

動的近傍競合学習を用いたコミュニティ分割手法

小野寺 大地† 鈴木 育男 山本 雅人 古川 正志

†北海道大学大学院 情報科学研究科

1 はじめに

コミュニティ分割問題はネットワーク上からコミュニティとなる部分集合を抽出し、特性を解析することを目的としている。コミュニティ分割を行う手法として、Girvan, Newman らによる Modularity を用いた手法 (CNM 法: Clauset Newman Moore) や, "edge-betweenness" を用いた手法が知られている [1]。しかし, これらの手法は, リンク数が多いネットワークや, ネットワーク全体のリンク密度が高い問題に対し計算時間が非常に大きいことやコミュニティ間が密に結合している場合に分割が困難である。

本研究では, ネットワークのコミュニティ分割問題に対し, 自己組織化マップ (self-organizing map: SOM) [2] に基づく動的近傍競合学習を用いたアプローチを提案する。自己組織化マップでよく用いられるトポロジではなく, 対象となるネットワークのトポロジを採用し, 更に学習過程でトポロジを変化させることでリンク構造を考慮したコミュニティ分割を実現する。

2 提案手法

2.1 自己組織化マップ

自己組織化マップは T. Kohonen が開発した教師なし競合学習法である。高次元の入力データ群を環状, 格子状などのトポロジ構造の低次元空間へ写像する手法である。例えば, 高次元空間を二次元平面へ写像する場合, 似た特徴値をもつ入力データ同士が近傍に位置するため, 特徴を分類する手法として用いられる。

提案手法の学習アルゴリズムは競合学習を基礎としているが, 学習するトポロジは異なり, 解析対象のネットワークトポロジを用いる。また, コミュニティ内でのリンク密度が高いという性質から, パス長が短いと同一コミュニティの可能性が高いことを考慮し, 全てのノードの学習の際の学習率はパス長が短いほど高くする (図 1)。このような学習により, コミュニティ内のノードは似た特徴値を獲得し, その獲得した特徴値を用いて, コミュニティ分割を行う (図 2)。

2.2 動的近傍競合学習

動的近傍競合学習は, 学習結果にしたがってトポロジ構造が変化する学習法である。

まず通常の競合学習と同様に学習を行い, 学習終了後にノードが獲得した特徴値を比較する。次に特徴値

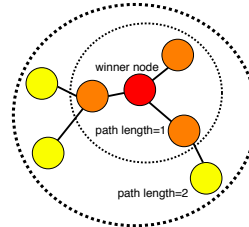


図 1: 勝者ノードからの近傍学習。勝者ノードとパス長が近いノードほど学習率が大きく, より勝者ノードと似た値をもつ。

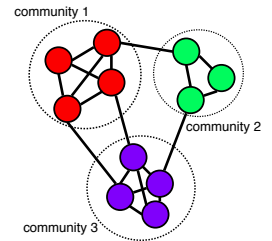


図 2: 競合学習をネットワークに適用した場合。リンクが密なノード群は似た値をもつ。

が離れているノード間のリンクを削除し, 再び学習を開始する。これにより, 特徴値が離れているノード間ではリンク切断後に学習が行われず, リンクが密な部分集合内で更に特徴値が近くなる。結果として, 獲得した特徴値とリンク構造の二点によりコミュニティの抽出が可能となる。

2.2.1 学習アルゴリズム

ネットワークのすべてのノード集合を N とする。また, ステップ t におけるそれぞれのノード $v \in N$ は, k 次元の特徴ベクトル $\mathbf{y}_v(t) = (y_1, y_2, \dots, y_k)$ をもつ。1 試行回数あたりの学習回数を T とし, 各ステップを $t (0 \leq t < T)$, リンク切断を含めた試行回数を TT , リンクの切断回数を ω とする。学習時の入力信号 $\mathbf{x} = (x_1, x_2, \dots, x_k)$ としたときの学習アルゴリズムを以下に示す。

1. すべての $v \in N$ に特徴ベクトル $\mathbf{y}_v(t)$ をランダムに生成する。
2. 入力信号 \mathbf{x} をランダムに生成する。
3. すべてのノード $v \in N$ の中から以下の式で勝者ノード v_{max} を選択する。

$$v_{max} = \arg \max_v \left(\frac{\mathbf{x} \cdot \mathbf{y}_v(t)}{\|\mathbf{x}\| \|\mathbf{y}_v(t)\|} \right) \quad (1)$$

4. v_{max} と, その近傍ノード集合を以下の式により学習する。ここで, h_c は, ステップ数とともに減衰する学習係数である。

$$\mathbf{y}_v(t+1) = \mathbf{y}_v(t) + h_c(t)(\mathbf{x} - \mathbf{y}_v(t)) \quad (2)$$

5. $t < T$ なら h_c の値を小さくし, 2. に戻り, 繰り返す。
6. 以下の式を満たすノード v, v' 間のリンクを切断する。ただし ω は $0 \leq \omega \leq 1$ の値をとるしきい値である。

$$\omega > \frac{\mathbf{y}_v(t) \cdot \mathbf{y}_{v'}(t)}{\|\mathbf{y}_v(t)\| \|\mathbf{y}_{v'}(t)\|} \quad (3)$$

Community Extraction Method based on Dynamic Neighborhood Competitive Learning

†Daichi ONODERA Ikuo SUZUKI, Masahito YAMAMOTO, Masashi FURUKAWA

†Hokkaido University Graduate School of Information Science

表 1: 理想値と CNM 法, 提案手法の比較. 提案手法のコミュニティ数の同定に CNM 法を用いた.

手法	理想値	CNM 法	提案手法
コミュニティ数	10	7	10
分割精度 s	1.00	0.696	0.990

7. $tt < TT$ なら, リンク切断後のネットワークに対し, 再び 1. から学習を再開する.

3 数値計算実験

コミュニティ分割の特徴を明確にするために, 対象とするネットワークは, ある程度コミュニティが明示的にわかるものでなければならない. そこで, 今回の実験では CS モデル (community structure model) を用いて解析を行う. また, 提案手法は CNM 法のようにコミュニティ番号を付与しない. そのため, コミュニティ番号の付与のため提案手法と CNM 法を組み合わせさせて評価する.

評価指標として分割精度 s を定義する. コミュニティ数 N のコミュニティ集合を $C = \{c_1, c_2, \dots, c_N\}$ とし, ノード v が属するコミュニティ番号を $c(v)$ とするとき, コミュニティ c_i は $c_i = \{v | \delta(c(v), i)\}$ と表される. ここで δ はクロネッカのデルタである. s は, c_i 内のノードと, 本来含まれるべきであるノードが一致した個数を全体のノードで割り, 正規化したものである. この値は正しい分割結果に近ければ 1 に近づき, 逆に正しくなければ 0 に近づく評価指標である.

3.1 CS モデルによるネットワークの生成

CS モデルでは, 予めコミュニティとなるノード群を決めておき, コミュニティ内でリンクを繋ぐ確率 p_{in} とコミュニティ間でリンクを繋ぐ確率 p_{out} を設定する. このときコミュニティ内のリンク密度を高めるために, $p_{in} \gg p_{out}$ とする. 結果, コミュニティ内ではリンク密度が高く, コミュニティ間ではリンクが少ない状態を実現することができる.

3.2 実験パラメータ

実験条件は以下のように設定した.

CS モデル

ノード数 100 のコミュニティを 10 個用意し, リンクの繋ぐ確率を, $p_{in} = 0.9, p_{out} = 0.01$ とした. 生成ネットワークを図 3 に示す.

自己組織化マップ

学習回数 T を 1000 回と設定し, リンクの切断回数 TT を 50 回, 近傍半径の初期値は最短パス長が 3 となるノードを学習する. その後徐々に近傍半径を小さくし, 最終的にパス長 1 まで減少させる. 学習率は $\alpha = 0.3$ から徐々に小さくし, 最終的には 0 とする. リンクの切断のしきい値を $\omega = 0.9$ とした.

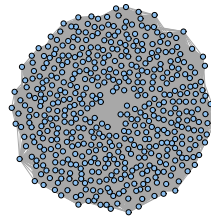


図 3: CS モデルにより生成したネットワーク.

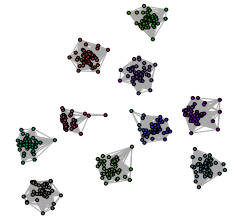


図 4: 図 3 に提案手法を適用した結果.

3.3 実行結果

3.2 のパラメータを用いて提案手法を生成ネットワークに適用した結果を図 4 に示す. また, コミュニティ分割について理想値, CNM 法, 提案手法についての分割コミュニティ数と, 分割精度 s を表 1 に示す. 図 3 と同様の可視化手法を用いて可視化を行うと図 4 より, コミュニティ内のリンクが残るため, コミュニティ外のリンクが存在しなく, 各々のコミュニティごとに凝集している. また, 表 1 より, CNM 法のみ解析結果に比べコミュニティ数, 分割精度ともに理想値に近い値を得ていることがわかる.

4 考察

提案手法を用いることで, コミュニティ分割が可能であるといえる. また, CNM 法のみ解析に比べ提案手法と CNM 法の組み合わせは理想値に近い値となっている. これは, コミュニティ外リンクを優先的に削除することにより, ネットワーク中からコミュニティを抽出していることを意味する. 結果より, CNM 法を用いて解析が容易なコミュニティ構造をもつネットワークへ変化しているといえる.

5 おわりに

CNM 法のみコミュニティ分割に比べ提案手法と CNM 法の組み合わせは, CNM 法で解析が難しいネットワークへ適用が可能であることが確認された. 今後の課題として, リンクを切断するしきい値を明確にし, CNM 法を用いないコミュニティ番号付与の方法を検討する必要がある.

参考文献

- [1] M. E. J. Newman and M. Girvan (2004): Finding and evaluating community structure in networks. Phys. Rev. E, Vol. 69, p.026113.
- [2] T. コホネン (2000): 自己組織化マップ, 徳高平蔵, 岸田悟, 藤村喜久郎 (訳), シュプリンガー・フェアラーク.