이학석사학위논문

기억 앙상블 학습을 위한 확률적 회귀형 신경망 모델과 오차 기대값 역전파 알고리즘

Expected Error Backpropagation Through Time for Memory Ensembles in Stochastic Recurrent Neural Network Models

2008 년 2 월

서울대학교 대학원

협동과정 뇌과학전공

김 권 일

기억 앙상블 학습을 위한 확률적 회귀형 신경망 모델과 오차 기대값 역전파 알고리즘

지도교수 장 병 탁

이 논문을 이학석사학위논문으로 제출함

2007 년 12 월

서울대학교 대학원

협동과정 뇌과학전공

김 권 일

김권일의 석사학위논문을 인준함

2008 년 2 월

위 원 장 강 봉 균 (인)

부 위 원 장 장 병 탁 (인)

위 원 최 석 우 (인)

초 록

기억은 다양하게 분류되는데, 그 중에서 입력된 정보들끼리 서로 연결되며 저장되는 기억을 연합기억(associative memory)라 한다. 이 연합기억을 기억 앙상블(memory ensemble)이라는 수학적으로 표현된 정보의 집합으로 정의하고, 이를 기억하고 회상할 수 있으면서 뇌신경망의 특징들을 잘 살린 확률적 재귀형 신경망 (stochastic recurrent neural network) 모델을 고안하였으며, 이 모델을 학습하기 위한 오차 기대값 역전파 알고리즘(expected error backpropagation algorithm)을 유도하였다. 또한, 시뮬레이션을 통해 이 모델과 알고리즘이 단순한 기억, 조회뿐만 아니라 확률적인 회상과 이행 추론(transitive inference)까지 할 수 있음을 보였다.

……

주제어: 인공신경망, 기억 앙상블, 재귀형 신경망, 연합기억, 역전파 알고리즘

학번: 2006-20485

목 차

그림 목차

표 목차

1. 서 론

우리가 어떤 꽃을 처음 보게 되면 <그림1>과 같이 그 향기, 모습, 이름 등의 다양한 정보들을 한꺼번에 기억하게 된다. 일단 기억하고 나면 나중에 향기나 사진과 같이 그 꽃에 대한 부분적인 정보만 접하게 되어도 나머지 정보들까지 머리 속에 떠오르게 된다. 그 꽃에 관한 정보들이 서로 연결되어 기억되는 것이다. 개인적 경험에 대한 기억도 그 내용뿐만 아니라, 겪은 때와 장소까지 포함하여 저장된다. 이와 같이 입력된 정보들이 서로 연결, 연관되면서 저장되는 기억을 연합기억(associative memory)이라 한다.

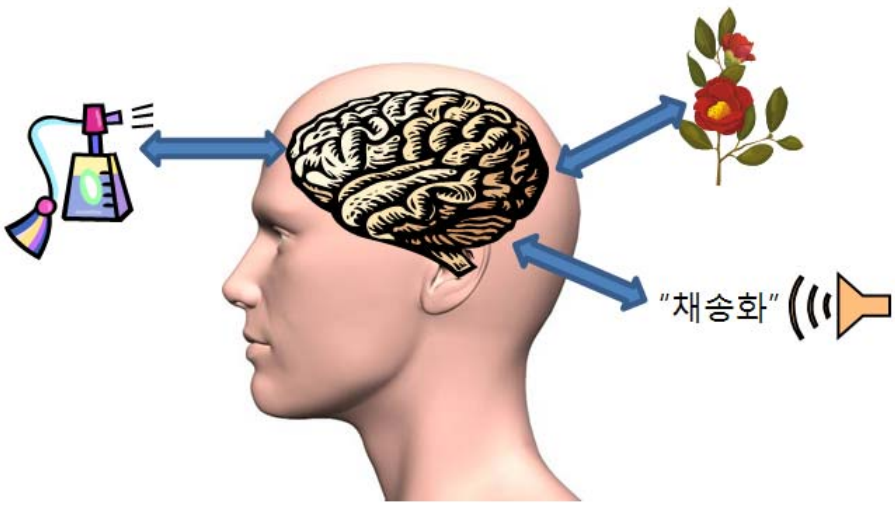

그림 1. 연합기억의 예. 꽃의 모습, 향기, 이름을 서로 연결하면서 기억한다.

기억 현상은 뇌 속의 신경세포간 연결(시냅스 synapse)의 변화로 설명되는데, 이를 수학적으로 모델링하여 연구하는 시도가 계속되고

있다. 특히 연합기억을 인공 신경망(artificial neural network) 모델에 적용하여 그 동안 많은 연구 및 응용 성과들을 내고 있다. 하지만 이들 모델은 뇌신경망의 여러 가지 특징들을 충분히 반영하지 못하고 있다.

본 논문은 연합기억과 뇌신경망의 특징을 보다 잘 살린 확률적 재귀형 신경망 (stochastic recurrent neural network) 모델과 오차 기대값 역전파 알고리즘(expected error backpropagation algorithm)을 소개한다.

2장과 3장에서 연합기억과 뇌신경망의 특징을 살펴보고, 대표적인 신경망 모델들의 특징과 단점을 소개한다. 4장에서는 연합기억을 수학적 표현으로 단순화한 개념인 기억 앙상블과 이를 저장하게 될 확률적 재귀형 신경망 모델을 소개하고 5장에서 이들을 학습시킬 오차 기대값 역전파 알고리즘에 대해 설명한 다음, 6장과 7장에서는 시뮬레이션 결과를 통해 이 모델과 알고리즘의 가능성에 대해 이야기하겠다.

2. 연합기억(Associative Memory)과 뇌신경망의 특징

학습이란 우리가 세상에서 지식을 습득하는 과정을 뜻하며, 기억이란 그 지식이 부호화되고, 저장되며, 이후에 조회되는 과정을 말한다.[1] 기억은 여러 기준에 따라 다양하게 분류되는데, 기억되는 정보 사이의 연관 유무에 따라 나눠본다면, 연합기억 (associative memory)과 비연합기억(nonassociative memory)으로 구분된다. 입력된 정보들이 서로 연결, 연관되면서 저장되는 기억을 연합기억이라 하는데, 여기에는 서술기억(declarative memory)과 고전적 조건화(classical conditioning), 절차기억(procedural memory)등이 포함된다.

서술기억은 의식적으로 저장하거나 회상해 낼 수 있는 기억이다. 사실 또는 경험에 대한 기억이 이 부류에 해당하며, 이는 의미기억(semantic memory)과 일화기억 (episodic memory)으로 다시 구분된다. 의미기억은 대상간의 관계 또는 단어 의미들간의 관계에 관한 지식을 뜻한다. 사물의 이름이나 성질과 같은 지식을 기억하는 것이다. 반면에 일화기억은 개인의 경험 즉 자전적 사건에 관한 기억으로서, 사건이 일어난 시간과 장소, 상황 등의 맥락을 함께 포함한다. 예를 들어, 오늘 식당에서 점심을 먹은 일, 몇 년 전에

[1] (Kandel, Schwartz, & Jessell, 2000)

친구와 영화를 본 일들에 대한 기억이다. 감각영역들을 통해 인지되면서 분해된 다양한 정보들이 해마(hippocampus)를 거치면서 서로 연결되어 서술기억을 형성하고, 이후 오랜 시간에 걸쳐 연합영역(associative cortices)으로 옮겨진다. 이런 기억 정보들 간의 연결은 의미기억과 일화기억 모두에서 일어난다. 의미기억의 경우 그 내용이 다른 지식들과 연관되면서 형성되고, 일화기억 역시 사건이 일어난 때와 장소가 연결되어 기억된다.[2] [3]

고전적 조건화는 조건 자극(conditioned stimulus)과 무조건 자극(unconditioned stimulus) 사이의 관계에 대한 학습이다. 의식적으로 저장, 회상할 수 없는 암묵적 기억(implicit memory)의 일종으로, '파블로프의 개' 실험이 대표적인 예다. 이 실험에서 개는 종소리라는 조건 자극과 잠시 후에 주어지는 음식이라는 무조건 자극 사이의 관계를 기억한 것이다. 이는 시간 차를 두고 입력되는 정보들이 서로 연결 되면서 기억된 것으로 볼 수 있다.

절차기억(procedural memory)은 동작(자전거 타는 법, 춤)이나 인지적 기술(읽기 기술)에 대한 학습을 통해 습득되는 기억을 말한다.[4] 이 역시 암묵적 기억의 일종인데, 상황과 동작, 동작과 동작 사이의 관계를 학습하는 것으로 볼 수 있다.

[2] (Kandel, Schwartz, & Jessell, 2000)

[3] (Eichenbaum, 2000)

[4] (Gazzaniga, Ivry, & Mangun, 2002)

이런 연합기억을 인공 신경망 모델을 사용해서 모델링 할 때는 다음과 같은 뇌신경망의 특징들을 고려해야 한다.

◆ 양방향성 (bidirectional network)

인공신경망을 사용한 모델링을 할 때 감각영역은 입력뉴런으로 단순화되는 경우가 많다. 하지만 O'Craven의 실험에 따르면 무엇을 볼 때뿐만 아니라 그것을 떠올릴 때도 시각영역은 활성화된다.[5] 따라서 감각 영역이 기억의 입력 채널일 뿐만 아니라 출력 채널이기도 하다는 것을 유추해 볼 때, 감각 영역을 입출력이 모두 가능한 모델로 구상하는 것이 더 적절할 것이다.

◆ 다양식성 (multimodal network)

지식이나 경험이 저장될 때 정보들은 다양한 감각기관을 통해 들어온다. 연합기억을 저장할 때 하나 이상의 정보들이 연관되는데,[6] 이것은 다수의 입출력 채널을 가질 수 있는 모델이 필요하다는 의미이다. 또한, 감각기관으로 의미 있는 정보가 항상 들어오는 것은 아니기 때문에 일부 채널이 비어있을 때에도 학습이 가능해야 할 것이다.

◆ 재귀성 (recurrent network)

[5] (O'Craven & Kanwisher, 2000)

[6] (Kandel, Schwartz, & Jessell, 2000)

재귀형 신경망(recurrent neural network)은 뉴런간에 피드백 회로(feedback loop)가 존재하는 신경망이며, 어떤 동적 시스템도 흉내 낼 수 있다.[7] 또한 피드백 회로로 인해 시간적 특성 처리, 단기기억 유지, 기억 용량 증대 등의 효과를 가진다. 뇌신경망의 기본 구조 역시 재귀형 신경망이며, 뇌의 동적 성질을 모델링하기 위해서는 재귀형 신경망이 필수적이다.[8]

◆ Spike trains

실제 뇌신경망에서 정보는 spike train으로 부호화되어 처리된다. 하지만 많은 인공 신경망 모델들이 spike train 대신에 발화율(firing rate)로 정보를 부호화하여 처리하는데, 이로 인해 수학적으로는 간단해지지만 spike train에 담겨있는 시간적 특성을 간과하게 된다. 반면에 spike train을 사용하는 경우에는 뉴런 사이에 정보가 전달될 때 계산 부담이 줄어드는 효과도 있다고 한다.[9]

◆ 확률적 동작 (probabilistic behavior)

뇌신경계는 동작할 때 발생하는 다양한 잡음들 때문에 확률적인 특성을 가진다. 특히 시냅스에서의 신호전달은 확률적인 특성을 띠는데, 시냅스전 말단(presynaptic terminal)에 도달하는 활동전위(action potential) 중에서 절반 이하 만이 시냅스후

[7] (Doya, Universality of Fully- Connected Recurrent Neural Networks, 1993)
[8] (Doya, Recurrent networks: learning algorithms, 2003)
[9] (Bohte & Kok, 2005)

(postsynaptic) 반응을 일으킨다고 한다. LTP(long-term potentiation)와 LTD(long-term depression)를 이런 전달확률의 변화로 설명하기도 한다.[10]

따라서 보다 실제에 가깝게 기억 현상을 연구하기 위해서는 결정론적 모델(deterministic model)이 아닌 확률적 모델(probabilistic model)이 필요하다. 또한 확률적 모델을 사용하는 경우에는 확률적 추론이 가능해진다는 이점도 있다.

◆ 이행 추론 (transitive inference)

이행 추론이란 A>B 와 B>C를 가지고 A>C를 유추하는 것을 말한다.[11] 다시 말해 이미 알고 있는 상관관계를 기반으로 하여, 아직 알지 못하는 상관관계를 추론하는 능력을 이행 추론(transitive inference)이라 하며, 이는 논리적 추리(logical reasoning)에 필수적인 요소이다.[12] 기억 모델이 단순한 저장, 조회 기능에만 머물지 않으려면 이런 이행추론 능력을 통해 저장된 기억을 사용해서 보다 고등한 기능을 실행할 수 있도록 해야 할 것이다.

[10] (Graham & Willshaw, 1999)

[11] (Eichenbaum, 2000)

[12] (Grosenick, Clement, & Fernald, 2007)

3. 연합기억을 위한 인공신경망 모델들

인공신경망 분야에서도 연합기억은 정보들이 서로 연관되면서 쌍을 이루어 기억된다는 개념으로 많이 연구되어 왔다.[13] 이 장에서는 연합기억을 저장하고 조회할 수 있도록 고안된 기존 모델들 중 유명한 것을 몇 가지 소개하겠다.

◆ Hopfield Network

Hopfield network는 John J. Hopfield가 제안한 연합기억 모델로서 입력이 기준값(threshold) 이상인 경우에만 발화하는 비교적 간단한 뉴런들끼리 완전하게 연결된(full-connected) 인공신경망이다. 신경망의 에너지 함수를 정의하고 에너지가 줄어드는 방향으로 수렴시키는 방식으로 학습시키는데, 연합기억뿐만 아니라 순회 세일즈맨 문제(traveling salesman problem)과 같은 최적화 문제에서도 좋은 성능을 보였다고 한다.[14]

하지만, 연결 가중치의 대칭성($w_{ij} = w_{ji}$)이나 뉴런 자신에게로의 feedback 회로를 금지한 제약($w_{ii} = 0$)은 실제 뇌신경망의 구조와 동떨어진 부분이다.

◆ Helmholtz Machine

Helmholtz machine은 Geoff Hinton에 의해 고안된 확률적 생성

[13] (Dayan & Abbott, 2001)

[14] (Haykin, 1998)

모델(probabilistic generative model)의 일종으로 비감독학습(unsupervised learning)을 통해 주어진 데이터의 확률 모델을 만들어 낸다. Helmholtz machine은 하나의 생성 네트워크 (generative network)와 인식 네트워크(recognition network)의 쌍으로 구성되는데, 인식 모델은 하나의 데이터 또는 패턴이 주어질 때 그 데이터에 내재된 특성들의 확률 분포를 추정하는데 이용되며, 생성 모델은 그 내부적 표현으로부터 입력 데이터를 추정함으로써 이러한 인식모델을 학습시키는데 사용된다.[15]

Helmholtz machine은 입력 뉴런에서 출력도 일어나는 양방향성을 가진 모델이지만, 재귀형 신경망이 아니라 두 개의 feedforward 신경망이 겹쳐져 있는 형태라는 점과 spike train을 처리하지 못한다는 점에서 뇌신경망의 특징을 만족시키지 못하고 있다.

◆ Liquid State Machine

Liquid state machine은 liquid라고 불리는 재귀형 신경망과 여기에 feedforward 형식으로 연결된 readout 뉴런들로 이루어진 인공신경망 모델이다. 이 모델의 특징은 liquid는 무작위로 생성하고, readout 뉴런들로의 연결가중치만 학습한다는 점이다. 단순해 보이지만 어떤 함수도 근사할 수 있으며, 다양한 분야에서 활발하게 연구되고

[15] (장정호, 김유섭, 장병탁, 2003)

있다.[16]

인공신경망에서 많이 사용하는 퍼셉트론(perceptron) 뿐만 아니라, 실제 신경세포를 상세하게 모델링한 뉴런 모델까지도 사용할 수 있기 때문에 연속적인(countinuous) 신호도 처리할 수 있다. 하지만 입력 뉴런과 출력 뉴런이 따로 존재하여 양방향성을 만족시키지 못한다.

[16] (Bohte & Kok, 2005)

4. 확률적 재귀형 신경망
(Stochastic Recurrent Neural Network)

이 논문에서는 그림 1 과 같이 각각의 감각영역을 통해 들어온 정보들이 서로 연결되면서 기억되는 현상에 적합한 모델을 제안하기 위해, 감각영역을 통해 들어오는 정보들은 이진 수열로, 뇌신경망은 피드백 회로를 가진 인공신경망(recurrent artificial neural network)으로 모델링 하였다.

4.1. 기억 앙상블 (Memory Ensemble)

동시에 입력된 정보들은 서로 연결되면서 기억되는데, 이렇게 한번에 기억되는 정보들의 집합을 *기억 앙상블* (memory ensemble) 이라 정의하자. 그림 1 에서는 꽃의 모습, 향기, 이름이 하나의 기억 앙상블이 될 것이다.

기억 앙상블들을 여러 개 기억해야 하는 상황을 생각해보자. $\mathbb{M}^{(s)}$ 를 s번째 기억 앙상블이라 한다면, 기억해야 할 모든 기억 앙상블들의 집합을 $\mathbb{M}=\left\{\mathbb{M}^{(1)},\mathbb{M}^{(2)},\cdots,\mathbb{M}^{(s)},\cdots,\mathbb{M}^{(S)}\right\}$ 로 정의할 수 있다. 이때 하나의 기억 앙상블 $\mathbb{M}^{(s)}$ 는 다양한 채널, 다시 말해 뉴런을 통해 입력되는 정보들(향기, 모습, 소리, 색깔 등)의 집합으로 정의된다.

$$\mathbb{M}^{(s)}=\left\{\mathbf{m}_k^{(s)} \mid k\in\mathbb{K}^{(s)},\mathbb{K}^{(s)}\subset\{1,\cdots,K\}\right\}$$

$\mathbf{m}_k^{(s)}$는 s 번째 기억 앙상블 안에서 k 번째 뉴런을 통해 전달되는 정보를 말한다. 이렇게 외부와 정보 교환을 할 수 있는 뉴런을 *단말 뉴런* (terminal neuron), 그렇지 못한 뉴런을 *은닉 뉴런* (hidden neuron)이라 하자. K는 신경망 내에 존재하는 단말 뉴런의 개수를 뜻한다.

그런데, 모습, 향기, 소리와 같은 정보들은 감각기관을 통해 신경세포로 전달되면서 spike train으로 변환되어 처리된다. 우리가 컴퓨터와 인터넷을 통해 문자, 이미지, 동영상 같은 다양한 정보들을 주고 받지만, 실제 그 내부에서는 디지털화 된 전자 신호로 바뀌어 처리되는 것과 마찬가지라 볼 수 있다. 이 논문에서는 spike train 대신에 수학적으로 간단하면서도 시간적인 성질을 살릴 수 있는 0 과 1로 이루어진 $1\times T$ 의 벡터를 사용해 기억 정보를 표현하겠다. 0은 뉴런이 발화하지 않은 상태, 1은 발화한 상태를 뜻한다. 그리고 이 정보는 $\tau_k^{(s)}$ 의 주기로 반복된다고 하자. 그러므로 다음과 같이 표현된다.

$$\mathbf{m}_k^{(s)} = \left[m_k^{(s)}(1) \quad m_k^{(s)}(2) \quad \cdots \quad m_k^{(s)}(t) \quad \cdots \quad m_k^{(s)}(T) \right]$$

$$m_k^{(s)}(t) \in \{0,1\} \qquad m_k^{(s)}(t) = m_k^{(s)}(t + \tau_k^{(s)})$$

예를 들어 학습될 기억 앙상블이 모두 5개라면 $S=5$ 이고, $\mathbb{M} = \left\{ \mathbb{M}^{(1)}, \mathbb{M}^{(2)}, \mathbb{M}^{(3)}, \mathbb{M}^{(4)}, \mathbb{M}^{(5)} \right\}$ 라 표현한다. 이들 중에서 2번째 기억 앙상블의 정보들이 1번, 2번 4번 뉴런을 통해 입출력 된다면

$\mathbb{M}^{(2)} = \left\{ \mathbf{m}_1^{(2)}, \mathbf{m}_2^{(2)}, \mathbf{m}_4^{(2)} \right\}$ 이 될 것이다. 이 때 1번 뉴런을 통하는 정보가 패턴 “11000”이 4회 반복되는 것이라면, 주기는 $\tau_1^{(2)} = 5$ 이고 정보는 $\mathbf{m}_1^{(2)} = \left[1\,1\,0\,0\,0\,1\,1\,0\,0\,0\,1\,1\,0\,0\,0\,1\,1\,0\,0\,0 \right]$ 라는 1×20 벡터로 표현된다.

4.2. 확률 뉴런 (Stochastic Neuron)

뉴런은 신경 세포의 모델로서 인공 신경망의 기본 구성 단위가 되고, 망의 성질에 큰 영향을 끼친다. 일반적으로 퍼셉트론 (perceptron) 모델이 널리 사용되는데, 필요한 망의 성질을 이끌어내기 위해 다양하게 변화시킬 수 있다. Activation function을 logistic sigmoid 함수에서 tanh 함수나, 선형 함수로 바꾸는 경우가 많다.

이 논문에서는 Helmholtz machine의 *확률 뉴런* (stochastic neuron)을 조금 변형해 신경망을 구성했다. 확률뉴런은 퍼셉트론처럼 모든 입력 값의 합이 sigmoid 함수를 통해 계산되어 나오는 0에서 1사이의 실수 값을 출력 값으로 삼지 않고, 이를 발화 확률로 생각하여 확률적으로 0 또는 1을 출력한다.

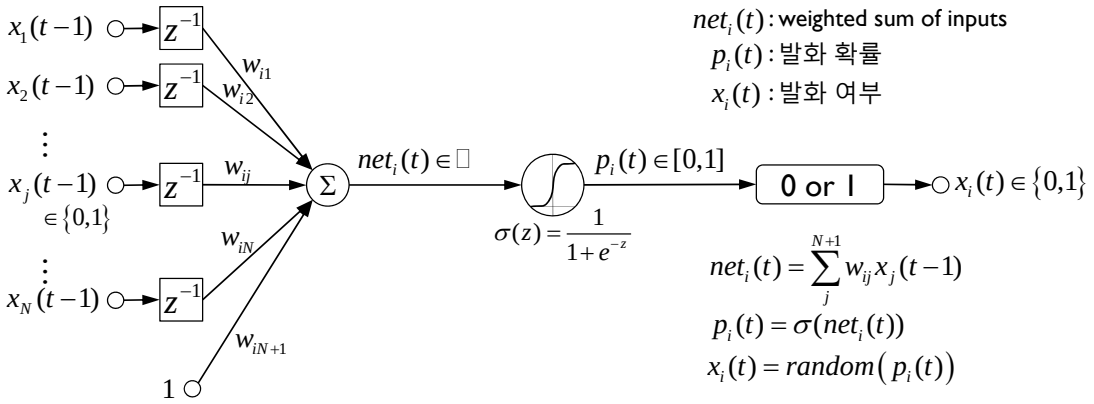

그림 2. 외부 입력이 없을 때의 확률 뉴런. 뉴런들의 이전 상태들을 weighted sum 하고, sigmoid 함수에 넣어 발화 확률로 만든 다음, 이 발화확률에 따라 확률적으로 뉴런의 발화 상태를 0또는 1로 결정한다.

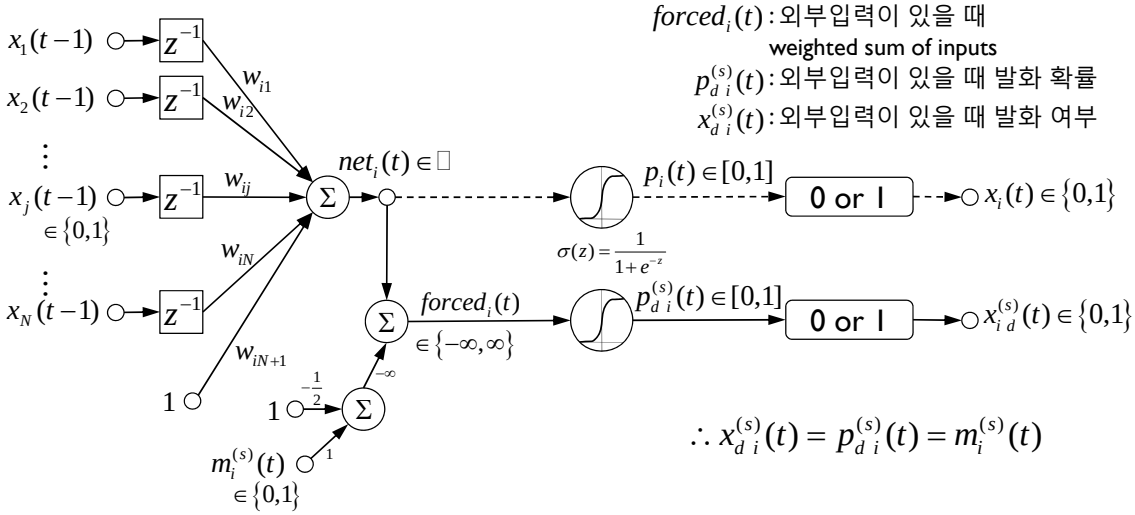

그림 3. 외부입력이 있을 때의 확률 뉴런. 외부입력 그대로 발화한다. 하지만 내부적으로 외부입력이 없다면 어떻게 동작할지도 계산해둔다.

이 논문의 확률 뉴런이 Helmholtz machine의 그것과 다른 점은 '신경망 내부의 뉴런에서 들어오는 입력이 아닌 외부 입력, 즉 기억 앙상블이 입력될 때에는 무조건 그 외부입력과 동일한 출력을 낸다는 것'이다. 이런 특징은 강제 학습 기법(teacher forcing technique)을 사용할 수 있게 해준다.[17]

4.3. 확률적 재귀형 신경망 모델 (Stochastic Recurrent Neural Network Model)

앞서 정의한 기억 앙상블의 학습을 위해 고안된 *확률적 재귀형 신경망* (Stochastic recurrent neural network)은 확률 뉴런들로 이루어진 인공신경망 모델이다. 뉴런들이 모두 같은 지연 시간(delay)를 가지고 있다고 가정하고 있으며(discrete and synchronous system), 각각의 뉴런들은 외부 입출력 유무에 따라 단말 뉴런(terminal neuron)과 은닉 뉴런(hidden neuron)으로 나뉜다. 단말 뉴런을 통해 기억 앙상블 신호가 입력되거나 기억된 정보를 조회 할 수 있고, 은닉 뉴런들로 인해 신경망은 다양한 내부 상태를 가질 수 있게 된다. 뇌신경망의 감각영역을 단말 뉴런들로, 연합영역을 은닉 뉴런들로 옮긴 셈이다.

[17] (Doya, Recurrent networks: learning algorithms, 2003)

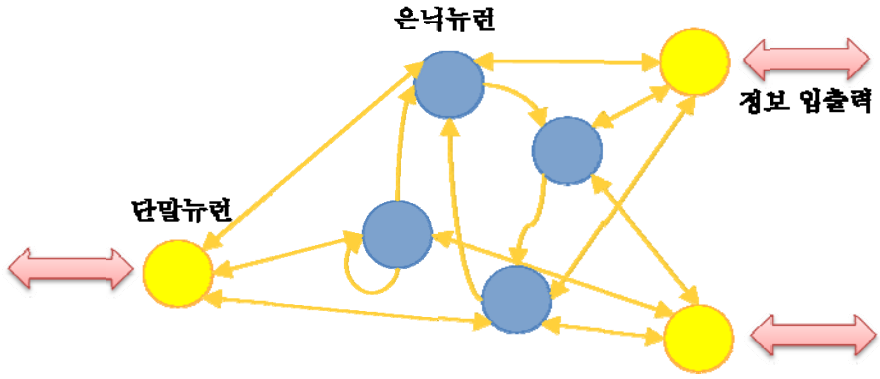

그림 4. 확률적 재귀형 신경망 모델. 확률 뉴런들로 구성된 재귀형 신경망 모델이며, 정보 입출력이 이루어지는 단말뉴런과 그렇지 않은 은닉뉴런으로 구분된다.

그렇다면 이 모델에서 학습과 조회는 어떤 것인가? 확률적 재귀형 신경망이 어떤 기억 앙상블의 모든 정보를 제대로 학습했다는 것은 그 기억 앙상블의 정보들 중에서 일부만 입력하여도 나머지 정보들이 출력된다는 것이다. 예를 들어 기억 앙상블 $\mathbb{M}^{(2)}=\left\{\mathbf{m}_1^{(2)},\mathbf{m}_2^{(2)},\mathbf{m}_4^{(2)}\right\}$ 를 제대로 학습했다면 1번 뉴런에 $\mathbf{m}_1^{(2)}$ 만 입력해도 2번 뉴런에서는 $\mathbf{m}_2^{(2)}$ 가, 4번 뉴런에서는 $\mathbf{m}_4^{(2)}$ 가 출력될 것이다. 이를 위해 학습 단계에서는 기억 앙상블의 모든 정보를 입력하면서 연결 가중치를 조절하고, 회상 단계에서는 기억 앙상블의 일부 정보만 입력한 다음 다른 정보들이 제대로 출력되는지를 관찰해야 한다.

확률적 재귀형 신경망 모델은 앞서 설명했던 뇌신경망의 특징을 만족시키도록 설계되었다.

◆ 양방향성: 뇌신경망의 감각영역 역할을 하는 단말 뉴런들을 통해 입력과 출력이 이루어진다. 단말뉴런의 발화 패턴은 기억을 저장할 때는 외부입력을 반영하고, 조회할 때는 신경망에 저장된 기억 내용을 표현한다.

◆ 다양식성: 다양한 외부 자극들이 감각세포를 거치면서 신경세포의 발화 신호로 바뀌어 처리되듯이. 어떤 종류의 정보도 이진 수열로 인코딩만 해주면 확률적 재귀형 신경망 모델로 학습할 수 있다. 또한 단말 뉴런의 개수를 조절해 많은 종류의 정보들을 한꺼번에 연관시켜 학습시킬 수도 있다.

◆ 재귀적 구조: 뇌신경망의 구조와 같이 피드백 회로를 가진다.

◆ Spike trains: 이 모델에서 사용되는 이진 수열은 spike train을 단순화한 것이지만 시간적 특성을 여전히 표현할 수 있다.

◆ 확률적 동작: 확률 뉴런을 사용함으로써 뇌신경망의 확률적 동작을 흉내 낼 수 있게 되었다.

◆ 이행 추론: 6장에서 실제 실행 결과로 이행 추론이 가능함을 보이겠다.

5. 오차 기대값 역전파 알고리즘
(Expected Error Backpropagation Algorithm)

5.1. 오차 역전파 알고리즘(Error Backpropagation Algorithm)

1974년 Paul Werbos에 의해 처음 개발된 오차 역전파 알고리즘(error backpropagation algorithm)은 가장 기본적인 인공신경망 지도학습(supervised learning: 원하는 목표값이 제공되는 학습) 알고리즘이다. 은닉 뉴런(hidden neuron)들이 들어간 multi-layer perceptron과 함께 등장하여 이전까지 풀 수 없었던 형태의 문제까지 해결하였고, 이후 다양한 분야에서 응용되고 있다.

오차 역전파 알고리즘은 오차함수(loss function)를 정하고, *연결가중치*(weight)에 대한 기울기를 구해서 오차함수를 최소화시키는 gradient descent 기법을 사용한다. 오차 역전파 알고리즘은 크게 정방향 단계(forward phase)와 역방향 단계(backward phase)로 나뉜다. 정방향 단계에서는 신경망에 학습 데이터를 입력하고 실행시켜서 각 뉴런에서의 실제 출력값을 구한다. 그런 다음 역방향 단계에서 실제 출력값과 목표 출력값의 차이를 가지고 연결 가중치의 변화량을 결정하게 된다. 오차함수는 목표 출력값 $d_i^{(s)}$와 실제 출력값 x_i간의 차이를 제곱합한 것이다.

$$\varepsilon = \frac{1}{2}\sum_{s=1}^{S}\sum_{i=1}^{N}(x_i - d_i^{(s)})^2$$

그리고 이를 chain rule을 이용해 각각의 연결가중치들에 대한

편미분 방정식을 유도한 다음에 아래 식에 따라 연결가중치를 조금씩 변화시키면서 학습한다.

$$\Delta w_{ij}(l) = \alpha \cdot \Delta w_{ij}(l-1) - \mu \cdot \frac{\partial \varepsilon}{\partial w_{ij}}$$

이런 방식을 gradient descent 라 부르는데, 학습 속도가 느리고 지역 최소점(local minimum)에 빠질 위험이 높다. 이런 위험을 줄이기 위해 이전 연결가중치 변화량을 일정 비율 반영하는 momentum 기법이 많이 사용된다. 그리고 연결 가중치들의 초기값, 학습의 속도를 결정하는 학습 계수(learning rate μ), 모멘텀 계수(momentum factor α)와 같은 파라미터들이 학습 성공 여부에 큰 영향을 끼친다.[18]

5.2. Backpropagation through Time Algorithm

Backpropagation through time algorithm은 feedforward형 (피드백 회로가 없는) 인공 신경망을 위한 오차 역전파 알고리즘을 피드백 회로가 있는 재귀형 신경망(recurrent neural network)에 적용한 학습 알고리즘이다. 연결가중치의 변화가 간접적으로 에러에 미치는 영향까지 감안하기 위해 'ordered partial derivatives'(Box 1) [19] 로 기울기를 계산한다. 음성 인식, 유량 예측, 태양 흑점 예측과 같이

[18] (Haykin, 1998)

[19] (KUMAR, RAJU, & SATHISH, 2004)

시간에 따른 변화를 학습하는데 유용하게 사용된다.[20]

Box 1. Ordered Partial Derivatives

순서가 정해진 N개의 식들에 의해 정의되는 N개의 변수들, $[Z_1, Z_2, ..., Z_N]$ 을 생각해보자. Z_i 는 순서상 앞에 오는 변수들, $[Z_1, Z_2, ..., Z_{i-1}]$ 에 의해 정의된다. 즉 Z_i 의 값이 결정되기 전에 $[Z_1, Z_2, ..., Z_{i-1}]$ 들이 먼저 결정되어야 한다는 것이다.

다음 식들이 그 예이다.

$$Z_1 = 1$$

$$Z_2 = 3Z_1$$

$$Z_3 = Z_1 + 2Z_2$$

이 경우 Z_1 이 변할 때 Z_3 가 변화하는 양, 즉 Z_3 를 Z_1 으로 편미분한 값을 구할 때, 세 번째 식뿐만 아니라 두 번째 식에 의해 간접적으로 미치는 영향까지 고려해야 하는 것이다.

Ordered partial derivatives의 공식은 다음과 같다.

$$\text{If } j \le i \quad \frac{\partial^+ Z_j}{\partial Z_i} = 0$$

$$\text{If } j = i+1 \quad \frac{\partial^+ Z_j}{\partial Z_i} = \frac{\partial Z_j}{\partial Z_i}$$

[20] (Werbose, 1990)

If $j > i+1$ $\dfrac{\partial^+ Z_j}{\partial Z_i} = \dfrac{\partial Z_j}{\partial Z_i} + \sum_{k=i+1}^{j-1} \left(\dfrac{\partial^+ Z_j}{\partial Z_k} \cdot \dfrac{\partial Z_k}{\partial Z_i} \right)$

or $\dfrac{\partial^+ Z_j}{\partial Z_i} = \dfrac{\partial Z_j}{\partial Z_i} + \sum_{k=i+1}^{j-1} \left(\dfrac{\partial Z_j}{\partial Z_k} \cdot \dfrac{\partial^+ Z_k}{\partial Z_i} \right)$

따라서 위의 예는 다음과 같이 계산된다.

$$\frac{\partial^+ Z_3}{\partial Z_1} = \frac{\partial Z_3}{\partial Z_1} + \frac{\partial Z_3}{\partial Z_2} \cdot \frac{\partial Z_2}{\partial Z_1} = 1 + 2 \cdot 3 = 7$$

5.3. 오차 기대값 역전파 알고리즘 (Expected Error Backpropagation Through Time Algorithm)

오차 기대값 역전파 알고리즘(expected error backpropagation through time algorithm)은 기억 앙상블을 확률적 재귀형 신경망에 학습시키기 위해 기존의 backpropagation through time 알고리즘을 바탕으로 고안되었다.

앞서 설명했던 것처럼 신경망이 기억 앙상블을 제대로 학습했다고 말하기 위해서는 그 기억 앙상블의 정보들 중 일부가 빠져서 입력되는 상황에서도 신경망의 뉴런들은 모든 정보들이 제대로 들어올 때와 똑같은 행동을 보여야 한다.

이를 위해 오차 기대값 역전파 알고리즘에서는 강제 학습(teacher forcing technique)을 시킨다. 춤을 가르치는 광경을 생각해보자. 선생님이 학생의 손을 맞잡고 춤을 춘다. 선생님의 움직임에 학생이 따라오게 되면서 학생은 적절한 동작과 자세, 힘의 배분을 익혀가게

된다. 이 때 선생님은 학생이 춤에 숙달되어감을 어떻게 알 수 있을까? 눈으로 학생을 살필 수도 있겠지만, 대개 맞잡은 손을 통해 느껴지는 저항으로 판단할 것이다. 처음에는 동작과 타이밍이 미숙해 선생님이 힘들게 끌고 다녀야겠지만, 숙달되면서 점점 저항이 줄어들고 마스터하게 되면 선생님과 학생은 완벽하게 호흡을 맞출 수 있을 것이다.

오차 기대값 역전파 알고리즘도 이와 같은 원리다. 학습할 때 각 단말뉴런에 기억 앙상블의 정보를 입력시키고, 해당 단말 뉴런들은 그 정보의 패턴과 똑같은 행동($x_{d\,i}^{(S)}(t)$)을 하도록 강제된다. 하지만, 내부적으로는 외부 입력이 없을 때 그 단말뉴런이 어떻게 행동할지($x_i(t)$)를 계산해서 그 차이가 줄어들도록 신경망의 연결가중치를 조금씩 조절해 나가는 것이다. 그리고, $x_{d\,i}^{(S)}(t)$ 와 $p_{d\,i}^{(S)}(t)$ 의 값이 같고, $x_i(t)$ 의 기대값이 발화확률 $p_i(t)$ 라는 점을 이용해서, $x_{d\,i}^{(S)}(t)$ 와 $x_i(t)$ 대신에 $p_{d\,i}^{(S)}(t)$ 와 $p_i(t)$ 를 사용해 알고리즘을 유도하였다.

$$p_i(t) = P[x_i(t) = 1]$$

$$\therefore E[x_i(t)] = p_i(t)$$

$$p_i(t) = \sigma\left(\sum_j^{N+1} w_{ij} x_j(t-1)\right) \quad \square \quad \sigma\left(\sum_j^{N+1} w_{ij} E\left[x_j(t-1)\right]\right)$$

$$\therefore p_i(t) = \sigma\left(\sum_j^{N+1} w_{ij}\, p_j(t-1)\right)$$

$$\varepsilon = \sum_{s=1}^{S} \varepsilon^{(s)} = \sum_{s=1}^{S} \sum_{t=0}^{T} \varepsilon^{(s)}(t)$$

$$= \frac{1}{2} \sum_{s=1}^{S} \sum_{t=0}^{T} \sum_{k \in \mathbb{K}^{(s)}} \left(E[x_k(t)] - x_{d\,k}^{(s)}(t) \right)^2$$

$$\therefore \varepsilon = \frac{1}{2} \sum_{s=1}^{S} \sum_{t=0}^{T} \sum_{k \in \mathbb{K}^{(s)}} \left(p_k(t) - p_{d\,k}^{(s)}(t) \right)^2$$

오차함수 역시 $p_{d\,i}^{(S)}(t)$ 와 $p_i(t)$ 로 정의하였다. 이제 이 오차함수를 최소화하는 방향으로 연결가중치를 줄여나가기 위해 오차함수를 연결가중치로 편미분한 공식을 유도해 보자. 연결가중치 변화에 따른 간접적인 영향까지 고려해야 하기 때문에 backpropagation through time 알고리즘과 마찬가지로 ordered partial derivatives를 사용해야 한다.

$$\frac{\partial^+ \varepsilon}{\partial w_{ij}} = \sum_{s=1}^{S} \frac{\partial^+ \varepsilon^{(s)}}{\partial w_{ij}} = \sum_{s=1}^{S} \sum_{t=0}^{T} \frac{\partial^+ \varepsilon^{(s)}(t)}{\partial w_{ij}}$$

$$\frac{\partial^+ \varepsilon^{(s)}(t)}{\partial w_{ij}} = \frac{\partial \varepsilon^{(s)}(t)}{\partial w_{ij}} + \sum_{t'}^{t} \sum_{k_{t'}}^{N} \frac{\partial \varepsilon^{(s)}(t)}{\partial p_{k_{t'}}(t')} \cdot \frac{\partial^+ p_{k_{t'}}(t')}{\partial w_{ij}}$$

$$\frac{\partial \varepsilon^{(s)}(t)}{\partial w_{ij}} = 0$$

$$\frac{\partial \varepsilon^{(s)}(t)}{\partial p_{k_{t'}}(t')} = \begin{cases} p_{k_t}(t) - p_{dk_t}^{(s)}(t) & \text{when } t = t' \text{ and } k_{t'} \in \mathbb{K}^{(s)} \\ 0 & \text{otherwize} \end{cases}$$

$$\therefore \frac{\partial^+ \varepsilon^{(s)}(t)}{\partial w_{ij}} = \sum_{k_t \in \mathbb{K}^{(s)}} \left\{ p_{k_t}(t) - p_{dk_t}^{(s)}(t) \right\} \cdot \frac{\partial^+ p_{k_t}(t)}{\partial w_{ij}}$$

여기서 시간 t일 때, k_i번 뉴런의 발화확률 $p_{k_i}(t)$의 연결가중치 w_{ij} 에 대한 ordered partial derivative는 다음과 같다.

$$\frac{\partial^+ p_{k_t}(t)}{\partial w_{ij}} = \frac{\partial p_{k_t}(t)}{\partial w_{ij}} + \sum_{k_{t-1}}^{N} \frac{\partial p_{k_t}(t)}{\partial p_{k_{t-1}}(t-1)} \cdot \frac{\partial^+ p_{k_{t-1}}(t-1)}{\partial w_{ij}}$$

$$\frac{\partial p_{k_t}(t)}{\partial w_{ij}} = \begin{cases} p_{k_t}(t)\left(1-p_{k_t}(t)\right) p_j(t-1) & \text{when } k_t = i \\ 0 & \text{otherwise} \end{cases}$$

$$\frac{\partial p_{k_t}(t)}{\partial p_{k_{t-1}}(t-1)} = p_{k_t}(t)\left(1-p_{k_t}(t)\right) w_{k_t k_{t-1}}$$

$$\frac{\partial^+ p_{k_0}(0)}{\partial w_{ij}} = \frac{\partial p_{k_0}(0)}{\partial w_{ij}}$$

$$= \begin{cases} p_{k_0}(0)\left(1-p_{k_0}(0)\right) & \text{when } k_0 = i \text{ and } j = N+1 \\ 0 & \text{otherwise} \end{cases}$$

$$p_{k_0}(0) = \sigma\left(w_{k_0 N+1}\right)$$

이렇게 구한 기울기를 아래 공식에 대입하여 매번 반복할 때마다 연결 가중치의 변화량 $\Delta w_{ij}(l)$을 결정하게 된다.

$$\therefore \Delta w_{ij}(l) = \alpha \cdot \Delta w_{ij}(l-1) - \mu \cdot \frac{\partial^+ \varepsilon}{\partial w_{ij}}$$

오차 역전파 알고리즘에서와 마찬가지로 오차 기대값 역전파 알고리즘의 학습도 정방향 단계와 역방향 단계로 나눌 수 있다. 정방향 단계에서는 기억 앙상블의 모든 정보들을 각 뉴런에 입력하면서 신경망을 동작시켜, 모든 뉴런들의 실제 출력값들(발화확률 $p_i(t)$)을 기록해둔다. 그리고 역방향 단계에서는

이 값들을 위의 식에 대입하여 연결 가중치의 변화량을 구해내는 것이다.

그런데, 여기서 학습성능을 향상시키기 위해 고려해야 할 것이 있다. 그것은 정방향 단계에서 확률뉴런이 원래의 공식 $x_i(t) = random\left(\sigma\left(\sum_j^{N+1} w_{ij} x_j(t-1)\right)\right)$ 대로 동작하게 되면 학습성능이 매우 저하된다는 것이다. 그 이유는 알고리즘은 조금씩 신경망을 변화시키면서 최적해를 찾아가야 하는데, 확률뉴런 동작의 무작위성 때문에 알고리즘이 탐색해야 하는 해공간(solution space)이 계속 변하기 때문이다. 이것은 마치 계속 변하는 미로에서 눈감고 더듬으며 길을 찾는 것과 같은 것이다. 이 경우 기억 앙상블 하나를 학습하는 것 같이 아주 쉬운 문제가 아니면 대부분 학습에 실패하였다. 그래서 해공간이 무작위적으로 변화하지 않도록 확률뉴런의 공식에서 random 함수를 임시로 빼놓고 학습을 하도록 하였다. 즉 $p_i(t) = \sigma\left(\sum_j^{N+1} w_{ij} p_j(t-1)\right)$ 으로 확률뉴런을 동작시키는 것이다. 물론 회상할 때는 원래대로 확률뉴런이 동작한다. 이 방법을 사용한 경우 신경망은 각 뉴런의 발화 기대값에 따라 동작하게 되는데, 학습성능이 월등하게 개선되었다.

6. 실행 결과

6.1. 정확도 검사

P. Werbose는 역전파 알고리즘이 제대로 동작하는지를 확인하기 위해 연결가중치를 조금 변화시켰을 때의 오차함수 값의 변화량과 알고리즘으로 구한 오차함수를 연결가중치로 편미분한 값을 비교하는 방법을 제안하였다.[21]

다음은 이 방법에 따라 확률적 재귀형 신경망을 임의로 생성하고 연결가중치 중에서 하나를 미세하게 변화시키면서 오차함수의 값과 오차함수를 연결가중치로 편미분한 값을 비교한 것이다.

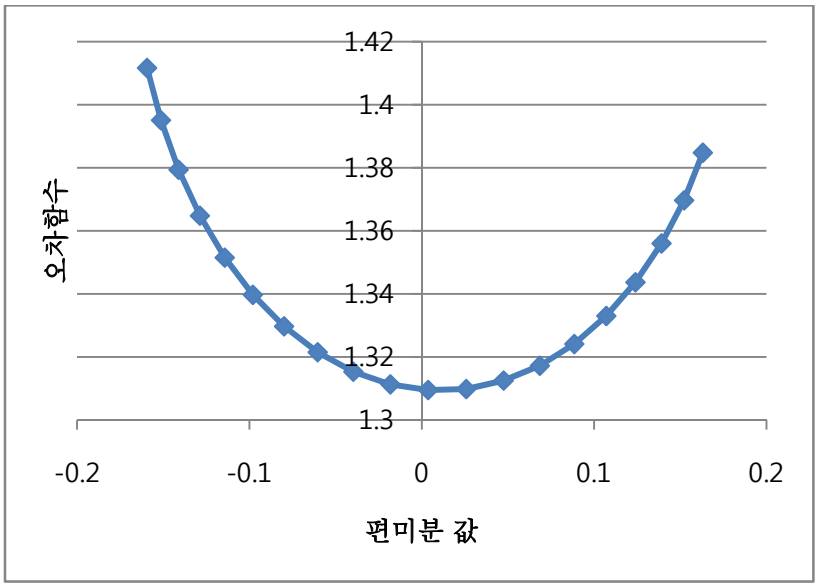

그림 5. 연결가중치 변화에 대한 오차함수와 편미분 값의 변화.

[21] (Werbose, 1990)

그림에서 알 수 있듯이 알고리즘으로 구한 편미분 값이 0이 되는 점과 오차함수의 값이 최소가 되는 점이 거의 일치한다. 따라서 알고리즘이 오차함수를 최소화하는 방향으로 연결 가중치를 제대로 변화시킬 있음을 확인할 수 있다.

6.2. 기본적인 학습 및 조회

이번에는 실제로 기억 앙상블을 학습하고 조회하는 실험을 해보도록 하겠다. 예를 들어 동물의 모습과 이름의 철자, 그리고 이름을 읽는 소리를 기억시키려 한다고 해보자. 다음 기억 앙상블이 그 예가 될 수 있을 것이다.

{{고양이의 모습, cat, kæt}, {소의 모습, cow, kau}}

이 정보들을 $\tau_1^{(s)} = 3$, $T = 15$ 인 이진 수열로 대치하여 수식으로 표현하면 다음과 같다.

$$\mathbb{M} = \left\{ \mathbb{M}^{(1)} = \left\{ \mathbf{m}_1^{(1)} = [0\ 0\ 1\ ...],\ \mathbf{m}_2^{(1)} = [1\ 1\ 0\ ...],\ \mathbf{m}_3^{(1)} = [0\ 1\ 1\ ...] \right\}, \right.$$
$$\left. \mathbb{M}^{(2)} = \left\{ \mathbf{m}_1^{(2)} = [1\ 0\ 1\ ...],\ \mathbf{m}_2^{(2)} = [0\ 1\ 0\ ...],\ \mathbf{m}_3^{(2)} = [1\ 0\ 0\ ...] \right\} \right\}$$

이 기억 앙상블들을 10개의 뉴런들로 이루어진 확률적 재귀형 신경망에 오차 기대값 역전파 알고리즘을 사용하여 학습시켰다. 이때 학습계수는 $\mu = 0.001$, momentum 계수는 $\alpha = 0.25$ 로 학습을 진행했다.

그 결과 아래 표와 같은 연결 가중치를 가진 신경망을 구하였다.

표 1. 동물 기억 앙상블 학습 결과 연결가중치

i \ j	1	2	3	4	5	6	7	8	9	10	11 (bias)
1	-1.16	1.26	-1.83	1.473	-2.55	-1.93	1.943	3.741	-2.45	-0.57	-0.02
2	1.176	-1.54	1.86	-1.52	2.354	2.226	-2.03	-3.65	2.514	0.642	0.169
3	0.297	2.652	-0.37	1.429	4.84	-4.01	1.553	-0.74	-2.85	-0.8	0.268
4	-0.1	0.435	0.078	-0.14	0.226	1.39	-0.32	-0.77	1.284	0.654	-0.22
5	2.202	-2.97	1.492	-0.28	-2.3	-1.51	-1.2	-0.64	-2.71	-2.16	1.234
6	2.533	-2.17	1.665	0.019	0.233	-1.11	-0.13	0.039	-1.14	-0.61	1.962
7	0.02	0.214	-0.17	-0.27	0.61	1.65	-0.56	-1.01	1.826	0.806	-0.22
8	-1.53	1.677	-0.94	-0.63	1.62	1.934	-0.49	-1.09	2.528	1.378	-1.54
9	2.599	-2.71	0.869	-0.39	0.057	-0.73	-0.47	-0.3	-0.92	-0.39	1.954
10	1.368	-1.56	0.1	-0.4	0.526	-0.25	-0.36	-0.4	-0.09	-0.14	0.838

이 신경망이 주어진 기억 앙상블을 잘 기억하고 있는지 확인하기 위해 정보를 조회해 보았다. 목표 패턴을 제외한 정보들을 입력하고 T=300 동안(패턴이 100번 반복될 시간 동안) 출력패턴들을 관찰하였고, 이를 100회씩 반복하였다.

표 2. 동물 기억 앙상블 조회 결과. 목표 패턴과 같은 패턴이 가장 많이 출력된다.

기억 쌍	단말 뉴런	목표 패턴	출력패턴		
			1 위	2 위	3 위
1	1	001	**001**	011	101
			54.72±7.23%	34.13±7.62%	9.11±4.01%
	2	110	**110**	100	010
			53.31±7.76%	37.10±8.41%	7.27±3.62%
	3	011	**011**	001	111
			50.15±5.24%	18.50±3.69%	15.51±3.37%
2	1	101	**101**	100	001
			76.07±5.93%	20.76±5.57%	2.09±1.28%
	2	010	**010**	011	110
			72.62±6.42%	24.04±6.18%	2.18±1.49%
	3	100	**100**	101	000
			55.66±5.20%	14.99±3.36%	10.31±2.76%

목표 패턴과 동일한 출력패턴이 가장 많이 관찰되었음을 알 수 있다. 즉 주어진 기억 앙상블들이 제대로 학습되었음이 확인되었다.

6.3. 확률적 회상

‘곰인형’을 ‘귀엽다’라고 기억하고, ‘강아지’도 ‘귀엽다’라고 기억한 경우를 생각해보자. 이때 ‘귀엽다’라는 입력이 들어온다면 ‘곰인형’과 ‘강아지’가 무작위로 떠오르게 될 것이다. 그리고 강아지보다 곰인형을 많이 경험할수록 ‘귀엽다’라는 입력에 곰인형을 떠올릴 확률이 더 높아질 것이다. 이런 확률적 회상을 할 수 있음을 보이기 위해 학습 빈도를 다르게 하여 학습시킨 다음,

회상할 때 출력되는 패턴들의 횟수를 세어 보았다.

$$\mathbb{M}=\left\{\mathbb{M}^{(1)}=\left\{\mathbf{m}_1^{(1)}=[1\ 0\ 0\ \ldots],\ \mathbf{m}_2^{(1)}=[1\ 0\ 1\ \ldots]\right\},\right.$$
$$\left.\mathbb{M}^{(2)}=\left\{\mathbf{m}_1^{(2)}=[1\ 1\ 0\ \ldots],\ \mathbf{m}_2^{(2)}=[1\ 0\ 1\ \ldots]\right\}\right\}$$

이 기억 앙상블들에서 2번 뉴런에는 항상 ‘101’이 입력되지만 1번 뉴런에 입력되는 패턴은 ‘100’ 또는 ‘110’이다. 그러므로 2번 뉴런에만 ‘101’이 입력될 때 1번 뉴런에서는 ‘100’나 ‘110’이 출력되어야 할 것이다. 그리고 학습할 때 더 많이 입력된 패턴일수록 회상할 때 더 자주 떠올라야 할 것이다.

이를 확인하기 위해 1번 뉴런에 ‘100’이 입력되는 기억 앙상블과 ‘110’이 입력되는 기억 앙상블의 비율을 다르게 하며 학습시켜 보고, 이후 2번 뉴런에만 ‘101’을 입력하면서 1번 뉴런에서 회상되는 패턴들을 세어 보았다.

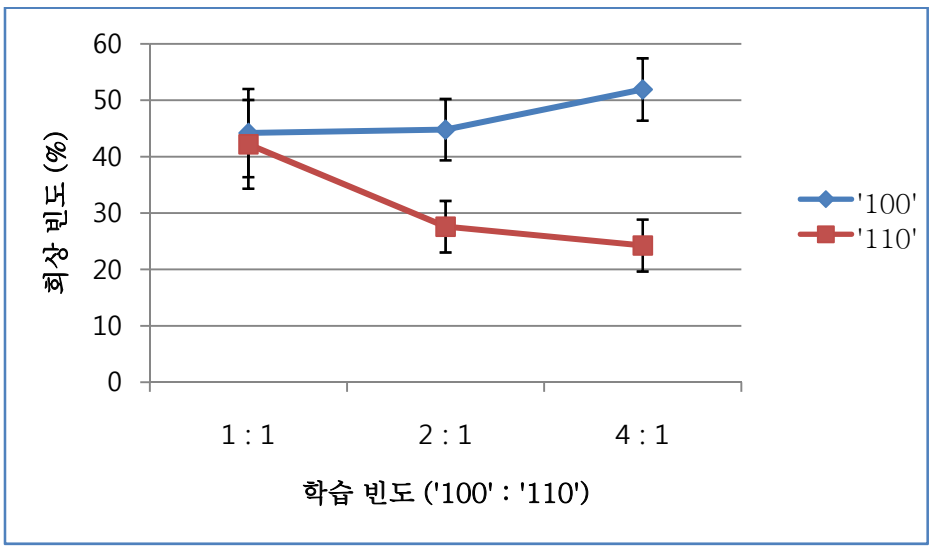

그림 6. 확률적 회상 빈도. 학습할 때의 빈도차이에 따라 회상빈도의

차이도 점점 커진다.

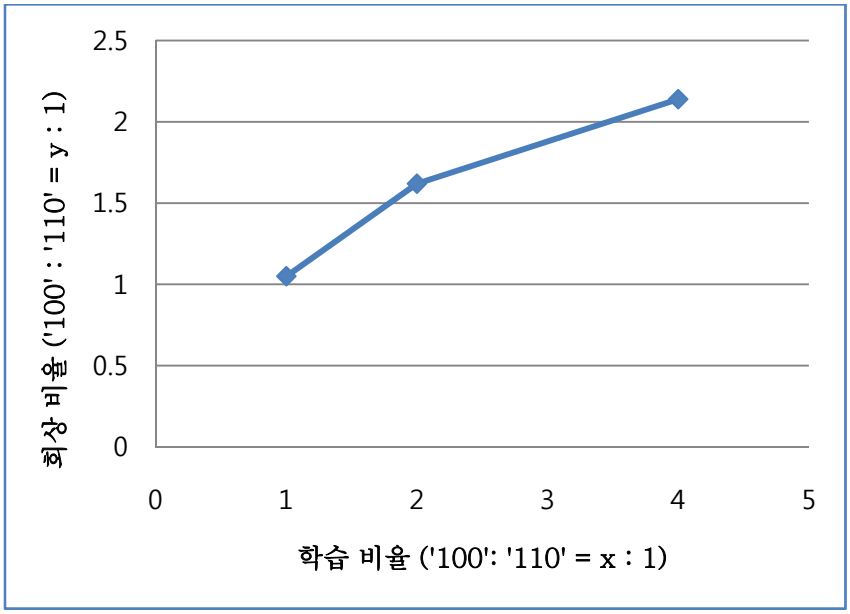

그림 7. 확률적 회상 비율. 학습할 때 기억 앙상블의 비율이 1:1일 때는 회상비율 역시 1:1이지만, 학습 비율이 커질수록 회상비율도 커진다.

그 결과 위 그림과 같이 학습할 때 어느 기억 앙상블을 자주 입력하느냐에 따라 회상할 때 출력되는 패턴들간의 비율이 달라졌다. 자주 학습된 패턴일수록 높은 확률로 회상되는 것을 알 수 있다.

6.4. 이행 추론

이행 추론은 이미 알고 있는 상관관계를 기반으로 하여, 아직 알지 못하는 상관관계를 추론하는 능력이다. A>B 와 B>C를 가지고 A>C를 유추하는 것을 말한다. 고양이의 냄새와 고양이를 연결시켜 기억하고, 고양이와 위험을 관련 지어 기억하는 쥐를 생각해보자.

고양이의 냄새로부터 위험을 빠르게 이행 추론해내지 못해서 도망치지 않는다면 근처에 있는 고양이에게 잡아 먹혀 버릴 것이다. 고양이의 냄새를 1번 뉴런에 입력되는 패턴 '100'에, 고양이를 2번 뉴런에 입력되는 패턴 '101'에, 그리고 위험을 3번 뉴런에 입력되는 패턴 '011'에 대응시켜 보자. 이를 기억 앙상블로 나타내면 아래 식과 같다.

$$\mathbb{M}=\left\{\mathbb{M}^{(1)}=\left\{\mathbf{m}_1^{(1)}=[1\ 0\ 0\ ...],\ \mathbf{m}_2^{(1)}=[1\ 0\ 1\ ...]\right\},\right.$$
$$\left.\mathbb{M}^{(2)}=\left\{\mathbf{m}_2^{(2)}=[1\ 0\ 1\ ...],\ \mathbf{m}_3^{(2)}=[0\ 1\ 1\ ...]\right\}\right\}$$

학습한 결과 연결가중치는 아래 표와 같다.

표 3. 이행 추론을 위한 학습 결과 연결가중치

i \ *j*	1	2	3	4	5	6	7	8	9 (bias)
1	-2.16	2.223	0.236	-0.84	2.86	-1.64	-0.06	-2.92	-0.8
2	-2.23	-1.36	4.915	-0.89	0.795	-0.02	-0.64	-1.1	0.156
3	2.848	-1.99	-0.71	1.312	-1.95	1.886	0.676	3.629	-0.13
4	-0.32	0.476	-0.58	-0.11	0.228	-0.58	-0.11	-0.77	-0.06
5	-0.39	-1.98	1.618	0	-0.82	0.62	-0.12	0.945	-0.53
6	0.28	1.146	-1.11	0.027	0.485	-0.13	-2.12	-0.59	0.263
7	-0.24	-0.13	-0.18	-0.26	0.063	-0.32	-0.42	-0.44	-0.15
8	0.154	2.274	-1.6	-0.01	0.748	-0.76	0.089	-1.15	0.714

이렇게 학습된 신경망에 '100'(고양이 냄새)만 1번 뉴런에 입력시키면서 회상시켜 보았다.

표 4. 이행 추론 확인을 위한 조회시 각 뉴런에서 출력되는 발화 패턴들.

1번 뉴런에 '100'을 입력하면 3번 뉴런에서 '011'이 출력된다.

뉴런	발화 패턴
1 번	100 100 100 100 100 100 100
2 번	101 101 101 101 101 101 101
3 번	111 011 011 011 011 011 011

위의 표와 같이 3번 뉴런에서 위험에 해당하는 패턴 '011'이 성공적으로 발화하고 있음을 확인할 수 있다. '고양이 냄새' 만으로 '위험'하다는 것을 떠올리는 이행 추론에 성공한 것이다.

7. 결 론

지금까지 연합기억과 뇌신경망의 특징들(양방향성, 다양식성, 재귀적 구조, spike trains 형태의 신호 처리, 확률적 동작, 이행 추론), 그리고 연합기억을 학습하기 위해 고안된 인공신경망 모델들에 대해 알아보았고, 확률적 재귀형 신경망 모델과 오차 역전파 알고리즘을 제시하였으며. 이 새로운 모델과 알고리즘이 단순한 기억과 회상뿐만 아니라 확률적 추론과 이행추론도 할 수 있음을 간단한 실험을 통해 보였다.

확률적 재귀형 신경망 모델은 연합기억을 위한 기존의 인공신경망 모델들이 충분히 반영하지 못하고 있는 뇌신경망의 특징들을 구현하고 있으며, 특히 확률적 동작과 이행 추론 능력을 통해서 단순한 기억·조회 기능을 넘어설 가능성을 지니고 있다. 필자는 앞으로 이 모델이 '창의성'을 흉내 낼 수 있을 것이라 기대한다. 창의성의 사전적 의미는 "새로운 것을 생각해 내는 특성"이다. 창의성의 정의를 '기존의 지식과 경험들을 재조합하여 지금까지 없었던 새로운 생각을 우연히 떠올리는 것'으로 한정한다면, 기억과 조회, 확률적 추론, 이행 추론이 가능한 확률적 재귀형 신경망 모델로 어느 정도는 가능할 것이다.

그러나 이런 창의성을 모사해 낼 수 있는 모델이 되기 위해서는 개선되어야 할 점들 역시 많다. 먼저 전반적으로 학습성능이

개선되어야 한다. Gradient descent 계열 학습 알고리즘의 한계로 최적해를 찾지 못할 확률이 높고, 학습할 때는 확률적 동작을 제한해야 하며, 속도도 비교적 느린 편이다. 이 때문에 보다 긴 패턴, 보다 많은 기억 앙상블을 저장하기 위해 뉴런 수를 늘려서 실험하기가 힘들다. 또한 회상할 때 확률적으로 동작하기 때문에 학습 패턴의 길이가 길어질수록 전체 패턴을 정확하게 회상할 확률이 급격하게 떨어진다. 예를 들어 매 순간 정확하게 발화할 확률이 90%라 하더라도, 패턴의 길이가 3이면 정확한 패턴을 출력할 확률은 약 73%(=0.9^3), 패턴의 길이가 10이라면 약 35%(=0.9^{10})까지 떨어지게 된다. 이런 현상을 보완하기 위해 통신기술에서 사용되는 parity bit와 같은 신호 검증, 수정 기술이 필요할 것이다. 그리고 실제 뇌신경망의 특징들을 더 잘 살려야 한다. 다양한 지연시간을 가지고(asynchronous), 뉴런 간의 연결이 성긴(sparse) 신경망 모델과 이에 적합한 알고리즘으로 개선한다면 실제 뇌에서 일어나는 다양한 현상들을 연구하는데 쓰일 수 있을 것이다.

참고문헌

강봉균. (2001). 기억과 시냅스 가소성. "한국뇌학회지", Vol. 1 (No. 1), 13-24.

장정호, 김유섭, & 장병탁. (2003). 헬름홀츠머신 학습 기반의 의미 커널을 이용한 문서 유사도 측정. "한국정보과학회 봄 학술발표 논문집 (B)", 제30권 1호, 페이지: 440-442.

Bohte, S. M., & Kok, J. N. (2005). Applications of spiking neural networks. *Information Processing Letters , 95*, 519-520.

Dayan, P., & Abbott, L. F. (2001). *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems.* Cambridge(MA): MIT Press.

Doya, K. (2003). Recurrent networks: learning algorithms. In M. Arbib (Ed.), *The handbook of brain theory and neural networks* (p. 955±60). Cambridge (MA): MIT Press.

Doya, K. (1993). *Universality of Fully- Connected Recurrent Neural Networks.* San Diego: University of California.

Eichenbaum, H. (2000). A cortical–hippocampal system for declarative memory. *Nature Reviews Neuroscience , 1*, 41-50.

Gazzaniga, M. S., Ivry, R. B., & Mangun, g. R. (2002). *Cognitive Neuroscience.* New York: W. W. Norton & Company, Inc.

Graham, B., & Willshaw, D. (1999). Probabilistic Synaptic Transmission in the Associative Net. *Neural Computation , 11*, 117-137.

Grosenick, L., Clement, T. S., & Fernald, R. D. (2007). Fish can infer social rank by observation alone. *Nature , 445* (7126), 429-432.

Haykin, S. (1998). *Neural Networks: A Comprehensive Foundation* (2nd ed.). Upper Saddle River, NJ, USA: Prentice Hall PTR.

Kandel, E. R., Schwartz, J. H., & Jessell, T. M. (2000). *Principles of Neural Science* (4th ed. ed.). New York: McGraw-Hill.

KUMAR, D. N., RAJU, K. S., & SATHISH, T. (2004). River Flow Forecasting using Recurrent Neural. *Water Resources Management , 18*, 143–161.

O'Craven, K. M., & Kanwisher, N. (2000). Mental Imagery of Faces and Places Activates Corresponding Stimulus-Specific Brain Regions. *Journal of Cognitive Neuroscience , 12* (6), 1013-1023.

Werbose, P. J. (1990). Backpropagation through time: what it does and How to do it. *Proceedings of the IEEE, Vol. 78, No. 10*, pp. 1550-1560.

Abstract

Kim, Kwonill

Interdisciplinary Program in Brain Science

The Graduate School Seoul National University

Associative memory is the aspect of memory that stores pairs of information that are related. It is mathematically defined as a set of information, a memory ensemble. To learn memory ensembles, the stochastic recurrent neural networks, in which the features of natural neural networks are fully concerned, and the expected error backpropagation learning algorithm are designed. The computer simulations show the noble model and algorithm can not only store and recall memory ensembles, but also carry out the probabilistic or transitive inference.

……

Keywords: artificial neural network, memory ensemble, recurrent neural network, associative memory, backpropagation through time

Student Number: 2006-20485