

Dynamic Time Warping を用いた 類似チャートパターンの探索

目次

- I. はじめに
- II. 先行研究
- III. DTW 距離の算出とその特徴
- IV. 終わりに

投資顧問業務部 ラップ運用 G 姫野 公佐

I. はじめに

将来の相場を予測する際に、現在のチャート形状と似ている過去の局面を参考情報にすることがある。株式市場の場合には、株価チャートの形状が類似していることで、必ずしも将来にわたって同じ経過を辿るとは限らないものの、株価チャートには各時点のファンダメンタルズや金融政策、政治動向、それらに対する市場参加者の投資家心理や実際の投資行動まで数多くの情報が反映された結果である。そのため、投資判断の参考情報には十分なり得ると考えられる。実際、テクニカル分析と呼ばれる株価予測の技法の一部には、特定の株価チャートが繰り返し発生する傾向を前提として手法が体系化されている例もあり、将来の相場を予測する際に現在のチャート形状と似ている過去の局面を探索する意義があると考えられる。

そのような2つのデータ同士の類似度を定量的に測る指標として多く用いられるものの一つに相関係数がある。しかしながら、データ期間が異なると計算ができない点や、人間の視覚的な直観と指標が一致しない場合があるなど、課題もある。

そうした課題を解決する手法として、主に音声解析分野で用いられてきた Dynamic Time Warping (DTW 距離) を金融時系列データに用いて分析を行う研究²がなされている。本稿では、主に金融時系列データに対して DTW 距離を用いた先行研究を紹介した上で、実際の DTW 距離の算出方法や相関係数等との指標値の比較を行う。その上で、実際の金融市場のデータを用いた類似チャートパターンの探索を行う。最後に結論と今後の課題を述べる。

¹ 例えば、ある特定の期間に2つの株価チャートが『上昇の後下落』というパターンは同一で人間の視覚的に類似度の高いチャート形状であっても、上昇と下落の開始時点が異なる場合には必ずしも高い相関係数とはならず、指標が示す結果と人間の感覚が異なる場合が想定される。

² 先行研究の詳細はII 先行研究に記載した。

Ⅱ. 先行研究

DTW 距離は、主に音声解析の分野で開発されたとされている。人間の言葉を音声波形として捉えた場合、同じ言葉であっても話し手や会話の場面等によって話す速度(波長)や声の大きさ(振幅)が異なることもある為、音声波形の波長や振幅に違いが生じる。そこで、波形そのものは同一であっても、波長や振幅が異なる音声データに対して柔軟に比較できる尺度³として DTW 距離が開発され、活用されるようになった。[1](参考文献、以下同様)

その後、金融時系列データ同士の類似度の比較に DTW 距離が適用されるようになった。[4]は、ある銘柄の株価の動きに連動する、いわゆる関連銘柄の発見方法として DTW 距離を用いる手法を提案した。また、株価予測に用いる研究として、[2]は日次の株価リターンを「月初=1」として指数化して DTW 距離を算出する Indexation DTW (IDTW) 距離を提案した。DTW 距離が「形状の一致」に注目した比較方法であるのに対し、IDTW 距離は「形状」に加えて「位置の一致」にも注目する方法である。例えば、同一の指数同士の比較であっても、時点が異なると指数そのものの水準が異なる為、形状は同一であっても距離そのものが大きくなることもある。このようなケースに対応するため、月初=1として指数化された IDTW 距離では、時間のズレによる水準の差も考慮して距離を算出する手法である。また[5]では、分析対象となる月の日次株価変動と IDTW 距離の近い過去の月⁴を抽出し、その翌月の月次リターンがプラスなら買い、マイナスなら売りとした場合の収益率を計算し、東証株価指数(TOPIX)のリターンを上回ることを報告した。

[6]は更に、分割最適化クラスタリングの手法である k -medoids 法に DTW 距離・IDTW 距離を用いて、過去の株価指数に関する月別の日次データを任意の数でクラスタリングを実施。対象となる月が属するクラスターにおいて、翌月のリターンで上昇が多い場合は買い、下落が多い場合は売りとした場合のバックテストを実施。この場合の収益率を計算したところ、東証株価指数(TOPIX)のリターンを上回ることを報告した。その他にも、国内株式市場以外の市場を対象とした分析もなされている。[3]は、4か国の株式市場と3通貨の為替市場⁵、仮想通貨市場に対して IDTW 距離を用いた実証分析を行い、異なる市場価格の予測に有効であることを示した。

以上のように、DTW 距離などの分析手法は、過去の類似したチャート形状を探索することに留まらず、株価予測に対しても一定の効果が実証されており、運用者にとってその有用性だけでなく、パフォーマンス向上に資する有効性も十分に期待されるものである。

³ 例えば、「2本のゴムひも(時系列データ)を伸び縮み(時間軸の調整)させて、どの程度まで形状を似せることができるか(最小の距離でマッチング)で類似性を評価する」とも考えられる。音声解析以外の分野での活用例として、「歩き方の分析(速く歩いても遅く歩いても、同じ人かどうかを見分ける)」、「心拍や脳波の比較(体調の変化を見つけるために、過去のデータと比較)」等がある。

⁴ 論文では、分析対象となる月と過去の IDTW 距離の近い月を順に1, 3, 6, 12, 18, 24, 30, 36ヵ月間抽出し、翌月の月次リターンの平均値を求めて分析を行っている。その結果、株価の上昇または下落が一致する正答率は3ヵ月が最も高く、正答率は64.17%となったことが示されている。

⁵ 4か国の株式市場は米国株式(S&P500)、フランス株式(CAC40 指数)、ドイツ株式(DAX)、イギリス株式(FTSE100)。3通貨の為替市場はドル円(USDJPY)、ユーロ円(EURJPY)、ポンド円(GBPJPY)。

Ⅲ. DTW 距離の算出とその特徴

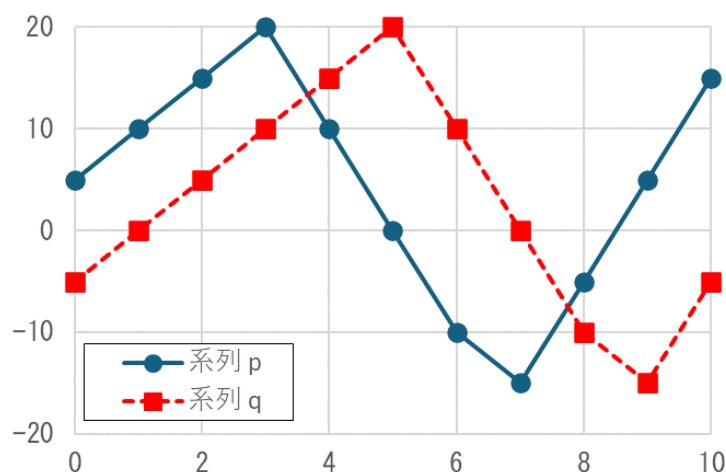
1. DTW 距離の算出例

以下で、DTW 距離、IDTW 距離の特徴を示し、株価分析の具体的事例を紹介する。
まず、以下2つの系列データを用いて、相関係数と DTW 距離の指標の違いを示す。具体例として、系列データ $p = \{p[0], p[1], \dots, p[i], \dots, p[10]\}$, $q = \{q[0], q[1], \dots, q[i], \dots, q[10]\}$ を用いて、相関係数と DTW 距離に基づく類似度を算出する。系列データ p, q の要素を①のように定義する。

$$\begin{aligned} p &= \{5, 10, 15, 20, 10, 0, -10, -15, -5, 5, 15\} \\ q &= \{-5, 0, 5, 10, 15, 20, 10, 0, -10, -15, -5\} \end{aligned} \quad \dots \textcircled{1}$$

この2つの系列データは、図表1のように位相は異なるものの波形は同一である。

図表1：系列 p, q の推移グラフ



まず、系列データ p, q の相関係数 ρ は、以下の②式で算出される。

$$\rho_{p,q} = \frac{\sum_{i=0}^{10} (p_i - \bar{p})(q_i - \bar{q})}{\sqrt{\sum_{i=0}^{10} (p_i - \bar{p})^2} \sqrt{\sum_{i=0}^{10} (q_i - \bar{q})^2}} \quad \dots \textcircled{2}$$

その結果、相関係数は 0.072 となる。相関係数が 1 に近いほど正の相関、-1 に近いほど負の相関、0 に近いほど無相関と一般的にいわれているため、0.072 という値は系列データ p, q が概ね無相関であることを示している。併せて以下の③式で定義される系列データ p, q の絶対値距離を算出する。

$$\sum_{i=0}^{10} |p[i] - q[i]| \quad \dots \textcircled{3}$$

結果は 145 となる。

次に、DTW 距離の算出を行う。DTW 距離は、2つの実数 $p[i]q[j]$ の距離⁶を $d(p[i], q[j])$ とする時、以下④式で算出される累積距離 $\gamma(i, j)$ を利用して作成する累積距離マトリクスを用いて計算できることが知られている。

$$\gamma(i, j) = d(p[i], q[j]) + \min \begin{cases} \gamma(i-1, j) \\ \gamma(i, j-1) \\ \gamma(i-1, j-1) \end{cases} \quad \dots \textcircled{4}$$

図表 2 では、まずマトリクスの縦軸左下から左上に p 系列のデータを、横軸左下から右下へ q 系列のデータを並べている。次に、各マスには、④式で算出される $(p[i], q[j])$ の「累積距離 $\gamma(i, j)$ 」を表示している。

次に、表左下マスにある両系列の始点($i = 0, j = 0$)での累積距離 $\gamma(0, 0)$ は、④式に $i = 0, j = 0$ における各要素をあてはめ算出。具体的には、 $p[0] = 5, q[0] = -5$ でありその距離 $d(p[0], q[0])$ は $5 - (-5) = 10$ 。④式右辺第 2 項は該当する値がない(*null*)であるので $\gamma(0, 0)$ は 10 となる(⑤式)。

$$\begin{aligned} \gamma(0, 0) &= d(p[0], q[0]) + \min[\gamma(p[-1], q[0]), \gamma(p[0], q[-1]), \gamma(p[-1], q[-1])] \\ &= 10 + \min[\text{null}, \text{null}, \text{null}] \\ &= 10 \end{aligned} \quad \dots \textcircled{5}$$

次に p 系列のみ時間を 1 つ進め、 $p[1] = 10, q[0] = -5$ であり、距離 $d(p[1], q[0])$ は $10 - (-5) = 15$ となり、累積距離 $\gamma(p[1], q[0])$ は式④に当てはめ $15 + 10 = 25$ となる(⑥式)。

$$\begin{aligned} \gamma(1, 0) &= d(p[1], q[0]) + \min[\gamma(p[0], q[0]), \gamma(p[1], q[-1]), \gamma(p[0], q[-1])] \\ &= 15 + \min[10, \text{null}, \text{null}] \\ &= 25 \end{aligned} \quad \dots \textcircled{6}$$

同様に、各時点の $p(i)$ と $q(j)$ を組み合わせた全ての累積距離を計算して表を作成している。

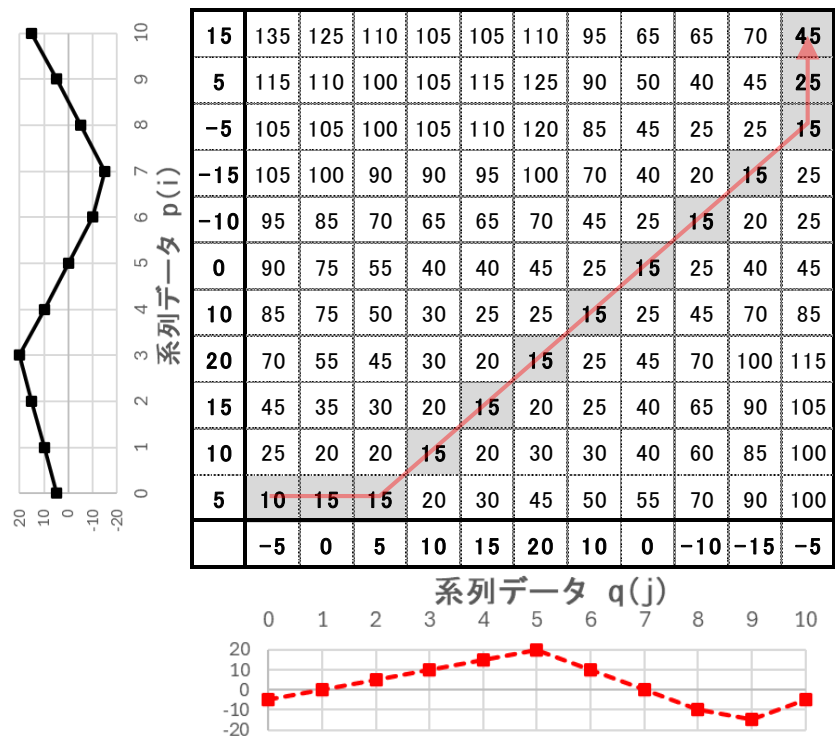
この時、時間の変化は縦軸で上(p の時間を進める)、横軸で右(q の時間を進める)、斜め右上(p, q 両方進める)のみを許容し、累積距離は、マトリクス始点($p(0), q(0)$)から、各時点

⁶ 距離関数は非負であればよく、通常ユークリッド距離やマンハッタン距離等が用いられる。

$(p(i), q(j))$ までの合計距離が最小となるルートで計算されることとなる。その結果として作成されたのが図表2のマトリクスである。始点から終点までを最短ルート(赤線で表示)でつないだ合計距離を DTW 距離と定義しており、この距離が小さい場合は「似ている」、大きい場合は「似ていない」と評価される。

今回のケースでは、右上のマスにおける累積距離 $\gamma(10,10)$ である 45 が DTW 距離となる。

図表 2：系列 p, q の DTW 距離算出の為の累積距離マトリクス



この値は前頁で算出した絶対値距離の 145 よりも小さく、同一の波形である系列データ p, q の特徴をより捉えていると考えられる。

このように、波形の類似性を捉えたい場合には、DTW 距離は相関係数や絶対値距離と比較すると優れた距離尺度である可能性がある。また、例示した系列データ p, q は同一の長さのデータとしたものの、DTW 距離は長さが異なるデータでも算出が可能であり、そのような点も相関係数や絶対値距離と比較した特徴であると考えられる。

2. 絶対値距離と IDTW 距離との比較

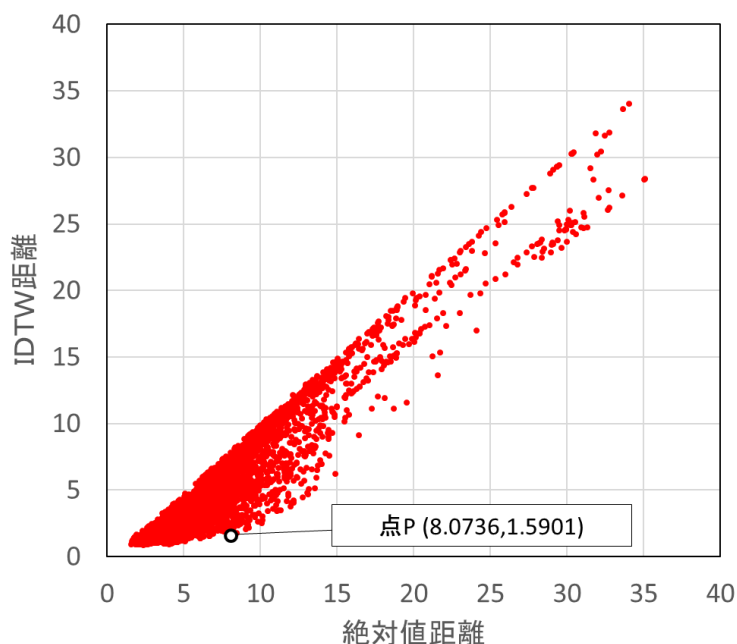
前項では、サンプルデータを用いて、絶対値距離と DTW 距離を比較し、より波形の変動パターンを捉えることができる可能性について述べてきた。ここでは、実際の金融市場のデータを用いて、絶対値距離と IDTW 距離との指標の比較を行い、その特徴を述べていく。

図表3は、TOPIX と S&P500 指数の日次リターンに対する期初を1とした 125 日間の絶対値距離を横軸、同指数同士の 125 日間の IDTW 距離を縦軸として、データ期間を期初から期末まで1日ずつスライドした結果をプロットした散布図である。データ期間は 2005 年

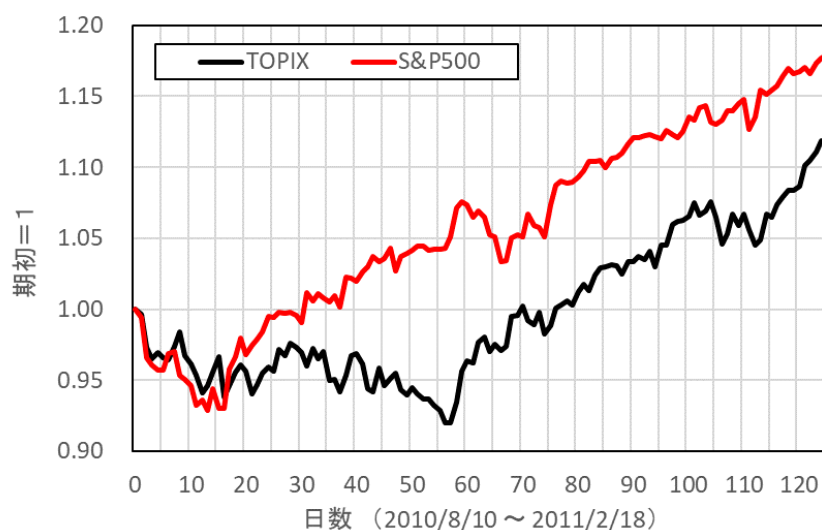
8月30日から2025年8月29日までの4,730日間とした。もし、2つの指数から算出される絶対値距離と IDTW 距離のそれぞれの距離尺度が似た傾向を示すのであれば、横軸の絶対値距離が小さい時に縦軸の IDTW 距離も小さく(類似度が高く)なる。反対に横軸の絶対値距離が大きい時に縦軸の IDTW 距離も大きく(類似度が低く)なる為、右肩上がりの対角線上に点が集まるはずである。また、対角線上以外の領域、中でも絶対値距離よりも IDTW 距離が小さくなる、対角線を境に右下の領域に点がプロットされる場合には、絶対値距離と比較して IDTW 距離が金融市場のデータの波形をより捉えている可能性がある。

結果を見ると、全体的に右肩上がりの対角線上に点がプロットされたものの、対角線を境に右下の領域にもプロットされた点が多数みられる。これは、横軸の絶対値距離よりも縦軸の IDTW 距離が小さいことを示しており、絶対値距離が捉えきれない金融市場のデータの波形を IDTW 距離が捉えている可能性がある。図表3左下の領域の中で、絶対値距離を IDTW 距離で除した値が最も低いプロット、すなわち絶対値距離と比較して IDTW 距離を用いた方が最も効果的と思われるケースを点P(絶対値距離=8.0736、IDTW=1.5901)とし、図表4にその対象期間125日分の東証株価指数と S&P500 指数の推移グラフを掲載した。各指数は期初を1として指数化している。その結果、対象期間における東証株価指数と S&P500 の指数は、35～55日目辺りでの上昇・下落の違いにより水準が異なるものの、期間全体として価格変動の形状は概ね一致している印象を受ける。一例ではあるが、DTW 距離(IDTW 距離)は絶対値距離と比較し、価格変動パターンの特徴を捉えられている可能性が高く、株価予想を行う上での実用性が期待できる手法と考える。

図表3：東証株価指数と S&P500 の絶対値距離と IDTW 距離
(125日ローリング日次リターンベース)



図表4：点P(図表3)における東証株価指数とS&P500の指数推移)



3. 東証株価指数の過去の類似チャートパターンの探索

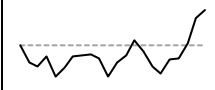
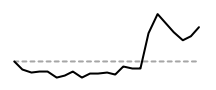
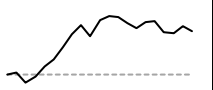
前項では、東証株価指数と S&P500 の2つの指数を用いて、相関係数と DTW 距離 (IDTW 距離) の特徴について述べてきた。ここでは、IDTW 距離を用いて、同一指数内における過去の類似局面の探索を行う。具体的には、2005 年 8 月 1 日から 2025 年 8 月 29 日までの東証株価指数の日次リターンデータについて、月初を 1 として月別に指数化し、241 ヶ月間の月毎の IDTW 距離を算出する。図表5には、算出例としてデータ期初3ヵ月(2005 年 8 月から 2005 年 10 月)とデータ期末3ヵ月間(2025 年 6 月から 2025 年 8 月)の計算結果を示した。

図表6には、前述の計算にて IDTW 距離が 0.0484 と最も小さかった組み合わせである 2017 年 7 月と 2023 年 2 月を含む5ヵ月間の株価チャートを表示した。類似度が最も高かった期間は、図表上に太字矢印にて記した。

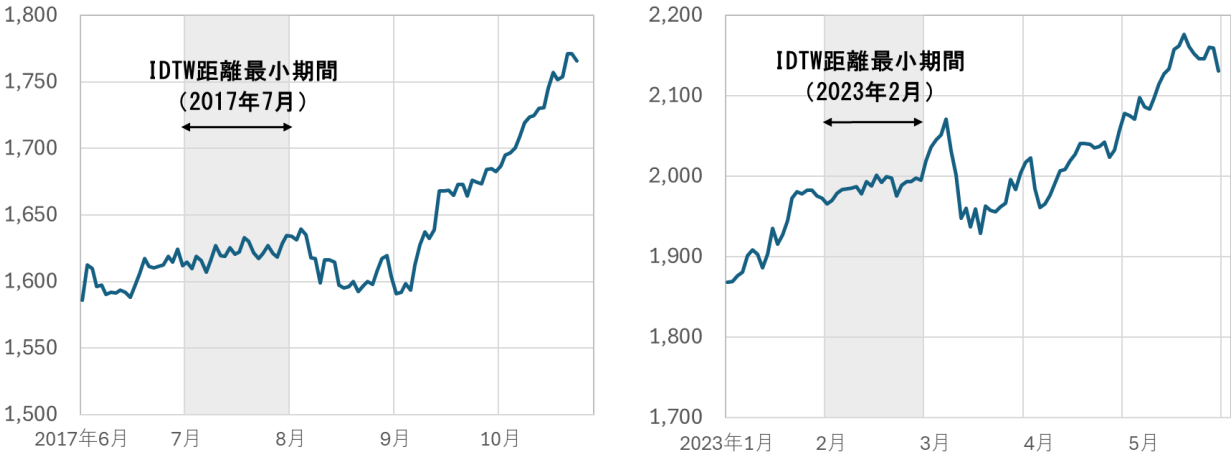
その結果、太字矢印で記された上段と下段の類似度が高い月同士の株価チャートの波形の類似性はもとより、その後約3ヵ月間にわたるチャート推移においても、小幅な下落後に上昇トレンドとなる波形も似たものとなっている。この期間の各月同士の IDTW 距離が小さい(類似度が高い)月同士のその後3ヵ月間となる 2017 年 8 月と 2023 年 3 月、2017 年 9 月と 2023 年 4 月、2017 年 10 月と 2023 年 5 月は、IDTW 距離がそれぞれ 0.2081、0.1234、0.0935 と小さく、類似度が高い傾向が継続⁷している。このケース以外でも同様の傾向が複数確認されており、DTW 距離は過去の類似局面の探索のみならず、先行研究で示された株価予測についても一定の効果がある可能性を確認した。

⁷ [5]では、IDTW に基づく戦略とモメンタム戦略との比較がなされている。両戦略の累積収益率の相関係数は 0.15 であり、IDTW 戦略の方がモメンタム戦略よりもパフォーマンスが上回ることを報告している

図表 5：東証株価指数の月次 IDTW 距離算出の例

	2025年6月	2025年7月	2025年8月
			
2025年8月	IDTW=0.3496	IDTW=0.4314	IDTW=0.5682
2025年9月	IDTW=0.9121	IDTW=0.7151	IDTW=0.5682
2025年10月	IDTW=0.1481	IDTW=0.2040	IDTW=0.5009

図表 6：類似度が高い月同士の東証株価指数のチャート
(2017 年 7 月と 2023 年 2 月)



IV. 終わりに

本稿では、DTW 距離の具体的な算出方法や、相関係数や絶対値距離との比較を通じた特徴について述べてきた。また、東証株価指数データを用いて、同一指数内における過去の類似チャートパターンの探索も行った。

実際の金融市場のデータを用いた分析では、IDTW 距離が絶対値距離よりも小さい値をとる事例が確認され、そのような事例では、IDTW 距離が人間の視覚的な類似性を捉えている可能性について言及した。また、月次ベースでの過去の類似チャートパターンの探索では、類似度の高い局面の発見のみならず、翌月以降も類似度の高い局面が継続する可能性も

示唆され、先行研究でも述べられた株価予測についても一定の効果がある可能性を確認した。

尚、本稿では類似チャートパターンの探索は主に月次ベースで行ったものの、期間は任意に設定が可能である。一方で、全てのパターンの探索には甚大な計算が必要となる為、期間の設定プロセスの明確化や実運用に資する運用戦略への適用について、今後検討を行って参りたい。

(2025 年 10 月 16 日 記)

※本稿中で述べた意見、考察等は、筆者の個人的な見解であり、筆者が所属する組織の公式見解ではない

【参考文献】

- [1] Sakoe, Hiroaki, and Seibi Chiba. "Dynamic programming algorithm optimization for spoken word recognition." *IEEE transactions on acoustics, speech, and signal processing* 26. 1 (2003): 43-49.
- [2] 今村光良, 中川慧, and 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." *人工知能学会全国大会論文集 第 31 回 (2017)*. 一般社団法人 人工知能学会, 2017.
- [3] 今村光良, 中川慧, and 吉田健一. "価格変動パターンによる証券/為替/仮想通貨市場の分析." *電気学会論文誌 C (電子・情報・システム部門誌)* 138. 8 (2018): 992-998.
- [4] 小沢育実. "関連銘柄同定のための時系列データ類似度尺度の提案." *情報処理学会第 78 回全国大会講演論文集* 2016. 1 (2016): 301-302.
- [5] 中川慧, 今村光良, and 吉田健一. "株価変動パターンの類似性を用いた株価予測." *人工知能学会全国大会論文集 第 31 回 (2017)*. 一般社団法人 人工知能学会, 2017.
- [6] 中川慧, 今村光良, and 吉田健一. "価格変動パターンを用いた市場予測 k-Medoids Clustering with Indexing Dynamic Time Warping の株式市場への適用." *電気学会論文誌 C (電子・情報・システム部門誌)* 138. 8 (2018): 986-991.

本資料について

- 本資料は、お客さまに対する情報提供のみを目的としたものであり、弊社が特定の有価証券・取引や運用商品を推奨するものではありません。
- ここに記載されているデータ、意見等は弊社が公に入手可能な情報に基づき作成したのですが、その正確性、完全性、情報や意見の妥当性を保証するものではなく、また、当該データ、意見等を使用した結果についてもなんら保証するものではありません。
- 本資料に記載している見解等は本資料作成時における判断であり、経済環境の変化や相場変動、制度や税制等の変更によって予告なしに内容が変更されることがありますので、予めご了承下さい。
- 弊社はいかなる場合においても、本資料を提供した投資家ならびに直接間接を問わず本資料を当該投資家から受け取った第三者に対し、あらゆる直接的、特別な、または間接的な損害等について、賠償責任を負うものではなく、投資家の弊社に対する損害賠償請求権は明示的に放棄されていることを前提とします。
- 本資料の著作権は三菱 UFJ 信託銀行に属し、その目的を問わず無断で引用または複製することを禁じます。
- 本資料で紹介・引用している金融商品等につき弊社にてご投資いただく際には、各商品等に所定の手数料や諸経費等をご負担いただく場合があります。また、各商品等には相場変動等による損失を生じる恐れや解約に制限がある場合があります。なお、商品毎に手数料等およびリスクは異なりますので、当該商品の契約締結前交付書面や目論見書またはお客さま向け資料をよくお読み下さい。

編集発行：三菱UFJ信託銀行株式会社 アセットマネジメント事業部
東京都港区東新橋1丁目9番1号
お問合せ：shisanunyo-joho_post@tr.mufg.jp