

# 지능형 정보시스템을 위한 개인성 학습 기법

김호준<sup>0</sup>, 박정선  
한동대학교 전산전자공학부  
hjkim@han.ac.kr

## Personality Learning Techniques for Intelligent Information System

Ho-Joon Kim<sup>0</sup>, Jung Sun Park  
School of Computer Science and Electronic Engineering, Handong University

### 요 약

본 연구에서는 정보시스템의 지능형 인터페이스를 위하여 사용자의 개인성을 학습하는 방법론으로서 신경망 이론의 활용가능성을 고찰한다. 입력형식의 유연성, 입력의 왜곡 및 소실가능성 등 시스템의 실용성과 연관하여 나타나는 자료의 특성을 수용하기 위하여, 학습과정에서 신호표현의 다양화와 부분 패턴의 의한 분류기능 등을 개선한 신경망모델을 제안한다. 이를 위하여 퍼지 양방향 연상기억장치와 구간연산으로 일반화된 다층 신경망모델을 결합하여 혼합형 분류모형을 제시하고 그 유용성을 고찰한다. 실험은 전공분야 선택을 위한 개인의 적성분석시스템을 대상으로 구현하였다.

### 1. 서 론

정보시스템에서 고객 개개인에 대한 취향과 관심도를 얼마나 효과적으로 반영하는가에 따라 고객이 느끼는 서비스의 편이성과 만족도가 달라질 수 있다. 이는 또한 일반 상품 정보시스템에도 해당될 수 있는데 예로서 서적, 음악, 영화 및 연극 등 문화상품이나 기호상품에 대한 만족도는 일반적인 상품과는 달리 개인적인 취향이나 선호 경향에 따라 상대적인 만족도의 차이를 보인다. 그럼에도 불구하고 이러한 정보를 제공하는 방법이 불특정 다수를 대상으로 하는 획일적인 방법에만 의존한다면 제공정보에 대한 고객의 만족도를 제고하기가 어렵다. 이에 본 논문에서는 고객의 개인적인 취향이나 관심에 따라 그 선호도가 민감하게 달라질 수 있는 분야에 대하여 개별 고객에게 가장 적합한 정보 및 상품을 능동적으로 제시하는 방법론으로서 개인성 학습기법을 연구하였다.

본 연구에서는 이러한 응용시스템이 갖는 특성중에 시스템의 성능과 실용성을 저해하는 다음 두가지의 요인을 중점적으로 고려한다.

- **완전한 자료획득의 어려움** : 시스템에 대한 사용자의 인식도와, 응답에 대한 적극성의 차이로 인하여 일부 입력항목에 대하여 자료의 누락 및 왜곡이 있을 수 있다.
- **자료표현의 단순화로 인한 정보 왜곡** : 분류에 사용되는 학습데이터의 형태는 이진값이나 모호성을 가지는 연속값 외에도 시계열표현, 구간형식의 표현 등 다양한 형태로 나타나는데 이들을 단순화 시킴으로써 성능저하가 발생할 수 있다.

이에 본 연구에서는 데이터의 표현형태와 연산기능을 확장함으로써 이러한 단점을 개선한다. 다시 말해서 기존의 단순 이진값으로 결정되는 데이터 표현 및 연산방식을 개선하여, 미지 데이터(unknown data)를 포함하여 처리될 수 있도록 하고, 단순 이진값이 아닌 모호한 응답이나, 구간 범위 형식의 데이터에 대한 처리가 지원되도록 함으로써 시스템의 표현능력, 연산능력 및 신뢰도의 개선을 시도하였다.

본 연구의 의의는 크게 두 가지로 대별된다. 첫번째는 양방향 연상기억장치(BAM: Bidirectional Associative Memory)[1-3]와 결합한 모듈구조의 신경망 학습모형이다. BAM의 연상기능은 입력 패턴에 포함될 수 있는 데이터 값의 왜곡을 보상하거나 미지 데이터(unknown data)를 추정할 수 있게 하며 상호 연상기능을 통하여 부분 패턴만으로도 원하는 정보를 분류해 낼 수 있게 한다.

두 번째로는 이러한 모델에 대하여 이진값 또는 모호한 실수 표현 및 구간 표현형식을 수용하는 네트워크의 활성화 특성 및 학습 방법을 제시한다. 이에 기존의 BAM[1]과 퍼지연상기억장치(FAM)[3]를 이용하는 방법론을 제시하며, 이어서 분류를 위한 신경망에서 추세분석 기능 등을 포함한 구간 데이터 연산기능 및 학습 방법을 제시한다. 그리고 전형적인 패턴 분류 신경망인 다층 퍼셉트론[4]과 추세분석 기능을 위한 FIR 네트워크[5]에 대하여 구간연산으로 일반화한 동작특성 및 학습방법을 보이고 그 유용성을 고찰한다. 실험으로는 실제 응용 시스템으로서 개인의 적성 분석시스템에 적용한 결과를 보인다.

2. 시스템 모형

본 논문에서는 그림.1과 같은 구조의 학습모형을 고려한다. 이는 그림에 보인 바와 같이 양방향 연상기억장치(BAM)와 분류를 위한 모듈구조의 다층 신경망이 상호 결합된 형태이다. 획득된 개인정보로부터 추출된 일련의 특징집합은 학습자료로서 BAM의 양 계층의 노드값을 초기화한다. 이후 BAM의 동작은 입력의 부실로 인한 누락 특징이 상호 연상기능에 의해 그 추정치가 생성되도록 하며 이들은 분류신경망의 입력이 되어 대상 정보항목에 대한 적합도를 생성하게 된다. 이러한 모델에서 각 네트워크 구조는 특징표현의 종류에 따라 서로다른 모델을 모듈구조로서 구성하고 분류결과를 최종적으로 결합하는 과정이 이루어진다. 다시 말해서 입력의 특징형태로서 단순 이진값 및 모호한 표현의 연속적인 값과 더불어 구간형태의 표현을 수용하는 신경망을 각각 별도로 고려하게 된다. 또한 이 시스템이 시간에 따라 동적으로 변하는 성향을 반영하기 위하여 시계열(time series) 예측 신경망을 적용하여 구현하였는데 이는 구간연산으로 일반화된 FIR 네트워크 모델을 사용하였다.

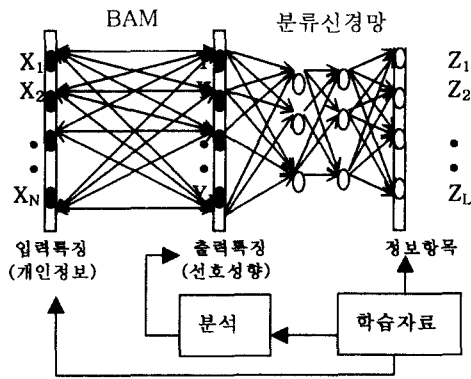


그림.1 신경망 기반 개인성 학습 모형

3. 시스템 특성 분석

본 논문의 핵심내용은 앞 절에서 제시한 학습모형에서 누락특징의 추정과정과 정보 분류과정에서 보다 일반화된 데이터 표현 능력과 그에 대한 연산능력을 제공함에 있다. 구체적으로 단순 이진값 데이터와 모호함을 반영하는 연속적인 실수값의 데이터는 물론, 구간(interval)형식의 데이터에 대한 연산기능 및 그에 대한 신경망의 학습방법을 정의한다.

(1) 소실 및 누락 정보의 복원

앞 절에서 언급한 바와 같이 실제 환경에서 시스템 사용자가 소정의 입력내용을 충실하게 응답하지 않을 가능성을 고려하여야 한다. 이의 영향은 입력특징의 누락이나 왜곡의 형태로 나타나게 되는데 이들에 대한 추정치를 BAM의 특성을 통하여 생성하게 된다. 이진값의 데이터는 전형적인 BAM의 특성으로부터 지원되며, 연속적인 실수형태의 연상기능은 FAM(Fuzzy Associative Memory)[3] 모

델을 기반으로하여 구현하였다. BAM에서 양 계층의 각 노드값을 X와 Y로 표현할 때 이들의 연상과정은 두 패턴집합간 상관관계 행렬 M과의 내적 연산을 통해 이루어진다. 이때 M과 이의 전치행렬 M<sup>T</sup>는 각각 식(1)과(2)로서 산출된다.

$$M = \sum_i X_i^T Y_i \quad (1)$$

$$M^T = \sum_i (X_i^T Y_i)^T = \sum_i Y_i^T X_i \quad (2)$$

본 논문에서 채택한 FAM에서 패턴 연상을 위한 상관관계 행렬의 연산은 퍼지 연산에 의해 이루어진다. 즉 Y = X•M 연산에서 Y의 j번째 성분은 식(3)에 의해 산출된다.

$$y_j = \max_i \{ \min(x_i, m_{ij}) \} \quad (3)$$

(2) 분류 신경망

시스템에서 분류신경망과 관련한 본 논문의 의의는 네트워크의 활성화 특성과 학습과정을 구간표현의 데이터를 수용하도록 일반화 하였다는 점이다. 우선 BAM의 최종 출력으로부터 주어지는 입력에 대하여 정보분류 기능은 다층퍼셉트론(MLP)을 대상으로 하는데, 여러 회의 자료획득을 통하여 얻어진 정보에서 특징의 편차범위를 구간으로 표시하여 그대로 분류과정에 적용할 수 있도록 하였다. 이는 전형적인 MLP의 특성을 구간연산기능으로 일반화함으로써 구현될 수 있다. 식(4), (5) 및 (6)은 이를 위한 신경망의 활성화 특성이다. 여기서 첨자 U와 L은 구간표현의 상한과 하한을 의미한다.

$$x_j(k) = f(Net_j(k)) = f([Net_j(k)^U, Net_j(k)^L]) \quad (4)$$

$$Net_j^U = \sum_{i, w_{ji} \geq 0} w_{ji} x_i^U + \sum_{i, w_{ji} < 0} w_{ji} x_i^L \quad (5)$$

$$Net_j^L = \sum_{i, w_{ji} \geq 0} w_{ji} x_i^L + \sum_{i, w_{ji} < 0} w_{ji} x_i^U \quad (6)$$

정보시스템에서 어떤 사용자군의 성향 분석 등과 같이 시간에 따라 동적으로 변화하는 문제는 정적인 일반 분류기법과는 달리 일종의 시계열(time series) 예측문제로 고려될 수 있다. 이에 본 논문에서는 시계열 예측 신경망으로서 FIR 네트워크구조를 기반으로 하여 구간표현의 자료를 수용하는 연산기능을 정의한다. 다시 말해서 각 계층간 임의 노드의 연결은 하위 노드의 활성화 값이 일정횟수의 주기로 이동(shift)되어 상위노드에 입력되는 형태로서 노드 활성화의 동적 특성을 반영한다. 즉 임의 노드에 입력되는 하위계층으로부터의 입력함은 식(7)과 식(8)로 정의된다.

$$Net_j(k)^U = \sum_{t=0, w_{jt} \geq 0}^T w_{jt} \cdot x_j(k-t)^U + \sum_{t=0, w_{jt} < 0}^T w_{jt} \cdot x_j(k-t)^L \quad (7)$$

$$Net_j(k)^L = \sum_{t=0, w_{jt} \geq 0}^T w_{jt} \cdot x_j(k-t)^L + \sum_{t=0, w_{jt} < 0}^T w_{jt} \cdot x_j(k-t)^U \quad (8)$$

이러한 활성화 특성을 기반으로 구간연산을 고려한 학습 알고리즘이 정의된다. 이는 일종의 수정된 역전파(BP) 알고리즘으로 다음과 같은 수식으로 표현된다. 즉 시점 k에서의 총 에러합을 e<sup>2</sup>(k)라 할 때, 노드 i

와  $j$  간의  $n$ 번째 가중치에 대한 변화분은 식(9)로서 산출된다.

$$\Delta w_{ij}(n) = -\eta \frac{\partial e^2}{\partial w_{ij}(n)} \quad (9)$$

$$\frac{\partial e^2}{\partial w_{ij}(n)} = \begin{cases} \delta_j(k)^t \cdot x_i(k-n)^t + \delta_j(k)^{tt} \cdot x_i(k-n)^{tt}, & \text{if } w_{ij}(n) \geq 0 \\ \delta_j(k)^t \cdot x_i(k-n)^t + \delta_j(k)^{tt} \cdot x_i(k-n)^{tt}, & \text{if } w_{ij}(n) < 0 \end{cases} \quad (10)$$

여기서 상위 노드  $j$  가 출력노드인 경우  $\delta_j(k)$  의 상한값과 하한값은 식 (11)과 (12)로 정해지며, 노드  $j$  가 중간노드인 경우 식 (13)과 (14)로 정해진다.

$$\delta_j(k)^t = (d_j(k)^t - x_j(k)^t) \cdot f'(Net_j(k)^t) \quad (11)$$

$$\delta_j(k)^{tt} = (d_j(k)^{tt} - x_j(k)^{tt}) \cdot f'(Net_j(k)^{tt}) \quad (12)$$

$$\delta_j(k)^t = \left( \sum_{m, w_{jm} > 0} \delta_m(k)^t \cdot w_{jm}(k) + \sum_{m, w_{jm} < 0} \delta_m(k)^{tt} \cdot w_{jm}(k) \right) \cdot f'(Net_j(k)^t) \quad (13)$$

$$\delta_j(k)^{tt} = \left( \sum_{m, w_{jm} > 0} \delta_m(k)^{tt} \cdot w_{jm}(k) + \sum_{m, w_{jm} < 0} \delta_m(k)^t \cdot w_{jm}(k) \right) \cdot f'(Net_j(k)^{tt}) \quad (14)$$

4. 구현 및 고찰

본 연구에서는 제안된 이론에 대하여 개인의 전공분야 선택을 위한 적성 분석 시스템을 대상으로 실험하였다. 이는 사용자의 설문으로부터 추출되는 입력특징으로부터 선택된 범위의 대상 전공들에 대한 각각의 적합도를 산출하는 방식으로 이루어진다. 누락된 특징의 생성을 위하여 FAM구조의 연상기억모델을 사용하였으며, 총 4가지 유형의 120개 문항을 적용하였다. 학습패턴의 생성을 위하여 총 40명의 학생 및 교수로부터 반복 획득한 학습데이터를 사용하였다.

그림.2는 시스템의 인터페이스와 실험결과와 한 예를 보인 것이다. 기존의 많은 적성검사 방식이 단순 이진값으로 응답하는 것과는 달리 응답자의 주관에 따른 모호한 실수값을 백분위로서 선택할 수 있도록 하였고, 이들 입력은 FAM 신경망의 초기 입력으로 설정된다.

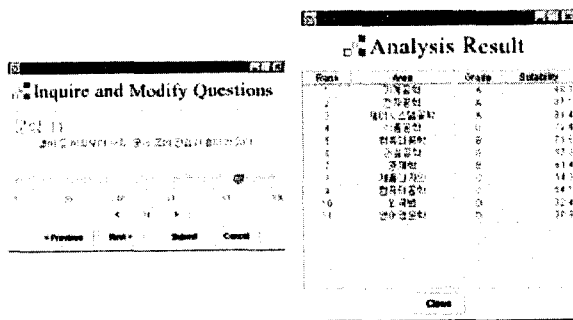


그림. 2 시스템 인터페이스 및 실험 결과의 예

대상전공은 총 15개 분야에 대해 적합도를 평가하도록 하였고, 사용자는 초기설정시 이들에 대해 선택적으로 적용할 수 있도록 하였다.

구간연산 신경망은 기존의 단일값을 취하는 신경망에 비

해 보다 일반화된 연산능력을 갖는다. 왜냐하면 만일 입력 특징에 단일값의 형태를 적용하기를 원한다면 구간표현에서 상한과 하한값을 동일값으로 일치시키면 된다. 이 경우에도 활성화 특성 및 학습 알고리즘은 동일한 연산을 적용할 수 있다.

이러한 방법의 또 다른 이점은 측정치의 왜곡이나 모호함을 보상하는 분류를 가능하게 한다는 점이다. 다시 말해서, 일부 문항에 대해서는 환경적 상황이나 사용자의 일시적인 심리상태에 따라 응답치의 편차를 보일 수 있는데, 단 한번의 측정치에 의존하여 판단하는 것 보다, 반복 응답후 응답치의 범위를 구간으로 표현하여 적용함으로써 정보의 왜곡을 최소화 할 수 있다는 점이다.

5. 결론

본 논문에서는 보다 진보된 지능형 정보시스템의 구축을 위하여 사용자의 개인성을 학습하기 위한 방법론을 제안하였다. 이는 정보검색시스템뿐만 아니라 특정 상품에 관하여 사용자의 취향이나 선호도를 반영하여 능동적으로 정보를 제시할 수 있도록 하는 이른바 지능형 상품정보 시스템을 위한 기반기술로서 의미를 갖는다. 본 연구에서는 개인성의 학습과정에서 필요한 데이터 표현의 다양화와 이에 대한 처리기법을 포함하여 학습 모델 및 응용 시스템의 구축에 관한 내용을 고찰하였다. 특히 BAM을 이용한 누락데이터의 보완기법 및 부분 특징에 의한 분류기법을 소개하였으며, 분류과정에서 정보표현 능력의 확장을 위하여 구간연산 기능을 고려한 신경망 활성화 특성 및 학습방법을 고찰하였다. 향후의 연구 방향으로는 자료획득방법의 다양화를 통하여 학습자료의 충실도를 제고하는 연구와 모듈구조 신경망의 의사결정 융합 방법론의 개선을 통하여 분류성능의 신뢰도를 개선하는 연구가 이루어져야 하겠다.

\* 이 연구는 과학기술부 뇌과학 연구개발사업으로 수행되었음.

6. 참고 문헌

[1] Bart Kosko, "Bidirectional Associative Memories," IEEE Transaction on Systems, Man, and Cybernetics, Vol.18, No.1, pp.49-60, 1988.  
 [2] Xinhua Zhuang, Yan Huand, and Su Sing Chen, "Better Learning For Bidirectional Associative Memory," Neural Networks, Vol.6, pp.1131-1146, 1993.  
 [3] Bart Kosko, "Fuzzy Association Memories," Neural Networks and Fuzzy System, Prentice Hall, pp.298-337, 1992.  
 [4] James L. McClelland, David E. Rumelhart, "Learning Internal Representation by Erroe Propagation," Parallel Distributed Processing, Massachusetts Institute of Techology, 1986.  
 [5] Eric A. Wan, 1996. "Time Series Prediction by Using a Connectionist Network with Internal Delay Lines," Time Series Prediction, pp. 195-217, 1996.