

## はじめに

神経細胞は自分がつながっている他の神経細胞から一定以上の強さの入力を受け取ると発火し、後続の神経細胞に情報を伝える。このような個々の神経細胞の働きについての理解に比べ、これらの神経細胞の通信が総体として、どの様に情報を表現しているかについては未知の部分が多い。しかし、近年の研究からある順序を伴った複数の神経細胞の発火、つまりシーケンスの繰り返しとその役割の一部を担っていることが示唆されている [1][2][3]。こうした発見で用いられている手法は比較的小規模なデータに対して有用な手法であるが、近年の計測技術の向上に伴い得られるようになってきた大規模なデータに対しての適用は困難である。そこで本研究では、従来手法の問題点を解決するアルゴリズムを提案する。

## 従来手法とその問題点

これまでのシーケンス検出には主にテンプレートマッチング [2][3] と主成分分析 [4] や独立成分分析 [5] を用いた解析手法が用いられてきた。前者は刺激・運動からみて特定のタイミングの複数の神経活動を平均化したテンプレートを作成し、テンプレートと神経活動との相関係数を計算することにより特定の刺激・運動に関連したシーケンスを検出する手法である。一方後者は、神経活動を特定のピンで集計したスパイク行列に独立成分解析等を適用することでシーケンスを検出する。テンプレートマッチングでは観察者が恣意的にテンプレートを作成しなければならない他、シーケンスの伸び縮みに対応できないという問題点がある。またスパイク行列を用いた手法では検出できるシーケンスの統計分布に制限がある、雑音に対してロバストでない等の問題点が存在する。

## 編集類似度を用いた神経活動の比較

本研究では編集類似度を用いた神経活動の比較を行うことにより、シーケンスの検出を目指す。比較の対象となる神経活動は特定の長さの時間窓をスライドさせることにより神経活動を表す行列を切り取ったものを用いた。編集類似度とは2つの文字列を一方の文字列を他の文字列に一致させるのに必要なギャップ文字の挿入、文字の一致を取る等の文字列操作の回数を元に定義する尺度である。本研究では動的計画法 [6] を用いたアルゴリズムを神経活動を表すバイナリの行列に対して適用可能なよう拡張した手法を用いる。動的計画法はある問題を連続する複数の小問題に分割し、小問題の解を利用して元の問題の解を求める手法である。これは2つの神経活動の類似度を比較するために、各々の部分神経活動同士の類似度を計算した結果を用いることに対応する。具体的な計算は、図1左に示すようなテーブルの各マス目の値を左下から右上に埋めていくことで行う。図左の#記号は空の部分神経活動を表している。なお、図右に示す2つの行列は例として比較する神経活動であり、 $V = (v_1 \dots v_5)$ 、 $U = (u_1 \dots u_5)$  に対応する列ベクトルが特定の時刻の各神経細胞の発火の様子を表している。このベクトルの行は神経細胞の番号に対応しており、ベクトル中の0, 1はそれぞれ発火の有無を表現している。また、ここで各マスの計算に用いた更新式は

$$s_{i,j} = \max \begin{cases} s_{i-1,j} \\ s_{i,j-1} \\ s_{i-1,j-1} + \mathbf{v}_i \cdot \mathbf{u}_j \end{cases} \quad (1)$$

である。 $s_{i,j}$  はテーブル中の  $i$  行  $j$  列目の類似度を表す。この更新式を図示した編集グラフを図1に示す。また図1に示すようにテーブルの右上の値に至る道筋を逆に辿ることにより、2つの神経活動間での共通成分を求めることができる。図右の神経活動ベクトル中の赤く色付けされた発火が共通成分である。

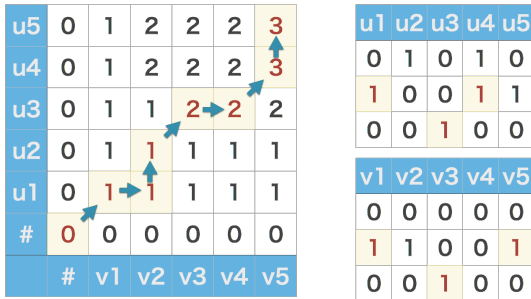


図1: 編集類似度計算の例

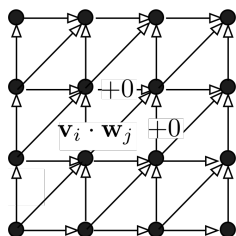


図2: 図1に対応する編集グラフ

図1の例はギャップ文字の挿入に対して罰則を設けず、文字の一致に対して対応するベクトルの内積分の加算を行う単純な文字列操作への重み付けによるものであった。しかし、このような単純な重み付けをタイムウィンドウを用いて切り出してきた神経活動データに対して適用した場合、シーケンスの検出に必要な複数神経細胞間の時間差を正しく考慮できないという問題点が存在した。本研究ではこのような問題点を解決するべくギャップ文字の挿入に対して指数関数的に罰則を科す手法を提案した。これは後藤の手法 [7] を任意の関数に対して適用可能な形に拡張したものである。さらに、この導入により生ずる、シーケンスの出現位置に依存して意図しない罰則が課されるという問題点を解決するため、局所配列比較の手法 [8] を取り入れた。実際に本研究で用いている更新式は以下の通りである。

$$s_{i,j}^{\downarrow} \leftarrow \begin{cases} 1 & s_{i-1,j} - \exp(a) \geq s_{i-1,j-1}^{\downarrow} - \exp(as_{i-1,j}^{\downarrow}) \\ s_{i,j-1}^{\downarrow} + 1 & s_{i-1,j} - \exp(a) < s_{i-1,j-1}^{\downarrow} - \exp(as_{i-1,j}^{\downarrow}) \end{cases} \quad (2)$$

$$s_{i,j}^{\rightarrow} \leftarrow \begin{cases} 1 & s_{i,j-1} - \exp(a) \geq s_{i,j-1}^{\rightarrow} - \exp(as_{i,j-1}^{\rightarrow}) \\ s_{i-1,j}^{\rightarrow} + 1 & s_{i,j-1} - \exp(a) < s_{i,j-1}^{\rightarrow} - \exp(as_{i,j-1}^{\rightarrow}) \end{cases} \quad (3)$$

$$s_{i,j} \leftarrow \max \begin{cases} 0 \\ s_{i-1,j-1}^{\downarrow} - \exp(as_{i-1,j}^{\downarrow}) \\ s_{i,j-1}^{\rightarrow} - \exp(as_{i,j-1}^{\rightarrow}) \\ \mathbf{v}_i \cdot \mathbf{w}_j - \text{base} \end{cases} \quad (4)$$

(5)

ここで、 $s_{i,j}^{\downarrow}, s_{i,j}^{\rightarrow}$  はそれぞれ、 $i$  行  $j$  列目のマスにおける  $V, W$  に対しての最適なギャップ文字の挿入回数を表す。また  $\text{base}$  はシーケンスが存在しない、すなわちすべての神経細胞が独立に発火していると仮定した場合における  $\mathbf{v}_i \cdot \mathbf{u}_j$  の期待値である。この更新式に対応する編集グラフを図1に示す。左右のグラフがそれぞれ  $s_{i,j}^{\downarrow}, s_{i,j}^{\rightarrow}$  に、中央のグラフが  $s_{i,j}$  に対応している。

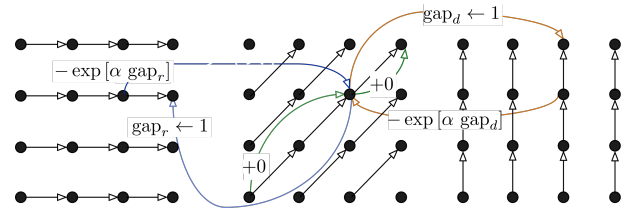


図3: 編集グラフ

## 手法の概説

前節では、異なる時間窓から抽出された2つの神経活動を比較するための類似度を提案した。この類似度が、正しく似通った発火列に対して高い類似度を与えているならば、各神経活動に対応する点は距離空間上での近接点となる。この場合、検出した繰り返し出現するシーケンスはこの空間中において密に分布している点の集合に対応する。図1に模式図を示す。ただし、本来この空間を正しく構成するためには任意のペアの類似度を解散する必要があるが、計算時間の観点から [9] を用いて類似度が高くなりうるペアのみを比較している。こうした密度によって定義されるクラスタを抽出する手法として本研究では OPTICS [10] 及び OPTICS 処理結果からのクラスタの自動抽出手法 [11] を用いた。さらに、このアルゴリズムは1つの時間窓に対して複数種類のシーケンスのみが含まれる状況では異なるクラスタを同一視してしまうという問題が存在するため、ここで得られた結果をグラフ上からのコミュニティ検出の手法である COPRA [12] で更に処理することによりクラスタを抽出している。なお、OPTICS, COPRA にはクラスタリングの結果を制御するパラメタがいくつか存在するが、これは評価指標 [13] が最良となる組み合わせを用いている。また、ここで得られたクラスタ中に存在する発火列を抽出するため、[14] を用いた。このアルゴリズムにより抽出された発火列が検出された繰り返しシーケンスである。

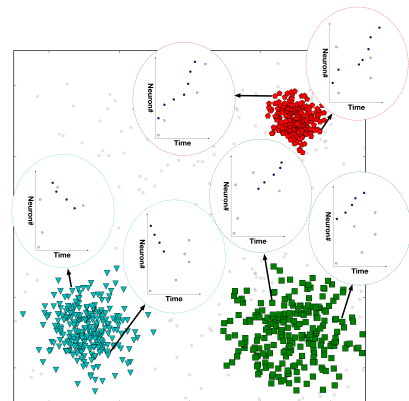


図4: 提案手法の概略図

## 人工データを用いた提案手法と既存手法の比較

人工データを用いて独立成分解析を用いた手法 [15] との比較を行った。データは 100 個の神経細胞を 1Hz でポアソン発火させたスパイクデータに対して、シーケンスを埋め込むことで生成した。埋め込んだシーケンスは 3 種類であり、神経細胞をランダムに選択し、ランダムなタイミングで発火させた。また各シーケンスの時間幅を 0.1 秒から 0.9 秒まで変化させつつクラスタリングを行った。なお、シーケンス中の各発火には  $\pm 10(\text{ms})$  のジッターが導入してある。結果の評価には下式の F 値を用いた。 $n_{i,j}$  はクラスタリングアルゴリズムが生成した  $k$  番目のクラスタにおいて  $m$  番目のクラスに属するデータの個数である。

$$\text{Purity} = \frac{1}{N} \sum_{i=1}^L \max n_{i,j} \quad (6)$$

$$\text{InversePurity} = \frac{1}{N} \sum_{i=1}^K \left[ \frac{\sum_{j=1}^L n_{i,j}}{\sum_{i=1}^K n_{i,j}} \max_j (n_{i,j}) \right] \quad (7)$$

$$F = \frac{2 \cdot \text{Purity} \cdot \text{InversePurity}}{\text{Purity} + \text{InversePurity}} \quad (8)$$

各時間幅での F 値の推移を図 2 に示す。

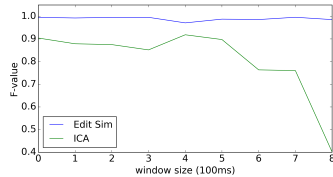


図 5: 既存手法と提案手法の比較

## 提案手法の実データへの適用

提案手法をラットの海馬 CA1 から記録されたデータ [16] に対して適用した。これはラットが線形迷路を移動中に計測されたものであり、22 個の神経細胞から約 20 分間記録されたものである。なお、このデータは Collaborative Research in Computational Neuroscience [17] にて公開されているものである。図 2 に検出されたコミュニティ (クラスタ) のうちラットの行動との関連が明らかであった 3 種を示す。図の上半分が各コミュニティの出現時刻であり、各時刻におけるラットの迷路上での位置を表している。図 2 に [14] を用いて検出したコミュニティ内を代表する発火を示す。この図において高い値を記録している神経細胞番号及び時刻はコミュニティ内で頻出している発火成分である。図 2 には実際のコミュニティ内の神経活動の例を示した。

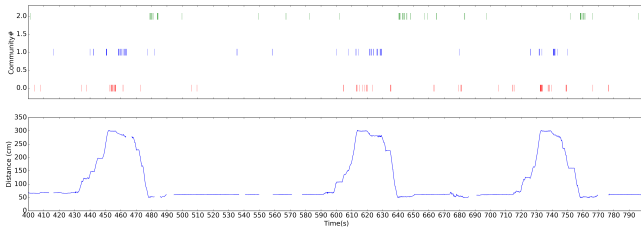


図 6: ラットの運動にロックしたシーケンス

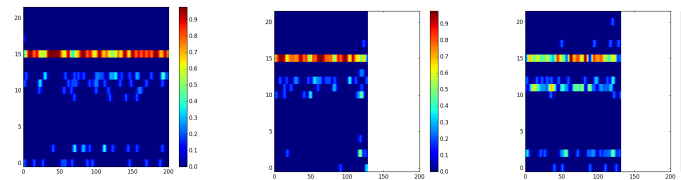


図 7: 各コミュニティを代表する神経活動

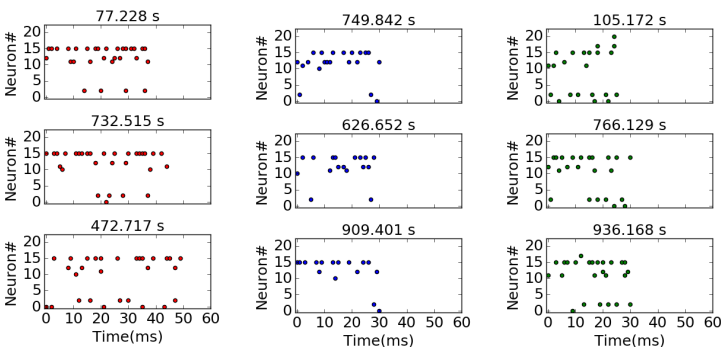


図 8: 検出されたシーケンス

## まとめと考察

本研究では編集類似度を用いた新しい神経活動の類似性比較手法を提案した。さらに、提案尺度と神経活動データからなる距離空間において、密度ベースクラスタリングアルゴリズムである OPTICS とコミュニティ検出手法の COPRA を組み合わせることにより、ロバストなクラスタリング結果が得られることをシミュレーションデータを用いた既存手法との比較により示した。このような性能差の原因は、既存手法では特定の時間長さのビンの中に存在する発火をすべて集計した情報を元にシーケンスの検出を行うため、検出したいシーケンスの時間構造が潰れてしまう、信号に対する雑音の比率が高くなりすぎる等の問題点が存在するためであると考えられる。この様子を図 2 に示した。一方提案手法では図 2 に示すようにこうした平均化を行わないため性能が悪化しないものと考えられる。また、ラットの海馬から記録された実データを提案手法に適用したところ、3 種類の行動に関連したシーケンスの検出に成功した。これらのシーケンスは発火している神経細胞の種類は非常に似通っているものの、それぞれの発火タイミングに差異が見られるものであった。こうしたシーケンスは既存手法での検出が困難なものであり、提案手法の有用性を示すものである。

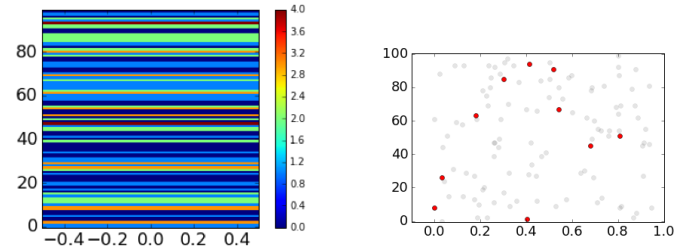


図 9: 既存手法の問題点

図 10: 提案手法で扱っているデータ

## 今後の課題

今後の課題として、まず今回の海馬データへの適用の結果得られた他の、今回用いたラットの行動のデータとの関連性が不透明クラスタに対する考察が挙げられる。これらのクラスタに対応するデータをレポジトリ中から発見することを目指す。第二に、提案手法において手動での決定が必要となっているギャップ導入時の指数関数的に増大する罰則項の重みを自動的に決定する手法の提案を目指す。具体的には神経活動データ中にシーケンスが全く存在しない場合、すなわち各神経細胞が独立に発火しているという仮定のもとでの編集類似度の期待値を用いた手法を検討する。また、本手法は比較的複雑な計算を多用するため長時間に渡る計算が必要であるという問題点を解決するため、現在行っているようなアルゴリズムレベルでの高速化のみでなく GPGPU 等を用いたハード面での高速化を検討する。

## 参考文献

- M Abeles, H Bergman, E Margalit, and E Vaadia. Spatiotemporal firing patterns in the frontal cortex of behaving monkeys. *Journal of neurophysiology*, 70(4):1629–1638, October 1993.
- Edgar J Bermudez Contreras, Andrea Gomez Palacio Schjetnan, Arif Muhammad, et al. Formation and Reverberation of Sequential Neural Activity Patterns Evoked by Sensory Stimulation Are Enhanced during Cortical Desynchronization. *Neuron*, 79(3):555–566, August 2013.
- David R Euston, Masami Tatsuno, and Bruce L McNaughton. Fast-forward playback of recent memory sequences in prefrontal cortex during sleep. *Science*, 318(5853):1147–1150, November 2007.
- Adrien Peyrache, Karim Benchenane, Mehdi Khamassi, et al. Principal component analysis of ensemble recordings reveals cell assemblies at high temporal resolution. *Journal of Computational Neuroscience*, 29(1-2):309–325, June 2009.
- M Laubach, M Shuler, and M A Nicolelis. Independent component analyses for quantifying neuronal ensemble interactions. *Journal of Neuroscience Methods*, 94(1):141–154, December 1999.
- S B Needleman and C D Wunsch. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of molecular biology*, 48(3):443–453, March 1970.
- O Gotoh. An improved algorithm for matching biological sequences. *Journal of molecular biology*, 162(3):705–708, December 1982.
- T F Smith and M S Waterman. Identification of common molecular subsequences. *Journal of molecular biology*, 147(1):195–197, March 1981.
- Edith Cohen, Mayur Datar, Shinji Fujiwara, et al. Finding interesting associations without support pruning. *Knowledge and Data Engineering, IEEE Transactions on*, 13(1):64–78, 2001.
- Mihael Ankerst, Markus M Breunig, Hans-Peter Kriegel, and Jörg Sander. OPTICS: Ordering Points To Identify the Clustering Structure. *ACM Sigmod Record*, 28(2):49–60, June 1999.
- Jörg Sander, Xuejin Qin, Zhiyong Lu, et al. Automatic Extraction of Clusters from Hierarchical Clustering Representations. In *Advances in Knowledge Discovery and Data Mining*, pages 75–87. Springer Berlin Heidelberg, Berlin, Heidelberg, April 2003.
- Steve Gregory. Finding overlapping communities in networks by label propagation. *New Journal of Physics*, 12(10):103018, October 2010.
- V Nicosia, G Mangioni, V Carchiolo, and M Malgeri. Extending the definition of modularity to directed graphs with overlapping communities. *Journal of Statistical Mechanics: Theory and Experiment*, 2009(03):P03024, March 2009.
- G J Barton and M J Sternberg. A strategy for the rapid multiple alignment of protein sequences. Confidence levels from tertiary structure comparisons. *Journal of molecular biology*, 198(2):327–337, November 1987.
- Vitor Lopes-dos Santos, Sidarta Ribeiro, and Adriano B L Tort. Detecting cell assemblies in large neuronal populations. *Journal of Neuroscience Methods*, 220(2):149–166, November 2013.
- Kenji Mizuseki, Anton Sirota, Eva Pastalkova, and György Buzsáki. Theta oscillations provide temporal windows for local circuit computation in the entorhinal-hippocampal loop. *Neuron*, 64(2):267–280, October 2009.
- K Mizuseki, A. Sirota, E Pastalkova, et al. Multiple single unit recordings from different rat hippocampal and entorhinal regions while the animals were performing multiple behavioral tasks. *CRCNS.org* (<http://dx.doi.org/10.6080/K09G5JR2>), 2013. Online; accessed 1-February-2016.