

脳の情報処理機構の解明に向けたデータマイニング技術の開発

Data mining for understanding information processing in the brain



小林 亮太 (Ryota KOBAYASHI, Ph.D.)

国立情報学研究所 情報学プリンシプル系 助教

総合大学院大学 複合科学研究科 情報学専攻 助教

(Assistant Professor, Principles of Informatics Research Division, National Institute of Informatics; Department of Informatics, Graduate University for Advanced Studies (Sokendai))

日本神経回路学会 日本神経科学学会 日本物理学会 Society for Neuroscience

受賞: EPFL-Brain Mind Institute Neuron Modeling Award (2008) INCF Prize (2009) 船井情報科学振興財団 船井研究奨励賞 (2010)

研究専門分野: ニューラルネットワーク 計算論的神経科学 機械学習 データマイニング

あらまし 脳は多数の神経細胞が神経回路を構成することにより、外界の認識、運動制御、意思決定のような高度な情報処理を行う。神経回路がどのような仕組みで情報処理を行っているかを理解することは脳科学における重要な目標の1つである。米国では、2013年から脳計測の技術革新を目標とした10年間のプロジェクト“Brain Initiative”が始まり、計測されたデータから重要な情報を自動的に抽出するアルゴリズムの開発が望まれている。本稿では、脳の情報処理機構を調べるための時系列データのマイニング（解析）アルゴリズムについて紹介する。信号処理、機械学習技術と神経細胞のモデリング技術を融合させることにより、直接測定することが困難である脳の状態（神経細胞の入力信号、神経回路構造）を測定可能なデータから同定することが可能になりつつある。

1. はじめに

脳は多数の神経細胞が神経回路を構成することにより、外界の認識、運動制御、意思決定のような高度な情報処理を行う。神経回路がどのような仕組みによって情報処理を実現しているかを知ることは脳科学における重要な目標の1つである。

脳は多数の神経細胞（ヒト：1,000億、マウス：数1,000万、ミツバチ：100万）から構成されているにもかかわらず、神経細胞集団の詳細な活動を計測する事は困難であった。しかし、多電極アレイ、カルシウムイメージングなどの計測技術の発展により神経細胞集団の計測が可能になりつつある。2013年から米国において脳計測の技術革新を目標とした10年のプロジェクト“Brain Initiative”[1]が始まり、計測技術が今後ますます発展することが期待される。一方、センサー、画像などのデータから有用な情報を抽出する技術（データ活用技術）を開発することは情報学における主要な課題の1つであり、信号処理、機械学習、データマイニング等の分野で研究が進められてきた。大きなノイズを含み、システムの状態が変動する脳計測データの解析技術を研究することは、データ活用技術を発展させる上で興味深い対象の一つである。また、このような研究を進めることで脳科学に貢献できるだけでなく、脳を参考にした新しい人工知能技術の開発に応用されることも期待できる。そこで、筆者らは脳計測データのデータマイニング（データ解析）技術の開発に着手した。

脳計測データのデータマイニング技術の開発は、欧米で活発に進められている。米国では、マサチューセッツ工科大学、コロンビア大学、カーネギーメロン大学などのグループが中心となり、統計学と機械学習技術の融合によるデータ解析技術の開発および実データ解析への応用が進められている[2][3]。欧州では、スイス連邦工科大学ローザンヌ校（EPFL）のグループが中心となり、脳活動、記憶の形成についての数理モデルの研究が進められている[4]。筆者らは実験で計測される神経細胞の活動を高精度に再現する数理モデルの開発を進めてきた[5][6][7]。特に、筆者らが開発したMAT（Multi-timescale Adaptive Threshold）モデルは、2007～2009年にかけてEPFLにおいて開催され

脳の情報処理機構の解明に向けたデータマイニング技術の開発

Data mining for understanding information processing in the brain

た神経細胞活動の予測コンテストにおいて優勝を果たした[8][9]。本稿では、筆者らが開発した脳の情報処理を調べるための時系列データの解析アルゴリズムについて紹介する。信号処理、機械学習技術と神経細胞のモデリング技術を融合させることにより、直接測定するのが困難である脳の状態を推定することが可能になりつつある。

本稿の構成を説明する。2章では、開発されたアルゴリズムの概要を理解するために必要な神経細胞、シナプスについての基本知識を説明する。3章では、神経細胞の入力信号を推定するアルゴリズム[10][11]について紹介する。4章では、神経回路構造を推定するアルゴリズム[12]を紹介する。5章ではまとめと将来展望について述べる。

2. 脳の情報処理機構の解明に向けて

ヒトの脳は1,000億個以上の神経細胞から構成されている。神経細胞は、脳の情報処理における基本的な素子であると考えられている。神経細胞は他の神経細胞からの入力電流を電気パルス（スパイク）に変換する。スパイクは電流に変換され、電流はシナプスと呼ばれる電線のようなものを流れて他の神経細胞へ送られる。本章では、3, 4章の内容を理解するために必要となる神経細胞、シナプスについて解説する。

2.1 神経細胞

神経細胞は数千~一万もの神経細胞からの入力電流（シナプス電流）をスパイクと呼ばれる電気パルスに変換する（図1）。スパイクはシナプス電流に変換されて他の神経細胞へ送られる。一方、多くの実験により、スパイクの頻度やタイミングは動物の認知、行動、情動に関連することが示されてきた。スパイクは脳内の情報を表現していると考えられているが、スパイクがどのように脳内の情報を表現しているのかということにはよくわかっていない。

脳活動のシミュレーションや計測されたデータを詳細に解析するため、神経細胞の数理モデルが開発されている。ここでは、代表的な神経細胞モデルである積分発火モデルとホジキンハクスレーモデルを紹介する。積分発火モデルでは、神経細胞の膜電位 V は RC 回路で記述される（図2A）。

$$C \frac{dV}{dt} = -gV + I(t), \quad \dots\dots (1)$$

ただし、 C は膜容量、 g はコンダクタンス（抵抗の逆数）、 $I(t)$ は入力電流である。

このモデルでは、入力電流を強くすると神経細胞がスパイクを生成するという実験事実を再現できない。そこで、閾値 θ を導入し、膜電位 V が閾値を越えればスパイクを生成させ、膜電位を $V_r (< \theta)$ にリセットする。

$$V(t) > \theta \rightarrow \text{時刻} t \text{ にスパイクを生成, } V(t) \rightarrow V_r \quad \dots\dots (2)$$

式(1), (2)は積分発火モデルと呼ばれ、脳の理論的研究において広く用いられている。近年、拡張された積分発火モデルが神経細胞のスパイクパターンを高精度に再現できることが示されている[4][6][7][8]。過去のスパイク履歴に応じて変動する閾値を導入した MAT (Multi-timescale Adaptive Threshold) モデル[7]と呼ばれる神経細胞モデルは、神経細胞が生成する90%のスパイクを正確に再現できる。

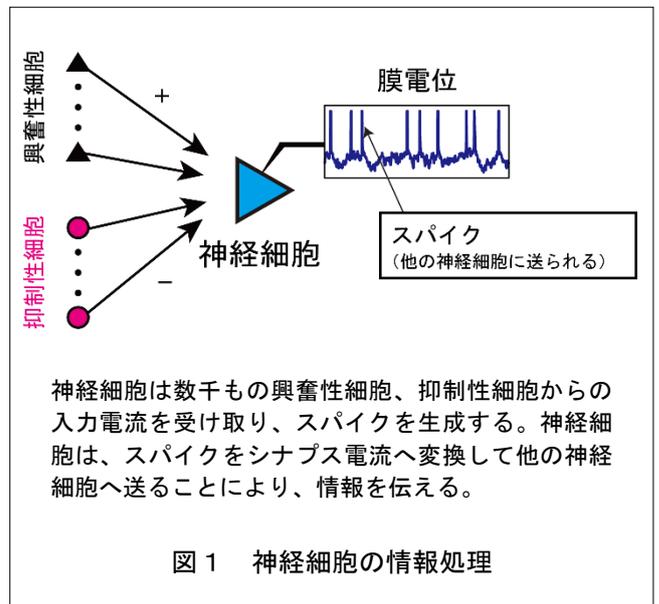
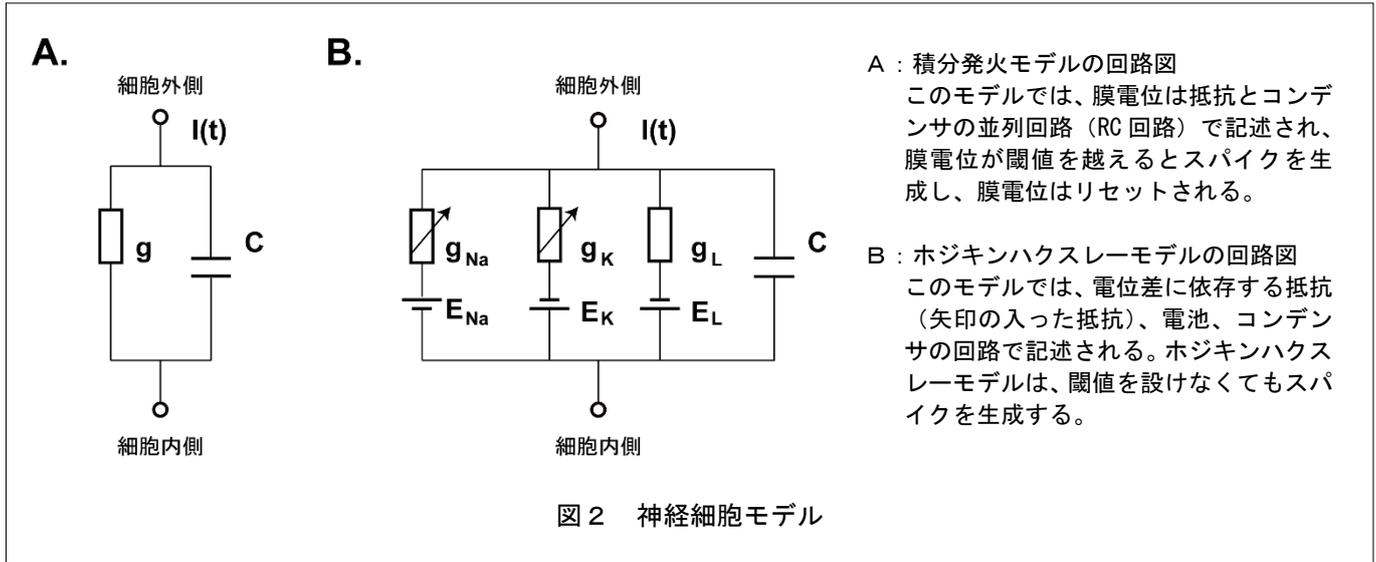


図1 神経細胞の情報処理

積分発火モデルでは、閾値という現象論的なパラメータを導入することによりスパイクを生成する神経細胞モデルを構築した。一方、スパイクを生成させる生理学的なメカニズムを明確にした数理モデルとしてホジキンハクスレーモデル[13]が挙げられる（図2B）。

脳の情報処理機構の解明に向けたデータマイニング技術の開発

Data mining for understanding information processing in the brain



ホジキンとハクスレーは膜電位に依存して開閉する Na、K イオンチャネルを組み込むことにより神経細胞の電気生理学的な性質を再現した。

$$C \frac{dV}{dt} = -g_{Na} m^3 h (V - E_{Na}) - g_K n^4 (V - E_K) - g_L (V - E_L) + I(t),$$

$$\frac{dm}{dt} = \alpha_m(V)(1 - m) - \beta_m(V)m,$$

$$\frac{dh}{dt} = \alpha_h(V)(1 - h) - \beta_h(V)h,$$

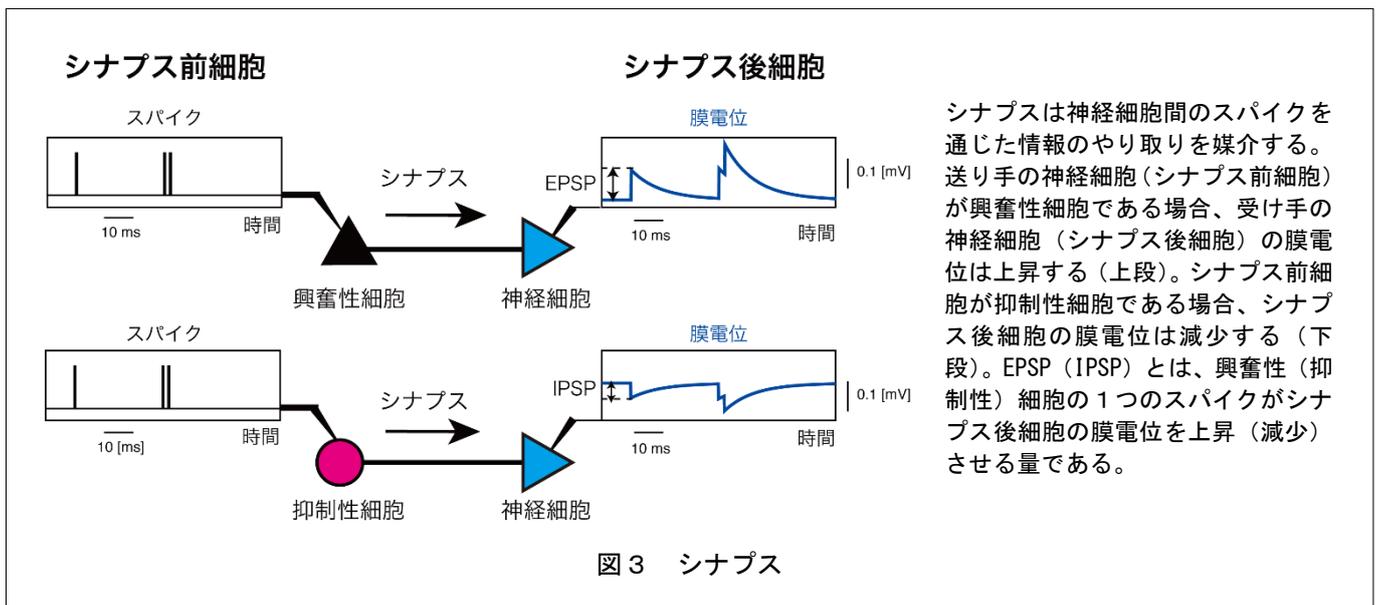
$$\frac{dn}{dt} = \alpha_n(V)(1 - n) - \beta_n(V)n,$$

ただし、 m, h, n はイオンチャネルの状態を表す変数

(ゲート変数)、 E_{Na}, E_K, E_L は各イオンの平衡電位である。ホジキンハクスレーモデルは積分発火モデルと全く異なる数理モデルのように見えるが、ホジキンハクスレーモデルの入出力特性は積分発火モデルで近似できることが示されている[4][6]。

2. 2 シナプス

神経細胞は「シナプス」と呼ばれる電線のようなものを用いて情報のやり取りを行う (図 3)。シナプスでつながった 2 つの神経細胞のうち、送り手の細胞をシナプス前細胞、受け手の細胞をシナプス後細胞と呼ぶ。



脳の情報処理機構の解明に向けたデータマイニング技術の開発

Data mining for understanding information processing in the brain

つながった2つの神経細胞のうち、送り手の細胞をシナプス前細胞、受け手の細胞をシナプス後細胞と呼ぶ。神経細胞は、シナプス後細胞の膜電位を増加させる興奮性細胞、シナプス後細胞の膜電位を減少させる抑制性細胞の2種類からなる。興奮性(抑制性)細胞の1つのスパイクがシナプス後細胞の膜電位を上昇(減少)させる量をEPSP(IPSP)と呼ぶ。シナプスが電線と異なる点として、情報が双方向に伝わるとは限らないことが挙げられる。神経細胞間に少なくとも一方向のシナプスがつながっている確率は10%、双方向にシナプスでつながっている確率は4%という報告がある[14]。脳は、神経細胞を頂点(ノード)、シナプスを枝(エッジ)とするグラフと考えることもできる。神経細胞間のシナプス構造(グラフ構造)を同定することは脳の情報処理を理解する上で非常に重要である。4章では、多数の神経細胞から計測されたスパイクデータからシナプス結合を推定する方法について説明する。

3. 神経細胞の入力信号の推定

3.1 問題設定

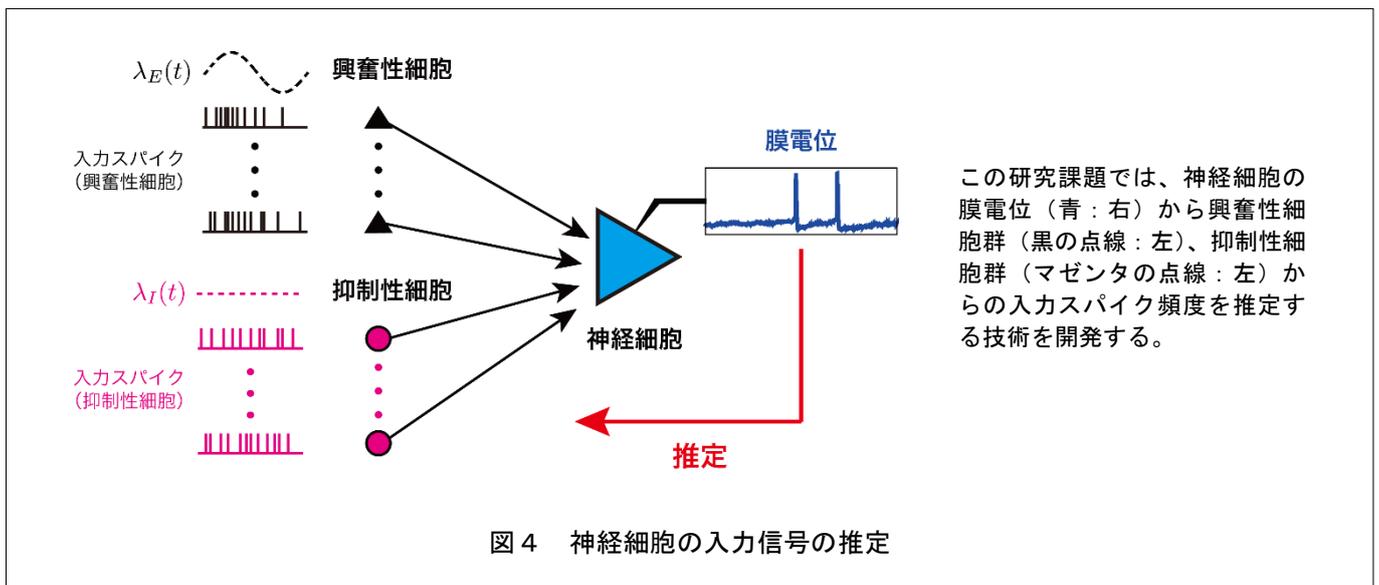
脳内の神経細胞の入力電流は多数の興奮性、抑制性細胞からの電流の総和である。神経細胞が興奮性細胞群から受け取る入力スパイク頻度(興奮性入力)、抑制性細胞群から受け取る入力スパイク頻度(抑制性入力)を調べることは、脳の情報処理機構を理解するために

重要である。例えば、聴覚野では興奮、抑制性細胞群が同期的に音に応答することが発見された[15]が、音の認識の詳細なメカニズムは明らかではない。一方、現在の実験技術では神経細胞の興奮性、抑制性入力を直接計測することは不可能である。そこで、筆者らは膜電位データから神経細胞の興奮、抑制性入力を推定する問題(図4)について考えた。

3.2 推定アルゴリズム

神経細胞の入力信号を推定するアルゴリズムは2つの手順からなる(図5A)。まず、神経細胞の入力電流の平均、分散を膜電位から推定する。この問題は不良設定問題になってしまい推定できない。計測時間を T (秒)、サンプリング間隔 h (秒)で計測を行うとしよう。この場合、計測されるデータ数は $N(= T/h)$ 個、推定すべきデータ数は $2N$ 個となり、推定データ数が計測データ数より多くなるため推定結果は1つに定まらない。そこで、神経細胞の入力電流の平均、分散は急激に変動しないことを仮定し、経験ベイズ法を用いて推定を行った。次に、推定された入力電流の平均、分散から興奮、抑制性入力を見積る。

開発したアルゴリズムの有効性を検証するため、神経細胞モデルのシミュレーションデータを解析した。シミュレーションデータを解析することによりアルゴリズムの推定誤差を評価できる。本アルゴリズムは、



脳の情報処理機構の解明に向けたデータマイニング技術の開発

Data mining for understanding information processing in the brain

A.

A : アルゴリズムの概要

- 1) シナプス電流の平均 $\mu(t)$, 分散 $\sigma^2(t)$ の推定。

$$\frac{dV}{dt} = -g_L(V - E_L) + \mu(t) + \sigma(t)\xi(t)$$

- 2) 興奮性入力 $\lambda_E(t)$, 抑制性入力 $\lambda_I(t)$ の推定。

$$\mu(t) = a_E\lambda_E(t) - a_I\lambda_I(t)$$

$$\sigma^2(t) = a_E^2\lambda_E(t) + a_I^2\lambda_I(t)$$

B : アルゴリズムの適用例

神経細胞モデルのシミュレーションから得られた膜電位データに推定アルゴリズムを適用した。上段は膜電位、中段、下段は興奮性入力、抑制性入力である。赤は推定値、灰は真値を表す。

B.

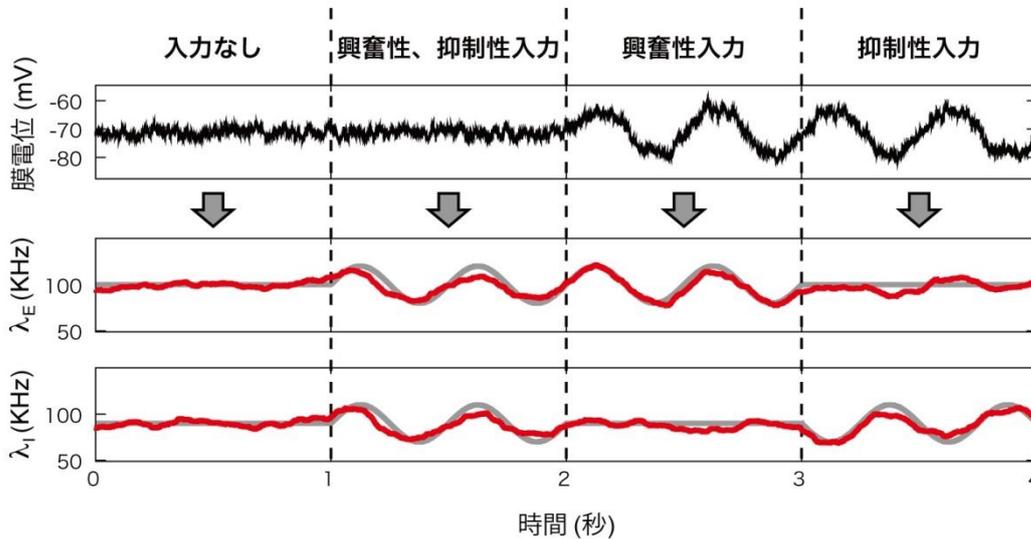


図5 神経細胞の入力信号の推定アルゴリズム

定常な入力と興奮、抑制性の釣り合った入力、興奮性のみ入力と抑制性のみ入力を膜電位データのみから正しく同定できる (図 5B)。詳細は論文[10]を参照して頂きたい。さらに、筆者らはこのアルゴリズムを拡張し、神経細胞の入力信号とイオンチャネル状態(ゲート変数)を同時に推定する手法を開発した[12]。そして、シミュレーションデータ、神経スライスからの計測データ[7]を解析してアルゴリズムの妥当性を検証した。現在は、開発したアルゴリズムを麻酔下モルモットからの計測データ[16]に適用し、神経細胞レベルで音の認識がどのように行われているかについて研究を進めている。

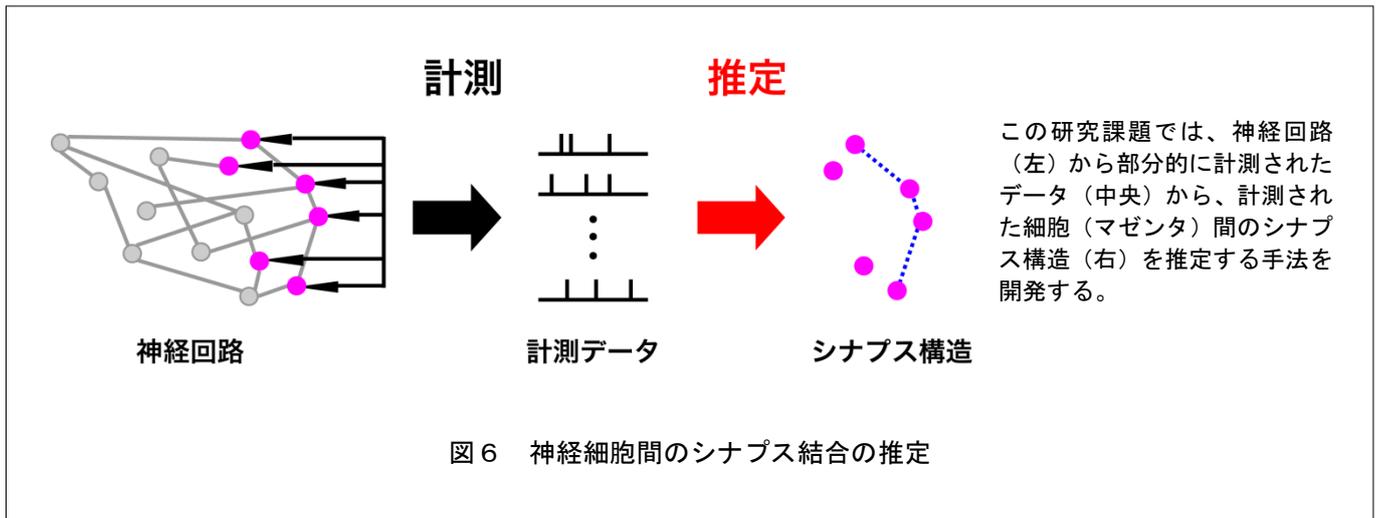
4. 神経シナプス結合の推定

4. 1 問題設定

脳は、多数の神経細胞が回路(神経回路)を構成することにより機能を発揮する。神経回路の構造を明らかにすることは脳科学における重要な研究課題の一つである。神経回路構造はシナプス結合により特徴づけられるため、計測データからシナプス結合を決定することが重要となる。近年、多電極アレイやカルシウムイメージングなどの計測技術の発展により、多細胞の神経活動を記録することが可能となりつつある。本章では、複数の神経細胞から計測されたスパイクデータから計測された細胞間のシナプス結合を推定する研究を紹介する(図6)。

脳の情報処理機構の解明に向けたデータマイニング技術の開発

Data mining for understanding information processing in the brain



スパイクデータとは各細胞のスパイク生成時刻のデータである。シナプス結合の強さ (EPSP, IPSP) はシナプスによって異なるが、ここではシナプス結合の有無のみを推定することを目的とする。

4. 2 推定手法

スパイクデータからシナプス結合を推定する手法と提案手法の検証結果について簡単に紹介する。詳細に興味のある読者は論文[12]を参考にして頂きたい。筆者らは、単一神経細胞の確率モデルである Escape Rate モデル[4][6]を多細胞系に拡張した Coupled Escape Rate モデル (CER) を開発し、最尤法を用いることでシナプス結合を推定した。推定手法の有効性を検証するため、詳細な神経回路モデルのコンピュータシミュレーションを行い、人工的に生成したスパイクデータから推定を行った。神経回路モデルでは、4,096 個の興奮性細胞、1,024 個の抑制性細胞から構成され、神経細胞は 2 次元格子状に配置された。神経回路構造として、局所的なシナプスを多くもつレギュラー回路、少数のランダムなシナプスをもつスモールワールド回路、多くのランダムなシナプスを持つランダム回路のシミュレーションを行った。開発手法 (CER) と既存手法 (Transfer Entropy: TE) の推定精度の比較を行った結果を表 1 に示す。3 つの神経回路において、開発手法は推定精度が向上している事がわかる。また、神経回路構造がランダムに近くなるとシナプス結合の推定は非常に難しくなることがわかる。現在は、

さらなる精度の向上を目指した手法の改良および統計的検定を行うための手法開発を進めている。

表 1 シナプス結合の推定精度 (シミュレーションデータ)

	レギュラー	スモールワールド	ランダム
提案手法	0.74	0.57	0.38
既存手法	0.65	0.37	0.10

開発手法 (CER) と既存手法 (TE) でシナプス結合の推定精度 (Matthews correlation coefficient: 推定値と真値の相関) を比較した。レギュラー、スモールワールド、ランダムの 3 つの構造の神経回路のシミュレーションから得られたスパイクデータからシナプス結合を推定した。

5. まとめ

本稿では、脳の情報処理機構を調べるための時系列データ解析アルゴリズムについて紹介した。まず、膜電位データから変動する興奮、抑制性入力信号を推定するアルゴリズムについて紹介した。このアルゴリズムは、確率微分方程式の推定アルゴリズムと考えることもできる。確率微分方程式は、統計物理学、物理化学、分子生物学、生態学、金融工学分野等において広く用いられている。今後は、脳から計測したデータの解析を進めるとともに、その他の分野への応用についても検討を進めたい。次に、多細胞から計測されたスパイクデータからシナプス結合を推定する手法について

脳の情報処理機構の解明に向けたデータマイニング技術の開発

Data mining for understanding information processing in the brain

で紹介した。脳科学の発展には、計測された大量のデータから重要な情報を抽出する技術を開発することがますます重要となるだろう。今後は、手法の改善を進めるとともに、実データの解析や実験を行っている研究者との議論を並行して進め、脳科学に貢献できるアルゴリズムの開発を進めたい。

謝辞

本稿は、北野勝則、坪 泰宏 (立命館大学)、篠本 滋 (京都大学)、Petr Lansky (チェコ科学アカデミー) との研究成果[10][11][12]をまとめたものである。共同研究者に感謝いたします。

参考文献

- [1] Insel TR, Landis SC, Collins FS. “The NIH brain initiative”, *Science*, 340, 687-688, 2013.
- [2] Brown EN, Kass RE, Mitra PP. “Multiple neural spike train data analysis: state-of-the-art and future challenges”, *Nature Neurosci.*, 7, 456–461, 2004.
- [3] Paninski L, Ahmadian Y, Ferreira DG, Koyama S, Rad RR, Vidne M, Vogelstein J, Wu W. “A new look at state-space models for neural data”, *J. Comput. Neurosci.*, 29, 107–126, 2010.
- [4] Gerstner W, Kistler WM, Naud, R, Paninski, L. “Neuronal Dynamics: From Single Neurons to Networks and Models of Cognition”, Cambridge University Press, 2014.
- [5] Kobayashi R, Shinomoto S. “Predicting spike times from subthreshold dynamics of a neuron”, *Advances in Neural Information Processing Systems* 19 (NIPS), 2006.
- [6] Kobayashi R, Shinomoto S. “State space method for predicting the spike times of a neuron” *Phys. Rev. E*, 75, 011925, 2007.
- [7] Kobayashi R, Tsubo Y, Shinomoto S. “Made-to-order spiking neuron model equipped with a multi-timescale adaptive threshold”, *Front. Comput. Neurosci.*, 3, 9, 2009.
- [8] Jolivet R, Kobayashi R, Rauch A, Naud R, Shinomoto S, Gerstner W. “A benchmark test for a quantitative assessment of simple neuron models”, *J. Neurosci. Methods*, 169, 417–424, 2008.
- [9] Gerstner W, Naud R. “How good are neuron models?”, *Science*, 326, 379–380, 2009.
- [10] Kobayashi R, Shinomoto S, Lansky P. “Estimation of time-dependent input from neuronal membrane potential”, *Neural Comput.*, 23, 3070–3093, 2011.
- [11] Kobayashi R, Tsubo Y, Lansky P, Shinomoto S. “Estimating time-varying input signals and ion channel states from a single voltage trace of a neuron”, *Advances in Neural Information Processing Systems* 24 (NIPS), 2011.
- [12] Kobayashi R, Kitano K. “Impact of network topology on inference of synaptic connectivity from multi-neuronal spike data simulated by a large-scale cortical network model”, *J. Comput. Neurosci.*, 35, 109–124, 2013.
- [13] Hodgkin AL, Huxley AF. “A quantitative description of membrane current and its application to conduction and excitation in nerve” *J. Physiol.*, 117, 500–544, 1952.
- [14] Song S, Sjöström PJ, Reigl M, Nelson S, Chklovskii DB. “Highly nonrandom features of synaptic connectivity in local cortical circuits”, *PLoS Biology*, 3, e68, 2005.
- [15] Wehr M, Zador AM. “Balanced inhibition underlies tuning and sharpens spike timing in auditory cortex”, *Nature*, 426, 442–446, 2003.
- [16] He J. “Slow oscillation in non-lemniscal auditory thalamus”, *J. Neurosci.*, 23, 8281–8290, 2003.

この研究は、平成22年度SCAT研究助成の対象として採用され、平成23～25年度に実施されたものです。