

Improved Movie Recommendation System based-on Personal Propensity and Collaborative Filtering

Doo-Soon Park[†]

ABSTRACT

Several approaches to recommendation systems have been studied. One of the most successful technologies for building personalization and recommendation systems is collaborative filtering, which is a technique that provides a process of filtering customer information based on such information profiles. Collaborative filtering systems, however, have a sparsity if there is not enough data to recommend. In this paper, we suggest a movie recommendation system, based on the weighted personal propensity and the collaborating filtering system, in order to provide a solution to such sparsity. Furthermore, we assess the system's applicability by using the open database MovieLens, and present a weighted personal propensity framework for improvement in the performance of recommender systems. We successfully come up with a movie recommendation system through the optimal personalization factors.

Keywords : Recommender Systems, Collaborative Filtering, Sparsity, Personal Propensity

개인성향과 협업 필터링을 이용한 개선된 영화 추천 시스템

박 두 순[†]

요 약

추천 시스템들에 대한 여러 방법들이 연구되고 있다. 개인화와 추천 시스템 중에서 가장 성공적인 방법은 협업 필터링이다. 협업 필터링은 고객들의 프로파일 정보를 기반으로 추천을 하므로 데이터가 충분하지 않다면 항목을 추천하는데 있어서 희박성의 문제가 제기된다. 본 연구에서는 희박성의 문제를 해결하는 방법으로 가중치를 가진 개인 성향을 협업 필터링에 활용하는 방법을 제안한다. 본 연구에서 가중치를 가진 최적의 개인 성향을 찾기 위해 공개 데이터인 MovieLens Data를 이용하여 성능 평가하였다. 실험 결과 본 연구에서 제안한 가중치를 가진 개인 성향들로 구축된 시스템이 기존의 개인 성향들을 이용한 시스템보다 향상된 성능을 보였다.

키워드 : 추천시스템, 협업 필터링, 희박성, 개인 성향

1. 서 론

최근 스마트폰의 확산과 언제 어디서든 인터넷에 접속이 가능한 유비쿼터스 환경에서는 사용자들이 원하는 정보를 손쉽게 찾아보기를 원하지만 정말 사용자가 원하는 정보를 찾기는 매우 어렵다. 이를 해결하기 위한 방법으로 추천 시스템이 등장하였다. 추천 시스템은 고객에게 그가 좋아할 만한 서비스나 항목을 추천해주는 서비스로서 Amazon이나 CD Now 등 인터넷 쇼핑몰에서 많이 사용되고 있다[1]. 추천 시스템 중에 가장 성공적인 방법이 협업 필터링

(Collaborative Filtering)으로 알려져 있고, 웹페이지, 영화, 논문, 신문기사 추천 등의 다양한 적용사례를 가지고 있다.

협업 필터링 기법은 다른 사람들의 의견을 통하여 항목들을 추천하거나 예측하는 것이다[2]. 협업 필터링 기법은 영화추천 시스템에서 가장 대중적으로 사용하는 방법으로서 사용자들에게 수집한 다양한 정보들을 바탕으로 프로파일을 작성하고 이러한 프로파일을 바탕으로 하여 추천하고자 하는 대상 즉 목표고객과 유사한 구매 선호도를 가진 다른 사용자가 구매한 항목이나 선호할 것이라고 예측되는 항목들을 자동으로 추천하는 방법이다[3].

협업 필터링 기법은 목표 고객의 프로파일 정보를 기반으로 추천을 하므로 고객 및 구매 데이터가 충분히 축적되어야 한다. 데이터가 충분하지 않다면 항목을 추천을 하는데 있어서 정확도가 떨어지게 되어 희박성의 문제가 제기된다.

이러한 희박성 문제를 해결하는 방법들은 여러 방법들이

※ 이 논문은 2012년도 순천향대학교 교수 연구년제에 의하여 연구하였음.

† 중신회원 : 순천향대학교 컴퓨터소프트웨어공학과 교수

논문접수 : 2013년 9월 24일

수정일 : 1차 2013년 10월 28일

심사완료 : 2013년 10월 28일

* Corresponding Author : Doo-Soon Park(parkds@sch.ac.kr)

존재하지만 본 논문에서는 인구통계학적 방법으로 개인 성향을 파악하여 협업 필터링에 적용하는 방법을 사용한다. 희박성 문제들을 해결하기 위한 개인성향과 협업 필터링에 관련된 많은 연구들이 제시되었다[3].

본 논문에서는 실제적인 공개 데이터인 MovieLens를 가지고 희박성 문제를 해결하기 위해 개인 성향에 가중치를 부여하여 개선된 영화 추천 시스템을 제시한다. 본 연구의 구성은 다음과 같다. 제 2장에서는 관련 연구에 대해서 기술한다. 제 3장에서는 데이터 준비와 영화추천시스템

구축에 대해서 설명하고, 제 4장에서는 협업 필터링에서 개인성향을 고려한 추천 시스템을 설명하며 제 5장에서는 결론을 설명한다.

2. 관련 연구

본 장에서는 개인성향과 협업 필터링을 이용한 영화 추천에 필요한 추천 시스템, 개인화(personalization) 기법, 희박성 등을 기술한다.

2.1 추천시스템

추천시스템은 고객이 관심을 가지는 상품에 관한 정보나 인구 통계학적 정보, 과거 구매행동 분석을 토대로 고객의 요구에 맞는 항목을 추천해주는 시스템이다[4]. 전자상거래에서는 개인화 서비스가 강조되고 있으며, 개인화 서비스가 중요한 이유는, 고객의 입장에서 상품 검색의 노력을 줄일 수 있고 기업 입장에서는 적합한 상품 추천으로 인하여 전자상거래 사이트에 대한 고객 충성도를 높여 줄뿐만 아니라 고객과의 유대감을 형성 시킬 수 있기 때문이다[5].

2.2 개인화 기법

추천시스템을 위한 개인화 기법으로는 다음과 같은 방법들이 있다.

1) 내용기반 추천 방식(content-based filtering)

사용자의 선호 정보를 제공받아 그와 유사한 항목을 추천해주는 기법[6]으로 항목의 내용 정보가 풍부하고 분석하기 용이한 텍스트 문서, 뉴스, 웹 페이지 등의 항목 추천에 적합하다[7]. 그래서 주로 검색 사이트나 쇼핑몰에서 주요 알고리즘으로 사용되고 있다.

2) 규칙 기반 필터링

사용자에게 관심 분야, 선호도 등을 질의하여 그 사용자의 정보 프로파일을 획득해 내는 기법이다. 취미와 특정 사항에 대한 선호도 등을 질문하게 되고 답을 통해 그 사람의 프로파일을 획득한다. 이러한 정보를 바탕으로 고객의 심리 정보와 고객의 선호도 정보에 알맞은 정보 및 상품을 추천 또는 제공하는 방법이다[5].

3) 인구 통계학적 추천(Demographic)

나이, 성별, 지역, 교육 수준, 직업 등의 사용자 정보를 활용하여 추천을 수행하는 시스템이다[8]. 인구 통계학적 속성은 다양한 종류의 항목이나 항목 범주에 대한 사용자의 성향을 분석하기 쉽다는 장점을 갖는다.

4) 협업 필터링 시스템(collaborative Filtering; CF)

추천할 대상 항목에 대한 사용자의 평가를 입력받아 분석한 후, 비슷한 사용자끼리 묶어서 좋아할 만한 항목을 추천해주는 시스템이다[6]. 협업 필터링 추천 시스템은 평가 정보를 활용하므로 사용자나 항목에 대한 정보가 없어도 추천을 수행할 수 있는 장점을 갖는다.

5) 학습 에이전트 기반 추천 시스템

웹 사이트 방문기록 및 횟수, 접속 장소, 시간 등 일종의 로그 파일 분석을 통하여 사용자의 속성, 습관, 개인의 선호를 추적하는 학습 에이전트를 이용하는 개인화 기법이다[9].

추천시스템은 추천 시스템이 얼마나 고객이 만족할 만한 항목을 추천할 수 있는지에 초점이 맞춰져 있다. 일반적으로 항목과 항목 간의 유사점을 분류하는 방식에 협업 필터링이라는 방법을 사용한다.

협업 필터링 시스템에서 일반적으로 사용되어 지는 k-최근접 이웃 모델(k-nearest neighborhood model)은 활성 사용자(active user)가 다른 사용자들과의 거리를 계산하고 거리가 가장 가까운 k명의 사용자들을 이웃으로 선별한다. 사용자들 간의 거리를 구하는데 피어슨 상관 계수(Pearson correlation coefficient), 제곱 평균 차이(mean-square-difference), 벡터 유사도(vector similarity) 등이 사용되어진다. 사용자들 간의 거리가 구해지면 항목에 대한 예측 점수는 Equation (1)과 같이 다른 사용자들의 평가 점수를 거리 비율에 따라 합해줌으로써 얻어진다.

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

$p_{a,i}$ 는 활성 사용자 a의 항목 i에 대한 예측을 나타낸다. n은 이웃 사용자의 수이며 $r_{u,i}$ 는 사용자 u의 항목 i에 대한 평가 점수이고 $w_{a,u}$ 는 활성 사용자 a와 이웃 간의 유사도 가중치이다[10].

Breese[11]는 협업 필터링 알고리즘은 모든 평가들, 항목들, 사용자들이 메모리에 저장되어 있는 것을 요구하는 memory-based 알고리즘과 오프라인 평가 패턴들의 요약들을 주기적으로 생성해야 하는 model-based 알고리즘으로 구분하였다. memory-based 방법은 구현하는데 쉽다는 장점 때문에 Amazon과 같은 상업적인 시스템에 효율적으로 사용

되었다[12]. 하지만 이 방법은 공통 항목(common item)을 기본으로 하고 있기 때문에 데이터가 희박하다거나 공통된 항목이 적다고 하면 신뢰도가 떨어진다. memory-based 방법의 단점들을 극복하기 위한 방법이 model-based 방법이다.

협업 필터링은 여러 분야에서 널리 연구 및 적용되고 있다. 기사 추천시스템인 GroupLens[10], 음악 추천 시스템인 Ringo[13] 및 World Wide Web에서 사용자와 관련된 정보를 찾아주는 PHOAKS[14] 등이 개발되었으며, Amazon, CDNow, Drugstore, MovieFinder 등에서 협업 필터링 기법을 사용하여 상품을 추천하고 있다.

2.3 희박성

희박성은 고객이 항목에 부여한 평가치의 개수가 부족하여 추천의 성과가 떨어지는 문제점이다. 취급하는 항목의 종류가 많은 Amazon이나 e-Bay 같은 대형 인터넷 쇼핑몰에서는 고객들이 실제로 경험하고 평가하는 항목들의 개수가 전체의 1%도 안 될 뿐만 아니라 평가를 하지 않는 고객들도 많다. 즉, 협업 필터링 과정 중 제 1단계에서 준비되어야 하는 고객*항목 행렬의 대부분이 비어 있게 되므로 협업 필터링이 제대로 작동되지 않는다.

희박성의 문제를 완화하기 위한 많은 연구들이 진행되고 있다. Breese[11]는 평가치가 비어있는 칸에 default 평가치를 부여하는 방법인 default voting을 사용하였고, Sarwar[4]는 고객들 간이 아닌 항목들 간의 유사도를 계산하여 추천에 활용하는 방안을 제안하였다. 개인 성향 중에서는 장르, 나이, 혈액형, 성별, 지역, 성격 등을 이용하여 추천하였다[1, 4]. 다른 방법으로는 차원 감소법이 있다. 이는 직접적으로 고객*항목 행렬의 차원을 감소시키는 방법으로서 고객이나 항목을 군집화하여 군집화된 그룹을 기본단위로 하여 협업 필터링 기법을 적용하는 방법이다[1]. 그리고 협업 필터링과 내용 기반 필터링을 결합한 하이브리드 기법은 데이터가 희박하여 협업 필터링 기법으로는 추천이 힘들 때, 고객들의 인구 통계정보나 거래내역기록, 항목의 속성 정보를 활용하여 추천하는 기법이다.

3. 데이터 준비와 영화 추천 시스템의 구축

기존의 협업 필터링 방법을 이용한 영화 추천 시스템에서 희박성 문제를 해결하기 위한 방법으로 여러 가지의 개인 성향들을 사용하였다. 개인 성향 중에서는 가장 많이 사용한 것은 장르를 비롯하여 나이, 혈액형, 성별, 지역, 성격 등이 사용되었다[1, 4]. 본 논문에서는 사용된 6개의 개인 성향 중에서 추천시스템 구축에 영향을 크게 미치는 요인을 분석하여 가중치를 부여하여 개선된 영화추천 시스템을 제안한다.

본 논문의 영화추천 시스템을 위해 사용한 데이터는 공개 데이터인 MovieLens로서 미네소타 대학의 GroupLens Research Project에서 1997년 9월부터 1998년 4월 까지 수집된 데이터로 1682개의 영화에 대한 943명의 고객들의 평가치가 구성되어 있으며 총 10,000개 레코드가 있다[1, 10].

MovieLens 데이터는 User Id와 Movie Id, 각 고객들이 관람한 영화에 대해 1점에서 5점까지의 5점 척도로 부여한 평가치로 구성되어 있고, 고객*영화 행렬 칸의 총 개수는 1586126(=943*1682)개 인데, 이 중에서 10000개의 칸에 평가치가 부여되어 있어서 희박성이 93.7%이다.

본 논문에서 사용한 MovieLens 데이터에서 개인화 요인은 주어지는 것이 아니기 때문에 제공된 정보를 바탕으로 요인들을 추출해야 할 필요성이 있다. MovieLens에서 제공하는 개인정보의 일부는 Fig. 1과 같다[3].

```
1|24|M|technician|85711
2|53|F|other|94043
3|23|M|writer|32067
4|24|M|technician|43537
5|33|F|other|15213
6|42|M|executive|98101
7|57|M|administrator|91344
8|36|M|administrator|05201
9|29|M|student|01002
10|53|M|lawyer|90703
11|39|F|other|30329
```

Fig. 1. MovieLens' user data(part)

6개의 개인 성향 중 가장 많이 사용되는 장르는 미국의 최대 영화 데이터베이스인 IMDB(Internet Movie Database, <http://us.imdb.com>) 기준에 따라 18개의 장르로 구분된다. 18개 장르는 Action, Adventure, Animation, Children, Comedy, Crime, Documentary, Drama, Fantasy, Film-Noir, Horror, Musical, Mystery, Romance, Sci-Fi, Thriller, War, Western 등이다[1].

나이는 전체적인 분포가 균일하게 이루어지도록 클러스터링 방법을 사용하였으며 최종적으로 8개 그룹으로 구성하였다. 지역은 미국을 기준으로 수집된 자료이므로 미국의 zip-code를 이용하여 지역을 구분하였다. 미국의 zip-code는 USPS.COM에서 제공하는 페이지에서 지역을 구분할 수 있다. 데이터의 가공 중에서 지역이 조회되지 않는 일부의 인원이 있었으며 이 인원들은 지역에 대한 개인화요인을 파악할 수 없으므로 이러한 인원을 제외한 총 906명의 데이터를 변화시켰다. 변화시킨 데이터 중에서 엘리버마에 속해있는 인원은 3명에 불과하고 아칸소 주에 속한 인원은 1명이 존재하는 등 일부 주의 데이터가 충분하지 못하여 미국 통계청에서 구분하는 4개의 지역(북부, 남부, 중서부, 서부)로 지역 데이터를 변경하였다.

성격의 경우 정의할 수 있는 판단 기준이 필요하였으며 MBTI(Myers-Briggs Type Indicator)의 16가지 경우로 분류 하였으며 16개의 경우는 ISTJ, ISFJ, INFJ, IMTJ, ISTP, ISFP, INFP, INTP, ESTP, ESFP, ENFP, ENTP, ESTJ, ESFJ, ENFJ, ENTJ 이다[15].

성별은 남녀로 구분하였고, 혈액형은 A형, B형, AB형, O형으로 구성하였다.

협업 필터링을 위한 사용자 평점 행렬은 어느 사용자가

196	242	3	영화번호	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
186	302	3	사용자	1	5	3	4	3	3	5	4	1	5	3	2	5	5	5	5
22	377	1		2	4								2						
244	51	2		3									4		4	4			
166	346	1		4										4	2	5	3		
298	474	4		5	4	3													
115	265	2		6	4				2	4	4				4	2	5	3	
253	465	5		7				5		5	5	4	3	5					
				8						3									
				9					5	4									
				10	4		4		4	4		4	5	3				4	
				11					4	5			2	2				5	
				12			5											5	
				13	3	3	5	1	2	4	3		1	5	5	4			

Fig. 2. rating data matrix(part)

어떤 영화를 보고 내린 평점 자료가 존재하므로 평점 자료 일부와 구성된 행렬의 일부는 Fig. 2과 같다[3].

4. 협업 필터링에서 개인 성향을 고려한 추천시스템

본 논문에서는 기존 협업 필터링 기법에 개인의 특징을 결정하는 개인화 요인들로 나이, 선호장르, 혈액형, 성별, 지역, 성격을 고려하여 목표 고객이 가장 선호할 것이라 예측되는 영화목록을 추천하기 위한 추천 시스템을 개발한다. 본 연구에서 제안하는 추천시스템의 구조는 Fig. 3과 같다.

본 논문에서 사용자가 평가한 평점 데이터가 충분한 경우에는 기존의 협업 필터링 방법으로 K-means clustering하고 추천리스트를 작성하여 추천 영화 목록을 제공한다. 하지만 사용자의 평가한 평가 데이터가 충분하지 못한 경우 개인의 특징을 반영하는 개인화 요인을 고려하여 평가치가 존재하는 다른 고객들과 K-means clustering 기법으로 고객들을 군집화한다. 고객과 군집화된 고객들이 본 영화 들 중 평점 데이터를 이용해 추천 목록을 구성한다.

추천을 위한 개인성향은 매우 다양한 종류가 있을 수 있다. 또한 어떠한 항목을 추천하는가에 따라 그 비중이 달라지기도 한다. 본 연구에서는 영화를 항목으로 추천해주는 시스템으로서 개인의 영화 선호도를 파악하기 위한 개인성향을 알아야 함으로 기존의 개인 성향에 사용된 장르, 성격, 나이, 성별, 혈액형, 지역 등 사용자에게 가장 적합한 영화를 추천해 줄 수 있는 개인성향들을 분석한다. 분석하기 위한 도구로는 XLMiner를 사용한다.

6개의 개인 성향을 가지고 구현할 수 있는 모든 가짓수는 $2^6 = 64$ 개에서 아무것도 없는 경우를 제외하면 Fig. 4와 같은 63가지 경우의 수가 있다[3]. 이 방법들 중에서 영화추천 시스템에 가장 영향을 주는 방법들을 찾기 위해 63가지에 대해 MovieLens 데이터에 적용하여 추천되는 영화들을 분석한다.

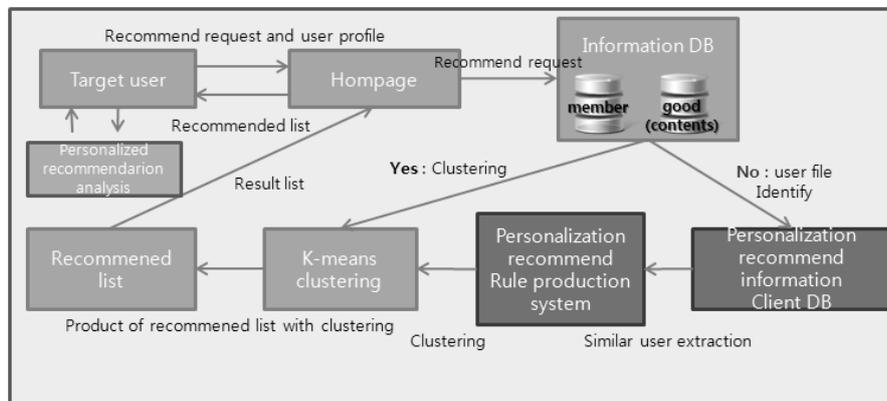


Fig. 3. movie recommender system

1Element	2Element	3Element	4Element	5Element	6Element
1 Age	7 Age	Sex	22 Age	Sex	Genre
2 Sex	8 Age	Genre	23 Age	Sex	Blood type
3 Genre	9 Age	Blood type	24 Age	Sex	Character
4 Blood type	10 Age	Character	25 Age	Sex	Region
5 Character	11 Age	Region	26 Age	Genre	Blood type
6 Region	12 Sex	Genre	27 Age	Genre	Character
	13 Sex	Blood type	28 Age	Genre	Region
	14 Sex	Character	29 Age	Genre	Blood type
	15 Sex	Region	30 Age	Blood type	Region
	16 Genre	Blood type	31 Age	Character	Region
	17 Genre	Character	32 Sex	Genre	Blood type
	18 Genre	Region	33 Sex	Genre	Character
	19 Blood type	Character	34 Sex	Genre	Region
	20 Blood type	Region	35 Sex	Blood type	Character
	21 Character	Region	36 Sex	Blood type	Region
			37 Sex	Character	Region
			38 Genre	Blood type	Character
			39 Genre	Blood type	Region
			40 Genre	Character	Region
			41 Blood type	Character	Region

Fig. 4. combinations of 6 personal propensities

추천되는 영화는 10개씩을 추천하도록 하였으며 Table 1 은 63개의 모든 경우의 수에 대해서 추천된 영화 목록의 결과이다[3]. 이를 추천하기 위해서 사용된 방법은 (Algorithm 1)이다.

Table 1. list of recommendation movies(part)

Age	Sex	Preferred genre
"Saint of Fort Washington, The" (1993)	"Great Day in Harlem, A" (1994)	"Wrong Trousers, The" (1993)
"Star Kid" (1997)	"Saint of Fort Washington, The" (1993)	"Close Shave, A" (1995)
'Santa with Muscles (1996)'	'Marlene Dietrich: Shadow and Light (1996)'	'Casablanca (1942)'
⋮	⋮	⋮
Blood type	Personality	Region
"Prefontaine" (1997)	"Casablanca" (1942)	"Someone Else's America" (1995)
"Star Kid" (1997)	"Maya Lin: A Strong Clear Vision" (1994)	'Entertaining Angels: The Dorothy Day Story (1996)'
""Great Day in Harlem, A (1994)""	""Close Shave, A (1995)""	"Great Day in Harlem, A" (1994)
⋮	⋮	⋮
Age, genre, Blood type, personality, region	sex, genre,blood type, personality	sex,blood type, region
"Casablanca" (1942)	"Santa with Muscles "(1996)	"Great Day in Harlem, A" (1994)
"Close Shave, A" (1995)	"Close Shave, A" (1995)	"Casablanca" (1942)
"Wrong Trousers, The" (1993)	"Wrong Trousers, The" (1993)	"Close Shave, A" (1995)
⋮	⋮	⋮
age, blood type, region	sex, personality, region	age, sex, region
"Star Kid" (1997)	"Someone Else's America" (1995)	"Star Kid" (1997)
"Pather Panchal" (1955)	"Pather Panchali" (1955)	"Pather Panchali" (1955)
"Casablanca" (1942)	"Wrong Trousers, The" (1993)	"Casablanca" (1942)

63가지의 방법들에 대해서 10개씩 추천된 영화들을 분석하여 6개의 개인성향중 영화 추천에 가장 영향을 주는 요인을 찾기 위해 Table 1로부터 각 방법에서 추천된 영화들의 수를 계산하였으며 추천된 영화 중 5번 이상 추천되어야만 의미 있는 영화 추천으로 판단하여 분석한다. 63가지의 방법들에 대한 추천된 영화의 수를 그래프로 표시하면 Fig. 5와 같다.

Algorithm 1. movie recommendation system based-on personal propensity

- 1 : INPUT: user*movie matrix U_k :user, R:rating table, N:recommended user
- 2 : OUTPUT: RL : recommended movie list
- 3 : Procedure:
- 4 : identify target user Profile
- 5 : added a personal propensity
- 6 : target user clustering
- 7 : recommended movie list with 10 movie

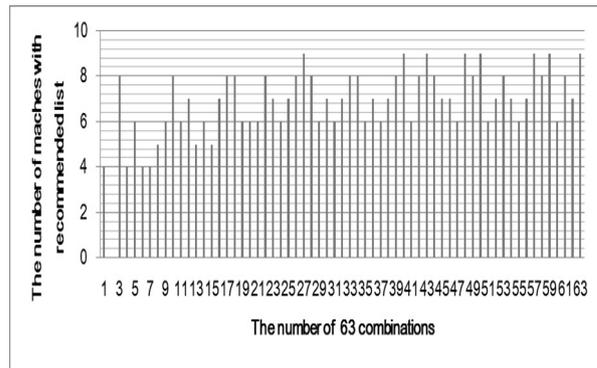


Fig. 5. analysis bar graph for 63 combinations

추천 영화에 대한 가중치 계산은 개인성향중에 가장 영향을 주는 주요 요인들을 분석하기 위해 본 논문에서는 6개의 개인성향에 대해 부분 집합의 관계를 이용하여 분석한다.

예를 들어 나이를 기본으로 생각한다면 다음으로는 나이와 나머지 속성 5 가지를 포함 관계로 계산한다. 이를 순서쌍을 이용해서 표현하면 (나이, 선호장르), (나이, 혈액형), (나이, 지역), (나이, 성별), (나이, 성격) 등 5가지의 방법에 대해서 영화 추천 중 몇 가지가 공통으로 나타나는 지를 분석한다. 다음으로는 (나이, 선호장르, 혈액형), (나이, 선호장르, 지역), (나이, 선호장르, 성별), (나이, 선호장르, 성격)에 대해서 분석하였으며 같은 방법으로 (나이, 선호장르, 혈액형, 지역), (나이, 선호장르, 혈액형, 성별), (나이, 선호장르, 혈액형, 성격), (나이, 선호장르, 혈액형, 성별, 지역), (나이, 선호장르, 혈액형, 성격, 지역), (나이, 선호장르, 혈액형, 성격, 성별), (나이, 선호장르, 혈액형, 성격, 성별, 지역) 등 63가지의 조합을 분석한다.

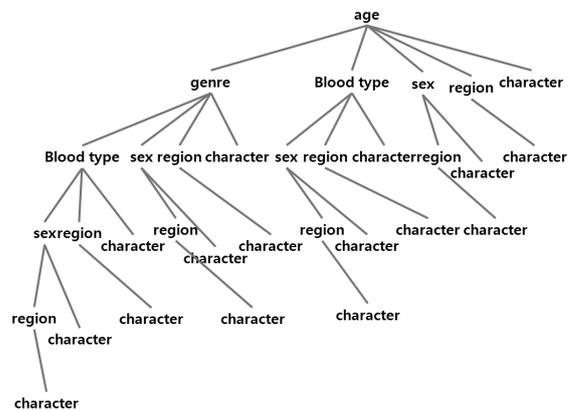
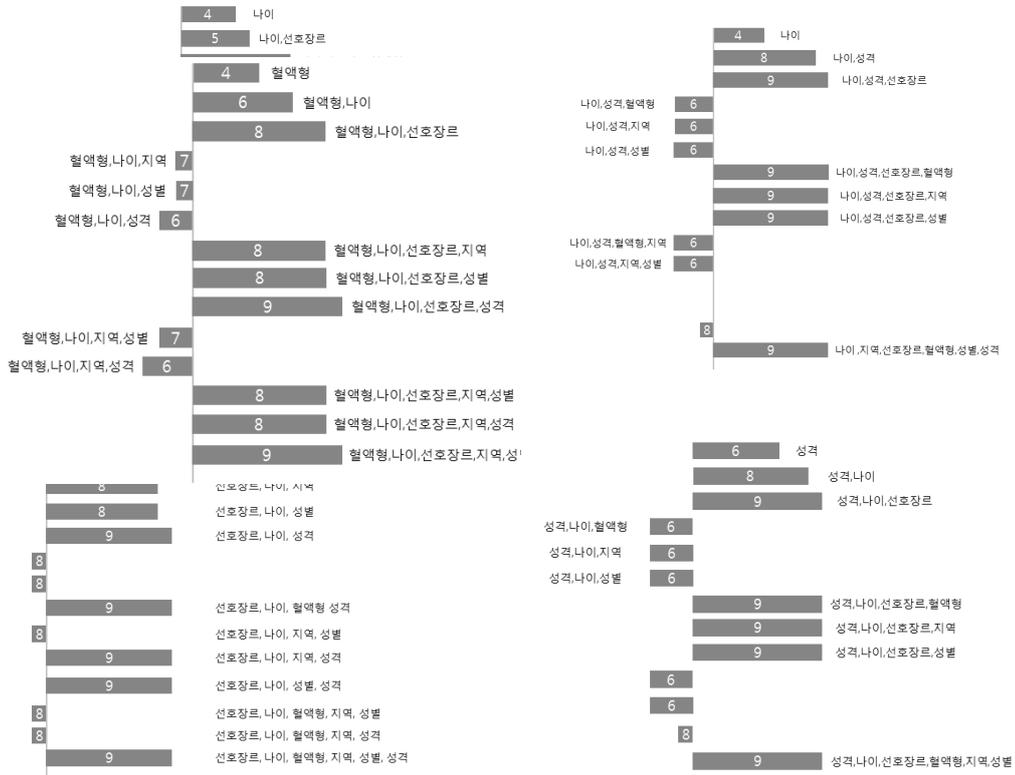


Fig. 6. combination of ages



르, 지역, 성별), (나이, 선호장르, 지역, 성격), (나이, 선호장르, 성별, 성격), (나이, 선호장르, 혈액형, 지역, 성별), (나이, 선호장르, 혈액형, 지역, 성격), (나이, 선호장르, 혈액형, 지역, 성별, 성격) 등에 따른 개인성향의 가중치 분석을 수행한다. 나이를 기본으로 하는 경우 모든 방법들은 Fig. 6과 같이 16가지의 방법이 있다.

같은 방법으로 다음으로는 성별, 성격, 혈액형, 선호장르, 지역 등 총 30가지(=16*6) 방법에 대해 분석을 수행하였다. 분석한 내용을 가시화 그래프로 몇 가지만 보면 다음과 같다.

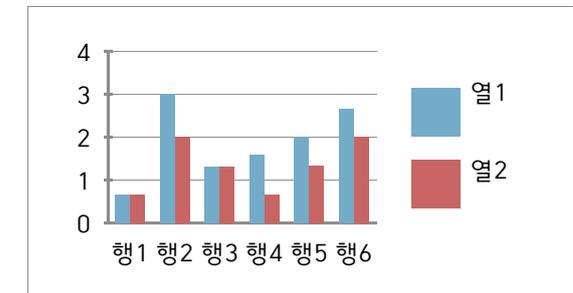
각 방법들에 대한 성능 평가는 MAE를 통해 성능 평가를 시행한다. MAE는 실제 목표 값과 출력 값의 차이를 평균한 것으로 Equation (2)와 같이 나타낼 수 있다.

$$MAE = \frac{\sum_{k=1}^n |y_k - t_k|}{n} \quad (2)$$

이때 y_k , t_k 는 각각 출력 값과 목표 값을 나타내고 n 은 샘플의 총 수를 나타낸다.

본 논문에서는 방법들 중에 3개의 개인 성향을 가지고 최대값에 도달한 (나이, 선호장르, 성격)과 (지역, 선호장르, 성격) 방법에 대해서 부분 집합을 변화시키면서 만들 수 있는 12가지의 방법에 대해 MAE 값을 구했으며 결과는 다음과 같다.

이중에서 MAE가 가장 작은 값인 (나이, 선호장르, 성격)과 (지역, 선호장르, 성격) 방법을 택하였다.



이젠 가중치를 계산하기 위해 (나이, 선호장르, 성격)과 (지역, 선호장르, 성격)에 도달하는 방법 중 추천된 영화가 일치하는 개수를 가지고 가중치를 계산한다. 각 개인성향에 대한 가중치에 대한 계산은 Equation (3)과 같다.

$$\text{각 개인 성향 가중치}_{k(=1,4)} = \frac{\sum_{i=1}^{12} m_k}{\sum_{k=1}^4 \sum_{i=1}^3 m} \quad (3)$$

m_k 는 각 개인 성향으로 k 를 포함하는 추천 영화의 수를 나타내고 m 은 각 성향에 대해 추천된 영화의 수이다.

지역에 대한 가중치(m_1) = 99/576, 성격에 대한 가중치(m_2) = 191/576, 나이에 대한 가중치(m_3) = 88/576, 선호장르에 대한 가중치(m_4) = 198/576 이다. 그러므로 가중치 계산을 해보면 선호장르가 34%, 성격이 33%, 지역이 17%, 나이가 15% 이다.

전반적으로는 나이는 성격, 성별은 선호장르, 성격은 나이와 선호장르에 효과가 있음을 보여주고 있다. 결국은 (선호장르, 성격, 지역, 나이)등 가중치를 가진 4가지 개인 성향을 가지고 영화 추천을 하면 6개의 개인 성향을 모두 사용한 경우와 같은 결과가 나왔다. 또한, (선호장르, 성격, 지역)과 (선호장르, 성격, 나이) 등 가중치를 부여한 3개의 개인 성향을 가져도 6개의 개인성향을 모두 사용한 경우와 같은 결과가 나왔다. 가중치를 부여한 방법은 (Algorithm 2)와 같다.

Algorithm 2. movie recommendation system with personal propensities	
1 :	INPUT: user*movie matrix U_k :user, R:rating table, N:recommended user
2 :	OUTPUT: RL : recommended movie list
3 :	Procedure:
4 :	identify target user Profile
5 :	added a personal propensity which is (genre, character, region) or (genre, character, age)
6. :	added weighting
7 :	target user clustering
8 :	recommended movie list with 10 movie

Fig. 7은 개인 성향을 통해 영화 목록을 추천받아 목표 고객에게 보여주는 화면이다.



Fig. 7. list of movie recommendation based on personal propensities and collaborative filtering

영화 추천 결과 6개의 개인 성향을 모두 사용한 경우 10개의 영화 추천 중 9개가 일치하였으며, 가중치를 사용한 (선호장르, 성격, 지역) 혹은 (선호장르, 성격, 나이) 등 3개의 개인 성향을 사용한 경우에도 9개가 정확하게 일치함을 보였으며, 선호장르 한 가지만 고려하여도 9개중 8개가 일치하였다. 반면에 (선호장르, 성격)의 개인 성향을 이용한 경우에도 9개중 8개가 일치한다.

결론적으로 시스템의 구성도에 따라서 (선호장르, 성격, 지역) 혹은 (선호장르, 성격, 나이) 등 3개의 성향을 가중치를 부여하여 추천 시스템을 구성하는 것이 가장 적절한 방법임을 실험결과로 알 수 있다.

5. 결 론

유비쿼터스 환경에서는 정보들은 기하급수적으로 증가하면서 사용자가 원하는 정보를 찾기가 점점 어려워지고 있다. 추천 시스템 중에서 가장 성공적인 방법이 협업 필터링이다. 그러나 협업 필터링에는 희박성의 문제가 있으며 이러한 문제점을 해결하는 방법으로 가중치를 부여한 개인성향을 이용한 추천시스템을 제안하였다. 본 논문에서는 개인화 추천을 위해 많은 개인 성향이 필요하지만 가중치를 주어서 가장 적은 개수의 개인 성향만을 가지고도 성능이 매우 우수한 영화 추천시스템을 제안하였다. 본 논문에서 제안한 시스템은 (선호장르, 성격, 지역) 혹은 (선호장르, 성격, 나이) 등에 가중치를 부여하여 6개의 개인성향을 모두 사용한 추천과 같은 결과를 얻었다.

본 시스템의 특징은 사용자가 시스템을 이용하기 전에 사용자 정보를 입력할 때 추천에 필요한 개인화 요인을 입력함으로써 충분한 평가 데이터가 없더라도 최적의 영화 추천 리스트를 제안하는 것으로 희박성을 해결하는 것이 특징이라 할 수 있다.

참 고 문 헌

- [1] J. Lee, S. Park, "Performance Improvement of a Movie Recommendation System using Genre-wise Collaborative Filtering," Journal of Intelligence Information Systems, Vol.13, No.4, pp.65 - 78, 2007.
- [2] Xiaoyuan Su and Taghi M. Khoshgoftaar, "A Survey of Collaborative Filtering Techniques," Advances in Artificial Intelligence, Vol.2009, Article ID 421425, pp.1 - 19, 2009.
- [3] W. Jeong, S. Kim, D. Park, and J. Kwak, "Performance Improvement of a Movie Recommendation System based on Personal Propensity and Secure Collaborative Filtering," J Inf Process Syst, Vol.9, No.1, pp.707 - 722, 2013.
- [4] B. Sarwar, G. Karypis, J. Konstan, J. Riedl, "Item based Collaborative Filtering Recommendation Algorithms" Proceedings of the 10th International World Wide Web Conference, pp.285-295, 2001.
- [5] T. Hong, H. Lee, B. Su, "A Web Personalized Recommender Systems Using Clustering - based CBR", Journal of Intelligence Information Systems, Vol.11, No.1, pp.107-121, 2005.
- [6] S. Lee, H. Lee, and H. Kim, "Dynamic Recommender on user taste tendency model : Focusing on Movie Recommender Systems," K.I.S.S. Journal, Vol.31, No.2, pp.153-163, 2004.
- [7] R. J. Mooney and L. Roy, "Content-based Book Recommending Using Learning for Text Categorization," In Proceedings of ACM SIGIR '99 Workshop Recommender Systems: Algorithms and Evaluation, 1999.
- [8] J. Le and O.R. Zaiane, "Combining Usage, Content, and Structure Data to Improve Web Site Recommendation," Fifth

- International Conference on Electronic Commerce and Web Technologies (EC-Web '04), pp.305-315, 2004.
- [9] R. J. Mooney and L. Roy, "Content-based Book Recommending Using Learning for Text Categorization," In Proceedings of ACM SIGIR '99 Workshop Recommender Systems: Algorithms and Evaluation, 1999.
- [10] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom and J. Riedl, "Grouplens : An Open Architecture for Collaborative Filtering of Netnews," Proceedings of the ACM Conf. on Computer Supported Cooperative Work, pp.175-186, 1994.
- [11] J. S. Breese, D. Heckeman and C. Kadie, "Empirical analysis of predictive algorithms for collaborative filtering," Proceedings of the Fourteenth Annual Conference on Uncertainty in Artificial Intelligence, pp.43-52, 1998.
- [12] G. Linden, B. Smith, and J. York, "Amazon.com recommendations:item-to-item collaborative filtering," IEEE Internet Computing, Vol.7, No.1, pp.76-80, 2003.
- [13] U. Shardanand and P. Maes, "Social information filtering: algorithms for automating 'word of mouth'," Proceedings of the SIGCHI conference on Human factors in computing system, pp.1210-217, 1995.
- [14] L. Terveen, W. Hill, B. Amento, D. McDonald, J. Creter, "PHOAKS: a system for sharing recommendations," Communications of the ACM Vol.40, No.3, pp.59-62, 1997.
- [15] S. Lee, D. Park, M. Hong, H. Lee, Y. Jeong, "A Personalized Movie Recommendation System Based on the Blood Types and Collaborative Filtering," Proceedings of Korea information Processing Society Fall Annual Conference, Vol.16, No.2, 2009.



박 두 순

e-mail : parkds@sch.ac.kr

1988년 고려대학교 전산학전공(이학박사)

2011년~현 재 한국정보처리학회 부회장

2009년~2012년 한국정보처리학회 영문

논문지 편집위원장

1985년~현 재 순천향대학교 컴퓨터

소프트웨어공학과 교수

관심분야: 병렬처리, 데이터마이닝, 멀티미디어 정보처리