

논문 2010-47CI-4-3

확장된 협업 필터링을 활용한 선호 요소 가변 추천 시스템

(Preference Element Changeable Recommender System based on
Extended Collaborative Filtering)

오 정 민*, 문 남 미*

(Jung-Min Oh and Nam-Mee Moon)

요 약

모바일 환경은 작은 단말 화면, 제한된 검색 기능으로 인해 보다 정확하고 개인화된 정보 제공이 필요하다. 본 논문은 유용한 정보 제공을 위한 필터링 기법으로 활용되는 추천 시스템 중 협업 필터링을 이용하여 모바일 상에서 사용자의 관심 그룹을 선택적으로 반영하는 추천 시스템을 구성한다. 1차 단계로 사용자의 선호 정보와 인구통계학적 특성을 동시에 고려하여 관심 그룹을 형성하고 2차 단계로 사용자가 관심 그룹의 추천 여부를 스스로 선택하도록 함으로써 최종 추천 리스트를 재구성한다. 이는 지금까지 일반적으로 추천 리스트를 제공하였던 것에서 벗어나 사용자의 선호를 보다 적극적으로 고려한 양방향적 유동적 추천 리스트 제공이 가능해짐을 의미한다. 마지막으로 사용자의 선택 여부에 따른 추천 케이스를 도출하여 iPhone 환경에 적용한 결과를 제시한다.

Abstract

Mobile devices wide spread among users after the release of Apple's iPhone, especially in Korea. Mobile device has their own advantages in terms of weight, size, mobility and so on. But, on the contrary, mobile device has to provide more accurate and personalized information because of a small screen and a limited function of information retrieval. This paper presents a user's preference element changeable recommender system by employing extended collaborative filtering as a technique to provide useful information in a mobile environment. Proposed system reflects user's similar groups by simultaneously considering users' information with preferences and demographic characteristics. Then we construct list of recommenders by user's choice. Finally, we show the implementation of a prototype based on iPhone.

Keywords : Recommender System, Collaborative filtering, Interactive mobile, Personalized, Web 2.0

I. 서 론

지하철이나 버스 등의 대중교통을 이용하면서 혹은 길거리에서 누군가를 기다리며 모바일 기기를 손에 들고 엔터테인먼트를 즐기는 풍경은 이제 너무나 익숙한 모습이 되었다. 국내에서는 별다른 재미를 보지 못해왔

던 스마트폰도 애플의 아이폰 등장 이후 가파른 성장세를 보이며 사용자 사이에 광범위하게 보급되었으며 이를 뒷받침하는 언제 어디서나 네트워킹이 가능한 모바일 환경 기술도 하루가 다르게 발전하고 있다.

웹 2.0의 도래로 사용자 스스로 콘텐츠의 적극적 생산자 역할을 수행하게 되면서 UCC(User Created Content) 유통의 대표 사이트 격인 유튜브는 매 1분마다 24시간 분량의 동영상 업로드가 일어난다고 발표한 바 있다^[1]. 이는 언제 어디서나 손쉽게 동영상을 업로드하고 다운로드할 수 있는 모바일 환경이 조성됨에 따라 동영상 소비가 더욱 적극적으로 선호되고 있음을 의미한다^[2].

* 정희원, 호서벤처전문대학원 IT응용기술학과
(Department of IT Application Tech, Hoseo Graduate School of Venture)

※ 이 논문은 2010년도 정부(교육과학기술부)의 재원으로 한국과학재단의 지원을 받아 수행된 연구임
(No.2010-0000487).

접수일자: 2010년6월1일, 수정완료일: 2010년7월7일

스마트폰의 보급과 네트워크 환경의 발달로 다양한 콘텐츠에 대한 사용자 요구는 커지고 있으나 모바일 기기의 작고 가벼운 사이즈로 인한 이동성의 편리함만큼 작은 화면 사이즈 및 제한적 정보 검색 방식이 여전히 문제로 지적된다. 하루 수백 개가 생산되는 콘텐츠 중에서 자신이 관심 있는 콘텐츠의 존재 여부를 확인하는 것 자체가 모바일 상에서는 쉽지 않다. 다른 기기에 비해 보다 개인화된 서비스가 중시되고 사용자에게 적절한 정보 중심으로 콘텐츠를 제공해야 할 필요성이 높은 성격의 기기임에도 불구하고 이러한 제약 사항으로 인해 정보 누수 현상이 발생한다. 따라서 이를 극복하기 위한 정보 필터링 기법으로 추천 시스템에 대한 관심이 점차 높아지고 있다.

추천 시스템(Recommendation System)은 특정 사용자에게 사용자가 선호할 것이라 예상되는 새로운 콘텐츠 혹은 아이템을 제공하는 시스템을 의미하는 것으로^[3~4], 90년대 온라인 전자상거래 사이트를 중심으로 발전되어 왔다. 사용자의 선호도를 기초로 하여 콘텐츠를 추천하기 때문에 정보 검색 시스템, 검색 엔진과 구별되며 '개별화(individualized)'된 '흥미 있고 유용한(interesting and useful)' 결과를 얻을 수 있다^[8].

웹을 중심으로 발전되어 온 추천 시스템은 모바일 환경으로 확장되며 유비쿼터스 환경을 인식하는 지능형 에이전트에 대한 연구에 좀 더 집중화 되는 경향을 보이고 있다. 그러나 사용자 간의 소통과 개입이 활발한 웹 2.0 시대의 콘텐츠 추천은 암시적인 사용자의 행동 패턴 정보 수집 뿐 아니라 보다 직접적이고 명시적인 사용자 선택권을 제시할 필요성이 있다.

이에 본 논문은 모바일 상에서 협업 필터링 기법을 통해 사용자의 선호도가 반영된 추천 목록을 제시하는 시스템을 구성하되 지금까지 일반적으로 제시되던 추천 시스템의 한계를 탈피하여 관련도가 높은 유사 그룹의 가중치를 사용자가 직접 조절하는 양방향적 선택형 추천 시스템을 제안한다.

II. 관련 연구

추천 시스템은 검색어를 직접 입력하는 정보 검색 시스템과는 달리 사용자의 선호도가 반영된 결과를 얼마나 정확하게 제시하는지가 핵심이다. 이를 위해 사용자가 부여한 평점, 관심 콘텐츠 저장 등의 명시적 프로파일과 구매 이력, 행동 패턴 등의 잠재적 프로파일을 고

려하여 필터링에 반영하여야 한다^[6]. 추천 시스템은 추천 방식에 따라 콘텐츠 기반 기법, 협업 필터링 기반 기법, 이 두 가지를 혼합한 하이브리드 기법으로 구분되는데 그 중에서 가장 성공적인 콘텐츠 추천 기법이자 보편적인 기법은 협업 필터링 기법으로 알려져 있다^[7].

콘텐츠 기반 기법은 콘텐츠를 직접 분석하여 콘텐츠와 콘텐츠, 콘텐츠와 사용자 선호도 간의 유사도에 따라 콘텐츠를 추천해주는 방식으로, 콘텐츠가 가진 특성이 가장 핵심적 추천 요소이다^[8]. 협업 필터링은 사용자들의 프로파일 정보를 활용하여 목표 사용자가 높게 평가할 것으로 예측되는 콘텐츠나 서비스를 추천하는 것으로, 목표 사용자의 평가 값과 다른 사용자의 평가 값을 이용한다^[9]. 협업 필터링을 활용하여 콘텐츠를 추천하는 과정은 크게 사전 평가 데이터 매트릭스 구성(1단계), 사용자 연관 정보 및 이웃 데이터 구성(2단계), 최근접 이웃에 따른 추천 상품 결정(3단계)으로 구성된다.

과정을 좀 더 자세히 살펴보면 그림 1에서와 같이 1 단계에서는 n명의 사용자들이 m개의 콘텐츠를 소비하고 평가한 값들을 토대로 n by m의 사용자*콘텐츠 매트릭스를 만든다. 2단계에서는 1단계의 매트릭스를 기반으로 사용자 간의 유사도를 계산한 후 최근접 이웃을 구성한다. 유사도 계산 지수로는 일반적으로 Pearson Correlation Coefficient와 Cosine이 사용된다. 3단계에서는 최근접 이웃의 평가를 이용하여 목표 사용자가 평가하기 전인 콘텐츠의 평가를 예측하고 그에 따른 Top-N 추천 리스트를 제공하게 된다.

각 추천 시스템은 장단점을 내포하고 있는데, 협업 필터링은 희박성(Sparsity), 확장성(Scalability)이 주요 한계로 지적된다^[6, 8, 10~11]. 희박성은 1단계에서 매트릭스에 넣어야 할 평가치가 부족하여 추천의 성과가 떨어지는 것을 말한다. 이는 월드 스타트와도 유사한 개념으로 사용자의 선호도를 판단할 수 있는 데이터가 부족하여 발생한다. 확장성은 데이터가 누적됨에 따라 사용자 수와 콘텐츠 수가 늘어나고 그에 따라 최근접 이웃

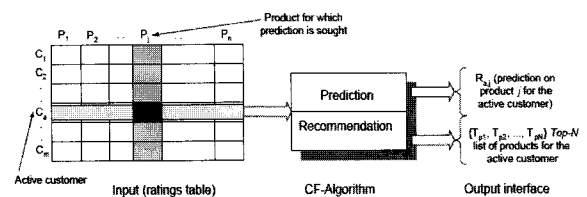


그림 1. 협업 필터링의 일반적 프로세스
Fig. 1. The Collaborative Filtering Process.

을 도출하기 위한 연산량이 기하급수적으로 늘어나는 것을 말한다. 데이터가 증가할수록 추천 목록을 생성하기까지 많은 시간이 소요되며 이로 인해 시스템의 효율성이 떨어지게 된다.

III. 사용자 선택형 협업 필터링 추천 시스템

1. 추천 시스템 개요

본 연구에서 제안하는 사용자 선택형 협업 필터링 추천 시스템(User-Selectable Recommendation System)은 사용자가 협업 필터링 2단계에서 도출되는 유사 사용자 그룹(이하 SG) 중 3단계의 추천 목록 구성에 관여할 그룹을 직접 선택하도록 선택권을 부여함으로써 일방적으로 제시되던 추천 목록을 양방향으로 확장하는데 목적을 둔다^[12]. 이는 사용자의 선택 결과에 따라 유연한 추천 목록을 제시함과 동시에 사용자가 자신의 관심 집단을 직접적으로 선택하게 함으로 관심도에 따라 선별된 추천 목록 제시가 가능하다는 장점이 있다. 다음의 그림 2에서 사용자 선택형 콘텐츠 추천 과정을 제시한다.

콘텐츠가 아닌 선호 장르(①, 이하 PG)를 기반으로 사용자 그룹 매트릭스를 구성함은 평점 수가 부족하여 추천의 성과가 떨어지는 희박성(Sparsity)의 문제를 보완하기 위하여 차원 감소법을 적용한 결과이다. 차원 감소법은 1단계에서 확보되어야 하는 사용자와 콘텐츠의 매트릭스 차원을 직접적으로 감소시키는 방법으로

사용자 혹은 콘텐츠를 군집화한 후 군집화된 그룹을 매트릭스의 기본 단위로 재구성하여 협업 필터링 기법을 적용하는 방법이다. 이를 통해 사용자가 콘텐츠에 부여한 평가값이 부족하여 추천 목록을 제시할 수 없는 희박성을 방지할 수 있다^[13].

이후 사용자와 동일한 PG를 공유하는 SG(②)를 사용자 정보에 기반하여 각 세 그룹으로 구성한다. 이들은 사용자와 PG가 같으며 연령대가 같거나(SG1), PG가 같으며 성별이 같거나(SG2), PG가 같으며 지역이 동일한 그룹(SG3)이다. 구성된 SG 중 사용자의 결정에 의하여 취사 선택된 SG를 합하여(③) 마지막 단계의 최근접 이웃을 구성하고(④) 그에 따른 추천 목록을 제시한다(⑤).

2. 추천 서비스 시나리오

제안하는 시스템에 따른 서비스 시나리오는 멀티미디어 콘텐츠 중 영화를 대상으로 하여 사용자 선택에 따른 콘텐츠 추천 리스트를 제공받는 것을 목표로 구성한다. 표 1에서 시나리오의 세부 내용을 제시한다.

세부 시나리오 2에서 사용자가 시청하고 평점을 부여한 콘텐츠에 따라 PG가 도출되고 사용자 정보에 기반한 고려 집단을 교차 분석하여 SG가 도출되면 이를 기반으로 4의 사용자 선택에 따른 추천 리스트가 제공된다. 이 시나리오는 사용자의 연령, 성별, 지역 정보가 사용자 개인 정보로 사전에 저장되어 있고, 콘텐츠 소비시 지역 정보 데이터가 업데이트됨을 전제로 한다.

표 1. 제안 서비스 시나리오

Table 1. The proposed service scenario.

목표	사용자 선택에 따라 추천 리스트를 재구성하여 제공한다
전제 상황	<ul style="list-style-type: none"> • 사용자는 그 동안 자신이 좋아하는 장르(분야)를 집중적으로 보고 평점을 부여해왔다. • 사용자는 이 장르들 중 나와 같은 연령을 가진 사람들이 추천하는 리스트가 궁금하다.
세부 시나리오	<ol style="list-style-type: none"> 1. 사용자가 모바일 기기에서 콘텐츠를 시청할 수 있는 Application을 실행한다. 2. Application에서 추천 콘텐츠의 편집 메뉴를 통해 '연령 추천'을 On으로 체크한다. 3. On으로 체크된 메뉴에 따라 추천 콘텐츠 리스트가 제공된다. 리스트는 사용자 '연령'과 사용자 '선호 장르'를 모두 고려한 리스트이다. 4. 사용자가 추천 리스트를 보고 마음에 드는 콘텐츠 하나를 선택하여 시청한다.

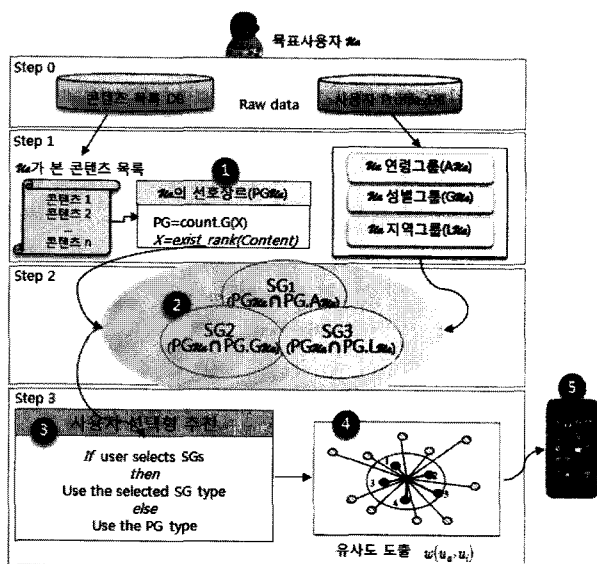


그림 2. 추천 시스템 프로세스

Fig. 2. Overall Process.

IV. 시스템 설계 및 결과

1. 데이터 구성

시스템 설계에 활용한 데이터는 미네소타 대학의 GroupLens Research Project에서 수집한 MovieLens 데이터 세트이다. MovieLens는 사용자들이 시청한 영화에 대한 선호도를 기반으로 영화를 추천하는 가상의 커뮤니티 사이트이자 추천 시스템으로, 데이터 세트는 1997년 9월에서 1998년 4월까지 수집된 1682편의 영화에 대한 943명의 사용자 평가(ratings)로 구성되어 있다 [14~15]. 원본 데이터는 영화와 사용자가 각기 구분되어 있어 사용자의 PG를 분석하기에 적합하지 않다. 따라서 전처리 과정으로 사용자 데이터에 PG를 추가하여 데이터 세트를 재구성하였다. 더불어 사용자 정보로 활용될 Zip-Code가 그룹화 되어야 지역에 따른 추천 목록을 제공할 수 있으므로 Zip-Code의 앞자리 숫자 2개의 규칙성을 활용하여 0~10까지의 코드로 변환하는 과정을 거쳐 다음 표 2와 같이 데이터를 수정하였다.

시스템 설계 및 어플리케이션 구현 환경은 Apache Tomcat 5.5, iPhone SDK(Xcode,Interface Builder) 3.2.1, JAVA SDK 1.6, Eclipse 3.5.2, My-Sql 5.1 이다.

표 2. 무비렌즈 수정 데이터
Table 2. MovieLens Modified User Data.

MovieLens Original DB				
Element	User 1	User n
ID	1	943
Age	7	73
Gender	Female / Male			
Occupation	21 groups (administrator, artist, doctor, educator, engineer, entertainment, executive, healthcare, homemaker, lawyer, librarian, marketing, none, other, programmer, retired, salesman, scientist, student, technician, Writer)			
Zip-code	1002,1040,1080,1331,1375,1581,1602 ...			

MovieLens Modified DB				
Element	User 1	User n
ID	1	941 (연속값 2명 제외)
Age	7	73
Gender	Female / Male			
Occupation	21 groups (administrator, artist, doctor, educator, engineer, entertainment, executive, healthcare, homemaker, lawyer, librarian, marketing, none, other, programmer, retired, salesman, scientist, student, technician, Writer)			
Zip-code	0,1,2,3,4,5,6,7,8,9,10			
P-Genre	PG1, PG2, PG3			

2. 정보 필터링 및 그룹화

데이터 세트에 따른 사용자들의 영화 시청수는 최소 20개에서 최대 737개까지이다. 시나리오에서 '사용자'로 언급된 목표 사용자(u_a)는 시청수를 기준으로 중앙값 근처의 사용자로 임의 선정한다. u_a 의 PG는 Drama, Action, Comedy 3개가 도출되었는데, 이는 빈도 분석에 따른 결과이다. 표 3과 같이 무작위로 10명의 사용자가 시청한 장르의 빈도 분석을 시행한 결과, 25% 이상의

표 3. 사용자 선호 장르 빈도 분석
Table 3. User's Genre Frequency Analysis.

영화 시청수	User 1	User 2	User 3	User 4	User 5	User 6	User 7	User 8	User 9	User 10									
104	49%	32	58%	81	44%	218	34%	48	46%	82	47%	14	61%	70	48%	12	50%	72	40%
66	31%	25	46%	47	26%	188	30%	28	27%	56	32%	13	57%	45	31%	10	42%	56	31%
41	19%	17	31%	36	20%	156	25%	25	24%	33	19%	11	48%	40	27%	8	33%	49	27%
25	12%	10	19%	34	18%	121	19%	25	24%	33	19%	5	22%	32	22%	7	29%	40	22%
24	11%	6	11%	25	14%	105	17%	22	21%	29	17%	4	17%	27	18%	5	21%	33	18%
22	10%	6	11%	20	11%	75	12%	12	12%	28	16%	4	17%	15	10%	5	21%	32	18%
21	10%	6	11%	16	9%	73	11%	12	12%	27	15%	3	13%	13	10%	4	17%	31	17%
20	9%	5	9%	15	8%	65	10%	10	10%	19	11%	3	13%	10	7%	3	13%	20	11%
14	7%	4	7%	14	8%	52	8%	7	7%	19	11%	2	9%	9	6%	1	4%	13	7%
13	6%	3	6%	14	8%	45	7%	7	7%	14	8%	1	4%	9	6%	1	4%	13	7%
13	6%	2	4%	12	7%	41	6%	6	6%	14	8%	0	0%	7	5%	1	4%	8	4%
12	6%	2	4%	9	5%	33	5%	3	3%	12	7%	0	0%	5	3%	1	4%	8	4%
10	5%	1	2%	8	4%	31	5%	2	2%	9	5%	0	0%	4	3%	1	4%	7	4%
6	3%	1	2%	7	4%	16	3%	2	2%	3	2%	0	0%	3	2%	0	0%	6	3%
6	3%	1	2%	7	4%	16	3%	2	2%	3	2%	0	0%	1	1%	0	0%	5	3%
4	2%	1	2%	5	3%	13	2%	2	2%	2	1%	0	0%	1	1%	0	0%	4	2%
3	1%	0	0%	3	2%	13	2%	0	0%	1	0%	0	0%	1	1%	0	0%	3	2%
1	0%	0	0%	1	1%	12	2%	0	0%	1	1%	0	0%	0	0%	0	0%	1	1%
0	0%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%
Up to 25%	2	3	2	1	3	0	2	0	3	3	4	3							

		Step 1	Step 2	Step 3						
				A	B	C	D	E	F	G
Sampling (500 user)	People	941	472	144	68	383	15	87	41	9
	%	100	50.2	15.3	7.2	40.7	1.6	9.2	4.4	1.0
u_a	People	941	432	136	146	234	25	47	39	6
	%	100	54.1	14.5	15.5	24.9	2.7	5.0	4.1	0.6

그림 3. 사용자 선택에 따른 SG 감소율
Fig. 3. SG deduction Percentage

분포율을 가진 장르는 평균 2.7개 장르가 도출되었다. 단발적 호기심이 아닌 지속적 선호도를 가진 장르라 할 수 있으려면 1/4 이상의 분포는 되어야 하므로 상위 평균 3개의 장르를 사용자의 PG로 적용한다.

각 사용자의 PG 데이터를 활용하여 u_a 의 SG를 도출한다. SG는 사용자 정보 유형에 따라 연령, 성별, 지역과 사용자의 PG가 교차 분석된 그룹이다. 각 SG는 사용자의 선택에 따라 통합 그룹으로 재구성될 수 있으므로 최종적으로 도출되는 SG는 그림 3과 같이 모두 7개이다. A는 PG&연령, B는 PG&지역, C는 PG&성별, D는 PG&연령&지역, E는 PG&연령&성별, F는 PG&지역&성별, G는 PG&연령&지역&성별 그룹이다.

Step이 진행됨에 따라 u_a 의 비교 그룹수가 축소되어 Step 3에서 G는 전체 사용자 대비 최대 0.6%까지 축소되었다. 500명을 샘플링하여 본 결과도 비슷한 결과를 보이며 G 단계에서는 1%로 줄어들음을 확인할 수 있다. 이는 곧 사용자가 어떠한 선택을 하느냐에 따라 보다 더 정확하게 타겟팅 된 SG가 도출될 수 있음을 의미한다.

3. 선택 케이스별 SG 유사도 도출

최종적으로 도출된 SG의 사용자 u_i 와 목표 사용자 u_a 간의 유사도 $w(u_a, u_i)$ 는 피어슨 상관계수(Pearson correlation coefficient)를 이용하여 계산한다. 계산식은 (1)과 같다.

표 4. u_a 의 SG1(A) 피어슨 상관계수
Table 4. u_a SG1(A)'s Pearson Correlation.

U_a	U_1	U_2	U_3	U_4	U_5	U_6	U_7	U_8	U_9	U_{10}	U_{11}	U_{12}	U_{13}	U_{14}
U_1	0.77	0.75	-0.20	0.76	0.63	0.66	0.52	0.44	0.73	0.91	0.61	0.36	0.50	0.14
U_2	0.77	1.00	(a)	-1.00	(a)	0.74	(a)	0.58	0.91	0.96	(a)	0.76	1.00	(a)
U_3	0.75	1.00	0.50	(a)	(a)	0.49	1.00	0.40	1.00	(a)	-1.00	(a)	(a)	0.42
U_4	0.20	(a)	-0.50	(a)	(a)	-0.27	(a)	0.46	(a)	(a)	(a)	(a)	(a)	-0.20
U_5	0.70	1.00	(a)	(a)	0.50	0.97	(a)	0.54	(a)	0.41	0.56	0.75	0.56	(a)
U_6	0.63	(a)	(a)	(a)	0.50	0.33	(a)	0.25	(a)	(a)	0.87	0.50	0.17	(a)
U_7	0.66	0.74	0.49	-0.27	0.97	0.33	0.76	0.31	0.76	0.96	-0.52	0.55	0.69	-0.15
U_8	0.52	(a)	1.00	(a)	(a)	0.76	0.96	(a)	(a)	(a)	(a)	(a)	(a)	(a)
U_9	0.44	0.58	0.40	0.46	0.54	0.25	0.31	0.90	0.33	0.97	0.76	0.43	-0.41	0.02
U_{10}	0.78	0.91	1.00	(a)	(a)	0.76	(a)	0.33	1.00	(a)	0.94	(a)	(a)	(a)
U_{11}	0.91	0.96	(a)	(a)	(a)	0.96	0.97	1.00	(a)	1.00	(a)	(a)	(a)	(a)
U_{12}	0.61	(a)	-1.00	(a)	0.41	(a)	-0.52	(a)	0.76	(a)	(a)	(a)	-0.50	0.46
U_{13}	0.36	0.76	(a)	(a)	0.56	0.87	0.55	(a)	0.43	0.94	1.00	(a)	1.00	-0.12
U_{14}	0.50	1.00	(a)	(a)	0.75	0.50	0.69	(a)	-0.41	(a)	(a)	-0.50	1.00	0.50
U_{15}	0.14	(a)	0.42	-0.20	0.56	0.17	-0.15	(a)	0.02	(a)	(a)	0.46	-0.12	0.50

(a) Correlation is significant at the 0.05 level (2-tailed).
* Correlation is significant at the 0.01 level (2-tailed).
Cannot be computed because at least one of the variables is constant.

$$w(u_a, u_i) = \frac{\sum_{j=1}^m (r_{a,j} - \bar{r}_a)(r_{i,j} - \bar{r}_i)}{\sqrt{\sum_{j=1}^m (r_{a,j} - \bar{r}_a)^2} \sqrt{\sum_{j=1}^m (r_{i,j} - \bar{r}_i)^2}} \quad (1)$$

(j: 아이템, $r_{a,j}$: j에 대한 U_a 의 평가치, \bar{r}_a : U_a 평가치 평균)

표 4는 u_a 의 SG1(A)을 대상으로 피어슨 상관계수를 구한 값의 일부이다. 본 값은 95% 신뢰 수준에서 도출되었으며 유의한 값을 진하게 표시하였다. 상관계수 값을 기초로 하여 SG에서의 최근접 이웃이 생성된다.

4. 선택 케이스별 추천 리스트 도출

Top-N에 의해 최근접 이웃이 선정되면 추천이 가능한 영화의 후보 리스트를 구성하는 단계로 넘어간다. 추천 후보 리스트는 최근접 이웃들이 본 영화 중 u_a 가 보지 않은 영화가 되며, 최근접 이웃의 영화별 평점(S_i)과 Top-N의 가중치($w(u_a, u_i)$)가 반영되어 예측 평가치 $P_{a,i}$ 가 도출된다^[16].

$$P_{a,i} = \frac{\sum_{u=1}^n (w(u_a, u_u) * S_{i,n})}{\sum_{u=1}^n (w(u_a, u_u))} \quad (2)$$

$P_{a,i}$ 를 반영한 u_a 의 추천 후보 리스트를 다음의 표 5에서 보인다. 사용자가 동일 연령 그룹의 추천 리스트를 보고자 하는 경우, SG1의 최근접 이웃이 본 영화별로 평점과 가중치가 모두 고려되어 최종 추천 순위가 도출된다.

평점과 가중치를 고려한 최종 $P_{a,i}$ 가 동일한 경우는 영화의 개봉일이 최근인 것에 우선순위를 부여한다.

본 논문에서는 앞서 그림 2의 ③ 선택 단계에 따라 사용자의 의도가 추천 리스트에 반영되어야 하므로 케

표 5. u_a 가 SG1(A) 선택시 추천 후보 리스트
Table 5. u_a SG1(A)'s Recommendation Nominee List.

M ID	96	10A	168	153	94	10A													
94	0.74	5	3.88	5	3.68	5	3.68	5	3.68	5	3.68	5	3.68	5	3.68	5	3.68	5	3.68
293	0.71	5	3.33	4	2.82	5	3.32	4	2.82	4	2.82	5	3.33	5	3.33	5	3.33	5	3.33
297	0.70	5	2.81	5	2.81	5	2.81	5	2.81	5	2.81	5	2.81	5	2.81	5	2.81	5	2.81
344	0.65	5	3.25	5	3.25	5	3.25	5	3.25	5	3.25	5	3.25	5	3.25	5	3.25	5	3.25
758	0.65	5	3.24	5	3.24	5	3.24	5	3.24	5	3.24	5	3.24	5	3.24	5	3.24	5	3.24
201	0.64	5	3.32	5	3.22	4	2.50	5	3.22	4	2.50	5	3.22	3	1.93	3	1.93	3	1.93
145	0.63	5	3.15	5	3.15	4	2.52	5	3.15	4	2.52	5	3.15	4	2.52	5	3.15	4	2.52
334	0.63	4	2.51	4	2.51	5	3.14	5	3.14	4	2.51	4	2.51	4	2.51	4	2.51	4	2.51
896	0.59	5	2.97	5	2.97	4	2.38	4	2.38	4	2.38	4	2.38	4	2.38	4	2.38	4	2.38
249	0.59	5	2.97	5	2.97	4	2.37	4	2.37	4	2.37	4	2.37	4	2.37	4	2.37	4	2.37
Prediction Sum(1)			22.69		21.65		21.15		18.15		18.15		23.49		23.42		23.42		23.42
Weight Sum(2)			4.86		5.20		4.59		4.81		4.81		5.20		5.20		5.20		5.20
(U1/2)			4.69		4.78		4.82		4.59		4.52		4.52		4.52		4.52		4.52

표 6. 케이스별 추천 리스트 비교표
Table 6. Recommender List by Case A, C, G.

	User 1	User 2	User 3
	M ID	M ID	M ID
A (Age)	127	64	187
	56	98	684
	134	11	64
	168	9	183
	153	127	56
C (Age+Gender)	56	57	9
	98	58	566
	127	97	121
	79	132	118
	174	134	385
G (Age+Gender+location)	56	56	28
	60	57	132
	79	58	133
	96	64	157

이스에 따라 리스트가 어떻게 달리 제시되는지 확인할 필요성이 있다. 이를 위해 목표 사용자 외에 추가적으로 2명의 추천 리스트를 도출, 표 6에서 결과값을 비교한다. 사용자는 모두 남자이며, User 1, 2는 20대, User 3은 40대이다. 동일한 20대임에도, 혹은 같은 남자임에도 선택에 따라 다양한 결과 리스트가 도출된다.

5. iPhone 선택형 추천 화면

제안한 사용자 선택형 콘텐츠 추천 시스템에 따른

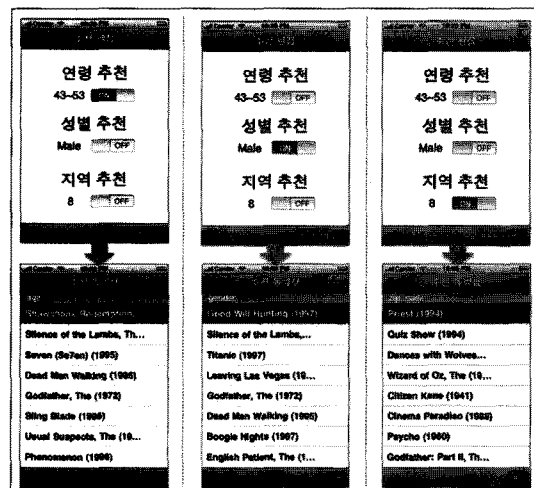


그림 4. iPhone 추천 화면 예시
Fig. 4. iPhone Recommendation View.

Prototype을 그림 4에서 제시한다. 기기는 스마트폰인 iPhone을 대상으로 하였으며 사용자 선택의 결과에 따라 추천 리스트가 달라짐을 확인할 수 있다.

V. 결 론

본 논문에서는 선호 장르를 기반으로 유사 사용자 그룹인 SG를 도출하고 사용자의 선택에 따라 추천 리스트를 유연성 있게 보여주는 협업 필터링 추천 시스템을 제시한다. 연령, 성별, 지역에 따라 사용자는 총 7개의 추천 케이스를 선택할 수 있으며 선택 결과에 따라 추천 리스트가 유동적으로 재구성된다. 제안된 시스템에 따른 추천 결과는 케이스별로 iPhone 환경에서 구현하여 Prototype을 제시한다.

본 논문이 갖는 의의는 다음과 같다. 첫째, 콘텐츠 생성 및 소비가 활발한 모바일 환경에서 사용자가 관심 있는 콘텐츠 소비 그룹을 쉽게 선택하고 변경할 수 있도록 함으로써 보다 직접적이고 관련도가 높은 추천 리스트 제공이 가능하다. 둘째, 사용자의 선택에 따라 불필요한 사용자 그룹을 제외하고 비교 그룹군을 최소화 시킴으로써 추천 시스템의 효율성을 높일 수 있다. 셋째, 웹에서의 추천 리스트를 모바일 상에서 보여주는 단순 방식을 탈피하여 모바일의 이동 환경을 고려한 '지역'을 추천 요소로 반영함으로써 향후 기기의 특징이 반영된 추천으로서의 확장성을 갖는다.

이러한 의의에도 불구하고 본 논문이 갖는 한계도 있다. 사용자의 선택권을 연령, 성별, 지역으로만 한정함으로써 직업 등 실질적으로 사용자가 관심을 가질 그룹 요소를 채택하였는지 확인하기 어렵다. 또한 데이터 세트의 한계로 지역 정보는 가공된 정보를 사용하였으므로, 지역 추천 결과의 정확성을 확신할 수 없다. 앞으로 이동성이 반영된 소비 정보를 활용한다면 지역에 따른 추천 리스트의 특징 또한 살펴볼 수 있을 것이다.

향후 연구로는 모바일 환경에서 사용자 선호도와 연관성이 높은 집단적 특징을 확인하고 소셜 네트워크를 활용한 추천 방식을 연구하고자 한다.

참 고 문 헌

- [1] The Official YouTube Korea Blog, <http://youtubekrblog.blogspot.com/2010/03/1-24.html>
- [2] 美 유튜브(YouTube), 모바일서비스 전면 확대, 구본경, KOTRA Global Window, 2008.01
- [3] J. BEN SCHAFER et al., E-Commerce Recommendation Applications, *Data Mining and Knowledge Discovery*, 5, pp.115~153, 2001.
- [4] 이경중 외, 사용자 선호도와 태그 간 상관도 분석을 통한 태그 기반 협력적 필터링 기법, *한국정보과학회 Vol.34, No.2*, pp.72~77, 2007.
- [5] 홍종규, 모바일 환경에서의 효율적인 멀티미디어 콘텐츠 추천 시스템 구조, 충남대학교 학위 논문, 2006.
- [6] Frederic P. Miller et al., Collaborative Filtering, *Alphascript Publishing*, pp.13~16, 2009.
- [7] 김재경 외, 모바일 전자상거래 환경에 적합한 개인화된 추천 시스템, *경영정보학연구 제15권 제3호*, pp.223~241, 2005.
- [8] ROBIN BURKE, Hybrid Recommender Systems: Survey and Experiments, *User Modeling and User-Adapted Interaction 12*: pp.331~370, 2002.
- [9] Paul Resnick et al., GroupLens: An Open Architecture for Collaborative Filtering of Netnews, *Proceedings of ACM 1994 Conference on Computer Supported Cooperative Work*, Chapel Hill, NC: Pages 175-186
- [10] Badrul M. Sarwar et al., Recommender Systems for Large-scale E-Commerce: Scalable Neighborhood Formation Using Clustering, *Proceedings of the Fifth International Conference on Computer and Information Technology*, 2002.
- [11] SERKAN ULUCAN, A RECOMMENDATION SYSTEM COMBINING CONTEXT-AWARENESS AND USER PROFILING IN MOBILE ENVIRONMENT, *MIDDLE EAST TECHNICAL UNIVERSITY*, pp. 20~24, 2005.12.
- [12] 오정민, 송주홍, 문남미, 모바일 환경 기반의 선택형 협업 필터링 추천 시스템, *한국멀티미디어학회 춘계학술발표대회 논문집*, 제13권, 1호, 2010.05
- [13] 이재식, 박석두, 장르별 협업 필터링을 이용한 영화추천시스템의 성능 향상, *한국지능정보시스템학회논문지 제13권 제4호*, PP.65~78, 2007.
- [14] MovieLens, <http://en.wikipedia.org/wiki/MovieLens>
- [15] MovieLens, <http://movielens.umn.edu/>
- [16] Badrul Sarwar et al., ItemBased Collaborative Filtering Recommendation Algorithms, *ACM 1581133480/01/0005*, 2001.

저 자 소 개



오 정 민(정회원)
 1999년 2월 숙명여자대학교
 경영학과 학사 졸업
 2008년 2월 서울벤처정보대학원
 대학교 디지털미디어학과
 석사 졸업
 2008년~현재 호서벤처전문
 대학원 IT응용기술학과
 박사과정

<주관심분야 : HCI, UI/UX, 모바일, 3-View, 메
 타데이터>



문 남 미(정회원)-교신저자
 1987년 2월 이화여자대학교
 대학원 컴퓨터학과
 석사 졸업
 1998년 2월 이화여자대학교
 대학원 컴퓨터학과
 박사 졸업

1999년~2003년 이화여자대학교 조교수
 2003년~2008년 서울벤처정보대학원대학교
 디지털미디어학과 교수
 2008년~현재 호서대학교 벤처전문대학원
 IT응용기술학과 교수

<주관심분야 : 디지털데이터방송비즈니스 모델,
 HCI, T-Commerce, 메타데이터>