

# 협동적 필터링과 SOM 신경망을 결합한 추천시스템 모델

이미희<sup>†</sup>, 우용태<sup>\*\*</sup>

## 요 약

추천시스템은 사용자가 제공한 선호, 관심, 구매경험과 같은 정보를 근거로 하여 다른 사용자에게 가장 알맞은 정보를 제공하는 일련의 가치교환 과정인 개인화를 가능하게 하는 시스템으로 고객의 선호도를 정확히 분석하고, 정제하여 정확한 예측력으로 고객이 원하는 가장 적절한 상품을 추천 해줄 수 있어야 한다. 대부분의 추천시스템들이 협동적 필터링 기법을 적용하고 있어 본 논문에서는 협동적 필터링 기법의 연산 수행량을 개선한 새로운 결합 모델인 SOM(Self-Organizing Map) 신경망 회로와 결합한 추천시스템을 제안하였다. 먼저, 사용자 그룹을 인구통계학적인 특징으로 세그먼트하고 SOM 신경망회로를 이용하여 item 특징에 대한 선호도를 입력 값으로 학습하여 클러스터를 생성하였다. 임의의 사용자에게 대한 추천은 선호도가 유사한 클러스터를 결정하여 협동적 필터링 기법을 적용하였으며, 기존의 협동적 필터링 기법의 연산 수행량과 비교 분석하였다. 또한 영화를 대상으로 한 실험을 통하여 추천효율이 향상되었음을 나타내었다.

## A Recommender System Model Combining Collaborative Filtering and SOM Neural Networks

Mi-Hee Lee<sup>†</sup>, Young-Tae Woo<sup>\*\*</sup>

## ABSTRACT

A recommender system supports people in making recommendations finding a set of people who are likely to provide good recommendations for a given person, or deriving recommendations from implicit behavior such as browsing activity, buying patterns, and time on task. We proposed new recommender system which combined SOM(Self-Organizing Map) neural networks with the Collaborative filtering which most recommender systems hat applied First, we segmented user groups according to demographic characteristics and then we trained the SOM with people's preferences as ito inputs. Finally we applied the classic collaborative filtering to the clustering with similarity in which a recommendation seeker belonged to, and therefore we didn't have to apply the collaborative filtering to the whose data set. Experiments were run for EachMovies data set. The results indicated that the predictive accuracy was increased in terms of MAE(Mean-Absolute-Error).

**Key words:** SOM(Self-Organizing Map) neural networks(SOM 신경망), Collaborative filtering(협동적 필터링), recommender systems(추천시스템)

## 1. 서 론

일대일마케팅 기법은 소비자들의 특성 분석에 따

른 개인별 차별화를 통하여 고객들의 잠재가치를 높이고 궁극적으로 기업의 이윤을 극대화하기 위한 새로운 마케팅 전략이다[1,2]. 이 전략은 사용자별로 차

※ 교신저자(Corresponding Author) : 이미희, 주소 : 마산시 합성2동 1번지(630-764), 전화 : 055)250-3126, FAX : 055)297-5181, E-mail : mihil@csc.ac.kr

접수일 : 2008년 2월 4일, 완료일 : 2008년 7월 16일

<sup>†</sup> 마산 창원대학 인터넷컴퓨터과 학과장

\*\* 창원대학교 컴퓨터공학과 교수

(E-mail : ytwoo@sarim.changwon.ac.kr)

※ 이 논문은 2006년도 창원대학교 연구비에 의하여 연구되었음

별화 된 맞춤 서비스를 제공하기 위한 핵심적인 기법 중의 하나인 개인화[3] 전략이다. 개인화 전략의 핵심 기술은 고객들의 취향과 구매 이력을 분석하여 개인별로 차별화 된 정보를 자동적으로 필터링하여 추천하기 위한 시스템이다[2,4,5]. 추천시스템은 고객의 선호도를 정확하게 분석, 정제하여 정확한 예측력으로 고객이 원하는 가장 적절한 상품을 추천할 수 있는 기능이 필요하다. 추천시스템은 크게 인구통계학적 정보를 이용한 기법[3,6,7], 내용기반 추천 기법[8,9], 항목기반 추천 기법[10,11], 그리고 협동적 필터링(collaborative filtering) 기법[12-15] 등이 연구되고 있다.

최근에 가장 널리 사용되는 협동적 필터링 기법은 임의의 사용자와 유사한 선호도를 가지는 다른 사용자의 평가정보를 이용하여 상품을 추천하는 방식이다[5]. 그러나 현재까지 국내외에서 개발된 추천시스템들의 대부분은 상관계수를 기반으로 한 협동적 필터링 기법에 의한 추천시스템들이 대부분인 관계로 상관계수에 의한 문제가 발생한다. 협동적 필터링 기법은 사용자의 평가가 적은 경우에는 사용자간의 유사도를 구하기 어려운 관계로, 예측의 정확도가 떨어지는 문제가 발생한다. 또한, 사용자의 평가만을 근거로 다른 사용자와의 유사도를 측정함으로써 동일한 상품을 평가하지 않은 다른 사용자와의 관계를 완전히 배제하게 되어 사용자와 매우 유사한 관계에 있을 수도 있는 다른 사용자와의 유사성을 반영하기 어렵다. 뿐만 아니라 사용자의 선호도를 직접적으로 반영하지 못하므로 실제 사용자와 유사성이 거의 없고 선호도가 다른 사용자가 이미 평가한 작은 수의 상품에 의해 유사성이 있는 것으로 잘못된 추천을 하게 될 가능성이 있다[5,13].

본 논문에서는 기존의 협동적 필터링 기법과 SOM(Self-Organizing Map) 신경회로망을 결합한 새로운 추천시스템 모델을 제안하였다[16-20]. SOM 신경회로망은 다차원의 데이터를 2차원으로 사상시키는 신경회로망의 하나로, 복잡한 다차원 데이터를 클러스터링 하기 위한 기법으로 널리 사용되고 있다. 또한 SOM 신경회로망은 클러스터링 결과의 구조를 미리 지정해 줄 수 있는 장점과 입력 데이터의 수가 많아도 빠른 시간 내에 결과를 얻을 수 있는 특징을 가진다. 하지만 SOM 신경회로망과 이를 결합한 연구는 아직까지 진행되지 않았다.

먼저, 본 모델에서는 사용자 그룹을 인구통계학적인 특징으로 세그먼트 하였다. 인구통계학적인 추천 기법은 사용자의 성별, 나이, 직업 등과 같은 사용자 유형별 특징을 분석하여 상품을 추천하는 방법이다[2]. 이 기법은 전통적인 추천시스템의 하나로, 단순한 형태의 정보필터링 기법으로 시스템 초기 구축 단계에서 많이 사용되는 추천시스템이다. 이 논문에서는 사용자를 성별과 나이로 세그먼트 한 후, 영화 데이터의 평가 값에서 장르별 평가 평균값을 입력값으로 하여 SOM 신경회로망을 이용하여 클러스터를 생성하였다. 임의의 사용자에 대한 추천은 선호도가 유사한 클러스터를 결정하여 협동적 필터링 기법을 적용하였으며, 기존의 협동적 필터링 기법의 연산 수행 량과 비교 분석하였다.

## 2. 관련연구

### 2.1 추천시스템

추천시스템은 사용자에게 대한 인구통계학적 정보, 가장 많이 팔린 상품, 사용자의 구매 패턴 등을 분석하여 개인의 선호도를 데이터 마이닝 기법에 의해 분석하여 사용자가 구매하고 싶은 상품을 쉽게 찾을 수 있도록 도와주는 시스템이다.

이 추천시스템은 다음과 같은 세 가지 측면에서 전자상거래의 판매를 촉진시키는 효과를 기대할 수 있다. 첫째, 추천시스템은 전자상거래 사이트에서 고객이 원하는 상품을 쉽게 찾을 수 있도록 도와줌으로써 상품 구매를 유도할 수 있다. 둘째, 추천시스템은 장바구니에 들어 있는 상품에 기반하여 사용자가 구매한 상품 외에 추가적인 상품을 추천하기 위한 교차 판매(cross-selling) 전략을 수립할 수 있다. 셋째, 추천시스템은 전자상거래 사이트와 사용자간의 가치 있는 관계를 지속적으로 유지하여 사용자의 로열티를 향상시킬 수 있다. 이러한 추천시스템은 전자상거래 분야에서 개인화 된 상품이나 콘텐츠를 추천하기 위한 일대일 마케팅 전략을 위한 핵심적인 기법이다[21].

### 2.2 추천시스템에 대한 기존 연구 방법

인구통계학적 추천(demographic-based recommendation) 기법은 사용자의 성별, 나이, 직업 등과 같은 인구통계학적 요소에 의해 사용자 유형별 특징

을 분석하여 상품을 추천하는 방법이다[3]. Kruwlich는 사용자 세그먼트를 위하여 성별, 생활방식, 거주지역 등과 같은 인구통계학적 인자에 대하여 학습 알고리즘을 적용하여 62개의 사용자 세그먼트를 구성하였다[6]. 또한 Winnow 알고리즘은 사용자 세그먼트를 위한 학습 방법으로 인자와 가중치 곱의 합으로 사용자간의 유사성을 판정할 수 있다[7].

내용기반 추천(content-based recommendation) 기법은 개인의 요구나 개인으로부터 입력된 모든 정보와 상품에 포함된 텍스트 정보를 이용하여 필터링하는 방식이다. 이 기법은 논리 연산자와 결합된 문자열 등을 검색에 포함시키거나 제거해야할 복잡한 문자열을 가진 텍스트 프로파일을 사용한다. 또한, 사용자 프로파일을 통해 과거 구매나 추천결과를 쉽게 반영할 수 있는 장점이 있으며 추천 속도도 빠르다. 그러나 이 기법은 상품에 대한 textual description 정보의 정확도를 판단하기 어렵고 상품과 사용자가 많은 경우에 효율성이 떨어지는 문제가 있다[12].

항목기반 추천(item-based recommendation) 기법은 상품간의 유사성을 이용하여 상품을 추천하는 방식이다. 이 기법은 Karypis, Badrule 등이 제안한 방식으로 상품간의 관계를 기반으로 하나의 상품에 대한 구매결과로부터 다른 상품에 대한 구매를 유도하기 위한 방법이다.

협동적 필터링 추천(collaborative filtering recommendation) 기법은 사용자들의 기초정보와 고객들의 선호도를 바탕으로 선호도에서 비슷한 패턴을 보이는 고객들을 분류하는 기법이다[22]. 즉 이 기법은 비슷한 취향을 가진 고객들에게 아직 구매하지 않은 상품들을 교차 추천하거나 분류된 고객의 프로파일에 따라 상품을 추천하는 형태이다. 협동적 필터링 기법은 정확한 개인별 추천과 예측하지 못한 상품을 추천할 수 있다는 장점을 가지고 있다. 또한, 이 기법은 추천시스템에 적용할 때 사용자의 의견을 반영할 수 있다는 장점으로 인해 최근에 많은 연구가 진행되고 있다[23].

2.3 협동적 필터링 기법

2.3.1 협동적 필터링 기법

협동적 필터링 기법을 이용한 추천시스템은 유사한 선호도를 가진 다른 사용자의 의견에 기반하여

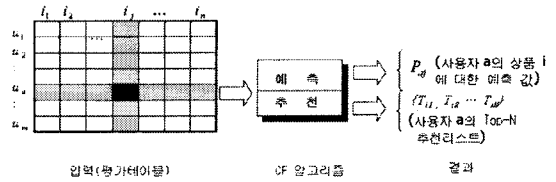


그림 1. 협동적 필터링 기법의 시스템 구조

상품에 대한 개인의 선호도를 예측하는 기법이다. 그림 1은 전형적인 협동적 필터링 기법의 프로세스 개념도이다[11].

평가테이블은 사용자 리스트  $U = \{u_1, u_2, \dots, u_m\}$ 와 상품 리스트  $I = \{i_1, i_2, \dots, i_n\}$ 로 구성되며, 각 사용자  $u_i$ 는 상품들의 평가 값에 대한 리스트  $I_{u_i}$ 를 가진다. 상품에 대한 평가(rating)는 점수로 표현되거나 구매 기록이나 접속 로그 분석, 웹 마이닝 등으로부터 유추될 수 있다.  $u_a$ 는 협동적 필터링 기법을 수행할 사용자(active user)이다. 또한 예측(prediction)  $P_{a,j}$ 는 상품  $i_j$ 에 대한 사용자  $u_a$ 의 선호도 예측 값이며,  $i_j \in I_{u_a}$ 이다. 추천 리스트  $\{T_{i1}, T_{i2}, \dots, T_{iN}\}$ 은 사용자  $u_a$ 가 구매하지 않은 상품 중에서 사용자에게 추천할 상품리스트  $I_r \subset I$ 이다. 즉,  $I_r \cap I_{u_a} = \emptyset$ 이다. 협동적 필터링 기법에서 사용자와 상품 데이터 집합  $m \times n$ 은 상품에 대한 평가 행렬 A로 표현할 수 있다. 각 항목  $a_{ij}$ 는 사용자  $i$ 의 상품  $j$ 에 대한 선호도 점수이다. 평가 값은 숫자로 표현되며 평가되지 않은 상품에 대한 평가 점수는 0으로 표현된다[11].

다음 표 1은 6개의 영화에 대한 4명의 사용자 평가 값의 예이다.

표 1에서 빈칸은 사용자가 보지 않았거나 평가하지 않은 영화를 의미한다. 예를 들어 철수의 슈렉에 대한 평가 값을 예측하기 위해 피어슨 상관계수를 사용한다. 철수와 영희의 상관관계를 계산하는 식은 다음 식(1)과 같다. 식(3)에서  $\bar{r}$ 와  $\bar{l}$ 은 사용자 철수,

표 1. 각 사용자의 영화에 대한 평가 값

영화 사용자	타이 타닉	쉬리	매트 릭스	반지의 제왕	해리 포터	슈렉
철수	1	5		2	4	?
영희	4	2		5	1	2
순이	2	4	3			5
길동	2	4		5	1	?

$K$ 와 영화,  $L$ 이 함께 평가한 상품에 대한  $K$ 와  $L$ 의 평균,  $i$ 는 각각의 상품을 의미한다[13].

$$r_{KL} = \frac{Cov(K, L)}{\sigma_K \sigma_L} = \frac{\sum_i (K_i - \bar{K})(L_i - \bar{L})}{\sqrt{\sum_i (K_i - \bar{K})^2} \sqrt{\sum_i (L_i - \bar{L})^2}} \quad (1)$$

- $L$  :  $L$  (영화) 사용자
- $K$  :  $K$  (철수) 사용자
- $\bar{K}$  :  $K$  평균값
- $\bar{L}$  :  $L$  평균값
- $i$  : item 수

철수의 평균이 3, 영화의 평균도 3으로 이 둘의 상관계수 값은 -0.8이다. 순이와 길동의 평균도 3으로 철수와 순이의 상관계수 값은 1, 철수와 길동은 0이 된다. 상관계수의 값이 1이면 완전한 양의 관계이고, -1이면 완전한 음의 관계이다. 상관계수의 값이 0이면 관계가 없는 것을 의미한다. 따라서 이 예에서 철수와 순이는 양의 상관관계가 있고, 철수와 영화는 음의 상관관계가 있다고 볼 수 있다. 상관관계를 포함한 평균값을 사용하여 계산하는 평가 예측 값은 식 (2)와 같이 표현된다.

$$prediction = \bar{u} + \frac{\sum_{i=1}^n (corr_i) \times (rating_i - \bar{i})}{\sum_{i=1}^n (corr_i)} \quad (2)$$

- $\bar{u}$ : 아이টে에 대한 사용자  $u$ 의 평균 평가 값
- $corr_i$ : 사용자  $i$ 와 예측될 아이টে에 가진 사용자 의 피어슨 계수
- $rating_i$ : 예측될 아이টে에 대한 사용자  $i$ 의 평가 값
- $\bar{i}$ : 일반적으로 아이টে에 대한 사용자  $i$ 의 평가 평균값
- $n$ : 사용자  $u$ 와 연관성을 가진 시스템내의 전체 사용자

2.3.2 협동적 필터링 기법의 문제점

상관계수 기반의 협동적 필터링 기법의 문제점은 크게 초기 평가의 문제점, 상관계수, 희소성(sparsity), 동의어(synonym), 모호 집단(gray sheep)에 대한 문제점 등이 있다.

초기 평가의 문제점은 시스템 구축 초기에 발생하

는 문제로 사용자로부터 충분한 평가 정보를 받지 못한 경우 상관계수를 구할 수 없거나 부정확하여 정확한 추천이 불가능한 문제이다. 협동적 필터링 기법에서 정확하고 효과적인 추천이 이루어지기 위해서는 충분한 사용자 평가 정보가 요구된다.

상관계수에 의한 문제는 사용자의 평가가 적은 경우에는 사용자간 유사도를 구하지 못하여 예측이 어렵거나 정확도가 떨어지는 문제이다.

희소성의 문제는 초기 시스템에서 발생하는 문제와 유사하지만 전자상거래 사이트에서 대량의 상품에 대해서 충분한 평가 결과를 구성하기 어려운 관계로 평가데이터가 희소성을 가진다. 이 문제를 해결하기 위하여 dimensionality reduction 기법을 사용하거나, 내용기반 software agent를 사용하여 자동으로 평가를 생성하여 데이터 셋의 밀도(density)를 높이는 방법이 제시되었다[6]. 그리고 묵시적인 기법으로 상품 및 콘텐츠에 대한 평가 정보를 자동으로 수집하는 기법에 대한 연구가 제시되었다[23].

동의어 문제는 상품명이 다른 유사한 성격의 상품에 대해 숨어있는 연관관계를 찾아서 추천에 반영하기 어려운 경우이다.

모호 집단의 문제는 일부 특이한 사용자들에 대한 문제이다. 대부분의 상품에 대해 평균적인 평가를 하여 좋은 것과 싫은 것이 분명하지 않거나 평가 결과가 일정하지 않은 사용자, 그리고 특이한 성향을 가져 다른 사용자와 상관관계를 거의 가지지 않는 사용자에 대해서는 적절한 추천을 하기 어렵다.

3. 협동적 필터링 기법과 결합 모델

협동적 필터링 기법에서 대량의 사용자와 아이টে에 의해 많은 시간이 소요되어 실시간 추천이 어렵다. 이에 따라 최근에는 협동적 필터링 기법과 다른 기법을 결합한 모델이 연구되고 있다.

3.1 기존의 결합모델 소개

3.1.1 P-tango 모델

협동적 필터링 기법은 기사를 평가한 사용자가 적을 경우에 추천이 부정확하다. 내용기반 시스템은 사용자가 제시한 명시적 키워드가 적거나 묵시적 키워드를 포함하는 기사가 너무 작을 때, 부정확한 예측을 초래한다. P-tango의 모델은 협동적 필터링 기법

과 내용기반 시스템에 각각 동등한 가중치를 모든 사용자에게 부여하였다[8]. P-tango의 모델은 사용자 평가를 할 때, 협동적 필터링 기법과 내용기반 시스템의 절대적 에러를 계산하였다. 또한 P-tango 모델에서는 과거 에러를 최소화하기 위하여 가중치를 조정하였다. 가중치는 처음에는 빠르게 조정되고, 평가 값과 예측의 수가 증가 할 때는 느리게 조정된다. P-tango 모델은 협동적 필터링 기법과 내용기반 시스템을 결합하여 모든 아이템과 사용자들에게 초기 예측을 가능하게 하였다. 하지만 P-tango 모델은 협동적 필터링 기법에 의존하는 점이 크므로 협동적 필터링 기법의 한계성을 완전히 극복하기 어려운 문제점이 있다.

### 3.1.2 협동적 필터링 기법과 개인정보 필터링 에이전트

협동적 필터링 기법의 단점은 추천 시 아이템의 내용을 고려하지 않는다는 점이다. Good등은 이러한 단점을 보완하기 위하여 협동적 필터링 기법이 개인정보 필터링 에이전트를 사용자들의 집단의견을 결합하는 새로운 모델을 제안하였다[15]. 이 기법은 정보 필터링 에이전트와 협동적 필터링 기법을 결합하여 사용자가 어느 항목이 가치가 있다는 것을 잘 정의함으로써 정보과다 현상을 완화시킬 수 있다. 하지만, 정보 필터링은 내용정보의 근원을 요구하고 예상치 못한 문제를 발견할 때 효율적인 추천이 어려운 문제점이 있다. Good가 제안한 결합모델은 다음과 같이 4가지 유형인 공동 회원의 의견을 사용하는 협동적 필터링 기법 유형과 단일 개인화된 에이전트형, 많은 에이전트들의 조합형 그리고 다양한 에이전트들과 공동회원 의견과의 조합형이 있다. 그러나 의견이 일치하는 사용자들이 서로 다르게 훈련한 에이전트의 의견에서도 이익을 얻을 것인지에 대한 의문이 제기되고 있다.

### 3.1.3 협동적 필터링 기법과 클러스터링 알고리즘

O'Connor는 사용자가 등급을 매긴 데이터에 기초하여 아이템의 집합을 분할하고 각각의 분할된 집합 영역 내에서 독립적으로 예측하기 위해, 클러스터링 알고리즘을 협동적 필터링 기법과 결합하는 모델을 제안하였다[16]. 이러한 분할을 통하여 협동적 필터링 기법의 질을 향상시키고 협동적 필터링 기법의 확장성을 증가시켰다. 그러나 분할 기법은 확장성은

좋아졌으나, 정확성은 나쁜 결과를 보였다. 또한 평가 데이터에 근거한 분할은 임의분할보다 더 정확한 예측 결과를 얻었다.

클러스터링과 분할 기법은 클러스터링 과정으로 유도하기 위하여 거리 미터법과 유사도 미터법을 사용한다. 아이템 사이의 유사도 계산은 미터법을 제공하기 위하여 필요하다. 각 아이템들 사이의 유사도 측정은 상관계수법을 이용한다. 사용자가 두 개의 영화에 대해 동일한 감정을 느꼈다면, 두 영화는 높은 상관성이 있다. 그러나 분할 기법은 음의 상관관계에 대해서는 고려하기 어려운 문제점이 있다[16].

### 3.1.4 기존 결합모델의 문제점

협동적 필터링 기법의 초기 평가의 문제점을 해결하기 위하여 P-tango는 온라인 신문 기사 예측을 내용기반 예측과 집합적 예측의 평균에 기초하여 협동적 필터링 기법을 결합하였다. 이 기법은 사용자와 평가 값이 증가하면 예측의 정확성을 얻을 수 있어 협동적 필터링 기법의 결합을 보완할 수 있었다. 하지만 이 기법은 협동적 필터링 기법에 의존하는 관계로 협동적 필터링 기법의 한계를 안고 있다.

협동적 필터링 기법의 단점중의 하나는 추천 시 아이템의 내용을 고려하지 않고, 새로운 아이템은 어떤 사용자가 아이템을 평가할 때까지 추천할 수 없다는 것이다. Good는 이러한 단점을 보완하기 위하여 협동적 필터링 기법이 개인 정보 필터링 에이전트를 사용자들의 집단의견을 결합하는 새로운 모델을 제안하였다. 하지만 서로 의견이 일치하는 사용자끼리 각각 다르게 훈련시키게 되었던 에이전트의 의견에서 이익을 얻을 것인지에 대한 의문이 남아있다. 또한 O'Connor가 제안한 분할 기법은 확장성은 좋아졌으나, 정확성은 나쁜 결과를 낳았다.

## 4. SOM 신경회로망과 협동적 필터링 기법을 이용한 결합 추천 모델

### 4.1 연구 배경 및 동기

전자상거래 사이트들은 개인별로 차별화 된 일대일 마케팅 전략을 적용하기 위해 eCRM 시스템을 활발하게 도입하고 있다. 추천시스템은 고객들의 상품에 대한 선호도와 구매이력을 분석하여 개인별로 차별화된 정보를 추천하기 위한 자동화된 정보 필터링

시스템이다. 이러한 추천시스템 중에서 협동적 필터링 기법은 비교적으로 정확한 개인별 추천과 예측하지 못한 상품을 추천할 수 있다는 장점으로 인해 가장 널리 사용되고 있다. 하지만 이 기법은 사용자와 유사한 선호도를 가지는 다른 사용자의 평가정보를 이용하여 상품을 추천하는 관계로 초기 평가의 문제, 희소성 문제, 동의어 문제, 모호 집단 문제가 발생한다[12-14]. 또한 상관계수에 기반 한 협동적 필터링 기법은 모든 사용자와 예측 대상자와의 유사도 측정을 하는 관계로 사용자 수가 많아질수록 연산 수행량이 증가하는 문제가 있다[24].

본 논문에서는 시스템 초기 구축 단계나 처음 방문한 사용자에게 대해서도 효과적으로 추천을 하기 위해 SOM 신경회로망과 협동적 필터링 기법을 결합한 추천 모델을 제안하였다. 제안 기법을 통하여 협동적 필터링 기법의 초기 평가 문제점과 사용자가 증가할수록 처리 시간이 증가하는 문제점을 개선하여 상관계수와 희소성의 문제를 해결하였다. 또한 본 논문에서는 SOM 신경회로망의 클러스터링 기법을 이용하여 대용량 데이터에 대한 처리 속도를 향상시켰다.

먼저, 본 연구에서는 전체 사용자를 성별과 나이와 같은 인구통계학적인 특성에 의하여 세그먼트 하였다. 각 세그먼트는 SOM 신경회로망을 이용하여 아이템 특성의 선호도에 따라 학습하여 클러스터를 생성하였다. 그리고 사용자가 평가하지 않은 아이템에 대한 평가 결과 예측은 임의의 사용자에게 선호도가 유사한 클러스터에 속한 사용자 그룹의 평가값을 토대로 협동적 필터링 기법을 적용하였다. 본 연구에서 제안한 시스템은 기존의 협동적 필터링 기법과 SOM 신경회로망을 결합한 모델로서, 연산 수행량에 대한 분석을 통하여 제안 모델의 연산 수행량의 개선과 추천 효율이 향상 될 수 있음을 보였다.

그림 2은 본 연구에서 제안한 결합 모델의 개념도이다.

본 연구에서 제안한 모델은 다음과 같은 단계로 사용자가 평가하지 않은 아이템에 대한 평가결과를 예측한다. 첫째, 전체 사용자를 성별과 나이로 세그먼트 하였다. 둘째, 각 세그먼트에 대해 사용자의 아이템 특성 평균값을 입력 값으로 구성하였다. 셋째, SOM 신경회로망을 이용하여 선호도 패턴을 학습하여 클러스터를 생성하였다. 넷째, 임의의 사용자에게 선호도가 유사한 클러스터에 속한 사용자 그룹

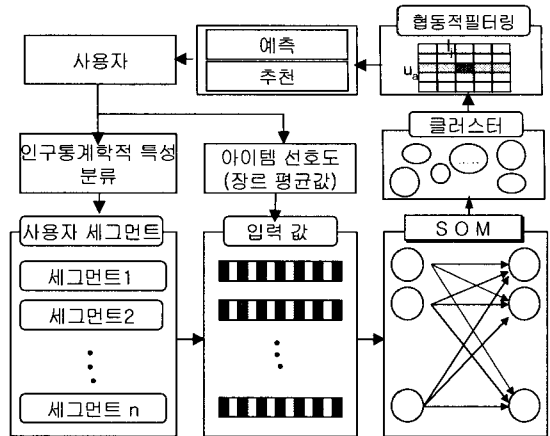


그림 2. 협동적 필터링 기법에 SOM 신경회로망을 결합한 추천시스템 개념도

의 평가 값을 토대로 협동적 필터링 기법을 적용하였다. 다섯째, 사용자가 평가하지 않은 아이템에 대한 평가 결과를 예측하였다.

#### 4.2 인구통계학적인 특성을 이용한 추천시스템

Krawlich 등은 사용자 유형별 세그먼트를 위하여 성별, 생활방식, 직업, 결혼여부, 거주 지역 등과 같은 인구통계학적인 인자에 대하여 클러스터링 기법을 연구하였다. Winnow 등은 62개의 사용자 세그먼트를 구성하는 연구를 하였다. 이러한 인구통계학적인 특성은 사용자의 피드백 정보가 없이도 추천이 가능하며, 시스템 초기 구축 단계에 적용할 수 있다. 이 기법은 시스템 구축 초기나 초기 고객에 대해서 적용할 수 있지만 다른 추천 기법에 비해 정확도가 떨어지므로 eCRM 시스템에서 개인화 전략에 적용하기 어렵다. 따라서 인구 통계학적인 정보에 의한 추천 기법은 다른 추천기법과 결합하여 추천 효율을 높이는 데 주로 사용되고 있다.

먼저 제안 시스템에서는 표본 집단에 대한 성별과 나이 인자에 동일한 가중치를 부여하여 두 인자가 일치하는 사용자끼리 세그먼트 하였다. 협동적 필터링 기법의 예측 값은 피어슨의 상관 계수식을 이용하여 계산하였다.

Hwang등은, 협동적 필터링 기법과 인구통계학 특성을 각각 다르게 적용시켰을 경우에 추천 효율이 다르게 나타날 수 있는 결과를 제시하였다. 이 연구에서 협동적 필터링 기법은 0.93, 협동적 필터링 기법

과 인구통계학 특성인 성별에 의한 오차는 1.00, 협동적 필터링 기법과 인구통계학 특성인 나이에 의한 오차는 0.74, 협동적 필터링 기법과 인구통계학 특성인 성별과 나이에 의한 오차는 0.59로 나타났다. 여기서 오차 값의 의미는 추천 기법별로 표본 집단에 대한 실제 평가 값과 예측 값과의 차이이다. 이 차이 값은 실제 평가 값과의 차이를 의미하는 오차이므로 값이 낮을수록 추천 효율이 우수함을 의미한다.

### 4.3 SOM 신경회로망을 이용한 클러스터링

먼저, 제안모델에서는 전체 사용자를 성별과 나이와 같은 인구통계학적인 특성에 의하여 세그먼트 하였다. 그림 3에서처럼 각 세그먼트에 대해 SOM 신경회로망을 이용하여 클러스터링 하였다. SOM 신경회로망의 입력 값은 영화 장르별 선호도 평가 평균값을 사용하였다. 영화 장르별 평균값은 사용자들의 영화에 대한 선호도를 0.0~1.0점으로 평가한 후 평균을 구하였다. 각 세그먼트별 SOM 신경회로망의 출력은 2개의 그룹으로 클러스터링 하였다.

클러스터링은 비슷한 특성을 갖는 다수의 데이터들을 하나로 묶어서 데이터 축약을 하는 작업이다. 클러스터링은 특별한 속성 정보나 배경 지식 없이 데이터로부터 직접 얻을 수 있는 비교 학습 기법이다. O'Connor에서 사용된 클러스터링 기법은 분할(partitioning) 기법이다. 분할 기법은 사용자가 등급을 매긴 데이터를 아이템 집합으로 분리하고 각각의 분할된 집합으로 클러스터링 하는 기법이다. 이 기법

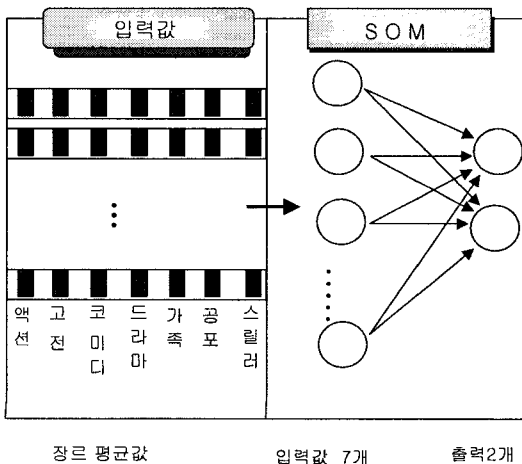


그림 3. SOM 신경회로망을 이용한 클러스터링에 대한 개념도

은 예측 데이터를 저장할 때 연관성 있는 데이터끼리 하나의 물리적인 공간에 저장되어질 수 있도록 클러스터링 한다. 이런 식으로 그룹화된 데이터는 검색 속도를 향상시키고 데이터 관리를 용이하게 한다. O'Connor가 제안한 분할 기법은 사용자와 다른 사용자간의 선호도를 측정하는 과정에서 상관계수의 절대치 값을 구한다. 하지만 이 절대치 값은 음의 상관관계에 대하여는 고려를 하지 못하는 단점이 있다.

본 논문에서 사용한 SOM 신경회로망 기법은 주어진 입력 데이터를 2차원의 노드로 사상시킴으로써 주어진 데이터의 개수를 노드의 개수만큼 축약시키는 역할을 한다. 또한 SOM 신경회로망 기법은 복잡한 다차원 데이터를 효과적으로 클러스터링 하기 위한 방법이다. 또한 클러스터링 결과의 구조를 사전에 지정해 줄 수 있는 특징이 O'Connor가 사용한 분할 기법과 다른 종류의 클러스터링 기법이다[18]. 따라서 본 연구에서는 사용자가 평가하지 않은 아이템을 대상으로 상품을 추천하기 위하여 SOM 신경회로망을 이용하여 클러스터링 그룹 내에서 협동적 필터링 기법을 사용하였다. 클러스터링 된 사용자를 대상으로 협동적 필터링 기법을 적용하는 경우는 전체 사용자를 대상으로 한 협동적 필터링 기법에 비해 연산 수행량이 줄어들 수 있다. 이러한 이유로 신경회로망 기법 중에 SOM을 사용하게 되었다

## 5. 협동적 필터링 기법과 제안 모델과의 성능 비교

### 5.1 기존의 협동적 필터링 기법에 의한 연산 수행량

사용자가  $N$  명이고 각 사용자의 아이템이  $M$  개라고 가정하면, 사용자와 아이템간의 관계는 그림 4와 같다.

	1	2	3	...	$M$ (아이템)
1					
2					
3					
4					
...					
$N$					

그림 4. 사용자와 아이템간의 관계

식 (2)에 의해 협동적 필터링 기법에 의한 연산 수행량은 다음과 같은 단계로 분석할 수 있다.

단계 1. 연산 수행량  $\bar{u}$  계산한다.

예를 들어 그림 9에서 5번 사용자의 3번 아이টে에 대한 예측값을 구한다고 가정하면,  $\bar{u}$ 는 5번 사용자의 3번 아이টে를 제외한 나머지 아이টে를 이용한 평균이다.

$$\text{즉, } \bar{u} = \frac{em_1 + em_2 + \dots + em_M}{M-1} \text{ 이다.}$$

따라서, [+] 연산은 M-1회 연산에서 자기 자신의 아이টে를 제외한 횟수이므로 M-2회가 된다. [/] 연산 1회가 요구된다. 따라서  $\bar{u}$ 의 총 연산 수행량은  $(M-2) + 1 = M-1$  이다.

단계 2. 연산 수행량  $\bar{i}$  를 계산한다.

$$\bar{i} = \frac{em_1 + em_2 + \dots + em_M}{M}$$

·  $\bar{i}$  : 5번 사용자가 아닌 다른 사용자의 평균값  
 [+] 연산은 다른 사용자의 평균값이므로 M-1회, [/] 연산 1회가 요구된다. 따라서  $\bar{i}$ 의 총 연산 수행량은  $(M-1) + 1 = M$  이다.

단계 3. 연산 수행량  $corr_i$  를 계산한다.

예측값을 구하는 해당 사용자와 5번 사용자가 아닌 다른 사용자 사이의 상관계수는 식 (3)과 같이 나타낼 수 있다.

$$corr_i = \frac{\sum_i (K_i - \bar{K})(L_i - \bar{L})}{\sqrt{\sum_i (K_i - \bar{K})^2} \sqrt{\sum_i (L_i - \bar{L})^2}} \quad (3)$$

- L : L 사용자
- K : K 사용자
- $\bar{K}$  : K 평균값
- $\bar{L}$  : L 평균값
- i : item 수

식 (5)에서 K는 5번 사용자, L은 일반 사용자라고 가정한다.

$\sum_i (K_i - \bar{K})(L_i - \bar{L})$ 에는 [-] 연산이  $2 \times M$  회 수행되고, [\*] 연산이 M 회 수행되며, [+] 연산 M 회가 요구된다.  $\sqrt{\sum_i (K_i - \bar{K})^2}$ 에는 [-] 연산이  $2 \times M$  회 수행되고, [\*] 연산이 M 회 수행되며, [+] 연산이 M 회가 수행되며,  $\sqrt{\sum_i (L_i - \bar{L})^2}$ 에는 [-] 연산이  $2 \times M$  회 수행되고, [\*] 연산이 M

회 수행되며, [+] 연산이 M 회 수행되고,  $\sqrt{\quad}$  연산이 1회가 요구된다. 따라서  $corr_i$ 의 총 연산 수행량은  $12M + 2$  회가 요구된다.

단계 4. 연산 수행량  $\sum_{i=1}^n (corr_i) \times (rating_i - \bar{i})$  계산한다.

$corr_i$  계산에  $12M + 2$  회가 요구되고,  $\bar{i}$  계산에 M 회가 요구되기 때문에  $(corr_i) \times (rating_i - \bar{i})$  계산에는  $12M + 2 + M + 2$  회가 요구된다. 따라서 총 연산은  $N \times (12M + 4) + N \times M$  회 수행된다.

단계 5. 연산 수행량  $\sum_{i=1}^n (corr_i)$  을 계산한다.

총 연산 수행량은  $N \times (12M + 2)$  이다.

단계 6. 총 예측값에 대한 연산 수행량을 계산한다.

단계 1에서 단계 5까지의 연산 수행량을 합산하면 다음과 같이 식 (4)로 표현할 수 있다.

$$\text{총연산수행량} = 25NM + 6N + M + 1 \quad (4)$$

- N : 사용자
- M : 아이টে

## 5.2 SOM 신경회로망 결합에 의한 연산 수행량

### 5.2.1 SOM 신경회로망 결합 모델에 의한 연산 수행량

SOM 신경회로망은 그림 5와 같이 나타낼 수 있으며, 다차원의 데이터를 2차원으로 사상시켜주는 신경회로망 기법의 한가지로서, 복잡한 다차원 데이터의 클러스터링에 알맞은 방법이다. 또한, SOM 신경회로망은 클러스터링 결과의 구조를 미리 지정해 줄 수 있는 장점과 입력 데이터의 수가 많아도 빠른 시간 내에 결과를 얻을 수 있고, 데이터의 크기에 영향을 적게 받는 특징이 있다.

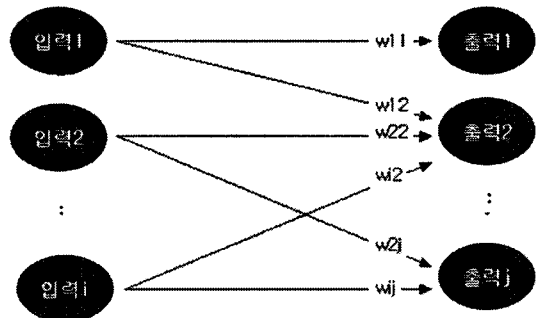


그림 5. SOM 신경회로망의 개념도



SOM 신경회로망에서 필요한 연산 수행량은 다음과 같이 계산된다.

첫 번째 출력노드의 연산 수행량은 다음 식 (5)처럼 계산된다. 이 결과는 각 출력노드의 가중치 값의 절대치를 계산해서 더한 값이다.

$$|i1-w11|+|i2-w21|+\dots+|i-nwn| \quad (5)$$

- $I$  : 입력패턴 수
- $n$  : 출력 node 수
- $w$  : 가중치 값

식 (7)에서 [-] 연산이  $i$ 번 발생하게 되며, [+] 연산이  $i-1$ 개의 연산이 일어난다. 마찬가지로,  $n$ 개 출력노드의 연산 수행량은  $[2i-1]n$ 번 발생한다.

다음 연산은 가중치 조절 식  $w_{ij}(t+i) = w_{ij}(t) + \eta(t)(x_i(t) - w_{ij}(t))$ 에서 [+] 연산이 1회, [x] 연산이 1회, [-] 연산이 1회 이루어진다. 승자노드와 연결된 가중치 벡터 값은 입력 값이  $i$ 개이므로 가중치 조절이  $i$ 번 이루어진다. 그러므로 가중치 조절에 대한 연산 수행량은 [+] 연산이  $i$ 회, [x]연산이  $i$ 회, [-] 연산이  $i$ 회가 된다. 즉 연산 수행량이  $3i$ 회 계산된다. 따라서 SOM 신경회로망에 대한 총 연산 수행량은  $[2i-1]n + 3i$  회이다.

### 5.2.2 결합 추천시스템의 연산 수행량

본 연구에서 제안한 결합모델은 SOM 신경회로망에 의해 클러스터링 된 사용자가 속한 출력노드에 협동적 필터링 기법을 적용하는 모델이다. 예를 들면, 출력노드가 10개면, 클러스터링 결과가 10개가 된다. 여기서,  $N$ 은 사용자의 전체 합이며, 각 클러스터 집합의 사용자 수는 그림 6과 같이 표현 할 수 있다.

여기서,  $P$ 는 타겟 사용자가 속하지 않은 클러스터의 수를 의미한다.

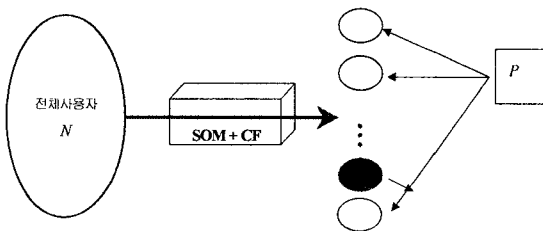


그림 6. SOM 신경회로망에 의해서 클러스터 된 집합의 사용자수

그러므로 SOM 신경회로망에 의해서 클러스터 된 집합의 사용자수는  $N-P$ 이다. 이에 따라 결합 추천시스템의 연산 수행량은 식 (6)과 같이 표현된다.

결합 모델의 연산 수행량=

$$25(N-P)M+6(N-P)+M+1+47 \quad (6)$$

따라서 기존의 협동적 필터링 기법과 본 연구에서 제안한 모델간의 결합 추천시스템의 연산 수행량의 차이는 식 (7)과 같이 표현된다.

연산 수행량의 차이=

$$(25NM+6N+M+1) - (25(N-P)M+6(N-P)+M+1+47) = 25NM+6N+M+1 - 25NM+25PM-6N+6P-M-48 (P \leq N) = (25M+6)P \quad (7)$$

식(7)은  $P$ 의 값에 따라서 연산 수행량이 증가한다. 각 클러스터의 연산 수행량은  $P$ 에 반비례하므로  $P$ 가 증가할수록 연산 수행량은 줄어들고  $P$ 가 작을수록 연산 수행량은 늘어난다. 즉,  $P$ 가 0과  $N$ 사이의 값이므로 다음 식(8)과 같이 나타낼 수 있다.

$$0 \leq (25M+6)P \leq (25M+6)N \quad (8)$$

식 (8)의 의미는  $P$ 에 따라서 연산 수행량이 줄어 들고 늘어나지만  $P=N$ 인 경우를 초과하지 않는다. 즉  $P=0$ 인 경우는 SOM 신경회로망에 의해 클러스터 된 집합에 모든 사용자가 포함된 경우를 의미하므로 협동적 필터링 기법과 동일한 경우이다.  $P=N$ 인 경우는 SOM 신경회로망에 의해 생성된 집합의 사용자가 하나도 존재하지 않는 경우를 의미하며 연산 수행량은 0이다.

본 논문에 적용된 결합모델의 연산 수행량은 다음과 같이 나타낼 수 있다. 먼저 연산 수행량은 인구 통계학적인 특징인 성별과 나이로 세그먼트 후, SOM 신경회로망을 통하여 클러스터링 하여 사용자가 속한 유사그룹을 추출 후 협동적 필터링 기법을 적용하여야 한다. SOM 신경회로망에 의해 분류에 대하여 협동적 필터링 기법을 적용시키는 경우의 연산 수행량은 기존의 협동적 필터링 기법보다 연산 수행량이 감소함을 알 수 있다.

## 6. 제안 모델과 기존 기법의 추천 효율 비교 분석

본 논문에서 제안한 기법의 효율성을 검증하기 위

하여 실험 결과를 비교 분석하였다. 전체 사용자를 대상으로 인구통계학적으로 분류하여 협동적 필터링 기법을 적용한 경우의 MAE와 SOM 신경회로망과 협동적 필터링 기법을 결합한 경우의 MAE를 식(9)를 통해 비교 분석하였다.

$$|\bar{E}| = \frac{\sum_{i=1}^n |e_i|}{n} \quad (9)$$

- n : 총 예측 회수
- $e_i$  : 실제 평가한 값과 예측한 값과의 오차 값의 절대치
- $|\bar{E}|$  : n개의  $e_i$ 값을 평가 개수 n개로 나눈 평균값의 절대치

6.1 인구통계학적 정보만을 고려한 실험

본 실험에서는 전체 사용자를 대상으로 인구통계학적으로 분류한 후 협동적 필터링 기법을 적용한 경우의 MAE를 분석하였다. 실험에서 사용한 인구통계학적인 정보는 나이와 성별이다. 그림 7은 인구통계학적으로 분류한 후 협동적 필터링 기법을 적용한 경우의 MAE 비교 값이다.

협동적 필터링 기법을 적용한 MAE 값보다, 나이와 성별로 세그먼트 한 후 협동적 필터링 기법을 적용한 MAE 값이 10개 그룹에서 6개 이상의 그룹에서 추천효율이 향상된 결과를 보여주고 있다.

6.2 인구통계학적인 정보와 SOM 신경회로망을 결합한 결과

전체 사용자를 대상으로 나이와 성별만으로 분류

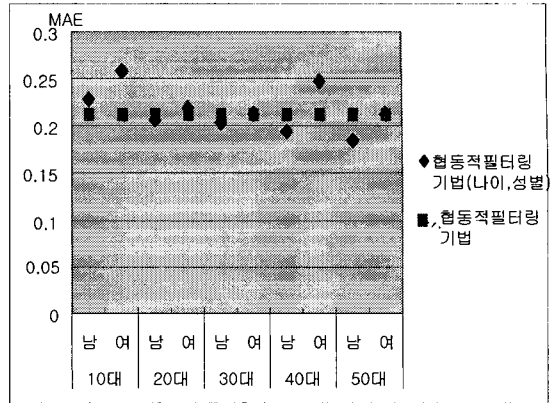


그림 7. 협동적 필터링 기법과 나이와 성별만을 고려한 협동적 필터링 기법의 MAE 값 비교

한 후 협동적 필터링 기법을 적용한 경우의 MAE를 분석하였다. 표 2는 나이와 성별을 고려한 MAE 값이다.

그림 8은 협동적 필터링 기법에 의한 MAE 값과 나이와 성별로 분류한 결과에 SOM 신경회로망을 적용시킨 후, 협동적 필터링 기법을 진행한 후의 MAE의 결과 값을 비교한 결과이다.

결과 비교에서처럼 협동적 필터링 기법의 MAE 값보다 제안 모델에서는 20개 그룹 중에서 15개 그룹이 낮은 MAE 값을 나타내어 제안 모델의 추천효율이 상당히 향상되는 결과를 보였다.

위 그림 9처럼 30대 여자 그룹은 2그룹에서 기존의 협동적 필터링 기법보다 매우 우수한 추천 효율을 보였고, 1그룹은 기존의 협동적 필터링 기법과 조금 향상된 추천 효율을 보였다. 남자 50대 2그룹에서는 기존의 협동적 필터링 기법보다 제안 모델이 매우

표 2. 제안 모델에서 나이와 성별을 고려한 MAE

나이	성별	클러스터링	MAE	나이	성별	클러스터링	MAE
10대	남	1	0.2308	40대	남	1	0.1490
		2	0.1627			2	0.2044
	여	1	0.2590		여	1	0.2748
		2	0.2087			2	0.1977
20대	남	1	0.1780	50대	남	1	0.2044
		2	0.2106			2	0.1616
	여	1	0.2243		여	1	0.0
		2	0.1800			2	0.3024
30대	남	1	0.2012	30대	여	1	0.2027
		2	0.1634			2	0.1746

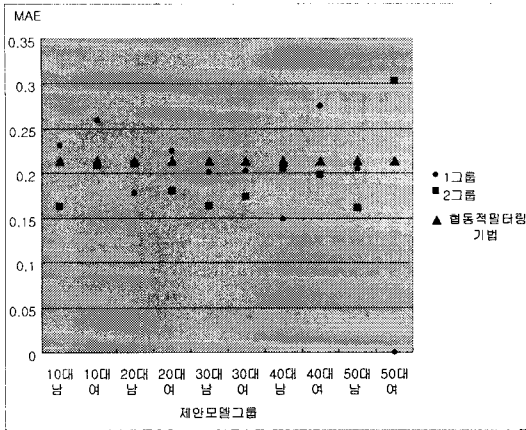


그림 8. 협동적 필터링 기법과 제안모델의 MAE 값 결과 비교

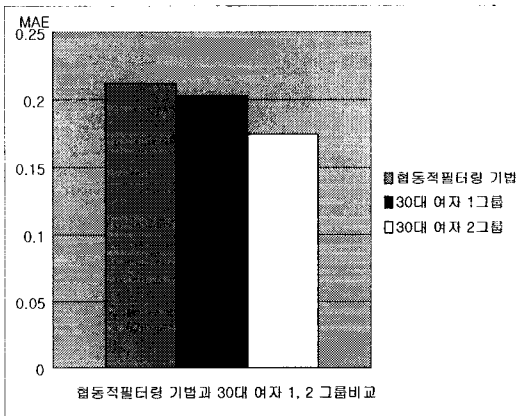


그림 9. 30대 여자 1·2그룹의 협동적 필터링 기법과 제안 모델의 실험 결과

우수한 추천 효율을 보였으나, 1그룹은 협동적 필터링 기법과 비교해서 추천 효율이 낮게 나타나는 결과를 보였다. 그 이유는 SOM 신경회로망의 특징이 입력패턴의 개수가  $i$ 개면, 출력노드의 개수가 입력패턴의 개수의 2배인  $2i$ 개로 정해져야 효율적인 클러스터링이 이루어지기 때문이다. 전처리 작업을 거친 데이터가 충분하지 못해 출력의 노드의 개수를 정하는데 한계가 있었다.

### 6.3 협동적 필터링 기법과 제안 모델의 수행 량 비교 분석

본 실험에서는 협동적 필터링 기법과 제안 모델의 연산 수행 량을 비교 분석하였다. 연산 수행 량은 5장의 일반적인 연산 수행 량 구하는 식에 근거하여 산

표 3. 협동적 필터링 기법과 제안 모델의 연산 수행 량 비교 분석

수행 종류	협동적 필터링 기법	SOM 신경회로망	제안기법 (남자10대 1그룹)
연산 수행 량 (회)	13,070,277	26,132	3,502,392

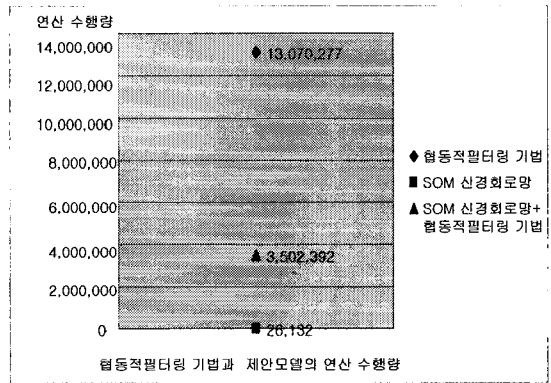


그림 10. 협동적 필터링 기법과 제안모델의 연산 수행 량 비교

표 4. 협동적 필터링 기법과 제안 모델의 수행시간 비교

수행 종류	협동적 필터링 기법	SOM 신경회로망	제안기법 (남자10대 1그룹)
수행시간 (초)	79	0.254	13

출하였다.

표 3에서처럼 제안 모델의 연산 수행 량이 협동적 필터링 기법보다 4배 정도 빠른 것으로 나타났다. 그림 10은 협동적 필터링 기법과 제안 모델의 연산 수행 량을 비교한 결과이다.

본 논문에서는 제안 모델에 대한 추천 효율 분석과 함께 수행시간을 비교 분석하였다.

협동적 필터링 기법의 수행시간과 SOM 신경회로망 수행시간, 그리고 제안 모델의 수행시간은 표 4에서처럼 제안 모델의 수행시간은 기존의 협동적 필터링 기법 수행시간보다 6배 이상 빠른 결과를 보였다.

## 7. 결 론

추천시스템은 사용자가 제공한 선호, 관심, 구매경험과 같은 정보를 근거로 하여 다른 사용자에게 가장

알맞은 정보를 제공하는 일련의 가치교환 과정인 개인화를 가능하게 하는 시스템으로 고객의 선호도를 정확히 분석하고, 정제하여 정확한 예측력으로 고객이 원하는 가장 적절한 상품을 추천 해줄 수 있어야 한다. 대부분의 추천시스템들이 협동적 필터링 기법을 적용하고 있어 본 논문에서는 협동적 필터링 기법의 연산 수행 량을 개선한 새로운 결합 모델인 SOM (Self-Organizing Map) 신경망 회로와 결합한 추천 시스템을 제안하였다.

협동적 필터링 기법은 임의의 사용자와 유사한 선호도를 가지는 다른 사용자의 평가정보를 이용하여 상품을 추천하는 방식이다. 그러나 대부분은 상관계수에 의한 문제가 발생한다. 사용자의 평가가 적은 경우에는 사용자간의 유사도를 구하기 어려운 관계로, 예측의 정확도가 떨어지는 문제가 발생한다. 또한, 사용자의 평가만을 근거로 다른 사용자와의 유사도를 측정함으로써 동일한 상품을 평가하지 않은 다른 사용자와의 관계를 완전히 배제하게 되어 사용자와 매우 유사한 관계에 있을 수도 있는 다른 사용자와의 유사성을 반영하기 어렵다. 뿐만 아니라 사용자의 선호도를 직접적으로 반영하지 못하므로 실제 사용자와 유사성이 거의 없고 선호도가 다른 사용자가 이미 평가한 작은 수의 상품에 의해 유사성이 있는 것으로 잘못된 추천을 하게 될 가능성이 있어 협동적 필터링 기법과 다른 기법을 결합한 hybrid 모델이 활발히 연구되었었다. 하지만 이런 연구 또한 협동적 필터링의 한계성을 극복할 수는 없었다.

본 논문에서는 사용자를 인구통계학적인 특성에 의해 세그먼트 한 후, 각 세그먼트별로 SOM 신경회로망을 적용하여 사용자를 클러스터링 하였다. SOM 신경회로망의 클러스터링은 영화 선호도를 기준으로 2개의 그룹으로 클러스터링 하였다. 또한 나이와 성별로 세그먼트 한 후 협동적 필터링 기법을 적용한 MAE 값이 전체 사용자에 대해 협동적 필터링 기법을 적용한 MAE 값보다, 10개 그룹에서 6개 이상의 그룹이 MAE 값이 낮게 나와 추천효율이 향상되었음을 보였다. 타겟 사용자가 속하는 그룹에만 협동적 필터링 기법을 적용한 MAE 값은 협동적 필터링 기법과 비교한 결과에서는 20개 그룹에서 15개 이상의 그룹이 MAE 값이 낮게 나와 추천효율이 상당히 향상됨을 보였다. 또한 SOM 신경회로망을 통과해서 분류된 집합을 가지고 협동적 필터링 기법을 적용시키는 경우

의 연산 수행 량이 SOM 신경회로망을 통과하지 않는 경우의 연산 수행 량 보다 훨씬 작음을 알 수 있다. 본 실험에서 제안 모델의 연산 수행 량이 협동적 필터링 기법을 행한 연산 수행 량과 비교해서 4배정도 빠른 것으로 나타났다. 그리고 수행시간을 비교한 결과 제안 모델의 수행시간이 기존의 협동적 필터링 기법 수행시간보다 6배 이상 빠르게 나타났다.

본 논문에서는 사용자를 클러스터링 할 때 나이와 성별만으로 세그먼트 한 후, 사용자가 평가한 영화의 장르 평가 평균값을 가지고 수행하였으나, 향후 연구에서는 좀 더 세분화된 인구통계학 적인 정보와 여러 가지 속성에 대한 선호도를 이용한다면 보다 동의어 문제와 모호집단에 대한 문제를 해결할 수 있을 뿐 아니라, 연산 수행속도를 개선시키기 위해서 다른 다양한 신경회로망을 적용하여 추천시스템의 성능을 향상시킬 수 있으리라 기대한다.

## 참 고 문 헌

- [1] 한국소프트웨어진흥원. <http://www.software.or.kr>, 2002.
- [2] J. B. Schafer and J. Konstan and J. Riedl, "Recommender systems in E-Commerce," *Proc. ACM Conference on Electronic Commerce*, pp. 158-166, 1999.
- [3] 한정기, "개인화(Personalization)의 핵심 기술," 2001. [http://www.personalization.co.kr/column\[010319\].htm](http://www.personalization.co.kr/column[010319].htm).
- [4] C. Lindquist, "Personalzation in E-Commerce," *Computerworld*, 1999.
- [5] J.B. Shcafer, J. A. Konstan, and J. Riedl, "E-Commerce Recommendation Applications," *GroupLens Research Project*, 1999.
- [6] B. Krulwich, "LIFESTYLE FINDER: Intelligent User Profiling Using Large-Scale Demographic Data," *Artificial Intelligence Magazine*, Vol.18, No.2, pp. 37-45, 1997.
- [7] M. J. Pazzani, "A Framework for Collaborative, Content-Based and Demographic Filtering," *Artificial Intelligent Review*, pp. 394-408, 1999.
- [8] M. Claypool, A. Gokhale, T. Miranda, P.

- Murnikov, D. Netes, and Matthew Sarti, "Combining Content-Based and Collaborative Filters in an Online Newspaper," *Proc. ACM SIGIR Wrokshop on Recommender Systems*, 1999.
- [9] C. Basu, H. Hirsh, and W. Cohen, "Recommendation as Classification: Using Social and Content-Based Information in Recommendation," *Proc. American Association for Artificial Intelligence*, pp. 714-720, 1998.
- [10] G. Karypis, "Evaluation of Item-Based Top-N Recommendation Algorithms," Technical Report CS-TR-00-46, *Computer Science Dept.*, Minnesota Univ., 2000.
- [11] B. M. Sarwar, G. Karypis, J. A. Konstan, and J. T. Riedl. "Item-Based Collaborative Filtering Recommendation Algorithms," *Proc. WWW Conference*, pp. 285-295, 2001.
- [12] B. M. Sarwar, G. Karypis, J. A. Konstan, and J. Riedl, "Analysis of Recommender Algorithms for E-Commerce," *Proc. ACM E-Commerce Conference*, pp. 158-167, 2000.
- [13] P. Resnick, N. Iacovou, M. Suchak, P. Bergstorm, and J. Riedl, "GroupLens: An Open Architecture for Collaborative Filtering of Netnews," *Proc. ACM 1994 Conference on Computer Supported Cooperative Work*, pp. 175-186, 1994.
- [14] B. M. Sarwar, J. A. Konstan, A. Borchers, J. Herlocker, B. Miller, and J. Riedl, "Using Filtering Agents to Improve Prediction Quality in the GroupLens Research Collaborative Filtering System," *Proc. ACM Conference on Computer Supported Cooperative Work*, pp. 345-354, 1998.
- [15] N. Good, J. B. Schafer, J. A. Konstan, A. Botchers, B. M. Sarwar, J. L. Herlocker, and J. Ricdl, "Combining Collaborative Filtering with Personal Agents for Better Recommendations," *Proc. Artificial Intelligence*, pp. 439-446, 1999.
- [16] M. O'Connor and J. Herlocker, "Clustering Items for Collaborative Filtering," Technical Report, Minnesota Univ., USA, 2000.
- [17] T. Kohonen, *Self-Organizing Maps, Information Sciences*. Springer, 1997.
- [18] S. Kaski, "Data Exploration Using Self-Organizing Maps," PhD thesis, Helsinki Univ., 1997.
- [19] T. Huntsberger and P. Ajjimarangsee, "Parallel Self-Organizing Feature Maps for Unsupervised Pattern Recognition," *Int. J. General Systems*, Vol.16, pp. 357-372, 1989.
- [20] T. Kohonen, *Self-Organizing Maps*. Springer, Berlin, 1997.
- [21] M. Claypool, D. Brown, P. Le, and M. Waseda, "Inferring User Interest," Technical Report WPI-CS-TR-01-97, 2001.
- [22] D. Billsus and M. Pazzani, "Learning Collaborative Information Filters," *Proc. the International Conference on Machine Learning*, pp. 46-53, 1998.
- [23] J. A. Konstan, B. N. Miller, D. Maltz, J. L. Herlocker, L. R. Gordon, and J. Reidl, "GroupLens: Applying Collaborative Filtering to Usenet News," *Proc. Communications of the ACM*, Vol.40, No.3, pp. 77-87, 1997.
- [24] D. Oard and J. Kim, "Implicit Feedback for Recommender Sytems," *Proc. AAAI Workshop on Recommender Systems*, pp. 80-82, 1998.



이 미 희

1981년~1985년 2월 경북대학교  
공과대학 전자공학과 공  
학사

1985년~1995년 경북대학교 공과  
대학 행정 및 실험실습  
조교

1987년~1990년 2월 경북대학교 산업대학원 공학석사  
1997년 3월~2008년 7월 마산 창원대학 모바일통신과  
조교수 근무

1999년~2004년 2월 창원대학교 공과대학 공학박사 취득  
2008년 8월~현재 마산 창원대학 인터넷컴퓨터과 학과장  
관심분야 : 웹프로그래밍, 통신프로그래밍, 데이터베이  
스, 추천시스템



우 용 태

1982년 경북대학교 전자공학과  
(공학사)

1984년 경북대학교 전자공학과  
(공학석사)

1995년 경북대학교 전자공학과  
(공학박사)

1987년~현재 창원대학교 컴퓨터  
공학과 교수

관심분야 : 데이터마이닝, 온톨로지 마이닝, 시맨틱웹