

# 修士論文概要書

Summary of Master's Thesis

Date of submission: 02/29/2020

専攻名 (専門分野) Department	電気・情報生命	氏名 Name	永山 瑞生	指導教員 Advisor	村田 昇
研究指導名 Research guidance	情報学習システム	学籍番号 Student ID number	5319E056-6		
研究題目 Title	NMF とバギングを用いたカルシウムイメージングデータのクラスタリング方法の一検討				

## 研究背景

脳内の情報伝達には複数のニューロンが関わっている。そのため、同時活動するニューロングループの推定が脳の機能解明につながる可能性がある。脳内の多くのニューロンを観察する方法に、蛍光イメージングの1つであるカルシウムイメージングがある。ニューロンで活動電位が発生（発火）すると細胞内の  $\text{Ca}^{2+}$  濃度が上昇する。カルシウムイメージングでは、この  $\text{Ca}^{2+}$  濃度上昇を蛍光で可視化する。蛍光イメージングを用いる利点として、(1) 高い空間分解能、(2) 広い観察範囲、(3) 遺伝子工学と併用して興奮性／抑制性ニューロンの同定などをした上での観察ができることが挙げられる。一方、時間分解能が電気生理学的手法よりも低いことが蛍光イメージングの欠点である。 $\text{Ca}^{2+}$  濃度の変化はニューロンの電気的变化よりも遅く、また、カルシウム感受性蛍光分子のキネティクスも影響する。さらに、カメラやレーザースキャンでのサンプリングレートは高くても 100[Hz] 程度であり、ニューロンの個々の発火を全て捉えるには不十分である。

本研究で扱うデータは、睡眠時と覚醒時のニューロンの活動を解析するために計測されたマウスのカルシウムイメージングデータである。8[Hz] のサンプリングレートで1時間おきに15分間のイメージングが計6回行われ、100～200個のニューロンが観測された。8[Hz] のサンプリングレートではニューロンのシナプス伝達を抽出することはできないため、ニューロンの機能的なコネクティビティを抽出することを目標とする。機能的なコネクティビティとは、つまり同時に活動するニューロンが持つ結合のことである。よって本研究では、ニューロンのクラスタリングの問題を解く。クラスタリングではグループに所属しないニューロンの推定とハードクラスタリングを目的とする。提案アプローチによってどれほどの情報が抽出できるかを人工データ実験によって確かめる。

## 提案アプローチ

ニューロンのクラスタリングには相互相関行列をクラスタリングする方法が考えられる。しかし、相互相関行列ではグループに所属しないニューロンを閾値によって決めることは難しい。提案アプローチではニューロン同士が同じグループになる確率値をバギングによって推定することで閾値を決めやすくする。また、提案アプローチでは nonnegative matrix factorization (NMF) を用いてニューロン間の隣接行列を推定する。NMF の基底数の推定は難しいため、複数の基底数の推定結果

を平均する。NMF の推定量を隣接行列とするのは、異なる基底数で推定量のサイズが変化しないためである。

提案アプローチは2段階に分かれる。まず、(1) ニューロン間の隣接行列  $\hat{A}(X, K)$  を推定し、(2) スペクトラルクラスタリングによってクラスタリングを行う。

ニューロン  $i$  の観測時系列は  $x_i: \in \mathbb{R}_+^J$  である。 $c_k: \in \mathbb{R}_+^J$  ( $k = 1, \dots, K$ ) をグループ  $k$  の活動の時系列とすると、仮定より  $x_i$  は  $c_i$  の重み付き和として表される：

$$x_i = \sum_{k=1}^K d_{ik} c_k + \eta_i. \quad (1)$$

ただし、 $d_{ik} \in \mathbb{R}^+$  で、 $\eta_i: \in \mathbb{R}^J$  はガウスノイズの時系列である。(1) は行列形式で以下のように表現できる：

$$X = DC + H. \quad (2)$$

ただし、 $X \in \mathbb{R}_+^{I \times J}$ ,  $D \in \mathbb{R}_+^{I \times K}$ ,  $C \in \mathbb{R}_+^{K \times J}$ ,  $H \in \mathbb{R}^{I \times J}$  である。また、 $D$  の要素  $(i, k)$  は  $d_{ik}$ ,  $C$  の  $i$  行は  $c_i$ ,  $H$  の  $i$  行は  $\eta_i$  である。

NMF は行列分解手法の一つであり、1つの非負行列を2つの非負行列の積として分解する。提案アプローチでは、NMF を用いて (2) を解いて  $D$  と  $C$  を求める。

NMF での推定量を  $\hat{A}(X, k) \in \{0, 1\}^{I \times I}$  とする。NMF の寄与率行列  $P \in \mathbb{R}^{I \times K}$  の要素を次のように定義する：

$$p_{ik} = \frac{\|d_{ik} c_k\|_1}{\sum_{l=1}^K \|d_{il} c_l\|_1}$$

要素  $p_{ik}$  はニューロン  $i$  に対する基底  $k$  の寄与率という意味である。 $\hat{A}(X, k)$  は、 $P$  の行ごとに最大値を1、それ以外を0とした行列  $G$  を用いて、

$$\hat{A}(X, k) = GG^\top,$$

と定義する。 $\hat{A}(X, k)$  は対称行列で、要素  $\hat{a}_{ij}$  はニューロン  $i$  とニューロン  $j$  が同じ基底に所属していると推定したら1、していないと推定したら0となる。

バギングを用いて隣接行列  $\hat{A}(X, K)$ ,  $K = \{K_{min}, \dots, K_{max}\}$  をブートストラップ推定量  $\hat{A}$  の平均として求める。データ  $X$  からブートストラップサンプル  $\{X^{*b}; b = 1, \dots, B\}$  を作成し、 $\hat{A}(X^{*b}, k)$  を推定する。NMF では基底数を決めなければならないが、確立された推定方法がなく難しい問題である。基底数を間違えた時のリスクを下げるため、異なる基底数  $K$

のモデル平均をとる． $\bar{A}(X, \mathcal{K})$  は  $\hat{A}(X^{*b}, k)$  の期待値として算出する：

$$\bar{A}(X, \mathcal{K}) = \frac{1}{|\mathcal{K}|} \sum_{k \in \mathcal{K}} \frac{1}{B} \sum_{b=1}^B \hat{A}(X^{*b}, k).$$

$\bar{A}(X, \mathcal{K})$  のクラスタリングにはスペクトラルクラスタリングを用いる．スペクトラルクラスタリングは，隣接行列からグラフラプラシアンを求め，その固有ベクトルを k-means などでもクラスタリングする手法である． $\bar{A}(X, \mathcal{K})$  は隣接行列なのでスペクトラルクラスタリングと相性が良いと考えられる．

## 実験結果

ニューロン 1000 個の活動のシミュレーションを行い，人工のカルシウムイメージングデータを作成して，提案アプローチがどこまでうまく働くかを検証した．その結果，NMF を使う有用性，バギングとモデル平均を用いることで精度の向上が確認できた．