F-005

時刻ごとのクラスタリング結果の選別に基づく時系列クラスタリング

Time-series clustering based on selecting clustering results at each point

吉川 侑汰 [1], 松井 藤五郎 [2], 武藤 敦子 [1], 犬塚 信博 [1], 島 孔介 [1], 森山 甲一 [1]

Yuta Yoshikawa, Tohgoroh Matui, Atuko Mutoh, Nobuhiro Inuzuka, Kosuke Shima, Koichi Moriyama

1 はじめに

クラスタリングとは，データ群に対して似た特徴を持つものでグループ分けを行い，そのグループごとの類似した特徴や性質を発見する手法である．クラスタリングは金融商品の分類に対しても活用されている．大坪ら [1] は時系列データに対して期間ごとに時系列クラスタリングを行い，その結果に対して時系列クラスタリングをするという時系列クラスタリングを 2 段階で行う手法を提案した．しかしこの手法は，1 段階目のクラスタリングが上手くいくことが前提となっているため，1 段階目の時系列クラスタリングが上手くいかない事例には適用できない．そこで本論文では，1 段階目のクラスタリング結果を選別することによってこの問題を解決する手法を提案する．また，人工データと実データを用いて提案手法の有効性を確認する．

2 従来手法

薮内ら [2] は時系列データに対して，UMAP を用いて次元圧縮をしたのちに kmeans 法によるクラスタリングを行うという時系列クラスタリングを提案した．本研究ではこれを一般的な時系列クラスタリングとする．

それに対して大坪ら [1] の従来の 2 段階時系列クラスタリングは，次のように行われる．

1. 時系列データを複数の期間に分割し，期間ごとに時系列クラスタリングを行う
2. 期間ごとに行われたクラスタリングの結果に対してクラスタの対応関係を調べ，対応するクラスタに同じ ID をつける
3. クラスタ ID を用いて階層クラスタリングを行う

1 段階目のクラスタリングが上手くいかなかった場合，2 段階目の階層クラスタリングが妥当なものにならない．

3 提案手法

提案手法では 1 段階目でのクラスタリング結果に対して Calinski-Harabasz 基準を用いてクラスタリング結果を選別する．ただし，選別によってクラスタリング結果が少なくなってしまうため，1 段階目のクラスタリングを期間ごとの時系列クラスタリングから時刻ごとのクラスタリングに変更する．

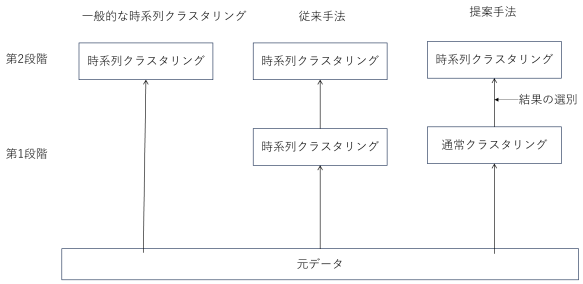

図 1: 手法の流れ

例えば，従来手法では月毎の 60ヶ月分の時系列データに対して，12ヶ月ずつの 5 期間に分割して時系列クラスタリングを行い，その結果に対して階層クラスタリングを行なうが，提案手法では，月毎にクラスタリングを行い，その結果を選別して階層クラスタリングを行う．本研究では各時刻でクラスタリングを行うことを通常クラスタリングと呼ぶこととする．

図 1 に一般的な時系列クラスタリング，従来手法，提案手法についてそれぞれの流れを図示している．

4 実験

一般的な時系列クラスタリングと従来手法と提案手法のクラスタリング結果を比較し提案手法の有効性を示すための人工データでの実験を行った．その後実際のファンドデータに提案手法を適用する実験を行った．また，1 段階目を通常クラスタリングに変更しただけの手法を提案手法 1，通常クラスタリングの結果を選別する手法を提案手法 2 とする．

4.1 人工データでの実験

それぞれのデータポイントがランダムに分布する時刻とグループごとに固まって分布する時刻が存在する時系列データのデータセットを人工的に作成し，それぞれの手法でクラスタリングをする実験を行う．実験の設定は以下の通りとする．

- 時系列データは 1 番から 15 番まで存在し，それぞれの時系列データの長さは 60 とする．
- 各時系列データは 12 期に 1 度クラスタごとに固まって分布する．

1: 名古屋工業大学, Nagoya Institute of Technology
2: 中部大学, Chubu University

- 時系列データのクラスタは 1〜5, 6〜10, 11〜15 の 3 クラスタ存在する．
- k-means 法，階層クラスタリングでのクラスタ数は 3 とする．

4.2 人工データでの実験結果

一般的な時系列クラスタリングでの結果を表 1 に，従来手法での 2 段階時系列クラスタリングの結果を表 2 に，提案手法 1 の結果を表 3 に，提案手法 2 の結果を表 4 に示す．

表 1: 一般的な時系列クラスタリング

クラスタ番号	所属クラスタ
1	1, 2, 3, 4, 5, 6, 7, 8, 9, 10
2	11, 15
3	12, 13, 14

表 2: 従来手法

クラスタ番号	所属クラスタ
1	1, 2, 3, 4, 7
2	5
3	6, 8, 9, 10, 11, 12, 13, 14, 15

表 3: 提案手法 1

クラスタ番号	所属クラスタ
1	1, 2, 3, 4, 5
2	6, 7, 8, 9
3	10, 11, 12, 13, 14, 15

表 4: 提案手法 2

クラスタ番号	所属クラスタ
1	1, 2, 3, 4, 5
2	6, 7, 8, 9, 10
3	11, 12, 13, 14, 15

また選別の結果残った時刻は $\{12, 24, 36, 48, 60\}$ であった．

4.3 実データでの実験

実データでの実験ではアクティブファンドの月次リターンに対してそれぞれの手法を適用しクラスタリング結果を比較する．実験の設定は以下の通りである．

- 1 番から 5 番がバリュー型，6 番から 10 番がグロース型，11 番から 15 番が高配当型
- 各月次リターンは 60 か月分
- データは 2018 年 1 月から 2022 年 12 月のものを使用する
- 従来手法の第 1 段階では分割を 12 か月ごとに行い，5 つのまとまりをそれぞれ次元圧縮して k-means 法でクラスタリングする
- k-means 法，階層クラスタリングでのクラスタ数はどちらも 3 とする

4.4 実データでの実験結果

一般的な時系列クラスタリングと従来手法，提案手法 1 と提案手法 2 でそれぞれ同じ結果となった．

一般的な時系列クラスタリング及び従来手法では {3, 6, 8, 9, 10}, {1, 2, 5, 13, 14, 15}, {4, 7, 11, 12} とクラスタリングされた．

提案手法 1 及び提案手法 2 では {3, 6, 8, 9, 10}, {1, 2, 4, 5, 11, 12, 13, 14, 15}, {7} とクラスタリングされた．

4.5 考察

表 1 から表 4 を見ると，一般的な時系列クラスタリングと従来手法に比べて，提案手法 1 と提案手法 2 の方がクラスタリング結果の精度が高く，また提案手法 1 よりも提案手法 2 の方がクラスタリング結果の精度が高いことがわかる．よって第 1 段階での通常クラスタリングの採用と第 1 段階後の結果の選別の両方がクラスタリング結果の向上に寄与していると言える．

また選別後に残っていた時刻はすべてクラスタ固有の値をとる時刻であるので，Calinski-Harabasz 基準によって適切に選別できていると言える．

実データでの実験では一般的な時系列クラスタや従来手法とは違う結果が得られた．提案手法 1 と 2 で同じ結果となったのは今回使用したデータでは 1 段階目のクラスタリング結果において各時刻の Calinski-Harabasz 基準の値に大きな差がなかったためだと思われる．

5 まとめ

本研究では従来の 2 段階時系列クラスタリングの問題点を解決するために，通常クラスタリングと結果の選別を導入した．人工データでの実験では提案手法の有効性を示し，実データでの実験ではアクティブファンドのクラスタリングを行った．しかし実データにおいては結果の選別の影響を確認できなかったため，特定の時刻でのみクラスタごとの特性が強く表れるようなデータで再度実験を行う必要がある．

参考文献

[1] 大坪 優希, 松井 藤五郎, 武藤 敦子, 島 孔介, 森山 甲一, 犬塚 信博. 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. 第 85 回情報処理学会全国大会, 7S-06 (2023).

[2] 籔内陽斗, 松井 藤五郎, 武藤 敦子, 島 孔介, 森山 甲一, 犬塚 信博. 長期リターンに対し UMAP を用いた投資信託クラスタリング, 第 35 回人工知能学会全国大会 2H3-GS-3b-05 (2021).