CF-005

多層ニューラルネットにおける 正負の結合重みに基づく大局構造抽出

渡邊千紘 †，平松薫 †，柏野邦夫 †

†NTT コミュニケーション科学基礎研究所 〒243-0198 神奈川県厚木市森の里若宮 3-1

概要

多層ニューラルネットは画像処理，音声認識，バイオインフォマティクスなど，様々な応用において非常に優れた認識・予測性能を実現してきた．しかし，多層ニューラルネットによる推論は，多くの非線形なパラメータの階層的につながった複雑な関係構造で表現されるため，人間がその内部表現を理解したり，知識を発見することは難しい．そこで，我々はネットワーク解析に基づき，学習後の多層ニューラルネットの大局的な構造を抽出する手法を提案した [19]．この手法では，重みの絶対値が大きい結合のリンク情報を用いて，似た結合パターンを持つユニットを同一コミュニティ (グループ) に分類することにより，元のニューラルネットを単純化したモジュール構造を抽出することを可能にしたが，一方で正負の結合を区別せずに扱う確率モデルを用いているため，複雑な隠れ構造を持つニューラルネットに対してはコミュニティ抽出の精度が十分ではなかった．本研究では，正負の結合を区別した新たな確率モデルに基づくモジュール構造抽出法を提案する．提案法を用いることにより，多層ニューラルネットに隠された構造を既存手法よりも精度良く発見することができ，またニューラルネットの推論について様々な知識が得られることを示す．

キーワード：多層ニューラルネットワーク，ネットワーク解析，コミュニティ抽出

1　はじめに

多層ニューラルネットは様々な分野における認識・予測の問題に対し，圧倒的な性能を実現している [1, 14]．例えば，画像処理 [13, 18] や音声認識 [3, 9]，バイオインフォマティクス [6, 4] の分野においては，多層ニューラルネットを用いることで，従来の手法を大幅に上回る精度で予測を行うことが可能になった．多層ニューラルネットにおいて，入出力データは非常に多くのユニットと，ユニット同士をつなぐ結合からなる階層的な構造で表現され，これにより実世界に存在する多くの複雑なデータを知覚し，認識することが可能となった．

このように，多層ニューラルネットでは多くの非線形なパラメータが絡み合うことにより，複雑な実データの表現が可能となった反面，その推論の構造を人間が理解することが困難であるという課題が存在する．この課題に対する 1 つの従来のアプローチとして，主に画像データを入力とする畳み込みニューラルネットに対し，各中間ユニットを活性化する入力画像を提示する手法 [22] や，出力値に応じて中間ユニット同士を分類し，各クラスタのセントロイドに近いユニットの出力を描画する手法 [7] など，可視化に基づくアプローチが提案されている．これらの手法では，各ユニットが持つ情報を詳細に見ることができる反面，ニューラルネット全体の推論を大局的に捉えることは難しい．一方，ニューラルネットのパラメータ圧縮や計算の高速化を目的として，その推論の構造を単純化して表現する手法が提案されており，これらの手法を用いてニューラルネットの推論の解釈性を向上することが考えられる．このような手法は，大まかに以下の 3 つのアプローチが存在する．

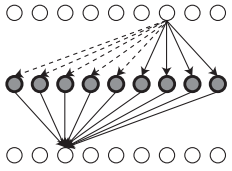

図 1.　結合重みの正負を区別しない既存手法の問題点．図は多層ニューラルネットの一部であり，実線は正の，点線は負の重みを表す．中央に描かれた層に含まれる 9 個のユニットは，隣接する層のユニットに対し，異なる符号の結合重みを持つにもかかわらず，既存手法では同一コミュニティに割り当てられてしまう．

(1) 結合の枝刈り：重みが小さい結合を冗長なものとみなし，削除する方法である．例えば，閾値以下の重みを持つ結合の枝刈りとニューラルネットの再学習を繰り返すことにより，パラメータを圧縮する手法 [8] や，さらに残った重みに対し Huffman 符号化を適用する手法 [11] などが提案されている．これらの手法により推論の解釈が可能な程度にニューラルネットを単純化しようとすると，非常に少数の結合を残すようにハイパーパラメータを設定する必要があり，汎化誤差が大きくなってしまう．逆に，汎化誤差を小さく保つには多数の結合を残す必要があり，推論の解釈が難しくなる．

(2) 低ランク近似：畳み込みニューラルネットにおける畳み込み処理の行列演算を，低ランクな行列の組み合わせで近似する手法である [5, 2]．これらの手法は畳み込み行列演算の高速化を目的としたものであり，ネットワーク構造自体の簡略化を実現するものではなかった．

(3) モデル構造学習：1 層のユニットを 1 つのグループとして捉え，一度に正規化を行うことで，効率的に最適なニューラルネットの構造を探す手法である [21]．層の数や畳み込みニューラルネットにおけるフィルタのパ

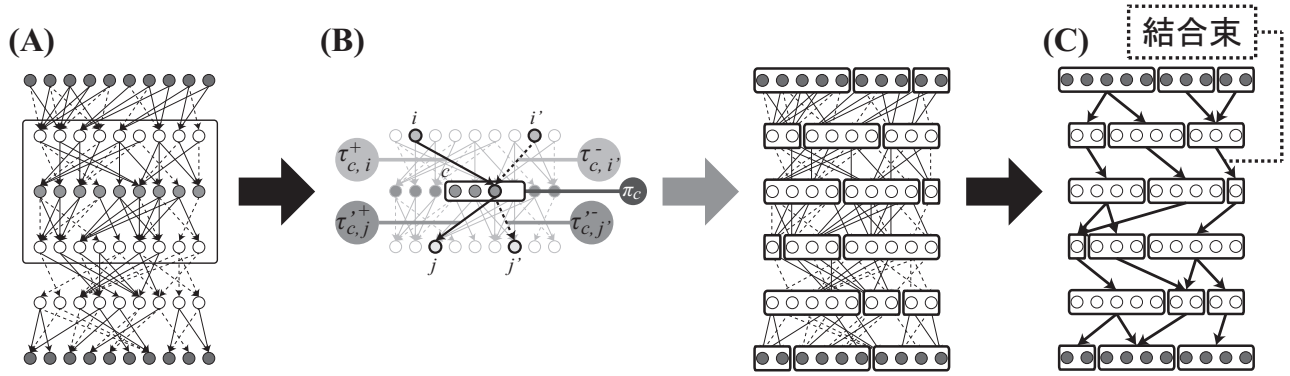

図 2. 提案手法：(A) バックプロパゲーションに基づく多層ニューラルネットの学習，(B) 学習したニューラルネットからのコミュニティ抽出，(C) 結合束の定義に基づくモジュール構造抽出により構成される．実線は正の，点線は負の重みを表す．

ラメータを含めて最適化を行う手法も存在する [10]．これらの手法においては，1 層のユニットをまとめて扱うため，(1) の手法と比べて一気に構造を単純化することができるが，一方で以下に説明する提案法のように，1 層の中での役割の違いによってユニットを分類したり，それらのグループ間の関係を見ることはできなかった．

そこで，我々は深層学習による推論を単純化し解釈しやすくするための新たなアプローチとして，学習した多層ニューラルネット (図 2 (A)) に対し，ネットワーク解析を行うことにより，その大局的な構造を抽出する方法 [19] を提案した．これは，多層ニューラルネットの各層において，隣接する層のユニットへの似た結合パターンを持つユニットの集合を検出し，同じグループ (ネットワーク解析においては，**コミュニティ**と呼ばれる) に割り当てるアルゴリズム (図 2 (B)) に基づき，さらに抽出されたコミュニティ間に存在する複数の結合を**結合束**を用いてまとめて表示する (図 2 (C)) ことにより実現された．この手法により，元の複雑なニューラルネットを単純化した表現 (本研究では，**モジュール構造**と呼ぶ) を得ることが可能になり，学習後のニューラルネットの推論の解釈可能性を向上させた．しかしながら，この手法では重みの絶対値の情報のみを用い，その正負を区別せずに扱っていたため，複雑な隠れ構造を持つニューラルネットに対しては適切なコミュニティ検出が行えない場合が存在した (図 1，4.1 節)．

そこで，本研究では，学習した多層ニューラルネットにおける正負の結合を区別し，異なる結合確率を表すパラメータを導入した新たな確率モデルに基づき，モジュール構造の抽出を行う手法を提案する．この手法を用いることで，複雑な隠れ構造を持つ多層ニューラルネットから生成された人工データを用いて実験を行うことにより，提案法が既存手法と比べてコミュニティ抽出の観点からより高精度に真の構造を発見できる場合があることを示す．また，実データを用いて学習した多層ニューラルネットに対し，提案法を適用してモジュール構造を抽出することにより，ニューラルネットが学習した入出力間の関係構造について，様々な知識を読み取ることができることを示す．

2　多層ニューラルネットの学習

本節で用いられる多層ニューラルネットや学習法は，既存研究 [19] と同じものである．$x \in \mathbb{R}^M$, $y \in \mathbb{R}^N$ をそれぞれ，入力，出力データとし，x, y の確率密度関数 $q(x, y)$ を $\mathbb{R}^M \times \mathbb{R}^N$ 上の関数とする．サンプルサイズ n の学習データの集合 $\{(X_i, Y_i)\}_{i=1}^n$ は，$q(x, y)$ に従い独立に生成されているものと仮定する．多層ニューラルネットでは，入力 $x \in \mathbb{R}^M$ とパラメータ $w \in \mathbb{R}^L$ から $\mathbb{R}^N$ への関数 $f(x, w)$ を用いて出力 y の推定を行う．

多層ニューラルネットのパラメータ $w = \{\omega_{ij}^d, \theta_i^d\}$ は，深さ d の層における i 番目のユニットと深さ $d+1$ の層における j 番目のユニットとの間の結合重み ω_{ij}^d と，深さ d の層における i 番目のユニットのバイアス θ_i^d からなる．D 層からなるニューラルネットは，シグモイド関数 $\sigma(x) = 1/(1 + \exp(-x))$ を用いて以下の関数で与えられる．

$$
\begin{aligned}
f_i(x, w) &= \sigma(\sum_j \omega_{ij}^{D-1} o_j^{D-1} + \theta_i^{D-1}), \\
o_i^{D-1} &= \sigma(\sum_j \omega_{ij}^{D-2} o_j^{D-2} + \theta_i^{D-2}), \quad \cdots, \\
o_i^2 &= \sigma(\sum_j \omega_{ij}^1 x_j + \theta_i^1).
\end{aligned}
$$

ここで，$\|\cdot\|$ を $\mathbb{R}^N$ 上のユークリッドノルムとすると，学習誤差 $E(w)$ と汎化誤差 $G(w)$ はそれぞれ，

$$
\begin{aligned}
E(w) &= \frac{1}{n} \sum_{i=1}^n \|Y_i - f(X_i, w)\|^2, \\
G(w) &= \int \|y - f(x, w)\|^2 q(x, y) dx dy.
\end{aligned}
$$

と定義される．学習データ集合とは独立に，$q(x, y)$ に従って生成されるテストデータ集合を $\{(X_j', Y_j')\}_{j=1}^m$

とすると，汎化誤差は以下の式で近似される．

$$G(w) \approx \frac{1}{m}\sum_{j=1}^{m} \|Y_j{}' - f(X_j{}', w)\|^2.$$

LASSO の手法 [12, 17] は，以下の $H(w)$ により定義される目的関数を最小化する手法であり，これを用いることで学習結果として疎なニューラルネットが得られる．

$$H(w) = n\left(\frac{1}{2}\,E(w) + \lambda \sum_{d,i,j} |\omega_{ij}^{d}|\right).$$

ただし，λ はハイパーパラメータである．パラメータは確率的最急降下法により学習される．$H_i(w)$ を i 番目のサンプル (X_i, Y_i) のみから計算される学習誤差とすると，パラメータは以下の式に従って更新される．

$$\begin{aligned}\Delta w &= -\eta \nabla H_i(w) \\ &= -\eta\left(\frac{1}{2}\nabla\{\|Y_i - f(X_i, w)\|^2\} + \lambda\,\mathrm{sgn}(w)\right). \quad (1)\end{aligned}$$

ここで，η を学習時間 t に対し $\eta(t) \propto \frac{1}{t}$ と定義することで，確率的最急降下法の収束が保証される．式 (1) は [19] の Algorithm 1 に従い計算することができ，この学習手法はバックプロパゲーション [20, 16] と呼ばれる．LASSO の手法を用いることにより，冗長な結合は重みの絶対値が 0 に近づくように学習される．

3 多層ニューラルネットにおける正負の結合重みに基づく大局構造抽出

我々は，ネットワーク解析に基づき，多層ニューラルネットからその大局的な構造を抽出し，解釈可能性を向上させるためのアルゴリズムを提案する．本手法は，多層ニューラルネットの各層に対し，似た結合パターンを持つユニットを分類するコミュニティ抽出法 [15] を適用した既存手法 [19] を拡張したものである．

3.1 正負の結合重みに基づくコミュニティ抽出法

入出力データを用いて学習した多層ニューラルネットの，深さ d の層について考える．ただし，入力層を $d = 1$ とする．深さ d の層と，隣接する層との間の結合関係をそれぞれ，隣接行列 $A^{+,d} = \{A_{ik}^{+,d}\}$, $A^{-,d} = \{A_{ik}^{-,d}\}$, $B^{+,d} = \{B_{kj}^{+,d}\}$, $B^{-,d} = \{B_{kj}^{-,d}\}$ で表す．ここで，$A^{+,d}(A^{-,d})$ の各要素 $A_{ik}^{+,d}(A_{ik}^{-,d})$ は，深さ $d-1$ の層における i 番目のユニットと，深さ d の層における k 番目のユニットとの間にある結合重みが ξ 以上 ($-\xi$ 以下) ならば 1, そうでなければ 0 として定義する．同様に，$B^{+,d}(B^{-,d})$ の各要素 $B_{kj}^{+,d}(B_{kj}^{-,d})$ を，深さ d の層における k 番目のユニットと，深さ $d+1$ の層における j 番目のユニットとの間にある結合重みから定義する．簡単のため，$A^{+,d}, A^{-,d}, B^{+,d}, B^{-,d}$ を $A^{+}, A^{-}, B^{+}, B^{-}$ と表記する．

提案手法は，同一コミュニティに属するユニットであれば，隣接する層に対して一定の正負の結合確率を持つという仮定に基づく．深さ d において，各コミュニティ c にユニットが属する確率を $\pi = \{\pi_c\}$, コミュニティ c のユニットに入力される正 (負) の結合の結合元が深さ $d-1$ の層における i 番目のユニットである確率を $\tau^{+} = \{\tau_{c,i}^{+}\}(\tau^{-} = \{\tau_{c,i}^{-}\})$, コミュニティ c のユニットから出力される正 (負) の結合の結合先が深さ $d+1$ の層における j 番目のユニットである確率を $\tau'^{+} = \{\tau_{c,j}'^{+}\}(\tau'^{-} = \{\tau_{c,j}'^{-}\})$ とする (図 2 (B))．これらのパラメータは，確率の和が 1 であるから，

$$\sum_c \pi_c = 1. \quad \sum_i \tau_{c,i}^{+} = 1. \quad \sum_i \tau_{c,i}^{-} = 1.$$
$$\sum_j \tau_{c,j}'^{+} = 1. \quad \sum_j \tau_{c,j}'^{-} = 1. \qquad (2)$$

多層ニューラルネットの結合関係から与えられる隣接行列 $A^{+}, A^{-}, B^{+}, B^{-}$ の尤度を最大化するパラメータ $\pi, \tau^{+}, \tau^{-}, \tau'^{+}, \tau'^{-}$ を求める．ここで，深さ d の層における k 番目のユニットが属するコミュニティを g_k として，集合 $g = \{g_k\}$ を導入し，$A^{+}, A^{-}, B^{+}, B^{-}$ と g の尤度を最大化する．$A \equiv \{A^{+}, A^{-}\}, B \equiv \{B^{+}, B^{-}\}, \tau \equiv \{\tau^{+}, \tau^{-}\}, \tau' = \{\tau'^{+}, \tau'^{-}\}$ とおくと，

$$\Pr(A, B, g|\pi, \tau, \tau') = \Pr(A, B|g, \pi, \tau, \tau')\Pr(g|\pi, \tau, \tau').$$

ここで，

$$\Pr(A, B|g, \pi, \tau, \tau') = \prod_k \Big\{\prod_i \left(\tau_{g_k,i}^{+}\right)^{A_{i,k}^{+}} \left(\tau_{g_k,i}^{-}\right)^{A_{i,k}^{-}}\Big\} \Big\{\prod_j \left(\tau_{g_k,j}'^{+}\right)^{B_{k,j}^{+}} \left(\tau_{g_k,j}'^{-}\right)^{B_{k,j}^{-}}\Big\},$$

$$\Pr(g|\pi, \tau, \tau') = \prod_k \pi_{g_k}.$$

上式を代入することにより，与えられたパラメータに対する A, B, g の対数尤度 $\mathcal{L}$ は以下の式で与えられる．

$$\begin{aligned}\mathcal{L} &= \ln \Pr(A, B, g|\pi, \tau, \tau') \\ &= \sum_k \Big\{\ln \pi_{g_k} + \sum_i (A_{i,k}^{+} \ln \tau_{g_k,i}^{+} + A_{i,k}^{-} \ln \tau_{g_k,i}^{-}) \\ &+ \sum_j (B_{k,j}^{+} \ln \tau_{g_k,j}'^{+} + B_{k,j}^{-} \ln \tau_{g_k,j}'^{-})\Big\}.\end{aligned}$$

対数尤度 $\mathcal{L}$ の隠れ変数 g に関する期待値 $\bar{\mathcal{L}}$ は，以下の式で与えられる．

$$\begin{aligned}\bar{\mathcal{L}} &= \sum_{g_1} \cdots \sum_{g_l} \Pr(g|A, B, \pi, \tau, \tau')\mathcal{L} \\ &= \sum_{k,c} q_{k,c}\Big\{\ln \pi_c + \sum_i (A_{i,k}^{+} \ln \tau_{g_k,i}^{+} + A_{i,k}^{-} \ln \tau_{g_k,i}^{-}) \\ &+ \sum_j (B_{k,j}^{+} \ln \tau_{g_k,j}'^{+} + B_{k,j}^{-} \ln \tau_{g_k,j}'^{-})\Big\}. \qquad (3)\end{aligned}$$

ただし，l を深さ d の層におけるユニット数とし，$q = \{q_{k,c}\}$ を以下で定義した．

$$q_{k,c} = \Pr(g_k = c|A,B,\pi,\tau,\tau') = \frac{\Pr(A,B,g_k=c|\pi,\tau,\tau')}{\Pr(A,B|\pi,\tau,\tau')}. \quad (4)$$

これは，k 番目のユニットがコミュニティ c に属する確率を表す．$\bar{\mathcal{L}}$ を最大化するパラメータ q,π,τ,τ' は，EM 法により求められる．

定理 3.1. パラメータ q,π,τ,τ' が $\bar{\mathcal{L}}$ を最大化するならば，以下の等式を満たす．

$$q_{k,c} = \Big\{\pi_c\Big[\prod_i \big(\tau^+_{g_k,i}\big)^{A^+_{i,k}}\big(\tau^-_{g_k,i}\big)^{A^-_{i,k}}\Big]\Big[\prod_j \big(\tau'^+_{g_k,j}\big)^{B^+_{k,j}} \big(\tau'^-_{g_k,j}\big)^{B^-_{k,j}}\Big]\Big\}/\Big\{\sum_s \pi_s\Big[\prod_i \big(\tau^+_{s,i}\big)^{A^+_{i,k}}\big(\tau^-_{s,i}\big)^{A^-_{i,k}}\Big] \Big[\prod_j \big(\tau'^+_{s,j}\big)^{B^+_{k,j}}\big(\tau'^-_{s,j}\big)^{B^-_{k,j}}\Big]\Big\}, \quad (5)$$

$$\pi_c = \frac{\sum_k q_{k,c}}{l},$$
$$\tau^+_{c,i} = \frac{\sum_k q_{k,c}A^+_{i,k}}{\sum_{k,i} q_{k,c}A^+_{i,k}},\quad \tau^-_{c,i} = \frac{\sum_k q_{k,c}A^-_{i,k}}{\sum_{k,i} q_{k,c}A^-_{i,k}},$$
$$\tau'^+_{c,j} = \frac{\sum_k q_{k,c}B^+_{k,j}}{\sum_{k,j} q_{k,c}B^+_{k,j}},\quad \tau'^-_{c,j} = \frac{\sum_k q_{k,c}B^-_{k,j}}{\sum_{k,j} q_{k,c}B^-_{k,j}}. \quad (6)$$

証明. 式 (4) の分子と分母はそれぞれ，Kronecker のデルタ $\delta_{i,j}$ を用いて以下のように書き直せる．

$$\Pr(A,B,g_k=c|\pi,\tau,\tau') = \sum_{g_1}\cdots\sum_{g_l}\delta_{g_k,c}\Pr(A,B,g|\pi,\tau,\tau') = \sum_{g_1}\cdots\sum_{g_l} \delta_{g_k,c}\prod_h\Big\{\pi_{g_h}\Big[\prod_i\big(\tau^+_{g_h,i}\big)^{A^+_{i,h}}\big(\tau^-_{g_h,i}\big)^{A^-_{i,h}}\Big] \Big[\prod_j\big(\tau'^+_{g_h,j}\big)^{B^+_{h,j}}\big(\tau'^-_{g_h,j}\big)^{B^-_{h,j}}\Big]\Big\}$$
$$= \Big\{\pi_c\Big[\prod_i\big(\tau^+_{c,i}\big)^{A^+_{i,k}}\big(\tau^-_{c,i}\big)^{A^-_{i,k}}\Big]\Big[\prod_j\big(\tau'^+_{c,j}\big)^{B^+_{k,j}} \big(\tau'^-_{c,j}\big)^{B^-_{k,j}}\Big]\Big\}\Big\{\prod_{h\neq k}\sum_s \pi_s\Big[\prod_i\big(\tau^+_{s,i}\big)^{A^+_{i,h}}\big(\tau^-_{s,i}\big)^{A^-_{i,h}}\Big] \Big[\prod_j\big(\tau'^+_{s,j}\big)^{B^+_{h,j}}\big(\tau'^-_{s,j}\big)^{B^-_{h,j}}\Big]\Big\},$$
$$\Pr(A,B|\pi,\tau,\tau') = \sum_{g_1}\cdots\sum_{g_l}\Pr(A,B,g|\pi,\tau,\tau') = \prod_k\sum_s \pi_s\Big[\prod_i\big(\tau^+_{s,i}\big)^{A^+_{i,k}}\big(\tau^-_{s,i}\big)^{A^-_{i,k}}\Big] \Big[\prod_j\big(\tau'^+_{s,j}\big)^{B^+_{k,j}}\big(\tau'^-_{s,j}\big)^{B^-_{k,j}}\Big].$$

したがって，$q_{k,c}$ は式 (5) で与えられる．

また，与えられた $\{q_{k,c}\}$ に対して $\bar{\mathcal{L}}$ を最大化する π,τ,τ' は，Lagrange の未定乗数法により求められる．

$$f = \bar{\mathcal{L}} - \alpha\sum_c \pi_c - \sum_c\Big(\beta^+_c\sum_i\tau^+_{c,i} + \beta^-_c\sum_i\tau^-_{c,i}\Big) - \sum_c\Big(\gamma^+_c\sum_j\tau'^+_{c,j} + \gamma^-_c\sum_j\tau'^-_{c,j}\Big),$$
$$\frac{\partial f}{\partial\pi_c} = \frac{\partial f}{\partial\tau^+_{c,i}} = \frac{\partial f}{\partial\tau^-_{c,i}} = \frac{\partial f}{\partial\tau'^+_{c,j}} = \frac{\partial f}{\partial\tau'^-_{c,j}} = 0. \quad (7)$$

とすると，以下の等式が導かれる．

$$\frac{\partial\bar{\mathcal{L}}}{\partial\pi_c} = \alpha,\quad \frac{\partial\bar{\mathcal{L}}}{\partial\tau^+_{c,i}} = \beta^+_c,\quad \frac{\partial\bar{\mathcal{L}}}{\partial\tau^-_{c,i}} = \beta^-_c,$$
$$\frac{\partial\bar{\mathcal{L}}}{\partial\tau'^+_{c,j}} = \gamma^+_c,\quad \frac{\partial\bar{\mathcal{L}}}{\partial\tau'^-_{c,j}} = \gamma^-_c. \quad (8)$$

式 (3) と式 (8) より，以下の等式が得られる．

$$\pi_c = \frac{1}{\alpha}\sum_k q_{k,c},\quad \tau^+_{c,i} = \frac{1}{\beta^+_c}\sum_k q_{k,c}A^+_{i,k},$$
$$\tau^-_{c,i} = \frac{1}{\beta^-_c}\sum_k q_{k,c}A^-_{i,k},\quad \tau'^+_{c,j} = \frac{1}{\gamma^+_c}\sum_k q_{k,c}B^+_{k,j},$$
$$\tau'^-_{c,j} = \frac{1}{\gamma^-_c}\sum_k q_{k,c}B^-_{k,j}. \quad (9)$$

式 (9) と式 (2) の正規化条件より，Lagrange の未定乗数 α, $\{\beta^+_c\}$, $\{\beta^-_c\}$, $\{\gamma^+_c\}$, $\{\gamma^-_c\}$ が求まり，式 (9) は式 (6) のように書き直せる． □

定理 3.1 より，最適なパラメータ π,τ,τ' と対応するコミュニティ割り当ての確率 q は，式 (5) と式 (6) に基づいて繰り返し推定される．本稿では，深さ d の層における k 番目のユニットが属するコミュニティは，$q_{k,c}$ を最大化するコミュニティ c として定義する．

3.2 多層ニューラルネットにおけるモジュール構造

前節で得られたコミュニティ構造から，既存手法 [19] に基づいて多層ニューラルネットのモジュール構造を定義する．これは，隣接する層のコミュニティ間に存在する複数の結合をまとめて 1 本の結合として表す (本稿では，これを結合束と呼ぶ) ことで，多層ニューラルネットの構造を単純化することによる (図 2 (C))．この手法では，コミュニティ間に存在する結合の本数の割合に基づいて結合束の有無が定義される．

4 実験

人工データ (4.1 節) と実データ (4.2 節) を用いて学習した多層ニューラルネットに対し，3 節の手法を適用し，モジュール構造を抽出した．以下に実験設定を記す．

1. 入出力データの正規化，パラメータの初期値設定，隣接行列の定義，モジュール構造抽出結果の可視化は [19] の実験における設定 (1), (2), (3), (4), (6),

(7) を用いている．

2. 学習した多層ニューラルネットの各層に対し，反復数 300 の EM アルゴリズムを 100 回行い，最終反復において尤度の期待値が最大となった回の結果をコミュニティ抽出の結果として定義した．
3. ハイパーパラメータの設定は以下の通りとした．
   - 学習データ数 n : 5000 (実験 1), 1905 (実験 2).
   - テストデータ数 m : 5000 (実験 1), 0 (実験 2).
   - 入力層，隠れ層，出力層のユニット数 : $\{45, 45, 45, 45\}$ (実験 1), $\{24, 20, 20, 18\}$ (実験 2).
   - 深層学習における 1 つのデータに対する平均の学習回数 a_1 : 2000.
   - ステップサイズ $\eta = 0.8 \times \frac{a_1 \times n}{a_1 \times n + 5 \times t}$.
   - 入力層から出力層までの各層間における LASSO のハイパーパラメータ λ : $\{1.0 \times 10^{-5}, 1.0 \times 10^{-5}, 2.5 \times 10^{-6}\}$ (実験 1), $\{1.0 \times 10^{-6}, 1.0 \times 10^{-6}, 1.0 \times 10^{-6}\}$ (実験 2).
   - 結合削除の閾値 ξ : 0.1 (実験 1), 0.2 (実験 2)
   - 1 層のコミュニティ数 : 3 (実験 1), 5 (実験 2).
   - 結合束を定義する手法 : [19] の *Method* 2 (実験 1), 3 (実験 2).
   - 結合束の有無を決める閾値 ζ : 0.2
   - 正規化後の入力データの最小値・最大値 : $\{-3, 3\}$ (実験 1), $\{-1, 1\}$ (実験 2).

4.1 実験 *1* : 入出力データが持つ隠れ構造の抽出

本節では，人工データを用いた実験によって，提案法が多層ニューラルネットの隠れ構造を適切に発見できることを示す．具体的には，ランダムに生成したモジュール構造を持つニューラルネットを用いて人工データを生成し，それを用いて学習した新たな多層ニューラルネットに提案法を適用することで，真のモジュール構造が復元されることを示す．実験の手順は以下の通りである：(1) 真のモジュール構造の生成，(2) 生成されたモジュール構造に基づく入出力データの生成，(3) 生成されたデータを用いた多層ニューラルネットの学習，(4) 学習されたニューラルネットからのモジュール構造抽出．

(1) 図 3 に基づき，真のモジュール構造を生成する．まず，図 3 左における実線の結合束で結ばれたコミュニティ間の結合重みを $\mathcal{N}(0.5, 1)$ に従い独立に生成する．これは，出力に影響を与えない入力や，入力に依存しない出力ができることを防ぐための処理である．次に，図 3 右に示されるように，ランダムな位置に r 本の点線の結合束を配置し，これらが結ぶコミュニティ間の結合重みを $\mathcal{N}(-0.5, 1)$ に従い独立に生成する．ただし，これらの結合重みのうち，絶対値が 1 以下のものは削除し重み 0 とした．各層におけるユニットのバイアスは $\theta_j^d \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 0.5)$ に従い生成した．(2) 45 次元の入力データを $x_j^n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 3)$ に従い生成し，これを

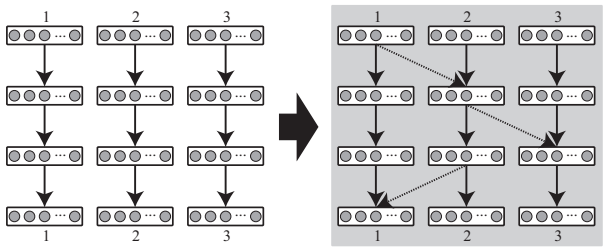

図 3. 実験 1 における真のモジュール構造の生成方法．各コミュニティは 15 個のユニットを含む．実線の結合束で結ばれたコミュニティ間の結合重みは $\mathcal{N}(0.5, 1)$ から，点線の結合束で結ばれたコミュニティ間の結合重みは $\mathcal{N}(-0.5, 1)$ から独立に生成される．

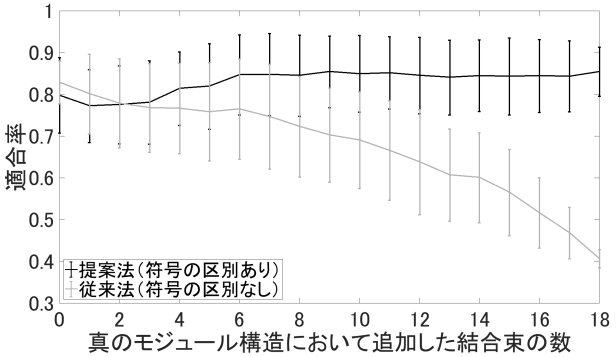

図 4. 追加した結合束の本数 r と適合率の関係．

上記のニューラルネットに入力したときの出力に対し，$\mathcal{N}(0, 0.05)$ に従う独立な雑音を加えたものを出力データとした．(3) これらのデータを用いて新たな多層ニューラルネットを学習し，(4) 提案法を適用してモジュール構造を抽出した．

真のモジュール構造における結合束の本数 r が $0, \cdots, 18$ の各場合に対し，上記に述べた，真のモジュール構造のランダム生成と，生成したモジュール構造の復元を 300 回繰り返す．提案手法および正負の結合を区別しない確率モデルを用いた既存手法 [19] により抽出されたコミュニティ構造の，真の構造に対する**適合率**を図 4 に示す．ここで，適合率は以下のように定義される．まず，真のコミュニティ構造と，提案法もしくは従来法を用いて抽出されたコミュニティ構造において，入力層の各ユニット u が属するコミュニティの番号をそれぞれ c_u, c'_u とおき，これらの間の**適合数**を求める．ここで，真のコミュニティ構造と，学習後のニューラルネットから抽出されたコミュニティ構造におけるコミュニティ番号の対応は未知である．本評価法においては，提案法・従来法のそれぞれにおいて，考えられる全ての対応関係の組み合わせのうち，適合数 $\sum_u \delta_{c_u, c'_u}$ が最大となるものにおける適合数を用いるものとする．ただし，$\delta_{i,j}$ を Kronecker のデルタとする．同様に，出力層に関しても提案法・従来法のそれぞれを用いた場合の適合数を求める．最後に，入出力層における，全ユニット数に対する適合数の和の割合を計算し，300 回の反復における平均値を適合率とする．

図 4 より，r が小さい (≤ 2) ときは符号を区別しな

い従来法の方が，$r \geq 3$ では提案法の方が真のコミュニティ構造に対する適合率が高くなることが分かった．これは，ニューラルネットが単純な隠れ構造を持つ場合には，重みの符号の情報を用いなくとも真の構造を抽出可能だが，コミュニティ間が密につながった複雑な構造を持つ場合には，重みの正負の情報を用いることで，より真のコミュニティ構造を見つけやすくなる場合があるからであると考えられる．

4.2 実験 *2*：市区町村データの推論からの知識発見

実データを用いて学習した多層ニューラルネットに対し，提案法を適用しモジュール構造を抽出した．ここで用いたデータは，日本における各市区町村の特徴を表すデータ [23] であり，今回は主に人口や事業所数などの社会環境における基本的な情報を入力データとし，婚姻件数や完全失業者数などのより詳細な状況を表す情報を出力データとした．[19] の実験設定と同様，欠損値のある市区町村のデータは削除し，また入出力データに強い偏りがあるため関数 $\log(1+x)$ で変換して用いた．

上記のデータを用いて学習したニューラルネット (図 5) から，モジュール構造を抽出した結果を図 6 に示す．図 6 から，多層ニューラルネットによる推論の構造について様々な知識が得られる．以下に，その一例を記す．

1. 入力層において，人口に関する複数の項目が同じコミュニティ (C1) に割り当てられており，これらの値は出力の推論を行う際に似た役割を果たすものと考えられる．
2. 出力層において，高齢者の多さを表す項目や死亡数，核家族世帯数が同じコミュニティ (C2) に割り当てられており，これらの値は同じような情報から推論されるものと考えられる．同様に，離婚件数，第 2 次産業従業者数，完全失業者数が同じ同じコミュニティ (C3) に割り当てられている．
3. 出生数と婚姻件数は同じコミュニティ (C4) に割り当てられており，これらの値は主に人口 (C1) や転入・転出者数，世帯数 (C8) などの値から推論されることが分かる (C4←C5←C6←C1, C4←C5←C7←C8)．これは，出生数の増加が人口の増加につながることや，出生・婚姻などのライフイベントに応じて一世帯の住所変更を行う場合が多いことが原因ではないかと考えられる．
4. モジュール構造から，中間層における各ユニットの役割についても手がかりを得ることができる．例えば，コミュニティ C9 に含まれる 8 個のユニットは，主に転入・転出者数，世帯数の情報 (C8) から死亡数，核家族世帯数，単独世帯数，65 歳以上の世帯員のいる核家族世帯数，高齢夫婦世帯数，高齢単身世帯数，第 1 次産業就業者数 (C2) の推論を行う際にのみ用いられていることが分かる．
5. 上記と同様に，コミュニティ C10 に含まれる 4 個のユニットは，主に人口に関する情報や就業者数，他市区町村への通勤者数 (C1) から製造業従業者数と第 2 次産業就業者数 (C11) の推論を行う際にのみ用いられていることが分かる．

5 考察

提案法により抽出されるモジュール構造は，ハイパーパラメータ ξ や ζ の設定に依存する．目的に応じて，これらのハイパーパラメータを最適化することは今後の課題である．また，本稿では適合率の計算方法として，入出力層におけるコミュニティ割り当ての整合性のみに基づくものを用いた．抽出されたモジュール構造に対し，中間層におけるコミュニティ割り当てや，各層のコミュニティ間をつなぐ結合束の有無も考慮して妥当性を評価するためには，新たな評価法が必要となる．さらに，提案法を画像や音声などのメディアデータや，より大規模なデータに適用することを検討していきたい．

6 結論

多層ニューラルネットは画像や音声など様々な実データを用いた課題において，既存手法を圧倒する認識・予測性能を実現してきた．しかし，その推論は非常に多くの非線形なパラメータが絡み合った複雑な階層構造により表現されるため，人間が理解し知識を発見することが難しい．本研究では，正負の結合を区別した新たな確率モデルに基づくネットワーク解析を行うことにより，学習した多層ニューラルネットから大局的な構造を抽出する手法を提案した．この手法では，隣接する層と似た正負の結合パターンを持つユニットを同一コミュニティに割り当て，コミュニティ間の複数の結合をまとめて表示することにより，元のニューラルネットを単純化した構造を抽出することを可能とした．提案法を用いることにより，多層ニューラルネットに潜む隠れ構造を適切に発見できることが示された．また，提案法により抽出されたモジュール構造が，実データに対する多層ニューラルネットの推論に関して様々な知識を得るための手がかりとなりうることが示された．

参考文献

[1] Y. Bengio, A. Courville, and P. Vincent. Representation learning: a review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 35, No. 8, pp. 1798–1828, 2013.
[2] C. Tai et al. Convolutional neural networks with low-rank regularization. In *International Conference on Learning Representations*, 2016.
[3] G. Hinton et al. Deep neural networks for acoustic modeling in speech recognition. *IEEE Signal Processing Magazine*, Vol. 29, No. 6, pp. 82–97, 2012.
[4] H. Xiong et al. The human splicing code reveals new insights into the genetic determinants of dis-

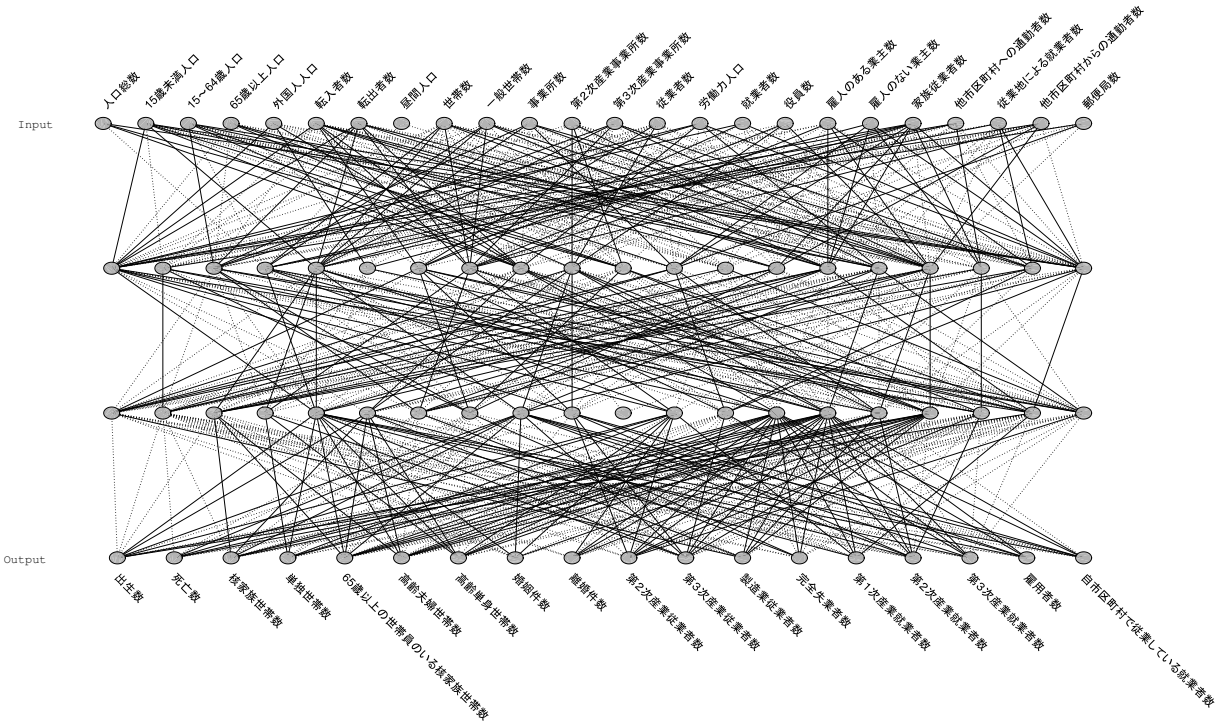

図 5. 市区町村の特徴データ [23] から学習した多層ニューラルネット.

ease. *Science*, Vol. 347, No. 6218, 2015.

[5] M. Denil et al. Predicting parameters in deep learning. In *Advances in Neural Information Processing Systems*, pp. 2148–2156, 2013.

[6] M. Leung et al. Deep learning of the tissue-regulated splicing code. *Bioinformatics*, Vol. 30, No. 12, pp. i121–i129, 2014.

[7] M. Liu et al. Towards better analysis of deep convolutional neural networks. *IEEE Transactions on Visualization and Computer Graphics*, Vol. 23, pp. 91–100, 2017.

[8] S. Han et al. Learning both weights and connections for efficient neural network. In *Advances in Neural Information Processing Systems*, pp. 1135–1143, 2015.

[9] T. Sainath et al. Deep convolutional neural networks for LVCSR. In *International Conference on Acoustics, Speech and Signal Processing*, 2013.

[10] J. Feng and T. Darrell. Learning the structure of deep convolutional networks. In *International Conference on Computer Vision*, 2015.

[11] S. Han, H. Mao, and W. Dally. Deep compression: Compressing deep neural networks with pruning, trained quantization and huffman coding. In *International Conference on Learning Representations*, 2016.

[12] M. Ishikawa. A structural connectionist learning algorithm with forgetting. *Journal of Japanese Society for Artificial Intelligence*, Vol. 5, No. 5, pp. 595–603, 1990.

[13] A. Krizhevsky, I. Sutskever, and G. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, 2012.

[14] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, Vol. 521, pp. 436–444, 2015.

[15] M. Newman and E. Leicht. Mixture models and exploratory analysis in networks. *Proceedings of the National Academy of Sciences*, Vol. 104, No. 23, pp. 9564–9569, 2007.

[16] D. Rumelhart, G. Hinton, and R. Williams. Learning representations by back-propagating errors. *Nature*, Vol. 323, pp. 533–536, 1986.

[17] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B*, Vol. 58, pp. 267–288, 1994.

[18] J. Tompson, A. Jain, Y. LeCun, and C. Bregler. Joint training of a convolutional network and a graphical model for human pose estimation. In *Advances in Neural Information Processing Systems*, 2014.

[19] C. Watanabe, K. Hiramatsu, and K. Kashino. Modular representation of layered neural networks. arXiv:1703.00168, 2017.

[20] P. Werbos. *Beyond regression : new tools for prediction and analysis in the behavioral sciences*. PhD thesis, Harvard University, 1974.

[21] M. Yuan and Y. Lin. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society Series B*, Vol. 68, No. 1, pp. 49–67, 2006.

[22] M. Zeiler and R. Fergus. Visualizing and understanding convolutional networks. In *European Conference on Computer Vision. Lecture Notes in Computer Science*, Vol. 8689, pp. 818–833, 2014.

[23] 政府統計の総合窓口 e Stat. 統計でみる市区町村のすがた 2016. `http://www.e-stat.go.jp/SG1/estat/GL08020103.do?_toGL08020103_&tclassID=000001073038&cycleCode=0&requestSender=search`, 2016.

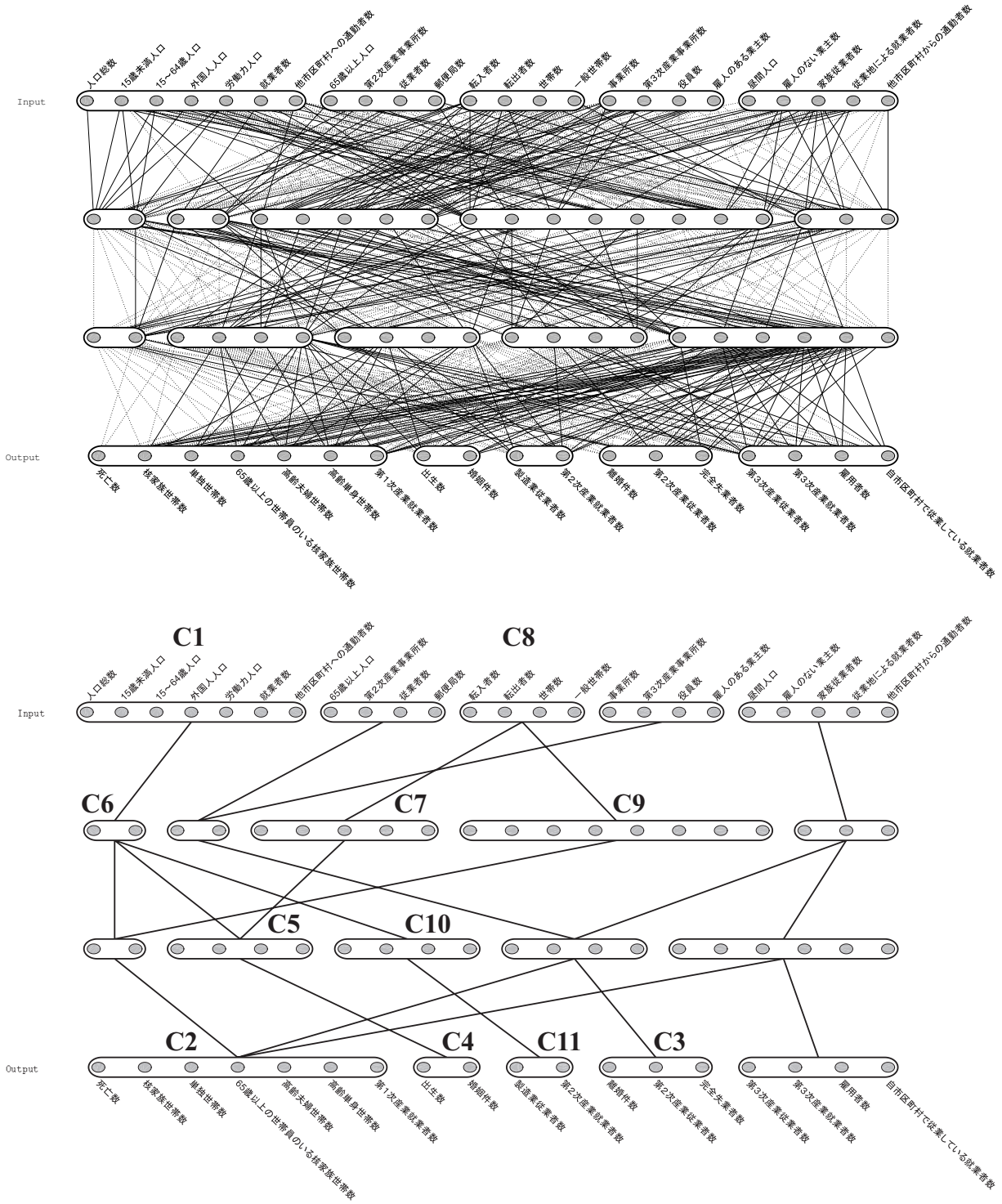

図 6. 上：提案法を用いてコミュニティ抽出を行った結果．下：結合束を定義することにより抽出されたモジュール構造．実線は正の，点線は負の重みを表す．