

# 퍼지적분을 이용한 다중 신경망의 결합방법†

조성배<sup>○</sup>, 김진형

한국과학기술원 전산학과 및 인공지능연구소

## A Fusion Method of Multiple Neural Networks with Fuzzy Integral

Sung-Bae Cho and Jin H. Kim

Center for Artificial Intelligence Research and Computer Science Department, KAIST

### 요약

본 논문에서는 신경망을 효과적으로 학습시켜 좀더 신뢰도 높은 결과를 얻기 위하여 상호 독립적으로 학습된 신경망의 결과를 Sugeno의 퍼지적분 기법으로 결합하는 방법을 제시한다. 이론적으로는 주어진 문제를 해결할 수 있는 신경망은 항상 존재한다는 것이 증명되어 있지만, 현실적인 문제에서 단순한 구조의 신경망은 학습시키기 어려울 뿐만 아니라 학습된 신경망이 제대로 일반화하지 못하는 경향이 있다. 이를 해결하기 위해서 주어진 문제에서 가능한 특성중 특정부분만을 고려하여 결정을 내리도록 각 신경망을 학습시킨 후, 이들의 결과를 일종의 퍼지척도로 보고 퍼지적분에 의해 통합하고자 한다. 제시된 방법의 유용성을 입증하기 위하여 온라인 필기문자를 인식하는 문제에 적용한 결과, 개별적인 신경망은 물론이고 그 밖의 여러 결합 방법과 비교하여 보다 우수한 결과를 얻을 수 있었다.

## 1 서론

신경망을 실제문제에 적용하기 위해서는 주어진 학습 데이터에 대해서 적당한 네트워크 구조를 선택한 후 가능한 한 많은 데이터를 처리할 수 있도록 학습 알고리즘을 수행시키는 것이 보통이다. 이러한 방식을 따르면 네트워크의 크기에 제한이 없고 무한한 양의 데이터가 제공되는 경우에 주어진 문제를 해결할 수 있는 신경망을 항상 구축할 수 있다는 것이 이론적으로 증명되어 있다 [1]. 하지만 실제적으로는 사용할 수 있는 자원에 제한이 있기 때문에 충분한 기능을 하도록 신경망을 학습시키기란 그리 쉬운 일이 아니다.

최근에는 이와 같은 문제를 극복하기 위해서 다양한 유형의 모델화 신경망이 연구되고 있다 [2]. 그 중 일부 연구자들은 동일한 입력에 대하여 서로 다른 조건하에서 여러개의 신경망을 학습시킨 후 각 신경망의 결과를 적절하게 결합하여 보다 나은 최종결과를 내도록 하는 방법을 사용하기도 한다. Mori 등[3]은 간지문자를 인식하기 위해서 주어진 문제를 여러개의 부분제로 나누고 각각에 신경망을 할당한 후 이들을 결합하여 대규모 신경망을 구축하였다. 또,

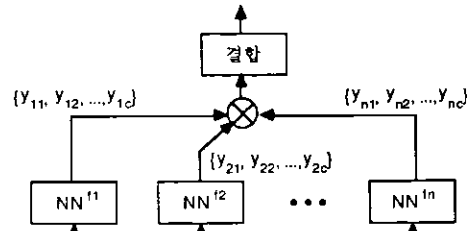
Jacobs 등[4]도 모델화된 다중 신경망을 제안하였는데, 이것은 여러개의 독자적인 네트워크가 강하게 연결된 시스템으로 각 네트워크가 전체 학습영역의 일부분을 처리하도록 학습된 것이다. 조성배 등 [5]은 이러한 다중 신경망을 체계적으로 연구하고자 시도하고 있다.

다중 신경망은 신경망의 학습이 많은 국소 최적점을 갖는 최적화 문제에서 연결강도  $w$ 를 선택하는 것이라고 볼 때, 각 신경망이 국소 최적점 하나에 매핑되는 것으로 그 의미를 해석할 수 있겠다. 여러개의 국소 최적점 중에서 전역적인 최적점을 찾는 방법들은 수행조건에 따라서 다른 "최적" 파라미터 ( $w$ )를 생성할 수 있는데, 이것은 초기 연결강도 ( $w^0$ )나 학습 데이터의 순서 등에서 기인하는 임의성 때문이다. 이러한 임의성은 각 네트워크의 오차를 다르게 하는 경향이 있어서 결과적으로 생성된 연결강도  $w$ 의 값이 달라지게 되는 것이다. 이와 같이 서로 다른 연결강도는 학습 데이터 내에 내재한 서로 다른 형태의 일반화 방식을 나타낸다고 볼 수 있다. 따라서 입력공간의 여러 부분집합들에 대해 다른 형태로 일반화한 각 신경망을 효과적으로 결합하면 보다 신뢰도 높은 결과를 얻을 수 있을 것이다.

본 논문에서는 퍼지적분 개념에 근본을 둔 새로운 다중 신경망 결합방법을 제안한다. 퍼지적분은 다양한 정보원에서 제공되는 대

† 본 논문은 한국과학재단의 목적기초 연구비 (90-01-00-01) 지원에 의한 것임.

상의 평가치는 물론이고 그 정보원의 예상된 중요도를 결합과정에서 동시에 고려한다. 즉, 이것은 주어진 가설에 대한 증거를 그 증거의 중요도에 대한 시스템의 예상치와 결합할 수 있도록 한다. 이러한 접근방법은 각 신경망에서 출력하는 인식값과 그 값을 낸 신경망의 주관적인 신뢰도를 효과적으로 결합시킬 수 있는 방법으로서 문제에 대한 엔지니어의 사전지식(a priori knowledge)을 결합시킬 수 있는 가능성을 제공한다.



(그림 2) 일반적인 다중 신경망의 구조

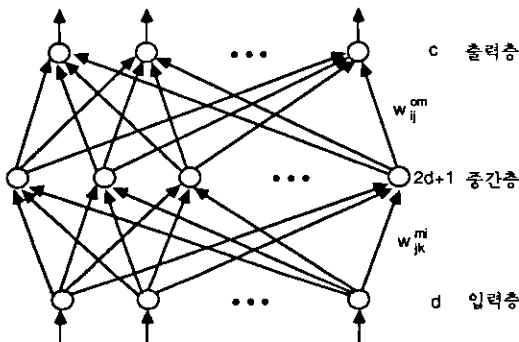
## 2 다중 신경망 분류기

신경망은 입력집합과 출력집합 사이를 매핑시키는 도구로서 생각할 수 있는데, 수학적으로 보면 이것은  $I$ 에서  $O$ 로 매핑하는 함수  $F$ 의 역할을 한다. 즉  $F: I \rightarrow O$ , 또는  $y = F(x)$ , 단  $y \in O, x \in I$ . 그런데, 분류문제란 결국 특징공간을 적절한 클래스 공간으로 매핑시키는 것이라 보면, 신경망을 일종의 분류기로 정형화할 수 있다.

(그림 1)은  $d$ 개의 입력노드와  $2d+1$ 개의 중간노드, 그리고  $c$ 개의 출력노드로 이루어진 이층구조의 신경망을 보여주고 있다. 여기에서  $d$ 는 특징의 수이고,  $c$ 는 클래스의 수이며,  $2d+1$ 은 적당히 선택된 수이다. (이 수는 Kolmogorov의 정리에 의해서 편의상 정해진 임의의 수이다.) 이 네트워크는 각 층간이 완전 연결된 형태를 하고 있는데, 이 네트워크는 비선형 의사결정식에 의해서 작동한다고 볼 수 있다. 즉, 주어진 입력  $X = \{x_1, x_2, \dots, x_d\}$ 와 클래스 집합  $\Omega = \{\omega_1, \omega_2, \dots, \omega_c\}$ 에 대해서 각각의 출력노드는 이 값이 자신의 클래스에 속할 가능성  $y_i$ 를 다음과 같은 비선형식에 의해서 추정한다.

$$y_i = f \left\{ \sum_{k=1}^{2d+1} w_{ik}^{om} f \left( \sum_{j=1}^d w_{kj}^{mi} x_j \right) \right\} \quad (1)$$

여기에서  $w_{kj}^{mi}$ 와  $w_{ik}^{om}$ 은 각각 입력층과 중간층 사이, 중간층과 출력층 사이의 연결강도이며  $f$ 는  $f(x) = 1/(1 + e^{-x})$ 와 같은 비선형 함수이다. 그리고 나서 가장 큰 값을 내는 노드를 해당 클래스로 선택한다. 여기에서 가장 중요한 부분은 연결강도의 값인  $w_{ik}^{om}$ 과  $w_{kj}^{mi}$ 를 어떻게 결정할 것인가이다.



(그림 1) 이층구조로 이루어진 다중 신경망

다중 신경망 구조의 기본 아이디어는 특정한 특징만을 고려하는  $N$ 개의 상호 독립적으로 학습된 신경망들을 구축하여 주어진 입력패턴에 대해서 각각의 신경망이 나름대로의 분류과정을 수행하도록 한 후, 여러가지 결합방법을 사용하여 총괄적인 분류결정을 내리는 방법을 이용하자는 것이다 [4, 6]. 여기에서 결국 핵심이 되는 것은 주어진 각 신경망의 결과를 어떻게 결합할 것인가 하는 문제이다.

다중 신경망 구조에서 각 노드의 출력값은 학습패턴이 그 클래스에 속할 사후확률(a posteriori probability)을 추정한 값이라고 볼 수 있다. 이때, (그림 2)에서 보여주는 것과 같이 각 신경망의 출력값의 집합을  $y = \{y_1, y_2, \dots, y_n\}$ 이라고 하고,  $y_1 = \{y_{11}, y_{12}, \dots, y_{1c}\}$ ,  $y_2 = \{y_{21}, y_{22}, \dots, y_{2c}\}$ , ...,  $y_n = \{y_{n1}, y_{n2}, \dots, y_{nc}\}$ 라고 하자. 물론 가장 중요한 것은 이러한 부분결과를 어떻게 결합하여 하나의 총체적인 결론을 도출할 것인가이다. 여기에서는 대표적인 방법으로 과반수 투표기법 (majority voting rule)과 Borda count에 대해서 간략히 알아본다.

과반수 투표기법은 네트워크들의 반수 이상이 내린 분류결정을 따르는 것이다. 만일 반수 이상이 일치하지 않는 경우가 발생한다면 그 결과는 오류로 간주한다. 이러한 방식의 성능을 알아보기 위하여 모든 신경망들이 오류를 낼 가능성이 독립적이며  $1-p$ 의 확률로 올바른 결론에 도달한다고 가정하자. 그러면  $N$ 개의 네트워크 중에서 정확하게  $k$ 개의 오류가 발생할 가능성은

$$\binom{N}{k} p^k (1-p)^{N-k} \quad (2)$$

이기 때문에 과반수 투표기법이 오류될 가능성은 다음과 같다.

$$\sum_{k > N/2}^N \binom{N}{k} p^k (1-p)^{N-k} \quad (3)$$

이 식은  $N$ 이 홀수인 경우 (또는 짝수인 경우),  $p < 1/2$ 이면  $N$ 이 증가함에 따라서 단조감소함을 귀납적으로 보일 수 있다. 다시 말해서 각 네트워크가 절반 이상 올바른 답을 낼 수 있고 그 답이 독립적이라면 좀더 많은 네트워크를 사용할수록 과반수 투표기법의 오류 가능성은 줄어든다.

Borda count는 특정한 클래스  $c$ 에 대해서  $c$ 보다 낮은 순위에 존재하는 클래스 수의 합이다. 즉,  $B_j(c)$ 는  $j$ 번째 네트워크에서 클래스  $c$ 보다 낮은 순위로 등급이 매겨진 클래스의 수라고 하면 클래스  $c$ 에 대한 Borda count는  $B(c) = \sum_{j=1}^n B_j(c)$ 이다. 최종 결정은 가장 큰 Borda count를 갖는 클래스를 선택함으로써 얻어진다.

### 3 퍼지적분을 이용한 신경망의 결합

퍼지적분은 Sugeno에 의해서 처음으로 개발된 것으로  $g_\lambda$ -퍼지적도에 의해서 정의되는 비선형 함수이다 [8, 9].  $Y = \{y_1, y_2, \dots, y_n\}$ 이 유한집합이고  $h: Y \rightarrow [0, 1]$ 이  $h(y_1) \geq h(y_2) \geq \dots \geq h(y_n)$ 을 만족하는 함수라 하면,  $Y$ 에 대한 퍼지적도  $g$ 의 퍼지적분 값  $e$ 는 다음과 같이 계산된다.

$$e = \max_{i=1}^n [\min(h(y_i), g(A_i))] \quad (4)$$

단,  $A_i = \{y_1, y_2, \dots, y_i\}$ .

여기에서  $g(A_i)$ 의 값은 다음과 같이 재귀적으로 계산된다.

$$g(A_1) = g(\{y_1\}) = g^1 \quad (5)$$

$$g(A_i) = g^i + g(A_{i-1}) + \lambda g^i g(A_{i-1}), \quad 1 < i \leq n. \quad (6)$$

단,  $\lambda$ 는 다음과 같은 식을 풀어서 얻을 수 있다.

$$\lambda + 1 = \prod_{i=1}^n (1 + \lambda g^i), \quad \lambda \in (-1, +\infty), \quad \lambda \neq 0. \quad (7)$$

이것은  $(n-1)$ 차 다항식을 풀어서  $-1$ 보다 큰 근을 구함으로써 쉽게 계산될 수 있다.

이러한 식을 다중 신경망의 결합에 적용하면,  $h(y_i)$ 는  $i$ 번째 신경망의 출력값이고  $g^i$ 는 이 신경망의 중요도, 즉 신뢰할 수 있는 정도를 나타낸다. 식 (4)에서 이 두값의 최소값은 실제 계산값  $h(y)$ 와 기대값  $g$  사이의 일치도를 나타내기 때문에, 결국 다중 신경망 구조를 위한 퍼지적분은 관측을 통해 얻은 값과 기대값 사이에서 최대한 일치되는 값을 탐색하는 것으로 해석할 수 있다. 여기에서 각 신경망의 중요도  $g^i$ 는 외부에서 엔지니어의 통찰력에 의해서 주관적으로 주어질 수도 있지만 학습 데이터를 통해서 통계적으로 얻어질 수도 있다. 결과적으로 다음의 알고리즘을 얻을 수 있다.

```

알고리즘 : 퍼지적분에 의한 신경망의 결합
calculate  $\lambda$ ; /* 각 부분 신경망의 중요도 */
for each class  $\omega_k$  do
    for each neural network  $y_i$  do
        calculate  $h_k(y_i)$ ;
        determine  $g_k(\{y_i\})$ ;
    end_for
    compute the fuzzy integral;
end_for
determine the final class;
    
```

최종단계에서 가장 큰 적분값을 갖는 클래스  $\omega_k$ 가 출력 클래스로 선택된다.

### 4 실험 결과

제시한 방법의 유용성을 알아보기 위해서 손으로 쓴 온라인의 영문자를 인식하는 문제에 적용하여 보았다. 여기에서 손으로 쓴 문자는 초당 80개의 점을 입력하는 Photron FIOS-6440 LCD 태블릿에 의해서 입력되었으며 SUN 워크스테이션 상에서 처리하였다. 실험대상이 되는 것은 13명의 필기자로 부터 수집된 숫자, 대문자 및 소문자를 분류하는 것이었다.

입력문자는 일련의 획으로 구성되는데, 각각의 획은 펜이 닿은 곳에서 부터 떨어질 때까지의 점들로 구성된다. 입력된 획에는 입력중에 발생한 잡음이나 필기자 자신의 오류로 발생하는 삐침등을 제거하기 위해서 여러가지 전처리 알고리즘을 적용하였다. 사용된 전처리 과정에는 다음과 같은 것이 있다. 즉, 찌친 점 제거(Wild Point Reduction), 점 축소(Dot Reduction), 획 꺾임 제거(Hook Analysis), 세점 평활화(Three Point Smoothing), 예각유지 필터링(Peak Preserving Filtering), N점 정규화(N Point Normalization)등. 전처리된 일련의 데이터 점들은 8방향의 선분들로 근사되어 입력에 사용되었다.

신경망을 학습시키기 위해서 각 클래스마다 40개의 데이터를 사용하였으며 인식실험을 위해서는 이와 별도로 500개의 데이터를 사용하였다. 먼저 비교를 위하여 10개의 입력노드와 20개의 중간노드, 그리고 10개 또는 26개의 출력노드로 이루어진 이층의 신경망을 세개 구현하였다. 각각의 네트워크는 서로 다른 수의 입력노드를 갖도록 한 것이다. 즉,  $NN_1, NN_2, NN_3$ 은 각각 10개, 15개, 20개의 입력노드를 갖는다.

먼저  $g_\lambda$ -퍼지적도인  $g$ 값의 변화에 따라서 퍼지적분이 어떠한 결과를 내는지 알아보았다. 그 결과는 <표 1>과 같다. 여기에서  $\lambda$ 는 식 (7)에 의하여  $0.006\lambda^2 + 0.11\lambda - 0.4 = 0$ 을 만족하여야 하는데 이 식에서  $-1$ 보다 큰 근의 값은  $\lambda = 3.109$ 이다.

<표 2>는 세개의 신경망과 여러가지 결합방법에 대한 인식률을 보여준다. 이 표에서 각 신경망의 중요도  $\lambda$ 는 학습 데이터에 대한 인식률에 의해서 지정하였다. 여기에서 보면 퍼지적분이 과반수 선택법이나 Borda count에 의한 선택법에 비해서 우수한 결과를 낼 수 있다. 그런데, 숫자와 대문자의 경우에는 결합방법이

표 1: 서로다른  $g$ 값에 따른 퍼지적분의 인식결과 (%)

경우	$g^1$	$g^2$	$g^3$	숫자	대문자	소문자
1	0.1	0.2	0.3	78.4	71.6	65.4
2	0.1	0.3	0.2	79.0	73.4	66.8
3	0.2	0.1	0.3	78.8	72.2	64.8
4	0.2	0.3	0.1	79.4	73.8	69.2
5	0.3	0.1	0.2	79.8	74.0	66.2
6	0.3	0.2	0.1	80.2	75.2	70.4

표 2: 퍼지적분과 기타 결합방법의 인식을 비교 (%). 여기에서 BC는 Borda count를 의미한다.

과제	NN <sub>1</sub>	NN <sub>2</sub>	NN <sub>3</sub>	과반수	BC	퍼지적분
숫자	77.4	76.0	76.2	77.8	78.2	81.2
대문자	73.2	66.8	70.8	74.0	77.4	77.6
소문자	59.0	69.8	57.4	65.4	67.8	66.0

우수한 성능을 내지만 소문자의 경우에는 그리 개선되지 않는 것을 볼 수 있다.

결합방법이 소문자의 경우에 하나의 네트워크만을 사용한 경우보다 좋지않은 결과를 낸 원인을 규명하기 위해서 각 네트워크가 오류낼 확률이  $e_1, e_2, e_3$ 인 세개의 네트워크를 고려해 보자. 여기에서  $e_1$ 이 가장 작은 오류라고 할 때 결합방법 (예를 들면, 과반수 투표기법)의 오류가 개별적인 신경망 중에서 가장 작은 오류를 내는 것보다 작게 될 조건은 다음과 같다.

$$e_1 > e_1e_2e_3 + (1 - e_1)e_2e_3 + e_1(1 - e_2)e_3 + e_1e_2(1 - e_3). \quad (8)$$

숫자와 대문자의 경우에는 이 조건이 만족되는 반면에 소문자를 인식하는 경우에는  $e_1$ 이 0.302인데 우측항의 값이 0.322가 되어 조건이 만족되지 않는다.

## 5 결론

본 논문에서는 다중 신경망 방식이라는 다층 신경망의 설계방법을 소개하고 이들을 효과적으로 결합하기 위한 기법으로 퍼지적분 기법을 둔 새로운 방법을 제안하였다. 이 방법의 가장 커다란 장점은 각 부분 신경망의 분류결과 뿐만 아니라 그러한 결과를 낸 신경망의 상대적인 중요도를 효과적으로 고려하여 총체적인 결정을 내릴 수 있다는 점이다. 이의 유용성을 입증하기 위해서 온라인 필기문자를 인식하는데 적용한 결과 좋은 성능을 얻을 수 있었다. 향후 좀더 연구하여야 할 내용은 각 부분 신경망의 인식도를 개선시키고 좀더 많은 수의 부분 신경망을 참여시켜 실제적으로 이용할 수 있도록 하는 것이다.

## 참고 문헌

- [1] K. Hornik, M. Stinchcombe, and H. White, "Multi-layer feedforward networks are universal approximators," *Neural Networks*, **2** (1989) 359-366.
- [2] F. Fogelman Soulie, "Neural network architectures and algorithms: a perspective," *Artificial Neural Networks*, (1991) 605-615 (Netherlands, Elsevier Science Publishers B.V.).
- [3] Y. Mori, and K. Yokosawa, "Neural networks that learn to discriminate similar Kanji characters," *Advances in Neural Information Processing Systems I*. (1989) 332-347 (Morgan Kaufmann).
- [4] R.A. Jacobs, M.I. Jordan, S.J. Nowlan, and G.E. Hinton, "Adaptive mixtures of local experts," *Neural Computation*, **3** (1991) 79-87.
- [5] S.-B. Cho, and J.H. Kim, "Two Design Strategies of Neural Network for Complex Classification Problems," *Proc. 2nd Int. Conf. Fuzzy Logic & Neural Networks*, Izuka, (July 1992) 759-762.
- [6] L.K. Hansen, and P. Salamon, "Neural network ensembles," *IEEE Trans. Pattern Analysis and Machine Intelligence*, **12** (1990) 993-1001.
- [7] S. Shlien, "Multiple binary decision tree classifiers," *Pattern Recognition*, **23** (1990) 757-763.
- [8] M. Sugeno, "Fuzzy measures and fuzzy integrals: A survey," *Fuzzy Automata and Decision Processes*, dm: North Holland, (1977) 89-102.
- [9] K. Leszczynski, P. Penczek, and W. Grochulski, "Sugeno's fuzzy measures and fuzzy clustering," *Fuzzy Sets and Systems*, **15** (1985) 147-158.
- [10] H. Freeman, "Computer processing in line drawing images," *Computing Survey*, **6** (March 1974) 57-98.