

## 신경망을 이용한 협력적 추천 시스템의 성능 향상

김종수 · 도영아 · 류정우 · 김명원

숭실대학교 컴퓨터학과

### A Collaborative Recommendation System Using Neural Networks for Increment of Performance

Jong-Su Kim · Young-A Do · Joung-Woo Ryu · Myung-Won Kim

Department of Computing, Soongsil University

**Abstract:** Personalization is a filtering technique to reduce information overload and it recently draws attention upon internet users and information providers. The collaborative recommendation is one of the widely used personalization methods. It recommends a user an article based on other users preferences for the target article or the preferences for other articles of the target user. The nearest neighbor method is one of the common collaborative recommendation methods, which recommends articles based on the behavior of the similar users. However, it suffers from poor performance mainly because it fails in properly computing similarity between users. The rule method extracts recommendation rules from the users preference data using data mining technique. It also suffers from simple and superficial association between users or articles without utilizing conceptual level of association such as commonality in contents and category.

In this paper we propose neural networks model to improve the accuracy of recommendations. We investigate various aspects of collaborative recommendation based on multi-layer perceptrons by varying parameters and the dimension of input data. We also combine the collaborative recommendation and the content-based recommendation within a neural network model to solve sparsity problem in pure collaborative recommendation. We finally demonstrate through experiments that the neural network based collaborative recommendation outperforms the existing methods.

Key words: Recommendation System, Neural Networks, Collaborative Filtering

**요약:** 개인화 기술은 과다한 정보를 줄이기 위한 여과 기술로서 최근 인터넷 사용자와 정보 제공자들로부터 관심의 대상이 되고 있다. 그 중 가장 많이 사용되고 있는 개인화 기술은 협력적 추천으로서, 이 방법은 사용자에게 어떤 항목(상품)을 추천해주는 데 있어서 그 사용자의 다른 항목들에 대한 선호도나 추천 항목에 대한 다른 사용자들의 선호도를 이용하여 추천해준다. 협력적 추천 방법 중 보편적으로 많이 사용되는 최근접 이웃 방법은 유사한 사용자들의 행위에 기반하여 항목들을 추천해 준다. 그러나 이 방법은 사용자들간의 단순한 유사도를 계산함으로써 성능이 떨어지는 문제점이 있다. 데이터마ining 기법을 적용하여 사용자 선호도 데이터로부터 추천 규칙을 생성하여 추천을 해주는 연관규칙 방법 또한 사용자나 항목간의 내용적, 분류적인 공통성 또는 개념에 의한 연관성을 고려하지 못하고, 단지 단순하고 표면적인 연관성만을 고려하는 문제점을 가지고 있다.

본 논문에서는 추천의 정확도를 향상시키기 위한 신경망 추천 모델을 제안한다. 신경망으로는 다층 퍼셉트론을 사용하여 우선 학습변수들과 입력차원의 변화에 따른 신경망에 기반한 협력적 추천의 성능에 대해 조사하고, 순수한 협력적 추천에서의 문제점인 희소성(sparsity)문제를 해결하기 위해 내용기반 추천과 협력적 추천을 병합한 모델을 신경망에서 보인다. 마지막으로 다른 협력적 추천 방법과의 비교를 통하여 본 신경망 추천 방법의 장점과 성능의 우수함을 보인다.

**주요어:** 추천 시스템, 신경망, 협력적 여과

교신저자 : 김종수

숭실대 컴퓨터학과, 156-743 서울시 동작구 상도 1동 1-1

Tel: 02-813-8755

Fax: 02-822-3622

E-mail: kjongsu@origio.net

## 서론

현재 웹사이트 상에서 이루어지고 있는 추천 서비스는 amazon, cdnow, GroupLens, Ringo, Firefly, NetPerception, WebWatcher, Letzia 등의 인터넷업체에서 사용자의 구매기록, 즉 취향에 맞는 항목(items) 및 정보를 사용하여 사용자에게 추천해주고 있다. 여기서 사용하는 추천 기술은 크게 협력적 추천(Collaborative filtering), 내용기반 추천(Content-based filtering), 인구통계적 추천(Demographic filtering)을 들 수 있다<sup>1</sup>. 협력적 추천은 특정 사용자와 유사한 선호도(rating)를 갖는 다른 사용자들의 목표 항목에 대한 선호도를 바탕으로 그 항목의 선호도를 추정하는 기술이며, 내용기반 추천은 사용자에게 있어 이전에 선호한 항목과 비슷한 속성을 갖는 항목은 사용자가 선호할 가능성이 높다고 보고 선호도가 표시된 항목들의 속성 정보를 이용하여 추천하는 기술이다. 이에 비해 인구통계적 추천은 사용자의 나이, 성별, 생활수준 등과 같은 인구통계적 정보를 바탕으로 항목이나 정보의 선호도를 추정한다. 기존의 협력적 추천 중에서 대표적인 방법으로는 최근접 이웃 방법(Nearest Neighbor Method)과 연관규칙 방법(Association Rule Method) 등이 있다<sup>2, 3, 4, 5</sup>. 최근접 이웃 방법은 적용하기가 용이한 반면 추천의 정확도가 크게 떨어지는 문제점이 있다. 그 이유는 한 사용자의 어떤 항목에 대한 선호도를 결정하는 데 있어서 사용자들과의 유사도를 계산하는 데 있어서 항목간의 중요도, 즉 가중치를 고려하지 못하는 것이다. 반면, 연관규칙 방법은 다른 항목에 대한 선호도 자료로부터 데이터 마이닝 기법을 적용하여 항목 선호에 대한 연관규칙을 추출하고 그 규칙을 사용하여 어떤 항목의 선호도를 추정한다. 따라서 항목들간의 중요도가 연관규칙의 지지도(support)나 신뢰도(confidence) 등으로 나타난다고 할 수 있으나 단순히 항목들간의 연관관계, 즉 표면적인 연관관계에 의하여 선호도를 결정함으로써 항목들간의 내용적인 공통성 또는 상위 개념에 의한 선호도가 고려되지 않음으로써 역시 성능이 떨어지는 문제점이 있다. 연관규칙은 그 특성상 선호도 또는 어떤 데이터가 연속적인 값으로 표시될 경우 이를 이진 값으로 변환해야 하는 문제점이 있다. 또 빈도수가 적은 선호도 항목일 경우 그 항목에 대한 규칙을 생성할 수 없다.

본 논문에서 제안한 추천 방법은 신경망을 이용하는

것으로 항목들간의 선호 상관관계나 사용자들간의 선호 상관관계를 신경망으로 학습시킴으로써 모델을 생성하고 그 모델을 사용하여 선호도를 추정한다. 이 방법은 기존의 추천 방법이 가지고 있는 문제점들을 다음과 같은 장점으로 해결할 수 있다. 첫째, 항목이나 사용자들간의 가중치를 학습할 수 있으므로 보다 정확한 선호도 계산이 가능할 뿐 아니라 신경망 은닉노드의 개념 형성 기능으로 보다 효율적인 선호도 산출이 가능하다. 둘째, 연속 수치형, 이진 논리형, 범주형 등의 자료 유형에 상관없이 데이터의 처리가 용이하다. 셋째, 다른 이질적(내용, 인구통계적 정보 등)인 종류의 데이터 및 정보를 통합하기 용이하다. 한편, 기존 방법에서는 내용기반 추천, 인구통계적 추천이 각각 다른 방법으로 수행되는 데 비하여 신경망 모델에서는 항목에 대한 내용 정보나 사용자의 인구통계적 정보를 단순히 노드에 추가하여 신경망을 학습시킴으로써 이들 방법들을 용이하게 통합할 수 있는 장점이 있다.

본 논문은 다음과 같은 순서로 구성되어 있다. 2절에서는 추천성능에 대한 평가 척도와 기존의 협력적 추천 방법들에 대하여 설명하고, 3절에서는 제안하는 신경망 추천 모델에 대해 설명하며, 4절에서는 EachMovie 데이터를 가지고 추천에 있어서 신경망 학습 파라미터가 미치는 영향에 대해서 분석하고, 내용기반추천과 협력적 추천을 병합한 신경망 추천 방법의 실험 결과를 보인다. 또한, 연관규칙 추천 방법과 실험 결과를 비교한다. 마지막으로 5절에서는 결론을 맺는다.

## 관련 연구

### 추천 성능에 대한 평가척도

추천에서의 성능 평가척도는 주로 기계학습에서 사용하는 accuracy와 정보검색에서 사용하는 표준 척도인 precision과 recall이 있다. 각각은 식 1, 2, 3, 4와 같이 정의된다.

따라서 추천 성능에 대한 평가에 있어 accuracy는 시스템에 의해서 분류된 항목들이 얼마나 옳게 분류되었는가를 나타내는 척도이며, precision은 추천된 항목들 중에 사용자가 선호하는 항목들의 비율을 나타내고, recall은 사용자가 선호하는 항목 수에 대한 추천된 항목 수의 비율을 나타낸다. 한편, [6]에서는 precision과

$$\text{accuracy} = \frac{(\text{추천된 항목 중에서 사용자가 선호하는 항목 수}) + (\text{추천되지 않은 항목 중에서 사용자가 선호하지 않는 항목 수})}{\text{선호도가 표시된 전체 항목 수}} \quad (\text{식 1})$$

$$\text{precision} = \frac{\text{추천된 항목 중에서 사용자가 선호하는 항목 수}}{\text{추천된 항목 수}} \quad (\text{식 2})$$

$$\text{recall} = \frac{\text{추천된 항목 중에서 사용자가 선호하는 항목 수}}{\text{전체 항목 중에서 사용자가 선호하는 항목 수}} \quad (\text{식 3})$$

$$F = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (\text{식 4})$$

recall을 가중치 통합한 F-measure를 제안하였다. 식 4는 precision과 recall을 같은 가중치로 하였을 경우의 F-measure이다.

### 기존의 협력적 추천 방법

현재 인터넷상에서 이루어지고 있는 개인화 서비스로는 사용자의 프로파일이나 구매기록(선호도)을 통해 취향에 맞는 항목 및 정보(컨텐츠)를 사용자에게 추천해 주는 것이 대표적이다. 이러한 추천 서비스에서 가장 중요한 단계는 선호도가 유사한 사용자 또는 항목들을 결정하는 데 있다. 따라서 이러한 결정 방법에 대한 연구가 활발히 진행되고 있으며 그 중 가장 대표적인 방법은 다음과 같다.

#### 최근접 이웃 방법(Nearest Neighbor Method)

가장 가까운 이웃을 찾아 새로운 사용자에게 대한 예측 및 분류작업을 하는 데 사용되는 방법이다. 예측 및 분류를 위해 먼저 고객 전체 자료로부터 학습자료를 생성하는 메모리 기반(Memory-based) 협력적 필터링 알고리즘이다. 이 방법은 새로운 사용자에게 대하여 전체 고객 자료로부터 가장 가까운 k개의 근접이웃(K-Nearest Neighbor)을 선택하여 다수결 원칙 또는 근접 정도에 따른 가중치 평균으로 분류 또는 예측 값을 계산하는 방법이다<sup>7</sup>.

예를 들어, 사용자  $a$ ,  $b$ 의  $m$ 개의 항목에 대한 선호도(rating)가  $a(r_{a1}, r_{a2}, \dots, r_{am})$ 와  $b(r_{b1}, r_{b2}, \dots, r_{bm})$ 로 주어졌다

면 두 사용자의 유사성은 다음과 같이 여러 가지 척도로써 나타낼 수 있다.

◆ 유클리드 거리(Euclidean distance) : 두 사용자를  $m$ 차원 공간에서의 두 벡터로 표시하여 두 사용자간의 유사성을 유클리드 거리를 사용하여 측정한다.

$$\text{dist}(a, b) = \sqrt{\sum_{i=1}^m (r_{ai} - r_{bi})^2} \quad (\text{식 5})$$

◆ 코사인 유사도 (cosine similarity) : 두 사용자를  $m$ 차원 공간에서의 두 벡터로 표시하여 두 사용자간의 유사성을 두 벡터 사이의 코사인(cosine) 각도를 사용하여 측정한다.

$$\cos \theta = \frac{a \bullet b}{\|a\|_2 \times \|b\|_2} \quad (\text{식 6})$$

(식 6)에서 “ $\bullet$ ”은 두 벡터의 내적(dot-product)을 의미하고, “ $\| \cdot \|_2$ ”은 벡터의 크기를 나타낸다.

◆ 상관관계 (correlation) : 두 속성간의 선형관계의 정도를 측정할 수 있는 척도인 피어슨 상관 계수(Pearson correlation coefficient)를 이용하여 두 사용자간의 유사성을 측정한다<sup>3</sup>.

$$\text{corr}_{ab} = \frac{\sum_i (r_{ai} - \bar{r}_a)(r_{bi} - \bar{r}_b)}{\sqrt{\sum_i (r_{ai} - \bar{r}_a)^2 \sum_i (r_{bi} - \bar{r}_b)^2}} \quad (\text{식 7})$$

단,  $\bar{r}_a$ 은  $a$ 의 평균, 즉 사용자  $a$ 가  $m$ 개의 항목에 부여한 등급의 평균을 나타낸다.

이와 같은 척도를 사용하여 유사한 사용자 또는 항목

을 측정하여 추천하는 방법은 적용하기 용이하다는 장점을 가지고 있지만, 사용자 또는 항목들간의 가중치를 고려하지 못함으로써 추천의 정확도가 떨어지는 단점이 있다. [1]에서는 피어슨 상관 계수를 이용하여 추천한 결과 67.9%의 정확성을 보이고 있다. 또 항목의 종류가 많은 데이터일 경우 희소성(sparsity)문제가 발생한다. 희소성 문제란 사용자가 선호도를 표시한 항목의 개수가 적을 경우 사용자 간의 유사성이 왜곡될 수 있는 문제를 말한다. 또한 항목의 종류뿐만 아니라 사용자 수가 많을 경우 알고리즘 수행속도가 느려지는 범위성(scalability)문제도 고려해야 한다<sup>2,9</sup>.

### 연관규칙 방법(Association Rule Method)

데이터 마이닝 기법 중 하나인 연관규칙 발견은 데이터들의 빈도 수와 동시 발생 확률을 이용하여 데이터와 데이터간의 관계를 찾는 방법이다. 여러 데이터 중에서 서로 관련이 깊은 데이터를 찾아 관련성을 규칙으로 표현한다. 추출된 규칙은 사람이 이해하기 쉬운 형태로 되어 있고 쉽게 적용할 수 있기 때문에 많은 연구가 진행되어 왔으며, 실제 상용시스템에서 여러 가지 알고리즘의 형태로 응용되고 있다.

연관규칙(R : A B)에 대하여 지지도는 생성된 연관규칙이 전체 아이템에서 차지하는 비율을 말한다. 즉, 데이터베이스에 속한 전체 트랜잭션 중 A와 B를 포함하는 트랜잭션의 비율을 의미한다.

$P(A \& B) = (\text{A와 B를 포함하는 트랜잭션 수}) / (\text{전체 트랜잭션 수})$

신뢰도는 연관규칙의 강도를 의미하며 전체부를 만족하는 트랜잭션이 결론부를 만족하는 비율, 즉 A를 포함하는 트랜잭션 중에서 B가 포함된 트랜잭션의 비율을 의미한다.

$P(B|A) = (\text{A와 B를 포함하는 트랜잭션 수}) / (\text{A를 포함하는 트랜잭션 수})$

빈발항목집합은 일정수준 이상의 지지도를 가지는 모든 항목집합을 추출하는 것이며 빈발항목집합이 추출되면 이를 이용하여 연관규칙을 생성한다. 연관규칙 알고리즘에는 기본적으로 Apriori 알고리즘이 사용되고 있으며, 이 알고리즘은 빈발항목집합을 계산하는 데 있어서 한 부분집합의 빈도수는 그것을 포함하는 집합의

빈도수보다 커야 한다는 조건을 바탕으로 한다. 예를들면 일정 빈도수 이상의 단일항목집합으로부터 2-항목집합의 후보들이 생성되고 이들 후보집합들은 다시 그 빈도수가 측정됨으로써 빈발항목집합의 여부가 판정된다. 이와 같은 방법을 반복함으로써 모든 가능한 일정 빈도수 이상의 빈발항목집합을 효율적으로 계산할 수 있다.

이러한 연관규칙이 최근 추천 시스템에서 많이 적용되고 있다. [8]에서는 수집된 사용자의 웹 페이지 네비게이션(navigation) 자료로부터 연관규칙을 추출하여 사용자에게 웹 페이지를 추천하는 시스템을 개발하였다. 그러나 기존에 사용하고 있는 연관규칙 알고리즘들은 추천서비스를 위해 적합하지 못하는데 그 이유는 기존의 알고리즘들은 최소 신뢰도에 대한 적당한 최소 지지도를 선택하는 방법과 사용자가 원하는 규칙의 개수를 제공하지 못하기 때문이다. 이는 너무 많은 규칙을 생성하여 계산 시간을 증가시키거나 혹은 너무 적은 규칙을 생성하여 추천의 precision 및 recall을 저하시키는 요인이 된다.

[5]에서는 최소 신뢰도에 대한 사용자가 원하는 규칙의 개수가 생성될 수 있게 최소 지지도를 자동적으로 조정할 수 있을 뿐만 아니라 미리 선택된 한 개의 항목만이 규칙의 결론부에 나타나도록 알고리즘을 확장하였다. [9]에서는 [5]에서 제안한 알고리즘을 이용한 협력적 추천 방법을 제안하였으며 추천에 사용된 규칙의 형태는 <그림 1>과 같다. EachMovie에 대한 추천 성능은 최소 신뢰도가 90%일 경우 accuracy 72%, precision

사용자 연관규칙	
IF [사용자 <sub>1</sub> : like] AND [사용자 <sub>2</sub> : like] THEN [목표 사용자 : like] (신뢰도 : 90%, 지지도 : 70%) : 한 항목에 대하여 사용자 <sub>1</sub> 이 좋아하고 사용자 <sub>2</sub> 가 좋아하면 목표 사용자도 좋아한다.	
항목 연관규칙	
IF [항목 <sub>1</sub> : like] AND [항목 <sub>2</sub> : like] THEN [목표 항목 : like] (신뢰도 : 90%, 지지도 : 60%) : 사용자가 항목 <sub>1</sub> 을 좋아하고 항목 <sub>2</sub> 를 좋아하면 목표 항목도 좋아한다.	

그림 1. 추천을 위한 연관규칙의 예

74%, recall 60%의 성능을 얻었다.

연관규칙의 경우 항목들이나 사용자들 간의 가중치가 연관규칙의 지지도나 신뢰도 등으로 나타난다고 할 수 있으나 단순히 표면적인 연관관계에 의하여 선호도를 결정하기 때문에 항목들간의 어떤 내용적인 공통성 또는 어떤 상위 개념에 의한 선호도가 고려되지 않음으로써 역시 정확도가 떨어지는 문제점이 있다. 또한, 연관규칙 특성상 선호도 또는 데이터가 연속적인 값으로 표시될 경우 이를 이진 값으로 변환해야 하며, 빈도수가 적은 선호도 항목일 경우 그 항목에 대한 규칙을 생성할 수 없는 문제점이 있다.

### 신경망 방법(Neural Networks Method)

신경망은 고등동물 두뇌의 정보처리 메커니즘을 모방한 새로운 계산 방법으로 특히 학습기능을 가짐으로써 기존의 컴퓨터(von Neumann computer)가 효율적으로 해결하기 어려운 문제, 즉 효율적인 알고리즘이 없는 비 알고리즘적인 문제를 용이하게 해결할 수 있고 잡음에 강해 패턴인식 및 제어 등 여러 문제에 많이 응용되고 있다.

Daniel Billsus and Michael J. Pazzani<sup>6</sup>는 SVD(Singular Value Decomposition)를 사용하여 신경망 추천 방법을 제안하였다. [6]에서 제안된 방법은 학습알고리즘에서 결측값(missing value)을 처리하기 위해 <표 1>과 같이 사용자 선호도를 학습데이터로 변환시킨다. 특히, <표 1>은 '항목<sub>5</sub>'에 대한 '사용자<sub>4</sub>'의 선호도를 예측하기 위해 '사용자<sub>4</sub>'의 선호도가 표시된 항목들만을 선택하여 변환시킨 학습데이터를 보여주고

있다. 사용자 선호도는 4등급(1,2:dislike, 3,4:like)으로 나타내고, 학습데이터는 사용자가 항목을 선호한다면 이진값 "10"으로 선호하지 않는다면 이진값 "01"로 나타낸다. 특히, 항목에 대한 선호를 알 수 없는 결측값의 경우 이진값 "00"으로 표시한다. 결측값이 표현된 학습데이터에 SVD를 적용하여 차원을 축약한 후, 항목에 대한 사용자들의 정보를 신경망의 입력으로 받고 목표 사용자('사용자<sub>4</sub>')의 정보를 신경망의 출력으로 하여 학습하는 방법이다.

실험에 사용된 EachMovie 데이터에 대한 세부적 설명은 4.1에서 기술하고 있다. [6]에서는 전체 사용자 중 앞에서부터 2,000명의 사용자만을 실험에 사용하였으며 사용자들이 선호를 표시한 영화(항목)는 1,410편이 된다. 목표 사용자는 선호도를 표시한 영화가 80편 이상이고 선호하는 영화의 비율이 0.75 이하를 보이는 사용자들 중에서 무작위로 선택하였고, 학습 데이터와 테스트 데이터의 수를 같게 하였다. 실험에서 입력 선호도 항목 수에 따른 예측 결과 값을 보여주고 또한, accuracy 68%, F-measure 69%의 추천 성능을 보여주고 있다. SVD는 데이터에서의 노이즈(noise)를 제거하고, 희소성 문제를 해결하는데 도움을 준다. 그러나, 계산 시간이 많이 걸리는 문제점이 있다.

위에서 언급한 최근접이웃, 연관규칙, 신경망 방법 외에도 베이지안 네트워크(Bayesian Networks), 클러스터링(Clustering), 잠재 변수(Latent Variable), 결정 트리(Decision Tree), 진화 알고리즘(Genetic Algorithm) 등과 같은 많은 기술들이 추천 시스템에 적용되어 왔다. 그 중 베이지안 네트워크에서는 노드가 항목에 대응하

표 1. 선호도의 결측값 표현을 위한 변환

	항목 <sub>1</sub>	항목 <sub>2</sub>	항목 <sub>3</sub>	항목 <sub>4</sub>	항목 <sub>5</sub>
사용자 <sub>1</sub>	4		3		
사용자 <sub>2</sub>		1		2	
사용자 <sub>3</sub>	3	4	2		4
사용자 <sub>4</sub>	4	2	1		?

사용자 선호도

	항목 <sub>1</sub>	항목 <sub>2</sub>	항목 <sub>3</sub>
사용자 <sub>1</sub> like	1	0	1
사용자 <sub>1</sub> dislike	0	0	0
사용자 <sub>2</sub> like	0	0	0
사용자 <sub>2</sub> dislike	0	1	0
사용자 <sub>3</sub> like	1	1	0
사용자 <sub>3</sub> dislike	0	0	1
목표 사용자 (사용자 <sub>4</sub> )	1(like)	0(dislike)	0(dislike)

'사용자<sub>4</sub>'에 대한 학습데이터

는데 각 노드의 상태는 그 노드에 대응하는 항목의 선호를 나타낸다. 한 항목은 여러 개의 부모 항목을 가지며 이들 항목들간의 조건부 확률을 결정 트리로서 나타낸다. 이와 같은 베이지안 네트워크는 항목들 간의 관련성을 학습을 통하여 결정된다. 이와 같이 생성된 네트워크 모델은 매우 규모가 적고, 빠르면서 최근접 이웃 방법과 유사한 성능을 갖는다<sup>3</sup>. 클러스터링 방법은 유사한 선호도를 가지는 사용자들을 그룹화 함으로써 수행된다. 클러스터 그룹이 생성되면 한 개인에 대한 예측은 그 클러스터에 속하는 다른 사용자들의 선호도를 평균화함으로써 이루어질 수 있다. 어떤 경우에는 최근접 이웃방법보다 성능이 낮으나, 클러스터링이 잘 되었을 경우 성능이 좋아질 수 있다<sup>3</sup>. 클러스터링 방법은 또한, 최근접 이웃 방법에서의 후보 집합을 만들기 위한 전 단계로서 사용될 수도 있다.

다음은 협력적 추천과 내용기반 추천을 병합한 여러 추천 방법들을 보여준다. GroupLens는 내용기반 추천을 지원하는 Usenet news를 위한 혼합 협력적 추천 시스템을 구현하였다<sup>10</sup>. 이 filterbot은 희소성을 해결하고 있으나 여전히 초기 평가 문제를 가지고 있다. Fab<sup>11</sup>은 웹 페이지를 추천하기 위한 내용기반 협력적 추천 통합 시스템을 구현하였다. Fab에서 사용자 프로파일은 내용기반 기술에 의해서 유지된다. 그 프로파일은 협력적 추천을 하기 위한 사용자들간의 유사도를 계산하기 위해 사용된다. 따라서 정확한 프로파일을 요구한다. 부정확한 프로파일은 협력적 추천의 예측력을 크게 줄이기 때문이다. Basu, Hirsh와 Cohen<sup>12</sup>은 사용자 영화 선호도를 예측하기 위해 선호도 정보와 내용 정보를 사용한 귀납학습 방법을 적용하였다. 그들은 협력적 추천이나 내용기반 추천을 단독적으로 사용하였을 때 보다 더 좋은 추천을 하기 위해 시도하였다. 그러나 더 좋은 결과를 얻기 위해 추가적인 혼합 요소들을 생성하기 위한 노력을 요구한다. ProfBuilder<sup>13</sup>는 내용기반 추천과 협력적 추천 둘 다를 사용하여 웹 페이지를 추천한다. 하지만 결합된 예측을 하지는 않는다. Michael J. Pazzani<sup>1</sup>는 내용기반을 통한 협력적 추천을 레스토랑 데이터를 사용하여 실험하였다. 협력적 추천은 사용자들 사이의 공통 선호도 항목이 많을 때 의미있는 상관관계를 얻을 수 있으므로 처음 도시를 방문한 사용자에게 대하여 공통 선호도를 가진 사용자들이 없을 경우 내용기반 프로파일을 사용하여 추천함으로써 해결하였다.

## 신경망 추천 모델

신경망을 이용한 추천 기술은 그 우수함이 보여지지만<sup>6</sup> 아직 연구가 부족한 편이다. 본 논문에서는 추천의 정확도를 향상시키기 위한 신경망 추천 모델을 제안한다. 또한, 내용기반 추천과 협력적 추천을 병합한 신경망 추천 모델을 제안한다.

### 기본 신경망 모델

신경망 모델은 다층퍼셉트론(Multi-Layer Perceptron)을 사용하고, 학습 알고리즘으로는 EBP(Error-BackPropagation) 알고리즘을 사용한다. 다층퍼셉트론은 <그림 2>와 같이 일반적으로 세 계층(입력층, 은닉층, 출력층)으로 구성되어 있고 각각의 노드는 다음 층의 모든 노드에 연결된 전향 신경망이다. 연결은 관련된 가중치를 갖는다. 여기서 은닉층 j번째 노드는 자신의 입력들에 대한 가중합,  $x_j = \sum_i w_{ij} I_i$ 를 갖고 비선형 활성화함수  $f(x) = \frac{1}{1+e^{-x}}$ 에 의하여 출력값이 결정된다.

신경망은 다음 식 8과 같이 입력(I)와 출력(O) 사이의 비선형 사상 함수를 제공한다.

$$o = F(I; w, A) \quad (\text{식 8})$$

여기서  $w$ 는 신경망의 가중치이며,  $A$ 는 신경망 구조를 의미한다. 입력이 제시될 때 신경망으로부터 목표 출력값을 생성하도록 하기 위하여 신경망의 가중치를 적절히 조정하는 것이다. 이것을 신경망의 학습이라고 말한다. 다층퍼셉트론의 학습 방법으로는 EBP 학습 알

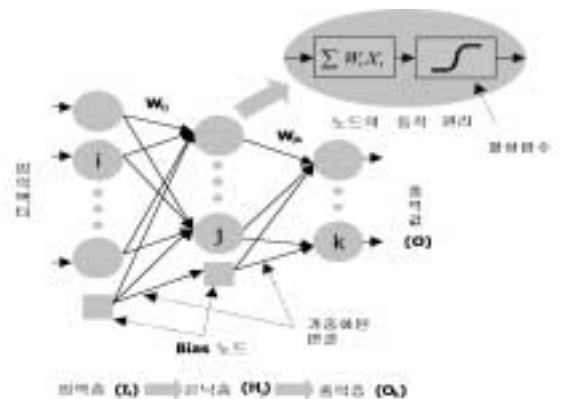


그림 2. 다층 퍼셉트론 구조

고리즘이 대표적이며 가장 많이 응용되고 있다.

EBP 학습의 목적은 신경망의 가중치들에 대하여 식 9를 반복적인 방법으로 E(error)를 최소화하도록 기울기 하강 방법을 사용하여 가중치를 결정하는 것이다.

$$E = \frac{1}{2} \sum_k (O_k - T_k)^2 \quad (\text{식 9})$$

학습 중, 어느 한 시점에서의 가중치의 변화량은 식 10과 같다.

$$\Delta W \propto -\frac{\partial E}{\partial W} \quad (\text{식 10})$$

따라서, 각 계층의 가중치 변경에 대한 일반식은 다음과 같다.

— 출력층과 은닉층의 사이의 가중치 변경

$$W_{kj}(t+1) = W_{kj}(t) + \eta \delta_k o_j$$

$$\delta_k = (t_k - o_k) o_k (1 - o_k) \quad (\text{식 11})$$

— 은닉층과 입력층 사이의 가중치 변경

$$W_{ji}(t+1) = W_{ji}(t) + \eta \delta_j o_i$$

$$\delta_j = \sum_k \delta_k W_{kj}(t) o_k (1 - o_k) \quad (\text{식 12})$$

여기서  $w_{ji}$ 는 노드  $i$ 에서  $j$ 로 가는 가중치이며,  $\eta$ 는 학습률(learning rate)이라고 한다.

### 사용자와 항목 신경망 모델

협력적 추천에는 사용자 모델과 항목 모델이 있다<sup>1, 2, 3</sup>. 협력적 추천에서의 사용자 모델은 사용자들 간의 연관성을 바탕으로 목표 사용자에게 새로운 항목을 추천한다. 예를 들면, <그림 3>에서 보는 바와 같이 사용자들

	사용자 <sub>1</sub>	사용자 <sub>2</sub>	사용자 <sub>3</sub>	사용자 <sub>4</sub>	사용자 <sub>5</sub>
영화 <sub>1</sub>	-	+	+	+	-
영화 <sub>2</sub>	+	+	+	+	+
영화 <sub>3</sub>	+	-	+	-	+
영화 <sub>4</sub>	-	+	-	+	-
영화 <sub>5</sub>	+	-	+	-	?

예측

그림 3. 사용자 모델

의 영화에 대한 선호도(+는 like, -는 dislike)를 나타낸다고 할 때 '사용자<sub>5</sub>'에게 새로운 영화 '영화<sub>5</sub>'를 추천할 것인가를 예측하기 위해 '사용자<sub>5</sub>'와 유사한 선호도를 보이는 다른 사용자, 즉 '사용자<sub>1</sub>'의 '영화<sub>5</sub>'에 대한 선호도를 참조하여 추정하는 방법이다.

이와는 반대로 항목 모델은 <그림 4>에서와 같이 서로 다른 사용자들이 평가한 항목들 간의 연관성을 바탕으로 하여 새로운 항목을 추천한다. 예를 들면, '사용자<sub>5</sub>'에게 새로운 영화 '영화<sub>5</sub>'를 추천할 것인가를 추정하기 위해 '영화<sub>5</sub>'와 가장 유사한 사용자들의 선호도를 보이는 다른 영화, 즉 '영화<sub>3</sub>'에 대한 '사용자<sub>5</sub>'의 선호도를 참조하여 추천한다.

사용자 신경망 모델은 사용자 모델의 경우, 목표 사용자와 다른 사용자들간의 선호도 연관성을 여러 항목에 대하여 신경망으로 학습하여 모델을 생성한 다음, 새로운 항목이 주어지면 그 항목에 대한 다른 사용자들의 선호도를 입력으로 하여 목표 사용자의 그 항목에 대한 선호도를 추정한다(그림 5). 이에 반하여 항목 신경망 모델은 목표 항목과 다른 항목들간의 연관성을 여러 사용자에게 대하여 신경망으로 학습하여 모델을 생성한다. 새로운 사용자가 주어지면 그 사용자의 다른 항목에 선호도를 입력으로 하여 그 사용자의 목표 항목에 대한 선호도를 추정한다(그림 6).

### 내용 및 인구통계적 정보를 고려한 신경망 모델

내용 및 인구통계적 정보를 고려한 신경망 모델은 <그림 7>에서 보여 주고 있다. 사용자 신경망 모델에 다른 사용자들의 선호도 정보와 함께 항목 내용, 즉 영

	사용자 <sub>1</sub>	사용자 <sub>2</sub>	사용자 <sub>3</sub>	사용자 <sub>4</sub>	사용자 <sub>5</sub>
영화 <sub>1</sub>	-	+	+	+	-
영화 <sub>2</sub>	+	+	+	+	+
영화 <sub>3</sub>	+	-	+	-	+
영화 <sub>4</sub>	-	+	-	+	-
영화 <sub>5</sub>	+	-	+	-	?

예측

그림 4. 항목 모델

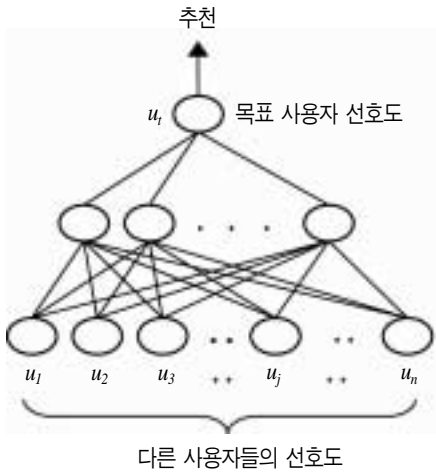


그림 5. 사용자 신경망 모델

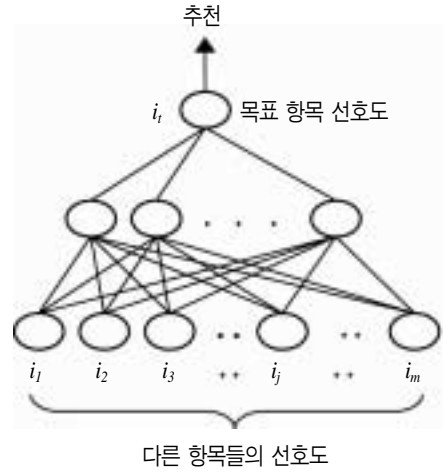


그림 6. 항목 신경망 모델

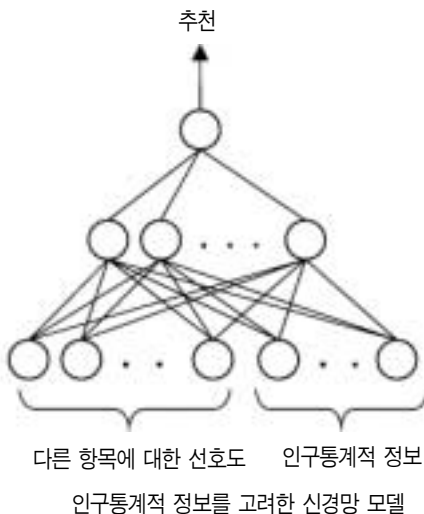
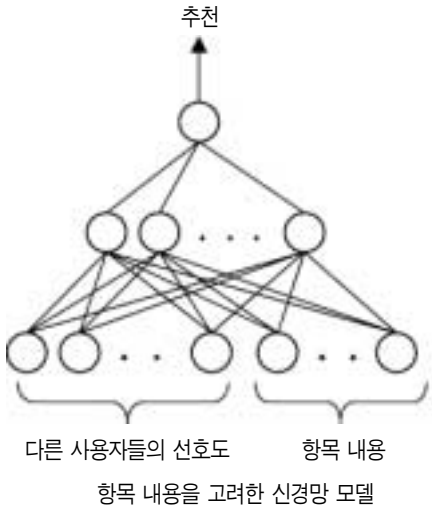


그림 7. 항목 내용 및 인구통계적 정보를 고려한 신경망 모델

화장르, 배우, 감독 등의 정보를 추가하여 학습하면 내용기반 추천과 협력적 추천을 병합한 모델이 생성된다. 또한, 항목 신경망 모델에 다른 항목들의 선호도 정보와 함께 인구통계적 정보, 즉 사용자의 성별, 나이, 주소 등의 정보를 추가하여 학습하면 인구통계적 추천과 협력적 추천을 병합한 모델이 만들어진다.

내용 및 인구통계적 정보를 고려하여 신경망을 구현하면 초기 사용자의 경우 협력적 추천에서 추천을 받을 수 없는 문제점을 해결할 수 있으며, 평가 정보가 적은 사용자가 상품을 추천받을 경우에도 어느 정도 성능을 향상시킬 수 있다.

제한한 신경망 추천 방법은 항목들간의 선호 상관관계나 사용자들간의 선호 상관관계를 신경망으로 학습 시킴으로써 모델을 생성하고 그 모델을 사용하여 선호도를 추정한다. 기존방법에 있어 최근접 이웃 방법은 적용하기는 용이하나 사용자 또는 항목들간의 가중치를 고려하지 못함으로 추천의 정확도가 떨어지는 문제점이 있으며, 연관규칙 또한 표면적인 연관관계에 의하여 선호도를 결정함으로써 상품들간의 어떤 내용적인 공통성 또는 어떤 상위 개념에 의한 선호도가 고려되지 않음으로써 역시 정확도가 떨어지는 문제점과 적용에 있어 연속 수치형인 경우 범주형으로 변환해야 하는 제약사항을 받는다는 문제점을 가지고 있다. 제한한 방법은 항목이나 사용자들간의 가중치를 학습할 수 있으므로 보다 정확한 선호도 계산이 가능할 뿐 아니라 신경망 은닉노드의 개념 형성 기능으로 보다 효율적인 선호



도 산출이 가능하며, 연속 수치형, 이진 논리형, 범주형 등의 자료 유형에 상관없이 데이터의 처리가 용이하다. 또한 <그림 7>과 같이 다른 이질적(내용, 인구통계적 정보 등)인 종류의 데이터 및 정보를 통합하기 용이하다.

## 실험 평가

### 실험 데이터

실험 데이터는 EachMovie<sup>14</sup> 데이터로써 72,916명의 사용자와 1,628개의 영화로 구성되어 있으며 각 고객이 본 영화에 대해서 평가한 선호도(rating)는 0, 0.2, 0.4, 0.6, 0.8, 1.0의 6단계 수치로 표현되어 있다. 본 실험에서는 최소 100편 이상 영화에 대한 선호도를 입력한 사용자 1,000명을 선택하였다. 또 선호도가 편향된 사용자를 제외하기 위하여 사용자의 영화에 대한 like 빈도(ratio)가 0.6~0.4인 사용자를 선택하였다. 학습 데이터(learning data)와 테스트 데이터(test data)는 각각 3:1로 나누어 실험하였다. 또 실험에 사용한 다층 퍼셉트론은 사용자 모델, 항목 모델 다같이 입력 노드 100개, 은닉노드 5개 그리고 출력노드 1개인 구조를 사용하였다. 특히 은닉 노드 수를 5개로 한 것은 신경망의 과다 학습(overfitting)을 피하고 일반화 성능을 높이기 위하여 가능한 한 작은 개수로 하였다.

또한, 연관규칙 추천과의 비교를 위해 같은 조건인 최소 100편 이상 영화에 대한 선호도를 입력한 사용자 1,000명을 학습 데이터로 사용하고, 70,000 이상인 사용자 ID에서 최소 100편 이상 영화에 대한 선호도를 입력한 사용자 100명을 무작위로 선택하여 테스트 데이터로 사용하여 실험하였다<sup>4, 5</sup>.

### 실험 결과

#### 학습 파라미터가 미치는 영향 분석

학습률과 학습회수(learning epoch)는 학습속도와 과다 학습에 밀접하게 관련된다. 학습률의 경우 보통 0.001~10 사이의 값을 사용하는데, 너무 큰 값의 학습률은 빠르게 학습이 진행될 수도 있지만 자칫하면 학습이 안 되고 진동하는 상황이 발생할 수 있다. 반대로 너무 작은 학습률은 오차가 작아지는 방향으로 학습이 이루어져서 최종적으로 오차 최소점에 수렴은 하지만

각 학습 단계에서의 가중치 변화량이 미세하여 전체 학습시간이 길어지는 단점을 들 수 있다. 실험 결과 학습률은 0.05에서 가장 적절하였다.

신경망에서의 학습은 MSE(Mean Square Error)가 0에 충분히 수렴할 때까지 학습을 시키는 것이 보통이다. 그러나 협력적 추천의 경우 사용자들 간의 항목에 대한 선호도의 평균적인 경향을 학습하는 것이 중요하며 소수의 특정 데이터에 대하여 정확히 학습시킬 필요는 없다. 왜냐하면 어떤 특정 선호도 정보가 절대적인 것이 아니기 때문이다. 10개의 사용자 모델에 대해 MSE가 0에 수렴할 때까지 학습을 시킬 경우 <그림 8>과 같이 과다 학습 문제가 발생한다. 여기서 학습회수는 100회 근방에서 테스트 데이터에 대한 accuracy가 가장 좋았고 그때의 평균 MSE는 0.04이다. 과다 학습은 신경망 출력 값이 학습 데이터에 너무 근접하게 학습됨으로써 잡음을 포함하여 모델이 형성되는 경우를 말한다. 과다 학습은 일반적으로 학습회수를 줄이거나 은닉층 노드 개수를 적게 하여 모델이 일반적 경향을 나타내도록 생성하여 해결할 수 있다. 따라서 협력적 추천기술은 특성상 적은 은닉노드의 개수와 학습 반복횟수를 사용하여 일반적인 신경망이 갖는 학습속도의 문제점을 보완할 수 있다. 또 추가 데이터에 대하여 추가 학습(incremental learning)할 경우 그전의 학습된 신경망 모델의 가중치를 초기 가중치로 하여 학습시킬 수 있다.

#### 선호도 정량화에 따른 결과

신경망의 장점 중 하나인 수치데이터 처리가 용이하다는 점을 이용하여 실제 선호도를 Billsus and Pazzani<sup>6</sup>와는 달리 <표 3>와 같이 정량화(quantization)하여 10개의 사용자 모델에 대해 실험을 한 결과 정량

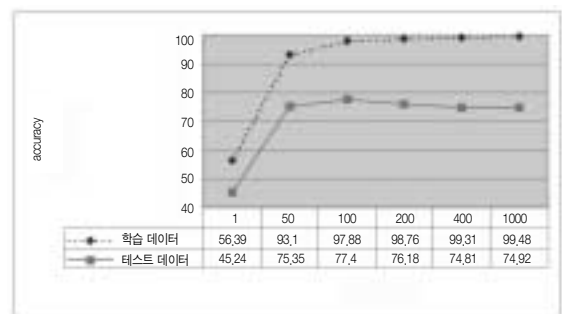


그림 8. 과다 학습 현상

표 3. 선호도 정량화

	선호도가 주어진 경우						선호도가 없는 경우	평균accuracy(%)
원래 선호도	0	0.2	0.4	0.6	0.8	1	표기 없음	
정량화1	-1	-0.6	-0.2	0.2	0.6	1	0	66.6
정량화2	-1			1			0	62.3
정량화3	-1		0		1		0	77.0

화3의 경우 accuracy가 가장 높았다. 즉, 평가에 도움이 안 되는 중간 선호도를 “0”으로 처리함으로써 accuracy를 높일 수 있었다.

#### 입력노드 수에 따른 결과

앞으로의 실험은 종료조건을 MSE는 0.04로 하고, 초기 가중치는 0.5 ~ -0.5, 학습률은 0.05, 선호도 정량화는 <표 3>의 정량화3으로 하여 실험한다. 10개의 항목 모델에 대해 입력 노드 개수를 점차 늘려 가면서 성능의 변화를 분석한 결과 <그림 9>에서와 같이 accuracy가 증가함을 알 수 있다. 그러나, 이것은 입력 데이터 수가 증가하면 전반적인 accuracy는 증가하나 입력 데이터 수에 비례하는 것은 아니다.

#### 사용자와 항목 모델 결과

<표 4>는 10개의 사용자 모델과 10개의 항목 모델을 가지고 사용자 모델, 항목 모델, 항목 내용 정보를 고려한 사용자 모델을 비교한 결과를 보여준다. 표에서 본 바와 같이 항목 모델보다는 사용자 모델이 더 좋다. 즉, 항목간의 연관성을 구하여 추천을 해주는 것보다는

사용자간의 연관성을 고려하여 항목을 추천해 주는 것이 더 우수하다는 것을 알 수 있다. 또 사용자 모델과 항목 내용 정보(여기서는 영화의 장르)를 고려한 모델에서는 항목 내용을 고려한 사용자 모델이 더 성능이 좋은 것을 알 수 있다.

또 6개의 항목 모델에 대해 성별, 나이를 고려한 항목 모델과 고려하지 않은 항목 모델의 accuracy를 비교한 결과 각각의 accuracy는 72%로 비슷함을 보였다. 이 경우 사용자 80% 이상이 성별이 남성이고 또 나이는 20대였기 때문에 이들 정보를 추가로 고려한 것이 특별한 성능의 향상에 기여하지 못한 것으로 생각된다.

#### 기존의 추천 방법과의 비교

Lin, Alvarez and Ruiz<sup>5</sup>는 연관규칙 방법을 이용하여 Billsus and Pazzani<sup>6</sup>와 같은 조건하에서 실험을 하였다. 그 결과 accuracy는 67%~ 69%로 서로 비슷함을 보였다. 따라서, 우리는 연관규칙 추천에서의 결과와 비교를 위해 같은 조건인 최소 100편 이상 영화에 대한 선호도를 입력한 사용자 1,000명을 학습 데이터로 사용하고, 70,000 이상인 사용자 ID에서 최소 100편 이상 영화에 대한 선호도를 입력한 사용자 100명을 무작위로 선택하여 테스트 데이터로 사용하여 실험하였다.

30개의 사용자 모델과 30개의 항목 모델에 대해 우리가 제안한 추천 방법과 기존의 최근접 이웃<sup>15</sup>, 연관규

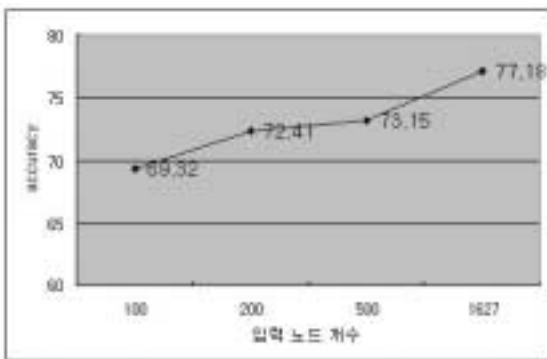


그림 9. 입력노드 수에 따른 accuracy 변화

표 4. 사용자와 항목 모델 실험 결과

	사용자 모델	항목 모델	장르를 고려한 사용자 모델
accuracy	77.0	71.7	79.7
precision	82.5	70.8	83.9
recall	83.2	73.3	86.3
F-measure	82.8	72.0	85.1

칙 추천 방법<sup>5,6</sup>을 비교한 결과 <표 5>와 같이 본 논문에서 제안한 신경망 추천이 기존의 추천들보다 더 우수함을 알 수 있다. 신경망 추천에서는 like와 dislike를 동시에 고려하고, 사용자나 항목들 간의 가중치의 차이를 학습할 수 있기 때문이다. 연관규칙 추천에서 recall이 현저히 낮은 이유는 사용자가 선호하는 항목의 개수는 많은데 비해 규칙에 의해 추천해 주는 항목의 개수가 적기 때문이다. 그러나, 신경망 추천에서는 precision과 recall이 동시에 높음을 알 수 있다. <표 4>의 accuracy 결과가 <표 5>의 accuracy 결과보다 낮게 나온 이유는 전자에서는 like 빈도(ratio)가 0.6 ~ 0.4인 데이터만을 가지고 실험을 하였기 때문이다. 하지만, 다른 기법의 accuracy보다 높다는 것을 알 수 있다. 즉, <표 4>에서는 선호가 분명하지 않는 사용자와 항목을 모델로 생성하여 실험한 결과를 보여주고 있으며, 이는 신경망이 다른 방법들보다 연관성을 학습하는데 효과적이라는 것을 나타내고 있다. 또한, 사용자 모델과 장르를 고려한 사용자 모델의 결과는 비슷한데 이는 장르 정보가 <표 4>의 결과와는 다르게 크게 영향을 미치지 못하고 있음을 알 수 있다.

## 결론 및 향후 연구

본 논문에서는 추천의 성능을 향상시키기 위해 신경망 추천에서의 학습 파라미터가 미치는 영향에 대해서 분석하고, 내용기반 추천과 협력적 추천을 병합한 신경망 추천 방법을 제안하였다. 실험 결과 제시한 신경망 추천 방법이 기존의 추천 방법들보다 우수한 성능을 보였다. 최근접 이웃 방법은 적용이 용이한 반면 항목간의 가중치를 고려하지 못하고, 추천의 정확도가 낮으며, 연관규칙 방법은 항목간의 가중치를 지지도나 신뢰도로 나타내어 최근접 이웃 방법보다는 정확도가 높으

나 이 또한 단순히 항목간의 표면적인 연관관계에 의해 선호도를 결정하고, 빈도수가 적은 선호도 항목의 경우 규칙생성이 어렵고, 연속적인 값의 처리에 어려움이 있다. 그러나 본 논문에서 제안한 신경망 추천 방법은 은닉노드에서 항목간의 가중치를 고려하며, 자료 유형에 상관없이 데이터 처리가 용이할 뿐만 아니라 항목에 대한 내용정보나 사용자의 인구통계적 정보를 고려하여 신경망을 학습시킴으로써 다른 추천 방법들 보다 좋은 성능을 보여주었다.

현재 추천에서 대용량 데이터에서의 범위성이 큰 문제가 되고 있다. 본 논문에서도 단순히 입력 노드개수를 무작위로 선택하여 실험을 하였으나 향후 연구로는 이러한 범위성 문제를 해결하기 위해 클러스터링 방법을 적용하거나 또는 최근접 이웃 방법을 취향이 유사한 사용자를 선택할 경우 추천 성능을 더욱 향상시킬 수 있을 것으로 생각된다. 또 다양한 이질적 정보를 고려하고 또 신경망 방법과 규칙을 사용한 방법의 통합에 대하여 연구하고자 한다.

## 감사의 글

본 연구는 과학기술부의 “뇌신경정보학 연구 사업”과 과학재단의 “특정기초 연구 사업”의 지원에 의하여 수행되었다.

## 참고문헌

1. Pazzani, M. J. A Framework for Collaborative, Content-Based and Demographic Filtering. *Artificial Intelligence Review* 13(5-6), pages 393-408. (1999)
2. Sarwar, B. M., Karypis, G., Konstan, J. A., and Riedl, J. Item-based Collaborative Filtering Recommender Algorithms. Accepted for publication at the WWW10

표 5. 기존의 추천 방법들과 신경망 추천과의 비교

	최근접이웃 추천	연관규칙 추천		신경망 추천		
		사용자모델	항목모델	사용자모델	항목모델	장르를 고려한 사용자 모델
accuracy	67.8	72.0	61.1	81.6	77.5	81.4
precision	60.3	75.1	75.4	77.4	76.3	78.0
recall	55.7	58.4	22.6	69.6	73.0	65.7
F-measure	57.9	65.7	34.8	73.3	74.6	71.3

- Conference . (2001)
3. Breese, J. S., Heckerman, D., and Kadie, C. Empirical Analysis of Predictive Algorithms for Collaborative Filtering. In Proceedings of the Fourteenth Annual Conference on Uncertainty in artificial Intelligence, pages 43-52. (1998)
  4. Lin, W., Ruiz, C., and Alvarez, S. A. Collaborative Recommendation via Adaptive Association Rule Mining. International Workshop on Web Mining for E-Commerce(WEBKDD2000), held in conjunction with the Sixth International Conference on Knowledge Discovery and Dat Mining(KDD2000). (2000)
  5. Lin, W., Ruiz, C., and Alvarez, S. A. A new adaptive-support algorithm for association rule Mining. Technical Report WPI-CS-TR-00-13, Department of Computer Science, Worcester Polytechnic Institute. (2000)
  6. Billsus, D., and Pazzani, M. J. Learning collaborative information filters. In Proceedings of the Fifteenth International Conference on Machine Learning, pages 46-53. (1998)
  7. Cohen, W. W. Web\_Collaborative Filtering: Recommending Music by Crawling The Web. The International Journal of Computer & Telecommunications Networking, V.33 N.1-6, pages 685-698. (2000)
  8. Fu, X., Budzik, J., and Hammond, K. J. Mining navigation history for Recommendation. In Proceedings of the 2000 International conference on intelligent user interfaces, pages 106-112. (2000)
  9. Sarwar, B. M., Karypis, G., Konstan, J. A., and Riedl, J. Analysis of Recommendation Algorithms for E-Commerce. In Proceedings of the ACM EC' 00 Conference, pages 158-167. (2000)
  10. Sarwar, B. M., Konstan, J. A., Rorchers, A., Herlocker, J., Miller, B., and Riedl, J. Using Filtering Agents to Improve Prediction Quality in the GroupLens Research Collaborative Filtering System. In Proceedings of the ACM Conference on Computer Supported Cooperative Work(CSCW). (1998)
  11. Balabanovic, M., and Shoham, Y. Content-based, collaborative recommendation. Communications of the ACM, 40(3), (1997)
  12. Basu, C., Hirsh, H., and Cohen, W. Recommendation as Classification: Using Social and Content-Based Information in Recommendation. In Proceedings of the American Association for Artificial Intelligence, pages 714-720. (1998)
  13. Ahmad, M., and Wasfi, A. Collecting User Access Patterns for Building User Profiles and Collaborative Filtering. In Proceedings of the 1999 International Conference on Intelligent User Interfaces, pages 57-64. (1999)
  14. P.McJones. Eachmovie collaborative filtering data set, <http://www.research.digital.com/SRC/eachmovie>. DEC Systems Research Center. (1997)
  15. Resnick, P., Iacovou, N., Sushak, M., Bergstrom, P., and Riedl, J. Grouplens: An Open Architecture for Collaborative Filtering of Netnews. In Proceedings of Computer Supported Cooperative Work Conference(CSCW), pages 175-186. (1994)