7S-06

動的クラスタリングの結果に基づく時系列クラスタリング

大坪優希 † 松井藤五郎 ‡ 武藤敦子 † 島孔介 † 森山甲一 † 犬塚信博 †
名古屋工業大学 † 中部大学 ‡

1 はじめに

データマイニングの手法としてクラスタリングが存在する．データに対してグループ分けを行い，グループごとの特徴の発見ができる．そして，時系列データをクラスタリングし，クラスタの変化を見ることでクラスタとしての傾向の変化を捉えることが可能となる．時系列データをクラスタリングする際，時刻ごとにクラスタ番号が変わってしまい，クラスター遷移を観測できない問題があった．米田らは，データラベルの同一性に基づいたクラスター一致率を提案し，クラスタの変化を見る手法である MONIC を改良した [1]．しかし，米田らの手法では時刻が離れたクラスタの変化は見ることができなかった．

本論文では，時系列データに対して動的クラスタリングを行い，その結果に基づいてグラフを作成することによって時系列データをクラスタリングする手法を提案する．提案手法では，前後だけでなくすべての期のクラスタと関係を見ることにより，時刻が離れたクラスタの変化を観測できる．

2 従来手法

米田らは，動的クラスタリングにおいて，クラスター遷移を抽出する方法を提案した [1]．これを本論文では FBL-MONIC (Forward-Backward Label-based MONIC) と呼ぶ．

データラベルの同一性に基づいたクラスター一致率を式 (1) に示す．ここで，時刻 t のクラスタ集合を $\Gamma_t = \{C_{t,1}, ...C_{t,k}\}$，クラスタ番号を a, b，クラスタ $C_{t,k}$ に含まれるデータラベル集合を $L_{t,k}$ とする．

$$overlap(C_{t,a}, C_{t',b}) = \frac{|L_{t,a} \cap L_{t',b}|}{|L_{t,a}|} \tag{1}$$

FBL-MONIC では，overlap 関数の値に応じてクラスターの生存，分裂，吸収，消滅という前向きのクラスター変化とクラスターの影響，構成という後ろ向きのクラスター変化を定義することによって，時刻 t と $t+1$ のクラスタリング結果に対してクラスターの遷移を抽出する．

従来手法では，動的クラスタリングの結果の変化に着目していたため，時刻が離れたクラスタの変化を観測できない問題があった．

3 提案手法

提案手法では，時系列データをクラスタリングする．(1) 時系列データを時刻ごとに動的クラスタリングし，(2) その結果で完全グラフを作成する．そして (3) Jaccard 係数を用いてグラフカットを行う．(4) 最終的にできた島ごとに対してクラスタのラベル付け直しを行い，(5) その結果を用いた階層クラスタリングを行う．

グラフカットに用いる Jaccard 係数の定義を式 (2) に示す．

$$Jaccard(C_{t,a}, C_{t',b}) = \frac{|C_{t,a} \cap C_{t',b}|}{|C_{t,a} \cup C_{t',b}|} \tag{2}$$

Jaccard 係数をクラスタ集合に適用すると，Jaccard 係数の値が大きいほどクラスタは似ていると言える．

クラスタラベル付け直しは，まず動的クラスタリングで得られた結果をもとに，図 1 のような完全グラフを作成する．次に，Jaccard 係数を用いて図 2 のようにグラフカットを行う．最後に，図 3 のようにグラフカットでできた島ごとにラベルをつけて完成とする．

最後に，時系列データに対して時刻ごとの島の異なり数を距離として階層クラスタリングを行う．

4 実験

時系列データの時刻ごとにクラスタリングを行った結果を模した疑似データを作成し，提案手法を適用した．本実験では，グラフカットの際に用いる Jaccard 係数の閾値を 0.7 とした．

4.1 実験データ

本実験のために作成した，時系列データを時刻ごとにクラスタリングした結果を模したデータを表 1 に示す．ラベルはデータラベル，期ごとに書かれている番号は，動的クラスタリングで得られたクラスタ番号を表す．

Time series clustering based on dynamic clustering results

Yuuki Ootsubo†, Tohgoroh Matsui‡, Atsuko Mutoh†, Kosuke Shima†, Koichi Moriyama†and Nobuhiro Inuzuka†
Nagoya Institute of Technology†, Chubu University‡

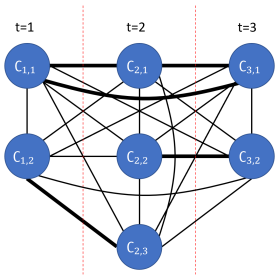

図 1: 完全グラフの例

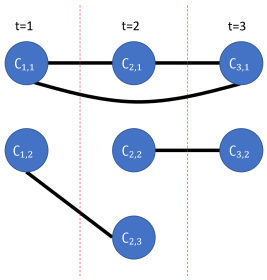

図 2: グラフカットの例

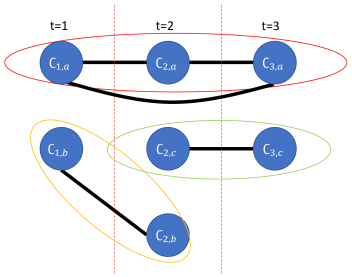

図 3: ラベル付け直しの例

表 1: 実験用疑似データ

ID	1 期	2 期	3 期	4 期	5 期
a	1	2	1	1	2
b	1	2	1	1	2
c	1	2	1	1	2
d	1	2	1	2	2
e	1	2	2	2	1
f	2	1	2	2	1
g	2	1	2	2	1
h	2	1	3	2	3
i	2	1	3	2	3
j	2	1	3	2	3

4.2 実験結果

実験用疑似データに，従来手法を用いてクラスタ番号を振り直したものを表 2 に，提案手法のクラスタラベル付け直しを適用したものを表 3 にそれぞれ示す．また，クラスタラベル付け直しの結果に対して階層クラスタリングを行った結果を図 4 に示す．図 4 の横軸はデータラベル，見やすさのため，縦軸を「異なり数 +1」としている．

提案手法では，従来手法では観測できていない 3 期のクラスタ番号 2 と 5 期のクラスタ番号 1 のような，時刻が離れた全く同じクラスタを観測することができた．

4.3 考察

階層クラスタリングでは，全体を通して同じクラスタに属している a, b, c と，1 度違うクラスタに属した d が近いものとされているため、適切にクラスタリングされていると考えられる．

今回の実験データでは起こらなかったが，提案手法を用いて時刻 t の一つのクラスタから $t+1$ の複数のクラスタがつながった島ができた場合，同じラベルを付けることができるため，異なり数が小さくなり階層クラスタリングを行ったときに有効であると考える．また，時刻が離れた同じクラスタを観測することができたが，提案手法を用いれば時刻が離れたクラスタ変化が起きた場合でも，Jaccard 係数の閾値を満たしていれば同じ島に属するため，時刻が離れたクラスタ変化を観測できる．

表 2: 従来手法で番号振り直し後

ID	1 期	2 期	3 期	4 期	5 期
a	1	1	1	1	1
b	1	1	1	1	1
c	1	1	1	1	1
d	1	1	1	2	1
e	1	1	3	2	2
f	2	2	3	2	2
g	2	2	3	2	2
h	2	2	2	2	3
i	2	2	2	2	3
j	2	2	2	2	3

表 3: 提案手法でラベル付け直し後

ID	1 期	2 期	3 期	4 期	5 期
a	1	1	1	1	1
b	1	1	1	1	1
c	1	1	1	1	1
d	1	1	1	2	1
e	1	1	3	2	3
f	2	2	3	2	3
g	2	2	3	2	3
h	2	2	4	2	4
i	2	2	4	2	4
j	2	2	4	2	4

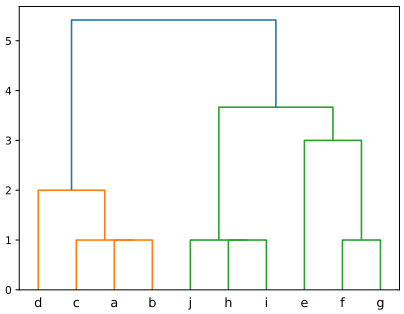

図 4: 階層クラスタリングの結果

5 まとめと今後の課題

本論文では，従来手法の問題である，時刻が離れたクラスタ変化を観測できない問題を解決する手法を提案した．実験では，時刻が離れた全く同じクラスタを観測することができた．

今後の課題として，グラフカットを行う際どの程度まで行うか，適切なカット数を調査する必要がある．

参考文献

[1] 米田一樹, 松井藤五郎, 武藤敦子, 森山甲一, 犬塚信博: 動的クラスタリングにおけるクラスタの変化分析, 情処研報 2019-MPS-122:6 (2019)