

# 하이브리드 SOM을 이용한 동적 웹 정보 추천 기법

윤 경 배\* · 박 창 희\*\*

## 요 약

최근, 사용자에게 가장 필요한 정보를 제공하기 위한 웹 정보 추천 시스템에 대한 연구가 인터넷 쇼핑몰등을 대상으로 활발히 진행되고 있다. 그 중 SOM(Self-Organizing Feature Maps)을 이용한 동적 웹 정보 추천 기법은 빠른 수행 속도와 간편하게 사용할 수 있는 장점이 존재하나, 모형에 대한 설명력 부족 및 최종적으로 구축된 모형에서 출력층의 각 노드가 한 개의 가중치 값들로 고정되는 단점이 존재한다. 본 논문에서는 이러한 단점인 모형에 대한 설명력 부족을 베이지안 추론 기법으로 해결하며, 하이브리드 SOM을 이용하여 최종적으로 구축된 모형에서 출력층의 각 노드가 한 개의 가중치 값들로 고정되는 것이 아니라 가중치가 속하게 되는 분포가 결정되도록 한다. 이러한 하이브리드 SOM을 이용하여 동적 웹 정보 추천 기법을 설계하고 구현하여 기존의 웹 정보 추천 기법과 성능 비교를 수행한 결과, 제안된 기법의 우수함이 입증되었다.

## Dynamic Web Recommendation Method Using Hybrid SOM

Kyungbae Yoon<sup>\*</sup> · Changhee Park<sup>\*\*</sup>

## ABSTRACT

Recently, provides information which is most necessary to the user the research against the web information recommendation system for the Internet shopping mall is actively being advanced the back which it will drive in the object. In that Dynamic Web Recommendation Method Using SOM (Self-Organizing Feature Maps) has the advantages of speedy execution and simplicity but has the weak points such as the lack of explanation on models and fixed weight values for each node of the output layer on the established model. The method proposed in this study solves the lack of explanation using the Bayesian reasoning method. It does not give fixed weight values for each node of the output layer. Instead, the distribution includes weight using Hybrid SOM. This study designs and implements Dynamic Web Recommendation Method Using Hybrid SOM. The result of the existing Web Information recommendation methods has proved that this study's method is an excellent solution.

**키워드 :** 하이브리드(Hybrid), 신경망(Neural Networks), SOM(Self-Organizing Feature Maps), 추천(Recomendation), 군집화(Clustering)

### 1. 서 론

인터넷의 대중화와 발전은 전자상거래와 같은 새로운 산업 구조를 도래 시켰으며, 넘쳐나는 정보의 홍수 속에서 이제는 인터넷을 떠나서는 살아갈 수 없을 정도로 인간 활동의 많은 부분을 차지하게 되었다. 그러나 정보구조(information architecture)에 대해 연구하는 Forrester Research의 연구에 의하면 “전체 인터넷 사이트의 51%는 웹 페이지에서 제시하는 내용을 쉽게 이해할 수 없게 구성되어 있고, 전체 사이트들 가운데 90% 이상이 적절치 못한 구조를 지니고 있다”고 보고하고 있다[7].

현재의 무질서한 인터넷 환경에서 특정 정보를 찾고자 무작정 웹을 서핑한다는 것은 매우 비효율적인 행위가 아닐

수 없다. 따라서 정보의 구조화를 통해 웹 사이트의 제정비를 피하고, 현재 인터넷 웹 사이트의 구조적 비효율성을 제거하여 사용자 하여금 좀더 쉽고 경제적으로 정보를 얻을 수 있는 방안 제시에 관한 연구가 절실하며, 이러한 인터넷 환경에서 사용자에게 꼭 필요한 정보를 제공해 줄 수 있는 정보 서비스 시스템의 필요성이 대두되고 있다[2, 3, 8].

웹 정보의 추천을 위해서는 사용자 집단의 분류가 필요하며, 이를 위해서 연관성 기반의 추천 기법, 통계적 및 군집화 기반의 사용자 모델링, SVM 등 여러 추천기법이 연구되고 있으며, 그 중 신경망(neural network)모형은 인간이 경험으로부터 학습해 가는 두뇌의 신경망 활동을 모방하여 자신이 가진 데이터로부터 반복적인 학습과정을 거쳐 패턴을 찾아내고 이를 일반화하기 때문에 향후 필요 정보를 예측하고자 하는 문제에 유용하게 이용할 수 있다[6].

본 논문에서는 신경망의 일종인 SOM(Self-Organizing Feature Maps)의 단점을 보완한 하이브리드 SOM을 이용하

\* 이 논문은 2004년도 김포대학의 연구비 지원에 의하여 연구되었음.

† 정 회 원 : 김포대학 컴퓨터계열 교수

\*\* 준 회 원 : 연세대학교 대학원 전자공학

논문접수 : 2003년 12월 1일, 심사완료 : 2004년 6월 14일

여 사용자 집단을 분류하고 분류된 사용자의 특성에 따라 일반화된 패턴을 찾아 동적으로 웹 정보를 추천하는 기법을 구현한다.

2. 동적 추천 요소 기법

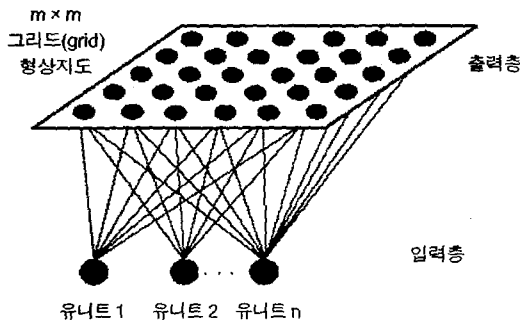
연관성 기반의 추천 기법, 통계적 및 군집화 기반의 사용자 모델링, SVM 등은 현재 사용되고 있는 대표적인 추천 요소 기법들이며 이는 <표 1>과 같다.

<표 1> 기존의 동적 추천 요소 기법

기존 연구 기법	장 점	단 점
연관성 기반의 추천 기법 [4]	<ul style="list-style-type: none"> <li>- 우수한 정확도</li> <li>- 연속형 데이터에 대한 높은 예측력</li> <li>- 변수간의 관계를 상관 계수를 통해 쉽게 파악</li> </ul>	<ul style="list-style-type: none"> <li>- 모형 구축을 위한 학습 시간이 오래 소요</li> <li>- 변수간의 비선형적인 관계 파악은 불가</li> <li>- 상관 계수가 크다고 반드시 연관성이 큰 것은 아님</li> </ul>
통계적/신경망 기반의 군집화 사용자 모델링 [5,9]	<ul style="list-style-type: none"> <li>- 모형 구축을 위한 학습 시간이 매우 빠름</li> <li>- 모형에 대한 해석이 용이 (통계적 군집화)</li> <li>- 복잡한 데이터 구조에 적용이 가능(신경망 기반의 군집화)</li> </ul>	<ul style="list-style-type: none"> <li>- 군집 결과에 대한 비수렴성(신경망 기반의 군집화)</li> <li>- 초기 설정이 주관적이고 이 설정에 의해 군집의 결과가 달라짐</li> </ul>
SVM (Support Vector Machine) [10]	<ul style="list-style-type: none"> <li>- 모형 구축을 위한 학습 시간이 매우 빠름</li> <li>- 데이터의 비선형적인 구조까지 모형화</li> <li>- 최소한의 데이터의 분석에 우수한 성능을 보임</li> </ul>	<ul style="list-style-type: none"> <li>- 모형에 대한 정확도가 다소 떨어짐</li> <li>- 모형의 결과에 대한 해석이 어려움</li> <li>- 다중 패턴의 분류가 어려움</li> </ul>

2.1 SOM을 이용한 군집화

인간의 뇌 구조를 가장 잘 모형화한 방법은 1980년대 초에 Kohonen[9]에 의해 제안된 자기 조직화 형상 지도(Self-Organizing feature Maps : SOM) 모형이다. SOM은 신경망 중에서도 학습 자료에 대한 결과값을 모르고 학습이 수행되는 자율 학습(unsupervised learning) 구조를 갖고 있다. 음성 인식, 문자 인식, 구문 분석 등 다양한 분야에 응용되는 SOM은 (그림 1)과 같이 입력층과 출력층으로 구성된 순방향 단층 신경망 구조를 갖는다.



(그림 1) SOM의 구조

그러나 SOM이 광범위하게 적용되고 많은 중요한 연구들이 이루어져 왔지만 아직도 해결되지 않은 여러 문제들이 있다[5]. 첫째로 SOM은 자료 공간(data space) 내에서 밀도 함수(density function)가 정의되고 있지 않다. 즉 학습 데이터의 분포와 가중치 분포간의 관계가 정형화(formalized)되어 지는 방법을 갖고 있지 못하다. 둘째로 SOM의 학습 알고리즘에는 오차(error)를 관리하는 목적 함수(objective function)를 최적화하는 과정이 없다. 따라서 이러한 목적 함수의 역할을 담당하는 확률적 분포를 이용하여 목적 함수를 구축하고 최적화 해가는 작업이 필요하다. 셋째로 SOM의 학습 알고리즘이 항상 수렴한다는 일반적인 보장이 없다. 특히 학습이 끝난 최종 모형이 국지적 최적값(local optimal value)로 빠지는 경우가 종종 있게 된다. 따라서 학습 데이터를 이용한 모형이 허용 한계 내에서 전역적 최적값(global optimal value)으로 수렴하게 하는 방안이 필요하다. 넷째로 SOM은 모형의 모수로 선택되어지는 최적값을 결정하는 이론적인 구조(theoretical framework)가 없다. 즉 SOM은 입력층으로부터 출력층으로의 연결을 통한 가중치의 갱신을 통한 모형이지만 새로운 입력에 대한 예측값을 계산할 때 예측값이 나오게 된 과정보다는 예측값 자체에 의미를 주기 때문에 예측 모형에 대한 설명이 절대적으로 부족하다. 예를 들어 학습률(learning rate)에 대한 초기값의 결정과 변화규칙, 이웃 함수(neighborhood function)의 감소율 등에 대한 함수, 가중치의 갱신 분포 등에 대한 이론적 구조가 전혀 없다. 이렇게 이론적 바탕이 부족한 신경망 모형에서 예측 결과값에 대한 모형적 해석을 가능하게 하는 방안이 필요하다. 다섯번째로 SOM은 다른 SOM 모형이나 이질적인 구조(architectures)의 학습 모형들과의 비교가 불가능하다. 이는 모형에 대한 명확한 수리적 규칙이 정의되지 않기 때문이다. 따라서 일반적인 모형 평가의 수리적 측도를 SOM 모형에도 적용시켜야 한다. 이러한 SOM의 해결할 수 없는 모형적 한계를 극복하는 방안이 하이브리드 SOM(Hybrid SOM)이다. 본 논문에서 제안하는 하이브리드 SOM은 SOM의 빠른 학습 능력을 유지하면서 군집 결과에 대한 수렴성을 보장하는 모형이다.

3. 하이브리드 SOM

학습 데이터의 군집화에 있어서 SOM은 신경망이 스스로 학습하여 최적의 군집을 형성하게 되며 이러한 경우 데이터의 특성을 잘못 이해하여 그릇된 결론에 도달할 위험이 있다. 또한, SOM에 의해서 최종적으로 구축된 모형은 한 개의 가중치 값으로 고정된 모수를 갖는 모형이 된다. 이 모형은 일반적으로 가장 좋은 모형이라는 보장은 없다. 최종 구축된 모형에 대한 최적값의 보장을 위해서 한 개의 모형으로 고정 시키지 않고 여러 가능성을 포함하는 모수에 대한 분포의 개념이 도입되어야 한다[1].

하이브리드 SOM 알고리즘은 SOM의 단점인 모형에 대한 설명 부족을 베이저안 추론 기법으로 해결하였고, 사전에 군집수를 결정해야 하는 계층적 군집화의 문제점에 대해서는 사전에 군집의 개수를 정하지 않는 SOM의 유연성을 따른다. 그리고 SOM에 의해 최종적으로 구축된 모형에서 출력층의 각 노드가 한 개의 가중치 값들로 고정되는데 비해 하이브리드 SOM은 고정된 가중치 값이 아닌 가중치가 속하게 되는 분포가 결정된다. 하이브리드 SOM은 가중치가 속하는 확률분포(probability distribution)의 모수가 학습 데이터에 의해 갱신된다. 이러한 갱신은 동일한 입력값에 대해서 항상 동일한 결과값만을 계산하지 않는다. 따라서 국지적 최적값에 빠진 경우에도 얼마든지 전역 최적값으로 빠져 나올 수 있다. 모형의 가중치 갱신은 주어진 학습 데이터를 학습하기 이전의 모형에 대한 믿음인 사전 확률 분포가 학습 데이터에 의한 우도 함수와의 결합을 통하여 사후 확률 분포가 되고 이 분포는 한 개의 학습 데이터에 의해 학습이 끝난 새로운 모형이 된다. 이러한 확률적 분포 갱신 학습이 SOM의 신경망 내에서 이루어진다.

(알고리즘 1)은 하이브리드 SOM의 초기화 단계를 나타내고 (알고리즘 2)는 실행단계를 나타낸다.

```

Algorithm : Initialize_HSOM (Input[i])

// HSOM의 초기화1: 네트워크 상수의 초기화
Initialize network parameter ;
    Input_layer(int i, int o, int init_neigh_size)
    num_inputs = i, num_outputs = o
    // 입력 벡터와 출력 벡터의 초기화
    neighborhood_size = init_neigh_size
    // 형상 지도의 이웃반경의 초기화
    weights = new float[num_inputs*num_outputs]
    // 가중치 지정
    outputs = new float[num_outputs]
    // 출력값 지정

// HSOM의 초기화 : 확률 분포의 초기화
Initialize bayesian distributions
// 베이저안 확률 분포 결정
// 초기 가중치의 분포를 평균과 분산이 각각 0과 1인 가우시안 분포로 결정
Initialization of the weight vector,  $w_j(0)$  to have probabilistic distribution,  $N(0, 1)$ .
// 학습율 함수의 결정
Initialization of the learning rate
 $\alpha(0), \alpha(t) \propto t^{-\alpha}; 0 < \alpha < 1$ .
// 이웃 반경 함수의 결정
Initialization of the neighborhood function  $K(j, j^*)$ ,
     $K$  decreases as to increase  $|j - j^*|$ 
    
```

(알고리즘 1) 하이브리드 SOM의 초기화 단계

(알고리즘 2)은 크게 3개의 알고리즘으로 이루어져 있다. 첫째는 초기의 가우시안 분포에서 각 형상 지도의 노드에 사용할 가중치들을 생성하게 된다. 초기의 사전분포는 평균이 0이고 분산이 1인 가우시안 분포를 사용한다. 이는 하이

브리드 SOM의 형상 지도에서 승리 노드의 결정에 있어서 유클리디안 거리를 사용하고, 또 입력 벡터의 값들이 평균이 0 분산이 1을 따르도록 초기화 되었기 때문이다. 따라서 처음부터 가중치의 갱신은 자신들이 속한 분포의 모수에 대한 갱신이 된다. 이 값들이 계속되는 학습에 의해 갱신되는 것이다. 둘째로 하이브리드 SOM의 승리 노드에 대한 결정이다. 이 단계에서는 주어진 반복 학습의 최대 범위를 지정하고, 최종 가중치 값이 갱신될 형상 지도의 노드를 결정하게 된다. 마지막으로 가중치의 분포가 갱신된다. 이때에는 가중치 분포인 가우시안 분포의 모수에 대한 하이퍼 파라미터에 대한 분포로부터 갱신이 이루어진다. (알고리즘 2)에서는 하이퍼 파라미터에 대한 분포로서 감마 분포를 사용하였다. 이는 베이저안의 사후 확률이 공액 분포가 되어 계산이 용이하게 되기 때문이다.

```

Algorithm : Train_하이브리드 SOM (Train[j])

// 하이브리드 SOM의 베이저안 확률 분포의 갱신
Determine the winner node // 승리 노드의 결정
// 가우시안 분포를 따르는 입력 벡터의 정규화
Normalization of input vector, Gaussian distribution with mean 0, variance 1
// 가중치의 hyper = parameter 결정
Choose the distribution of weights,  $w \sim f(\theta)$ 
// 최소 유클리디안 거리를 갖는 노드를 승자로 결정
Choose the winner node  $j^* = \arg \max y$ , using Euclidean criteria.

Update of parameters // 가중치 분포의 모수 갱신
 $w_j^{New} = w_j^{Old} + \alpha(j)K(j, j^*)(X - w_j^{Old})$ ,
    where,  $K(j, j^*)$ ; Neighborhood function
end

// 가중치 분포의 모수 갱신 작업의 반복
Replace old distribution by current.

// 하이브리드 SOM의 승리 노드 결정
Set Winner_index = 0,
    maxval = -1000000 // 최대 반복 회수 지정

Find the winner neuron // 승리 노드 결정

for (j = 0 ; j <= num_outputs ; j++)
    for (i = 0 ; i < num_inputs ; i++)
        // 최소 유클리디안 거리를 갖는 노드가 승리 노드가 됨
         $winner[i, j] = \operatorname{argmin} \|x(k) - w_j\|$ 
    end

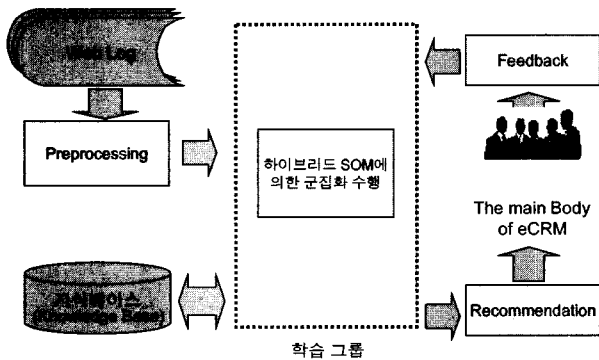
// 가중치 분포의 갱신
Set m = (int)alpha, delta = alpha-m
// 가우시안 분포로부터 모수 생성
while(count < data_size)
    v = rand()/32767.0, w = pow(y/m,delta)/(1 + (y/m-1) * delta)
end

// 수렴될 때까지 반복 학습
Repeat Until given criteria satisfaction.
    
```

(알고리즘 2) 하이브리드 SOM의 실행단계

#### 4. 하이브리드 SOM을 이용한 추천 시스템 설계

하이브리드 SOM을 이용한 동적 웹 정보 추천 시스템은 웹 서버에 접속하는 사용자에게 가장 필요한 정보를 예측하여 실시간으로 해당 정보를 추천해 주는 시스템으로 전체 구조는 (그림 2)와 같다.



(그림 2) 동적 웹 정보 추천 시스템의 구조

먼저 웹 서버로부터의 웹 로그 데이터를 정제하는 모듈이 학습 모듈과 직접적으로 연결되어 있다. 이 학습 모듈에서 클릭 스트림 데이터의 희소성은 제거된다. 지식베이스 모듈은 학습 모듈과 서로 정보를 주고 받게 된다. 전처리 모듈을 거친 학습 데이터에 대하여 하이브리드 SOM을 적용하여 최종적으로 구축된 군집 결과를 지식베이스에 저장하게 된다. 다음으로 처음 접속하는 새로운 사용자에게 대하여 웹 추천 정보는 지식베이스가 갖고 있는 할당 군집의 대표 선호도를 이용하게 된다. 그리고 실제 사용자의 웹 접속 시간을 통한 갱신 정보가 다시 지식베이스의 데이터를 갱신하게 된다. 이렇게 지식베이스 모듈과 학습 모듈은 서로 유기적으로 순환하는 구조를 갖게 된다. 새로운 사용자에게 대한 추천 모듈은 학습 모듈로부터 일방적인 정보를 받게 된다. 즉 학습 모듈로부터의 예측 정보를 통하여 새로운 사용자에게 대한 웹 페이지 추천을 하게 된다. 사용자에게 대한 추천 결과는 피드백 모듈을 통하여 다시 학습 모듈로 정보가 들어오게 된다. 이렇게 5개의 모듈이 서로 유기적으로 결합하는 형태로 작동하게 된다.

#### 5. 실험 및 평가

본 논문에서는 웹 로그 데이터로부터 제안 시스템의 성능을 평가하기 위해 Pentium-IV 2GHz CPU와 1G RAM, Windows XP 환경에서 마이크로 소프트 Visual C++ 6.0으로 구현하였다. 본 논문의 실험에서 사용한 데이터는 2000년도 KDD Cup 대회에 참가한 참가자들에게 문제로 주어졌던 웹

로그의 클릭 스트림 데이터로써 인터넷 쇼핑몰 Gazelle.com의 2개월간의 클릭 스트림 만을 모아 놓은 1.2GB의 텍스트 데이터이다[12].

<표 2>은 형상 지도의 차원을 3×3으로 결정하고 SOM과 하이브리드 SOM의 군집 결과를 비교한 결과이다.

<표 2> SOM과 하이브리드 SOM의 수렴성 비교

군 집 수	SOM	하이브리드 SOM
3	11	7
4	29	21
5	28	43
6	19	18
7	8	9
8	5	2

<표 2>은 총 100번의 군집화를 수행하는 동안 군집화의 결과로서 얻어진 최종 군집의 수를 나타낸다. SOM의 군집 결과를 살펴보면 최종 군집의 개수는 4개, 5개, 그리고 6개인 결과가 우열을 가릴 수 없게 나오고 있다. 그러나 하이브리드 SOM에서는 군집수가 5개인 것이 43번이나 나왔다. 이는 군집화 기법인 하이브리드 SOM의 경우 군집이 5개에 수렴하고 있음을 알 수 있다. 따라서 SOM의 비수렴성 문제는 하이브리드 SOM에 의해 해결할 수 있음을 알 수 있다. 5개의 군집수로 최종 하이브리드 SOM 군집을 수행한 결과, 군집에 해당되는 개체의 수와 추천 페이지에 대한 우선 순위는 <표 3>과 같다.

<표 3> 하이브리드 SOM 최종 군집 결과

(단위 : 초)

군집	개체수	1순위	2순위	3순위	4순위	5순위
Group1	2,423	page 37 (25.1)	page 46 (21.3)	page 8 (18.8)	page 52 (18.3)	page 42 (18.1)
Group2	1,989	page 201 (30.1)	page 123 (26.9)	page 169 (24.1)	page 137 (14.1)	page 144 (9.9)
Group3	1,903	page 83 (19.6)	page 102 (13.8)	page 116 (13.1)	page 78 (10.2)	page 126 (7.1)
Group4	1,471	page 219 (24.3)	page 194 (21.9)	page 83 (17.9)	page 261 (16.4)	page 137 (14.1)
Group5	866	page 19 (28.3)	page 63 (27.8)	page 23 (23.6)	page 42 (21.4)	page 62 (20.6)

<표 3>는 학습 데이터를 사용한 8,652개 ID의 사용자들 군집화한 결과를 나타낸다. 이는 각 군집에 최종적으로 할당된 사용자 수인 개체수와 상위 1순위부터 5순위까지의 웹 페이지 그리고 이 페이지에 대한 접속 가능 시간의 평균값을 나타내고 있다. 예를 들어 그룹 1에 속한 사용자는 모두 2,423명이고, 이 집단에서 가장 우선 순위가 높은 웹 페이지

는 페이지 37이 된다. 그리고 이 집단의 사용자들이 이 페이지에 머문 평균 시간은 25.1초였다. 따라서 새로운 사용자가 만약 군집 1에 속하게 된다면 이 사용자에게는 웹 페이지 37을 우선적으로 추천하게 된다. 다음의 추천 순위로는 페이지 46, 페이지 8, 페이지 52, 그리고 페이지 42의 순이 된다.

<표 4>은 ID가 8,653인 새로운 사용자가 접속을 시도하는 경우에 있어 기존의 5개의 군집과의 유사도를 계산한 경우를 나타낸다. 이때 유사도 계산은 각 웹 페이지에 대한 접속 시간을 이용한 유클리디안 거리를 이용한다.

<표 4> 새로운 사용자와 군집간의 유사도 계산 결과

(단위 : 초)

군 집	Group 1	Group 2	Group 3	Group 4	Group 5
거 리	2.49	3.76	1.42	4.33	2.85

이 결과에 의하면 ID가 8,653인 사용자는 가장 작은 유클리디안 거리가 계산된 군집 3에 할당된다. 이 새로운 사용자에 대한 최우선 웹 페이지의 추천은 <표 3>의 결과를 이용하여 군집 3의 가장 상위의 대표 선호도를 갖는 웹 페이지인 페이지 83을 추천하게 된다. 그리고 이때 이 사용자가 여기에 머문 시간은 19.6초로 예측할 수 있다.

실제로 ID가 8,653인 이 사용자는 페이지 83을 21.3초 동안 접속하였다. 하이브리드 SOM에 의해 추천한 값과 실제 값의 차이는 |21.3초 - 19.6초|가 된다. 이와 같은 방법으로 테스트 데이터에 대해 하이브리드 SOM에 의해 추천한 MSE (mean square error)와 기존 방법에 의한 MSE의 비교 결과는 <표 5>와 같다.

<표 5> 최종 성능 평가 결과

(단위 : 초)

구 분	피어슨 기반의 추천 시스템	하이브리드 SOM
MSE	1.63	0.94

<표 5>의 최종 성능 평가의 비교 결과를 보면 제안 시스템의 MSE 값이 비교 시스템인 피어슨의 추천 시스템에 비해 훨씬 작음을 알 수 있다. 즉 하이브리드 SOM을 이용한 추천 시스템이 기존의 추천 시스템에 비해 더 정확한 예측 결과를 보임을 알 수 있다.

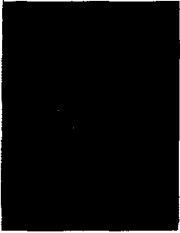
## 6. 결 론

본 논문에서 제안한 하이브리드 SOM을 이용한 동적 웹 정보 추천 시스템은 사용자별로 적합한 웹 정보를 추천하는 일련의 과정이 실시간으로 이루어질 수 있도록 설계되었으며, 이렇게 설계된 시스템은 기존의 추천 시스템보다 정확

도, 예측을 위한 학습 시간, 그리고 모형의 수렴성면에서 우수함을 실험을 통하여 관찰할 수 있었다. 그러나 예측의 정확성 문제와 최적 군집수 결정 문제, 그리고 모형의 수렴 조건의 주관적 결정에 대한 문제점이 여전히 내포되고 있으며, 따라서 향후에는 이러한 제약조건을 해결하기 위해 서로 동질적인 데이터 마이닝 기법의 결합을 통하여 다양한 측면에서의 희소성 데이터를 분석하고 추천할 수 있는 새로운 알고리즘의 개발이 필요하며, 클릭 스트림 이외의 다양한 웹 로그 데이터를 이용하여 본 논문에서 제안하는 시스템을 적용함으로써 동적 웹 추천 기법의 기능을 지속적으로 확장할 수 있도록 연구를 수행해 나가야 할 것이다.

## 참 고 문 헌

- [1] 윤경배, 최준혁, 왕창중, "하이브리드 SOM을 이용한 효율적인 지식베이스 관리", 정보처리학회논문지B, 제9-B권 제5호, pp.46-53, 2002.
- [2] 이준규, "인터넷 개인화 아이템 추천 알고리즘에 대한 연구", 연세대학교 전자계산공학과 석사논문, 2001.
- [3] 임영희, "후처리 웹 문서 클러스터링 알고리즘", 정보처리학회논문지B, 제9-B권 제1호, pp.7-16, 2002.
- [4] Basu, C et al., "Recommendation as classification : Using Social and Content-based Information in Recommendation," *Proceedings of the Workshop on Recommendation system, AAAI Press, Menlo Park California*, pp.1028-1031, 1988.
- [5] Bishop, C. M., Svensen, M., Williams, C. K. I., "GTM : A Principled Alternative to the Self Organizing Map," *Proceedings 1996 International Conference on Artificial Neural Networks, ICANN 96, Vol.II*, pp.165-170, Bochum, Germany, 1996.
- [6] Chun-Che Huang, "Using Intelligent Agents to Management Fuzzy Business Processes," *IEEE Transactions on System, Man, and Cybernetics-Part A : System and Humans*, Vol.31, No.6, pp.1045-1049, 2001.
- [7] Forrester Research, "http://www.educorner.com/courses/ia/", 2002.
- [8] Frank, "Data Mining," Morgan Kaufmann Publisher, 2000.
- [9] Teuvo Kohonen, "Self Organizing Maps," Springer, 1997.
- [10] V. Vapnik et al., "Support vector networks," *Machine Learning 20*, pp.273-297, 1995.
- [11] Robert Cooley, Bamshad Mobasher, and Jaideep Srivastava, "Web Mining : Information and Pattern Discovery on the World Wide Web," In *Proceedings of the 9th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'97)*, pp.1-3, November, 1997.
- [12] "http://www.ecn.purdue.edu/KDDCUP/", 2003.



### 윤 경 배

e-mail : kbyoon@kimpo.ac.kr  
1986년 인하대학교 수학과(학사)  
1994년 인하대학교 대학원 정보공학과  
(공학석사)  
1998년 서강대학교 경제대학원 정보기술  
경제학(경제학석사)

2003년 인하대학교 대학원 전자계산공학과(공학박사)

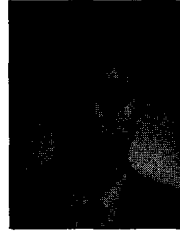
1986년~1987년 대우자동차(주) MIS 근무

1988년~1991년 LG-EDS(주) 기술연구소 근무

1992년~1997년 동부정보기술(주) 정보기술연구소 근무

1998년~현재 김포대학 컴퓨터계열 교수

관심분야 : 지식기반 데이터베이스, 데이터마이닝, CRM, 지문  
및 음성 인식, 인공지능 등



### 박 창 희

e-mail : chpark92@yonsei.ac.kr  
2000년 한양대학교 전자과(학사)  
2003년~현재 연세대학교 공학대학원 전자  
공학 석사과정  
관심분야 : 생체(지문, 홍채, 음성)인식, 데이  
터마이닝, 인공지능, 영상처리 등