

연관 사용자 군집과 베이지안 분류를 이용한 사용자 선호도 예측 방법

정경용* 김진현* 이정현*

*인하대학교 전자계산공학과

dragon@nlsun.inha.ac.kr ador@nlsun.inha.ac.kr jhlee@inha.ac.kr

User Preference Prediction Method

Using Associative User Clustering and Bayesian Classification

Kyung-Yong Jung* Jung-Hyun Lee*

*Dept. of Computer Science & Engineering, Inha University

요약

기존의 협력적 필터링 기술을 이용한 사용자 선호도 예측 방법에서는 아이템에 대한 사용자의 선호도를 기반으로 이웃 선정 방법(Nearest-Neighbor Method)을 사용하고, 피어슨 상관 계수에 의해 사용자의 유사도를 구하므로 아이템에 대한 내용을 반영하지 못할 뿐만 아니라 희박성 문제를 해결하지 못하였다. 본 논문에서는 기존의 사용자 선호도 예측 방법의 문제점을 보완하기 위하여 연관 사용자 군집과 베이지안 분류를 이용한 사용자 선호도 예측 방법을 제안한다. 제안한 방법에서는 협력적 필터링 시스템에서의 희박성(Sparsity)문제를 해결하기 위하여 ARHP 알고리즘을 사용하여 사용자들 장르별로 군집하며 새로운 사용자는 Naive Bayes 분류자에 의해 이들 장르 중 하나로 분류된다. 또한, 분류된 장르 내에 속한 사용자들과 새로운 사용자의 유사도를 구하기 위해 Naive Bayes 학습을 통해 사용자가 평가한 아이템에 추정치를 달리 부여한다. 추정치가 부여된 선호도를 기존의 피어슨 상관 관계에 적용할 경우 결측치(Missing Value)로 인한 예측의 오류를 적게하여 예측의 정확도를 높일 수 있다. 제안된 방법의 성능을 평가하기 위해서 기존의 협력적 필터링 기술과 비교 평가하였다.

1. 서론

대부분의 추천 시스템들은 아이템의 수가 많아질수록 사용자가 아이템에 관련된 정보를 얻는데 어느 정도 한계가 있기 때문에 같은 아이템에 대해서 두 사용자간에 선호도를 표시할 확률은 적어지게 되고, 상관관계를 비교 할 아이템의 수는 증가하게 된다. 또한 아이템의 속성에 대한 사용자의 선호도를 직접적으로 반영하지 못하는 문제점도 있다[8].

협력적 필터링 기술에서 고려하기 힘든 부분에 대해서 내용기반 필터링을 이용함으로써 문제점을 해결한다. 보다 좋은 성능을 얻기 위해서는 이러한 필터링 기법들을 결합하고 보완할 필요가 있다. 내용기반 필터링과 협력적 필터링을 결합하여 더 좋은 예측 결과를 얻고자 하는 연구가 최근에 이루어지고 있다[2].

본 논문에서는 기존의 협력적 필터링 시스템에서의 희박성(Sparsity) 문제를 해결하기 위하여 ARHP 알고리즘을 사용하여 사용자들 장르별로 군집하며 새로운 사용자는 Naive Bayes 분류자에 의해 이들 장르 중 하나로 분류된다. 또한, 분류된 장르 내에 속한 사용자들과 새로운 사용자의 유사도를 구하기 위해 Naive Bayes 학습을 통해 사용자가 평가한 아이템에 추정치를 달리 부여한다. 추정치가 부여된 선호도를 기존의 피어슨 상관 관계에 적용할 경우 결측치(Missing Value)로 인한 예측의 오류를 적게하여 예측의 정확도를 높일 수 있다. 제안된 방법의 성능을 평가하기 위해서 기존의 협력적 필터링 기술과 비교 평가하였다.

2. 관련 연구

2.1 협력적 필터링

협력적 필터링에서 사용자와 유사한 선호도를 가지는 이웃을 찾아내고 사용자간에 선호도를 표시한 아이템의 선호도를 예측하기 위해서 사용되는 유사도 기준값으로는 대표적으로 Correlation, Vector based similarity 등이 사용된다[2].

Spearman rank correlation coefficient는 순위가 동등한 개수가 많을 경우 정확도가 감소되는 경향이 있다. 명시적인 선호도인 경우에는 피어슨 상관 계수(Pearson correlation coefficient)의 경우보다 정확도가 떨어진다. Vector based similarity는 명시적인 선호도 데이터보다는 인터넷 사용자가 특정 아이템을 클릭하거나 본 횟수를 이용하는

암묵적인 선호도 데이터들 다룰 경우에 사용자간에 유사도 가중치로 사용한다. 위의 상관계수를 이용하여 유사한 선호도를 가지는 이웃들을 정하고 식(1)에 의해 예측 선호도 값을 계산한다.

$$p_{a,j} = \frac{\sum_i w(a,i)(v_{i,j} - \bar{v}_a)}{\sum_i w(a,i)} \quad \text{식(1)}$$

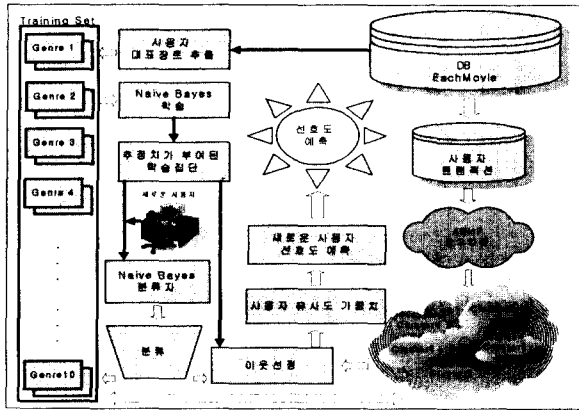
여기서 $p_{a,j}$ 는 사용자 a의 아이템 j에 대한 선호도를 예측한 값이고, \bar{v}_a 는 사용자 a의 선호도 평균값이다. $w(a,i)$ 는 사용자 a와 사용자 i의 유사도 가중치이고, n은 사용자 a와 다른 사용자들간의 유사도가 0이 아닌 사용자수이다. 유사도 가중치 $w(a,i)$ 는 위에서 언급한 식으로 구한다[8].

2.2 Association Rule Hypergraph Partitioning 알고리즘

ARHP 알고리즘은 연관 규칙과 Hypergraph Partitioning을 이용하여 트랜잭션 기반의 데이터 베이스에서 연관된 항목들을 클러스터링하는 방법이다[3][6]. Hypergraph $H=(V, H)$ 는 사용자들로 구성된 정점(vertex)들의 집합 V와 빈번한 항목 집합들을 나타내는 Hyperedge들의 집합 E로 구성된다. Hypergraph Partitioning을 위한 가중치로 연관 규칙의 신뢰도를 사용한다. Hypergraph의 Partitioning은 다단계의 과정을 걸쳐서 진행된다. 이를 Multilevel Hypergraph Partitioning이라고 하는데, Hypergraph를 축소하는 단계(Coarsening Phase), 축소된 Hypergraph에 대하여 Partitioning하는 단계(Partitioning Phase), Partitioning한 Hyperedge를 원래 크기의 Hypergraph로 확대하는 단계(Uncoarsening Phase)를 여러 번 반복하면서 이루어진다. 군집하기 위한 항목집합들의 모든 연관 규칙과 신뢰도를 구한 후, 연관 규칙에 포함되는 항목을 vertex로, 연관 관계를 Hyperedge로 매핑한다. 그리고 신뢰도를 Hypergraph Partitioning을 위한 가중치로 하여, 항목들의 장르를 구한다[6].

3. 제안된 사용자 선호도 예측 시스템

[그림 1]은 본 논문에서 제안하는 연관 사용자 군집과 베이지안 분류를 이용한 사용자 선호도 예측 방법이다.



[그림 1] 연관 사용자 군집과 베이저안 분류를 이용한 사용자 선호도 예측 방법

3.1 ARHP 알고리즘에 의한 연관 사용자 군집

사용자에 의해 선호도가 표시된 아이템들을 사용자 트랜잭션으로 재구성한다. 이를 연관 규칙 탐색 방법을 이용하여 사용자 트랜잭션 안에 빈번하게 동시에 출현하는 사용자들의 집합을 찾는다. 사용자들에 대한 Large 항목집합을 가지고 Apriori 알고리즘[1]을 이용하여 연관 규칙을 생성한다. 그리고 연관 규칙의 신뢰도를 가중치로 하여 Hypergraph Partitioning에 의해 군집시키는 ARHP 알고리즘을 적용함으로써 사용자들 간의 군집을 만든다.

3.2 사용자의 대표장르 추출

사용자가 선호도를 보인 아이템으로 사용자의 대표장르를 추출한다. 대표장르를 추출하기 위해서는 사용자의 장르별 아이템의 선호도 합을 구한 후 선호도의 합이 가장 큰 장르를 대표장르로 정한다. [알고리즘 1]은 사용자의 대표장르를 추출하는 알고리즘이다. 사용자의 대표장르 추출은 훈련집합을 구성할 때 사용한다.

```

[알고리즘 1] 사용자의 대표장르를 추출하는 알고리즘
Num_class ← # of item in GenreID;
MainGenreID ← Null;
MainGenreMaxSum ← 0;
For(j=1; j ≤ Num_class; j++){
    GenreMaxSum ← 0;
    For(each item){
        GenreMaxSum ← GenreMaxSum + Score;
        //아이템에 대해서 장르별 선호도의 합을 구한다.
    }
    If (GenreMaxSum > MainGenreMaxSum){
        MainGenreID ← GenreID of j th;
        MainGenreMaxSum ← GenreMaxSum;
        // 선호도의 합이 가장 큰 장르의 ID를 Return한다.
    }
}
Assign(MainGenreID); // 대표장르 결정
    
```

3.3 Naive Bayes 알고리즘을 적용한 사용자 유사도 가중치

추정치가 부여된 학습집단을 구축하기 위해서는 우선 훈련 집합을 만들어야 한다. 훈련집합은 [알고리즘 1]에 의해서 사용자의 대표 장르를 구한 후 장르별 사용자 군집을 기반으로 아이템들을 장르별로 수집한 데이터이다. 훈련집합의 아이템에 추정치를 부여하기 위하여 Naive Bayes 학습 알고리즘을 사용한다. 본 논문에서는 아이템의 발생여부만 사용하는 방법이 아닌 아이템의 출현빈도를 고려하는 다항 베이저안 학습법을 사용한다[5].

새로운 사용자를 장르별로 분류하기 위해 Naive Bayes 분류자를 사용한다. Naive Bayes 분류자는 추정치가 부여된 학습집단을 사용하여 식(2)에 의해 사용자를 장르별로 분류할 수 있다.

$$GenreID = \underset{Genre \in GenreTot}{\text{arg max}} P(GenreID) \prod_{V_{a,i}} P(nut_{i,GenreID}) \quad \text{식(2)}$$

각 아이템에 대한 선호도 값을 가지는 새로운 사용자는 $u_{new} = \{x \in p \mid nut_1(x), nut_2(x) \dots nut_n(x)\}$ 로 표현하며, $nut_n(x)$ 는 새로운 사용자가 선호도(p)를 표시한 아이템들이다. u_{new} 가 분류될 장르는 GenreID로, 전체 장르는 GenreTot로 표현한다. $P(nut_{i,GenreID})$ 는 사용자가 선호도를 보인 아이템들이 장르에 포함될 확률의 값을 나타낸다. 새로운 사용자

의 장르 결정은 확률 값이 가장 높은 장르에 할당한다

[알고리즘 2]는 아이템에 선호도를 표시한 새로운 사용자에게 식(2)을 사용하여 추정치가 부여된 학습 집단에서 사용자를 장르에 따라 분류하고, Naive Bayes 학습을 피어슨 상관관계수에 적용함으로써 사용자 a와 사용자 i의 유사도 가중치를 구할 수 있다.

사용자가 평가한 아이템의 선호도는 장르별로 추정치가 다르게 적용된다. 이를 위해 사용자가 선호도를 표시한 아이템에 추정치가 부여된 학습 집단을 적용하면 장르별 아이템에 사용자의 선호도 값을 곱한다. 이렇게 함으로써 결측치(Missing Value)로 인한 예측의 오류를 줄일 수 있다. 또한 아이템을 분류하여 분류된 장르에 따라 아이템에 대한 사용자의 선호도에 가중치를 달리 부여하여 사용자의 선호도만을 이용하는 것이 아닌 통계적인 값에 의해 가중치를 부여한다.

[알고리즘 2] Naive Bayes 알고리즘을 적용한 사용자 유사도 가중치

```

/* Naive Bayes 학습을 이용한 훈련집합의 아이템에 추정치 부여 */
TIS ← 전체 아이템의 수;
Data ← 훈련 집합;
For Each Class Variable{
    ut_k ← Item which class variable is GenreID;
    P(GenreID) ←  $\frac{|ut_k|}{|Data|}$ ; //라 클래스에서 아이템이 출현할 확률
    n ← Total Number in GenreID;
    For each ut_k in TIS{
        n_k ← frequency of ut_k in n;
        P(ut_k|GenreID) =  $\frac{n_k + 1}{n + |TIS|}$  //아이템인 ut_k에 추정치를 부여하기 위한 식
    }
}
/* Naive Bayes 분류자를 이용한 새로운 사용자의 장르 분류 */
Num_Class ← # of Item in GenreID;
MainGenreID ← Null;
MainMaxG ← 0;
For(j=1; j ≤ Num_Class; j++){ // Naive Bayes 분류자에 의한 사용자 분류
    For Each Item of Class{
        Prob ← Prob + Cal_W( ut_k ) * Calc_P( ut_k | GenreID );
    }
    G ← Calc_P(GenreID) * Prob;
    If(G > MainMaxG){
        MainGenreID ← GenreID of j th;
        MainMaxG ← G;
        // 가장 큰 G 값을 Return한다.
    }
}
Assign(MainGenreID); // 새로운 사용자 대표 장르 결정
/* 사용자 a와 사용자 i의 유사도 가중치 */
β_a ← 사용자 a의 가중치가 부여된 선호도의 평균값;
β_i ← 사용자 i의 가중치가 부여된 선호도의 평균값;
P_1 ← P_2 ← P_3 ← 0;
For Each Item k in I_a ∩ I_i{
    V_a,k ← 사용자 a가 아이템 k에 대해서 보여준 선호도;
    V_i,k ← 사용자 i가 아이템 k에 대해서 보여준 선호도;
    β_a,k ← P(ut_k|GenreID) * V_a,k;
    β_i,k ← P(ut_k|GenreID) * V_i,k;
    P_1 ← P_1 + (β_a,k - β_a)(β_i,k - β_i);
    P_2 ← P_2 + (β_a,k - β_a)^2;
    P_3 ← P_3 + (β_i,k - β_i)^2;
}
β(a,i) ← P_1 / √(P_2 * P_3); // 사용자 a와 사용자 i의 유사도 가중치
Assign(β(a,i));
    
```

3.4 새로운 사용자의 선호도 예측

새로운 사용자의 선호도 예측은 [알고리즘 2]에서 생성된 사용자 a와 사용자 i의 유사도 가중치를 기존의 협력적 여과 필터링 기술의 피어슨 상관관계수에 적용한다. 이는 사용자의 선호도만을 이용하는 것이 아닌 통계적인 값에 의해 가중치를 부여하기 때문에 예측의 정확도는 향상된다. 특정 아이템에 대한 이웃들의 선호도와 각 이웃들의 선호도 평균과의 거리를 이웃들과의 유사도 가중 평균함으로써 사용자의 아이템에 대한 선호도는 예측된다. 이를 수식으로 표현하면 식(3)과 같이 정의한다.

$$\hat{v}_{a,i} = \bar{v}_a + \frac{\sum_{k \in I_a \cap I_i} \beta(a,i) (v_{i,k} - \bar{v}_i)}{\sum_{k \in I_a \cap I_i} \beta(a,i)} \quad \text{식(3)}$$

$P_{a,k}$ 는 사용자 a의 아이템 k에 대한 추정치가 부여된 선호도를 예측한 값이고, \bar{v}_a 는 사용자 a의 가중치가 부여된 선호도 평균값이다. n은 사용자 a와 다른 사용자들간의 유사도가 0이 아닌 사용자 수이다. 식(1)의 $w(a,i)$ 는 사용자의 선호도만을 사용하여 사용자의 유사도 가중치를 계산하나 $\beta(a,i)$ 는 [알고리즘 2]에 의해서 계산된다.

4. 실험 및 성능평가

본 논문에서 제안한 사용자 선호도 예측 방법은 Visual C++ 6.0으로 구현되었으며, 실제 실험 환경은 Pentium III 450Mhz, 256MB Ram

환경에서 수행되었다. 컴팩 연구소에서 18개월 동안 협력적 필터링 알고리즘을 연구하기 위해서 영화에 대한 사용자의 선호도를 조사한 EachMovie 데이터[4]를 사용한다. 이 데이터는 총 72916명의 사용자와 1628종류의 영화에 대해서 0.0에서부터 1.0까지 0.2 간격으로 명시적으로 평가한 선호도로 구성되어 있다. 영화의 특징은 10개의 장르로 구분되어 있다.

실험 방법은 3장에서 제안된 방식을 3가지로 구분하였다. 첫 번째 방법(P+A)은 기존의 협력적 필터링 기술에 연관 사용자 군집만을 적용한 방법이고 두 번째 방법(P+B)은 아이템을 분류하여 분류된 카테고리에 따라 아이템에 대한 사용자의 선호도에 가중치를 달리 부여하여 아이템의 정보를 반영한다. 마지막 방법(P+A+B)은 연관 사용자 군집 안에서 두 번째 방법을 적용한다.

4.1 실험 방법 및 결과

본 논문에서는 EachMovie 데이터를 전처리하여 30861명의 사용자와 1612종류의 영화에 대해서 실험을 진행하였다. 이는 Naive Bayes 학습을 위한 훈련집합이다. 사용자 프랜잭션에서는 1601개의 연관 규칙과 신뢰도를 생성하였고, 연관 규칙의 평균 길이는 3이다. 500명의 사용자들에 대해 ARIIP 알고리즘 적용결과, 최소 지지도 30%를 만족하는 5개의 사용자 클러스터를 아래 [표 1]과 같이 생성되었다.

[표 1] 클러스터에 포함된 사용자 리스트

| 클러스터번호 | 클러스터에 포함된 사용자 | 사용자수 |
|--------|---|------|
| 1 | {5}[6][10][13][18][19][25][26][27][37][38][46][47][51][52][60][90].. | 114 |
| 2 | {1}[2][3][4][11][12][17][20][28][35][36][45][53][63][68][69][76].. | 89 |
| 3 | {9}[10][16][24][34][42][43][54][55][64][65][66][77][82][85][92][93].. | 104 |
| 4 | {7}[8][15][21][23][29][32][29][50][56][57][70][73][78][79][82][94].. | 105 |
| 5 | {14}[22][30][31][40][41][47][49][58][59][71][72][80][81][89][95].. | 88 |

훈련집합은 ARIIP 알고리즘을 의해서 연관된 사용자 군집을 기반으로 선호도를 표시한 아이템들을 장르별로 만든다. 10개의 장르별로 아이템을 분류한 것은 아래 [표 2]와 같다.

[표 2] 훈련 집합(Training Set)

| GenreID | 사용자가 선호도를 알지 못한 아이템 | 사용자 수 | 영화 수 |
|---------|--|-------|------|
| 1 | Action Goldeneye, Clueless, 12Monkeys, StarGate, StarWars.. | 13500 | 1502 |
| 2 | Animation Toy Story, Exit to Eden, Heavy Metal, Pocahontas.. | 125 | 163 |
| 3 | Art/Foreign Four Rooms, Birdcage, Antonia's Line, Birdcage.. | 385 | 961 |
| 4 | Classic Jumanji, Balto, HappyGilmore, Foreign Student.. | 249 | 1541 |
| 5 | Comedy Ace Ventura, Bronx Tale, Fatal Instinct, Catwalk.. | 4107 | 1545 |
| 6 | Drama Sabrina, Nixon, Ace Ventura, Clerks, Get Shorty.. | 11569 | 1604 |
| 7 | Family Casper, Apollo13, Bad Boys, Batman Forever, Gordy | 158 | 1248 |
| 8 | Horror Copycat, Screamers, Mary Reilly, Babe, Clueless.. | 74 | 453 |
| 9 | Romance American President, Nixon, swiss Family Robinson.. | 166 | 603 |
| 10 | Thriller Die Hard, Taxi Driver, Crimson Tide, The Net.. | 448 | 916 |

훈련집합의 아이템들은 추정치를 부여하기 위해서 Naive Bayes 알고리즘에 의해서 학습한다. 추정치가 부여된 학습집합에서 분류된 장르에 따라 아이템에 대한 사용자의 선호도의 가중치를 달리 부여하여 결측치 값(Missing Value)에 아이템의 정보를 반영한 것은 [표 3]과 같다. 새로운 사용자는 [알고리즘 2]에 의해서 장르가 분류되면, 분류된 장르 내에 속한 사용자들과 새로운 사용자의 유사도를 구하기 위해 추정치가 부여된 학습집합을 통해 사용자가 평가한 아이템에 추정치를 달리 부여한다.

[표 3] 추정치가 부여된 선호도와 예측된 선호도 값

| 아이템 | 클러스터 1 | | | | | |
|-----|--------|--------|--------|--------|--------|--------|
| | 5 | 6 | 10 | 127 | 13 | 18 |
| 1 | 0.0025 | 0.0038 | 0.0098 | 0.0035 | 0.0029 | 0.0097 |
| 19 | 0.0012 | 0.0015 | 0.0024 | 0.0018 | 0.0027 | 0.0075 |
| 21 | 0.0020 | 0.0037 | 0.0034 | 0.0023 | 0.0020 | 0.0037 |
| 35 | 0.0008 | 0.0013 | 0.0006 | 0.0019 | 0.0075 | 0.0026 |
| 45 | 0.0012 | 0.0020 | 0.0017 | 0.0026 | 0.0007 | 0.0047 |

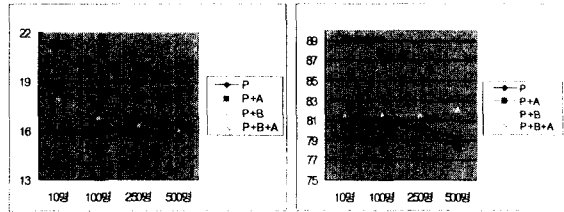
□ : 추정치가 부여된 선호도를 예측한 값

예측 알고리즘을 평가하는 여러 가지 방법 중에서 예측 값과 실제 값의 차이를 표시하는 MAE 방식과 예측할 수 있는 아이템의 전체 대비 비율인 Coverage 방식을 사용하여 성능평가 하였다[2].

[표 4]와 [그림 2]는 제안하는 방식을 10회 반복하여 실험한 예측 값들의 MAE / Coverage를 평균한 값들이다. 본 실험에서는 평가를 백분율로 표시한다.

[표 4] MAE / Coverage

| 알고리즘 | (단위%) | | | |
|------------|---------------|---------------|---------------|---------------|
| | 10명 | 100명 | 250명 | 500명 |
| 기존방식(P) | 18.76 / 81.22 | 18.43 / 80.87 | 18.12 / 80.25 | 17.91 / 79.22 |
| 방법1(P+A) | 18.81 / 81.33 | 17.01 / 81.33 | 16.79 / 80.01 | 16.45 / 81.34 |
| 방법2(P+B) | 17.99 / 81.54 | 16.89 / 81.59 | 16.32 / 81.6 | 16.12 / 82.13 |
| 방법3(P+A+B) | 18.21 / 82.23 | 17.59 / 82.34 | 16.11 / 82.82 | 15.41 / 84.22 |



[그림 2] 성능 비교

[표 4]와 [그림 2]를 보면 기존의 방식과 비교해 볼 때 정확도가 향상되었다. 방법 1과 방법 3은 처음에는 기존 방식과 정확도가 비슷하지만, 사용자가 많아질수록 정확도가 높아지는 것을 볼 수 있다. 사용자의 수가 적을 때에는 연관 사용자 군집의 의미가 크지 않으므로 이 결과와 예측과 부합한다고 할 수 있다. 방법 2는 기존방식과 비교하면 사용자의 수에 관계없이 정확도가 높은 것으로 나타난다. 이는 아이템에 대한 정보를 반영하여 통계적인 값에 의해 가중치를 부여하기 때문에 예측의 정확도는 향상된다.

Coverage는 연관 사용자 군집만을 적용했을 경우 다른 방법보다 심각하게 감소하는 경향이 보인다. 그러나 군집을 이용하여 연관할 수 있는 데이터 양이 줄어들기 때문에 연관 시간이 줄어드는 것을 알 수 있다. 전체적인 EachMovie 데이터의 예측 결과를 보면 연관 사용자 군집과 베이지안 분류를 이용한 방법이 기존의 피어슨 상관계수만을 적용했을 경우보다 우수한 결과를 보이나 연관 사용자 군집만을 적용한 경우에는 Coverage가 줄어든 문제가 있으므로 이를 보완해야 할 것으로 보인다.

5. 결론

본 논문에서는 연관 사용자 군집을 적용하여 사용자들을 군집하고, 이 군집에 속한 사용자들의 선호도를 기반으로 사용자의 선호도에 가중치를 달리 부여하여 아이템에 대한 정보를 반영하여 통계적인 값에 의해 가중치를 부여하여 사용자의 선호도를 예측하는 방법을 제안하였다. 제안한 방법에 대한 성능을 기존의 협력적 필터링 기술과 비교 실험한 결과 예측의 정확도가 향상되었기 때문에 본 논문에서 제안한 방식이 효과적임을 알 수 있었다.

향후 과제로는 사용자 유사도 간의 상관 계수에 대한 연구와 영화의 장르뿐만 아니라 아이템의 속성들을 선호도에 이용한다면 좋은 결과를 기대할 수 있을 것이다.

6. 참고문헌

[1] R. Agrawal and R. Srikant, "Fast Algorithm for Mining Association Rules," Proc. of the 20th VLDB Conference, pp. 487-499, 1994.
 [2] 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.
 [3] G. Karypis and V. Kumar, "Multilevel k-way Hypergraph Partitioning," DAC, pp.343-348, 1999.
 [4] P. McJones, EachMovie collaborative filtering dataset, url:http://www.research.digital.com/SRC/eachmovie,1997.
 [5] T. Michael, *Maching Learning*, McGraq-Hill, pp.154-200, 1997.
 [6] E. H. Han, et. al., "Clustering Based On Association Rule Hypergraphs," DMKD, 1997.
 [7] M. O'Connor and J. herlocker, "Clustering Item for Collaborative Filtering," 2000.
 [8] 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.