

長時間スパイク同定の為のオンラインスペクトラル分類

坂本 航太郎 統計的機械学習研究センター 特任助教

概要

長時間の多細胞同時記録実験におけるスパイク同定は重要な工程であるが、既存手法においては熟練した実験解析者による修正が必要であるなどの問題が残っていた。本研究では介入を必要としないより効率的なスパイクソーティング法としてオンラインスペクトラル分類法を提案した。

背景

スパイクソーティング

- 神経細胞は活動電位の一過性上昇であるスパイクによって情報伝達を行っており、脳組織に刺入した多重極電極により記録した多数の神経細胞のユニット活動（細胞外に生じる微弱な電気的変化、細胞外記録）を分離する解析をスパイクソーティング (spike sorting) と呼ぶ。
- 記録される波形は神経細胞と電極との間の相対的な距離に依存するため、また同一の神経細胞から出力される波形は一樣であることが知られているため、スパイクソーティングは、k-means法、混合ガウスモデル、スペクトラルクラスタリングなどの既存のクラスタリング手法を用いて可能である。
- スパイクソーティングには、(1) 信号のフィルタリング、(2) スパイクの検出、(3) 特徴量抽出、(4) クラスタリングというステップがあるが、ここでは (3)、(4) のみを取り扱う。

長時間記録時の問題点

短時間の記録についてのスパイクソーティングは前述の通り既に様々な手法が提案されているが、長時間記録のスパイクソーティングは以下のような問題点がある。長時間の神経細胞の活動を記録し解析することは睡眠や老化等の長期的かつ複雑な現象を解明するためにもきわめて重要である。

- 細胞と電極との相対的な距離が動いてしまい、クラスターがドリフト
- 長時間記録では検出されるサンプルに占めるノイズが増える
- 既存手法のMoDT法 [Shan, et al., 2017.] は全体のスパイク列を複数の短時間のバッチに区切ってその中で混合t分布をフィッティングしているが、(1) 我々の扱いたいデータでは望まれるクラスタリング結果を得られなかった、(2) ラベルの同一性に議論の余地がある、(3) 計算時間がかかる、等の問題がある
- 前回MoDT法を改良し、複数回MoDT法を施しながら解析者がその結果を修正することを繰り返す”Expert-in-the-Loop”の手法を提案している (hcMoDT) [Matsumoto, et al., 2019.]
- この手法(hcMoDT)においても、(1) 解析者の介入が必要、(2) ラベルの同一性の問題、(3) 得られた結果の精度感、等の未解決の問題点があり、効率的なよりよい手法が望まれている

提案手法

区間をスライドさせる

全体の時系列を区間に分け、区間をズラして逐次クラスタリングを実行する。時系列データ x_1, x_2, \dots と表す。 s 番目の区間 (時系列長 τ) のデータ点 $x_{s\delta+1}, x_{s\delta+2}, \dots, x_{s\delta+\tau}$ について類似度行列 $W^{(s)} = (w_{ij}^{(s)}) \in R^{\tau \times \tau}$, $s = 0, 1, \dots, n$ を計算する。区間を δ サンプル分前進させると $(s+1)$ 番目のデータ点は $x_{(s+1)\delta+1}, x_{(s+1)\delta+2}, \dots, x_{(s+1)\delta+\tau}$ となる。重なっている区間は $x_{(s+1)\delta+1}, x_{(s+1)\delta+2}, \dots, x_{s\delta+\tau}$ である。よって $w_{ij}^{(s)}$ for $i, j = \delta+1, \delta+2, \dots, \tau$ と $w_{ij}^{(s+1)}$ for $i, j = 1, 2, \dots, \delta$ は同じ成分となる。これをもとに $W^{(s+1)}$ における成分は $w_{ij}^{(s)}$ における $w_{ij}^{(s)}$ の $i, j = \delta+1, \delta+2, \dots, \tau$ 成分は同一として新しいデータ点のみ計算することとする。”通常の”ス

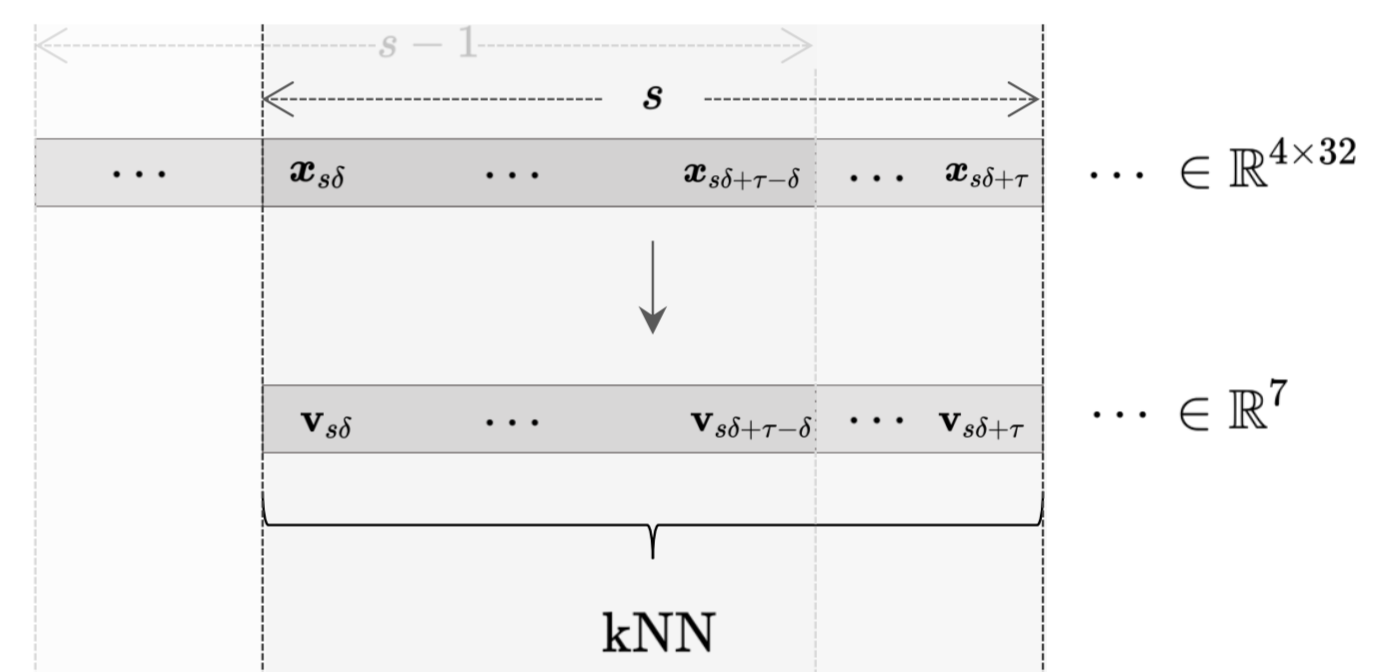
ペクトラルクラスタリングと同様に対角行列 $D^{(s)} = (d_{ij}^{(s)}) \in R^{\tau \times \tau}$ where $d_{ii}^{(s)} = \sum_{j=1}^{\tau} w_{ij}^{(s)}$, $L^{(s)} = D^{(s)} - W^{(s)}$ を構成する。

近似による効率化

連続する区間の $W^{(s)}$ と $W^{(s+1)}$ についてグラフラシアン固有値問題 $L^{(s)}v^{(s)} = \lambda^{(s)}D^{(s)}v^{(s)}$, $v^{(s)} \neq 0$ と $L^{(s+1)}v^{(s+1)} = \lambda^{(s+1)}D^{(s+1)}v^{(s+1)}$, $v^{(s+1)} \neq 0$ の固有値とそれぞれに対応した固有ベクトルの成分が類似することを期待する。計算上の都合で恒等な固有値問題を解く: $W^{(s)}v^{(s)} = \mu^{(s)}D^{(s)}v^{(s)}$, $v^{(s)} \neq 0$, $\mu^{(s)} = 1 - \lambda^{(s)}$, $s = 0, 1, \dots$ 。ベキ乗法あるいは Krylov-Schur 法を以下の様な初期ベクトルを使用して前の区間のラプラシアン固有ベクトルを利用して次の区間のラプラシアン固有ベクトルを近似する: $u^{(s,0)} = [u_{\delta+1}^{(s-1)}, u_{\delta+2}^{(s-1)}, \dots, u_{\tau}^{(s-1)}, \bar{u}^{(s-1)} \dots \bar{u}^{(s-1)}]^T$ 。

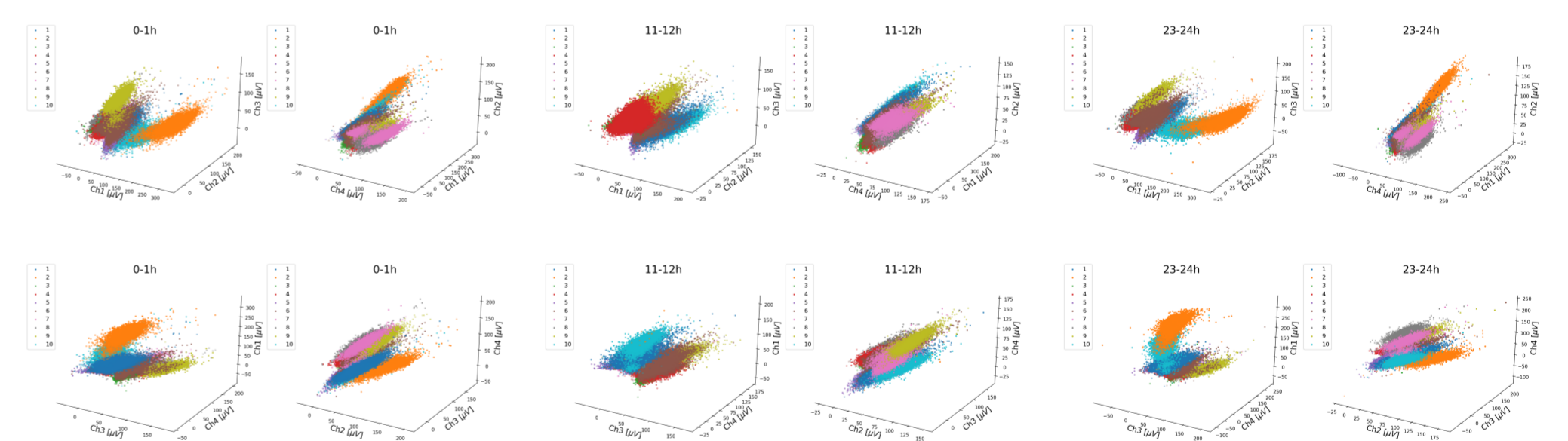
k近傍法による分類

前の区間の固有ベクトルの各成分と次の区間の固有ベクトルをk近傍法によって比較し、距離が近いものに同一のラベルを付与する。以下に示すのがアルゴリズムの概念図である。



評価実験

実際のマウスの細胞外記録のデータを用いて本提案手法の検証を行った。24時間分で4621906のスパイクが検出されおり、各サンプルは4本の電極x32のデータ点=128次元を持つ。10000点を含む区間を1000ずつズラして逐次計算した。以下にクラスタリングの結果を散布図に表示した (各電極で検出されたスパイクの最大の点)。



スパイクソーティングのクラスタリング結果の評価は一般には正解がないため難しいが提案されている評価指標 (ISI, isolation 距離, L-ratio) を計算し、hcMoDTの結果と比べて、使用が期待できる同数のクラスター数を確認した。またhcMoDTのラベリング結果と98.54%一致した。

今後の課題

- 後者2つ評価指標はガウス分布を仮定しているためより良い評価指標が望まれる
- 細胞が死滅するなどしてクラスターが消失するケースに対応する
- 生物学的な知見を用いてクラスタリングの質を上げる

本研究は東北大学松本すみ礼氏、筑波大学Kaspar Vogt先生、保國恵一先生、櫻井鉄也先生との共同研究です。