

# 협업적 필터링 및 퍼지시스템 기반 사용자 성향분석에 의한 영화평가 예측 시스템

## A Movie Rating Prediction System of User Propensity Analysis based on Collaborative Filtering and Fuzzy System

이수진 · 전태룡 · 백경동 · 김성신\*

Soojin Lee, Taeryong Jeon, Gyeongdong Baek and Sungshin Kim\*

부산대학교 전자전기공학과

### 요 약

지능형 추천 시스템은 사용자의 요청에 응답하는 수동적인 시스템이 아닌 사용자가 원하는 서비스를 제안하는 시스템으로서 최근 콘텐츠 서비스 분야에 많이 개발되고 있다. 이러한 지능형 추천 시스템은 콘텐츠 개인화 서비스에 응용되고 있으며 대표적인 추천기법으로 내용기반과 협업적 필터링 기법이 있다. 본 연구에서는 협업적 필터링 및 퍼지 시스템을 이용하여 추천 시스템의 기반 기술인 예측 시스템을 제안하였다. 제안한 예측 시스템은 사용자의 과거 영화평가 정보를 바탕으로 영화에 대한 평가점수를 예측한다. 영화평가 예측시스템의 성능은 영화 평가점수의 실제값과 예측값의 오차를 RMSE (root mean square error) 방법으로 계산한 후 기존의 영화평가 시스템 RMSE 값과 비교하여 평가하였다. 본 연구를 통해 제안한 영화평가 예측시스템이 추천 시스템의 기반 기술로서 활용이 가능하고 다른 멀티미디어 콘텐츠 서비스 추천에도 응용이 가능할 것으로 기대한다.

키워드 : 협업적 필터링, 퍼지 시스템, 예측 시스템, 추천, 영화 평가 예측

### Abstract

Recently an intelligent system is developed for the service what users want not a passive system which just answered user's request. This intelligent system is used for personalized recommendation system and representative techniques are content-based and collaborative filtering. In this study, we propose a prediction system which is based on the techniques of recommendation system using a collaborative filtering and a fuzzy system to solve the collaborative filtering problems. In order to verify the prediction system, we used the data that is user's rating about movies. We predicted the user's rating using this data. The accuracy of this prediction system is determined by computing the RMSE(root mean square error) of the system's prediction against the actual rating about the each movie and is compared with the existing system. Thus, this prediction system can be applied to base technology of recommendation system and also recommendation of multimedia such as music and books.

Key Words : Collaborative filtering, fuzzy system, prediction system, recommendation, movie rating prediction

## 1. 서 론

인터넷과 컴퓨터의 발달로 온라인을 통하여 활용할 수 있는 정보들이 다양해짐에 따라, 사용자에게 불필요한 내용을 제거하고 필요한 정보만을 제공해주는 정보 필터링 기법이 제안되며 추천 시스템(recommendation system)과 같은 응용 프로그램에 대한 관심이 집중되고 있다.

추천 시스템이란 사용자에게 새로운 아이템을 추천해주는 시스템으로서 뉴스, 기사, 영화, 도서, 음악 등의 분야에

응용할 수 있다. 이는 전자상거래(e-commerce)를 기점으로 사용자의 구매를 촉진하기 위해 이전의 구매 정보를 기초로 사용자가 선호할 만한 제품을 예측하고 정보를 제공하는 추천 시스템 개발로 이어지고 있다[1]. 그 대표적인 예로 통신 판매회사인 Amazon이나 온라인 영화 대여회사인 Netflix를 들 수 있다. 이들은 자사의 콘텐츠 추천 시스템을 구축하여 빠른 시간 내에 사용자의 기대감과 만족감을 이끌어 낼 수 있는 상품을 사용자에게 제시한다[2]. 이와 같은 추천 시스템을 개발하기 위한 기반 기술로서 데이터 마이닝, 패턴 인식, 정보 필터링 등 다양한 기법들에 대한 연구가 이루어지고 있으며 그 중 정보 필터링 기술에 의한 연구가 주를 이루고 있다. 1990년 중반 이후의 대표적인 추천 시스템 연구 방향은 내용기반 추천기법(content based recommendation)과 협업적 필터링기법(collaborative filtering)으로 구분된다[3, 4]. 내용기반 추천 기법은 정보 기술에 바탕을 둔 시스템으로서 콘텐츠를 직접 분석하여 아이템과 아이

접수일자 : 2008년 11월 1일

완료일자 : 2009년 2월 18일

\* Corresponding Author : sskim@pusan.ac.kr

이 논문은 2008년 정부(교육과학기술부)로부터 지원받아 수행된 연구임(지역거점연구단육성사업/차세대물류IT기술연구사업단)

템, 아이템과 사용자 선호도 간의 유사도를 분석한 후 이를 바탕으로 새로운 아이템을 추천해주는 시스템이다. 하지만 추천할 수 있는 콘텐츠의 범위가 좁으며 사용자의 선호도와 유사한 아이템만 찾아보면 제한적이거나 지나치게 특성화된 추천이 되기 쉽다[5-7]. 이에 반해 협업적 필터링은 특정 사용자와 유사한 성향을 나타내는 다른 사용자들을 분석하여 콘텐츠 선호도를 추정하는 시스템이다. 이 기법은 다른 사용자의 경험을 바탕으로 특정 사용자가 선호하는 콘텐츠를 예측할 수 있다. 하지만 새로운 아이템이 추가되었을 때, 그리고 유사한 그룹의 정보가 없을 때 이에 대한 사용자들의 평가가 쌓이기 전에는 콘텐츠 추천이 어려운 문제점을 가지고 있다[8-11].

본 연구에서는 추천 시스템의 기반 기술인 예측 시스템을 제안하기 위해 미국의 온라인 영화 대여 회사인 Netflix 에서 제공하는 데이터를 사용하였다. Netflix 영화평가 데이터를 바탕으로 사용자의 영화평가 점수를 예측하는 방법을 제안하였으며, 영화평가 예측 시스템을 구축하기 위해 사용자별 개인 성향을 반영한 영화 평가 모델을 제안하였다.

## 2. 영화평가 예측시스템

본 연구에서는 사용자의 영화 성향을 분석하기 위해 유사한 성향을 나타내는 집단들의 과거 영화평가 기록을 분석하여 영화 평가를 예측하는 시스템을 개발하였다.

### 2.1 데이터 재구성

제안한 영화평가 예측시스템은 Netflix 사에서 제공한 데이터를 사용하였다. 데이터의 전체 구조는 그림 1과 같으며 모든 데이터는 텍스트 파일 형식으로 구성된다. Netflix 사에서 제공한 데이터는 48만여 명의 사용자들이 17,770개 영화에 대해서 각자 평가한 정보이며, 영화에 대한 평가는 1~5점의 정수로 나타낸다. 그리고 시스템의 성능 평가를 위해 probe data와 qualifying data를 제공한다.

Training data는 약 48만 명의 전체 사용자가 총 17,770개의 영화에 대해 평가한 정보이며 영화 평가 예측 시스템의 기반 정보로 적용한다. 하지만 training data는 사용자 중심이 아닌 영화 중심으로 분류되었기에 사용자 성향을 분석하기에 구조적으로 적합하지 않다. 따라서 전처리 과정으로 사용자를 중심으로 데이터를 재구성하였다. 그림 2는 Netflix 사에서 제공한 training data를 개인 성향분석 목적에 맞게 재구성한 것을 보여준다. 여기서 영화 ID는  $mv_i$ 로 나타내며,  $i$ 의 범위는  $1 \leq i \leq 17,770$ 이다. 위의 17,770개의 영화 파일을 통해  $mv_i$  영화에 대한 사용자별 평가를 알 수 있다. 그리고 사용자 ID는  $u_k$ 로 나타내며,  $k$ 의 범위는  $1 \leq k \leq 2,649,429$ 이고 480,189명의 사용자가 분포되어 있다. Netflix 사에서 제공하는 데이터 중 probe data와 qualifying data는 영화 평가 예측 시스템의 성능을 평가하기 위해 제공되는 데이터이다. 이 데이터의 평가항목 즉, 임의의 영화에 대한 사용자의 평가를  $1.0 \sim 5.0$ 까지 예측하는 것이다. Probe data는 각 평가 항목에 대한 실제값이 제공되지만 qualifying data의 평가는 온라인을 통해 예측값의 오차를 RMSE(root mean square error)로 확인할 수 있다. 본 연구에서는 영화평가 예측시스템의 신뢰성 평가 및 성능 개선을 목적으로 probe 데이터를 적용하여 시스템의 성능을 평가하며 평가 항목에 대한 실제값과 예측값 사이의 오차가 최소가 되도록 시스템을 구성하였다.

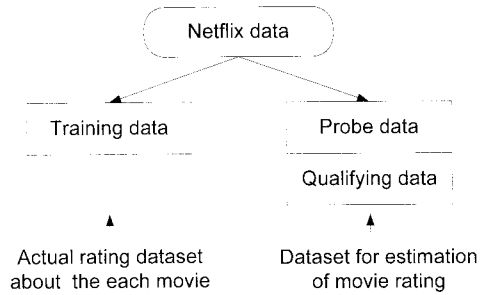
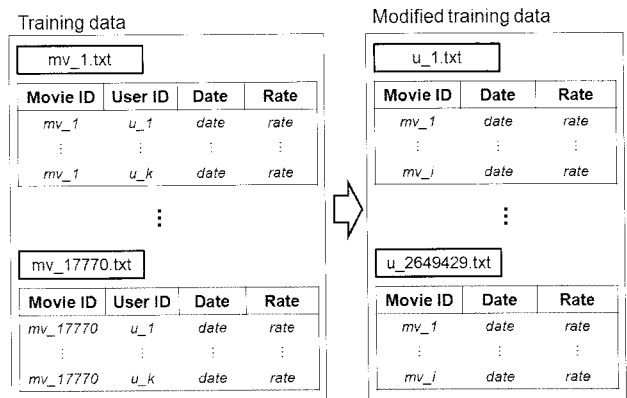


그림 1. Netflix에서 제공된 데이터 구조

Fig 1. Data structure by Netflix



$i$ : movie index ( $1 \leq i \leq 17770$ ),  $k$ : user index ( $1 \leq k \leq 2649429$ ).

그림 2. 사용자 중심으로 수정한 training data

Fig. 2. Modified training data based on user

### 2.2 영화평가 예측과정

영화에 대한 평가예측 항목인 임의의 영화에 대한 사용자의 평가 rate 예측은 그림 3과 같은 과정으로 수행하였다. 먼저 평가예측 데이터로 사용되는 probe data를 통해서 평가하고자 하는 영화인 target movie와 이를 시청한 사용자인 target user 정보를 각각 추출하였다. 그리고 target user에 의해 평가된 영화 평점 정보를 바탕으로 사용자 정보의 신뢰성을 판단하는 필터링 과정을 수행하였다. 필터링된 사용자는 개인별 과거 영화 평가들의 평균을 이용하여 영화에 대한 평가 예측이 이루어지며 필터링 되지 않은 사용자는 협업적 필터링 기법을 이용하여 개인 성향 분석에 의해 평가 예측이 이루어진다. 이는 target user와 유사한 성향을 가지는 그룹을 구성하여 유사성향 그룹, 즉 related user group과 target user와의 유사도를 계산하여 target user의 target movie에 대한 평가를 예측하였다. 또한 협업적 필터링이 가지는 단점을 보완하기 위하여 training data로부터 계산할 수 있는 각 사용자별 평가 평균, 사용자의 날짜별 평균, 영화별 평가 평균, 유사 성향 그룹과 target user 사이의 유사도에 의한 평가 방법을 이용하여 퍼지 추론과정을 수행한다. 위의 일련의 과정을 수행한 후 probe data의 영화 평가 항목은 실제 평가값과 추정된 평가값 사이의 RMSE의 오차를 계산하여 시스템 성능을 분석하였다.

### 2.3 사용자 필터링

본 연구에서는 사용자의 개인 성향을 분석하기 위해 유사한 성향을 가진 다른 사용자들을 related user group으로

분류하는 과정을 수행하였다. 이 때, related user group으로 분류하기 위해서는 사용자들이 영화에 대해 평가한 정보를 신뢰할 수 있어야 한다. 하지만 제공된 training data에는 의미가 없는 정보로 인해 신뢰성이 떨어지는 사용자까지 포함되어 있다. Netflix사에서 제공된 평가 기록 데이터를 살펴보면 한 사용자가 하루에 영화 평가 기록을 작성한 횟수를 기준으로 최소 1편을 평가한 사용자에서부터 최대 17,653편의 영화를 평가한 사용자까지 구성되어 있다. 이는 영화에 대한 정확한 평가가 이루어지지 않은 데이터도 상당수 포함되어 있음을 의미하며 따라서 신뢰성 있는 related user group을 구성하기 위해 필터링 작업이 필요하다. 본 연구에서는 두 가지 조건으로 의미 없는 데이터를 가지는 사용자를 필터링한다. 먼저 하루에 평가한 영화 편수가 전체 사용자가 평균적으로 평가한 영화 편수보다 많은 사용자를 필터링한다. 즉, 하루에 최대로 평가한 영화 수가 200편 이상인 사용자를 필터링 한다. 다음으로 사용자가 평가한 총 영화 편수와 하루에 최대로 평가한 영화 편수의 포션(portion)이 0.25 이상이면 필터링 하였고 포션을 계산하는 방법은 다음과 같다.

$$portion = \frac{\max(n_{u,d})}{n_{u,i}} \quad (1)$$

여기서  $n_{u,i}$ 는 선택된 사용자( $u$ )가 영화( $i$ )를 평가한 총 개수이며  $n_{u,d}$ 는 사용자가 하루에 평가한 영화 개수이다. 위와 같이 필터링 된 사용자는 다른 사용자와 성향을 비교하지 않고 사용자의 과거 영화 정보를 이용하여 개인의 성향을 분석한다. 즉, 필터링 된 사용자가 평가한 모든 영화에 대한 rate의 평균을 이용하여 사용자의 성향을 분석한다.

#### 2.4 개인 성향별 사용자 그룹화

개인 성향을 분석하기 위해 구성되어지는 related user group은 신뢰성을 가지는 즉, 필터링 되지 않은 사용자들을 대상으로 이루어진다.

Related user group 구성은 사용자와 영화의 관계 분석 과정이 바탕이 되며 이는 사용자-영화 프로파일로서 그림 4와 같이 나타낸다. 사용자-영화 프로파일은 임의의 영화  $i$ 에 따른 사용자( $u$ )의 영화 평가점수( $R$ )가 있으며 빈 공간은 시청하지 않은 영화에 대한 항목을 의미한다. 그리고 영화  $i$ 의 범위는  $1 \leq i \leq 17,770$  이고 사용자  $k$ 의 범위는  $1 \leq k \leq 2,649,429$ 이다. 위의 프로파일을 통해 평가 예측 대상인 target user와 연관이 있는 모든 사용자들에 대한 영화 평가 기록을 분석해서 related user group을 구성한다. 사용자-영화 프로파일의 정보를 통해 그림 5와 같이 target user와 관련 있는 다른 사용자를 찾아 분류 한다. 사용자를 그룹으로 분류하는 과정은 다음과 같이 표현할 수 있다. 먼저 영화 평가 예측 항목으로 구성된 probe data에서 평가 대상이 되는 target user와 target movie의 정보를 추출한다. 다음으로 target user 이외의 모든 사용자를 대상으로 target movie에 대한 평가 기록이 있는 사용자를 추출한다. 마지막으로 추출한 사용자들 중에서 target user가 과거에 평가한 영화와 최소 한 편 이상 같은 정보를 가진 사용자들을 related user group으로 분류한다. 그림 4의 사용자-영화 프로파일에서 target user는 3이고 target movie는 4일 때, related user group은 1, 4,  $k-1$ 이다.

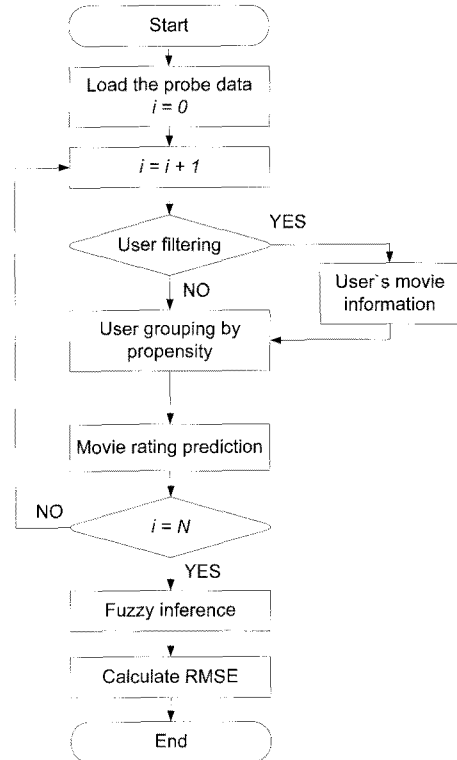


그림 3. 영화 평가 예측 알고리즘  
Fig. 3. Flowchart of movie rating prediction

		user index ( k )							
		1	2	3	4	...	k-2	k-1	k
movie index ( i )	1	3	1		2				
	2	3		3	3			4	2
	3		4	4				5	
	4	4		?	3			2	4
	...								
	i-2	2		2	1			3	2
	i-1		4						
i	4	2	5	5			3		

그림 4. 사용자-영화 프로파일  
Fig. 4. A profile of user-movie

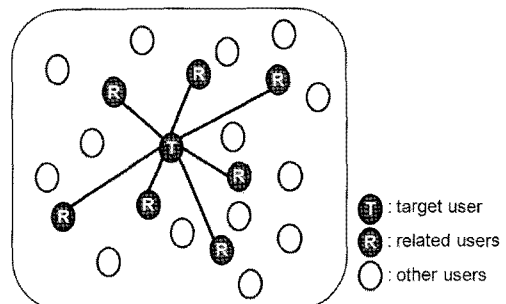


그림 5. 관련 있는 사용자 그룹화  
Fig. 5. Related user grouping

2.5 협업적 필터링을 이용한 영화평가 예측

본 연구에서는 협업적 필터링을 이용하여 영화에 대한 평가를 예측하였다. probe data를 통해 target user와 target movie가 설정되면 target user와 유사 성향을 보이는 related user group과의 관계 분석을 통해 target user의 target movie에 대한 평가 rate를 예측한다. related user group과 target user와의 관계는 유사도에 의해 추론하며 유사도는 식(2)로 계산하였다.

$$\mu = \frac{N(r_{t,i}, r_{u,i})}{n_i} \quad (2)$$

$r_{t,i}$ 는 target user가  $i$ 영화에 대해 평가한 rate이고,  $r_{u,i}$ 는 related user가  $i$ 영화에 대해 평가한 rate이다. 그리고  $N(r_{t,i}, r_{u,i})$ 는  $r_{t,i}$ 와  $r_{u,i}$ 가 같은 개수를 나타내며  $n_i$ 는 target user가 전체에 대해 평가한 영화 목록 수이다. 유사도  $\mu$ 는 target user와 related user 간의 유사한 정도를 나타내며 유사도가 1.0에 가까울수록 target user와 비슷한 성향을 가진다고 할 수 있다.

식(2)에 의해 계산한 related user group과 target user와의 유사도는 영화 평가 예측에 적용한다. 유사도를 이용한 영화 평가 예측 수식은 다음과 같다.

$$Estimated\ rate = \frac{\sum_{i=1}^n \mu_i \times R}{\sum_{i=1}^n \mu_i} \quad (3)$$

식(3)은 무게 중심법을 이용하여 영화에 대한 평가를 예측하는 방법이다. 이는 target user와 related user group의 유사한 정도를 target movie에 대한 평가 예측에 비율적으로 반영하기 위한 것이다. 여기서  $\mu_i$ 는 related user  $i$ 에 대한 유사도이고  $R$ 은 related user가 target movie에 대해 평가한 실제 rate 값을 의미한다. 그리고  $n$ 은 target user와 관련 있는 related user 수를 나타낸다.

2.6 협업적 필터링 및 퍼지시스템 기반 영화평가 예측

유사도를 이용한 영화평가 예측 방법이란 협업적 필터링을 이용하여 다른 사용자와의 관계 분석을 통해 특정 사용자에 대한 성향을 분석하는 것이다. 그러나 협업적 필터링은 새로운 아이템이 추가되면 데이터 정보 수집 전까지 추천을 기대하지 못하며, 유사한 그룹의 정보가 없을 때 추천하기 어려운 문제점을 가지고 있다. 이러한 문제점을 보완하기 위해 본 연구에서는 퍼지추론 시스템을 제안하였다.

퍼지추론 시스템의 입력 요소는 다음과 같다. 먼저 협업적 필터링을 이용하여 다른 사용자와의 유사도 관계 분석을 통한 유사도 기반의 평가와 사용자의 과거 영화평가 정보를 바탕으로 사용자의 성향을 분석한 각 사용자의 성향별 평가 평균, 특정한 날에 대한 성향을 나타내는 사용자의 날짜별 평가 평균, 특정 영화에 대해 보편적인 성향을 나타내는 각 영화별 평가 평균을 퍼지 시스템의 입력 요소로 구성하였다. 위의 4가지 입력 요소를 이용하여 영화에 대한 평가 예측 값이 출력되는 퍼지 시스템을 설계한다. 퍼지 시스템의 출력인 영화에 대한 평가점수는 1.0~5.0점 사이의 실수이며 소수점 첫째 자리까지 계산한다. 각 입력요소에 대한 소속함수는 LR(low rate), MR(middle rate), HR(high rate) 세 구간으로 구성하며 가우시안 소속함수를 적용하였다. LR은 rate 평

균 1점, MR은 rate 평균 3점, HR은 rate 평균 5점이 기준이며 표준 편차 0.5의 분포를 가진다. 퍼지 시스템의 출력인 영화에 대한 평가 예측을 위해 Mamdani 규칙을 이용하여 추론하였으며 퍼지 규칙은 그림 6과 같다.

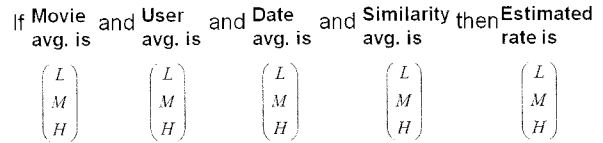


그림 6. 영화 평가 예측을 위한 퍼지 규칙  
Fig. 6. Fuzzy rule for movie rating prediction

3. 실험 및 결과

본 연구에서는 영화 평가 예측 시스템을 개발하기 위하여 Visual C++ 6.0을 이용하였으며 Netflix사에서 제공하는 17,770개의 영화와 48만 여명의 사용자들에 대한 과거 영화시청 및 평가점수 데이터인 training data를 이용하여 실험하였다. 이 때, 데이터의 양이 너무 방대하기 때문에 모집단의 특성을 반영한 표본을 대상으로 실험하였으며 표본 집단 추출은 확률적 표본 추출 방법 중 랜덤 표본(random sample) 방법을 적용하였다. 또한 시스템 성능 확인을 위해 표본의 크기를 증가시키며 각각 10회 반복 실험하였으며 신뢰도 구간으로서 모집단에 대한 오차를 추정하였다.

3.1 협업적 필터링을 이용한 결과

협업적 필터링을 이용하여 영화에 대한 평가를 예측한 결과는 표 1과 같다. 각 표본 집단별로 10회 반복 실험에 대한 시스템 오차를 RMSE로 나타내었다. 표 2는 각 표본 집단에 대한 RMSE 평균과 99%의 신뢰도로서 모집단에 대한 오차 값을 추정한 결과이다. 위의 실험을 통해 모집단에서 99% 신뢰도를 가지며  $0.9036 \leq m \leq 0.9913$ 의 오차를 가질 것이라고 추정할 수 있다. 그림 7은 실험 횟수와 표본 집단의 크기별 시스템 오차의 편차 및 중앙 값을 나타내기 위해 RMSE 값을 boxplot 그래프로 표현한 것이다. 가로축은 표본 집단의 크기인 사용자 수를 나타내며 세로축은 제안한 시스템의 RMSE 값의 분포를 나타낸다.

표 1. 협업적 필터링을 이용한 RMSE 결과  
Table 1. Results of RMSE using collaborative filtering

	100명	300명	500명	1000명
1	1.0399	0.9456	0.9008	0.8950
2	0.9933	0.9618	0.9104	0.9195
3	0.9155	0.9598	0.9271	0.9661
4	0.9746	0.9848	0.9256	0.9368
5	0.9588	0.8803	0.9214	0.9286
6	1.0755	0.9461	0.8797	0.9018
7	0.9812	0.9233	0.8833	0.9055
8	0.9427	0.9494	0.8832	0.9290
9	0.9529	0.9054	0.9131	0.9311
10	0.9559	0.9215	0.9121	0.9107

표 2. 협업적 필터링을 이용한 신뢰도 결과  
Table 2. Results of reliability using collaborative filtering

	100명	300명	500명	1000명
표본 평균	0.97903	0.93782	0.90568	0.92240
99% 신뢰도	$0.9668 \leq m \leq 0.9913$	$0.9333 \leq m \leq 0.9424$	$0.9036 \leq m \leq 0.9078$	$0.9207 \leq m \leq 0.9241$

3.2 협업적 필터링 및 퍼지 시스템을 이용한 결과

협업적 필터링 및 퍼지 시스템을 이용하여 영화에 대한 평가를 예측한 결과는 표 3과 같다. 협업적 필터링으로 실험했을 때와 마찬가지로 각 표본 집단에 대해 10 회 반복 실험하였다. 각 표본 집단에 대한 RMSE 평균과 99%의 신뢰도로서 모집단의 값을 추정한 결과는 표 4와 같다. 표본 집단에 대한 실험 결과를 바탕으로 모집단의 RMSE는 99% 신뢰도로서  $0.9036 \leq m \leq 0.9913$ 을 가질 것이라고 추정할 수 있다. 그림 8은 그림 7과 같이 영화에 대한 평가를 예측하였을 때, RMSE의 결과를 boxplot 그래프로 표현한 것이다. 두 실험의 결과를 통해 표본 집단이 증가할수록 RMSE의 분포는 좁아지고, RMSE 평균도 감소하는 것을 확인할 수 있다. 이것은 표본 집단의 수가 증가함에 따라 다른 사용자와의 유사성 정보 증가하기 때문에 보다 정확한 예측이 가능한 것이다. 현재 Netflix 사에서 제공한 데이터를 이용하여 개발된 시스템 중 가장 정밀한 영화 평가 예측을 나타낸 BellKor 팀의 시스템 예측 오차는 0.9841 이다 [12]. 본 연구에서 제안한 시스템은 모집단의 표본을 대상으로 실험하였으며 BellKor 팀은 모집단 전체를 대상으로 시스템 성능을 평가했기 때문에 직접적인 비교를 하기는 어렵다. 하지만 표본에서의 예측 오차를 통해 99%의 신뢰도 구간으로서 모집단에 대한 오차를 추정할 결과, 제안한 시스템이 기존의 영화 예측 시스템의 성능을 개선할 수 있음을 확인할 수 있다.

표 3. 협업적 필터링 및 퍼지시스템 기반 RMSE 비교  
Table 3. Results of RMSE using collaborative filtering and fuzzy system

	100명	300명	500명	1000명
1	0.8234	0.8117	0.8325	0.8172
2	0.7558	0.8166	0.8222	0.8034
3	0.7436	0.8280	0.8345	0.8183
4	0.8141	0.7972	0.8203	0.8194
5	0.8473	0.7837	0.8077	0.8329
6	0.7964	0.7992	0.7558	0.8082
7	0.8424	0.7816	0.7925	0.7910
8	0.7206	0.7797	0.7765	0.8039
9	0.8065	0.7946	0.8090	0.8480
10	0.7123	0.8034	0.8049	0.7931

표 4. 협업적 필터링을 이용한 신뢰도 결과  
Table 4. Results of reliability using collab. filtering

	100명	300명	500명	1000명
표본 평균	0.97903	0.93782	0.90568	0.92240
99% 신뢰도	$0.9668 \leq m \leq 0.9913$	$0.9333 \leq m \leq 0.9424$	$0.9036 \leq m \leq 0.9078$	$0.9207 \leq m \leq 0.9241$

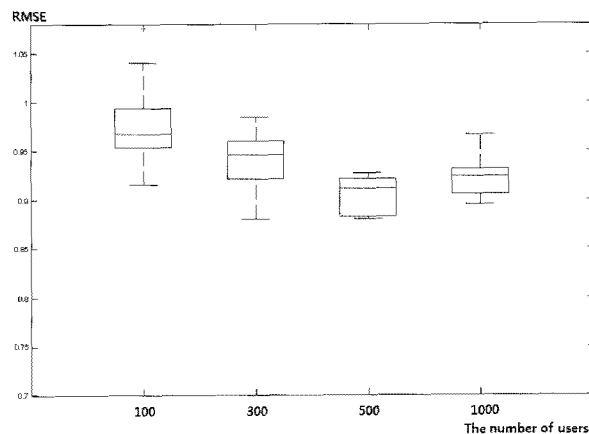


그림 7. 협업적 필터링을 이용한 RMSE 결과  
Fig. 7. Result of RMSE using collaborative filtering

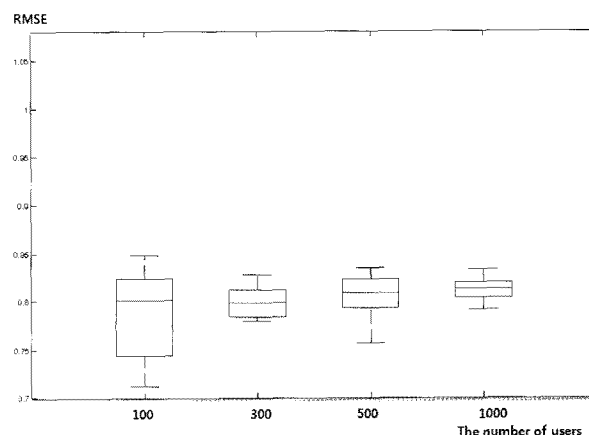


그림 8. 협업적 필터링 및 퍼지시스템 기반 RMSE 결과  
Fig. 8. Result of RMSE using collaborative filtering and fuzzy system

4. 결 론

본 연구에서는 콘텐츠 추천 시스템의 기반 기술인 예측 시스템을 제안하였다. 제안된 예측 시스템은 온라인 영화 대여 회사인 Netflix사에서 제공한 데이터를 바탕으로 개발되었으며 특정 영화에 대한 특정 사용자의 평가를 예측하고자 하였다. 이를 위해 협업적 필터링을 이용하여 전체 사용자 중 평가 예측 대상이 되는 사용자와 유사 성향을 가지는 related user group을 분류하고 그룹에 대한 성향 분석을 통해 영화에 대한 평가를 예측하였다. 또한 새로운 아이템 추가 시 데이터 정보 수집 전까지 추천을 기대하지 못하거나 유사한 그룹의 정보가 없을 때 아이템 추천이 어려운 협업적 필터링의 단점을 보완하기 위하여 퍼지 시스템을 적용하였다.

제안된 예측 시스템의 성능은 데이터를 제공한 Netflix사의 기준에 따라 RMSE 방법으로 계산하였으며 시스템 성능 개선을 목적으로 랜덤 표본 추출된 사용자를 대상으로 실험하였다. 또한 기존의 영화 평가 시스템과의 성능을 비교하기 위해 99% 신뢰도 구간으로 모집단에 대한 RMSE 오차 구간을 추정하였다.

본 실험을 통해 사용자의 성향을 분석하여 영화에 대한 평가를 예측할 수 있었으며 퍼지 시스템을 통해 협업적 필터링의 단점을 보완할 수 있음을 확인할 수 있었다. 이를 바탕으로 제안한 예측 시스템이 추천 시스템의 기반 기술로서 활용이 가능하며, 영화 이외도 음악, 도서와 같은 멀티미디어 추천이 가능할 것으로 기대한다.

Neighborhood-based Collaborative Filtering,"  
KDD 2007 Netflix Competition Workshop, 2007.

저 자 소 개

참 고 문 헌

- [1] Herlocker J, Konstan J, Terveen L, and Riedl J , "Evaluating Collaborative Filtering Recommender Systems," *ACM Transactions on Information Systems* 22, ACM Press, pp. 5-53, 2004.
- [2] G. Adomavicius and A. Tuzhilin, "Towards the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions," *IEEE Transactions on Knowledge and Data Engineering* 17, pp. 634-749, 2005.
- [3] Ansari,A., Essegaiier,S. and RKohli,R., "Internet Recommendation Systems," *Journal of Marketing Reserch*, vol.37, pp. 363-375, 2000.
- [4] Il Im, "Augmenting Knowledge Reuse Using Collaborative Filtering Systems," *A Dissertation Presented to the aculty of the graduate school USC (Information Systems)*, pp. 191, 2001.
- [5] Basu,C., Hirsh,H. and Cohen,W, "Recommendation as Classification : Using Social and Content-based Information in Recommendation," *Proc. of the Fifteenth National Conference on Artificial Intelligence(AAAI-98)*, pp. 714-720, 1998.
- [6] Lang, K., "NewsWeeder : Learning to Filter Netnews," *Inproceedings of the 12th International Conference on Machine Learning*, 1995.
- [7] Pazzani,M., "A Framework for Collaborative, Content-Based and Demographic Filtering," *Artificial Intelligent Review*13(5-6), pp. 393-408, 1999.
- [8] D. Goldberg, D. Nichols, B. M. Oki and D. Terry, "Using Collaborative Filtering to Weave an Information Tapestry," *Communications of the ACM* 35, pp. 61-70 ,1992.
- [9] Konstan,J., Miller,B., Maltz,D., Herlocker,J., Gordon, K. and Riedl,J. "GroupLens :Applying Collaborative Filtering to Usenet News," *Communications of the ACM*, vol. 40, no. 3, pp. 77-87, 1997.
- [10] Rensnick,P., Iacovou,N., Suchak,M., Nergstorm,P. and Riedl.,J. "GroupLens : An Open Architecture for Collaborative Filtering of Netnews," *Proc. of CSCW '94*, pp. 175-186, 1994.
- [11] Shardanand,U. and Maes,P., "Social information filtering : Algorithms for automating 'word of mouth'," *Proc. of ACM CHI '95 Conference on Human Factors in Computing Systems*, pp. 210-217, 1995.
- [12] Robert M. Bell and Yehuda Koren, "Improved



이수진(Soojin Lee)  
2007년 : 부산대학교 정보컴퓨터공학과 (공학사)  
2009년 : 부산대학교 전자전기공학과 (공학석사)  
2009년~현재 : (주)효성 예방진단기술팀

관심분야 : 데이터마이닝, 퍼지이론  
E-mail : pooh3887@pusan.ac.kr



전태룡(Taeryong Jeon)  
2007년 : 신라대학교 컴퓨터정보통신공학과 (공학사)  
2008년~현재 : 부산대학교 전자전기공학과 석사과정

관심분야 : 신경회로망, 지능제어  
E-mail : jtr@pusan.ac.kr



백경동(Gyeongdong Baek)  
2006년 : 부산대학교 전자전기정보컴퓨터공학부(공학사)  
2008년 : 동대학원 전기공학과(공학석사)  
2008년~현재 : 동대학원 전자전기공학부 박사과정

관심분야 : 지능제어, 고장진단  
E-mail : gdbaek@pusan.ac.kr



김성신(Sungshin Kim)  
1986년 : 연세대학교 전기공학과(공학석사)  
1996년 : Georgia Inst. of Technology. 전기및컴퓨터공학과 (공학박사)  
1998년~현재 : 부산대학교 전자전기공학부 부교수

관심분야 : 지능시스템, 데이터마이닝, 고장예측및진단  
Phone : +82-51-510-2374  
Fax : +82-51-513-0212  
E-mail : sskim@pusan.ac.kr