

時系列データ間の類似尺度を用いた 業種区分の再編とリスクコントロール

05001589 明治大学先端数理科学研究科 *長島 暉 NAGASHIMA Hikaru
01013460 明治大学総合数理学部 乾 孝治 INUI Koji

1. 緒言

1.1. 研究背景

近年、世界各国の大手企業において、本来の事業で培った技術を活用して、新たな分野の開拓・事業の幅の拡大を実現することにより、継続的な成長を目指す動きが見られるようになってきている。既存の日本国内の業種区分には、従来、日本株の分類とされてきた「東証33業種分類」や、これを17業種に再編した「TOPIX17業種区分」などがあるが、現状として、これらをはじめとした従来の業種区分には、急速に進む企業のビジネスモデルの変更・事業の多角化は反映されていない。

1.2. 先行研究

DTW(Dynamic Time Warping) 距離は、近年の計算機の性能向上により、その計算量の多さを克服することができるようになったため、金融をはじめとした広い分野で用いられはじめている。また、Paparrizos and Gravano(2015)[1] は、時系列データ間の類似尺度 SBD(Shape-based distance) とそれを用いた k -shape を提案した。

業種区分について、時系列データ間の類似尺度を用いて分析した研究は Tristan and Chin(2020)[2] による研究があるが、世界的にみても数が少ない。Tristan and Chin(2020)[2] は、分析手法に階層クラスタリングを用いたが、その結果の評価の難しさなどからビッグデータ解析においては一般的に非階層クラスタリングが用いられる。さらに、Megaro and Tierra(2020)[3] は、非階層クラスタリングの評価に Internal CVIs と Pihur et al.(2009)[4] に基づいたランク集計法を用いた。

1.3. 研究目的・検討項目

本研究の目的は、時系列データ間の類似尺度を用いた非階層クラスタリングによって、ビジネスモデルの変更・事業の多角化を反映した新たな(便宜的)業種区分を構築し、より優れたポートフォリ

オ運用、特にリスクコントロールを図ることである。またこれについての検討項目は、新たな業種区分、従来の業種区分(TOPIX17業種区分)、業種区分なしの場合について、アウトオブサンプル検証によって算出した複数の運用評価指標の結果により比較することである。

2. 研究手法

2.1. 使用データ

QUICK Workstation Astra Manager よりデータを取得。日経225採用銘柄である企業の株価データを用い、周期は日次、期間は2016~21年6月のうちの営業日1342日分とする。このうち、クラスタリングに2016~20年の5年分、アウトオブサンプル検証に2021年の半年分をそれぞれ用いる。

2.2. クラスタリングとその評価

本研究では、正規化+DTW+ k -medoids(DTW+ k -medoids(N))、標準化+DTW+ k -medoids(DT-W+ k -medoids(S))、 β 値正規化+DTW+ k -medoids(DTW+ k -medoids(β))のDTW距離を用いた3つのクラスタリング手法と、 k -shapeの合計4つのクラスタリング手法を用いる。なお、クラスタ数は、TOPIX17業種区分と比較するため、 $k=2\sim 17$ の範囲を考える。

その後、Silhouette Index や Calinski-Harabasz Index をはじめとした7つのIndexの値を算出し、Pihur et al.(2009)[4]が提案し、Megaro and Tierra(2020)[3]が用いたランク集計法により最適なクラスタ数を求める。

2.3. アウトオブサンプル検証

クラスタリングの結果を受けて、TOPIX17業種区分を集約した区分として日経平均6セクターを加えることとし、TOPIX17業種区分、新たな業種区分、業種区分なしの合計4区分7分類に対して、銘柄数が3と4のポートフォリオをそれぞれ10000個構築するモンテカルロ・シミュレーション

ンによるアウトオブサンプル検証を行う。このとき、ポートフォリオ1個当たりにつき10000シナリオの30日後の将来予測を行い、10000シナリオの平均的なReturn、Risk、Sharpe ratio、Treyner ratio、Sortino ratio、 β 値を1ポートフォリオずつ算出し、最終的には10000ポートフォリオの平均的な運用評価指標の値を算出する。

その後、1個当たり10000シナリオある10000個のそれぞれの運用評価指標のデータに対してJarque-Bera検定を行い、その結果に応じて適したTuckey型の多重比較を適用して有意差を測る。

3. 研究結果と考察

以下の表1は、アウトオブサンプル検証の結果の一部を表し、表内の各値は、TOPIX17業種区分の値を0とした場合に対する数値である。なお、全て年率換算後の値で、Mean ReturnとMean Riskの単位は%である。

表1：TOPIX17を基準としたシミュレーションの結果

Classification	M Return	M Risk	M β	M Sharpe
銘柄数 3				
日経平均6セクター	-0.898	-0.218	-0.015	-0.054
DTW+k-medoids(N)	-4.000	-0.356	-0.018	-0.221
DTW+k-medoids(S)	-3.018	-0.188	-0.013	-0.165
DTW+k-medoids(β)	-1.600	-0.089	+0.000	-0.085
k-shape	-6.254	-0.506	-0.028	-0.343
業種区分なし	-0.611	+0.022	-0.006	-0.035
銘柄数 4				
日経平均6セクター	-0.926	-0.199	-0.020	-0.058
DTW+k-medoids(N)	-5.188	-0.486	-0.024	-0.288
DTW+k-medoids(S)	-3.815	-0.403	-0.022	-0.212
DTW+k-medoids(β)	-2.109	-0.106	+0.001	-0.111
業種区分なし	-0.598	+0.072	-0.004	-0.034

また、以下の表2は、多重比較Steel-dwassの方法により得られたp値のうち、TOPIX17業種区分との比較のみを抜粋したものである。

表2：Steel-dwassの方法の結果(対TOPIX17)

Classification	M Return	M Risk	M β	M Sharpe
銘柄数 3				
日経平均6セクター	0.133	0.000***	0.000***	0.086 [†]
DTW+k-medoids(N)	0.000***	0.000***	0.000***	0.000***
DTW+k-medoids(S)	0.000***	0.016*	0.007**	0.000***
DTW+k-medoids(β)	0.010**	0.976	1.000	0.009**
k-shape	0.000***	0.000***	0.000***	0.000***
業種区分なし	0.709	0.998	0.648	0.716
銘柄数 4				
日経平均6セクター	0.041*	0.000***	0.000***	0.023*
DTW+k-medoids(N)	0.000***	0.000***	0.000***	0.000***
DTW+k-medoids(S)	0.000***	0.000***	0.000***	0.000***
DTW+k-medoids(β)	0.000***	0.819	0.813	0.000***
業種区分なし	0.622	0.575	0.918	0.638

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, [†] $p < 0.1$

表1と表2より、銘柄数 $N = 3, 4$ のときのいずれもMean ReturnとMean Sharpe ratioはTOPIX17業種区分が最も大きいことがわかる。その一方で、Mean RiskとMean β については、 $N = 3$ ときにはk-shape、 $N = 4$ ときにはDTW+k-medoids(N)がそれぞれ最も小さい値を示した。したがって、新たな業種区分によるリスクコントロールが図れるといえる。また、全ての運用評価指標でTOPIX17業種区分と業種区分なしの場合の間に有意差がなかったことから、TOPIX17業種区分の形骸化が考えられる。

4. 結言

時系列データ間の類似尺度を用いた非階層クラスタリングによるビジネスモデルの変更・事業の多角化を反映させた新たな業種区分を用いることで、リターンの向上は難しいが、より優れたリスクコントロールを図ることができる。また、TOPIX17業種区分の形骸化を考慮し、新たな業種区分の枠組みを作るべきである。

今後は、データの期間や周期の変更、ポートフォリオの銘柄数を増やした場合の結果への影響を検証すべきである。さらに、別の従来の業種区分の場合についても検証すべきである。

参考文献

- [1] Paparrizos, J., & Gravano, L. (2015). *k*-Shape: Efficient and Accurate Clustering of Time Series, SIGMOD(International Conference on Management of Data), 69-76.
- [2] Tristan, L., & Chin Sin, O. (2020). Portfolio Management: A Financial Application of Unsupervised Shape-based Clustering-Driven Machine Learning Method, International Journal of Computing and Digital Systems, 235-243.
- [3] Megaro, A., & Tierra, S. p. A. (2020). Multivariate time-series clustering for vehicle usage states identification, POLITECNICO DI TORINO.
- [4] Pihur, V., Datta, S., & Datta, S. (2009). RankAggreg, an R package for weighted rank aggregation, BMC Bioinformatics.