

# 역전파 신경망을 이용한 개인 맞춤형 상품 추천 시스템 구축

## Construction of Personalized Recommendation System Based on Back Propagation Neural Network

정귀임, 박상성, 신영근, 장동식  
고려대학교 정보경영공학부

Gwi-Im Jung(lily8424@korea.ac.kr), Sang-Sung Park(hanyul@korea.ac.kr),  
Young-Geun Shin(toctop@korea.ac.kr), Dong-Sik Jang(jang@korea.ac.kr)

### 요약

최근 고객 선호도에 맞는 정보 또는 상품을 예측하기 위한 연구들이 활발히 진행되고 있다. 고객의 만족도를 향상시키기 위해서 먼저 불필요한 정보들을 제거시켜야 하며 이러한 정보 필터링은 내용기반 필터링, 협업 필터링 등 여러 가지 기법을 통해 연구되고 있다. 본 논문에서는 기존 필터링 기법들의 문제점으로 나타나고 있는 희소성과 확장성을 해결하기 위해서 역전파 신경망을 이용하여 연구를 수행하였다. 신경망의 훈련 데이터는 설문조사를 통해 얻어진 데이터를 사용하였다. 최종적으로 설문조사를 통해 데이터를 수집하고 신경망 기반 추천시스템의 프로토타입을 제안하였고 기존 정보필터링 기법의 문제점을 개선하였다.

■ 중심어 : | 인공신경망 | 역전파 신경망 | 추천 시스템 | 개인화 서비스 |

### Abstract

Thousands of studies on predicting information and products that are suitable for customers' preference have been actively proceeding. In massive information, unnecessary information should be removed to satisfy customers' needs. This Information filtering has been proceeding with several methods such as content-based and collaborative filtering etc. These conventional filtering methods have scarcity and scalability problems. Thus, this paper proposes a recommendation system using BPN to solve them. Data obtained by survey questionnaire are used as training data of neural network. The recommendation system using neural network is expected to recommend suitable products because it creates optimal network. Finally, the prototype for recommendation system based on neural network is proposed to collect data and recommend appropriate methods through survey questionnaire. As a result, this research improved the problems of conventional information filtering.

■ keyword : | ANN(Artificial Neural Network) | BPN(Back Propagation Neural Network) | Recommendation System | Personalized Service |

\* 본 연구는 2007년도 두뇌한국 21 사업에 의하여 지원되었습니다.

\* 본 연구는 정보통신부 및 정보통신 연구진흥원의 대학 IT 연구센터 지원사업의 연구결과로 수행되었습니다. (IITA-2007-(C1090-0603-0025))

\* 본 연구는 정보통신부 및 정보통신 연구진흥원의 IT 신 성장 동력 핵심기술개발사업의 일환으로 수행하였습니다. [2007-S019-01. 정보투명성 보장형 디지털 포렌식 시스템 개발]

접수번호 : #071019-001

심사완료일 : 2007년 11월 27일

접수일자 : 2007년 10월 19일

교신저자 : 장동식, e-mail : jang@korea.ac.kr

## I. 서론

최근 인터넷의 급격한 발전은 기업의 경영환경을 상품 중심에서 사용자 중심으로 유도하고 있다. 이에 따라 전자상거래를 통한 마케팅 활동 또한 사용자의 요구에 초점을 두어 이루어지고 있으며 사용자 개인에게 맞춤형 제품을 추천하는 시스템이 연구되고 있다. 현재 인터넷 쇼핑물을 통해 사용자에게 제공되는 제품에 대한 정보는 점점 더 많아지고 있다. 하지만 사용자에게 불필요하거나 너무 많은 정보들이 제공되어 사용자가 이러한 정보들을 반영할 수 있기보다는 정보공해에 시달리고 있다. 따라서 이러한 정보화의 역기능을 보완하여 사용자의 요구에 적합한 정보 전달이 요구되며 사용자 개인에게 맞춤형 상품 추천시스템이 필요하다 [1][2].

개인화된 상품 추천시스템은 이미 미국에서는 Amazon.com[3]과 CD Now[4]에서 일반화되고 있으며 우리나라에서도 역시 개인화된 추천 서비스가 확대되고 있다. 추천시스템의 선행 연구로는 내용기반, 규칙기반, 협업 필터링 및 이외의 여러 기법을 이용한 연구들이 있다. 그 중에 내용기반과 협업 필터링 기법을 주축으로 연구되고 있으며 지금까지는 협업필터링 기법이 가장 성공적인 추천 기술로 여겨지고 있다. 하지만 선행기술에 쓰인 기법들이 정보의 확장성 및 희소성 등의 단점이 있다는 연구결과가 제시되면서 이러한 단점을 개선하기 위한 연구들이 진행되고 있다[5][6].

본 연구에서는 인공신경망(Artificial Neural Network: ANN)의 역전파 알고리즘을 적용한 협업필터링 기법을 이용하여 위에서 제시한 단점들을 보완하고 추천 능력을 개선할 수 있는 새로운 모델을 제안한다.

## II. 이론적 배경

추천시스템은 사용자에게 개인화된 서비스를 제공하는 시스템으로 e-CRM에서 실제적으로 가장 널리 알려진 수단이다. 추천시스템은 추천 방식에 따라 내용기반(content-based), 규칙기반(rule-based), 협업 필터

링(collaborative filtering) 등의 세 가지 종류로 분류할 수 있다.

### 1. 내용기반(Content-based) 추천시스템

내용기반 추천시스템은 사용자에게 있어 이전에 선호한 항목과 비슷한 속성을 갖는 항목은 사용자가 선호할 가능성이 높다고 보고 선호도가 표시된 항목들의 속성 정보를 이용하여 추천하는 기술이다. 내용 기반 추천시스템은 초기 연구에서 우수한 성능이 보인다고 평가되었다. 하지만 추천 상품의 속성을 멀티미디어 형식이 아닌 텍스트 형식으로만 이루어져야한다는 문제점과 특정 사용자의 과거 경험만을 바탕으로 하기 때문에 사용자간의 유사도가 전혀 고려되지 않는다는 문제점으로 사용자간의 상관관계에 따른 추천능력이 떨어진다는 연구결과가 제시되었다[7]. 이러한 문제점을 보완하고자 협업필터링 기법과 다른 추천 기법들이 연구되었다.

### 2. 규칙기반(Rule-based) 추천시스템

규칙기반 추천시스템은 사용자의 인터넷 쇼핑물 접촉의도를 파악하여 사용자의 의도에 맞는 추천을 위해 과거의 구매이력에 의한 연관관계 분석을 통한 추천기법이다. 규칙기반 추천 기법을 적용한 시스템으로는 사용자에게 적합한 개인화된 광고를 추천해주는 Broad-Vision의 One-to-One 시스템이 있다. 규칙기반 추천 기법은 가치 있는 규칙의 도출, 도출된 규칙의 유효성 평가와 규칙의 지속적인 업데이트와 같은 어렵고 복잡한 작업이 필요하다는 단점을 가지고 있다. 이를 해결하기 위해 규칙을 생성시키는 방법으로 연관성 규칙 기법이나 데이터 마이닝 기법을 적용하여 연구되고 있다[8].

### 3. 협업필터링(Collaborative filtering) 추천시스템

협업필터링은 사용자들의 선호도와 관심 표현을 바탕으로 선호도 관심에서 비슷한 패턴을 가진 사용자들을 식별해 내어 비슷한 취향을 가진 사용자들에게 관련

상품을 추천하는 기법이다. 협업 필터링의 초기연구는 Tapestry, GroupLens의 사례가 대표적이다[9].

Goldberg등에 의해 개발된 메일의 분류 실험 시스템인 Tapestry가 협업 필터링을 적용한 최초의 시스템이다. 이 밖에 협업필터링을 적용한 시스템으로는 음악 앨범 추천을 위한 링고시스템, 영화 추천용인 무비렌즈 시스템, 조크 추천용으로 제터 시스템, 그리고 온라인 라디오 추천용으로 플라이 캐스팅 등이 있다[10].

### 3.1 기존 협업필터링 적용시스템(NBCFA)

현재로서 가장 유명하고 대표적인 협력 시스템으로 알려진 GroupLens는 인터넷을 기반으로 형성된 토론 시스템인 usenet news 그룹의 기사를 추천하기 위해 최초로 자동화된 이웃 기반의 협업필터링 알고리즘(Neighborhood-Based Collaborative Filtering Algorithm: NBCFA)을 제안하였다[11]. NBCFA는 선정된 이웃과 추천 대상고객과의 상품에 대한 유사도라는 유사도로 나타낼 수 있다. 유사도를 구하는 방법으로는 피어슨 상관관계수, 벡터 유사도 등이 있다. GroupLens에서는 유사도 가중치로 피어슨 상관관계수를 사용하였으며 식(1)과 같다.

$$r_{uj} = \frac{\sum_{i=1}^m (R_{u,i} - \bar{R}_i)(R_{j,i} - \bar{R}_j)}{\sqrt{\sum_{i=1}^m (R_{u,i} - \bar{R}_i)^2 \cdot \sum_{i=1}^m (R_{j,i} - \bar{R}_j)^2}} \quad (1)$$

$R_{a,i}$  : 항목 I에 대한 사용자 a의 선호도

m : 사용자 a와 u 가 선호도를 표시한 항목의 개수

$\bar{R}_a$  : 각각 사용자 a의 선호도 평균

피어슨 상관관계수를 이용하여 고객 간의 유사도를 측정하여 적용된 usenet news의 최종 선호도 예측은 식(2)와 같이 계산되었다.

$$\hat{U}_x = \bar{U} + \frac{\sum_{J \in Raters} (J_x - \bar{J})r_{uj}}{\sum_{J \in Raters} |r_{uj}|} \quad (2)$$

$\bar{U}$  : 고객 u 자신의 선호도 평균

$J_x$  : 고객 j가 평가한 실제 선호도 평가지

Raters : 상품에 대해 선호도를 표시한 고객

NBCFA는 쉽게 적용할 수 있는 반면 데이터 수가 커질수록 많은 연산처리가 요구되는 확장성 문제(scalability problem)와 같은 단점이 있다[12][13].

### 3.2 기존 협업필터링 적용시스템(CMA)

협업 필터링의 다른 추천 기법으로는 GroupLens에서 사용된 NBCFA보다 향상된 예측력을 보이는 대응 평균 알고리즘(Correspondence Mean Algorithm: CMA)있다. CMA는 식(3)과 같이 나타낸다.

$$\hat{U}_x = \bar{U}_{match} + \frac{\sum_{J \in Raters} (J_x - \bar{J}_{match})r_{uj}}{\sum_{J \in Raters} |r_{uj}|} \quad (3)$$

NBCFA의  $\bar{U}$ 는 고객 u 자신의 선호도를 나타내는 반면 CMA의  $\bar{U}_{match}$ 는 고객 u 자신과 이웃 j가 공통의 선호도를 나타내기 때문에 고객 u의 선호도가 과도하게 반영되는 것을 개선할 수 있다[14].

기존 협업 필터링 방법은 비슷한 선호도를 가진 사용자들에게 관련 상품을 교차 추천하는 방법으로 유사 사용자들이 동일하게 평가한 상품에 대해서 또는 데이터가 많은 경우에 다른 기법에 비해 상대적으로 정확한 예측을 한다는 장점을 가지고 있다[15]. 반면 사용자들이 이질적인 평가 결과를 보이는 상품에 대해서는 예측력이 현저히 떨어지는 단점을 가지고 있으며 사용자의 데이터를 충분히 가지고 있다 할지라도 그 사용자가 구매하거나 평가를 내린 아이템이 적은 경우, 혹은 아이템의 종류가 너무 많은 경우에는 제대로 된 추천 결과를 내지 못하는 희소성 문제(sparsity problem)를 가지고 있다.

이상의 추천기법들 중에서 가장 성능이 뛰어난 협업 필터링의 문제점인 희소성과 확장성을 해결하고자 데이터 마이닝, SVM 알고리즘, 하이브리드 SOM, 베이지안 기법 및 신경망 등을 적용한 추천 방법들이 연구되

고 있다[16].

본 연구에서는 기존 추천 방법의 문제점인 가중치 적용, 한정된 데이터 처리와 협업필터링 추천 방법의 성능을 개선, 향상시키기 위해 ANN을 적용한 추천 방법을 제안하고 추천 시스템을 구축했다.

### III. 연구 모형

본 연구에서는 ANN을 적용한 추천시스템을 제안한다. ANN 추천시스템은 [그림 1]과 같이 ANN생성 단계와 ANN을 이용한 추천 단계로 나누어진다.

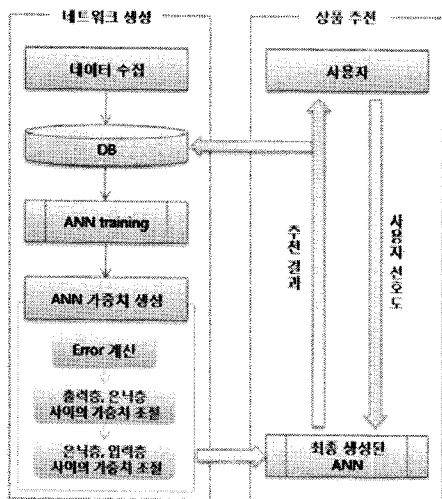


그림 1. 전반적인 연구모형

#### 1. ANN 생성 단계

사용자에게 상품을 추천하기 전에 추천시스템에 적용할 최적의 ANN을 생성하는 단계를 거친다. ANN을 훈련시키기 위한 사용자 선호도에 대한 데이터는 웹을 통해 수집하고 데이터베이스에 저장된다. 저장된 데이터는 ANN의 훈련 데이터로 사용된다. 훈련을 통해 오차를 최소화하는 반복학습이 이루어지고 반복학습 과정에서 오차를 최소화하기 위해 출력층과 은닉층, 은닉층과 입력층 사이의 연결 가중치 값이 갱신된다. 이 과정을 통해 입력 데이터에 맞는 최적의 연결 가중치가

정해지고 최종 가중치로 ANN이 생성된다[17].

#### 2. ANN 알고리즘

인공 신경망(Artificial Neural Network: ANN)은 다층 퍼셉트론(Multi-Layer Perceptron: MLP)의 효율적 학습 방법인 역전파 신경망(Back Propagation Neural Network: BPN)을 사용하였다. BPN은 입력층과 출력층 사이에 은닉층을 갖는 삼층전향구조로 되어있다. [그림 2]는 BPN의 구조를 보여준다.

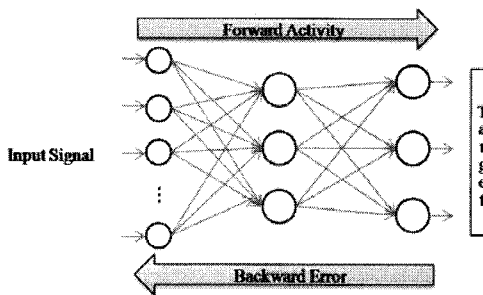


그림 2. BPN의 구조

BPN 학습방법은 초기 연결 가중치 값을 데이터에 적합한 값으로 변환하는 학습과정을 거치게 된다. 전향 단계는 신경망의 입력 패턴을 제시하고 각 노드에 대해서 입력함수와 활성화 함수를 이용하여 출력을 산출한다. 연결 가중치 값의 계산에 사용되는 활성화 함수는 식(4)와 같은 시그모이드 함수를 사용하였다.

$$\text{logsig}(x) = \frac{1}{1 + \exp^{-x}} \quad (4)$$

활성화 함수에 의한 은닉층 출력 값과 출력층의 출력 값은 식(5,6)과 같이 나타낼 수 있다.

$$h = \text{logsig}\left(\sum_{i=1}^n w_{ij}x_i\right) \quad (5)$$

$$y = \text{logsig}\left(\sum_{j=1}^m w_{jk}x_j\right) \quad (6)$$

$x_i$  = 입력변수,  $w_{jk}$  = 은닉층과 출력층 사이의 연결

가중치,  $w_{ij}$  = 입력층과 은닉층 사이의 연결가중치 후향 단계는 학습에서 중요한 요소인 연결 가중치 값의 갱신을 하는 단계로써, 목표값과 출력값의 차이를 식(7)을 통해 계산하여 오차를 구하고, 이를 출력층에서 입력층의 순서대로 오차 값을 최소로 하는 방향으로 층과 층 사이의 연결 가중치를 갱신한다. 이런 전·후향 단계를 거쳐 오차를 최소로 하는 최종적인 연결 가중치 값을 구해 재설정한다.

$$e = \frac{\sum_{n=1}^j (y - t)^2}{2} \quad (7)$$

### 3. 상품 추천 단계

상품 추천 단계는 실질적으로 사용자에게 정보를 제공하는 단계이다. 네트워크 생성 단계에서 입력데이터에 맞는 적합한 가중치가 생성되고 그 가중치를 적용하여 생성된 최종 ANN을 상품 추천 단계에 적용한다. ANN 알고리즘을 이용하여 구축된 상품 추천 시스템에 네트워크 생성 단계에서 만들어진 네트워크를 적용하여 최종적으로 신경망 네트워크 모형을 적용한 시스템을 완성한다. 웹을 통해 만들어진 ANN 추천시스템을 이용하여 사용자로부터 선호도를 입력 받고 입력받은 데이터는 추천 시스템을 거쳐 도출된 결과 값을 바탕으로 사용자에게 제공된다.

## IV. 실험 및 결과

본 장에서는 본 연구에서 제안한 ANN 추천시스템의 구축 과정과 그 결과를 설명하였다. 신희여행 상품을 아이템으로 하여 ANN 추천시스템을 구축하였으며 성능 평가를 위해 필요한 데이터를 수집하고 데이터 분석과 함께 ANN 추천시스템의 성능을 분석하였다.

### 1. ANN 생성 단계

#### 1.1 데이터 수집

ANN 추천시스템에 필요한 최종 네트워크 가중치 생

성과 성능 평가에 사용하기 위한 데이터를 웹 설문 조사를 통해 수집하였다. 데이터 수집 대상은 추천 상품의 주 고객층이 될 수 있는 20~30대인 본 연구자의 지인들로 구성하였다. 총 385개의 데이터를 수집하였으며 그 중 공란이 존재하거나 응답이 불성실하다고 생각되는 설문을 제외하고 350개의 데이터를 본 연구에 사용하였다. 350개의 데이터 중 80%는 ANN 추천 시스템의 훈련 데이터로 나머지 20%는 성능 평가를 위한 테스트 데이터로 사용하였다.

### 1.2 설문 문항 구성 및 데이터 분석

웹 설문문의 문항은 입력변수 11개와 출력변수 5개로 구성하였다. 입력변수는 사용자의 선호도를 휴양형, 휴양+관광형, 고급휴양형, 레저형, PIC/CLUB MED와 같이 5가지 요인으로 분류하여 각 요인을 설명하는 설문을 구성하였다. 각 요인에 대한 정의는 [표 1]과 같다.

표 1. 각 입력 요인의 정의

요 인	내 용
휴양형	주로 리조트 안에서 시간을 갖는 간단한 시내관광을 즐기는 것 외에는 자유시간이 주어지는 여행 일정이 여유로운 상품으로 이루어진다.
휴양+관광형	관광과 휴양, 두 가지를 같이 즐길 수 있는 상품으로 관광일정을 통해 여행자의 문화를 체험할 수 있는 상품으로 이루어진다.
고급휴양형	고급스러운 휴양지에서 최고의 서비스를 제공하는 상품으로 이루어진다.
레저형	여행일정이 활동적인 레저체험을 중심으로 계획된 상품으로 이루어진다.
PIC / CLUB MED	레저스포츠 및 엔터테인먼트를 즐길 수 있는 시설을 모두 갖춘 리조트 안에서 여행자가 자유롭게 여행 일정을 계획 할 수 있는 상품으로 이루어진다. 여행 기간 동안 주로 리조트 안에서 생활하게 된다.

설문 문항은 각 변수의 개념에 맞게 여행사에서 제공하는 정보를 참고로 하여 본 연구에 맞게 작성하였다. 출력 변수는 현재 여행 사이트에서 사용하고 있는 여행 상품 중에서 인기도가 높은 상품으로 선정하였다. 구성된 설문 문항을 [표 2]에 제시하였으며 웹 설문 화면은 [그림 3]과 같이 구현하였다.

표 2. 설문 문항

요인	문항	내용	요인	문항	내용
휴양형	NO1_1	액티비티가 적은 저렴한 비용의 여행상품을 선호하는 편이다.	고급 휴양형	NO3_1	최상의 숙박 시설을 갖춘 곳을 선호하는 편이다.
	NO1_2	조용하고 한적한 여행을 선호하는 편이다.		NO3_2	최상의 서비스를 제공하는 여행을 선호하는 편이다.
	NO1_3	여행 일정이 여유로운 타입을 선호하는 편이다.	레저형	NO4_1	여유로운 여행보다 활동적인 여행을 선호하는 편이다.
휴양 + 관광형	NO2_1	휴양보다는 관광을 즐길 수 있는 여행을 선호하는 편이다.		NO4_2	활동적인 레저체험을 마음껏 즐길 수 있는 여행을 선호하는 편이다.
	NO2_2	관광과 휴양을 함께 할 수 있는 여행을 선호하는 편이다.	PIC/CLUB MED	NO5_1	이동이 많지 않은 여행을 선호하는 편이다.
	NO2_3	시내 중심지와 가까운 리조트를 선호하는 편이다.		NO5_2	일정을 자유롭게 조절할 수 있는 여행을 선호하는 편이다.

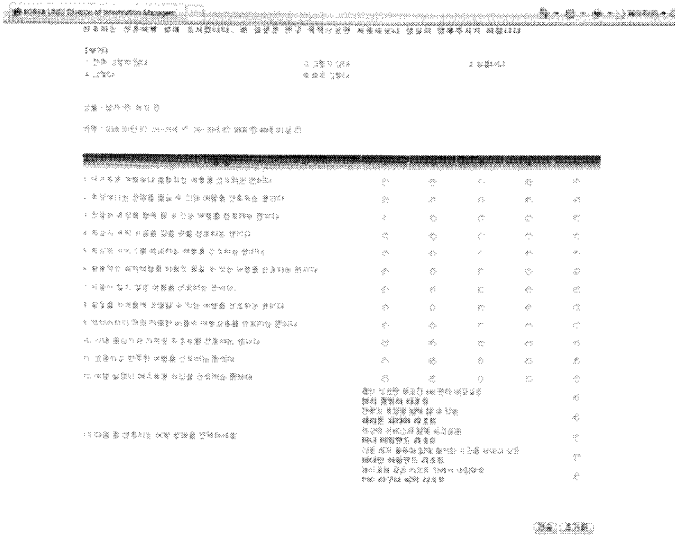


그림 3. 웹 설문 화면

이와 같이 구성된 웹 설문을 통해 수집된 데이터를 본 실험에 사용하기에 앞서 데이터 검증을 시행하였다. 데이터들의 내적 일관성 검증을 위한 신뢰성 분석과, 타당성 검증을 위한 요인분석을 수행 하였다. 신뢰성은 크론바하 알파 값을 계산하여 0.7을 넘으면 신뢰성이 있다고 보았다. 또한 요인분석은 요인추출방법으로 주 성분 분석(principal component method)을 사용하였고 회전방법으로는 베리맥스(varimax rotation)를 사용하였다. 요인분석에서의 요인 적재값(factor loading)은 0.4이상, 고유값(eigen value)은 1이상을 기준으로 하였다. 각 기준 값은 기존 논문에서 사용되어진 값을 참고 하였다. 데이터 분석 결과는 [표 3]과 같다.

표 3. 요인 분석 및 신뢰성 분석 결과

문항	적재값	고유값	Cronbach' a
NO1_1	0.828	2.902	0.882
NO1_2	0.852		
NO1_3	0.382		
NO2_1	0.931	2.105	0.967
NO2_2	0.938		
NO2_3	0.948		
NO3_1	0.956	1.939	0.946
NO3_2	0.959		
NO4_1	0.828	1.713	0.851
NO4_2	0.852		
NO5_1	0.964	1.493	0.791
NO5_2	0.692		

분석 결과 각 측정문항의 요인 적재 값은 NO1\_3 문항을 제외한 나머지 문항 모두 0.5이상으로 수용 가능한 것으로 나타났다. 또한 크론바하 알파값은 5개 변수 모두 0.7 이상으로 데이터의 신뢰성을 입증하였다. 적재 값이 0.382가 나온 NO1\_3인 '여행 일정이 여유로운 타입을 선호하는 편이다.' 문항을 제외한 나머지 문항을 본 실험에 사용하였다.

### 1.3. 네트워크 가중치 생성 프로그램

ANN 추천시스템에 사용될 최종 네트워크 가중치를 만들기 위해 별도의 네트워크 가중치 생성 프로그램을 구축하였다. 대량의 데이터를 웹상에서 훈련시키기 위해서는 많은 연산시간이 필요하기 때문에 이런 문제를 해결하기 위해 네트워크 가중치 생성 시스템을 통해 시간을 단축 시켰다. 네트워크 가중치 생성 시스템은 Visual C# 2005를 이용하여 구축하였으며 [그림 4]에는 C#으로 구현한 네트워크 가중치 생성 시스템을 나타내었다.

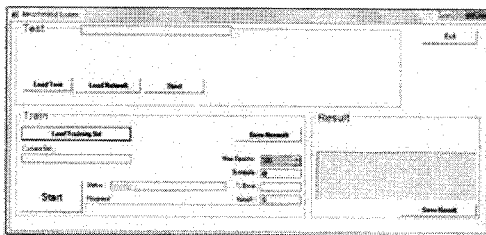


그림 4. 네트워크 가중치 생성 시스템

시스템을 통해 Learning rate 변화에 따른 성능을 평가해 보았다. 그 결과 Learning rate가 150일 때 가장 좋은 성능을 나타냈으며 그 밖에 Training Parameter는 최대반복수=300, 에러 임계치=0.0002 으로 설정하였다. ANN 가중치 생성 시스템을 통해 나온 최종 네트워크의 가중치는 [그림 5]와 같다.

입력층 i개, 은닉층 j개, 출력층 k개라고 할 때 [그림 5]에서 보여주는 weight[1, 0, 2]의 1은 출력층과 은닉층의 사이를 의미하고 0과 2는 은닉층 0번째 노드와 출력층 2번째 노드사이의 가중치 W02를 의미한다. 또한 weight[2, 0, 10] 역시 2는 입력층과 은닉층의 사이를 의

미하고 입력층 0번째 노드와 은닉층 10번째 노드의 가중치를 의미한다.

	1	2	3
1	Weight[1, 0, 0] =	-19	
2	Weight[1, 0, 1] =	10.29486	
3	Weight[1, 0, 2] =	16.82537	
4	Weight[1, 0, 3] =	189.4288	
5	Weight[1, 0, 4] =	-76.33733	
6	Weight[1, 0, 5] =	23	
7	Weight[1, 0, 6] =	37.83096	
8	Weight[1, 0, 7] =	50.10589	
9	Weight[1, 0, 8] =	172.2848	
10	Weight[1, 0, 9] =	-107.0098	
11	Weight[1, 0, 10] =	1	
12	Weight[2, 0, 0] =	18.35711	
13	Weight[2, 0, 1] =	-9.890671	
14	Weight[2, 0, 2] =	-26.48273	
15	Weight[2, 0, 3] =	2.04012	
16	Weight[2, 0, 4] =	-24.48093	
17	Weight[2, 0, 5] =	-22.32882	
18	Weight[2, 0, 6] =	26.44646	
19	Weight[2, 0, 7] =	-7.670854	
20	Weight[2, 0, 8] =	4.813305	
21	Weight[2, 0, 9] =	8.292356	
22	Weight[2, 0, 10] =	-20.52636	

그림 5. ANN 가중치

## 2. 상품 추천 단계

### 2.1 사용자 선호도 입력

위 데이터 분석 결과에서 요인 적재 값이 기준치 이하로 나타난 NO1\_3의 '여행 일정이 여유로운 타입을 선호하는 편이다.' 문항을 제외한 나머지 문항을 사용하여 사용자의 선호도 입력 문항을 재구성하였다. 사용자 선호도 입력 화면은 웹을 통해 구현 하였으며 [그림 6]과 같이 나타났다.

### 2.2 상품 추천 결과

네트워크 가중치 생성 프로그램에서 생성된 최종의 네트워크 가중치를 ANN 추천시스템 알고리즘에 적용하여 입력데이터에 적합한 ANN 추천시스템이 완성된다. 사용자의 선호도 입력 후 추천 알고리즘을 통해 사용자에게 제시되는 추천 결과 값은 위 변수 선정에서 5가지 요인으로 분류된 휴양형, 휴양+관광형, 고급휴양형, 레저형, PIC/CLUB MED에 해당하는 상품 3가지를 제시한다. [그림 7]은 추천 결과 값 중에서 'PIC/CLUB MED'에 해당되는 상품 군을 추천한 화면이다.

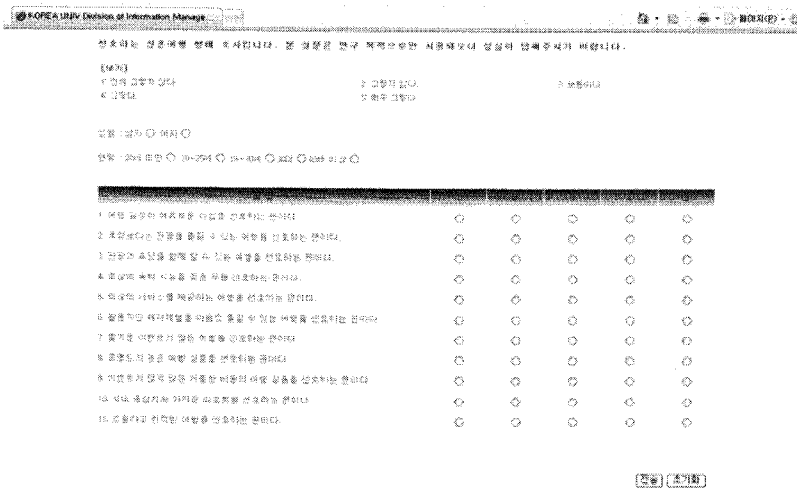


그림 6. 사용자 선호도 화면

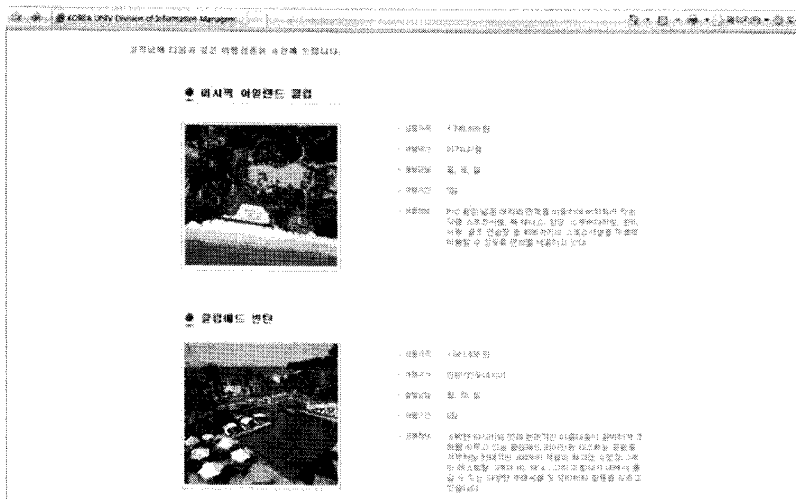


그림 7. 'PIC/CLUB MED'의 추천 결과 화면

### 3. ANN 추천 시스템 성능 평가

본 연구에서 제안한 ANN 추천시스템의 정확도측정을 위해 성능 평가를 시행하였다. ANN 추천시스템의 성능 평가는 본 연구에서 수집한 데이터를 사용한 평가와 함께 협업필터링 추천시스템 연구에서 주로 사용되는 Movielens 데이터를 사용하여 성능을 평가하였다. Movielens의 데이터를 사용하여 성능 평가 실험을 수행함으로써 ANN 추천시스템의 성능을 다시 한 번 검

증할 수 있으며 본 연구에서 사용된 여행상품의 데이터에 한정되지 않고 다른 데이터와의 유연성을 입증할 수 있다. 또한 많은 논문에서 검증된 데이터를 사용함으로써 정확한 결과를 도출할 수 있으며 선행 논문에서 제안한 추천기법과 비교평가가 가능하다.

MovieLens의 데이터는 GroupLens 연구소에서 공개한 영화에 대한 평가 자료로써 100K, 1million dataset의 두 가지로 구성되어있다. 100K 데이터는 943명의 사



용자를 대상으로 1682개의 영화에 대하여 1-5점 척도로 평가 되었으며 100,000개의 선호도 평가치로 구성되어 있다. 1million은 6040명이 3900개의 영화에 대하여 평가한 자료로 1,000,209개의 선호도 평가치로 구성되어 있다. 이 중 100K의 100,000개의 데이터를 실험에 사용하였다.

본 실험에서 사용한 측정지표는 통계적 정확도 측정 지표인 정규화 된 MAE(NMAE: Normalized Mean Absolute Error)를 사용하였다. MAE는 모든 테스트 대상에 대해서 평가치와 예측치 간의 오류를 구하고 이 오류의 절대값을 합한 후 테스트 대상의 수로 나누어 구할 수 있다. MAE가 낮을수록 예측의 정확도는 좋아지게 된다. 본 연구에서는 측정 데이터의 선호도 척도가 다르기 때문에 MAE를 식(9)와 같이 정규화 하여 분석하였다. MAE와 NMAE는 다음과 같이 표현된다[14].

$$MAE = \frac{\sum_{i=1}^n |x_i - \hat{x}_i|}{n} \quad (8)$$

$$NMAE = \frac{MAE}{Range\ of\ rating} \quad (9)$$

본 논문에서는 [18]의 연구에서 제시한 최근접 이웃을 추출하는 협업필터링 추천방법(CBR\_CF), 사용자의 선호도별로 클러스터링 된 소속 군집 내에서 사례기반 추론을 수행하는 협업필터링 방법(UC\_CBR\_CF), 사용자별 클러스터링 사례기반추론과 아이템 속성별 클러스터링에 기반을 둔 사례기반추론을 결합한 하이브리드 추천 시스템(UC\_IC\_CBR\_CF)과 [19]의 연구에서 제시한 K-means, 신경망 알고리즘인 SOM을 이용하여 실험한 성능 측정 결과를 비교대상으로 하여 분석하였다.

추천기법의 성능을 평가한 결과는 표 4에 제시하였다. 그 결과 Movielens 데이터로 평가한 ANN 추천 시스템의 성능이 0.051, 본 연구에서 수집한 Honeymoon 데이터로 평가한 ANN 추천 시스템의 성능이 0.075로 기존 추천방법보다 더 우수하다는 것을 알 수 있었다.

표 4. 추천시스템의 성능 평가 결과

추천 방법	NMAE
CBR_CF	0.123
UC_CBR_CF	0.107
UC_IC_CBR_CF	0.088
K-means	0.166
SOM	0.158
Movielens_ANN 추천시스템	0.051
Honeymoon_ANN 추천시스템	0.075

## V. 결론

본 논문에서는 기존 추천시스템의 단점을 개선하고 추천 정확성을 향상시키기 위해 인공신경망을 이용한 추천시스템을 제안하였다. 인공신경망을 이용한 ANN 추천시스템을 설계하고 기존 추천시스템의 추천 기법과 성능을 비교 분석하였다. 기존 추천 기법보다 우수하다고 평가되는 결과를 토대로 php와 C#을 이용하여 실제 적용 가능한 ANN 추천시스템의 구현을 통해 인터넷상에서 개인화 서비스를 행할 수 있도록 하였다. ANN 추천 시스템은 다음과 같은 우수성을 지닌다.

첫째, ANN 추천시스템은 기존 사용자 혹은 아이템 기반 추천 기법에서 가중치를 고려하지 않아 추천 정확도가 떨어진다는 단점을 개선하였다. 사용자간이나 아이템간의 가중치를 학습 할 수 있으므로 보다 정확한 선호도 계산이 가능하다. 또한 이질적인 정보를 쉽게 입력노드를 추가하여 처리할 수 있다.

둘째, ANN 추천시스템 구축에 있어서 C#을 이용하여 가중치 생성 시스템을 병합함으로써 고객그룹과 거래 데이터가 많아질수록 유사한 고객그룹을 찾기 위한 연산량이 기하급수적으로 증가하는 단점을 보완할 수 있다. 가중치 생성에 있어서 데이터의 양과 연산시간은 비례한다. 정확한 추천 결과를 위해 필요한 많은 데이터의 연산시간은 증가할 수밖에 없으며 이를 웹을 통해 구현하는 것 보다 C#을 이용하여 구현하는 것이 연산시간을 줄일 수 있다. 또한, 판매자의 시스템 활용 면에서 시스템 업데이트할 때 쉽게 사용할 수 있다.

셋째, 웹을 통한 ANN 추천시스템은 다른 추천시스

템보다도 실제 사용자에게 쉽고 정확한 개인화 서비스를 제공할 수 있으며 추천 상품의 종류에 관계없이 쉽게 적용할 수 있다.

향후 연구에서는 데이터 축적에 따라 자동적으로 네트워크 가중치가 업데이트 되도록 자동화된 시스템 연구가 필요하다.

**참고 문헌**

[1] M. J. Pazzani, "A Framework for Collaborative, Content-Based and Demographic Filtering," *Artificial Intelligence Review*, Vol.13, No.5-6, pp.393-408, 1999.

[2] J. B. Schafer, D. Frankowski, J. Herlocker, and S. Sen, "Collaborative Filtering Recommender Systems," *Lecture Notes in Computer Science*, Vol.4321, pp.291-324, 2007.

[3] <http://www.amazon.com/>

[4] <http://www.cdnw.com>

[5] M. Balabanovic and Y. Shohm, "Fab: Content Based, Collaborative Recommendation," *Communications of the ACM*, Vol.40, No.3, pp. 66-72, 1997.

[6] J. H. Ahn and T. W. Hong, "Collaborative filtering for recommender systems: a scalability perspective," *International Journal of Electronic Business*, Vol.2, No.1, pp.77-92, 2004.

[7] M. Claypool, A. Gokhale, T. Mirana, P. Murnikv, D. Netes, and M. Sartin, "Combing Content-Based and Collaborative Filters in an Online Newspaper," *Proc. of Workshop on Recommender Systems Implementation and Evaluation*, 1999.

[8] I. Chun and I. Hong, "The implementation of knowledge-based recommender system for electronic commerce using Java expert system library," *Proceedings of IEEE International Symposium on Industrial Electronics*, pp.1766-1770, 2001.

[9] D. Goldberg, D. Nichols, B. B. Oki, and D. Terry, "Using Collaborative filtering to Weave an information Tapestry," *Communications of the ACM*, Vol.35, No.12, pp.61-70, 1992.

[10] S. Upendra and M. Patti, "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.

[11] P. Resnick, N. Iacovou, M. Suchak, and P. Bergstrom, "GroupLens: An open architecture for collaborative filtering of netnews," *Proceedings of the ACM Conference on Computer Supported Cooperative Work*, pp.175-186, 1994.

[12] B. M. Sarwar, G. Karypis, J. A. Konstan, and J. Riedl, "Item-Based Collaborative Filtering Recommendation Algorithms," *Proceedings of the 10th International World Wide Web Conference*, pp.285-295, 2001.

[13] 김재경, 서지혜, 안도현, 조운호, "A personalized recommendation methodology based on collaborative filtering", *한국지능정보시스템학회 논문지*, 제8권, 제2호, pp.139-157, 2002.

[14] H. C. Lee, "An Exploratory Study for Decreasing Error of Prediction Value of Recommended System on User Based," *Journal of Korean Data & Information Society*, Vol.17, No.1, pp.77-86, 2006.

[15] J. Herlocker, J. A. Konstan, A. Borchers, and J. Riedl, "An Algorithmic Framework for Performing Collaborative Filtering," *Proceedings of the 1999 Conference on Research and Development in Information Retrieval*, 1999.

[16] K. Goldberg, T. Roeder, D. Gupta, and C. Perkins, "Eigentaste: a constant

time-collaborative filtering algorithm," Information Retrieval Journal, Vol.4, No.2, pp.133-151, 2001.

[17] W. Huang and R. Lippmann, "Neural net and traditional classifiers," Advances in Neural Information Processing Systems, Vol.1, pp. 387-396, 1988.

[18] 홍태호, 이희정, 서보밀, "클러스터링 기반 사례 기반추론을 이용한 웹 개인화 추천 시스템", 한국 지능정보시스템학회, 제11권, 제1호, pp.107-121, 2005.

[19] X. Zhang, J. Edwards, and J. Harding, "Personalised online sales using web usage data mining," Computers in Industry, Vol.58, Issues 8-9, pp.772-782, Dec. 2007.

**저자 소개**

**정 귀 임(Gwi-Im Jung)**

준회원



- 2006년 2월 : 충주대학교 산업경영공학과 (공학사)
- 2006년 2월 ~ 현재 : 고려대학교 정보경영공학부 석사과정

<관심분야> : 전문가시스템, 패턴인식

**박 상 성(Sang-Sung Park)**

정회원

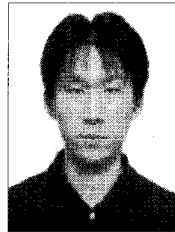


- 2006년 2월 : 고려대학교 산업시스템정보공학과 (공학박사)
- 2006년 5월 ~ 현재 : 고려대학교 BK21 사업단연구교수

<관심분야> : 컴퓨터 비전, 패턴인식, 지식관리 전문가시스템 응용

**신 영 근(Young-Geun Shin)**

준회원

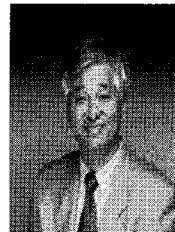


- 2005년 2월 : 고려대학교 산업시스템정보공학과 (공학사)
- 2005년 9월 ~ 현재 : 고려대학교 정보경영공학부 석박사 통합과정

<관심분야> : 패턴인식, 스케줄링, 인공지능

**장 동 식(Dong-Sik Jang)**

정회원



- 1979년 : 고려대학교 산업공학과 (공학사)
- 1985년 : 텍사스 주립대학 산업공학과 (공학석사)
- 1988년 : 텍사스 A&M 산업공학과 (공학박사)

• 1989년 ~ 현재 : 고려대학교 정보경영공학부 교수  
<관심분야> : Computer Vision, 최적화이론, 컴퓨터 알고리즘