

유전알고리즘을 이용한 신경망의 구성 및 다양한 학습 알고리즘을 이용한 신경망의 학습

Constructing Neural Networks Using Genetic Algorithm and Learning Neural Networks Using Various Learning Algorithms

양 영 순¹ 한 상 민²
Yang, Young-Soon Han, Sang-Min

Abstract

Although artificial neural network based on backpropagation algorithm is an excellent system simulator, it has still unsolved problems of its structure-decision and learning method. That is, we cannot find a general approach to decide the structure of the neural network and cannot train it satisfactorily because of the local optimum point which it frequently falls into. In addition, although there are many successful applications using backpropagation learning algorithm, there are few efforts to improve the learning algorithm itself. In this study, we suggest a general way to construct the hidden layer of the neural network using binary genetic algorithm and also propose the various learning methods by which the global minimum value of the learning error can be obtained. A XOR problem and line heating problems are investigated as examples.

1. 서론

근래에 공학의 넓은 분야에서 인공신경망(artificial neural network, ANN) 기법이 도입되고, 그에 따른 많은 성과가 나타나고 있는 것은 인공신경망이 갖고 있는 많은 장점에 기인한다고 볼 수 있다. 즉, 인공신경망은 기존의 노이만 방식과 달리 각 연산요소(신경세포)의 병렬처리로 빠른 연산을 수행할 수 있고, 정보의 분산표현 방식으로 잡음, 파손 등에 강한 시스템을 구축할 수 있으며, 생물학적인 뇌와 비슷한 원리로 동작하여 기계학습을 구현하는데 용이하다. 기존의 인공지능 분야와 비교해볼 때, 이러한 학습의 구현이라는 측면에서 인공신경망은 다른 여타의 인공지능 분야보다 뛰어난 강점을 지닌다고 할 수 있다. 인공신경망 특히, 역전파 모델에서 구현된 기계학습은 인간이 지능을 통해 어떤 것을 학습하는 것과 매우 흡사하다. 즉, 인간은 어떤 사실들을 학습하여 자신의 지식으로 축적시킨 뒤 학습된 사실들과 유사한 것에 대해 학습된 것과 비교 검토하여 답을 구할 수 있는 능력을 가지고 있는데, 역전파 모델도 이와 같은 방식으로 학습을 수행할 수 있다. 이러한 것은 넓은 범위에서 볼 때 시스템 매핑(system mapping) 또는 시스템 시뮬레이션(system simulation)의 범주로 이해될 수 있는데, 이러한 특성이 역전파 모델을 공학적으로 성공하게 만든 이유라고 할 수 있다.

본 연구에서도 역전파 모델을 시스템 매핑의 도구(tool)로 간주하여 기존의 역전파 모델이 가지고 있는 장

1. 정회원 · 서울대학교 조선해양공학과 교수
2. 서울대학교 조선해양공학과 박사과정

단점을 평가한 뒤, 단점을 해결하기 위한 방안을 제시하는 것으로 중심과제를 삼을 것이다. 역전파 모델이 갖고 있는 문제점은 크게 두 가지로 구분될 수 있는데, 첫째는 신경망의 구조가 학습성능에 매우 큰 영향을 미치고 있음에도 불구하고 아직까지 신경망 구조 결정을 위한 일반적인 접근 방법이 없어서 시행착오 방법을 사용한다는 것이고, 둘째는 역전파 모델이 사용하는 학습규칙인 델타규칙(delta rule)이 신경망의 학습을 항상 보장하지는 못한다는 것이다(즉, 국소 최적점에 빠진다). 델타규칙을 일반적인 최적화 기법의 관점에서 보면 신경망의 에러를 최소화하기 위해 최대 경사법(steepest descent method)을 이용하는 규칙으로 볼 수 있는데, 신경망 분야에서 델타규칙에 의한 역전파 신경망 모델이 제안된 이후에 신경망을 이용한 다수의 성과들이 나타났지만 학습 알고리즘인 델타규칙 자체를 개선하려는 노력은 별로 없는 것으로 보인다. 최대 경사법을 사용하는 델타규칙은 신경망의 학습을 항상 보장하지는 못하므로, 이를 극복하여 보다 좋은 학습 성취도를 얻을 수 있는 학습 알고리즘을 개발하는 것은 매우 필요한 일이라고 볼 수 있다. 그러므로, 본 연구에서는 이진수형 유전알고리즘의 이산적 성질을 이용하여 신경망의 구조를 최적화하는데 있어서의 일반적인 방법을 제시하고, 최적화된 구조에서 신경망에 대한 미세 학습을 수행하기 위해 여러 가지의 학습 알고리즘을 개발하여 이를 적용하는 것에 주안점을 두었다. 예제로는 신경망 문제에 자주 등장하는 대표적인 비선형 문제인 XOR 문제와 보다 실제적인 예제로써 선상가열(line heating)문제를 적용하였다.

2. 순차신경망

2.1. 순차신경망의 개요

순차신경망(feed-forward neural network)은 입력과 출력의 정보흐름이 입력층(input layer), 은닉층(hidden layer), 출력층(output layer)의 방향으로 일정하게 흐르는 형태의 신경망을 말하는데, 정보 흐름의 측면에서 회귀 신경망(recurrent neural network)과는 대조가 된다. 결국 순차신경망은 어떤 입력에 대한 출력을 내는 형태로 동작하는데, 이러한 측면 때문에 순차신경망은 시스템 매핑(system mapper)로 자주 사용된다. 순차신경망은 입력층, 은닉층, 출력층의 신경세포(neuron)들과 그들간의 연결(connection)들로 이루어져 있는데, 각 신경세포는 바로 이전에 있는 모든 신경세포들의 활성화값(activation value)에 대해 가중합을 구하고 여기에 자신의 활성화함수를 취하여 바로 다음에 연결되어 있는 신경세포에게 활성화값으로 넘겨준다. 또, 각 연결들은 자신의 연결강도(connection weight)를 가지고 있어서 연결의 처음 끝단에 하나씩 붙어 있는 신경세포의 활성화값과 연결강도가 곱해지면서 신경세포의 가중합이 구해진다.

2.2. 순차신경망의 문제점

2.2.1. 신경망의 구조 결정방법

신경망의 학습은 학습방법 뿐만 아니라 신경망의 구조와도 밀접한 관련을 가지고 있으므로, 주어진 시스템을 시뮬레이션하기 위해 어떤 구조의 신경망을 채택하느냐에 따라 소기의 성과를 거둘 수도 있고 그렇지 못할 수도 있다. 순차신경망인 역전파 인공신경망은 입력층과 출력층이 결정되어 있으므로 신경망의 구조란 은닉층의 구조를 의미한다. 은닉층에 신경세포가 부족하면 신경망의 학습이 수행되지 못하고(에러가 줄지 못한다), 은닉층의 신경세포가 최적의 개수보다 많으면 학습은 가능하나 학습을 위한 시간(신경망의 학습은 본래 시간이 많이 소비되는 일)이므로 신경망의 학습시간을 줄이기 위한 연구도 활발히 수행되고 있는 실정이다)이 많이 소비되는 단점이 있게 된다. 기존에는 적절한 은닉층의 구조를 결정하기 위해 은닉층의 신경세포 수를 변화시켜 가면서 학습실험을 하는 시행착오 방법이 사용되어 왔고, 그 결과 어떤 특정한 문제에 적합한

신경망의 구조들이 제시되어 왔으나, 다른 문제에 대해서는 또 다른 시행착오가 필요하므로 신경망 구조 결정을 위한 일반적인 방법론이 필요하다고 할 수 있다. 본 연구에서는 신경망의 구조를 최적화하기 위해서 이진수형 유전알고리즘을 도입하였는데, 이진수형 유전알고리즘은 이산화된 변수를 다루는데 있어서 좋은 성능을 발휘하므로 topology 최적화와 같은 문제에 잘 적용이 될 수 있다. 신경망의 구조 최적화도 결국 topology 최적화의 관점에서 문제의 접근이 가능하다고 할 수 있다.

2.2.2. 신경망의 연결강도 결정방법(학습방법)

순차신경망에서의 학습은 목적함수가 신경망의 에러이고 설계변수가 신경망의 연결강도인 비제약 최적화 문제(unconstrained optimization problem)로 볼 수 있다. 기존에는 신경망의 에러를 최소화하여 학습을 달성할 수 있는 방법으로써 결정론적(deterministic) 최적화 기법(초기 설계점에 따라 최적점이 하나로 결정되는 방법)이 이용되었다. 즉, 역전파 신경망 모델이 사용하는 학습방법 중 하나인 일반화된 델타규칙(generalized delta rule)은 최적화 기법으로써 결정론적 방법인 최대 경사법(steepest descent method)을 사용한 것이라고 볼 수 있다. 순차신경망을 학습시키는 방법은 결정론적 최적화 기법 외에 스토캐스틱 최적화 기법(stochastic optimization technique)이 있는데, 이는 결정론적 방법처럼 설계점의 탐색방향(search direction) 및 이동길이(step length)가 현재의 설계점 위치에 따라 결정되는 것이 아니고 어떤 조건에 의해 임의(random)로 선택될 수 있는 방법을 말한다. 일반적으로 스토캐스틱 최적화 기법은 넓은 영역의 설계영역을 다룰 수 있으므로 전역 최적해를 구하는 면에서 결정론적 방법보다 월등히 뛰어나다고 할 수 있는 반면에, 많은 양의 함수값 평가를 수행해야 하므로 계산 시간이 많이 걸린다는 약점이 있다. 특히, 설계점이 최적점 부근에 도달했을 때 더 이상 스토캐스틱 방법은 효율적이지 못한 탐색 방법이 되는데, 실제로 대부분의 스토캐스틱 탐색 방법은 국부 탐색(local search)에 매우 수렴 속도가 느린 것으로 알려져 있다. 이와는 대조적으로 결정론적 최적화 기법은 현재의 위치에서 탐색을 수행할 방향 및 이동거리를 항상 알고 있으므로 빠른 시간 내에 탐색을 수행할 수 있는 장점이 있다. 그러나, 결정론적 최적화 기법은 국소 최적점에 빠질 수 있는 가능성을 항상 가지고 있으므로, 이 방법이 순차신경망의 학습규칙으로 사용된다면 어떠한 문제에 대해서는 학습이 수행되지 못할 수 있다는 위험요소를 내포하고 있는 것으로 볼 수 있다. 결국 전역 최적점을 찾으면서 시간 비용의 측면에서도 바람직한 학습방법은 이와 같은 두 가지의 최적화 기법을 적절히 연결시킨 스토캐스틱-결정론적 하이브리드 방법(stochastic-deterministic hybrid learning method)이라고 할 수 있다. 즉, 스토캐스틱 방법으로 전역 최적해가 존재하는 부근까지 탐색을 수행한 뒤 결정론적 방법으로 국부 탐색을 효율적으로 수행하는 학습 모델이 가장 바람직하다고 할 수 있다. 본 연구에서 신경망의 구조를 결정하기 위해 이진수형 유전알고리즘을 이용하였다고 언급하였는데, 이진수형 유전알고리즘은 신경망의 구조 결정뿐만 아니라 전역 최적해 부근까지 탐색을 수행해 주는 전역적 탐색능력도 매우 우수하므로 하이브리드 방법의 첫 번째 과정인 스토캐스틱 모델의 역할을 충분히 수행해 줄 수 있다.

3. 유전알고리즘

3.1. 유전알고리즘의 원리 및 적용가능성

유전알고리즘(genetic algorithm, GA)는 자연진화의 법칙인 적자생존(survival of fittest)과 자연도태(natural selection)의 원리를 근간으로 정립된 최적화 알고리즘이다. 자연진화의 법칙에 의하면 일정한 환경에 무리를 지어 살고있는 생명체들은 그 환경에 적합한 형질을 가진 개체(individual)가 생존할 확률을 크게 가지므로 다음 세대를 만들어내는 번식(reproduction)과정에 더욱 많이 참여할 수 있게 되고, 번식을 통해 두

부모 개체의 형질은 서로 섞여서 자손 개체에게 전해지게 되는데(crossover), 매우 드물게 자손 개체의 형질은 부모와 전혀 다른 것이 될 수도 있다(mutation). 이러한 과정을 통해서 개체의 형질은 보다 좋은 방향으로 진화해 나가고 부적합한 형질을 가진 개체들은 진화의 과정에서 점차 도태되어, 결국 주어진 환경에 가장 적합한 형질의 개체들이 절대다수를 형성하게 된다. GA는 이러한 진화 법칙의 artificial version으로서, 설계 영역에 다수의 설계점을 분포시켜 설계점들이 개체군(population)을 형성하게 한 뒤 목적함수 값과 제한조건의 위반 정도에 따라 각 설계점에 적합도(fitness)를 부여하고 적합도가 클수록 번식(reproduction)의 과정에 참여할 확률을 크게 하여 교배(crossover)와 돌연변이(mutation)가 일어날 수 있도록 한다. 이러한 방식으로 적합도가 큰 설계점에 비슷한 형질을 가진 설계점들이 다음 단계에서 보다 많이 형성되어 계산이 진행될수록 전체 설계점들은 좋은 방향으로 탐색이 수행된다. GA는 여러 개의 설계점들이 집단을 이루어 동시에 설계영역을 탐색하므로 설계점이 한 점에서 한 점으로 이동하는 기존의 국부적인 탐색방법보다 더 넓게 설계영역의 정보를 활용할 수 있어서 전역 최적점에 수렴할 확률이 매우 크다. 그러므로, 만일 GA가 신경망의 에러를 감소시키는 학습규칙으로 적용된다면 기존의 델타규칙보다 신경망의 학습 성취도를 크게 향상시켜줄 것이다. 또한, GA는 기존의 방법과는 달리 미분값이나 그 외의 다른 정보는 필요 없고 오직 목적함수와 제한조건의 값만을 사용하는 직접 탐색방법이므로 실세계에 존재하는 복잡하고 다양한 문제에 비교적 쉽게 적용이 가능하다. 그러므로, GA는 본 연구에서 다루고 있는 신경망의 학습규칙으로도 쉽게 적용될 수 있다.

개체의 형질을 어떻게 표현하는가에 따라 GA는 이진수형 유전알고리즘(binary genetic algorithm, BGA)과 실수형 유전알고리즘(real genetic algorithm, RGA)으로 나눌 수 있는데, BGA는 설계변수의 실제 값에 대응된 이진수 문자열(binary string)을 개체의 유전자형(genotype)인 염색체로 사용한다. 이러한 이진 문자열은 실제의 염색체에 유전인자가 나열되어 있는 것과 같은 형태이므로 염색체상의 교배와 돌연변이 과정을 표현하기가 매우 단순하고 편리하다. 또한 이러한 문자열은 이산적 성질을 가지고 있으므로 이산적 설계변수를 가지고 있는 최적화 문제와 시스템의 구조를 최적화하는 topology 최적화 문제에 효과적으로 사용할 수 있는데, 신경망의 구조를 최적화하는 경우도 BGA의 이러한 이산적 성질을 이용하면 좋은 성과를 얻을 수 있을 것이다. 한편, RGA는 설계변수의 값을 직접 사용한다. 즉, BGA가 개체의 표현형과 그것에 대응된 유전자형을 동시에 사용하는 반면, RGA는 오직 표현형만을 사용한다. 표현형만을 가지고 교배와 돌연변이를 수행해야 하므로 BGA와 같이 단순하게 구현되지는 않고 표현형에 대한 적절한 조작이 필요하다. 설계변수 값을 사용하므로 BGA에서처럼 유전자형 구현을 위한 추가의 메모리 공간이 필요 없고, 유전자형과 표현형의 매핑을 위한 계산이 필요 없어 계산시간을 상당히 줄여 줄 수 있다. 또한, BGA의 이산적 성질에 기인한 설계변수의 정밀도 문제를 근본적으로 해결함으로써 BGA의 약점인 국부 탐색 능력을 향상시켜 준다.

3.2. 이진수형 유전알고리즘을 이용한 은닉층 최적화

3.2.1. 유전자형과 표현형

신경망 구조 최적화의 관점에서 볼 때 BGA의 표현형은 각 신경세포들이 어떤 값의 연결강도로 연결되어 망을 구성하고 있는 신경망의 구조 자체를 나타내야 한다. 다음의 그림 1과 같은 신경망이 주어졌을 때, 신경세포들과 연결들에 대해 그림의 순서대로 번호를 매긴다면 신경망의 구조는 $[w_0 w_1 w_2 w_3 w_4 w_5 w_6 w_7 w_8]$ 과 같이 연결강도의 나열로 표현할 수 있다. 연결강도 변수 하나에 4개의 이진수를 사용할 경우, 대응되는 유전자형은 다음의 그림 2와 같이 나타낼 수 있다(여기서, b_0, b_1, \dots 등은 0 또는 1의 이진수이다). 결국, 그림 1은 신경망의 구조를 나타내는 표현형이고, 그림 2는 교배, 돌연변이 등의 과정을 수행할 때 그 대상이 되는 유전자형이다. BGA가 신경망의 유전자형에 대해 교배, 돌연변이 등의 유전 연산을 수행하여 새로운 신경망 집단을 만들게 되면, 새로운 집단의 적합도를 구하기 위해 각 유전자형에 대응된 표현형(실제 연결강도의 값들)을 얻어야 하는데 이때 유전자형에서 표현형으로의 매핑이 필요하다.

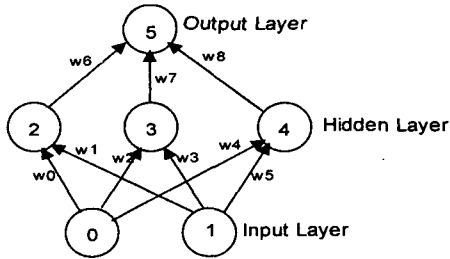


그림 1. 신경망에 대한 표현형

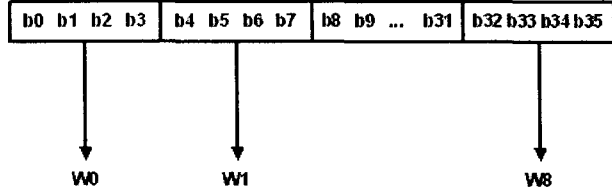


그림 2. 신경망에 대한 유전자형

3.2.2. 신경망 은닉층 최적화의 원리

유전자형에서 표현형으로의 매핑을 통해 신경망의 연결강도를 구하되, 본 연구에서는 적절한 기준에 합당하면 매핑된 연결강도 값이 0 이 될 수 있게 하였다. 연결강도 값이 0 이라는 것은 그 연결로 이어진 신경세포들 사이에서 어떠한 정보 교환도 일어나지 않음을 의미하므로, 이는 곧 그 신경세포들 사이의 연결을 제거할 수 있음을 의미한다. 이러한 사실은 연결강도가 0 인 연결이 은닉층과 출력층 사이에 존재할 때 더욱 중요한 의미를 가지는데, 그 이유는 은닉층과 출력층 사이에 존재하는 연결의 연결강도가 0 이면 연결을 제거할 수 있을 뿐만 아니라 그 연결의 아래에 이어진 은닉층의 신경세포까지도 제거할 수 있기 때문이다. 다음의 그림 3에서 연결 w_6 의 연결강도가 0 이므로 신경세포 2의 활성값은 신경세포 5의 입력으로 전달되지 못한다. 이는 신경세포 2의 활성이 출력층 신경세포 5의 활성값에 영향을 주지 못한다는 것을 의미하므로, 결국 신경세포 2는 출력층에서 계산되는 출력값과 목적패턴사이의 에러인 신경망의 에러에 영향을 주지 못하게 된다. 그러므로, 그림 5에서처럼 신경세포 2가 제거되기 전이나 제거된 후나 그 둘은 동일한 신경망이라고 볼 수 있다. 결국 제거된 후의 신경망은 $[0 \ 0 \ w_2 \ w_3 \ w_4 \ w_5 \ 0 \ w_7 \ w_8]$ 과 같은 표현형을 갖게 된다.

$w_6 = 0$ 이므로 신경세포 2에서 5로 정보의 흐름이 없다

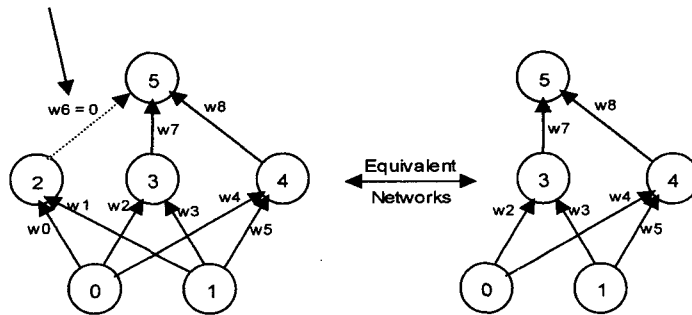


그림 3. BGA를 이용한 신경망 구조최적화의 원리

4. 다양한 학습규칙의 적용

4.1. 결정론적 모델(Deterministic Model)

결정론적 모델로 본 연구에서 검토한 학습방법은 모멘텀을 고려하지 않은 최대 경사법(steepest descent method without momentum, SD)과 모멘텀을 고려한 최대 경사법(steepest descent method with momentum,

SDM) 및 공액 구배법(conjugate gradient method, CG)이다.

4.1.1. 모멘텀을 고려하지 않은 최대 경사법

흔히 역전파 신경망이라고 불리는 신경망 모델이 바로 모멘텀을 고려하지 않은 최대 경사법인 SD를 학습 알고리즘으로 사용하는데, 이것이 바로 역전파 신경망의 일반화된 델타규칙(generalized delta rule)이다. 이는 신경망의 에러를 목적함수로 각 연결강도를 설계변수로 취하고 목적함수의 최대 경사 방향을 따라가면서 탐색을 수행하는 방법을 말하는데, 순수한 최적화 기법과는 달리 목적함수의 그레디언트를 중앙차분법과 같은 방법으로 구하기보다는 그레디언트가 구해지는 과정을 보다 조직화하여 신경망의 학습 알고리즘이 되게 만든 것을 말한다. 이를 식으로 표현하면 다음과 같다.

$$w_{ji}^{new} = w_{ji}^{old} + \alpha \delta_j a_i$$

여기서, α : 학습률, δ_j : 신경세포 j의 델타, a_i : 신경세포 i의 활성화값

4.1.2. 모멘텀을 고려한 최대 경사법

모멘텀을 고려한 최대 경사법 즉, SDM을 이용하여 신경망의 학습을 수행하는 알고리즘은 앞에서의 SD와 동일한데 다만 모멘텀이라는 새로운 항이 추가된 것이 다를 뿐이다. 이는 다음과 같은 식으로 표현된다.

$$w_{ji}^{new} = w_{ji}^{old} + \Delta w_{ji}$$

$$\text{여기서, } \Delta w_{ji} = \eta \delta_j a_i + \xi \Delta w_{ji}^{old}$$

여기서, η : 학습률(상수), ξ : 모멘텀률(상수), a_i : 신경세포 i의 활성화값

δ_j : 신경세포 j의 델타, $\Delta w_{ji}^{old} = w_{ji}^{old} - w_{ji}^{older}$: 모멘텀

4.1.3. 공액 구배법

최대 경사법이 탐색을 수행하는 방향은 항상 현재의 설계점에서 구한 목적함수의 음의 그레디언트 방향인데, 이렇게 탐색을 수행하는 것은 효율성이 떨어지는 약점이 있다. 즉, 최대 경사법으로 초기 설계점으로부터 최소점까지의 계산과정을 살펴보면 이미 탐색을 수행한 방향과 중복되는 방향에서 계산을 수없이 수행한 것을 알 수 있다. 이는 탐색이 수행되는 동안에 현재의 탐색방향을 구하기 위해서 이전에 구했던 탐색방향에 대한 고려는 전혀 없이 현재의 설계점에서 목적함수의 그레디언트만을 사용했기 때문에 나타나는 결과이다. 이와 대조적으로 공액 구배법은 현재의 탐색방향을 구할 때 현재의 탐색방향이 목적함수의 Hessian에 대해 이전에 구했던 탐색방향들과 서로 독립인 방향 즉, 공액 구배(conjugate direction) 방향이 되도록 계산을 수행한다. 이런 방법으로 탐색방향을 구하면 탐색방향들끼리 서로 중복이 없어져서 탐색의 효율을 크게 향상시킬 수 있다. 공액 구배법을 적용하여 신경망의 학습규칙을 정식화하면 다음과 같이 나타낼 수 있다.

$$w_{ji}^{new} = w_{ji}^{old} + \Delta w_{ji}$$

$$\text{여기서, } \Delta w_{ji} = \eta [\delta_j a_i + \gamma \Delta w_{ji}^{old}]$$

여기서, η : 학습률(1차원 탐색을 수행), a_i : 신경세포 i의 활성화값

δ_j : 신경세포 j의 델타, $\Delta w_{ji}^{old} = w_{ji}^{old} - w_{ji}^{older}$: 모멘텀

$$\gamma = \frac{[\delta_j a_i - \delta_j^{old} a_i^{old}] \cdot \delta_j a_i}{\delta_j^{old} a_i^{old} \cdot \delta_j^{old} a_i^{old}}, \text{ 여기서, } \delta_j^{old}, a_i^{old} : \text{이전 스텝에서 델타와 활성화값}$$

4.2. 스토캐스틱 모델(Stochastic Model)

본 연구에서 스토캐스틱 모델로 신경망의 학습을 수행하기 위해 시뮬레이티드 어닐링(simulated annealing, SA)과 실수형 유전알고리즘(real genetic algorithm, RGA)을 사용하였는데, 두 가지 모두 자연계의 현상을 모사한 최적화 알고리즘이다. 스토캐스틱 모델은 목적함수의 그래디언트를 사용하지 않으므로 역전파와 같은 과정을 필요로 하지 않고, 오직 feed-forward 과정만 존재한다.

4.2.1. 시뮬레이티드 어닐링(SA)

SA는 금속의 어닐링 과정을 모사한 최적화 알고리즘이다. 금속을 용융점까지 가열한 후 천천히 온도를 내려주면 그 금속의 전체 에너지는 점점 낮아져 결국 최소점에 도달하게 되며 그 중간단계에 어떠한 상태 i 에서 다음 상태 j 와의 에너지 차이 $\Delta E = (E_i - E_j)$ 에 따라 다음의 확률한도 내에서 상태 이동이 허용될 수 있는데 이를 Metropolis 기준이라고 부른다. 이는 다음과 같은 식으로 표현된다.

$$P_{accept} = \text{MIN}(1, \exp[\frac{E_i - E_j}{k_B T}])$$

여기서, T : 금속의 온도, k_B : 볼츠만 상수, E_i, E_j : 현재, 다음 상태의 에너지

이식은 다음 상태의 에너지가 커진다고 하더라도 위 식의 조건을 만족시킨다면 상태 이동이 가능하다는 것을 의미하는데, 이를 최적화 기법에 응용하면 목적함수의 증가도 적절한 기준에 의해 허용할 수 있어서 국소 최적점으로부터 빠져 나올 수 있는 가능성을 부여할 수 있게 된다.

4.2.2. 실수형 유전알고리즘(RGA)

RGA는 집단 탐색을 수행하여 보다 넓은 설계영역의 정보를 활용함으로써 전역 최적점에 도달할 수 있는 확률이 매우 크므로, 신경망의 학습을 수행시키는데 있어서 다른 어떤 알고리즘보다 학습성취 가부의 문제에 대해 안전하게 사용할 수가 있다. 또한, RGA는 국부 탐색 능력이 상대적으로 높아서 미세한 학습까지도 가능하고, 목적함수인 에러의 그래디언트를 사용하지 않으므로 feed-forward 과정만으로 간단하게 학습을 수행할 수 있다. BGA에서 RGA로 초기값을 넘겨주는 일은 BGA에서 얻어진 가장 좋은 개체(best individual)를 사용하여 RGA의 초기 개체군을 형성시키는 것으로 가능하다.

5. 연구 결과

본 연구에서는 예제로써 XOR 문제와 선상가열 문제 1, 2를 적용하여 보았는데, 다음과 같은 세 가지의 경우에 대해 결과를 비교·검토하여 보았다.

- ① 스토캐스틱-결정론적 하이브리드 모델 즉, 은닉층 최적화-미세 탐색 모델
- ② 스토캐스틱 단독 모델 및 ③ 결정론적 단독 모델

예상했던 것과 같이 ①의 경우가 학습능력 및 시간효율 측면에서 가장 우수한 것임을 알 수 있었고, ②는 시간효율의 측면에서 ③은 학습능력의 측면에서 문제가 있는 것으로 드러났다.

5.1. 스토캐스틱-결정론적 하이브리드 모델을 채택한 경우

초기의 신경망의 구조로써 XOR 문제는 2-4-1, Line Heating 1 문제는 2-4-1, Line Heating 2 문제는 3-5-1 을 주고 BGA를 수행시켜본 결과 다음의 표 1과 같은 최적의 신경망 구조를 얻을 수 있었다.

	Given Structure	Optimized Structure
XOR	2-4-1	2-2-1
Line Heating 1	2-4-1	2-2-1
Line Heating 2	3-5-1	3-3-1

표 1. BGA에 의해 신경망의 구조가 최적화된 결과

그리고, 각 학습방법들을 적용하여 미세 학습을 시켜 다음의 표 2와 같은 결과를 얻을 수 있었다. BGA는 구조를 최적화 시키는 것과 동시에 학습에러에 대한 전역 최적점 부근까지 탐색을 수행하므로 BGA와 연결되어 다른 어떤 방법의 최적화 기법을 사용하여 탐색을 수행하여도 국소 최소에 빠지지 않고 미세 탐색을 잘 수행하는 것을 알 수 있다. 특히 BGA-CG 모델이 시간효율 면에서 가장 우수한 것을 알 수 있다.

	XOR			Line Heating 1			Line Heating 2		
	Error	Iteration	Locality	Error	Iteration	Locality	Error	Iteration	Locality
BGA RGA	0.0000003	44	No	0.001594	10070	No	0.012966	20470	No
BGA SA	0.0000160	840	No	0.004127	10110	No	0.036373	20140	No
BGA CG	0.0000000	19	No	0.0000994	430	No	0.000999	6550	No
BGA SDM	0.0002218	950	No	0.0000990	2050	No	0.024245	20890	No
BGA SD	0.0002424	980	No	0.0001000	4840	No	0.032772	21280	No

표 2. 하이브리드 모델('BGA+미세 학습' 모델)의 결과

다음의 그림은 하이브리드 모델에서 신경망이 학습된 예이다.

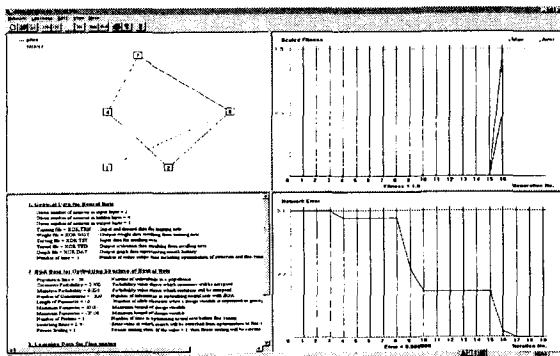


그림 4. XOR 문제에 대한 BGA-CG 방법의 결과

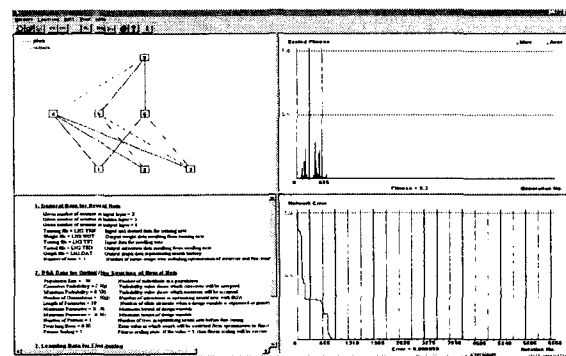


그림 5. LH2 문제에 대한 BGA-CG 방법의 결과

그림 4, 5에서 왼쪽 상단의 그림은 신경망의 구조가 최적화된 상황을 보여주고 있는데, 은닉층의 신경세포 수가 각각 4개와 5개에서 2개씩 줄어든 것을 나타내고 있다.

5.2. 각 학습 알고리즘을 단독 수행한 경우

신경망의 구조를 BGA 수행 결과 얻어진 것으로 고정하여(XOR는 2-2-1, LH1은 2-2-1, LH2는 3-3-1로 고정) 신경망의 학습을 수행시켜 보았는데, 학습 알고리즘으로는 스토캐스틱 모델(RGA, SA)과 결정론적 모델(CG, SDM, SD)을 각각 단독으로 적용하였다. 예상했던 것과 같이 스토캐스틱 모델은 국소 최적점에 빠지지 않고 올바른 학습을 수행하였지만 학습수행을 위한 반복계산이 너무 많이 필요하였고, 결정론적 모델은 빠른 학습을 수행할 수 있었지만 자주 국소 최적점에 빠지는 것을 확인할 수 있었다. XOR 문제의 경우 CG, SDM, SD 방법 모두 국소 최적점에 빠지는 것을 알 수 있었는데, 특히 SD 방법의 경우 학습에러가 각각 3.130417과 4.172400인 두 개의 국소 최적점에 빠지는 것을 자주 볼 수 있었다.

선상가열 1, 2 문제의 경우는 SD 방법만 국소 최적점에서 벗어나지 못했다. 즉, RGA, SA는 물론이고 CG, SDM도 좋은 학습성능을 보여 주었다. 여기서 주목할 만한 사실은 선상가열 1, 2 문제의 경우는 CG 방법이 BGA-CG 방법보다 더 좋은 결과를 준다는 것인데, 선상가열 1, 2 문제와 같이 비선형성이 별로 없는 자료를 학습시키는 경우에는 CG 방법만으로 신경망을 학습시키는 것도 괜찮은 방법임을 알 수 있었다. 그러나, CG 방법 단독으로 신경망의 학습을 수행하는 것도 학습을 수행하기 이전에 최적의 신경망 구조를 미리 알고 있어야 하므로, BGA에 의한 신경망 구조 최적화의 결과가 필요하다. 각 방법들에 의한 신경망 학습결과는 다음과 같이 표 3에 요약되어 있다.

	XOR			Line Heating 1			Line Heating 2		
	Error	Iteration	Locality	Error	Iteration	Locality	Error	Iteration	Locality
RGA	0.0000000	11	No	0.000101	10000	No	0.008884	20000	No
SA	0.0000300	21	No	0.005663	10000	No	0.035858	20000	No
CG	0.0000000	5	Yes 3.125	0.000083	88	No	0.000993	180	No
SDM	0.0019570	1000	Yes 3.127548	0.000268	10000	No	0.001000	18560	No
SD	0.0052230	1000	Yes 3.130417 4.172400	0.001969	10000	Yes 0.119271	0.001829	20000	Yes 0.042960

표 3. 단독 모델의 결과

다음의 그림은 단독 모델에서 신경망이 학습된 예이다.

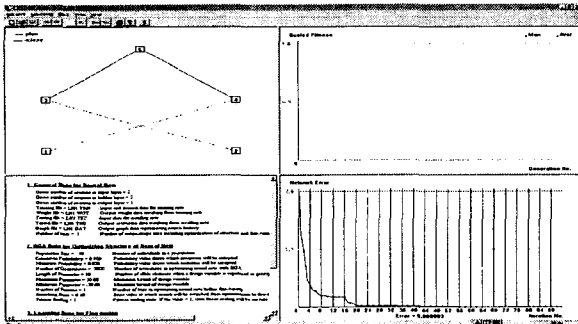


Fig. 6. LH1 문제에 대한 CG 방법의 결과

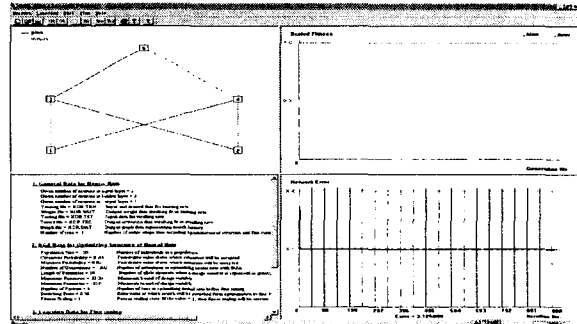


그림 7. XOR 문제에서 국소 최적점에 빠진 경우 - CG

앞의 그림 6은 선상가열 문제 1에서 은닉층의 신경세포수를 2개로 고정시키고 CG 방법으로 학습을 수행한 결과인데 상당히 좋은 학습성취도를 보여주고 있다. 그리고, 그림 7은 XOR 문제에서 학습을 수행한 결과 국소 최적점에 빠져 더 이상 학습이 진행되지 못하고 있는 상황을 보여주고 있다.

6. 결론

본 연구에서는 시스템 시뮬레이터로 사용되는 순차 신경망의 미해결 문제점인 신경망 구조 결정의 문제와 신경망 학습방법의 문제를 해결하고자 하였고, 예제로써 XOR 문제와 선상가열 문제를 적용하여 보았다. 그 결과 다음과 같은 결론을 얻을 수 있었다.

첫째, 이진수형 유전 알고리즘(BGA)은 신경망 은닉층 구조 최적화를 위해 매우 만족스러운 결과를 준다. 즉, BGA가 갖고 있는 변수의 이산적 성질은 신경망 은닉층 구조 최적화를 위해 효과적으로 사용될 수 있었는데, 본 연구의 예제로써 적용된 XOR 문제와 선상가열 1, 2 문제의 경우에는 BGA를 수행해 본 결과 입력층의 신경세포 개수와 같은 수만큼의 신경세포를 갖는 은닉층으로도 충분히 학습된 신경망을 얻을 수 있었다. 또한, BGA로 신경망의 최적화를 수행한 후에 실수형 유전 알고리즘(RGA), 시뮬레이티드 어닐링(SA), 공액 구배법(CG), 모멘텀을 고려한 최대 경사법(SDM), 모멘텀을 고려하지 않은 최대경사법(SD)으로 미세 학습을 시켜본 결과 모두 성공적인 학습이 수행됨을 알 수 있었다. 이는 BGA가 신경망 은닉층 구조 최적화 뿐만 아니라 학습에러의 전역 최적점 부근까지 전역적 탐색을 효과적으로 수행하기 때문인데, BGA의 이러한 특성 때문에 스토캐스틱-결정론적 하이브리드 모델을 자연스럽게 얻을 수 있었다.

둘째, 신경망의 학습을 위해 본 연구에서 적용한 방법 중 BGA-CG 하이브리드 모델이 가장 좋은 결과를 주었다. 학습 성취도 측면(전역적 최적해가 얻어진 정도 및 에러가 줄어든 정도)에서는 BGA-RGA, BGA-SA, BGA-SDM, BGA-SD 모두 충분히 학습된 신경망을 얻을 수 있었으나 BGA-CG보다 훨씬 많은 반복계산이 필요하였다. 즉, BGA-CG는 전역적인 학습 및 미세 학습을 효과적으로 수행할 뿐 아니라 시간 효율적인 면에서도 가장 우수하다고 할 수 있다.

결국, 본 연구에서는 기존의 순차 신경망의 문제를 극복하기 위해 'BGA에 의한 신경망의 은닉층 결정 후 CG에 의한 미세 학습의 수행'이라는 하이브리드 학습 모델을 제안할 수 있었고 그 방법을 적용해본 결과 만족할만한 신경망의 학습 모델이라는 것을 알 수 있었다.

참고문헌

- [1] 양영순, 김기화, "실수형 Genetic Algorithm에 의한 최적설계", 한국전산구조공학회 논문집, 제8권 제2호, 1995.
- [2] 安居院猛, 長尾智晴, ツエネテイツクアルゴリズム, 昭晃堂, 1993.
- [3] David, E. Goldberg, Genetic Algorithm in Search, Optimization and Machine Learning, Addison Wesley Publishing Company, California, 1989
- [4] James, A. Freeman, David, M. Skapura, Neural Networks (Algorithms, Applications, and Programming Techniques), Addison-Wesley Publishing Company, California, 1991
- [5] Philip, D. Wasserman, Neural Computing (Theory and Practice), ANZA Research, Inc., Van Nostrand Reinhold, New York, 1989
- [6] Limin, F., Neural Networks in Computer Intelligence, McGraw-Hill Inc., 1994.
- [7] Simon, H., Neural Networks, A Comprehensive Foundation, Macmillan Publishing Company, 1994