

신경망의 웨이트 분포 분석

고진욱, 이철희

연세대학교 전기·컴퓨터 공학과

Tel: (02) 361-2779, E-mail: chulhee@bubble.yonsei.ac.kr

Analyzing weight distribution of neural networks

Jinwook Go and Chulhee Lee

Department of Electrical and Computer Engineering, Yonsei University

ABSTRACT

In this paper, we analyze weight distributions of neural networks. If we construct a vector containing all weights of a neural network, then training process can be viewed as finding a solution point in the weight space. In order to obtain insight into the training process of neural networks, we investigate the distribution of the solution points in the weight space. Experiments provide some interesting results, showing that solution points tend to form clusters in the weight space and the information may be used to speed up the training process.

1. 서론

신경망은 학습과정을 통해 주어진 문제를 해결할 수 있도록 웨이트가 변경되고 최종적으로 최적의 웨이트를 얻게된다. 그러나 신경망을 다른 초기 웨이트로부터 학습시킬 때, 출력 패턴은 유사하지만 최종 웨이트는 서로 다른 경우가 발생한다. 이러한 신경망의 특징은 하나의 문제라 할지라도 그 해(solution)가 유일하지 않고 무수히 많이 존재할 수 있다는 것을 의미한다. 신경망을 구성하는 전체 웨이트를 구성요소로 갖는 벡터를 웨이트 벡터라고 할 때, 이러한 웨이트 벡터는 웨이트 공간(weight space)상의 한 점(point)으로 생각할 수 있다. 이런 관점에서 웨이트 공간에서 주어진 문제의 해가 될 수 있는 웨이트 점(weight point)들이 밀집되어 분포된 영역을 해영역(solution region)이라 하고, 학습 과정을 해영역 밖에서 추측된 웨이트 점을 해영역의 내부로 이동시키는 것으로 생각할 수 있다 [1].

본 논문에서는 이러한 해영역의 분포를 조사하기 위해 동일한 문제를 서로 다른 초기 웨이트로부터 학습시키고 그 결과 얻어진 최종 웨이트들을 조사하였다.

그 결과 해영역은 웨이트 공간의 여러 곳에 분포되지만, 그 해영역들이 서로 무리(group)를 지으며 분포되는 특징을 보여주었다. 이러한 특징은 서로 다른 문제의 해영역들이 서로 인접할 가능성을 제시해준다. 즉, 이전의 학습 결과는 신경망의 지식으로 저장될 수 있고, 또한 새로운 문제의 학습에 도움을 줄 수 있는 기억력의 역할을 담당할 수 있다.

2. 웨이트 공간에서의 해영역 분포

본 논문에서는, 해영역의 분포를 구하기 위해 2 차원의 2 클래스로 구성된 58개의 패턴 분류 문제를 조사하였다. 신경망 구조는 2 차원 2 클래스 문제를 최대 정확도로 분리할 수 있는 가장 간단한 구조로서, 2개의 입력층 뉴런, 3개의 은닉층 뉴런, 2개의 출력층 뉴런, 그리고 2개의 바이어스(bias)로 구성된다. 그러므로 선택된 신경망의 전체 웨이트 개수는 17개이고, 학습 과정은 17차원 웨이트 공간에서 하나의 웨이트 점(17차원 웨이트 벡터)을 찾는 과정으로 생각할 수 있다. 입력 데이터는 실제 원격 탐사된 데이터로부터 추정된 통계로부터 정규분포를 가정하여 발생된 데이터를 사용하였고, 학습 알고리즘은 오류 역전파(backpropagation) 알고리즘을 사용하였다 [2].

각 패턴 분류 문제에 대해, 1000개의 다른 초기 웨이트로부터 신경망을 학습시켜 1000개의 최종 웨이트를 얻었다. 초기 웨이트는 균일 분포를 갖는 -0.5에서 0.5 사이의 값이고, 학습은 가우시안 최대우도(Gaussian ML) 분류기로 계산된 분류 정확도와 비교하여 그 차이가 2% 이하일 때 종료된다. 본 논문에서는 각 문제에 대한 해(solution)가 되는 웨이트 점들의 집합을 웨이트 그룹으로 정의한다. 표 1에 58개의 웨이트 그룹(58000개 웨이트 점들)으로부터 추정된 공분산 행렬(covariance matrix)의 고유치(eigenvalue)의 크기, 고유치의 비율, 누적 값을 나타내었다. 가장 큰 9개가 99% 이상을 차지하고 있다. 이는 대부분의 웨이

신경망의 웨이트 분포 분석

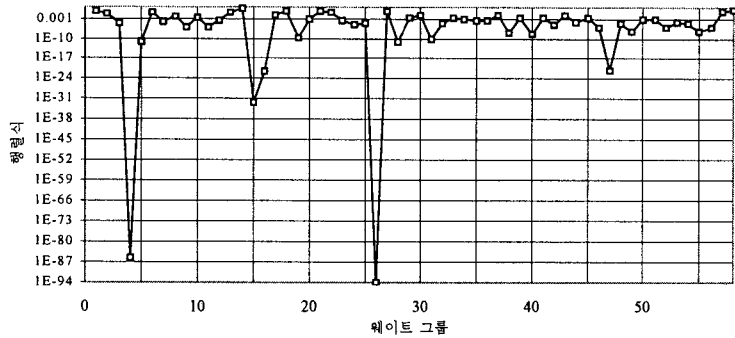


그림 1. 58개 웨이트 그룹의 행렬식(determinant).

트들이 가장 큰 몇 개의 고유치에 대한 고유벡터(eigenvector) 방향을 따라 길게 늘어진 분포를 갖고 있고, 나머지 고유벡터 방향으로서는 거의 변동이 없음을 의미한다.

표 1. 58개 웨이트 그룹의 고유치의 크기, 비율, 그리고 누적값.

	Eigenvalue	Proportion	Accumulation
1	110.884	27.047	27.047
2	110.406	26.931	53.978
3	105.423	25.715	79.693
4	16.111	3.930	83.623
5	15.914	3.882	87.505
6	14.597	3.561	91.065
7	11.352	2.769	93.834
8	11.240	2.742	96.576
9	10.873	2.652	99.228
10	2.656	0.648	99.876
11	0.152	0.037	99.913
12	0.151	0.037	99.950
13	0.149	0.036	99.986
14	0.023	0.006	99.992
15	0.015	0.004	99.996
16	0.014	0.003	99.999
17	0.004	0.001	100.000

그림 1은 웨이트 그룹의 행렬식(determinant)을 보여 준다. 행렬식은 고유치들의 곱과 같다. 그림에서 웨이트 그룹의 행렬식의 크기가 현저한 차이를 나타냄을 알 수 있다. 이는 웨이트 그룹들이 분포된 실제 체적에 있어서 많은 차이가 있음을 가리키고, 몇 개의 웨이트 그룹은 매우 길게 늘어진 형태로 분포됨을 알 수 있다. 예를 들면, 가장 큰 행렬식은 4.05이고, 가장 작은 행렬식은 10^{-94} 이다.

그러나, 웨이트 그룹 내에 분포된 모든 웨이트 점들이 주어진 문제에 대해 최대 정확도를 제공하지는 않는다. 본 논문에서는, 웨이트 그룹내의 균일한(homogeneous) subgroup을 찾기 위해 Euclidean 거리를 이용한 clustering을 적용하였다. 한가지 고려할 사

항은 신경망의 은닉층 뉴런들의 순서를 조합(combination)함으로써 하나의 웨이트 점을 동일한 출력값을 나타내는 6개의 서로 다른 웨이트 점들(웨이트 벡터들)로 나타낼 수 있다는 점이다. 이것은 17차원 웨이트 공간에서 6개(은닉층 뉴런 개수의 계승(factorial))의 동등한(equivalent) 영역들이 존재함을 가리키고, 효율적인 clustering을 위해 웨이트 점들을 같은 영역에 속하도록 할 필요가 있다. 이를 위해 하나의 웨이트 점의 조합으로부터 얻은 6개의 웨이트 점들 각각에 대해, 입력층의 바이어스와 은닉층 뉴런들을 연결하는 웨이트들의 합을 구하고, 그 중 가장 큰 값을 갖는 하나의 웨이트 점을 선택한다. 이러한 방법으로 재배열된 1000개의 웨이트 점들을 clustering에 사용하였고, 그 결과 각 웨이트 그룹에 대해 평균 51개의 subgroup을 찾을 수 있었다. 그림 2는 웨이트 그룹의 clustering 결과를 보여준다. 동일한 문제의 해 영역이라도 웨이트 공간의 여러 곳에 분포될 수 있고, 또한 문제에 따라 개수가 다소 차이가 있음을 알 수 있다. 표 2는 한 웨이트 그룹의 subgroup들에 포함된 웨이트 점들의 개수를 보여준다. 많은 수의 다른 초기 웨이트 점들이 같은 subgroup에 포함됨을 알 수 있다.

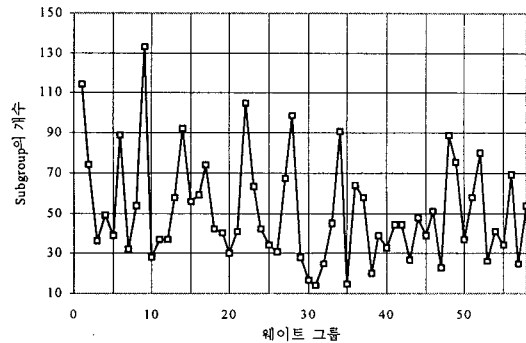


그림 2. 58개 웨이트 그룹의 clustering 이후의 subgroup의 개수.

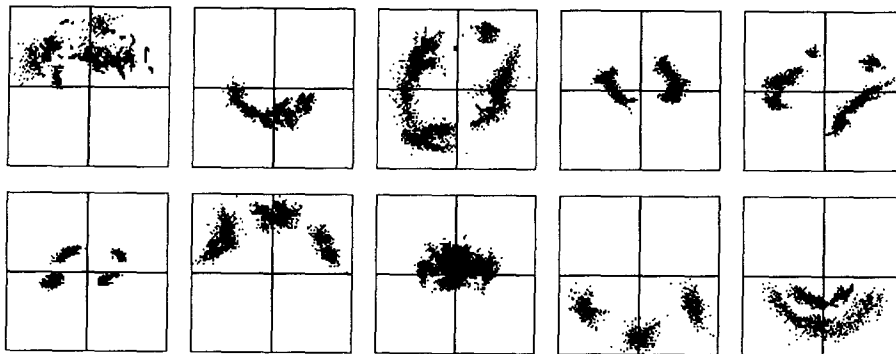


그림 3. 확장 후의 웨이트 그룹들의 분포.

표 2. 각 subgroup에 포함된 웨이트 개수.

웨이트 개수	subgroup 개수	웨이트 개수	subgroup 개수
231	1	25	1
135	1	24	1
104	1	19	1
56	1	17	1
53	2	14	1
49	1	11	1
39	1	10	2
36	1	5	1
31	1	3	2
29	1	2	4
28	1	1	7

표 3. 웨이트 subgroup의 고유치.

	Eigenvalue	Proportion	Accumulation
1	0.233	98.183	98.183
2	0.003	1.253	99.436
3	0.001	0.554	99.990
4	2.30E-05	0.010	100.000
5	1.77E-07	0.000	100.000
6	1.65E-09	0.000	100.000
7	6.58E-11	0.000	100.000
8	1.11E-11	0.000	100.000
9	2.15E-12	0.000	100.000
10	7.35E-13	0.000	100.000
11	4.76E-13	0.000	100.000
12	9.55E-14	0.000	100.000
13	7.48E-14	0.000	100.000
14	3.44E-29	0.000	100.000
15	1.03E-29	0.000	100.000
16	1.44E-31	0.000	100.000
17	1.49E-43	0.000	100.000

표 3은 표 2에서 가장 많은 웨이트 점들이 포함된 subgroup의 231 개의 웨이트 점들로 추정된 공분산 행렬의 고유치를 보여준다. 가장 큰 3개의 고유치가 99% 이상의 비율을 차지하고 있고, 그 체적이 매우 작음을 알 수 있다.

subgroup이 분포된 영역의 모든 웨이트 점들은 최대 정확도를 제공하지만 체적이 매우 작다. 이러한

subgroup을 다음과 같은 과정으로 최대 체적을 갖는 동질의 subgroup으로 만들 수 있다. 먼저, 가장 작은 고유치에 대한 고유벡터 방향으로 고유치의 증가를 통해 subgroup을 확장시키고, 정규 분포를 가정하여 [3] 1000 개의 테스트용 웨이트 점들을 발생시킨다. 이 웨이트 점들로 시험한 정확도와 subgroup의 최대 정확도의 차가 2%이하인 웨이트 점들의 개수가 995이상이면, 확장을 유효하다고 생각하고, 그렇지 않으면 고유치를 증가시키지 않는다. 같은 과정을 나머지 고유치들에 대해서 반복한다. 마지막으로, 변경된 고유치들을 사용하여 1000개의 웨이트를 발생시키고, 이 중 최대 정확도와와의 차가 2% 이하인 웨이트 점들만으로 새로운 평균 벡터와 공분산 행렬을 추정한다. 이러한 모든 과정은 고유치들이 더 이상 증가하지 않을 때까지 반복된다. 본 논문에서는 약 3000개의 subgroup들을 확장시

표 4. 확장 후, 웨이트 subgroup의 고유치와 확장 전과 후의 고유벡터간의 각도.

	Eigenvalue	Proportion	Accumulation	Angles of Eigenvectors (degree)
1	36.679	69.837	69.837	8
2	11.642	22.167	92.003	74
3	2.984	5.643	97.646	104
4	0.763	1.452	99.098	78
5	0.342	0.652	99.750	77
6	0.099	0.188	99.938	43
7	0.019	0.037	99.975	136
8	0.007	0.014	99.989	60
9	0.005	0.009	99.997	126
10	6.85E-04	0.001	99.999	50
11	4.32E-04	0.001	100.000	86
12	1.84E-04	0.000	100.000	78
13	4.00E-05	0.000	100.000	12
14	9.39E-21	0.000	100.000	69
15	8.05E-21	0.000	100.000	62
16	3.34E-21	0.000	100.000	119
17	8.80E-37	0.000	100.000	51

켰고, 확장 과정동안 평균적으로 약 5만개의 웨이트 점들이 조사되었다. 그림 3은 10개 웨이트 그룹의 확장 과정 이후의 분포를 각 웨이트 그룹의 가장 큰 2개의 고유치들에 대한 고유벡터들의 평면으로 매핑(mapping)시킨 그림이다. 웨이트 그룹마다 여러 가지 다른 분포를 형성하고 있고, subgroup들이 서로 무리를 지으며 모여있음을 알 수 있다. 표 4는 확장 과정 이후의 표 3의 subgroup의 고유치를 보여준다. 가장 큰 4개의 고유치가 99% 이상의 비율을 차지하고 있고, 체적은 약 10^{124} 배 증가되었다. 또한 확장 전과 후의 고유벡터간의 각도가 차이를 보이므로 확장 과정동안 subgroup의 분포가 변했음을 알 수 있다.

3. 웨이트 분포를 이용한 신경망의 초기화

본 논문에서는 신경망의 해역의 분포를 조사하였다. 다음으로 이러한 웨이트 분포에 대한 정보를 신경망의 지식으로 사용하여, 새로운 문제에 대해 빠른 학습이 가능한지 조사하였다. 먼저 웨이트 분포를 지식으로 사용하기 위해, 48개의 패턴 분류 문제들로부터 찾은 해역들에 포함된 약 백만 개의 웨이트 점들을 clustering하여 100개의 cluster들을 구하였다. 그림 3은 100개 cluster의 중심 벡터들로부터 추정된 공분산 행렬의 가장 큰 2개의 고유치들에 대한 고유벡터들의 평면으로 매핑된 100개의 cluster들의 분포를 나타낸다 [4]. 많은 cluster들이 일정한 위치에 밀집되어 있는 것을 볼 수 있다.

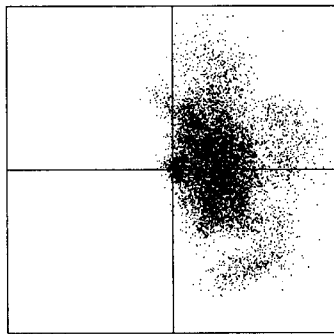


그림 3. 100개 cluster들의 분포.

웨이트 분포에 대한 지식이 새로운 문제를 학습하는데 유용한 정보를 제공하는지를 시험하기 위해, clustering에 참여한 48개의 패턴 분류 문제와는 전혀 다른 406개의 새로운 2 클래스 패턴 분류 문제에 대해 다음과 같은 실험을 하였다. 100개의 cluster들의 중심 벡터를 초기 웨이트로 사용하여 학습 과정 없이 시험

하였을 때, 가장 높은 정확도를 제공하는 cluster 중심 벡터를 선택하고, 또한 100개의 랜덤 초기 웨이트 중 같은 방식으로 가장 높은 정확도를 제공하는 하나의 웨이트 벡터를 선택한다. 이 두 웨이트들을 각각 초기 웨이트로 사용하여 100번, 500번, 그리고 1000번 iteration을 수행하였다. 표 5는 406개의 문제의 각 iteration에서의 분류 정확도의 차이들을 평균한 것을 보여준다. cluster 중심 벡터를 사용한 초기화가 학습이 수행됨에 따라 정확도의 차이는 줄어들지만, 계속적으로 성능의 향상을 보임을 알 수 있다. 더욱이 406개의 분류 문제 중 87개는 학습과정 없이 직접 시험하였을 때, 문제에 대한 최대 정확도와와의 차가 2% 보다 작은 결과를 보여 주었다. 그러나 151개의 문제에 대해서는 랜덤 초기 웨이트와 비교하여 더 낮은 정확도를 나타내었다. 실험 결과를 종합해 볼 때, 웨이트 분포에 대한 지식은 어느 정도 새로운 문제의 학습 능력을 향상시킬 수 있음을 보여주었고, 이러한 웨이트 분포에 대한 지식이 늘어갈수록 신경망의 학습 성능을 더 향상시킬 수 있을 것으로 기대된다.

표 5. 406개의 새로운 문제에 대한 분류 정확도의 차이들의 평균값 (cluster 중심벡터 - 랜덤 웨이트).

	학습 데이터	시험 데이터
학습과정 없이 직접 시험	3.59 %	3.52 %
100번 iteration	2.15 %	1.74 %
500번 iteration	1.02 %	0.65 %
1000번 iteration	0.74 %	0.53 %

4. 결론

본 논문에서는 신경망의 해역의 분포를 분석함으로써, 학습 성능을 향상시킬 수 있는 가능성을 탐구하였다. 실험을 통해 웨이트 분포에 대한 지식이 새로운 문제의 학습 능력을 향상시킬 수 있음을 보여 주었다. 이러한 해역의 분포에 대한 정보가 증가하면 할수록, 신경망은 새로운 문제를 좀 더 효율적으로 학습할 수 있을 것으로 예측된다.

참고문헌

- [1] John A. Richards, *Remote Sensing Digital Image Analysis*. New York: Springer-Verlag, 1993
- [2] R. Lippmann, An introduction to computing with neural nets, *IEEE Acoust. Speech, Signal Processing Mag.*, Apr. 1987.
- [3] I. Bellido and E. Fiesler, Do backpropagation trained neural networks have normal weight distributions?, *ICANN93*, pp. 772-775, 1993.
- [4] K. Fukunaga, *Introduction to Statistical Pattern Recognition*. New York: Academic, 1990.