

회귀 신경망 모델의 은닉 뉴런 간 연결구조에 따른 성능과 동역학적 특성 분석

윤상웅¹ 장병탁^{1,2}

¹서울대학교 협동과정 뇌과학전공

²서울대학교 컴퓨터공학부

{swyoon, btzhang}@bi.snu.ac.kr

Analyzing Performance and Dynamics of Recurrent Neural Network Given Various Structures of Hidden Neuron Connections

Sangwoong Yoon¹ Byoung-Tak Zhang²

¹Interdisciplinary Program in Neuroscience, Seoul National University

²School of Computer Science & Engineering, Seoul National University

요약

시계열 데이터를 다룰 수 있는 기계학습모델인 회귀 신경망은 되먹임 연결을 허용하기 때문에 앞먹임 신경망에 비해 훨씬 다양한 구조를 가질 수 있다. 본 연구에서는 은닉 뉴런 간의 네트워크 구조에 초점을 맞추어 그것이 회귀 신경망의 정보처리 능력에 미치는 영향을 탐구하고자 한다. 이를 위해 회귀 신경망 모델 중 하나인 Echo State Network을 기준으로 하여, 여러 가지 잘 알려진 네트워크 모델에 따라 은닉 뉴런 간 연결을 구성하고 각각의 경우에 시계열 학습 능력과 동역학을 분석하였다. 그 결과, 은닉 뉴런의 네트워크 구조에 따라 모델의 성능이 큰 폭으로 변하는 것이 관찰되었으며, 그러한 현상은 신경망 동역학이 가지는 임계도(criticality)의 변화와 잘 일치했다. 본 연구의 결과는 기존 회귀 신경망 연구에서 주된 관심사였던 신경망 연결 가중치뿐만 아니라 신경망의 연결 구조가 모델의 성능에 중요한 영향을 미친다는 사실을 보여주며, 성능 향상을 위한 중요한 단서가 될 수 있다.

1. 서론

회귀신경망(Recurrent Neural Network, RNN)은 과거 상태에 대한 정보를 기억할 수 있기 때문에 시계열 데이터를 자연스럽게 다룰 수 있는 모델이지만 모델의 학습이 무척 어렵다는 것이 알려져 있다. 그에 따라 다양한 최적화 기법들이 제안되어왔고, 최근에 상당한 진전이 보고된 바 있다[1]. 한편, 그러한 대부분의 연구에서 연결 가중치의 업데이트만이 주된 관심사였다. 본 연구에서는 은닉 뉴런 간 연결 구조에 초점을 맞추어, 연결구조가 다양한 네트워크 모델을 따를 때 RNN의 성능과 뉴런의 발화 동역학이 어떻게 변하는지 분석한다.

2. 연구 방법

2.1 Echo State Network

Jaeger가 제안한 Echo State Network(ESN)은 회귀 신경망의 일종으로, 독특한 접근 방법을 통해 회귀 신경망의 고질적인 가중치 학습 문제를 회피하고 성능 향상을 성취하였다[2]. (그림 1)은 본 연구에 사용된, 가장 단순한 형태의 ESN을 보여준다. ESN은 보통의 Artificial

Neural Network과 같이 입력(input), 은닉(hidden), 출력(output) 뉴런으로 이루어져 있으며, 그 순서대로 정보를 전달한다. ESN에서 입력-은닉, 은닉-은닉 가중치는 임의로 초기화되며, 오직 은닉-출력 가중치만 감독 학습(supervised learning)으로 업데이트된다. 대신 일반적으로 입력, 출력에 비해 아주 많은 수의 은닉 뉴런이 사용된다.

ESN의 경우 은닉 뉴런 연결 가중치의 학습이 일어나지 않으므로 학습 알고리즘과 무관하게 연결 구조의 영향을 분석할 수 있다. 또한 ESN의 구성은 일반적인 RNN과 실질적으로 동일하며, ESN에 대해 성립하는 사실들이 non-ESN RNN에도 쉽게 일반화될 수 있기 때문에 본 연구의 모델로서 ESN을 선정하였다.

본 연구에서 사용된 ESN은 입력 뉴런 2개(시계열 입력과 상수), 은닉 뉴런 200개, 출력 뉴런 1개를 가지고 있으며, 은닉 뉴런 간의 연결을 2.2에 소개될 네트워크 모델에 따라 구성하였다. 은닉 뉴런의 연결 가중치는 (-1, 1) 사이에서 임의로 초기화 하였고, 가중치 행렬의 spectral radius가 0.7을 갖도록 scaling하였다.

2.2 네트워크 모델

자연계의 기저에 존재하는 네트워크를 이해하기 위해 여러 가지 네트워크 모델이 제안되었고 성공적으로 적용

되어 왔다. 예를 들어, 인터넷 연결망은 척도 없는 네트워크(scale-free network) 모델로 설명될 수 있으며, 지인들의 연결망은 작은 세상 네트워크(small-world network)로 분류할 수 있다. 각 네트워크 모델은 특정한 성질을 가지는 네트워크를 구성하는 방법을 제시하며, 노드의 수 등 몇 가지 매개변수를 가진다.

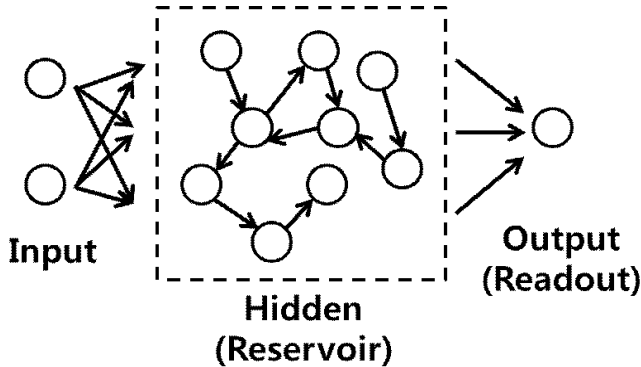


그림 1. Echo State Network의 구조. 은닉 뉴런은 reservoir, 출력 뉴런은 readout이라고 불리기도 하나, 본 논문에서는 은닉 뉴런과 출력 뉴런으로 통일하여 지칭하였다.

본 연구에서는 널리 연구되었거나 자연계에 존재한다고 알려진 네트워크 모델을 바탕으로 은닉 뉴런 간 연결을 구성하였다. Echo State Network에서 기본적으로 사용하고 있는 희소 임의 네트워크, 통상의 RNN에서 널리 사용되는 완전 연결(fully connected) 네트워크, 가장 단순한 형태의 네트워크인 환형 격자(circular lattice), 작은 세상 성질(small-world property)을 가진다고 알려진 Watts-Strogatz 모델, 노드 차수 분포가 power law를 따르며 척도 없는 네트워크인 선호적 연결(preferential attachment) 모델에 대해 분석을 수행했다. 네트워크 모델의 선정에 있어서 [3]에 소개된 내용을 참고하였으며, 각 네트워크 모델의 구현을 위해 CONTEST Toolbox를 사용하였다[4].

2.3 비선형 시계열 데이터

모델의 성능평가를 위해 H. Jaeger가 제안한 비선형 시계열 학습문제를 사용하였다[5]. 과거의 10단계의 입력과 출력으로부터 다음 단계의 출력을 계산하는 문제로서, 이를 성공적으로 수행하기 위해 모델은 비선형 연산 능력과 과거 입출력에 대한 기억능력을 갖추어야 한다. 이 시계열은 (식 1)과 같이 주어진다. 식에서 $I(n)$ 은 n 번째 시점의 입력값으로 0에서 1까지 임의의 값을 가지며, $O(n)$ 은 n 번째 시점의 출력값이다.

$$O(n+1) = 0.3O(n) + 0.05O(n) \left[\sum_{i=0}^9 O(n-i) \right] + 1.5I(n-9)I(n) + 0.1$$

식 1. 본 연구에서 학습 성능 지표로 사용되는 비선형 시계열

본 연구에서 수행한 실험에서는 길이 4400의 시계열을 만들고, 앞뒤 절반으로 나누어 학습 데이터와 시험 데이터로 각각 사용했다. 학습이 이루어진 정도는 시험 데이터에 대한 표준 근제곱평균 오차 (normalized root mean square error, NRMSE)로 측정하였다. 은닉-출력 가중치의 학습과 NRMSE 계산 시, 초기조건에 의한 영향을 무시하기 위해 학습, 시험 데이터의 시작하는 200개의 데이터 포인트를 제외하였다.

3. 은닉 뉴런 연결구조에 따른 시계열 학습 능력

2.1에서 설명한 구조와 2.2에서 설명한 네트워크 모델들에 따라 여러 가지 형태의 은닉 뉴런 연결 구조를 가지는 ESN을 만들고 시계열 학습 성능을 실험했다. 각 네트워크 모델의 매개변수는 시행착오를 통해 최고의 성능을 내는 것으로 선택되었으며, 각 네트워크 모델의 경우에 대해 50회씩 반복 실험을 하고 그 평균과 표준편차를 (그림 2)에 도시하였다. 편의를 위해 이하에서 각 ESN은 자신의 은닉 뉴런 연결 구조로 지칭되었다.

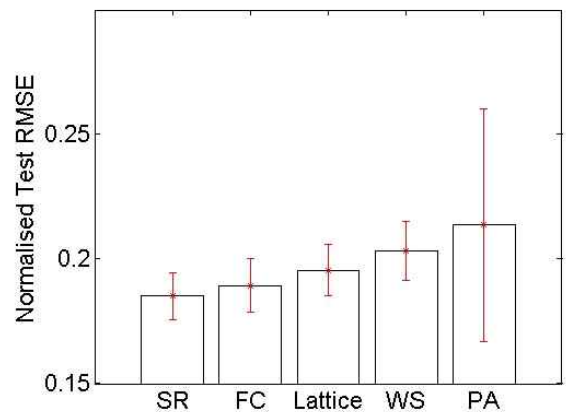


그림 2. 은닉 뉴런 연결구조에 따른 성능.

SR: Sparse Random, FC: Fully Connected, WS: Watts-Strogatz, PA: Preferential Attachment.

ESN의 가장 처음 제안된 형태인 희소 임의 네트워크가 가장 안정되고 좋은 성능을 보여주었으며, 나머지 네트워크는 평균적 성능이 떨어지거나(FC, Lattice, WS) 불안정했다.(PA) 성능의 분산은 입력-은닉, 은닉-은닉 뉴런 간 가중치의 임의 초기화에 기인한 것으로 생각된다.

희소 임의 네트워크가 완전 연결 네트워크에 비해 통계적으로 유의미하게 뛰어난 성능을 보여주었는데(Wilcoxon rank sum test, $p < 0.001$), 많은 non-ESN

RNN이 은닉 뉴런 간의 완전 연결 구조를 사용하고 있다는 점에서 흥미롭다. 이 결과는 희소 임의 네트워크의 성질을 RNN의 성능을 향상시키는 데에 사용할 수 있다는 것을 시사한다. 한편, 선호적 연결 네트워크의 성능은 매우 불안정했으나 일부 아주 뛰어난 경우가 관찰되었다. 이런 특별한 성능을 보이게 하는 요인을 규명하는 것도 유의미한 후속 연구가 될 것이다.

4. 은닉 뉴런 연결구조에 따른 동역학 특성

(그림 2)에서 드러난 은닉 뉴런의 네트워크 구조에 따른 성능 차이의 원인을 알아보기 위해, 은닉 뉴런 발화 동역학의 임계도(criticality)를 측정하였다. 계의 동역학은 임계도가 작을수록 질서 상태에 가깝고 클수록 혼돈 상태에 가까운데, 질서 상태와 혼돈 상태의 경계에서 계의 계산능력이 최대가 되며 그것은 ESN의 은닉 뉴런들에 대해서도 마찬가지라고 알려져 있다[6]. 측정 방법으로 Lazar와 동료들이 수행했던 섭동 분석을 현재 상황에 맞게 Hamming distance 대신 Euclidean distance를 사용하는 방식으로 변형하여 적용하였다[7].

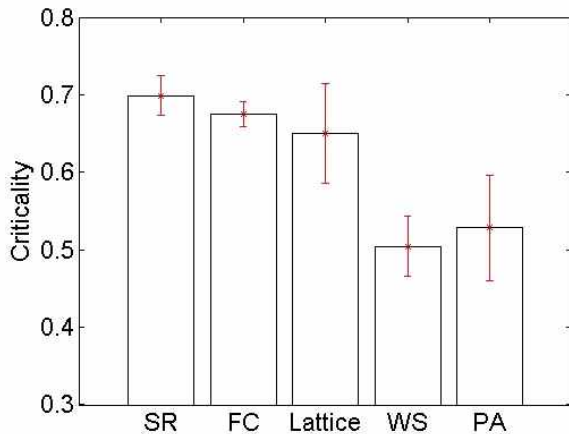


그림 3. 은닉 뉴런 연결 구조에 따른 임계도.
도표의 약어는 그림 2와 같음.

알려진 결과와 마찬가지로, 시계열 예측 능력이 좋을수록 은닉 뉴런 발화 패턴의 임계도가 1에 가까웠다. 성능이 좋지 않은 네트워크들은 낮은 임계도를 보였는데, 이는 발화 패턴이 정보를 처리하기에 충분히 복잡하지 않다는 것을 의미한다. 다만 일부 선호적 연결 네트워크의 경우 희소 임의 네트워크보다 성능이 우수함에도 임계도는 더 낮았다. 낮은 임계도를 갖지만 더 좋은 성능을 보이는 예외 상황이 [7]에도 보고된 바 있으나, 완전하게 이해되지는 않았다. 선호적 연결 네트워크로 구성된 ESN이 보이는 이러한 흥미로운 현상들에 대해 추가적인 분석을 수행해볼 가치가 있을 것이다.

5. 결 론

본 논문에서는 RNN의 일종인 ESN에서 은닉 뉴런 간

연결 구조가 성능과 동역학에 미치는 영향을 보고하였다. 희소 임의 네트워크는 가장 좋은 성능과 가장 질서-혼돈 경계에 가까운 동역학을 보여주었으며, 나머지 형태의 네트워크에서는 보다 낮은 성능과 충분히 복잡하지 않은 동역학이 관찰되었다. 다만 선호적 연결 네트워크는 때때로 보다 좋은 성능을 보여주었는데, 이때는 하임계적 동역학이 관찰되었다. 희소 임의 네트워크의 성질과 선호적 연결 네트워크 중 특이한 사례들에 대해 더 연구하여 RNN에 적용한다면 성능향상을 기대해볼 수 있을 것이다.

감사의 글

이 논문은 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구이며 (NRF-2010-0017734-Videome), 정부(산업통상자원부)의 재원으로 한국산업기술평가관리원 지원 (KEIT-10035348-mLife, KEIT-10044009)을 일부 받았음.

참고 문헌

- [1] I. Sutskever, Training Recurrent Neural Networks, University of Toronto, PhD thesis, 2013.
- [2] H. Jaeger, The “echo state” approach to analysing and training recurrent neural networks, German National Research Center for Information Technology GMD Technical Report 148, 2001.
- [3] M. E. J. Newman, Networks: An Introduction, Oxford University Press, 2010.
- [4] A. Taylor, and D. J. Higham, CONTEST: A Controllable Test Matrix Toolbox for MATLAB, *ACM Transactions on Mathematical Software*, 35, 4, 26, 2009.
- [5] H. Jaeger, Adaptive Nonlinear System Identification with Echo State Networks, *Advances in Neural Information Processing Systems*, 593-600, 2002.
- [6] V. Gómez, A. Kaltenbrunner, V. López, H. J. Kappen, Self-organization using synaptic plasticity, *Advances in Neural Information Processing Systems*, 513-520, 2008
- [7] A. Lazar, G. Pipa, and J. Triesch, SORN: a Self-Organizing Recurrent Network, *Frontiers in Computational Neuroscience*, 3, 23, 2009.