

# ランダムウォークに基づいたグラフ構造モデリング

野田 淳史<sup>†1,a)</sup> 石田 諒<sup>†1</sup> 日野 英逸<sup>†1</sup> 龍野 正実<sup>†2</sup> 赤穂 昭太郎<sup>†3</sup> 村田 昇<sup>†1</sup>

概要：グラフ構造とは、データの変数間の関係を表す構造であり、鉄道路線、インターネット、神経回路網などが代表例として挙げられる。これらのグラフの変数間の依存関係が得られた際に、変数間の繋がり  
の強度を推定する問題は実用上重要であるが、ある変数間の依存関係には複数の変数から受けた影響が含まれているため、直接的な繋がり  
は分かりにくい。そこで本稿では、繋がり  
の有無のみを表現した行列でグラフを簡潔に表現し、そのグラフ上で情報が遷移する様子をモデル化することで、変数間の依存関係を近似する。モデルには、遷移の回数  
を表すパラメータとグラフ構造を表すパラメータが含まれており、それらを推定するアルゴリズムを提案する。最後に、提案手法の優位性を実験的に示す。

キーワード：グラフ構造, glasso, graph Laplacian, diffusion kernel, ランダムウォーク

## 1. はじめに

グラフ構造とは、データやものの関係を抽象化した数理的構造を指す。グラフ構造を持つ対象は、鉄道路線、インターネット、送電網、神経回路網など多岐に渡り、グラフ構造に基づいたパターン認識やデータマイニングが盛んに研究されている [1–3]。グラフはノード（頂点）と、ノード同士を結ぶエッジ（枝）から構成される。鉄道路線の例では、ノードとは駅であり、エッジとは線路である。各駅は線路を用いて乗客を授受しており、いくつかの駅を介することで、直接線路で結ばれていない駅まで行くことが出来る。

エッジを辿ることで離れたノードとの関係を表すことが出来る一方で、多数のノードやエッジを介した関係まで考慮すると、ノード間の直接的な関係を見失うことがある。例えば、神経回路網で複数のニューロンの発火が数十ミリ秒の範囲で観測された場合、ニューロン同士が連続して発火し相関が高く見えても、それらのニューロンは直接は繋がっておらず、別のニューロンを介して繋がっているために連続して発火したように見えるという擬相関が考えられる。本研究では、グラフのノード同士の関係性を表す量だけが既知である時に、ノード間のエッジの有無を推定する

問題を扱う。ノード同士の関係性を表す量とは、共起頻度や相関係数などがあるが、以降ではまとめて依存関係と呼ぶことにする。

こうしたグラフ構造の推定問題へのアプローチの一つに、データが多次元正規分布に従って生成されていると仮定し、その精度行列（分散共分散行列の逆行列）がノード間の結合強度を表しているとみなして推定する方法がある。多くの場合、グラフのノード同士の少数の本質的な結合を推定することが重要であり、精度行列にスパース性を仮定する。スパースな分散共分散行列の逆行列を求めることから、このアプローチは Sparse Inverse Covariance Selection (SICS) と呼ばれている。SICS の枠組みでは、尤度と精度行列の事前分布の積を最大化することで精度行列を推定する。代表的な解法アルゴリズムとしては、COVSEL [4], graphical lasso (glasso) [5], SINCO [6] などがある。これらの手法には計算時間が早いというメリットがあるが、データの生成に多次元正規分布を仮定していることや、分散共分散行列を考えているため有向グラフの解析には用いることが出来ないといったデメリットがある。

一方で、神経回路網の解析のために提案された情報幾何学的なグラフ構造推定手法 [7] がある。この手法ではデータの生成モデルとして指数分布族を仮定して、いくつかの近似を用いるものの解析的に結合強度を求めることができる。しかし柔軟なモデルを仮定しているため、十分な量のデータが得られれば良いが、そうでない時はノイズに過剰に適合してしまうことが問題となる。また結合強度の傾向（情報の伝わりやすさ）に関してハイパーパラメータを持ち、そのパラメータの決定方法がないという難点がある。

<sup>†1</sup> 早稲田大学  
Waseda University

<sup>†2</sup> レスブリッジ大学  
The University of Lethbridge

<sup>†3</sup> 独立行政法人産業技術総合研究所  
National Institute of Advanced Industrial Science and Technology

a) a24.noda@fuji.waseda.jp

本稿では、SICS のようにデータの生成分布を多次元正規分布に限定せず、かつモデルを簡略化する事でノイズの影響をさほど受けずに、真に重要な情報だけを抽出することが出来る手法を提案する。

## 2. 問題設定

本節ではまず、グラフ構造とグラフの種類に関して説明し、次にノード同士の依存関係に関する情報を持つ観測データからグラフの本質的な結合のみを推定する枠組みを提案する。例として図 1 のようなグラフを考える。

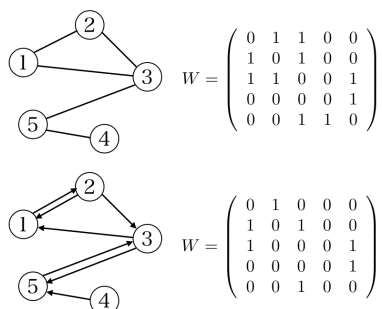


図 1 グラフ構造の例 (上: 無向グラフ, 下: 有向グラフ)  
Fig. 1 Example of graph structure

図 1 における各頂点をノード ( $i = 1, 2, \dots, n$ ) と呼び、ノード間を繋ぐ線のことをエッジと呼ぶ。エッジとノードの集合をそれぞれ  $V, E$  とする。ノード  $V$  とエッジ  $E$  の組  $G = (V, E)$  でグラフ  $G$  を定義する。

グラフにはエッジに向きがある場合とない場合の 2 種類がある。ノード間の繋がりだけに興味がある場合には、図 1 (上) のようにエッジに向きがない無向グラフを用いる。一方、ノード間で影響を及ぼしあう方向にも興味がある場合には、図 1 (下) のようにエッジに向きがある有向グラフを用いる。図 1 の行列  $W$  に関しては graph Laplacian の節で述べる。

次に、主な変数の一覧を表 1 に示す。

表 1 変数の一覧  
Table 1 List of variables

記号	説明
$n$	ノードの数
$t_{ij}$	観測データ
$T$	$t_{ij}$ を $ij$ 成分に持つ行列
$\theta_{ij}$	観測データの平均
$\Theta$	$\theta_{ij}$ を $ij$ 成分に持つ行列
$p(t_{ij}; \theta_{ij})$	$\theta_{ij}$ を平均とする $t_{ij}$ の確率密度関数
$\lambda_{ij}$	$i$ と $j$ の直接的な繋がりを表すパラメータ
$\Lambda$	$\lambda_{ij}$ を $ij$ 成分に持つ行列

ここで、 $T$  はノード  $i$  と  $j$  の依存関係の特徴づける量  $t_{ij}$  を  $ij$  成分とする  $n \times n$  行列である。例えば、鉄道路線の場

合は、乗客は皆往復で同じ経路を辿るという仮定の下で、一日に駅  $i$  と駅  $j$  の間を往復する人数を無向グラフで表すことが出来る。また、神経回路網の場合は、単位時間内にニューロン  $i$  の発火直後にニューロン  $j$  の発火があった回数を有向グラフで表すことが出来る。 $T$  の対角成分はノード間の依存関係として定義出来ないため、便宜上 0 で初期化するものとし、グラフ構造の推定には用いない。 $T$  にはノイズが加わって観測されると考え、そのノイズに確率分布を仮定してモデル化する。 $\theta_{ij}$  を  $t_{ij}$  の平均とし、 $t_{ij}$  は  $\theta_{ij}$  に平均 0 の独立同一分布に従うノイズが加わって観測されたと考える。簡単のため、ノイズは正規分布に従うとし、観測データ  $t_{ij}$  の密度関数を、

$$\phi(t_{ij}; \theta_{ij}, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2\sigma^2} (t_{ij} - \theta_{ij})^2 \right\} \quad (1)$$

とする。ここでは式 (1) における平均  $\theta_{ij}$  のみを推定し、分散  $\sigma^2$  は  $ij$  によらず一定であることを仮定し、以下では、

$$\phi(t_{ij}; \theta_{ij}, \sigma^2) = \phi(t_{ij}; \theta_{ij}) \quad (2)$$

で表す。パラメータ  $\theta_{ij}$  を  $ij$  成分に持つ行列を  $\Theta$  とすると、観測データ  $T$  の同時分布は、

$$\phi(T; \Theta) = \prod_{i \neq j} \phi(t_{ij}; \theta_{ij}) \quad (3)$$

と表すことが出来る。

ノード  $i$  と  $j$  の直接的な繋がりを表すパラメータ  $\lambda_{ij}$  を  $ij$  成分に持つ行列を  $\Lambda$  とし、 $\Theta$  は  $\Lambda$  を写像

$$f: \mathbb{R}^{n \times n} \rightarrow \mathbb{R}^{n \times n} \\ \Lambda \mapsto f(\Lambda) = \Theta \quad (4)$$

で写すことで得られると仮定する。 $\Lambda$  を求める際には、 $\Lambda = f^{-1}(\Theta)$  のように逆変換可能であることが望ましい。逆変換が可能な写像としては後述する行列指数関数や Cayley 変換などがある。式 (4) の変換が、

$$\theta_{ij} = c_0 + c_{ij} \lambda_{ij} + \sum_k c_{ij}^k \lambda_{ik} \lambda_{kj} + \sum_{kl} c_{ij}^{kl} \lambda_{ik} \lambda_{kl} \lambda_{lj} + \dots \quad (5)$$

の様に表せると仮定する。このとき、パラメータ  $\theta_{ij}$  と  $\lambda_{ij}$  の関係を模式的に図 2 に示す。

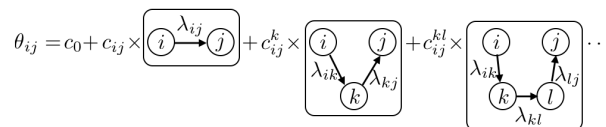


図 2  $\theta_{ij}, \lambda_{ij}$  の解釈  
Fig. 2 Interpretation of  $\theta_{ij}, \lambda_{ij}$

式 (5) の第 2 項はノード  $i$  と  $j$  の直接的な繋がりによる情報伝達を表し、第 3 項は一つのノードを介して伝わる情報

伝達を表す．このように  $i$  と  $j$  の依存関係の強さを表す  $\theta_{ij}$  は、様々なルートからの情報流入の和であると考えられる．また、複数のノードを介してノード  $i$  からノード  $j$  に情報が伝わる際の減衰係数をまとめて

$$C = \{c_0\} \cup_{ij} C_{ij} \quad (6)$$

$$C_{ij} = \{c_{ij}\} \cup \{c_{ij}^k | k = 1, 2, \dots\} \cup \{c_{ij}^{kl} | k, l = 1, 2, \dots\} \cup \dots \quad (7)$$

で表す．式 (5) からわかるように、 $\Lambda$  と  $C$  が分かれば  $\Theta$  を表現することが出来る．

パラメータ  $\Theta$  には、式 (5) のように複数の経路を伝わる影響が含まれていてノード間の直接的な関係が明瞭でないため、パラメータ  $\Lambda$  を求めることとする．さらに、 $\Lambda$  に対して制約を課しスパースな  $\Lambda$  を求めることで、主要な繋がりのみを推定することが出来る．

### 3. グラフ構造のモデル

本節では、ノード間の繋がりの有無のみを表した隣接行列  $W$  から作られる graph Laplacian  $L$  を説明し、グラフ上でのランダムウォークモデルとしての解釈を紹介する．また、このモデルの下での観測データ  $T$  の意味付けを行う．

#### 3.1 Graph Laplacian

本節では、行列を用いてグラフ  $G = (V, E)$  を表記する方法である graph Laplacian を導入する．まず、無向グラフを考える．隣接行列  $W$  を、

$$w_{ij} = \begin{cases} 1 & i \text{ と } j \text{ が結合している時} \\ 0 & \text{それ以外} \end{cases} \quad (8)$$

と定義する． $w_{ij}$  は隣接行列  $W$  の  $ij$  成分であり、エッジは同一の強度で繋がれているものとしている．ただし、隣接行列  $W$  の対角成分は自身へのエッジなので存在せず、常に 0 である．本稿では問題を解きやすくするため、主要な繋がりを表すパラメータ行列  $\Lambda$  がグラフの隣接行列  $W$  によって、

$$\Lambda = W \quad (9)$$

と表せることを仮定する．このように実数値を取り得る  $\Lambda$  を整数値のみを取る  $W$  で表すことで、ノード間の繋がりが明瞭になり、ノイズに過剰適合することなくグラフ構造を求めることが出来ると考えられる．

次に、次数行列  $D$  を、

$$d_{ij} = \begin{cases} \sum_{k \neq i, k=1}^n w_{ik} & i = j \\ 0 & i \neq j \end{cases} \quad (10)$$

と定義する．ただし、 $d_{ij}$  は  $D$  の  $ij$  成分である．

隣接行列  $W$  と、 $W$  から求められる次数行列  $D$  を用いて graph Laplacian  $L$  を、

$$L = D - W \quad (11)$$

で定義する． $D$  と  $W$  の関係から、graph Laplacian  $L$  には、行和と列和が 0 になるという性質がある．Graph Laplacian [8] を用いた研究は、その固有値問題を解くことでグラフを分割するクラスタリング手法 [9, 10] や、多様体学習 [11] の研究を始めとして、盛んに行われている．

Graph Laplacian は、式 (8) で  $i$  と  $j$  が結合している時を 1 としていることからわかるように、対称行列(無向グラフ)を仮定している．一方、graph Laplacian を非対称行列(有向グラフ)に拡張したものとして、digraph Laplacian [12] がある．その際の隣接行列  $W$  は、

$$w_{ij} = \begin{cases} 1 & i \text{ から } j \text{ へのエッジがある時} \\ 0 & \text{それ以外} \end{cases} \quad (12)$$

となる．次数行列  $D$  の定義は式 (10) と同じで、 $L$  の定義は式 (11) と同じであるため、digraph Laplacian  $L$  の場合は、行和のみが 0 になる．

#### 3.2 グラフ上のランダムウォーク

Digraph Laplacian  $L$  が推定出来れば隣接行列  $W$  が求まるので、すなわちグラフ構造が求まったことになる．また、 $L$  の  $ij$  成分には  $i$  から  $j$  へエッジがあれば  $-1$  が、無ければ 0 の値が入っており、 $L$  は 1 つ隣のノードとの関係を表している．本節では、グラフ上での情報の遷移を  $L$  を用いて表現する．

単位時間内にノード間で遷移が起こる回数の平均を  $\beta$  とする．十分大きな  $k$  を考え、時間間隔  $\frac{1}{k}$  の間に遷移が起こる確率を考えると  $\frac{\beta}{k}$  となる．したがって、微小な時間間隔  $\frac{1}{k}$  の間の遷移を表す行列は、

$$I - \frac{\beta}{k}L = \begin{pmatrix} 1 - \frac{\beta}{k}d_{11} & \frac{\beta}{k}w_{12} & \dots & \frac{\beta}{k}w_{1n} \\ \frac{\beta}{k}w_{21} & 1 - \frac{\beta}{k}d_{22} & & \vdots \\ \vdots & & \ddots & \\ \frac{\beta}{k}w_{n1} & \dots & & 1 - \frac{\beta}{k}d_{nn} \end{pmatrix} \quad (13)$$

となる．式 (10) の次数行列  $D$  の定義より、式 (13) の行の和が 1 となることから、 $(I - \frac{\beta}{k}L)$  は遷移確率行列となっている．この行列で表現される遷移の様子を図 3 に示す．

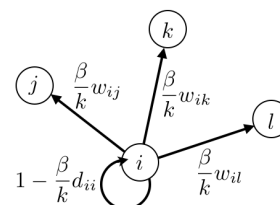


図 3  $(I - \frac{\beta}{k}L)$  で表現される 1 回の遷移  
 Fig. 3 One transition represented by  $(I - \frac{\beta}{k}L)$

したがって、 $i$  番目が 1 でその他の要素が全て 0 の  $n$  次元ベクトルに行列  $(I - \frac{\beta}{k}L)$  を右から掛けることで、1 回の遷移でノード  $i$  から各ノードへ情報が拡散する割合が分かる。単位時間内の微小な時間間隔  $\frac{1}{k}$  の間の遷移確率が  $(I - \frac{\beta}{k}L)$  なので、単位時間内の遷移確率は  $(I - \frac{\beta}{k}L)^k$  となる。遷移がいつでも自由に起こる状態を考えるために、遷移回数  $k$  について極限を取ることで、遷移確率行列を、

$$\lim_{k \rightarrow \infty} \left( I - \frac{\beta L}{k} \right)^k = e^{-\beta L} \quad (14)$$

と定義出来る。右辺は行列の指数関数のマクローリン展開であり、 $L$  が graph Laplacian (対称行列) の時、式 (14) の右辺は diffusion kernel [13] と呼ばれる。式 (14) のように、 $(I - \frac{\beta}{k}L)$  を無限に掛け合わせることで、 $e^{-\beta L}$  の  $ij$  成分はノード  $i$  から始まったランダムウォークが定常状態でノード  $j$  にいる確率と解釈することが出来る。

### 3.3 観測データの定義

本節では、観測データの生成モデルについて定義する。式 (14) を式 (5) と対応付けるために、 $\theta$  と  $e^{-\beta L}$  のスケールを調整するパラメータである  $\alpha$  を式 (14) に掛けてテイラー展開すると、

$$\alpha \cdot e^{-\beta L} = \alpha \cdot \left( I - \beta L + \frac{\beta^2 L^2}{2!} - \frac{\beta^3 L^3}{3!} + \frac{\beta^4 L^4}{4!} - \dots \right) \quad (15)$$

となり、第 2 項から順に、 $-\alpha\beta$  が  $c_{ij}$  に、 $\alpha\beta^2/2!$  が  $c_{ij}^k$  にこのように対応している。また、式 (15) の  $L$  の  $ij$  成分が式 (5) の  $\lambda_{ij}$  と対応している。そこで、観測データ  $T$  の  $ij$  成分  $t_{ij}$  を

$$t_{ij} = [\alpha \cdot e^{-\beta L}]_{ij} + \varepsilon \quad \varepsilon \sim \mathcal{N}(0, \sigma^2) \quad (16)$$

と定義する。行列  $M$  に対して  $[M]_{ij}$  は  $M$  の  $ij$  成分を表すものとする。ただし、 $\sigma^2$  は正規分布の分散であり、観測ノイズの大きさを意味するパラメータである。今節までのまとめとして、変数間の関係を図 4 に示す。

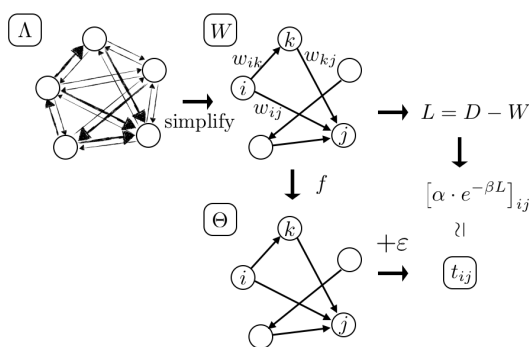


図 4 変数間の関係

Fig. 4 Relation between variables

再び神経回路網を例に挙げると、 $t_{ij}$  は、ニューロン  $i$

が発火して生じる信号が  $W$  のグラフ上で複数の経路を伝わってニューロン  $j$  に及ぼす影響の総和となっている。

## 4. アルゴリズム

本節では、観測データから式 (16) のパラメータ  $\alpha, \beta$ , graph Laplacian  $L$  を推定する反復アルゴリズムを提案する。なお、前述したように、観測データの対角成分はノード間の依存関係を表す値ではないため、便宜上 0 で初期化している。一方、式 (16) で定義したモデルは対角成分を持つため、 $f$  の逆関数を用いて  $\alpha, \beta, L$  を推定する場合には適切に  $T$  の対角成分を補間する必要がある。そのため、反復アルゴリズムでパラメータ  $\alpha, \beta, L$  を更新しつつ、 $T$  の対角成分を更新していく。

はじめに  $T$  の対角成分を任意の自然数  $r$  で置き換えて初期化し、 $T'$  とする。この際、 $T'$  が正定値となるように  $r$  を選ぶ。グラフのエッジ総数が  $m$  であるとし、 $m = 1$  で初期化する。

次に、 $T'$  の対角成分を計算する。無向グラフの場合、 $T'$  は正定値対称行列なので対角化し、固有値の対数を取ればよい。一方、有向グラフの場合、対角化する代わりにジョルダン標準形を用いて計算を行う [14]。 $[\log(T')]_{ij}$  の絶対値を降順にソートし、

$$w_{ij} = \begin{cases} 1 & |[\log(T')]_{ij}| \geq \zeta \\ 0 & \text{otherwise} \end{cases} \quad (17)$$

のように、 $|[\log(T')]_{ij}|$  が閾値  $\zeta$  以上だった所にエッジがあるとして  $W$  を作る。閾値  $\zeta$  は、降順にソートした  $|[\log(T')]_{ij}|$  の  $m$  番目の値とする。  $W$  から  $L$  を作り、求めた  $L$  に対して誤差関数

$$J(\alpha, \beta, L) = \sum_{i=1}^n \sum_{\substack{j=1 \\ i \neq j}}^n \left( t'_{ij} - [\alpha \cdot e^{-\beta L}]_{ij} \right)^2 \quad (18)$$

を最小化する  $\alpha, \beta$  を任意の最適化手法で求める。本研究では Nelder Mead 法 [15] を用いた。推定した  $\alpha, \beta, L$  を用いて  $T$  の対角成分を

$$T' = T + \text{diag}(\alpha \cdot e^{-\beta L}) \quad (19)$$

のように更新する。  $\text{diag}(\cdot)$  は行列の非対角成分を全て 0 にし、対角成分のみを残す関数とする。対角成分を更新する反復のカウントを  $k$  とし、任意に定めた定数  $k'$  まで  $k$  を更新する。上記の処理を  $m = 1, 2, \dots, n(n-1)$  に対して行い、作成した  $n(n-1)$  個の  $L$  の中で最も誤差関数を小さくする  $L^m$  を  $L$  の推定値とする。無向グラフの場合は、 $[\log(T')]_{ij}$  ( $i > j$ ) に対して降順にソートして  $w_{ij}$  ( $i > j$ ) を作り、 $w_{ji}$  ( $j > i$ ) =  $w_{ij}$  ( $i > j$ ) として対称行列にすればいいため、 $m$  は最大で  $n(n-1)/2$  で済む。アルゴリズムを Algorithm 1 にまとめる。下記の  $W^m, D^m, L^m, \alpha^m, \beta^m$  はそれぞれ、真のグラフのエッジ総数を  $m$  とした時の  $W, D, L, \alpha, \beta$  の推定値である。

**Algorithm 1** Proposed algorithm

---

**input:**  $T, r, k'$   
**initialize:**  $T' = T + rI_n$   $I_n$ : 単位行列  
**for**  $m = 1$  to  $n(n-1)$  **do**  
  **for**  $k = 1$  to  $k'$  **do**  
    **step1:**  
    式 (17) に従って  $[\log(T')]_{ij}$  から  $W^m$  を作る .  

$$L^m = D^m - W^m$$
  
    **step2:**  
     $L^m$  に対して, 誤差関数を最小化する  $\alpha^m, \beta^m$  を求める .  

$$(\alpha^m, \beta^m) = \arg \min_{\alpha, \beta} J(\alpha, \beta, L^m)$$
  
    推定した  $\alpha^m, \beta^m, L^m$  を用いて  $T$  の対角成分を更新する .  

$$T' = T - \text{diag}(T) + \text{diag}(\alpha^m \cdot e^{-\beta^m L^m})$$
  
  **end for**  
**end for**  

$$\hat{m} = \arg \min_{m=\{1,2,\dots,n(n-1)\}} J(\alpha^m, \beta^m, L^m)$$
  
**return**  $L^{\hat{m}}$

---

**5. 実験**

本節では, 実際のグラフ構造である隣接行列  $W$  を決めた上で推定を行った結果を示す.  $W$  のサイズは  $20 \times 20$  として,  $\alpha, \beta$  は  $[0, 1]$  上の一様分布から生成する. 無向グラフは提案手法と glasso の比較を行い, 有向グラフは提案手法のみで推定を行う.

**5.1 無向グラフの推定**

本節では, 観測データが式 (16) から生成されるという仮定に基づき人工的に生成した観測データと, 隣接行列  $W$  を精度行列とする多次元正規分布から生成した観測データに対して, それぞれ提案手法と glasso で推定を行う. glasso は非対称行列に用いることが出来ないため, 隣接行列  $W$  は対称行列とする.

**5.1.1 提案モデルの仮定に従うデータを用いた実験**

はじめに,  $W$  の密度を, 各成分  $w_{ij} (i \neq j)$  が 1 になる確率で定義し, 密度が 0.4 の  $W$  を用意する. 次に, 式 (16) に従って観測データ  $T$  を生成する. グラフのエッジの有無の正答率を, 推定した隣接行列の  $ij$  成分を  $\hat{w}_{ij}$  とし,

$$\text{accuracy (\%)} = \left( 1 - \frac{\sum_{i=1}^n \sum_{j=1}^n |w_{ij} - \hat{w}_{ij}|}{n^2} \right) \times 100 \quad (20)$$

で定義して, 正答率を縦軸に, 観測ノイズの大きさ (ノイズ  $\varepsilon$  の標準偏差  $\rho$ /観測データ  $T$  の平均) を横軸に取った図を, 図 5 に示す.

この際, glasso では正則化項のパラメータ  $\rho$  を決める必要がある. また, 求めた精度行列に対して閾値を定め, 精度行列の  $ij$  成分を 0 か 1 に振り分け隣接行列としている. この時,  $\rho$  と閾値をそれぞれ  $[0, 1]$  上でグリッドサーチし,

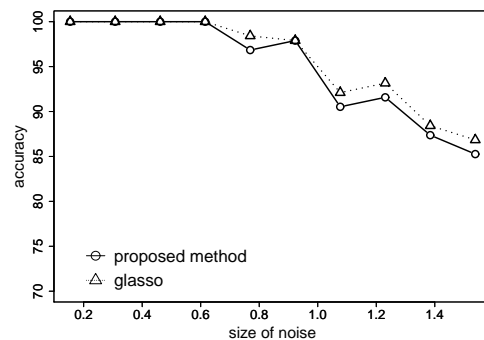


図 5 提案手法と glasso の正答率 (密度 = 0.4)  
Fig. 5 Accuracy rate of proposed method and glasso

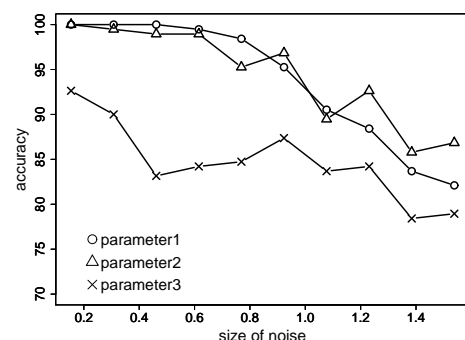


図 6 複数の  $\rho$  と閾値での正答率 (密度 = 0.4)  
Fig. 6 Accuracy rate of various  $\rho$  and threshold

最も正答率が高くなるパラメータを採用している. つまり, 図 5 ではそれぞれのノイズに対して本来未知である正解  $W$  を用いて最適な  $\rho$  と閾値を決めているため, glasso の正答率は, glasso が推定し得る理想的な最良値となっている. また, 図 6 に示すように, あるノイズをのせた際に最適だった  $\rho$  と閾値は, 別のノイズをのせた際には一般には最適とは限らない. また, 図 6 の parameter3 で示したように, 不適切な  $\rho$  と閾値を選ぶと常に正答率が低いこともある. このように, ノイズによって最適な  $\rho$  と閾値が異なるため, glasso を用いる際には  $\rho$  と閾値の選び方が重要となるが, 実データを解析するにはその値を決めるのが困難である. 一方, 提案手法ではパラメータのチューニングが不要であり, かつ glasso が推定し得る理想的な最良値に近い正答率を達成をすることが出来る.

**5.1.2 多次元正規分布に従うデータを用いた実験**

密度を 0.4 とした  $W$  が正定値となるよう適当な実数で  $W$  の対角成分を埋め,  $W$  を精度行列とした多次元正規分布からベクトル  $s$  を  $N$  個作成する. なお, 多次元正規分布の平均は 0 とする.  $N$  個のベクトル  $s_1, \dots, s_N$  から計算した経験共分散行列  $S = \frac{1}{N} \sum_{i=1}^N s_i s_i^T$  を観測  $T$  として推定を行う. この際の観測ノイズの標準偏差の大きさと正答率の関係を図 7 に示す.

ここでも, 図 5 と同様に, glasso では最適な  $\rho$  と閾値を

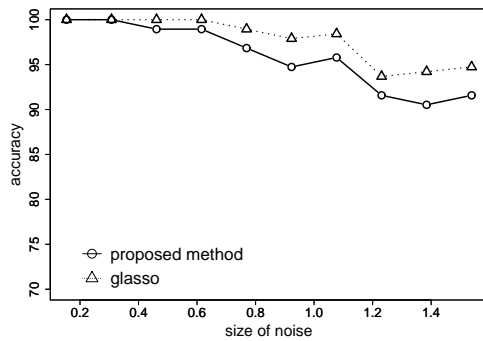


図 7 提案手法と glasso の正答率 (密度 = 0.4)

Fig. 7 Accuracy rate of proposed method and glasso

グリッドサーチしている。図 7 より、提案手法は、他のモデル（ここでは多次元正規分布）から生成されたデータに対してもパラメータのチューニングをすることなく、ノイズが小さい場合には概ね正しく推定することが出来る。

### 5.2 有向グラフの推定

本節では、非対称な隣接行列  $W$  である表現される有向グラフの推定を行う。密度が 0.2, 0.4, 0.6 の  $W$  を用意し、それぞれの  $W$  に対して  $\alpha, \beta$  は固定する。この際の観測ノイズの標準偏差の大きさと正答率の関係を図 8 に示す。

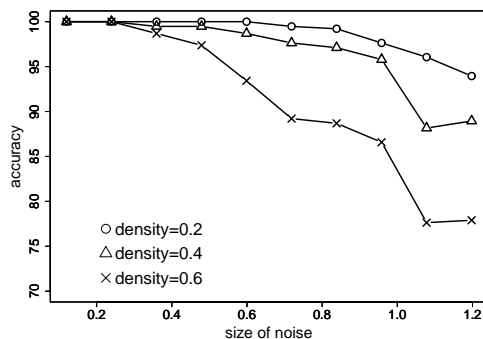


図 8 提案手法の正答率 (密度 = 0.2, 0.4, 0.6)

Fig. 8 Accuracy rate of proposed method

図 8 のように、提案手法では有向グラフに対しても無向グラフと同様に推定が出来る。また、グラフが密になるほど正答率が下がる。

## 6. まとめ

本稿では、ノード間の依存関係が得られた際に、ノード間のエッジの有無を推定する手法を提案した。多くの場合主要な繋がりのみが重要なため、digraph Laplacian  $L$  を用いてグラフ構造を簡潔に表現し、 $L$  で規定されるグラフ上で遷移確率行列を用いてノード間の依存関係を近似した。

グラフ構造推定のための既存手法の一つである glasso では、正則化などのパラメータをチューニングする必要があ

るが、提案手法ではパラメータチューニングをする必要がないという利点がある。提案手法では、glasso が最も正答率が高くなるようにパラメータをチューニングした際の最良値と同等の推定が出来ることを実験的に示した。また、glasso は無向グラフにしか適用出来ないのに対して、提案手法は有向グラフを扱うことが出来る。そのため、神経回路網のように、ニューロン同士の影響の及ぼし方に向きがあるという仮定が妥当な対象にも適用が出来る。今後は、提案手法を用いてニューロンの発火データを解析し、脳内のグラフ構造を推定していく予定である。

### 参考文献

- [1] H. Bunke and K. Riesen: *Recent advances in graph-based pattern recognition with applications in document analysis*, Pattern Recognition 44(5):1057-1067, 2011.
- [2] D. Conte, P. Foggia, C. Sansone and M. Vento: *Thirty years of graph matching in pattern recognition*, International Journal of Pattern Recognition and Artificial Intelligence 18(3):265-298, 2004.
- [3] T. Washio and H. Motoda: *State of the art of graph based data mining*, ACM SIGKDD Explorations Newsletter 5(1):59-68, 2003.
- [4] O. Banerjee, L. El Ghaoui and A. d'Aspremont: *Model selection through sparse maximum likelihood estimation for multivariate gaussian for binary data*, Journal of Machine Learning Research, 9:485-516, 2008.
- [5] J. Friedman, T. Hastie and R. Tibshirani: *Sparse inverse covariance estimation with the graphical lasso*, Biostatistics, 9(3):432-441, 2007.
- [6] K. Scheinberg and I. Rish: *SINCO - a greedy coordinate ascent method for sparse inverse covariance selection problem*, RC24837(W0908-046), 2009.
- [7] M. Tatsuno, J. M. Fellous and S. Amari: *Information-Geometric Measures as Robust Estimators of Connection Strengths and External Inputs*, Neural Computation 21, 2309-2335, 2009.
- [8] F. R. K. Chung: *Spectral Graph Theory*, American Mathematical Society, 1997.
- [9] J. Shi and J. Malik: *Normalized Cuts and Image Segmentation*, IEEE Transactions on pattern analysis and machine intelligence, 22(8), 2000.
- [10] M. Belkin and P. Niyogi: *Laplacian eigenmaps and spectral techniques for embedding and clustering*, In proceedings of NIPS, 2002.
- [11] H. Sahbi, P. Etyngier, J. Audibert and R. Keriven: *Manifold learning using robust graph Laplacian for interactive image search*, In Proceedings of CVPR, 2008.
- [12] Y. Li and Z. Zhang: *Ramdon Walks on Digraphs, the Generalized Digraph Laplacian and the Degree of Asymmetry*, In Proceedings of WAW.74-85, 2010.
- [13] R. Kondor and J. Lafferty: *Diffusion Kernels on Graphs and Other Discrete Input Spaces*, In Proceedings of ICML, 2002.
- [14] N. J. Higham: *Functions of Matrices: Theory and Computation*, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2008.
- [15] J. Lagarias, J. Reeds, M. Wright and P. Wright: *Convergence Properties of the Nelder-Mead Simplex Method in Low Dimensions*, SIAM Journal on Optimization, 9(1):112-147, 1998.