

## 모바일 전자상거래 환경에 적합한 개인화된 추천시스템\*

김 재 경\*\*, 조 윤 호\*\*\*, 김 승 태\*\*, 김 혜 경\*\*

### A Personalized Recommender System for Mobile Commerce Applications

Jae Kyeong Kim, Yoon Ho Cho, Seung Tae Kim, Hye Kyeong Kim

In spite of the rapid growth of mobile multimedia contents market, most of the customers experience inconvenience, lengthy search processes and frustration in searching for the specific multimedia contents they want. These difficulties are attributable to the current mobile Internet service method based on inefficient sequential search. To overcome these difficulties, this paper proposes a MOBILE COntents Recommender System for Movie (MOBICORS-Movie), which is designed to reduce customers' search efforts in finding desired movies on the mobile Internet. MOBICORS-Movie consists of three agents: CF (Collaborative Filtering), CBIR (Content-Based Information Retrieval) and RF (Relevance Feedback). These agents collaborate each other to support a customer in finding a desired movie by generating personalized recommendations of movies. To verify the performance of MOBICORS-Movie, the simulation-based experiments were conducted. The results from this experiments show that MOBICORS-Movie significantly reduces the customer's search effort and can be a realistic solution for movie recommendation in the mobile Internet environment.

**Keywords :** Recommender Systems, Mobile Commerce, Collaborative Filtering, Relevance Feedback

---

\* 이 연구는 2003년도 경희대학교 지원에 의한 결과임(KHU-20031070).

\*\* 경희대학교 경영대학 e-비즈니스 전공

\*\*\* 국민대학교 e-비즈니스학부

## I. 서론

최근 들어 휴대폰 및 PDA 등의 모바일 단말기의 급속한 진화와 광범위한 보급으로 인하여 모바일 단말기를 통해 언제 어디서나 웹(world wide web)에 접속하여 다양한 콘텐츠를 이용할 수 있는 모바일 웹 서비스(mobile web service)가 빠르게 확산되고 있다.

1992년 2월 NTT DoCoMo의 i-모드를 시작으로 모바일 웹 서비스는 전 세계적으로 급증하고 있으며, 우리나라도 1999년 SK텔레콤의 '엔택', KTF의 '퍼스넷', LG텔레콤의 '이지아이' 서비스가 시작된 이래 모바일 웹 서비스 시장은 매년 급성장을 거듭하고 있다. 2003년에는 주문형비디오(VOD: Video On Demand), 멀티미디어 메시징 서비스(MMS: Multimedia Messaging Service) 등의 활성화로 모바일 웹 서비스 시장은 2조 원대를 넘어설 것으로 예상되며[한국전산원, 2003], 향후에도 사용 요금 인하, 지능형 복합 단말기 출시, 다양한 어플리케이션의 보급 확대 등으로 모바일 웹 서비스를 이용하는 고객은 점점 늘어날 것으로 전망된다. 이에 다양한 모바일 콘텐츠가 지속적으로 개발되고 있으며, 2003년 9월 무선 인터넷망의 개방과 함께 기존 이동통신사뿐 아니라 온라인 포털업체까지 모바일 콘텐츠 시장에 진입한 상태이다.

현재 모바일 웹 서비스에서 제공되는 콘텐츠는 약 2,000여 종이 넘으며, 캐릭터/벨소리, 게임, e메일, 동영상 등의 멀티미디어 콘텐츠의 이용도가 전체 이용률 중 68.7%를 차지해 가장 인기가 높다[한국인터넷정보센터, 2003]. 또한, 현재 모바일 웹 서비스를 이용하는 고객들이 향후 제공받고자 하는 서비스 또한 음악듣기, 벨소리/멜로디, 게임 등의 멀티미디어 콘텐츠가 대부분으로 보고되고 있다. 이와 같이 모바일 웹 서비스 시장은 멀티미디어 관련 콘텐츠와 함께 빠르게 성장할 것으로 보인다.

하지만 모바일 웹 서비스 시장의 외형적인 성

장에도 불구하고, 작은 LCD 화면, 입력 방법의 제약, 접속 브라우저 한계 등으로 모바일 웹 환경은 기존 유선 웹 환경에 비해 정보 검색이 어렵다[Buchanan *et al.*, 2001]. 이러한 이유로 현재 모바일 웹 서비스를 이용하는 고객들은 자신이 원하는 멀티미디어 콘텐츠를 검색하는데 많은 노력과 비용을 소요하고 있다. 현재 모바일 상에서 제공되는 영화 서비스를 받으려면, 영화 관련 콘텐츠(e.g. 최신 영화 개봉관, 베스트 비디오)를 클릭 한 모바일 고객은 먼저 텍스트 형태의 영화 목록을 보고 그 중 하나를 선택한 후, 보다 자세한 정보를 다음 화면을 통해 확인하게 된다. 해당 영화가 마음에 들면 '구매'(즉, 바로 보기)를 선택하고 그렇지 않으면, 이전 단계의 화면으로 되돌아가서 자신이 원하는 영화가 나올 때까지 동일한 과정을 반복한다. 이와 같이 단순 반복적인 시행착오를 통해 자신이 원하는 영화를 검색하도록 하는 현행 서비스 방식은 검색시간이 오래 걸리며, 이로 인해 통신 요금에 부담을 느낀 고객은 종종 해당 사이트의 접속을 종료한다. 결국, 이동통신사와 영화 제공업체는 그만큼의 이익 창출 기회를 잃게 된다.

따라서 기존 유선 웹 환경과 달리 다양한 정보를 동시에 제공하지 못하는 모바일 영화 서비스에서 고객이 보다 적은 노력과 비용으로 자신이 원하는 영화를 검색할 수 있도록 지원하는 추천시스템(recommender system)은 필수적인 요소이다. 하지만 현재까지 인터페이스가 불편한 모바일 웹 환경에서 불확실하고 애매한 속성을 가진 멀티미디어 콘텐츠를 추천하는 시스템을 연구한 사례는 찾아보기 힘들다.

본 연구에서는 모바일 고객이 보다 적은 노력과 비용으로 자신이 원하는 멀티미디어 콘텐츠를 찾을 수 있도록 지원하는 개인화된 추천시스템, MOBICORS-Movie(MOBile Contents Recommender System-Movie)를 제시한다. MOBICORS-Movie는 다양한 모바일 멀티미디어 콘텐츠 중에서 스트리밍(streaming) 방식의 영화

를 추천하는 시스템이다. MOBICORS-Movie는 대표적인 추천 기법인 협업필터링(collaborative filtering)과 적합성 피드백(relevance feedback)을 이용한 내용기반 정보검색(content based information retrieval)을 모바일 웹 환경에 맞게 결합하였다. 제시된 MOBICORS-Movie 시플레이션을 통하여 다양한 관점에서 그 성능을 분석하였다.

## II. 문헌연구

### 2.1 추천시스템의 협업필터링

추천시스템은 통계적 기법과 지식 탐사 기술(knowledge discovery technology)을 이용하여 고객 요구에 가장 적합한 상품을 추천해주는 시스템으로서, 고객들의 편의를 도모하고 교차판매(cross selling) 및 매출 증대에 초점을 맞춘 시스템이다[Sarwar, et al., 2000]. 이러한 시스템은 개인이 처리하기에 너무 많은 양의 정보가 산재해 있는 온라인 환경에서 유용하며, Amazon.com과 CDNow와 같은 전자상거래 사이트에서 없어서는 안 될 부분이다[Schafer et al., 1999].

현재까지 추천시스템을 구현하기 위한 다양한 기법들이 개발되어 왔으며, 이 중에서 협업필터링(CF: Collaborative Filtering)은 가장 성공적인 상품 추천 기법으로 알려져 있다. 협업필터링은 고객들의 상품에 대한 평가를 이용하는 정보 필터링 기법의 하나로, 일상생활에서 가족, 친구, 동료들의 경험을 통한 구전 효과를 자동화한 것이다. 즉, 해당 고객과 선호도가 유사한 고객들이 과거에 좋아했던 상품을 추천한다. 일반적으로 협업필터링 기반 추천은 크게 (1) 입력 데이터 구성, (2) 유사 선호도 집단 형성, (3) 추천 상품 결정 단계로 나누어진다[Sarwar et al., 2000].

이러한 협업필터링은 다양한 인터넷 비즈니스 분야에서 적용되고 있으나, 협업필터링이 가

지고 있는 다음과 같은 근본적인 문제인 입력데이터의 희박성, 신상품 추천 문제, 그리고 추천 알고리즘의 확장성 문제로 모바일 환경에서 영화를 추천하기에는 부족하다.

- (1) 입력 데이터의 희박성: 협업필터링 기반 추천시스템은 고객의 선호도 데이터를 많이 확보할수록 추천의 정확도가 높아진다. 그러나 모바일 비즈니스의 성장과 함께 모바일 웹 사이트에서 취급하는 콘텐츠가 많아짐에 따라 고객의 직접 평가나 구매정보 분석을 통하여 수집되는 선호도 데이터가 존재하지 않은 콘텐츠의 개수가 상대적으로 많아진다. 따라서 고객-상품 행렬은 희박 행렬(Sparse Matrix)일 수밖에 없으며, 유사집단을 탐색하는 과정에서 아주 적은 수의 선호도 데이터를 사용하므로 고객들 간의 유사도 측정시 신뢰성이 떨어지게 된다. 이러한 현상은 결국 추천결과의 정확도를 떨어뜨리게 하는 주요인으로 작용한다[Balabanovic and Shoham, 1997; Sarwar et al., 2000; Cho et al., 2004; Kim et al., 2002; Kim et al., 2003; Melville et al., 2002].
- (2) 신상품 추천 문제: 협업필터링은 상품에 대한 고객의 선호도 데이터를 기반으로 추천하므로 신상품의 경우와 같이 어느 고객도 평가하지 않은 상품은 그 상품의 선호도를 알 수 없기 때문에 누군가가 선호도를 입력하거나 구매하기 전에는 그 상품을 추천할 수 없게 된다[Balabanovic and Shoham, 1997; Melville et al., 2002]. 따라서 멀티미디어 콘텐츠 기술의 발전과 고객 취향의 급속한 변화에 따라 수시로 새로운 멀티미디어 콘텐츠가 제공되는 모바일 웹 환경에 협업필터링을 적용할 경우, 매우 치명적인 약점이 될 수 있다.
- (3) 추천 알고리즘의 확장성: 추천시스템은 하루가 다르게 늘어나는 고객과 상품을 처리할

수 있는 확장성(scalability)을 가져야 한다. 즉 추천시스템에 적용되는 알고리즘은 현재 서비스를 이용하고 있는 다수의 고객들에게 추천을 제공하는 동시에, 그들의 선호도를 실시간으로 학습하여 유사 선호도 집단을 빠르게 탐색할 수 있어야 한다[Sarwar *et al.*, 2000; Cho *et al.*, 2004; Kim *et al.*, 2002; Kim *et al.*, 2003]. 모바일 단말기를 가지고 있는 사람이라면 누구든지 언제 어디서나 접속이 가능한 유비쿼터스(ubiquitous) 컴퓨팅과 함께 적절한 장소에서 적절한 시간에 고객이 원하는 콘텐츠를 제공해 줄 수 있는 추천 능력은 모바일 비즈니스의 성공과 직결되는 문제이다. 따라서 협업필터링 기반 추천시스템에서 추천 알고리즘의 확장성 향상은 모바일 웹 서비스를 더욱 실질적이고 유용하게 만드는 필수적인 부분이 될 것이다.

이와 같은 협업필터링의 한계점들을 해결하기 위하여 많은 연구자들은 추천 대상 고객이 선호하는 상품과 유사한 특성을 가진 상품을 추천하는, 내용 기반 필터링(content based filtering)을 협업필터링과 결합하여 사용하였다[Balabanovic and Shoham, 1997; Basu and Cohen, 1998; Melville *et al.*, 2002; Smyth and Cotter, 2000]. 비록 이러한 혼합 추천(hybrid recommender) 방식은 각자의 영역에서 향상된 결과를 가져왔지만, 아직까지 영화에 대한 고객의 불확실한 선호도와 인터페이스가 불편한 모바일 환경을 동시에 고려한 연구는 없는 상황이다. 즉 영화에 대한 고객의 선호도는 일반 상품과 달리 고객의 현재 마음 상태에 따라 상대적으로 쉽게 변할 수 있으며, 다른 영화를 검색하는 과정 중에 충분히 바뀔 수 있다. 게다가 관심 영화의 상세 정보를 한 화면에 하나씩 보여주는 모바일 웹 환경에서, 고객이 진정으로 원하는 내용(즉, 줄거리)의 영화를 찾는다는 것은 쉽지 않은 일이다.

따라서 모바일 웹 서비스에 적용되는 영화 추천시스템은 고객과의 상호작용을 통해 고객이 선호하는 내용 속성을 실시간으로 학습하는 능력을 갖추어야 한다.

## 2.2 정보검색시스템의 검색 모델과 적합성 피드백

정보검색시스템(Information Retrieval System)이란 사용자가 필요로 하는 정보를 수집하여 내용을 분석한 뒤 찾기가 쉬운 형태로 조직하여, 정보에 대한 요구가 발생했을 때 해당 정보 또는 정보가 포함된 문헌을 찾아 제공하는 시스템이다[Baeza-Yates and Riberiro-Neto, 1999]. 정보검색시스템에서 고려되는 주요 문제는 어떤 문헌이 질의와 유사하며 어떤 문헌이 유사하지 않은가를 측정하는 것이며, 또한 사용자가 원하는 정보를 찾기 위한 효율적인 질의를 작성하는 일이다. 적합성 피드백(RF: Relevance Feedback) 기법은 최초로 시험적 질의(tentative query)로 검색을 수행한 후, 이전의 검색 결과에 대한 평가에 기반을 두어 다음 번 질의를 개선시켜 나가면서 이상적인 최적 질의를 추정하는 기법을 말한다[Salton, 1989]. 적합성 피드백을 통해 질의를 확장해 가는 과정은 다음 식 (1)과 같은 Standard\_Rocchio식으로 표현할 수 있다[Baeza-Yates and Riberiro-Neto, 1999].

$$q_{i+1} = \alpha q_i + \frac{\beta}{|V_r|} \sum_{v_j \in V_r} v_j - \frac{\gamma}{|V_n|} \sum_{v_j \in V_n} v_j \quad (1)$$

식 (1)에서,  $q_{i+1}$ 은 새로 확장된 피드백 질의 벡터,  $q_i$ 는 확장되기 이전 단계의 질의 벡터를 의미한다.  $V_r$ 과  $V_n$ 은 각각 초기 검색된 문서 집합 중에서 적합하다고 판단된 문서들의 집합과 부적합하다고 판단된 문서들의 집합을 의미하며,  $|V_r|$ 과  $|V_n|$ 은 각각 해당 문서 집합 내의 문서 개수이다.  $\alpha$ ,  $\beta$ ,  $\gamma$ 는 각각 이전 단계의

질의, 적합 문서 집합, 부적합 문서 집합 간의 중요도를 조절하는 상수이다. 위에서 설명한 Standard\_Rocchio 식[Rocchio, 1971]은 적합 문서나 부적합 문서의 정보를 각 문서 집합의 크기로 정규화하여 질의 확장에 적용하는 방법이다. 이외에도 로치오식의 변형으로 정규화 과정없이 조율 상수로만 중요도를 조정하는 Ide\_Regular 식[Ide, 1971]과, 질의를 확장하는데 적합 문서는 모두 사용하지만 부적합 문서에 대해서는 최상위 문서 하나만을 사용하는 Ide\_Dec\_Hi 식[Ide, 1971]이 있으나, 여러 가지 실험 결과 세 가지 식 모두 비슷한 결과를 보인다고 알려져 있다[Baeza-Yates and Riberiro-Neto, 1999].

적합성 피드백의 주요 장점은 간단하면서도 좋은 결과를 보인다는 점이다. 여기서 간단하다는 것은 수정된 용어 가중치가 검색된 문헌 집합을 통해 바로 계산되기 때문이며, 결과가 좋다는 것은 사용자의 정보 요구가 잘 반영된 질의 벡터가 많은 실험을 통해 관찰되었기 때문이다[Baeza-Yates and Riberiro-Neto, 1999]. 따라서 본 연구의 MOBICORS-Movie는 모바일 영화 서비스를 이용하는 고객들의 불편함을 최소화하면서 효과적으로 그들이 선호하는 내용 속성을 학습하기 위하여 적합성 피드백을 이용한다.

협업필터링과 달리 피드백 메커니즘을 가지고 있는 적합성 피드백은 고객이 선호하는 내용 속성을 실시간으로 학습할 수 있으며, 벡터 공간 모델을 이용하기 때문에 질의와 유사한 내용 속성을 가진 새로운 영화 또한 검색할 수 있다. 하지만 적합성 피드백은 먼저 고객이 시스템에 질의를 제공해야 하며, 이 초기 질의는 보다 많은 정보 요구를 포함할수록 최적 질의를 보다 빠르게 추정해낸다. 하지만 인터페이스가 제한적인 모바일 웹 환경과 영화에 대한 불확실한 선호도로 인해 초기 질의는 고객이 진정으로 원하는 영화와 동떨어질 수 있으며, 잘못된 초기 질의로 인해 고객은 자신이 원하는 영화를 찾기 위해 오히려 더 많은 노력을 해야 할 것이다. 따

라서 MOBICORS-Movie는 협업필터링을 이용하여 선호도가 유사한 고객들의 의견이 반영된 영화들을 추천 대상 고객에게 제시한 후, 고객에 의해 적합하다고 판단되어진 영화의 내용 속성을 초기 질의로 고객이 선호하는 내용 속성을 포함한 영화를 검색하고, 검색된 영화에 대한 고객의 평가를 이용하여 다음 번 질의를 개선해 나가면서 이상적인 최적 질의를 추정한다.

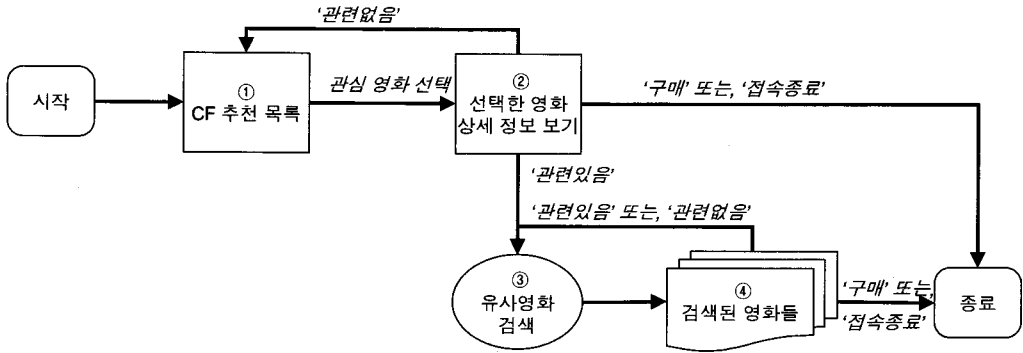
### III. MOBICORS-Movie

#### 3.1 MOBICORS-Movie의 개요

MOBICORS-Movie는 다양한 정보를 동시에 제공하지 못하는 모바일 웹 환경과 영화에 대한 고객 선호도가 불확실하다는 점을 고려하여, 고객이 보다 적은 노력과 비용으로 자신이 원하는 영화를 찾을 수 있도록 지원하는 모바일 영화 추천시스템이다. MOBICORS-Movie는 <그림 1>과 같이 모바일 고객과의 상호작용을 통해 고객이 원하는 영화를 추천한다.

먼저, MOBICORS-Movie는 모바일 영화 서비스를 이용하는 고객에게 초기 CF 추천 목록(①)을 제시한다. 이 CF 추천 목록은 추천 대상 고객이 앞으로 구매할 가능성이 큰 영화들의 제목으로 구성되며, 대상 고객의 특정 영화에 대한 구매 가능성은 대상 고객과 유사한 성향을 지닌 이웃 고객들(neighbors)의 선호도 정보를 통해 예측된다.

고객은 CF 추천 목록 중 관심 있는 영화를 선택한 후, 보다 자세한 정보(포스터, 주연, 장르, 줄거리 등)를 다음 화면을 통해 보고 자신이 원하는 영화인지 확인한다(②). 이 때 고객은 현재 보이는 영화가 적합한지(구매하고자 하는 영화와 유사하면 메뉴에서 '관련있음'을 선택) 또는, 부적합한지(구매하고자 하는 영화와 유사하지 않으면 메뉴에서 '관련없음' 선택)를 판단한다. 고객이 '관련없음'을 선택하면 MOBICORS-Movie는



<그림 1> MOBICORS-Movie의 영화 추천 과정

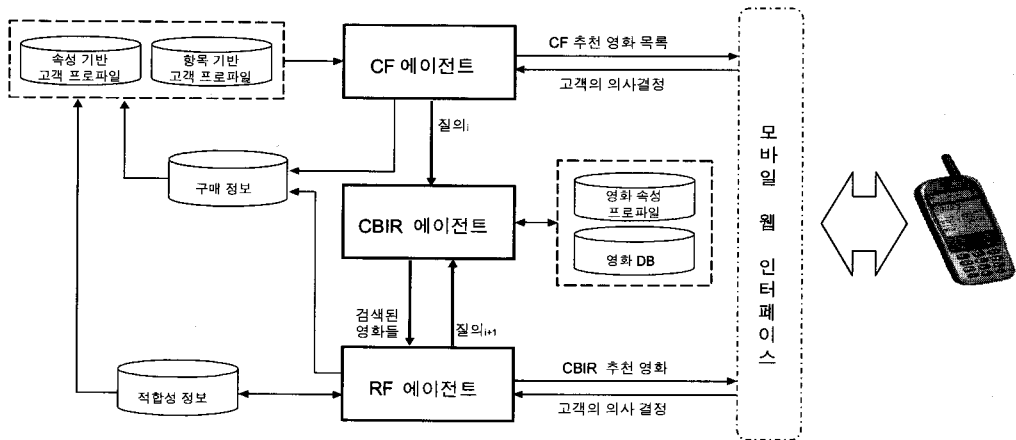
CF 추천 목록을 다시 보여주며, 고객이 '관련있음'을 선택하면 해당 영화와 유사한 내용(즉, 줄거리)을 가진 영화를 검색한다(③). 검색된 영화들(④)은 한 화면에 하나씩 상세 정보와 함께 대상 고객에게 추천되고 대상 고객은 추천 받은 영화들에 대한 적합성 판단('관련있음' 또는, '관련없음')을 MOBICORS-Movie에 제공한다. MOBICORS-Movie는 고객의 피드백 정보를 이용하여 고객이 원하는 내용 속성을 학습하여 이를 기반으로 다시 새로운 영화를 검색한다(③). 이러한 과정 중에서 고객은 언제든지 바로보기('구매')를 선택하거나 접속을 종료할 수 있다.

이렇게 MOBICORS-Movie는 고객들의 선호

도 정보를 분석하여 추천 대상 고객을 위한 초기 CF 추천 목록을 생성하고 고객과의 상호작용을 통해 고객이 선호하는 내용의 영화만을 추천함으로써, 모바일 영화 서비스를 이용하는 고객의 검색 노력을 줄여준다.

### 3.2 MOBICORS-Movie의 구조도

MOBICORS-Movie는 고객이 원하는 영화만을 추천하기 위하여, 협업필터링을 이용하여 초기 CF 추천 목록을 생성하고 내용 기반 정보검색(CBIR: Content Based Information Retrieval)으로 고객이 선호하는 내용 속성을 포함한 영



<그림 2> MOBICORS-Movie의 구조도

화를 검색하며 적합성 피드백을 통해 고객이 선호하는 내용 속성을 실시간으로 학습한다. MOBICORS-Movie에 적용된 기법과 그에 해당하는 기능을 <그림 2>에 표현한 구조도와 함께 설명하면 다음과 같다.

<그림 2>에서 보는 바와 같이 MOBICORS-Movie는 (1) CF 에이전트, (2) CBIR 에이전트, (3) RF 에이전트의 세 개의 에이전트들로 구성되어 있다. 이들 에이전트들은 서로 협력하면서 모바일 고객이 보다 적은 노력과 비용으로 자신이 원하는 영화를 찾을 수 있도록 지원한다.

(1) CF 에이전트: CF 에이전트는 추천 대상 고객과 유사한 성향을 가진 고객들의 선호도를 분석하여 초기 CF 추천 목록을 생성한다. 고객들의 선호도 정보는 영화 내용 속성(키워드)에 대한 선호도를 나타내는 속성 기반 고객 프로파일(feature-based customer profile)과 개별 영화에 대한 선호도를 나타내는 항목 기반 고객 프로파일(item-based customer profile)에 저장되어 있으며, 고객의 구매 행위와 적합성 판단에 따라 실시간으로 갱신된다. CF 에이전트는 속성 기반 고객 프로파일을 이용하여 추천 대상 고객과 유사한 성향을 가진 이웃 고객들을 탐색하고 탐색된 이웃 고객들의 개별 영화에 대한 선호도(항목 기반 고객 프로파일)를 분석하여 추천 영화를 결정한다. 이웃 고객들을 탐색할 때 사용되는 속성 기반 고객 프로파일은 신규 영화와 고객이 하루가 다르게 늘어나는 모바일 웹 환경을 고려한 것이다. 즉 시스템 내의 모든 영화에 대한 고객 선호도가 아닌 시스템 내에 미리 정의된 영화 키워드에 대한 고객 선호도를 이용함으로써, 선호도가 아직 존재하지 않는 새로운 영화로 인해 발생하는 입력 데이터의 희박성(sparsity) 문제를 해결하고 늘어나는 영화의 개수와 상관없이 이웃 고객 탐색을 가능하도록 하여 추천

알고리즘의 확장성(scalability)을 유지한다.

(2) CBIR 에이전트: CBIR 에이전트는 고객의 내용 속성 선호도를 나타내는 질의와 유사한 영화를 검색(내용 기반 정보검색)한 후, 검색된 영화들을 RF 에이전트에 전달한다. CBIR 에이전트에 입력되는 질의는 초기 CF 추천 목록 중 고객에 의해 선택된 영화의 내용 속성(질의<sub>i</sub>)이거나 RF 에이전트에 의해 대상 고객의 내용 속성 선호도가 반영된 질의(질의<sub>i+1</sub>)이다. 질의와 시스템 내의 영화들 사이의 유사도는 키워드 공간상의 코사인 척도(cosine measure)로 계산되며, 이 과정에서 질의와 유사한 속성을 가진 새로운 영화의 추천 또한 가능하므로 협업필터링의 신상품 추천 문제는 해결될 수 있다..

(3) RF 에이전트: 모바일 웹 환경에서의 화면 크기의 제약으로 인하여, RF 에이전트는 CBIR 에이전트에 의해 검색된 영화들을 한 화면에 하나씩 순차적으로 추천 대상 고객에게 보여준다. 고객은 추천받은 영화들에 대한 적합성 판단('관련있음' 또는, '관련없음')을 RF 에이전트에 제공하며, RF 에이전트는 고객의 적합성 판단을 이용하여 고객의 내용 속성 선호도가 반영되도록 질의를 갱신한다. 갱신된 질의(질의<sub>i+1</sub>)는 CBIR 에이전트에 전달되고 CBIR의 내용 기반 정보검색은 다시 수행되며, 적합성 피드백 과정에서 생성된 고객의 적합성 판단은 실시간으로 속성 기반 고객 프로파일과 항목 기반 고객 프로파일을 갱신한다. 이 과정에서 고객의 선호도를 암시적으로 추정함으로써 협업필터링의 희박성 문제는 해결된다.

## IV. MOBICORS-Movie 추천절차

### 4.1 사용 표기 및 의미

MOBICORS-Movie는 고객의 선호도를 암시적

으로 추정하여 속성 기반 고객 프로파일(feature-based customer profile)과 항목 기반 고객 프로파일(item-based customer profile)에 실시간으로 저장하고, 저장된 선호도 정보를 이용하여 추천 영화를 결정한다. 추천 영화는 CF 추천 절차와 CBIR-RF 추천 절차 통해 결정되며, 고객은 자신이 원하는 영화가 추천되면 언제든지 바로보기('구매')를 선택할 수 있다. MOBICORS-Movie의 프로파일 관리와 두 추천 절차를 설명하기 위해 사용하는 표기들과 각각의 의미는 다음과 같다.

- $i$  :  $i$ 번째 고객
- $j$  :  $j$ 번째 영화
- $k$  :  $k$ 번째 키워드
- $l$  : 선호도가 유사한 고객의 수
- $m$  : 초기 CF 추천 목록에 포함된 영화의 수
- $n$  : CBIR 에이전트가 검색한 영화의 수
- $pls()$  : 구매 가능성 점수
- $q$  : 질의
- $sim()$  : 유사도 계산척도
- $w()$  : 추천 영화에 대한 대상 고객의 선호도
- $B$  : 유사 선호도 집단
- $F$  : 속성 기반 고객 프로파일

- $L$  : 전체 키워드의 수
- $M$  : 전체 고객의 수
- $N$  : 전체 영화의 수
- $P$  : 항목 기반 고객 프로파일
- $V$  : 영화 속성 프로파일
- $V_r$  : 적합한 영화로 판단된 영화들의 집합
- $V_n$  : 부적합한 영화로 판단된 영화들의 집합
- $|V_r|$  : 집합  $V_r$ 의 영화 수
- $|V_n|$  : 집합  $V_n$ 의 영화 수

## 4.2 프로파일의 생성과 관리

### 4.2.1 영화 속성 프로파일

시스템 내에 존재하는 모든 영화의 내용 속성은 해당 영화의 내용을 대표하는 키워드로 구분되어 있으며, 영화-키워드 행렬  $V=(v_{jk})$ 로 표현된다. 즉,  $j$ 번째 영화가  $k$ 번째 키워드를 포함한다면  $v_{jk}$ ( $V$ 의  $j$ 번째 행  $k$ 번째 열의 값)는 '1'이며, 포함하지 않는다면 '0'이다. 따라서 행벡터  $V_j$ 는  $j$ 번째 영화에 존재하는 모든 키워드의 값을 '1'로 표현함으로써,  $j$ 번째 영화의 내용 속성을 나타낸다. 예를 들어  $N=9$ 이고  $L=6$ 이라면 영화 속성 프로파일은 <표 1>과 같다.

<표 1> 영화 속성 프로파일  $V=(v_{jk})$

영화 \ 키워드	Betrayal	Crime	Cult-favorite	Drama	Police	Twist-in-the-end
Godfather	1	1	0	1	1	1
Pulp Fiction	0	1	1	1	0	0
Reservoir Dogs	0	1	1	0	1	1
Fight Club	0	1	1	1	0	1
Fargo	0	1	1	0	1	0
Psycho	0	0	1	0	1	1
Goodfellas	1	1	0	1	1	0
Casablanca	1	0	0	1	1	0
Vertigo	0	0	0	1	1	1



### 4.2.2 속성 기반 고객 프로파일

속성 기반 고객 프로파일은  $L$ 개의 영화 키워드에 대한  $M$ 명 고객의 선호도 집합으로 구성되며, 이는 고객-키워드 행렬  $F = (f_{ik})$ 로 표현된다. 즉,  $f_{ik}$ ( $F$ 의  $i$ 번째 행  $k$ 번째 열의 값)는  $i$ 번째 고객이  $k$ 번째 키워드를 선호하는 정도를 나타내며, 선호도가 긍정적일 때는 양의 값을 가지며 부정적일 때는 음의 값을 가진다. 예를 들어  $M=6$ 이고  $L=6$ 이라면 속성 기반 고객 프로파일은 <표 2>와 같다.

고객  $i$ 의  $L$ 개의 키워드에 대한 선호도를 의미하는, 행벡터  $F_i$ 는 추천 영화에 대한 고객의 의사결정에 따라 다음 식 (2)와 같이 갱신된다.

$$F_i = \begin{cases} F_i + 2V_j & \text{고객 } i \text{가 영화 } j \text{를 '구매' 한 경우} \\ F_i + V_j & \text{고객 } i \text{가 영화 } j \text{에 대하여 '관련있음'을 선택한 경우} \\ F_i - V_j & \text{고객 } i \text{가 영화 } j \text{에 대하여 '관련없음'을 선택한 경우} \\ F_i & \text{고객 } i \text{가 영화 } j \text{를 아직 보지 않은 경우} \end{cases} \quad (2)$$

<표 2> 속성 기반 고객 프로파일  $F = (f_{ik})$

고객 \ 키워드	Betrayal	Crime	Cult-favorite	Drama	Police	Twist-in-the-end
Kim	5	10	0	5	-10	0
Cho	10	5	-5	5	5	10
Kang	5	15	5	10	-5	0
Lee	0	10	0	10	0	-5
Sang	-5	5	-5	5	0	15
Ryu	10	15	5	0	10	5

<표 3> 갱신된 속성 기반 고객 프로파일  $F = (f_{ik})$

고객 \ 키워드	Betrayal	Crime	Cult-Favorite	Drama	Police	Twist-in-the-end
Kim	5	13	3	6	-8	0
Cho	10	5	-5	5	5	10
Kang	5	15	5	10	-5	0
Lee	0	10	0	10	0	-5
Sang	-5	5	-5	5	0	15
Ryu	10	15	5	0	10	5

### 4.2.3 항목 기반 고객 프로필

항목 기반 고객 프로필은  $N$ 개의 영화에 대한  $M$ 명의 고객의 선호도 집합으로 구성되며, 이는 고객-영화 행렬  $P = (p_{ij})$ 로 표현된다. 즉,  $p_{ij}$  ( $P$ 의  $i$ 번째 행  $j$ 번째 열의 값)는  $i$ 번째 고객이  $j$ 번째 영화를 선호하는 정도를 나타내며, 선호도가 긍정적일 때는 양의 값을 가지며 부정적일 때는 음의 값을 가진다. 예를 들어  $M=6$ 이고  $N=9$ 라면 항목 기반 고객 프로필은 <표 4>와 같다.

고객  $i$ 의  $N$ 개의 영화에 대한 선호도를 의미하는, 행벡터  $P_i$ 는 추천 영화에 대한 고객의 의사 결정에 따라 다음 식 (3)와 같이 갱신된다.

식 (3)에서,  $i$ 는 1에서  $M$ 까지  $j$ 는 1에서  $N$ 까지며,  $M$ 과  $N$ 은 각각 전체 고객과 영화의 수이다.  $w(i,j)$ 는 고객  $i$ 의 추천 영화  $j$ 에 대한 선호도를 의미하며, 고객  $i$ 의 행위에 따라 항목 기반 고객 프로필  $P=(p_{ij})$ 의 해당 원소 값에 더해지거나 감해진다. 항목 기반 고객 프로필의 갱신 또한 이전의 연구들[Cho *et al.*, 2004; Lawrence *et al.*, 2001]과 같이 '구매'를 선택한 영화에 대해서 두 배의 가중치를 적용하였다. 예를 들어 고객 Kim이 <표 1>의 Pulp Fiction에 대하여 '관련있음'을 선택하고 Fargo를 '구매'하였다면, <표 4>의 항목 기반 고객 프로필은 <표 5>와 같이 갱신된다.

$$P_i = P_i + W(i, j) \quad w(i, j) = \begin{cases} +2 & \text{고객 } i \text{가 영화 } j \text{를 '구매' 한 경우} \\ +1 & \text{고객 } i \text{가 영화 } j \text{에 대하여 '관련있음'을 선택한 경우} \\ -1 & \text{고객 } i \text{가 영화 } j \text{에 대하여 '관련없음'을 선택한 경우} \\ 0 & \text{고객 } i \text{가 영화 } j \text{를 아직 보지 않은 경우} \end{cases} \quad (3)$$

<표 4> 항목 기반 고객 프로필  $P = (P_{ij})$

영화 고객	Godfather	Pulp Fiction	Reservoir Dogs	Fight Club	Fargo	Psycho	Goodfellas	Casablanca	Vertigo
Kim	5	15	0	15	0	-10	5	0	-5
Cho	15	-5	5	5	-5	0	10	5	5
Kang	15	10	15	15	15	0	5	15	0
Lee	10	5	5	10	5	-5	5	0	0
Sang	5	0	10	15	-5	5	0	-5	15
Ryu	10	-5	5	0	0	-5	5	-5	-10

<표 5> 갱신된 항목 기반 고객 프로필  $P = (P_{ij})$

영화 고객	Godfather	Pulp Fiction	Reservoir Dogs	Fight Club	Fargo	Psycho	Goodfellas	Casablanca	Vertigo
Kim	5	16	0	15	2	-10	5	0	-5
Cho	15	-5	5	5	-5	0	10	5	5
Kang	15	10	15	15	15	0	5	15	0
Lee	10	5	5	10	5	-5	5	0	0
Sang	5	0	10	15	-5	5	0	-5	15
Ryu	10	-5	5	0	0	-5	5	-5	-10

### 4.3 CF 추천 절차

CF 에이전트에 의해 수행되는 CF 추천 절차는 이웃 고객들의 선호도 정보를 이용하여 추천 대상고객이 구매할 가능성이 높은 영화들로 초기 CF 추천 목록을 구성하며, CF 추천 목록에 포함되는 영화들은 유사 선호도 집단(neighborhood) 형성 단계와 추천 영화 결정 단계를 통해 결정된다.

#### 4.3.1 유사 선호도 집단 형성

CF 추천 목록을 생성하기 위하여, CF 에이전트가 수행하는 첫 단계는 영화 내용에 대한 선호도가 유사한 이웃 고객들로 구성된 유사 선호도 집단을 형성하는 것이다.

추천 대상 고객과 다른 고객들과의 유사 선호도는 속성 기반 고객 프로파일의 분석을 통해 이루어진다. 즉, 두 고객  $a$ 와  $b$ 사이의 유사 선호도  $sim(a,b)$ 는 속성 기반 고객 프로파일  $F$ 의 두 행벡터의 코사인 각도를 이용하여 다음 식 (4)와 같이 계산할 수 있다.

$$sim(a, b) = \cos(a, b) = \frac{a \cdot b}{|a| \times |b|} = \frac{\sum_{k=1}^L f_{ak} \times f_{bk}}{\sqrt{\sum_{k=1}^L (f_{ak})^2} \times \sqrt{\sum_{k=1}^L (f_{bk})^2}} \quad (4)$$

식 (4)에서  $k$ 는 키워드,  $L$ 은 전체 키워드의 수이며,  $f_{ak}$ 와  $f_{bk}$ 는 각각  $k$ 번째 키워드에 대한 고객  $a$ 와  $b$ 의 선호도이다. 두 고객의 유사 선호도가 가장 높을 때는 '1', 가장 낮을 때는 '-1'이며, CF 에이전트는 계산된 유사 선호도 값이 큰 상위  $l$ 명의 고객으로 유사 선호도 집단을 형성한다. 예를 들어 <표 3>의 속성 기반 고객 프로파일에서 추천 대상 고객이 Kim이라고 할 때, Kim과 다른 고객과의 유사 선호도는  $sim(Kim, Cho) =$

$0.30, sim(Kim, Kang) = 0.96, sim(Kim, Lee) = 0.73, sim(Kim, Sang) = 0.18, sim(Kim, Ryu) = 0.47$ 이며,  $l = 2$ 라면 Kim의 유사 선호도 집단은 영화 내용에 대한 선호도가 가장 유사한 Kang과 Lee로 구성된다.

#### 4.3.2 추천 영화 결정

CF 에이전트가 수행하는 다음 단계는 항목 기반 고객 프로파일에 기록된 이웃 고객들의 영화에 대한 선호도를 분석하여 초기 CF 추천 목록을 생성하는 것이다.

CF 추천 목록은 구매 가능성이 가장 높은 영화들로 구성되며, 추천 대상 고객  $c$ 가 영화  $j$ 를 구매할 가능성은 다음 식 (5)와 같은 구매 가능성 점수(purchase likeliness score)를 통해 계산된다.

$$pls(c, j) = \frac{\sum_{i \in B} (p_{ij} - \bar{p}_i) \cdot sim(c, i)}{\sum_{i \in B} sim(c, i)} \quad (5)$$

식 (5)에서  $p_{ij}$ 는 항목 기반 고객 프로파일  $P$ 에 기록된  $i$ 번째 고객의  $j$ 번째 영화에 대한 선호도,  $\bar{p}_i$ 는  $i$ 번째 고객의 모든 영화에 대한 평균 선호도이며,  $sim(c,i)$ 는 대상 고객  $c$ 와 이웃 고객  $i$ 와의 유사 선호도이다. 구매 가능성 점수는 이웃 고객들의 선호도가 긍정적인 영화에 대해서는 양의 값을 가지며, 부정적인 영화에 대해서는 음의 값을 갖는다. CF 에이전트는 추천 대상 고객이 이전에 구매하지 않은 영화들 중에서, 구매 가능성 점수  $pls(c,j)$ 가 높은  $m$ 개의 영화들을 순위화하여 초기 CF 추천 목록을 구성한다. 여기서 만약 선호도가 아직 기록되지 않은 신규 고객이 접속한 경우, 초기 추천 목록은 다운로드 횟수가 가장 많은 베스트셀러(bestseller)로 구성된다. 예를 들어 추천 대상 고객이 Kim이고 고객들의 영화에 대한 선호도 정보가 <표 5>의 항목 기반 고객 프로파일에 기록되어 있다면,

CF 에이전트는 Kim의 이웃 고객들(Kang, Lee)의 영화에 대한 선호도를 분석하여 Kim이 구매할 가능성이 높은 영화를 추천 영화로 결정한다. 단 여기서 Kim이 이미 구매한 Fago는 제외된다. Kim의 특정 영화에 대한 구매 가능성 점수는  $pls(Kim, Godfather) = 5.48$ ,  $pls(Kim, Pulp Fiction) = 0.48$ ,  $pls(Kim, Reservoir Dogs) = 3.33$ ,  $pls(Kim, Fight Club) = 5.48$ ,  $pls(Kim, Psycho) = -9.52$ ,  $pls(Kim, Goodfellas) = -2.37$ ,  $pls(Kim, Casablanca) = 1.17$ ,  $pls(Kim, Vertigo) = -7.37$ 이며,  $m = 3$ 이라면 CF 추천 목록은 구매 가능성 점수가 높은 Godfather, Fight Club, Reservoir Dogs 순으로 구성된다.

#### 4.4 CBIR-RF 추천 절차

CBIR-RF 추천 절차는 반복적인 탐색 작업을 통해 추천 대상 고객이 원하는 내용을 포함한 영화를 검색한다. 반복적인 탐색작업은 유사한 내용 속성을 가진 영화를 검색하는 CBIR 에이전트와 고객의 내용 속성 선호도를 학습하는 RF 에이전트에 의해 수행된다.

##### 4.4.1 CBIR 에이전트의 유사 영화 검색

CBIR 에이전트는 입력되는 질의와 시스템 내의 영화를 비교하여 대상 고객이 선호하는 내용 속성을 가진 영화를 검색한다. 질의와 영화는  $L$  차원의 키워드 벡터로 표현되며, 고객의 내용 속성 선호도를 나타내는 질의  $q$ 와 시스템 내의 영화  $j$ 의 유사도  $sim(q, j)$ 는 두 벡터의 코사인 각도를 이용하여 다음 식 (6)과 같이 계산할 수 있다.

$$sim(q, j) = \cos(q, j) = \frac{q \cdot j}{|q| \times |j|} = \frac{\sum_{k=1}^L q_k \times v_{jk}}{\sqrt{\sum_{k=1}^L (q_k)^2} \times \sqrt{\sum_{k=1}^L (v_k)^2}} \quad (6)$$

여기서  $k$ 는 키워드,  $L$ 은 전체 키워드의 수이며,  $q_k$ 와  $v_{jk}$ 는 각각  $k$ 번째 키워드에 대한 질의  $q$ 와 영화  $j$ 의 가중치이다. 영화 속성 프로파일  $V=(v_{jk})$ 는  $j$ 번째 영화가  $k$ 번째 키워드를 포함하면 '1', 포함하지 않으면 '0'의 값을 가진다. CBIR 에이전트에 입력되는 질의와 시스템 내의 영화의 유사도가 가장 높을 때는 '1', 가장 낮을 때는 '0'이며, 질의  $q$ 는 CF 추천 목록에서 고객이 적합 영화('관련 있음')로 선택한 영화의 내용 속성이거나 RF 에이전트에 의해 고객의 내용 속성 선호도가 반영된 갱신된 질의이다.

CBIR 에이전트는 유사도를 계산한 후, 내용 속성 유사도 값이 큰 상위  $n$ 개의 영화를 검색하여 RF 에이전트에 전달한다. 여기서 추천 대상 고객이 이전에 구매한 영화와 CF 추천 목록에 포함된 영화는 유사도 계산에서 제외된다. 예를 들어 시스템 내의 모든 영화가 <표 1>과 같이 존재하고 추천 대상 고객 Kim이 CF 추천 목록(Godfather, Fight Club, Reservoir Dogs) 중 Godfather를 적합 영화('관련 있음')로 선택하였다면, CBIR 에이전트는 Godfather의 내용 속성을 초기 질의로 하여 시스템 내의 다른 영화와의 유사도를 계산한다. 단 여기서 Kim이 이미 구매한 영화 Fago와 CF 추천 목록에 포함되었던 Fight Club과 Reservoir Dogs는 제외된다.

계산된 유사도는  $sim(\text{질의}i, \text{Pulp Fiction}) = 0.52$ ,  $sim(\text{질의}i, \text{Psycho}) = 0.52$ ,  $sim(\text{질의}i, \text{Goodfellas}) = 0.89$ ,  $sim(\text{질의}i, \text{Casablanca}) = 0.77$ ,  $sim(\text{질의}i, \text{Vertigo}) = 0.77$ 이며,  $n = 3$ 이라면 영화의 내용 속성이 가장 유사한 Goodfellas, Casablanca, Vertigo는 RF에이전트에 전달된다.

##### 4.4.2 RF 에이전트의 질의 갱신

RF 에이전트는 CBIR 에이전트에 의해 검색된 영화들을 한 화면에 하나씩 상세 정보와 함께 추천 대상 고객에게 보여준 후, 고객의 적합성 판단('관련 있음' 또는, '관련 없음')을 이용하여

<표 6> 질의, 영화 벡터 행렬

영화 \ 키워드	Betrayal	Crime	Cult-favorite	Drama	Police	Twist-in-the-end
Godfather( $q_i$ )	1	1	0	1	1	1
Goodfellas $\in V_r$	1	1	0	1	1	0
Casablanca $\in V_r$	1	0	0	1	1	0
Vertigo $\in V_n$	0	0	0	1	1	1
갱신된 질의( $q_{i+1}$ )	2	1.5	0	1	1	0

여 질의를 갱신한다. 고객에 의해 적합하다고 판단(‘관련있음’)된 영화에 포함된 키워드 가중치는 더하고 부적합하다고 판단(‘관련없음’)된 영화에 포함된 키워드 가중치는 감하는 과정을 반복하면서, 질의와 구매 영화의 유사도가 증가하도록 질의를 개선시켜 나간다.

적합성 피드백을 통해 질의를 개선시키는 과정은 식 (7)과 같은 Standard\_Rocchio 식을 따른다.

$$q_{i+1} = q_i + \frac{1}{|V_r|} \sum_{v_j \in V_r} v_j - \frac{1}{|V_n|} \sum_{v_j \in V_n} v_j \quad (7)$$

여기서,  $q_{i+1}$ 은 갱신된 질의 벡터,  $q_i$ 는 갱신되기 전의 질의 벡터를 의미한다.  $V_r$ 과  $V_n$ 은 각각 검색된 영화 집합 중에서 적합하다고 판단된 영화들의 집합과 부적합하다고 판단된 영화들의 집합,  $|V_r|$ 과  $|V_n|$ 은 각각 해당 영화 집합 내의 영화 수이다.

일반적으로 초기 CF 추천 목록에서부터 시작된 이전 단계 질의 또한 의미 있는 정보를 포함하고 있으므로  $\alpha = 1$ 로 고정하였으며, 긍정적 피드백(‘관련있음’)과 부정적 피드백(‘관련없음’)에 대한 조절 상수는 각각  $\beta = 1$ ,  $\gamma = 1$ 로 고정하였다. RF 에이전트를 통해 추천 대상 고객의 내용 속성 선호도가 반영된 새로운 질의는 다시 CBIR 에이전트로 전달되며, 이러한 반복과정을 통해 갱신되는 질의는 점점 해당 고객이 진정으로 원하는 내용 속성을 포함하는 영화와 유사해진다. 만약 추천 대상 고객 Kim이 RF 에이전트

로부터 추천 받은 Goodfellas와 Casablanca는 적합영화(‘관련있음’)로 선택하고 Vertigo는 부적합 영화(‘관련없음’)로 선택하였다면, 초기 질의( $q_i$ )의 키워드 가중치는 <표 6>의 갱신된 질의( $q_{i+1}$ )와 같이 갱신된다.

<표 6>의 갱신된 질의( $q_{i+1}$ )에서 적합 영화에 보다 많이 포함된 Betrayal, Crime, Drama, Police의 가중치는 증가하고, 부적합 영화에만 포함된 Twist-in-the-end의 가중치는 감소하였다. 이렇게 Kim의 내용 속성 선호도가 반영된 갱신된 질의( $q_{i+1}$ )는 CBIR 에이전트로 전달되고 CBIR 에이전트의 유사 영화 검색은 다시 시작된다.

#### 4.5 예제: Kim을 위한 영화 추천

MOBICORS-Movie는 모바일 고객의 추천 영화에 대한 의사결정을 바탕으로 속성 기반 고객 프로파일과 항목 기반 고객 프로파일에 저장된 고객 선호도 정보를 실시간으로 갱신하고, CF 추천 절차와 CBIR-RF 추천 절차를 통해 고객이 원하는 영화만을 추천함으로써 모바일 고객의 영화 검색 노력을 줄여준다.

시스템 내의 모든 영화가 <표 1>과 같이 존재하고, 추천 대상 고객 Kim의 선호도 정보가 <표 2>의 속성 기반 고객 프로파일과 <표 4>의 항목 기반 고객 프로파일에 저장되어 있다면, Kim을 위한 MOBICORS-Movie의 영화 추천은 다음과 같다.

먼저 Kim이 이전의 방문에서 Pulp Fiction에 대해 '관련있음'을 선택하고 Fargo를 '구매'하였다면, <표 2>의 속성 기반 고객 프로파일과 <표 4>의 항목 기반 고객 프로파일은 각각 <표 3>과 <표 5>와 같이 갱신된다. 그리고 Kim이 모바일 영화 서비스를 이용하기 위해 다시 방문하였다면, CF 에이전트는 <표 3>의 속성 기반 고객 프로파일을 분석하여 Kim과 영화 내용에 대한 선호도가 가장 유사한 Kang과 Lee로 유사 선호도 집단( $l=2$ )을 형성하고, <표 5>의 항목 기반 고객 프로파일에 기록된 Kang과 Lee의 영화에 대한 선호도를 분석하여 구매 가능성 점수가 높은 Godfather, Fight Club, Reservoir Dogs로 초기 CF 추천 목록( $m=3$ )을 구성한다. 여기서 Kim이 이미 구매한 영화 Fargo는 구매 가능성 점수를 계산할 때 제외되며, 초기 CF 추천 목록의 영화들은 구매 가능성 점수가 높은 영화부터 순위화되어 있다.

CF 에이전트는 초기 CF 추천 목록을 Kim에게 제시하고, Kim은 CF 추천 목록 중 자신이 원하는 영화가 있으면 '구매'를 선택하고 그렇지 않으면 구매하고자 하는 영화와 가장 유사한 영화를 적합 영화('관련있음')로 선택한다. 여기서 만약 Kim이 Godfather를 적합 영화로 선택하였다면, CBIR 에이전트는 Godfather의 내용 속성을 초기 질의로 하여 시스템 내의 다른 영화와의 유사도를 계산한다. 단 Kim이 이미 구매한 영화 Fargo와 CF 추천 목록에 포함되었던 Fight Club과 Reservoir Dogs는 CBIR 에이전트의 유사 영화 검색에서 제외된다.

CBIR 에이전트는 Godfather와 내용 속성이 가장 유사한 Goodfellas, Casablanca, Vertigo를 검색( $n=3$ )하여 RF 에이전트에 전달하고, RF 에이전트는 검색된 영화들을 한 화면에 하나씩 상세 정보와 함께 Kim에게 추천한다. 그러면 Kim은 추천 영화를 구매하거나 또는 추천 영화에 대한 적합성 판단을 시스템에 제공한다. 여기서 만약 Kim이 Goodfellas와 Casablanca를 적합

영화('관련있음')로 선택하고 Vertigo를 부적합 영화('관련없음')로 선택하였다면, Kim의 내용 속성 선호도를 나타내는 질의는 <표 6>의 갱신된 질의( $q_{i+1}$ )와 같이 갱신된다. 즉, 적합 영화에 많이 포함된 Betrayal, Crime, Drama, Police의 가중치는 증가하고, 부적합 영화에만 포함된 Twist-in-the-end의 가중치는 감소하였다.

이렇게 Kim의 내용 속성 선호도가 반영된 갱신된 질의는 다시 CBIR 에이전트로 전달되며, CBIR 에이전트의 유사 영화 검색과 RF 에이전트의 질의 갱신 과정은 Kim이 '구매'를 선택하거나 접속을 종료할 때까지 반복된다.

## V. 성능 평가

본 장에서는 MOBICORS-Movie의 성능을 다양한 관점에서 분석하기 위하여 시뮬레이션을 실시하였다. 실험에 사용된 영화와 키워드 정보(키워드 유/무)는 IMDb([www.imdb.com](http://www.imdb.com))<sup>1)</sup>에서 제공하는 영화 정보를 이용하였다. 영화는 고객들로부터 좋은 평가를 받은 상위 250개가 사용되었으며, 영화의 내용을 대표하는 키워드는 IMDb에서 정의한 키워드 중 출현 빈도가 높은 상위 100개가 사용되었다. MOBICORS-Movie의 성능을 평가하기 위하여 1000명의 고객을 설정하였으며, 고객들의 초기 선호도 정보는 고객당 다섯 개까지의 영화를 임의로(random) 구매하도록 하고, 구매된 영화 정보를 이용하여 속성 기반 고객 프로파일(feature-based customer profile)과 항목 기반 고객 프로파일(item-based customer profile)을 생성하였다.

MOBICORS-Movie의 성능 평가는 vps(view per success)를 통해 정량적으로 측정되었다. vps란 고객이 영화를 구매하기까지 본 영화의 개수이며, 이는 곧 고객이 자신이 원하는 영화를 검색

1) IMDb(Internet Movie Database): 최대 규모의 영화 데이터베이스를 보유한 미국의 영화 전문 인터넷 사이트.

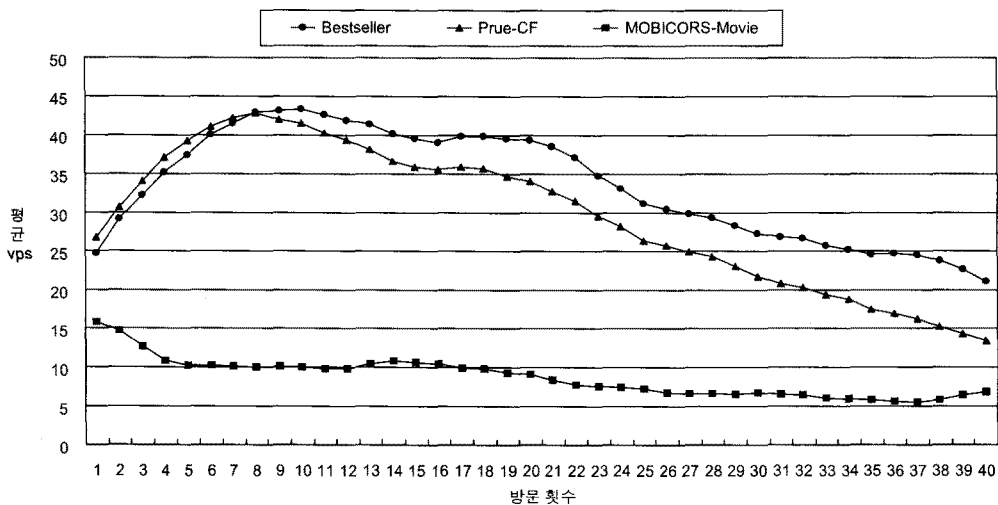
하기까지 소요한 노력의 양을 나타낸다. 예를 들어, vps가 15라는 의미는 고객이 MOBICORS-Movie가 추천한 14개의 영화를 검색한 후, 15번째 만에 원하는 영화를 구매하였다는 것을 의미한다. 따라서 낮은 vps의 결과는 MOBICORS-Movie가 효과적인 추천 절차를 가지고 고객이 원하는 영화를 보다 빨리 검색했다는 것을 의미한다.

한편, 시뮬레이션에서 시스템이 인식하는 키워드와 고객이 인식하는 키워드가 완전히 동일하다면 시스템이 고객의 선호도를 100% 정확하게 알고 있다는 의미이므로 VPS는 항상 1이 된다. 그러므로 시뮬레이션 실험에서는 고객과 시스템이 인식하고 있는 키워드를 일정부분은 같게 하고 일정부분은 다르게 하여 vps가 항상 1이 되지 않게 하면서 보다 현실적인 성능평가가 가능하게 하였다. 즉, 고객과 시스템이 다르게 인식하는 키워드의 개수는 각각 25개(전체 키워드의 25%)를 설정하였으며, 동일하게 인식하는 키워드는 50개(전체 키워드의 50%)로 하였다. 여기서 고객이 인식하는 75개의 키워드는 고객의 의사결정(적합성 판단 또는, 구매)에 사용되었으며, 시스템이 인식하는 75개(전체 키워드의

75%)의 키워드는 CF 에이전트의 유사 고객 집단 탐색, CBIR 에이전트의 유사 영화 검색, RF 에이전트의 질의 갱신에 사용되었다.

시뮬레이션을 위한 프로그램은 Visual Basic 6.0과 ADO 컴포넌트를 사용하여 구현하였으며, 250개의 영화와 100개의 키워드 정보는 MS-Access에 저장되어 관리되었다. 실험은 3.1GHz 인텔 펜티엄 IV 프로세서와 1GHz램이 장착된 Windows 2000 Server에서 수행되었다.

초기 추천 목록을 베스트셀러(bestseller) 추천, 순수 협업필터링(pure-CF), MOBICORS-Movie로 바꾸어가며 방문 횟수(총 40회)에 따른 전체 고객의 평균 vps의 변화를 관찰하였다. <그림 3>에서 보는 바와 같이, 순수 협업필터링은 베스트셀러 추천에 비해 나은 성능을 나타내며 MOBICORS-Movie는 베스트셀러 추천과 순수 협업필터링에 비해 초기 방문부터 상당히 나은 성능을 나타낸다. 초기에는(8회 방문까지) 베스트셀러 추천이 순수 협업필터링에 비해 약간 나은 성능을 보인다. 이는 고객들의 구매 기록이 충분히 축적되지 않은 상태에서 순수 협업필터링의 유사 선호도 집단이 형성되었기 때문이다. 또한, 우리의 실험 설정에서 고객들의 초기 선



<그림 3> MOBICORS-Movie와 기존 추천 방법의 성능 비교

호도 정보로 고객 당 다섯 개까지의 영화를 임의 구매하도록 하였기 때문에 방문 초기에는 고객 선호도 정보가 축적되는 비율보다 상대적으로 새로운 음악을 구매하는 비율의 증가가 더 높아서, CF나 Best Seller의 경우 평균 vps가 증가(성능 감소)하게 된다. 하지만 9회 방문 이후부터는 축적된 구매 기록으로 추천 대상과 유사한 성향을 가진 고객들로 유사 선호도 집단이 구성되면서, 순수 협업필터링과 베스트셀러 추천의 평균 vps는 계속 감소하기 시작하며, 순수 협업필터링 추천은 전체 고객의 구매 기록을 이용하는 베스트셀러 추천에 비해 나은 성능을 보이기 시작한다.

반면 MOBICORS-Movie의 평균 vps는 1회 방문부터 상당히 낮으며, 베스트셀러 추천과 순수 협업필터링과 달리 방문 횟수가 늘어날수록 점점 낮아진다. 이는 적합성피드백을 통하여 초기 입력데이터 희박성으로 인하여 발생하는 추천 성능저하의 문제를 해결한 것으로 분석된다. 즉, 현재 고객의 적합성 판단 정보를 학습하여 CBIR 에이전트의 질의를 갱신하고, 해당 질의와 유사한 속성을 가진 영화를 검색하여 추천하게 된다. 이 과정에서 데이터 희박성으로 인하여 발생할 수 있는 신뢰성이 떨어지는 초기 CF 추천 목록을 보완하게 된다.

한편, 방문 횟수가 늘어나면서, MOBICORS-Movie의 유사 선호도 집단이 보다 유사한 성향의 고객들로 정교하게 구성되고 이로 인해 초기 CF 추천 목록은 고객이 원하는 영화를 포함하거나 또는 고객이 원하는 영화와 유사한 내용의 영화들로 구성된 것으로 분석된다. 즉 MOBICORS-Movie는 고객의 선호도를 암시적으로 추정하고(implicit rating), 시스템 내에 미리 정의된 키워드에 대한 고객 선호도(속성 기반 고객 프로파일)를 이용하여 이웃 고객들을 탐색하기 때문에, 충분한 양의 고객 선호도 정보를 분석하여 유사 선호도 집단을 구성한다. 따라서 인터페이스가 제한적인 모바일 환경에서 MO-

BICORS-Movie은 베스트셀러 추천과 순수 협업필터링에 비해 보다 정확한 초기 추천 목록을 고객에게 제시함으로써 고객의 검색 노력을 줄여준다.

## VI. 결 론

본 연구에서는 인터페이스가 불편한 모바일 웹 환경과 불확실한 고객 선호도를 고려하여, 모바일 웹 서비스를 이용하는 고객이 보다 적은 노력과 비용으로 자신이 원하는 콘텐츠를 찾을 수 있도록 지원하는 개인화된 추천시스템 MOBICORS-Movie를 제시했다. MOBICORS-Movie는 스트리밍(streaming) 방식의 영화를 추천하는 시스템으로써, 대표적인 추천 기법인 협업필터링의 한계점들(입력 데이터의 희박성, 신상품 추천 문제, 추천 알고리즘의 확장성)을 보완하기 위하여, 적합성 피드백을 이용한 내용 기반 정보검색을 모바일 웹 환경에 맞게 응용하여 결합하였다.

MOBICORS-Movie는 다음과 같은 특징을 가지고 고객이 원하는 영화를 추천한다; (1) MOBICORS-Movie는 추천 대상 고객과 유사한 성향을 지닌 이웃 고객들의 선호도 정보를 분석하여 초기 CF 추천 목록을 생성한다, (2) MOBICORS-Movie는 고객이 선호하는 내용 속성과 유사한 영화를 검색하여 추천한다, (3) MOBICORS-Movie는 추천 영화에 대한 고객의 적합성 판단을 이용하여 고객의 내용 속성 선호도를 실시간으로 학습한다.

또한 MOBICORS-Movie의 성능을 분석하기 위하여 실제 모바일 고객의 의사결정 과정과 정황을 모방한 시뮬레이션을 수행하였다. 실험한 결과, MOBICORS-Movie는 모바일 환경에서 베스트셀러 추천과 순수 협업필터링에 비해 상당히 나은 성능을 보이며, 다양한 경우에 우수한 효과성을 보여주고 있다. 따라서 MOBICORS-Movie는 향후 모바일 서비스에서 응용할 수 있는 멀티미디어 추천에 사용될 수 있는 기반 시스템이



될 것으로 기대된다.

본 연구의 한계점은 다음과 같으며 이는 향후 연구에서 개선되어야 할 점들이다.

- (1) MOBICORS-Movie는 모바일 영화 서비스를 이용하는 고객의 검색 노력을 줄이기 위해 고안한 추천시스템이다. 하지만 모바일 상거래의 성장과 함께 주문형 비디오(VOD: Video On Demand), 주문형 오디오(MOD: Music On Demand), 멀티미디어 메시징 서비스(MMS: Multimedia Messaging Service)와 같은 새로운 멀티미디어 콘텐츠가 활성화되고 있다. 따라서 향후의 연구에서는 보다 다양한 콘텐츠를 대상으로 한 추천시스템을 개발할 필요가 있다.
- (2) MOBICORS-Movie 내의 모든 영화들은 영화의 내용을 대표하는 키워드의 집합으로 표현된다. 하지만 영화는 영상과 음향으로 구성되어 있기 때문에, 키워드만으로 영화의 모든 것을 표현하는데 한계가 있다. 따라서 향후의 연구에서는 영상과 음향 자체의 속성을 이용하여 검색을 수행하는, 내용 기반의 멀티미디어 검색(CBMR: Content Based Multimedia Retrieval)을 추천시스템에 활용할 필요가 있다.
- (3) MOBICORS-Movie의 영화 속성 프로파일

과 속성 기반 고객 프로파일은 출현 빈도가 높은 상위 100개의 키워드로 정의되었다. 하지만 키워드들 간의 상관관계를 고려하지 않았기 때문에 영향력이 큰 몇 개의 키워드로 시스템 내의 영화와 고객의 내용 속성 선호도가 구분될 수도 있다. 따라서 향후의 연구에서는 신뢰성(reliability)과 타당성(validity)을 갖춘 실험을 위해 키워드간의 상관관계를 사전에 분석해야 할 것이다.

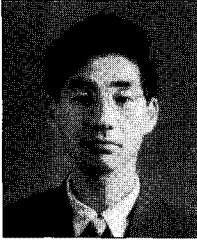
- (4) MOBICORS-Movie는 추천 영화에 대한 고객의 적합성 판단을 이용하여 고객이 선호하는 내용 속성을 실시간으로 학습한다. 하지만 시간의 변화에 따른 고객 선호도 변화를 고려하지 않았다. 즉 일 년 전에 구매한 영화나 지금 구매한 영화나 고객 프로파일을 갱신하는 가중치는 동일하다. 따라서 향후의 연구에서는 시간의 관점을 가지고 고객의 선호도를 학습하는 방법을 개발할 필요가 있다.
- (5) MOBICORS-Movie는 모바일 고객의 의사결정 과정과 정확도를 모방한 시뮬레이션을 수행하였다. 하지만 실제 모바일 고객들은 보다 다양한 상황에서 의사결정을 내릴 것이다. 따라서 향후의 연구에서는 실제 고객을 대상으로 한 실험을 통해 MOBICORS-Movie의 효과성을 검증할 필요가 있다.

## 〈참 고 문 헌〉

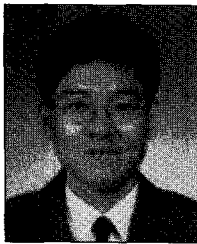
- [1] 2003 한국인터넷 백서, 한국전산원, 2003.
- [2] 무선인터넷 이용 현황 및 실태 조사, 한국인터넷정보센터, 2003.
- [3] Baeza-Yates, R. and Riberiro-Neto, B., *Modern Information Retrieval*, ACM Press, New York, NY, 1999.
- [4] Balabanovic, M. and Shoham Y., "Fab: Content-based Collaborative Recommendation," *Communication of the ACM*, Vol. 40, No. 3, 1997, pp. 66-72.
- [5] Basu, C. and Cohen, W., "Recommendation as Classification: Using Social and Content-Based Informations in Recommendations," *Proceeding of the 15th National Conference on Artificial Intelligence*, 1998, pp. 714-720.
- [6] Buchanan G., Jones M., Thimbleby H., Farrant S., and Pazzani M., "Improving

- Mobile Internet Usability," *Proceedings of the Web 2001 Conference*, Hong Kong, ACM Press, 2001, pp. 673-680.
- [7] Cho, Y.H. and Kim, J.K., "Application of Web usage Mining and Product Taxonomy to Collaborative Recommendations in e-commerce," *Expert Systems with Applications*, Vol. 26, 2004, pp. 233-246.
- [8] Herlocker, J.L., Konstan, J.A., Borchers, A., and Riedl, J., "An Algorithmic Framework for Performing Collaborative Filtering," *In Proc. Conference on Research and Development in Information Retrieval*, 1999, pp. 230-237.
- [9] Ide, E., *New Experiments in Relevance Feedback*, In *The SMART Retrieval System*, Prentice Hall, 1971, pp. 337-354.
- [10] Kim, J.K., Cho, Y.H., Kim, W.J., Kim, J.R., and Suh, J.Y., "A Personalized Recommendation Procedure for Internet Shopping Support," *Electronic Commerce Research and Applications*, Vol. 1, No. 3, 2002, pp. 301-313.
- [11] Kim, J.K. and Cho, Y.H., "Using Web Usage Mining and SVD to improve E-Commerce Recommendation Quality," *PRIMA 2003*, 2003, pp. 86-97.
- [12] Konstan, J.A., Miller, B.N., Maltz, D., Herlocker, J.L., Gordon, L.R., and Riedl, J., "GroupLens: Applying Collaborative Filtering to USENET news," *Communications of the ACM*, Vol. 40, No. 3, 1997, pp. 77-87.
- [13] Lawrence, R.D., Almasi, G.S., Kotlyar, V., Viveros, M.S., and Duri, S.S. "Personalization of supermarket product recommendations," *Data Mining and Knowledge Discovery*, 2001, Vol. 5, No. 1, pp. 11-32.
- [14] Melville, P., Mooney, R.J., and Nagarajan, R., "Content-boosted Collaborative Filtering," *In Proc. SIGIR 2001 Workshop on Recommender Systems*, 2001.
- [15] Rocchio, J.J., *Relevance Feedback in Information Retrieval in the SMART Retrieval System - Experiments in Automatic Document Processing*, Prentice Hall, Englewood Cliffs, New Jersey, USA, 1971, pp. 313-323.
- [16] Salton, G., *Automatic Text Processing, The Transformation, Analysis, and Retrieval of Information by Computer*, Addison-Wesley, 1989.
- [17] Sarwar, B., Karypis, G., Konstan, J.A., and Riedl, J., "Analysis of Recommendation Algorithms for e-commerce," *In Proc. ACM E-Commerce*, 2000, pp. 158-167.
- [18] Schafer J.B., Konstan, J.A., and Riedl, J., "Recommender Systems in E-commerce," *Proceedings of the ACM Conference on Electronic Commerce*, 1999, pp. 158-166.
- [19] Smyth, B. and Cotter, P., "A Personalized TV Listings Service for the Digital TV Age," *Journal of Knowledge-Based Systems*, Vol. 13, No. 2-3, 2000, pp. 53-59.

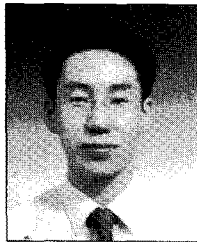
◆ 저자소개 ◆



김재경 (Kim, Jae Kyeong)  
서울대학교에서 산업공학 학사, 한국과학기술원에서 경영정보시스템 전공으로 산업공학 석사 및 박사학위를 취득하였으며 현재 경희대학교 경영대학 교수로 재직하고 있다. 미국 미네소타 주립대학교 경영정보학과에서 교환교수를 역임하였으며, 경희대학교 경영대학 정보센터 소장을 맡고 있다. 주요 연구분야로는 비즈니스 인텔리전스, 상품추천시스템, e비즈니스 모형 및 전략 등이며, 주요저서로, e비즈니스 시대의 경영정보시스템, e비즈니스 시스템, 비즈니스 인텔리전스, e-Business, 의사결정분석 및 응용 등이 있으며, Applied Artificial Intelligence, Artificial Intelligence Review, Electronic Commerce Research and Applications, European Journal of Operational Research, Group Decision and Negotiation, International Journal of Human-Computer Studies 등의 외국 학술지 및 다수의 국내 학술지에 논문을 게재하고 있다.



조윤희 (Cho, Yoon Ho)  
현재 국민대학교 경상대학 e-비즈니스학부 조교수로 재직 중이다. 서울대학교 계산통계학과(전산학전공)를 졸업하고, KAIST 경영정보공학과에서 석사, KAIST 경영공학과에서 박사학위를 취득하였으며, LG전자(주)에서 6년간 주임연구원으로 재직하였다. 주요 연구분야는 Mobile Business, CRM, Data mining 등이며, IEEE Intelligent Systems, Expert Systems with Applications, Electronic Commerce Research and Applications, Computers and Industrial Engineering, International Journal of Internet and Enterprise Management 등에 논문을 게재하였다.



김승태 (Kim, Seung Tae)  
강남대학교 경영학과에서 학사, 경희대학교 경영학과에서 e비즈니스 전공으로 석사학위를 취득하고 현재 DNI컨설팅회사에서 CRM/DBM 전략수립, 마케팅 실행과제 발굴, 로열티 프로그램 개발 등 마케팅 컨설팅 업무를 수행하고 있다.



김혜경 (Kim, Hye Kyeong)  
경희대학교 물리학과에서 학사, 경영학과에서 e-Business 전공으로 석사학위를 취득하고, 현재 동 대학원 박사과정에 재학 중이다. 주요 관심분야는 데이터 마이닝, 상품추천시스템, 유비쿼터스 컴퓨팅 등이다. Expert Systems: The International Journal of Knowledge Engineering and Neural Networks, Lecture Notes in Computer Science, International Conference on Electronic Business 등에 논문을 게재하였다.

◆ 이 논문은 2005년 1월 21일 접수하여 1차 수정을 거쳐 2005년 8월 30일 게재확정되었습니다.