] 정경용, 김진현, 이정현, "연관 사용자 군집과 베이지안 분류를 이용한 사용자 선호도 예측 방법," 제28회 한국정보과학회 추계학술발표 논문집(II) -우수논문, pp.109-111, 2001.

[7] R. Agrawal, and R. Srikant, "Fast Algorithm for Mining Association Rules," Proc. of the 20th VLDB Conference, pp.487-499, 1994.

[8] E. H. Han, et al., "Clustering Based On Association Rule Hypergraphs," Proc. of SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery, May, 1997.

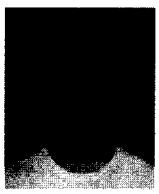
[9] K. Y. Jung, J. H. Choi, K. W. Rim, J. H.Lee, "Development of Design Recommender System using Collaborative Filtering," LNCS 2911, 6th International Conference of Asian Digital Libraries(ICADL'03), Springer-Verlag, pp.100-110, 2003.12.

[10] M. J. Zaki, S. Parthasarathy, M.Ogihara, and W. Li, "New Algorithms for Fast Discovery of

Association Rules," In Proc.of the 3rd IEEE Conference on Knowledge Discovery and Data Mining, pp.283-286, 1997.
[11] P. McJones, EachMovie collaborative filtering dataset, URL:http://www.research.digital.com/SRC/eachmovie, 1997.
[12] D. Billsus, M. J. Pazzani, "Learning Collaborative Information Filters," Proc. of ICML, pp.46-53, 1998.
[13] J. S. Breese, D. Heckerman, and C. Kadie, "Empirical Analysis of Predictive Algorithms for Collaborative Filtering," Proc. of the 14th Conference on Uncertainty in Artificial Intelligence, 1998.
[14] J. Herlocker, J. Konstan, A. Borchers, and J. Riedl, "An Algorithm Framework for Performing Collaborative Filtering," Proc. 2000 ACM-SIGMOD Int. Conf. on Management of Data, 1999.
[15] G. Karypis, "Evaluation of Item-Based Top-N Recommendation Algorithm," Technical Report CS-TR-00-46, Computer Science Dept., University of Minnesota, 2000.
[16] R. Cooley, et al., "Data Preparation for Mining World Wide Web Browsing Patterns," Knowledge and Information Systems, Vol. 1, No. 1, 1999.
[17] 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.
[18] J. Konstan, B. Miller, D. Maltz, J. Herlocker, L. Gordon, and J. Riedl, "GroupLens: Applying Collaborative Filtering to Usenet News," Communications of the ACM, Vol. 40, No. 3, pp.77-87, 1997.
[19] K. Y Jung, Y. J. Na, J. H. Lee, "FDRAS: Fashion Design Recommender Agent System using the Extraction of Representative Sensibility and the Two-Way Filtering on Textile," LNCS 2736, 14th International Conference on Database and Expert Systems Applications, Springer-Verlag, pp.631-640, 2003.
[20] 정경용, 류중경, 강운구, 이정현, "내용 기반 여과와 협력적 여과의 병합을 통한 추천 시스템에서 조화 평균 가중치", 정보과학회논문지 : 소프트웨어 및 응용, 제30권 제3호, pp.239-250, 2003.4.
[21] 정경용, 김진수, 김태용, 이정현, "선호도 재계산을 위한 연관 사용자 군집 분석과 Representative Attribute- Neighborhood을 이용한 협력적 필터링 시스템의 성능향상", 한국정보처리학회(B), 제10-B권, 제3호, pp.287-296, 2003.6.
[22] K. Y. Jung, J. J. Jung, J. H. Lee, "Discovery of User Preference in Personalized Design Recommender System through Combining Collaborative Filtering and Content Based Filtering," LNAI 2843, 6th International Conference on Discovery Science (DS'03), Springer-Verlag, pp.320-327, 2003.
[23] K. Y. Jung, Y. J. Na, J. H. Lee, "Creating User-Adapted Design Recommender System through Collaborative Filtering and Content Based Filtering," LNAI 2902, EPIA'03 International Workshop on Extraction of Knowledge from Data Bases (EKDB'03), Springer-Verlag, pp.204-208, 2003.

정 경 용

2000년 인하대학교 전자계산공학과(공학사). 2002년 인하대학교 전자계산공학과(공학석사). 2002년~현재 인하대학교 전자계산공학과 박사과정. 2001년~현재 에이플러스전자(주) 선임연구원. 2003년~현재 가천길대학 뉴미디어과 겸임교수 관심분야는 웹 마이닝, 기계학습, 정보검색, CRM, 협력적 필터링, 자연어처리, 전자상거래



김 진 현

2000년 인하대학교 전자계산공학과(공학사). 2000년~2001년 (주)웹플러스 사원 2003년 인하대학교 전자계산공학과(공학석사). 2003년~현재 (주)동양시스템즈 사원. 관심분야는 데이타마이닝, 기계학습, 정보검색, 추천시스템

정 헌 만

1996년 서울산업대학교 전자계산공과(공학사). 2001년 인하대학교 전자계산공학과(공학석사). 2001년~현재 인하대학교 전자계산공학과 박사과정. 관심분야는 웹 서비스, 시맨틱웹, RDF, 협력적 필터링, Virtual Reality

이 정 현

1977년 인하대학교 전자공학과 졸업. 1980년 인하대학교 대학원 전자공학과(공학석사). 1988년 인하대학교 대학원 전자공학과(공학박사). 1979년~1981년 한국전자기술연구소 시스템 연구원. 1984년~1989년 경기대학교 전자계산학과 교수 1989년~현재 인하대학교 컴퓨터공학부 교수. 관심분야는 연어처리, HCI, 정보검색, 음성인식, 컴퓨터구조