
이미지 기반 필터링을 이용한 개인화 아이템 추천

Personalized Item Recommendation using Image-based Filtering

정경용
상지대학교 컴퓨터정보공학부

Kyung-Yong Chung(kyjung@sangji.ac.kr)

요약

유비쿼터스 컴퓨팅의 발달로 인하여 다양하고 폭넓은 정보가 디지털 형태로 빠르게 생산 및 배포되고 있다. 사용자가 이러한 정보과잉 속에서 자신이 원하는 정보를 단시간 내에 검색하는 것은 그리 쉬운 일이 아니다. 본 논문에서는 이미지 기반 필터링을 이용한 개인화 아이템 추천 기법을 제안한다. 피상적인 내용 분석이라는 단점을 개선하기 위하여 사용자가 관심을 가지는 이미지 데이터로부터 특징을 추출하는 이미지 기반 필터링을 사용하였다. 제안한 방법에 대해 MovieLens 데이터에서 내용 기반 필터링과 협력적 필터링과의 비교 실험을 통해 성능을 평가하였다. 실험 결과, 제안한 방법이 기존의 다른 방법보다 우수함을 확인하였다.

■ 중심어 : | 추천 시스템 | 이미지 검색 | 데이터마이닝 | 유비쿼터스 컴퓨팅 |

Abstract

Due to the development of ubiquitous computing, a wide variety of information is being produced and distributed rapidly in digital form. In this excess of information, it is not easy for users to search and find their desired information in short time. In this paper, we propose the personalized item recommendation using the image based filtering. This research uses the image based filtering which is extracting the feature from the image data that a user is interested in, in order to improve the superficial problem of content analysis. We evaluate the performance of the proposed method and it is compared with the performance of previous studies of the content based filtering and the collaborative filtering in the MovieLens dataset. And the results have shown that the proposed method significantly outperforms the previous methods.

■ keyword : | Recommender System | Image Retrieval | Data Mining | Ubiquitous Computing |

1. 서론

이미지 프로세싱 기술과 이미지 데이터의 저장 방식의 급속한 발전에 힘입어 이미지 데이터베이스는 하나의 중요한 유비쿼터스 컴퓨팅 시스템으로 간주되고 있

다. 현재 이미지 데이터를 검색하고 관리할 수 있는 기능을 제공하는 상용화된 데이터베이스 시스템이 있다 [1][2][4]. 그러나 이러한 상용 시스템이 이미지 데이터 검색에 사용하는 방법은 이미지 데이터와 정보 검색 분야에서의 관련된 키워드를 사용하는 방식으로서 사용

* "본 논문은 2007년도 상지대학교 교내 연구비 지원에 의한 것입니다."

접수번호 : #071012-001
접수일자 : 2007년 10월 12일

심사완료일 : 2007년 02월 28일
교신저자 : 정경용, e-mail : kyjung@sangji.ac.k

자가 이미지 데이터를 데이터베이스에 입력할 때 적절한 키워드를 같이 입력해야만 한다. 이러한 방식은 이미지에 대응하는 적절한 키워드를 만들기가 쉽지 않으며 더구나 이런 키워드를 기억하거나 알고 있어야만 원하는 이미지를 검색할 수 있다는 단점을 가진다[1][2]. 이에 비해 키워드와 관계없이 사용자가 시각적인 예를 이용해 검색하면 원하는 정보를 쉽게 검색할 수 있다. 즉, 이미지 데이터의 내용을 기반으로 검색할 수 있는 시스템이 필요하다[2][3]. 본 논문에서는 대부분의 사용자가 유사한 내용 정보 또는 동일한 이미지를 연속적으로 검색하는 습관을 갖는 특성을 이용하여 사용자가 관심을 가지는 이미지 데이터로부터 특징을 추출하는 이미지 기반 필터링을 이용한 개인화 아이템 추천 기법을 제안한다. 이미지 기반 필터링은 특징 거리 척도를 사용하여 비교되는 이미지의 특징을 자동으로 추출하기 위해 이미지 픽셀에 색 히스토그램을 적용하여 사용한다[3]. 그리고 이미지 데이터에 대한 선호도 차이와 색 히스토그램 거리의 정적 상관 관계를 선형적인 예측 모델에 결합하여 개인화 아이템을 추천한다. 기존의 내용 기반 필터링을 이용한 추천 시스템은 문서의 텍스트를 대상으로 단어의 위치와 출현 빈도를 고려해 파라미터들의 확률값을 추정하는 다항 모델을 기반으로 Naive Bayes 분류자를 이용하여 프로파일을 학습하고 추천을 한다. 본 논문에서는 웹문서에서 많은 부분을 구성하는 이미지 데이터를 내용 기반 필터링에 포함하기 위해서 이미지의 내용 기반 검색을 위한 기술에 중점을 둔다.

2. 관련연구

유비쿼터스 컴퓨팅에서 이미지 데이터베이스 시스템에 대한 관심이 크게 늘어나고 있다. 이미지 프로세싱 기술과 이미지 데이터의 저장 방식의 급속한 발전에 힘입어 이미지 데이터베이스는 하나의 중요한 유비쿼터스 컴퓨팅 시스템으로 간주되고 있다.

2.1 유사 이미지 데이터 추출과 패턴인식

입력된 이미지 데이터와 동일한 이미지 데이터를 검

색하는 것은 패턴인식 분야에서 연구가 되어 왔다. 그러나 이미지 기반 필터링을 위한 유사 이미지 데이터 추출 방법과 일반적인 패턴 인식 방법에는 몇 가지 중요한 차이점이 있다.

첫째, 이미지 기반 필터링에 저장되는 이미지 데이터는 일반적으로 매우 다양한 실세계의 정보를 담고 있어서 표준 패턴을 정해 놓고 최근접법분류(NNC: Nearest Neighbor Classification)을 통해 목표가 되는 이미지 데이터를 찾아가는 일반적인 패턴 매칭 방식의 적용이 불가능하다는 점이다. 따라서 특정 다양한 형태의 이미지의 데이터를 대상으로 하는 이미지 기반 필터링의 경우 데이터의 인식 작업이 패턴인식 시스템보다 매우 어렵다. 또한 질의를 처리할 때도 표준 패턴 또는 모델을 미리 정의하기 불가능하며 하나의 결과를 추출하기 보다는 유사한 후보를 질의 결과로 추출하게 되고 최종인식이나 의미적 해석은 주로 사용자가 한다.

둘째, 이미지 데이터는 패턴인식 시스템이 다루는 데이터의 양에 비해 매우 많다는 점이다. 실제로 대부분의 패턴인식 시스템은 이미지 기반 필터링에 비해 인식할 이미지 데이터를 적게 가지므로 이미지 데이터를 검색하는데 간단한 순차 검색 방법을 사용한다. 이미지 기반 필터링은 많은 이미지 데이터를 저장하게 되므로 순차적 검색 방법은 매우 비효율적이며 따라서 적절한 이미지 데이터 자동 분류 방법이 필요하다.

2.2 이미지 데이터 검색 기법

이미지 데이터 검색은 검색 방식에 의해서 분류될 수 있는데, 이미지 검색방식에 의해 분류해 보면 명세 기반 방식, 브라우저를 이용한 검색, 비의미적 특성을 이용한 검색[4], 공간 제약을 이용한 검색, 의미적 특성을 이용한 검색으로 분류될 수 있다.

명세 기반 방식은 사람이 인식할 수 있는 내용을 처리할 수 있고 의미를 쉽게 모델링할 수 있는 장점이 있지만 사람이 일일이 명세를 작성해야 하며 관점에 따라 명세가 달라질 수 있다는 단점이 있다. 브라우저를 이용한 검색은 사용자에게 친숙한 인터페이스를 제공하여 이미지를 쉽게 검색할 수 있도록 하는 방식이다. 비의미적 특성을 이용한 검색은 가장 전통적이고 많이 활

용된 방법으로 이미지의 색상, 모양, 질감 등을 이용한다. 단점은 색상, 모양, 질감 등의 낮은 단계의 특성을 가지고, 이미지의 내용을 표현하기 힘들다는 점이다. 공간 제약을 이용한 검색은 이미지에 있는 객체나 영역의 위치와 관계를 이용해서 이미지를 검색하는 방식이다. 이 방식은 이미지를 객체나 영역으로 얼마나 잘 분할하느냐에 따라 영향을 많이 받는데, 영상분할은 해결되지 않은 분야 중에 하나이다. 의미적 특성을 이용한 검색은 이미지가 가지고 있는 내용을 최대한 객관적으로 표현하고, 검색하려고 시도하는 방식이다. 단점은 데이터의 표현이 어렵다는 점이다[5].

2.3 개인화 아이템 추천

추천 시스템은 선호도를 기반으로 하여 개인화 아이템을 추천한다. 정보 검색 시스템으로부터 추천 시스템을 구별할 수 있는 기준은 “개인화”와 “흥미 있고 유용한”과 같은 단어가 된다. 아이템 추천에서 사용되는 기법은 다양한 기준에 의해 여러 가지로 분류할 수 있다. 아이템 추천에서 사용되는 기법은 다양한 기준에 의해 여러 가지로 분류할 수 있다. 개인화는 추천하는데 기초가 되는 데이터의 출처와 그 데이터가 입력되어 어떻게 이용되는지를 기준에 따라 내용 기반 필터링, 사례 기반 필터링, 규칙 기반 필터링, 협력적 필터링(사용자 기반 협력적 필터링과 아이템 기반 협력적 필터링)의 범주로 분류된다. 여기서 최근 개인화 아이템 추천에서 가장 많이 이용되고 있는 기법은 협력적 필터링과 내용 기반 필터링이다[2][3].

3. 이미지 기반 필터링을 이용한 추천

피상적인 내용 분석이라는 단점을 개선하기 위해 사용자가 관심을 가지는 이미지 데이터로부터 특징을 추출하고 선호도와 이미지 색 정보 사이의 정적 상관관계를 이용하는 이미지 기반 필터링에 대해서 기술한다.

[그림 1]은 이미지 기반 필터링의 구성도이다. 내용 정보 데이터베이스를 구축하기 위한 웹 로봇 에이전트에 의한 정보 추출은 MovieLens 데이터[6]에서 제공하

는 URL을 사용해서 영화에 대한 웹문서를 수집한다. 이미지 기반 필터링에서는 이미지 데이터에 대한 선호도와 색 히스토그램 거리의 정적 상관관계를 선형적인 예측 모델에 결합하여 개인화 아이템을 추천한다. 이는 사용자가 아이템에 대해서 선호도를 평가할 때 이미지의 색 정보가 유사한 아이템을 평가하고 있다는 것을 의미한다. 서로 유사한 색 히스토그램 거리를 가지고 있는 이미지를 군집한 후 군집내의 총 이미지 수에 대한 각 이미지 개수의 비율을 나타내는 가중치 벡터를 사용하여 이미지 기반 필터링에서 아이템 추천을 하였다. 이미지 기반 필터링에서 개인화 아이템을 추천함으로써 유사한 아이템만을 제공하여 과도하게 특수화되는 경향을 개선한다.

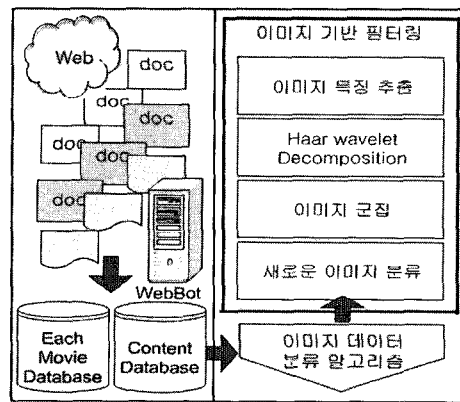


그림 1. 이미지 기반 필터링의 구성도

3.1 이미지 특징 추출을 위한 색 히스토그램

이미지의 내용을 표현하기 위해 사용되어진 특징을 살펴보면 이미지에 대한 명세, 객체의 모양, 질감 등을 이용하는 비의미적 방법[7], 의미적 특성, 브라우징, 공간제약[5], 색 히스토그램 방법 등이 있다[2]. 이 중에서 이미지 데이터를 표현하기 위하여 가장 일반적으로 사용되는 방법이 색 히스토그램이다. 색 히스토그램은 멀티미디어 이미지 데이터의 특성을 표현하기 위하여 고차원 또는 저차원으로 표현이 가능하다. 색 히스토그램을 고차원으로 이용할 경우 이미지에 대한 많은 색상 정보를 유지함으로써 이미지간의 유사도 검색할 때 정밀도가 높은 장점을 가진다. 그러나 R*-트리[8]와 같은

색인 구조에서는 많은 겹침이 발생하여 검색 성능을 저하시키고, 고비용의 유사도 계산을 해야 하는 단점이 있다. 반면, 저차원의 색 히스토그램을 사용할 경우에는 색인 구조에 효율적일뿐 아니라 저비용의 계산으로 유사도를 측정할 수 있으나, 이미지 데이터에 대한 색상 정보의 손실로 검색할 때 정밀도가 떨어지는 단점이 있다. 이러한 저차원 색 히스토그램으로의 변환을 통해 손실되는 이미지 데이터의 색상 정보를 보완하기 위해서 2차원 Haar 웨이블릿 변환을 사용한다[9]. 웹문서에서 많은 부분을 구성하는 이미지 데이터는 색 히스토그램을 계산하고 양자화된 색을 가진 픽셀의 비율로 구성된다. 이미지 데이터들을 비교하기 위해서 색 히스토그램간의 유클리드 거리를 계산한다. 두개의 이미지 데이터를 비교하기 위한 색 히스토그램의 거리는 $Cdis(P_k, P_l)$ 로 표현하고 식(1)과 같이 정의한다. 여기서 $P_k(j)$ 는 j 색을 가지는 이미지 k 의 픽셀 수의 비율을 의미한다. $Cdis$ 는 0에서부터 7까지의 유효범위를 가진다.

$$Cdis_1(P_k, P_l) = \sum_j |P_k(j) - P_l(j)| \quad (1)$$

색 히스토그램은 픽셀의 배열을 고려하지 않고 이미지 특징 추출을 하기 위해서 2차원 Haar 웨이블릿 변환을 사용하여 서브 이미지를 생성한다. 원래의 이미지를 그대로 사용하면 복잡도가 증가하기 때문에 계산량이 많아지게 되고 상당히 많은 표본수가 필요하게 된다. 이러한 이유로 이미지 데이터의 특징 벡터를 추출하기 위해서 차원을 감소시켜 사용한다.

이미지 데이터의 특징 추출을 위해 서브 이미지는 이진 값으로 양자화 되므로 원래의 이미지 데이터의 픽셀은 길이가 9인 이진 벡터를 가지게 된다[3][10]. 길이가 $512(=2^9)$ 벡터의 히스토그램은 이미지 데이터의 분석과 유사도를 계산하기 위한 특징 벡터가 된다. 본 논문에서는 Haar 웨이블릿을 기저로 하는 웨이블릿 변환을 총 4번하게 되는데 웨이블릿 변환을 4번을 하는 이유는 5회 이상에서는 정보의 손실이 많아 적정 횟수를 4회로 결정한다. [그림 2]는 Haar 웨이블릿 변환을 총 4번 반복한 것을 나타내는 그림이다. 이미지는 1차원 신호가 x 축과 y 축으로 나열되어 있는 2차원 신호로 간주할 수

있다. 이를 분석하기 위해서는 이미지를 각각 x, y 방향으로 Low pass 필터와 High pass 필터에 통과시키면, 한 이미지에 대해서 네 부분의 분할된 성분이 추출된다. 저주파 성분을 L로 표시하고 고주파 성분을 H로 표시할 때, LL부분은 x, y 양방향 모두 Low pass 필터에 통과시킨 성분이고, HH는 모두 High pass 필터에 통과시킨 성분이라는 것을 의미한다. LH는 x 방향으로 Low pass 필터, y 방향으로 High pass 필터에 통과시킨 성분을 의미하고 HL은 x 방향으로 High pass 필터, y 방향으로 Low pass 필터를 통과시킨 성분을 의미한다. 이와 같은 이미지 데이터의 다중 분할을 통해 이미지 데이터의 크기를 1/4로 줄인 상태에서 원래의 이미지 데이터의 정보가 대부분 유지되는 LL 부분을 얻을 수 있다.

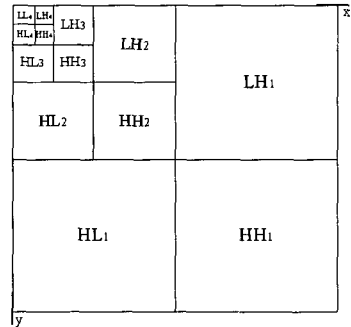


그림 2. 이미지 데이터의 다중 분할

3.2 선호도와 이미지 색 정보 사이의 관계

선호도와 이미지 색 정보 사이의 관계를 분석하기 위해서 MovieLens 데이터[6]에서 선호도와 이미지 데이터를 사용하여 기술한다. 선호도와 이미지 색 정보 사이의 상관관계를 분석하기 위해서 200명의 사용자가 평가한 선호도와 평가한 아이템을 대표하는 이미지 데이터 1,000개를 대상으로 두 단계 실험을 하였다. 첫 번째 단계는 선호도 데이터에서 사용자가 두개의 서로 다른 이미지 데이터를 평가할 때의 선호도 차이를 측정한다. 식(1)에 의해서 이미지 데이터간의 색 히스토그램 거리를 측정한다. 두 번째 단계는 이미지 데이터를 10개의 동일한 길이 간격을 가진 색 히스토그램 거리의

범위에 따라 군집한다. 각각의 군집에서 평균 선호도 차이를 측정한다.

상관 계수 분석에 대한 실험은 변수간의 관련성을 분석하기 위해 사용하는 방법으로서, 하나의 변수가 다른 변수와 어느 정도의 관련성이 있는지를 분석하는 기법이다. 이미지 데이터에 대한 선호도 차이와 색 히스토그램 거리의 상관관계를 파악하기 위해서 실험을 진행하였다. 실험 결과, 선호도 차이와 색 히스토그램 거리 간에 명확한 정적 상관관계가 있다는 것을 알 수 있다. 이는 각각의 아이템에 대해서 선호도를 평가할 때, 색 정보가 다른 아이템보다는 색 정보가 유사한 아이템을 선호하고 있다는 것을 의미한다. 선호도 차이와 색 히스토그램 거리간의 상관관계에 있어서 강하게 또는 약하게 나타나는 개인적인 차가 존재하겠지만, 이는 10번의 반복적인 실험에 의한 200명 사용자의 평균값으로 나타낸 실험 결과이다.

3.3 이미지 기반 필터링에서 예측

이미지 데이터에 대한 선호도와 색 히스토그램 거리의 정적 상관관계를 이용하는 이미지 기반 필터링에 대해서 기술한다. 사용자 u 가 이전에 평가했던 모든 이미지에 대해 색 히스토그램 거리를 기반으로 $Group_j(i)$ 에 군집한다. [표 1]은 이미지 기반 필터링에서 군집을 위한 색 히스토그램 거리의 범위를 나타낸다.

표 1. 색 히스토그램 거리의 범위

군집	범위	군집	범위
Group ₁	0.0 < $Cdis$ ≤ 0.5	Group ₆	4.0 < $Cdis$ ≤ 4.5
Group ₂	0.5 < $Cdis$ ≤ 1.0	Group ₇	4.5 < $Cdis$ ≤ 5.0
Group ₃	1.5 < $Cdis$ ≤ 2.0	Group ₈	5.0 < $Cdis$ ≤ 5.5
Group ₄	2.0 < $Cdis$ ≤ 2.5	Group ₉	5.5 < $Cdis$ ≤ 6.0
Group ₅	3.0 < $Cdis$ ≤ 3.5	Group ₁₀	6.5 < $Cdis$ ≤ 7.0

색 히스토그램 거리의 범위에 따른 이미지의 군집 중 하나를 군집 $Group_j(i)$ 이라고 할 경우 군집 j 는 식(2)로 표현한다. 모든 이미지는 색 히스토그램 거리에 의해서 10개의 군집 중에 하나의 군집에 속하게 되고 각각의

군집은 서로 유사한 색 히스토그램 거리를 가지고 있는 이미지이다. 식(2)에서 군집 j 는 이미지 데이터로 구성된 특징 벡터이다. 군집 j 에 속한 이미지의 수는 m 이며, $Image_m^j$ 는 군집 j 에 속한 이미지라고 정의한다.

$$Group_j(i) = \{Image_1^j, Image_2^j, \dots, Image_m^j\} \quad (2)$$

군집에 속한 이미지는 색 히스토그램 거리로부터 군집된 유사한 이미지의 빈도에 따라 가중치가 부여된다. 각각의 가중치는 모든 군집내의 이미지의 총 수에 대한 각 이미지의 수의 비율이다.

이미지 데이터에 대한 선호도 차이와 색 히스토그램 거리의 정적 상관관계를 이미지 기반 필터링에 결합하기 위하여 선형적인 예측 모델을 식(3)과 같이 정의한다. $P^{image}(u, i)$ 는 사용자 u 를 위한 이미지 i 에 대한 이미지 기반 필터링에서 예측을 나타낸다. 예측은 군집 j 의 평균 선호도의 가중치가 부여된 합으로 계산된다. $r_{u,i}$ 는 이미지 i 에 대해서 사용자 u 의 선호도를 나타내고, $Group_j(i)$ 는 군집 j 로 색 히스토그램 거리가 동일한 이미지 데이터의 집합이다. $WGroup_j(i)$ 는 군집 j 의 각 이미지에 가중치가 부여된 군집이다.

$$P^{image}(u, i) = \sum_{i \in Group_j(i), j \in \{1, 10\}} r_{u,i} \cdot \frac{1}{|WGroup_j(i)|} \quad (3)$$

4. 성능평가

4.1 실험 환경 및 실험 데이터

본 논문에서 제안한 유비쿼터스 컴퓨팅에서 이미지 기반 필터링을 이용한 개인화 아이템 추천 기법을 MS-Visual Studio C++ 6.0으로 구현하였으며, 실제 실험 환경은 PentiumIII 450MHz, 512MB RAM 환경에서 수행하였다. 실험 데이터로는 컴팩 연구소에서 18개월 동안 협력적 필터링 알고리즘을 연구하기 위해서 아이템에 대한 선호도를 조사한 MovieLens 데이터를 사용한다[6]. 이는 아이템에 관한 평가 데이터가 0, 0.2, 0.4,

0.6, 0.8, 1의 6단계의 수치로 표현되어 있다. 여기서 6단계는 입력 허용 범위를 나타낸다. 본 논문에서는 유비쿼터스 컴퓨팅에서 이미지 기반 필터링을 이용한 개인화 아이템 추천의 예측 정확도 향상을 위한 실험을 위해서 MovieLens 데이터를 전처리하여 사용한다. 계산상의 편의와 메모리 절약을 위해 0-1까지 0.2간격으로 표현된 데이터에 5를 곱하여 0-5까지 1간격으로 변환하여 사용하였다. 최소 100회 이상 선호도를 평가한 사용자 4,798명을 추출하여 이 가운데 1,000명을 기존 사용자 군으로 두고 나머지 사용자 중에 무작위로 테스트 사용자 100명을 선택하여 총 1,628개의 아이템 중 테스트 사용자가 선호도를 평가한 임의의 10개 아이템에 대해서 선호도를 예측하고 실제 선호도와 비교, 평가하였다.

4.2. 분석 및 성능 평가

추천의 성능을 평가하기 위한 방법으로 MAE(Mean Absolute Error)를 사용한다[10][11]. 본 논문에서는 평가를 위해 이미지 기반 필터링을 이용한 아이템 추천 방법(IbF), 텍스트에서 단어의 위치와 출현 빈도를 고려해 확률값을 추정하는 Naive Bayes 분류자를 기반으로 하는 내용 기반 필터링을 이용한 아이템 추천 방법(CbF)[1][8], 기존의 사용자 기반의 협력적 필터링에서 많이 쓰이는 피어슨 상관관계수[10]를 이용한 추천 방법(P_Corr)을 사용자의 수를 변화시키면서 성능을 비교하였다. 그리고 아이템에 대한 평가한 횟수를 증가시켰을 경우 제안된 방법(IbF)을 Pazzani 방법, Billsus 방법 [11], Good 방법[1]의 MAE 성능을 비교 평가를 하였다.

[그림 3]은 사용자 수에 따른 IbF, CbF, P_Corr의 MAE를 나타낸 것이다. 사용자의 수가 많아짐에 따라 IbF의 성능은 높아지나 CbF, P_Corr을 이용한 방법은 큰 차이가 없음을 나타낸다. 예측의 정확도에서는 IbF가 CbF과 P_Corr보다 우세함을 알 수 있다. 이는 기존의 내용기반 필터링에서 단어의 빈도와 출현 횟수만으로 하는 피상적인 내용 분석의 단점을 해결하였고 관심을 가지는 이미지 데이터로부터 특징을 포함시켜 추천에 반영하였기 때문에 MAE의 결과 성능이 뛰어난 것을 알 수 있다. [그림 4]에서 초기 평가 문제를 갖는 Billsus

방법은 평가의 수가 작을 경우 MAE 값은 크나, 평가의 수가 많아짐에 따라 MAE의 결과가 서서히 낮아짐을 보인다. 반면, 회박성 문제를 갖는 Good 방법은 평가의 수가 증가할지라도 MAE의 결과가 크게 낮아지지 않음을 보인다. Pazzani 방법과 제안한 IbF는 평가의 수가 증가할수록 점차 MAE의 결과가 낮아짐을 보이며, 특히 IbF는 평가의 수가 증가할수록 상당히 낮아짐을 보인다. 결론적으로 IbF 방법이 추천의 정확도가 가장 높은 것이 MAE로 증명되었다.

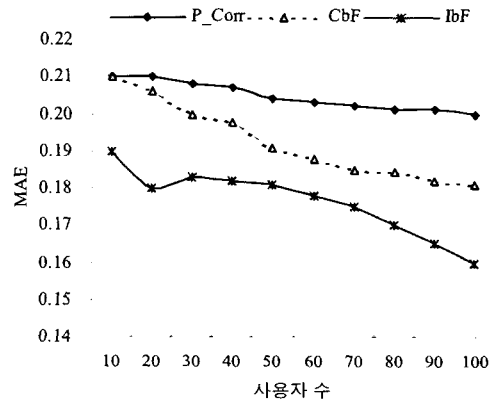


그림 3. 사용자 수에 따른 MAE

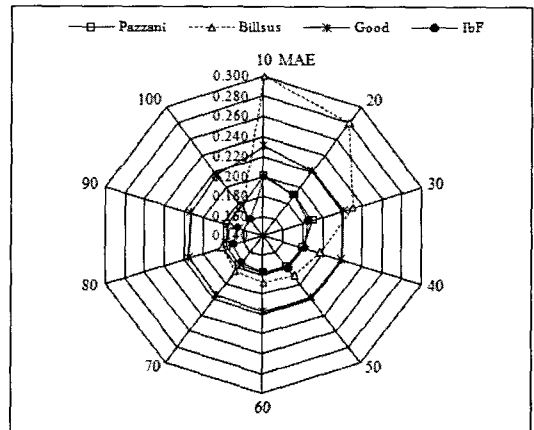


그림 4. n번째 평가에서의 MAE 평가

5. 결론

협력적 필터링 기반의 추천 시스템은 사용자의 관심이나 선호도에 맞는 아이템에 대한 예측은 잘 수행할 수 있으나 특정 아이템의 내용 분석에 대한 사용자의 정보 요구를 해결하는 것은 효율적이지 못하다. 본 논문에서는 피상적인 내용 분석이라는 단점을 개선하기 위하여 사용자가 관심을 가지는 이미지 데이터로부터 특징을 추출하는 이미지 기반 필터링을 이용한 개인화 아이템 추천을 제안한다. 제안한 방법을 MovieLens 데이터에 적용하여 실험 및 평가를 실시하였는데, 내용 기반 필터링을 이용한 추천 방법과 사용자 기반의 협력적 필터링을 이용한 추천 방법과의 비교 실험을 통해 성능을 평가하였다. 그리고 Pazzani 방법, Billsus 방법, Good 방법과도 성능 평가를 하였다. 실험 결과, 제안한 방법이 기존의 다른 방법보다 우수함을 확인하였다. 향후 연구로는 선호도 차이와 색 히스토그램 거리의 정적 상관관계를 적용한 이미지 기반 필터링을 제안하였으나 다른 속성과의 상관관계에 관한 연구가 진행되어야 할 것이다.

참고 문헌

[1] B. Sarwar, G. Karypis, J. Konstanb, "Analysis of Recommendation Algorithms for E-Commerce," The ACM E-Commerce 2000 Conf., 2000.

[2] 고수정, 이정현, "Apriori 알고리즘에 의한 연관 단어 지식 베이스에 기반한 가중치가 부여된 베이즈안 자동 문서 분류", 한국멀티미디어논문지, 제4권, 제2호, pp.171-181, 2001.

[3] K. Y. Jung, "Content-based Image Filtering for Recommendations," LNAI 4203, pp.312-321, Springer-Verlag, 2006.

[4] 광노현, 정진완, "Wavelet을 이용한 이미지 저장 및 검색 시스템", In Proc. of the Korean Database Conference 2005, pp.243-249, 2005.

[5] 최기호, 문의정, 염성주, 김우생, 이원규, "내용을 기반으로 한 이미지 검색 데이터베이스 시스템",

정보과학회지, 제13권, 제1호, pp.8-18, 1995.

[6] <http://www.cs.umn.edu/Research/GroupLens/>

[7] C. Faloutsos, W. Equitz, M. Flickner, W. Niblack, D. Petkovic, and R. Barber, "Efficient and Effective Querying by Image Content," Jour. of Intelligent Information Systems, Vol.3, pp.231-262, 1994.

[8] N. Beckmann, H. P. Kriegel, R. Schneider, and B. Seeger, "The R*-tree: An Efficient and Robust Access Method for Points and Rectangles," In Proc. of ACM SIGMOD Int. Conf. on Management of Data, pp.322-331, 1990.

[9] 한정운, 김병곤, 이재호, 임해철, "이미지 내용 기반 검색을 위한 이미지 타일 평균 RGB 방법", 한국정보과학회: 학술대회지, pp.296-298, 1999.

[10] A. Kohrs and B. Merialdo, "Improving Collaborative Filtering with Multimedia Indexing Techniques to Create User-Adapting Web Sites," In Proc. of the ACM Int. Conf. on Multi-Media, pp.27-36, 1999.

[11] M. J. Pazzani, A Framework for Collaborative, Content-based and Demographic Filtering, Artificial Intelligence Review, pp.393-408, 1999.

저자 소개

정 경 용(Kyung-Yong Chung)

정회원



- 2000년 2월 : 인하대학교 전자계산공학과(공학사)
- 2002년 2월 : 인하대학교 컴퓨터정보공학과(공학석사)
- 2005년 8월 : 인하대학교 컴퓨터정보공학과(공학박사)

• 2006년 3월 ~ 현재 : 상지대학교 컴퓨터정보공학부 교수

<관심분야> : 데이터마이닝, 지능시스템, 감성공학