

적합도 공유에 의해 증분화된 진화 신경망의 결합

(Fusion of Evolutionary Neural Networks Speciated by Fitness Sharing)

안 준 현^{*} 조 성 배^{**}

(Joon-Hyun Ahn) (Sung-Bae Cho)

요약 진화 신경망은 기존의 경험적 지식 대신에 진화 알고리즘의 전역 탐색 능력을 사용해서 최적의 신경망을 찾는다. 하지만 실제계의 복잡한 문제는 하나의 신경망으로 해결하기 어려운 경우가 많기 때문에 최근에 하나 이상의 신경망을 결합한 다중 신경망에 관한 연구가 활발히 진행되고 있다. 본 논문에서는 진화과정 중 상호보완 가능한 다양한 신경망을 얻기 위한 증분화 방식을 제안한다. 또한 적합도 공유를 통해 증분화된 진화 신경망의 결과를 효과적으로 결합하기 위해 추상 레벨, 순위 레벨, 측정치 레벨의 여러 결합 방법을 이용한 다중 신경망 시스템을 개발한다. UCI 데이터베이스의 벤치마크 문제 중 호주 신용카드 승인 데이터에 대하여 실험한 결과, 증분화를 사용해 탐색한 신경망을 결합한 경우는 더 높은 인식률을 보였으며 Borda 결합의 경우 0.105의 오류율을 보여 제안한 방법이 효과적임을 알 수 있었다.

키워드 : 진화 신경망, 신경망 결합, 증분화, 적합도 공유

Abstract Evolutionary artificial neural networks (EANNs) are towards the near optimal ANN using the global search of evolutionary algorithms instead of trial-and-error process. However, many real-world problems are too hard to be solved by only one ANN. Recently there has been plenty of interest on combining ANNs in the last generation to improve the performance and reliability. This paper proposes a new approach of constructing multiple ANNs which complement each other by speciation. Also, we develop a multiple ANN to combine the results in abstract, rank, and measurement levels. The experimental results on Australian credit approval data from UCI benchmark data set have shown that combining of the speciated EANNs have better recognition ability than EANNs which are not speciated, and the average error rate of 0.105 proves the superiority of the proposed EANNs.

Key words : Evolutionary Algorithm, Fitness Sharing, Neural Network

1. 서론

신경망은 여러 응용 분야에서 널리 사용되고 있는데, 특히 인공지능이나 인지 모델링과 같은 지능 시스템에서 병렬 계산, 입력 데이터의 잡음에 대한 견고함, 뛰어난 학습 능력과 생물학적 신경망과의 유사성으로 인해 많이 사용된다. 하지만 응용 분야에 적절한 신경망은 설계자의 경험에 기반한 시행착오를 통해서 설계된다[1].

이런 단점을 보완하기 위해 진화 알고리즘의 전역 탐색

능력을 신경망 공간에 적용해서 최적의 신경망을 찾는 진화 신경망에 관한 연구가 수행되었다[2]. 진화 신경망의 목표는 어떻게 효과적으로 진화시켜 마지막 세대에 최적화된 신경망을 얻을 수 있는가이다. 따라서 최적화된 하나의 신경망만을 사용하기 때문에 진화와 학습 과정에서 다른 신경망들이 얻은 정보를 모두 무시해 버린다. 또한 최적의 신경망일지라도 단일 신경망이 실제계의 복잡한 문제에 대해서 완벽한 해를 제공하기는 쉽지 않다[3,4].

최근에는 하나 이상의 신경망을 결합한 다중 신경망에 관한 연구가 활발히 진행되고 있다. 다중 신경망은 마지막 세대에 남은 신경망들의 결과를 결합해서 더 좋은 일반화 능력을 가진 시스템을 구축한다. 하지만 다중 신경망이 좋은 성능을 보이기 위해서는 다양한 신경망들이 생성되어서 각 신경망이 서로를 보완해야 한다. 만약 신경망들이 서로 똑같은 결과만을 생성한다면 결합

* 이 논문은 2000년도 연세대학교 학술연구비의 지원에 의하여 이루어진 것임.

† 비회원 : (주)캐직하우스테크놀로지
jhahn@candy.yonsei.ac.kr

** 통신회원 : 연세대학교 컴퓨터과학과 교수
sbcho@csai.yonsei.ac.kr

논문접수 : 2001년 3월 5일

심사완료 : 2001년 9월 4일

으로 성능 향상을 얻기 힘들며 다중 신경망을 구축하는 의미도 없어진다. 신경망이 서로를 보완하기 위해서는 각 신경망이 오류를 범하는 부분이 서로 달라서 한 신경망이 범하는 오류를 다른 신경망이 해결할 수 있어야 한다[3,5].

기존 연구에서는 다양한 신경망을 생성하기 위해 신경망의 초기 가중치나 위상 구조 등을 다양하게 하거나, 각 신경망을 서로 다른 데이터나 알고리즘으로 학습시키는 방법을 사용했다. 이런 방법들은 몇몇 문제에 대해서는 좋은 성능을 보이기도 하지만 근본적으로 많은 시행착오를 거치거나 문제에 대한 사전 지식을 사용해야 한다는 단점이 있다.

본 논문에서는 이를 극복하면서 다양한 신경망을 생성하기 위해 종분화 방식을 도입한다. 종분화는 유전자 알고리즘에서 비슷한 해를 가지는 개체를 모아 종을 이루도록 하기 위하여 적합도가 특정 개체에 쏠리는 것을 방지함으로써 다양한 해를 찾도록 한다. 또한 종분화된 다양한 신경망을 여러가지 방법으로 결합하여 그 우수성을 입증하도록 한다.

본 논문의 구성은 다음과 같다. 먼저 2장에서는 진화 신경망과 다중 신경망의 결합에 관한 관련 연구들을 살펴보고, 3장에서는 다중 신경망을 진화시키는 방법과 종분화를 위한 적합도 공유에 대해서 알아본 후 이를 신경망의 진화에 적용하는 방법을 제안한다. 4장에서는 본 논문에서 사용한 다중 신경망의 결합 방법들을 소개하고, 5장에서는 제안한 방법의 성능 검증에 위한 실험과 그 결과를 분석한다.

2. 관련 연구

2.1 진화 신경망

초기 신경망 자동 설계에 관한 연구에서는 다양한 구성적(constructive) 알고리즘이나 가지치기(pruning) 알고리즘을 사용하였다. 구성적 알고리즘은 최소의 신경망에서 시작하여 학습하면서 필요에 따라 새로운 은닉층과 노드 및 연결 정보를 추가하는 방법이다. 가지치기 알고리즘은 반대로 불필요한 은닉층, 노드, 연결 정보를 제거하면서 신경망을 설계하는 방식이다. 하지만 두 방식 모두 신경망의 전체적인 구조 영역이 아니라 주어진 환경에서 제한된 영역만을 탐색하므로 최적화된 신경망을 찾기 어렵다.

이런 한계를 극복하기 위하여 진화 알고리즘이 도입되었다. 진화 신경망은 진화 알고리즘을 사용하여 자동으로 신경망을 설계하는 방법으로 신경망의 가중치, 위상 구조, 은닉 노드 수, 학습 알고리즘 등 신경망 학습

시 결정해야 하는 인자들을 진화를 통해 찾아서 최적의 신경망을 결정한다.

Harp 등은 학습 알고리즘의 인자를 결정하는데 진화적 접근 방식을 이용했다[6]. 역전파 알고리즘의 인자를 신경망 구조 정보와 함께 염색체로 표현하고 이를 진화시켰다. 이 방법은 신경망의 학습 인자 정보와 신경망 구조 정보간의 상호 작용을 통해 보다 근본적으로 역전파 알고리즘과 진화 알고리즘을 결합하였다.

Lee는 신경망을 학습시키기 전에 최적에 가까운 초기 가중치 집합을 결정하기 위해 유전자 알고리즘을 사용하고, 역전파 알고리즘을 이용해서 최종적으로 학습을 시켰다[7]. 또한 유전자 알고리즘과 역전파 알고리즘을 함께 사용한 경우가 좀더 효과적인 학습이 가능함을 보였다.

Castillo 등은 G-Prop-II라는 진화 신경망 시스템을 제안했다[8]. 이 시스템은 유전자 알고리즘과 역전파 알고리즘을 사용해서 다층 퍼셉트론을 학습시킨다. 유전자 알고리즘으로 은닉층 크기, 초기 가중치 값, 학습 인자를 모두 다층 퍼셉트론의 인자 공간에서 탐색해서 최적화된 다층 퍼셉트론을 찾는다.

Yao와 Liu는 EPNet이라는 진화 신경망을 제안하였다. EPNet은 Fogel이 제안한 진화 프로그래밍(evolutionary programming)을 이용해서 신경망을 결정한다[9]. EPNet은 교차연산을 사용하지 않고 신경망의 구조와 가중치 값을 변경시키는 다섯 가지 돌연변이 연산자를 사용한다. 혼합 학습, 노드 삭제, 연결 삭제, 연결 추가 그리고 노드 추가의 다섯 가지 연산자를 통해 신경망의 기능적인 진화를 유도한다. 따라서 부모 세대가 학습한 기능적인 정보를 자식 세대에까지 최대한 보존하도록 한다. EPNet은 좋은 성능을 보이지만, 진화 과정에서 다른 신경망들이 학습한 정보를 모두 무시하고 하나의 신경망만을 사용하므로 성능에 한계가 있다.

2.2 다중 신경망의 결합

여러 분야에서 다양한 해결책을 결합하여 일반화 능력이 뛰어나고 신뢰도 높은 시스템이 구축되고 있다. 신경망 분야에서도 마찬가지로 여러 개의 신경망을 사용함으로써 하나의 신경망이 잘못된 결과를 도출하더라도 다른 신경망에 의해 보완될 수 있도록 하는 연구가 활발하다[4]. 신경망을 결합하는 기본 개념은 각각의 신경망으로부터 나온 출력값을 이용하여 결합 인식기의 출력값을 결정하는 것이다. 이때, 어떠한 결합방법이 사용되었는가와 어떠한 신경망을 결합대상으로 하였는가가 전체 인식기의 성능을 결정한다. 이러한 신경망의 결합은 결과의 형태에 따라 세 가지 레벨로 나눌 수 있다[10]. 하나의 결과만을 출력하는 추상 레벨(abstract level)과 순위의

형태로 출력하는 순위 레벨(rank level), 그리고 순위와 더불어 신뢰값까지 함께 출력하는 측정치 레벨(measurement level)등으로 분류된다[11].

추상 레벨은 각 신경망이 하나의 결과만을 출력하는 방식으로 대표적인 방법으로는 다수결 방법, 베이시안 방법[12] 등이 있다. 이 방법의 경우, 제공되는 정보가 적기 때문에 신경망의 과거 성향이 결합에 크게 관여하게 된다. 순위 레벨은 각 신경망이 인식에 참여하는 모든 클래스의 순위를 결과로 출력한 후 모든 클래스들의 순위를 비교하여 결과를 산출하는 방식이다. 이 방법은 보다 많은 정보를 바탕으로 결합을 수행하고, 과거의 성향이 결합에는 거의 사용되지 않는다는 특성이 있다. Borda 함수, 가중화 Borda함수, Condorect 함수를 통한 결합이 순위 레벨에 속한다. 측정치 레벨의 경우, 각 인식이 클래스들에 대한 순위와 더불어 각 순위에 대한 신뢰값까지 함께 출력하는 방식이기 때문에, 가장 많은 정보로 결합을 수행한다. 그러나, 각 인식이 별로 제공되는 신뢰값이 일관성을 유지해야 한다는 어려움이 있다. 대표적인 방법으로 평균[13]과 가중 평균 방법이 있다.

Opitz와 Shavlik은 다중 신경망의 결합에 진화 알고리즘을 적용한 ADDEMUP 알고리즘을 제안했다[3]. ADDEMUP은 신경망들 각각이 좋은 성능을 보이면서 상호 보완되도록 신경망 결합을 구축하는 알고리즘이다. ADDEMUP은 각 신경망의 적합도를 다음과 같이 정의한다.

$$Fitness_i = Accuracy_i + \lambda Diversity_i,$$

$$\text{단, } Diversity_i = \sum_x (o_i(x) - \bar{o}(x))^2$$

여기서 $o_i(x)$ 와 $\bar{o}(x)$ 는 각각 신경망 i 의 출력과 신경망 집단의 출력값을 나타낸다.

유전자 알고리즘을 사용해서 새로운 신경망을 생성하고 그 신경망의 적합도를 인식률과 다양성으로 평가해서 신경망 집단에 포함할지를 결정한다. 즉 신경망 집단에 새로이 추가할 신경망은 높은 인식률을 가지면서 기존의 신경망 집단과는 다른 기능을 갖도록 한다. ADDEMUP은 앞의 신경망 결합 연구들이 이미 생성된 신경망들을 어떻게 결합시킬지에 초점을 맞춘데 반해, 적합도 값을 이용해서 다양한 신경망의 생성을 유도했다.

Yao와 Liu는 EPNet의 마지막 세대 신경망을 투표, 순위 기반 선형 결합 그리고 RLS 알고리즘을 이용한 선형 결합 방법을 이용해서 결합하였다[14]. 최고 성능의 단일 신경망보다 더 좋은 성능을 얻을 수 있었지만, 간단한 결합 방법만을 사용하였고 집단내의 모든 신경망을 결합에 사용하였다는 한계가 있다.

3. 종분화 신경망의 진화

제안하는 종분화된 진화 신경망은 다음과 같은 알고리즘으로 생성된다(그림 1 참조).

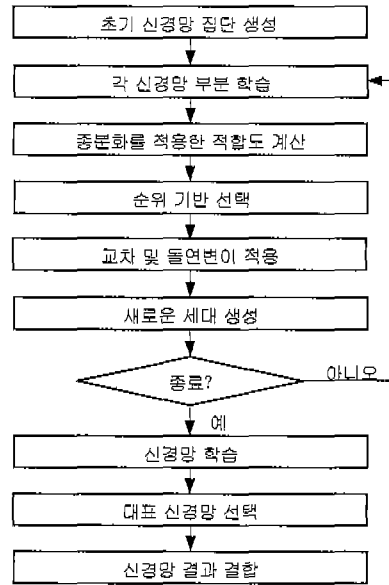


그림 1 적합도 공유를 적용한 신경망 진화 알고리즘

- 단계 1 (초기 신경망 집단 생성) : 임의의 가중치로 초기화된 신경망의 집단을 생성한다.
- 단계 2 (각 신경망의 부분 학습) : 각 신경망을 역전파 알고리즘으로 부분 학습시킨다.
- 단계 3 (종분화를 적용한 적합도 값 계산) : 검증 데이터로 테스트하여 각 신경망의 적합도를 종분화 방식으로 계산한다.
- 단계 4 (순위 기반 선택) : 적합도가 구해지면 인식률이 높은 상위 50% 개체를 선택한다.
- 단계 5 (교차 및 돌연변이 적용) : 신경망 집단에 교차와 돌연변이 연산을 적용한다.
- 단계 6 (새로운 세대 생성)
- 단계 7 (종료 조건 검사) : 종료 조건이 만족되지 않으면 단계 2로 되돌아간다.
- 단계 8 (신경망 학습) : 마지막 세대의 신경망 집단을 충분히 학습시킨다.
- 단계 9 (대표 신경망 선택) : 신경망 집단을 분석해서 각 신경망 중들로부터 대표 신경망을 선택한다.

- 단계 10 (신경망 결과 결합) : 테스트 데이터에 대한 대표 신경망의 인식 결과를 각 레벨의 적절한 방법으로 결합한다.

3.1 신경망의 진화

신경망을 진화시키기 위해서는 이를 진화에 적합한 형태로 인코딩하여야 한다. 신경망 인코딩은 유전자 알고리즘에서 개체를 표현하는데 일반적으로 사용하는 이진 표현, 트리, 링크드 리스트, 행렬 등의 방법이 있다. 본 논문에서는 간단하면서도 진화 연산자를 적용하기 쉬운 행렬을 사용한다. N개의 노드를 가진 신경망은 N x N 크기의 정방행렬에 연결 정보와 가중치를 표시해서 인코딩한다. 행렬의 위상단에는 노드간 연결 정보를 1과 0으로 표시하고, 각 연결 정보에 대칭하는 좌하단에는 가중치를 나타낸다. 그림 2는 전체 노드 수가 네 개인 신경망과 해당 유전자형의 예를 보여준다.

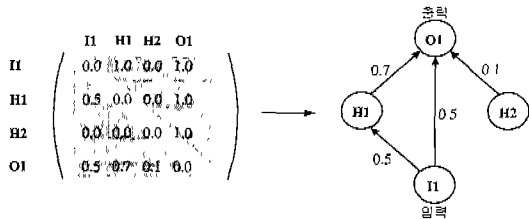


그림 2 행렬을 이용한 신경망 인코딩 방법(좌측의 행렬 표현은 우측의 신경망과 대응)

(1) 교차 연산

교차 연산은 두 개의 개체를 선택해서 그 중 일부분을 교환하는 것으로 두 개의 신경망을 선택한 후 그 연결 정보와 가중치를 서로 교환한다. 우선 교차 연산을 적용할 두 신경망을 선택하고 교차점을 은닉 노드 중에서 하나 선택한다. 그리고 두 신경망에서 이 노드의 연결 정보와 가중치 정보를 서로 교환한다. 교차점은 두 신경망 상의 같은 위치에 있는 은닉노드로 선택된다. 그림 3은 은닉노드 H2를 교차점으로 한 교차 연산의 예를 보여준다. 이 교차 연산으로 H2 노드 이후에 있는 모든 은닉 노드의 연결 정보와 가중치 정보가 교환된다.

(2) 돌연변이 연산

돌연변이 연산은 임의로 선택된 개체의 일부분을 바꾸는 것으로 연결 삭제와 연결 생성의 두 형태로 작용한다. 연결 삭제는 임의의 연결 정보를 선택해서 실제 연결이 있는 경우 이 연결을 삭제하고, 반대로 연결 생성이 없는 경우 새로운 연결을 생성한 후 임의의 값으로 연결 강도를 설정한다. 그림 4는 돌연변이 연산의 예

를 보여준다.

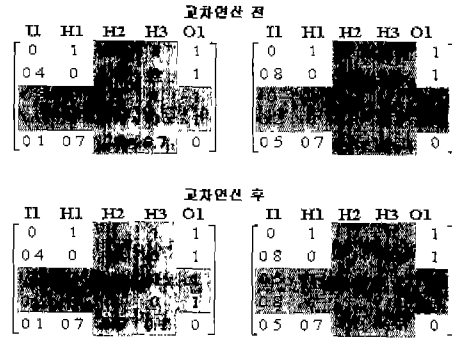


그림 3 교차 연산(임의의 하나의 은닉노드층을 선택하여 그 층을 중심으로 교차)

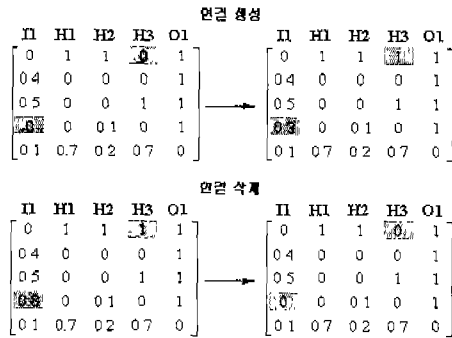


그림 4 돌연변이 연산(연결이 존재하는 경우 삭제하고, 연결이 없는 경우 임의의 연결가중치로 생성)

3.2 적합도 공유

신경망 집단의 종분화를 위해 적합도 공유를 사용한다. 적합도 공유는 적합도가 집중된 개체의 적합도를 줄이고 그 주위의 개체들과 적합도를 공유해서 유전자 알고리즘의 탐색 반경을 넓힌다. 결과적으로 탐색하지 않았던 영역을 탐색해서 다양한 해를 찾으려 한다.

각 개체의 적합도를 f_i 라 하고, $sh(d_{ij})$ 를 공유 함수라고 할 때 적합도 공유값 fs_i 는 다음과 같이 계산된다.

$$fs_i = \frac{f_i}{\text{population size} \sum_{j=1} sh(d_{ij})}$$

공유 함수 $sh(d_{ij})$ 는 두 개체 간의 공유도를 나타내는 것으로 서로 다른 개체는 작은 값을 가지고 비슷한 개체들은 큰 값을 가지게 된다. 공유 함수는 각 개체간의 유사성 차이(d_{ij})를 이용해서 다음과 같이 구한다.

$$sh(d_{ij}) = \begin{cases} 1 - \frac{d_{ij}}{\sigma_s}, & 0 \leq d_{ij} < \sigma_s, \\ 0, & d_{ij} \geq \sigma_s. \end{cases}$$

σ_s 는 공유 반경을 나타내는 값으로 적합도 공유를 할만큼 비슷한 개체인지를 구분하는 기준이 된다.

그림 5는 적합도 공유의 예를 보여주는데, 개체 i 는 이와 유사한 개체 a, b, c 와 적합도를 공유한다. 이 개체들은 개체 i 와 d_{ia}, d_{ib}, d_{ic} 만큼 떨어져 있으며, 이 값들은 모두 공유 반경 σ_s 보다 작다. 적합도는 공유 함수 $sh(d_{ij})$ 에 따라 거리에 반비례해서 공유하게 된다.

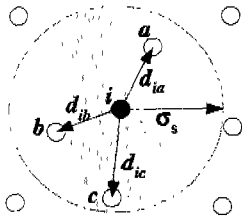


그림 5 적합도 공유의 예(원은 적합도 공유반경을 표현)

3.3 신경망의 증분화

신경망의 적합도로 검증 데이터의 인식률을 사용하였다. 적합도 공유를 적용하는데 있어서 문제가 되는 것은 개체간의 유사성을 어떻게 비교할 것인가와 공유 반경 σ_s 를 어떻게 정하는가이다.

(1) 유사도 측정 및 공유 반경 설정

아직까지 두 신경망을 비교하는데 최적인 기준이 없기 때문에 각 신경망 출력값의 평균을 계산해서 유사도를 계산했다. 입력 데이터 i 에 대한 신경망 i 의 출력값 o'_i 는 다음과 같이 정의할 수 있다.

$$o'_i = (o'_{i1}, o'_{i2}, \dots, o'_{iN})$$

이 때 N 은 신경망 i 의 출력 노드 수이다.

학습 데이터가 총 T 개 있다면 학습 데이터에 대한 신경망 i 의 j 번째 출력 노드의 출력값 평균 \bar{o}_{ij} 는 다음과 같이 계산된다.

$$\bar{o}_{ij} = \frac{\sum_{t=1}^T o'_{ij}}{T}$$

이 때 신경망 i 의 출력값 평균 \bar{o}_i 는 다음과 같다.

$$\bar{o}_i = (\bar{o}_{i1}, \bar{o}_{i2}, \dots, \bar{o}_{iN})$$

신경망 i, j 간의 유사도는 다음과 같이 각 신경망의 출력값 평균의 유클리드 거리로 측정한다.

$$d_{ij} = \sqrt{\sum_{n=1}^N (\bar{o}_{in} - \bar{o}_{jn})^2}$$

따라서 두 신경망이 유사할수록 d_{ij} 의 값은 작아진다.

공유 반경은 두 개체가 적합도를 공유하는지를 결정하는 유사도의 기준값이다. 일정한 차이가 나는 다중해를 가지는 경우 그 거리를 공유 반경으로 정했을 때 가장 효과적인 적합도 공유를 할 수 있다. 공유 반경을 설정하기 위해서 초기 신경망 집단내의 모든 개체간의 유사도를 계산 한 후 평균값을 구하여 이등분 한 후 이용한다.

$$\sigma_s = \frac{1}{2(i \times j)} \sum_i \sum_j \sqrt{\sum_{n=1}^N (\bar{o}_{in} - \bar{o}_{jn})^2}$$

(2) 종결정 및 대표 신경망 선택

증분화된 신경망은 유사한 특징을 가진 신경망들이 하나의 종으로 모여있으므로 결합시에 각 종을 대표하는 신경망을 선택해서 이들의 결과를 결합한다. 신경망 집단에서 종을 구분하기 위해서 단일 연결 클러스터링(single linkage clustering) 분석법을 사용했다[15]. 단일 연결 클러스터링은 전체 집단 중에서 가장 유사한 개체부터 하나의 클러스터로 묶어나가는 전형적인 계층적 클러스터링 기법이다. 이 분석법을 적용해서 종을 나누기 위한 계통도(dendrogram)를 그렸다. 그림 6은 단일 연결 클러스터링을 적용했을 때 얻을 수 있는 신경망 집단의 계통도 예를 보여준다. X축은 20개의 신경망 집단의 인덱스를 나타내며 Y축은 유사도 값을 나타낸다.

종의 결정은 유사도 값의 레벨에 따라 선택한다. 예를 들어 그림 6에서 유사도 레벨을 0.06으로 한 경우 3개의 종이 구성된다. 18번과 19번 신경망, 1번과 2번 신경

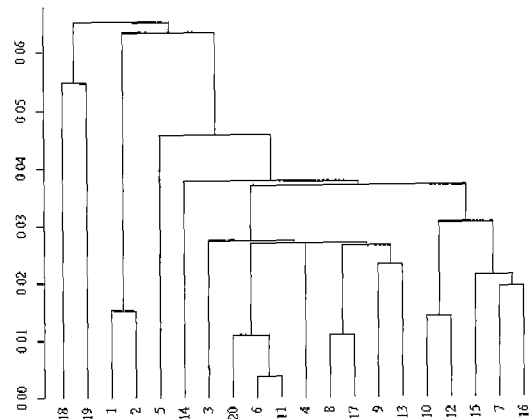


그림 6 신경망 계통도의 예

망, 그리고 나머지 신경망이 각각 하나의 종을 이루게 된다. 각 종의 대표 신경망은 종을 이루는 신경망 중에서 가장 성능이 좋은 신경망으로 선택했다. 최적화된 종의 수를 결정하기 위해서 유사도 레벨별로 종의 개수를 변화시켜가면서 결과를 뽑고 그 중에서 가장 좋은 성능을 보이는 경우를 최종 종의 수로 결정하였다.

4. 다중 신경망의 결합

신경망 집단의 결과를 결합하기 위해 추상 레벨에서는 투표 방법과 게이팅 방법을 사용하였고, 순위 레벨에서는 Borda 함수와 Condorect 함수를 사용하였다. 측정치 레벨에 속하는 방법으로는 평균 방법과 가중 평균 방법을 사용하였다.

다중 신경망이 M 개의 클래스($A = \{1, \dots, M\}$)로 이루어진 문제에서 입력된 미지의 패턴 x 가 어느 클래스에 속하는지를 결정하는 것으로 볼 때, 신경망의 결합은 K 개의 신경망이 출력하는 결과 e_k 를 이용하여 결합된 결과 $F(e_k(x))$ 를 결정하는 것이다. 이때 C_i 는 i 번째 클래스를 나타낸다.

4.1 추상 레벨

(1) 투표 방법

가장 많은 신경망이 선택한 결과를 다중 신경망의 결과로 사용하는 것으로 다음과 같은 방식으로 구한다.

$$F(e_k(x)) = j \quad \text{if } S(x \in C_j) = \max(S(x \in C_i))$$

$$S(x \in C_i) = \sum_{k=1}^K G_k(x \in C_i), i \in A$$

$$G_k(x \in C_i) = \begin{cases} 1 & \text{if } e_k(x) = i \text{ and } i \in A \\ 0 & \text{otherwise} \end{cases}$$

이 방법은 별도의 계산 과정이나 기억 장소없이 다중 신경망의 결과를 결합할 수 있다. 그러나 각 신경망의 성능에 관계없이 동일한 가중치로 결합하기 때문에 상대적으로 성능이 좋지 않은 신경망이 포함되어 있을 경우, 결합 결과에 큰 영향을 미칠 수 있다.

(2) 게이팅 방법

학습 데이터에 대한 정보를 테스트 데이터에 적용해서 가장 적합한 신경망을 선택하는 방법이다. 알고리즘은 다음과 같다.

- 단계 1 : 학습시 각 신경망은 자신이 올바르게 인식한 데이터 목록 생성

- 단계 2 : 테스트 시에는 입력 데이터와 가장 유사한 학습 데이터 검색

- 단계 3 : 선택된 학습 데이터를 올바르게 인식한 신경망 선택

- 단계 4 : 선택된 신경망으로 테스트 데이터 분류

4.2 순위 레벨

(1) Borda 함수

클래스 i 에 대하여 신경망 k 가 출력한 순위 r_k^i 에 따라 $M - r_k^i$ 를 구하고 이 값들을 모든 신경망에 대해 합산한 것을 Borda 점수로 정한 후, 이 점수가 가장 큰 클래스를 결과로 결정하는 방식이다. Borda 함수는 다음과 같다.

$$F(e_k(x)) = \max_{i \in A} (B_i(e_k(x)))$$

$$\text{단, } B_i(e_k(x)) = \sum_{k=1}^K (M - r_k^i(x))$$

(2) Condorect 함수

각 클래스에 대하여 자신을 제외한 다른 모든 클래스와 순위를 비교하여 높은 순위의 신경망의 개수를 합하여 최소 값을 구한 후 그 값을 그 클래스의 Condorect 값으로 정하고, 이 값들을 비교하여 높은 값을 가지는 클래스가 상위의 순위가 되는 방식이다. Condorect 함수는 다음과 같다.

$$F(e_k(x)) = \max_{i \in A} (Con_i(e_k(x)))$$

$$\text{단, } Con_i(e_k(x)) = \min_{j \in M-i} \#(K : r_k^j > r_k^i)$$

여기서, $\#(K : r_k^j > r_k^i)$ 는 클래스 i 가 클래스 j 보다 높은 순위라고 판별한 신경망의 총수이다.

4.3 측정치 레벨

(1) 평균

비교적 간단한 방법으로 각 클래스 i 마다 주어진 신경망의 출력값 m_k^i 를 모두 합하여 평균을 구해서 값이 가장 큰 클래스를 다중 신경망의 결과로 사용하는 방법이다. 수식으로 표현하면 다음과 같다.

$$F(e_k(x)) = \max_{j \in A} S_j \quad \text{단, } S_i = \left(\sum_{k=1}^K m_k^i \right) / K$$

(2) 가중 평균

평균 방법을 보완한 것으로, 각 신경망의 인식을 값을 이용해서 각 신경망마다 가중치를 부여한 후, 출력값에 가중치를 곱해서 평균을 구한다. 신경망의 인식률이 높을수록 더 높은 가중치를 부여하기 위해 신경망의 오류율을 이용해서 가중치를 결정하였다. 신경망 i 의 오류율이 E_i 일 때 가중치 w_i 는 다음과 같이 구한다.

$$w_i = \frac{1 - E_i}{\sum_k (1 - E_k)}$$

5. 실험결과

5.1 실험 환경

실험을 통해 신경망의 증분화를 분석하고 결합의 성

능을 기존 연구 결과들과 비교해 보았다. 실험을 위하여 UCI의 벤치마크 데이터 중에서 호주 신용카드 승인 (Australian Credit Approval) 데이터를 사용했다. '+', '-' 두 클래스를 가진 문제로 총 데이터 수는 690개이며 각 데이터는 14개의 속성을 가지고 있다. 총 데이터 중 307개는 '+', 383개는 '-' 클래스이다. 학습 데이터, 검증 데이터, 테스트 데이터는 각각 346개, 172개, 172개씩 사용하였다.

각 세대의 신경망 수는 20개이고 총 200세대를 진화시켰다. 각 신경망은 초기 5개의 은닉 노드를 가진 전방향 신경망이고 학습은 역전파 알고리즘을 사용하였다. 학습률은 0.1이며 학습 데이터는 최초 부분 학습시 400회, 부분 학습시 200회, 최종 학습시에 1000회씩 사용하였다. 교차 연산은 0.3, 돌연변이 연산은 0.1의 비율로 적용했다. 적합도 공유 변경은 실험을 통해 0.003의 값으로 정하였다. 실험 결과는 신경망 집단을 다르게 초기화해서 총 10회 수행한 것의 평균이다.

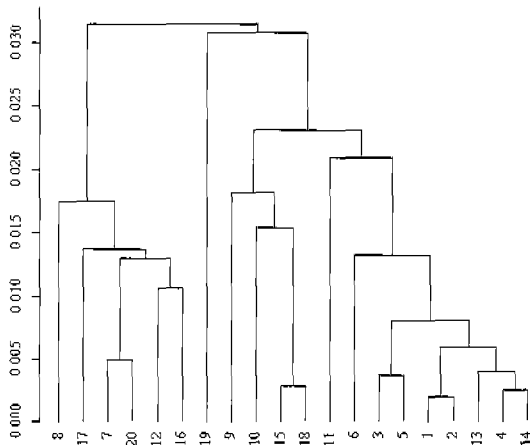


그림 7 증분화된 경우의 신경망 계통도

5.2 실험 결과 및 분석

그림 7은 증분화를 사용한 경우 진화로 얻어진 최종 세대의 신경망 계통도를 나타내며, 표 1은 각 신경망의 구조 및 테스트 데이터에 대한 성능을 보여준다. 유사도 차이를 0.020 이상으로 했을 때, 신경망은 (8, 17, 7, 20, 12, 16), (19), (9, 10, 15, 18), (11, 6, 3, 5, 1, 2, 13, 4, 14)의 네 종류로 분화하였으며 각 종에서 가장 성능이 좋은 대표 신경망을 선택하여 최종적으로 신경망 1, 9, 12, 19로 다중 신경망을 구축하였다.

마지막 세대 신경망 집단의 인식률을 비교해 보면 증

분화를 사용하지 않은 기존 방식의 경우 학습, 검증, 그리고 테스트 데이터에 대해서 각각 0.9116, 0.8415, 0.8401의 인식률로 증분화를 사용했을 때의 0.9042, 0.8381, 그리고 0.8273에 비해 더 나은 성능을 보였다. 그러나, 표 2, 3, 4에서 볼 수 있듯이 전체 신경망을 평균 방법으로 결합한 경우 기존 방식과 증분화 방식은 비슷한 결과를 보였으나, 종으로부터 대표 신경망을 선택해서 결합한 경우는 전체적으로 향상된 결과를 보였다. 증분화를 하지 않은 경우 결합을 통해서 거의 성능 향상을 얻지 못했다. 반면 증분화한 경우 신경망의 진화 과정에서 상대적으로 인식률이 낮은 신경망들이 생성되었음에도 불구하고 결합을 통해 높은 성능 향상을 얻었다. 이는 신경망의 진화 과정에서 증분화가 서로 다른 특성을 지닌 다양한 신경망 객체들을 생성시켰으며 이들의 결합을 통해 높은 성능 향상을 얻었다는 것을 의미한다.

표 1 각 신경망의 구조 및 성능

| 신경망 | 입력 | 은닉 | 출력 | 인식률 |
|-----|----|----|----|--------|
| 1 | 14 | 5 | 2 | 0.8314 |
| 2 | 14 | 5 | 2 | 0.843 |
| 3 | 14 | 5 | 2 | 0.8198 |
| 4 | 13 | 5 | 2 | 0.8372 |
| 5 | 14 | 5 | 2 | 0.8256 |
| 6 | 14 | 5 | 2 | 0.843 |
| 7 | 13 | 5 | 2 | 0.8547 |
| 8 | 14 | 5 | 2 | 0.8314 |
| 9 | 14 | 5 | 2 | 0.8547 |
| 10 | 14 | 5 | 2 | 0.8488 |
| 11 | 14 | 5 | 2 | 0.8314 |
| 12 | 14 | 5 | 2 | 0.8081 |
| 13 | 14 | 5 | 2 | 0.843 |
| 14 | 14 | 5 | 2 | 0.8372 |
| 15 | 14 | 5 | 2 | 0.8488 |
| 16 | 14 | 5 | 2 | 0.843 |
| 17 | 14 | 5 | 2 | 0.8198 |
| 18 | 14 | 5 | 2 | 0.8198 |
| 19 | 14 | 5 | 2 | 0.8605 |
| 20 | 14 | 5 | 2 | 0.8198 |

표 2 기존 방식(전체 결합)의 인식률: 호주 신용카드 데이터

| 기존방식 (전체결합) | 투표 | 게이팅 | Borda | Condorect | 평균 | 가중 평균 |
|-------------|--------|--------|--------|-----------|--------|--------|
| 평균 | 0.8332 | 0.8500 | 0.8343 | 0.8384 | 0.8390 | 0.8395 |
| 표준편차 | 0.0153 | 0.0215 | 0.0092 | 0.0109 | 0.0145 | 0.0137 |
| 최대 | 0.8547 | 0.8895 | 0.8430 | 0.8547 | 0.8547 | 0.8547 |
| 최소 | 0.8140 | 0.8256 | 0.8198 | 0.8256 | 0.8198 | 0.8198 |

표 3 종분화 방식(전체 결합)의 인식률: 호주 신용카드 데이터

| 종분화방식 (전체결합) | 투표 | 케이팅 | Borda | Condorect | 평균 | 가중 평균 |
|-----------------|--------|--------|--------|-----------|--------|----------|
| 평균 | 0.8349 | 0.8337 | 0.8349 | 0.8400 | 0.8337 | 0.8337 |
| 표준편차 | 0.0120 | 0.0151 | 0.0120 | 0.0134 | 0.0157 | 0.0157 |
| 최대 | 0.8430 | 0.8488 | 0.8430 | 0.8547 | 0.8488 | 0.8488 |
| 최소 | 0.8140 | 0.8140 | 0.8140 | 0.8198 | 0.8081 | 0.8081 |

표 4 종분화 방식(대표 결합)의 인식률: 호주 신용카드 데이터

| 종분화방식 (대표결합) | 투표 | 케이팅 | Borda | Condorect | 평균 | 가중 평균 |
|-----------------|--------|--------|--------|-----------|--------|----------|
| 평균 | 0.8430 | 0.8337 | 0.8430 | 0.8454 | 0.8430 | 0.8430 |
| 표준편차 | 0.0117 | 0.0151 | 0.0117 | 0.0098 | 0.0109 | 0.0109 |
| 최대 | 0.8547 | 0.8488 | 0.8547 | 0.8547 | 0.8547 | 0.8547 |
| 최소 | 0.8314 | 0.814 | 0.8314 | 0.8314 | 0.8256 | 0.8256 |

5.3 기존 연구 결과와의 비교

표 5는 Borda 결합 방식을 사용한 종분화 신경망의 결과와 다른 연구 결과[16]의 오류율을 비교한 값이다. 제시한 결합 방법 중에서는 Borda 함수와 투표 방식으로 결합한 경우가 상대적으로 기각률이 높긴하였지만 가장 좋은 결과를 보였으며, 기존 연구 결과보다 낮은 오류율을 보였다. 표 6은 각 결합방법별로 총 10회의 실험에 대한 평균 기각률을 정리해 놓은 것이다. 표 7은 당뇨병 데이터에 대한 대표적인 기존연구와의 비교결과이다.

표 5 기존 연구 결과와의 비교 : 호주 신용카드 데이터

| 알고리즘 | 오류율 | 알고리즘 | 오류율 |
|----------|-------|----------|-------|
| 제안한 방법 | 0.105 | IndCART | 0.152 |
| EPNet | 0.115 | Backprop | 0.154 |
| Cal5 | 0.131 | C4.5 | 0.155 |
| JTrule | 0.137 | SMART | 0.158 |
| DIPOL92 | 0.141 | Baytree | 0.171 |
| Discrim | 0.141 | k-NN | 0.181 |
| Logdisc | 0.141 | NewID | 0.181 |
| CART | 0.145 | LVQ | 0.197 |
| RBF | 0.145 | ALLOC80 | 0.201 |
| CASTLE | 0.148 | CN2 | 0.204 |
| NaiveBay | 0.151 | Quadisc | 0.207 |

표 6 결합방식별 평균 기각률: 호주 신용카드 데이터

| 종분화방식 (대표결합) | 투표 | 케이팅 | Borda | Condorect | 평균 | 가중 평균 |
|-----------------|------|------|-------|-----------|------|----------|
| 평균 | 4.1% | 2.1% | 3.7% | 1.5% | 2.3% | 2.4% |

표 7 기존 연구 결과와의 비교 : 당뇨병 데이터

| | 제안하는 방법 | Logdisc | EPNet | DIPOL92 | Discrim |
|-----|---------|---------|---------|---------|---------|
| 오류율 | 0.218 | 0.223 | 0.224 | 0.224 | 0.225 |
| | SMART | RBF | Itrule | BP | Cal5 |
| 오류율 | 0.232 | 0.243 | 0.245 | 0.248 | 0.25 |
| | CART | CASTLE | Quadisc | | |
| 오류율 | 0.255 | 0.258 | 0.262 | | |

6. 결론

본 논문에서는 신경망의 진화에 종분화를 적용해서 다양한 신경망들로 구성된 다중 신경망을 구축하고 유용성을 보이기 위해 다양한 결합 방법을 이용하였다. 신경망의 종분화를 위하여 적합도 공유 방법을 이용했으며, 각 신경망의 유사도를 비교하기 위해서 출력값의 평균을 이용했다. 종분화된 신경망의 경우 평균적인 신경망의 인식률이 종분화하지 않은 경우보다 낮았으나, 이들을 결합함으로써 더 높은 인식률을 얻을 수 있었다. 종분화된 신경망의 결합은 추상 레벨, 순위 레벨, 그리고 측정치 레벨의 다양한 방법을 적용해 보았으며, 그 중에서 Borda 결합이 가장 낮은 오류율을 보였다. 향후 연구에서는 신경망 간의 유사도에 대한 연구를 보충하고 종분화된 신경망에 대한 깊이 있는 분석을 시도할 것이다.

참고 문헌

- [1] K. Balakrishnan and V. Honavar, "Evolutionary Design of Neural Architectures - A Preliminary Taxonomy and Guide to Literature," *Tech. Rept. CS TR95-01*. Department of Computer Science, Iowa State University, Ames, 1995.
- [2] X. Yao, "Evolving Artificial Neural Networks," *Proceedings of the IEEE*, 87(9):1423-1447, September 1999.
- [3] D. W. Opitz and J. W. Shavlik, "Actively Searching for an Effective Neural Network Ensemble," *Connection Science*, Vol 8, Nos 3 & 4, pp. 337-353, 1996.
- [4] A. J. C. Sharkey, "On Combining Artificial Neural Nets," *Connection Science*, Vol. 8, pp.299-313, 1996.
- [5] A. J. C. Sharkey and N. E. Sharkey, "Combining Diverse Neural Nets," *The Knowledge Engineering Review*, 12:3, 231-247, 1997.
- [6] S. A. Harp, T. Samad and A. Guha, "Toward the Genetic Synthesis of Neural Networks," in *Proc. 3rd Int. Conf. Genetic Algorithms and Their Applications*, J. D. Schaffer, Ed. San Mateo, CA:

Morgan Kaufmann, pp 379-384, 1989.

- [7] S.-W. Lee, "Off-line Recognition of Totally Unconstrained Hand-Written Numerals Using Multilayer Cluster Neural Network," *IEEE Trans, Pattern Anal. Machine Intell.*, vol. 18, pp. 648-652, 1996.
- [8] P. A. Castillo, V. Rivas, J.J. Merelo, J. Gonzalez. A. Prieto and G. Romero, "G-Prop-II: Global Optimization of Multilayer Perceptrons Using GAs," *Proceedings of the 1999 Congress on Evolutionary Computation*, vol. 3, pp. 2022-2027, May 1999.
- [9] X. Yao and Y. Liu, "A New Evolutionary System for Evolving Artificial Neural Networks," *IEEE Transactions on Neural Networks*, vol 8, pp. 694-713, May 1998.
- [10] L. Xu, A. Krzyzak and C. Y. Suen, "Methods of Combining Multiple Classifiers and Their Applications to Handwriting Recognition," *IEEE Trans. on Systems, Man and Cybernetics*, vol. SMC-22, no. 3, pp. 418-435, 1992.
- [11] 백종현, *다중 인식기의 다단계 결합을 통한 무제약 필기숫자 인식*, 연세대학교 대학원 박사학위 논문, 1996.
- [12] J.C. Bioch, O.V.D. Meer, and R. Potharst, "Classification Using Bayesian Neural Nets," *IEEE International Conference on Neural Networks*, vol. 3, pp. 1488-1493, 1996.
- [13] M. Perrone and L. N. Cooper, "When Networks Disagree: Ensemble Methods for Hybrid Neural Networks," *Neural Networks for Speech and Image Processing*, Chapman Hall, 1993.
- [14] X. Yao and Y. Liu, "Making Use of Population Information in Evolutionary Artificial Neural Networks," *IEEE Transactions on Systems, Man and Cybernetics*, Part B: Cybernetics, 28(3):417-425, June 1998.
- [15] A. D. Gordon, *Classification: Methods for the Exploratory Analysis of Multivariate Data*, Chapman and Hall, 1981.
- [16] D. Michie, D. J. Spiegelhalter and C. C. Taylor, *Machine Learning, Neural and Statistical Classification*, 1994.



조 성 배

1988년 연세대학교 전산학과(학사).
1990년 한국과학기술원 전산학과(석사).
1993년 한국과학기술원 전산학과(박사).
1993년 ~ 1995년 일본 ATR 인간정보통신연구소 객원 연구원. 1998년 호주 Univ. of New South Wales 초청연구원. 1995년 ~ 현재 연세대학교 컴퓨터과학과 부교수. 관심분야는 신경망, 패턴인식, 지능정보처리.



안 준 현

1998년 2월 연세대학교 컴퓨터과학과 졸업(학사). 2001년 2월 연세대학교 컴퓨터과학과 석사과정 졸업. 2001년 3월 (주)매직하우스테크놀로지. 관심분야는 신경망, 종분화, 패턴인식, 진화 알고리즘