

협업적 여과 시스템의 성능 향상을 위한 장르 패턴 기반 사용자 클러스터링

최자현*, 하인애*, 홍명덕*, 조근식**

GGenre Pattern based User Clustering for Performance Improvement of Collaborative Filtering System

Jahyun Choi*, Inay Ha*, Myung-Duk Hong*, Geun-Sik Jo**

요약

협업적 여과 시스템은 사용자에게 대한 클러스터링을 구축한 후, 구축된 클러스터를 기반으로 사용자에게 아이템을 추천한다. 그러나 사용자 클러스터링 구축에 많은 시간이 소요되고, 사용자가 평가한 아이템이 피드백 되었을 경우 재구축이 쉽지 않다. 본 논문에서는 영화 추천 시스템에서의 사용자 클러스터링의 재구축 시간을 단축시키기 위해서 빈발 패턴 네트워크를 이용하여 사용자가 선호하는 장르 패턴을 추출하고, 추출된 패턴을 통해 사용자 클러스터링을 구축한다. 구축된 사용자 클러스터링을 협업적 여과에 적용하여 사용자에게 영화를 추천한다. 사용자 정보가 피드백 될 때, 전통적 협업적 여과는 사용자 클러스터링을 재구축하기 위해 모든 이웃 사용자를 재탐색하여 클러스터링 한다. 하지만 빈발 패턴 네트워크를 이용하여 장르 패턴 기반의 사용자 클러스터링을 적용한 협업적 여과는 사용자 클러스터링을 재구축시 사용자 탐색 공간을 국한시킴으로써 탐색 시간을 줄일 수 있다. 제안하는 장르 패턴 기반의 사용자 클러스터링을 통해 사용자 정보가 피드백 된 후 사용자 클러스터를 재구축시 소요되는 시간을 줄일 수 있고, 전통적인 협업적 여과 시스템과 유사한 성능의 추천이 가능하게 되었다.

▶ Keyword : 협업적 여과, 빈발패턴네트워크, 사용자 클러스터링

Abstract

Collaborative filtering system is the clustering about user is built and then based on that clustering results will recommend the preferred item to the user. However, building user clustering is time consuming and also once the users evaluate and give feedback about the film then rebuilding the system is not simple. In this paper, genre pattern of movie recommendation systems

• 제1저자 : 최자현 • 교신저자 : 최자현

• 투고일 : 2011. 08. 30, 심사일 : 2011. 09. 25, 게재확정일 : 2011. 10. 20

* 인하대학교 컴퓨터·정보공학과(Department of Computer · Information Engineering, Inha University)

** 인하대학교 컴퓨터·정보공학과(School of Computer · Information Engineering, Inha University)

is being used and in order to simplify and reduce time of rebuilding user clustering. A Frequent pattern networks is used and then extracts user preference genre patterns and through that extracted patterns user clustering will be built. Through built the clustering for all neighboring users to collaborative filtering is applied and then recommends movies to the user. When receiving user information feedback, traditional collaborative filtering is to rebuild the clustering for all neighbouring users to research and do the clustering . However by using frequent pattern Networks, through user clustering based on genre pattern, collaborative filtering is applied and when rebuilding user clustering inquiry limited by search time can be reduced. After receiving user information feedback through proposed user clustering based on genre pattern , the time that need to spent on re-establishing user clustering can be reduced and also enable the possibility of traditional collaborative filtering systems and recommendation of a similar performance.

▶ Keyword : Collaborative Filtering, Frequent Pattern Networks, User Clustering

I. 서 론

오늘날 방대한 양의 데이터베이스와 다양한 형태의 웹 응용프로그램을 기반으로 한 전자상거래의 발전으로 추천시스템의 필요성이 더욱 증대되고 있다. 고객맞춤화나 개인화 서비스는 인터넷 서비스 제공자들의 중요한 성공요인으로 인식되고 있다[1]. 추천시스템이 사용자를 대신하여 사용자에게 적합한 아이템을 빠른 시간 내에 추천하고 그 추천 내용 또한 정확하다면 사용자는 만족감을 얻을 수 있다. 하지만 아무리 정확하다고 해도 추천 시간이 오래 걸린다면 그 추천 시스템은 문제가 있다. 그래서 추천시스템이 효율적으로 사용자에게 아이템을 추천해 줄 필요가 있다. 추천시스템을 위한 여러 방법 중 대표적인 것이 협업적 여과(Collaborative Filtering)[2]이다. 협업적 여과는 사용자에 대한 클러스터링을 구축 후, 구축된 클러스터를 기반으로 사용자에게 아이템을 추천한다. 하지만 사용자 클러스터링 구축에 많은 시간이 소요되고, 사용자가 평가한 영화가 피드백 될 때마다 사용자 클러스터링을 재구축하는데 시간이 많이 소비된다. 재구축 시간은 협업적 여과를 추천시스템에 활용하는데 많은 방해요소가 된다. 본 논문에서는 사용자 클러스터링의 재구축 시간을 단축하기 위해 빈발 패턴 네트워크(Frequent Pattern Networks)[3]를 이용하여 사용자 클러스터링을 구축하고 협업적 여과에 적용하였다. 빈발 패턴 네트워크는 아이템 사이에 존재하는 연관규칙을 찾아 클러스터를 생성한다. 빈발 패턴 네트워크로 생성된 클러스터로 탐색 공간을 국한시킴으로써 협업적 여과에서 사용자 유사도를 계산하는 시간을 줄일 수 있다. 본 논문에서 제안하는 사용자 클러스터링은 사용자가 평가한 영화의 장르를 분석하여 빈발 패턴 네트워크를 구

성하고 생성된 패턴을 통해 사용자가 선호하는 장르 패턴을 추출할 수 있다. 추출된 장르 패턴과 동일한 사용자를 클러스터링하고, 이를 기반으로 협업적 여과를 수행한다. 빈발 패턴 네트워크는 규칙의 조건부에 속성이 하나 추가될 때마다 훈련 데이터를 재탐색할 필요가 없어 사용자가 평가한 영화가 피드백 되었을 경우, 빈발 패턴 네트워크를 이용하면 정보를 쉽게 누적시켜 패턴을 추출할 수 있게 된다[3]. 제안된 방법을 통해 사용자 클러스터링의 재구축시 소요되는 시간을 감소시키면서 효율적으로 추천을 할 수 있다.

본 논문의 구성은 다음과 같다. 2장에서는 영화 장르를 고려하여 목표 사용자에게 영화를 추천한 연구와 연관 규칙을 이용하여 영화를 클러스터링을 하여 사용자에게 영화를 추천한 연구에 대해 기술한다. 3장에서는 협업적 여과 시스템의 성능 향상을 위한 장르 패턴 기반 사용자 클러스터링에 대해 기술하고, 제안하는 시스템의 실험 및 결과에 대한 부분으로 실험 데이터에 대한 내용과 제안방법에 대한 성능 평가를 수행한다. 4장에서는 결론 및 향후 연구 방향에 대해 기술한다.

II. 관련 연구

2. 영화의 장르 속성을 이용한 추천시스템

협업적 여과 시스템[4]은 사용자들이 아이템들을 경험하고 부여한 선호 점수들을[5] 기반으로 목표 사용자가 평가하지 않은 아이템들 중에서 높게 평가할 것이라고 예상되는 아이템을 추천한다.

영화 추천시스템을 위해 영화 장르를 고려한 연구가 진행되고 있다[4][6][7]. [4]은 사용자-아이템 매트릭스를 사용자-분류 매트릭스로 전환하여 사용자가 선호하는 장르를 분석

한다. 선호 장르가 같은 사용자 집단을 이용하여 사용자 클러스터링을 하고, 목표 사용자에게 영화를 추천한다. 예를 들어, 영화 “아저씨”가 “Drama, Thriller”로 되어 있는 장르에 속해 있다면, 이 두 장르로 구성되어 있는 영화를 평가한 사용자들을 클러스터링 한다. 목표 사용자가 가장 많이 평가한 장르로 구성되어진 클러스터 내의 영화를 목표 사용자에게 영화를 추천한다. 즉, 영화마다 장르 구성이 달라 장르 구성별 사용자 클러스터링의 수는 늘어나게 된다. 영화의 종류가 많아질수록 클러스터링에 소요되는 시간이 오래 걸린다는 문제점이 있다. [7]은 영화가 지닌 장르 조합을 기반으로 장르 상관도를 계산하고, 사용자의 선호하는 장르 조합과 각 영화가 지닌 장르 조합의 상관도를 적용하여 영화의 평균 선호도를 재계산한다. 평균 선호도가 높은 영화들 중 상위 N개를 목표 사용자에게 추천한다. 하지만 시대별 장르의 조합이 달라짐에 따라 영화 추천 효율이 저하되는 문제점이 있다. [8]은 전통적인 협업적 여과를 이용하여 사용자 클러스터링을 구축한 후, 구축된 클러스터를 이용하여 이웃 사용자가 평가한 영화가 속해 있는 장르의 점수를 계산한다. 장르의 점수가 높은 상위 N개의 장르를 선정하고, 선정된 장르에 속해 있는 영화를 목표 사용자에게 추천하였다. 이 논문의 경우, 단일 장르에 소속된 경우엔 추천 성능이 높지만, 다중 장르의 경우에는 추천 성능이 저하됨을 알 수 있었다. 그리고 위와 같은 연구는 사용자가 평가한 영화가 피드백 될 경우, 피드백 된 정보를 반영하여 사용자 클러스터링을 구축하는 것이 용이하지 않다. 본 연구에서는 재구축이 용이하게 빈발 패턴 네트워크를 이용하여 사용자 클러스터링을 한다.

2.1 연관 규칙을 이용한 영화 추천 시스템

연관규칙은 데이터들의 빈도수와 동시 발생 확률을 이용하여 한 항목의 그룹과 다른 항목의 그룹 사이에 존재하는 연관성을 규칙의 형태로 표현한 것이다[8]. 연관규칙을 찾기 위한 알고리즘들 중 가장 많이 쓰이는 Apriori 알고리즘[9]이 있으며, 이 알고리즘은 최소지지도를 만족하는 모든 빈발 항목 집합을 생성한다. 하지만 패턴의 길이가 길 경우, 생성된 후보 항목 집합(Candidate Set)이 최소지지도를 만족하는가를 확인하기 위해 지속적으로 데이터베이스에 접근한다. 빈발 항목 집합을 생성하는데 걸리는 시간비용을 줄이기 위해 빈발 패턴 트리, 빈발 패턴 네트워크[3] 알고리즘이 연구되었다.

영화 추천 시스템에 연관규칙을 적용한 연구가 진행되고 있다. [2]은 빈발 패턴 트리를 이용하여 영화간의 연관 규칙을 생성하고, 생성된 규칙을 이용하여 영화를 클러스터링 한다. 구축된 클러스터를 통해 사용자에게 영화를 추천한다. 하지만 클러스터링 구축 시간이 전통적 협업적 여과보다 오래

걸린다는 문제점이 있다. 이는 빈발 항목 생성을 위해 빈발 패턴 트리의 삽입, 삭제가 반복적으로 일어나기 때문에 연산 시간이 길어지기 때문이다. 본 연구에서는 빈발 항목 생성에 드는 시간을 줄이기 위해 빈발 패턴 네트워크를 이용하였다.

3. 협업적 여과 시스템의 성능 향상을 위한 장르 패턴 기반 사용자 클러스터링

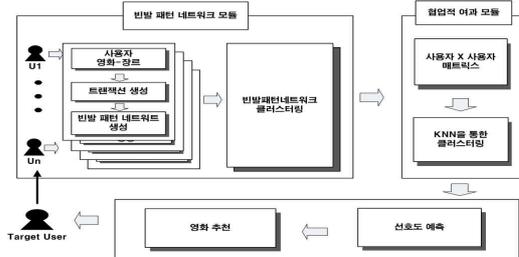


그림 1. 시스템 구조
Fig. 1. System Architecture

본 논문에서 제안하는 구조는 그림 1과 같이 빈발 패턴 네트워크 모듈, 협업적 여과 모듈, 선호도 예측 및 추천으로 구성되어있다. 빈발 패턴 네트워크 모듈에선 목표 사용자가 평가한 영화 장르를 분석하여 트랜잭션을 구성한다. 구성된 트랜잭션을 기반으로 빈발 패턴 네트워크를 구축하고, 구축된 빈발 패턴 네트워크를 이용하여 목표 사용자가 선호하는 장르 패턴을 추출한 후, 추출된 장르 패턴 기반의 사용자 클러스터를 구축한다. 그리고 협업적 여과 모듈에서는 구축된 사용자 클러스터를 기반으로 최근접 이웃을 선정하고, 선정된 이웃 사용자를 이용하여 목표 사용자가 평가하지 않은 영화의 선호도를 예측한다. 목표 사용자의 예측된 선호도 점수 중에서 점수가 높은 상위 N개의 영화를 목표 사용자의 추천 목록으로 생성하여 목표 사용자에게 영화를 추천한다.

3.1. 장르 패턴 기반의 사용자 클러스터링

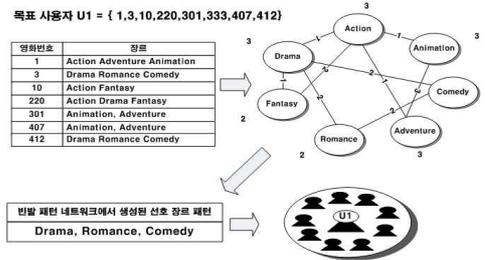


그림 2. 장르 빈발 패턴 추출
Fig 2. Genre Frequent Pattern Extraction

그림 2와 같이 데이터베이스에 축적된 목표 사용자 UI이 평가한 영화가 속한 장르들을 이용하여 사용자 클러스터링을 구축하고, 구축된 사용자 클러스터를 통해 협업적 여과에 적용한다.

빈발 패턴 네트워크의 구성은 다음과 같다.

아이템을 표현하기 위한 정점(Vertex)은 아이템의 이름을 갖고, 정점의 빈도수는 아이템 발생 빈도수를 표현한다. 두 아이템 집합의 표현은 간선(Edge)으로 구성하고, 간선의 빈도수는 두 정점(아이템)에 동시 발생하는 빈도수를 나타낸다.

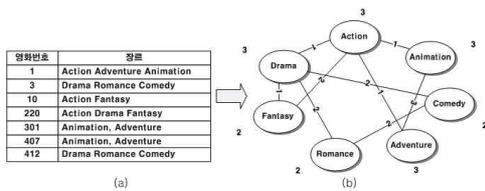


그림 3. 빈발 패턴 네트워크
Fig. 3. Frequent Pattern Network

그림 3은 빈발 패턴 네트워크의 간단한 예이다. (a)는 데이터베이스에 목표 사용자가 평가한 영화가 속해 있는 장르의 종류가 트랜잭션으로 구성되어 축적되고, (b)는 트랜잭션을 빈발 패턴 네트워크를 적용했을 때의 모습이다. 아이템은 원으로 표현하고, 두 아이템을 잇는 선은 간선으로 표현한다. 아이템 위의 숫자는 트랜잭션에 나타나는 아이템의 빈도수이다. 예를 들어 Action 위의 숫자 3은 트랜잭션에 Action이라는 이름을 가진 아이템이 3번 존재함을 의미한다. Action과 Fantasy를 잇는 간선 위의 숫자 2는 트랜잭션에 Action과 Fantasy가 동시에 발생한 경우가 2번 존재함을 의미한다.

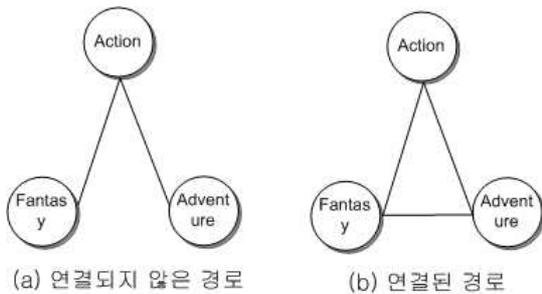


그림 4. 연결된 경로
Fig. 4. Connected Path

장르 패턴의 구성은 그림 4의 (b)와 같이 연결된 경로만을 패턴으로 구성한다. (a)는 네트워크의 Fantasy와 Adventure 사이의 간선이 존재하지 않기 때문에 연결된 경로가 아니다. 연결된 경로가 되려면 (b)와 같이 Fantasy와 Adventure

사이의 간선이 존재해야 한다. 즉, 연결된 경로가 되려면 그래프는 완전히 연결되어야 하고, 이렇게 연결된 경로는 하나의 패턴이 될 수 있다.

표 1은 그림 3의 빈발 패턴 네트워크를 통해서 생성된 장르 패턴 목록이다. 장르 패턴의 빈도수가 2이상인 장르 패턴만이 목록에 추가된다.

표 1의 장르 패턴 목록을 통해서 목표 사용자 UI이 가장 선호하는 장르 패턴을 추출할 수가 있다.

표 1. 장르 패턴 목록
Table 1. Genre Pattern List

장르 패턴	빈도수
Drama, Romance, Comedy	2
Animation, Adventure	3
Action, Fantasy	2

다음은 사용자가 선호하는 장르 패턴을 선택하는 방법이다.

첫째, 선호 장르 패턴을 선택할 때 장르 패턴의 차수가 높은 것을 우선적으로 선택을 한다. 이것은 가장 높은 차수를 가진 장르를 포함하는 패턴이 많은 장르와 연결되어 사용자가 선호할 패턴이 될 확률이 높기 때문이다. 여기서 ‘차수’는 장르 패턴을 구성하는 장르의 개수를 말한다.

둘째, 선택해야 하는 장르 패턴의 차수가 같다면 장르 패턴의 빈도수가 높은 것을 선택한다.

셋째, 장르 패턴의 차수가 같고, 장르 패턴의 빈도수도 같다면 장르 패턴을 구성하는 장르들의 빈도수를 합하여 장르 빈도수가 높은 것을 선택하여 적용한다.

넷째, 장르 패턴을 구성하는 장르들의 빈도수를 합한 값도 같다면 먼저 발생한 패턴을 선택 적용한다.

빈발 패턴 네트워크를 통해 선택된 장르 패턴은 사용자가 가장 선호한다는 것을 의미한다.



그림 5. 장르 패턴 기반의 사용자 클러스터링
Fig. 5. User Clustering based on Pattern of Genre

장르 패턴 기반의 사용자 클러스터링을 구축할 때, 선호 장르 패턴이 같은 이웃 사용자들과 클러스터링을 구축한다. 예를 들어, 그림 5의 빈발 패턴 네트워크에서 생성된 선호 장르 패턴 (Drama, Comedy, Romance)을 가진 이웃 사용자들을 클러스터링한다. 또한 선호 장르 패턴의 부분집합인 (Drama, Romance), (Romance, Comedy), 그리고 (Drama, Comedy)의 패턴을 가진 이웃 사용자들이 있다면 같은 클러스터에 포함 시킨다. 즉, 목표 사용자 UI이 선호하

는 장르 패턴의 부분집합을 가진 이웃 사용자들은 모두 목표 사용자 U1과 클러스터링 된다.

사용자 클러스터링은 사용자 개개인마다 사용자가 선호하는 장르 패턴 기반의 이웃 사용자들과 클러스터링이 구축된다.

사용자 클러스터 구축 후 사용자 정보가 피드백이 되었을 경우 다음과 같이 진행된다.

그림 6과 같이 목표 사용자 U1의 평가 정보가 피드백 되었을 경우, 데이터베이스에 피드백 정보를 누적하고, 추가된 정보로 (a)와 같이 트랜잭션을 생성한 후, 빈발 패턴 네트워크에 추가한다. 빈발 패턴 네트워크는 규칙의 정보가 추가될 경우 모든 트랜잭션을 재탐색을 하지 않고, 구축된 빈발 패턴 네트워크에 추가된 정보만 누적된다. (b)와 같이 추가된 트랜잭션 Animation과 Drama 사이의 간선 edge(Animation, Drama)가 추가되었고, 아이템(장르)의 발생 빈도수도 증가된다.

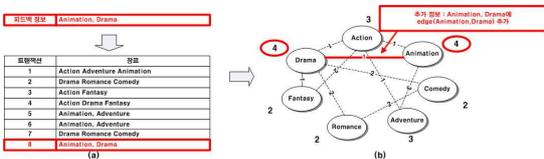


그림 6. 빈발 패턴 네트워크 재구축
Fig. 6. Frequent Pattern Network Rebuilding

장르 패턴의 빈도수가 2이상이면 선호 장르 패턴 목록에 추가된다. 만약 선호 장르 리스트 중에 추가된 정보들로 인하여 최상위의 선호 장르 패턴이 바뀌면 사용자 클러스터링을 다시 구축한다.

3.2. 장르 패턴 기반의 사용자 클러스터링을 적용한 협업적 여과

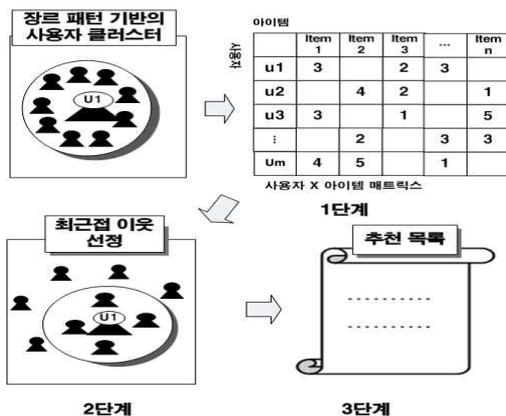


그림 7. 협업적 여과 3단계
Fig. 7. Collaborative Filtering 3 Step

협업적 여과는 장르 패턴 기반의 사용자 클러스터를 통해 수행된다. 협업적 여과는 다음 3단계를 거쳐 목표 사용자에게 영화를 추천한다.

1단계, 목표 사용자가 선호하는 장르 패턴 기반으로 클러스터링 된 사용자들이 평가한 선호도 점수를 이용하여 사용자×영화 매트릭스를 구성한다.

$$w(A, B) = \frac{\sum_{i=1}^q (R_{Ai} - \bar{R}_A)(R_{Bi} - \bar{R}_B)}{\sqrt{\sum_{i=1}^q (R_{Ai} - \bar{R}_A)^2 \sum_{i=1}^q (R_{Bi} - \bar{R}_B)^2}} \quad (1)$$

사용자×영화 매트릭스를 기반으로 식(1)과 같이 피어슨 상관 계수(Pearson correlation coefficient)를 이용하여 클러스터 내 사용자의 유사도를 계산한다.

2단계, 클러스터 중 목표 사용자와 유사도가 높은 K명의 최근접 이웃을 선정한다.

3단계, 선정된 최근접 이웃의 선호 점수를 기반으로 사용자가 평가하지 않은 영화에 대해 식(2)를 이용하여 선호도 점수를 예측한다.

$$R_{A,i} = \bar{R}_A + \frac{\sum_{j=1}^k w(A_i)(R_{ji} - \bar{R}_j)}{\sum_{j=1}^k |w(A_j)|} \quad (2)$$

목표 사용자의 예측된 선호도 점수 중에서 점수가 높은 상위 N개의 영화를 목표 사용자의 추천 목록으로 생성하여 목표 사용자에게 영화를 추천한다.

3.3. 협업적 여과의 성능 측정

본 논문에서는 추천 평가 지표를 고려하여 추천시스템의 성능을 측정을 하였다.

Top-N 기법에 의해 생성된 추천 목록의 평가에 사용되는 두 가지 측정지표에 재현율(Recall)과 정확도(Precision)가 사용된다.

재현율은 목표 사용자가 실제로 경험하여 평가한 아이템들 중에서 추천시스템이 생성한 추천목록에 속하는 아이템의 비율로 식(3)과 같이 계산된다.

$$Recall = \frac{|T \cap R|}{|T|} \quad (3)$$

T집합은 사용자가 실제로 경험하여 평가한 아이템들 중에서 추천시스템의 성능측정을 위해 따로 모아둔 아이템들의 집합이고, R집합은 추천시스템에 의해 생성된 추천 목록의 집합이다. 한 아이템은 T집합이나 R집합 또는 두 집합에 모두 속할 수 있다.

정확도는 추천시스템이 생성한 추천목록의 아이템들 중에

목표 사용자가 실제로 경험하여 평가한 아이템에 속하는 비율로 식(4)와 같이 계산된다.

$$Precision = \frac{|T \cap R|}{|R|} \quad (4)$$

재현율과 정확도는 추천 목록의 개수에 따라 달라진다. 즉, 추천 목록의 개수가 증가하면 Recall은 분자가 증가하여 그 값이 커지고, Precision은 분모가 증가하여 그 값이 작아진다. 이러한 상충관계 때문에 두 성능지표를 동시에 고려하는 F1이 식(5)와 같이 계산된다.

$$F1 = \frac{2 \times Recall \times Precision}{Recall + Precision} \quad (5)$$

3.4. 실험 및 결과

본 연구의 대상은 미네소타 대학의 GroupLens Research Project에서 수집된 MovieLens 데이터를 사용하였다. MovieLens 데이터는 User ID와 MovieID, 그리고 각 고객들이 관람한 영화에 대해 1점에서 5점까지의 점수를 부여한 선호도 점수로 구성되어 있다. 1692편의 영화들에 대해 943명의 사용자로부터 생성된 100K MovieLens 데이터를 이용하였다.

각 영화는 미국 최대 영화 데이터베이스인 IMDb(Internet Movie Database, <http://us.imdb.com>)의 기준에 따라 18개의 장르에 속해 있다. 장르의 종류는 표 2와 같다. 한 영화는 최소 1개에서 최대 5개까지의 장르에 속해 있다.

표 2. 장르 종류
Table 2. Kind of Genre

Action	Documentary	Mystery
Adventure	Drama	Romance
Animation	Fantasy	Sci-Fi
Children	Film-Noir	Thriller
Comedy	Horror	War
Crime	Musical	Western

실험은 MovieLens의 데이터 중 80%의 훈련데이터(Training Data Set)와 20%의 검증데이터(Validation Data Set)로 구분하여 활용하였다. 모든 선호도 점수는 이미 주어져 있는 상태이고, 구현하는 추천시스템의 성능측정을 위해 일부 아이템의 선호도 점수를 모른다고 가정하고, 그 아이템의 선호도 점수를 예측한 후에 실제 평가한 선호도 점수와 비교한다.

빈발 패턴 네트워크로 장르 패턴 기반의 사용자 클러스터가 구축되었던 후 최근접 이웃의 수인 k를 결정하기 위해 k를

증가시키면서 MAE의 변화를 관찰하였다. 본 연구는 k를 5에서 100까지 5씩 증가 시키면서 최근접 이웃의 수의 변화에 따른 평균 MAE의 변화를 관찰한 결과 k의 값이 25일 때, 평균 MAE가 제일 작았다. 따라서 본 연구에서는 최근접 이웃의 수를 25명으로 설정하였다. 그리고 성능비교를 위해 전통적인 협업적 여과의 최근접 이웃의 수를 25로 고정하여 실험하였다.

실험 평가는 제안된 방법을 적용한 추천 성능 F1과 사용자 클러스터링 재구축에 소요되는 시간 비교를 통해 이루어졌다.

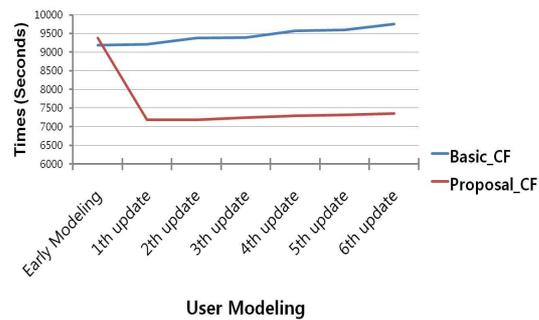


그림 8. 사용자 클러스터링 재구축 시간 비교
Fig. 8. Compares to rebuilding Time of User clustering

본 논문에서 제안하는 방법을 통해 그림 8과 같이 초기 사용자 클러스터링 구축을 보면, 빈발 패턴 네트워크를 구축하는데 소요되는 시간이 있어 전통적인 협업적 여과 기법과 비슷한 시간이 들었다. 사용자가 평가한 영화가 피드백 됨에 따라 사용자 클러스터링의 재구축에 소요되는 시간이 전통적인 협업적 여과 기법에 비해 평균 23%로 감소됨을 볼 수 있다. 이는 빈발 패턴 네트워크 특성인 규칙의 조건부에 속성이 하나 추가될 때마다 훈련 데이터를 재탐색할 필요가 없어 사용자가 평가한 영화가 피드백 되었을 경우, 빈발 패턴 네트워크를 이용하면 정보를 쉽게 누적시켜 패턴을 추출할 수 있기 때문이다. 추출된 패턴을 통해 사용자 클러스터링을 구축한다. 이를 협업적 여과에 적용한다. 협업적 여과의 사용자 클러스터링을 재구축 시 탐색공간을 줄여 재구축 시간을 줄일 수 있었다.

표 3은 전통적인 협업적 여과 시스템과 논문에서 제안된 시스템의 추천 성능을 나타낸다.

표 3. Basic_CF와 Proposal_CF 성능 비교
Table 3. Comparing Basic_CF and Proposal_CF

	Recall	Precision	F1
Basic_CF	0.319	0.835	0.462
Proposal_CF	0.322	0.843	0.466

F1의 성능지표를 통해 제안한 방법이 전통적인 방법과 유사한 성능을 나타낸다는 것을 알 수 있었다. 추천 성능의 개선이 나타나지 않은 것은 빈발 패턴 네트워크를 이용하여 사용자 장르 패턴을 추출할 때, 탐욕적인 방법을 이용하여 가장 선호되는 패턴만 적용하였기 때문이다. 이는 [4], [7]과 같이 장르 패턴에 가중치를 적용하여 장르 패턴을 선택한다면 추천 성능이 향상 될 수 있다.

IV. 결론

본 논문에서는 전통적인 협업적 여과의 사용자 클러스터링 재구축이 쉽지 않다는 문제점을 보완하기 위해 장르 패턴 기반의 사용자 클러스터링을 제안하였다. 장르 패턴 기반의 사용자 클러스터링은 빈발 패턴 네트워크를 통해 사용자가 평가한 영화의 장르를 분석하여 사용자가 선호하는 장르를 추출하고, 추출된 장르 패턴을 기반으로 사용자 클러스터링을 구축하여 협업적 여과 시스템에 활용하였다. 본 논문에서 제안하는 시스템은 초기 사용자 클러스터링 구축 시 사용자가 선호하는 장르 패턴을 추출하기 위한 빈발 패턴 네트워크를 구축하는데 시간이 소요되기 때문에 전통적인 협업적 여과 기법과 비슷한 시간이 소요되었다. 하지만, 사용자가 평가한 영화가 피드백 뒤에 따라 사용자 클러스터링의 재구축에 소요되는 시간이 전통적인 협업적 여과 기법에 비해 평균 23%로 감소됨을 확인할 수 있었다. 대용량의 데이터를 가진 영화 추천 시스템에 장르 패턴 기반의 사용자 클러스터링을 적용하여 재구축 시간의 23%를 줄일 수 있다면, 추천 시스템은 효율적으로 사용자에게 영화를 추천해 줄 수 있다. 나아가 음악, 책 등과 같은 다른 도메인을 추천할 때도 같은 방법을 적용할 수 있다. 즉, 아이템이 영화의 장르처럼 특정 카테고리에서 조합을 구성으로 가지는 경우에 그 조합이 특정한 연관 규칙을 가진다면 효율적으로 아이템을 추천해 줄 수 있을 것이다.

향후 연구로는 제안하는 시스템의 추천 성능 향상을 위해 빈발 패턴 네트워크를 이용하여 사용자가 선호하는 장르 패턴을 추출할 때, 발견된 장르 패턴에 가중치를 적용하여 장르 패턴을 선택하는 것이 필요하고, 결정된 패턴을 활용하여 추천 성능을 향상 시킬 수 있을 것이다.

Acknowledgments

본 연구는 정보통신산업진흥원의 IT/SW 창의연구과정의 연구결과로 지식경제부와 에스케이씨앤씨 주식회사에 의해

지원된 과제로 수행되었음”
(NIPA-2010-C1810-1006-0002)

참고문헌

- [1] A. Ansari, S. Essegiaer, and R. Kohli, "Internet Recommendation Systems," *Journal of Marketing Research*, Vol. 37, No. 3, pp. 363-375, Aug 2000
- [2] D. Cho, K. Chung, K. Rim and J. Lee, "Method of Associative Group Using FP-Tree in Personalized Recommendation System," *Journal of Korea Contents*, Vol. 7, No. 10, pp. 19-26, Oct 2007
- [3] K. Oh, J. Jung, I. Ha, G. Jo, "Discovering Association Rules using Item Clustering on Frequent Pattern Network," *Journal of The Korea Intelligent Information System Society*, Vol. 14, No. 1, pp.1-17, March 2008
- [4] J. Wang, N. Zhang, and J. Yin, "Collaborative Filtering Recommendation Based Fuzzy Clustering of User Preferences," *Fuzzy Systems and Knowledge Discovery*, Vol 6, pp. 1946-1950, 10-12 Aug. 2010
- [5] G. Lekakos and G. M. Giaglis, "Improving the prediction accuracy of recommendation algorithms: Approaches anchored on human factors," *Interacting with Computers*, Vol. 18, Issue 3, pp. 410-431, May 2006
- [6] S. Choi, Y. Han, "A Content Recommendation System Based on Category Correlations," *ICCGI '10 Proceedings of the 2010 5th International Multi-conference on Computing in the Global Information Technology*, pp. 66-70, 11 Nov 2010
- [7] S. Lee, S. Park, "Performance Improvement of a Movie Recommendation System using Genre-wise Collaborative Filtering", *Journal of The Korea Intelligent Information System Society*, Vol. 13, No. 4, pp. 65-78, Dec 2007
- [8] N. Pasquier, Y. Bastide, R. Taouil, and L. Lakhal, "Discovering Frequent Closed Itemsets for

Association Rules,” In Proceedings of the 7th International Conference on Database Theory, pp. 398-416, 1999

[9] R. Agrawal and R. Srikant, “Mining Sequential Patterns,” 11th International Conference on Data Engineering, pp. 3-14, 1995

[10] P. Resnick, N. Iacovou, M. Suchak, and P. Bergstrom, J. Riedl, “GroupLens: an open architecture for collaborative filtering of netnews,” ACM conference on Computer supported cooperative work, pp. 175-186, 1994

저 자 소개



최 자 현
 2009년 : 상지대학교 컴퓨터공학과(공학사)
 2010년~현재 : 인하대학교 컴퓨터정보공학과 석사과정
 관심분야 : 지능형 전자상거래, Data Mining
 Email : cjhun81@eslab.inha.ac.kr



하 인 애
 2005년 : 수원대학교 컴퓨터학과(공학사)
 2007년 : 인하대학교 컴퓨터정보공학과(공학석사)
 2007년~현재 : 인하대학교 컴퓨터정보공학과 박사과정
 관심분야 : 지능형 전자상거래, Web 2.0, Semantic Web, Recommendation System
 Email : inay@eslab.inha.ac.kr



홍 명 덕
 2008년 : 서울디지털대학교 컴퓨터정보공학과(공학사)
 2011년 : 인하대학교 컴퓨터정보공학과(공학석사)
 2011년~현재 : 인하대학교 컴퓨터정보공학과 박사과정
 관심분야 : 지능형 전자상거래, Semantic Web, CSP, 온톨로지
 Email : hmdgo@eslab.inha.ac.kr



조 근 식
 1982년 : 인하대학교 전자계산학과(공학사)
 1985년 : Queens Colleg/CUNY Computer Science(공학석사)
 1991년 : City University of New York, Computer Science(공학박사)
 1991년~현재 : 인하대학교 컴퓨터정보공학과 교수
 관심분야 : 인공지능, CSP, 온톨로지, Semantic Web, Web 2.0, 지능형 전자상거래
 Email : gsjo@inha.ac.kr