

Kullback-Leibler 엔트로피를 이용한 종분화 신경망 결합의 성능향상

論 文

51D-4-4

Performance Improvement of Ensemble Speciated Neural Networks using Kullback-Leibler Entropy

金 京 中* · 趙 誠 培**

(Kyung-Joong Kim and Sung-Bae Cho)

Abstract - Fitness sharing that shares fitness if calculated distance between individuals is smaller than sharing radius is one of the representative speciation methods and can complement evolutionary algorithm which converges one solution. Recently, there are many researches on designing neural network architecture using evolutionary algorithm but most of them use only the fittest solution in the last generation. In this paper, we elaborate generating diverse neural networks using fitness sharing and combining them to compute outputs then, propose calculating distance between individuals using modified Kullback-Leibler entropy for improvement of fitness sharing performance. In the experiment of Australian credit card assessment, breast cancer, and diabetes in UCI database, proposed method performs better than not only simple average output or Pearson Correlation but also previous published methods.

Key Words : 진화 신경망, Kullback-Leibler 엔트로피, 적합도 공유, 다중 신경망

1. 서 론

신경망의 구조를 진화 알고리즘을 이용하여 설계하려는 연구가 다양하게 진행되고 있다[1,2,3]. 기존의 오류 역전파 알고리즘을 이용한 신경망의 구조 설계는 사전에 노드의 수를 결정해 놓아야 하고, 노드간의 가중치를 탐색하기 위해서 여러 함수가 미분 가능하여야 한다는 문제점이 있다. 반면에, 진화 알고리즘을 이용하여 신경망의 구조를 탐색하는 경우 노드의 수를 사전에 결정하지 않아도 되고 목적함수가 미분 가능하지 않아도 된다.

최근, 진화 알고리즘의 하나인 진화 프로그래밍(evolutionary programming)을 사용한 Yao의 EPNET은 다양한 벤치마크 데이터에 대해 좋은 성능을 내고 있다[4]. 하지만, 기존의 연구는 병렬적으로 탐색한 마지막 세대의 다양한 해중에서 가장 높은 적합도를 가지는 개체만을 탐색하는 단점이 있다. 이러한 단점을 해결하기 위해, 최종세대의 여러 개의 해를 종합적으로 이용하는 방법이 연구되고 있다. 다양한 신경망 개체들을 생성하고 결합을 수행한다면 더 높은 일반화 성능을 지닌 인식을 생성할 수 있다.

그러나, 진화 알고리즘은 현실세계의 다양한 문제에 적용이 가능한 탐색기법으로 활용도가 높은 반면, 일반적으로 하나의 해로 수렴하는 해솔림(genetic drift)의 단점이 있어서 다양한 해가 존재하는 문제에 대해서는 최대한의 성능을 발휘할 수

없다. 이러한 문제를 해결하기 위해 종분화(speciation)와 같은 기법들이 등장하였다[5].

종분화는 해의 분화를 촉진하여 진화 알고리즘의 마지막 세대에 서로 다른 다양한 해가 살아남도록 하는 기법으로, 적합도 공유와 크라우딩이 대표적인 방법이다. 적합도 공유 기법은 공유반경안에 존재하는 해들의 적합도를 상호 공유하도록 하여, 밀집도가 높은 지역에 위치하는 개체들의 생존율을 낮히고 밀집도가 낮은 지역에 위치하는 개체들의 생존율을 높이는 방법이다[6]. 크라우딩은 DeJong에 의해 1975년에 제안되었으며 "crowding factor model"로도 불린다[7].

기존연구에서는 진화 알고리즘을 이용하여 신경망 개체를 탐색하고, 적합도 공유 방법을 적용하여 해의 분화를 촉진한다[8]. 마지막 세대에는 해솔림(genetic drift)이 일어나지 않고, 서로 다른 성격의 개체들이 군집을 이루게 된다. 클러스터링 방법을 이용하여 종을 분류하고, 각 종의 대표되는 개체를 택하여 결합 신경망을 구성한다. 그림 1은 제안하는 방법을 설명해 주고 있다. 결합 신경망은 다양한 UCI 벤치마크 데이터에 대한 테스트를 수행하여 성능을 평가한다.

적합도 공유 알고리즘은 두 개체의 유사도를 평가하기 위한 방법을 필요로 하지만, 일반적으로 두 개의 신경망의 유사도를 평가하기 위한 명확한 방법은 존재하지 않는다. 각 신경망의 출력노드의 출력값을 평균하여 유클리디안 거리를 계산하는 출력값 평균이 쉽게 이용되는 방법이고, Pearson Correlation 기법도 사용된다. 본 논문에서는 정보 이론에 기반한 수정된 Kullback-Leibler 엔트로피를 이용하여 두 신경망 사이의 유사도를 측정한다[9]. 수정된 Kullback-Leibler 엔트로피를 유사도 측정기준으로 사용했을 때, UCI 벤치마크 데이터에 대하여 출력값 평균이나 Pearson Correlation을 사용했을 때보다 유사하거나 더 나은 성능을 얻을 수 있었다.

* 準 會 員 : 延世大學校 컴퓨터科學科 博士課程

** 正 會 員 : 延世大學校 컴퓨터科學科 副教授

接受日: 2001年 10月 16日

最終完了: 2002年 2月 8日

본 논문의 구성은 다음과 같다. 2장에서는 본 연구의 배경이 되는 다양한 연구들에 대해 서술한다. 3장에서는 적합도 공유 방법을 이용하여 신경망을 진화 시키는 방법에 대해 설명한다. 4장에서는 호주 신용카드 데이터, 유방암 데이터, 당뇨병 데이터 등의 UCI 벤치마크 데이터에 대한 실험결과를 서술한다. 5장에서는 실험결과를 바탕으로 Kullback-Leibler 엔트로피 유사도 기준의 우수성에 대해 결론을 내린다.

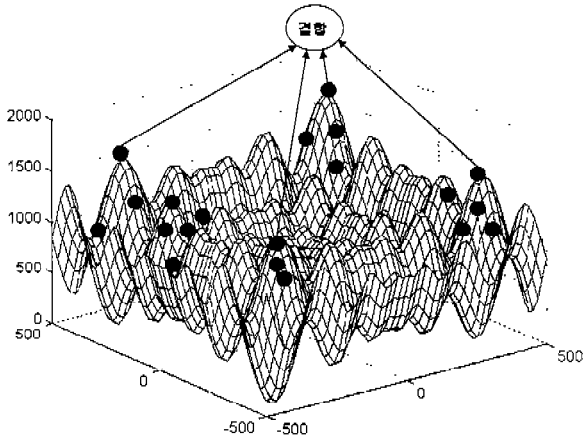


그림 1 제안하는 종분화 신경망 결합 모델
Fig. 1 Proposed ensemble model of speciated neural networks.

2. 관련연구

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 정보 등을 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 학습시 결정해야 하는 인자들을 진화를 통해 찾아서 최적의 신경망을 결정한다.

Harp 등은 학습 알고리즘의 인자를 결정하는데 진화적 접근 방식을 이용했다[10]. 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[11]. 또한 유전자 알고리즘과 역전파 알고리즘을 함께 사용한 경우

가 좀더 효과적인 학습이 가능함을 보였다.

Castillo 등은 G-Prop-II라는 진화 신경망 시스템을 제안했다[12]. 이 시스템은 유전자 알고리즘과 역전파 알고리즘을 사용해서 다중 퍼셉트론을 학습시킨다. 유전자 알고리즘으로 은닉층 크기, 초기 가중치 값, 학습 인자를 모두 다중 퍼셉트론의 인자 공간에서 탐색해서 최적화된 다중 퍼셉트론을 찾는다.

Yao와 Liu는 EPNet이라는 진화 신경망을 제안하였다[4]. EPNet은 Fogel이 제안한 진화 프로그래밍(evolutionary programming)을 이용해서 신경망을 결정한다[4]. 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 보존시키도록 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[13]. 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[14]. 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[15].

추상 레벨은 각 신경망이 하나의 결과만을 출력하는 방식으로 대표적인 방법으로는 투표 방법, 베이지안 방법[16], BKS 방법[17] 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 Borda 함수, Condorect 함수를 통한 결합이 순위 레벨에 속한다. 측정치 레벨의 경우, 각 인식기가 클래스들에 대한 순위와 더불어 각 순위에 대한 신뢰값까지 함께 출력하는 방식이기 때문에, 가장 많은 정보로 결합을 수행한다. 그러나, 각 인식기 별로 제공되는 신뢰값이 일관성을 유지해야 한다는 어려움이 있다. 대표적인 방법으로 평균[18]과 가중 평균 방법이 있다. 그 이외에도 학습 데이터에 대한 정보를 테스트 데이터에 적용해서 가장 적합한 신경망을 선택하는 게이팅 방법이 있다.

2.3 출력값 평균과 Pearson Correlation 방법

적합도 공유방법을 이용할 경우, 개체간의 유사도를 평가하기 위한 방법이 필요하며 출력값 평균과 Pearson Correlation 방법이 손쉽게 이용될 수 있다. 출력값 평균은 학습 데이터에 대한 각 신경망의 출력값의 평균을 계산해서 각 신경망간의 유사도를 비교한다. 입력 데이터 t에 대한 신경망 i의 출력값 o_i^t 는 다음과 같이 정의할 수 있다.

$$o_i^t = (o_{i1}^t, o_{i2}^t, \dots, o_{iN}^t)$$

이 때 N은 신경망 i의 출력 노드 수이다.

학습 데이터가 총 T개 있다면 학습 데이터에 대한 신경망 i의 j번째 출력 노드의 출력값 평균 \bar{o}_{ij} 는 다음과 같이 계산된다.

$$\bar{o}_{ij} = \frac{\sum_{t=1}^T o_{ij}^t}{T}$$

이 때 신경망 i의 출력값 평균 \bar{o}_i 는 다음과 같이 구한다.

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

신경망 i, j 간의 유사도는 출력값 평균의 유클리디언 거리를 측정했다. 신경망 i와 j의 유사도는 다음과 같이 구한다.

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

두 신경망이 유사할 수록 dij의 값은 작아진다.

두 변수간의 관계를 보여주는데 사용되는 Pearson Correlation을 신경망의 유사도를 측정하는데 적용했다. 신경망 a와 u의 유사도는 다음과 같이 계산할 수 있다.

$$W_{au} = \frac{\sum_{i=1}^m (r_{ai} - \bar{r}_a) * (r_{ui} - \bar{r}_u)}{\sqrt{\sum_{i=1}^m (r_{ai} - \bar{r}_a)^2 * \sum_{i=1}^m (r_{ui} - \bar{r}_u)^2}}$$

이 때, r_{ai} 는 신경망 a의 i번째, \bar{r}_a 는 신경망 a의 출력값 평균이다.

3. 종분화된 신경망의 진화

제안하는 종분화된 진화 신경망은 다음과 같은 알고리즘으로 생성된다(그림 2 참조). 종분화를 위해 적합도 공유방법이 이용되며 두 신경망의 유사도를 평가하기 위해 정보이론에 기반한 Kullback-Leibler 엔트로피 기법이 이용된다.

- 단계 1 (초기 신경망 집단 생성) : 임의의 가중치로 초기화한 신경망의 집단을 생성한다.
- 단계 2 (각 신경망의 부분 학습) : 각 신경망을 역전과 알고리즘으로 부분 학습시킨다.

- 단계 3 (종분화를 적용한 적합도 값 계산) : 검증 데이터로 테스트하여 각 신경망의 적합도를 적합도 공유를 적용하여 계산한다.
- 단계 4 (순위 기반 선택) : 적합도가 구해지면 인식률이 높은 상위 50% 개체를 선택한다.
- 단계 5 (교차 및 돌연변이 적용) : 신경망 집단에 교차와 돌연변이가 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 종들로부터 대표 신경망을 선택한다.
- 단계 10 (신경망 결과 결합) : 테스트 데이터에 대한 대표 신경망의 인식 결과를 각 레벨의 적절한 방법으로 결합한다.

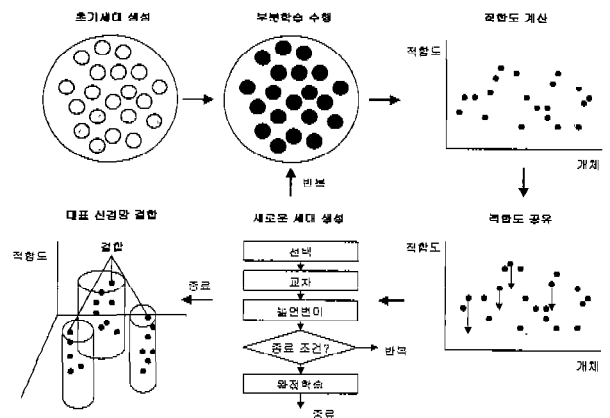


그림 2 종분화된 신경망 결합을 위한 전체 알고리즘
Fig. 2 Algorithm for evolutionary neural network speciation and ensembles.

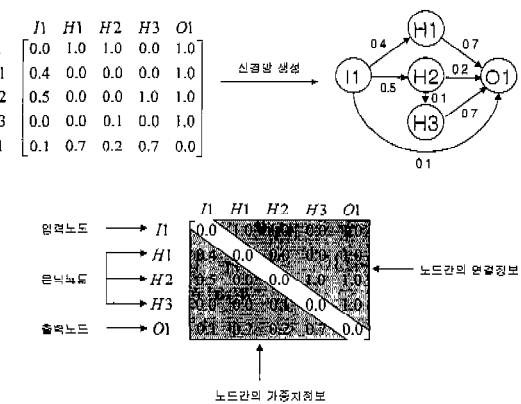


그림 3 행렬을 이용한 신경망의 표현
Fig. 3 Neural network representation using matrix.

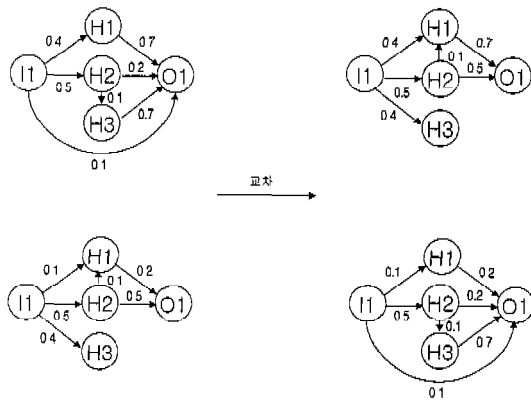


그림 4 은닉노드 H2를 기준으로 수행한 교차연산
Fig. 4 Crossover operator.

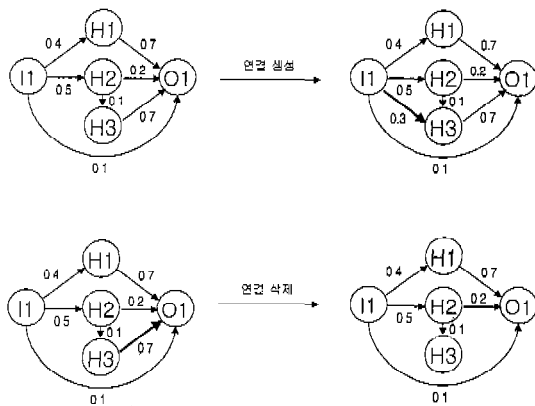


그림 5 연결생성 또는 연결삭제를 수행한 돌연변이 연산
Fig. 5 Mutation operator.

3.1 신경망의 진화

신경망을 진화시키기 위해서는 이를 진화에 적합한 형태로 인코딩하여야 한다. 신경망 인코딩은 유전자 알고리즘에서 개체를 표현하는데 일반적으로 사용하는 이진 표현, 트리, 링크트 리스트, 행렬 등의 방법이 있다. 본 논문에서는 간단하면서도 진화 인산자를 적용하기 쉬운 행렬을 사용한다. N개의 노드를 가진 신경망은 N×N 크기의 정방행렬에 연결 정보와 가중치를 표시해서 인코딩한다. 행렬의 우상단에는 노드간 연결 정보를 1과 0으로 표시하고, 각 연결 정보에 대칭하는 좌하단에는 가중치를 나타낸다. 그림 3은 전체 노드 수가 다섯 개인 신경망과 해당 유전자형의 예를 보여준다.

3.1.1 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 4는 은닉노드 H2를 교차점으로 한 교차 연산의 예를 보여준다. 이 교차 연산으로 H2 노드 이후에 있는 모든 은닉 노드의 연결 정보와 가중치 정보가 교환된다.

3.1.2 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 연결이 없는 경우 새로운 연결을 생성한 후 임의의 값으로 연결 강도를 설정한다. 그림 5는 돌연변이 연산의 예를 보여준다.

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 찾도록 한다.

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

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

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 6은 적합도 공유의 예를 보여주는데, 개체 A는 이와 유사한 3개의 개체와 적합도를 공유한다. 반면, B는 이와 유사한 7개의 개체와 적합도를 공유하고 C는 다른 개체와 적합도를 공유하지 않는다. 상대적으로 B의 적합도가 A보다 크게 감소하게 된다.

공유 반경은 두 개체가 적합도를 공유하는지를 결정하는 유사도의 기준값이다. 일정한 차이가 나는 다중 해를 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.

종분화된 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) 분석을 사용했다[19]. 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.

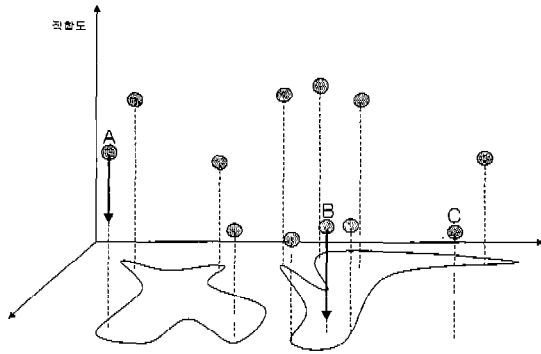


그림 6. 적합도 공유의 예 (A와 B는 공유반경 이내의 개체들과 적합도를 공유)
Fig. 6 Example of fitness sharing.

3.3 Kullback-Leibler 엔트로피를 이용한 신경망의 종분화

신경망의 적합도로 검증 데이터의 인식률을 사용하였다. 적합도 공유를 적용하는데 있어서 문제가 되는 것은 개체간의 유사성을 어떻게 비교할 것인가이다. 개체간의 유사도를 평가하기 위해 수정된 Kullback-Leibler 엔트로피 방법이 사용되었으며 비교를 위해 출력값 평균 방법, Pearson Correlation 방법이 이용되었다.

Kullback-Leibler 엔트로피 방법은 다양한 응용분야에서 적용되고 있는 거리 계산방법으로 정보이론에 기반하고 있다 [20,21,22]. 신경망의 출력은 단순히 0과 1사이의 임의의 값이 아니라, 분류기의 베이시안 사후 확률값이라 할 수 있다. 이 성질을 이용하면 수정된 Kullback-Leibler 엔트로피값을 사용해서 두 신경망의 차이를 측정할 수 있다. 이 엔트로피값은 두 분포 p와 q 사이의 거리를 다음과 같이 정의한다.

$$D(p, q) = \sum_{i=1}^m p_i \log \frac{p_i}{q_i}$$

여기서 m은 신경망의 출력 노드의 수를 나타낸다.

하지만, 이 엔트로피값은 대칭성이 없기 때문에 $(D(p, q) \neq D(q, p))$ 실제 거리를 나타내지 못한다. 이 문제를 해결하기 위해 대칭적인 엔트로피를 다음과 같이 정의할 수 있다.

$$D(p, q) = \frac{1}{2} \sum_{i=1}^m (p_i \log \frac{p_i}{q_i} + q_i \log \frac{q_i}{p_i})$$

p와 q를 두 신경망의 출력값 확률 분포라고 하고, 각 신경망이 m개의 출력 노드를 가지고, n개의 데이터로 학습된다고 할 때, 유사도는 다음과 같이 측정될 수 있다.

$$D(p, q) = \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^n (p_{ij} \log \frac{p_{ij}}{q_{ij}} + q_{ij} \log \frac{q_{ij}}{p_{ij}})$$

이 때 p_{ij} 는 j번째 학습 데이터에 대한 신경망의 i번째 출력값을 뜻한다. 두 신경망이 비슷할수록 엔트로피 값은 더 작아지게 된다. 그림 7은 서로 다른 형태의 확률분포를 가지는 P와 Q를 보여준다.

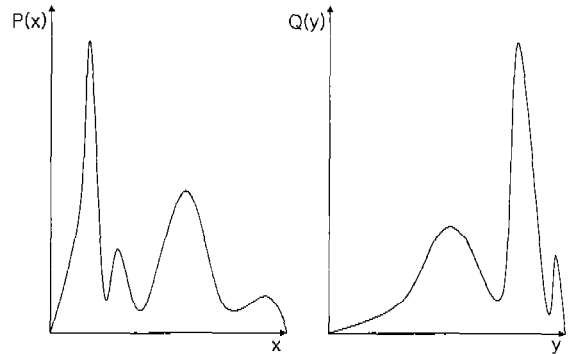


그림 7 두 가지 서로 다른 확률 분포
Fig. 7 Two different distributions of probability.

4. 실험결과

4.1 실험 환경

실험을 통해 신경망의 종분화를 분석하고 결합의 성능을 기존 연구 결과들과 비교해 보았다. 실험을 위하여 UCI의 벤치마크 데이터 중에서 호주 신용카드 승인(australian credit approval) 데이터를 사용했다 [23]. '+' , '-' 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.

유방암 데이터는 Wisconsin 대학병원의 W. H. Wolberg가 제공한 데이터로 본래 현미경으로 관측한 세포 정보에 기반해서 양성과 악성을 분류할 목적으로 제작되었다 [23]. 총 데이터 수는 699개이며 각 데이터는 9개의 속성을 가지고 있다. 총 데이터 중 458개는 양성, 241개는 악성 클래스이다. 학습 데이터, 검증 데이터, 테스트 데이터는 각각 349개, 175개, 175개씩 사용하였다. 당뇨병 데이터는 총 768개로 이 중 500개는 클래스 1, 268개는 클래스 2이다 [23]. 학습 데이터, 검증 데이터, 테스트 데이터는 각각 384개, 192개, 192개씩 사용하였다.

각 세대의 신경망 수는 20개이고 총 200세대를 진화시켰다. 신경망의 형태는 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 반경은 실험을 통해 0.003의 값으로 정하였다. 실험 결과는 신경망 집단을 다르게 초기화해서 총 10회 수행한 것의 평균이다.

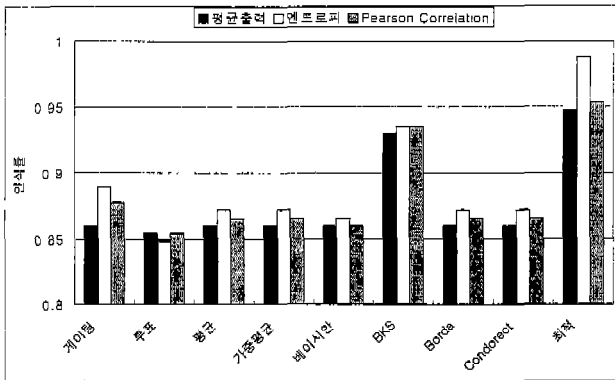


그림 8 종분화 신경망 결합 결과: 호주 신용카드 승인 데이터
Fig. 8 Recognition rates of speicated neural network ensembles: Australian credit card approval.

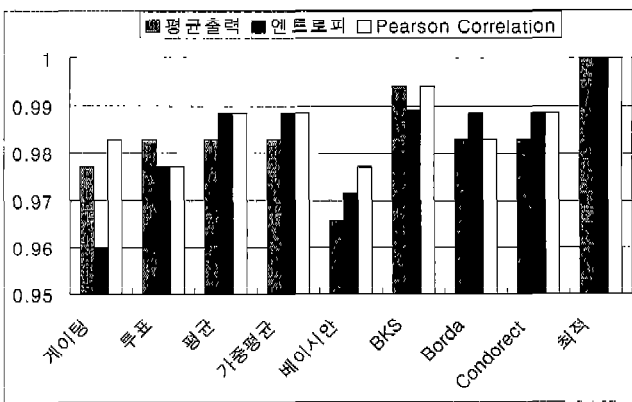


그림 9 종분화 신경망 결합 결과: 유방암 데이터
Fig. 9 Recognition rates of speicated neural network ensembles: Breast cancer).

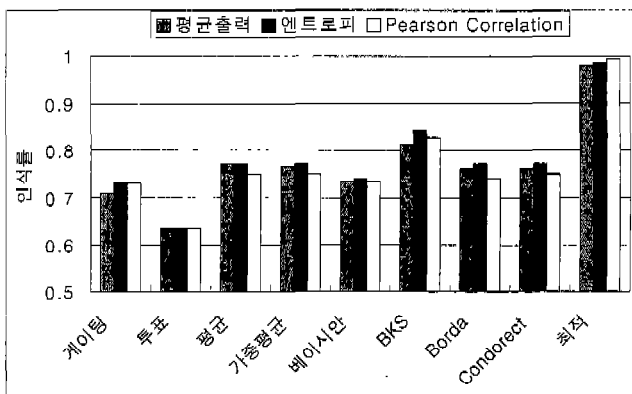


그림 10 종분화 신경망 결합 결과: 당뇨병 데이터
Fig. 10 Recognition rates of speicated neural network ensembles: Diabetes.

4.2 실험결과

제안한 Kullback-Leibler 엔트로피 방법이외에 출력값 평균 방법과 Pearson Correlation 방법이 사용된다. 그림 8, 9, 10은 Kullback-Leibler 엔트로피 방법, 평균출력력방법, Pearson Correlation 방법을 다양한 결합방법을 사용해서 비교한 결과이다. 최적 결합은 하나의 신경망이라도 올바른 결과를 내면 전

체 신경망이 올바른 결과를 낸 것으로 인정하며 비교를 위해 사용한다. 호주 신용카드 데이터와 당뇨병 데이터에 대해서 Kullback-Leibler 엔트로피 방법은 다른 두 가지 방법에 비해 향상된 성능을 보였다. 유방암 데이터에 대해서 Kullback-Leibler 엔트로피 방법은 일부분 향상된 성능을 보였으나 다른 두방법과 유사하거나 낮은 성능을 보이는 부분도 있었다.

4.3 결과 분석

종분화된 신경망을 분석하기위해 호주 신용카드 승인 데이터를 사용해서 실험을 수행하였다. 신경망 집단을 단일 연결 클러스터링(single linkage clustering) 분석법[19]을 사용해서 종을 구분했다. 그림 11과 12는 종분화를 사용하지 않은 경우와 사용한 경우의 단일 연결 클러스터링을 기법을 이용한 덴드로그램(dendrogram)을 나타낸다. 두 그래프 모두 호주 신용카드 승인 데이터에 대해서 유사도 비교 기준으로 수정된 Kullback-Leibler 엔트로피를 사용한 실험 중 임의로 선택한 결과이다. 종분화를 하지 않은 경우를 보면 클러스터간 최대 유사도 차이가 30 이상이 되지 않고 있지만, 종분화를 한 경우 최대치가 80을 넘으며 클러스터간 유사도의 차이 역시 더 큰 값을 보여주고 있다. 이는 종분화를 사용하였을 경우, 마지막 세대에 남은 신경망들이 특정 개체와 유사하지 않은 다양한 신경망들이 생성되었다고 볼 수 있다.

4.4 기존 연구결과와의 비교

표 1은 호주 신용카드 데이터에 대한 BKS방법을 이용한 종분화 신경망 결합 결과와 기존 연구와의 비교이다. 표 2는 기존의 연구 결과 중에 대표적인 FNNCA[24]와 HDANNS, EPNet의 결과를 비교하였다. FNNCA는 Setino와 Hui가 제안한 컨스트럭티브 알고리즘으로 구축한 신경망이고, HDANNS[25]는 실행과 오류 과정을 거쳐 사람이 직접 설계한 최적화 신경망이다. 그리고 EPNet[4]은 Yao와 Liu가 제안한 진화 신경망 알고리즘의 결과이다. FNNCA나 EPNet보다는 더 좋은 성능을 보였으나 HDANNS 방식에 비해서는 약간 안 좋은 성능을 보였다. 하지만 HDANNS 방식이 많은 비용을 들여서 찾아낸 최적의 해임을 감안한다면 제안한 시스템의 결과는 만족할 만한 수준이라고 할 수 있겠다. 표 3은 당뇨병 데이터에 대한 기존 연구결과와의 비교이다.

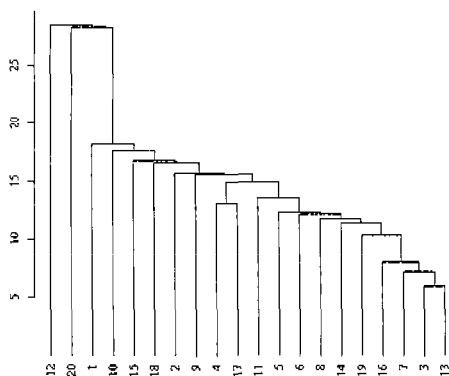


그림 11 종분화하지 않은 경우의 덴드로그램
Fig. 11 Dendrogram of non-speicated case.

5. 결 론

본 논문에서는 정보이론에 기반한 수정된 Kullback-Leibler 엔트로피를 사용하여 신경망개체간의 유사도를 평가하고 이를 바탕으로 한 적합도 공유방법으로 다양한 신경망개체를 생성한 후 결합하였다. 중분화된 신경망의 결합은 추상 레벨, 순위 레벨, 그리고 측정치 레벨의 다양한 방법을 적용해 보았으며 기존의 출력값 평균과 Pearson Correlation 방법을 사용한 유사도 평가와 비교하여 다양한 UCI 벤치마크 데이터에 대해 향상된 성능을 얻을 수 있었다. 향후 연구에서는 신경망 간의 유사도에 대한 연구를 보충하고 중분화된 신경망에 대한 깊이 있는 분석을 시도할 것이다.

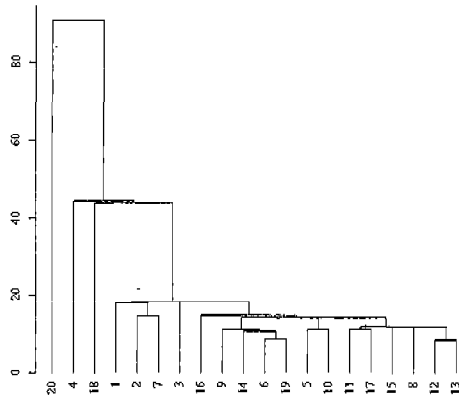


그림 12 중분화한 경우의 덴드로그램
Fig. 12 Dendrogram of speciated case.

표 2 기존 연구 결과와의 비교 : 호주 신용카드 데이터
Table 2 Comparison of recognition rates: Australian credit card approval.

	평균출력	P. Correlation	엔트로피
오류율	0.089	0.092	0.096
	EPNet	Cal5	ITrule
오류율	0.115	0.131	0.137
	DIPOL92	Discrim	Logdisc
오류율	0.141	0.141	0.141
	CART	RBF	CASTLE
오류율	0.145	0.145	0.148
	NaiveBay	IndCART	BP
오류율	0.151	0.152	0.154

표 3 기존 연구 결과와의 비교 : 유방암 데이터
Table 3 Comparison of recognition rates: Breast cancer.

	엔트로피	평균출력	P. Correlation
오류율	0.0118	0.0125	0.0125
	HDANNS	EPNet	FNNCA
오류율	0.01149	0.01376	0.0145

표 4 기존 연구 결과와의 비교 : 당뇨병 데이터
Table 4 Comparison of recognition rates: Diabetes.

	엔트로피	P. Correlation	평균출력
오류율	0.202	0.205	0.218
	Logdisc	EPNet	DIPOL92
오류율	0.223	0.224	0.224
	Discrim	SMART	RBF
오류율	0.225	0.232	0.243
	ITrule	BP	Cal5
오류율	0.245	0.248	0.25
	CART	CASTLE	Quadisc
오류율	0.255	0.258	0.262

감사의 글

이 연구는 과학기술부가 지원하는 뇌과학연구 프로그램에 의하여 지원 받은 것임.

참 고 문 헌

- [1] D. Montana and L. Davis, "Training feedforward neural networks using genetic algorithms," Proc. of Eleventh Int'l Joint Conf. on Artificial Intelligence, pp. 762-767, San Mateo, CA, 1989.
- [2] D. B. Fogel, L. J. Fogel, and V. W. Porto, "Evolving neural networks," Biological Cybernetics, Vol. 63, pp. 487-493, 1990.
- [3] X. Yao, "Evolving artificial neural networks," Proceedings of the IEEE, vol. 87, no. 9, pp. 1423-1447, September 1999.
- [4] 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.
- [5] T. Back, D. B. Fogel, and Z. Michalewicz, Handbook of Evolutionary Computation, IOP Publishing and Oxford University Press, New York and Bristol (UK), 1997.
- [6] D. E. Goldberg, Genetic Algorithms in Search, Optimization, and Machine Learning, Addison-Wesley, Reading Massachusetts, 1989.
- [7] K. A. De Jong, "An analysis of the behavior of a class of genetic adaptive systems," Doctorial dissertation, University of Michigan, 1975.
- [8] J.-H. Ahn and S.-B. Cho, "Speciated neural networks evolved with fitness sharing technique," Proceedings of Congress on Evolutionary Computation, Vol. 1, pp. 390-396, May 2001.
- [9] S. Kullback and R. A. Leibler, "On information and sufficiency," Ann. Math. Stat., 22, pp. 79-86, 1951.
- [10] 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.

[11] 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.

[12] 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.

[13] A. J. C. Sharkey, "On combining artificial neural nets," Connection Science, Vol. 8, pp. 299-313, 1996.

[14] 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.

[15] 백종현, 다중 인식기의 다단계 결합을 통한 무제약 필기숫자 인식, 연세대학교 대학원 박사학위 논문, 1996.

[16] 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.

[17] A. Khotanzad, and C. Chung, "Hand written digit recognition using BKS combination of neural network classifiers," Proceedings of the IEEE Southwest Symposium on Image Analysis and Interpretation, pp. 94-99, 1994.

[18] 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.

[19] A. D. Gordon, Classification: Methods for the Exploratory Analysis of Multivariate Data, Chapman and Hall, 1981.

[20] K. Viele and C. Srinivasan, "Parsimonious estimation of multiplicative interaction in analysis of variance using Kullback-Leibler information," Journal of Statistical Planning and Inference, Vol. 84, pp. 201-219, 2000.

[21] J. A. Garcia, J. Fdez-Valdivia, X. R. Fdez-Vidal, and R. Rodriguez-Sanchez, "Information theoretic measure for visual target distinctness," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 23, pp. 362-383, April 2001.

[22] M. N. Do and M. Vetterli, "Texture similarity measurement using Kullback-Leibler distance on wavelet subbands," Proc. IEEE Int. Conf. on Image Proc. (ICIP), Vancouver, Canada, Sep. 2000.

[23] UCI Machine Learning Repository, <http://www1.ics.uci.edu/~mllearn/MLRepository.html>

[24] R. Setino and L. C. K. Hui, "Use of a quasi-newton method in a feedforward neural network construction

algorithm," IEEE Trans. on Neural Networks, Vol. 6, no. 1, pp. 273-277, 1995.

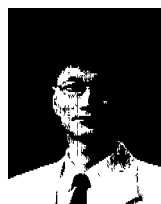
[25] L. Prechelt, "Proben1 - a set of neural network benchmark problems and benchmarking rules," Tech. Rep. 21/94, Fakultat fur Informatik, Universitat Karlsruhe, 76128, Germany, 1994.

저 자 소 개



김경중 (金京中)

1996.3-2000.2 연세대학교 컴퓨터과학학사
 2000.3-2002.2 연세대학교 컴퓨터과학석사
 2002.3-현재 연세대학교 컴퓨터과학과박사
 과정 재학중
 전화 : 02-2123-3877
 팩스 : 02-365-2579
 E-mail : uribyl@candy.yonsei.ac.kr



조성배 (趙誠培)

1984.3-1988.2 연세대학교 전산과학 학사
 1988.3-1990.2 한국과학기술원 전산학 석사
 1990.3-1993.8 한국과학기술원 전산학 박사
 1993.10-1995.4 ATR 인간정보통신연구소
 객원연구원
 1995.9-현재 연세대학교 컴퓨터과학과 교수
 전화 : 02-2123-2720
 팩스 : 02-365-2579
 E-mail : sbcho@cs.yonsei.ac.kr