

## 1 はじめに

大規模なデータの中から有用な知識を見つけ出す手法としてデータマイニングがある。中でも複数のテーブルから構成されるデータベースを対象とするものを MRDM と呼び、ILP の枠組みで行われてきた。MRDM を社会ネットワークのような、事例が独立していない開いた構造に対応させた手法として、花火節を用いたパターンマイニング [1] がある。

しかし、この手法ではネットワーク特有の特徴について考慮されておらず、その点について改善を図るため距離の概念を導入したマッチング [2] が提案された。本研究では、距離ではなくノード同士の繋がり方に着目した誘導部分グラフ同型に基づくマッチングを提案し、より有意義なパターンのみをマイニングすることを目指す。

## 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では、開いた構造を持つデータベースに対して、ILP 分野におけるエゴセントリックネットワークと類似した構造である近傍を抽出し単位花火節 (1-shell の候補パターン) を生成する。その中でも最小サポートを超える支持度を持つものを 1-shell の頻出パターンとする。k-shell の候補パターンは、(k-1)-shell の頻出パターンの重ね合わせによって生成される。このように、近傍からボトムアップに頻出パターンが現れなくなるまで探索を繰り返す。

## 3 従来のマッチング手法

よく扱われるマッチング方法として、 $\theta$  包摂がある。2つの節  $S_1, S_2$  に対して、 $S_1 \theta \subseteq S_2$  となる代入  $\theta$  が存在するとき、 $S_1$  は  $S_2$  にマッチするという。もう一つの方法として、OI 包摂がある。これは  $S_1$  が  $S_2$  の部分グラフ同型であればマッチする。また、距離に基づくマッチングとして変数深度、 $\theta$ +項の深度、OI+項の深度がある。変数深度では、パターン上でターゲットからの距離を定義し、距離が異なるノードには異なる代入が存在しなければならない。 $\theta$ +項の深度、OI+項の深度では、 $\theta$ 、OI それぞれのマッチング方法に対し、パターン上での距離とネットワーク上での距離の一致を要求する。

## 4 誘導部分グラフ同型マッチング

ノード同士の繋がり方に着目したマッチング方法として、誘導部分グラフ同型に基づくマッチングを提案する。

**定義 1** 誘導部分グラフ同型マッチングある節  $C$  とデータベース  $D$  において、以下の条件を満たす代入

$\theta = \{A_1/B_1, A_2/B_2, \dots, A_n/B_n\}$  が存在するとき、 $C$  は  $D$  にマッチするという。

- $C \theta \subseteq D$
- $B_i \neq B_j (i \neq j)$
- $\{B_1, B_2, \dots, B_n\}$  から誘導される誘導部分グラフと節  $C$  が同値

## 5 実験

Zachary の空手クラブネットワーク [3] に対し、最小サポート 8% で実験を行った。その結果、提案手法で頻出であると判断されたパターンの一部を図 1 に、他手法では頻出になるが、提案手法では頻出にならなかったパターンの一部を図 2 に示す。提案手法では同 shell のパターンについて Apriori 性が成り立たない場合があり、その事実が正しく反映されている。また、各手法における頻出パターンの数を表 1 に示す。



図 1: 頻出パターン

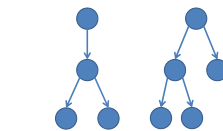


図 2: 頻出でないパターン

表 1: 最小サポート 8%:頻出パターン

	$\theta$	変数深度	$\theta$ +項	OI	OI+項	誘導
1-shell	3	3	3	5	5	5
2-shell	12	7	5	62	31	18
3-shell	29	7	0	-	0	0
4-shell	14	1	0	-	0	0
5-shell	0	0	0	-	0	0

## 6 まとめ

本研究では、誘導部分グラフ同型に基づくマッチングを提案した。実験結果から、従来手法では頻出とされるパターンをいくつか排除できていることが確認できた。しかしながら、排除したパターンが本当に有意義でないかどうかはまだ分かっていない。実データへの適用及びパターンの意味的な考察が今後の課題である。

### 参考文献

- [1] 西尾典晃. 関係的知識発見におけるパターンの表現に関する研究, 名古屋工業大学平成 24 年度修士論文, 2013.
- [2] 山崎誠治. 開いた構造のパターンマイニングにおける論理的手法に関する研究, 名古屋工業大学平成 25 年度修士論文, 2014.
- [3] W. W. Zachary. An information flow model for conflict and fission in small groups. *Journal of Anthropological Research* 33, 452-473, 1977.