

사용자의 재생 시간을 이용한 멀티미디어 추천 시스템☆

A Multimedia Recommender System Using User Playback Time

권 형 준* 정 동 근** 홍 광 석***

Kwon Hyeong-Joon Chung Dong-Keun Hong Kwang-Seok

요 약

본 논문에서는 사용자의 재생 시간을 이용한 멀티미디어 추천 시스템을 제안한다. 제안하는 시스템은 사용자에 의해 요청된 멀티미디어 콘텐츠와 그것이 재생된 시간을 기록하고, 기록된 데이터를 가지고 퍼지 연관규칙 탐사 방법을 이용하여 사용자가 관심을 보일 만한 멀티미디어 콘텐츠와 사용자에 의해 재생된 시간에 기반하여 선호 등급을 예측한다. 제안하는 방법은 사용자의 선호 정보를 별도로 입력 받지 않고도 예측된 선호 등급에 따라서 추천 목록에 대한 선호 정도를 예측할 수 있으며, 거짓된 선호 정보의 유입을 방지하는 장점이 있다. 유효성 검증을 위해 제안하는 시스템을 구현하고 실험한 결과, 사용자로부터 입력 받은 선호 정보를 포함하지 않은 트랜잭션으로부터 사용자가 높은 선호도를 보일 것이라 예상되는 추천 목록을 선별하여 추천 시스템에 적용할 수 있음을 확인하였다.

Abstract

In this paper, we propose a multimedia recommender system using user's playback time. Proposed system collects multimedia content which is requested by user and its user's playback time, as web log data. The system predicts playback time based preference level and related contents from collected transaction database by fuzzy association rule mining. Proposed method has a merit which sorts recommendation list according to preference without user's custom preference data, and prevents a false preference. As an experimental result, we confirm that proposed system discovers useful rules and applies them to recommender system from a transaction which doesn't include custom preferences.

☞ keyword : 추천 시스템(Recommender Systems), 연관 규칙(Association Rules), 개인화(Personalization)

1. 서 론

UCC(User Created Contents)는 멀티미디어 정보의 생성 및 배포의 근원을 소비자 중심으로 바꾸었다. 대표적인 동영상 UCC 제공 서비스인 유튜브는 사용자에 의해 등록된 동영상 개수가

5500만개를 돌파했으며 국내 동영상 UCC기반 사업자들도 독자적인 개인화 서비스를 내세워 사용자들을 유치하기 위해 박차를 가하고 있다 [1][2]. 이와 같은 현상은 참여, 개방 및 공유로 대표되는 웹 2.0의 개념과 무관하지 않다.

정보 제공자와 수요자의 벽을 허무는 웹 2.0 시대가 도래하면서 온라인 상에 공유되는 멀티미디어 데이터의 개수가 기하급수적으로 증가하였다. 따라서 양질의 정보가 묻히는 현상이 발생하고, 사용자는 자신의 취향에 맞는 멀티미디어 정보를 검색하기 위해 많은 시간을 소비해야 하는 문제점을 해결하는 것이 새로운 과제로 대두되었다. 이에 따라, 웹 환경에서 개인의 성향에 맞춘 정보를 추천하는 추천 시스템

* 준 회 원 : 성균관대학교 대학원 전자전기컴퓨터공학과 박사과정

katsyuki@skku.edu

** 정 회 원 : 을지대학교 의료산업학부 의료전산화전공 교수
chung@eulji.ac.kr(교신저자)

*** 정 회 원 : 성균관대학교 정보통신공학부 교수
kshong@yurim.skku.ac.kr

[2008/12/05 투고 - 2008/12/08 심사 - 2008/01/13 심사완료]

☆ 이 논문은 성균관대학교의 2008학년도 삼성학술연구비에 의하여 연구되었음.

에 관한 연구가 지속적으로 이루어지고 있다. 최근에는 유사도 측정 기반의 협업 추천 방법들에 관한 연구가 활발하지만, 협업 추천 방법은 사용자로부터 입력 받은 각 멀티미디어에 대한 선호 정보를 반드시 필요로 한다. 이에 본 논문에서는, 사용자의 선호 정보를 이용하지 않고 사용자에 의해 재생된 시간을 기반으로 연관 규칙의 탐사를 수행하는 멀티미디어 추천 시스템을 제안한다. 제안하는 추천 시스템은 웹 기반 정보 시스템에서 사용자의 관심도를 측정하기 위해 사용하는 웹 로그 데이터 중 사용자가 웹 페이지에 머문 시간(Duration Time)을 이용하여 성향을 분석하거나 관심도를 측정하는 기존의 연구들에 근거하였다[3][4][5]. 사용자가 흥미를 보인 웹 페이지의 로그를 분석하면 오랜 체류 시간을 보이는 것처럼 사용자가 흥미 있어 했던 멀티미디어 정보는 오랜 재생 시간을 가질 것이다. 여기서 두 가지 의문점이 제기된다. 첫째, 웹 서버의 로그에 기록된 웹 페이지 체류 시간을 웹 페이지에 포함된 멀티미디어 정보의 관심도로 사용해도 되는지에 관한 것이다. 멀티미디어의 재생은 중지되었으나 웹 페이지의 체류 시간은 더 길어질 여지가 있고, 웹 페이지 내에서 별도의 웹 컨트롤에 의해 실행된 시간은 별도로 웹 서버의 로그에 기록되지 않는다. 이에 따라 본 논문에서는 웹 페이지의 관심도 측정이 아니라 멀티미디어에 대한 관심도 측정의 목적으로 재생 시간을 별도로 기록하는 경우를 가정하였다. 둘째, 최대 재생 시간이 각각 다른 멀티미디어 정보에서 사용자의 재생 시간이 길고 짧음을 어떻게 판단할 수 있는지에 관한 것이다. 본 논문에서는 문제의 해결을 위해 퍼지 연관규칙 탐사 방법을 사용하였다[6][7]. 사용자의 재생 시간을 퍼지 계층으로 확장하여 퍼지 숫자로 변환하고, 각 멀티미디어의 최대 재생 시간과 길고 짧게 재생한 정도에 대해 등급을 부여하여 연관규칙의 탐사를

수행함으로써 사용자가 관심을 보일 만한 정보의 선호 등급을 예측하였다. 제안하는 추천 시스템은 협업 추천 시스템과는 다르게 사용자의 선호 정보를 별도로 입력 받지 않으므로 선호 정보 입력 절차가 불필요하다. 이는 추천 시스템에 사용되는 데이터를 더욱 많이 확보할 수 있는 장점을 가지며, 거짓된 선호도의 입력이 자동으로 방지되는 장점이 있다. 또한, 개인별로 다른 선호도 기준을 미디어의 최대 재생 시간의 관점에서 정규화 함으로써 보다 공정한 추천 목록 생성이 가능하다. 더불어, 멀티미디어 정보의 추천에 있어서 기존의 웹 페이지에 체류한 시간을 기반으로 관심도를 측정하는 방법보다 더 정확한 선호 정보의 예측이 가능하다. 제안하는 추천 시스템은 멀티미디어 데이터를 제공하는 정보 시스템에서 동작하는 정보 시스템에 특히 유용하게 적용될 수 있다. 동영상 UCC, DTV, 음성 웹 등 재생 가능한 미디어를 제공하는 장치 및 시스템에 응용 가능하며, 정지영상 및 텍스트 위주의 정보를 제공하는 웹 기반 시스템에도 제안하는 추천 방법을 적용할 수 있음은 물론이다.

본 논문의 구성은 다음과 같다. 2장에서는 추천 시스템에 이용되는 기법들인 협업 필터링과 연관규칙의 탐사 방법을 소개하고, 멀티미디어의 재생 시간을 사용자 관심도로 사용하는 근거가 되는 웹 페이지 체류 시간에 관한 기존 연구들을 소개한다. 3장에서는 퍼지 개념 계층에서의 연관규칙 탐사에 기반을 두는 제안하는 추천 시스템의 설계 및 구현 방법을 자세히 설명한다. 제안하는 추천 시스템의 성능 평가를 위한 실험 결과를 4장에 보이며, 5장에서 결론을 맺는다.

2. 관련 연구

2.1 추천 시스템

사용자 개인이 관심을 보일만한 정보를 시스

템이 자동적으로 제공함으로써 정보 제공의 개인화와 깊은 연관성을 가지는 추천 시스템은 사용자가 관심을 보일만한 아이템의 목록을 만들어 사용자에게 추천하는 기능을 수행한다 [8][9][10][11][12]. 추천 시스템의 개발 방법에 대한 최근의 연구들을 살펴보면, 데이터 마이닝 기법 중 하나인 연관규칙의 탐사를 이용하는 방법, 아이템의 속성에 근거하는 내용 기반 추천 방법, 사용자의 선호도를 이용하는 협업 추천 방법 등이 있다[8]. 협업 추천은 과거에 사용자가 보인 성향이 미래에도 지속될 것이라는 가정에 근거하여 수행되는 방법이다[9]. 그 중에서도 사용자 기반 협업 추천 방법은 사용자가 선호할 만한 아이템을 예측하기 위해서 유사한 선호도를 가지는 다른 사용자들의 아이템 선호도에 근거하여 추천 아이템 목록을 만든다 [10]. 과거에 많은 사용자들에 의해 수집된 아이템에 대한 선호도를 토대로 사용자 간의 유사도를 계산하여 유사한 사용자들의 집합을 만들거나 유사도가 높은 순서대로 정렬한다. 사용자들의 유사도 계산이 종료된 후, 추천 대상 사용자와 가장 유사한 k명의 사용자들의 선호 정보에 근거하여 특정 아이템에 대한 사용자의 선호도를 예측한다. 아이템 기반 협업 추천 방법은 사용자에게 의해 과거에 입력된 선호도와 특정 사용자에게 대한 선호도를 예측하고자 하는 아이템들의 유사성을 계산하여 가장 유사한 아이템을 선별하는 방법이다[11]. 유사도 계산 방법에는 코사인 계수, 자카드 계수, 다이스 계수, 내적 계수와 유클리디안 거리, 카이 제곱 등 여러 가지 방법이 있으나 모두 일관된 유사도 순위 결과를 도출하며 최근에는 대용량의 데이터 세트에서 가장 뛰어난 성능을 보이는 피어슨 상관계수를 주로 이용한다고 알려져 있다[12].

정보 시스템의 지능화 및 개인화를 위한 수단과 함께 마케팅을 목적으로 널리 사용되고 있는 연관규칙의 탐사는 데이터 마이닝 기법

중 하나로써 아이템 간의 조건과 결과로 표현되는 패턴을 찾아내는 것이다[12][13][14][15]. 즉, ‘아이템 X가 발생하면 Y도 발생한다’ 와 같은 규칙을 찾아내기 위해 사용된다. 이는 추천 시스템에도 유용하게 이용될 수 있으며, X와 Y에는 각각 한 개의 아이템 뿐만 아니라 여러 개의 아이템을 포함하는 집합이 될 수 있다. 연관규칙의 탐사는 데이터 세트 내에서 특정 아이템이 발생한 비율인 지지도의 계산 결과에 따라서 빈발 집합을 찾아내고, 찾아낸 빈발 아이템들을 $R: X \Rightarrow Y$ 로 표현되는 연관규칙에서 X와 Y에 대입하여 X와 Y가 동시에 발생한 조건부 확률인 신뢰도에 따라서 각 아이템들 간의 연관규칙을 발견한다. 대표적인 알고리즘으로는 apriori, FP tree 등이 있는데, 계산하는 데에 소요되는 시간이 다를 뿐이며 지지도와 신뢰도를 활용하는 기본적인 연관규칙의 탐사 구조와 결과는 동일하다[14].

최근에 연관규칙의 탐사를 추천 시스템에 효과적으로 활용한 예를 살펴보면, 음성 웹에 응용한 사례가 있다[15]. 사용자가 입력한 단어의 집합으로부터 연관규칙을 탐사하고 그 결과에 근거하여 콘텐츠를 추천함으로써 눈으로 검색 결과를 확인할 수 없는 음성 웹에서의 탐색의 불편함을 해소하기 위한 방법으로 사용되었다. 발견된 연관규칙은 사용자에게 의해 요청된 X에 대해 Y를 추천하는 방법으로 추천 시스템에 적용될 수 있으나, 협업 추천을 이용한 방법처럼 사용자의 선호도를 예측할 수는 없다. 그러나 사용자의 선호도 정보가 없더라도 사용자가 선호할 만한 아이템의 예측이 가능하고 협업 추천 방법보다 상대적으로 계산량이 적다는 장점이 있다.

2.2 체류 시간과 사용자 관심도

웹 로그는 사용자의 웹 사이트 사용성을 평

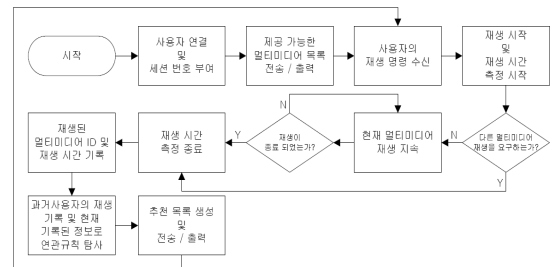
가하는 것의 기초 자료가 된다. 과거에는 페이지 뷰, 방문자 수, 순 방문자 수, 회원의 숫자 등을 로그에 기록하고 분석함으로써 웹 사이트의 가치를 평가하거나 사용자의 사이트 이용 내역을 분석함으로써 사용자의 관심사를 파악하였다. 그러나 Ajax 등과 같은 비동기 통신과 RIA, Macromedia Flash 등의 멀티미디어 기반 환경이 등장하면서 앞서 설명했던 전통적인 측정 지표로는 사용자의 관심 정도를 측정하기에 무리가 있어서 체류 시간이 정보에 대한 사용자의 관심 정도를 측정하기 위한 좋은 지표가 되고 있다. 이에 기존의 웹 페이지 체류시간과 웹 사이트 체류 시간을 바탕으로 사용자의 관심도 측정 및 성향 분석 시스템이 제안된 바 있다 [3][4][5]. 하지만 웹 페이지의 체류 시간과 웹 페이지 내에 포함된 동영상 멀티미디어의 재생 시간은 다르다. 본 논문에서는, 체류 시간이 사용자의 관심도를 반영한다는 기존의 연구들에 근거하여 동영상의 재생 시간을 사용자의 관심도로 삼는다.

협업 추천 방법은 선택의 후보가 많은 것에 대한 추천 시스템의 개발에 유용하게 이용되고 있으나 아이템에 대한 사용자들의 선호 정보가 입력되지 않은 경우 사용자 혹은 아이템 사이의 유사도를 계산할 수 없어서 추천 능력의 저하를 초래한다. 또한, 사용자가 선호 정보를 입력하더라도 거짓된 선호 정보가 입력되면 유사도 계산은 가능하지만 계산 결과의 신뢰성이 저하된다. 본 논문에서 제안하는 추천 방법은 사용자의 선호도를 입력 받는 절차를 필요로 하지 않으면서도 사용자가 시스템에 노출한 자연스러운 관심의 정도인 재생 시간의 등급별 예측이 가능하다. 즉, 제안하는 시스템은 사용자들에 의해 요청된 멀티미디어 데이터와 재생된 시간으로 이루어진 데이터베이스 트랜잭션을 이용한다는 점에서 협업 추천 방법을 참고하였고, 추천 목록을 생성하는 방법에서는 연관

규칙의 탐사를 사용하므로 협업 추천 방법과 데이터 마이닝이 결합된 하이브리드형 추천 시스템이라고 할 수 있다.

3. 제안하는 시스템

본 논문에서 제안하는 추천 시스템은 사용자에 의해 재생된 시간을 퍼지 계층의 데이터 세트로 변환하여 사용자의 선호도로 활용한다. 이렇게 만들어진 선호도를 기반으로 수량을 고려할 수 있는 퍼지 기반 연관규칙 탐사를 수행하여 추천 목록을 만든다. 사용자가 수동으로 선호도를 입력하지 않아도 재생 시간을 선호 정보로 활용하므로 앞서 설명한 협업 추천 방법의 문제점인 선호 정보가 없는 아이템에 관한 문제점을 해결할 수 있다. 또한, 멀티미디어의 최대 재생 시간과 사용자에 의해 재생된 시간에 기반하여 사용자의 선호도를 결정함으로써 선호 정보 수집 여부가 사용자에게 노출되지 않으므로 거짓된 선호도가 입력되는 것에 방지할 수 있다.



(그림 1) 제안하는 시스템의 실행 순서도

그림 1은 제안하는 시스템이 실행되는 순서를 보인다. 제안하는 추천 시스템은 사용자에 의해 요청된 미디어의 재생이 종료될 때 마다 사용자가 요청한 멀티미디어와 사용자의 재생 시간을 바탕으로 기존에 시스템을 이용했던 사용자들의 데이터로부터 사용자가 선호할 것이

라 예상되는 미디어를 예측한다. 멀티미디어의 최대 재생 시간과 비교하여 사용자에게 의해 재생된 시간이 높은 비율을 차지한다면 많은 관심을 보였다고 판단하며, 예측할 수 없는 경우는 관심 정보를 추천하는 단계를 생략한다.

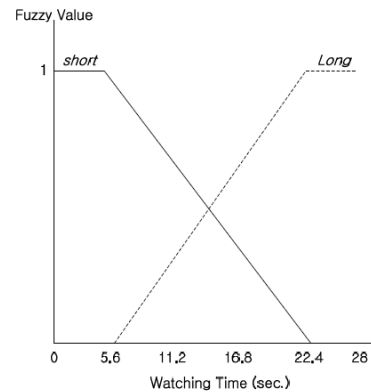
제안하는 추천 시스템은 멀티미디어를 제공하는 시스템에서 효과적이며, 동영상 UCC와 같은 다량의 멀티미디어를 주로 제공하는 시스템에 더욱 유용하다. 본 논문에서는 표 1의 데이터 세트를 가지고 제안하는 시스템의 추천 목록 생성 방법을 단계적으로 설명한다. Session ID는 사용자의 연결 번호, 5(14)에서 5는 멀티미디어 데이터의 고유번호이며 14는 재생된 시간(초)을 의미한다.

(표 1) 데이터 세트의 예

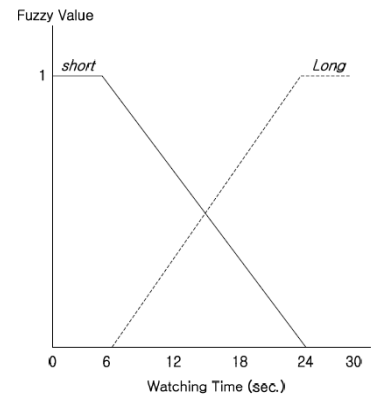
Session ID	Requested Multimedia ID (Playback time)
1	5(14), 1(3), 3(9), 2(7)
2	5(9), 3(11), 2(9)
3	1(9), 3(7), 4(13)
4	4(12), 1(16), 2(17)
5	1(12), 3(14), 2(16), 5(12)
6	1(9), 2(7), 5(15), 4(4)

첫 번째 단계는 사용자들의 멀티미디어 재생 정보가 기록된 데이터 세트에 퍼지 소속 함수를 적용하여 퍼지 계층의 데이터 세트로 변환하는 것이다. 퍼지 소속 함수는 각 멀티미디어의 최대 재생 시간에 따라서 자동으로 생성되며, 시스템 관리자가 설정한 개수에 따라 임의의 구간들로 나누어진다. 1번 미디어의 최대 재생 시간이 28초, 2번 미디어의 최대 재생 시간이 30초라고 가정하고 소속 함수를 생성한 예는 그림 2와 같다. 멀티미디어의 총 재생 시간을 다섯 구간으로 나눈 후, 첫 번째 구간을 짧게 재생한 정도로 판단하고 마지막 구간을 길

게 재생한 정도로 판단하였다. 가운데 위치한 구간들은 길게 재생한 정도와 짧게 재생한 정도를 판단하기 어려운 애매모호한 구간으로 규정하였다.



(1)



(2)

(그림 2) 소속 함수의 예

표 1의 Session ID 2번의 재생 기록을 살펴보면, 사용자는 총 재생 시간이 30초인 2번 멀티미디어를 9초간 재생하였다. 이를 퍼지 데이터 세트로 변환하기 위해서는 그림 2의 소속 함수 (2)를 적용하여 짧게 재생한 정도와 길게 재생한 정도를 계산한다. 짧게 재생한 정도는 0.7976, 길게 재생한 정도는 0.2024임을 알 수 있다. 표 2는 데이터 세트의 예제인 표 1을 퍼

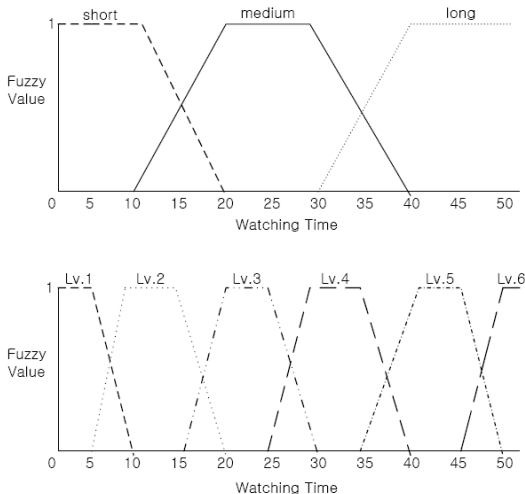
지 데이터 세트로 변환한 결과를 보인다.
(표 2) 변환된 퍼지 계층의 트랜잭션

도를 계산하고, 시스템 관리자에 의해 결정되는

Session ID	1 (short)	1 (long)	2 (short)	2 (long)	3 (short)	3 (long)	4 (short)	4 (long)	5 (short)	5 (long)
1	1	0	0.9167	0.0883	0.619	0.381			0.037	0.963
2			0.7976	0.2024	0.4603	0.5397			0.5	0.5
3	0.8333	0.1667			0.7778	0.2222	0.4306	0.5694		
4	0.4444	0.5556	0.3214	0.6786			0.5	0.5		
5	0.6667	0.3333	0.381	0.619	0.2222	0.7778			0.2222	0.7778
6	0.8333	0.1667	0.9167	0.0833			1	0	0	1
Sum.	3.7777	1.2223	3.3334	1.6716	2.0793	1.9207	1.9306	1.0694	0.7592	3.2408
Supp.	0.755	0.244	0.666	0.484	0.519	0.480	0.643	0.356	0.189	0.812
Maximum Playtime	30 seconds		28 seconds		21seconds		24 seconds		18 seconds	

소속 함수는 그림 3과 같이 더욱 세분화 될 수 있다. 소속 함수의 정밀한 세분화는 더욱 신뢰성 있는 추천 목록의 생성을 유도할 수 있지만 소속 함수의 세분화를 정밀화 할수록 추천 목록의 생성을 위한 계산량이 증가한다.

최소 지지도 임계치 이상을 만족하는 아이템을 선별한다. μ 는 전체 퍼지 집합, t 는 사용자 재생 기록 데이터, D 는 해당 아이템을 포함하는 사용자 재생 기록 데이터 개수, i 는 사용자 재생 기록 데이터의 일련번호를 나타낸다.



(그림 3) 소속 함수의 구간 세분화

두 번째 단계는 변환된 퍼지 계층의 데이터 세트를 식 (1)에 따라 모든 아이템에 대한 지지

$$\text{supp}(item) = \frac{\sum \mu_{item}(t_i)}{D} \quad (1)$$

지지는 시스템 관리자에 의해 설정된 최소 지지도 임계치에 의해 해당 아이템이 데이터 세트에서 얼마나 빈발하게 발생하는지 판단하는 척도로써, 2Long에 대한 지지도를 계산하면 $(0.0883 + 0.2024 + 0.6786 + 0.619 + 0.833) / 5 = \text{약 } 0.484$ 가 된다. 표2의 트랜잭션에 대해서 최소 지지도 임계치를 0.3으로 설정하고 높은 선호도를 가진 아이템을 선별하면 1Short(0.755), 2Short(0.666), 2Long(0.484), 3Short(0.519), 3Long(0.480), 4Short(0.643), 5Long(0.812)가 된다. 선별된 아이템들은 시스템을 이용한 사용자들이 시스템에 빈발하게 노출했던 성향이라고 볼 수 있다. 또한, 두 개 이상의 아이템을 동시에 선호한 정도를 고려할 수 있는데, 두 가지 이상의 아

아이템에 대한 지지도 계산 방법은 식 (2)와 같이 두 개의 아이템 중 낮은 값을 취해서 합산한다.

$$\text{supp}(item1, item2) = \frac{\sum \min(\mu_{item1}(t_i), \mu_{item2}(t_i))}{D} \quad (2)$$

마지막 단계로써, 지지도 계산에 의해 선별된 빈발 집합들의 모든 아이템들을 연관규칙을 나타내는 R: X→Y에서 X와 Y에 대입하여 식 (3)에 따라 신뢰도를 계산하고, 시스템 관리자에 의해 설정된 최소 지지도 임계치 이상의 규칙을 탐사한다.

$$\text{conf}(item1 \rightarrow item2) = \frac{\text{supp}(item1, item2)}{\text{supp}(item1)} \quad (3)$$

아이템 2Long과 3Long의 지지도는 $(0.0883 + 0.2024 + 0.619) / 3 = 0.3032$ 이며, 이를 2Long의 지지도 $(0.0883 + 0.2024 + 0.6786 + 0.619 + 0.0883) / 5 = 0.3353$ 로 나누어 R: 2Long→3Long의 신뢰도를 계산하면 0.9042가 된다. 이 때, 두 아이템 모두를 포함하는 트랜잭션만을 고려해야 한다. 2Long과 3Long을 모두 포함하는 데이터의 개수는 3개이므로 2Long과 3Long의 신뢰도를 계산할 때 사용되는 총 데이터의 개수는 3개이다. 결론적으로 말하면, 2번 멀티미디어를 선호한 사용자들이 3번 멀티미디어를 선호한 정도는 0.9042라고 말할 수 있다.

최소 지지도 임계치를 0.9로 설정했다면, R: X→Y로 표현되는 연관규칙에서 R: 2Long→3Long은 유용한 규칙이라고 할 수 있을 것이며, 제안하는 추천 시스템은 R: 2Long→3Long에 따라서 2번 멀티미디어를 선호한 사용자에게 3번 멀티미디어를 추천해 줄 것이다. 그러나 R: 2Long→3Short을 고려하면, 규칙의 신뢰도는 $0.1709 / 0.3353 = 0.5096$ 으로써 최소 지지도 임계치를 만족하지 않으므로 2번 멀티미디어에 높은 선호도를 보인 사용자들은 3번 멀티미디어

에 낮은 선호도를 보였다는 연관 규칙을 신뢰할 수 없어서 2번 멀티미디어에 많은 관심을 보인 사용자에게 3번 멀티미디어를 추천하지 않을 것이다. 하지만 결과부가 Short인 아이템은 사용자가 많은 관심을 보이지 않은 아이템이 결과부에 포함된 규칙이므로 제안하는 시스템에서 추천 목록으로써 고려할 필요가 없다. 즉, 결과부에는 사용자가 높은 선호도를 보인 Long 아이템들만을 대입하고, 조건부에는 모든 아이템을 대입하여 추천 목록을 생성한다.

제안하는 방법을 정리하면, 1) 수집된 멀티미디어의 고유번호와 재생된 시간을 수집한 트랜잭션을 퍼지 계층으로 변환한다. 이 때, 아이템은 미리 설정한 소속 함수에 따라서 하나의 아이템이 여러 개의 퍼지 아이템으로 나뉘어진다. 2) 퍼지 트랜잭션에서 사용자에게 의해 빈번하게 요청되었던 아이템을 선별하기 위해 지지도를 계산하고, 임의의 최소 지지도 임계치 보다 높은 지지도를 가지는 아이템들의 짝을 지어 신뢰도를 계산하여 임의의 최소 신뢰도 이상을 가지는 연관 규칙을 탐사한다. 3) 발견된 연관 규칙에서 결과부에 높은 재생 등급을 가진 규칙들을 선별한다. 제안하는 시스템에서 연관 규칙 2Long→3Long이 최종적으로 선별된 규칙 중 하나라면, 2번 콘텐츠를 오랜 시간 재생한 사용자에게 3번 콘텐츠를 추천할 것이며, 2Short→3Long이 최종적으로 선별된 규칙 중 하나라면, 2번 콘텐츠에 많은 관심을 보이지 않은 사용자에게 3번 콘텐츠를 추천해 줄 것이다.

최소 지지도 임계치를 낮게 설정할 수록 발견되는 연관 규칙의 개수가 많아지지만 발견된 규칙의 신뢰도는 점점 낮아진다. 즉, 추천 성능이 낮을 것이다. 반대로, 최소 지지도 임계치를 높게 설정할수록 적은 개수의 연관 규칙을 발견하지만 발견된 규칙의 신뢰도는 높아지며 더욱 좋은 추천 성능을 보인다.

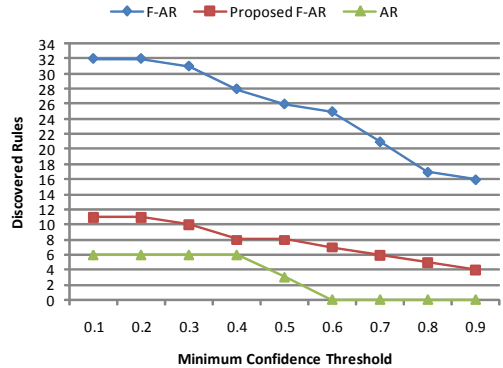
4. 실험 및 결과

대부분의 협업 추천 시스템의 성능 측정에 관한 연구는 무비렌즈에서 제공하는 사용자 선호도 데이터 세트를 이용하여 실시된다[15]. 무비렌즈 데이터 세트는 평점 1점과 5점의 사이에서 사용자가 수동으로 입력한 선호도를 기반으로 구축된 데이터 세트이므로 제안하는 추천 시스템의 성능을 판단하기에는 적합하지 못하다. 제안하는 추천 시스템은 연관규칙 탐사 기반의 추천 시스템으로써, 협업 추천 방법과의 비교 실험보다 데이터 마이닝 계열의 연관규칙 탐사 방법과 비교하는 것이 성능 평가에 적합하다. 제안하는 추천 시스템의 성능을 평가할 만한 공신력 있는 데이터 세트가 없어 본 논문의 표 1의 데이터 세트를 이용하여 상대적인 성능을 관측하였다.

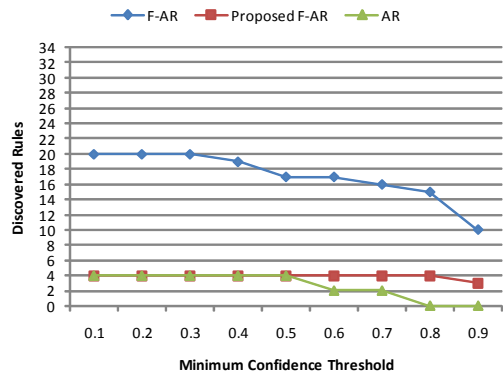
본 실험은 연관규칙 탐사 방법으로 발견할 수 있는 규칙의 개수, 제안하는 방법을 통해 발견할 수 있는 규칙의 개수, 제안하는 방법을 통해 발견할 수 있는 유용한 규칙(useful rules)의 개수를 도출하여 비교하는 것으로 실시되었다. 그림 4, 5 및 6에서, x축은 최소 신뢰도 임계치, y축은 발견된 규칙의 개수를 나타낸다. 각각의 실험 결과들은 최소 지지도 임계치를 0.3, 0.5 및 0.7로 설정하고 기존의 연관규칙 탐사 방법을 이용한 추천 방법과 퍼지 연관규칙 탐사 방법을 이용한 제안하는 추천 방법의 성능을 보인 결과이다. 퍼지 연관규칙 탐사 방법을 F-AR, 연관규칙 탐사 방법에 의한 연관규칙 탐사 방법을 AR, 제안하는 방법은 Proposed F-AR로 표기하였다.

그림 4, 5 및 6에서 보는 바와 같이, 최소 지지도 임계치를 낮게 설정할 수록 발견되는 규칙의 개수는 증가하지만 발견된 규칙에 대한 신뢰도는 하락한다. 보다 정확한 추천을 위해서는 최소 지지도 및 신뢰도의 임계치를 높게 설

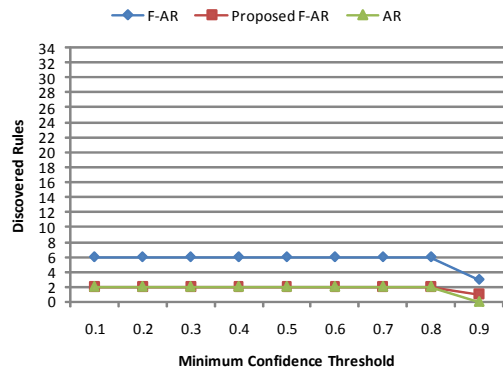
정하는 것이 바람직하나, 추천 목록의 양이 충분하지 못할 수 있기 때문에 보유한 데이터 세트에 따라서 적절한 최소 임계치의 설정이 요구된다.



(그림 4) 제안하는 시스템의 실험 결과 최소 신뢰도 임계치 0.3



(그림 5) 제안하는 시스템의 실험 결과 최소 신뢰도 임계치 0.5



(그림 6) 제안하는 시스템의 실험 결과 최소 지지도 임계치 0.7

그림 5를 살펴보면, F AR방법은 최소 신뢰도 임계치 0.7에서 16개의 규칙을 발견하였다. 발견된 규칙들에 대한 신뢰도(IF A→B)와 추천 여부(THEN) 그리고 규칙에 대한 지지도(CONF.)와 신뢰도(SUPP.)의 계산 결과를 표 3에 나타내었다. 각 멀티미디어에 대한 사용자의 재생 시간과 F AR에 의해 발견된 16개의 규칙 중에서 제안하는 추천 시스템에 적용될 규칙은 IF A→B에서 B에 Long을 포함하는 4, 8, 10, 14번 4개의 규칙이다. 일반적으로 IF A and B then C로 나타내는 것이 일반적이나, 퍼지 연관규칙에서는 A와 B의 위치가 서로 변경되면 결과값이 달라지기 때문에 IF A → B then C로 표현하는 것이 적합하다.

(표 3) 그림 5의 최소 신뢰도 임계치 0.7에서 발견된 규칙에 대한 세부 내역

No.	IF A→B	THEN	CONF.	SUPP.
1	5Long→4Short	no	1.000	1.000
2	5Long→2Short	no	0.828	0.725
3	5Long→1Short	no	0.889	0.888
4	4Short→5Long	yes	1.000	1.000
5	4Short→3Short	no	1.000	0.375
6	4Short→2Short	no	0.866	0.650
7	4Short→1Short	No	0.851	0.532
8	3Short→5Long	yes	1.000	0.380
9	3Short→2Short	no	1.000	0.380
10	3Short→1Short	No	1.000	0.571
11	2Short→5Long	yes	0.852	0.725
12	2Short→4Short	no	1.000	0.650
13	2Short→1Short	no	0.971	0.655
14	1Short→5Long	yes	1.000	0.888
15	1Short→4Short	no	0.718	0.532
16	1Short→2Short	no	0.907	0.655

최소 지지도 및 신뢰도의 임계치 변화를 통한 실험을 통해서 퍼지 계층의 연관규칙 탐사를 이용한 추천 시스템은 전통적인 연관규칙

탐사 방법보다 더욱 많은 규칙을 발견할 수 있음을 확인할 수 있었다. 특히, 제안하는 방법에 의해 사용자 선호 정보가 없는 데이터 세트를 이용하는 추천 시스템에 유용하게 사용될 수 있는 규칙을 선별한 결과도 전통적인 연관규칙 탐사 방법보다 더욱 많음을 확인하였다. 즉, 연관규칙의 탐사에 의해 발견된 규칙들을 이용하여 추천 시스템에 적용하는 것은 추천 목록에 대한 사용자의 선호 정도를 예측할 수 없으나, 재생 가능한 멀티미디어 기반의 정보 시스템에서는 제안하는 방법을 통해서 사용자의 선호 정보를 별도로 입력 받지 않더라도 로그에 기록된 재생 시간을 이용하여 추천에 유용한 규칙들을 발견할 수 있음을 확인하였다.

5. 결 론

본 논문에서는 퍼지 연관규칙 탐사 기반의 멀티미디어 추천 시스템을 제안하였다. 제안하는 추천 시스템은 사용자에게 의해 멀티미디어가 재생된 시간을 바탕으로 사용자의 선호 정도를 측정하였다. 제안하는 방법은 사용자가 수동으로 선호도를 입력하는 절차를 생략하면서도, 재생 시간 기반의 선호도 측정 여부를 노출하지 않기 때문에 거짓된 선호 정보의 유입을 미연에 방지하여 더 신뢰성 있는 추천 목록을 생성할 수 있다. 또한, 기존의 추천 시스템들이 사용자 개인마다 다른 기준에 의해 입력된 선호도를 이용하는 것과 다르게 제안하는 추천 시스템은 사용자의 자연스러운 행동인 재생 시간을 선호도로 이용함으로써 추천 목록을 생성함에 있어서 더 공정한 계산이 가능하다.

참 고 문 헌

- [1] 오세근, “웹 2.0 경제와 동영상 기반의 UCC 패러다임”, 정보통신연구진흥원 주간기술동

- 향, 통권 1285호, pp.24 - 36, 2007.
- [2] 김재영, “통신 방송 융합과 UCC”, 정보통신 연구진흥원 주간기술동향, 통권 1294호, pp.22 - 30, 2007.
- [3] 김창근, 정연홍, 김일, “웹페이지 관심도 분석에 관한 연구”, 한국해양정보통신학회논문지, Vol.11 No.4, pp.687 - 695, 2007.
- [4] 김병진, 최현우, 김용성, “사용자 관심도를 반영한 동적 웹 문서 추천 시스템”, 한국정보과학회 가을학술발표논문집, pp.136 - 138, 2001.
- [5] 강제호, 정연홍, 김일, 최봉준, 박규석, “웹페이지 체류시간을 이용한 성향분석 시스템의 설계 및 구현”, 한국인터넷정보학회 춘계학술발표집, pp.130 - 135, 2002.
- [6] Miguel Delgado, Nicolas Marin, Daniel Sanchez and Maria Amparo Vila, “Fuzzy Association Rules: General Model and Applications”, IEEE Transactions on Fuzzy Systems, VOL.11 NO.2, pp.214 - 225, 2003.
- [7] F. Berzal, I. Blanco, D. Sanchez, and M. A. Vila, “A New Framework to Assess Association Rules” Lecture Notes in Computer Science, Vol.2189, pp.95 - 104, 2001.
- [8] Herlocker, J. L., Konstan, J. A., Terveen, L. G. and Riedl, J. T., “Evaluating Collaborative Filtering Recommender Systems”, ACM Transactions on Information Systems, Vol.22 No.1, pp.5 - 53, 2004.
- [9] Goldberg David, David Nichols, Brain M. Oki, Douglas Terry, “Using Collaborative Filtering to Weave an Information Tapestry”. Communications of the ACM, Vol.35 No.12, pp.61 - 70, 1992.
- [10] Herlocker, J., Konstan, J., Borchers, A. and Riedl, J., “An Algorithmic Framework for Performing Collaborative Filtering”, Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval, pp.230 - 237, 1999.
- [11] Khanhquan Truong, Fuyuki Ishikawa, Shinichi Honiden, “Improving Accuracy of Recommender System by Item Clustering”, IEICE Transactions on Information and Systems, Vol.E90 D, pp.1363 - 1373, 2007.
- [12] Gediminas Adomavicius and Alexander Tuzhilin, “Toward the Next Generation of Recommender Systems: A Survey of the State of the Art and Possible Extensions”, IEEE Transactions on Knowledge and Data Engineering, Vol.17 No.6, pp.734 - 749, 2005.
- [13] R. Agrawal, T. Imielinski and A. Swami, “Mining association rules between sets of items in large databases”, Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, pp.207 - 216, 1993.
- [14] J.Han, J.Pei, and Y.Yin, “Mining Frequent Patterns without Candidate Generation”, Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data, pp.1 - 12, 2000.
- [15] 권형준, 홍광석, “사용자의 성향을 고려하는 지능형 음성 웹”, 정보처리학회논문지:B Vol. 15 B No.4, pp.347 - 354, 2008.
- [16] GroupLens Research, <http://www.grouplens.org>, Department of Computer Science and Engineering at the University of Minnesota.

◎ 저 자 소 개 ◎



권 형 준

2005년 서울보건대학 전산정보처리과 졸업(학사)
2005년 (주)블루엠 소프트웨어개발팀 사원
2008년 성균관대학교 대학원 전자전기컴퓨터공학과 졸업(석사)
2008년~현재 성균관대학교 대학원 전자전기컴퓨터공학과 박사과정
관심분야 : 데이터 분석, 기계 학습, HCI
E-mail : katsyuki@skku.edu



정 동 군

1986년 성균관대학교 전자공학과 졸업(학사)
1988년 성균관대학교 대학원 전자공학과 졸업(석사)
1996년 성균관대학교 대학원 전자공학과 졸업(박사)
1990년~현재 을지대학교 의료산업학부 의료전산학전공 교수
관심분야 : 유비쿼터스 응용, 임베디드 시스템, 신호 및 제어
E-mail : tchung@eulji.ac.kr



홍 광 석

1985년 성균관대학교 전자공학과 졸업(학사)
1988년 성균관대학교 대학원 전자공학과 졸업(석사)
1992년 성균관대학교 대학원 전자공학과 졸업(박사)
1990~1993년 서울보건대학 전산정보처리과 전임강사
1993~1995년 제주대학교 정보공학과 전임강사
1995년~현재 성균관대학교 정보통신공학부 교수
관심분야 : 오감인식, 융합 및 재현, HCI
E-mail : kshong@yurim.skku.ac.kr