

사용자 선호도와 시각적 기술자를 이용한 사용자 프로파일 기반 이미지 추천 알고리즘

김 덕 환[†] · 양 준 식^{**} · 조 원 희^{**}

요 약

정보 기술과 인터넷의 발전은 멀티미디어 콘텐츠의 양에 있어서 폭발적인 성장을 가져 왔으며 이러한 멀티미디어 콘텐츠 양의 증가는 이용자의 요구에 맞는 멀티미디어 콘텐츠 추천에 대한 필요성을 더 증가 시켰다. 현재까지 일반상품과 멀티미디어 콘텐츠 추천을 위한 기법에는 협업필터링 (CF: Collaborative Filtering)이 있다. 하지만 기존의 CF 기법은 이미지가 갖고 있는 시각적 특징을 제대로 표현하지 못하고 있으며, 입력 데이터의 희박성 (Sparsity) 문제와 신상품 추천 문제 그리고 선호도의 동적인 변화 문제를 포함하고 있기 때문에 이미지 콘텐츠 추천에는 적합하지 않다. 이와 같은 기존의 CF기법의 단점을 해결하기 위해서 본 논문에서는 새로운 이미지 추천 방법으로 FBCF (Feature Based Collaborative Filtering) 기법을 제안한다. FBCF 기법은 시각적 특징을 선호도에 따라 군집화한 새로운 사용자 프로파일 구성방법을 제시하며, 선호도 피드백을 통하여 구매자의 현재 성향을 추천에 반영할 수 있다. 실제 모바일 이미지 데이터를 사용한 실험에서 FBCF 기법이 기존의 CF 기법보다 400% 향상된 성능을 보임을 확인할 수 있다.

키워드 : 군집화, 이미지 세그멘테이션, 추천 시스템, 협업 필터링, 영역 기반 이미지 검색

Image recommendation algorithm based on profile using user preference and visual descriptor

Deok-Hwan Kim[†] · Junsik Yang^{**} · Won-Hee Cho^{**}

ABSTRACT

The advancement of information technology and the popularization of Internet has explosively increased the amount of multimedia contents. Therefore, the requirement of multimedia recommendation to satisfy a user's needs increases fastly. Up to now, CF is used to recommend general items and multimedia contents. However, general CF doesn't reflect visual characteristics of image contents so that it can't be adaptable to image recommendation. Besides, it has limitations in new item recommendation, the sparsity problem, and dynamic change of user preference. In this paper, we present new image recommendation method FBCF (Feature Based Collaborative Filtering) to resolve such problems. FBCF builds new user profile by clustering visual features in terms of user preference, and reflects user's current preference to recommendation by using preference feedback. Experimental result using real mobile images demonstrate that FBCF outperforms conventional CF by 400% in terms of recommendation ratio.

Key Words : Clustering, Image Segmentation, Recommendation System, Collaborative Filtering, Region-based Image Retrieval

1. 서 론

최근 유선인터넷에서 무선인터넷으로 더 나아가 유비쿼터스 네트워크로 정보통신 기술이 진화하면서 이미지, 음악, 동영상 등 수많은 멀티미디어 콘텐츠들이 유무선 통신망을 통해 사용자들에게 제공되고 있다[3]. 그러나 휴대 단말기를

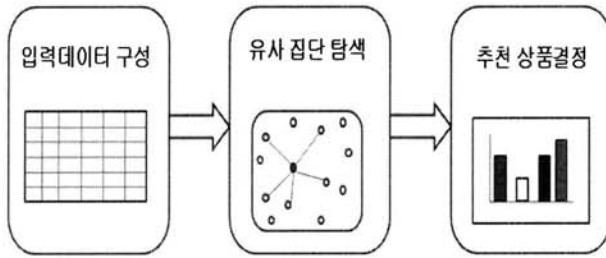
통해 멀티미디어 서비스를 제공받는 모바일 인터넷 환경에서는 작은 LCD 화면, 입력 장치의 제한, 접속 브라우저 능력의 한계 등으로 인하여 기존 유선 웹 환경과 달리 자유로운 정보 검색이 어렵다[4]. 따라서 원하는 멀티미디어 콘텐츠를 얻을 때까지 많은 노력과 비용을 요구하고 있어 실제로 고객들이 경험하는 서비스 만족도는 극히 낮은 것으로 보고되고 있다[2, 3].

향후 멀티미디어 서비스의 발전은 보다 개인적인 취향에 맞는 개별적인 서비스가 주종을 이룰 것으로 전망 되고 있다[1]. 이에 따라 각 이용자의 취향을 적기에 파악하여 개개인의 취향과 부합하는 멀티미디어 콘텐츠만을 제공함으로써

* 이 논문은 2007년 정부(교육과학기술부)의 재원으로 한국학술진흥재단의 지원을 받아 수행된 연구임.(KRF-2007-313-D00632)

† 종신회원 : 인하대학교 전자공학부 교수(교신저자)

** 준 회원 : 인하대학교 대학원 전자공학과 석사과정
논문접수: 2008년 3월 18일
수정일: 1차 2008년 5월 22일
심사완료: 2008년 6월 9일



(그림 1) CF 기반 추천 프로세스

고객이 보다 적은 노력으로 원하는 콘텐츠를 찾을 수 있도록 지원하는 개인화된 멀티미디어 추천 시스템의 필요성이 관련 산업에서 대두되고 있다[17].

추천 시스템은 통계적 기법과 지식탐사기술 (Knowledge Discovery Technology)을 이용하여 사용자의 취향에 가장 부합하는 상품을 추천해주는 시스템으로써, 고객들의 편의를 도모하고 교차판매 (Cross Sell) 및 매출 증대에 초점을 맞춘 시스템이다[5]. 현재까지 추천 시스템을 구현하기 위한 다양한 기법들이 개발되어 왔는데, 이 중에서 협업필터링 이하 CF (Collaborative Filtering)이 가장 성공적인 추천 기법으로 알려져 있으며 Amazon.com, CDNow 등과 같은 e-비즈니스 사이트에서 널리 활용되고 있다. CF 기반 추천 시스템은 상품을 추천하고자 하는 고객과 취향이 유사한 고객들의 의견을 반영하여 추천 대상 고객이 아직 구매하지 않은 상품에 대한 선호도를 예측한 후 선호도가 높을 것으로 예측되는 상품을 추천하는 시스템이다.

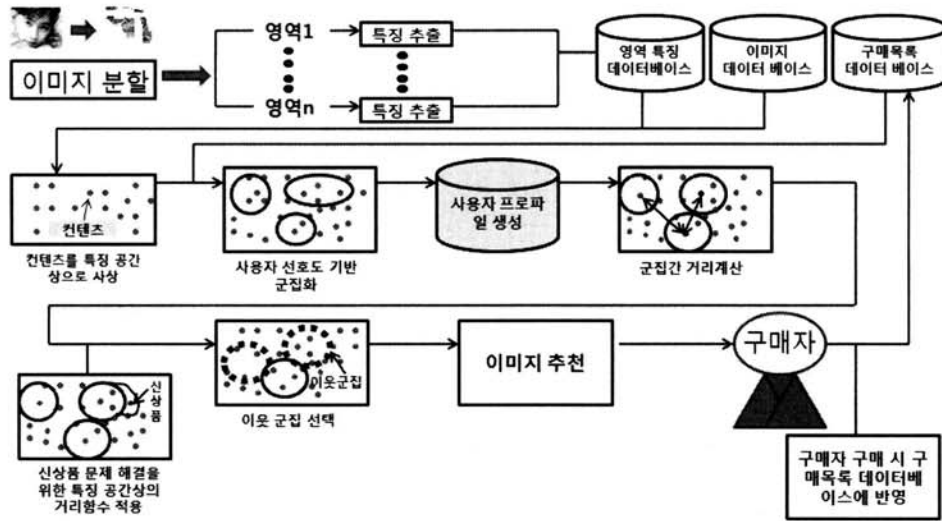
일반적으로 CF 기반 추천 프로세스는 (그림 1)과 같이 크게 입력 데이터 구성, 유사 군집 탐색, 추천 상품 결정 단계로 나누어볼 수 있는데, 이들 각각은 다음과 같은 역할을 수행 한다[5, 11].

- (1) 입력 데이터 구성: CF 기반 추천 시스템에서의 입력 데이터는 보통 n 개의 상품에 대한 n 명의 고객 선호도 집합으로 구성되며, 이는 n 고객-상품 행렬 P 로 표현 된다. 예를 들어, 구매 데이터로부터 선호도를 측정할 경우, 행렬 P 의 i 번째 행 i 번째 열의 값 p_{ij} 는 i 번째 고객이 j 번째 상품을 구매 했으면 1의 값을 그렇지 않으면 0의 값을 각각 가지게 된다.
- (2) 유사 선호도 군집 탐색: 유사 선호도 군집 탐색은 CF 기반 추천 시스템에서 가장 중요한 단계로써 고객-상품 행렬 P 를 이용하여 각 고객 u 에 대하여 가장 선호도가 유사한 l 명의 유사 군집 (Neighborhood)을 찾는 과정이다. 일반적으로 고객 간의 유사도 측정방법으로써 피어슨 상관계수 (Pearson Correlation), 코사인 프로젝션 (Cosine Projection) 등이 사용되고 있다.
- (3) 추천 상품 결정: 상품 추천을 위한 마지막 단계로써 설정된 유사 군집으로부터 N 개의 추천 상품을 결정한다. 추천 상품들을 선택하기 위한 기준으로써 일반적으로 빈발 구매 상품 추천 (Most-frequent Item Recommendation)이

사용된다. 빈발 구매 상품 추천은 대상 고객의 유사 군집에 대한 구매 이력 데이터를 분석하여 상품별 구매 빈도가 높은 상위 N 개의 상품을 추천하는 방법이다.

이러한 협업필터링은 가장 성공적인 추천기법으로 알려져 있으며, 다양한 인터넷 비즈니스 분야에서 적용되고 있으나, 다음과 같은 세 가지 근본적인 문제점으로 인하여 멀티미디어 콘텐츠 추천에는 적합하지 않다.

- (1) 입력 데이터의 희박성 (Sparsity) 문제 : CF 기반 추천 시스템은 고객의 선호도 데이터를 많이 확보할수록 추천의 정확도가 높아진다. 그러나 온라인 멀티미디어 서비스의 성장과 함께 유무선 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, 유사군집을 탐색하는 과정에서 아주 적은 수의 선호도 데이터를 사용하므로 고객들 간의 유사도 측정 시 신뢰성이 떨어지게 된다. 이러한 현상은 결국 추천결과의 정확도를 실추 시키게 하는 주요인으로 작용 한다[5, 11, 12, 13].
- (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[12, 13]. 따라서 멀티미디어 콘텐츠 기술의 발전과 고객 취향의 급속한 변화에 따라 수시로 새로운 멀티미디어 콘텐츠가 제공되는 웹 사이트에 협업필터링을 적용할 경우 매우 치명적인 약점이 될 수 있다. 이에 대한 해결책으로써 내용기반 필터링 (Content-based Filtering)을 협업 필터링과 혼합한 하이브리드 (Hybrid) 추천 방법이 근래에 등장하기 시작했다 [7, 12, 13, 16]. 내용기반 필터링은 내용이 유사한 상품은 사용자가 유사한 평가를 하리라는 가정에 의한 것이다. 즉, 하이브리드 추천 방법은 평가하지 않은 상품들의 평가 값을 내용기반 필터링으로 미리 예측함으로써 예측된 평가 값과 실제 사용자가 평가한 값들을 함께 고려하여 두 사용자의 유사 선호 경향을 분석한다. 따라서 두 사용자 간에 공통으로 평가한 상품이 없으면 두 사용자의 선호 경향이 유사한지 판별할 수 없게 되거나 평가 결과가 없는 상품을 추천할 수 없는 협업필터링의 근본적인 문제를 해결함으로써 추천의 정확도를 높일 수 있다. 그러나 현재까지의 영화 및 음악 추천 시 개발된 하이브리드 추천 방법은 정보검색 (Information Retrieval) 분야의 연구결과를 바탕으로 키워드 기반의 내용 속성을 사용하는 연구만이 진행되어 왔다. 키워드 기반의 내용속성 추출방법은 사람의 주관적인



(그림 2) 사용자 선호도 추천 알고리즘의 전체적인 개요

판단에 의해 속성 정보가 입력되므로 다양한 멀티미디어의 속성 (예: 이미지속성 - 색상, 질감, 모양 등, 음악속성 - 음정, 박자, 빠르기 등, 비디오속성 - 대표 프레임의 색상, 질감, 모션픽처)을 정확하고 객관적으로 측정하기가 어려워 멀티미디어의 내용속성에 대한 사용자의 선호도를 분석하는데 한계가 있다.

- (3) 선호도의 동적인 변화 문제: CF 기반 추천 시스템은 오로지 과거의 구매자의 구매 목록에서만 국한되어 사용되어지며 구매자의 가장 최근 선호도를 반영하지 않는다. 하지만 온라인 멀티미디어 상의 대부분을 차지하는 상품들이 음악이나 이미지는 짧은 기간 동안에 선호도의 동적인 변화를 가진다. 즉 이미지의 시각적인 특성상 그 의미를 표현하기 어렵기 때문에 같은 사용자라 하더라도 시간과 기분에 따라 선호도가 변화할 수 있다. 따라서 과거에 누적된 프로파일에서 추출되는 선호도정보를 기반으로 하는 CF 기반 추천 시스템은 동적인 변화를 가지는 구매자의 선호도를 만족시키기 어렵다 [17]. 이러한 현상은 추천결과의 정확도를 실추시키게 하는 주요인으로 작용한다.

이에 대한 해결책으로써 과거의 구매목록 뿐만 아니라 현재 구매자의 선호도를 학습하는 방법을 사용할 수 있다. 즉 학습된 선호도를 기반으로 사용자 프로파일을 재구성 하고, 이를 기반으로 상품들을 추천함으로써 동적인 변화 문제를 해결할 수 있다.

가전제품, 의류 등 물리적 상품의 경우에는 상품에 관한 정보가 부족하거나 주관적이어서 내용 분석을 하기가 어렵지만, 이미지, 오디오, 동영상 등의 멀티미디어 검색 (CBMR: Content-Based Multimedia Retrieval)에서는 일반 상품과 달리 내용 속성을 자동적으로 추출하여 검색에 사용하고 있다[6, 7, 8, 9, 10]. 따라서 개별 고객이 선호하는 멀티미디어 정보를 다차원 특징 공간상의 점들 (하나의 선호도 군집)로 표현한 후 추천 대상 고객과 다른 고객 간의 기하학적 거리

를 계산하여 선호도 유사군집을 구하면 위에서 논의한 협업 필터링의 문제점을 해결할 수 있다. 본 논문에서는 멀티미디어 추천 시 발생하는 협업필터링의 근본적인 문제점을 해결하기 위하여 CBMR 기법을 협업필터링과 결합한 하이브리드 방식의 멀티미디어 추천 알고리즘을 제안하고자 한다.

본 논문의 구성은 다음과 같다. 1장에서는 논문의 요약과 서론을 제시한다. 2장에서는 멀티미디어 추천 기법 중 모바일 배경화면 이미지 추천알고리즘의 개요와 모델 구축 관계 그리고 기존의 CF 시스템이 갖고 있는 문제점 해결을 나타낸다. 3장에서는 FBCF 기법을 사용하여 실험을 하고 결과를 논의한다. 4장에서는 결론에 대하여 기술한다.

2. 멀티미디어 추천 기법

2.1 모바일 배경화면 이미지 추천 알고리즘의 개요

본 절에서는 과거 T기간 동안 모바일 단말기를 이용하는 사용자들의 배경화면이미지 구매내역과 이미지들의 구매 패턴이 주어질 때 대상고객이 좋아할 이미지들을 시각적 속성에 기반을 두어 추천하는 알고리즘의 개요를 기술한다. 특히, 기존의 CF와 달리 새로운 고객 프로파일 구성방법을 제시한다. 추천 알고리즘은 모델 구축(model building phase) 및 이미지 추천(image recommendation phase)의 두 가지 단계로 나누어진다.

(그림 2)는 전체적인 과정을 보여준다. 모델 구축 단계는 고객 구매목록 데이터베이스로부터 신뢰할 수 있는 모델을 구축하기 위하여 주기적인 시간 단위로 한 번씩 수행된다. 반면에, 이미지 추천단계는 대상고객이 구매할 확률이 높은 컨텐츠를 추천하기 위하여 사용된다.

모델 구축단계는 다음의 3단계로 나누어진다: 이미지 분할단계, 시각적 속성 추출단계, 구매목록 데이터베이스를 이용하여 사용자가 구매한 이미지들을 특징공간에서 군집하는 단계. 첫 번째는 데이터베이스 내에 있는 모든 이미지들에

대하여 이미지 분할 방법을 수행한다. 하나의 이미지는 여러 개의 의미 있는 영역들로 분할된다. 두 번째는 이미지의 영역들로부터 색상, 질감, 모양 등의 다양한 시각적 특징들을 추출한다. 영역에 기반을 둔 지역화된 특징은 개체를 잘 표현할 수 있기 때문에 이미지의 전체 픽셀로부터 추출한 지역화된 특징보다 사용자의 상위수준 개념을 확실히 파악할 수 있다.

마지막으로는 구매목록 데이터베이스를 분석하여 사용자가 구매한 이미지들을 특징공간에서 군집하여 사용자 프로파일을 구축한다. 이미지는 특징벡터 공간에서 각각 특징벡터로 표현될 수 있으며, 마찬가지로 영역들도 특징공간에서 점들로 나타낼 수 있다. 즉, 사용자가 구매한 이미지들은 특징공간에서 여러 개의 점들로 표현되며 이들을 그룹 하여 사용자별로 군집을 구성하게 된다.

일반적인 CF 기법과는 달리 FBCF 기법에서는 사용자가 구매한 이미지들을 벡터 공간상의 특징벡터들로 표현하여 구매한 군집의 집합을 사용자 프로파일로 사용하여 사용자 간의 선호도를 벡터 공간상의 군집간 거리함수로 측정할 수 있게 한다.

추천 단계는 모델구축 단계에서 만들어진 특징군집들의 집합을 이용하여 대상고객의 이웃들을 탐색한다. 특징군집들의 집합에 대한 k-최근접 탐색방법을 수행하여 대상고객의 군집과 최근접한 이웃군집을 형성한다. 최종적으로, 이웃군집의 군집들에 속한 이미지들과 대상고객의 군집반경에 포함된 신상품들 중 구매확률이 상위-N번째에 들어가는 이미지들을 추천하게 된다. 즉, 멀티미디어의 특성을 이용하여 유사한 속성을 가진 콘텐츠들을 추천할 수 있는 장점을 가지고 있다.

본 논문에서 제안하는 멀티미디어 콘텐츠 추천 알고리즘의 개요는 다음과 같다. 일반적으로, 이미지가 갖고 있는 의미를 저 수준에서 시각적으로 표현하기 어렵기 때문에 불확실성과 애매성이라는 문제점이 이미지 검색의 성능을 일반적으로 낮게 하는 요인이다. 이미지 콘텐츠 추천 알고리즘은 이미지 콘텐츠의 영역특징 데이터베이스, 구매목록 데이터베이스, 사용자 프로파일 p등의 입력을 받아 추천된 이미지 콘텐츠 목록 R를 출력한다.

첫 번째, 영역 특징 데이터베이스 및 구매목록 데이터베이스 구축함으로써 데이터베이스내의 모든 이미지들에 대하여 이미지분할방법을 적용하여 영역들을 구성한다. p차원 특징 공간 R^p 에서 배경화면 이미지들을 구성하는 영역들의 특징 벡터 $x_i = \{x_{i1}, x_{i2}, \dots, x_{ik}\}, i=1, \dots, N$ 를 각각 추출하고 영역을 구축한다. 모든 사용자에 대하여 이미지의 구매내역을 구매목록 데이터베이스에 저장한다.

두 번째, 사용자 프로파일 생성으로써 사용자들이 구매한 이미지들을 그룹 하여 특징군집들의 집합으로 사용자 프로파일을 구성하는 과정이다. 특징군집은 각각 임의의 사용자 a가 구매한 적이 있는 이미지들의 영역을 묶은 군집으로 나타낸다. 군집의 중심과 분산, 유효반경, 군집에 속한 사용자 정보와 데이터들을 사용자 프로파일에 저장한다.

세 번째, 이웃군집 형성으로써 대상고객 c와 임의의 다른 사용자 a의 프로파일에 있는 콘텐츠들은 각각 군집을 형성한다. 이웃군집을 구하기 위하여 특징군집들의 집합을 이용한 k-최근접 질의를 수행한다. 기본적으로 대상고객 c와 다른 사용자 a의 군집간 거리 T_{ca} 를 이용하여 계산하며 대상고객 c를 위한 이웃 군집 $H = \{h_1, h_2, \dots, h_L\}, c \in H$ 를 T_{ca} 값의 오름차순으로 L개를 구한다.

네 번째, 추천 목록 생성으로써 이웃군집이 과거에 구매한 적이 있는 이미지들 x에 대한 대상고객 c의 구매확률 $PLS(c, x)$ 을 계산하여 구매빈도가 높은 상위 k개의 콘텐츠 $R = \{x^1, x^2, \dots, x^k\}$ 를 추천한다.

2.2 모델구축단계

본 절에서는 사용자가 구매했던 이미지들에 대한 시각적 속성을 특징 벡터로 표현하여 사용자 프로파일을 생성한다. CF 알고리즘의 주요 요소인 사용자 프로파일은 콘텐츠에 대한 사용자의 선호 여부에 관한 정보를 포함한다. 사용자 프로파일의 생성은 (a) 이미지를 분할하고 특징 정보를 추출하는 과정과 (b) 사용자가 구매한 콘텐츠를 특징군집들의 집합으로 표현하는 과정으로 각각 구성된다.

2.2.1 이미지 분할 과정

이미지 분할을 위하여 버클리 대학의 Malik등이 제안한 normalized cuts 세그멘테이션 방법을 수정한 알고리즘을 사용하였다. Normalized cuts는 점들의 집합을 부분집합들로 구분하기 위해 그래프 이론방법을 적용한다[15]. 이 방법은 이미지를 영역으로 분할하는 경우에 적용될 수 있다. 이미지의 각 픽셀을 특징공간의 점으로 취급하고 선택된 특징에 따라 매우 유사한 픽셀들을 묶어준다.

(그림 3)은 SKT의 배경화면 이미지 다운로드 서비스에서 제공하고 있는 원본 이미지와 normalized cuts 세그멘테이션 방법을 이용하여 영역을 구별한 이미지의 예제를 보여 준다 [14]. 본 논문에서는 normalized cuts 이미지 분할방법을 사



(그림 3) 세그멘테이션을 이용하여 추출된 영역들

용하여 이미지에 포함된 객체를 구별하는 영역을 만들었다. 즉, 이미지는 객체를 포함하는 다수의 영역들로 표현된다.

2.2.2 이미지 콘텐츠의 특징 추출 과정

이미지의 모양을 표현하기 위하여 영역의 크기 RS , 영역의 위치, 2차 모멘트 등의 속성을 사용한다[18]. 영역의 크기 RS 는 이미지의 높이 H 와 폭 W 이고 영역의 넓이가 A 일 때 이미지의 크기에 의해 정규화 된 영역의 면적으로 구해진다. $RS = \frac{A}{W \times H}$ 영역의 위치는 이미지 내의 객체의 상대적인 위치를 나타낸다.

축척불변성(scale invariance)을 유지하기 위하여 각 좌표들은 이미지의 높이와 폭에 의해 각각 정규화 한다. 먼저 영역의 중심 좌표는 행과 열에 따라 영역 픽셀의 분포를 나타낸 히스토그램의 평균값 x_{cm} , y_{cm} 이 구해지며 영역의 위치 x_{loc} , y_{loc} 는 x_{cm} , y_{cm} 를 폭과 높이에 관하여 정규화한 값이다. 2차 모멘트는 영역의 중심 좌표를 기준으로 한 영역의 픽셀들의 표준편차이다. 영역의 픽셀 개수를 N 으로 할 때 x 축과 y 축 상의 영역의 2차 모멘트는 다음의 식 (1), (2)와 같이 구해진다 :

$$(2차\ 모멘트)_x = \frac{1}{(W/2)} \sqrt{\frac{\sum_{n=1}^N (x_n - x_{cm})^2}{N}} \quad (1)$$

$$(2차\ 모멘트)_y = \frac{1}{(H/2)} \sqrt{\frac{\sum_{n=1}^N (y_n - y_{cm})^2}{N}} \quad (2)$$

Compactness는 영역의 면적을 둘레의 길이의 제곱으로 나눈 비율이다. 영역의 모양이 원일 때 가장 높은 값을 가지며, 오목한 모양일 때 작은 값을 갖는다. Compactness는 식 (3)으로 구할 수 있다.

$$Compactness = \frac{A}{P^2} \quad (3)$$

침도 (convexity)는 영역의 블록된 정도를 나타낸다. 영역의 면적을 철폐 (convex hull)의 면적으로 나누어 구할 수 있다. 침도는 식 (4)로 구할 수 있다.

$$Convexity = \frac{A}{A_{hull}} \quad (4)$$

색상은 객체를 표현하는 유용한 특성중의 하나이다. 영역의 색상은 $L*a*b$ 색상공간에서의 픽셀들의 평균과 표준편차를 계산한 6차원의 속성 값들로 나타낸다. 질감은 영역 안에 있는 명암 패턴의 변화를 나타낸다. 각각 다른 규모(scale)와 방향을 갖는 선형 필터뱅크를 사용하여 필터에 대한 영역내의 픽셀들의 반응을 평균하였다. 필터뱅크의 짝수 부분은 가우시안 커널의 2차 미분을 이용하여 홀수부분은 힐버트 변환을 이용한다.

2.2.3 사용자 프로파일 생성

특징군집들의 집합을 구성하는 목적은 이웃 군집을 보다 효율적으로 찾고자 하는 것이다. 특징군집은 임의의 사용자 a 가 구매한 적이 있는 이미지들의 영역들로 나타낸 점들을 포함하며, 군집의 중심, 분산, 유효반경과 군집의 이미지를 구매한 사용자에 대한 정보를 포함한다. $UP = \{C_1, C_2, \dots, C_k\}$ 로 표현한다. 여기서, C_i 는 i 번째 사용자가 구매한 이미지들을 구성하는 영역들의 집합 $\{x_{i1}, x_{i2}, \dots\} = \{\{x_{i11}, x_{i12}, \dots, x_{i1m}\}, \{x_{i21}, x_{i22}, \dots, x_{i2m}\}\}$ 이다. n_i 는 i 번째 사용자가 구매한 이미지들의 개수이다. $\{x_{i11}, x_{i12}, \dots, x_{i1m}\}$ 는 이미지 x_{i1} 을 구성하는 영역들의 집합이다. 군집 C_i 는 군집의 중심 \bar{x}_i , 가중치 공분산 행렬 S_i , 유효 반경 γ 과 사용자에 대한 정보를 포함한다.

$\bar{x} = (x_1, \dots, x_p) \in \mathbb{R}^p$, $S_i = \sum_{k=1}^{n_i} v_k (x_k - \bar{x}_i)(x_k - \bar{x}_i)'$ 평균 벡터는 초평면 타원의 위치를 결정한다. 반면에 공분산 행렬은 형태와 방향을 나타낸다. 각 군집의 상대적인 가중치는 각 군집에 있는 점들의 적합성 점수들의 합으로 정해진다. 일반적으로 군집을 타원으로 표현할 수 있다. 유효 반경은 새로운 이미지 x 가 주어진 군집에 속하는 지 여부를 결정하는 임계값이다. 임의의 점 x 가 타원 안에 위치하면 식 (5)와 같은 성질을 만족 한다[3].

$$(x - \bar{x}) \left(\frac{1}{n} S \right)^{-1} (x - \bar{x}) < \frac{(n-1)p}{(n-p)} F_{p, n-p}(\alpha) \quad (5)$$

데이터가 정규 분포를 따른다고 가정 하고 α 를 유의수준이라고 하자. 주어진 유의수준에 대하여, 데이터의 $100(1-\alpha)\%$ (일반적으로 95~99%)가 타원 내에 위치하며 그 $F_{p, n-p}(\alpha)$ 거리함수는 자유도가 $p, n-p$ 인 분포를 따른다. α 가 감소할수록, 주어진 유효 반경은 증가한다. 타원의 바깥에 위치한 점들은 이상치 (outlier)로 인식되며 새로운 군집을 형성한다.

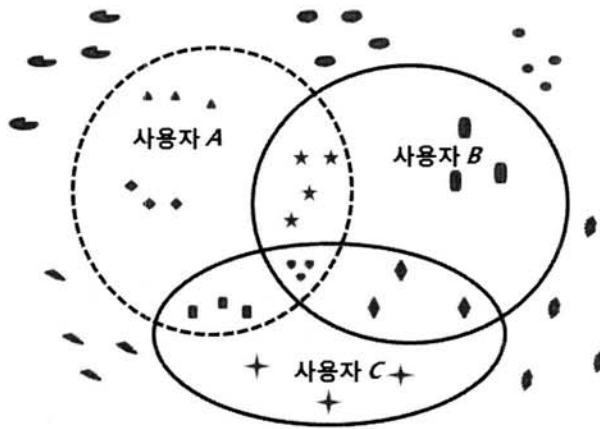
i 번째 사용자가 구매한 이미지들의 영역의 집합을 표현한 군집 C_i 의 크기를 n_i , 평균을 $\bar{x}_i = \sum_{x \in C_i} x / n_i$, 분산을 $S_i = \sum_{x \in C_i} (x - \bar{x}_i)(x - \bar{x}_i)'$ 라고 하자. 두 군집 C_i 와 C_j 의 공분산은 $S_{p_{ij}} = (S_i + S_j) / (n_i + n_j - 2)$ 이다.

본 논문에서는 이웃군집 형성에 적합한 군집간 거리 함수로써 다음과 같이 Hotelling의 T_{ij}^2 함수를 사용한다. 두 군집 C_i 와 C_j 간의 거리함수는 식 (6)과 같이 정의된다.

$$T_{ij}^2 = \frac{n_i n_j (n-2)}{(n_i + n_j)^2} (m_i - m_j)' S_{p_{ij}}^{-1} (m_i - m_j) \quad (6)$$

2.2.4 이웃군집 형성

기존의 CF 알고리즘은 Cosine 함수나 Person 계수 등을 사용하여 사용자간 상관도를 계산한다. 그러나 이 방법은 목표 사용자와 선호도가 비슷한 유사군집을 찾기 어렵다. 본 논문에서는 기존의 구매정보나 웹로그 정보를 이용하여 상관관계에 의해 이웃을 찾아 추천하는 것과 다르게 멀티미



(그림 4) 사용자가 선호하는 이미지 콘텐츠의 집합

디어 상품을 특징공간에서 표현하기 때문에 유사한 속성을 가진 상품들을 추천할 수 있는 장점을 가지고 있다.

(그림 4)는 사용자 'A', 'B', 'C'가 각각 선호하는 이미지 상품들의 집합을 2차원 특징공간상에 군집으로 각각 표현한 것이다. 특징 공간상에 구매 데이터베이스의 모든 이미지들의 영역들을 점들로 나타낼 수 있다. 각 사용자가 구매한 이미지들 집합은 군집을 형성한다. (그림 4)에 보여준 것과 같이, 사용자 'A'가 구매한 이미지 집합은 5개의 이미지들이며 16개의 영역으로 구성된다. 이들 중 사용자 'A'와 'B'가 함께 구매한 이미지는 2개, 사용자 'A'와 'C'가 구매한 이미지는 2개이며, 'A', 'B'와 'C'가 함께 구매한 이미지는 1개이다.

이미지 상품들이 다차원 특징 공간에서 점들로 표현될 수 있기 때문에, 본 논문에서는 목표 사용자와 다른 사용자간의 거리를 계산하여 실제 유사군집을 구할 수 있다. 군집간 거리 함수로 유클리디언 거리 함수가 주로 사용된다. 이 함수는 단순하고 계산하기 쉬우며 군집이 균일하게 분포하며 군집의 형태가 원일 때 잘 동작한다. 그러나 각 사용자의 선호하는 정도는 같지 않으며 그들의 분포는 (그림 4)에서 보여준 바와 같이 서로 다르다.

대상고객 c 와 임의의 다른 사용자 a 의 프로파일에 있는 상품들은 각각 군집을 형성한다. 이웃군집을 구하기 위하여 특징군집트리를 이용한 k -최근접 질의를 수행한다. 기본적으로 대상고객 c 와 다른 사용자 a 의 군집간 거리는 T_{ca}^0 를 이용하여 계산하며 대상고객 c 를 위한 이웃 군집 $H=(h_1, h_2, \dots, h_L), c \in H$ 를 T_{ca}^0 값의 오름차순으로 L 개를 구한다. 구해진 거리 값에 따라서 오름차순으로 유사도 군집을 정렬하여 L 유사도 군집을 선택할 수 있다. 최종적으로, 목표 사용자 c 를 위한 유사도 군집 $H=(h_1, h_2, \dots, h_L), c \in H$ 를 결정한다.

2.2.5 추천 목록 생성

상품 추천을 위한 마지막 단계로써 설정된 유사군집으로부터 빈발 구매 상품 추천 방법을 사용하여 구매빈도가 높은 상위 k 개의 상품 $R=\{x_1, x_2, \dots, x_k\}$ 를 추출한다. 상품 x 에 대한 목표 사용자 c 의 구매확률 (PLS: Purchase Likeness

Score)은 다음과 같이 정의된다.

$$PLS(c, x) = \frac{\sum_{a \in H} R_x \times sim(c, a)}{\sum_{a \in H} sim(c, a)} \tag{7}$$

식(7)에서 사용자 a 는 유사도 군집 H 에서 구해지며 R_x 는 유사군집이 이미지 x 를 구매한 빈도이다. 식 (8)은 목표 사용자 c 와 이웃 사용자 a 간의 유사한 정도를 구하는 함수이며, 2.3절에서 정의된 군집간 거리함수에서 구한 값의 역수를 취해 정규화 하여 구한다. 사용자 u, w 는 각각 목표사용자의 유사군집 H 에 속해 있다.

$$sim(c, a) = \frac{Max_{u, w \in H}[d(u, w)] - d(c, a)}{Max_{u, w \in H}[d(u, w)] - Min_{u, w \in H}[d(u, w)]} \tag{8}$$

2.3 기존의 CF 시스템이 갖고 있는 문제점 해결

2.3.1 입력 데이터의 희박성 문제 해결을 위한 멀티미디어 특징 표현 방법

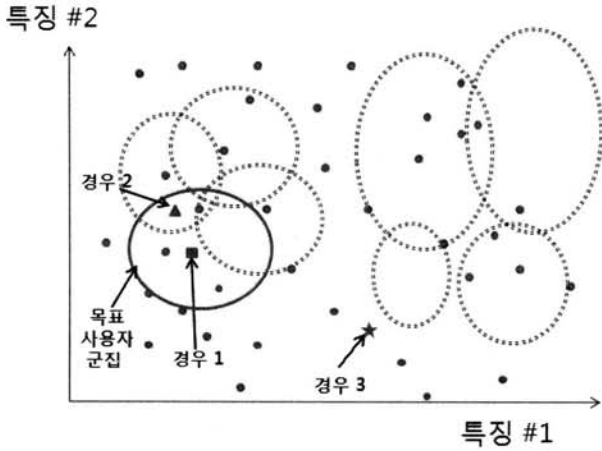
아주 적은 수의 선호도 데이터 사용으로 유사도 측정 신뢰성이 떨어지게 된다. 본 논문에서는 사용자 프로파일을 고객-상품 행렬 대신에 다차원 속성공간내의 특징벡터들로 표현한다. 따라서 사용자가 선호하는 상품들이 속성공간의 점들로 표현되며, 이들은 군집을 이루게 된다. 입력 데이터가 적은 경우 군집내의 점들이 적어지게 되어 군집간 거리함수 계산할 때 문제가 발생할 수도 있다[19].

이와 같은 입력 데이터의 희박성 문제를 해결하기 위하여 2.1절에서 설명한 바와 같이 한 상품을 여러 개의 특징벡터, 즉 여러 개의 객체들로 표현하는 방법을 개발하고자 한다. 예를 들어, 이미지의 경우 여러 개의 영역 (region)으로 식별되며, 각각 영역의 특징벡터를 추출하여 공간상에 여러 개의 점으로 표현할 수 있다. 사용자가 한 이미지를 구매하는 경우 그 이미지를 표현하는 여러 개의 영역 특징 벡터들이 입력되므로 데이터의 희박성 문제를 해결 할 수 있다. 즉, 하나의 상품에 대하여 속성공간상에 다수의 점들로 표현하는 시도이다.

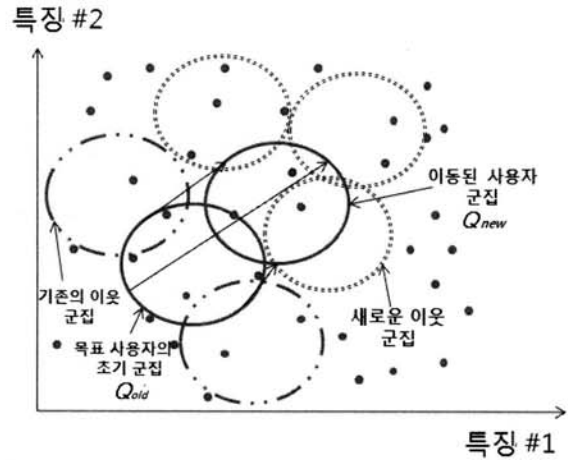
2.3.2 신상품 추천 문제 해결을 위한 특징공간상의 거리함수 적용

신상품은 구매된 적이 없으며, 선호도를 갖지 않고 있다. 만약 신상품이 선호도 (rating)를 갖는다면 추천될 수 있을 것이다. 본 논문에서는 이 문제를 해결하기 위하여 신상품에 가상의 선호도 (virtual rating)를 부여하여 신상품을 추천하는 방법을 제안 한다. 가상의 선호도 부여에는 세 가지 부여 방법이 있다. 첫 번째는 최대값을 부여하는 방법, 두 번째로는 평균값을 부여하는 방법, 세 번째로는 최소값을 부여하는 방법이다. 이 세 가지 방법 중 본 논문에서는 첫 번째 방법인 최대값을 부여하는 방법을 적용하여 실험하였다.

(그림 5)에서 '경우 1', '경우 2'와 '경우 3'은 신상품이며



(그림 5) 사용자가 선호하는 이미지 상품 이웃관계



(그림 6) 구매자 선호도 반영 기법의 이동되는 군집

다차원의 특징 벡터들로 표현되므로 특징 공간간상에 점들로 나타내어진다. '경우 1'은 신상품 중에서 선택된 사용자에게 구매된 상품을 나타내며 '경우 2'는 신상품 중에서 구매되지는 않았지만 선택된 사용자 군집 안에 포함된 상품을 나타낸다. '경우 3'은 신상품 중에서 구매되지도 않았으며 선택된 사용자의 군집 안에 포함되지 않은 상품을 나타내며 이 경우에는 베이지안 분류방법을 사용하여 g 개의 군집 C_1, \dots, C_g 가 주어질 때 새로운 상품 '경우 3'이 포함될 군집을 결정할 후, 해당 군집의 유효 반경 (effective radius) 거리 내에 '경우 3'이 위치하면 목표 사용자 군집에 포함 시키는 방법을 제안한다. 신상품 x_{new} 를 위한 군집 C_i 의 베이지안 분류 함수는 식 (9)과 같이 정의된다.

$$\hat{d}_i(x_{new}) = -\frac{1}{2}(x_{new} - \bar{x}_i)'S_{pooled}^{-1}(x_{new} - \bar{x}_i) + \ln(w_i) \quad (9)$$

식 (9)에서 w_i 는 i 번째 군집의 정규화 된 가중치이며, 가장치는 사용자의 선호도의 합으로 구해진다.

$\hat{d}_1(x_{new}), \hat{d}_2(x_{new}), \dots, \hat{d}_g(x_{new})$ 중에 가장 큰 값을 갖는 군집 C_k 를 선택한 후 x_{new} 가 그 군집의 유효 반경 안에 위치하는 지를 검사한다.

즉, $(x_{new} - \bar{x}_k)' \left(\frac{1}{n} S \right)^{-1} (x_{new} - \bar{x}_k) < \frac{(n-1)p}{(n-p)} F_{p, n-p}(\alpha)$ 이면 목표 사용자 군집에 포함 시키게 된다.

통계적으로, 군집의 유효반경은 자유도 p , $n-p$ 신뢰수준 α 인 F 분포를 따른다.

2.3.3 구매자의 동적인 선호도 반영을 위한 선호도 피드백 추천 방법

이미지는 시각적인 특성상 그 의미를 표현하기 어렵기 때문에 같은 사용자라 하더라도 시간과 기분에 따라 선호도가 변화 할 수 있다. 또한 CF 기반 추천 시스템은 오로지 과거의 구매자의 구매 목록의 리스트로만 추천 리스트를 만든다. 본 논문에서는 이 문제를 해결하기 위하여 선호도 피드백을 반영한 FBCF-PF (Preference Feedback) 방법을 제안한다. 지정된 구매자 프로파일 P 는 각각 긍정적인 특징 군

집과 부정적인 특징 군집으로 형성 된다. 다른 구매자들 또한 각각 긍정적인 특징 군집과 부정적인 특징 군집으로 형성 된다. 부정적인 특징 군집은 짧은 기간 동안의 구매자의 동적인 선호도를 나타낸다. 긍정적인 특징 군집은 구매자 자신이 구매한 콘텐츠 목록을 포함하며 긴 기간 동안의 구매자의 선호도를 나타낸다. 기존의 CF 방법은 구매자의 선호도를 축적된 구매자의 구매 목록 프로파일로부터 학습시킨다. 또한 구매자의 구매목록 중 구매자가 현재 구매한 새로운 구매 목록은 방출시킨다. 이와 같이 CF기법은 오로지 과거의 구매자의 구매 목록에서만 국한되어 사용되어지며 추천 구매자 세대의 가장 최근 선호도를 반영하지 않는다는 단점이 있다. 하지만 FBCF-PF 방법은 서로 영향을 미치는 선호도 피드백을 통하여 구매자의 현재 선호도를 반영한다.

(그림 6)은 가장 최근의 선호도를 반영하기 위하여 선호도를 피드백 하는 과정을 보여준다. 즉 피드백을 통해 사용자 군집 중심을 다시 계산하며 사용자 군집의 중심이 이동된다. 사용자 군집의 이동시 이웃군집은 다시 설정된다. FBCF-PF기법은 rocchio 공식을 사용하여 사용자 군집의 중심이 움직인다. 새로운 의미의 특징 군집 벡터 Q_{new} 은 구매자의 구매 이미지와 선호도 피드백 추천을 시행함으로써 얻어지는 추천 목록 군집의 구매 이미지들을 사용하여 구할 수 있다. Q_{new} 을 구하는 식 (10)은 다음과 같다.

$$Q_{new} = Q_{old} + \beta \sum_{i=1}^{n_1} \frac{R_i}{n_1} - \gamma \sum_{i=1}^{n_2} \frac{S_i}{n_2} \quad (10)$$

식 (10)에 Q_{new} 는 구매자의 특성을 반영한 T 시간 동안의 이동된 새로운 사용자 군집이며 Q_{old} 는 구매자의 특성이 반영되지 않은 $T-1$ 시간 동안의 기존의 선택된 사용자 군집을 의미한다. R_i 는 구매자가 구매한 각각의 이미지의 특징 벡터를 의미 하며, n_1 은 구입 이미지 영역 개수이다. S_i 는 T 시간 동안의 추천 목록 중에 구매자가 아직 구매하지 않은 각각의 특징 벡터이다.

이렇게 구해진 Q_{new} (이동된 사용자 군집)에 포함되어 있는 상품들을 추천하는 것은 사용자의 가장 최근의 선호도를

반영한 추천 시스템의 완성을 의미하는 것이다.

3. 실험 및 결과

실험은 본 논문에서 제안한 FBCF 기법과 FBCF-PF 기법을 기존의 방법인 CF 기법과 비교 및 평가하였다. 실험 환경은 인텔 코어 2 듀오 2.4Ghz CPU 가 탑재된 PC를 사용하며, 마이크로소프트사의 윈도우즈 XP 환경에서 C++ 언어를 이용하여 추천시스템을 구현하였다.

3.1 실험 환경

실험은 시각적 특징을 선호도에 따라 군집화한 새로운 사용자 프로파일 구성방법으로 추천하는 FBCF 기법과, 선호도 피드백을 통하여 구매자의 현재 성향을 추천에 반영할 수 있는 FBCF-PF 기법의 성능을 평가 하는데 초점을 두었다. 세부적인 실험은 다음의 두가지 목표에 대해 수행 한다. 첫째, 추천율의 성능을 평가하기 위하여 FBCF 기법과 기존의 CF 기법을 비교하고, 고객이 구매 하지 않은 신상품에 선호도를 할당하여 신상품이 추천가능한 FBCF 기법과 신상품 추천 문제를 가지는 기존의 CF 기법의 신상품 추천 성능을 비교 한다. 둘째, 구매자의 동적인 변화 문제를 해결한 FBCF-PF 기법과 기존의 CF 기법의 추천 성능을 비교 한다. FBCF-PF 기법은 동적으로 변화는 군집의 횟수를 초기에는 1로 설정 하고 3까지 변화 할 수 있게 설정 하였다.

실험에는 SKT(SK Telecom)에서 제공한 실제 데이터를 사용했으며 <표 1>과 같이 사용자의 구매목록 데이터 25,680개, 배경화면 이미지 데이터 5,326개 및 고객정보 데이터 476개로 이루어져 있다. 구매목록 데이터는 구매한 고객 ID, 이미지를 구매한 날짜, 구매한 이미지 ID로 구성되며, 배경화면 이미지 데이터는 배경화면 이미지의 ID, 배경화면 이미지의 이름, 배경화면 이미지의 발매 날짜로 구성되며, 고객정보 데이터는 고객 ID, 고객이 배경화면 이미지를 첫 구매한 날짜, 마지막 구매날짜 및 배경화면 이미지의 총 구매 횟수로 구성된다. 영역특징 데이터는 배경화면 이미지 ID와 배경화면의 이미지가 분할되어 영역별로 나뉘어져 있는 영역 ID, 색상을 특징으로 하는 6차원의 특징데이터, 모양을 특징으로 하는 6차원의 특징데이터, 질감을 특징으로 하는 8차원의 특징데이터로 구성된다. 실험에서는 2004년 6월부터 8월까지의 데이터로 한정하였다.

실험의 신뢰성을 높이고 양질의 추천을 하기 위해서는 학습 데이터를 여과 할 필요가 있기 때문에 고객중 SKT로부터 배경화면 이미지를 40개 이상 구매한 고객들로만 한정 하였다.

제한한 방법의 성능 평가를 위해서 추천율과 신상품 추천율을 사용한다. 식 (11)과 같이 추천율은 추천된 이미지의 수에 구매된 이미지의 수를 나눈값으로 정의 한다.

$$\text{추천율} = \frac{\sum_{i=1}^n h_i}{n} \tag{11}$$

n 은 추천을 수행하는 횟수이며, h_i 는 한번 시행 했을때의 추천율(구매된 이미지의 수/추천된 이미지의 수)을 의미 한다.

신상품은 추천일을 기준으로 15일, 30일, 45일을 기준으로 하여 신상품임을 판단한다. 실험에서는 30일을 신상품으로 판단한다. 식 (12)와 같이 신상품 추천율은 다음과 같이 계산 한다.

$$\text{신상품 추천율} = \frac{\sum_{i=1}^n hnew_i}{n} \tag{12}$$

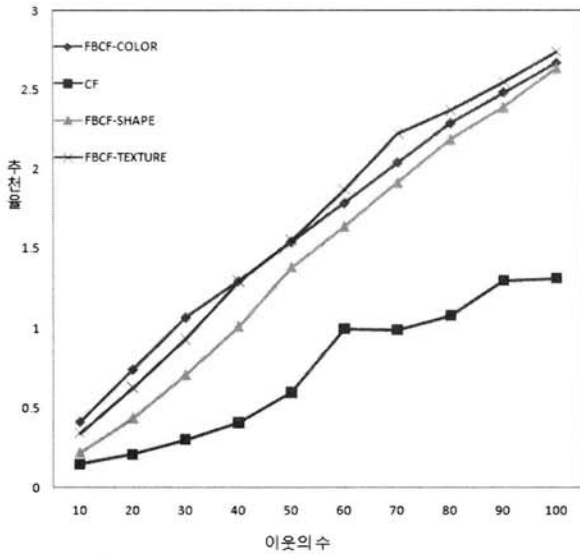
n 은 추천을 수행하는 횟수이며, $hnew_i$ 는 한번 시행 했을 때의 신상품 추천율(구매된 신상품 이미지의 수/추천된 이미지의 수)을 의미 한다.

실험에서 초기에 이웃의 개수는 10개로 설정 하였고, 최대 100개의 이웃으로 추천을 시행하여 총 10번의 추천을 시행 하여 평균을 결정 한다. 학습기간은 2004년 6월부터 2004년 8월까지로 한정 하였고, 이 기간을 6 구간으로 나누고 이를 15일씩 구분하여 군집화를 실시하였다. 앞서 설명한 학습 기간은 군집화를 시행할 때에 군집에 들어가는 이미지의 출시 날짜를 고려하여 학습 기간과 매칭이 되는 것을 가지고 군집을 생성하고 6번의 추천을 시행하여 평균을 결정 한다. 위의 실험들은 총 10번씩 반복 한다. 따라서 이웃을 기반으로 하는 실험의 n 은 100이고, 출시 날짜를 기반으로 하는 실험의 n 은 60이다.

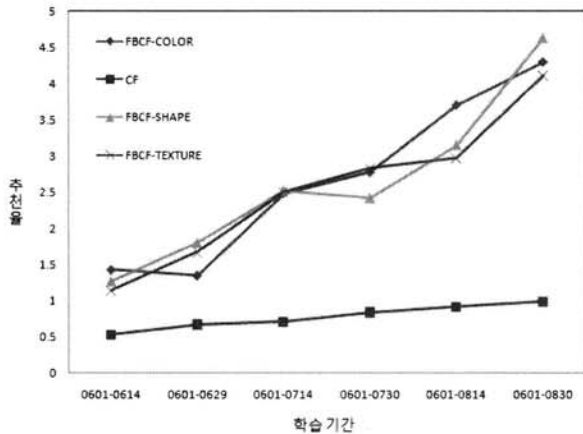
또한, 영역특징점을 색상, 모양 및 질감으로 하는 FBCF 기법과 기존의 CF 기법과의 추천율 성능평가를 시행 하고, FBCF-PF 기법의 성능을 평가 하기 위하여 동적인 군집 형성의 횟수를 1에서 3까지 변화하여 기존의 CF 기법과의 성능 평가를 시행 한다.

<표 1> 실험에 사용한 데이터

데이터	속성	수량
구매목록 데이터	구매날짜, 고객 ID, 이미지 ID	25,680
배경화면 이미지 데이터	이미지 ID, 이미지 이름, 이미지 발매 날짜	5,326
고객정보 데이터	고객 ID, 첫 구매날짜, 마지막 구매 날짜, 구매 횟수	476
영역특징 데이터	이미지 ID, 영역 ID, 색상(6차원),모양(6차원),질감(8차원)	57,692



(그림 7) 이웃의 수와 특징점의 추천율 효과

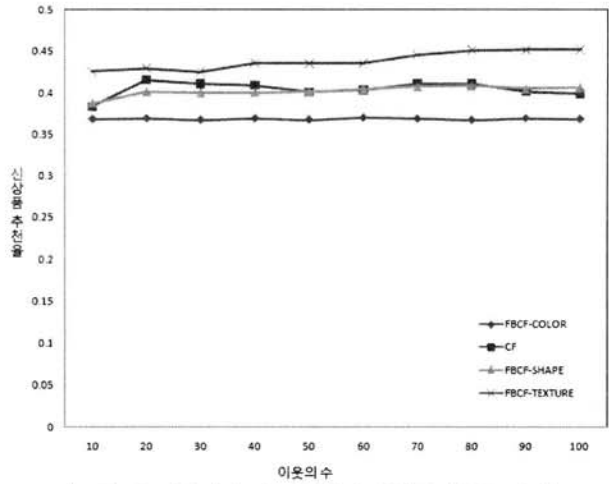


(그림 8) 학습 기간과 특징점의 추천율 효과

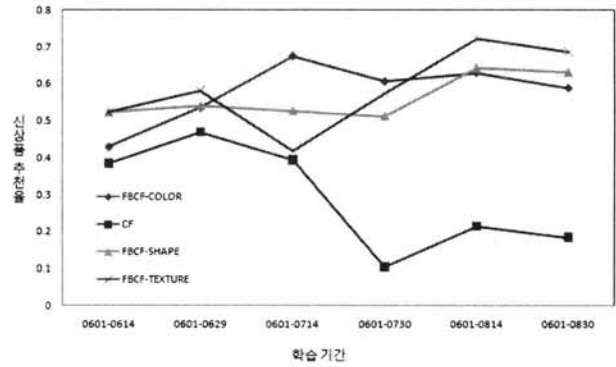
3.2 FBCF 기법을 이용한 실험 및 분석

(그림 7)은 영역특징점을 색상, 모양 및 질감으로 하는 FBCF와 기존의 CF와의 추천율 성능평가를 한다. 이웃의 수와 특징점에 따른 추천율의 결과는 질감을 영역 특징점으로 하는 FBCF 기법이 기존의 CF 기법보다 최대 250% 향상된 성능을 보여 준다. 또한 FBCF 기법은 이웃의 수가 10개에서 100개로 증가 하였을 때 추천의 성공률이 최대 700% 향상되었다. 특징별 성능은 영역특징점을 질감으로 하는 것이 영역 특징점을 모양으로 한 것 보다 최대 157% 높은 성능을 보여 준다. FBCF 기법은 이웃이 많아짐에 따라 성향이 비슷한 이웃들의 이미지들을 기반으로 추천을 함으로서 추천 성공률이 높게 나타났다. 이웃의 수가 증가 하면 추천해 줄 수 있는 목록이 늘어나므로 추천 성공률이 높아진다.

(그림 8)은 배경화면 이미지 구매 기록을 군집으로 만드는 학습 기간을 변화함에 따른 추천율 변화량을 실험한 결과이다. 학습 기간이 늘어날수록 군집도 커지며 이웃의 군집도 커진다. 기존의 CF 기법의 결과와는 최대 400% 향상된 성능을 보여 주었다.



(그림 9) 이웃의 수와 특징점의 신상품 추천율 효과



(그림 10) 이웃의 수와 특징점의 신상품 추천율 효과

신상품에 대한 실험은 기존의 CF방식의 신상품에 대한 추천 문제를 해결하기 위해 시행한다. 신상품 추천을 위한 실험은 아직 구매 된 적이 없는 신상품에 가상 선호도를 부여하여 신상품도 기존의 상품과 같이 추천을 받을 수 있게 하는 방법을 이용한다. 본 실험에서는 가상의 선호도에 신상품이 속한 군집의 선호도값들 중 최대값을 부여하는 방법을 적용하였다.

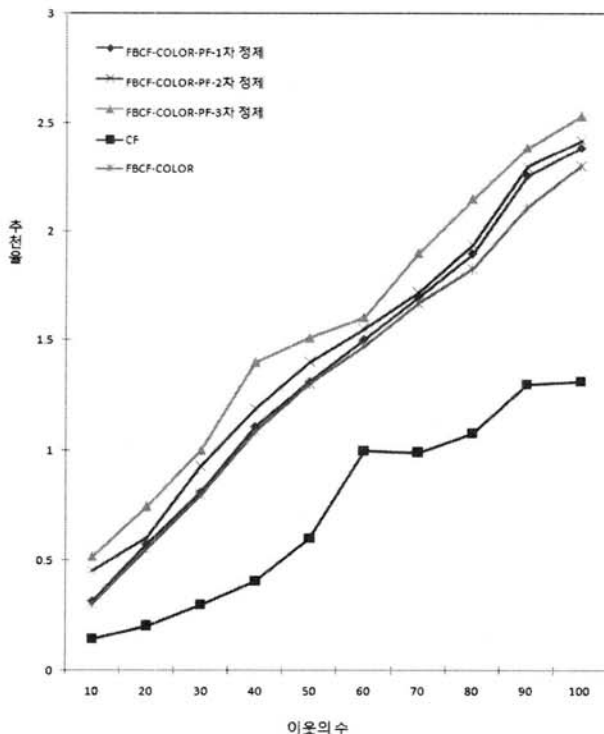
(그림 9)는 이웃의 수의 변화에 따른 특징점을 기반으로 한 신상품 추천율을 나타내었고, (그림 10)은 학습기간의 변화에 따른 특징점을 기반으로 한 신상품 추천율을 나타내었다. 질감을 영역특징점으로 하는 FBCF방식은 CF방식보다 신상품추천율이 80% 좋게 나타난다. 그 이유는 다음과 같다. CF에서 추천될 수 있는 신상품은 선택된 사용자에게 이미 구매된 상품이거나 신상품 중에서 구매되지는 않았지만 선택된 사용자 군집 안에 포함된 상품일 경우뿐이다. 반면에 FBCF는 앞의 경우를 포함하여 아직 선택된 사용자의 군집 안에 포함 되지 않은 상품도 추천할 수 있다.

(그림 10)은 학습 기간에 따른 신상품 추천율의 변화를 알아보기 위한 실험이다. FBCF는 학습 기간이 증가함에 따라 추천율은 증가 하지만 CF 방식은 신상품 추천율이 불규칙적으로 변화 하는 것을 알 수 있다. CF 방식에서는 신상품에 관한 정보가 부족하기 때문에 적절하게 신상품을 추천

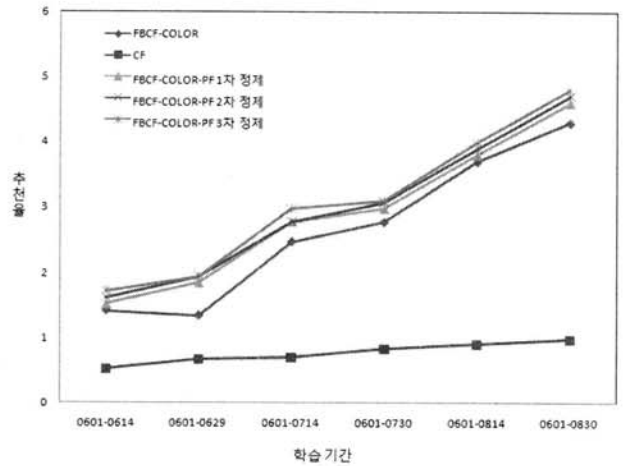
하지 못할 수도 있다. 이러한 CF 방식을 개선한 FBCF의 실험 결과는 학습 기간이 증가할수록 신상품 추천율이 점차 증가함을 보여준다.

3.3 FBCF-PF를 이용한 실험 및 분석

이미지는 시각적인 특성상 그 의미를 표현하기 어렵기 때문에 같은 사용자라 하더라도 시간과 기분에 따라 선호도가 변화 할 수 있다. 이 문제를 해결 하기 위해서는 가장 최근의 선호도를 반영하여 추천을 시행해야 한다. 즉 피드백을 통해 사용자 군집 중심을 다시 계산하며 사용자 군집의 중심을 이동 시키고 이웃군집을 다시 설정 하여 사용자의 동적인 선호도를 반영할 수 있다. 사용자의 동적인 선호도를 반영한 실험은 사용자 피드백의 횟수를 달리 하여 총 3번을 진행한다. 영역특징점은 색상을 이용하였으며 이웃의 수와 학습 기간의 증가에 따른 추천율의 변화를 측정한다. 즉 사용자의 동적인 피드백의 횟수를 인자로 두어 추천율을 측정한다. (그림 11)에서는 이웃의 숫자와 동적인 선호도를 반영하여 추천율의 변화를 알아본다. CF 방식이 가장 낮은 추천율을 보였으며, 동적인 선호도를 반영하지 않은 색상 기반의 영역특징점 추천율이 다음으로 나타났다. 동적인 선호도를 1번 적용한 것과 2번 적용한 것 3번 적용한 것 순으로 추천율의 성능을 보여 주었다. CF와 동적인 선호도를 3번 적용한 실험은 동적인 선호도를 3번 적용한 것이 200% 향상된 성능을 보여 주었다. 색상 기반의 영역특징점을 적용한 FBCF와 동적인 선호도를 세 번 적용한 것은 동적인 선호도를 세 번 적용한 것이 110% 향상된 성능을 보였다.



(그림 11) 이웃수와 동적인 선호도를 반영한 추천율 효과



(그림 12) 학습 기간과 동적인 선호도를 반영한 추천율 효과

(그림 12)는 학습 기간에 따른 추천율의 변화를 알아보기 위한 실험이다. (그림 12) 또한 (그림 11)과 같이 사용자의 동적인 선호도를 적용하여 실험을 진행 하였다. 사용자의 동적인 선호도 피드백을 세 번까지 적용 하였고 세 번 적용한 실험이 가장 높은 추천율을 보였다. CF 방식이 가장 낮은 추천율을 보였으며, 동적인 선호도를 반영하지 않은 색상 기반의 영역특징점의 추천율이 다음으로 나타났다. 동적인 선호도를 1번 적용한 것과 2번 적용한 것 3번 적용한 것 순으로 추천율의 성능을 보여 주었다. CF와 동적인 선호도를 3번 적용한 실험은 동적인 선호도가 약 190% 정도 높은 결과를 보였다. 색상 기반의 영역특징점 방식과 동적인 선호도를 세 번 적용한 것과는 110% 정도 높은 결과를 보임으로서 멀티미디어 추천 방식에 사용자의 멀티미디어 추천 방식에 사용자의 동적인 선호도를 적용한 추천 시스템이 향상된 성능을 가짐을 알 수 있었다.

4. 결론

본 논문에서는 협업 필터링 기법의 널리 알려진 문제점인 입력데이터의 희박성 문제와 신상품 추천 문제 그리고 구매자의 동적인 선호도 반영을 위한 선호도 피드백 추천 문제를 해결 하는 이미지 추천 알고리즘을 제안하였다. 입력데이터의 희박성 문제를 해결하기 위해 멀티미디어 특징을 이용한 사용자 프로파일 구성방법과 신상품 추천 문제를 해결하기 위한 특징 공간상의 거리함수 방법 그리고 구매자의 동적인 선호도 반영문제 해결을 위한 선호도 피드백 추천 방법 등을 제안한 FBCF 기법은 기존의 CF 기법 보다 높은 수준의 추천율을 제공한다.

멀티미디어 시장 규모가 빠르게 증가함에 따라 제안된 FBCF기법을 실제 추천시스템에 적용하여 기존의 CF기법에 의한 멀티미디어 콘텐츠 적용의 한계성을 극복하고 멀티미디어 콘텐츠 추천을 위한 킬러 애플리케이션으로 사용할 수 있을 것이다.

제안된 알고리즘을 적용한 실험 결과는 CF 기법보다 FBCF 기법이 최대 400% 향상된 성능을 보였으며 CF 기법에 비

해 FBCF 기법의 신상품 추천율이 최대 80% 향상된 성능을 보였다. 또한 FBCF 기법에 비해 선호도 피드백이 적용된 FBCF-PF는 110% 향상된 성능을 보였다.

향후 연구로는 추천된 이미지와 사용자가 구매한 이미지 간의 격차를 해소하여 사용자의 선호도를 더욱 많이 반영할 수 있는 군집으로 바뀌어나가는 양질의 추천 시스템의 알고리즘을 설계 할 것이다. 또한 이미지를 비롯하여 음악, 영상, 책 및 영화의 영역까지 추천의 범위를 넓혀가는 방향으로 연구를 진행하고자 한다.

참 고 문 헌

- [1] 인터넷백서, 한국인터넷정보센터, 2007.
- [2] G. Buchanan, M. Jones and H. Thimbleby, "Improving mobile internet usability," Proceedings of the 10th World Wide Web Conference, 2001.
- [3] 무선인터넷 산업분석, 대우증권, 2005.
- [4] Ovum Forecasts: Global Mobile Markets 2003-2007, Ovum, 2007.
- [5] C. Y. Kim, J. K. Lee, Y. H. Cho, and D. H. Kim, "VISORS: A visual content recommender for the mobile web," IEEE Intelligent Systems, 19(6), pp.32-39, Nov., 2004.
- [6] B. Sarwar, G. Karypis, J. Konstan and J. Riedl, "Analysis of recommendation algorithms for e-commerce," Proceedings of the ACM E-Commerce Conference, pp.158-167, 2000
- [7] Y. H. Cho, J. K. Kim and S. H. Kim, "A personalized recommender system based on Web usage mining and decision tree induction," Expert Systems with Applications, 23(3), pp.329-342, 2002.
- [8] M. Balabanovic and Y. Shoham, "Fab: content-based collaborative recommendation," Communication of the ACM, 40(3), pp.66-72, 1997.
- [9] P. Melville, R. J. Mooney, and R. Nagarajan, "Content-boosted collaborative filtering for improved recommendations," Proceedings of the 18th National Conference on Artificial Intelligence, pp.187-192, 2002.
- [10] J. T. Foote, "Content-based retrieval of music and audio," Proceedings of SPIE, pp.138-147, 1997.
- [11] J. K. Kim, Y. H. Cho, M. Y. Kang, and H. K. K, "A hybrid multimedia contents recommendation procedure for a new item problem in m-commerce," 2006.
- [12] M. Flickner, H. Sawhney, W. N. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker, "Query by image and video content: the QBIC system," IEEE Computer Magazine, 28(9), pp.23-32, 1995.
- [13] Y. Ishikawa, R. Subramanyay and C. Faloutsos, "MindReader: querying databases through multiple examples," Proceedings of the 24th VLDB Conference, pp.218-227, 1998.
- [14] K. Chakrabarti, K. Porkaew and S. Mehrotra, "Query refinement for multimedia similarity retrieval in MARS," Proceedings of the 7th ACM Multimedia Conference, pp. 235-238, 1999.
- [15] R. Brunelli and O. Mich, "Image retrieval by examples," IEEE Transactions on Multimedia, 2(3), pp. 164-171, 2000.
- [16] J. Shi and J. Malik, "Normalized cuts and image segmentation," IEEE Transactions On Pattern Analysis And Machine Intelligence, 22(8), Aug., 2000.
- [17] D. Comaniciu and P. Meer, "Roubst analysis of feature spaces: color image segmentation, Proceedings of Conference on Computer Vision and Pattern Recognition, pp.750-755, 1997.
- [18] D. H. Kim, J. W. Song, J. h. Lee, and B. G. Choi, "Support vector machine learning for region-based image retrieval with relevance feedback," ETRI Journal, 29(5), pp.700-702, Oct., 2007.
- [19] Y. H. Cho, C.Y. Kim, and D. H. Kim, "Image recommendation in the mobile internet," Proceeding of the conference on PRICAI, pp.963-964, 2004.



김 덕 환

e-mail : deokhwan@inha.ac.kr
 2003년 2월 한국과학기술원 컴퓨터(공학박사)
 2006년~현 재 인하대학교 전자공학부 부교수
 관심분야: 시각정보처리, 데이터마이닝,
 임베디드 시스템



양 준 식

e-mail : juneseek@iesl.inha.ac.kr
 2008년 2월 인하대학교 컴퓨터정보공학부
 (학사)
 2008년 3월~현 재 인하대학교 전자공학과
 석사과정
 관심분야: 추천 시스템, 시스템 소프트웨어



조 원 희

e-mail : chowon@iesl.inha.ac.kr

2007년 2월 명지대학교 정보통신공학부

통신공학과(학사)

2007년 9월~현재 인하대학교 전자공학과

석사과정

관심분야: 추천시스템, 시스템 소프트웨어