

科学研究費助成事業 研究成果報告書

平成 28 年 6 月 17 日現在

機関番号：32613

研究種目：若手研究(B)

研究期間：2014～2015

課題番号：26870577

研究課題名(和文) スパース正則化モデルによる神経回路の光計測データからの信号抽出

研究課題名(英文) Sparse MAP inference of cell shapes and spike trains from calcium imaging data

研究代表者

竹川 高志 (Takekawa, Takashi)

工学院大学・情報工学部・准教授

研究者番号：50415220

交付決定額(研究期間全体)：(直接経費) 2,400,000円

研究成果の概要(和文)：カルシウムイメージングの大規模化により、観測データから細胞位置とスパイク列を取り出す手順を自動化する必要性が生じている。本研究において独立な空間変数と時間変数を仮定した単純な確率モデルにスパース性を導入する定式化を行い、その解が反復二次計画問題によって少ない制御パラメータで効果的に得られることを示した。提案手法では、記録データ中の重なりのある複数の細胞の活動を正しく分離して推定することができるため、従来手法における近接した神経細胞に対して同期発火を誤検出してしまう欠点を解決することができた。同時に推定を高速に行う数値解法を提案し大規模データへの適用を可能とした。

研究成果の概要(英文)：Calcium imaging has become popular as data acquisition technique from neural circuits and the scale of imaging data is rapidly increasing. In this paper, we present an automatic cell sorting algorithm which detects cell positions and spike timings simultaneously from large scale calcium imaging data. Our algorithm is based on a simple probabilistic model with random variables for spatial and temporal activities of each cell. The model also contains spatial and temporal baseline as random variables and these variables can be effectively inferred by iterative quadratic programming problems with few control parameters for sparseness. Present inference procedure can detect separately spiking activities from from spatially overlapped neurons which activities contains cross-talk components. Faster implementation of the approach have been also developed by tuning of numerical solver routine using the problem specific characteristic.

研究分野：計算論的神経科学

キーワード：カルシウムイメージング 信号検出 信号源検出 スパイク時系列 モデル推定 大規模最適化問題

1. 研究開始当初の背景

神経細胞集団がどのように情報処理を行っているかを理解するためには複数の神経細胞の活動電位を同時に記録する必要がある。近年、複数細胞の同時記録手法としてカルシウムイメージング技術が急速に発展し広く使われるようになってきている。しかし、実際にはノイズの含まれた動画データから、各細胞の位置と活動の時系列を抽出する必要がありデータの大規模化に伴い抽出作業のコストが多大なものとなってきている。

また、マニュアルによる抽出作業では重なった細胞の扱いなど問題点があり、マニュアル作業の問題点を解決することと高速に自動処理を行う必要性が生じている。そのためすでにいくつかの手法が提案されていたが、既存手法では特に大規模データに対して細胞やスパイクの取りこぼしが多く、推定したスパイク時系列を用いた解析には不安が残る状態であった。

実際に、研究開始当初から現在に至るまで多くの研究室ではそれぞれ独自に主にマニュアル作業によって細胞の検出と活動の推定を行っており、人的資源を多く費やしながらか十分な信頼性を確保することができていない。このため、神経科学上の成果を発表する際にも最も基本的な解析である細胞の活動推定自体が問題となり、本質的な議論が困難である状況も生じている。

2. 研究の目的

カルシウムイメージングデータからのスパイク時系列抽出では ROI (region of interest) と呼ばれるニューロンの空間領域での形状を求めてから ROI 内でのカルシウム濃度の変化を計算し、スパイク時刻を推定するものが多い。しかし 3 次元の細胞を 2 次元で記録しているため別々のニューロンの ROI 同士が重なっている場合があり、この場合本来は存在しない隣り合った別のニューロンの活動が含まれてしまうことがある。このことは、神経回路ではニューロン間の同期的な活動が重要であると考えられるため、大きく誤った結論を導いてしまう危険性がある。

これまでに提案された ROI の重なりを考慮した主な自動ソーティング手法としてはヒューリスティックな手法[1]とモデルベースのもの[2][3][4]が提案されている。ヒューリスティックな手法ではパラメータのチューニングが困難であること、大規模なデータに対して細胞数の推定が不正確となり十分な精度が得られないことが問題となる。

モデルベースの手法では非負拘束条件を利用して解を得る手法が主流であるが、カルシウムイメージングデータの観測値が相対値であるため、非負拘束条件を正しく適応するた

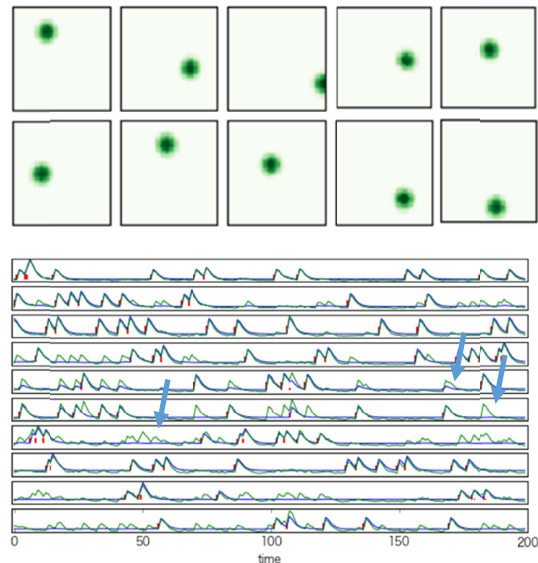


図 1 提案手法によるスパイク検出の例。緑がカルシウム濃度変化、赤がスパイク。隣り合った細胞のスパイクが分離できている (矢印部分)。

めには、細胞のスパイクとは無関係なベースラインの推定が必要であり、実験状況によってベースラインも時間的・空間的に異なるので注意が必要となる[4]。

また、モデルベースの場合、空間的な細胞同士の重なりを考慮することになりサイズの大きいデータに対しては、細胞数の更新を行いながら大規模な最適化問題を繰り返し解く必要が生じるため、数値計算の実行速度を大幅に向上させる必要がある。

本研究では、以上のような背景にもとづいて、検出精度の向上と実用的な速度での大規模データへの適用を目的とし、確率モデルに対する適切な事前知識の導入と高速な数値解法の構築、さらに数値解法において収束性と実行速度を向上させる様々な手法について検証した。

3. 研究の方法

細胞形状とスパイク時系列と観測データに関する Vogelstein らにより提案された確率モデル[2]とその改良[3]をもとにして、細胞形状と発火率および記録における空間あるいは時間的な偏りや変化であるベースラインに関する事前知識を導入した上で、MAP 推定を行う手順を設計した。

細胞の大きさとスパイク発火頻度などに関する事前知識を導入することで高精度な推定を行う理論的枠組みを提案し、実装と検証を行った。特に、ベースラインの推定に関するモデル化を大きく変更し、事前知識と MAP 推定によるより適切なペナルティ項を導入することで収束性と検出精度を同時に向上させ

の方法について検証を行った (図 1) .

同時にアルゴリズムを実装し高速化を行い、人工データによる精度の検証と実際の大規模データに適用した際の実用性と有効性を確認した .

4 . 研究成果

既存研究[2][3]における基本モデルでは観測値 F_{lt} は、ベースライン b_{lt} , 細胞数 K , 各細胞の活動 $a_{kl}v_{kt}$, ノイズ (誤差) ε_{lt} により構成される . 各細胞の活動は空間と時間の成分 a_{kl} , v_{kt} に分離できる . データのピクセル数を L , フレーム数を T とする . 時間成分はスパイク時系列 u_{kt} と Autoregressive (AR) model で表される . a_{kl} , u_{kt} はそれぞれ非負である .

$$F_{lt} = b_{lt} + \sum_{k=1}^K a_{kl}v_{kt} + \varepsilon_{lt},$$

$$v_{kt} = \sum_{\tau=1}^{N_{\tau}} \gamma_{\tau}v_{kt-\tau} + u_{kt}, \quad a_{kl} \geq 0, \quad u_{kt} \geq 0$$

このモデルについて、誤差が最小になるように変数 a と変数 v について別々に繰り返し最適化することが基本方針となるが、モデル自体の解の性質に加えて数値計算上の収束および計算速度も実用上の問題となる .

本研究において得た成果は (1) a および u に関する事前分布に関する改良 , (2) ベースラインに対する事前分布の導入 , (3) 導入した事前分布を含む MAP 推定の効率的な解法の提案と実装 , (4) 計算機上での数値計算の効率化 , である . 特に (2) については、事前分布の導入により推定精度の向上と収束に対する性質の向上が同時に実現することができた . また、細胞数の推定には、多数の細胞候補からベイズ推定における関連度自動決定の性質を用いて絞り込みを行う方式を採用し、良好な結果を得ている (図 2) . 以下、各項目について解説を行う .

(1) 細胞形状 a およびスパイク u に対して単純な L1 正規化項の導入により解の性質を改良する提案はすでに行われていた . これをさらに押し進め、細胞形状とスパイクへの事前知識としてはカルシウム色素の反応の強さとは無関係に大きさと発火率のみが影響することを考慮して、以下のように反応強度 s を導入した .

$$F_{lt} = b_{lt} + \sum_{k=1}^K s_k a_{kl} v_{kt} + \varepsilon_{lt},$$

$$s_k \geq 0, \quad \max_l a_{kl} = \max_t u_{kt} = 1$$

その結果、細胞数の推定において精度の向上が見られた .

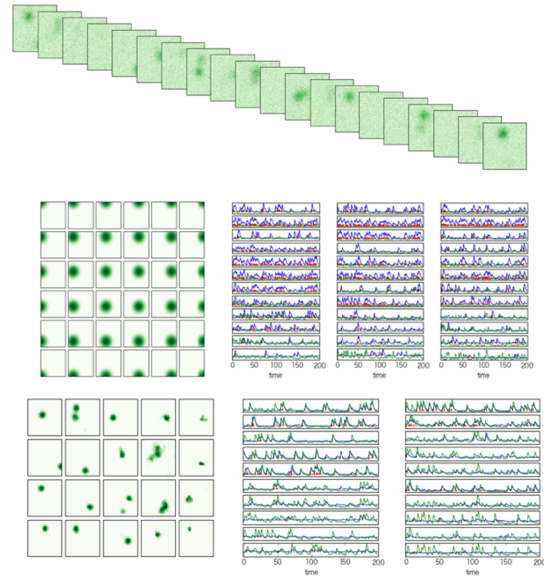


図 2 イメージングデータの例 (上) とアルゴリズムの動作の様子 . (中左) 初期条件として細胞候補の形状を設定 . (中右) 初期条件から細胞の活動状況を推定 . (下左) 活動状況から細胞形状を再び推定 . 1 回の反復ですでに細胞候補を 36 から 20 に絞り込むことができている .

(2) 細胞形状とスパイクだけでなくベースラインの分散についても正規化項を導入することで、推定される細胞形状の安定化とスパイク時系列がより安定してスパースな推定結果が得られるようになった . また、ベースラインへの正規化項により数値計算において反復回数が少なく収束する効果があることを発見した .

(3) 単純な最尤法 + 正規化項のモデルに対して、細胞サイズと発火頻度に関する事前知識を指数分布と正規分布による事前分布として再定義し MAP 推定を行うと、下記の非線形最適化問題となる .

$$\min_{\sigma^2, b, a, v} \frac{LT}{2} \log \sigma^2 + \frac{E}{2\sigma^2} + \frac{P_B}{2\sigma^2} + \frac{P_A}{a_0} + \frac{P_U}{u_0},$$

$$E = \sum_{l,t=1}^{LT} \varepsilon_{lt}^2, \quad P_A = \sum_{k,l=1}^{KL} a_{kl}, \quad P_U = \sum_{k,t=1}^{KT} u_{kt},$$

$$P_B = \sum_{l=1}^L (b_l - \bar{b}_L)^2 + \sum_{t=1}^T (b_t - \bar{b}_T)^2$$

もとの二次計画問題に対して非線形部分が導入されるが、逐次二次計画法を用いると計算量はほぼ変わらず非線形項導入の効果を得ることができた .

(4) 逐次二次計画法による各反復では、二次計画問題を解くことになる . 実験技術の向

上により、データサイズが大きくなり解くべき二次計画問題の次元数も増大している。そのため、単純にデータサイズに依存して計算量が増大するだけでなく細胞候補数や初期条件を適切に設定することの困難さから数値解法における収束条件の悪化などが問題となった。そこで、逐次二次計画法において各ステップを厳密に解かず主双対内点法で1ステップのみ更新を行う手法を導入した。さらに、主双対内点法の更新ベクトルを問題固有の行列の特徴を利用した行列演算の最適化を行った上でCG法を用いることで大幅な高速化を実現することができた。

研究背景で述べたように、カルシウムイメージング記録を行っている多くの研究室において独自に行なっている解析に対して信頼できる共通のツールを提供することは、神経科学全体の発展に大いに寄与することになる。今回開発した手法は、これまで提案されている手法よりも大幅に精度が向上しており、今後の標準的な手法として非常に有力な候補である。

また、実際には(1)～(4)はそれぞれ独立しているわけではなく、それぞれの相補的な特徴により全体として精度の向上と計算時間の削減が実現できている。今回の問題解決のために導入した、L1正規化項とL2正規化項の組み合わせ、信号強度を無視した信号の性質に対する事前知識の導入、大規模二次計画問題に対する逐次二次計画法と主双対内点法の組み合わせなどは、それぞれ他の分野への応用が可能である。

以上の成果の主要部分については、国際学会において発表し、詳細について論文投稿および特許申請を準備中である(2016年4月に特許申請を完了)。また、実際にカルシウムイメージング記録を行っている複数の研究室と共同研究を行い、提案手法の有用性を確認している。

<引用文献>

- E.A. Mukamel, A. Nimmerjahn and M.J. Schnitzer, "Automated analysis of cellular signals from large-scale calcium imaging data", *Neuron* 63, 747-760, 2009.
- Vogelstein, J., Packer, A., Machado, T., Sippy, T., Babadi, B., Yuste, R. & Paninski, L. "Fast non-negative deconvolution for spike train inference from calcium imaging", *J. Neurophys.* 104, 3691-3704, (2010).
- T. Takekawa, et. al, "Spike train inference from calcium imaging data with non-negative constraints", Annual

Meeting of the Japan Neuroscience Society (Neuroscience 2012), Nagoya, Sep 2012.

R. Maruyama, et. al., "Detecting cells using non-negative matrix factorization on calcium imaging data", *Neural Network* 55, 11-19, 2014.

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計 1 件)

Takashi Takekawa, Keisuke Ota, Masanori Murayama and Tomoki Fukai, "Spike detection from noisy neural data in linear-probe recordings", *European Journal of Neuroscience*, 39, 1943-1950, doi: 10.1111/ejn.12614, 2014. 査読あり

[学会発表](計 6 件)

Takashi Takekawa, Masaaki Sato, Noriaki Ohkawa, Kaoru Inokuchi, Yasunori Hayashi, Tomoki Fukai, "Sparse MAP inference of cell shapes and spike trains from calcium imaging data with non-negative constraints", 45th annual meeting of Society for Neuroscience (SfN), Chicago, Oct 2015.

Takashi Takekawa, Masaaki Sato, Noriaki Ohkawa, Kaoru Inokuchi, Yasunori Hayashi, Tomoki Fukai, "Automatic inference system of spike trains and cell shapes from large-scale calcium imaging data using non-negative sparse model", Cell Symposia: Engineering the Brain, Chicago, Oct 2015.

竹川高志, 佐藤正晃, 大川宜昭, 井ノ口馨, 林康紀, 深井朋樹, "繰り返し逐次二次計画法によるカルシウムイメージングからの細胞位置およびスパイク時刻の自動検出システム", 第38回日本神経科学大会, 神戸, 2015年7月.

竹川高志, "混合 Normal Inverse Gaussian 分布モデルの変分ベイズ解法", 第17回情報論的学習理論ワークショップ, 名古屋, 2014年11月.

竹川高志, 太田桂輔, 村山正宜, 深井朋樹, "電極が密に配置されたシリコンプローブによるノイズの多い非定常な記録データからのスパイクソーティング", 第37回日本神経科学大会, 横浜, 2014年9月.

竹川高志, "最新の実験技術による大規模データを活かすソーティング技術", 第

53 回日本生体医工学会大会, 仙台, 2014
年 6 月. 招待講演

〔図書〕(計 0 件)

〔産業財産権〕

出願状況(計 0 件)

取得状況(計 0 件)

〔その他〕

ホームページ等

[http://er-web.sc.kogakuin.ac.jp/Profiles/11/
0001061/profile.html](http://er-web.sc.kogakuin.ac.jp/Profiles/11/0001061/profile.html)

6 . 研究組織

(1) 研究代表者

竹川 高志 (TAKEKAWA, Takashi)

工学院大学・情報学部・准教授

研究者番号：50415220

(2) 研究分担者

なし

(3) 連携研究者

なし