

複雑ネットワークに適用するファジィクラスタリング手法の検討

上杉 一矢[†]木村 昌臣[‡]芝浦工業大学工学部情報工学科[†]芝浦工業大学工学部情報工学科[‡]

1. はじめに

現実世界に存在するネットワークは多様であり、巨大で複雑な構造を有している。これを複雑ネットワークと言い、多くのノードとエッジで構成されているために、ノード間の関係の把握は一般に容易でない。この問題の対応策として、クラスタリングという、エッジが疎となる部分で複雑ネットワークを分割し、エッジが互いに密に張られているノード群(クラスタ)を抽出する手法が開発されている。現在、複雑ネットワークには、ノードが一つのクラスタにのみ所属するハードクラスタリング手法が採られている。しかし、ノードが複数のクラスタに繋がっているならば、片方ではなく両方に所属していると考えたほうが便利であるため、複数のクラスタへの所属を考慮する必要がある。そこで、本研究では一つのノードが複数のクラスタに所属可能なファジィクラスタリングという方法に着目する。この方法を利用した既存研究として、座標に対してクラスタリングを行う方法がある[1]。その方法の適用にあたり、ネットワークのノードをノード間の距離を元に座標に変換するが、この段階で特定のノードからの距離が同じ場合に座標も同じになってしまう場合、エッジが繋がっていなくても同じクラスタに含まれることがあり、ネットワークの繋がりを正しく反映させることが困難になる。そこで、本研究では Newman のモジュラリティの考え方を拡張することにより、複雑ネットワークに直接適用するファジィクラスタリング手法を提案する。

2. 提案手法

本研究では、ファジィクラスタリングによって分割された状態の良さを表す評価関数を定義し、その定義式を元にネットワーク構造に適用するクラスタリング手法の提案を行う。提案手法を実装するシステムを作成した後、得られたクラスタの精度についての検討を行う。

2.1 評価関数の定義

参考文献[2]で提案された固体物理で用いられるスピングラス系のハミルトニアンと類似した形をもつ評価関数(1)を、本研究に適した表現に再定義して用いる(k_i : i のエッジ数, M : 総エッジ数)。

$$H = -\frac{1}{2M} \sum_{i \neq j} \left(A_{ij} - \frac{k_i k_j}{2M} \right) \delta(\sigma_i, \sigma_j) \quad (1)$$

従来の評価関数(1)は、ノード i と j が隣接しており、かつ i が所属するクラスタの番号 σ_i と j の σ_j が等しい時、 H が小さくなるため、クラスタ分割を H についての最適化問題として扱える。式(1)の $\delta(\sigma_i, \sigma_j)$ は、 $\sum_c \delta(\sigma_i, c) \delta(\sigma_j, c)$ と書き直せ、ここに表れる $\delta(\sigma_i, c)$ は、 σ_i がある一つのクラスタにのみ所属する確率と見なすことができるのでこれを一般化すると、ノード i が複数のクラスタに所属する確率 $P(\sigma_i = c)$ の特別な場合と見なすことができる。この確率が帰属度に相当する。このような一般化を行うことで、 $\delta(\sigma_i, \sigma_j)$ は $\sum_c P(\sigma_i = c) P(\sigma_j = c)$ と置き換えることができる。この置き換えた式を用いて評価関数を式(2)として再定義する。

$$H_F = -\frac{1}{2M} \sum_{i \neq j} \left(A_{ij} - \frac{k_i k_j}{2M} \right) \sum_c P(\sigma_i = c) P(\sigma_j = c) \quad (2)$$

2.2 ファジィクラスタリング手法の提案

最適化問題を扱うにあたり、問題解決に遺伝的アルゴリズム(GA)を使用する。GAは、生物の環境適応のための生物進化の現象を工学的にモデル化することで最適な個体(染色体)を生成するアルゴリズムである。染色体はノード番号と各ノードの帰属度で構成される。通常のGAでは、遺伝子部分にあたる帰属度を"1", "0"の2進数で表現するが、交叉及び突然変異の処理で、実数値の連続性を活かす探索を行うために実数値GAを用いる。なお、交叉ではBLX- α 、突然変異にはWrightの突然変異法、淘汰にはエリート保存戦略を用いる。BLX- α とは、Eshelmanらにより提案された交叉法である。親個体となる染色体の遺伝子の実数値ベクトルを基準として、変数ごとに親の2点間の区間 I を α 倍した領域からランダム

Examination of fuzzy clustering technique for applying to complex network

[†]Kazuya Uesugi, [‡]Masaomi Kimura

[†]Shibaura Institute of Technology, [‡]Shibaura Institute of Technology

に子個体を生成する。Wright の突然変異法は、実数値に対して摂動を行う突然変異法であり、実数値が取り得る範囲から、実数値を境として分割し、同確率でどちらかの範囲を選択する。実数値の範囲を超えないような制御を行う摂動幅を用いて、選択した範囲からランダムに値を選択する。エリート保存戦略は、適応度の高い(Hが小さい)染色体から順に次の世代に残していく方法である。

これに加え、最適解をより早く得るために、淘汰の直前に最急降下法により変化させた染色体を発生させる。また、交叉、突然変異の処理を行った後では、帰属度の総和が 1 に保たれていないので、正規化を行うことにより総和を 1 に保つ。

3. 評価実験

本手法の評価のため、小規模ネットワークに対してクラスタリングを行い、作成したシステムが最適解、もしくは最良解の導出を行うかどうかを調べる。クラスタリングを行うネットワークは、あるノードが複数のクラスタに所属するようなネットワークである必要があるため、ノード数が 7 のネットワーク(図 1)を使用する。評価方法としては、評価関数式(2)による、実際にクラスタリングされたネットワークの分割状態と適応度を最小にするネットワークの分割状態を比較し、世代ごとの適応度の変化に対してそれぞれ評価する。

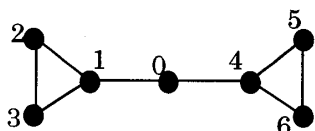


図 1 実験用ネットワーク

4. 結果・考察

提案手法による結果を以下の表 1, 図 2 に示す。

表 1 提案手法により得られた結果
(適応度を最小とする帰属度)

クラスタリングされたネットワークの適応度	-0.514633(-0.515625)
各ノードの帰属度 (クラスタ 1)	ノード 0 : 0.3848(0.5)
	ノード 1 : 0.999(1.0)
	ノード 2 : 0.999(1.0)
	ノード 3 : 0.999(1.0)
各ノードの帰属度 (クラスタ 2)	ノード 0 : 0.6152(0.5)
	ノード 4 : 0.999(1.0)
	ノード 5 : 0.999(1.0)
	ノード 6 : 0.999(1.0)

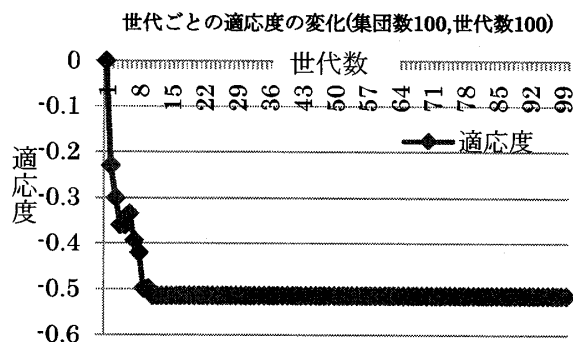


図 2 世代ごとの最小適応度の変化

まず、本手法で得られる帰属度はノード 1~3 とノード 4~6 が別のクラスタに属し、ノード 0 が双方のクラスタに属す場合、ハードクラスタリングによる値(-0.514618)より小さい値をとることが分かる。実数値 GA により得られた結果については上記の割り当てによる帰属度に到らないものの、ハードクラスタリングによる値を下回った。図 1 のネットワークの構造と得られた結果を比較すると本手法からファジィなネットワーク構造分割を得られることを示唆していると考えられる。

又、図 2 から適応度の収束の様子を示唆しているが、数ステップで適応度が低い分割に達していることが見て取れる。一方、表 1 からは最小の適応度を得るための探索手法について更なる工夫が必要であることが分かった。

5. 今後の課題

実験結果の考察より、システムの実装に改善する余地があることがわかった。本研究では複雑ネットワークに適用するファジィクラスタリング手法を提案した。

参考文献

- [1] 山崎恭史, 木村昌臣:
トピックマップ構造におけるスコープ自動抽出方法の提案, 情報処理学会第 71 回全国大会論文集, Vol.71, No2, pp2.183-2.184
- [2] Jorg Reichardt and Stefan Bomholdt:
Statistical mechanics of community detection, PHYSICALREVIEWW74, pp.016110_1-016110_14(2006)