

1 はじめに

ネットワークの中からよく現れるつながり方や構造を発見する方法はパターンマイニングといわれている。パターンマイニングは社会ネットワークから有益な情報を見つける有効な手段である。ILP[1] によるネットワークパターンマイニングのひとつとして花火節を用いたパターンマイニング [2] がある。これはネットワークのあるノードと隣接するノードの集合から導出される誘導部分グラフであるエゴネットからパターンを生成しパターンマイニングを行っていくものである。この手法は計算量が多く大規模な社会ネットワークに適用するのは未だ現実的ではない。そこで本研究では包摂関係を用いてアルゴリズム内でのネットワークのマッチング回数を減らすことを提案する。

2 花火節を用いたパターンマイニング

花火節を用いたパターンマイニングははじめに頻出パターンを見つけないネットワークからエゴネットを抽出し単位花火節を生成する。単位花火節のうち最小サポートを超える支持度をもつものを深度 1 の頻出パターンとする。深度 $k(k \geq 2)$ の候補パターンは深度 $k-1$ の重ね合わせによって生成される。候補パターンのうち最小サポートを超える支持度をもつものを深度 k の頻出パターンとする。これを新たな頻出パターンが現れなくなるまで繰り返す。

3 包摂関係による候補パターンの削減

花火節を用いたパターンマイニングでは単位花火節同士が包摂関係になっている。それにより、単位花火節から構成される花火節もまた包摂関係になっているものが存在する。そこで生成された候補パターン同士を比較することにより候補パターンを削減が可能である。ある候補パターンの支持度が最小サポートを満たさなかった場合、その拡大パターンも頻出でないといえることから、まだネットワークとマッチングを行っていない候補パターンと比較を行い、拡大パターンである候補パターンをネットワークとマッチングすることなく候補から除外する。

4 実験

Zachary の空手クラブネットワーク (ノード数 34) [3] に対して従来法と提案手法のマッチング回数および実行時間の比較を行った。候補の数が大幅に増える深度 2, 深度 3 では包摂関係になっているものが多く、多くの候補パターンが頻出でないとして識別された。マッチング回数が減ったことに伴い実行時間が減少した。これはネットワークとのマッチングより候補パターンの比較のほうが簡単であったためだと考える。

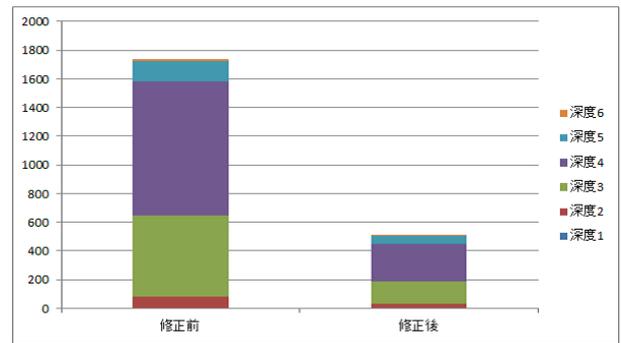


図 1: 最小サポート 5 % でのマッチング回数

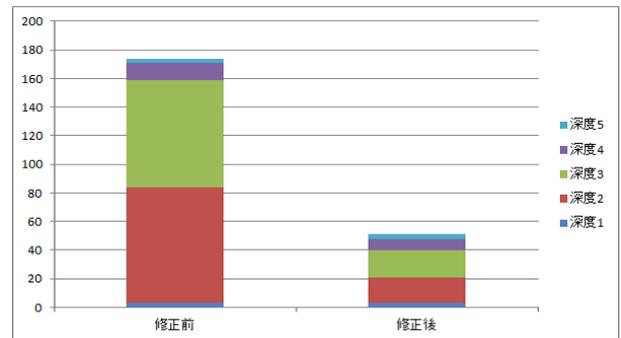


図 2: 最小サポート 10 % でのマッチング回数

表 1: アルゴリズム修正前後の実行時間

(ms)	修正前	修正後
5 %	5875.3	3797.3
10 %	224.7	48.4

5 まとめ

本研究では、花火節を用いたパターンマイニングのマッチングの回数を減らすことを提案した。提案手法を実装して従来手法と比較を行い、マッチング回数の削減及び実行時間が短縮したことを確認した。今後の課題として花火節を用いたパターンマイニングのさらなる改善が必要と考える。特にまだネットワークからエゴネットを生成の段階での計算量が多いのでエゴネット抽出の効率化が必要である。そして、花火節により挙げられた頻出パターンから実際の社会ネットワーク分析を行っていきたいと考えている。

参考文献

- [1] 古川 康一, 尾崎 知伸, 植野研. 帰納論理プログラミング. 共立出版, 2001.
- [2] Noriaki Nishio, Atsuko Mutoh and Nobuhiro Inuzuka, "On Computing Minimal Generators in Multi-Relational Data Mining with respect to theta-Subsumption", CEUR Workshop Proceedings (Late Breaking Papers of 22nd Inductive Logic Programming Conference), pp. 50-55, 2012
- [3] W. W. Zachary. An information flow model for conflict and fission in small groups. *Journal of Anthropological Research* 33, 452-473, 1977.