

멀티미디어 추천시스템을 위한 속성 생성 기법

김형일[†], 엄정국^{**}

요 약

멀티미디어 추천시스템은 사용자의 선호도를 분석하여 멀티미디어 상품을 사용자에게 추천하는 시스템이다. 다양한 추천 기법들에서 가장 널리 사용되는 기법은 협동적 여과 방식이다. 그러나 협동적 여과는 정보 부족 문제와 초기 시작 문제가 존재한다. 선호도 정보가 적게 존재하면 유사 사용자 추출이 어려우며, 이러한 문제는 시스템을 처음 사용하는 새로운 사용자에게 더욱 심각한 문제를 발생시킨다. 본 논문에서는 정보 부족 문제를 해결하고 추천 정확도를 향상시키기 위해 사용자와 상품에 대한 속성 생성 기법을 제안한다. 본 논문에서 제안한 기법은 속성의 분포를 이용하여 추가 속성을 생성하고, 추가 속성을 포함한 변형된 데이터를 이용하여 상품을 추천한다. 여러 실험을 통해 제안된 기법의 효과를 확인하였다.

A Feature Generation Method for Multimedia Recommendation System

Hyungil Kim[†], Jeongkook Eom^{**}

ABSTRACT

Multimedia recommendation systems analyze user preferences and recommend items(multimedia contents) to a user by predicting the user's preference for those items. Among various kinds of recommendation methods, collaborative filtering(CF) has been widely used and successfully applied to practical applications. However, collaborative filtering has two inherent problems: data sparseness and the cold-start problems. If there are few known preferences for a user, it is difficult to find many similar users, and therefore the performance of recommendation is degraded. This problem is more serious when a new user is first using the system. In this paper, we propose a method of generating additional feature of users and items into CF to overcome the difficulties caused by sparseness and improve the accuracy of recommendation. In our method, we first generate additional features by using the probability distribution of feature values, then recommend items by applying collaborative filtering on the modified data to include additional features. Several experimental results that show the effectiveness of the proposed method are also presented.

Key words: Feature Generation Method(속성 생성 기법), Multimedia Recommendation System(멀티미디어 추천시스템), Data Mining(데이터 마이닝), Information Filtering(정보 여과)

1. 서 론

멀티미디어 추천시스템은 사용자의 선호도를 분석하고 상품들에 대한 사용자의 선호도를 예측하여

각 사용자의 취향에 맞는 멀티미디어 상품들(영화, 음악 등)을 추천한다. 멀티미디어 추천시스템에서 사용하는 상품 추천 방식은 크게 내용기반 추천, 인구통계학적 추천, 협동적 여과 추천으로 나뉜다.

※ 교신저자(Corresponding Author): 김형일, 주소: 충남 천안시 쌍용동 456번지 (330-718), 전화: 041)570-1467, FAX: 041)572-6833, E-mail: hkim@kornu.ac.kr
접수일: 2007년 10월 16일, 완료일: 2008년 1월 2일
[†] 정희원, 나사렛대학교 멀티미디어학과 교수

^{**} 정희원, 나사렛대학교 멀티미디어학과 교수
(E-mail: jkeom@kornu.ac.kr)
※ 이 논문은 2007년도 나사렛대학교 학술연구비 지원에 의해 연구되었음

내용기반 추천은 상품을 설명하는 또는 상품을 구성하는 단어들의 집합을 이용하여 상품을 정보화하고 이들 정보를 학습하여 추천에 이용하는 기법이다 [1-4]. 내용기반 추천은 상품의 내용정보를 활용하기 때문에 단어들로 구성된 책, 기사, 메일 등에 적합한 추천 기법이다. 내용기반 방법에서 자주 사용하는 방법은 단순빈도 방식과 상대빈도 방식이다.

단순빈도 방식에서 자주 활용되는 방식은 단어빈도(Term Frequency) 방식으로, 단어빈도 방식은 특정 단어의 출현빈도를 이용하여 상품 특징값을 생성한다. 멀티미디어 상품에 단어빈도와 같은 내용기반 방식을 적용할 때는 상품의 설명문이 활용된다. 일반적으로 설명문에는 배우, 감독, 유형, 줄거리 등이 나타나며, 이와 같은 설명문을 활용하여 상품의 특징 생성에 활용한다.

상대빈도 방식에서 많이 활용되는 방식은 역문헌빈도(Inverse Document Frequency)이다. 역문헌빈도 방식은 문헌빈도가 낮은 단어에 높은 가중치를 부여하는 방식으로, 문헌에 빈번하게 출현하는 단어는 중요도가 낮다는 가정에서 출발한 방식이다. 즉, 모든 상품에 자주 출현하는 단어는 중요도가 떨어진다는 것이 전제가 되는 방식이다.

인구통계학적 추천은 사용자들이 갖는 인구통계학적 정보를 토대로 유사 사용자 집단을 형성한 후 유사 사용자 집단에서 많이 구매한 상품을 추천하는 기법이다[5]. 이 기법의 기본 전제는 인구통계학적 정보가 서로 비슷한 사용자들의 경우에 유사한 성향을 나눌 가능성이 높다는 것이고, 인구통계학적인 정보로는 나이, 성별, 직업, 교육 수준 등이 이용될 수 있다.

영화 데이터를 분석하면 일반적으로 남자들은 액션 유형을 선호하고 여자들은 드라마 유형의 영화를 선호하는 경우가 많다. 이러한 성별 정보에 따라 액션 유형의 영화가 출시되면 여자보다는 남자들에게 새로 출시된 영화를 추천하는 것이 옳은 방법일 것이다. 이와 같이 인구통계학적인 정보를 바탕으로 사용자들의 구매 데이터를 분석하여 새로운 상품이 출시되었을 때 상품 추천에 활용하면 효과적이다. 인구통계학적 정보는 특징값 생성이 쉽고 사용자에게 특별한 정보를 요구하지 않아도 상품을 추천할 수 있다는 장점이 있으나, 개인화된 특성을 나타내기는 어렵다는 단점이 있다.

협동적 여과 추천은 사용자의 상품 구매내역을 다른 사용자들과 비교하여 상품을 추천하는 방법이다 [6-10]. 즉, 사용자들의 상품 구매내역을 이용하여 사용자들 사이의 유사성을 파악하고, 유사도 정보를 기반으로 상품의 선호도를 예측하여 상품을 추천한다. 협동적 여과의 장점은 유사 사용자의 평가를 기반으로 상품을 추천하기 때문에 상품에 대한 내용정보가 없어도 상품 추천이 가능하다는 장점이 있다. 웹에서 사용자의 성향을 파악하기 위해서는 사용자의 목시적 행위 결과인 구매이력만을 활용해야 하는 경우가 빈번히 발생하며, 이와 같은 경우에 내용정보를 활용하지 않는 협동적 여과 기법은 우수한 추천 기법이다. 그리고 협동적 여과 기법은 사용자의 구매내역이나 평가정보만을 활용하여 상품의 선호도를 예측하기 때문에 내용정보가 적은 영화, 음악, 이미지 등과 같은 멀티미디어 상품들에 적합한 추천 기법이다.

협동적 여과 방식은 추천 대상이 되는 사용자와 유사한 성향을 갖는 유사 사용자들을 추출한 후, 유사 사용자들의 상품 평가값을 이용하여 특정 사용자의 상품 선호도를 예측하는 방식을 따른다. 이와 같은 방식을 따르므로 협동적 여과 방식에서는 유사 사용자 추출이 매우 중요하다. 유사 사용자의 수가 많을수록 특정 사용자의 성향 분석은 쉬워지나, 연산 시간이 오래 걸린다는 단점이 있다.

다양한 추천 기법 중에 협동적 여과는 상용화된 추천시스템에 성공적인 적용이 이루어진 기법이다. 그러나 협동적 여과 기법은 정보 부족 문제(sparseness problem)가 있다. 정보 부족 문제란 객체들의 수량에 비해 평가정보가 극히 소량인 경우를 말하며, 이러한 정보 부족은 유사 사용자 추출을 어렵게 하는 문제를 발생시켜 추천 정확도를 크게 저하시킨다. 일반적으로 웹 사이트에 나타난 상품의 종류는 수 백 만 종류에 이르는 경우가 많으며, 적은 상품 종류를 소유한 사이트라 할지라도 수 천 종류에 이르는 경우가 빈번하다. 또한 가입한 회원의 수도 수 백 만에 이르는 경우도 매우 흔하다. 그러나 사용자가 웹 사이트에서 구매한 상품은 극히 소량일 경우가 많으며, 회원으로는 가입되어 있지만 상품을 전혀 구매하지 않는 사용자들도 빈번히 발생한다. 이러한 상황에서 구매 데이터를 형성하면 구매정보가 잘 나타나지 않는 정보 부족 문제가 발생한다. 정보 부족 문제를 해결하기 위해서는 정보를 확장하는 것이 중요한 대안일 것이다.

본 논문에서는 정보 부족 문제를 해결하기 위해 객체 속성의 분포를 이용하여 새로운 속성을 만드는 기법을 제안한다. 제안한 기법은 객체들의 속성을 이용하여 평가 데이터를 분석하고 분석된 정보를 기반으로 추가 속성을 생성한다. 생성된 추가 속성은 기본 데이터와 결합하여 새로운 추가 속성 데이터를 형성한다. 이렇게 형성된 추가 속성 데이터는 멀티미디어 추천시스템에 적용되어 상품 추천에 이용된다. 본 논문에서 제안한 추가 속성을 생성하는 기법을 속성 생성 기법이라 한다. 속성 생성 기법은 평가 데이터의 속성 분포를 이용하여 추가 속성을 생성하기 때문에 상품들의 연관성을 잘 나타내어 추천 정확도를 향상시킨다.

본 논문의 2장에서 추천 기법에서 가장 널리 사용되는 협동적 여과를 중심으로 관련 연구를 기술하고, 3장에서는 본 논문에서 제안한 속성 생성 기법에 대해 설명한다. 4장에서는 제안한 기법의 실험 결과를 설명하고, 5장에서 결론과 향후 연구에 대해 설명한다.

2. 관련 연구

내용기반 추천은 두 가지 추천 방식을 주로 사용한다. 하나의 방식은 상품간의 특성에 대한 유사성을 이용하는 상품 상관관계 추천 방식이고, 다른 하나는 사용자 신상정보와 상품 특성의 유사성을 이용하는 속성기반 추천 방식이다. 두 가지 추천 방식 중에서 속성기반 추천 방식을 많이 활용하며, 이 추천 방식은 추천하고자 하는 상품의 자체 내용(예 : 설명문)과 사용자의 신상정보간의 유사성을 고려하여 상품을 추천하는 방식이다.

내용기반 추천시스템의 장점은 신문기사나 논문과 같이 내용정보가 풍부한 상품에 적용했을 경우, 비교적 정확한 추천을 할 수 있다는 것이다. 그러나 영화나 음반과 같은 멀티미디어 정보처럼 내용정보가 빈약한 상품의 경우, 상품의 특징을 적절히 표현할 수 없기 때문에 올바른 상품 추천이 어렵다는 단점이 있다. 또한, 사용자의 신상정보가 특정 내용에 대해 높은 가중치를 가지고 있을 경우 특정 내용으로 사용자의 취향이 고정되는 일이 발생할 수 있다. 예를 들어, 특정 사용자의 구매이력에서 '모험'이라는 영화 유형이 '역사'나 '전쟁'에 비해 높은 가중치를 가질 경우 추천시스템은 '모험'에 대한 상품만을 추천

하게 된다. 결국 사용자의 취향이 다른 유형에 있음에도 불구하고 내용기반에서는 이를 정확히 판단할 수 없다.

인구통계학적 추천이란 사용자의 나이, 성별, 직업 등과 같은 인구통계학적 정보를 상품 추천에 활용하는 방법이다. 인구통계학적 추천은 인구통계학적 정보를 기초로 사용자 집단을 형성한 후, 사용자 집단의 상품 선호도를 분석하여 상품을 추천하는 기법이다.

예를 들면, 나이가 '30대', 직업이 '교사'인 A 사용자가 '책1'을 구매하였을 경우, 나이가 '30대', 직업이 '교사'인 B 사용자에게 '책1'을 추천하는 것을 말한다. 이 때 B 사용자에게 A 사용자가 구매한 상품을 추천하는 이유는 B 사용자와 A 사용자와의 인구통계학적 정보인 직업과 나이가 동일함으로 같은 성향을 갖는 사용자라고 가정하기 때문이다.

인구통계학적 추천 방법은 나이, 성별, 직업 등 사용자가 가지는 기본적인 정보가 유사한 사용자들은 비슷한 구매 성향을 가진다는 전제를 기본으로 한다. 인구통계학적 추천은 상품에 대한 사용자의 평가가 없는 초기 사용자의 경우 활용성이 높다는 장점이 있지만, 사용자 집단의 성향을 이용하기 때문에 개인화된 특성을 표현할 수 없어 데이터 이용을 극대화하지 못한다.

협동적 추천은 상품을 추천하고자 하는 사용자와 비슷한 구매 성향을 나타내는 유사 사용자들의 구매 이력을 이용하여 상품을 추천하는 방식을 취한다. 협동적 여과 기법은 상품의 내용 자체를 분석하는 것이 아니기 때문에 영화, 음악, 영상 등의 멀티미디어 상품들도 추천할 수 있다는 장점이 있다. 또한 유사한 사용자들의 성향을 분석하여 상품을 추천하기 때문에 다른 추천 기법에 비해 매우 높은 추천 정확도를 나타낸다.

협동적 추천의 절차는 다음과 같다. 먼저 사용자들 사이의 유사도를 계산하기 위해 사용자의 평가정보를 읽어 신상정보 벡터를 생성하고, 사용자 유사도 계산에는 일반적으로 피어슨 상관계수(pearson correlation coefficient)를 사용한다. 피어슨 상관계수는 공통된 상품에 대해 현 사용자의 평가와 타 사용자의 평가를 비교하는 방식으로 구성되어 있다. 사용자의 평가는 사용자가 평균적으로 평가한 점수와 상품의 평가점수의 차를 구하여 상품에 대한 선호도

를 예측한다. 사용자의 평가 평균에 대한 상대적 평가를 사용하는 이유는 사용자 평가점수 범위가 서로 다르기 때문이다.

예를 들어, 특정한 상품을 평가할 때 평가값의 범위를 0부터 10사이의 값을 사용한다고 가정하자. 사용자 A는 상품 I가 매우 마음에 들어 평가값을 5점을 주었고, 사용자 B는 상품 I가 마음에 들지 않아 5점의 평가값을 주었다. 이러한 경우 5점은 두 가지 선호도를 나타낸다. 사용자 A의 5점은 좋다는 의미의 선호도이고, 사용자 B의 5점은 나쁘다는 의미의 선호도이다. 그러므로 같은 점수라 하더라도 사용자의 평가값 평균을 동시에 활용하여야 정확한 선호도를 예측할 수 있다. 사용자 유사도가 계산되고 나면 상품에 대해 유사한 사용자들의 평가에 대한 가중 평균을 계산하여 사용자에 대한 상품 선호도를 예측한다.

GroupLens[8]와 같은 협동적 여과기반 추천시스템에서는 사용자의 선호도를 예측하는데 상관관계(correlation) 기법을 사용하였고, 다양한 변형된 기법들이 추천시스템의 정확도 향상을 위해 제안되었다.

Billsus와 Pazzani[6]는 충분한 정보가 없을 때 사용자의 선호도 예측을 위해 속성 추출 기술을 적용하였고 사용자-상품 선호도 행렬의 차원을 줄이기 위하여 SVD(Singular Value Decomposition)를 사용하였으며, Breese 등[11]은 사용자로부터의 평가정보가 없는 상품에 대해 기본 평가값을 사용하였고 각 상품에 얼마나 많은 사용자가 평가를 수행했느냐에 따라 서로 다른 가중치를 적용하였다.

Herlocker 등[7]은 다양한 유사도 계산 방식과 유사도 가중치 방법에 대한 실험을 수행하였다. 유사도 계산에는 피어슨 상관계수, 스피어만 상관계수, 벡터 유사도, 엔트로피를 이용하였고, 선호도를 구하는 방법으로 평균 가중치, 유사 사용자의 상품 선호도의 가중치 합, z 평균 방법을 이용하였다.

Shardanand[9]는 사용자 사이의 유사도가 아닌 상품-상품 유사도를 이용하는 방법을 제안하였고, Basu[12], Claypool[13] 등은 협동적 여과기반 추천과 내용기반 추천을 결합하는 복합 추천을 제안하였으며, Pazzani[14]는 사용자의 내용기반 신상정보 사이의 유사도로 사용자 사이의 유사도를 계산함으로써 협동적 여과와 내용기반 여과를 혼합하는 시도를 하였다.

Condliff 등[15]은 베이지안 혼합 효과 모형에서

사용자의 평가값과 사용자와 상품의 속성값을 통합하는 시도를 하였고, Popescul 등[16]은 three-way aspect 모형에서 내용정보와 사용자와 상품의 동시 발생 데이터를 결합한 통합 확률 모형을 제안한 바 있다.

정보 부족 문제를 해결하기 위한 연구로는 평가값을 생성하는 내용기반 소프트웨어 에이전트를 사용한 Good 등[17]의 연구가 있으며, 선호도가 나타난 상품의 평가정보와 내용정보를 이용하여 선호도가 나타나지 않은 상품에 대해 선호도 예측값을 생성한 후, 원시 데이터에 선호도 예측값을 추가하여 가상 데이터를 생성하고 협동적 추천을 수행하는 Melville 등[18]의 연구가 있다. 그러나 이들의 방법은 추천하고자 하는 상품에 대하여 기술한 내용정보 등을 사용하는데 비해, 본 논문에서 제안한 기법은 객체의 속성 분포를 이용하여 추가 속성을 생성하여 평가 데이터를 변경하는 방법을 취한다는 점에서 차이가 있다.

3. 속성 생성 기법

일반적으로 멀티미디어 데이터들은 내용기반 형식으로 존재하지 않기 때문에 해당 데이터가 어떠한 내용을 담고 있는지 알기가 어렵다는 것이 문제이다.

예를 들면, 사용자가 영화를 감상하기 위해 영화를 내려 받은 경우, 영화를 보지 않고서는 영화가 어떠한 내용을 담고 있는지 파악할 수 없다. 우리가 알 수 있는 정보는 영화를 내려 받을 때 참고한 제목, 유형, 주연 등에 관한 영화에 대한 간단한 특징정보가 전부일 것이다.

영화나 음악과 같은 멀티미디어 데이터를 추천하는 추천시스템들은 사용자들에게 올바른 추천을 수행하기 위해 상품에 대한 사용자들의 평가정보를 활용한다. 이와 같은 상품에 대한 사용자 평가정보 집합을 평가 데이터라 하며, 평가 데이터의 형식은 일반적으로 표 1에서와 같이 2차원으로 표현된다. 평가 데이터에서 행의 첫 열은 상품들을 평가한 전체 사용자들을 나타내고, 첫 행은 평가에 참여한 모든 상품들이 나타난다. 여기서 '1'은 사용자가 영화를 구매했다는 의미를 표현한 것이다.

사용자가 영화에 대해 성향을 자세히 표시하여 주면 좋은 평가 데이터를 만들 수 있으나, 일반적으로 사용자들에게서 자세한 평가정보를 얻기란 쉬운 일

표 1. 구매 데이터 예제

영화 사용자	영화1	영화2	영화3	...	영화n
사용자1	1	0	0	...	0
사용자2	0	1	0	...	1
사용자3	0	1	1	...	0
:	0	0	1	...	0
Um	0	0	0	...	1

이 아니다. 영화나 음악을 제공하는 인터넷 업체에서 사용자들에게 구매한 상품에 대해 자세한 평가를 요구하면 사용자들은 불편함을 느껴 평가 요구에 응하지 않는 경우가 많다. 이와 같은 이유로 대부분의 유료 업체에서는 사용자 이탈을 방지하기 위해 사용자에게 상품에 대한 평가를 요구하지 않는 경우가 많으며, 요구한다고 하더라도 사용자에게 불편을 끼치지 않을 정도로 요구한다.

멀티미디어 상품일 경우에는 상품의 활용 정보가 빈약하여 사용자의 평가정보가 더욱 절실히 필요하지만, 이러한 현실에서 멀티미디어 데이터에 대한 평가정보 획득은 더욱 어려운 문제로 남아 있다. 이와 같은 이유로 인터넷 업체들은 상품에 대한 구매정보를 추천시스템의 평가 데이터로 활용한다.

사용자가 특정 상품을 구매했다면 사용자는 구매한 상품을 좋아하기 때문에 구매했을 것이며, 이러한 사실을 전제로 평가 데이터를 구성한다면 생성하기 어려운 평가 데이터를 쉽게 만들 수 있다. 그러나 간단한 평가 데이터를 활용하여 우수한 추천을 이루기는 어렵다는 것이 추천시스템의 현실이다. 그래서 본 논문에서도 실험에 사용한 평가 데이터를 구매내역만 나타나게 변경하여 실험에 활용하였다.

이와 같은 평가 데이터에서 사용자 정보에는 나이, 직업, 성별 등이 표현되어 있으며, 상품 정보에는 유형, 주연, 감독 등이 표현되어 있다. 사용자의 나이, 직업, 성별 등은 사용자를 표현할 수 있는 특징정보들이고, 유형, 주연, 감독 등은 상품의 특징정보들이다. 이와 같은 특징정보들을 이용하여 사용자와 상품과 같은 객체들을 벡터로 표현할 수 있다.

예를 들면, n개의 속성으로 이루어진 객체들은 n차원의 특징 벡터 $V = \langle \text{feature}_1, \text{feature}_2, \text{feature}_3, \dots, \text{feature}_n \rangle$ 로 표현 된다. 특징 벡터를 이용하여 평가 데이터의 객체들을 다시 표현하면 사용자의 경우에

는 <나이, 직업, 성별>이라는 특징 벡터를 구성할 수 있고, 상품의 경우에는 <유형, 주연, 감독>이라는 특징 벡터를 구성할 수 있다. 이와 같은 특징 벡터의 속성들을 이용하여 평가 데이터를 재구성하면 사용자들의 성향을 쉽게 파악할 수 있다.

본 논문에서는 사용자나 상품과 같은 객체들의 특징정보로 나타난 속성을 이용하여 평가 데이터를 재구성하여 멀티미디어 추천에 활용하는 방법을 제안한다. 본 논문에서 제시한 속성 생성 기법은 평가정보에 나타난 특징정보를 활용하여 평가 데이터를 재구성하는 방법이다. 속성 생성 기법은 사용자의 선호정보를 풍부하게 함으로써 사용자의 특성 파악에 도움이 된다.

속성 생성 기법의 기본 개념은 객체들의 특징정보를 이용하여 평가 데이터 구조를 속성 분포를 기반으로 파악한 후, 파악된 분포 정보를 이용하여 평가 데이터를 갱신한다. 갱신된 평가 데이터에는 속성 생성 기법에 의해 생성된 추가 속성이 포함된다. 평가 데이터에 추가 속성이 포함된 데이터를 추가 속성 데이터라 한다.

추가 속성 데이터는 복잡한 구조로 이루어지지만, 개념만 이용하여 약식으로 표현하면 표 2와 같다. 표에서 빗살무늬로 나타난 부분이 평가 데이터이고, 음영이 나타난 부분이 속성 생성 기법에 의해 생성된 추가 속성이다. 추가 속성이 평가 데이터와 결합하여 추가 속성 데이터를 생성한다. $a_1, a_2, a_3, \dots, a_n$ 은 평가 데이터에서 사용된 속성들이고, A_1 은 속성 생성 기법에 의해 생성된 속성으로 평가 데이터에 추가되어 추가 속성 데이터를 생성할 때 사용된다. 평가 데이터를 추가 속성 데이터로 변경하게 되면 객체의 연관성이 뚜렷하게 나타나기 때문에 정보 추출에 효과적으로 사용된다. 표 2에 나타난 평가 데이터에서 O1과 O5의 연관성은 뚜렷하게 보이지 않지만, 추가 속성 데이터에서는 O1과 O5의 연관성이 뚜렷하게

표 2. 추가 속성 데이터 예제

	a1	a2	a3	...	an	A1
O1	1	0	0	...	0	1
O2	0	1	0	...	0	0
O3	0	1	1	...	0	0
O4	0	0	1	...	0	0
O5	0	0	0	...	1	1

나타난 것을 확인할 수 있다.

U_a 는 사용자 객체이고 C_k 가 상품 객체일 때, 각 객체들의 특징정보를 활용한 추가 속성 A_c 을 생성하는 속성 생성 기법을 $Appending(A_c)$ 라 한다. 이 때, 사용자와 상품 객체들의 특징들로 구성된 특징 벡터를 $V_i(= \langle \text{feature}_1, \text{feature}_2, \text{feature}_3, \dots, \text{feature}_n \rangle)$ 라 한다.

이러한 전제에서 속성 생성 기법을 설명하면 다음과 같다. 영화에 대한 선호도 데이터에 표현된 유형이 $C_1, C_2, C_3, \dots, C_n$ 이라 하면 전체 집합인 선호도 데이터는 유형에 의해 모두 나뉠 수 있으므로 각 유형은 상호 배타적인 사건이다. 이러한 전제에서 사용자 U 가 특정 유형 C 를 좋아할 사후 확률은 $\Pr[C|U]$ 이다. 그러므로 유형에 대한 확률은 식 1과 같이 나타낼 수 있다. 식 1을 사용자 관점에서 확률을 전개하면 식 2와 같다. 사용자 U_i 가 특정 유형 C_j 를 좋아할 확률은 $\Pr[C_j|U_i]$ 이고, 사용자 U_i 에 대한 특징정보가 나이(Age), 직업(Job), 성별(Sex)이라면 식 2는 나이, 직업, 성별에 대한 속성 확률로 표현될 수 있으며, 이와 같은 경우에 사용자를 기반으로 일반화하면 식 3과 같다. 식 3을 다시 상품을 기반으로 일반화하면 식 4와 같다. 식 4에 속성 상수 ρ 을 결합하여 속성 생성 기법을 생성하고, 속성 생성 기법은 식 5와 같다. 속성 상수 ρ 는 속성 생성 적용도를 나타내는 상수이다.

$$\Pr[C_1|U_a] + \Pr[C_2|U_a] + \Pr[C_3|U_a] + \dots + \Pr[C_n|U_a] \quad (1)$$

$$= \Pr[C_1|U_1] \cdot \Pr[C_1|U_2] \cdot \Pr[C_1|U_3] \cdot \dots \cdot \Pr[C_1|U_m] + \Pr[C_2|U_1] \cdot \Pr[C_2|U_2] \cdot \Pr[C_2|U_3] \cdot \dots \cdot \Pr[C_2|U_m] + \dots + \Pr[C_n|U_1] \cdot \Pr[C_n|U_2] \cdot \Pr[C_n|U_3] \cdot \dots \cdot \Pr[C_n|U_m] \quad (2)$$

$$= \prod_{a=1}^m \Pr[C_1|U_a] + \prod_{a=1}^m \Pr[C_2|U_a] + \prod_{a=1}^m \Pr[C_3|U_a] + \dots + \prod_{a=1}^m \Pr[C_n|U_a] \quad (3)$$

$$= \sum_{k=1}^n \prod_{a=1}^m \Pr[C_k|U_a] \quad (4)$$

$$Appending(A_c) = \rho \sum_{k=1}^n \prod_{a=1}^m \Pr[C_k|U_a] \quad (5)$$

멀티미디어 데이터를 이용하여 사용자 선호도를 측정하기 위해 많은 추천시스템들은 사용자의 구매 이력과 같은 묵시적인 정보를 활용한다. 그러나 이와 같은 묵시적인 정보들을 이용하여 평가 데이터를 생성하면 다양한 특성을 나타내지 못하여 다양한 분석

을 수행할 수 없다. 다양한 분석을 수행할 수 없을 경우에 발생하는 가장 큰 문제점은 추천시스템의 추천 정확도 저하이다. 이러한 정확도 문제점을 해결하기 위해서는 단순한 평가 데이터를 다각도로 분석하고 다양한 정보를 추출할 수 있도록 평가 데이터를 재구성해야 한다. 평가 데이터의 재구성을 위해 본 논문에서는 속성 생성 기법을 제안한다.

속성 생성 기법은 사용자의 선호도 정보가 부족한 평가 데이터의 정보량을 풍부하게 만들어 추천시스템의 정확도를 향상시킨다. 속성 생성 기법에서는 각 객체들의 특징정보들을 활용하여 평가 데이터를 분석한 후, 분석된 결과를 활용하여 새로운 속성을 생성한다. 속성 생성 기법에 의해 생성된 추가 속성은 평가 데이터를 다양한 방법으로 범주화하는 특성을 갖는다. 평가 데이터를 다양한 방법으로 범주화하게 되면 데이터의 분류 방법이 다양해지므로 사용자의 선호도 정보를 다각적으로 분석할 수 있다는 장점이 있다. 또한 속성 생성 기법이 적용된 데이터에는 기존의 데이터보다 정보량이 풍부하며, 풍부한 정보량은 멀티미디어 추천시스템의 정확도 향상에 많은 도움을 준다.

4. 실험 결과

4.1 실험 방법

실험에서 사용한 실험 데이터는 GroupLens의 MovieLens 데이터를 활용하였다[19]. MovieLens 데이터는 6,040명의 사용자가 3,883개의 영화들에 대해 평가를 수행한 데이터로 사용자의 명시적 평가가 이루어진 데이터 집합이다. 사용자들은 5단계로 이루어진 평가값을 활용하여 영화에 대한 평가를 수행하였으며, 사용자의 평가가 나타난 수는 1,000,209건으로 정보 부족이 심한 데이터이다. MovieLens 데이터는 사용자의 나이, 성별, 직업 등과 같은 사용자 정보와 영화 제목, 유형 등의 상품 정보를 포함하고 있으며, 영화 유형은 총 18개이다.

실험을 위해 MovieLens 데이터에 표시된 사용자의 영화에 대한 평가값을 변형하여 이용하였으며, 평가값을 변형한 방법은 다음과 같다. 사용자의 영화에 대한 평가값이 사용자가 평가한 전체 평가값들의 평균보다 큰 경우에는 평가값을 1로 변환하였고, 사용자의 영화에 대한 평가값이 사용자가 평가한 전체

평가값들의 평균보다 작거나 같은 경우에는 0으로 변환하였다. 이와 같은 방법으로 데이터를 변환하여 실험 데이터로 활용하였으며, 실험 데이터에 나타난 '1'은 해당 상품을 '구매'하였다는 의미이다. 웹상에서는 상품 구매 여부에 관한 간단한 정보만을 추출할 수 있기 때문에 이와 같은 데이터 변환 작업을 수행하였다.

본 실험에서 사용한 실험용 데이터의 크기는 상품의 수가 1,000개이고 사용자의 수가 1,000명이며, 데이터 추출은 임의추출 방식을 취했다. 실험 데이터에 나타난 평가 총수는 17,157개이고 정보 표현율은 1.7%로 실험 데이터는 매우 심각한 정보 부족 현상을 나타낸다.

실험에 사용된 데이터는 속성 추가량에 따라 4개의 실험 데이터 집단으로 나뉘며, 각 데이터 집단은 속성 데이터I, 속성 데이터II, 속성 데이터III, 속성 데이터IV이다. 속성 데이터I은 추가 속성이 한 개이고, 속성 데이터II, 속성 데이터III, 속성 데이터IV에서 사용한 속성들은 각각 두 개, 세 개, 네 개다. 속성 생성 기법에서 사용한 속성은 나이, 성별, 직업, 유형이다. 네 가지 종류로 속성 생성 기법 나누어 실험한 이유는 속성 생성량에 따른 추천 정확도 변화를 알아보기 위해서다.

최종적으로 추출된 실험 데이터는 10개의 동일한 크기의 데이터 집단으로 나누어 하나의 집단은 실험 데이터로 사용하고, 다른 9개의 집단은 훈련 데이터로 사용한다. 이와 같은 방법을 서로 다른 10개의 데이터 집단에 모두 적용하여 실험하는 10회 교차검증 (10-fold cross validation) 방식을 실험 방법으로 채택했다.

사용자 유사도 측정에는 벡터 유사도를 이용하였다. 벡터 유사도는 사용자의 상품에 대한 선호도를 n차원의 벡터로 표현하고, 벡터와 벡터의 거리를 이용하여 유사도를 측정하는 방식으로 공식을 식 6에 나타내었다. $similarity_{v,w}$ 는 사용자 v 와 w 의 벡터 유사도를 나타낸다. $r_{v,i}$ 는 사용자 v 가 상품 i 에 대하여 평가한 값이고, m 은 속성의 개수를 나타낸다.

$$similarity_{v,w} = \frac{\sum_{i=1}^m r_{v,i} \times r_{w,i}}{\sqrt{\sum_{i=1}^m r_{v,i}^2} \sqrt{\sum_{i=1}^m r_{w,i}^2}} \quad (6)$$

실험에 사용된 추천 기법은 협동적 여과 기법과 나이브 베이저안 기법이다. 두 기법 모두 기본 데이터와 속성 데이터를 이용하여 추천 결과를 생성하고 기본 데이터와 속성 데이터의 추천 정확도를 비교한다. 추천의 정확도 비교에는 적중률을 사용하였으며, 사용자와 추천 상품의 수를 변화시키면서 기본 데이터와 속성 데이터를 이용한 추천 성능을 비교 실험하였다. 적중률은 시행 횟수로 적중된 총수를 나누어 계산하며, 계산식은 식 7과 같다.

$$Hit\ ratio = \frac{total\ number\ of\ hits}{total\ number\ of\ tests} \quad (7)$$

4.2 추천 방법

실험에서 사용한 추천은 두 가지 방식으로 나뉘며, 하나는 기본 데이터를 활용한 일반 추천이고, 다른 하나는 속성 생성 기법이 적용된 속성 데이터를 활용한 추천이다.

일반 추천에서는 다음과 같은 방법으로 추천이 이루어진다. 첫 번째 단계에서는 상품을 추천받을 사용자의 평가 데이터를 이용하여 사용자들간의 유사도를 계산한 후, 유사한 사용자들을 추출한다. 추출된 유사 사용자들의 선호도 정보를 이용하여 상품을 추천받을 사용자의 선호도를 예측한 후, 상품에 대한 예측값 목록을 생성하여 상품을 추천한다.

속성 생성 기법이 적용된 추천의 흐름을 그림 1에 나타내었고, 속성 생성 기법이 적용된 추천은 다음과 같은 방식으로 추천이 이루어진다. 상품을 추천받을 사용자의 평가정보와 유사도를 계산할 사용자들의 평가정보를 읽은 후, 속성 생성 기법을 적용한다. 속성 생성 기법에는 사용자기반과 상품기반이 있으며, 사용자 임의적으로 선택적으로 이용할 수 있다. 사용자기반은 사용자 관점에서 속성을 생성하는 방식이고, 상품기반은 상품 관점에 속성을 생성하는 방식이다. 속성 생성 기법이 적용된 후, 벡터 유사도를 활용하여 사용자들의 유사도를 계산한다. 유사 사용자 추출은 유사도 분석에서 계산한 사용자들간의 유사도를 기준으로 상위 U명을 추출한다. 추출한 U명의 유사 사용자의 평가를 이용하여 상품을 추천받을 사용자의 예상 평가 벡터를 생성한다. 평가 벡터에서 평가값이 높은 I개의 상품을 추출하여 추천 목록을 생성한 후, 상품을 추천한다.

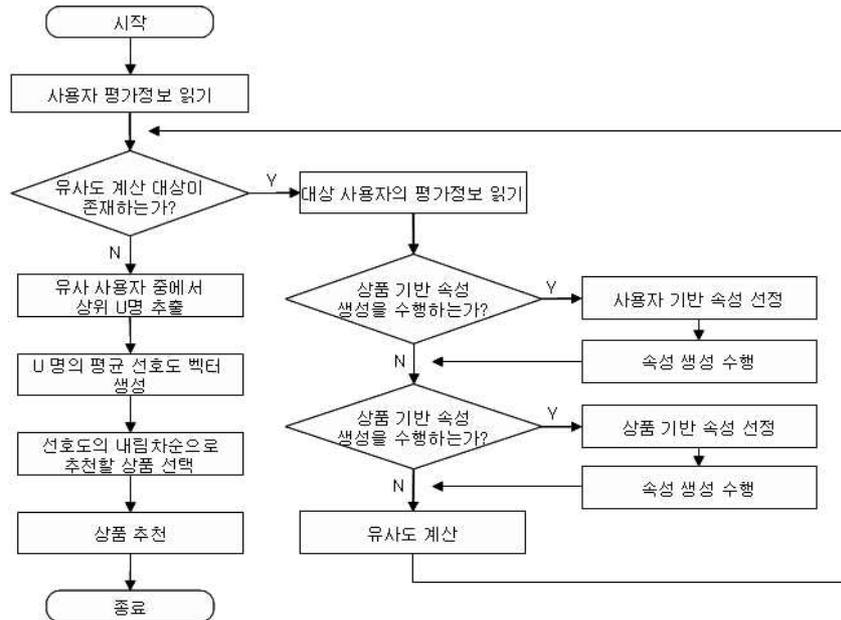


그림 1. 속성 생성 기법이 적용된 추천 흐름도

4.3 성능 측정

표 3에 나타난 결과는 기본 데이터와 속성 데이터의 정확도에 관한 실험 결과이며, 속성 데이터는 속성 데이터IV를 사용하였다. 표에 나타난 I는 추천한 상품의 수를 의미하고, U는 사용자의 수를 의미한다. 표에 나타난 정확도는 기본 데이터의 추천 정확도이고, 향상률은 기본 데이터에 대한 속성 데이터의 정확도 증가율이다. 속성 데이터가 기본 데이터보다 정확도가 증가한 경우에는 '▲'를 증가율 앞에 표시하고

감소한 경우에는 '▽'를 증가율 앞에 표시하였다.

본 실험에서 기본 데이터를 사용한 추천 정확도는 상품의 추천 상품의 수량이 증가하거나 유사 사용자의 수가 증가할 때 비례적으로 증가하였다. 이와 같은 결과가 발생하는 이유는 추천 상품의 수량이 증가하면 추천된 상품의 수량이 증가한 것이므로 사용자가 좋아할 상품일 확률이 높기 때문이다. 그러므로 추천 상품의 증가는 정확도 증가에 중요한 역할을 한다. 그러나 추천 상품의 수량 증가는 사용자에게 선택이라는 추가적인 작업을 발생시키므로 많은 수의 상품 추천은 좋지 못한 방법이다. 그래서 본 실험에서도 많은 수의 상품 추천은 실험에서 배제하였다.

기본 데이터의 실험 결과를 보면 유사 사용자의 수가 증가하면 추천 정확도가 증가하는 것을 알 수 있으며, 이와 같은 결과가 발생하는 이유는 정보가 부족한 데이터에서 유사 사용자를 활용하게 되면 정보가 풍부해지기 때문이다. 그러나 많은 수의 유사 사용자를 이용하게 되면 사용자들의 유사도 계산에 오랜 시간이 걸린다는 단점이 발생한다.

기본 데이터에 대한 속성 데이터의 정확도 향상률을 보면 사용자의 수가 적을수록 향상률이 높은 것을 알 수 있으며, 이와 같은 우수한 결과가 발생하는 이유는 속성 생성 기법이 적용된 속성 데이터가 기본

표 3. 기본 데이터와 속성 데이터IV의 정확도 비교

U		I					
		2	3	6	7	10	20
3	기본 데이터의 정확도	5.13	7.17	12.05	13.41	16.60	23.49
	속성 데이터의 향상률	▲68%	▲50%	▲33%	▲31%	▲26%	▲19%
5	기본 데이터의 정확도	6.68	8.88	14.63	16.01	19.84	28.73
	속성 데이터의 향상률	▲51%	▲47%	▲27%	▲26%	▲20%	▲14%
10	기본 데이터의 정확도	8.55	11.52	18.14	19.82	24.32	34.96
	속성 데이터의 향상률	▲34%	▲27%	▲18%	▲17%	▲14%	▲7%
20	기본 데이터의 정확도	9.46	12.72	20.37	22.21	27.38	38.66
	속성 데이터의 향상률	▲23%	▲18%	▲10%	▲10%	▲7%	▲4%
30	기본 데이터의 정확도	10.07	13.56	21.08	23.17	28.58	40.05
	속성 데이터의 향상률	▲19%	▲13%	▲9%	▲8%	▲6%	▲4%
50	기본 데이터의 정확도	10.09	13.77	20.96	23.03	28.49	40.04
	속성 데이터의 향상률	▲16%	▲10%	▲9%	▲7%	▲6%	▲5%

데이터에 비해 정보량이 풍부하기 때문이다. 추천 상품 수량의 변화에 따른 정확도 향상률에서도 추천 상품 수량이 적을수록 향상률이 높다는 것을 알 수 있으며, 이와 같은 우수한 결과가 발생하는 이유는 속성 생성 기법이 적용된 속성 데이터는 사용자의 성향이 잘 나타나 있기 때문이다. 그러므로 속성 생성 기법을 활용하게 되면 빠른 시간에 상품을 추천할 수 있다는 장점이 있다.

속성 데이터IV의 결과를 보면 원시 데이터보다 매우 우수한 정확도를 나타낸다. 본 실험에서 I=2에서 68%의 정확도 향상을 나타내었고, 가장 낮은 정확도 향상도 16%를 나타내었다. 이와 같은 실험 결과를 보더라도 정보 부족이 심각한 데이터에서는 정보량을 증가시키는 것이 추천 정확도 향상에 중요한 요소로 작용한다는 것을 알 수 있다.

표 4부터 표 6은 속성 데이터I, 속성 데이터II, 속성 데이터III의 실험 결과이다. 속성 데이터I은 하나의 추가 속성만을 활용한 데이터이고, 속성 데이터II는 두 개의 추가 속성을 활용한 데이터이며, 속성 데이터III는 세 개의 추가 속성을 활용한 데이터이다.

속성 데이터I의 실험 결과를 보면 I=2와 I=3인 경우에만 추천 정확도의 향상을 나타내었고, 속성 데이터II의 실험 결과에서는 거의 모든 영역에서 성능 향상을 나타내었다. 특히 속성 데이터II의 실험 결과에서도 유사 사용자나 추천 상품의 수가 적을수록 높은 정확도 향상을 나타내었다.

속성 데이터III의 결과에서는 I(추천 상품)=2와 U(사용자)=3에서 54%의 우수한 성능 향상을 나타내었으며, U=3에서 평균 정확도 향상률은 30%이고

표 4. 기본 데이터와 속성 데이터의 정확도 비교

U		I					
		2	3	6	7	10	20
3	기본 데이터의 정확도	5.13	7.17	12.05	13.41	16.60	23.49
	속성 데이터의 향상률	▲4%	▲1%	0%	0%	0%	0%
5	기본 데이터의 정확도	6.68	8.88	14.63	16.01	19.84	28.73
	속성 데이터의 향상률	0%	0%	0%	0%	0%	0%
10	기본 데이터의 정확도	8.55	11.52	18.14	19.82	24.32	34.96
	속성 데이터의 향상률	▲1%	0%	0%	0%	0%	0%
20	기본 데이터의 정확도	9.46	12.72	20.37	22.21	27.38	38.66
	속성 데이터의 향상률	▲3%	▲2%	0%	0%	0%	0%
30	기본 데이터의 정확도	10.07	13.56	21.08	23.17	28.58	40.05
	속성 데이터의 향상률	▲1%	0%	0%	0%	0%	0%
50	기본 데이터의 정확도	10.09	13.77	20.96	23.03	28.49	40.04
	속성 데이터의 향상률	▲1%	▽1%	0%	0%	0%	0%

표 5. 기본 데이터와 속성 데이터II의 정확도 비교

U		I					
		2	3	6	7	10	20
3	기본 데이터의 정확도	5.13	7.17	12.05	13.41	16.60	23.49
	속성 데이터의 향상률	▲29%	▲22%	▲14%	▲12%	▲10%	▲7%
5	기본 데이터의 정확도	6.68	8.88	14.63	16.01	19.84	28.73
	속성 데이터의 향상률	▲19%	▲17%	▲10%	▲10%	▲7%	▲5%
10	기본 데이터의 정확도	8.55	11.52	18.14	19.82	24.32	34.96
	속성 데이터의 향상률	▲14%	▲10%	▲7%	▲7%	▲4%	▲2%
20	기본 데이터의 정확도	9.46	12.72	20.37	22.21	27.38	38.66
	속성 데이터의 향상률	▲10%	▲8%	▲4%	▲3%	▲1%	▲1%
30	기본 데이터의 정확도	10.07	13.56	21.08	23.17	28.58	40.05
	속성 데이터의 향상률	▲6%	▲4%	▲2%	▲2%	▲1%	0%
50	기본 데이터의 정확도	10.09	13.77	20.96	23.03	28.49	40.04
	속성 데이터의 향상률	▲5%	▲2%	▲2%	▲2%	▲1%	▲1%

표 6. 기본 데이터와 속성 데이터III의 정확도 비교

U		I					
		2	3	6	7	10	20
3	기본 데이터의 정확도	5.13	7.17	12.05	13.41	16.60	23.49
	속성 데이터의 향상률	▲54%	▲41%	▲27%	▲24%	▲20%	▲14%
5	기본 데이터의 정확도	6.68	8.88	14.63	16.01	19.84	28.73
	속성 데이터의 향상률	▲38%	▲34%	▲21%	▲20%	▲15%	▲10%
10	기본 데이터의 정확도	8.55	11.52	18.14	19.82	24.32	34.96
	속성 데이터의 향상률	▲28%	▲21%	▲14%	▲13%	▲10%	▲5%
20	기본 데이터의 정확도	9.46	12.72	20.37	22.21	27.38	38.66
	속성 데이터의 향상률	▲20%	▲15%	▲9%	▲8%	▲5%	▲3%
30	기본 데이터의 정확도	10.07	13.56	21.08	23.17	28.58	40.05
	속성 데이터의 향상률	▲15%	▲11%	▲7%	▲6%	▲4%	▲3%
50	기본 데이터의 정확도	10.09	13.77	20.96	23.03	28.49	40.04
	속성 데이터의 향상률	▲13%	▲8%	▲6%	▲6%	▲4%	▲4%

I=2에서 평균 정확도 향상률은 21.6%로 속성 데이터가 기본 데이터에 비해 높은 추천 정확도를 나타냈다. 이러한 우수한 추천 성능 향상 결과는 모든 실험 구간에서 유지되고 있다. 이러한 결과를 보더라도 속성 생성 기법을 평가 데이터에 적용하면 적은 수의 사용자와 추천 상품을 이용하여도 우수한 추천 성능을 나타낸다는 것을 알 수 있다.

그림 2는 상품 감소에 따른 속성 생성 기법의 정확도 향상률 결과이고, 그림 3은 사용자 감소에 따른 속성 생성 기법의 정확도 향상률 결과이다. 비교된 기법들은 속성 생성 기법I, 속성 생성 기법II, 속성 생성 기법III, 속성 생성 기법IV이다. 그림 2를 보면 상품의 수가 적을수록 빠르게 정확도 향상을 이루는 것을 알 수 있다. 가장 우수한 향상은 속성 생성 기법 IV로 68%의 정확도 향상을 나타내고 가장 낮은 향상

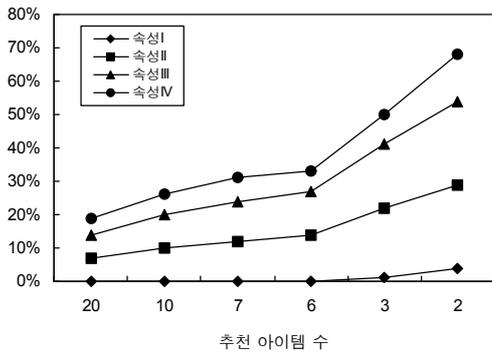


그림 2. 상품 감소에 따른 속성 생성 기법의 정확도 향상률

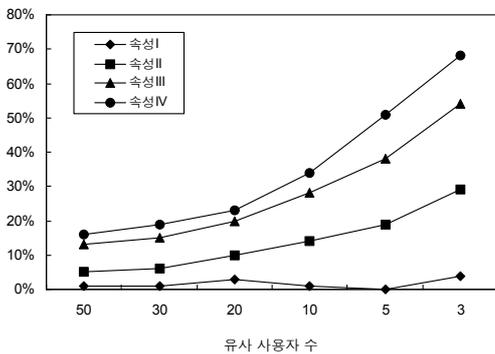


그림 3. 사용자 감소에 따른 속성 생성 기법의 정확도 향상률

들은 속성 생성 기법I로 4%의 정확도 향상을 나타내어 평균 38%의 정확도 향상률을 나타내었다. 이와 같은 결과는 추가 속성이 증가함에 따라 속성 생성 기법의 효율성이 증가된다는 것을 의미한다. 그림 3에서도 사용자의 수가 적을수록 정확도 향상이 높아지는 것으로 나타난다.

본 실험 결과는 속성 생성 기법을 활용하면 사용자들의 특성 파악에 효과적이기 때문에 적은 수의 사용자와 추천 상품을 사용하여도 높은 정확도 향상을 이룰 수 있다는 것을 의미한다.

표 7의 실험 결과는 나이브 베이지안을 적용하여 추천 실험을 수행한 결과이다. 기본 데이터를 활용한 결과에는 매우 낮은 추천 정확도를 나타내었지만, 속성 데이터I를 활용한 경우에는 평균 17.8%의 정확도 향상률을 나타내었고, 속성 데이터II를 활용한 경우에는 평균 36.1%의 정확도 향상률을 나타내었다. 속성 데이터III와 속성 데이터IV를 활용한 경우에는 각각 평균 52.5%와 67%의 정확도 향상률을 각각 나타내었다.

표 7. 나이브 베이지안을 적용한 기본 데이터와 속성 데이터의 정확도 비교

나이브 베이지안 적용	I					
	2	3	6	7	10	20
기본 데이터의 정확도	0.83	1.12	1.75	1.88	2.31	3.37
속성 데이터I의 향상률	▲16%	▲16%	▲17%	▲18%	▲19%	▲21%
속성 데이터II의 향상률	▲34%	▲33%	▲35%	▲37%	▲38%	▲40%
속성 데이터III의 향상률	▲48%	▲47%	▲51%	▲54%	▲56%	▲59%
속성 데이터IV의 향상률	▲57%	▲57%	▲64%	▲70%	▲74%	▲80%

본 실험 결과에서 추가 속성이 늘어날수록 정확도는 매우 큰 폭의 향상을 나타내었으며, 나이브 베이지안을 적용한 추천시스템에서도 속성 생성 기법을 적용하면 추천 정확도 향상에 많은 도움을 줄 수 있다는 것을 확인하였다.

5. 결 론

멀티미디어 추천시스템의 단점인 정보 부족 문제를 해결하기 위해 본 논문에서는 속성 생성 기법을 제안한다. 속성 생성 기법은 사용자나 상품에 나타난 속성정보들을 이용하여 평가 데이터를 분석하여 추가 속성이라는 새로운 속성을 생성시키는 기법으로 평가 데이터에 추가 속성을 포함하여 평가 데이터를 속성 데이터로 변경하는 것을 최종 목적으로 한다. 추가 속성이 포함된 속성 데이터를 멀티미디어 추천시스템에서 활용하면 추천 성능이 향상되며, 실험 결과에서도 속성 생성기법의 적용률이 높아질수록 높은 추천 정확도를 나타내었다. 협동적 여과 기법에서 속성 생성 기법이 가장 높게 적용된 속성 데이터IV가 가장 우수한 추천 성능을 나타내었으며, 속성 데이터IV는 일반 데이터에 비해 I=2에서는 평균 35%의 성능 향상을 나타내었고 U=3에서는 평균 37.8%의 성능 향상을 나타내었다. 나이브 베이지안을 적용한 실험 결과에서도 평균 정확도 향상률이 속성 데이터I, 속성 데이터II, 속성 데이터III, 속성 데이터IV에서 각각 17.8%, 36.1%, 52.5%, 67%로 우수하게 나타났다. 본 실험으로 속성 생성 기법은 정보 부족 문제를 완화하여 멀티미디어 추천시스템의 정확도를 향상시킨다는 것을 확인할 수 있다.

향후 연구로 속성들의 분포 특성에 따른 속성 생성의 효율에 대한 연구와 속성들의 연관성에 따른 속성 생성의 활용도에 대한 연구가 필요하다.

참 고 문 헌

- [1] R. Armstrong, D. Freitag, T. Joahims, and T. Mitchell, "WebWatcher: A Learning Apprentice for the World Wide Web," *Proceedings of the 12th National conference on Artificial Intelligence*, 1995.
- [2] M. Balabanovic and Y. Shoham, "Fab : Content-Based Collaborative Recommendation," *Recommendation Communications of the ACM*, Vol. 40, No. 3, pp. 66-77, 1997.
- [3] H. Lieberman, "Letizia : An Agent that Assists Web Browsing," *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, 1995.
- [4] M. Pazzani, J. Muramatsu, and D. Billsus, "Syskill & Webert: Identifying Interesting Web Sites," *Proceedings of the 13th National Conference on Artificial Intelligence*, 1996.
- [5] B. Krulwich, "Lifestyle Finder: Intelligent User Profiling Using Large-Scale Demographic Data," *Artificial Intelligence Magazine*, Vol. 18, No. 2, 1997.
- [6] D. Billsus and M. J. Pazzani, "Learning Collaborative Information Filters," *Proceedings of the 15th International Conference on Machine Learning*, 1998.
- [7] J. Herlocker, J. Konstan, A. Borchers, and J. Riedl, "An Algorithmic Framework for Performing Collaborative Filtering," *Proceedings of the ACM SIGIR-99*, 1999.
- [8] J. Konstan, B. Millr, D. Maltz, J. Herlocker, L. Gordon, and J. Riedl, "GroupLens: Applying Collaborative Filtering to Usenet News," *Communications of the ACM*, Vol. 40, No. 3, pp. 77-87, 1997.
- [9] U. Shardanand and P. Maes, "Social Information Filtering: Algorithms for Automating 'Word of Mouth'," *Proceedings of the Conference of Human Factors in Computing Systems*, 1995.
- [10] L. Terveen, W. Hill, B. Amenta, D. McDonald, and J. Creter, "PHOAKS: A System for Sharing Recommendations," *Communications of the ACM*, 1997.
- [11] J. Breese, D. Heckerman, and C. Kadie, "Empirical Analysis of Predictive Algorithms for Collaborative Filtering," *Proceedings of the Fourteenth Annual Conference on Uncertainty in Artificial Intelligence*, 1998.
- [12] C. Basu, H. Hirsh, and W. Cohen, "Recommendation as Classification: Using Social and Content-Based Information in Recommendation," *Proceedings of the 15th National Conference on Artificial Intelligence*, 1998.
- [13] M. Claypool, A. Gokhale, T. Miranda, P. Murnikov, D. Netes, and M. Sartin, "Combining Content-Based and Collaborative Filters in an Online Newspaper," *Proceedings of the ACM SIGIR Workshop on recommendation Systems*, 1999.
- [14] M. Pazzani, "A Framework for Collaborative, Content-Based and Demographic Filtering," *Artificial Intelligence Review*, pp. 393-408, 1999.
- [15] M. Condliff, D. Lewis, D. Madigan, and C. Posse, "Baysian Mixed-Effect Models for Recommendation Systems," *Proceedings of recommendation Systems Workshop at SIGIR-99*, 1999.
- [16] A. Poposcul, L. Ungar, D. Pennock, and S. Lawrence, "Probabilistic Models for Unified Collaborative and Content-Based Recommendation in Sparse-Data Environments," *Proceedings of the 17th Conference on Uncertainty in Artificial Intelligence*, 2001.
- [17] N. Good, J. B. Shafer, J. A. Konstan, A. Borchers, B. Sarwar, J. Herlocker, and J. Riedl, "Combining Collaborative Filtering with Personal Agents for better Recommendations," *Proceedings of the Sixteenth National Conference on Artificial Intelligence*, 1999.
- [18] P. Melville, R. Mooney, and R. Nagarajan, "Content-Boosted Collaborative Filtering," *Proceedings of the SIGIR-2001 Workshop on*

recommendation Systems, 2001.

[19] GroupLens, <http://www.grouplens.org/>



김형일

1996년~1998년 (주)경기은행
2004년 동국대학교 대학원 컴퓨터공학과(공학박사)
2005년~2006년 동국대학교 컴퓨터공학과 IT교수(정보통신부)
2007년~현재 나사렛대학교 멀티

미디어학과 교수

관심분야: 지능형시스템, 기계학습, 지능형 로봇, 임베디드시스템, 지능형 에이전트



엄정국

1980년 동국대학교 통계학과(이학사)
1983년 동국대학교 통계학과(이학석사)
1990년 동국대학교 통계학과(이학박사)
1996년~현재 나사렛대학교 멀티

미디어학과 교수

관심분야: 수리통계, 회귀분석, 수치해석, 데이터마이닝