

<https://doi.org/10.7236/JIIBC.2020.20.1.155>
JIIBC 2020-1-22

최적화문제를 해결하기 위한 완화(Relief)법

A Relief Method to Obtain the Solution of Optimal Problems

송정영*, 이규범**, 장지걸*

Jeong-Young Song*, Kyu-Beom Lee**, Jigeul Jang*

요약 일반적으로 최적화문제는 간단하게 해결하기 어렵다. 그 이유는 주어진 문제가 단순하면 바로 해결되지만, 복잡할수록 그 경우의 수는 방대하기 때문이다. 본 연구는 인공지능망 최적화에 대한 연구이다. 여기에서 우리가 다루고 있는 것은, 인공지능망을 구축하기 위한 완화법으로써, 최적화하는 방법이다. 주요 논제로는, 신경망 네트워크 전체의 안정성과 불안정성, 경비 절감, 에너지 절감과 같은 비결정적인 문제를 다루고 있다. 이를 위하여, 우리는 연상기억 모델 즉, 국소적 최소인 기억정보가 가짜인 정보를 선택하지 않는 방법을 제시한다. 그리고, 시뮬레이티드 어닐링법으로써, 이것은 가급적 낮은값을 가지고 있는 그 방향을 예측하고 그 이전의 낮은값과 결합해 나가서 더 낮은값으로 반복 수정해 나가는 방법이다. 그리고, 비선형 계획문제는, 방대한 조합상태의 수를 목적함수 합의 최소화를 위하여 적절한 최소하방법을 적용하여 입출력을 확인하여 수정해 나가는 방법이다. 결국 본 연구는 최적화문제를 해결하기 위한 이론적인 접근 방법으로써 완화법으로서의 접근가능한 유용한 방법을 제시하였다. 따라서, 본연구는 새롭게 인공지능망을 구축할 때, 효율적으로 적용 할 수 있는 좋은 제안이 될 것으로 생각한다.

Abstract In general, optimization problems are difficult to solve simply. The reason is that the given problem is solved as soon as it is simple, but the more complex it is, the very large number of cases. This study is about the optimization of AI neural network. What we are dealing with here is the relief method for constructing AI network. The main topics deal with non-deterministic issues such as the stability and unstability of the overall network state, cost down and energy down. For this one, we discuss associative memory models, that is, a method in which local minimum memory information does not select fake information. The simulated annealing, this is a method of estimating the direction with the lowest possible value and combining it with the previous one to modify it to a lower value. And nonlinear planning problems, it is a method of checking and correcting the input / output by applying the appropriate gradient descent method to minimize the very large number of objective functions. This research suggests a useful approach to relief method as a theoretical approach to solving optimization problems. Therefore, this research will be a good proposal to apply efficiently when constructing a new AI neural network.

Key Words : AI network, Optimal solution, Relief method

*정회원, 배재대학교 컴퓨터공학과

**정회원, 배재대학교 컴퓨터공학과

접수일자: 2019년 10월 1일, 수정완료: 2020년 1월 4일

게재확정일자: 2020년 2월 7일

Received: 1 October, 2019 / Revised: 4 January, 2020 /

Accepted: 7 February, 2020

**Corresponding Author: wsclkb7@hanmail.net

Dept. of Computer Engineering, Pai Chai Univ., Korea

I. 서 론

최근에 인공지능이 화제에 오르면서 인공지능경망을 구축할 경우 뉴-론과 뉴-론의 커넥션에 대한 응용이 많이 등장하고 있다. 이 원리는 오래전부터 알려져 있지만, 최근에 여러 분야에 응용됨에 따라서 이 완화법의 다양성에 대하여 그 유용성을 인식하게 되었다. 본 논문에서는 여러 가지로 응용 가능성이 있는 완화법에 대하여 분석하고, 파생되는 효과에 대하여 논하고자 한다.

II. 완화법 문제설정 및 접근

주어진 어떤 상태에 대하여 그 평가가 일방적으로 주어지는 시스템이 있다. 여기에서, 시스템에서 일어날 수 있는 상태를 N개의 정수 또는 실수로 표현 할 수 있는 그룹을 나열해 본다. 정수 그룹의 열 데이터를 N개의 비트 열(列)로 하고, 실수 그룹으로서는 N차원 벡터로 생각할 수 있는데, 다음 식으로 표현 할 수 있다.

$$s=(s_1, s_2, \dots, s_N) \quad (1)$$

이 상태에 대하여 주어진 평가 또는 코스트(cost)를 s의 함수로써 다음과 같이 표현한다.

$$E(s) \quad (2)$$

이 코스트는 때에 따라서 에너지로도 표현 할 수 있기 때문에 E로 표현 하였다. 주어진 시스템이 상태 s를 결정하는 것이 어느 정도 바람직한가라는 평가를 이 E로 표현되는 것을 생각하여, 코스트 또는 에너지로 설명을 한다고 할 때, 우리의 관심사는 보다 더 바람직한 상태로 갖게 하고, 더 낮은 값을 갖도록 하는 것이다^[1-3].

주어진 상태 s에 대한 에너지 E는 N의 제곱 승수 정도의 시간과 공간을 사용하면 얻을 수 있을 것이다. 상태 요소 s_i 는 +1 또는 -1 를 선택한다고 하면, 에너지 E는 다음과 같이 된다.

$$E=-1/2 * (\sum_{i \neq j} J_{ij} s_i s_j) \quad (3)$$

여기에서 J_{ij} 는 i번째와 j 번째 요소(노드)간의 [상호작용]이라고 불린다. 따라서, 대칭성, 즉,

$$J_{ij} = J_{ji} \quad (4)$$

을 만족한다. 시스템은 $N(N-1)/2$ 개 실수 조합 $\{ J_{ij} \}$ 가 주어짐에 따라서 결정된다. 임의의 어떤 상태 $s=(s_1, s_2, \dots, s_N)$ 에 대한 에너지를 산출하는데는 $O(N^2)$ 만큼 필요하다. 위의 식 (3)에서 에너지를 최소화 하는 s_0 를 기저 상태라고 하여, 이 기저상태를 구하는 문제는 조합 최적

화문제에서도 NP(비결정적문제)로서 아주 어렵다는 것은 잘 알려져 있다^[4-8]. 즉, N 제곱근만큼 시간이 걸리기 때문에 확실하게 기저상태를 구하는 것은 일반적으로 알려져 있지 않다. 물론 특수한 상호작용의 조합 $\{ J_{ij} \}$ 을 선택하면 간단하게 구해지는 경우가 있다. 예를 들면, J_{ij} 가 모두 단조증가 또는 단조감소라고 가정하면, 상태 $s_0 = (1,1,1,\dots, 1)$ 과 $(-1, -1, -1, \dots, -1)$ 이 기저 상태가 된다. 그러나, J_{ij} 는 때에 따라서 증가와 감소가 뒤섞여있는 체계로서 이러한 체계의 기저상태를 구하는 것은 아주 어렵고, 사실상 곤란하다. 조합 최적화문제를 풀기 위하여 열거법으로 접근하여 풀어야 하는데, 여기에서 생각할 수 있는 몇가지 방법으로는 분기 한정법(Branch and Bound Method)과 동적 계획법(Dynamic Programming) 등으로 접근하여 해결 할 수는 있지만 그 내용으로 들어가 보면, 결국 열거법으로 풀어야 하는 문제가 되어 버린다. 단순하게 열거하는 것으로 한다면, 2^N 만큼 걸려서, 각각의 상태를 에너지 계산으로 하여 N^2 만큼 걸리게 된다. 따라서 계산에 필요한 시간은

$$N^2 2^N \quad (5)$$

로 되어 고성능 컴퓨터를 사용한다고 해도 $N \sim 50$ 정도의 시간이 걸리게 된다. NP 문제에서도 위에서 언급한 합리화된 계산을 적용한다고 하면 위의 (5)식 만큼은 걸리지 않는다고 해도 걸리는 시간이 지수 의존성을 피해가기는 어렵고, $N \sim 100$ 을 초과하는 해를 얻는 예는 거의 없다. 이러한 기저 상태를 구하는 문제는 분할 문제에서도 일어날 수 있는 문제이지만, 이것은 LSI 설계에도 일어날 수 있고, 현실에서는 꽤 커다란 N 시스템의 최적화가 요구 된다^[9-11]. 이런 경우, 만족할 만한 정확한 해는 거의 불가능하고 오히려 근사해법이 차라리 더 좋을 수 있다. 즉, 정확한 최소는 아니더라도 충분히 적은 에너지가 되는 상태를 구하는 것이다. 여기에는, 가장 단순한 근사 해법으로서 랜덤법이 있다. 이것은 랜덤상태를 만들어서 그 에너지를 산출하는 과정을 반복하는 것이다. 이것은, 주어진 범위에서 최선의 해를 찾아내는 근사 해이다. 만약, 주어진 문제를 분석하여, M개 상태가 만들어지는 경우 그 계산량은 다음과 같다.

$$MN^2 \quad (6)$$

이 방법의 특징은 문제가 복잡한 것과는 별도로 일단 해가 구해진다는 것이다. 단, 제한된 계산 시간에 따라서 가능한 상태의 수 M은 전체상태의 수 2^N 과 비교하여 일반적으로 아주 작은 것으로서 별로 좋은 결과를 기대하기는 어렵다^[12-14].

랜덤법과 비교하여 약간 효율이 좋을 것으로 생각되는

것이 완화법이다. 어떤 방법으로든 현재 얻어진 상태를 출발점으로 하고 그 근방의 상태에서 에너지를 구한다. 출발점 에너지로부터 보다 더 낮은 에너지 상태를 발견하면, 출발점을 이동하는 작업을 반복적으로 계속하여 더 이상 에너지가 낮아지지 않으면 정지한다. 우리는 이것을 근사해라고 한다. 여기에서 근방이라고 하는 것은 예를 들면 해밍거리가 작은 것으로 할 수 있다. 1-근방 탐색은 에너지 요소 1개를 변경하여 에너지 변화를 구하는 것으로, 복잡도 계산량은 $O(N)$ 이라고 할 수 있다. 왜냐하면, 한 상태 변경 ΔS_i 에 따른 에너지 변화 ΔE 는

$$\Delta E = - \Delta S_i \sum_{j(\neq i)} J_{ij} S_{ij} \quad (7)$$

로 주어지기 때문에 국소적인 '장'

$$f_i = \sum_{j(\neq i)} J_{ij} S_{ij} \quad (8)$$

을 계산하는 것만으로 상태 변화의 가부가 판단되기 때문이다. 주어진 어떤 한 상태가 국소점의 에너지에 도달했다는 것을 판단하기 위해 모든 국소적 '장'이 상태의 부호에 일치 하는 것을 확인 할 필요가 있기 때문에 적어도 $O(N^2)$ 의 계산량이 필요하게 된다. 이 완화가 종료되기까지 L 회 정도의 상태변경을 했다고 하면, $L < N$ 이라고 할 경우, 계산량은 최대 다음과 같은 식으로 주어진다.

$$LN^2 \quad (9)$$

그러면 여기에서 얻어진 상태는 어느 정도 영역에서 최소로 되어 있는가. 물론 계산이 정지 했다고 하는 사실에서 최종상태는 1-근방 상태에서 최소로 되어 있다는 것은 보장된다. 그리고 적어도 도중까지 계산된 모든 상태에서는 최소이다. 따라서, 얻어진 상태는 적어도 $O(L + N)$ 개 중에서는 최소라는 결론은 내릴 수 있다. 그러나 상태는 에너지곡면을 거리 L 만큼 움직이고 있기 때문에 결과적으로는 같은 방법으로 계산하여 L -근방 전체에 걸쳐서 탐색 한 것으로 생각 할 수 있다. 이렇게 하면, 얻어진 국소적 최소는 대략 $O(N/L)^2$ 개의 상태에서 최소라고 평가 할 수 있다. $O(M)$ 개의 상태중에서 최소인 랜덤법과 비교하여 압도적으로 유효함을 알 수 있다.

그렇다면, 현실적으로 완화법에 의해 얻어진 결과는 어느 정도 의미가 있는 것인가. 위의 (3)식에 있어서 $\{J_{ij}\}$ 에 대하여 증가 또는 감소가 섞여 있는 경우, 1-근방에서 정의된 국소적 최소 수 N_m 은 해석적으로 평가 할 수 있다. 그 통계적 평균은 다음 식으로 주어진다.

$$\langle N_m \rangle \sim e^{0.2N} \quad (10)$$

여기에서, 모든 상태의 수 $2^N \sim e^{0.7N}$ 중에서 $e^{0.2N}$ 이 국소적 최소이기 때문에, 하나의 국소점과 만나는 영역의

넓이는 평균 $e^{0.5N}$ 보다 크다. 그럼에도 불구하고 국소적 최소의 개수는 지수 개 만큼이기 때문에 충분히 줄었다고는 말할 수 없다. 완화법은 적용가능하지만 결국 국소적 탐색이기 때문에 당연히 한계가 있다.

완화법을 이용하기 위한 방법은 다음 2가지로 나누어 고찰 적용할 수 있다. 첫째는 더 이상 줄일수 없을때까지 최적화를 목표로하여 완화법에 동반되는 국소성의 한계를 조금씩 더 늘여가는 방법이다. 둘째는 국소성을 역으로 이용하는 방법이다. 전자는 그 예로써 시뮬레이티드 어닐링(Simulated Annealing)이라고 할 수 있고, 후자의 예는 연상기억 모델을 들 수 있다. 먼저 연상기억모델을 확인한다.

III. 연상기억 모델

연상기억모델의 접근과 적용에는 여러 가지가 있지만, 본 연구에서는 최적화 문제를 중심으로 전개하고자 한다. 연상기억 모델은 홉 필드가 자기상기형 연상기억 모델을 만들어서 스핀 시스템과 비교하여 보여 주었지만, 이 모델의 상태를 기억정보로 놓으면 근방의 상태는 기억정보로 향하여 완화된다. 잘못된 정보를 입력해도 올바른 정보를 찾아 낼 수 있는 이른바 오류 정정 능력을 얻게 된다. 여기에서 문제점은 어떻게 하면 그와 같은 네트워크를 설계 하는가 이다. 여기에서 홉필드 모델(Hop Field Model)이라고 하는 것은 기억하고 있는 정보 하나를 $S_i^{(m)}$, $m=1,2,\dots,M$ 으로 하는, 이른바 위에서 기술한 식 (3)의 결합으로서 다음과 같이 놓을 수 있다.

$$J_{ij} = \sum_{m=1}^M S_i^{(m)} S_j^{(m)} \quad (11)$$

그러면 이 모델은 대칭결합을 가지고 있는 스핀 시스템으로서, 기억정보의 수 M 이 요소수 N 과 비교하여 충분히 작은 경우 그것들 각각은 '거의' 안정적으로 국소적 최소로 볼 수 있다. 여기에서 '거의'라고 하는 것은 국소적 최소 상태가 엄밀하게 기억정보와 일치하는 것과는 다르며 아주 조금 다른 경우가 있다는 의미이다. 또한, M/N 이 일정한 값 (예를들면, α) 이하이면서 이와 비슷한 경우는 위의 식 (1)과 같은 상황이 성립하면서, 각 기억정보에는 가깝지 않은 다수의 국소적 최소가 발견된다. 이것을 [가짜 기억정보]이다. 그리고, M/N 이 α 를 초과하면 기억정보 안정성이 보장되지 않고 기억시키려고 하는 내용과 거의 관계없는 다수의 국소적 최소가 존재하는 상황이 된다.

이와같이 식 (11)에 대한 모델의 특징과 한계가 주어 지게 되면, 다음과 같은 효율이 좋은 네트워크를 설계할 수 있다. 즉, 모델을 현실과 가깝고 사용가능한 것으로 하기 위하여 기억용량을 크게 할 것, 주어진 상태를 정확하게 안정화 할 것, 범위를 확실한 기준으로 분할 할 것, 가 짜 기억을 제거 할 것, 기억내용을 계층적인 구조로 줄 것 등을 고려할 필요가 있다. 여기에는 여러 가지로 연구 가 진행되고 있고, 그 결과도 나와 있지만, 범위를 확실한 기준으로 분할 할 것에는 충분한 연구가 진행되지 않고 있다. 이 이외에도 네트워크 설계 원리의 새로운 패러다 임으로서 홉 필드 머신 학습법(Hop Field ML)이 있다. 이 학습법은 주어진 상태를 확률적으로 발생시키는 머신 으로 확률을 제어하는 방법이다.

IV. 시뮬레이티드 어닐링

서론에서 최적화 문제를 해결하는 근사 해법으로 랜덤 법, 완화법에 대하여 서술 했다. 완화법은 랜덤법과 비교 하여 유효하기는 하지만 근사라는 관점에서는 아직 불충 분하다. 완화법에 동반되는 국소성을 해결하기 위해서는 일단 완화한 영역을 탈출하여 다른 영역으로 들어가는 경우를 생각하지 않으면 안된다. 골짜기에서 어떻게 더 낮은 골짜기를 선택하는가에 대하여는 일반적인 방법이 없다. 가급적 낮은 방향으로 내려 가면 다음 골짜기가 더 낮을 것이라고 예측하는 것은 바람직한 생각이다. 이와 같이 큰 흐름의 예측과 방향을 적당히 결합해 가는 방법 이 시뮬레이티드 어닐링이다.

상태를 변경하여 그 에너지변화를 dE 로 하여 측정한다. 이 값이 음수이면 에너지가 감소되었으므로 변경된 상태를 선택하고, 양수이면 더 감소하기 위하여 다음 식 으로 확률을 계산한다.

$$e^{-dE/T} \quad (12)$$

이식에서 보는 바와 같이 에너지의 변화 즉, 올라가고 내려가는 것은 파라메타 T 에 의해서 조절된다. 이 T 는 한 변수에 대응하는 변수이지만, $T=0$ 에 가까워지는 경 우에는 완화법과 동일하고, $T=\infty$ 에 가까워지는 경우에는 한가지 종류의 랜덤법으로 볼 수 있다. 주어진 T 를 기준으로 하여 시스템의 확률은 다음과 같은 상태로 주 어진다.

$$P(s) \propto e^{-E(s)/T} \quad (13)$$

이 확률분포가 볼츠만 머신이다. 충분하게 T 가 작은 경우 낮은 에너지상태의 확률은 높게 나타나지만, 국소적

최소구간의 에너지 장벽을 넘어가는 데는 꽤 시간이 걸 린다. 낮은 값에서 평형상태로의 완화를 촉진하기 위해서 는 높은 값에서 낮은 값으로 천천히 낮추는 방법을 생각 할 수 있다. 서서히 냉각시킨다는 의미에서의 어닐링이라 는 방법으로 낮은 에너지 상태를 얻어내기 위한 방법이다.

V. 비선형적인 문제

실수(Real Number)로 형성된 한 집합에 대하여 코스 트가 주어진 경우의 최적화에 대하여는 선형계획법등에 서는 코스트가 선형함수로서 하한이 없기 때문에, 범위로 제한을 둔다. 그러나, 그 범위에 엄밀한 제한을 두지 않 고, 모든 제약은 부드러운 코스트 함수로 표현되는 경우 의 최적화도 그 본질은 조합 최적화와 거의 비슷하다. 다 수의 국소적 최소를 가질 경우 결정적인 해결방법은 없 고, 어디까지나 국소적 탐색이 기본이다. 조합 최적화와 비교하여 얼핏 보기에 유리한 것은, 함수의 연속성을 탐 색에 이용할 수 있다는 것이지만, 불리한 것은 상태수를 계산할 수 없을 정도로 무한이라서 엄밀한 의미에서 수 령하지 않기 때문에 계산량 평가가 어렵다.

조합 최적화 해법의 경우에는 별로 변형시키지 않고 응용할 수 있다. 서론에서 기술한 완화법은 경사하강법 (Gradient Descent Method)으로 대신 할 수 있다. 동 시에 경사하강법적용 과정에서 백색잡음을 입력시킴으로 써 다른 효과를 기대 할 수 있고, 그 내용을 확인할 수 있다.

코스트 함수의 연속성을 더욱 유용하게 사용하기 위하 여 이 경사하강법을 조금 변형하여, 한층 더 높은 수렴성 을 주는 뉴-턴법, 공역법등을 도입하는 것이 효력있다. 수렴속도는 별도로 하고, 한가지 예로써, 최적화 문제에 적용 가능한 오류역전파법(Back Propagation Method) 에서 일어 날 수 있는 구체적인 문제에 적용하여 오류역 전파법을 일반적인 형태로 표현하면 다음과 같다.

실수 파라메타의 그룹 $S=(S_1, \dots, S_N)$ 이 있다고 할 경우, 다음과 같은 함수식 (14), 그리고 목적신호함수 (15)에 대하여 목적함수로서의 코스트는 식 (16)과 같이 놓을 수 있다.

$$y = f_s(x) \quad (14)$$

$$z = F(x) \quad (15)$$

$$C(s, x) = (z - y)^2 \quad (16)$$

그러면, 여러개의 입력에 대하여 출력과의 오류 총합 을 총 코스트로서 식(17)으로 표현 할 수 있다.

$$E(s) = \sum C(s, x) \quad (17)$$

오류역전파법에서는 이 코스트를 감소시키는 방법으로서 경사하강법을 사용한다. 다층구조의 피드 포워드형(Feed Forward) 네트워크에 있어서 노드간의 결합을 제어 파라메타 S_i 를 사용하여 조절한다. 계산법으로서는 제 1층 결합강도를 변화시키기 위하여, 제 2층 결합의 출력과 교사신호의 차이를 알아내야 하기 때문에 역으로 계산할 수 밖에 없다.

오류역전파법을 확인하기 위하여, 실제응용문제에 적용시키는 시도는 많이 있다. 역전파법이 이용되는 곳은 Bool 함수적인 것이 많고, 활성화 함수로서 이용되는 것은 시그모이드 함수이다. 이런 경우 코스트함수는 실제로 현장에서 많은 분야에 사용되고 있다. 참값 주변 코스트함수는 움푹 패인 곳으로 형성되어 있으며, 일반적으로 가늘고 긴 협곡처럼 되어 있다. 넓고 평탄한 꼭대기 정상에서 가늘고 긴 골짜기가 있는 상태로 생각할 수 있다. 꼭대기 정상에서 있는 사람은 어디에 골짜기가 있는지 알 수 없다. 골짜기에 들어가서 서 있는 경우, 그 골짜기는 가늘고 긴 모양으로 최소를 만날 때 까지 긴 길을 따라서 걸어갈 수 밖에 없다. 골짜기가 깊은 경우 경사하강법을 적용하려 하면 수치적 불안정을 보여주어야 하거나 또는 목적함수 값보다 커야한다. 경사하강법은 다음 (18)식에 따라서 상태를 변화시켜 나가는 방법이지만, 유한 step 폭을 가지고 있는 각 단계별로 항상 반드시 에너지가 감소 한다는 보장은 해 주지 않는다.

$$\Delta S_i \propto \partial E / \partial S_i \quad (18)$$

step 폭이 움푹 들어간 스케일에 대하여 충분히 작지 않은 경우, 애석하게도 수렴하지 않고 진동하는 경우가 있다. 이 진동을 피하기 위하여 임의의 값을 입력하는 방법이 있지만 이것도 만능이라고는 볼 수 없다. 골짜기의 중심에 도달해도, 실제로 부드러운 경사가 있어서, 골짜기를 따라 긴 길이 있을 수 있다. 이러한 사태가 아주 높은 차원의 공간에서 일어나는 현상이기에 최소치를 찾는 것은 쉽지 않다. 실제로 비선형 최적화문제에서는 이와 같은 이유에서 수렴속도가 아주 늦어진다. 따라서, 일부 연구자들은 이와 같은 사태를 극소적 최소문제와 혼동하여 사용하기도 한다.

한편, 오류 역전파법을 적용하는데 있어서 다음 두가지 문제를 고려해야 한다. 첫 번째로는, 최적화 문제는 식 (17)을 최소화하는 방법으로 진행 되지만, 오류 역전파법에서는 [샘플데이터로부터의 학습]을 전제로 하여, 하나하나 출력력의 예에 따라서 순차적으로 파라메타를

수정해 나간다. 여러개의 학습예에 따라서 (16)식을 코스트라고 생각하는 것이다. 오류 역전파의 네트워크 구조는 코스트함수를 (16)식으로 하면, 국소적 최소값은 존재하지 않을 것으로 예상된다. 따라서, E=0에 도달 할 때까지는 정지하지 않고, 끊임없이 상태를 변경하도록 해야 한다. 참 해를 발견 할 때까지 상태 이동과정에는 관심을 가지고 연구해야 할 가치가 있다. 두 번째 문제는, 식 (17)의 합을 구하는 것이 문제가 된다. 입력력관계가 bool 함수라고 할 경우, 입력 x 가 L bit 라고 할 때, 만일 $L = 100$ 이라고 하면, 계산가능한 출력력관계는 2^{100} 을 넘는 수가 되어 모든 가능한 예를 조사하는 것은 거의 불가능하다. 일반적으로 학습은 가능한 출력력관계의 일부만 실행하고, 학습이 [완료]된 머신은 학습하지 않은 출력력 관계를 실현시켜야 한다. 이것은 통계학에서 생각할 수 있는 추정, 검정의 문제이다. 시스템의 일반화 능력은 시스템의 파라메타 수가 적으면 적을수록 높다고 알려져 있다. 이것은 언뜻 보기에 이상하다고 생각할지 모르지만, 파라메타수가 적으면, 시스템의 잠재능력이 낮음을 표현하고 있다는 것을 의미한다. 파라메타수를 적게 하는 것은 문제 영역을 한정시키는 것과 마찬가지로, 파라메타 수가 적은 시스템이 문제를 해결하는 경우는, 문제자체가 처음부터 단순한 것이었다고 말 할 수 있다. 따라서, 오류 역전파 관련 연구는 최적화 문제를 일반화시켜 나가는 방법중의 하나로서 여러분야에 적용됨으로써 점차 발전해 가고 있다. 학습이 어느정도의 신뢰성을 줄 것인가 또는 필요에 따라 파라메타의 수를 줄이는 것이 최적화에 어느 정도 기여하는 가는 중요한 문제이다.

VI. 결 론

본 논문에서는, 최적화문제를 해결하기 위한 한가지 방법으로서의 완화법에 관련한 몇 가지 문제를 가져와서 고찰하고 그 해결방법을 제시했다. 즉, 국소적 최소인 기억정보가 더 작은 정보가 있을지 모른다는 가짜 정보를 선택하지 않는 방법을 제시한 연상기억모델과, 가급적 낮은 값을 가지고 있는 그 방향을 예측하여 이전의 낮은 값과 결합하여 더 낮은 값으로 반복 수정해 나가는 시뮬레이션 이터드 어닐링법, 그리고, 비선형 계획문제는 방대한 조합상태의 수를 목적함수합의 최소화를 위하여 적절한 최소하강법을 적용하여 출력력의 값을 반복 수정해 나가는 방법으로 해결하는 방법이다. 넓은 의미에서의 완화법은 아직 어려운 문제를 가지고 있음에도 불구하고, 의외로

많은 문제를 해결하는데 있어서 유용하게 사용되고 있다. 앞으로 이 완화법이 동적인 문제를 풀어 나가는 방법으로 적용되기를 기대한다. 그런 가운데 지금도 이와 같은 완화법을 실제 현장에서 응용하는데 있어서, 보다 더 적은 에너지 또는 코스트를 위하여 고민해야 할 필요가 있고, 더 나아가 향후 더 훌륭한 접근 방법이 연구되기를 기대한다.

References

- [1] Ermao Cai, Da-Cheng Juan, Dimitrios Stamoulis, and Diana Marculescu, "Neuralpower: Predict and deploy energy-efficient convolutional neural networks", In Asian Conference on Machine Learning, pp. 622-637, 2017.
- [2] Ermao Cai and Diana Marculescu, "Temperature effect inversion-aware power-performance optimization for finfet-based multicore systems", IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, Vol. 36, No. 11, pp. 1897-1910, 2017.
- [3] Zhuo Chen and Diana Marculescu, "Distributed reinforcement learning for power limited many-core system performance optimization", In Proceedings of the 2015 Design, Automation & Test in Europe Conference & Exhibition, EDA Consortium, pp. 1521-1526, 2015.
- [4] Matthieu Courbariaux, Yoshua Bengio, and Jean-Pierre David, "Training deep neural networks with binary weights during propagations", In Advances in Neural Information Processing Systems, pp. 3123-3131, 2015.
- [5] Ruizhou Ding, Zeye Liu, Rongye Shi, Diana Marculescu, and RD Blanton, "Filling the gap between conventional deep neural networks and binarized networks", In Proceedings of the on Great Lakes Symposium on VLSI 2017, pp. 35 - 40, 2017.
- [6] Kyuseung Han, Jae-Jin Lee, Jinho Lee, Woojoo Lee, and Massoud Pedram, "Optimizing ultralow power nocs exploiting the temperature effect inversion", IEEE Transactions on Computer Aided Design of Integrated Circuits and Systems, Vol. 37, No. 2, pp. 458-471, 2018.
- [7] Dimitrios Stamoulis, Ermao Cai, Da-Cheng Juan, and Diana Marculescu, "Power-and memory-constrained hyper-parameter optimization for neural networks", In Design, Automation & Test in Europe Conference & Exhibition, IEEE, pp. 19-24, 2018.
- [8] Chen Zhang, Peng Li, Guangyu Sun, Yijin Guan, Bingjun Xiao, and Jason Cong, "Optimizing fpga-based accelerator design for deep convolutional neural networks", In Proceedings of the 2015 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays, FPGA '15, pp. 161-170, 2015.
- [9] N. Boyd, G. Schiebinger, and B. Recht, "The Alternating Descent Conditional Gradient Method for Sparse Inverse Problems", SIAM Journal on Optimization, Vol. 27, No. 2, pp. 616-639, 2017.
- [10] B. Efron and T. Hastie, "Computer Age Statistical Inference", Institute of Mathematical Statistics Monographs, Cambridge University Press, 2016.
- [11] Z. Harchaoui, A. Juditsky, and A. Nemirovski, "Conditional gradient algorithms for norm-regularized smooth convex optimization", Mathematical Programming, pp. 1-38, 2014.
- [12] P. Jain, P. Netrapalli, and S. Sanghavi, "Low-rank matrix completion using alternating minimization", Proceedings of the forty-fifth annual ACM symposium on Theory of computing, ACM, pp. 665-674, 2013.
- [13] P. Netrapalli, P. Jain, and S. Sanghavi, "Phase retrieval using alternating minimization", Advances in Neural Information Processing Systems, pp. 2796-2804, 2013.
- [14] E. Van den Berg and M. P. Friedlander, "Sparse optimization with least-squares constraints", SIAM Journal on Optimization, Vol. 21, No. 4, pp. 1201-1229, 2013.

저 자 소 개

송 정 영(정회원)



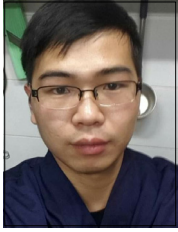
- 1984년 2월 한남대학교 컴퓨터공학과 졸업
- 1992년 3월 와세다대학 전기전자정보공학연구과 졸업(공학석사)
- 1995년 3월 와세다대학 전기전자정보공학연구과졸업(공학박사)
- '95.3 - '97.2 청운대학교전임강사
- '11.9 - '12.8 아이다호주립대 방문교수
- 1997.3 - 현재 배재대학교 컴퓨터공학과 교수
- 관심분야 : 패턴처리(영상, 문자, 음성), 기계학습

이 규 범(정회원)



- 1987년 2월 한남대학교 컴퓨터공학과 졸업
- 1999년 8월 한남대학교 컴퓨터공학과 졸업(공학석사)
- 2004년 2월 배재대학교 컴퓨터공학과 졸업(공학박사)
- '03.3 - '16.2 한남대학교 겸임교수 및 외래교수
- '00.9 - 현재 배재대 겸임교수 및 외래교수
- 관심분야 : 기계학습, 인공지능, 소프트웨어 복잡도, 빅데이터

장 지 걸(정회원)



- 2013년 9월 배재대학교 컴퓨터공학과 대학원입학
- 2018년 8월 배재대학교 컴퓨터공학과 졸업(공학석사)
- 관심분야 : 중국어 음성인식, 패턴인식

※ 이 논문은 2019학년도 배재대학교 교내학술 연구비 지원에 의하여 수행된 것 임.