

# 複数の応答信号を教師データとするネットワークの サイズ非依存な構造的特徴に基づく分類

佐藤 進也<sup>1,a)</sup>

受付日 2018年1月19日, 採録日 2018年7月10日

**概要:** 本論文では, サイズに非依存なネットワークの構造的特徴を教師データから学習し分類問題を解く手法を提案する. 本手法は, ネットワーク中で信号が伝播する様子のあるノードで観測しそこから構造的特徴を把握する既存手法を応用したものである. 提案手法では, ネットワークから一定の条件を満たすノードの集合を抽出し, その各要素で観測した情報を学習と推定に用いる. 1つのネットワークから複数の構造的特徴を抽出することで学習と推定に利用可能なデータ量を増やせること, また, 伝播させる信号のパラメータも学習のハイパーパラメータとし, 分類性能の調整に使えることが本手法の特徴である. 評価のため, Barabási-Albert (BA) モデルにより生成されたネットワークと, BA モデルにより生成されたものの一部を改変して得られたネットワークの識別問題に対して, 既存手法と提案手法を適用し分類性能を比較した. その結果, 提案手法で既存手法を上回る正解率が得られることを確認した.

**キーワード:** 複雑ネットワーク, ネットワーク分類, 構造解析, 機械学習, パラメータ探索, 信号処理, ネットワーク埋め込み

## Classification of Networks Based on Size-independent Structural Characteristics Using Multiple Response Signals as Training Data

SHIN-YA SATO<sup>1,a)</sup>

Received: January 19, 2018, Accepted: July 10, 2018

**Abstract:** This paper proposes a method for solving classification problems by learning size independent structural characteristics of networks from training data. This method makes use of the existing technique that grasps structural characteristics of a network by monitoring the ways signals are propagated through links from a node. The method extracts a set of nodes that satisfy a certain condition from the network, and uses information obtained at each node in the set for training and inference. It is a feature of the proposed method that we can increase the amount of data that can be used for learning and inference by extracting multiple structural features from one network. Another feature is that we can use the parameters of signals to be propagated through the network as hyperparameter of learning and to use it for classification performance tuning. We conducted an evaluation experiment for comparing classification performances where the existing methods and the proposed method were respectively applied to the classification problem of networks generated by the Barabási-Albert (BA) model and the networks obtained by modifying part of the networks generated by the BA model. We confirmed that the proposed method yields better accuracy than the existing methods.

**Keywords:** complex network, network classification, structural analysis, machine learning, parameter tuning, signal processing, network embedding

### 1. はじめに

一般に, 多くのシステムは相互関係を持つ複数の要素から構成される. 互いにリンクしあうページで構成される Web や, 相互接続されたニューロンで構成される脳など

<sup>1</sup> 日本工業大学先進工学部情報メディア工学科  
Department of Information Technology and Media Design,  
Faculty of Advanced Engineering, Nippon Institute of Technology,  
Minamisaitama, Saitama 345–8501, Japan

<sup>a)</sup> shin-ya.sato@acm.org

をその例としてあげることができる。これらの要素をノード、関係をリンクで表すことで、ネットワークとしてとらえることができる。この、ネットワークとして抽象化し解析するアプローチは多くの成果をもたらす、ネットワーク解析手法の研究もさかんに行われてきた [1], [2]。その一連の研究の中で、ネットワークの構造に着目したとき異なるシステムに共通性を見い出せること、システムの機能や特徴をネットワークの構造と対応付けて説明できることが分かってきた [3], [4], [5]。たとえば、多くのシステムを抽象化して得られるネットワークにはスケールフリーという構造的特徴が共通して認められる。この構造的共通性は、システムの形成プロセスに共通性がある可能性を示唆している [6]。別の例としては、脳の機能を神経細胞ネットワークの構造的特徴（フィードフォワードなど）に対応付けて説明する取り組みをあげることができる [5]。現在では、ネットワークの構造的特徴把握は、システムの機能、特徴を理解するための有効な方法の1つとなっている。

ネットワークの構造を比較する方針として、サイズ（ノード数）を含めて違いを詳細に区別するアプローチが考えられる。たとえば、2つのネットワークがともにフィードフォワードネットワークであっても、サイズが異なれば違うものとしてとらえる。一方、サイズが異なってもつながり方のパターンが共通しているのであれば同一視するという考え方もある。これは、たとえば、脳というシステムを神経細胞ネットワークの構造から理解するという立場からすると、サイズがどうであれ、フィードフォワードネットワークは同様な情報処理機能を果たすであろうから同一視するという考え方である。本論文では、後者の立場をとり、サイズに非依存な構造を識別する方法について議論する。特に、ネットワークの構造を識別する問題のうち、教師あり学習による分類問題として定式化できるものを考える。これは、すなわち、ネットワークの分類例と未分類のネットワーク  $N$  が与えられたとき、 $N$  がいずれの類（クラス）のいずれに属するかを推定するという問題である。本論文では、この分類問題を解くための新しい手法を提案する。

ネットワークのクラスには様々なものが考えられる。完全グラフのクラスや2次元格子のクラスといったものが考えられるが、本論文では、より実ネットワークに近いネットワークのクラスを題材にし、提案手法の分類性能を評価する。具体的には、Barabási-Albert (BA) モデル [6] により生成されたネットワークと、BA モデルにより生成されたものの一部を改変して得られたネットワークの分類問題を扱う。これらを識別できるということは、改変の有無を識別できるということを意味する。提案手法を使って、この種の分類問題に取り組んだ結果、既存手法の上回る正解率が得られた。本論文では、その実験の内容と結果について述べる。

本論文では以下の順序で議論をすすめる。2章では関連研究について、3章では提案手法についてそれぞれ述べる。提案手法の有効性を確認するために行った実験について4章で述べ、その結果について5章で議論する。

## 2. 関連研究

### 2.1 NetSimile

サイズに非依存な構造を把握する手法としては、Berlingerioらによる NetSimile [7] がある。これは、クラスタ係数や次数など7種類のネットワーク指標の分布に関する5種類の統計量から得られる35次元の「シグネチャ」を Canberra 距離 [8] で比較することで類似性を判断するものである。それぞれのネットワーク指標は、構造的特徴のある側面をとらえたものであるため、単一の指標が一致したことがネットワーク全体として構造的に類似していることを示しているわけではない。そこで、NetSimile では、複数の指標の統計量をもとに総合的に類似性を判断している。本論文では、NetSimile の分類性能と比較し、提案手法の有効性を評価する。

### 2.2 ネットワークの応答信号に基づく構造解析

サイズに非依存な構造を把握するもう1つの手法として、ネットワークの構造的特徴を時系列で表現する手法 [9] が提案されている。これは、ネットワークのあるノードから信号（時系列） $\{s_i\}$  を入力し、リンク経由で伝播させ、その結果得られる信号（応答信号） $\{r_i\}$  ( $i = 1, 2, \dots, n$ ) のスペクトルを計算し、そこから構造的特徴を読み取るものである。以降、本論文ではこの手法を応答スペクトル法と呼ぶ。

図1は応答スペクトル法の処理手順を示したものである。破線の枠はデータ、実線の枠は処理を示す。応答スペクトル法では、トピックモデルにより、応答信号の周波数成分と構造的特徴が対応付けられている。そして、応答信号における特定の周波数成分の検出により、対応する構造が識別される。また、応答信号中に混在する複数の周波数成分をそれぞれ分離することにより、対応する構造を識別することができる。提案手法は、この研究で導入された応答信号を分類に用いる。クラスと応答信号の対応関係を学習し、得られたモデルに基づいて分類を行う。応答スペクトル法は混在している複数の構造の把握を主要な目的の1つ

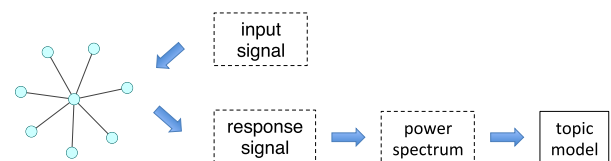


図1 応答スペクトル法によるネットワーク構造解析の手順  
Fig. 1 Procedure of network structure analysis by the response spectrum method.

として開発された手法であるが、提案手法ではクラス分類を目的としている。混在している個々の構造を把握する必要がないため、スペクトルの計算およびトピックモデルによる潜在トピックの抽出といった処理は行わず、応答信号を直接教師データとして用いる。

### 2.3 ネットワーク埋め込み

ネットワーク埋め込みとは、ネットワークの各頂点  $v$  を、他頂点との構造的関係を維持しつつ  $n$  次元空間の点  $\mathbf{z}(v)$  に対応させることである。 $v$  をベクトル  $\mathbf{z}(v)$  にすることにより、機械学習の適用が可能となり、頂点のクラスタリングや分類、可視化への応用が行われている [10]。これまでに様々なネットワーク埋め込みの手法が提案されている。 $\mathbf{z}(v)$  の導出にノードの隣接情報のみを使用するもの [11], [12] や、ノードの属性を利用するもの [13], [14] もある。

本論文の提案手法で用いている応答信号も数値の列であるので、隣接情報のみに基づくネットワーク埋め込みの一手法と考えられる。また、逆に、ネットワーク埋め込みを使ってネットワーク全体の特徴を把握することもできる。文献 [15] では、ネットワーク  $N$  の部分グラフ  $S$  に対し、その特徴を表すベクトル  $\mathbf{z}(S)$  を以下のように定義している：

$$\mathbf{z}(S) = \sum_{v \in S} \mathbf{z}(v)$$

上式で、 $S = N$  とすれば、ネットワーク全体の特徴を表す  $\mathbf{z}(N)$  が得られる。4 章では、近年開発されたネットワーク埋め込みの手法のうち、隣接情報のみによりベクトルが得られる node2vec [12] を使って分類実験を行った結果を示し、提案手法と比較する。

## 3. 提案手法

本章では、提案手法について説明する。

### 3.1 ネットワークの応答信号

まず、提案手法の導入に必要な応答信号の定義 [9] を示す。信号は時系列（すなわち数量の列）であり、ネットワークにおける信号の伝播とは、この数量が、時間経過にともないリンクを介して伝わっていくこととして定式化される。 $N$  を無向ネットワークとし、そのノードの集合を  $\{v_i\}$  とする。また、 $\mathbf{A} = (a_{ij})$  を隣接行列とし、確率行列  $\mathbf{B}$  を

$$\mathbf{B} = (b_{ij}), b_{ij} = a_{ij} / \sum_i a_{ij}$$

とする。各ノードで時刻  $t$  に受け取る数量を表すベクトルを  $\mathbf{q}_t$  とすると、信号の伝播は

$$\mathbf{q}_t = \mathbf{B}\mathbf{q}_{t-1}$$

と定義される。

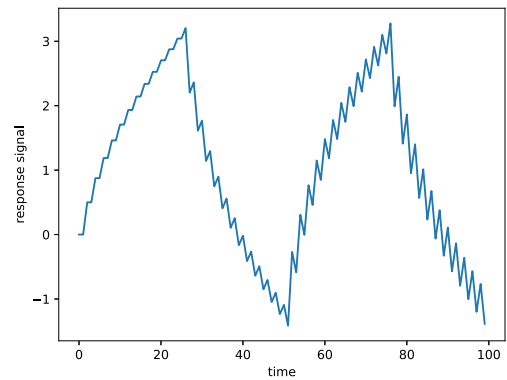


図 2 応答信号の例

Fig. 2 Example of response signal.

ノード  $v_k$  から信号  $\mathbf{s} = \{s_t\}$  が入力されたとき、 $v_k$  は、近傍ノードから受け取った数量とともに入力  $s_t$  を次時刻に近隣ノードに伝達する。これは、 $k$  番目の要素だけが 1 である単位ベクトル  $\mathbf{e}_k$  を使って、

$$\mathbf{q}_t = \mathbf{B}(\mathbf{q}_{t-1} + s_t \mathbf{e}_k)$$

と書ける。ノード  $v_m$  で観測される信号を  $\mathbf{r} = \{r_t\}$  で表すと、定義から  $r_t$  は

$$r_t = \mathbf{e}_m \mathbf{q}_t$$

と書ける。 $\mathbf{r}$  はネットワーク  $N$ 、入力信号  $\mathbf{s}$ 、入力ノード  $v_k$ 、そして観測ノード  $v_m$  に依存するので、厳密に言えば、 $\mathbf{r}(N, \mathbf{s}, v_k, v_m)$  と書くべきである。ノード  $v_k$  で観測される、入力  $\mathbf{s}$  に対する応答信号  $\epsilon(N, \mathbf{s}, v_k)$  は、上記の式を使って、

$$\epsilon(N, \mathbf{s}, v_k) = \mathbf{r}(N, \mathbf{s}, v_k, v_k)$$

と定義される。矩形波を入力とした場合の応答信号の例を図 2 に示す。矩形波は、周期  $p$  をパラメータとして以下のように定義される信号である：

$$s_t = \begin{cases} 1 & t \bmod p < p/2 \text{ である場合} \\ -1 & \text{その他の場合} \end{cases}$$

### 3.2 提案手法の詳細

#### 3.2.1 学習の基本方針

ネットワーク  $N$  から応答信号  $\epsilon(N, \mathbf{s}, v_k) = \{r_0, r_1, \dots\}$  が得られたとしよう。提案手法では、この応答信号の、あらかじめ決めた範囲  $I = [\alpha, \alpha + w - 1]$  の部分列  $\xi(N, \mathbf{s}, v_k, I) = (r_\alpha, \dots, r_{\alpha+w-1})$  で  $N$  を特徴づけ、学習に用いる。すなわち、 $\xi(N, \mathbf{s}, v_k, I)$  と  $N$  の属するクラスの対応関係を分類器に学習させる。この方式を、図 1 と対比させて図式化したものを図 3 に示す。以降、 $\xi(N, \mathbf{s}, v_k, I)$  を特徴ベクトルと呼ぶ。

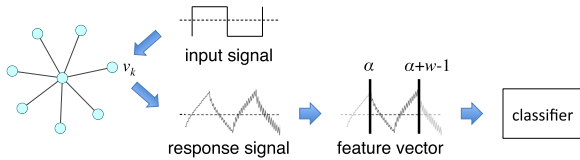


図 3 提案手法の基本的な手順

Fig. 3 Basic procedure of the proposed approach.

表 1 分類正解率の比較

Table 1 Comparison of accuracy of classification.

手法	kNN	NN	SVM
正解率	0.62	0.72	0.74

### 3.2.2 分類器

分類器には既存の機械学習手法を利用する。より分類性能の高い手法を採用するため、先行研究 [9] と同じ分類実験により、分類器の比較を行った。この実験は、完全グラフのクラス、2次元格子のクラスなどの7クラスのいずれかに属する52のネットワークから異なる74の応答信号を取得し\*1、それぞれどのクラスに属するネットワークから取得したかを推定するという問題である。本項ではその結果を示す。

使用した分類器は、k近傍法 (kNN)、ニューラルネットワーク (NN)、サポートベクトルマシン (SVM) の3種類である。入力信号は先行研究の実験条件と同じく周期50の矩形波、 $\alpha = 10$ ,  $w = 100$ とした。表1は、実験結果をまとめたものである。

k近傍法では、ある要素  $x$  の属するクラスを、 $x$  に近い  $k$  個の要素が属しているクラスから推定する。ここでは、 $x$  からの距離で重み付けをし、クラスを決定する方式を採用した。たとえば、 $k = 2$  の場合、 $x$  に近い2つの要素が異なるクラスに属している場合、距離の近い方を採用する。距離の尺度としては、L1 ノルム、L2 ノルム、相関係数の3つをそれぞれを用いた。相関係数は類似性が高いものほど大きな値となるため、その最大値である1との差を距離とした。実験の結果、相関係数を用い、 $k = 6$  の場合に最も高い正解率0.62が得られた。

ニューラルネットワークによる学習では、ユニット数  $u$  のAffine層を  $L$  層重ねたネットワークを使用し、確率  $d$  でドロップアウトを行った。図4は学習の進行状況を示したグラフであり、100エポックごとの正解率の変動を示したものである。比較のため、グラフにはハイパーパラメータ  $u$ ,  $L$ ,  $d$  を変えた場合を示した。 $L < 5$  の範囲では正解率の変動パターンが互いに類似しており、いずれも同程度の正解率 (0.72) に達しているが、さらに層の数が大きくなると、明らかな正解率の低下が認められる。なお、Affine型

\*1 異なるノードから同一の応答信号が得られる場合がある。たとえば、完全グラフではすべてのノードで同じ応答信号が得られる。このような重複を除いた結果、52のネットワークから74の信号が得られる。

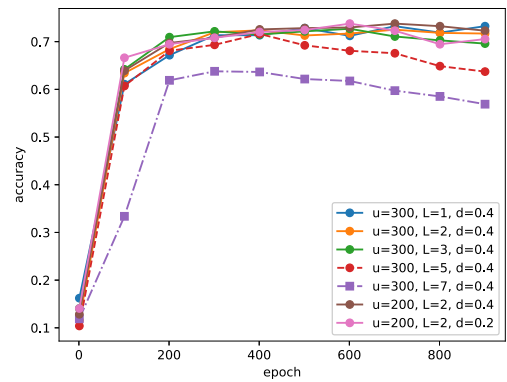


図 4 ニューラルネットワークによる分類の正解率

Fig. 4 Accuracy of classification by neural networks.

ネットワークの代わりにリカレントネットワーク (LSTM, GRU) を使い、応答信号を時系列として学習させたが、いずれの正解率もAffine型の結果には及ばなかった。

サポートベクトルマシンによる分類では、ガウシアンカーネルを用いた。この場合、誤分類の許容度を定める  $C$  と、カーネル関数のパラメータ  $\gamma$  という2つのハイパーパラメータにより分類性能を調整できる。グリッドサーチにより  $C$  と  $\gamma$  を探索した結果、 $C = 100$ ,  $\gamma = 0.008$  のときに最も高い正解率0.74が得られた。

以上の結果から、最も優れた性能を示したSVMを分類器として採用することにした。

### 3.2.3 信号観測ノードの選択

特徴ベクトル  $\xi(N, s, v_k, I)$  の値は信号観測ノード  $v_k$  に依存する。 $v_k$  を複数選び、それぞれの特徴ベクトルを導出すれば、そこからネットワーク  $N$  の構造に関するより多くの情報が得られる可能性がある。これらの情報は、学習においてはより多くの教師データとして、推定時には、アンサンブル学習の手がかりとして利用することができる。この考え方に基づき、本研究では、 $N$  の一部のノードからなる集合  $S$  を選び、 $\{\xi(N, s, v, I)\}$ ,  $v \in S$  を学習に使う。 $N$  のクラスを推定する際には、 $\xi(N, s, v, I)$  それぞれからクラスを推定し、多数決により結論を決める。なお、 $S$  の要素の選び方としては、中心性の高い一定数のノードを採用するという方法が考えられる。

### 3.2.4 学習のハイパーパラメータ

特徴ベクトル  $\xi(N, s, v_k, I)$  の値は、 $v_k$  以外にも、 $s$ ,  $I = [\alpha, \alpha + w - 1]$  に依存する。より具体的には、入力信号  $s$  のパラメータである、波形、周期、振幅、そして、応答信号のうち特徴ベクトルとして採用する範囲を定める  $\alpha$ ,  $w$  の値に応じて特徴ベクトルの値も変化し、その結果、分類の正解率も変化する可能性がある。つまり、これらのパラメータは、学習のハイパーパラメータとしてとらえることができる。



## 4. 評価実験

本章では、提案手法を評価する実験の内容と結果について述べる。

### 4.1 実験内容

本実験では、より実ネットワークに近いネットワークのクラスを題材にし、提案手法の分類性能を評価する。具体的には、BA モデルにより生成されたネットワークと、BA モデルにより生成されたものの一部を改変して得られたネットワークの分類問題を扱う。これらを識別できるということは、改変の有無を識別できるということを意味する。

なお、この問題は、以下に述べるリンクスパム [16] の状況の模擬を意図して作成した。Web ページの著作において、ページの作者は情報を取捨選択し、参照する価値があると判断したページを選んでリンクを張る。複数の Web ページ作者による、このような著作プロセスの結果として形成される Web ページのネットワークにはスケールフリー性があることが知られている [17], [18]。また、「リンクされている」という事実は、ページの価値の定量化にも使われている [19]。この定量化手法は、Web 検索のランキングに利用されるようになったが、それを逆にとり、特殊なリンク生成により間接的に Web 検索のランキングを操作する「検索エンジンを騙す」行為も行われるようになった。本来の自然な著作プロセスから逸脱して生成されるリンクがリンクスパムと呼ばれるものである。

### 4.2 データセット

実験に用いる 2 つのネットワークのクラス（集合）は、以下のようにして作成した。それぞれをクラス 1、クラス 2 と呼ぶ。クラス 1 の要素は、BA モデルに従って生成したネットワークである。BA モデルは、逐次的にノードとリンクを追加してネットワークを構成する生成モデルであり、新しいノードから、優先的選択によりすでに存在する  $\ell$  個のノードに対してリンクを生成する。本実験では、 $\ell = 5$  とし、ノード数が 30, 50, 70, 90, 110 のネットワークをそれぞれ 10 ずつ、合計 50 のネットワークを作成した。

クラス 2 に属するネットワークは、次のようにして作成した。まず、クラス 1 と同様にノード数 30, 50, 70, 90, 110 のネットワークを BA モデルにより作成した。次に、それぞれから  $h$  ノードを無作為に選び、それらがスターネットワークを形成するようにリンクを生成した。具体的には、 $h$  ノードから 1 つのノード  $s_0$  を無作為に選び、 $s_0$  と残りの  $(h - 1)$  ノードそれぞれの間リンクを生成することでスターネットワークを作成した。 $h$  には、 $6 \leq h \leq 15$  を満たす 10 個の値を選んだ。この手順により、クラス 1 と同数の 50 のネットワークを作成した。

クラス 2 のネットワークは、クラス 1 のネットワークに

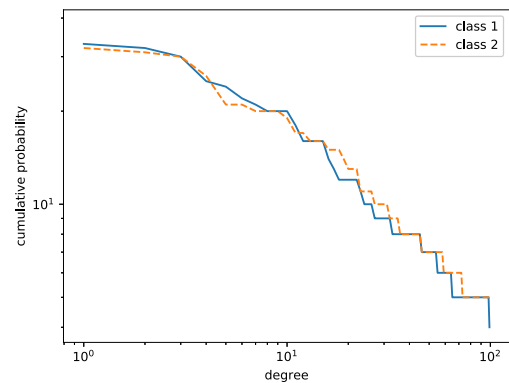


図 5 分類対象ネットワークの度数分布の比較

Fig. 5 Comparison of degree distributions of networks to be classified.

加工を施したネットワークである。その構造的な違いを既存のネットワーク指標を用いて判別することは容易ではない。たとえば、図 5 に示した、クラス 1, 2 の要素の度数分布は互いに似通っている。クラスター係数は、それぞれ 0.22, 0.21 であり、リンクを追加していないクラス 1 のネットワークでむしろ高い値となっている。

### 4.3 評価方法

前節で述べたネットワークの分類を、提案手法, NetSimile, node2vec, そして応答スペクトル法を使って行い、正解率を比較した。

提案手法では、3.2.2 項の結果に基づき分類器に SVM を使用した。また、入力信号には、先行研究と同様に振幅 1 の矩形波を採用した。応答信号を観測するノードには、度数中心性の高い上位  $m$  ノードを選んだ。本手法のハイパーパラメータは、SVM の  $C$  と  $\gamma$ 、入力信号の周期  $p$ 、応答信号の区間を定める  $\alpha$ 、 $w$ 、観測ノード数  $m$  である。これらの最適な値をグリッドサーチにより探索した。

NetSimile は、2.1 節で述べたように、シグネチャの Canberra 距離で構造的な近さを推定する。そこで、この Canberra 距離に基づく  $k$  近傍法により分類を行った。クラス推定に利用する近隣データの数  $k$  を変化させ最も高い正解率を探った。

node2vec により、ネットワークの各ノード  $v$  を  $n$  次元ベクトル  $\mathbf{z}(v)$  に対応付けることができる。本実験では、文献 [15] に倣い、ネットワーク  $N$  に  $n$  次元ベクトル  $\mathbf{z}(N) = \sum_{v \in N} \mathbf{z}(v)$  を対応付け、ネットワークの分類をベクトルの分類に帰着させた。ベクトルの分類にはガウシアンカーネルを用いた SVM を使用した。また、 $2 \leq n \leq 160$  の範囲で高い正解率をもたらす  $n$  を探索し、その結果を本手法の正解率とした。

応答スペクトル法については、文献 [9] で行われた実験と同じ条件、すなわち周期 50 の矩形波を入力信号とし、 $\alpha = 10$ 、 $w = 100$  という条件下での正解率を調べた。

表 2 分類正解率の比較

Table 2 Comparison of accuracy of classification.

手法	正解率	標準偏差	ハイパーパラメータ
NetSimile	0.60	0.00	$k = 8$
node2vec	0.53	0.05	$C = 1100.0,$ $\gamma = 0.00001, n = 124$
応答スペクトル法	0.49	0.04	$\alpha = 10, w = 100$
提案手法	0.69	0.01	$C = 0.5, \gamma = 10.0,$ $p = 2, \alpha = 10,$ $w = 200, m = 11$

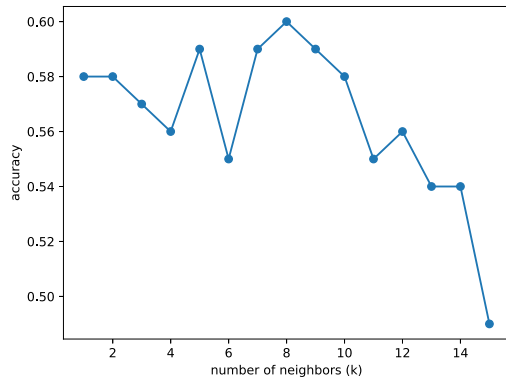


図 6 NetSimile による分類の正解率

Fig. 6 Accuracy of classification using NetSimile.

それぞれの分類手法の評価には leave-one-out 検証を用いた。すなわち、前節で述べた 100 のネットワークを 99 と 1 に分割し、99 を教師データに用い、残りの 1 つのネットワークのクラスを推定する。すべての分割の組合せに対して推定を行い、その結果の平均値を当該手法の 1 検証試行あたりの正解率とした。検証試行は複数回行い、その平均を最終的な正解率とし標準偏差を求めた。検証試行の回数は 20 回とした。なお、提案手法では、各ネットワークから  $m$  個の特徴ベクトルを導出して学習あるいは推定に使用するため、実質的には、他の手法の  $m$  倍のデータを使って学習と推定を行うことができる。

#### 4.4 評価結果

各手法により得られた正解率の最大値と、そのときの標準偏差およびハイパーパラメータの値を表 2 にまとめた。また、NetSimile、提案手法のそれぞれにおいて、 $k$  ( $k$  近傍法においてクラス推定に利用する近隣データの数)、 $m$  の変化とともに正解率の変動を図 6 と図 7 に示す。

### 5. 考察

#### 5.1 正解率の比較

本実験により、3 つの手法の中では、提案手法の正解率が最も高いことが確認できた。

提案手法において、 $m = 1$  のときに 0.46 と低い値であった正解率は、 $m$  の増加とともにゆらぎながらも増加し、

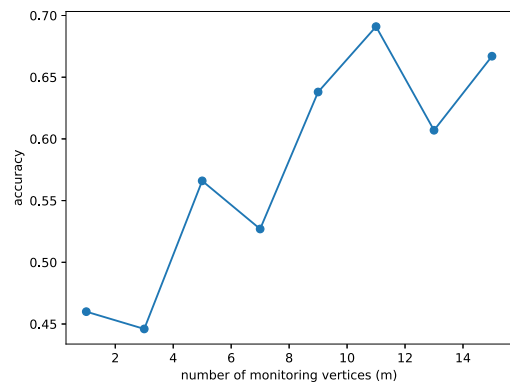


図 7 提案手法による分類の正解率

Fig. 7 Accuracy of classification by the proposed method.

$m = 11$  で最高値 0.69 が得られた。これは、複数ノードで応答信号を観測し学習に用いるアプローチが効果的であることを示している。本手法のもう 1 つの特徴はハイパーパラメータの多さである。これは、学習性能をチューニングする手段が多いということであるが、現在はグリッドサーチを使用しており、探索にコストがかかることが本手法の欠点となっている。

NetSimile の正解率は 6 割にとどまったが、NetSimile の分類性能が全般的に低いわけではないことに注意を促したい。たとえば、3.2.2 項において分類器の比較のために利用した、52 個のネットワークの 7 クラスへの分類問題に NetSimile を適用した場合には 9 割を超える高い正解率が得られることを確認している。この事実を考慮すると、今回の結果は問題の難しさを示していると考えられる。

node2vec の正解率は、無作為にクラスを選んだ場合よりもわずかに高い値であった。ネットワーク埋め込みの手法は、基本的に、各ノード間の関係を把握することを目的として開発されたものである。もともとネットワーク全体の特徴を把握する手法との比較は不利であることが予想されたが、それを裏付ける結果となった。

応答スペクトル法の正解率は 4 手法の中で一番低いものとなった。この正解率の低さも手法の問題に対する不適合が原因と考えられる。2.2 節で述べたように、応答スペクトル法は、ネットワークに混雑する構造を把握することを主要な目的の 1 つとして開発された手法である。一方、実験に使用したネットワークの生成法である BA モデルでは、ノードとリンクの追加を繰り返すことでネットワークを生成するが、そこで追加されるノードとリンクはスターネットワークを形成している。つまり、分類対象の 2 種類のネットワークには改変の有無という違いはあるものの、改変によって付け加えられた構造はもともとのネットワークの局所的構造とサイズを度外視すれば同じものとなっている。このような状況では、応答スペクトル法では二者の識別が困難であることを実験結果が示している。

表 3  $r(h)$  の平均の比較Table 3 Comparison of average of  $r(h)$ .

手法	$r(h)$ の平均	
	$6 \leq h \leq 10$	$11 \leq h \leq 15$
NetSimile	0.52	0.68
node2vec	0.49	0.58
提案手法	0.62	0.83

## 5.2 $h$ と正解率の関係

評価実験は、BA モデルにより生成されたネットワークと、BA モデルにより生成されたものの一部を改変して得られたネットワークを識別するというものであり、その改変の内容は、無作為抽出した  $h$  ノードがスターネットワークになるようリンクを追加するというものであった。 $h$  と正解率の関係を調べるため、改変されたネットワークのうち、改変されたことを正しく認識できたものの割合  $r(h)$  を調べた。 $r(h)$  の変動の傾向を把握する 1 つの方法として、 $h$  がとりうる値の範囲を  $6 \leq h \leq 10$  と  $11 \leq h \leq 15$  の 2 区間に分け、それぞれの範囲の  $r(h)$  の平均を比較した。表 3 は、その結果をまとめたものである。いずれの手法でも、 $h$  の値が大きい範囲で  $r(h)$  の平均がより高くなっていることが分かる。 $h$  の値が大きいということは改変の規模が大きいということであり、それゆえ識別が容易である（分類問題としてよりやさしい）ことが予想されるが、この結果はその予想を支持するものとなっている。また、これらの 2 区間それぞれで、各手法の  $r(h)$  の平均を比較すると、いずれの場合も提案手法が最も優れていることが分かる。

## 5.3 手法と問題の適合性

5.1 節でも触れたが、先行研究 [9] の評価および本論文で分類器を選択するためにも行った 7 クラスへの分類問題では、NetSimile の正解率は 9 割を超えており、提案手法を上回る性能を示している。NetSimile はクラスタ係数や度数分布など 7 種類のネットワーク指標に基づいて構造の類似性を判断している。よって、2 種類のネットワークを比較する際、その違いがこれらの基本的な指標で識別しやすいものである場合には NetSimile は優れた性能を示すことが期待できる。7 クラスへの分類問題で分類対象となったネットワークは、そのネットワーク全体を、完全グラフや 2 次元格子などといった 1 つの構造的特徴として把握できるものであった。これらのネットワークは基本的なネットワーク指標での識別が比較的容易なものであり、このことが、NetSimile の高い正解率に寄与していると考えられる。

逆に、これらの指標で識別が難しい場合には、NetSimile の性能は低下すると考えられる。たとえば、もし、異なる 2 つのネットワークを NetSimile が着目している 7 つの指標で比較し違いが見い出せなければ、NetSimile はそれらは同一のネットワークと判断してしまうことになる。本論文

の評価実験で比較する 2 種類のネットワークには、4.2 節でも述べたとおり、クラスタ係数や度数分布に明確な違いが認められない。NetSimile が判断基準として用いている指標で違いを把握することが困難であることが、NetSimile の正解率が低い理由の 1 つとしてあげられる。一方、提案手法は特定の指標に依存していない。このことが、提案手法の優位性をもたらしていると考えられる。

このように、分類対象とするネットワークの特性により手法の適/不適があるため、問題ごとに適切な手法を選ぶ必要がある。さらに様々な分類問題で評価を行い、手法選択の判断基準を明確にすることが今後の課題である。

## 6. むすび

本論文では、サイズに非依存なネットワークの構造的特徴を教師データから学習し分類問題を解く手法を提案し、分類性能を NetSimile などの既存手法と比較することで、その有効性を確認した。提案手法では、各ネットワークに属するノードの集合  $S$  から応答信号を取得し、特徴ベクトルを作成して学習と推定に用いる。このようにして、学習と推定に用いることができるデータの量を  $|S|$  倍に増やせること、また、入力信号のパラメータも学習のハイパーパラメータとし、分類性能の調整に使えることが本手法の特徴である。

本手法ではハイパーパラメータが多い分探索にコストがかかるという問題がある。効率的な探索方法の開発が課題である。また、本論文で評価実験に用いた問題だけでなく、さらに様々な分類問題で評価を行うことで、提案手法の適用範囲を明らかにすることも今後の課題である。

本手法は、ネットワークの分類例をもとに分類基準を学ぶ。学習は構造的特徴を手がかりに行われるが、評価実験のデータセットがそうであったように、例となる分類は構造的特徴から直接的に導かれるものである必要はない。1 章で述べたように、多くのシステムでは、その機能と要素間の関係を示すネットワークの間には対応関係がある。機能の違いからシステムの分類例を作り、その背景にある、対応する構造の違いを本手法で学習するというアプローチが考えられる。機能的な観点からの分類への応用が期待される。

## 参考文献

- [1] Strogatz, S.H.: Exploring complex networks, *Nature*, Vol.410, No.6825, pp.268–276 (2001).
- [2] Costa, L., Rodrigues, F., Traverso, G. and Boas, P.: Characterization of Complex Networks: A Survey of Measurements, *Advances in Physics*, Vol.56, No.1, pp.167–242 (2007).
- [3] Kleinberg, J.: Authoritative sources in a hyperlinked environment, *J. ACM*, Vol.46, No.5, pp.604–632 (1999).
- [4] Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D. and Alon, U.: Network motifs: Simple



- building blocks of complex networks, *Science*, Vol.298, No.5594, pp.824–827 (2002).
- [5] Bullmore, E. and Sporns, O.: Complex brain networks: Graph theoretical analysis of structural and functional systems, *Nature Reviews Neuroscience*, Vol.10, No.3, pp.186–198 (2009).
- [6] Albert, R. and Barabási, A.-L.: Statistical mechanics of complex networks, *Reviews of Modern Physics*, Vol.74, pp.47–97 (2002).
- [7] Berlingerio, M., Koutra, D., Eliassi-Rad, T. and Faloutsos, C.: Network similarity via multiple social theories, *Proc. 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pp.1439–1440, ACM (2013).
- [8] Lance, G.N. and Williams, W.T.: Computer Programs for Hierarchical Polythetic Classification (“Similarity Analyses”), *The Computer Journal*, Vol.9, No.1, pp.60–64 (1966).
- [9] 佐藤進也：ネットワークの応答信号に基づく構造解析法の提案, 情報処理学会論文誌, Vol.58, No.6, pp.1226–1233 (2017).
- [10] Hamilton, W.L., Ying, R. and Leskovec, J.: Representation Learning on Graphs: Methods and Applications, *IEEE Data Eng. Bull.*, Vol.40, No.3, pp.52–74 (2017).
- [11] Perozzi, B., Al-Rfou, R. and Skiena, S.: DeepWalk: Online Learning of Social Representations, *Proc. 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp.701–710, ACM (2014).
- [12] Grover, A. and Leskovec, J.: node2vec: Scalable Feature Learning for Networks, *Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp.855–864 (2016).
- [13] Yang, Z., Cohen, W.W. and Salakhutdinov, R.: Revisiting Semi-supervised Learning with Graph Embeddings, *Proc. 33rd International Conference on Machine Learning - Volume 48*, JMLR.org, pp.40–48 (2016).
- [14] Huang, X., Li, J. and Hu, X.: Label Informed Attributed Network Embedding, *Proc. 10th ACM International Conference on Web Search and Data Mining*, pp.731–739, ACM (2017).
- [15] Duvenaud, D.K., Maclaurin, D., Iparraguirre, J., Bombarell, R., Hirzel, T., Aspuru-Guzik, A. and Adams, R.P.: Convolutional Networks on Graphs for Learning Molecular Fingerprints, *Advances in Neural Information Processing Systems 28*, Cortes, C., Lawrence, N.D., Lee, D.D., Sugiyama, M. and Garnett, R. (Eds.), pp.2224–2232, Curran Associates, Inc. (2015).
- [16] Gyöngyi, Z. and Garcia-Molina, H.: Link Spam Alliances, *Proc. 31st International Conference on Very Large Data Bases*, pp.517–528, VLDB Endowment (2005).
- [17] Albert, R., Jeong, H. and Barabási, A.-L.: Diameter of the World-Wide Web, *Nature*, Vol.401, No.6749, pp.130–131 (1999).
- [18] Kumar, R., Raghavan, P., Rajagopalan, S. and Tomkins, A.: Trawling the Web for Emerging Cyber-communities, *Proc. 8th International Conference on World Wide Web*, pp.1481–1493, Elsevier North-Holland, Inc. (1999).
- [19] Brin, S. and Page, L.: The Anatomy of a Large-scale Hypertextual Web Search Engine, *Proc. 7th International Conference on World Wide Web*, Amsterdam, The Netherlands, pp.107–117, Elsevier Science Publishers B.V. (1998).



佐藤 進也 (正会員)

1986年東北大学理学部数学科卒業。1988年同大学大学院修士課程修了。同年日本電信電話株式会社入社。NTT未来ねっと研究所等を経て、2014年日本工業大学工学部情報工学科教授、2018年より先進工学部情報メディア工学科教授。博士（情報理工学）。協調作業支援、Web情報検索・マイニング、複雑ネットワーク等の研究に従事。訳書『スモールワールド』（ダンカン・ワッツ著、東京電機大学出版局、共訳）。ACM、ISOC、電子情報通信学会、人工知能学会各会員。