

# 株価分析のための時系列クラスタリング入門

白田 由香利

学習院大学経済学部 〒171-8588 東京都豊島区目白 1-5-1

E-mail: yukari.shirotaATgakushuin.ac.jp

あらまし 株価変動分析においても機械学習の時系列データクラスタリング手法が広く使われている。代表的な距離定義は k-Shape 法の SBD, 及び DTW などであるが, これらは入力データの標準化を前提とするので, 重要な分散(リスク)情報を捨て去ってしまうという問題がある。株価分析においては, 現在も, 1952 年にハリー・マーコウィッツが創案した「収益率の平均 (平均リターン) とリスク(リターンの標準偏差)の散布図」による分析法が広く使われている。この手法ではリスク情報を捨てずに用いているが, より詳細な変動パターンを知るためには(V字回復, S字型など), 機械学習クラスタリングが必要となる。そこで我々は Amplitude-based clustering 法という, 分散情報を残したまま変動パターンを解析する新アルゴリズムを開発し, 株価変動パターン分析手法を研究している。現在我々は, 第 1 段階で Amplitude-based clustering を行い, 成長率と変動パターンによって荒く分類し, 第 2 段階で詳細を分析したいクラスターに対してのみ k-Shape 法を行うという二刀流のアプローチを取っている。次元圧縮としては UMAP を使っている。本稿では, マーコウィッツの散布図の作り方から説明を始め, 上記二刀流の手法による分析方法を示す。クラスタリング結果について UMAP 上で考察するプロセスも解説する。事例として, インド企業などの株価分析結果を示す。本稿ではクラスタリングの結果をグラフィクスにより可視化し, 見るだけで違いが分かるように解説する。

**キーワード** 株価分析, マーコウィッツの現代ポートフォリオ理論, 時系列データクラスタリング, k-Shape 法, Amplitude-based clustering.

## 1. 始めに

本稿ではクラスタリングについては習得しているが, 株価データのクラスタリングについて初めてのユーザーを対象として, 株価データの時系列データクラスタリング手法を説明する。株価分析においては, 現在も, 1952 年にハリー・マーコウィッツが創案した「**収益率の平均** (平均リターン) と**リスク**(リターンの標準偏差)の散布図」による分析法[1][2]が広く使われている。株価分析においてリスク情報は重要である。しかし, マーコウィッツの手法では, 1 企業のデータは, 平均リターンとリスクという 2 つの数字のみで表現されるので, 株価変動パターンにおける変動のパターン情報は得られないという問題がある。

昨今の時系列データのクラスタリングは, 変動パターンによるクラスタリングを可能とし, その点ではマーコウィッツの手法を凌駕する。しかし, k-Shape 法及び DTW による手法では, 入力データの標準化を前提とするため, リスク情報が失われてしまうので, 有効ではない。我々は機械学習クラスタリング技術によりマーコウィッツの手法を拡張したい。そこで我々はリスク情報を残した手法 Amplitude-based clustering 法を開発した。現在は, 第 1 段階で Amplitude-based clustering を行い, 成長率と変動パターンによって荒く

分類し, 第 2 段階で詳細を分析したいクラスターに対してのみ k-Shape 法を行うという二刀流のアプローチを取っている。次元圧縮としては UMAP [3][4]を使っている。

本稿では, 分析の 3 つプロセス (1)株価データの加工, (2)時系列データクラスタリング, (3)次元圧縮による評価, 株価分析の流れ全体を説明する。Amplitude-based clustering 手法はデータ間の距離の分散が大きくデータ解釈などが k-Shape 法と異なるので, その用法について詳しく解説する。

第 2 節は株価データの加工, 及びマーコウィッツの現代ポートフォリオ理論を簡単に説明する。第 3 節では, 時系列データクラスタリングの距離と手法について概説する。第 4 節では, Amplitude-based clustering 法を簡単に説明し, その活用法を解説する。第 5, 6 節では, 事例をあげてデータ分析の具体例を示す。最終節はまとめである。

本稿で用いた株価データはすべてビューロー・ヴァン・ダイク社の企業情報データベース Orbis を用いて検索した。グラフィクスは Wolfram Research 社の Mathematica を用いて作成した。

## 2. マーコウィッツの現代ポートフォリオ理論

本節では、1952年にハーリー・マーコウィッツによって発表された論文[1]を始めとする現代ポートフォリオ理論及び、シャープによる Capital Asset Pricing Model(CAPM) [5] [6]を説明する。

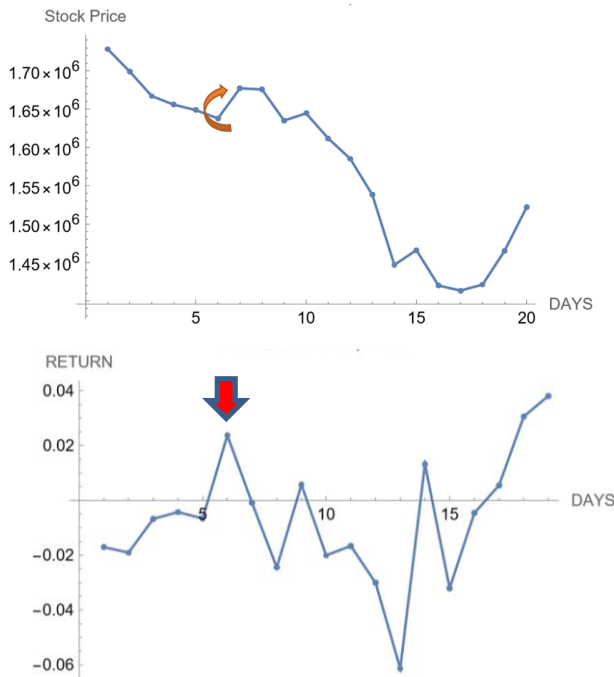


図 1: 株価変動とその対数リターングラフ

まず株価の収益指標，リターンは以下のように株価前日比の自然対数で表すことが一般的である(j 日目の株価を  $S_j$  で表している) [7].

$$\log_e \frac{S_j}{S_{j-1}}$$

テイラー展開により, 前日価格と今日の価格が類似している場合, 以下に示すように対数を用いない通常の収益率とほぼ等しくなる.

$$\log_e \frac{S_j}{S_{j-1}} \cong \frac{S_j - S_{j-1}}{S_{j-1}}$$

図 1 にリターン変動の例を示した. 上図は株価の変動そのままであり, 下図はリターンの変動を表す. 対数関数であるので, 前日比が 1 以上で正の値を取り, 1 以下で負の値を取る. また, 価格上昇と下落が殆ど同じ値となる. 例えば, 上昇 100 円→110 円で 0.095, 下落 110 円→100 円で -0.095 となり, 絶対値は等しい. これは対数の  $\log \frac{A}{B} = \log A - \log B$  という性質で証明できる.

収益率  $\frac{A-B}{B}$  では 0.1, -0.09 のように値が異なってしまう.

マーコウィッツのポートフォリオ理論では, リターンを用いるが, クラスタリングの際は, 株価そのま

まの値を使う場合と, リターン値を使う場合の両方がある. 人間が目視でクラスタリング結果を確認する際は株価の方が, 解釈が容易であるためである.

図 2 にマーコウィッツのリスク-平均リターン図の例を示す. 1 つの点が 1 つの銘柄を示す. リスクは当該期間内のリターンの標準偏差である. 図 3 に 2 銘柄から構成されるポートフォリオのリスク-平均リターン図を示した. A と B のバランス比によってリスク, リターン平均は変わる. リスクは, A と B の相関係数が -1 に近いほど小さくでき, A や B の単体より小さい値になることはあるが, 平均リターンについては, A あるいは B の平均リターンを上回ることはいない.

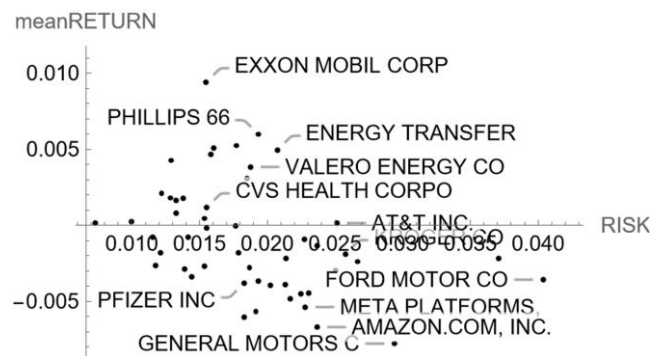


図 2: マーコウィッツのリスク・リターン平面の例

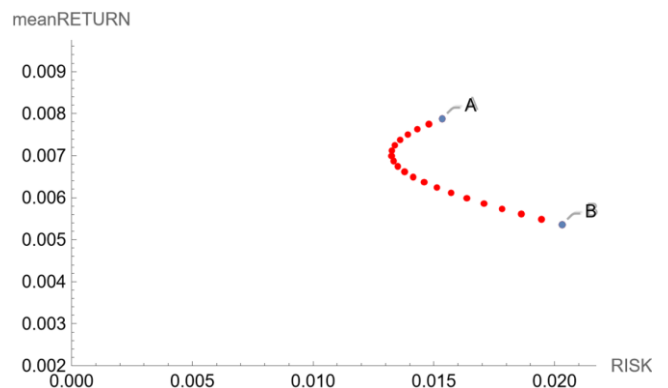


図 3: 曲線の上半分が効率的フロンティア

マーコウィッツは主要論文”Portfolio Selection”において「平均リターンだけを基準とするのでは分散投資の原理に従っていない. 所与の水準の分散に対して期待リターンが最も高いポートフォリオを選択すべき」と提案した[8]. この理論を平均分散行動原理と呼ぶ. 例えば, 図 3 においてリスク 0.015 に対して 2 つのポートフォリオの可能性があるが, 同じリスクであれば平均リタンの高い方を選択すべきである. 平均リターンが大きい方のポートフォリオ群を「効率的である」と呼び, 平均リターン及びリスクの曲線の上半分を効率的フロンティアと呼ぶ. ここまでをまとめると, 理

論的には、リスクの高い資産に集中投資をするよりも、リスクの低いポートフォリオに投資するほうが望ましく、効率的フロンティアの部分を選択したほうがよい。しかしこの理論的帰結はいくつかの過程に依拠しており現実的な条件の下では妥当ではない[8]。キンローらは3つの修正すべき仮定をあげているが[8]、本稿ではその内容にはふれない。

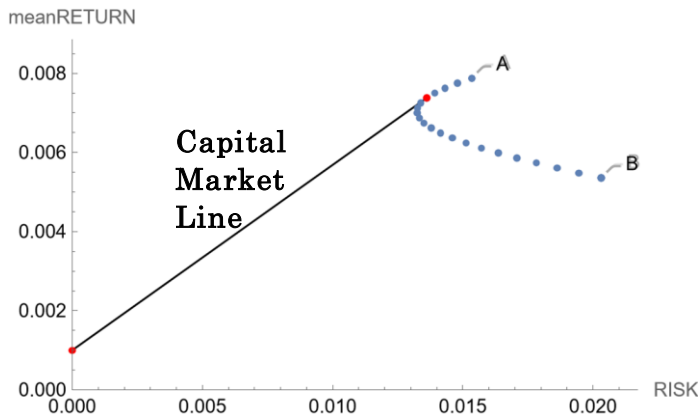


図 4: Capital Market Line

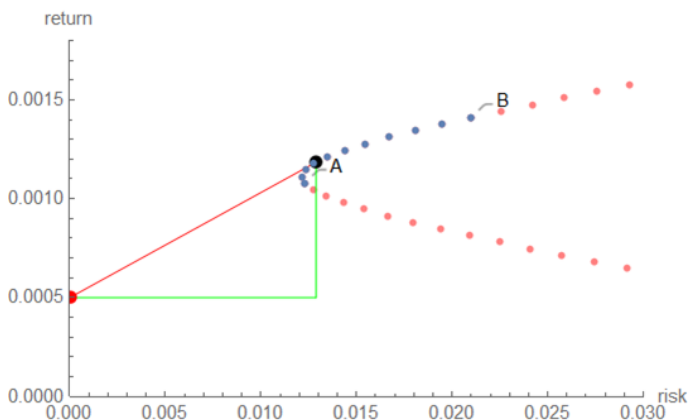


図 5: Sharpe ratio 最大化問題

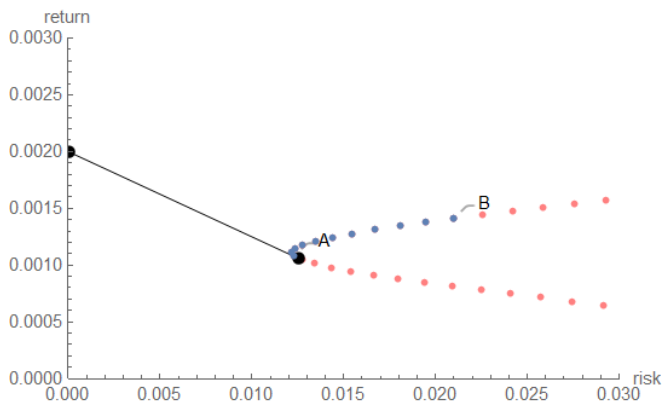


図 6: 接点ポートフォリオの解がない場合

次に、シャープが発展させて構築した **capital asset pricing model (CAPM, 資本資産評価モデル)** を説明する[5]。CAPM では、無リスク資産と、効率的ポートフォリオの組合せを考える。無リスク資産とは、リスクが殆どなく、安全な資産を指す。先進国の短期の国債を用いることが多い。図 4 のリスク 0 の y 軸上の点が無リスク資産に相当する。

無リスク資産とポートフォリオを組み合わせの場合、フロンティア上のすべての点の中で、リスク増分 ÷ 平均リターン増分 (Sharpe ratio) を最大とする点が 1 点存在する。このポートフォリオを接点ポートフォリオと呼ぶ。この計算は、A と B のバランス比  $t$  を媒介変数として、Sharpe ratio を最大化する問題として数学的に解ける(図 5 参照)。解が存在しない場合もある(図 6 参照)。2 種類の資産を 3 以上の複数資産にしてポートフォリオを構築する場合でも、ラグランジェの未定乗数法[9]を用いて解を計算できる。Sharpe ratio の解法事例としては以下の教材を参照して頂きたい[10]。

<https://shirotaabc.sakura.ne.jp/hosomichi/PDF/SharpRatio.pdf>

<https://shirotaabc.sakura.ne.jp/hosomichi/PDF/SharpRatioQA1.pdf>

<https://shirotaabc.sakura.ne.jp/hosomichi/PDF/portFchanges.pdf>

解が存在する場合、合理的な投資家は、無リスク資産と接点ポートフォリオに分散投資する。リスク回避型の投資家であれば無リスク資産の比率を増大させ、リスクを取る投資家は接点ポートフォリオの比率を増大させる。

この節で示したように、ポートフォリオ構築においてはリスクが重要であり、株価変動の分散データが重要である。よって k-Shape 法のようなデータ標準化を前提にするクラスタリング手法では、株式分析に有用とは言えない。これに関しては第 5 節で後述する。

### 3. 時系列クラスタリングの手法と距離

本節では、クラスタリング手法のアルゴリズムと距離定義を概説する。初めにアルゴリズム、次に距離を説明する。

#### 3.1 アルゴリズム

クラスタリングの代表的アルゴリズムとして、k-means 法と階層型クラスタリング法がある[11]。まず k-means 法のアルゴリズムを説明する。始めに  $k$  個の点をランダムに選び、一時的なクラスター重心とする。次に、各データの最近傍重心を選び、当該データをそのクラスターのメンバーとする。そして新たなクラス

ターメンバーにおける、そのクラスター重心を再計算する。これをクラスター間のメンバー移動がなくなるまで繰り返す。

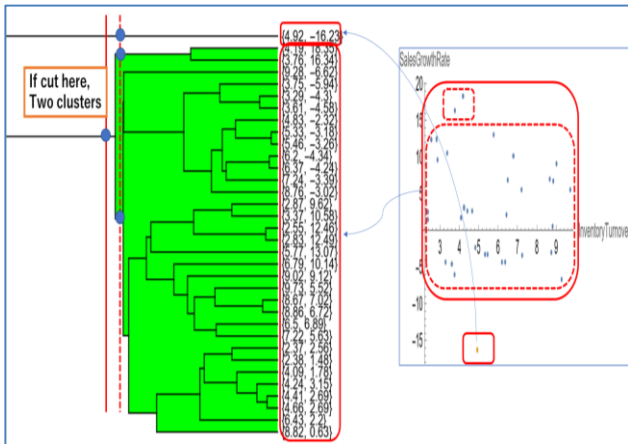


図 7:階層型クラスタリングの結果のデンドログラム

次に階層型クラスタリング手法を説明する。図 7 に階層型クラスタリングによって生成されたクラスター間の距離関係を表わすデンドログラムを示した。近接するデータとしてクラスターを生成し、次第に周囲のデータを取り込みクラスターは大きくなる。最終的には全データは一つのクラスターとなる。

3.2 距離

次にクラスタリングの代表的アルゴリズムとして、k-means 法と階層型ク時系列データクラスタリングで広く使われている距離 3 種類を説明する。

(A) 相関係数を用いた距離  $d_{i,j} = \sqrt{\frac{1}{2}(1 - \rho_{i,j})}$

(B) k-Shape 法の SBD(Shape-based Distance)

(C) DTW(Dynamic Time Warping)距離

上記(A)は、Prado が階層型リスクパリティ法(HRP 法) [12-14]で使っている距離である。HRP はポートフォリオ構築のためのクラスタリング手法として先駆的に普及した。Prado 氏は第 1 回目の ACM AI in Finance で基調講演も行い[15], ポートフォリオ構築用に HRP を拡張した多くの階層型クラスタリング法[16] [17] [18] [19]が発表された。(A)の距離定義は、相関係数を使っていることから、データ標準化を行っており、その時点でリスク情報は消去されている。換言すると HRP 法のクラスタリングにおいては、リスク情報は使われていない。HRP 法を用いた自動車製造業の株価分析例が [20] [21] [22] [23]に紹介されている。

次に k-Shape 法の SBD を説明する[24, 25]. SBD は、信号処理における相互相関関数 Cross Correlation 関数の考えに基づく。関数 f(), g()に対して以下のような積

$$(f \times g)(w) = \int f(t) \times g(t + w)dt$$

分を考え、期間 w だけ g()をずらした時、積分値が最大となるような w を発見する。離散的な共通の時間列に対する 2 つの時系列 x, y に対しては以下のように、Cross-Correlation\_w, CCw を定義する(図 8 参照)。

$$CC_w(x, y) = \max_w \sum_{i=1}^m x_i \times y'_i$$

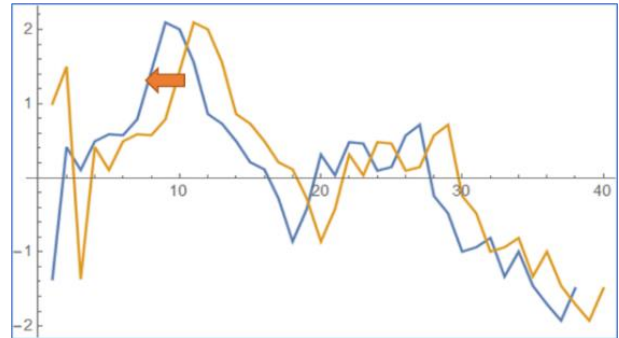


図 8: k-shape 法では片方のパターンを w だけずらす

SBD は CCw を標準偏差で割り標準化した後、1 から減じることで距離として 0 から 1 の範囲の値を作る。SDB の特長は、ピークのシフトや、スケーリングに堅固であることである。

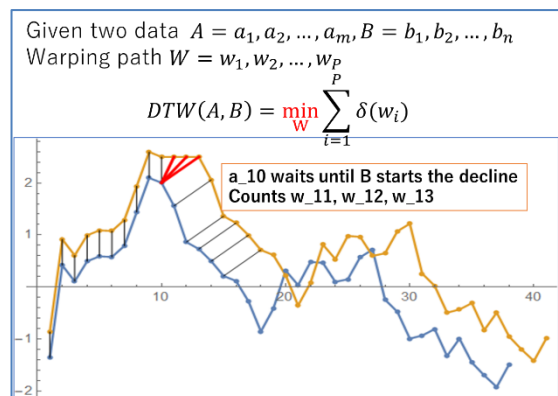


図 9:DTW の warping path の決定的ようす

次に DTW 距離を説明する。DTW の特長は、時間軸を調節することである。図 9 に示すように、片方が別のデータが下落し始めるのを待っている。2 つの時系列の対応点を全体距離が最小になるように決める[26]. DTW は時間のずれに堅固で、信号のゆがみにも堅固であるが、k-shape 法に比較して実行時間が長くなるという欠点がある。オーダーは O(mn)である。株価クラスタリングにおいては、株価データは細かい変動の繰り返しであるため、本来次の山であるべき頂点と対応付けられてしまう可能性があり(山谷の位相のマッチがずれる), 性能では k-shape 法のほうが若干優れている[27-29].

#### 4. Amplitude-based Clustering

本節では Amplitude-based clustering を説明する。

我々はデータ標準化を行わないクラスタリング手法を求めていた。株式データではマーコウィッツの理論においてもリスク情報は重要であり、経済の応用においては、GDP の成長率の比較など各種の分散情報を必須とするグラフがある。それらのクラスタリングのため白田と Chakraborty は Amplitude-based clustering 手法を開発した[30].距離はユークリッド距離であり 2 銘柄の距離を定義した後、距離の距離を列データ対列データとして定義する。距離距離データを用いることで全体の中での 2 銘柄の距離が定義できる。ここで  $G$  はリターン値を、 $T$  は全営業日数、 $N$  は銘柄数を示す。

$$ED_{ij} = \sqrt{\sum_{k=1}^T (G_{i,k} - G_{j,k})^2}$$

$$\widetilde{ED}_{ij} = \sqrt{\sum_{n=1}^N (ED_{n,i} - ED_{n,j})^2}$$

距離距離行列を入力として階層型クラスタリングを行い、その後、准対角化を行う。准対角化では、銘柄の順序を入れ替えて対角線上に可能な限り距離が小さいセルが並ぶようにする(図 10 参照)。目視でも、白いセルの集合としてクラスターが認識できる。k-shape 法の結果と異なり、分散の値が大きいため、クラスター数を距離の閾値で決めるというアプローチは適切ではない。分析者の興味の対象として、分析対象クラスターを選択すべきである。図 10 に示すように、青い線で囲まれたクラスターに注目する、あるいは、そのクラスターを分割した黒い線のクラスターへ分割して分析することも可能である。

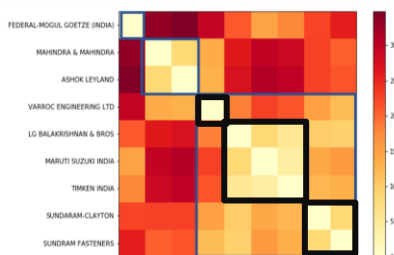


図 10: Amplitude-base clustering の准対角化後の距離距離行列

Amplitude-based clustering 法は距離定義及びアルゴリズムに新規性はないが、用法としてクラスタリングの解釈法に新規性がある。用法として以下の 3 つがある。

- (用法 1) 成長率あるいは分散が最大の銘柄を探す
- (用法 2) 距離の閾値を設定し、閾値以下のクラスター

一(複数)を抽出し、パターンが類似している銘柄群を作成する。

(用法 3) 興味のある銘柄を含むクラスターを抽出する(距離の閾値は自分で決める)

上記用法については次節で具体例を用いて説明する。現在、我々は以下の二刀流アプローチで株価分析を行っている。第 1 段階で Amplitude-based clustering を行い、成長率と変動パターンによって荒く分類し、第 2 段階で詳細を分析したいクラスターに対してのみ k-Shape 法を行う。分散が同程度の銘柄のクラスターを一度作ってから、そのクラスターに対して k-shape 法を適用することは意味がある。

#### 5. 事例 1:自動車数製造業の株価分析

本節では株価クラスタリング事例として、2008 年金融危機における世界 135 社の自動車製造業の時価総額の変動パターンをクラスタリングした例を示す[31]. 株価でなく時価総額を選んだ理由は、株価の場合、分割が起こると 1 株の株価が減少し、それまでのデータとの整合性のため調整を行う必要があり、扱いが煩雑であるからである。時価総額とは、1 株の株価×発行株数であるため、値に連続性がある。特に長期の株価成長率を分析したい場合などは時価総額が適している。データの取得期間は 2007 年 1 月から 2011 年 12 月の月次データである。サブプライムショックにより製造業の殆どの株価は下落したが、その後多くの企業で株価は上昇に転じた。本論文[31]では、その回復力レジリエンスを比較分析した。時価総額データは比較のため、2007 年 1 月データを 1 としてインデックス化した(図 11 参照)。時価総額の回復率が大きく異なることが分かる。

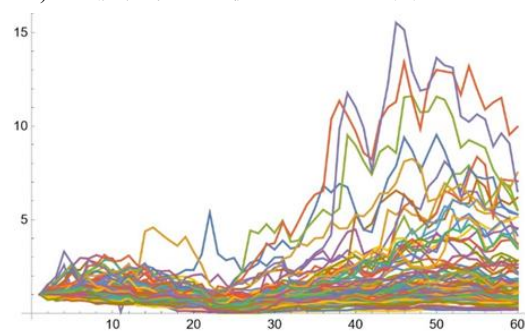


図 11: インデックス化された時価総額

図 12 に Amplitude-based clustering の結果を示す。本分析では、距離距離の閾値を 4 として、類似パターンをもつクラスターを抽出した(図 13 参照)。結果、8 個のクラスターが抽出できた。図 13 では、我々が注目している日本の大企業を含むクラスターが茶色で示されている。我々は系列会社が近傍に位置しているかを分析する目的で、この日本企業に注目し、類似パターン

をもつ企業群を調べた。結果は、系列会社が多く含まれていた。図 14 には距離距離行列を入力して UMAP した結果を示す。図 14 のクラスター0(青色の点)とは、当該 8 個のクラスター以外の銘柄のことであり、デンドログラムにおいて右端と左端に存在する。デンドログラムにおいて成長率の大きい銘柄は左端に存在している。

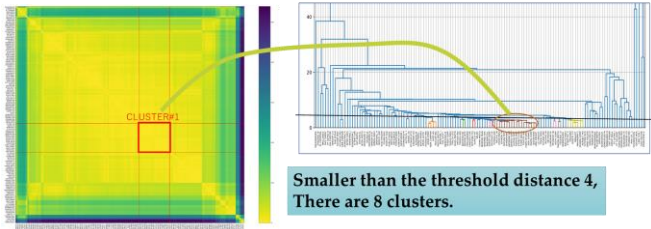


図 12：距離距離の閾値を 4 として、類似パターンをもつクラスターを抽出

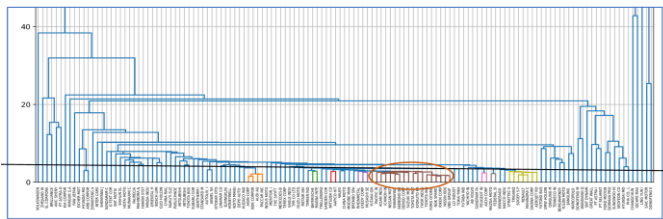


図 13：距離距離の閾値 4 で抽出できた 8 個のクラスター

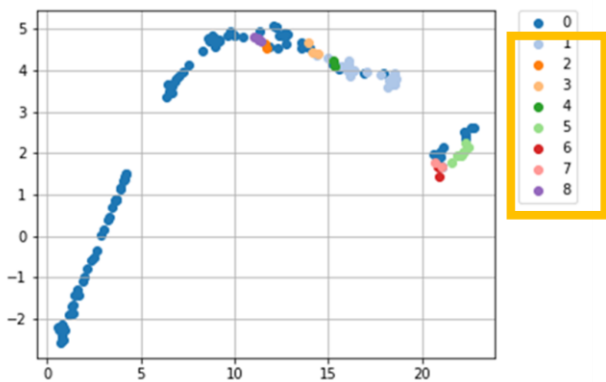


図 14：距離距離行列を UMAP した結果

Amplitude-based clustering では標準化を行わないため、銘柄によって分散は大きく異なる(図 11 参照)。従って銘柄間の距離が大きくなるため、UMAP の結果は図 14 のように曲線状になった。抽出された 8 個のクラスターを色で識別しているが、類似パターンの銘柄は UMAP 上でも近傍に位置することが目視で確認できる。図 15 に 2 次元に次元圧縮した UMAP の主成分軸の意味解釈を行う。図 15 に、UMAP の x 軸の値とその銘柄の変動パターンを示した。図中、矢印は 2008 年 10 月を示す。ここで下落が見られる。X の値が小さ

いほど 10 月の金融危機下落時からの回復力が大きいことが分かる。よって、UMAP による x 軸は、レジリエンスの弱さを示す主成分と解釈できる。図 14 で x 軸上の左の方に位置する銘柄の方がレジリエンスが高いと解釈できる。

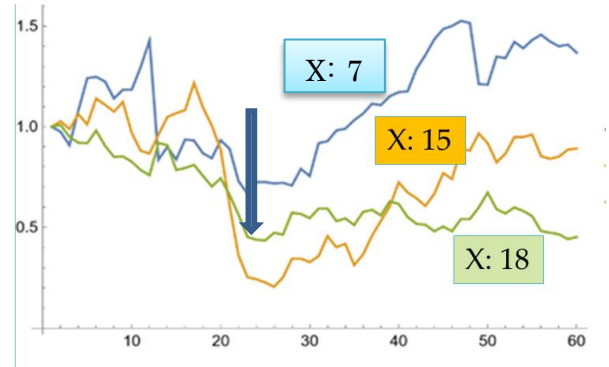


図 15:UMAP の x 値が小さいほどレジリエンスが高い

本事例では Amplitude-based clustering の用法を示し、次元圧縮の結果からレジリエンスの強さを表す主成分を抽出することに成功した。また、大企業とその系列会社の変動パターンが類似していることを確認できた。

### 6. 事例 2:インド企業の時価総額分析

本節では、長期に渡るインド企業の収益変化のクラスタリングを行う。対象は 2020 年度の収益 TOP50 社の 2010 年から 2019 年の時価収益(Profit/Loss before Tax)年次データの変動分析を行った[32]。インド企業の収益規模拡大は著しい。その中でも IT サービス業 HCL, Infosys 等の成功は顕著であるので、これらの収益の成長率を分析したかった。図 16, 17 に Amplitude-based clustering の結果を示す。閾値を設定せず、興味の対象クラスター A,B,C,D,E を設定している。

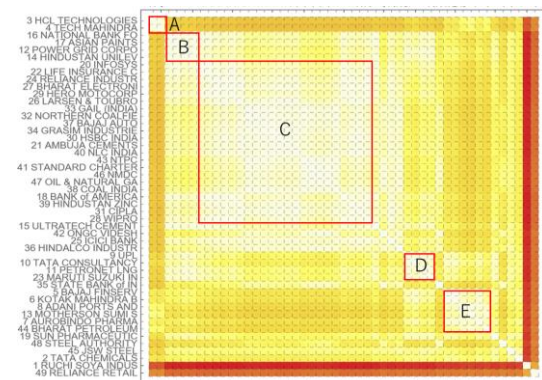


図 16：インド企業の Amplitude-based clustering 結果

クラスター A が最も分散の大きい企業である(図 18 参照)。A の 2 社は、Tech Mahindra 及び HCI であり、成長率約 6 倍という急成長をしている。図 19 にクラスター B の 4 社の成長パターンを示した。パターンの類似性が高く、右肩上がりに順調に成長しているよう

すが見られる. この4社にITサービス業はなく, 業種は各種多様であった.

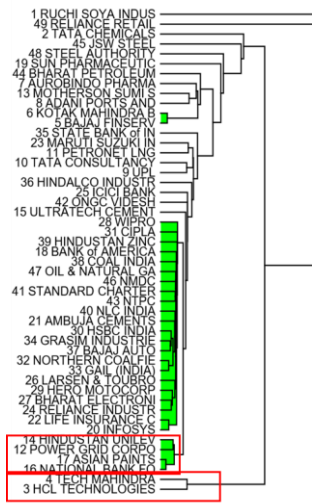


図 17: インド企業の Amplitude-based clustering 結果

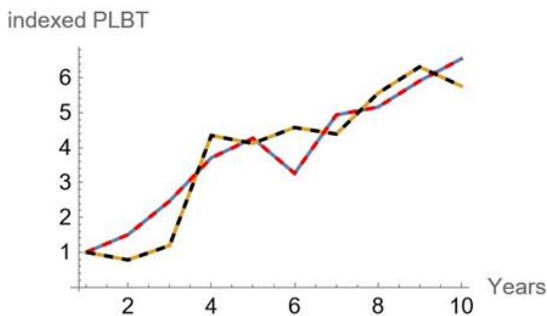


図 18: クラスターAは最も収益成長率が高い2社

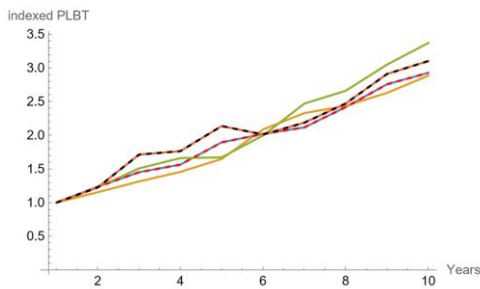


図 19: クラスターBは2番目に収益成長率が高い4社

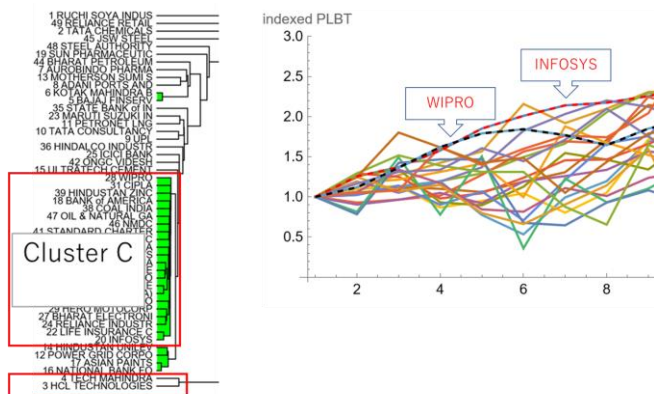


図 20: クラスターCは多くの企業を含む

図 20 にクラスターCの収益変動パターンを示した. ITサービス企業, INDFOSYS, WIPRO がクラスターCの上位に位置していることが見える. インドのITサービス業の躍進が確認できた. クラスターCは類似した分散(距離閾値)の間に多数の企業が存在する. クラスターCの場合, k-shape法によってより詳細なクラスタリングをすべきである. 次にクラスターEを見ると, その変動パターンはS字型をしている企業が多いことが分かる(図 21 参照). 2019年度のデータで下落を示す企業が3社ある. このように, S字型, V字型回復のような形状でグループ分け可能な点がマーコウィッツの理論にない利点である.

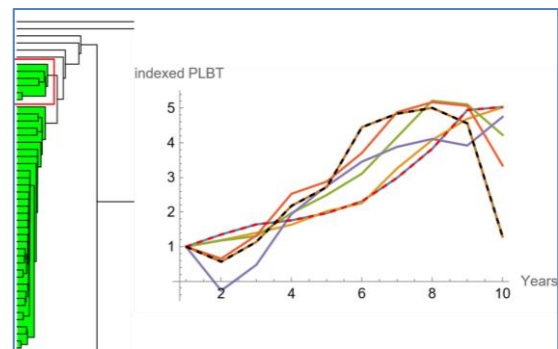


図 21: クラスターEの変動パターン

まとめると, インド企業の収益性の急成長のようすを時系列クラスタリングで評価した. 想定したようにITサービス業の成長は顕著であった. また, パターンのS形状などの企業群のクラスター等も発見でき, 順調な発展だけではない企業のようすを分析できた.

### 7. まとめ

本稿では, 分析の3つプロセス(1)株価データの加工, (2)時系列データクラスタリング, (3)次元圧縮による評価, 株価分析の流れ全体を事例を用いて説明した. 株価分析の基本技術としてマーコウィッツの現代ポートフォリオ理論を説明した後, その分析を発展あるいは補完する意味で, 機械学習クラスタリングでは変動パターンによるクラスタリングが可能となることを示した. マーコウィッツ理論が示すようにリスク(分散)情報は重要である. しかし, k-shapeなどのクラスタリングは, データ標準化を前提とするため, 分散の分析はできない. そこで Amplitude-based clustering手法を用いた. 本稿では, Amplitude-based clusteringの具体的な用法を示した. 株価分析における目的は, 成長が顕著である銘柄の発見, 特定銘柄と類似する変動をする銘柄の発見等である. 企業の詳細な内部データがなくても, 株価データ変動のクラスタリング分析から, 系列会社の動き, 変動の将来予測など多くの知見が得られる. 企業分析に株価クラスタリングは有効手

段であるので、企業分析者はぜひとも機械学習クラスタリングを活用して頂きたい。

## 参 考 文 献

- [1] H. Markowitz, "Portfolio Selection, ," *Journal of Finance*, vol. 7, no. 1, pp. 77-91, 1952.
- [2] M. Rubinstein, "Markowitz's" portfolio selection": A fifty-year retrospective," *The Journal of finance*, vol. 57, no. 3, pp. 1041-1045, 2002.
- [3] L. McInnes. "UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction — umap 0.5 documentation," 2023/2/2.
- [4] A. Coenen, and A. Pearce, "Understanding UMAP; 2019," URL <https://pair-code.github.io/understanding-umap>.
- [5] W. F. Sharpe, "Capital asset prices: A theory of market equilibrium under conditions of risk," *The journal of finance*, vol. 19, no. 3, pp. 425-442, 1964.
- [6] W. F. Sharpe, "The sharpe ratio," *Streetwise—the Best of the Journal of Portfolio Management*, vol. 3, pp. 169-185, 1998.
- [7] 富島佑允, *投資と金融がわかりたい人のためのファイナンス理論入門 プライシング・ポートフォリオ・リスク管理: CCC メディアハウス* 2018.
- [8] ウィリアム・キンロー, マーク・クリッツマン, and デービッド・ターキントン, *誤解だらけのアセットアロケーション: 実務家のためのガイド: 東洋経済新報社*, 2020.
- [9] 白田由香利, *悩める学生のための経済・経営数学入門: 3つの解法テクニックで数学アレルギーを克服!*: 共立出版, 2009.
- [10] 白田由香利. "数学グラフィクス奥の細道," <https://shirotaabc.sakura.ne.jp/hosomichi/>.
- [11] S. Aghabozorgi, A. S. Shirkhorshidi, and T. Y. Wah, "Time-series clustering—a decade review," *Information Systems*, vol. 53, pp. 16-38, 2015.
- [12] D. P. M. Lopez, "Building diversified portfolios that outperform out of sample," *The Journal of Portfolio Management*, vol. 42, no. 4, pp. 59-69, 2016.
- [13] M. L. De Prado, *Advances in financial machine learning*: John Wiley & Sons, 2018.
- [14] M. M. L. de Prado, *Machine learning for asset managers*: Cambridge University Press, 2020.
- [15] M. L. d. Prado, Keynote speech "Advances in Financial Machine Learning", in ACM ICAIF (AI in Finance), online, 2020.
- [16] M. Molyboga, "A modified hierarchical risk parity framework for portfolio management," *The journal of financial data science*, vol. 2, no. 3, pp. 128-139, 2020.
- [17] J. Pfitzinger, and N. Katzke, "A constrained hierarchical risk parity algorithm with cluster-based capital allocation," *Stellenbosch University, Department of Economics*, 2019.
- [18] H. Lohre, C. Rother, and K. A. Schäfer, "Hierarchical Risk Parity: Accounting for Tail Dependencies in Multi - asset Multi - factor Allocations, " *Machine Learning for Asset Management: New Developments and Financial Applications*, pp. 329-368, 2020.
- [19] L. Cao, "Ai in finance: challenges, techniques, and opportunities," *ACM Computing Surveys (CSUR)*, vol. 55, no. 3, pp. 1-38, 2022.
- [20] 白田由香利, "株価変動パターンの機械学習クラスタリング," in 日本経営数学会第44回(通算64回)研究大会, 東京, 2022, pp. pp. 4-7.
- [21] 白田由香利, 村上朱音, 佐倉環, and 久保山哲二, "コロナ期のグローバル市場の指標変動パターンのクラスタリング," *学習院大学計算機センター年報*, vol. 43, 2022.
- [22] 村上朱音, "COVID-19 パンデミック拡大初期における世界的自動車製造業の株価変動クラスタリング分析," *学習院大学経営学博士前期課程修論*, 2022年3月.
- [23] Y. Shirota, K. Yamaguchi, A. Murakami, and M. Morita, "An analysis of political turmoil effects on stock prices: a case study of US-China trade friction." *Proc. of ACM ICAIF (AI in Finance)*, pp. 1-7, 2020.
- [24] J. Paparrizos, and L. Gravano, "k-shape: Efficient and accurate clustering of time series." pp. 1855-1870.
- [25] J. Paparrizos, and L. Gravano, "Fast and accurate time-series clustering," *ACM Transactions on Database Systems*, vol. 42, no. 2, pp. 1-49, 2017.
- [26] F. Petitjean, A. Ketterlin, and P. Gançarski, "A global averaging method for dynamic time warping, with applications to clustering," *Pattern recognition*, vol. 44, no. 3, pp. 678-693, 2011.
- [27] H. Ito, A. Murakami, N. Dutta, Y. Shirota, and B. Chakraborty, "Clustering of ETF Data for Portfolio Selection during Early Period of Corona Virus Outbreak," *Gakushuin Journal of Economics*, vol. 58, no. 1, pp. 99-114, 2021.
- [28] Y. Shirota, and B. Chakraborty, "Automakers Stock Price Movement Comparison under COVID-19." pp. 375-379.



- [29] Y. Shirota, and A. Murakami, "Long-term Time Series Data Clustering of Stock Prices for Portfolio Selection." pp. 1-6.
- [30] Y. Shirota, and B. Chakraborty, "Amplitude-Based Time Series Data Clustering Method," *Gakushuin Economics Papers*, vol. 59, no. 2, pp. 127-140, 2022.
- [31] S. Matsubishi, and Y. Shirota, "Resilience Evaluation of Automakers After 2008 Financial Crisis by UMAP," *International Journal of Machine Learning*, 2023.
- [32] Y. Shirota, B. Chakraborty, and H. Yoshiura, "An Analysis of Indian Companies' Profit Growth Pattern by Amplitude-based Clustering," in 5th IEEE Intl. Conf. on Knowledge Innovation and Invention (ICKII) 2022 Hualien, Taiwan and online, 2022, pp. in printing.