

Special Thema | 신경회로망과 반도체

1. 서론

동성수 교수
(용인송담대학 디지털전자과)

신경망은 인간의 두뇌를 모방한 회로망으로써 기존의 컴퓨터가 잘 하지 못하는 연산이나 추론 등의 능력을 보완하기 위해 쓰여진다. 대부분의 신경망 알고리즘은 소프트웨어로 구현되어 컴퓨터나 마이크로프로세서를 이용하여 응용되고 있으나 더욱 빠른 속도를 얻기 위해 전용칩으로도 구현되고 있다. 따라서 신경망과 반도체와의 관계는 매우 밀접하다. 즉 더욱 효율적인 반도체 공정을 위해서 신경망 알고리즘이 동원되기도 하고, 더욱 효율적인 응용을 위해 반도체 칩으로도 제작되는 것이다. 기술적 진보와 경제적 측면의 상황에 따라 신경망 기술은 암흑기도 있었고 중흥기도 있었으며 앞으로도 그러할 것이다. 이 글은 신경망과 관련된 기본 내용을 다시 다루어 봄으로써 신경망의 의미를 되새기고자 하는데 목적이 있다.

2. 신경망이란?

우리의 기억 속에 아직도 자리 잡고 있는 영화, 터미네이터에서는 신경망 칩이 인공두뇌의 핵심으로 작용하여 놀라운 역할을 수행하는걸 보여주었다. 이는 컴퓨터의 발전과 기술 환경의 빠른 변화 속에서 인간과 같이 보고, 듣고, 말하고, 생각하고, 스스로 판단하게 하는 인간과 같은 기계를 만들어 힘들고 복잡한 일에서 벗어나고자 하는 사람들의 욕구를 나타낸 것이라고 할 수 있다.

그러나 신경회로망의 개발이 인공두뇌를 만드는 것을 의미하는 것은 아니다. 기존의 컴퓨터는 수치계산 등 잘 정의된 일을 처리하는 데는 놀랍게 발전되었지만 연상, 추론, 인식 등 인간이 쉽게 처리하는 특정 분야에서는 아직 많은 문제점을 가지고 있다. 즉, 문제가 명확하게 정의되어 있고 프로그래머에 의해 문제 해결 방법인 알고리즘이 제시된 경우는 뛰어난 성능을 발휘하지만 문제가 복잡하거나 잘 정의되어 있지 않은 문제는 해법을 찾기 어렵다. 이러한 문제점을 해결하기 위해서 인간 두뇌의 정보 처리 방식을 모방한 신경회로망이 나오게 되었다.

신경회로망은 이와 같이 알고리즘을 찾기 어려운 문제에 대해 신경망을 학습시킴으로써 문제를 스스로 해결할 수 있게 한다. 또한 기존의 컴퓨터가 순차적으로 정보를 처리하는데 반해 신경망은 인간의 신경세포와 같이 정보를 병렬적으로 처리하기 때문에 정보의 양이 많거나 복잡한 문제에 대해서 효율적이며 잡음이 있거나 애매한 정보도 효과적으로 처리할 수 있다. 이러한 장점을 활용하여 신경회로망은 기존 컴퓨터가 해결하기 힘든 패턴인식, 음성인식, 최적화 문제, 적응 제어 등에 더 효과적이다. 그 외에도 기업 진단, 신용 평가, 주식 시세 예측, 운행 스케줄링, 프로세스 제어, 고장 진단, 반도체 공정 상태의 진단과 제어 등 여러 가지 응용 분야가 있다.

3. 신경망의 특징

생명체의 신경조직으로부터 착안하여 모델화한 정보처리 시스템인 신경회로망은 단순한 소자들의 병렬, 분산 연결 구조를 가지고 있고 외부의 입력에 의해 동적반응을 일으켜 필요한 출력을 생성시킨다. 이러한 신경망은 먼저 병렬 분산 처리 방식으로 데이터를 처리한다는 특징이 있다. 따라서 영상과 같은 2차원 데이터도 빠른 시간 안에 처리가 가능하다. 다음 특징으로는 연상기억능력을 들 수 있다. 이는 주소를 가지고 저장 장소를 찾는 방식이 아니고 내용의 일부나 관련된 암시를 가지고 전체의 내용을 찾아내는 방식이다. 입력 데이터에 포함될 수 있는 잡음에 강한 능력을 보여준다. 세 번째 특징으로 고장극복의 능력을 들 수 있으며 마지막 특징으로는 학습능력을 들 수 있다. 이러한 특징들로 인해 연상이나 추론이 가능한 것이다.

기존 컴퓨터의 해결 방식과 다른 방식의 신경망은 인간 두뇌 형태의 시스템을 제작하는 것이 아닌 특정 응용 분야에 활용되는 보조 처리 장치의 역할로 이해하는 것이 타당할 것이다.

4. 신경망의 구조

인간 두뇌의 기본 단위인 신경세포(Neuron)의 구조는 그림1과 같다. 인간의 두뇌는 평균 10의 11승개

의 신경세포들로 이루어져 있으며, 각 신경세포는 수상돌기들을 통해서 약 만 개까지의 다른 신경세포들과 연결되어 있다.

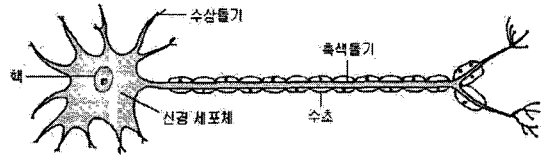


그림 1. 신경 세포의 구조.

신호는 전기화학적인 과정들을 통해 전달되는데 그림2와 같이 수상돌기와 축삭돌기의 연결부인 시냅스(Synapse)가 기본적인 신호전달 지점이다. 수상돌기들은 다른 신경세포들로부터 온 신호들을 모으고 세포체가 그 신호들을 합하여 그 결과를 축삭돌기를 통해 다른 신경세포로 전달한다. 중요한 것은 그 결과 값이 시냅스를 통해서 넘어갈 수 있는지에 대한 여부이다. 신경세포들 간의 연결 상태 즉, 시냅스를 통해 다른 신경세포에 정보를 전달하는지 여부 자체가 학습되어진 지식이라 할 수 있고, 이런 연결 상태들은 학습에 따라 변할 수 있으므로 여러 가지 정보들을 계속 습득할 수 있다[1].

여러 신경망 모델들은 신경세포 즉 뉴런과 시냅스들 간의 결선 방법과 학습의 규칙에 따라 나누어

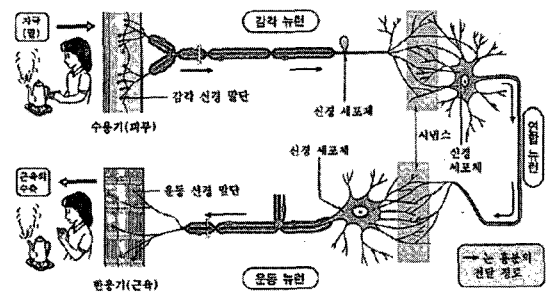


그림 2. 신경의 전달 예.

진다. 뉴런이란 생체의 신경세포와 마찬가지로 인접한 다른 많은 뉴런들로부터 입력신호들을 받아서 하나의 결정된 신호(정 또는 부)를 출력한다. 시냅스는 뉴런의 출력을 받아 저장되어 있거나 공급되어지는 가중치(Weight)를 곱해서 다른 뉴런으로 보낸다. 여기서 가중치를 연결강도라고 부른다. 뉴런의 출력은 그 뉴런의 입력부에 연결된 다른 모든 뉴런들의 출력 값(즉, 외부로부터의 모든 입력 값)에 해당 가중치를 곱한 값들을 모두 더해서 뉴런의 전달함수를 통과시킴으로써 얻어진다.

$$Y = f\left(\sum_{i=1}^n W_i X_i + B X_0\right)$$

이러한 작용을 하는 간단한 신경망 모델을 다음의 그림3에 표현하였다.

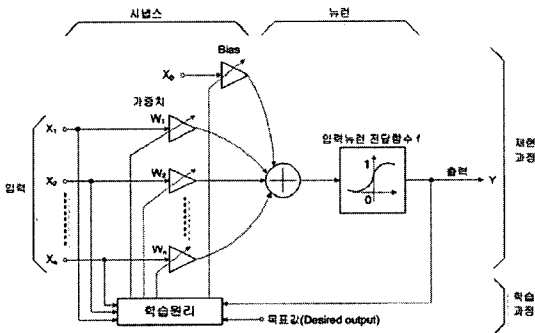


그림 3. 간단한 신경망의 구조.

여기서 설명한 입출력간의 연산이 신경망의 재현(Recall)과정이며, 적절한 가중치의 값을 구하는 과정이 학습(Learning)과정으로서 여러 가지 방법들이 있다. 위 그림에서 보여 주는 것은 표본의 입출력 데이터를 이용하여 표본에 속하지 않는 입력 패턴에 대한 가장 적절한 출력이 나올 수 있도록 가중치(W_i)를 점진적으로 선정하는 과정이다. 그림은 단층 뉴런 전방결선(Feedforward) 신경망의 경우이나 그 외에도 다수의 뉴런 층을 연속하여 연결할 수 있으며, 되먹임 루프의 결선방식에 따라서는 되먹임형(Feedback), 순환형(Recurrent) 또는 수평형(Lateral)

결선을 포함하는 모델들이 있다[2]. 이를 그림4에 간단히 표현 하였다.

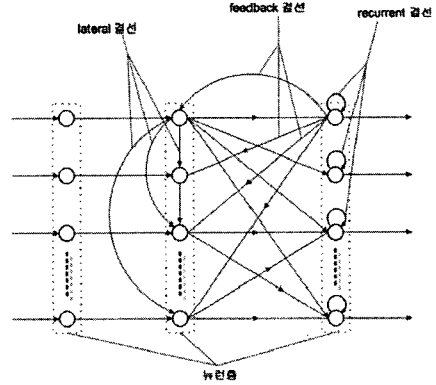


그림 4. 여러 가지 결선 방식.

그림4는 여러 가지 모델에서 가능한 결선 방식을 나타내고 있다.

5. 동작 원리

신경망의 대표적 두 가지 모델에 대하여 간단한 예제를 들어 그 동작원리를 다음과 같이 설명하였다. 첫째로 패턴인식기를 구현하는 예를 들어보자. 두 개의 2진 입력신호(X_1, X_2)가 서로 다를 때(즉 “10” 또는 “01”), 출력으로는 ‘1’을 발생 시키는 다층 신경망을 그림5와 같이 구성할 수 있다.

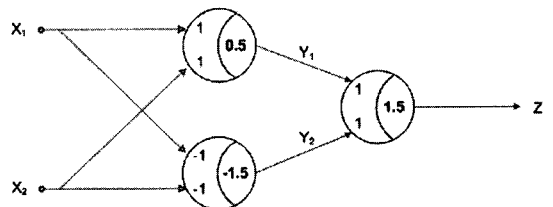


그림 5. XOR 연산을 수행하는 2층 신경망.

이들의 동작을 식으로 표현하면

$X_1 + X_2 \geq 0.5$ 일 때 $Y_1 = 1$ 이고, 논리 연산 표현으로는 $Y_1 = X_1 + X_2$ 이다.

그리고 $-X_1 - X_2 \geq -1.5$ 일 때 $Y_2 = 1$ 이고, 논리 연산 표현으로는 $Y_2 = \bar{X}_1 + \bar{X}_2$ 이다.

그리고 $Y_1 + Y_2 \geq 1.5$ 이면 $Z = 1$ 이 출력 된다.

즉, $Z = Y_1 Y_2 = X_1 \bar{X}_2 + \bar{X}_1 X_2$ 이 되어 Z 는 X_1, X_2 의 XOR 연산을 수행 한다. 이러한 연산을 그림6과 같이 2차원 평면상에 표시하였다. 그림에서 그림자 영역 내에 속하는 (X_1, X_2) 에 대하여 출력을 '1'로 발생하는 패턴 분류기 기능임을 알 수 있다.

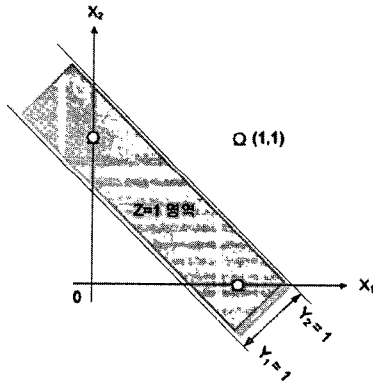


그림 6. 2층 신경망에 의한 패턴 분류.

이와 같은 방법에 의하면 각층의 뉴런수를 증가 시킴에 따라서 보다 다양한 패턴도 인식할 수 있게 된다.

두 번째 예제로는 되먹임 결선형 모델 중에서 전 반결선(Fully-connected)인 홉필드(Hopfield) 회로망의 예를 들어 신경망의 또 다른 특성인 에너지 최소화 기능을 설명하고자 한다[3]. 그림7은 3개의 뉴런으로 구성된 홉필드 회로망이다.

각 연결부의 가중치 W_{ij} 는 i 번째 뉴런의 출력 값(가중치로 들어오는 입력)이 곱해져서 j 번째 뉴런의 입력으로 더해져 들어가는 연결강도를 나타내며, 각 뉴런 V_i 는 문턱(Threshold)값 T_i 를 가지고 있다고 하자.

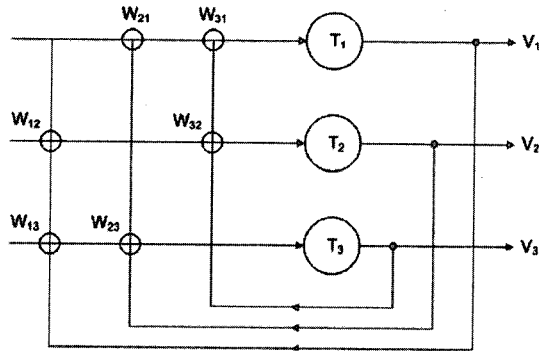


그림 7. 3개 뉴런의 홉필드 신경망.

$$\text{그러면 } V_i = \begin{cases} 1 & \text{if } \sum W_{ij} V_j \geq T_i \\ 0 & \text{otherwise} \end{cases}$$

여기에 다음의 W 값들과 T 값들을 적용하면 이 회로의 출력상태는 그림8과 같이 변화 시킬 수 있다. 각 상태의 값(원 안의 값)은 각 뉴런 V_i 의 상태를 의미한다.

$$\begin{aligned} W_{12} = W_{21} &= -0.5 & T_1 &= -0.1 \\ W_{13} = W_{31} &= 0.2 & T_2 &= 0 \\ W_{23} = W_{32} &= 0.6 & T_3 &= 0 \end{aligned}$$

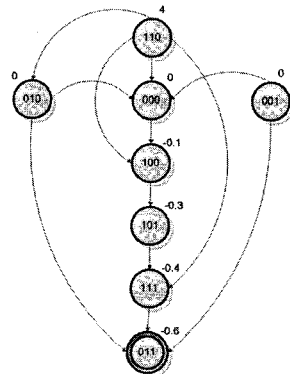


그림 8. 주어진 홉필드 회로망의 상태 천이도.

홉필드(J. Hopfield)는 그의 회로망에서 에너지를 다음과 같이 정의하였다.

$$E = -\frac{1}{2} \sum_i \sum_j W_{ij} V_i V_j + \sum_i V_i T_i$$

또한 그 에너지는 회로망의 모든 상태 변화에 대하여 단조 감소함을 보였다. 그림8의 각 상태위에 표시된 숫자는 그 상태(State)에서의 에너지 값이며, 앞의 문제에서 실제로 에너지 값이 단조 감소하여 최소치 -0.6에 도달했을 때 회로망은 "011"의 안정 상태를 유지하게 됨을 알 수 있다. 다시 말하면 적절히 주어진 W와 T의 값에 의하여 홉필드 회로망은 어떠한 초기값에서 시작하더라도 일정시간 내에 에너지 최소점에서 안정상태에 도달한다. 이러한 특성에 의하여 홉필드 회로망은 기존의 알고리즘으로는 계산량이 너무 커서 최적해를 구하기 어려운 문제들의 실시간 풀이에 사용될 수 있다.

6. 학습 원리

신경망의 학습 원리에 대하여 헵(D. Hebb)은 1949년 다음과 같이 언급하였다[4]. "하나의 신경 세포가 다른 신경 세포에 연결되어 있으면서 계속적으로 그 세포의 활성화(Firing)에 기여한다면 그 두 세포간의 결속력(상호 영향력)의 효율을 증가시키기 위한 신진대사의 변화(연결 가중치 증가)가 일어나게 된다."

이 현상을 수식으로 표현한다면

$$W_{ij}(t+1) = W_{ij}(t) + cX_i(t)Y_j(t), \text{ 여기서 } c \text{는 학습율}$$

가 된다. 그림9와 같이 n개의 뉴런이 m개의(앞의 n개 중 일부를 포함해도 좋음) 뉴런 상태에 영향을 미친다고 할 때 학습과정이란 알려져 있는 여러 쌍의 벡터 (X^k, Y^k)로부터 이들을 모두 만족시키는 연결강도 n×m 매트릭스 W를 구하는 일이다.

재현(Recall)과정이란 X와 W로부터 Y를 구하는 과정으로서 Y=WX에 의하여 항상 구해진다. 반면 P쌍의 (X, Y)로부터 W를 구하는 일은 일반적으로 불확정성(Indeterministic) 알고리즘이 된다. 그러나 P

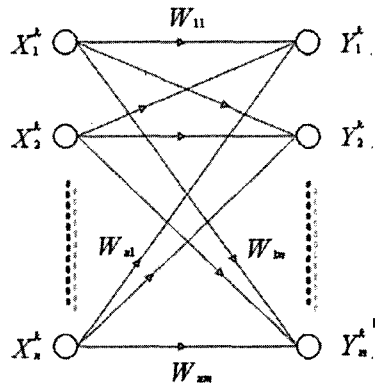


그림 9. 뉴런간의 상호 연결.

쌍의 입력 벡터가 서로 단위 직교(Orthonormal)하다면 W_{ij}는 헵의 법칙에 의하여

$$W_{ij} = \sum_{k=1}^P X_i^k Y_j^k \text{ 이 되며,}$$

충분한 P가 주어진다면 학습되지 않은 패턴 입력에 대해서도 연상 재현이 잘 이루어지게 된다. 이에 대한 예로서 다음의 사상을 행하는 신경망의 연결 강도를 학습률에 의하여 계산하여 보자.

$$\begin{aligned} X_1 &= (1 \ -1 \ -1) & Y_1 &= (-1 \ -1 \ 1) \\ X_2 &= (-1 \ 1 \ -1) & Y_2 &= (-1 \ 1 \ -1) \\ X_3 &= (-1 \ -1 \ 1) & Y_3 &= (1 \ -1 \ -1) \end{aligned}$$

$$W = \sum_{K=1}^3 X_K^T Y_K \text{ 에 의하면}$$

$$W = \begin{pmatrix} -1 & -1 & 3 \\ -1 & 3 & -1 \\ 3 & -1 & -1 \end{pmatrix} \text{ 이 된다.}$$

재현과정은 입력 X₁ = (-1 1 -1)에 대한 출력 Y₁ = F(X₁W) = F(-3 5 -3) = (-1 1 -1)로서 학습시킨 사상을 행한다.

단, 여기서 뉴런 전달함수 F는 문턱(Threshold)함수이다.

헵의 가설에서부터 여러 가지의 학습법칙들이 제안되고 있으며, 그 중에서 델타 룰(Delta Rule)을 소개하면, 연결강도 W 는 현재의 출력과 목표치와의 차이(오차)를 줄이는 방향으로 조절된다.

$$\begin{aligned} \text{즉 } W(t) &= W(t-1) + cE(t)X^T(t) \\ E(t) &= t(t) - W(t-1)X(t) \\ t(t) &= \text{목표 출력값} \end{aligned}$$

델타 룰은 가중치 공간에서 오차를 줄이는 방향으로 이동하는 Gradient Descent를 구현하게 된다.

$$\Delta W_{ij} = \frac{dE_i}{dW_{ij}}$$

델타 룰의 개념을 다층 신경망에 확대 적용시키려면 변형이 필요하다. 다층 신경망의 출력과 목표값과의 오차에 의하여 출력층의 연결강도는 정해진다. 그러나 출력층이 아닌 은둔층(Hidden Layer)의 경우 목표 값이 지정되지 않으므로 오차가 직접 산출되지 않는다. 이 경우 은둔층의 연결강도를 산출해 내는 방법으로 잘 알려진 오차역전파(Error Back-Propagation) 학습방법이 있다[5]. 은둔층의 연결강도를 산출하기 위하여 은둔층 출력의 오차값 대신에 출력층 오차와 출력층 연결강도의 곱을 사용하여 델타 룰을 적용시킨다. 즉,

$$W_h(t) = W_h(t-1) + c(E_o(t)W_o(t))X_h^T(t)$$

이와 같은 방법으로 출력층으로부터 입력층에 이르는 모든 은둔층의 연결강도를 오차 역전 시켜가며 구하게 된다. 이상에서 설명된 학습방법 이외에도 여러 가지 학습 룰들이 있으며, 응용분야에 따라 그 효용성이 다르게 평가된다.

7. 신경망 칩

단순하게 지능과 신경망의 크기를 비교해 보자. 거머리는 천개 정도의 신경망 연결을 가지고 있으며, 이 연결은 초당 만개를 처리한다고 한다. 파리는 일억 개의 신경망 연결을 가지고 있으며 매초 십억

개씩 처리한다. 바퀴벌레는 십억 개의 신경망 연결을 천억 개씩 처리하며, 벌은 백억 개의 연결고리를 초당 일조 개씩 처리한다고 한다. 반면에 사람은 무려 일경 개의 신경망 연결을 매초 천조 개씩 처리한다고 한다. 이러한 신경망을 컴퓨터의 소프트웨어로 구현한다고 할 때 한 개의 신경망 연결은 한 개의 메모리와 한 개의 곱셈을 필요로 한다. 따라서 신경망 연결 수와 연결의 처리 속도를 컴퓨터의 주기억 용량과 계산 속도로 환산하면 된다. 현재 반도체 기술로 메모리 용량은 크게 문제가 될 것 같지 않으나 계산 속도에는 문제가 있어 보인다. 즉, 슈퍼컴퓨터가 백만 개의 신경망 연결을 매초 일억 개씩 처리한다면 이는 파리 수준보다도 못한 것이다. 이는 반도체 칩인 트랜지스터의 처리 속도 보다 신경망 연결 개수에 기인한다고 할 수 있다. 따라서 기존의 컴퓨터로는 인간 수준의 신경망에 접근하기 힘들며 신경망 하드웨어가 필요한 이유가 여기에 있다.

현재의 기술 수준에서 신경망을 하드웨어로 만드는 가장 현실적인 방법은 반도체를 사용하여 집적회로를 만드는 것이다. 이제까지의 연구를 통해 몇 가지 분류로 나누어 볼 수 있다. 학습 관점에서 나누어 보면 학습 값이 고정된 칩과 학습이 가능한 칩으로 분류할 수 있으며 학습 가능한 경우도 칩 자체에서 가능한 경우와 칩 외부에서 학습이 가능한 칩이 있다. 용도로 구분하자면 여러 가지 용도에 사용할 수 있도록 한 범용 칩과 특정한 용도에서 사용할 수 있는 특수용도 칩으로 분류할 수 있는데, 현실적으로 특수용도가 일반적이다. 구조로 분류하자면 디지털 칩, 아날로그칩, 디지털과 아날로그 혼용칩으로 분류할 수 있다.

8. 고찰

이제까지의 신경회로망에 대한 대부분의 연구는 기존의 컴퓨터를 이용하여 신경망의 동작이나 학습에 관한 소프트웨어 시뮬레이션이다. 상용화된 신경망 소프트웨어는 매우 다양하며 고객관계 분석, 신용 평가, 컴퓨터 네트워크 관리, 전자 상거래 내용관리 등 많은 응용에 이용되고 있다. 물론 언어인식, 안면인식, 신호처리 분야 등 공학 분야에는 말할 것도

없다.

상대적으로 신경회로망 칩에 대한 연구는 적다고 볼 수 있으나 상용화 시켜 쓰는 칩도 있다. 그러나 소프트웨어 보다는 그 범위가 매우적이다. 지난 90년대 초까지는 칩 및 보드기반의 신경망 하드웨어가 많은 가능성을 보였다. 개발되는 신경망 하드웨어가 소프트웨어 시뮬레이션의 기능을 크게 향상 시켜 기존의 마이크로프로세서를 대체할 수 있을 것이라고 기대 하였다. 그러나 현재 상황은 그렇지 않다. 어찌된 일 일까?

반도체 기술의 발전과 경제적인 측면에서 볼 때, 기존 마이크로프로세서의 속도 향상과 가격의 하락이 이루어졌기 때문이다. 신경망 칩을 제작하기 위한 시간적, 경제적 노력 보다 빠르게 실행되는 마이크로프로세서에 신경망 알고리즘을 구현한 소프트웨어를 내장시켜 동작 시키는 것이 시간적으로나 경제적으로 훨씬 유리하기 때문이다. 그러나 그것은 일정한 범위에 한정된 응용일 것이다.

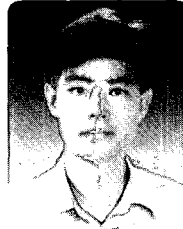
신경망 기술이, 주어진 환경에서 학습을 통해 인간과 같은 수준의 판단을 할 수 있어서 미래를 예측할 수 있다는 믿음과 그것의 구현은 스스로 학습하며 진화해 가는 칩이 중심이 될 것 이라는 공학적 희망은 언젠가 이루어질 것이라 생각된다. 신경망의 역사에서 암흑기를 지나 부흥기가 도래했던 것처럼.

참고 문헌

- [1] Geoffrey E. Hinton, "Connectionist Learning Procedures," Artificial Intelligence, Vol. 40, p. 185, 1989.
- [2] Tony Robinson, "Practical Network Design and Implementation," Neural Network Summer School, Cambridge University, 1992.
- [3] J. J. Hopfield, "Neurons with Graded Response Have Collective Computational Properties Like Those of Two-State Neurons," Proceedings of the National Academy of Sciences USA, Vol. 81, p. 3088, 1984.
- [4] D. O. Hebb, "The organization of Behavior", Wiley, 1949.
- [5] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representation by error

propagation," Parallel Distributed Processing: Explorations in the Microstructure of Cognition., Vol. I: Foundations., chapter 8., Bradford Books/MIT Press, Cambridge, MA, 1986.

저자|역력



성명 : 동성수

◆ 학력

- 1990년 인하대 전기공학과 공학사
- 1992년 인하대 대학원 전기공학과 공학석사
- 2005년 인하대 대학원 전기공학과 공학박사 수료

◆ 경력

- 1992년 ~ 1995년 나우정밀연구소 전임연구원
- 1996년 ~ 2000년 삼성전자 정보통신 선임연구원
- 2001년 ~ 현재 용인송담대학 디지털전자과 조교수

