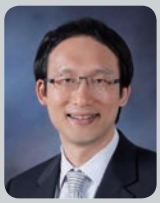
대규모 데이터 기반의 뇌 신경망 (Brain Network) 연구 동향

김창업
서울대학교 의과대학
신경면역 정보저장
네트워크 연구센터



김상정
서울대학교 의과대학
신경면역 정보저장
네트워크 연구센터

Ⅰ. 서 론

최근 신경과학 분야의 연구 풍경이 빠른 속도로 변화하고 있다. 관찰 기술의 제약으로 오랜 시간 소규모 신경세포 수준의 연구에 집중해왔던 신경과학계가 뇌 신경망(brain network) 수준의 데이터를 생산, 분석할 수 있게 되면서 본격적인 '빅데이터' 시대로의 진입을 준비하는 모양새다. 미국의 국가과학프로젝트로 2014년 시작된 BRAIN Initiative는 대규모 신경활성 기록 기술을 비롯하여 신경세포집단 전체를 낱낱이 밝힐 수 있는 기술들을 개발하고, 이를 통해 뇌 신경망의 동역학을 이해해보자고 독려 중이다[1]. 이미 '빅데이터' 시대로 진입한 유전체학, 천체물리학 등의 다른 분야와 마찬가지로 새로운 흐름은 연구자들의 관점과 접근 방식에 근본적인 변화를 유도하고 있으며, 이에 따라 데이터 표준화, 공개 협력연구, 대용량 데이터의 처리와 분석 등, 다양한 주제로 매우 흥미로운 논의들이 진행되고 있다. 본 글에서는 그 중에서 고차원 시계열 데이터(high-dimensional time series data)의 특성을 갖는 신경망의 데이터 분석방법에 초점을 맞추어 최근의 동향을 소개하고자 한다.

뇌 신경망 데이터는 크게 거시수준(large-scale level) 데이터와 중간(meso), 혹은 미시수준(micro-scale level)의 데이터로 구분될 수 있다. 전자는 주로 인간을 대상으로 한 뇌영상 연구(fMRI, PET, MEG, EEG 등)를 통해 얻어지며 뇌의 넓은 범위를 포함하지만 개별 세포수준의 정보가 아닌 영역(region) 수준에서의 데이터를 제공한다. 후자는 개별 신경세포 수준의 정밀도로 여러 신경세포의 활성을 기록하지만 동시에 기록할 수 있는 신경세포의 숫자는 일반적으로 수백 개를 넘지 못한다. 또한 침습적인 기술 특성상 사람이 아닌 동물을 대상으로만 실험이 이루어지고 있다. 양 분야 모두 뇌의 작동원리를 이해하는 것

을 목표로 하지만, 산의 정 반대방향에서 각자 터널을 파며 들어가고 있는 형국이어서 아직 두 연구 진영간의 갭은 상당히 큰 상황이라 할 수 있다. 때문에 신경망 분석이라는 공통된 주제에 대해서도 각자의 영역에서 다소 상이한 방법들이 발전해나가고 있는 상황이다. 향후 가속화될 대규모 신경활성 기록 기술 및 관련 테크놀로지의 발전으로 이러한 갭이 메꾸어 질 것으로 기대된다. 본 글에서는 기능적 자기공명영상(fMRI), 양전자방출단층촬영술(PET) 등을 중심으로 이뤄지고 있는 인간 뇌영상 기반의 뇌신경망 연구와 multi-electrode array recording, two-photon calcium imaging등을 주요 수단으로 이뤄지고 있는 동물의 신경망 연구를 차례로 살펴본다.

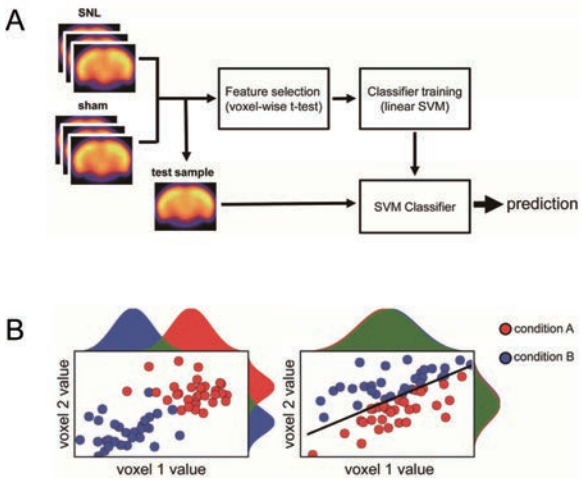
뇌 신경망 데이터는 거시수준 데이터와 중간, 혹은 미시수준의 데이터로 구분될 수 있다.

Ⅱ. 관련연구

1. 인간 대상의 뇌 영상(human brain imaging) 연구

A. MVPA(multi-variate pattern analysis)

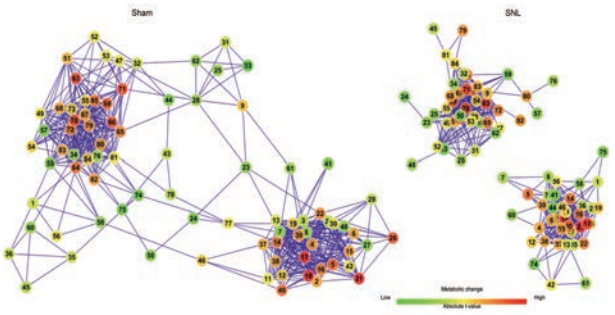
90년대 이후 fMRI, PET 등, 인간 대상의 뇌영상 연구가 폭발적으로 증가하면서 거시수준의 뇌 이해에 많은 진전이 있었다. 2000년대 중반까지 수행된 일반적인 연구들은 특정 행동, 질병 등의 조건에서 활성화되는 뇌 영역을 찾기 위해 수만 개에 이르는 voxel(volumetric pixel)의 활성도를 행동, 혹은 질병상태간에 통계적으로 비교하였다. 그러나 이러한 분석은 기본적으로 voxel간의 독립성을 가정하고 이루어지므로, 실제 긴밀하게 연결되어 작동하고 있을 뇌 신경망에 대한 이해는 제한될 수 밖에 없었다. 이러한 접근법을 MVPA와 대비하여 univariate analysis라고 한다. 반면 MVPA는 조건에 관여하는 voxel들의 '패턴'을 찾고자 한다. 구체적인 방법으로서 기계학습의 감독 학습(supervised learning) 기술을 이용하며, 데이터를 훈련군(training set)과 테스트군(test set)으로 나누어 교차검증을 시행, voxel들의 패턴 정보로부터 해당 조건을 성공적으로 예측할 수 있는지 평가한다(〈그림 1A〉). 성공적인 예측은 이용된 voxel들의 패턴, 즉 뇌 영역들의 활성패턴이 해당 조건에 관여하고 있음을 시사하게 된다. 이때 이용된 voxel들에 대해 독립적으로 univariate 분석을 수행할 경우 해당 조건과 통계적으로 유의한 관련을 보이지 못할 수도 있다(〈그림 1B〉). 즉 뇌의 각 영역들은 독립적으로 기능하지 않으며 뇌의 기능은 뇌활성의 분산된 패턴(distributed pattern)으로 나타남을 시사한다. 하지만 MVPA를 이용한 분석은 관련된 뇌영역들의 활성 패턴을 조사할 뿐, 상호 연결된 뇌 신경망이 어떻게 작동하고 있는지에 대한 정보를 주지는 않는다. 또한 가장 높은 예측력을 보인 뇌영역들의 집합이 반드시 생리학적으로 그 기능에 관여하는 활성패턴이라고 할 수 있는 지에 대해서도 비판의 여지가 있다.



〈그림 1〉 A. MVPA 를 이용한 뇌영상 분석 과정. B. Univariate analysis만으로도 두 조건이 구분되는 경우(왼쪽)와 MVPA를 적용하지 않으면 구분할 수 없는 경우(오른쪽)

B. 정적 기능적 연결성(static functional connectivity) 분석

뇌 영역간의 기능적 연결성은 뇌 활성 데이터의 변이량으로부터 계산될 수 있다. 시간, 혹은 개체에 따른 활성의 변화가 두 영역간에 유사하다면(covary), 두 영역 사이에 기능적인 연결관계가 존재한다고 보는 것이다. 그간의 많은 연구를 통해 기능적 연결성이 구조적 연결성(structural connectivity)을 반영하고 있으며[2], 특정 행



〈그림 2〉 뇌 영상 데이터로부터 구성된 graph

위나 질병 등의 상태를 반영하고 있음이 확인 되었다[1]. 기능적 연결성은 상관 계수(correlation coefficient)나 Granger causality 등으로 측정되는데, 전자가 두 영역간의 단순 상관성을 의미한다면, 후자는 인과관계(causality)를 추론한다[3].

기능적 연결성 연구는 초기 관심 영역간의 연결성을 조사하는 연구에서 시작하여 seed-based analysis와 brain network analysis로 발전하였다. Seed-based analysis는 관심 영역(주로 특정 행위, 혹은 질병 상태에서 높은 활성을 보인 영역)의 voxel을 seed로 설정하고, seed voxel과 나머지 전체 voxel들간의 연결성을 조사한다. 관심 영역과 뇌 전체 영역간의 연결관계를 파악할 수 있으나, 관심영역 설정 단계에서 연구자의 편견이 들어갈 수 있는 단점이 있으며, 뇌의 신경망을 전체적으로 조망할 수는 없다는 한계가 있다. Brain network analysis는 뇌 영역들을 node로, 정의한 연결성을 edge로 표현하여 그래프(graph)를 구성하고, 복잡계 네트워크(complex network) 이론을 적용하여 신경망의 topology를 분석한다(〈그림 2〉). 허브(hub), 모듈 구조, small-worldness, network efficiency와 등을 조사함으로써 뇌 신경망이 갖는 특성 뿐 아니라, 특정 행동이나 질병상태에서 뇌신경망의 topology가 어떻게 변화하는지 알 수 있다. 그러나 graph를 구성하기 위해 node를 어떻게 정의해야 하는지, 연결성 계산시 역치값의 설정은 어떻게 이루어져야 하는지 등, 명확한 기준이 없는 분석과정들이 결과에 큰 영향을 미칠 수 있음이 지적되고 있으며, 이 부분은 아직 극복해야 할 과제로 남아있다[4].

C. 동적 기능적 연결성(dynamic functional connectivity) 분석

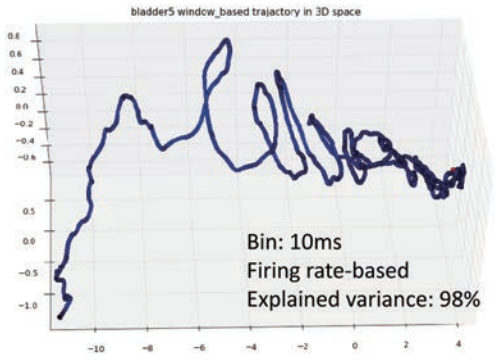
MVPA와 기능적 연결성에 기반한 연구들이 뇌의 거시적 작동원리에 대해 많은 새로운 사실을 밝혀낸 것은 사실이지만, 데이터 수집기간 동안 뇌의 평균적인 활동양상을 조사함으로써 정적인 신경망(static network)을 연구할 수 밖에 없다는 한계점이 있다. 그러나 뇌는 복잡한 동역학적 특성을 갖고 있음이 이미 알려져 있으므로 신경망의 연결패턴 역시 동적으로 변화하고 있다고 가정하는 것이 합당하다. 최근 동적 기능적 연결성(dynamic functional connectivity)의 탐구가 시작되면서 인간 뇌 신경망의 보다 완전한 시공간적 이해에 접근하고 있다. 주로 안정시(resting state) fMRI의 시계열 데이터에 일정한 time window를 설정, window 내 데이터로부터 네트워크를 구성하고, 이 window를 overlapping 하는 방식으로 이동시키며 연속적인 네트워크를 구성하는 sliding window analysis가 시도된다[5]. 지금까지의 연구에서 뇌 신경망의 연결성 변화는 연속적인 변화가 아니라, 동역학계의 fixed point와 같이 구분되는 특정 패턴이 반복적으로 나타나고 유지되는 multistable state의 양상을 띠는 것으로 드러나고 있다. 또한 이러한 동적 연결패턴의 변화가 질환상태와 관련되며, 연결패턴의 변화와 함께 네트워크 속성도 동적으로 변화한다는 사실도 확인되었다[6]. 그러나 측정된 연결성의 변화는 심장, 호흡의 변화, 하드웨어의 불안정성 등, 다양한 노이즈에 기인할 수 있으므로 결과의 분석과 해석에 많은 주의가 필요하다. 동적 기능적 연결성 연구는 아직 탐색적인 연구 단계에 있으며 신뢰할 수 있는 분석방법을 확립하기 위해선 다소간의 시간이 필요할 것으로 보인다.

2. 동물 대상의 대규모 신경활성 기록(large-scale neural activity recording) 연구

A. 기초적인 tuning curve를 이용한 분석

동물 대상의 신경망 연구는 인간 대상의 뇌 영상 연구와 반대로, 세포수준의 정밀성을 확보한 상태에서 스케일을 넓혀나가고 있다. 전통적으로 여러 신경세포의 활

성을 기록하기 위하여 multi-electrode array (MEA) recording이 이용되어왔으며 이를 이용하여 local field potential(LFP)을 기록하거나 개별 세포들의 spike를 기록, 분리해낼 수 있다. LFP 기록은 기록자극 근처의 통합된 전기적 활성을 나타낼 뿐, 개별 세포수준의 정보를 주지는 않는다. Spike 기록은 개별세포 수준의 활동전위를 기록할 수 있으나 기록된 세포들의 공간정보가 매우 제한적이라는 한계가 있다. Tuning curve는 일련의 자극(빛 자극의 방향 혹은 팔을 움직이는 각도 등)에 대한 신경세포의 관련 활성(firing rate)을 Gaussian curve, cosine curve 등으로 모델링함으로써 얻어진다. 즉, 개별 세포가 특정 자극에 얼마나 반응하는지에 대한 선호도로 이해될 수 있으며, tuning curve를 구한 후엔 역으로 신경세포들의 활성 데이터로부터 주어진 자극을 추론할 수 있다. 이때 신경세포들이 서로 독립적으로 기능한다고 가정하고 개별 세포들의 tuning curve 정보와 활성도를 단순 선형결합하여 자극을 추론하는 방식이 가장 기초적인 분석방식이다.



〈그림 3〉 MEA로 기록된 신경세포집단의 궤도(trajectory)

최근 GLM이나 MAP 추정과 같은 통계 신호처리를 이용한 신경세포간 기능적 연결성에 대한 연구가 활발히 진행되고 있다.

B. 신경세포간 기능적 연결성 (functional connectivity between neurons) 분석

고전적으로 신경세포의 spike train data로부터 연결성을 추론하기 위하여 cross-correlogram이나 joint PSTH(peri-stimulus time histogram) 분석이 이루어져왔으나 이런 방식으로 추론된 연결성은 기록되지 않은 숨겨진 변수나(신경세포 혹은 외부자극) 관찰된 변수들간의 복잡한 연결관계에서 오는 간접적 상관성에 의하여 교란되는 한계가 있다. 최근엔 GLM(generalized linear model)과 같은 모델을 기반으로 한 신경망 분석이 활발하게 연구, 적용되고 있다[7]. GLM은 신경세포들의 tuning curve와 상호 연결관계, 불응기(refractory period)등을 parameter를 갖는 모델로서, 신경세포들의 활성기록을 변수로 받아 확률모형에 따라 예측을 수행한다. 모델 적합을 위해 maximum likelihood methods가 주로 이용되나 데이터에 비해 parameter가 많은 경우(신경세포의 수가 늘어나면 연결이 늘어나고 필연적으로 parameter가 증가한다), 모델 적합에 사용된 데이터는 잘 설명하지만 새로운 데이터는 설명하지 못하는 과적합(overfitting) 문제가 발생하게 된다. 이를 막기 위해 basis function을 이용하여 parameter space를 제한하거나, 연결관계에 대한 기존 지식을 포함, Bayes' rule을 적용하여 MAP(maximum a posteriori) 추정을 할 수도 있다.

C. 차원축소(dimensionality reduction)와 동역학(dynamics) 분석

동시에 측정된 여러 신경세포들의 활성기록은 다차원 시계열의 데이터(high dimensional time series data)로서, 차원축소기법을 적용하여 2차원, 혹은 3차원 공간 상에 시간에 따라 evolve하는 궤도(trajectory)를 표현할 수 있다[8](〈그림 3〉). 이 경우 모델 구성을 위한 parameter 적합이 필요 없으므로 과적합 문제를 피할 수 있을 뿐 아니라, 동시기록 신경세포의 숫자가 증가하는 경우 모델 기반 분석에 비하여 효율적으로 신경망의 동역학을 탐색할 수 있는 수단이 된다. 차원축소를 위하여 PCA(principal component analysis)와 같은 선형적 축소기법이 주로 이용되며, 궤도로 표현된 신경망의 시간에 따른 변화양상은 동역학적 관점에서 분석될 수 있다.

Ⅲ. 향후 연구 및 결론

신경망의 복잡하고 역동적인 패턴이 어떻게 인지, 기억, 정서를 만들어내는지 이해하기 위해서는 실제 신경세포로 구성된 신경망의 데이터를 획득하고 분석하는 것이 필수적이다. 최근의 신경과학 연구자들은 과거 어느 때보다 실제 신경망을 잘 관찰할 수 있는 기술과 분석, 처리 능력을 갖추고 있으며 관련 기술혁신의 속도는 매우 가파르다[9]. 향후 신경망 연구는 과거와 달리 대규모의 실제 데이터를 중심으로 전개되리라 예상되며 이에 따라 실험가와 분석가, 이론가, 공학자간의 긴밀한 협력 연구가 보다 중요해지리라 전망된다.

참고문헌

[1] http://www.braininitiative.nih.gov/2025/index.htm.

[2] Greicius, M.D., et al., Resting-state functional connectivity reflects structural connectivity in the default mode network. Cereb Cortex, 2009. 19(1): p. 72-8.

[3] Roebroeck, A., E. Formisano, and R. Goebel, Mapping directed influence over the brain using Granger causality and fMRI. Neuroimage, 2005. 25(1): p. 230-42.

[4] Smith, S.M., The future of FMRI connectivity. Neuroimage, 2012. 62(2): p. 1257-66.

[5] Hutchison, R.M., et al., Dynamic functional connectivity: promise, issues, and interpretations. Neuroimage, 2013. 80: p. 360-78.

[6] Zalesky, A., et al., Time-resolved resting-state brain networks. Proc Natl Acad Sci U S A, 2014. 111(28): p. 10341-6.

[7] Stevenson, I.H., et al., Inferring functional connections between neurons. Curr Opin Neurobiol, 2008. 18(6): p. 582-8.

[8] Cunningham, J.P. and B.M. Yu, Dimensionality reduction for large-scale neural recordings. Nat Neurosci, 2014. 17(11): p. 1500-9.

[9] Stevenson, I.H. and K.P. Kording, How advances in neural recording affect data analysis. Nat Neurosci, 2011. 14(2): p. 139-42.

김창업

- 2007년 2월 동국대학교 한의학과 학사
- 2014년 2월 서울대학교 의과대학 의학박사 (생리학)
- 2014년 3월~현재 서울의대 신경면역 정보저장 네트워크 연구센터 Post Doc. 정보의학인증의(CPBMI)

〈관심분야〉
Neural coding, High-dimensional time series data, Systems medicine

김상정

- 1990년 서울대학교 의과대학 학사
- 1996년 서울대학교 의과대학 의학박사 (생리학)
- 2010년~현재 서울대학교 의과대학 생리학 교실 교수
- 2011년~현재 신경면역 정보저장 네트워크 연구센터 센터장
- 2014년~현재 서울대학교 신경과학 연구소 소장
- 2014년~현재 서울대학교 의과대학 생리학교실 주임교수
- 2015년~현재 서울대학교 대학원 의과학과 학과장

〈관심분야〉
Learning and memory, Chronic pain, In vivo brain imaging