

2020 年度修士論文

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

5319E056-6 永山 瑞生

2021 年 2 月 5 日

早稲田大学 先進理工学研究科
電気・情報生命専攻
情報学習システム研究室
指導教員 村田昇

修士論文概要書

Summary of Master's Thesis

Date of submission: 02/05/2020

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 Ca^{2+} 濃度が上昇する。カルシウムイメージングでは、この Ca^{2+} 濃度上昇を蛍光で可視化する。カルシウムイメージングでは多数のニューロンを観測できる一方、時間分解能が電気生理学的手法よりも低いという欠点がある。 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。

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、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問題を解く。本研究では簡単のため、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。

(1) ニューロン間の隣接行列の推定
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。式 (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 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 が同じ基底に所属すると推定したら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 $\bar{A}(X, K)$ 、 $K = \{K_{\min}, \dots, K_{\max}\}$ 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。 $\bar{A}(X, K)$ は $\hat{A}(X^{*b}, K)$ の期待値

として算出する：

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

(2) スペクトラルクラスタリング

隣接行列 $\bar{A}(X, \mathcal{K})$ 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リアンを求め、その固有ベクトルを k-means などで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。クラスタ数はスペクトラルクラスタリングで用いられる固有値ギャップによって推定する。

実験結果

ニューロン 1000 個についてニューロングループを定義し、人工のカルシウムイメージングデータを作成した。100 個の興奮性ニューロンについて提案アプローチによってクラスタリングを行い、先行研究手法と相互相関行列を用いた時との比較を行った。その結果、提案アプローチの方がニューロングループが活動する時間が少ない場合や、グループに所属しないニューロンがある場合に精度が良かった。

提案アプローチを用いて 1 匹のマウスの実データ解析を行った結果、睡眠状態では大規模クラスタが推定された。

目次

1	序論	1
1.1	カルシウムイメージングデータ	1
1.2	関連研究	3
1.2.1	脳データの解析	3
1.2.2	カルシウムイメージングの解析例	4
1.3	目的	4
2	準備	7
2.1	非負行列因子分解	7
2.1.1	NMF の基底数	7
2.2	バギング	8
2.2.1	ブートストラップ法	8
2.3	スペクトラルクラスタリング	10
3	NMF とバギングを用いたクラスタリング方法	13
3.1	隣接行列の推定	13
3.1.1	数理モデル	13
3.1.2	NMF による \hat{A} の推定	15
3.1.3	バギング	16
3.1.4	モデル平均	17
3.1.5	NMF の一意性	17
3.2	\bar{A} を用いたクラスタリング	19
3.2.1	評価	19
4	実験	21
4.1	人工データ実験	21
4.1.1	シミュレーション	21
4.1.2	\bar{A} の推定に関する実験	23
4.1.3	\bar{A} のクラスタリングに関する実験	26
4.2	実データ解析	33
4.2.1	実データ	33
4.2.2	提案アプローチのクラスタリング結果	35
4.2.3	相互相関行列との比較	35
4.3	補遺	38
4.3.1	ネットワーク構造のパラメータ値の決め方	38
4.3.2	スパイクシミュレーションの設定	38
5	結論	43
付録 A	考案した NMF の制約	51
1	スケール除去	51
2	重複除去	51
3	時間方向への制約	52

付録 B	NMF のモデルエビデンス	55
1	ラプラス近似	55
2	ブートストラップによる近似	56
3	実験	56

第1章 序論

脳の機能解明のためにニューロンの活動解析が行われている。ニューロンは複数で活動することによって情報処理を行っているため、集団に注目して解析することが重要である。ニューロンの観察方法として、パッチクランプ法、細胞内記録法、細胞外記録法などの電気生理学的な手法が挙げられる。これらの手法は十分な時間分解能かつ細胞レベルでニューロンを観察することができる。しかし、電気生理学的な手法では電極から近いニューロンしか電位変化の計測できないため、観察できるニューロンの数は数十から多くても数百程度である。

より多くのニューロンを観察するために、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という手法が用いられる。ニューロンで活動電位が発生（発火）すると細胞内の Ca^{2+} 濃度が上昇する。カルシウムイメージングでは、この Ca^{2+} 濃度上昇を蛍光で可視化する。具体的には、 Ca^{2+} と結合すると蛍光強度が変化する蛍光分子を細胞内に発現させておき、 Ca^{2+} 濃度を蛍光強度として蛍光顕微鏡で観察する。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な手法よりも低いことが蛍光イメージングの欠点である。 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。これより、カルシウムイメージングデータからはニューロン活動の相関しか取ることができないと考えられる。

ニューロン集団の解析にはクラスタリングやネットワーク構造推定などが考えられる。相関しか取れないカルシウムイメージングデータからは向きのある有向グラフの推定は難しく、向きのない無向グラフの推定しかできない。クラスタリングは、同じグループ内のニューロン同士がエッジを持つ無向グラフとしても考えられる。結果の解釈はクラスタリングの方が行いやすい。ニューロンの集団解析で考えられる最も単純な方法は、相互相関行列を用いてクラスタリングを行う方法である。しかし、ニューロンが同時活動する時間が短い場合には同じグループ内のニューロンの相関がグループ外のニューロンと比べて高くなり、意味のある情報が抽出できないと思われる。また、どのグループにも所属しないニューロンの存在も考えられるが、相互相関行列ではこのようなニューロンの判定は難しい。

本研究では、低い時間分解能のカルシウムイメージングデータからニューロンをクラスタリングするアプローチを提案し、人工データ実験を通してどの程度の情報が抽出できるかを確認する。提案アプローチでは、ニューロンが同じグループに所属する確率値を無向グラフの隣接行列として推定し、それをスペクトラルクラスタリングによってクラスタリングする。グループに所属しないニューロンは、他のニューロンとグループになる確率が低いものとして推定することができる。隣接行列の推定には数値モデルに基づいて non-negative matrix factorization (NMF) を用いることで、相互相関行列や他の行列分解手法よりも精度の良いクラスタリングができると考えられる。

1.1 カルシウムイメージングデータ

本節では、カルシウムイメージングデータがどのようなデータなのかを説明し、どの程度の情報が取り出せるかを議論する。

ニューロンは1[ms]単位で活動電位が発生する（発火）。活動電位は細胞内外のイオン濃度

が局所的に変化することによって生じる。活動電位は細胞体から軸索を伝わり、シナプスを介して結合している別のニューロンに伝わる。哺乳類の皮質ニューロンにおいて、ニューロンからニューロンへシナプスを介して活動電位が伝わるには数十 [ms] かかる [1]。このように活動電位を伝えることによってニューロン間で情報がやり取りされる。

ニューロンが他のニューロンとコネクションを持つ状態のことをコネクティビティという。脳のコネクティビティには、anatomical connectivity と functional connectivity と effective connectivity の3種類がある [2]。Anatomical connectivity はニューロンが物理的に繋がっていることを指す。Functional connectivity を持つ2つのニューロンは活動に相関があり、機能が同じことを指す。例えば、ある物を見た時、それに特異的に反応するニューロン群は functional connectivity を持つ。Effective connectivity は、1つのニューロンから別のニューロンへ情報が伝わることを指す。脳の機能解明のためには functional connectivity と effective connectivity の推定が必要である。脳内にはセルアセンブリというニューロン群が存在すると言われており [3]、セルアセンブリ内のニューロンはある刺激の元で同時に活動することが知られている。セルアセンブリは functional connectivity を持っていると考えられるので、functional connectivity を推定することはセルアセンブリの推定につながる。Effective connectivity は情報伝達の順番が分かるため、これも脳の機能解明につながる。例えば、リアクチベーション現象があることが知られている。リアクチベーション現象とは、ニューロンが同じ順番で発火することが繰り返されることであり、記憶の定着などにも関わっていると言われている。データによってどのコネクティビティの情報を取り出せるかは異なり、電気生理では effective connectivity が推定できるが、サンプリングレートの低いカルシウムイメージングデータでは functional connectivity までしか推定ができない。

1個1個のニューロンを1[ms]単位で計測できればニューロン全ての活動を計測できるが、そのような技術は存在しない。ニューロンの計測方法には様々なものがあり、それぞれ計測可能な時間分解能と空間分解能が異なる。計測方法別の分解能については [4] の Fig 1 が分かりやすい。EEG, PET, fMRI は脳の一部のニューロンの活動によって生じた電位、血流、代謝量の変化を計測する。これらの手法ではニューロン単位の計測は行えないが、脳全体を計測することができる。電気生理学的な手法やカルシウムイメージングではニューロン単位で膜電位の変化やカルシウムイオン濃度の変化を計測する。これらの手法ではニューロン1個1個を計測できるが脳全体を計測することはできない。

カルシウムイメージングとはニューロン内のカルシウムイオン濃度を可視化することでニューロンの活動を計測する手法である。観察したい個体のニューロンにカルシウムイオンと結合する蛍光タンパク質を発現させると、細胞内のカルシウムイオン濃度に応じて蛍光強度が変化する。ニューロンが発火するとカルシウムイオンが流入するため蛍光強度が上昇し、その後徐々に蛍光強度は減少する。ニューロンを蛍光顕微鏡で観察することで図 1.1 のような蛍光強度を反映した画像が得られる。その画像から個々のニューロンの蛍光強度のデータを取得できる。

電気生理学的な手法と比べた時のカルシウムイメージングの利点として、ニューロンの位置情報が分かるため同じニューロンを複数回観察できることとより多くのニューロンの測定ができることが挙げられる。また、ニューロンには興奮性ニューロンと抑制性ニューロンの二種類があり、カルシウムイメージングではこの種類を見分けることができる。二種類の蛍光タンパク質を用いて、片方の蛍光タンパク質を抑制性ニューロンのみに発現させることで興奮性・抑制性が分かる。

カルシウムイメージングが電気生理学的な手法より劣る点として、時間分解能が挙げられる。カルシウムイメージングのサンプリングレートは測定機器に限界があり [5]、通常 1-50[Hz] 程度である。ニューロンの発火は約 1[ms] で、シナプスを介して発火が伝わるのは 40[ms] 以内 [6] なので、ニューロンの発火を個別に観測することはできず、低いサンプリングレートでは発火の伝達も捉えることができない。また、カルシウムイオン濃度の変化は発火のタイムスケールより長い。発火後に蛍光強度が変化し始めるのは数 [ms] 後、蛍光強度がピークに達するのは数 100[ms] 後である。一方、蛍光強度が元の値に戻るまでには、数 100[ms] か

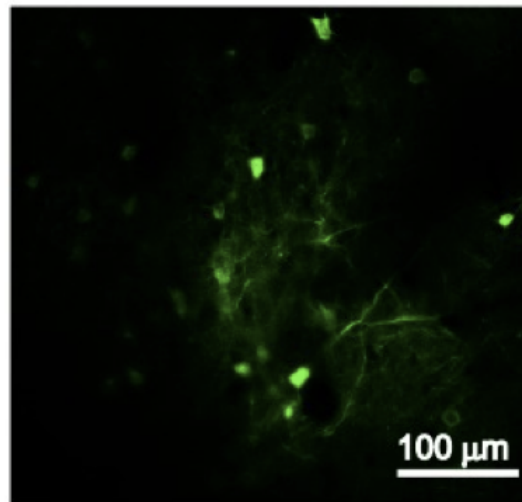


図 1.1: カルシウムイメージングで観察できる画像の例.

ら数 1000[ms] かかる [7]. また, 蛍光タンパク質の性能によっても時間遅れが生じる.

本研究で扱うデータのサンプリングレートは 8[Hz] で, シナプス伝達一つを見るには不十分である. 顕微鏡で観測すると活動電位が伝わる順番は分からなくなるため, カルシウムイメージングデータからはニューロンの活動の相関の情報しか得ることができない. これより, このデータではニューロン間のシナプス伝達を推定するのではなく, 機能的に同じニューロン, つまり同時に活動するニューロングループを推定する問題が適していると考えられる. 機能的に同じニューロンは, 3つの脳のコネクティビティのうち functional connectivity で繋がっているニューロンを指す.

1.2 関連研究

1.2.1 脳データの解析

カルシウムイメージングは時間解像度が低く空間解像度が高いデータが取得できる手法である. 同じ特徴である fMRI に対するデータの解析手法を紹介する.

fMRI データの解析は model based な手法と data-driven な手法がある [8]. Model based な手法は cross-correlation analysis (CCA) や coherence analysis (CA) などが挙げられ, 各指標に基づいて脳部位 (ROI) の機能的な結合 (前述の functional connectivity) を推定する. CCA では 2つの ROI の相関を使う. CA では, 2つの ROI の周波数ドメインでの相関を見る. fMRI の BOLD 信号の周期は 10[s] ほどなので, 0.1[Hz] 以下の周波数ドメインでの相関を見る. 結合の有無は各指標の閾値で決めなければならない. Data-driven な手法は更に行列分解とクラスタリングに分けることができる. 行列分解には principal component analysis (PCA) や singular value decomposition (SVD), independent component analysis (ICA) などが挙げられ, BOLD 信号を複数の成分に分解することができる. 同じ成分の重みが多い ROI は結合があるとする. ICA は成分数を事前に与える必要がある. クラスタリングは fuzzy clustering analysis (FCA) や hierarchical clustering analysis (HCA) などがある. クラスタリングではデータ間の距離を定義する必要がある. FCA はある ROI が複数のクラスタに存在できるクラスタリング手法で, クラスタ数を事前に決める必要がある. HCA はクラスタ数を決めずに ROI やクラスタが近いか遠いかを見ることができる.

本研究では, 行列分解手法の 1つである NMF を用いて隣接行列を推定する. 数理モデルを仮定すると, カルシウムイメージングデータには ICA や PCA より NMF の方がデータに

適している。

1.2.2 カルシウムイメージングの解析例

カルシウムイメージングデータの解析には二種類考えられる。まずは、カルシウムイメージングデータからスパイクを推定し、そのスパイク列を解析する方法である。Vogelstein らは逐次モンテカルロ法を用いてカルシウムイメージングデータからスパイク推定を行なった [9]。しかし、カルシウムイメージングデータでも低いサンプリングレートで計測されたデータではこの方法は使えない。

もう一つは、データから直接ニューロンの活動を解析する方法である。Mishchencko はベイズ推定を用いたニューロンの結合推定を行なった [10]。しかし、カルシウムイメージングのサンプリングレートが 30[Hz] 以上でないと意味のある結合推定結果は得られないと報告している。また、Stetter らは transfer entropy (TE) を用いて培養された興奮性ニューロンの結合推定を行なった [11]。この手法では、2つのニューロン間の TE を計算し、向きのあるエッジを推定している。また、ニューロンのネットワーク構造を仮定した人工のカルシウムイメージングデータを作成し、推定精度を議論している。しかし、本研究で扱うデータはサンプリングレートが低いため、向きの情報は取り出せない。Ikegaya らは蛍光強度データの一次微分を閾値で切ることによって発火のタイミングの情報を取り出し、テンプレートマッチングによってリアクチベーション現象を解析した [12]。この研究では 1[Hz] のデータからもリアクチベーション現象を確認できていた。しかし、本研究ではニューロン集団を推定したため、この方法はあまり適さない。

Molter らはカルシウムイメージングデータからニューロングループを抽出する方法を 8 つ人工データ実験と共に試した [13]。手法は大きく 2 つに分けられ、ニューロンペアの相関を見るものと、全てのニューロン活動の状態を見るものである。前者では固有値分解によって相関行列を作成した後、ICA や Promax rotation によってグルーピングを行う。後者では、ニューロンの活動から SVD, k-means クラスタリング、スペクトラルクラスタリングなどを用いてグルーピングを行う。後者の方法では、各グループの時間方向の活動を平均をとるなどして、グループの活動としていた。人工データは、ニューロンをポアソン分布にしたがって発火させ、発火からカルシウムイメージングの観測データに変換していた。同じグループに所属するニューロンは発火確率を同じ時間帯にあげることで表現していた。用いられていた手法では、ニューロンが複数グループに所属する場合やグループに所属しない場合も対応ができる。最も良いと結論づけられていた手法は ICA-CS と SGC という手法だったが、どちらの手法も閾値などのパラメータを決めなければならない。

Ghandor らは学習に関係するニューロン (engram cell) 群について NMF を用いて解析を行った [14]。実験は複数セッションに分けて行われており、それぞれのセッションで NMF を用いてニューロングループとその活動に分解していた。基底数は AICc で決めていた。セッションごとに推定されたグループが近いかを cosine similarity で計った結果、engram cell では non-engram cell よりも繰り返し活動するグループが多かった。NMF によってニューロングループがどれほど抽出できるかは論じられていなかった。

1.3 目的

カルシウムイメージングデータは上述のように多くのニューロンを観測できる利点がある一方、時間分解能が低いという欠点を持つ。本研究で扱うデータは低いサンプリングレートで観測されたデータである。ニューロンは複数で活動することで情報伝達を行っており、睡眠時に多数のニューロンが同時に活動する現象も確認されている。これより、本データでは同時に活動するニューロングループを推定して解析を行うことが適当である。

本研究の目的は、カルシウムイメージングデータから同時に活動するニューロングループを推定することである。紹介した関連研究は、本研究で扱いたいデータのサンプリングレ

トでは使えないものや、パラメータを適切に決めなければならないものだった。提案アプローチでは、数理モデルに基づいてNMFを用いてニューロンの隣接行列を推定し、スペクトラルクラスタリングを用いて隣接行列からクラスタリングを行う。隣接行列はバギングを用いて、2つのニューロンが同じグループになる確率値をつける。そのため、グループに所属しないニューロンの推定や、どれくらい結果を信用して良いのかがわかる。その際も閾値を決めなければいけないが、確率値という分かりやすい指標のため、既存手法より簡単だと思われる。NMF自体も基底数を事前に決めなければいけないが、大体の基底数を決めればその基底数周りの結果を平均することでパラメータによる結果の劇的な変化を抑えることができる。また、人工データ実験によって提案アプローチがカルシウムイメージングデータからニューロングループの情報をどの程度取り出せるかを確かめる。

本論文の構成は以下の通りである。第2章では本研究で使う手法の説明をする。第3章では数理モデルを元にしたクラスタリングのアプローチを説明する。第4章では人工データの作成方法と人工データ実験の結果、実データ解析の結果を述べる。第5章で本研究をまとめる。また、付録Aにカルシウムイメージングデータのために考案したNMFの制約を、付録BにNMFのモデルエビデンスの推定を試みた結果を載せる。

第2章 準備

本章では、提案アプローチで用いる NMF、バギング、スペクトラルクラスタリングについて説明する。

2.1 非負行列因子分解

非負行列因子分解 (nonnegative matrix factorization; NMF) [15] は行列分解の手法の一つである。NMF は非負行列 $X \in \mathbb{R}_+^{I \times J}$ を2つの非負行列の積として分解する：

$$X \approx DC, \quad 1 \leq K \leq \min\{I, J\}.$$

ただし、 $D \in \mathbb{R}_+^{I \times K}$ 、 $C \in \mathbb{R}_+^{K \times J}$ 、 $K < \text{rank}(X)$ である。このモデルを図 2.1 に示す。基底数 K は事前に決めなければならないパラメータである。

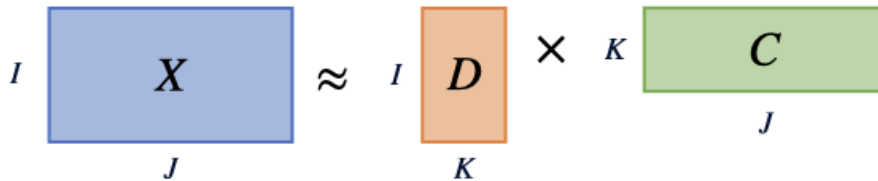


図 2.1: NMF のモデル。

NMF では目的関数を最小化することによって D と C を求める。目的関数はデータに対して置く仮定によって変わる。

2.1.1 NMF の基底数

NMF の基底数を決めるのは難しい問題である。NMF の基底数の決め方にはいくつかのアプローチがある。まずは、専門家や解析者の知識に基づいて決めることである。この方法はデータに対して十分な知識がない時には使えない。

次に BIC [16] や AIC [17] を用いる方法である。これらは漸近理論に基づいた近似を行った情報量基準である。NMF はデータが増えるほどパラメータ数 $(I + J) * K$ が増えるという特徴があり、これらを用いるのは本来不適切である。

次に R の NMF パッケージにも組み込まれている Brunet ら [18] の方法を紹介する。彼らは NMF の推定結果からノード同士が同じ基底に所属するかしないかを表す connectivity matrix $M \in \{0, 1\}^{I \times I}$ を作成する。本論文の推定量 A と同じ意味の行列である。初期値を 20-100 回変化させて M の平均 $\bar{M} \in [0, 1]^{I \times I}$ を計算する。彼らは真の基底数ではこの推定量が 0 か 1 に寄るようになると仮定して、最も \bar{M} が安定する基底数を求める。安定度は $1 - \bar{M}$ とその cophenetic correlation coefficient の Pearson correlation から計算する。

Ubaru ら [19] はブートストラップを用いて NMF を行っており、 D がそれほど変わらない基底数を採用している。推定した D^{*b} 同士の相互相関行列について dissimilarity [20] を測りその平均が最小となる基底数を用いる。

Hutchins らはテストデータに対する RSS (residual sum of squares) が真の基底数以降になるとあまり下がらなくなると論じている [21]。

Bayesian NMF ではギブスサンプリングなどを用いてモデルエビデンスを計算している [22].

上記で述べた方法の他にも様々な方法が考案されているが、全てのデータに当てはめられるような枠組みは存在しない。

2.2 バギング

アンサンブル学習の一つにバギング [23] がある。バギングはある予測器にブートストラップサンプルを入力して、その結果の平均を用いることで精度を向上させる方法である。

2.2.1 ブートストラップ法

ブートストラップ法 [24] とは、データによって推定量がどれだけばらつくかを近似する方法である。

例えば正規分布の平均値の推定を考える。用いる推定量は標本平均である。平均 μ の正規分布から 20 点サンプルし、そのデータから平均値を推定する。これを図 2.2 に示す。サンプルされたデータ点を軸上の黒線で示しており、平均の推定量を $\hat{\mu}$ とする。モンテカルロ法

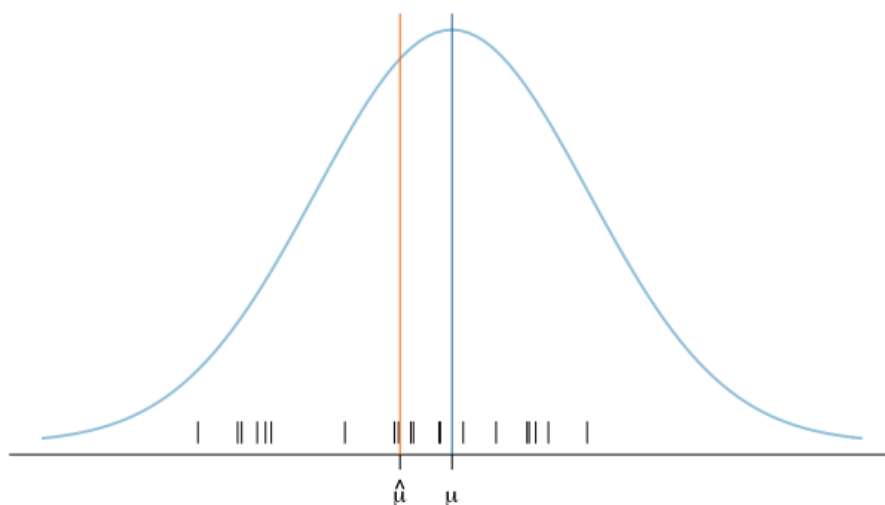


図 2.2: 正規分布の平均値の推定。平均 μ の正規分布（水色線）からサンプルされたデータ点を軸上に黒線で図示している。標本平均を $\hat{\mu}$ で示している。

で 20 点のデータを複数作り、それぞれで平均値を推定すると推定量 $\hat{\mu}$ の分布が得られる。推定量の分布を図 2.3 に示す。図 2.2 でサンプルされたデータ点から復元抽出を行い 20 点のデータを複数作成する。このサンプルをブートストラップサンプルという。ブートストラップサンプルそれぞれで平均値を推定すると推定量 $\hat{\mu}$ の分布が得られる。この時、真の平均は μ である。推定量 $\hat{\mu}$ の分布を図 2.4 に示す。モンテカルロ法から作られたサンプルによる推定値の分布も示しているが、分布の平均 μ とデータ点の平均 $\hat{\mu}$ の差分ずれていることがわかる。モンテカルロ法から作られたサンプルとブートストラップサンプルそれぞれの推定値の平均のヒストグラムを、それぞれの真の値を 0 としてずらしたものを図 2.5 に示す。真の値での推定値のばらつきはほとんど同じである。これよりブートストラップ法では、データの分布によってどれだけ推定量がばらつくかを、観測データから得ることができる。

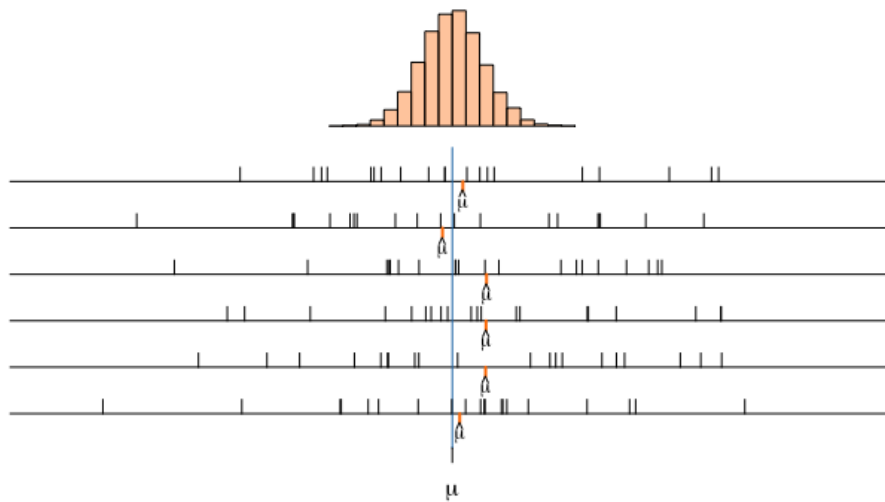


図 2.3: モンテカルロ法で 20 点のデータを複数作り, それぞれで平均の推定量 $\hat{\mu}$ を推定した図. 上部のヒストグラムは $\hat{\mu}$ の分布.

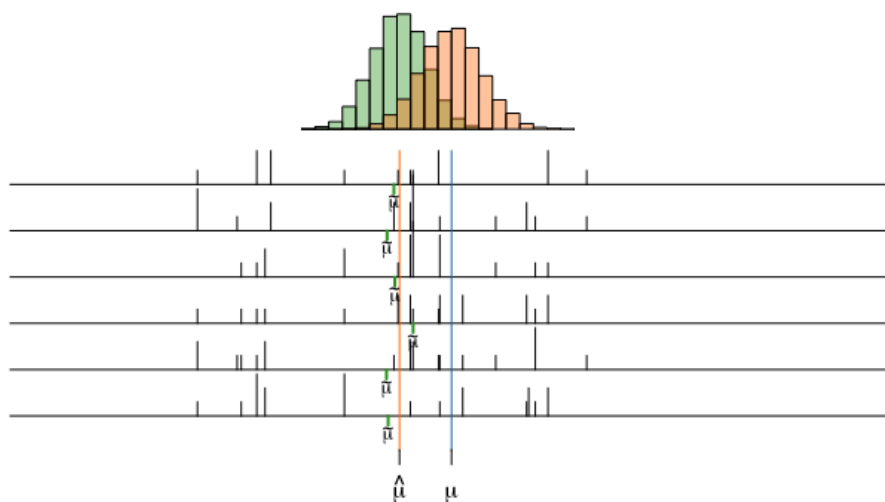


図 2.4: 図 2.2 のサンプルからブートストラップサンプルを複数作成し, それぞれで平均の推定量 $\tilde{\mu}$ を推定した図. 上部のヒストグラムは $\hat{\mu}$ の分布 (オレンジ) と $\tilde{\mu}$ の分布 (緑).

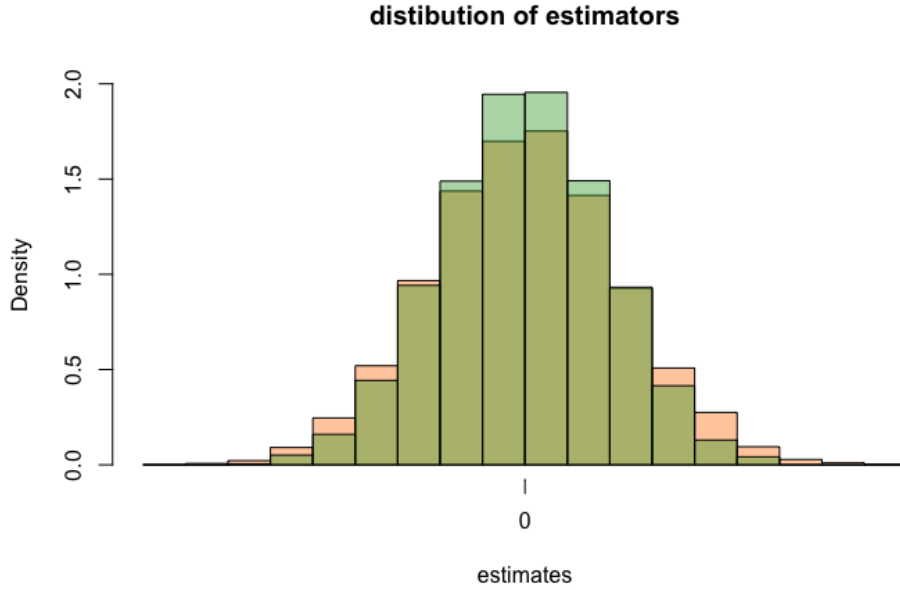


図 2.5: 正規分布の平均値の推定.

2.3 スペクトラルクラスタリング

スペクトラルクラスタリングはクラスタリング手法の一つで、類似度行列からクラスタリングを行う。類似度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 などでクラスタリングする。スペクトラルクラスタリングはグラフカットとして解釈することができる。

具体的な手順を説明する。類似度行列を $Q \in \mathbb{R}^{I \times I}$ とする。類似度行列 Q から重み付き無向グラフの隣接行列を $R \in \mathbb{R}_+^{I \times I}$ 作成する。隣接行列 R は Q をそのまま用いる場合や、 k 近傍や ε 近傍以外を 0 にすることで作成する。またそのグラフの次数行列を $O \in \mathbb{R}_+^{I \times I}$ とする。次数行列 O は対角行列でその対角成分は以下のように定義する：

$$o_{ii} = \sum_{j=1}^I r_{ij}.$$

グラフラプラシアンにはいくつか種類がある。まず、正規化されていないグラフラプラシアンは以下のように定義されている：

$$L = O - R.$$

次に、正規化されたグラフラプラシアンには 2 種類ある：

$$\begin{aligned} L_{sym} &= O^{-\frac{1}{2}} L O^{-\frac{1}{2}} = I - O^{-\frac{1}{2}} R O^{-\frac{1}{2}}, \\ L_{rw} &= O^{-1} L = I - O^{-1} R. \end{aligned}$$

いずれかのグラフラプラシアンを使って Algorithm 1 のアルゴリズムでクラスタリングを行う。Luxburg は正規化されたグラフラプラシアンの方が良く、 L_{sym} より L_{rw} の方が良いと述べている [25]。

クラスタ数の決め方には様々な方法がある。クラスタリング手法によらない方法には、安定性を見る方法 [26] や Gap 統計量 [27] がある。スペクトラルクラスタリングに特化した方法としては、固有値ギャップを見る方法 [25] がある。固有値ギャップは固有値を小さい順にプロットした際に、急激に固有値が大きくなる箇所であり、その順位をクラスタ数とする。

Algorithm 1 スペクトラルクラスタリングのアルゴリズム

Input: 類似度行列 Q , クラスタ数 k **Output:** クラスタリング結果 B_1, \dots, B_k

- 1: 隣接行列 R を Q そのままを用いるか, Q の k 近傍や ε 近傍以外を 0 にして作成
 - 2: グラフラプラシアン L を計算する
 - 3: グラフラプラシアン L の固有ベクトルを求め, 最初の k 個 $\mathbf{u}_1, \dots, \mathbf{u}_k$ を取り出す
 - 4: 行列 $U \in \mathbb{R}^{I \times k}$ を $\mathbf{u}_1, \dots, \mathbf{u}_k$ を列に持つ行列とする
 - 5: 行列 U の行を k-means でクラスタリングを行いクラスタを得る: B_1, \dots, B_k
-

第3章 NMFとバギングを用いたクラスタリング方法

本章では、ニューロン同士が同じグループである確率を推定し、その情報を元にクラスタリングを行う方法について説明する。

3.1 隣接行列の推定

3.1.1 数理モデル

カルシウムイメージングデータに対していくつかの仮定をおいた。

仮定 1

グループが K 個存在し、同じグループ内のニューロンは同じ時刻に活動する。ニューロンは複数のグループに所属することができる。観測時間内ではグループに属するニューロンは変化しない。

仮定 2

複数のグループが同時に活動する時、属するニューロンは被らない（ニューロンが属するグループは同時には活動しない）。

序論で述べたように、脳内にはセルアセンブリが存在すると言われているため、グループが K 個存在するという仮定は妥当である。また、シナプス結合は短期間で変化するものではないため、仮定 1 は妥当である。仮定 2 の概念図を図 3.1 に示す。仮定 2 は、主に線形モデルで扱うための仮定であり後述する。脳内のニューロンの働きを考えても、ニューロンはある時刻で 1 つの機能しか持たないと考えられるため、所属する複数グループが同時に活動することは考えにくい。逆に、同時に活動する複数グループは、同じ機能を持った 1 つのグループであると考えられる。

これらの仮説をもとに、数理モデルを構築する。観測データ X は $X \in \mathbb{R}_+^{I \times J}$ とする。ただし、 \mathbb{R}_+ を非負の実数の集合、 J を観測時系列の長さとする。ある行列 A の i 行を $a_{i:}$ 、 j 列を $a_{:j}$ 、 (i, j) 要素を a_{ij} または $[A]_{ij}$ と表記する。よって、ニューロン i の観測時系列は $x_{i:} \in \mathbb{R}_+^J$ である。

行ベクトル $c_{k:} \in \mathbb{R}_+^J$ ($k = 1, \dots, K$) をグループ k の活動の時系列とすると、仮定より $x_{i:}$ は $c_{k:}$ の重み付き和として表される：

$$x_{i:} = \sum_{k=1}^K d_{ik} c_{k:} + \eta_{i:}. \quad (3.1)$$

ただし、 $d_{ik} \in \mathbb{R}^+$ 、 $\eta_{i:} \in \mathbb{R}^J$ は i.i.d. なガウスノイズの時系列である。カルシウムイメージングのノイズはポアソン分布に従う光子ノイズであるが、光子数が多い場合はガウス分布で近似できる [28]。仮定 2 を置くことによって、式 (3.1) の線形モデルを考えることができる。仮定 2 が無い場合、あるニューロンの蛍光強度は複数グループからの影響によって上限なく上昇できてしまう。実際は、ニューロンの蛍光強度の最大値は蛍光タンパク質の量で決まるため、蛍光強度は最大値で飽和する。その場合式 (3.1) には飽和を組み込まなければいけな

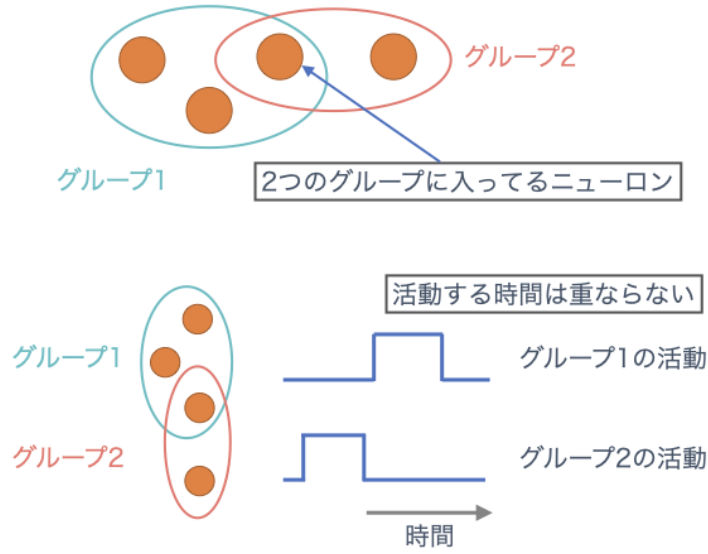


図 3.1: 仮定 2 の概念図. 所属するニューロンに被りがある 2 つのグループは活動時間が被らない.

い. 本論文では, 複数グループが同時に活動する場合はそれらも同じグループとみなし, 単純な線形モデルで表す.

式 (3.1) は行列形式で以下のように表現できる:

$$X = DC + H. \quad (3.2)$$

ただし, $D \in \mathbb{R}_+^{I \times K}$, $C \in \mathbb{R}_+^{K \times J}$, $H \in \mathbb{R}^{I \times J}$ 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 のモデルを図 3.2 に示す.

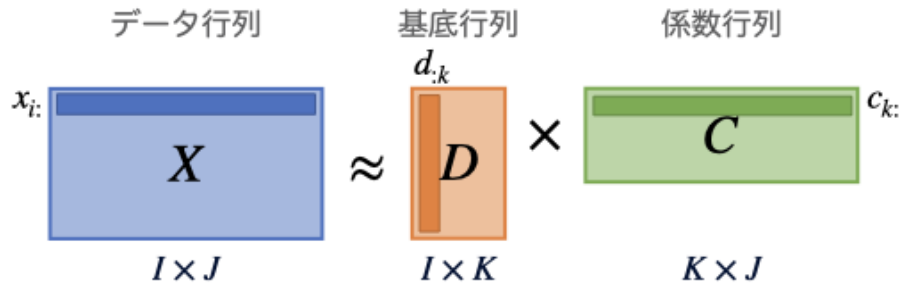


図 3.2: NMF のモデル.

ノイズ H は正規分布に従うと仮定しているため, 最小化する NMF の目的関数は以下である:

$$\arg \min_{D \geq 0, C \geq 0} \|X - DC\|_F^2. \quad (3.3)$$

ただし, 行列 $M \in \mathbb{R}_+^{I \times J}$ のフロベニウスノルム $\|M\|_F$ は以下のように定義される:

$$\|M\|_F = \sqrt{\sum_{ij} m_{ij}^2}.$$

3.1.2 NMF による \hat{A} の推定

X , D , C は上述の行列であり、式 (3.2) は NMF の問題として解くことができる。

NMF の寄与率行列 $P \in [0, 1]^{I \times K}$ の要素を次のように定義する：

$$p_{ik} = \frac{\|d_{ik}c_{k:}\|_1}{\sum_{l=1}^K \|d_{il}c_{l:}\|_1}.$$

また、概念図を図 3.3 に示す。ニューロン i の観測系列 $x_{i:}$ は $d_{ik}c_{k:}$ ($k = 1, \dots, K$) で再構成

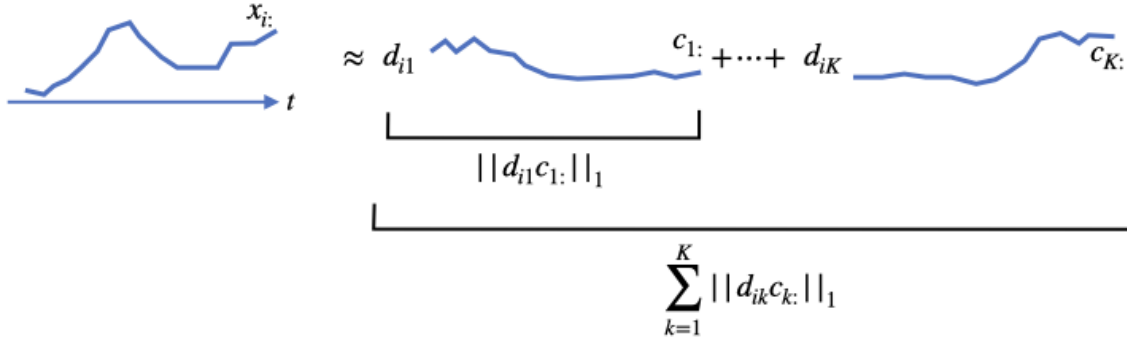


図 3.3: 寄与率の概念図。

されている。観測系列は非負なので、 $\|d_{ik}c_{k:}\|_1$ は基底 k の情報が $x_{i:}$ の再構成のうちどれくらいを占めているかを示す。よって、要素 p_{ik} は、ニューロン i の観測系列 $x_{i:}$ の再構成にどれくらい基底 k が寄与したかという量である。NMF の目的関数に C の行和が 1 という制約を入れると $P = D$ となる。

NMF での推定量 $\hat{A}(X, K) \in \{0, 1\}^{I \times I}$ とする。 $\hat{A}(X, K)$ は寄与率行列 P から作成する。まず、 P と同じサイズの行列 $G \in \{0, 1\}^{I \times K}$ を作る。 G は、 P の各行について最大値のみを 1、それ以外を 0 とした行列である：

$$g_{ij} = \begin{cases} 1 & (j = \arg \max_j p_{ij}) \\ 0 & (\text{otherwise}) \end{cases}.$$

推定量 $\hat{A}(X, K)$ を

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

と定義する。これより \hat{A} 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たら 1、していなかったら 0 となる。この推定量は cluster ensemble でも用いられている pairwise similarity[29] と似たものになっている。

この推定量を使う利点は 2 つある。1 つ目は、NMF では D の行と C の列は毎回同じ順番ではないため、その対応を考えなくて良い点である。2 つ目は、基底数の情報がなくなるため、異なる基底数の結果と比較できるところである。2 つ目の利点があるからこそ後述のモデル平均を行うことができる。

\hat{A} のみでもクラスタリングは行える。しかし、この $\{0, 1\}$ の行列には「本当は 0 だが偶然 1 になった要素」やその逆が存在する。全てのニューロンを 1 つの基底に割り振っているため、本来どのグループにも所属しないニューロンの判定もできない。そこで、 \hat{A} の要素に信頼度をつけることを次節で考える。

3.1.3 バギング

推定量 \hat{A} の要素の信頼度はバギングによって求める．提案アプローチではブートストラップサンプル $\{X^{*b}; b = 1, \dots, B\}$ から NMF を用いて $\hat{A}^{*b} = \hat{A}(X^{*b}, K)$ を推定する．これより \hat{A}^* の分布が求まるので， \hat{a}_{ij}^* が 1 となる確率が以下のようにして求まる：

$$Pr(\hat{a}_{ij}^* = 1) = E[\hat{a}_{ij}^* | X].$$

提案アプローチではこの確率値を使ってグループに所属しないニューロンを推定することができると思われる．全ての j について $Pr(\hat{a}_{ij}^* = 1)$ が閾値を下回れば，ニューロン i はどのニューロンとも同じグループになる確率が低いと考えられる．よって，ニューロン i はグループに所属しないと推定することができる．

ブートストラップサンプルの作成にはいくつかの方法が考えられる．簡単な 3 種類の方法について説明する．列のサンプリングによるブートストラップは，データ行列 X の各列がデータサンプルとしてみなせるので，列をサンプリングして X^{*b} を作る方法である．ブロックブートストラップでは，複数列を塊としてサンプリングする．この方法は時系列データでのブートストラップで用いられており，時間方向に制約の入った NMF などでは有効である．残差型ブートストラップは一回 NMF のモデルを推定し，推定後の残差をサンプリングして推定した $\hat{D}\hat{C}$ に足す方法である．

$$X^{*b} = \hat{D}\hat{C} + H^{*b}.$$

ただし， \hat{D} と \hat{C} は最尤推定量で， H^{*b} は $X - \hat{D}\hat{C}$ をリサンプリングした行列である．手順を図 3.4 に示す．ニューロンごとにノイズの大きさが異なることが想定される場合は，行ごとにリサンプリングを行うのが適切だと考えられる．NMF のモデル式 (3.2) では H は i.i.d. なノイズという仮定を置いているので，残差型ブートストラップが一番モデルに沿ったブートストラップ方法と言える．

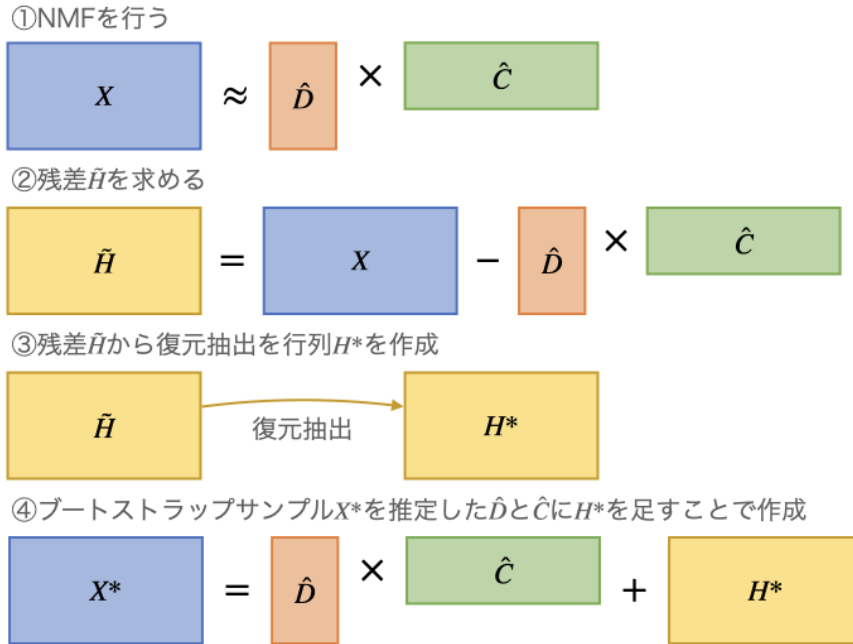


図 3.4: 残差型ブートストラップの手順.

3.1.4 モデル平均

NMF の基底数を決めるのは難しいので、複数の基底数の結果のモデル平均をとることで、基底数を間違えた時のリスクを軽減する。モデル平均とは異なるモデルの推定量の平均をとって精度向上を測る方法である。 \bar{A} は基底数によって行列サイズが変化しないので、異なる基底数の推定結果の平均をとることができる。その場合の推定量は：

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

$$\mathcal{K} = \{K_{min}, \dots, K_{max}\}.$$

ただし、平均する最小の基底数を K_{min} 、最大の基底数を K_{max} とする。

NMF の基底数を求めるのは難しい問題なので、モデル平均によって、異なる基底数の推定結果を平均して精度をあまり落とさないようにする。今回扱うデータから推定される \hat{A} は基底数が異なる時に大きく変化しないと考えられる。

アンサンブル学習では、それぞれのモデルにある程度の推定精度があり [30]、答えに多様性がある方が精度が上がる [31] と言われている。本論文での使い方は異なり、真の基底数が分からない時に推定を間違えるリスクを下げるという使い方である。

3.1.5 NMF の一意性

NMF の推定には一意性がなく、ある正則行列 Q を考えた時、

$$\begin{aligned} X &= DC \\ &= DQRC \\ &= D'C', \\ R &= Q^{-1}, \\ D' &= DQ, \\ C' &= RC, \end{aligned} \tag{3.4}$$

のように別の D' と C' が推定される可能性がある。

NMF に一意性がある条件は [32] などでもまとめられている。しかし、一意性を持たせるにはかなり条件が狭められる。

NMF と同じように非負行列を分解する手法に nonnegative rank factorization (NRF) [33] がある。NRF ではデータ行列 X を $X = DC$ に分解できる最小の基底数を nonnegative rank $\text{rank}_+(X)$ と定義し、 $\text{rank}(X) = \text{rank}(X)_+$ となる行列 X を対象とする。全ての非負行列は NMF できるが、NRF ができるとは限らない。NRF の計算は NP 困難であるが、[33] では NRF の計算と、NRF を持たない行列に対して MNRS という分解の計算方法を提示している。しかし、 X にノイズが乗っている場合に NRF が存在する条件は調査されていない。

式 (3.4) の場合の D' と C' を用いて作られる寄与率行列 P' と元の寄与率行列 P の関係を考える。簡単のため、 X に行和 1 の正規化を加え C に行和 1 の制約を加える。この時 D の行和も 1 となり、 $P = D$ となる。式 (3.4) より、 $P' = PQ$ となる。この時、 G の作り方には一意性がないため推定量 A にも一意性がない。

$K = 3$ の場合の図を用いて説明をする。図 3.5 に C の空間内の $x_{i\cdot}$ と D の空間内の $d_{i\cdot}$ を表す。 $d_{i\cdot}$ の和は 1 であるが、軸は図 3.6 のように動くことができる。 G を作成した際にニューロンがどの基底に所属するかを図 3.7 に示した。軸の変化によってニューロンが所属する基底が変化することがある。その結果、推定量 \hat{A} も変化するため、NMF は \hat{A} の推定に関して一意性はない。第 3 章の実験を通して一意性がなくても精度に向上は見られることを確かめる。

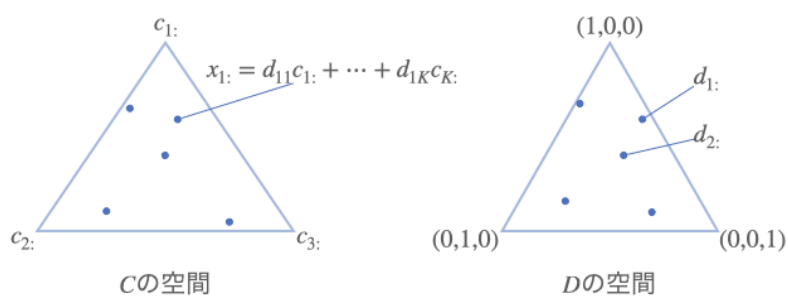


図 3.5: C の空間と D の空間.

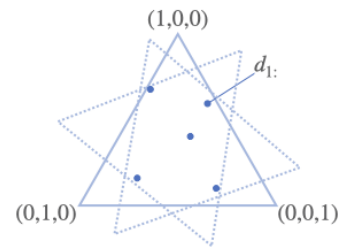


図 3.6: 変換 Q によって D の空間の軸は回転・移動する.

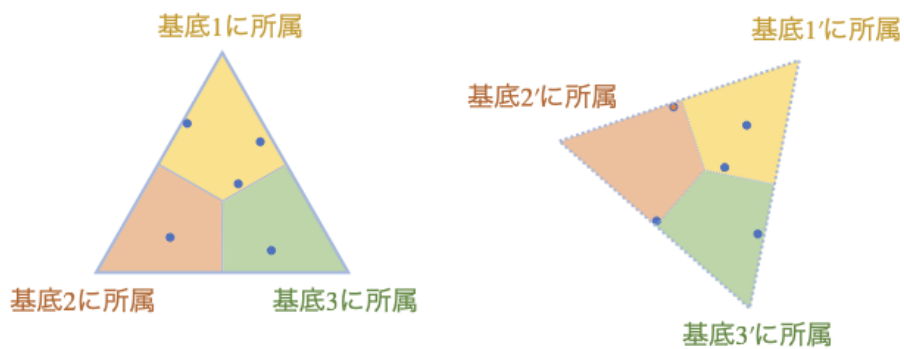


図 3.7: D の空間の軸の変化によってニューロンが所属する基底が変化する.

3.2 \bar{A} を用いたクラスタリング

\bar{A} 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．他のクラスタリング手法も使えるが， \bar{A} をグラフとして見るができるため，スペクトラルクラスタリングを用いる．ただし，Algorithm 1 の 1 段階目で類似度行列から隣接行列を求める際に， \bar{A} をそのまま使用する．

クラスタ数の決定には固有値ギャップを用いる．クラスタ数が明瞭な類似度行列について固有値ギャップと Gap 統計量を試した結果，固有値ギャップの方が安定してクラスタ数を決定できたためである．Gap 統計量の reference set は， \bar{A} の固有値分解を行い固有値ベクトルを一様分布から生成することで作成した．この場合 Gap 統計量を用いるには，固有値空間で一様であることを仮定しなければならないため，相性が悪かったと考えられる．固有値ギャップは目で見て判断しなければならないが，本実験では固有値の差分を取って大津の二値化にかけ，差分の大きいグループの最大固有値の順位を固有値ギャップとして計算する．

3.2.1 評価

クラスタリングの評価方法には [13] で用いられていた Best Match score を用いる．二つのクラスタリング結果 $\mathcal{B} = \{B_1, \dots, B_{|\mathcal{B}|}\}$ と $\mathcal{B}' = \{B'_1, \dots, B'_{|\mathcal{B}'|}\}$ があった場合を考える．ただし， B_k はクラスタ k に所属するニューロンの集合である．Best Match score は

$$\text{Best Match score} = \frac{1}{|\mathcal{B}| + |\mathcal{B}'|} \left(\sum_{B \in \mathcal{B}} \max_{B' \in \mathcal{B}'} \frac{|B \cap B'|}{|B \cup B'|} + \sum_{B' \in \mathcal{B}'} \max_{B \in \mathcal{B}} \frac{|B \cap B'|}{|B \cup B'|} \right),$$

である．これは Jaccard 係数のクラスタ数の平均である．2つのクラスタリング結果が近いほど Best Match score は高くなる．評価には \mathcal{B} をクラスタリング結果， \mathcal{B}' を真のクラスタ集合として計算する．

第4章 実験

本章では、人工データ実験を通して提案アプローチの性質調査と実データ解析の結果を述べる。

4.1 人工データ実験

4.1.1 シミュレーション

ニューロン集団のカルシウムイメージングデータをシミュレーションによって作り、解析手法を評価する。シミュレーションでは1) ニューロンのネットワーク構造を作成し、2) スパイクのシミュレーションを行い、3) 蛍光強度の観測データに変換する。

ネットワーク構造

シミュレーションに用いるニューロンの個数を N として、ニューロンのネットワーク構造を $S \in \{0, 1\}^{N \times N}$ とする。 s_{ij} はニューロン i からニューロン j へ活動電位が伝わるかを表している。本節では S の作り方を説明する。

ニューロンのネットワーク構造には small world network[34] を用いる。Small world network はノード数、張り替え確率、初期次数を決めることによってネットワークを作成するアルゴリズムである。初期次数は、ニューロンが平均何個のニューロンとシナプス結合を持つかという変数である。張り替え確率は、初期次数によって作成された規則的なグラフのエッジをランダムに張り替える確率である。そのため、エッジのうち何割が遠くのニューロンとつながっているかを表す変数である。張り替え確率と初期次数はニューロンのコネクション割合のデータに基づいて決定した。パラメータ値などは 4.3.1 節に記載している。

スパイクシミュレーション

スパイクのシミュレーションに Izhikevich モデル [35] を用いる。このモデルは Hodgkin-Huxley モデルをもとにしており、計算コストが低い。Izhikevich モデルでは、あるニューロンの膜電位が閾値を超えると発火したとみなし、あらかじめ定義したニューロンのネットワーク構造に従って結合を持つニューロンの膜電位を上昇させる。

このシミュレーションで設定しなければならないのは、個々のニューロンの特徴パラメータ、重み付きのネットワーク構造、外部からのランダムな入力である。ニューロンの特徴パラメータは [35] のものを用いた。

重み付きネットワーク構造を $W \in \mathbb{R}^{N \times N}$ とすると、ニューロン i から j へ結合があった場合（つまり $s_{ij} = 1$ の場合）、 w_{ij} はニューロン i が発火した時にニューロン j の膜電位をどれだけ上昇させるかという数値である。重み付きネットワーク W は、前節で作成した S の非ゼロ要素を乱数で置き換えることで作成する。重み付きネットワーク構造 W の作成方法は2種類ある：

1. 全ての重みを同じ分布から生成する
2. 同じグループへの興奮性ニューロンからの入力は強めにする

具体的な数値については 4.3.2 節に記載している．1 番目の方法で実験を行った際に，重みの大きい別のグループのニューロンからの影響が見られたため，2 番目の方法で作り直すことを考えた．

最後に外部からのランダムな入力について説明する．ニューロンには観測範囲外からの入力がある（以降，外部入力とする）．そのため，シミュレーション中も外部からの電位を乱数としてニューロンの電位に足す．本論文では同時に活動するニューロンを推定するのが目的の 1 つであり，ニューロンの活動も外部入力の大きさに表現する．ある時間帯にあるニューロングループが活動する時，そのニューロングループには平均値を上げた外部入力を足し，それ以外のニューロンには平均 0 の外部入力を足す．具体的な数値については 4.3.2 節に記載している．こうすることで，ニューロングループの活動のみ上がる（つまり蛍光強度が上がる）．

カルシウムイメージングモデル

スパイクデータからカルシウムイオン濃度を計算する [9] のモデルを用いる：

$$[\text{Ca}^{2+}]_{i,t} - [\text{Ca}^{2+}]_{i,t-1} = -\frac{\Delta}{\tau}([\text{Ca}^{2+}]_{i,t-1} - [\text{Ca}^{2+}]_b) + An_{i,t} + \sigma_c \sqrt{\Delta} \epsilon_{i,t}.$$

ただし， $[\text{Ca}^{2+}]_{i,t}$ をニューロン i の時刻 t でのカルシウムイオン濃度， $[\text{Ca}^{2+}]_b$ をカルシウムイオン濃度のベースライン， Δ を時間幅， τ は時定数， A は 1 つのスパイクでのカルシウムイオン濃度の上がり幅， $n_{i,t} \in \{0, 1\}$ はニューロン i の時刻 t でのスパイク， σ_c はノイズの分散， $\epsilon_{i,t}$ は標準正規分布に従う確率変数である．この人工データでは，蛍光強度がある値から上昇しない飽和状態は考えないこととする．

次に，同論文のモデルを使ってカルシウムイオン濃度 $[\text{Ca}^{2+}]_{i,t}$ をカルシウムイメージングで計測される蛍光強度 $F_{i,t}$ に変換する：

$$F_{i,t} = \alpha[\text{Ca}^{2+}]_{i,t} + \beta + \sigma_F \epsilon_{i,t}.$$

ただし， α は強度， β はバイアス， σ_F はノイズの分散である．表 4.10 に使用したパラメータを示す．何種類かの蛍光タンパク質の性能を調べた論文に [36] がある．この論文の，1 秒間に 10 回発火した時の decay time（蛍光強度が上がり切ってから半分の強度になるまでの時刻）から $\tau = 2.3$ とし，SN 比から $\sigma_c = 0.5$ とした．

パラメータ値	
$[\text{Ca}^{2+}]_b$	0.1
Δ	0.001
τ	2.3
A	5.0
σ_c	0.5
α	1.0
β	10
σ_F	1.0

表 4.1: カルシウムイメージングモデルでのパラメータ値

観測モデル

実データは 8[Hz] でサンプリングされたデータなので，シミュレーションした蛍光強度を 125[ms] ごとに足し合わせる：

$$x_{i,t'} = \sum_{t=1}^{125} F_{i,t}.$$

ここで、 t' はサンプリング後の時刻を表す。

上記の方法で作成した人工データ時系列と実データをそれぞれ図 4.1 と図 4.2 に示す。人工データと実データでノイズの乗り方が異なるように見えるが、その理由は分からなかった。

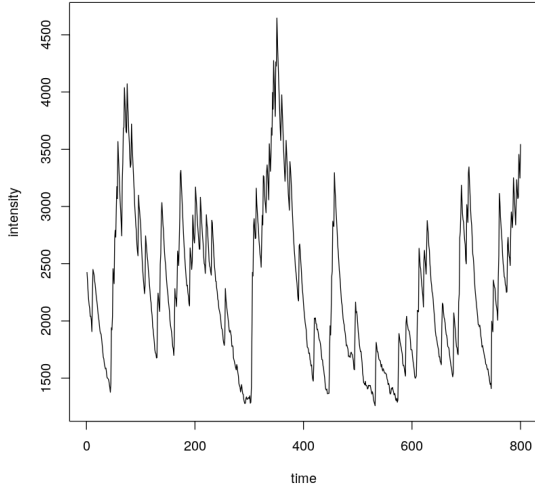


図 4.1: 人工データの時系列.

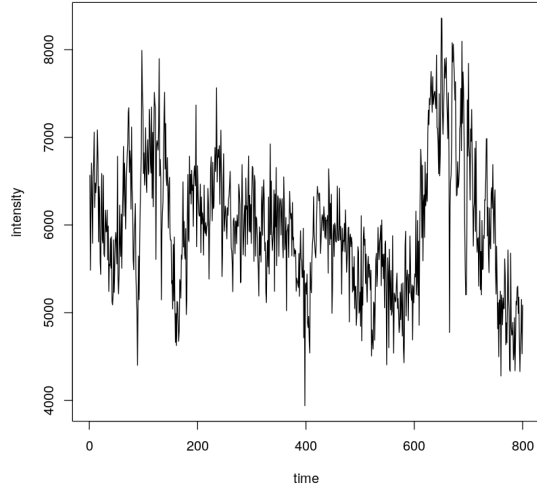


図 4.2: 実データの時系列.

実験設定

これ以降の実験での共通設定を説明する。800 個の興奮性ニューロンと 200 個の抑制性ニューロンについてネットワーク構造 S を作成し、 S は全実験を通して固定である。全 1000 個のニューロンのうち、解析に用いるのは固定された 100 個の興奮性ニューロンのみとする。1 つのグループに所属するニューロン数は 50~200 個とする。ニューロングループは同時に 2 つまで活動でき、シミュレーション時間内でどれだけ活動するかは実験によって異なる。例えば、100[s] のシミュレーションでグループが活動する時間を 5[s] とした時に、グループが活動する機会は 20 回ある。その 20 回の内グループが活動できない時間を設定している実験もある。シミュレーションの安定性から、記載のシミュレーション時間より 5[s] 長くシミュレーションを行い、最初の 5[s] は解析から除外している。

4.1.2 \bar{A} の推定に関する実験

本節では \bar{A} の推定に関する結果について述べる。人工データは 5[s] ごとに 1 つか 2 つのグループが活動するデータを 100[s] シミュレーションさせて作成した。人工データの実験設定は表 4.2 に示す。

隣接行列 \bar{A} は基底数 8 から 12 までをそれぞれ 30 回ブートストラップを行って NMF を行った結果を用いる。NMF の更新則は Nesterov 更新 [37] を用いる。NMF は初期値に依存するため、本節では「NMF1 回の結果」は「初期値を 20 回変えて NMF を行い目的関数が最小となった結果」とする。20 回は少ないかもしれないが (1000 回ほどこの操作を行っている論文もある)、実験サイクルを回すためにこの値にした。以上の NMF の設定を表 4.3 に示す。

シミュレーション時間	100[s]
グループの活動時間	5[s]
グループの活動可能回数	20
W の種類	1 番目
$N_{e_{plus}}$	0.8
$N_{i_{plus}}$	0.2
シミュレーション回数	100

表 4.2: 人工データの設定

1 回の結果を出す NMF の回数	20
基底数	8 - 12
ブートストラップサンプル数	30
ブートストラップ方法	列のサンプリング

表 4.3: NMF の設定

$\hat{A}(X, K)$ の一意性

前章で述べた通り, NMF には一意性がないため $\hat{A}(X, K)$ にも一意性がない. 具体的には, 1 つの人工データについて初期値のみを変化させて NMF を行うと推定される $\hat{A}(X, K)$ は同じではない. A の上三角の要素について 1 と推定された頻度を調べる. 収束性と前章で述べたように寄与率が D の空間で取りうる値の範囲を狭めるため, C の行和を 1 とする制約を入れている. X に正規化を加えなかった結果を図 4.3 に, X に行和 1 の正規化を加えた結果を図 4.4 に示す. X に行和 1 の正規化を加えると, D の自由度は下がるはずである.

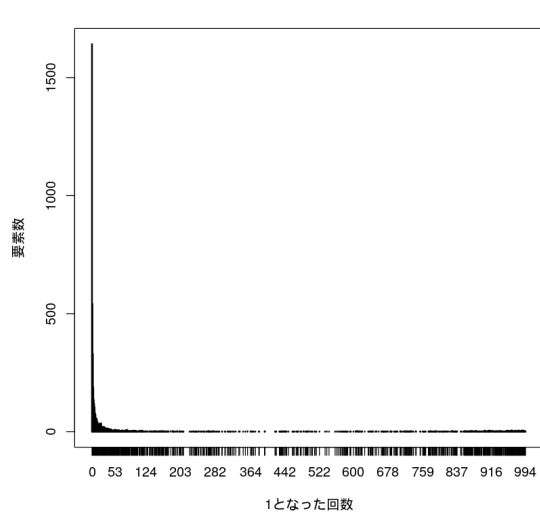


図 4.3: X を正規化せずに初期値を 1000 回変化させて NMF から $\hat{A}(X, K)$ を推定し, 各要素について 1 と推定された頻度.

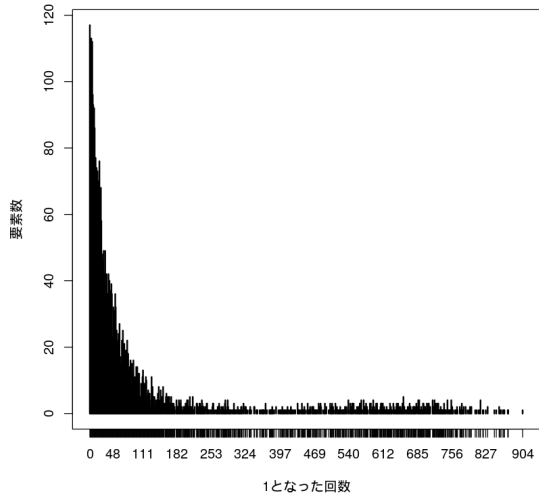


図 4.4: X を行和 1 で正規化して初期値を 1000 回変化させて NMF から $\hat{A}(X, K)$ を推定し, 各要素について 1 と推定された頻度.

$\hat{A}(X, K)$ に一意性がある場合は, ある $\hat{A}(X, K)$ の要素が 1 と推定される回数は 0 か 1000 になる. 結果より, そのようになっていないので $\hat{A}(X, K)$ に一意性がないのがわかる. また,

図 4.4 より X を正規化しない方がばらつきが小さい。これは、正規化をしない方が X の中の大きな値に推定が引っ張られ、同じ局所解に陥りやすくなっているからだと考えられる。

NMF を用いる妥当性

NMF を用いて隣接行列を求める性能を検証するため、NMF, PCA, ICA, logistic regression, glasso の性能の比較を行った。Logistic regression と glasso については筆者の卒業論文 [38] を参照されたい。どちらも時間窓をスライドさせてネットワークを推定する。今回の実験では時間窓を 40, スライド幅を 20 とした。Glasso のハイパーパラメータを $\rho = 0.3$ とした。一回でもエッジが張られたニューロン同士は同じグループとして推定量 \hat{A} と同じ行列を作成した。

PCA と ICA は NMF と同じ行列分解の手法 [39] である。PCA と ICA では D に相当する行列で NMF と同じように推定量 \hat{A} を作成する。3つの手法の基底数は 10 とした。

人工データ 100 個から、 $\hat{A}(X, K)$ を閾値 0.5 で切って $\{0, 1\}^{I \times I}$ の行列にした時の F1 score を比較した。図 4.5 より、NMF の精度が最も高いことが分かる。NMF の非負制約がデータ

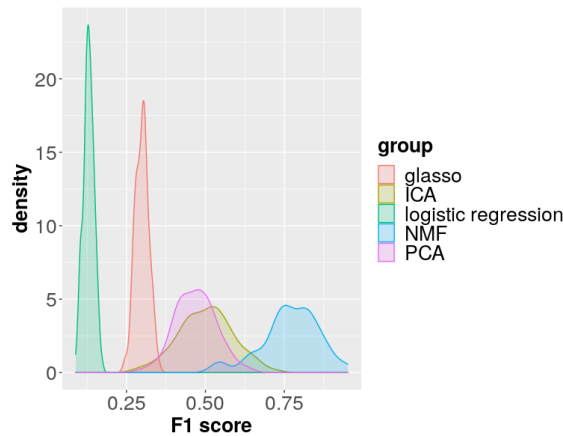


図 4.5: NMF, PCA, ICA, logistic regression, glasso の F1 score の密度分布

の生成モデルに合っているためだと思われる。Glasso と logistic regression については精度が低いですが、各窓ごとのニューロンネットワークの活動を反映している可能性があるため、この実験のみで有用性は判断できない。

バギングの効果

ブートストラップ法で $\bar{A}(X, K)$ を推定する有用性を確認するために、1回 NMF を行って \hat{A} を推定した結果、30回初期値を変えて \hat{A} を推定して平均を取った結果、30回ブートストラップを行って \bar{A} を推定した結果の F1 score を図 4.6 に示す。NMF の基底数は真の基底数 10 を用いた。図 4.6 より、ブートストラップを行った方が精度が高くなることがわかる。

モデル平均の精度

モデル平均を行うことでそれほど精度は落ちないことを検証するために、基底数別に推定された \bar{A} とモデル平均をとった \bar{A} の F1 score を算出した。その結果を図 4.7 に示す。これより、真の基底数周りの \bar{A} の平均をとることで精度は保たれることがわかる。

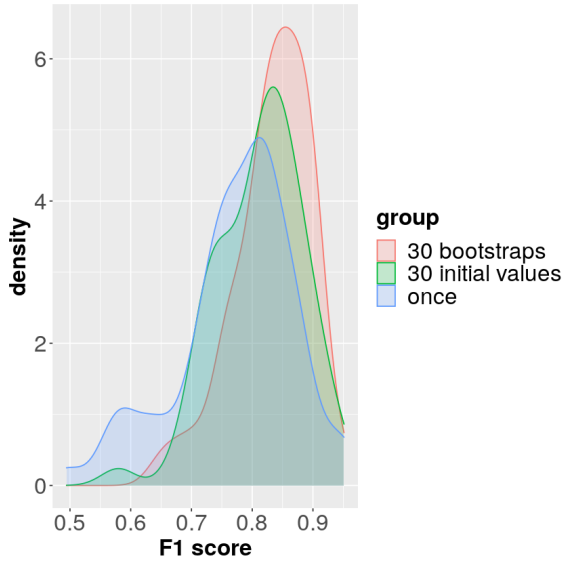


図 4.6: NMF を 1 回行った時の $\hat{A}(X, 10)$, 30 回初期値を変えた $\hat{A}(X, 10)$ の平均, 30 回ブートストラップを行った $\bar{A}(X, K)$; $K = 10$ それぞれの F1 score の分布.

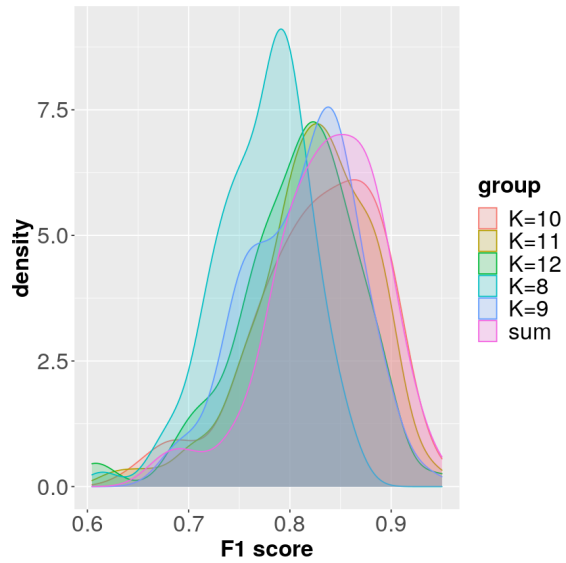


図 4.7: 基底数ごとにブートストラップを行った時の $\bar{A}(X, K)$ の F1 score と全ての基底数についてモデル平均をとった $\bar{A}(X, K)$ の F1 score の分布.

NMF の基底数

NMF の基底数を決める方法をいくつか試した．人工データ 86 個について Brunet らと Ubrau らの方法で基底数を決めた時に各基底数が何回選ばれるかを 図 4.8 と図 4.9 に示す．Brunet らの方法では初期値を変化させていたが，本実験ではブートストラップの結果で代用している．Brunet らの方法では真の基底数 10 に近い基底数が多く選ばれているが，小さい基底数も選ばれている．Ubaru らの方法では小さい基底数が選ばれる傾向にあった．

また，1 つの人工データについて AIC と AICc を計算した結果を 図 4.10 と図 4.11 に示す．AIC と AICc は最小となるモデルを選択する情報量基準だが，基底数が大きくなるごとに減少する傾向があった．他の人工データについても同様の傾向にあった．

4.1.3 \bar{A} のクラスタリングに関する実験

得られた類似度 \bar{A} についてスペクトラルクラスタリングを行い，精度を確かめる．比較手法には [13] で性能の良かった ICA-CS を用いる．ICA-CS のパラメータは論文の中で推奨されていたものを用いた．また，相関行列を k 近傍法 ($k = 20$) で隣接行列にしたものと ε 近傍法 (ε は相互相関行列の上から 20[%] の値) で隣接行列にしたものとも比較を行う．実データを扱う際には k や ε をどう設定すればいいかわからない．また，グループ内のニューロン数に大きな差がある場合は k 近傍法ではうまく推定できないことが考えられる．今回は比較のために k と ε の値を有利だと思われる値にした．

クラスタリング実験

人工データ実験の設定を表 4.4 に，NMF の設定を表 4.5 に示す．

作成した人工データの種類を表 4.6 に示す．900[s] のシミュレーションではニューロングループは 5[s] 活動する機会が 180 回与えられる．その全てでいずれかのグループを活動させる

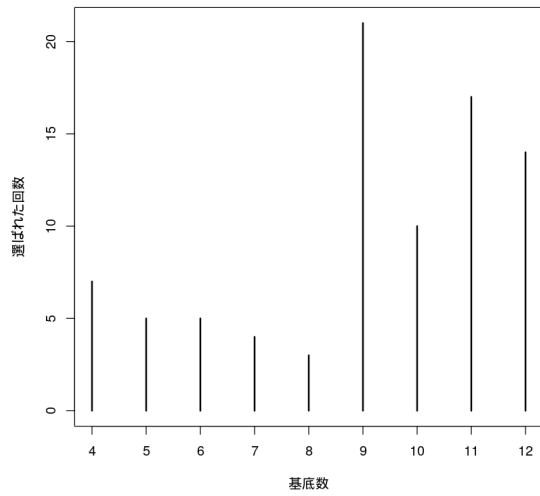


図 4.8: Brunet らの方法で基底数を決めた時に
各基底数が選ばれた回数 (真の基底数は
10).

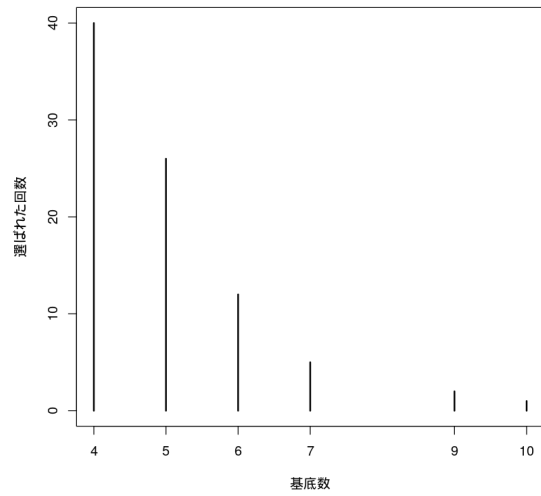


図 4.9: Ubaru らの方法で基底数を決めた時に
各基底数が選ばれた回数 (真の基底数は
10).

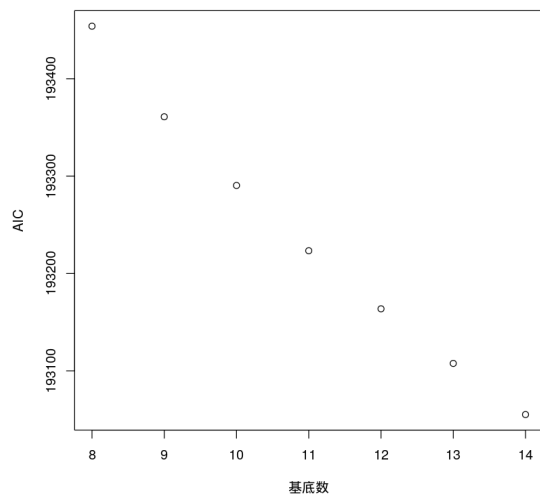


図 4.10: あるデータについて AIC を計算した
時の結果.

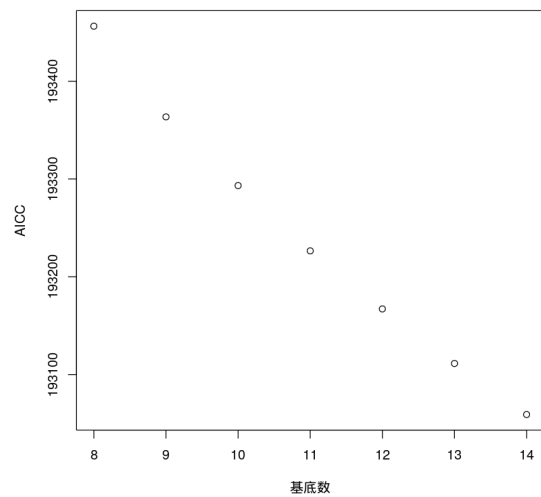


図 4.11: あるデータについて AICc[40] を計算
した時の結果.

シミュレーション時間	900[s]
グループの活動時間	5[s]
グループの活動可能回数	60, 30
W の種類	2 番目
Ne_{plus}	0.5
Ni_{plus}	0.2
シミュレーション回数	50

表 4.4: 人工データの設定

1 回の結果を出す NMF の回数	20
基底数	8 - 12
ブートストラップサンプル数	30
ブートストラップ方法	列のサンプリング

表 4.5: NMF の設定

わけではなく、活動できる時間帯をランダムに決める。人工データ non-active120 では活動可能回数を 60 回, non-active150 では活動可能回数を 30 回とした。1 回の活動の機会ではグループは 1 つか 2 つ活動できる。人工データは non-active120 をベースとして、グループに所属しないニューロンがある場合の提案アプローチの効果を確かめるために non-active120-0group を設計している。また、活動回数を減らした時の提案アプローチの効果を確認するために non-active150 を設計している。

人工データの種類	グループ数	活動回数 (180 回中)	特徴
non-active120	10	60	ニューロンは必ず 1 つのグループに所属する
non-active120-0group	9	60	ニューロンの約 10% は グループに所属しない
non-active150	10	30	ニューロンは必ず 1 つのグループに所属する

表 4.6: 人工データの種類

評価方法は 5 つ用いる。1 つ目は、 \bar{A} をそのままクラスタリングして BMS を計算する。残りの 4 つは、ニューロンを 2 種類の閾値で除去し、それぞれ 2 種類の方法で BMS を評価する。提案アプローチでは、ニューロン i が他の全てのニューロンと同じグループになる確率が閾値を下回った場合にニューロン i を取り除くが、その閾値を 0.5 と 0.6 として設定した。ニューロンが取り除かれた場合、 \bar{A} から該当するニューロンの行と列を削除してクラスタリングを行う。これらの結果の評価として、真のグループから除去したニューロンを取り除いた場合と取り除かない場合で BMS を計算する。

表 4.6 の 3 種類のデータについて、 \bar{A} と相関行列のスペクトラルクラスタリング、ICA-CS でクラスタリングを行った。真のクラスタとの Best Match score (以降 BMS) を算出した結果を図 4.12 に示す。相互相関行列に k 近傍法を用いた場合を correlation- k 、 ε 近傍法を用いた場合を correlation- ε としている。提案手法の評価方法は、 \bar{A} をそのままクラスタリングしたものを proposed-unremove、閾値 0.5 でニューロンを除去して真のグループから該当ニューロンを取り除いて評価したものを proposed0.5ch、真のグループのまま評価したものを proposed0.5 としている。閾値を 0.6 にした場合も同じ表記である。ICA-CS は 3 種類の人工データ全てにおいて BMS が低い。これは、パラメータが適切ではなく多くのニューロ

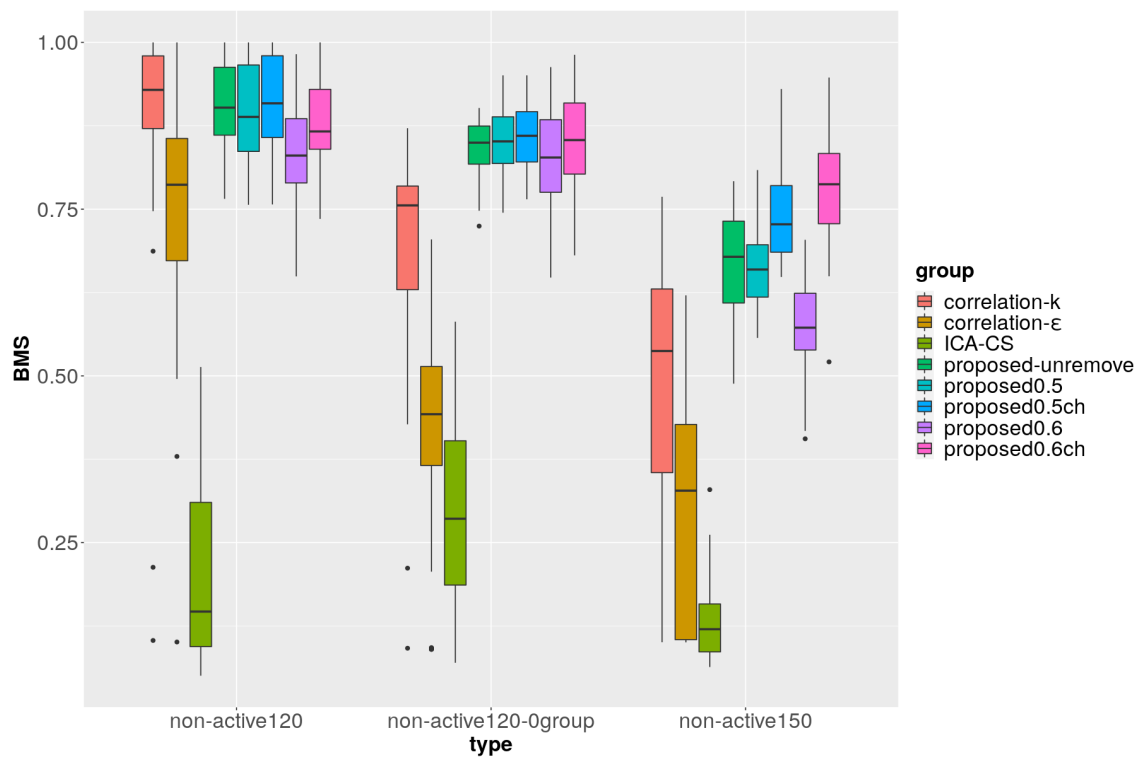


図 4.12: 3 種類的人工データについて提案アプローチ, 相関行列, ICA-CS でクラスタリングをした結果.

ンがクラスタに属していないからだと思われる．相互相関行列に ε 近傍法を用いた場合は k 近傍法を用いた場合より BMS が低くなっている．相互相関行列を閾値で切るとニューロンごとにエッジの数が多いものと少ないものができてしまい，それが影響を与えていると考えられる．non-active120 では，相互相関行列に k 近傍法を用いた場合の方が提案アプローチより BMS が高くなっている．しかし，non-active120-0group では提案アプローチより相互相関行列の BMS の下り幅が大きくなっている．これは，ランダムに活動するニューロンの存在が相互相関行列によるクラスタリングに悪影響を与えているためだと思われる．グループが活動する時間帯を減らした non-active150 でも相互相関行列の BMS の下り幅が提案アプローチよりも大きい．ニューロンがグループに依らずランダムに活動する時間が多いことが相互相関行列によるクラスタリングに悪影響を与えていると思われる．これより，ランダムに活動するニューロン数や全ニューロンがランダムに活動する時間に対して，提案アプローチの方が相互相関行列よりもロバストだと考えられる．

人工データ non-active120-0group で閾値 0.5 と 0.6 で取り除いたニューロンの precision と recall をそれぞれ図 4.13 と図 4.14 に示す．プロットの大きさは，その座標にある点の数を表している．プロットの色は，そのデータ内で取り除かれたニューロン数を示している．閾値 0.5 の方が取り除いている数は少ないが precision は高い．また，図 4.12 でも proposed0.5 の方が proposed-unremove より BMS が高い．それに対して proposed0.6 の BMS は proposed-unremove より下がっている．よって，閾値は 0.6 より 0.5 を採用すべきである．

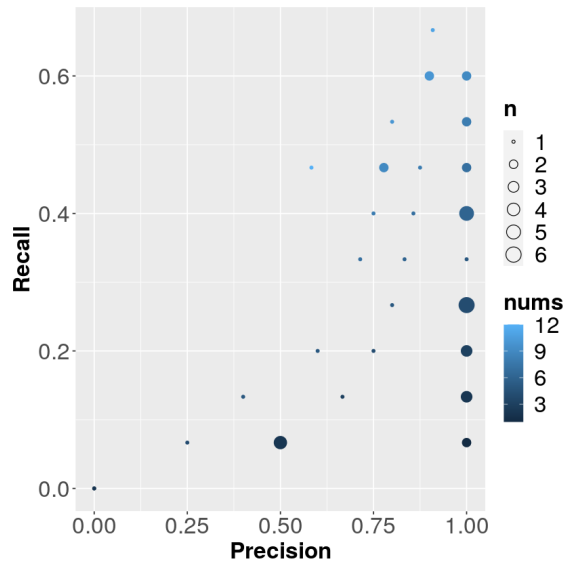


図 4.13: non-active120-0group で閾値 0.5 でニューロンを切った場合の precision と recall.

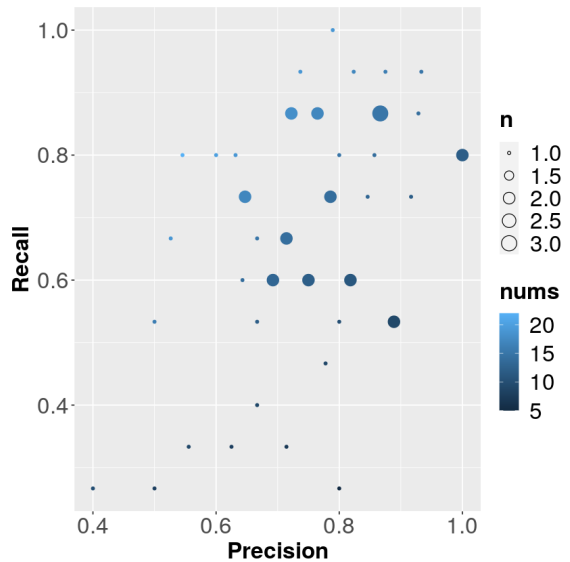


図 4.14: non-active120-0group で閾値 0.6 でニューロンを切った場合の precision と recall.

固有値ギャップと大津の二値化によって推定されたクラスタ数を比較する．3 種類の人工データそれぞれについて相互相関行列で推定されたクラスタ数と提案アプローチで推定されたクラスタ数を図 4.15～4.17 に示す．提案アプローチの方が安定して真のクラスタ数を当てられていることがわかる．図 4.12 で correlation-k の BMS の低さやばらつきの大きさは，推定されたクラスタ数によるばらつきも関係していると思われる．

グループの活動時間数に対する提案アプローチのロバスト性

グループの活動時間数に対して提案アプローチの方が相互相関行列よりもロバストかを確認するための実験を行った．人工データ実験の設定を表 4.7 に示す．グループは 100 回活動

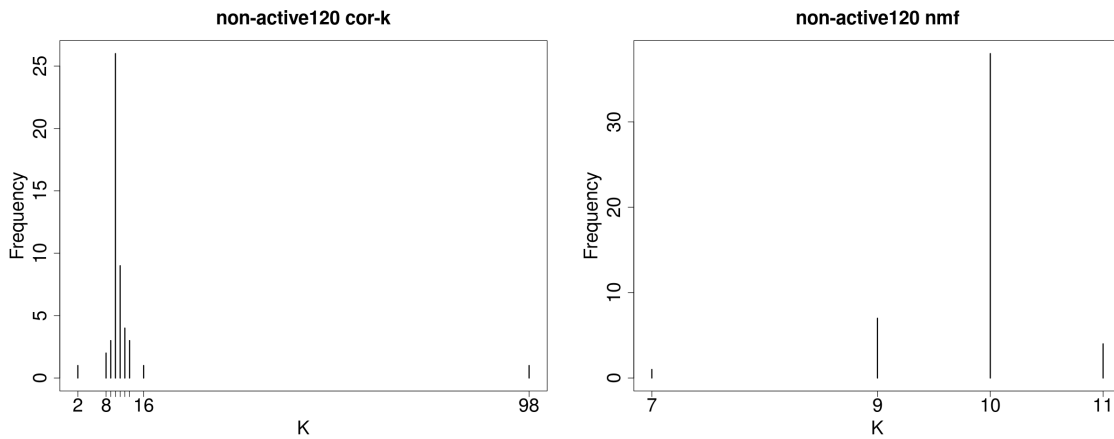


図 4.15: non-active120 で推定されたクラスタ数. 左は相互相関行列に k 近傍法を用いた場合. 右は提案アプローチでニューロン除去しなかった場合. 真のクラスタ数は 10.

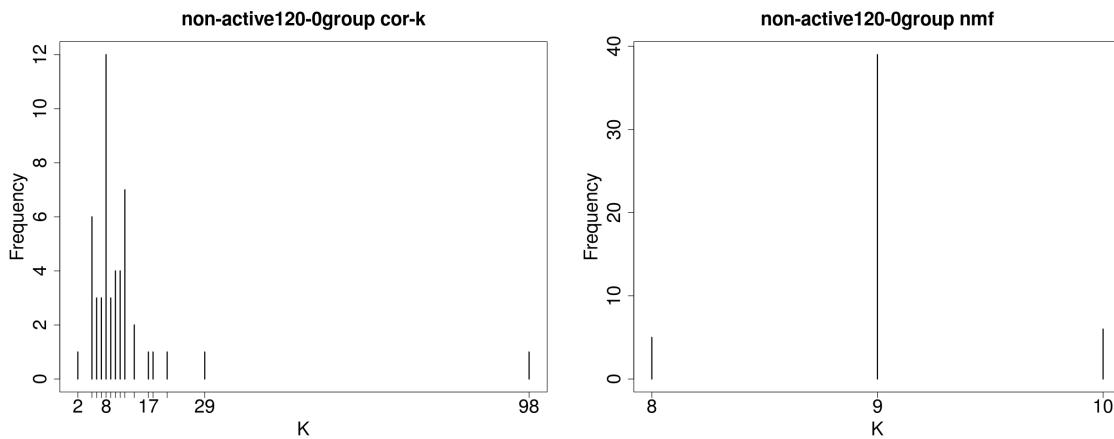


図 4.16: non-active120-0group で推定されたクラスタ数. 左は相互相関行列に k 近傍法を用いた場合. 右は提案アプローチで閾値 0.5 でニューロン除去した場合. 真のクラスタ数は 9.

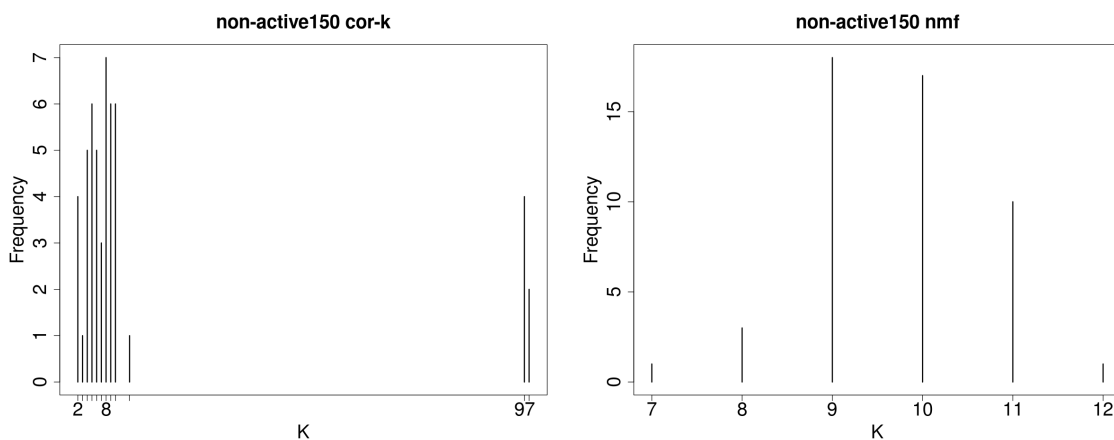


図 4.17: non-active150 で推定されたクラスタ数. 左は相互相関行列に k 近傍法を用いた場合. 右は提案アプローチでニューロン除去しなかった場合. 真のクラスタ数は 10.

する機会があるが、その中で活動時間帯を 70, 60, 50, 40 回と減らした 4 種類の人工データを用いた。

シミュレーション時間	100[s]
グループの活動時間	1[s]
グループの活動可能回数	40, 50, 60, 70
W の種類	2 番目
N_{eplus}	0.8
N_{iplus}	0.4
シミュレーション回数	50

表 4.7: 人工データの設定

提案アプローチと相互相関行列のクラスタリング結果を図 4.18 に示す。前節の実験よりもシミュレーション時間とグループの活動時間を短くしたため、全体的に BMS が低くなっている。全ての人工データにおいて提案アプローチの BMS の分散の方が小さい。グループの活動回数が多い non-active30 では相互相関行列の BMS の中央値の方が高いが、グループの活動回数が少ない non-active60 では提案アプローチの BMS の中央値の方が高くなっている。これより、活動回数が少ない方が提案アプローチの方がロバストだと考えられ、図 4.12 でも同じことが起こっていると考えられる。

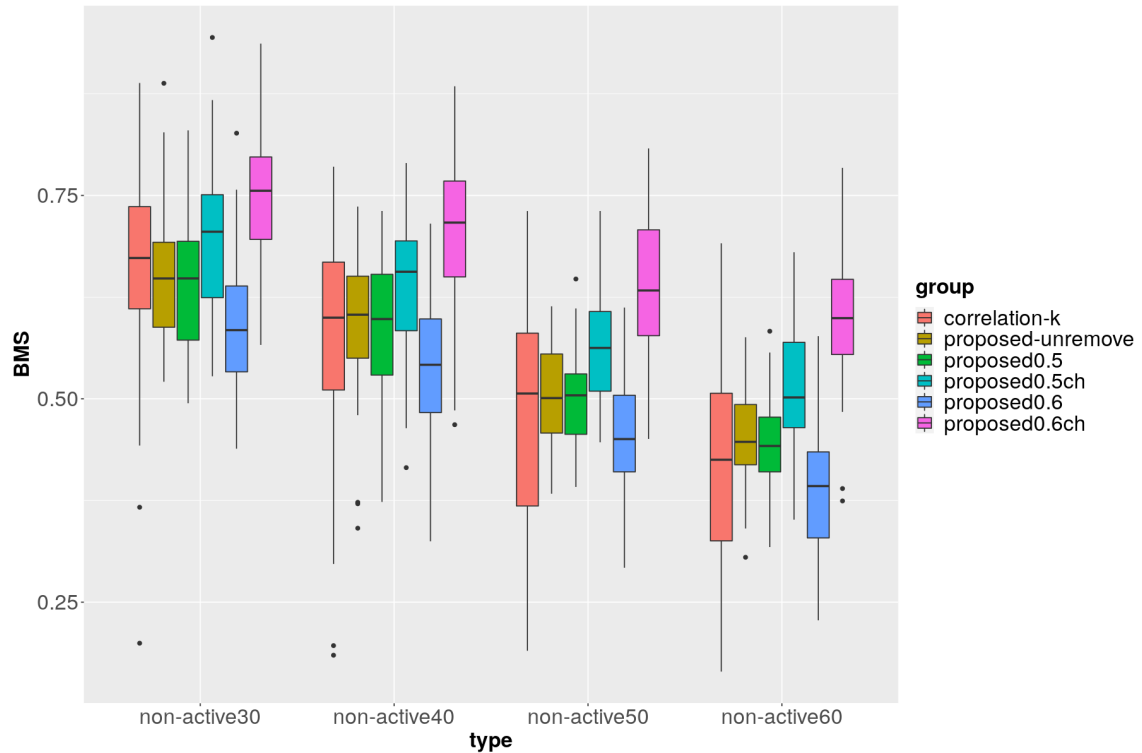


図 4.18: 4 種類の人工データについて提案アプローチと相関行列でクラスタリングをした結果。

モデル平均の効果

モデル平均の効果を確認するため、表 4.6 の non-active120 の人工データについて基底数が 1 つの場合と 5 つモデル平均した場合を比較した。基底数が 1 つの場合は $K = \{8, \dots, \{22\}$ として \bar{A} を求め、スペクトラルクラスタリングを行った。その時の BMS を図 4.19 に示す。基底数が 5 つの場合は $K = \{8, \dots, 12\}, \dots, \{18, \dots, 22\}$ とした。その時の BMS を図 4.20 に示す。また、それぞれの推定クラスタ数を図 4.21 と図 4.22 に示す。図 4.19 の基底数 8 の BMS より図 4.20 の基底数 8 から 12 までもモデル平均した時の BMS の方が高いことがわかる。図 4.19 の基底数 22 の場合も同様に図 4.20 の基底数 18 から 22 のモデル平均の方が BMS が高い。しかし、図 4.19 の基底数 18 の BMS より良くなっているとは言えない。あくまでもモデル平均は、真から最も遠い基底数の BMS より良くなるだけであり、リスク回避の役割が大きい。推定クラスタ数についても同様のことがいえる。

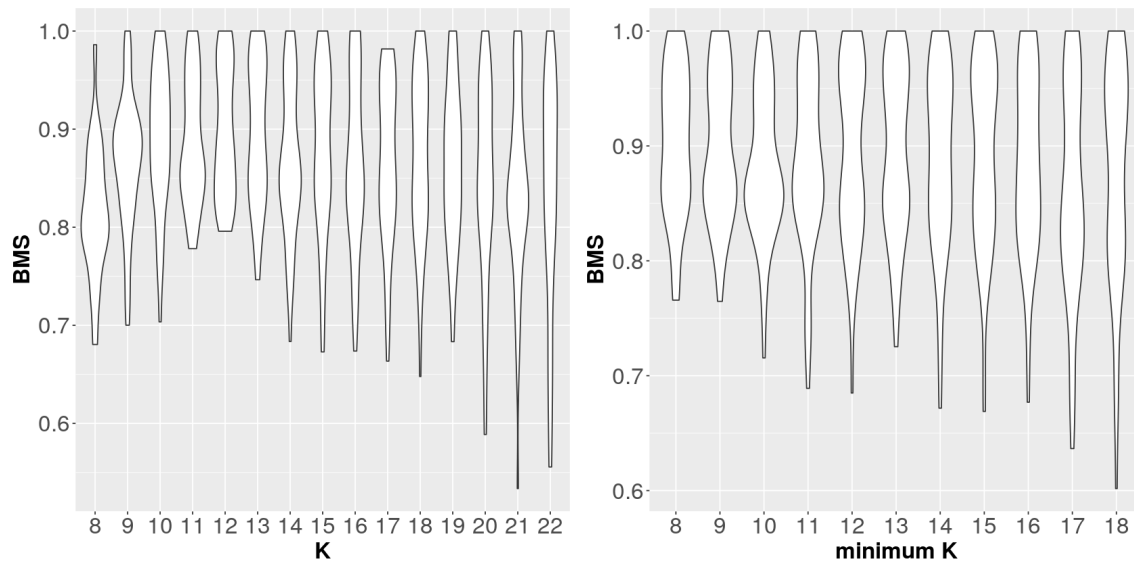


図 4.19: 基底数 1 つの場合の BMS の分布。 図 4.20: 基底数 5 つの場合の BMS の分布。

4.2 実データ解析

本節では実データを解析する。

4.2.1 実データ

本研究で用いるデータは筑波大学の柳沢研究所で計測されたカルシウムイメージングデータである。本データは、2 光子多細胞カルシウムイメージングによって 1 匹のマウスの大脳皮質 1 次運動野第 2・3 層のニューロンのイメージング画像を得た後、人手で ROI がつけられ、ニューロンごとの数値データに直されたものである。通常の睡眠状態と睡眠を阻害した状態の計測が別日に行われた。ただし、観測ニューロンは同じである。1 時間おきに 15 分間のイメージングがそれぞれ 6 回と 5 回行われた。イメージングのサンプリングレートは 8[Hz] である。用いられた蛍光タンパク質は GCaMP6s である。実験系は図 4.23(A) の系で行われた [41]。観測されたニューロン数は 154 であった。時間方向には 4[s] ごとのマウスの状態 (wake, REM, NREM) のラベルが付いており、ニューロンごとに興奮性か抑制性のラベルがついている。通常睡眠データと阻害睡眠データの時系列方向のラベルをそれぞれ図 4.24 と図 4.25 に示す。

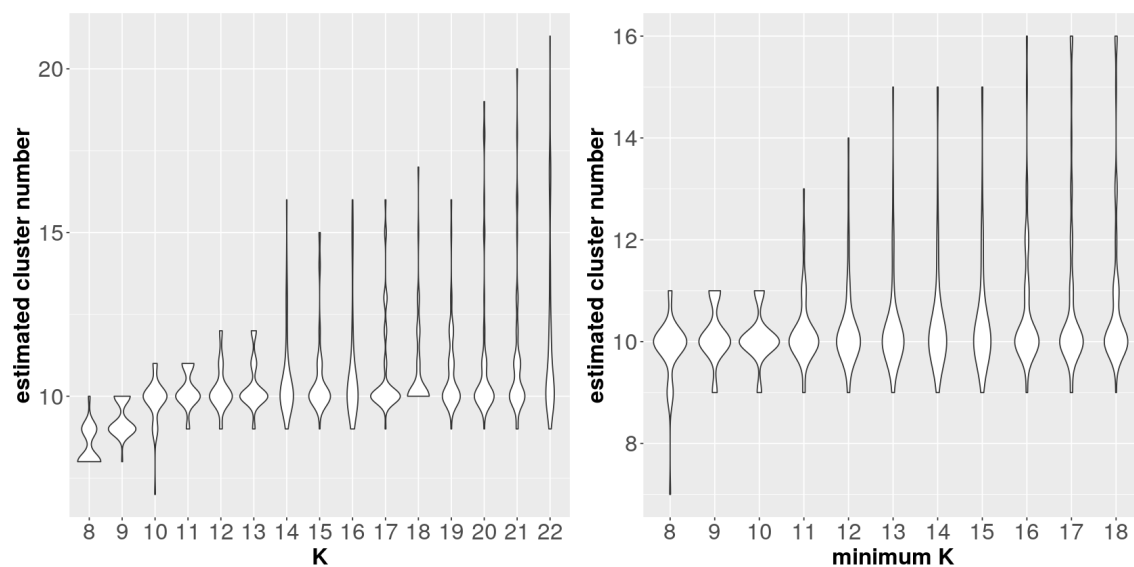


図 4.21: 基底数 1 つの場合の推定クラスタ数の分布. 図 4.22: 基底数 5 つの場合の推定クラスタ数の分布.

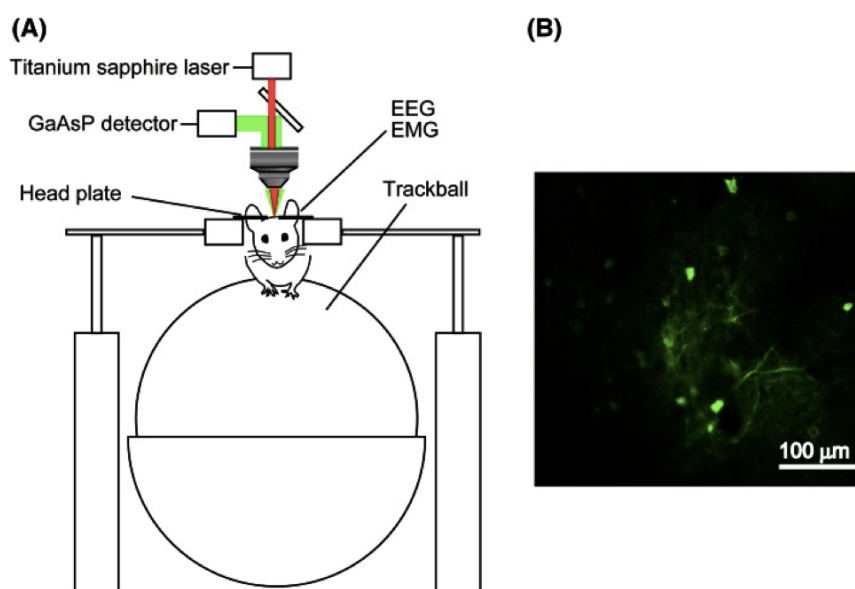


図 4.23: (A) カルシウムイメージングの測定系. マウスは頭を固定されたままボールの上で活動することができる. (B) カルシウムイメージング画像.

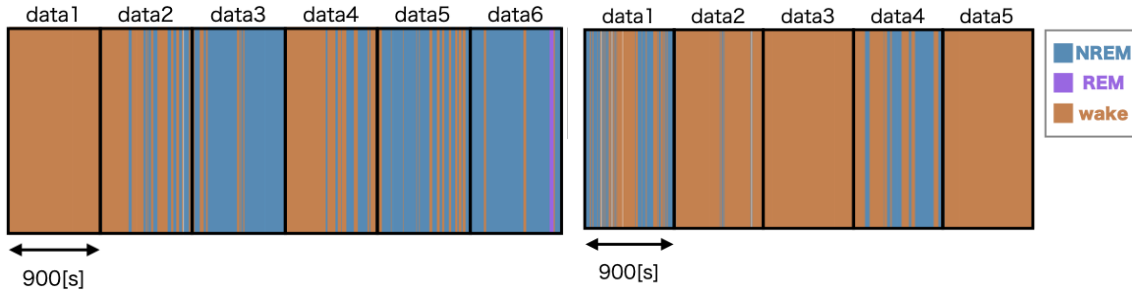


図 4.24: 通常睡眠データのラベル

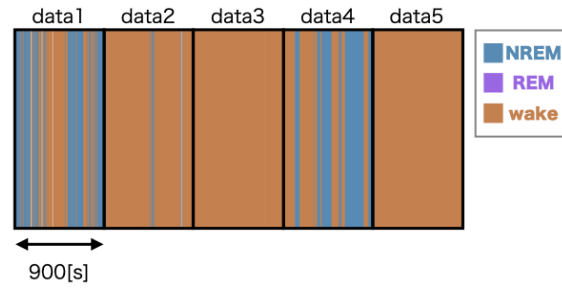


図 4.25: 阻害睡眠データのラベル

4.2.2 提案アプローチのクラスタリング結果

15 分のデータごとに提案アプローチを用いてクラスタリングを行う。NMF を一回行い、それを最尤推定の結果とする。最尤推定結果から BIC を計算し、目視で基底数 4 つを決める。NMF の設定を表 4.8 に示す。 \bar{A} の閾値を 0.5 としてグループに所属しないニューロンを除外する。クラスタ数は人工データ実験の時と同様に固有値ギャップからプログラムを使って求めるが、あまりに大きすぎるクラスタ数の場合（例えば 152 など）は目視で決定した。

1 回の結果を出す NMF の回数	100
基底数	BIC 最大周りの 4 つの数
ブートストラップサンプル数	100
ブートストラップ方法	残差型ブートストラップ

表 4.8: NMF の設定

推定された通常睡眠データのクラスタ内ニューロン数を図 4.26 に、阻害睡眠データのクラスタ内ニューロン数を図 4.26 に示す。図 4.24 と図 4.25 と見比べると、睡眠時に大規模クラスタが存在していることがわかる。

推定されたクラスタにデータ間で被りがあるかを見るため、各データペアでクラスタ間の Jaccard 係数を計算した。通常睡眠データ内で比較した結果を図 4.28 に示す。例えばデータ 1 で 11 個のクラスタが推定され、データ 2 で 16 個のクラスタが推定されたが、図 4.26 の “n1-n2” はデータ 1 の各クラスタとデータ 2 の各クラスタの Jaccard 係数を計算し、データ 1 のクラスタそれぞれで最大となった係数をプロットしている。同様に、阻害睡眠データ内で比較した結果を図 4.29 に、通常睡眠データと阻害睡眠データで比較した結果を図 4.30 に示す。この結果から、“n3-n6”、“n6-n3”、“n3-d4”、“n6-d4” で Jaccard 係数が大きくなっていることがわかる。これらのクラスタは、図 4.26 と図 4.27 のクラスタ内ニューロン数が多いクラスタであった。また、この 3 つのデータ（通常睡眠のデータ 3 とデータ 6、阻害睡眠のデータ 4）は NREM 睡眠が半分ほどを占めているデータである。これより、この 3 つのクラスタ全てに所属するニューロンが NREM 睡眠中に同時活動している可能性がある。他にも Jaccard 係数が 0.4 近いクラスタペアもあるが、クラスタ内ニューロン数が少なかったため今回は議論しない。

4.2.3 相互相関行列との比較

前節で提案アプローチによって捉えた睡眠中の大規模クラスタを、相互相関行列から抽出できるかを検証する。15 分間のデータそれぞれについて相互相関行列を計算し、相互相関行列をそのまま用いた場合と k 近傍法で k の数を変化させた場合と ϵ 近傍法で ϵ を変化させた場合にスペクトラルクラスタリングを行う。 k の値は 10, 20, 30, 40, 50 として、閾値 ϵ

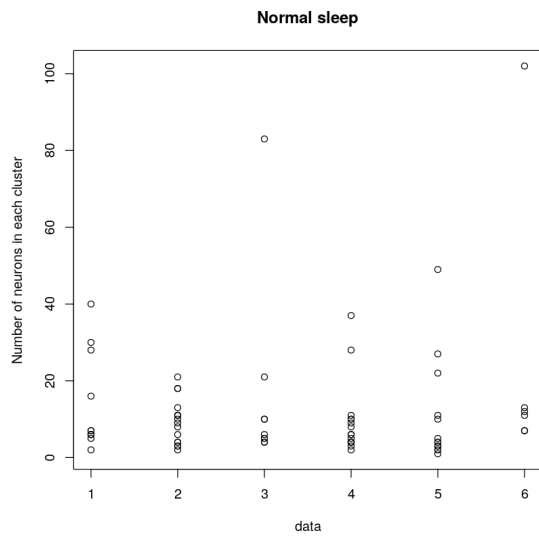


図 4.26: 通常睡眠データから推定されたクラスタに含まれるニューロン数.

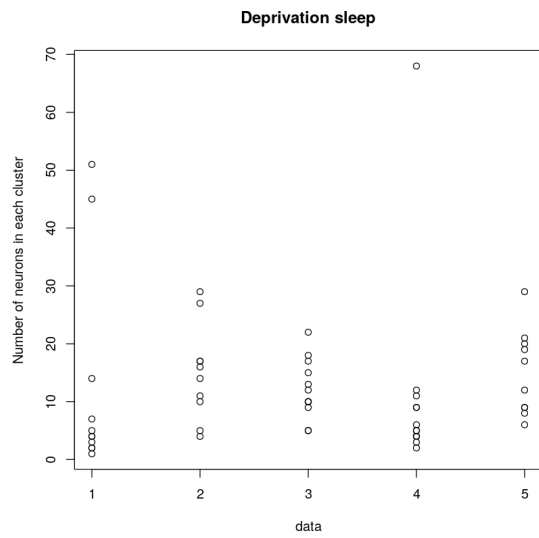


図 4.27: 阻害睡眠データから推定されたクラスタに含まれるニューロン数.

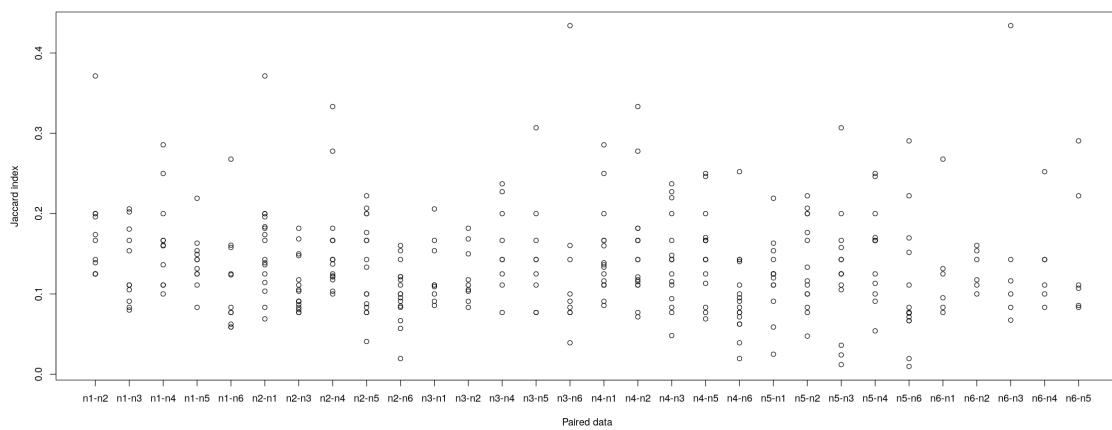


図 4.28: 通常睡眠データペアで各クラスタの Jaccard 係数を計算した図. 横軸のラベルはデータペアを表し, “n1-n2” はデータ 1 とデータ 2 のクラスタ間で Jaccard 係数を計算した.

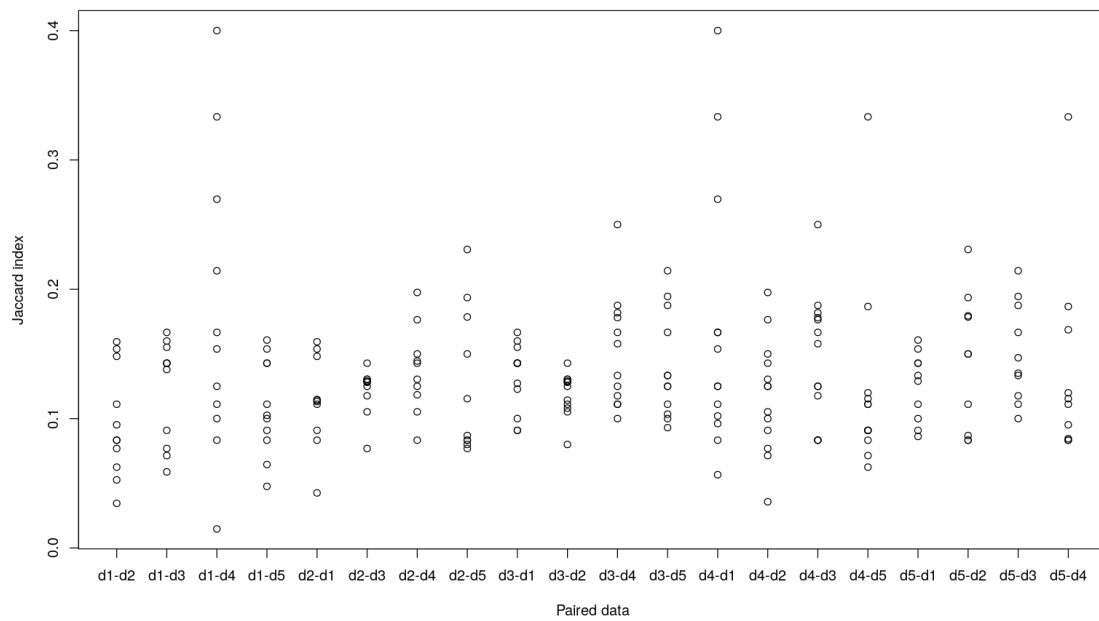


図 4.29: 阻害睡眠データペアで各クラスタの Jaccard 係数を計算した図.

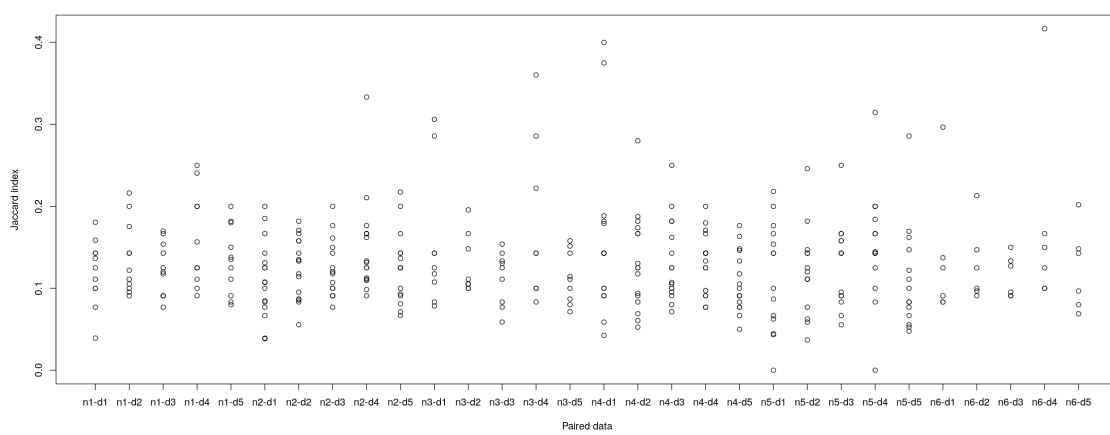


図 4.30: 通常睡眠データと阻害睡眠データで各クラスタの Jaccard 係数を計算した図.

は相互相関行列の上から 20, 30, 40, 50[%] の値とした。クラスタ数は前節同様プログラムを用いて自動で決定したが、クラスタ数があまりに大きすぎる場合に目視で判断できるデータに対しては目視で決定した。

相互相関行列をそのまま用いた場合と k 近傍法を用いた場合に推定されたクラスタのクラスタ内ニューロン数を図 4.2.3 に示す。相互相関行列そのままの場合や k が小さい場合はうまくクラスタ数が決められず 1 つのニューロンのクラスタができています。また、 k を大きくした場合でも、睡眠中のデータで大規模クラスタは見受けられない。

ε 近傍法を用いた時の推定クラスタ内ニューロン数を図 4.2.3 に示す。閾値が上から 30[%] や 40[%] の時は睡眠中データでの大規模クラスタは見られるが、20[%] と 50[%] の時には見られない。閾値によって結果が変化してしまうため、閾値を決める必要のない点で提案アプローチの方が良いと考えられる。

4.3 補遺

4.3.1 ネットワーク構造のパラメータ値の決め方

ニューロンのネットワーク構造を small world network によって表すために、初期次数と張り替え確率を実データから決める。今回はこの値はニューロンのコネクションの割合と相互のコネクションの割合から決める。興奮性ニューロン同士の 6.7% であり、そのうち双方向のコネクションの割合は 24% である [42]。発達中マウスの興奮性ニューロンから抑制性ニューロンへのコネクティビティと抑制性ニューロンから興奮性ニューロンへのコネクティビティはどちらも 78% であった [43]。成熟したマウスではより少ないと思われるが、データが見つからなかったため、40% とした。相互のコネクションの割合がランダムにエッジを作るよりも高いのは、近いニューロンにコネクションが作られやすいからだと考えられる。これらのデータを実現するように初期次数と張り替え確率を調整した。用いたパラメータを表 4.9 に示す。抑制性ニューロン同士のコネクティビティは分からないため、興奮性ニューロンと同じにしている。

結合の種類	初期次数	張り替え確率
同種類のニューロン間	$0.0335N$	0.3
興奮性ニューロンと抑制性ニューロン間	$0.2N$	0.3

表 4.9: ネットワーク構造のパラメータ値

実際のネットワーク構造の作り方を説明する。ネットワーク構造は興奮性ニューロン同士の結合、抑制性ニューロン同士の結合、興奮性ニューロンと抑制性ニューロン間の結合の 3 つに分けて作成する。まず、全ニューロンのうち抑制性ニューロンと興奮性ニューロンのインデックスを決めておく。全てのニューロンについて表 4.9 に従ってネットワークを作成し、それぞれに対応する隣接行列の上三角または下三角行列を取り出して結合する。作成したいのは向きのある有向グラフなので、上三角行列と下三角行列を分けて作成する。

4.3.2 スパイクシミュレーションの設定

スパイクシミュレーションで設定しなければならないのは、個々のニューロンの特徴パラメータ、重み付きのネットワーク構造、外部からのランダムな入力である。これらについて説明する。

まず、個々のニューロンの特徴パラメータについて説明する。このモデルではニューロンごとに 4 つのパラメータを設定する必要がある。そのパラメータでニューロンを特徴づける。本論文では興奮性ニューロンには regular spiking neurons、抑制性ニューロンには fast

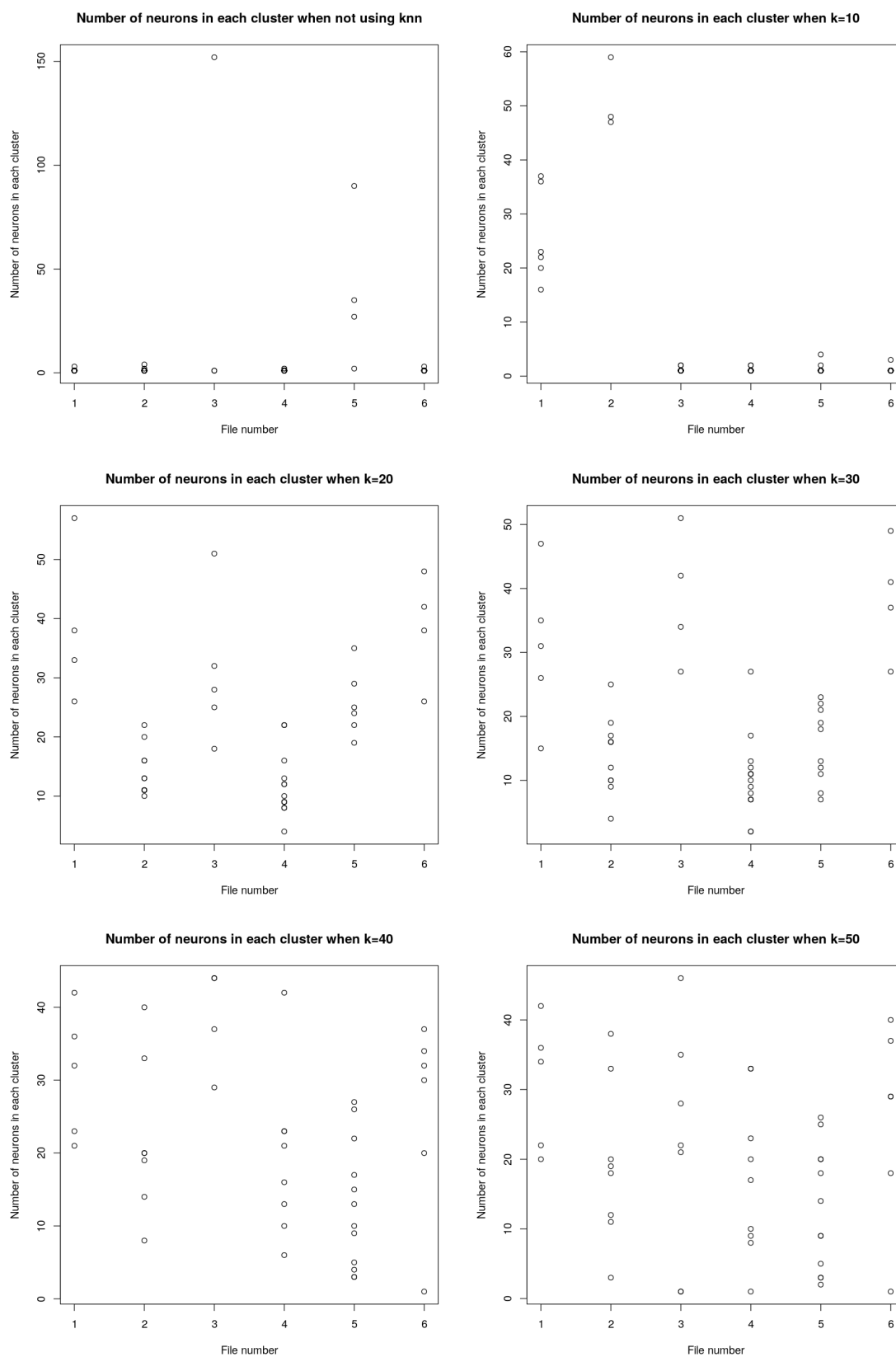


図 4.31: k 近傍法の k を変化させた時のクラス内ニューロン数. 左上は k 近傍法を用いず相互相関行列のまま用いてスペクトラルクラスタリングを行ったもの.

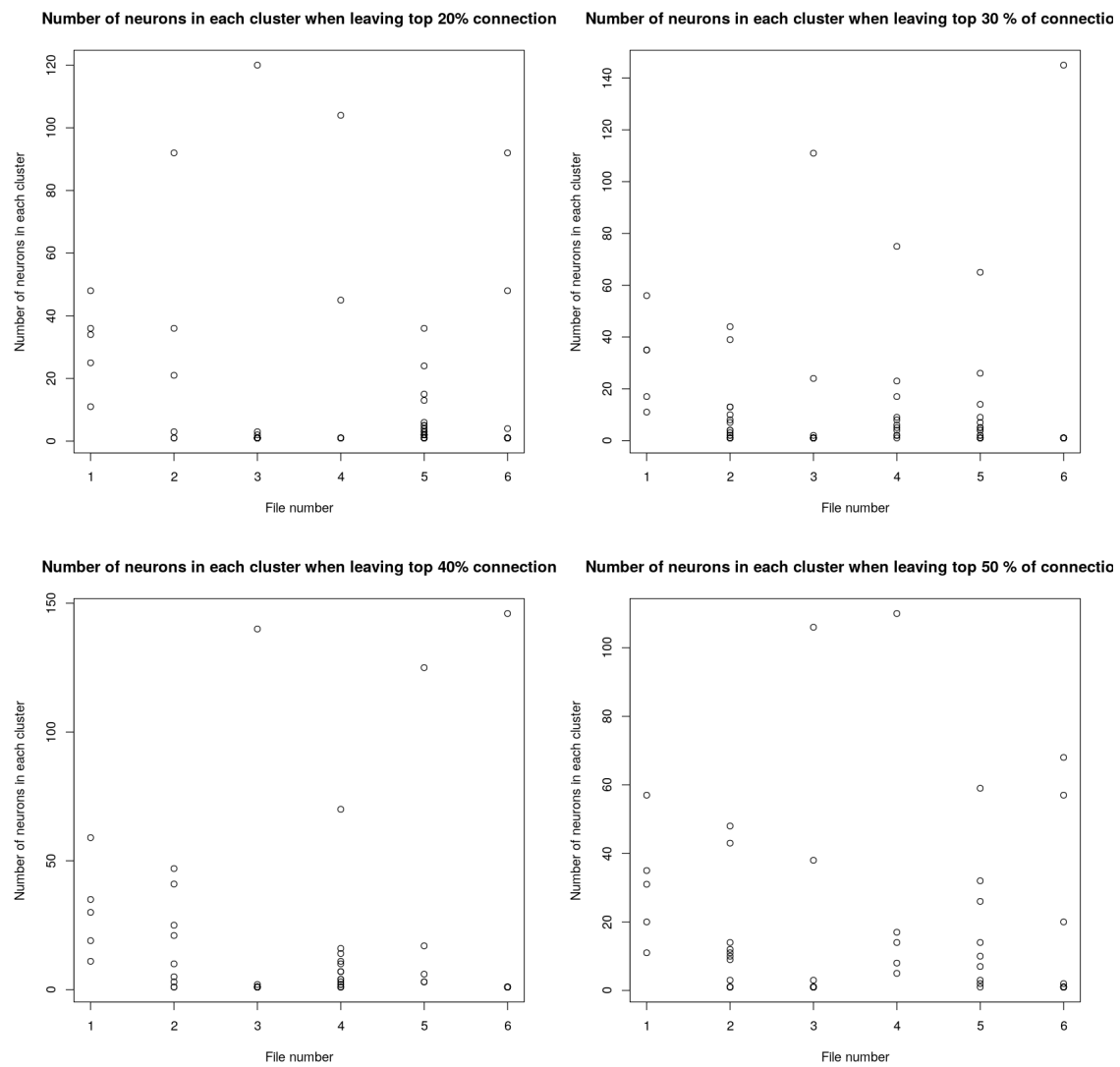


図 4.32: ϵ 近傍法の ϵ を変化させた時のクラスタ内ニューロン数. ϵ の値は相互相関行列の上から何 [%] までの値とした.

spiking neurons を用いる。それらのパラメータを 表 4.10 に示す。ただし、 r_e と r_i は 0 から 1 の一様分布に従う確率変数である。

ニューロンの種類	a	b	c	d
興奮性ニューロン	0.02	0.2	$-65 + 15r_e^2$	$8 - 6r_e^2$
抑制性ニューロン	$0.02 + 0.08r_i$	$0.25 - 0.05r_i$	-65	2

表 4.10: Izhikevich モデルのパラメータ値

次に、重み付きのネットワーク構造 $W \in \mathbb{R}^{N \times N}$ について説明する。ニューロン i から j へ結合があった場合、 w_{ij} はニューロン i が発火した時にニューロン j の膜電位をどれだけ上昇させるかという数値である。 W は、前節で作成した S の非ゼロ要素を数値で置き換えることで作成する。重み付きネットワーク構造 W の作成方法は 2 種類ある：

1. 全ての重みを同じ分布から生成する
2. 同じグループへの興奮性ニューロンからの入力が強めにする

1 番目の方法では、興奮性ニューロンからの結合は標準偏差 $\sigma_w = 1.5$ の対数正規分布の 10 以下の分布、抑制性ニューロンからの結合は一様分布 $U(-10, 0)$ からサンプルする。

$$w_{ij} = \begin{cases} \left\{ \frac{1}{\sqrt{2\pi\sigma_w^2}} \exp\left\{-\frac{(\ln w)^2}{2\sigma_w^2}\right\} \middle| w \in [0, 10] \right\} & (s_{ij} = 1, i \text{ は興奮性ニューロン}) \\ U(-10, 0) & (s_{ij} = 1, i \text{ は抑制性ニューロン}) \\ 0 & (s_{ij} = 0) \end{cases}$$

2 番目の方法では、同じグループの興奮性ニューロンからの結合は一様分布 $U(7, 10)$ 、異なるグループの興奮性ニューロンからの結合は標準偏差 $\sigma_w = 1.5$ の対数正規分布の 7 以下の分布、抑制性ニューロンからの結合は一様分布 $U(-10, 0)$ からサンプルする。対数正規分布を用いる理由は [44] のデータに基づく。

$$w_{ij} = \begin{cases} U(7, 10) & (s_{ij} = 1, i \text{ は興奮性ニューロン}, i \text{ と } j \text{ は同じグループ}) \\ \left\{ \frac{1}{\sqrt{2\pi\sigma_w^2}} \exp\left\{-\frac{(\ln w)^2}{2\sigma_w^2}\right\} \middle| w \in [0, 7] \right\} & (s_{ij} = 1, i \text{ は興奮性ニューロン}, i \text{ と } j \text{ は異なるグループ}) \\ U(-10, 0) & (s_{ij} = 1, i \text{ は抑制性ニューロン}) \\ 0 & (s_{ij} = 0) \end{cases}$$

最後に外部からのランダムな入力について説明する。ニューロンには観測範囲外からの入力がある（以降、外部入力とする）。そのため、シミュレーション中も外部からの電位を乱数としてニューロンの電位に足す。本論文では、ニューロンの活動も外部入力の大きさに表現する。活動していない興奮性ニューロンと抑制性ニューロンにはそれぞれ、 $\mathcal{N}(0, 9)$ と $\mathcal{N}(0, 0.01)$ に従う乱数を足す。活動している興奮性ニューロンと抑制性ニューロンの外部入力は、それぞれの正規分布の平均を N_{eplus} と N_{iplus} とする。これらを表 4.11 に示す。活動していないニューロンへの外部入力は [35] で用いられていたものを採用した。ただし、興奮性ニューロンの活動時の外部入力は変化させた実験もある。

実際の脳でもこのように外部からの入力によってニューロンの活動を制御していると考えられる。あるニューロングループを活動させる別の方法として、そのグループのハブとなるニューロンにのみ強い外部入力を与える方法も考えられる。

実際のマウスのニューロンの発火頻度を [45] を元に表 4.12 に示す。表 4.6 の non-active120 データの興奮性ニューロンと抑制性ニューロンの発火頻度を図 4.33 と図 4.34 に示す。実際のデータと人工データでオーダーは合っていることがわかる。表 4.12 のデータは電気生理のデータなので、スパイク頻度の高いニューロンが観測される。そのため、多少数値が異なってもオーダーが合っていれば良いと思われる。

ニューロンの種類	活動時の外部入力	活動していない時の外部入力
興奮性ニューロン	$\mathcal{N}(N_{eplus}, 9)$	$\mathcal{N}(0, 9)$
抑制性ニューロン	$\mathcal{N}(N_{iplus}, 0.01)$	$\mathcal{N}(0, 0.01)$

表 4.11: シミュレーションに用いる外部入力の値

ニューロンの種類	覚醒時 (Hz)	ノンレム睡眠時 (Hz)	レム睡眠時 (Hz)
興奮性ニューロン	0.76 ± 1.53	0.69 ± 0.86	0.88 ± 1.33
抑制性ニューロン	5.59 ± 7.25	4.69 ± 5.62	4.25 ± 9.43

表 4.12: ニューロンごとの発火頻度の中央値

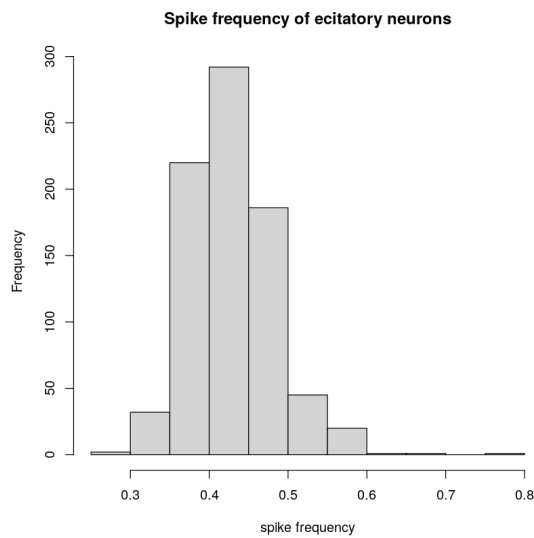


図 4.33: 興奮性ニューロンのスパイク頻度のヒストグラム。

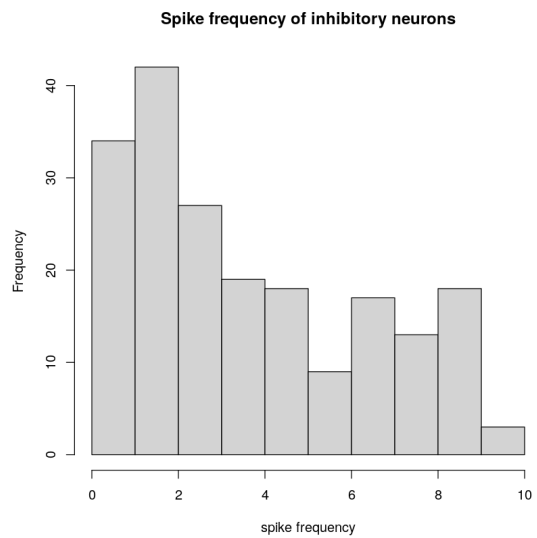


図 4.34: 抑制性ニューロンのスパイク頻度のヒストグラム。

第5章 結論

本論文では、サンプリングレートの低いカルシウムイメージングデータからニューロンのクラスタリングを行うことを目的に、バギングと NMF を用いた隣接行列推定がクラスタリングに有効であることを確かめた。データの生成モデルを立てた上で、ニューロン同士の結合推定に NMF を使い、バギングによって結合確率を推定した。NMF では基底数を決めることが難しいが、モデル平均によってその問題を緩和した。推定した隣接行列はスペクトラルクラスタリングを用いてクラスタリングを行った。クラスタ数は NMF の基底数とは別に固有値ギャップから推定した。人工データ実験の結果、真の基底数周りで隣接行列を推定できれば、ランダムに活動するニューロンがある場合やニューロンがランダムに活動する時間が多い場合に相互相関行列よりもロバストにグループ推定できることを確かめた。また、グループに所属しないニューロンもある程度推定できることも確かめられた。提案アプローチで実データ解析を行った結果、NREM 睡眠の多いデータで大規模クラスタが推定された。相互相関行列を用いたクラスタリングでは、パラメータによっては睡眠中の大規模クラスタが推定された。しかし、結果はパラメータによって大きく変化したため、提案アプローチの方が解析方法として扱いやすいのではないかとと思われる。

今後の展望としては、ニューロンが2つのグループに所属する場合に提案アプローチを拡張できるかを確かめる必要がある。また、付録 B で述べる方法で NMF の真の自由度が測れるかどうか実験を通して確かめたい面白い問題である。

謝辞

本研究を進めるにあたり指導教員の村田昇先生には多くのご指導をいただきました。また、今回用いた技術に関する問題の面白さを知ることができました。深く感謝いたします。赤穂先生、日野先生、有竹さんにも貴重なお時間を割いて助言をいただき、質問に答えていただきました。深く感謝いたします。また、貴重なデータを提供してくださり、カルシウムイメージングやニューロンに関する知見について教えてくださった筑波大学柳沢研究室の上田助教に感謝いたします。

研究室の先輩、同期、後輩にはゼミなどを通じて多くのアドバイスをいただきました。特に同期の皆様には細かい相談や精神面でも多く支えていただきました。ありがとうございました。

参考文献

- [1] E. M. Izhikevich, J. A. Gally, and G. M. Edelman, “Cerebral Cortex V 14 N 8 Spike-timing Dynamics of Neuronal Groups”, *Cortex August*, vol. 14, pp. 933–944, 2004.
- [2] O. Sporns, “Brain connectivity”, *Scholarpedia*, vol. 2, no. 10, p. 4695, Oct. 2007.
- [3] K. D. Harris, *Cell Assemblies of the Superficial Cortex*, Oct. 2012.
- [4] T. J. Sejnowski, P. S. Churchland, and J. A. Movshon, “Putting big data to good use in neuroscience.”, *Nature neuroscience*, vol. 17, no. 11, pp. 1440–1, Nov. 2014.
- [5] 中村 健, “神経細胞内局所的カルシウム濃度変化のリアルタイムイメージング法”, *Folia Pharmacol. Jpn*, vol. 121, no. 5, pp. 357–364, 2003.
- [6] G. Q. Bi and M. M. Poo, “Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type.”, *The Journal of neuroscience: the official journal of the Society for Neuroscience*, vol. 18, no. 24, pp. 10 464–10 472, 1998.
- [7] 平 理一郎, “脳神経計算原理の解明を目指した 2 光子多細胞イメージングの情報技術展開 [II・完] ”, *Animal Genetics*, vol. 101, no. 9, pp. 926–931, 2018.
- [8] K. Li, L. Guo, J. Nie, G. Li, and T. Liu, “Review of methods for functional brain connectivity detection using fMRI”, *Computerized Medical Imaging and Graphics*, vol. 33, no. 2, pp. 131–139, Mar. 2009.
- [9] J. T. Vogelstein, B. O. Watson, A. M. Packer, R. Yuste, B. Jedynek, and L. Paninski, “Spike Inference from Calcium Imaging Using Sequential Monte Carlo Methods”, *Biophysical Journal*, vol. 97, no. 2, pp. 636–655, Jul. 2009.
- [10] Y. Mishchencko, J. T. Vogelstein, and L. Paninski, “A Bayesian approach for inferring neuronal connectivity from calcium fluorescent imaging data”, *The Annals of Applied Statistics*, vol. 5, no. 2B, pp. 1229–1261, 2011.
- [11] O. Stetter, D. Battaglia, J. Soriano, and T. Geisel, “Model-Free Reconstruction of Excitatory Neuronal Connectivity from Calcium Imaging Signals”, *PLoS Comput Biol*, vol. 8, no. 8, p. 1 002 653, 2012.
- [12] Y. Ikegaya, G. Aaron, R. Cossart, D. Aronov, I. Lampl, D. Ferster, and R. Yuste, “Synfire chains and cortical songs: temporal modules of cortical activity.”, *Science (New York, N.Y.)*, vol. 304, no. 5670, pp. 559–64, Apr. 2004.
- [13] J. Mölter, L. Avitan, and G. J. Goodhill, “Detecting neural assemblies in calcium imaging data”, *BMC Biology*, vol. 16, no. 1, p. 143, Nov. 2018.
- [14] K. Ghandour, N. Ohkawa, C. C. A. Fung, H. Asai, Y. Saitoh, T. Takekawa, R. Okubo-Suzuki, S. Soya, H. Nishizono, M. Matsuo, M. Osanai, M. Sato, M. Ohkura, J. Nakai, Y. Hayashi, T. Sakurai, T. Kitamura, T. Fukai, and K. Inokuchi, “Orchestrated ensemble activities constitute a hippocampal memory engram”, *Nature Communications*, vol. 10, no. 1, pp. 1–14, Dec. 2019.
- [15] D. D. Lee and H. S. Seung, “Learning the parts of objects by non-negative matrix factorization”, *Nature*, vol. 401, no. 6755, pp. 788–791, Oct. 1999.

-
- [16] I. Wasserman, “Bayesian model selection and model averaging”, *Journal of Mathematical Psychology*, vol. 44, no. 1, pp. 92–107, Mar. 2000.
 - [17] H. Akaike, “A New Look at the Statistical Model Identification”, *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716–723, 1974.
 - [18] J. P. Brunet, P. Tamayo, T. R. Golub, and J. P. Mesirov, “Metagenes and molecular pattern discovery using matrix factorization”, *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 12, pp. 4164–4169, Mar. 2004.
 - [19] S. Ubaru, K. Wu, and K. E. Bouchard, “UoI-NMF cluster: A robust nonnegative matrix factorization algorithm for improved parts-based decomposition and reconstruction of noisy data”, in *Proceedings - 16th IEEE International Conference on Machine Learning and Applications, ICMLA 2017*, vol. 2017-December, Institute of Electrical and Electronics Engineers Inc., 2017, pp. 241–248.
 - [20] S. Wu, A. Joseph, A. S. Hammonds, S. E. Celniker, B. Yu, and E. Frise, “Stability-driven nonnegative matrix factorization to interpret Spatial gene expression and build local gene networks”, *Proceedings of the National Academy of Sciences of the United States of America*, vol. 113, no. 16, pp. 4290–4295, 2016.
 - [21] L. N. Hutchins, S. M. Murphy, P. Singh, and J. H. Graber, “Position-dependent motif characterization using non-negative matrix factorization”, *BIOINFORMATICS ORIGINAL PAPER*, vol. 24, no. 23, pp. 2684–2690, 2008.
 - [22] A. T. Cemgil, “Bayesian inference for nonnegative matrix factorisation models”, *Computational Intelligence and Neuroscience*, vol. 2009, 2009.
 - [23] L. Breiman, “Bagging Predictors”, Tech. Rep., 1996, pp. 123–140.
 - [24] B. Efron, “Bootstrap Methods: Another Look at the Jackknife”, *The Annals of Statistics*, vol. 7, no. 1, pp. 1–26, Jan. 1979.
 - [25] U. Von Luxburg, “A Tutorial on Spectral Clustering”,
 - [26] A. Ben-Hur, A. Elisseeff, and I. Guyon, “A stability based method for discovering structure in clustered data”, Tech. Rep.
 - [27] R. Tibshirani, G. Walther, and T. Hastie, “Estimating the number of clusters in a data set via the gap statistic”, Tech. Rep.
 - [28] L. Sjulson and G. Miesenböck, “Optical Recording of Action Potentials and Other Discrete Physiological Events: A Perspective from Signal Detection Theory”, *Physiology*, vol. 22, no. 1, pp. 47–55, Feb. 2007.
 - [29] T. Boongoen and N. Iam-On, *Cluster ensembles: A survey of approaches with recent extensions and applications*, May 2018.
 - [30] J. Kittler, M. Hatef, R. P. Duin, and J. Matas, “On combining classifiers”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 3, pp. 226–239, 1998.
 - [31] L. I. Kuncheva and D. P. Vetrov, “Evaluation of stability of k-means cluster ensembles with respect to random initialization”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 11, pp. 1798–1808, 2006.
 - [32] X. Fu, K. Huang, N. D. Sidiropoulos, and W. K. Ma, “Nonnegative Matrix Factorization for Signal and Data Analytics: Identifiability, Algorithms, and Applications”, *IEEE Signal Processing Magazine*, vol. 36, no. 2, pp. 59–80, 2019.

-
- [33] B. Dong, M. M. Lin, and M. T. Chu, “Nonnegative rank factorization—a heuristic approach via rank reduction”, *Numerical Algorithms*, vol. 65, no. 2, pp. 251–274, Feb. 2014.
 - [34] D. J. Watts and S. H. Strogatz, “Collective dynamics of ‘small-world’ networks”, *Nature*, vol. 393, no. 6684, pp. 440–442, Jun. 1998.
 - [35] E. M. Izhikevich, “Simple Model of Spiking Neurons”, *IEEE TRANSACTIONS ON NEURAL NETWORKS*, vol. 14, no. 6, 2003.
 - [36] T. W. Chen, T. J. Wardill, Y. Sun, S. R. Pulver, S. L. Renninger, A. Baohan, E. R. Schreiter, R. A. Kerr, M. B. Orger, V. Jayaraman, L. L. Looger, K. Svoboda, and D. S. Kim, “Ultrasensitive fluorescent proteins for imaging neuronal activity”, *Nature*, vol. 499, no. 7458, pp. 295–300, 2013.
 - [37] N. Guan, D. Tao, Z. Luo, and B. Yuan, “NeNMF: An optimal gradient method for nonnegative matrix factorization”, *IEEE Transactions on Signal Processing*, vol. 60, no. 6, pp. 2882–2898, Jun. 2012.
 - [38] 永山瑞生, “マウスのカルシウムイメージングデータを用いた睡眠・覚醒状態の解析”, 2018.
 - [39] A. Cichocki, R. Zdunek, A. H. Phan, and S.-i. Amari, *Nonnegative Matrix and Tensor Factorizations*. Chichester: Wiley Publishing, 2009.
 - [40] M. R. Symonds and A. Moussalli, *A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike’s information criterion*, Aug. 2011.
 - [41] T. Kanda, N. Tsujino, E. Kuramoto, Y. Koyama, E. A. Susaki, S. Chikahisa, and H. Funato, “Sleep as a biological problem: an overview of frontiers in sleep research”, *The Journal of Physiological Sciences*, vol. 66, no. 1, pp. 1–13, Jan. 2016.
 - [42] J.-S. Jouhanneau, J. Kremkow, A. L. Dorn, and J. F. Poulet, “In Vivo Monosynaptic Excitatory Transmission between Layer 2 Cortical Pyramidal Neurons”, *Cell Reports*, vol. 13, no. 10, pp. 2098–2106, Dec. 2015.
 - [43] C. Holmgren, T. Harkany, B. Svennenfors, and Y. Zilberter, *Pyramidal cell communication within local networks in layer 2/3 of rat neocortex*, Aug. 2003.
 - [44] S. Song, P. J. Sjöström, M. Reigl, S. Nelson, and D. B. Chklovskii, “Highly non-random features of synaptic connectivity in local cortical circuits”, in *PLoS Biology*, vol. 3, 2005, pp. 0507–0519.
 - [45] B. O. Watson, D. Levenstein, J. P. Greene, J. N. Gelinas, and G. Buzsáki, “Network Homeostasis and State Dynamics of Neocortical Sleep”, *Neuron*, vol. 90, no. 4, pp. 839–852, May 2016.
 - [46] M. Babaei, S. Tsoukalas, M. Babaei, G. Rigoll, and M. Datcu, “Discriminative Nonnegative Matrix Factorization for dimensionality reduction”, *Neurocomputing*, vol. 173, pp. 212–223, Jan. 2016.
 - [47] V. C. Cheung, K. Devarajan, G. Severini, A. Turolla, and P. Bonato, “Decomposing time series data by a non-negative matrix factorization algorithm with temporally constrained coefficients”, in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, vol. 2015-Novem, Institute of Electrical and Electronics Engineers Inc., Nov. 2015, pp. 3496–3499.
 - [48] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight, “Sparsity and smoothness via the fused lasso”, *Tech. Rep.*, 2005, pp. 91–108.
-

付録 A 考案した NMF の制約

本節では、データの特徴を考慮して考案した NMF の制約とモデルエビデンスの計算方法を紹介する。

1 スケール除去

ニューロンごとに発現している蛍光タンパク質の量や細胞の大きさが異なる。そのため、ニューロンごとのバイアスが観測データに載っていると考え、バイアスを除去する方法を試した。この時の数理モデルは以下ようになる：

$$X = DC + H + B.$$

ただし、 $B \in \mathbb{R}_+^{I \times J}$ は行ごとに同じ数値が入ったバイアス行列である。

バイアスの推定方法は、 D に 1 列を足し、 C に $\mathbf{1}$ の 1 行を足して NMF を更新する。 D の列にバイアスが推定されることを期待した。

簡単な人工データ実験を行った結果、足したバイアスよりも大きいバイアスが推定されてしまうことがわかった。また、そもそもの数理モデルが異なると考え直した。

蛍光タンパク質の量や細胞の大きさが異なるということは、各ニューロンはスケールされていると考えられ、以下のように表される：

$$X = A(DC + H).$$

ただし、 $A \in \mathbb{R}_+^{I \times I}$ は対角行列である。

ナイーブな求め方は、 A は単位行列で初期化し、NMF 一回の更新ごとに X をニューロンごとに残差の四分位範囲で割り、その値を A にかけていく。簡単な人工データ実験で乗法更新則を用いた場合は A は真の値に近いものが推定された。

2 重複除去

本研究で置いた仮定では、あるニューロンが複数のグループに所属する時、グループの活動は被らないとしている。しかし、NMF の推定時にそのような制約は入れていないので、NMF で推定した結果この仮定が破られているようであれば制約は入れなければならない。

そこで、以下の目的関数を考えた：

$$\arg \min_{D \geq 0, C \geq 0} \|X - DC\|_F^2 - \lambda \sum_{k=1}^K \sum_{l \neq k}^K (\|d_{:,l} - d_{:,k}\|_1 \|c_{l,:} - c_{k,:}\|_1). \quad (1)$$

更新則を導出する。参考にしたのは [46] である。式 (1) の Lagrange 関数 L は、

$$\begin{aligned} L = & \text{Tr}(X^T X) - 2\text{Tr}(X^T DC) + \text{Tr}(C^T D^T DC) - \text{Tr}(\Phi_C C^T) \\ & - \text{Tr}(\Phi_D D^T) - \lambda \text{Tr}(F^T C H^T S^T D F), \end{aligned}$$

であり, KKT 条件は,

$$\begin{aligned}\frac{\partial L}{\partial C} &= \frac{\partial L}{\partial D} = 0 \\ D &\geq 0 \\ C &\geq 0 \\ \Phi_C &\geq 0 \\ \Phi_D &\geq 0 \\ \Phi_C C &= \Phi_D D = 0\end{aligned}$$

である. ただし, Φ_D と Φ_C はそれぞれ $D \geq 0$, $C \geq 0$ に対する Lagrange 乗数で, $F \in [0, 1]^{K \times (K-1)!}$ は 2 つの時間の組み合わせを表現した以下のような行列である:

$$F = \begin{pmatrix} 1 & 1 & \dots & 0 & \dots & 0 \\ -1 & 0 & \dots & 1 & \dots & 0 \\ 0 & -1 & \dots & -1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & \dots & 1 \\ 0 & 0 & \dots & 0 & \dots & -1 \end{pmatrix}.$$

また, $S = \text{sign}(DF)$, $H = \text{sign}(F^T C)$ とおく.

D を求めるには以下の式を解く:

$$\frac{\partial L}{\partial D} = -2XC^T + 2DCC^T - \Phi_D - \lambda SHC^T FF^T = 0.$$

これを求めると D の要素の更新は以下である:

$$d_{ik} \leftarrow d_{ik} \frac{2[XC^T]_{ik} + \lambda[SHC^T FF^T]_{ik}^+}{2[DCC^T]_{ik} - \lambda[SHC^T FF^T]_{ik}^-}.$$

ただし, $[\cdot]^+$ は行列の中の正の要素, $[\cdot]^-$ は負の要素である.

同様に, C の要素の更新は以下の通りである:

$$c_{kj} \leftarrow c_{kj} \frac{2[D^T X]_{kj} + \lambda[FF^T D^T SH]_{kj}^+}{2[D^T DC]_{kj} - \lambda[FF^T D^T SH]_{kj}^-}.$$

しかし, ハイパーパラメータ λ の最適な決め方は分からない. また, λ が大きいと D と C はスパースになるが, 本来 2 つの基底に所属するニューロンの $d_{i\cdot}$ が 1 つの基底以外 0 に近くなることが実験を通してわかった. この手法は改善の余地がある上, 実際のデータを NMF で分解した際にどれくらい仮定に沿っていないかを確認した上で使う必要がある.

3 時間方向への制約

カルシウムイメージングデータはスパイク情報を反映するのが遅く, 一度上がった蛍光強度は緩やかに下がっていく. そのため, NMF で分解を行う際も, 行列 C の時間方向に前時刻の値と近くなるような制約を入れることでより正確なニューロングループの抽出が行えると考えられる. C の偶数列を前後の列の平均とする NMF も提案されている [47]. しかし, これはかなりスムーズになる制約だと考えられる. そこで, 以下のような制約を加えた目的関数が考えられる:

$$\arg \min_{D \geq 0, C \geq 0} \|X - DC\|_F^2 + \lambda \sum_t \|c_{:t} - c_{:t-1}\|_1.$$

これは fused lasso [48] と同じような制約である.

更新則は, D は式 (3.3) と同じだが C は異なり以下である:

$$c_{kj} \leftarrow c_{kj} \frac{2[D^T X]_{kj} - s_{kj}}{2[D^T DC]_{kj}}.$$

ただし, $s_{:j} = \text{sign}(c_{:j} - c_{:j-1})$ であり, 1 列目のみ $s_{:1} = \mathbf{0}$ である.

重複除去制約と同じく, λ の決め方が分からないため使用は断念した. NMF は初期値依存の問題があるため, λ を変化させると解空間も変化し, 結果は滑らかに変化しない. そのため, λ を変化させた効果を比較するのは難しい.

付録B NMFのモデルエビデンス

NMFの基底数を決める際にブートストラップから計算されたモデルエビデンスを用いることを考える。BICは最尤推定量の尤度からモデルエビデンスを近似して扱っている。モデルエビデンスはブートストラップ法によっても近似計算できると考えられる。

基底数 k のNMFのモデルを \mathcal{M}_k とおく。ノイズ行列 H の各要素が正規分布 $\mathcal{N}(0, \sigma^2)$ に従う \mathcal{M}_k の尤度は以下である：

$$p(X|Y_k, \mathcal{M}_k) = \prod_{i,j} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{([Y_k]_{ij} - x_{ij})^2}{2\sigma^2}\right). \quad (1)$$

ただし、 $Y_k = D_k C_k$ で、 Y_k はモデル \mathcal{M}_k における推定量である。対数尤度は以下のようになる。

$$\log p(X|Y_k, \mathcal{M}_k) = -\frac{IJ}{2}(\log 2\pi + 2\log \sigma) - \frac{1}{2\sigma^2} \sum_{i,j} ([Y_k]_{ij} - x_{ij})^2.$$

今回の問題を Bayesian model averaging の枠組みに当てはめると、異なる基底数のNMFから求めた \bar{A} をモデルの事後確率で重み付て足し合わせることになる。この時の式は、

$$p(\bar{A}|X) = \sum_k p(\bar{A}|\mathcal{M}_k, X) p(\mathcal{M}_k|X),$$

で $p(\bar{A}|X)$ を求めることになる。

モデルの事後確率は

$$p(\mathcal{M}_k|X) = \frac{m_k p(\mathcal{M}_k)}{\sum_l m_l p(\mathcal{M}_l)},$$

で表される。ただし、

$$m_k = \int p(X|Y_k, \mathcal{M}_k) p(Y_k|\mathcal{M}_k) dY_k,$$

である。これはモデルエビデンスや marginal likelihood と呼ばれる。また、 $p(\mathcal{M}_k)$ はモデルが正しい確率である。

ある条件下で、母数の事後確率の密度関数はブートストラップによる最尤推定量の分布と同じになる。そのため、ブートストラップサンプルから推定した Y_k の分布をもとに事後確率を計算すれば良い。

現在は $p(\mathcal{M}_k)$ に関する知識はなく無情報なため、モデルエビデンスを推定することで各モデルの重みが求まる。

1 ラプラス近似

モデルエビデンスの計算には事前分布を決めなければならず、積分も計算しなければいけない。そこでラプラス近似により m_k は以下のような \hat{m}_k で近似できる。

$$\log \hat{m}_k = \log p(X|\hat{Y}_k, \mathcal{M}_k) - \frac{d_k}{2} \log n.$$

ここで、 d_k は母数の数 ($I \times K + K \times J$), n は観測データ数 (J) である. この近似を使って Bayes 因子のログをとったものが BIC である. Bayes 因子は 2 つのモデルエビデンスの比である. NMF ではデータの増加とともにパラメータ数も増加するため, BIC を用いて基底数を決めるのは本来適さない.

実験では, $\log p(X|\hat{Y}_k, \mathcal{M}_k)$ は初期値を変えて NMF を行い, 尤度が最も大きくなった対数尤度を用いる. また, $\sigma^2 = \text{Var}(X - Y_k)$ として計算する.

2 ブートストラップによる近似

ブートストラップの推定量の分布は最尤推定量の分布を近似する (ブートストラップサンプルが生成されたパラメータ分バイアスは乗る). これを簡単なデモンストレーションで確認する.

人工的に D と C を作成し, 標準正規分布に従うノイズを加えて X を作る. 事後確率 $p(\mathcal{M}_k|X)$ は

$$p(\mathcal{M}_k|X) = \frac{p(X|Y_k, \mathcal{M}_k)p(\mathcal{M}_k)}{p(X)},$$

である. 事前分布 $p(\mathcal{M}_k)$ を無情報とした時, 事後確率は尤度に比例する. そこで, データ X のある 2 要素 (i, j) と (i', j') について尤度の分布を図示する. ノイズに標準正規分布を用いたので, $\sigma = 1$ として式 (1) を 2 つの要素のある範囲について計算する. 求めたいのは事後確率の分布なので, 「 $[Y_k]_{ij} = 5$ かつ $[Y_k]_{i'j'} = 8$ の時の尤度を計算する」ことを 5 や 8 の値をずらしながら行う. その結果を図 1 に示す.

次に, ブートストラップの推定量が図 1 を近似することを確かめる. ブートストラップサンプルを作成し NMF を行い, 図 1 の 2 要素についてヒートマップを作成する. その結果をブートストラップのサンプル数別に図 2~4 に示す. これより, 図 4 は図 1 の分布を近似できていることがわかる.

これより, ブートストラップの推定量の分布は最尤推定量の分布を近似するので, m_k は以下のように近似できる.

$$\begin{aligned} m_k &= \int p(X|Y_k, \mathcal{M}_k)p(Y_k|\mathcal{M}_k)dY_k \\ &\sim \frac{1}{B} \sum_b \prod_{i,j} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{([Y_k^b]_{ij} - X_{ij}^b)^2}{2\sigma^2}\right). \end{aligned}$$

ここで, Y_k^b はブートストラップサンプル b から計算された Y_k で, B はブートストラップサンプル数である. また, $\sigma^2 = \text{Var}(X^b - Y_k^b)$ として計算する.

3 実験

人工データセット 86 個について 2 つの近似方法でモデルエビデンスを計算した. その結果を図 5 と図 6 に示す. 同じデータの結果を線で結んでいる. ラプラス近似を行った場合は真の基底数 10 付近で最大値をとっている. ブートストラップによる近似を行った場合は基底数に比例してモデルエビデンスの値も大きくなっている. また, 同じデータでも基底数によってオーダーが大きく異なるが, これは尤度計算に用いる分散をデータから計算しているためである.

また, 図 6 のログを取り, 図 5 とプロットした図を図 7 に示す. これより, ブートストラップによる近似を用いた場合のモデルエビデンスの値は大きいことが分かる. ラプラス近

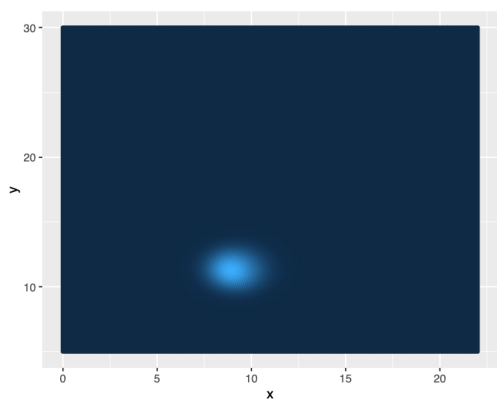


図 1: 人工データの 2 つの要素の尤度分布

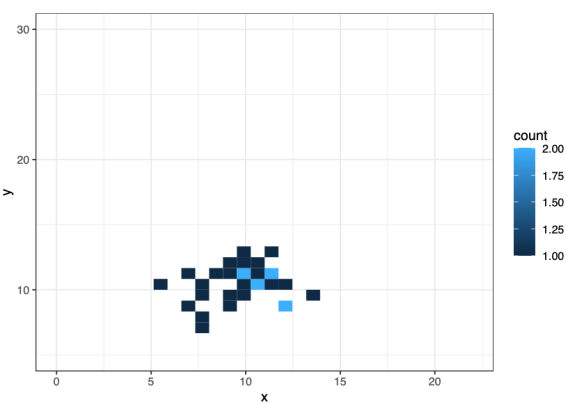


図 2: 30 個のブートストラップサンプルから推定された 2 つの要素の値の分布

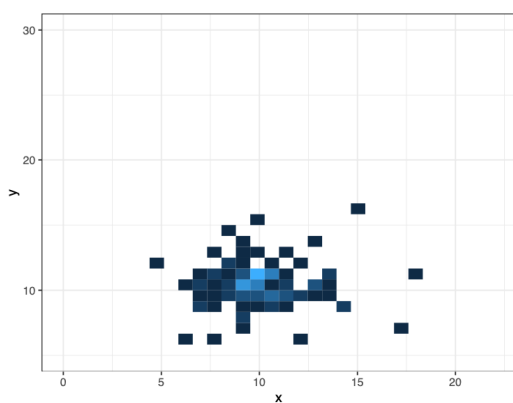


図 3: 100 個のブートストラップサンプルから推定された 2 つの要素の値の分布

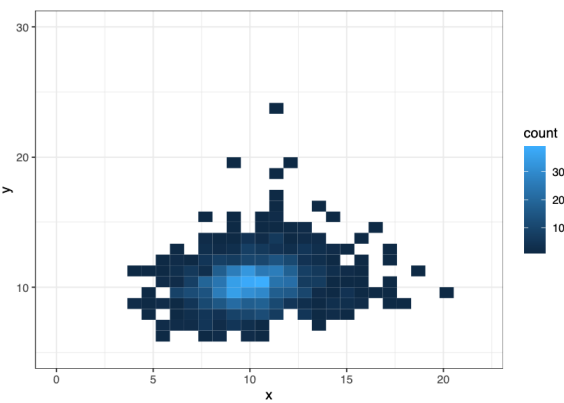


図 4: 1000 個のブートストラップサンプルから推定された 2 つの要素の値の分布

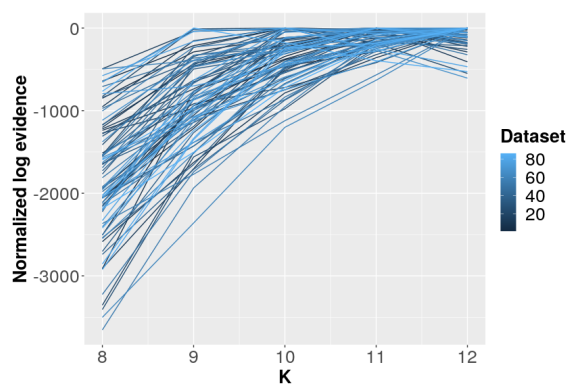


図 5: ラプラス近似によって計算されたログエビデンス. 各データについて最大値が 0 となるように定数を足した.

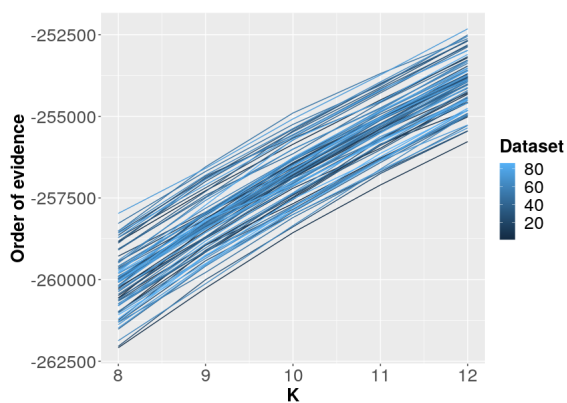


図 6: ブートストラップ近似によって計算されたエビデンスのオーダー (10 を底とした指数).

似では $\frac{d_k}{2} \log n$ というパラメータ数による補正が加わっているため、ブートストラップによるモデルエビデンスの方が高い値を持つ。逆に、図 7 の 2 つの近似手法のモデルエビデンスの差が過剰に足された補正の値だと考えることもできる。つまり、 $\frac{d_k}{2} \log n$ から 2 つのモデルエビデンスの差を引いた数値が NMF の真の自由度であるとも考えられる。これを確かめるためには数値実験を行う必要がある。

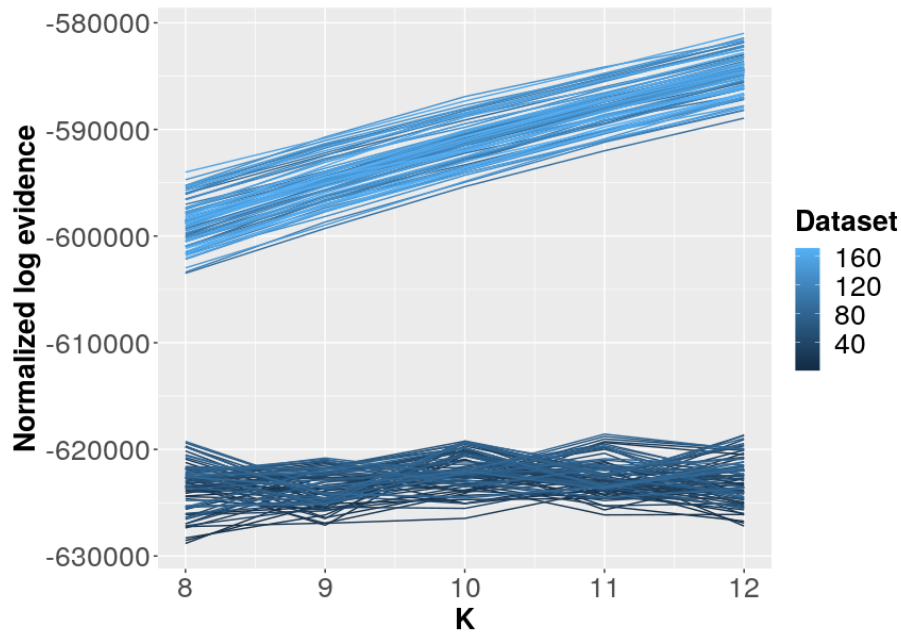


図 7: 2 つの近似方法によって計算されたログエビデンス。

研究活動の不正行為 定義確認書

2021 年 1 月 18 日

早稲田大学大学院先進理工学研究科長 殿

早稲田大学大学院先進理工学研究科

電気・情報生命専攻 2 年本人氏名 永山 瑞生

(自署)

私は、裏面に記載のある研究活動の不正行為に関する定義を理解し、修士論文提出において、不正行為または不正行為と疑われる行為を一切行っていないことを誓約します。

なお、当該行為を行なうと、厳重な処分（無期停学・当該学期成績無効・修士論文不合格等）を受けること、学位取得後であっても学位取消となることを十分に認識した上で、論文を執筆しました。

研究倫理教育受講確認書

2021 年 1 月 18 日

本人氏名 永山 瑞生

私は、修士論文の執筆にあたり、以下の研究倫理教育を受講しました。

✓	GEC 設置科目「研究倫理概論」の受講
	専攻設置科目の受講 科目名 ()
	その他 () ※受講前に指導教員を通じて専攻主任の承認を受けること。 ※受講証明を添付すること。