

価格変動パターンを用いた市場予測

～IDTW Based k -medoids clustering の株式市場への適用

Financial Markets Prediction with Using Price Fluctuation patterns

中川 慧[†] 今村 光良[†] 吉田 健一[‡]

Kei Nakagawa Mitsuyoshi Imamura Kenichi Yoshida

1. はじめに

様々な資産の価格を予測するために様々な方法が提案されている。ファイナンスの分野では、資産価格の変動を説明する様々なファクター(特徴量)を発見してきた[1,2]。

本研究では、資産価格の予測のため、現在の価格変動が過去のいつの時点に似ているかという価格変動パターンを、特徴量として抽出する。抽出にあたっては、指数化した価格変動に対して Dynamic Time Warping(DTW)距離[3]を計測する Indexing DTW(IDTW)[5]を用いた非類似度行列に対して k -medoids クラスタリング[4](IDTW Based k -medoids clustering)を行う。当該手法による時系列クラスタリングにより、予測に有効な価格変動パターンを可視化して把握することが可能となる。

今まで金融市場において十分に活用されてこなかった価格変動パターンの有効性を報告する。

2. IDTW Based k -medoids clustering

先行研究[4,5]において、IDTW を用いて適切にスケールリングされた月間の日次株価変動パターンが予測に有効な特徴量であることが報告されている。本研究では背後にある価格変動パターンを時系列クラスタリングによって具体的に抽出し、さらなる予測精度の向上を目指す。

月間の株価変動はデータ数が月毎に異なるため、単純なベクトル空間上のユークリッド距離を用いたクラスタリング手法(例えば、 k -means 法)は適用できない。よってデータ数が一定でない場合の類似度の測定、クラスタリング方法を適切に組み合わせる必要がある。

本研究では以上を考慮し、IDTW を用いて作成された非類似度行列に対して、 k -medoids クラスタリングを行うことを提案する。月間の株価のような長さの異なる金融時系列データに対して、自然なクラスタリングが可能となることが本手法の利点である。

2.1 Dynamic Time Warping (DTW)

現在の株価変動に近い過去の株価変動を抽出する際に、時系列データ間の類似度(距離)を計測する必要がある。時系列データ間の類似度を測るため、これまで多くの尺度が提案されてきた。よく使用される単純な尺度として相関係数やユークリッド距離が挙げられる。

ただし相関係数はデータ間の線形関係しか捉えることができず、ユークリッド距離は、人間の直観に反する結果を生じてしまう場合がある[7]。人間は時系列データの形を柔軟に認識できるのに対し、ユークリッド距離では時間方向の対応が固定化されるためである。また相関係数やユークリッド距離では 2 つの時系列の長さが異なる場合に計算ができないという問題点がある。

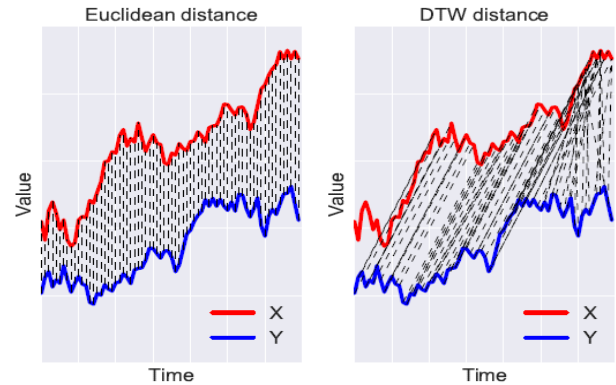


図 1:ユークリッド距離と DTW 距離の比較

DTW 距離[3]はユークリッド距離における当該問題点を克服する尺度である。すなわち長さの異なる時系列データの比較を可能とし、2つの時系列に対して時間軸を伸縮させながら最適な対応付けを行う距離尺度である(図1)。以下に DTW 距離を算出するアルゴリズムを記載する。 x, y はそれぞれ長さ N, M を持つ時系列データである。 w は伸縮を制限するウィンドウであり、本研究では5とした。距離関数 d についてはマンハッタン距離 $d(x, y) = |x - y|$ とした。

Algorithm 1 DTW distance

```

1: procedure DTW( $x, y, w = 5$ )
   //Initialize Matrix  $D$ 
2:   Var  $D(N, M)$ 
3:    $D(1, 1) = 0$ 
4:   for  $i = 2$  to  $N$  do
5:     for  $j = 2$  to  $M$  do
6:        $D(i, j) = 1$ 
7:     end for
8:   end for
   //Calculate DTW

```

distance

```

9:   for  $i = 2$  to  $N$  do
10:    for  $j = \max(1, i - w)$  to  $\min(M, i + w)$  do
11:

```

$$D(i, j) = d(x_{i-1}, y_{j-1}) + \min \begin{pmatrix} D(i, j-1), \\ D(i-1, j), \\ D(i-1, j-1) \end{pmatrix}$$

```

12:   end for
13: end for
14: return  $D(N, M)$ 
15: end procedure

```

[†]一般社団法人情報処理学会, IPSJ[‡]一般社団法人電子情報通信学会, IEICE

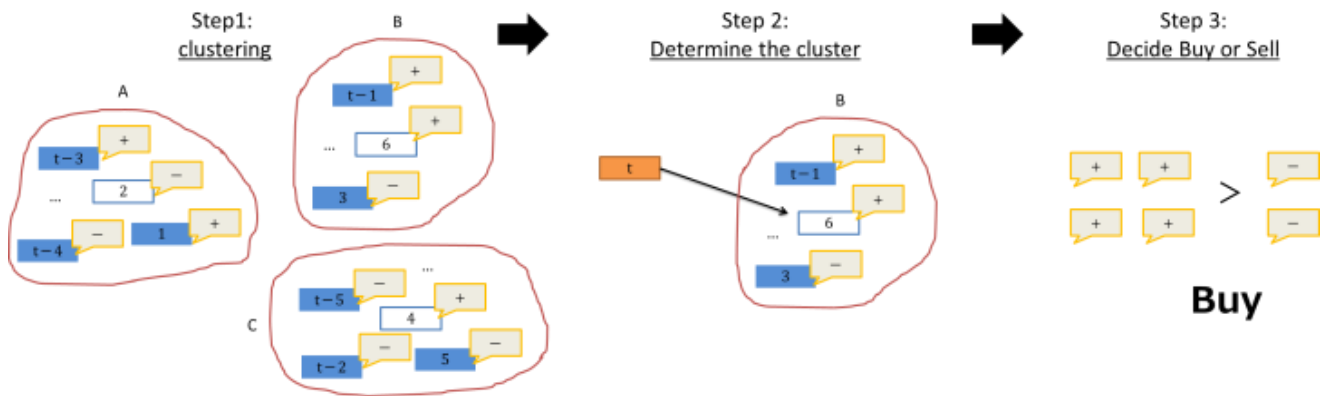


図 2:分析手順

2.2 Indexing Dynamic Time Warping (IDTW)

株価は計測期間(日次/週次/月次)によって変動幅が異なり、また時期によって株価水準も大きく異なる。そのため、株価に DTW を適用するためには(1)適切な計測期間と(2)標準化の方法を決める必要がある。

株価の季節性は投資家において広く認識されており、とりわけ 5 月に株を売却すると良いとされる"Sell in May" が有名である。実際に世界の株式市場で"Sell in May" 効果が確認されている[8]。つまり投資家は"月"単位で株価の変動を認識していると言える。そこで(1)については、ひと月内の日次の株価を計測期間とする。

また実務的には、異なる期間の株価を比較する際には標準化、あるいは収益率換算して比較するよりも、期初を 1 とし指数化して比較の方が自然である。この点を考慮して、(2)については月初の値を 1 として、月内の日毎の株価推移を、前日との比を使って表わす(指数化)。

以上の点を考慮して先行研究[5,6]では、株価の類似度を計測する尺度として IDTW を提案し、IDTW によって抽出された価格変動パターンが予測に有効な特徴量であることを検証した。以下に IDTW 距離算出の具体的なアルゴリズムを記載する。

Algorithm 2 IDTW distance

```

1: procedure IDTW(x, y)
    //Scaling Data
2:  Var  $I_x, I_y$ 
3:   $I_{x_1} = 1, I_{y_1} = 1$ 
4:  for  $i = 1$  to  $N$  do
5:     $I_{x_i} = I_{x_{i-1}} \frac{x_i}{x_{i-1}}$ 
6:  end for
7:  for  $i = 1$  to  $M$  do
8:     $I_{y_i} = I_{y_{i-1}} \frac{y_i}{y_{i-1}}$ 
9:  end for
    //Apply DTW
10: return  $DTW(I_x, I_y)$ 
11: end procedure
    
```

2.3 IDTW Based k -medoids clustering

k -medoids 法とは、 k -means 法と類似した分割最適化クラスタリングの手法である。 k -means 法との具体的な相違点は、クラスターは centroid(重心)ではなく medoid で代表される点である。medoid とは、クラスター内のデータ点で、その点以外のクラスター内の点までの非類似度の総和が最小になる点である。直感的には、最もクラスターの中心付近に位置する点を表す。従って、medoid は必ずクラスタリングの対象となるデータ内に存在する。そのため、 k -medoids 法は分類するデータ間の非類似度行列を与えれば適用でき、任意の非類似度尺度(距離)に対して実行できる。すなわち、クラスタリングの対象が長さの異なる時系列のような、ベクトルで表現されたデータ以外であっても非類似度を定義できれば適用できる。

そこで本研究では、月間の株価のような長さの異なる金融時系列データに対して、IDTW を非類似度尺度に使用した k -medoids 法を適用することを提案する。IDTW を用いることで、金融時系列データに対して自然なクラスタリングが可能となる。具体的なアルゴリズムは以下の通り。 $\{x_1, \dots, x_N\}$ はそれぞれ長さの異なる時系列データ、 k はクラスター数である。停止条件(stopping criterion)はクラスターの割り当てが変化しなかった場合、または繰り返し回数上限(100回)に達した場合である。

Algorithm 3 IDTW Based k -medoids clustering

```

1: procedure IDTW Based  $k$ -clustering( $\{x_1, \dots, x_N\}, k$ )
2:  Randomize  $m_1, \dots, m_k$ 
3:  While stopping criterion has not been met do
    //Cluster Assignment
4:    for  $i = 1$  to  $k$  do
5:       $c_i := \{x_i | IDTW(x_i, m_k) \leq IDTW(x_i, m_l)\}$ 
6:    end for
    //Update Medoids
7:    for  $j = 1$  to  $k$  do
8:       $m_j := \min_{x_l \in c_j} \sum_{i=1}^N IDTW(x_l, x_i)$ 
9:    end for
10: return  $m_1, \dots, m_k$ 
11: end procedure
    
```

表 1: クラスター数毎の結果

	収益率		正答率	
	IDTW	DTW	IDTW	DTW
k = 2	1.13	0.23	58.87	53.23
k = 3	0.98	-0.17	57.26	51.61
k = 4	1.48	0.66	61.29	53.23
k = 5	1.62	0.64	63.71	54.03
k = 6	1.51	0.50	64.52	54.03
k = 7	1.31	0.05	61.29	51.61
k = 8	1.32	0.06	62.10	53.23
k = 9	1.47	0.45	62.90	54.84
k = 10	1.14	0.41	58.06	55.65
k = 11	1.04	0.79	58.87	59.68
k = 12	0.90	0.56	59.68	58.06

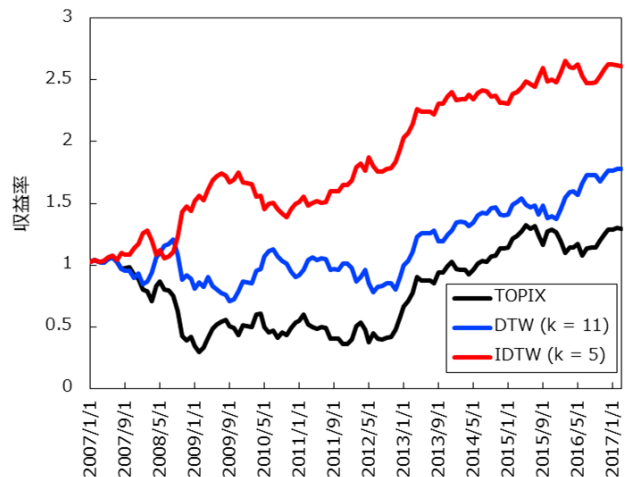


図 3: TOPIX、DTW 及び IDTW によるクラスタリングの収益率推移

3. 実証分析

3.1 分析手順

本章では、TOPIX 指数(配当込)を用いた実証分析により、類似した価格変動パターンのクラスタリングと、抽出したクラスターが特徴量として有効かどうかを確認するため予測精度の分析を行う。データは情報端末の Bloomberg から取得した。クラスタリング手法は、提案手法の有効性を確認するため、DTW Based k -medoids clustering(DTW)、及び IDTW Based k -medoids clustering(IDTW)の2つの手法を用いる。なお、クラスタリングの際にクラスター数 k を決める必要があるが、ここでは、 $2 \leq k \leq 12$ まで動かし、予測精度が最も良くなる、すなわち最も価格変動を説明するクラスター数を抽出する。予測精度は収益率と正答率の両面で評価する。

具体的な分析手順は下記の通り(図 2 も併せて参照)。データ期間は 1989 年 1 月から 2017 年 3 月までとし、学習期間は 2006 年 12 月まで、検証期間は 2007 年 1 月から 2017 年 3 月までの 10 年間とした。

分析手順

- Step 1: $t-1$ 月までの月間の日次価格変動をもとに、クラスター数 k を固定し、クラスタリングを行う。各月は翌月のリターンの上昇、下落をラベルとして保持する。
- Step 2: t 月の月間の価格変動が属するクラスターを特定する。
- Step 3: t 月が属するクラスターを求め、当該クラスターにおいて、上昇(下落)のラベルが多い場合は、月末の価格で買い(売り)、収益を計算する。同数の場合には、上昇と下落のラベルのそれぞれのリターンの平均値を取り、大きい方を採用する。
- Step 4: $t = t + 1$ に進めて Step 1 に戻る。

3.2 分析結果

各クラスター数の結果のサマリーが表 1 である。収益率、正答率ともに IDTW が DTW を上回る傾向がみられる。

IDTW のクラスター数については、 $k = 4, 5, 6$ 辺りにピークがあり、それ以降については収益率、正答率ともに減衰している。そのため、予測という観点からは TOPIX 指数の価格変動のクラスター数は 5 程度であると言える。一方で DTW によるクラスタリングでは、 $k = 11, 12, 13$ 辺りにピークがあり、IDTW と比べ、クラスター数が増加する。また、図 3 は TOPIX と DTW、IDTW の収益率が最も良い結果の収益率推移を示した図である。IDTW は期間に依らず右肩上がりの推移であり、TOPIX と DTW を上回る結果となった。

図 4 は、2017 年 3 月の分析手順における Step 1 の IDTW を用いた、 $k = 5$ のクラスタリング結果を示した図である。黒線が中心点として選ばれた価格変動であり、赤線が当該クラスターに属する価格変動を表す。また、表題のサンプルはクラスター内に含まれるサンプル数、かつここは翌月のリターンの上昇確率を表す。

図 4 のクラスターの解釈を左から順に述べる。価格変動が下落傾向に関わらず、翌月の上昇確率が高く、リバーサルを表している。次の 2 つのクラスターは強い価格上昇と下落を表しており、翌月も上昇と下落確率が高く、典型的なモメンタムを表している¹。4 つ目のクラスターは下落を伴うがほぼ横ばいの変動を示し、翌月も横ばいの変動である。最後は変動を伴わない横ばいの変動であるが、翌月に下落する可能性が高い。

一般に日本株式市場においては、モメンタムの有効性が確認できないことが多い[9]。しかし、本節で確認したように価格変動の類似度に基づきクラスターを適切に定義することで強いモメンタム効果が確認できた。

¹過去に価格が上昇(下落)した場合、将来も上昇(下落)が続く現象をモメンタムといい、一方で、過去に価格が上昇(下落)した場合、将来は反転して下落(上昇)がする現象をリバーサルという。

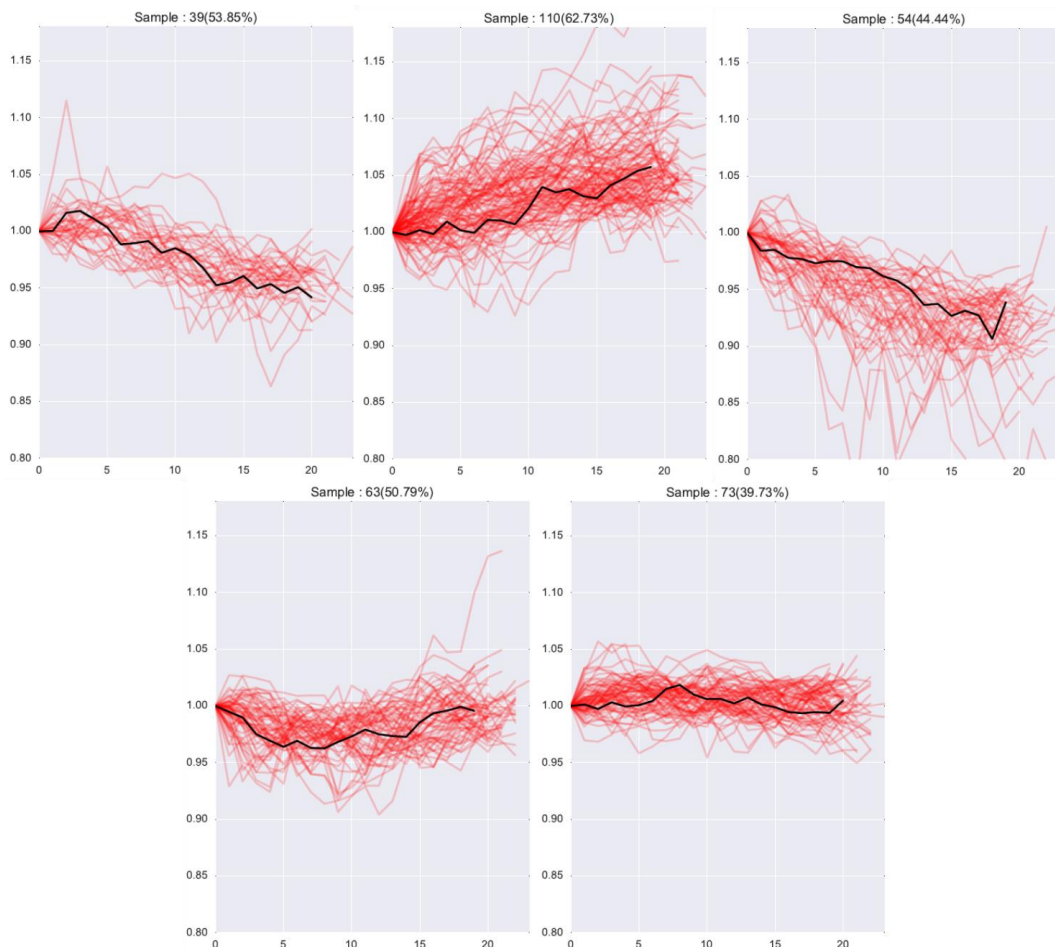


図4: クラスター数5の場合のIDTWによるクラスタリング結果
 左から、リバーサル、モメンタム、モメンタム、横ばい、横ばいのクラスターを形成している。

4. おわりに

本研究では、株価の価格変動パターンのクラスタリングを行い、予測のための特徴量として抽出した。抽出にあたっては、指数化した価格変動に対してDTW距離を計測するIDTWを用いた非類似度行列に対して k -medoidsクラスタリング(IDTW Based k -medoids clustering)を提案した。TOPIX指数を用いた実証分析とクラスターを可視化した結果、以下の点が確認できた。

- 収益率、正答率ともにIDTWによるクラスタリングがTOPIX指数、DTW両者を上回り、高い予測精度となった。
- IDTWによるクラスタリングでは、予測精度という点で価格変動パターンのクラスター数は5程度であることがわかった。
- 日本市場では有効性が確認できないモメンタム効果がクラスタリングによって抽出できた。

参考文献

- [1] Fama, Eugene F., and Kenneth R. French. "The cross-section of expected stock returns." *The Journal of Finance* 47.2 (1992): 427-465.
- [2] Harvey, Campbell R., Yan Liu, and Heqing Zhu. "... and the cross-section of expected returns." *The Review of Financial Studies* 29.1 (2016): 5-68.
- [3] Itakura, Fumitada. "Minimum prediction residual principle applied to speech recognition." *IEEE Transactions on Acoustics, Speech, and Signal Processing* 23.1 (1975): 67-72.
- [4] Kaufman, Leonard, and Peter Rousseeuw. *Clustering by means of medoids*. North-Holland, 1987.
- [5] 中川 慧, 今村 光良, 吉田 健一, "株価変動パターンの類似性を用いた株価予測" 2017年度人工知能学会全国大会(第31回) 論文集(2017).
- [6] 今村 光良, 中川 慧, 吉田 健一, "資産価格変動パターンの類似性に着目した金融市場予測の評価", 2017年度人工知能学会全国大会(第31回) 論文集(2017).
- [7] Keogh, Eamonn J., and Michael J. Pazzani. "Scaling up dynamic time warping for datamining applications." *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2000.
- [8] Bouman, Sven, and Ben Jacobsen. "The Halloween indicator," Sell in May and go away": Another puzzle." *The American Economic Review* 92.5 (2002): 1618-1635.
- [9] Asness, Clifford S., Tobias J. Moskowitz, and Lasse Heje Pedersen. "Value and momentum everywhere." *The Journal of Finance* 68.3 (2013): 929-985.