

脳ネットワークのモジュール分解

Modular decomposition of brain networks

岡本 洋^{*1 *2}
Hiroshi Okamoto

^{*1} 富士ゼロックス(株)研究技術開発本部
Research & Technology Group, Fuji Xerox Co., Ltd.

^{*2} 理化学研究所 脳科学総合研究センター[†]
RIKEN Brain Science Institute

Modular structure of brain networks is characterized by overlapping and hierarchy. Here we propose a machine learning algorithm to detect overlapping and hierarchical modular structure of a network. Modular decomposition of the neural network of *C. elegans* and a macaque cortical network is examined by use of this algorithm.

1. はじめに

脳における情報処理はネットワークを舞台として行われる。観測技術の進歩により、脳における構造・機能のネットワークが様々なレベルで明らかにされつつある[1]。*C. elegans* (線虫)の302個の神経細胞の間の配線構造は、電子顕微鏡観察によりすでに完全解明された[2, 3]。トレーシング法による皮質結線構造の解明も進められている[4]。区分(parcel)間の関係をfMRI信号の時間相関から定めることにより、parcelをノードとする機能ネットワークが構成される[1]。本論文では、これらの構造・機能のネットワークを「脳ネットワーク」と総称する。

ネットワーク科学では、ネットワークの中のノードが密に繋がったかたまり部分のことを「コミュニティ」あるいは「モジュール」と呼ぶ(本論文では「モジュール」の呼び方を採用する)。もう少し正確に言うならば、モジュールとは次のようなノードの集団である: この集団の中ではリンクの密度が高く、その外部とのリンクはより疎である。モジュール構造は脳ネットワークを含む実世界における様々なネットワークの一般的性質である。ネットワークから効果的・効率的にモジュール構造を検出するアルゴリズムの開発は、近年のネットワーク科学における中心課題の一つである[5]。

脳ネットワークにおける個々のモジュールには、何らかの情報処理機能が対応するであろう。一つの神経細胞あるいは一つの脳領域は、一般に複数の機能に関与する。従って、脳ネットワークのモジュール構造には重なりがあるはずである。さらに、大脳皮質における初期から高次領野への階層構造は、より大きなモジュールが複数のより小さなモジュールから構成されることを示唆する。このように、脳ネットワークのモジュール構造は重なりと階層で特徴づけられる[1]。

脳における情報処理の過程を明らかにするためには、脳ネットワークのモジュール構造を知ることが重要である。すでに多くのモジュール検出アルゴリズムが提案されている[5]。しかしながら、脳ネットワークの特徴であるところの、重なりと階層のあるモジュール構造に適用できるものは、これまでほとんどなかった[1]。

我々は、マルコフ連鎖のベイズ定式化に基づくことにより、重なりと階層のあるモジュール構造を検出する機械学習アルゴリズムを構築することに成功した。このアルゴリズムを用いて、*C. elegans*の神経細胞ネットワークおよびサル皮質ネットワークのモジュール分解を行った。

2. 方法

マルコフ連鎖とは、ネットワーク状の確率状態遷移である。マルコフ連鎖を、リンクをたどりながらネットワーク上をランダムに歩き回るエージェント(以下、Mr. Xと呼ぶ)の動きにたとえることができる。時刻 t に Mr. X がノード n にいる確率を $p_n(t)$ とする。ただし、 $n=1, \dots, N$ であり、 N はネットワークを構成するノードの総数である。 $p_n(t)$ の時間発展は

$$p_n(t) = \sum_{m=1}^N T_{nm} p_m(t-1). \quad (1)$$

で記述される。ここで、 $\mathbf{T} = (T_{nm}) \equiv (A_{nm} / \sum_{n=1}^N A_{nm})$ は遷移確率行列である。ただし、 $\mathbf{A} = (A_{nm})$ はネットワークの隣接行列であり、 A_{nm} はノード m と n の間のリンクの重みを表す。

ネットワークが複数のモジュールで覆われているとする。Mr. X は、しばらくあるモジュールに留まってその中を歩き回り、あるとき別のモジュールに移ってからしばらくそこに留まってその中を歩き回り、さらにまた別のモジュールに移ってからしばらくそこに留まってその中を歩き回り、... というように振る舞うであろう。Mr. X がモジュール k に滞在しているという条件の下で、彼が時刻 t にノード n にいる確率を $p_n^{(k)}(t)$ とする。すると、モジュール k は確率分布 $\vec{p}^{(k)} = (p_1^{(k)}, \dots, p_N^{(k)})$ で表わされる。そこで、(1)の定常解 $p_n^{(\text{stead})} \equiv \lim_{t \rightarrow \infty} p_n(t)$ を、 K 個のモジュール $\{\vec{p}^{(1)}, \dots, \vec{p}^{(K)}\}$ を用いて

$$p_n^{(\text{stead})} = \sum_{k=1}^K \pi_k p_n^{(k)} \quad (2)$$

と分解することを考える。ただし、 $\sum_{k=1}^K \pi_k = 1$ である。

Mr. X があるノードにいることが観測されたことを N 次元ベクトル $\vec{r}^{(d)}$ で表わす。Mr. X がいたノードに対応する成分のみが 1 であり、他の成分は全て 0 である。 d は観測を区別する番号である。次に、Mr. X の居場所が $\vec{r}^{(d)}$ と観測されたとき、彼がどのモジュールに滞在しているかを表す変数として、 K 次元ベクトル $\vec{z}^{(d)}$ を導入する。Mr. X が滞在するモジュールに対応する成分のみが 1 であり、他の成分は全て 0 である。Mr. X がどのノードに滞在しているかは直接には観測されないで、以下では $\vec{z}^{(d)}$ は潜在変数として扱われる。

Mr. X がモジュール k に滞在しているという条件の下で、時刻 t における彼の居場所が $\vec{r}^{(d)}$ と観測される確率を、多項分布で表わすことができる:

$$P(\vec{r}^{(d)} | \vec{p}^{(k)}(t)) = \prod_{n=1}^N [p_n^{(k)}(t)]^{r_n^{(d)}}. \quad (3)$$

連絡先: 岡本洋, 富士ゼロックス(株)研究技術開発本部,
〒220-8668 神奈川県横浜市みなとみらい6丁目1番.

E-mail: hiroshi.okamoto@fujixerox.co.jp

[†] 客員研究員

$\bar{p}^{(k)}(t)$ の事前分布には、多項分布(3)に共役な Dirichlet 分布

$$P(\bar{p}^{(k)}(t) | \bar{p}^{(k)}(t-1)) \sim \prod_{n=1}^N [p_n^{(k)}(t)]^{\alpha \sum_{m=1}^N T_{nm} p_m^{(k)}(t-1)} \quad (4)$$

を仮定するのが自然である。ここで、 α は Dirichlet 分布の精度を表すパラメタであり、次節で示すように、階層構造の検出で重要な役割を果たす。 $\alpha \rightarrow \infty$ において、(3)はマルコフ連鎖の元の式(1)に一致する。

Mr. X の居場所を D 回観測してデータ $\{\bar{\tau}^{(d)}\} (d=1, \dots, D)$ を得たとする。事後分布はベイズの定理から、次式で与えられる：

$$P(\{\bar{z}^{(d)}\}, \{\bar{p}^{(k)}(t)\} | \{\bar{\tau}^{(d)}\}) \sim \prod_{k=1}^K \left[\prod_{n=1}^N [p_n^{(k)}(t)]^{\beta \sum_{d=1}^D \sum_{k'} \tau_n^{(d)} + \alpha \sum_{m=1}^N T_{nm} p_m^{(k)}(t-1)} \right] \quad (5)$$

$\{\bar{\tau}^{(d)}\} (d=1, \dots, D)$ を学習データととらえ、 $\{\bar{p}^{(k)}\}$ および $\{\pi_k\}$ (すなわち、このデータが得られること背景にあるネットワークのモジュール構造)を定めたい。これは、機械学習における標準手法である EM アルゴリズムの考え方に従って実行できる。

$\ln P(\{\bar{z}^{(d)}\}, \{\bar{p}^{(k)}(t)\} | \{\bar{\tau}^{(d)}\})$ を潜在変数に関する事後分布 $P(\{\bar{z}^{(d)}\} | \{\bar{\tau}^{(d)}\})$ で平均化する：

$$Q \equiv E \left[\ln P(\{\bar{z}^{(d)}\}, \{\bar{p}^{(k)}\} | \{\bar{\tau}^{(d)}\}) \right]_{P(\{\bar{z}^{(d)}\} | \{\bar{\tau}^{(d)}\})} = \sum_{k=1}^K \sum_{d=1}^D \gamma_{dk} \ln \pi_k + \sum_{k=1}^K \sum_{n=1}^N \left[\beta \sum_{d=1}^D \gamma_{dk} \tau_n^{(d)} + \alpha \sum_{m=1}^N T_{nm} p_m^{(k)}(t-1) \right] \ln p_n^{(k)}(t) \quad (6)$$

ただし

$$\gamma_{dk} \equiv P(z_k^{(d)} = 1 | \bar{\tau}^{(d)}) = \frac{\pi_k \prod_{n=1}^N [p_n^{(k)}(t)]^{\beta \tau_n^{(d)}}}{\sum_{k=1}^K \pi_k \prod_{n=1}^N [p_n^{(k)}(t)]^{\beta \tau_n^{(d)}}} \quad (7)$$

である。 $\{\bar{p}^{(k)}\}$ および $\{\pi_k\}$ は、 Q を最大化するものとして、次で定められる：

$$p_n^{(k)}(t) = \frac{\alpha}{\alpha + \beta D_k} \sum_{m=1}^N T_{nm} p_m^{(k)}(t-1) + \frac{\beta D_k}{\alpha + \beta D_k} \frac{1}{D_k} \sum_{d=1}^D \gamma_{dk} \tau_n^{(d)}, \quad (8)$$

$$\pi_k = D_k / D. \quad (9)$$

ただし、 $D_k = \sum_{d=1}^D \gamma_{dk}$ である。

Mr. X は N 個のノードのどれかにいるのであるから、観測結果として得られるパターンは N 通りである。そこで、Mr. X がノード n で観測されたことを、第 n 成分だけが 1 で他の成分は全て 0 の N 次元ベクトル $\bar{\tau}^{(n)}$ で表す。観測回数 D が N に比べて十分大ならば、(7)-(9)式において、 γ_{dk} 、 $\tau_n^{(d)}$ および $\sum_{d=1}^D$ を、それぞれ、 $\tilde{\gamma}_{nk}$ 、 $\tilde{\tau}_n^{(i)}$ および $D \sum_{i=1}^N p_i^{(\text{stead})}$ と置き換えることができる。

ノード n がモジュール k に属する確率は、ベイズの定理から

$$p(k | n) = p_n^{(k)} \pi_k / \sum_{k=1}^K \pi_k p_n^{(k)} \quad (10)$$

となる。一般に、一つの n に対して $p(n | k)$ は複数の k で非ゼロ正となる。これは、提案アルゴリズムが定めるモジュール構造には重なりがあることを意味する。

3. 結果と議論

まず、提案アルゴリズムのモジュール分解性能を、モジュール構造が既知のネットワークを用いて確認した。Lancichinetti et al. の方法[6]を用いて六つのモジュール(これらを正解とみなす)を持つノード数 200 の人工ネットワークを生成した。正解モジュールと検出モジュールとの比較の便宜上、ノード n が属するモジュールを、(10)式を最大化する k で一意に定めた。正解モジュールと検出モジュールは完全に一致した(図 1 左)。提案アルゴリズムを社会ネットワーク分析における有名なベンチマーク課題「Zachary karate club ネットワーク分割」[7]に適用したところ、実際の分割が正確に再現された(図 1 右)。以上で、提案ア

ルゴリズムを用いてネットワークのモジュール分解を実行できることが確認できた。アルゴリズムの特長である重なりと階層のあるモジュール構造の検出については、脳ネットワークを用いて次に示す。

C. elegans の 302 個の神経細胞が構成するネットワーク[1, 2]に提案アルゴリズムを適用した。 α の値が十分大のとき、ネットワーク全体が一つのモジュールと判定された； α の値を小さくしてゆくと、モジュールは次々に分裂していった(図 2)。このように、Dirichlet 事前分布の精度 α を変化させることにより、モジュールの階層構造を導くことができる。さらに、 $\pi_k p_n^{(k)}$ と $\pi_{k'} p_n^{(k')}$ ($k \neq k'$) がともに非ゼロ正となる n が複数存在する(一例を矢印で示す)。すなわち、検出されたモジュールは重なりを持つ。講演では、サル皮質ネットワーク[4]から得た同様な結果についても報告する。

ネットワークから重なりと階層のあるモジュール構造を検出するアルゴリズムを提案し、これを脳ネットワークに適用した。今後は、検出されたモジュール構造が実際の神経情報処理機能にどう対応するかどうかを調べたい。

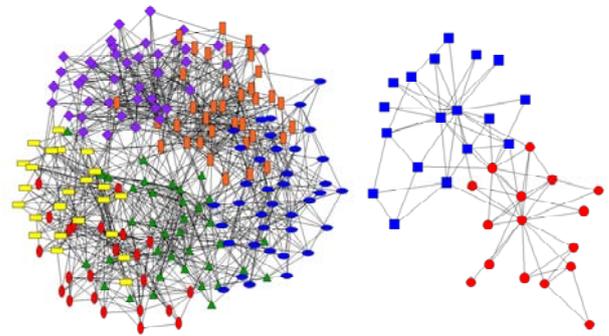


図 1:人工ネットワーク(左)および Zachary karate club ネットワーク(右)のモジュール分解。ノードの形が正解モジュールを、色が検出モジュールを表す。形と色が完全に一致していることに注目。

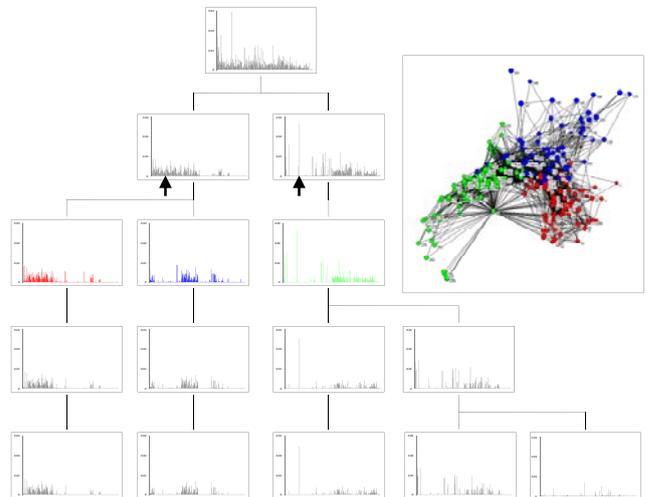


図 2:C. elegans 神経細胞ネットワークの階層モジュール分解。各グラフの横軸はノード番号、縦軸は $\pi_k p_n^{(k)}$ 。最上層から順に $\alpha = 400, 300, 200, 150, 140$ 。

参考文献

1. Meunier, D. et al. Front Neurosci 4, 1-11 (2010)
2. White, J.G. et al. Phil Trans R Soc London 314, 1-340 (1986)
3. Watts, D.J. & Strogatz, S.H. Nature 393, 440-442 (1998)
4. <http://cocomac.g-node.org/drupal/>
5. Newman, M.E.J. Nature Phys 8, 25-31 (2013)
6. Lancichinetti et al. Phys Rev E 78, 046110 (2008)
7. Zachary, W.W. J. Anthropol. Res. 33, 291-473 (1977)

謝辞: 本研究は JPSJ 科研費 23500379, 23300061 の助成を受けた。