

동적 사용자 모델을 이용한 개인화된 문맥광고*

강영길^o, 김성민, 이수원
송실대학교 대학원 컴퓨터학과

{dudrfl34^o, mabak}@mining.ssu.ac.kr, swlee@ssu.ac.kr

Personalized Contextual Advertisement Using a Dynamic User Model

Youngkil Kang^o, Seongmin Kim, Soowon Lee
Dept. of Computing, Graduate School, Soongsil University

요 약

문맥광고 또는 컨텍스트 기반 광고란 사용자들이 선택한 웹 콘텐츠 내용을 기반으로 하여 연관성이 있는 광고를 자동으로 선택하여 사용자에게 제공하는 광고기법이다. 즉, 웹 사이트를 방문하는 고객을 타겟으로 하여 그들이 찾고자 하는 것과 관련된 광고를 내보냄으로써 효과적인 광고가 이루어지도록 하는 것이다. 그러나 기존의 문맥광고는 사용자가 관심을 가지는 키워드가 아닌 광고주가 선택한 키워드를 중심으로 광고 내용을 선택하기 때문에 사용자의 실제적인 관심이 반영되지 않아 광고의 효과가 떨어지는 문제점을 가지고 있다.

이러한 문제를 해결하기 위해 본 논문에서는 사용자가 웹 콘텐츠를 선택할 때 마다 사용자의 선호도를 동적으로 학습하고, 학습된 선호도를 문맥광고에 활용하는 개인화된 문맥광고를 제안한다. 실험을 위해서 제안한 방법으로 광고를 생성해서 보여주는 웹 브라우저를 구현하여 기존의 문맥광고와 개인화된 문맥광고에 대한 사용자의 평가를 비교하였다. 실험 결과 본 논문에서 제안한 개인화된 문맥광고가 '콘텐츠의 내용과의 연관성', '사용자의 클릭여부' 등의 항목에서 기존의 문맥광고에 비해 우수하다는 결과를 얻을 수 있었다.

1. 서 론

인터넷 사용자의 증가와 더불어 인터넷 광고 시장은 급격하게 커지고 있다. 미국의 인터넷 광고 시장의 경우, 2004년 96억 달러에서 2005년에 125억 달러로 30% 이상 성장하여 급격하게 성장하고 있다. 특히, 한국의 경우에는 지난 5년 동안 인터넷 광고비가 300% 가량 성장하여 2006년 인터넷 광고의 시장규모는 약 9000억 원에 달할 것으로 추정된다[1].

인터넷 광고의 형태는 팝업(Pop-up), 배너(Banner), 동영상, 키워드 등 여러 가지 형태로 나타난다. 이 중에서도 큰 비중을 차지하는 것은 키워드 광고로서 전체 규모의 약 40%를 차지하고 있다. 키워드 광고에서 최근에 각광을 받고 있는 것은 문맥광고로서, 웹 사용자가 선택한 웹 콘텐츠의 내용에 포함된 키워드를 분석하여 연관성이 높은 광고를 싣는 광고기법이다. 이러한 문맥광고는 콘텐츠 내용에 자연스럽게 광고가 포함되어 이용자들이 갖는 광고에 대한 거부감을 줄일 수 있는 장점을 가진다. 또한, 관련 콘텐츠에 관심을 가지고 읽는 사람들에게 대한 타겟 마케팅으로 적극 활용할 수 있기 때문에 앞으로 높은 성장이 이루어질 것으로 기대된다.

문맥광고는 블로그(Blog), 뉴스(News), 커뮤니티 사이트(Community Site) 등 다양한 웹 콘텐츠에서 서비스가 제공되고 있다. 현재 서비스되는 문맥광고는 웹 콘텐츠에서 등장하는 키워드들 중에서 광고주가 구매한 키워드

들을 바탕으로 광고가 이루어진다. 따라서 광고의 내용과 웹 콘텐츠와의 연관성이 낮으며 사용자의 실제적인 관심이 반영되지 않아 광고의 효과가 떨어지게 되는 문제점을 가질 수 있다.

이러한 문제를 해결하기 위해 본 논문에서는 동적 사용자 선호도 모델을 이용한 개인화된 문맥광고를 제안한다. 제안 방법에서는 우선, 사용자가 웹 콘텐츠를 선택할 때 마다 사용자의 선호도를 동적으로 학습한다. 다음으로, 학습의 결과인 키워드에 대한 사용자 선호도와 웹 콘텐츠에서 추출된 키워드들이 갖는 문서내의 중요도를 이용하여 광고 키워드 적합도를 계산한다. 마지막으로, 가장 높은 광고 키워드 적합도를 가지는 키워드에 관련된 광고를 서비스함으로써 개인화된 문맥광고를 구현한다.

2. 관련 연구

사용자 선호도 모델에 대한 연구는 사용자의 선호도와 콘텐츠의 내용과의 연관성을 바탕으로 추천이 이루어지는 내용 기반 추천 분야에서 활발하게 이루어지고 있다. [2]에서는 TV 프로그램의 장르, 세부장르, 채널과 같은 구조적인 내용에 대한 선호도 모델을 제안하였다. 또한 [3]에서는 TV 프로그램의 구조적인 정보뿐만 아니라 제목과 줄거리에서 추출된 키워드들에 대한 선호도가 반영된 사용자 모델을 제안하였다.

그러나 위와 같이 콘텐츠의 내용에 대한 선호도 모델을 생성하는 방법은 사용자의 누적된 히스토리가 있어야만 생성할 수 있다는 단점을 가진다. 즉, 사용자 히스토

* 본 연구는 서울시 산학연 협력산업의 지원으로 이루어졌습니다.

리가 없는 시스템의 사용 초기에는 사용자 모델이 학습되어 있지 않기 때문에 추천이 이루어지지 않거나 성능이 떨어지게 되는 Cold Start Problem이 나타나게 된다.

이러한 내용기반 추천의 문제를 해결하기 위한 대안으로 선호도가 유사한 다른 사용자 모델을 활용하는 협력적 추천(Collaborative Filtering)[4]이나 특정 집단의 선호도를 나타내는 Stereotype의 활용[5]등의 방법들이 제시되고 있다.

3. 개인화된 문맥광고

본 논문에서 제안하는 개인화된 문맥광고 기법에서 사용자 모델은 웹 콘텐츠를 사용하는 사용자의 대다수가 익명사용자라는 전제로 세션이 생성되어 종료할 때 까지 유지된다. 사용자 모델의 학습은 사용자가 브라우저를 통해 웹 콘텐츠를 선택할 때 마다 동적으로 이루어지며, 학습된 선호도는 키워드의 문서에 대한 가중치와 함께 키워드의 광고 적합도 계산에 활용된다. 제안한 방법은 텍스트 위주의 웹 콘텐츠를 적용대상으로 하며, 본 논문에서는 웹 뉴스로 도메인을 한정한다.

3.1 사용자 모델

사용자 모델은 [표 1]과 같이 사용자가 본 웹 뉴스에 대하여 학습한 “선호 키워드 벡터”와 가장 최근에 선택한 웹 뉴스를 나타내는 키워드 벡터로 구성된다.

[표 1] 사용자 모델의 구성

선호 키워드 벡터	<키워드1, 선호도>, <키워드2, 선호도>
최근 콘텐츠 키워드 벡터	<키워드1, 빈도>, <키워드2, 빈도>

제안한 사용자 모델은 웹 뉴스를 이용하는 익명 사용자를 대상으로 한다. 여기서의 익명 사용자는 웹 사이트에 로그인을 하지 않은 상태에서 웹 뉴스를 이용하는 사용자들을 가리킨다. 현실적으로 웹을 이용하는 대다수의 사용자가 익명 사용자이기 때문에 사용자의 선호도를 학습하여 활용할 수 있는 기간은 하나의 세션이 생성되어 유지되는 동안이다. 따라서 사용자 모델의 생성 및 종료는 세션의 생성 및 종료와 같은 시간에 이루어진다.

과거의 선호도를 반영하지 않는 익명 사용자 선호도 모델은 계속 변화하는 사용자의 현재 관심사를 더욱 잘 반영한다. 실제로 사용자가 웹 뉴스를 선택할 때는 자신이 좋아하는 뉴스를 찾아서 선택하기보다는 웹 뉴스 사이트에 표시된 현재 이슈가 되는 뉴스를 선택하는 경우가 많기 때문에 이전에 학습된 과거의 관심사를 반영할 경우 현재의 관심사의 반영이 적어지는 경우가 발생한다. 반면에 제안한 사용자 모델은 웹 뉴스를 선택하는 현재 시점에서의 선호도만을 반영하게 된다.

3.2 사용자 모델의 학습

초기 사용자 모델의 생성은 사용자가 웹 브라우저를

실행하여 최초로 선택한 웹 뉴스의 키워드들을 선호 키워드에 추가하는 것으로 이루어진다. 이 때 기본 선호도는 0.5로 설정된다. 그리고 현재 선택한 웹 뉴스의 키워드를 최근 콘텐츠 키워드로 단어 출현 빈도(TF)와 함께 사용자 모델의 최근 콘텐츠 키워드 벡터에 저장한다.

사용자 모델이 생성된 후에는 사용자가 새로운 웹 뉴스를 선택할 때마다 사용자 모델의 학습이 이루어진다 [식 1]. 우선 현재 선택한 뉴스의 키워드 벡터(V_c)와 최근 콘텐츠 키워드 벡터(V_r)간의 코사인 유사도로 문서간의 유사도를 구한다. 이 값이 크다는 것은 연관성을 가지는 웹 콘텐츠를 사용자가 계속 선택하고 있다는 것으로서 현재 선택된 웹 콘텐츠가 높은 선호도를 가지는 것을 나타낸다. 따라서 현재 선택한 웹 뉴스의 키워드가 선호 키워드 벡터에 존재할 경우에는 [식 1]에 의해 그 키워드의 선호도에 미리 계산된 코사인 유사도 값을 더하여 선호도를 높인다. 만약, 선호 키워드 벡터에 존재하지 않는 키워드의 경우에는 초기 사용자 생성 시와 마찬가지로 키워드를 선호 키워드 벡터에 추가하고 초기 값을 0.5로 설정한다.

$$preference_k^u = preference_k^u + CosineSim(V_c, V_r)$$

, for all k both in V_c and V_r

u : 사용자, k : 키워드

V_c : 현재 선택된 콘텐츠의 키워드 벡터

V_r : 최근에 선택된 콘텐츠의 키워드 벡터

V_p : 선호 키워드 벡터

$preference_k^u$: 사용자 u의 키워드 k에 대한 선호도

$preference_k^u$: 갱신된 사용자 u의 키워드 k에 대한 선호도

[식 1] 키워드 선호도 업데이트

선호 키워드 벡터의 갱신이 끝나면 [식 2]에 의해 선호도가 가장 높은 키워드의 선호도 값으로 모든 키워드의 선호도를 나누어 선호도를 0과 1사이로 정규화한다. 정규화에 의해 선호도가 0.1보다 낮아진 키워드는 선호 키워드 벡터에서 삭제한다. 그리고 현재 보고 있는 웹 뉴스의 키워드를 최근 콘텐츠 키워드로 단어 출현 빈도(TF)와 같이 저장하여 다음의 선호도 학습에 활용한다.

$$preference_k^u = \frac{preference_k^u}{Max[preference_{k'}^u, \text{for all } k' \text{ in } V_p]}$$

u : 사용자, k : 키워드

V_p : 선호 키워드 벡터

$preference_k^u$: 사용자 u의 키워드 k에 대한 선호도

$preference_k^u$: 갱신된 사용자 u의 키워드 k에 대한 선호도

[식 2] 키워드 선호도 정규화

3.3 광고 키워드의 선택

웹 뉴스에 대한 광고 키워드 추출은 두 가지 경우로 나누어진다. 첫 번째는 웹 뉴스에서 추출한 키워드가 사용자 모델에 없는 경우이고, 두 번째는 웹 뉴스에서 추출한 키워드가 사용자 모델에 있는 경우이다. 전자의 경우는 초기에 사용자 모델이 충분히 학습되지 않았을 경우에 발생하게 되는데, 이런 경우에는 키워드의 단어 출현 빈도(TF)를 바탕으로 광고 키워드 후보를 추출하여 광고 내용을 구성한다.

후자의 경우는 우선 뉴스에 출현하는 키워드의 단어 출현 빈도(TF)를 [식 3]에 의해 단어 출현 빈도의 최대값으로 정규화 하여 0과 1사이의 값을 가지는 뉴스내의 키워드 가중치를 구한다. 그리고 [식 4]에 의해 사용자 모델의 키워드 선호도와 문서상의 키워드 가중치를 곱하여 키워드의 광고 적합도를 구한다.

$$weight(k,c) = \frac{tf(k,c)}{Max\{tf(k',c), \text{ for all } k' \text{ in } V_c\}}$$

- k : 키워드, c : 현재 선택된 콘텐츠
- V_c : 현재 선택된 콘텐츠의 키워드 벡터
- $tf(k,c)$: 콘텐츠 c 에서 키워드 k의 단어출현빈도
- $weight(k,c)$: 콘텐츠 c에서 키워드 k의 키워드가중치

[식 3] 키워드 가중치 계산

$$adSuitability(k,c,u) = weight(k,c) \times preference_k^u, \text{ for all } k \text{ both in } V_c \text{ and } V_p$$

- k : 키워드, c : 현재 선택된 콘텐츠
- V_c : 현재 선택된 콘텐츠의 키워드 벡터
- V_p : 선호 키워드 벡터
- $weight(k,c)$: 키워드 벡터 V_c 에서 키워드 k의 키워드 가중치
- $preference_k^u$: 사용자 u의 키워드 k에 대한 선호도
- $adSuitability(k,c,u)$: 사용자 u와 콘텐츠 c에 대한 키워드 k의 광고 적합도

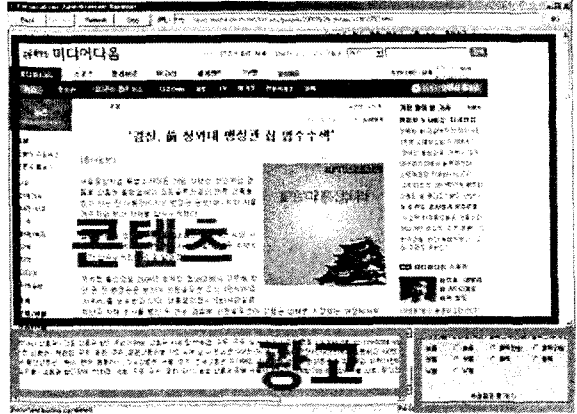
[식 4] 키워드의 광고 적합도 계산

4. 실험 및 평가

4.1 실험 시스템

본 논문에서는 실험을 위해 [그림 1]과 같이 사용자가 선택한 웹 뉴스에 대해 제안된 방법으로 광고를 생성하여 보여주는 웹 브라우저를 구현하였다. 이 웹 브라우저

는 중앙의 콘텐츠 영역에서 사용자가 웹 뉴스를 선택하면, 제안한 방법을 통해 광고를 생성하여 아래의 광고 영역으로 출력한다. 광고 영역의 옆에는 사용자의 평가를 수집하기 위한 설문 영역이 결합되어 있다.



[그림 1] 개인화 문맥광고를 위한 웹 브라우저

구현된 웹브라우저는 사용자가 웹 뉴스를 선택할 때마다 제안된 방법을 통해 사용자 모델을 학습하고, 현재 선택된 뉴스에서 가장 적합한 광고 키워드를 선정한다. 선정된 광고 키워드와 관련된 광고를 구하기 위해서 실험 시스템에서는 뉴스를 제공하고 있는 포털사이트의 검색엔진을 활용하였다. 즉, 광고 키워드에 대한 검색 결과 중에서 광고 내용을 추출하여 사용자에게 제시한다.

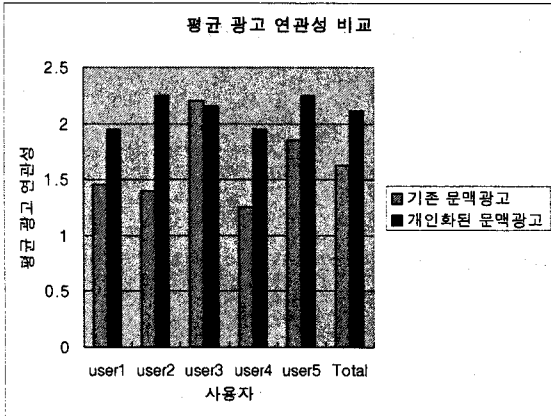
이러한 방법으로 광고를 생성한 이유는 포털사이트에서 제공하는 문맥광고에 어떤 광고주들이 있는지를 알 수가 없기 때문에 동일한 광고주들을 대상으로 하는 비교 실험이 불가능하기 때문이다. 때문에 일반 사이트를 광고하는 것 보다는, 광고의 의사가 있는 사이트를 광고하는 것이 의미가 있을 것이라고 판단하여 실험시스템의 광고주를 키워드 검색광고에 등록된 광고주들로 정하였다.

4.2 평가 방법 및 결과

본 논문에서 제안되는 방법을 평가하기 위해 비교대상으로 현재 문맥광고 서비스를 제공하고 있는 포털사이트 하나를 선정하였다. 그리고 포털사이트에서 제공하는 웹 뉴스들 중에서 5명의 사용자가 선택한 20개의 뉴스에 포함된 광고와 실험 시스템에서 보여주는 광고를 대상으로 지정된 평가 항목들에 대해 평가하였다. 평가 항목은 포털사이트에서 제공하는 광고와 뉴스 간의 연관성, 실험시스템에서 제공하는 광고와 뉴스 간의 연관성, 포털사이트에서 제공하는 광고의 클릭 여부, 실험 시스템에서 제공하는 광고의 클릭 여부의 4가지로 구성된다. 이 중에서 연관성 관련 항목의 값은 '높음', '보통', '낮음'의 3가지로 각각 3,2,1의 정수 값을 부여하였고, 클릭 여부는 '클릭', '클릭안함'의 2가지로 각각 1,0의 정수 값을

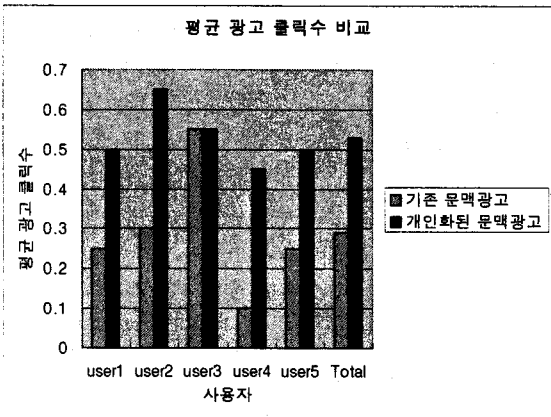
부여하였다.

아래의 [그림 2]는 사용자가 선택한 뉴스와 노출된 광고와의 연관성을 평가한 값의 평균을 비교한 그림이다. 그림에서와 같이 전체적으로 시스템에서 제공한 광고가 사용자가 선택한 뉴스의 내용과 연관성이 높다는 결과를 확인할 수 있었다.



[그림 2] 콘텐츠 내용과 광고와의 연관성 비교

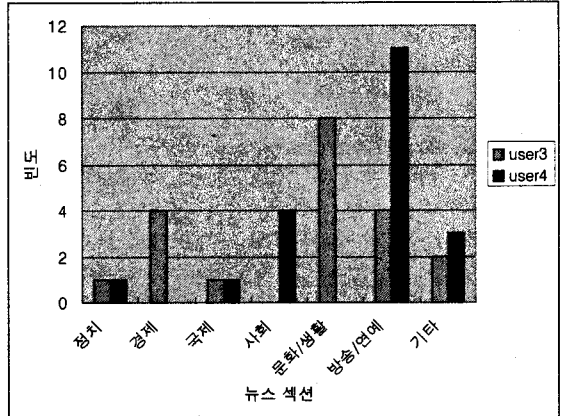
또한, [그림 3]의 광고 클릭 여부의 비교에서는 실험 시스템에서 제공하는 광고에 대한 클릭의 평균이 약 2배 이상 높게 나오는 결과를 얻을 수 있었다. 특히, user4의 경우에는 시스템에서 제공하는 광고의 클릭여부가 포털 사이트에서 제공하는 것에 비해 4배 이상 높게 나왔다. 반면에 user3의 경우에는 포털사이트에서 제공하는 광고와 실험시스템에서 제공하는 광고의 평가가 비슷한 것을 확인할 수 있었다.



[그림 3] 광고에 대한 사용자의 클릭 여부

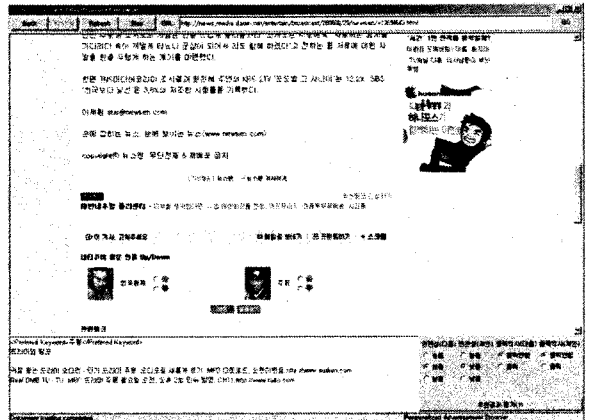
이렇게 사용자별로 다른 평가가 발생하는 이유는 사용자들이 주로 선택한 뉴스의 섹션들에서 차이가 나기 때문으로 추측된다. [그림 4]는 위의 실험 결과에서 가장

큰 차이를 보이는 user3과 user4의 뉴스 선택 패턴을 비교한 것이다. 그림을 보면, user4가 '방송/연예' 섹션의 뉴스를 집중적으로 선택한 반면에 user3은 다양한 섹션의 뉴스를 선택하였을 뿐 아니라, user4가 전혀 선택하지 않은 '문화/생활' 섹션의 뉴스를 가장 많이 선택한 것을 알 수 있다.



[그림 4] 사용자별 뉴스 선택 패턴 비교

특히, '방송/연예' 뉴스에 포함되는 광고는 뉴스와의 연관성이 높지 않았기 때문에 실험시스템에서 제공하는 광고를 선호하는 user4의 특성이 나타난 것으로 추측된다. 아래의 [그림 5]는 연관성이 낮은 광고가 보여 지는 '방송/연예' 뉴스의 예로서 '주몽'이라는 드라마에 대한 뉴스이다. 이 뉴스에 포함된 광고는 화장품 쇼핑물에 대한 광고로서 뉴스의 내용과는 연관성이 매우 낮은 반면에 실험 시스템에서는 드라마를 다시 볼 수 있는 사이트를 광고하여 user4의 클릭을 이끌어 냈다.



[그림 5] 연관성이 낮은 광고의 예

5. 결론 및 향후 연구

기존의 문맥광고는 개인의 선호도를 고려하지 않기 때문에 광고의 효과가 떨어지는 문제를 가진다. 이러한 문제를 해결하기 위해 본 논문에서는 사용자의 선호도를 동적으로 학습하여 사용자가 보는 뉴스의 내용과의 비교를 통해 최적의 광고 키워드를 추출하여 광고를 제공하는 방법을 제안하였다.

제안한 방법의 실험을 위해 현재 포털사이트에서 제공되는 문맥광고와 실험 시스템에서 제공되는 광고에 대한 사용자의 평가를 수집하였다. 그 결과, 뉴스와의 연관성과 클릭 여부의 비교에서 본 논문에서 제안한 방법이 기존의 문맥광고에 비해 우수하다는 결과를 얻을 수 있었다.

그러나 포털 사이트에서 제공되는 문맥광고의 광고주를 알 수 없기 때문에 정확한 비교평가가 이루어지지 않는 못했으며, 사용자의 뉴스 구독 형태에 따라 평가가 달라지는 문제를 가지고 있다. 따라서 동일한 광고주를 대상으로 하여 기존의 문맥광고와 논문에서 제안한 방법에 의한 광고의 비교를 진행할 필요가 있으며, 뉴스의 선택과 광고 내용과의 연관성에 관련한 연구를 통해 사용자의 다양한 뉴스 구독 형태를 반영할 수 있도록 사용자 모델을 개선할 필요가 있다. 또한, 초기에 사용자 모델의 내용이 없을 경우를 위한 유사 사용자 모델의 적용에 대한 연구도 필요하다.

6. 참조 문헌

- [1] 한국 인터넷 마케팅 협회, <http://www.imck.or.kr>
- [2] 유상원, 이홍래, 이형동, 김형주, TV 프로그램을 위한 내용기반 추천시스템, 정보과학회논문지:컴퓨팅의 실제, 제 9권, 제 6호, 2003.
- [3] 윤현호, 강영길, 이성진, 이수원, 김현, 사용자 선호도 기반의 TV 프로그램 추천 기법, 한국정보과학회 한국 컴퓨터종합학술대회 2005 VOL.32, 2005.
- [4] Weizhen Dai and Robin Cohen, Dynamic Personalized TV Recommendation System, Workshop on Personalization in Future TV, 2003
- [5] Kaushal Kurapati and Srinivas Gutta, TV Personalization through Stereotypes, Workshop on personalization in future TV, Workshop on Personalization in Future TV, 2002