

# 株価変動パターンの類似性を用いた株価予測

## Stock Price Prediction Using Similarity of Stock Price Fluctuation Patterns

中川 慧 \*1\*2      今村 光良 \*1\*3      吉田 健一 \*2

Kei Nakagawa      Mitsuyoshi Imamura      Kenichi Yoshida

\*1日興グローバルラップ株式会社

Nikko Global Wrap Ltd.

\*2筑波大学大学院 ビジネス科学研究科

University of Tsukuba Graduate School of Business Sciences

\*3筑波大学大学院 システム情報工学研究科

University of Tsukuba Graduate School of Systems and Information Engineering

Various methods to predict stock prices have been studied. A typical one is based on the time series analysis. The others are based on machine learning techniques using cross-sectional data as feature values. In the field of finance, “Value”, “Momentum”, etc. have been used as the feature for prediction. In this research, we use the pattern of stock price change as the feature. We employ a dynamic time warping (DTW) method to calculate the similarity of daily based stock price change. Since the level of the stock price differs depending on the measured period, we also developed a standardization method to compensate the difference of price levels. In this paper, we also show the advantage of the proposed method, named as Indexation DTW (IDTW), over conventional 12-month momentum strategy using Japanese stock market data.

## 1. はじめに

株価を予測する際に、現在の相場と似た過去の相場を参考にすることがある。本研究では、このような予測プロセスを定量的に再現し、予測に活用する手法の提案を行う。

具体的にはまず、現在の株価変動が過去のどの時点の株価変動に近いかを定量的に抽出する。抽出に際しては、時系列データ間の類似度を測るために、音声認識の分野等で使用されている Dynamic Time Warping (DTW) 距離 [7] を使用する。更に株価は計測期間 (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ため、適切な計測期間と株価の表現方法を決める必要がある。本研究では、日次の株価変動を月初の値との比で表現した月間の株価変動に対して DTW を適用する Indexation DTW (IDTW) による手法を提案する。

次に、IDTW を用いた実証分析を行い、抽出した過去の株価変動が月毎の株価予測に有効な特徴量であることを報告する。具体的には、IDTW が過去の株価変動を特徴量とするモメンタム戦略よりも投資戦略として有効である事を示す。

## 2. 先行研究

株価を予測するために様々な方法が提案されている。

代表的な手法は時系列解析に基づくものである。時系列解析のモデルとして、条件付き平均モデル、条件付き分散モデルがある。条件付き平均モデルの代表例として AR モデル、MA モデル、ARMA モデルがある。これらは株価の水準あるいは収益率の予測に用いられる。AR モデルは過去の株価の線形結合で将来の株価を予測するモデルである。MA モデルは過去の株価の攪乱項の線形結合で将来の株価を予測するモデルである。ARMA モデルは AR と MA を組み合わせたモデルである。一方で、条件付き分散モデルとして ARCH モデルや、ARCH を

さらに一般化した GARCH モデルが提案されている。これらは株価のボラティリティを予測するために用いられる [6]。

以上の時系列分析手法は、データ自身の過去の変動パターンを線形方程式でモデル化するという意味で恣意性の低い方法である。

一方で、クロスセクショナルなデータを特徴量として使用した、いわゆる機械学習に基づく株価予測も多く存在する。時系列解析と異なり、機械学習による予測は特徴量をうまく選択する必要がある。従来のファイナンス分野で、株価を予測するための特徴量としてはバリュエーションやモメンタム等がある [5]。最近ではアナリストレポートやニュースといったテキストデータを、特徴量として用いる試みが模索されている。例えば、[9] はアナリストレポートを対象にテキストマイニングを行い、アナリストレポートは数値情報とは異なる情報を株式市場に伝えると報告している。[12] は日経新聞の記事を用いて、共起解析、主成分分析および回帰分析の3段階からなる CPR 法により日本株式市場の長期的な動向を予測している。また、[13] は、ロイターニュースのテキスト情報と株価の関連性を Deep Learning を用いて解析し、将来の株価予測ができる可能性を報告している。

本研究では機械学習の枠組みで、上記のようなテキストデータではなく、現在の株価変動に近い過去の株価変動パターンが予測に有効な特徴量であることを示す。過去の株価変動を用いた特徴量としてモメンタムがあるが、終値の差により変化を捉えるモメンタムに対して、提案手法は株価の変化パターンに着目する点が異なる。

## 3. Indexation Dynamic Time Warping (IDTW) による特徴抽出

### 3.1 Dynamic Time Warping

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ユークリッド距離が

連絡先: 中川 慧, 日興グローバルラップ株式会社 (出向元; 三井住友アセットマネジメント株式会社), 〒103-0016 東京都中央区日本橋小網町 9-2

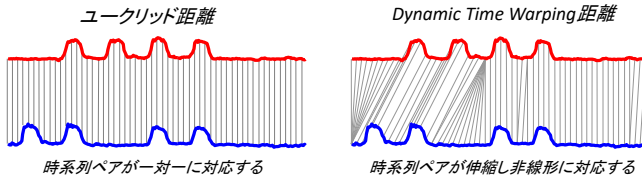


図 1: 時系列データの対応付け

挙げられる。ただし単純な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。加えて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 [8]。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。

DTW 距離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 に、ユークリッド距離と DTW に基づく類似度を算出する際の、時系列データペアの対応例を示す。

DTW 距離の問題点は、計算量が  $O(M \times N)$  となるため膨大である事であるが、近年の高速な計算機を利用すれば十分計算可能である。すなわち、 $M \times N$  の大きさの類似度行列を作り、動的計画法により  $x, y$  内の比較する点の組を示すワーピングパスを発見し、距離を計算する。ここで  $M, N$  は時系列データ  $x, y$  の長さを表す。具体的には以下の手順で計算を行う。

#### DTW 距離算出アルゴリズム

1.  $M \times N$  の大きさの類似度行列  $D$  を作る
2.  $D_{0,0} = 0$ , それ以外の類似度行列の値に  $\infty$  を代入し初期化する
3. 以下の式により  $i = 1, 2, \dots, M, j = 1, 2, \dots, N$  の全ての組み合わせの  $D_{i,j}$  を計算する
 
$$D_{i,j} = d(x_i, y_j) + \min(D_{i,j-1}, D_{i-1,j}, D_{i-1,j-1})$$
 $d$  は距離測度であり、マンハッタン距離  $d(x_i, y_j) = |x_i - y_j|$  を使用する
4. 類似度行列  $D$  の  $D_{M-1, N-1}$  を時系列データ  $x, y$  の DTW 距離とする

### 3.2 提案手法 - Indexation DTW (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。こうした株価の季節性については [2] で検証されており、実際に世界の株式市場で "Sell in May" 効果が確認されている。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。本研究では、この点を考慮して、(1) について、月間の日次の株価変動を距離計算の対象期間として DTW 距離を計算した。

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

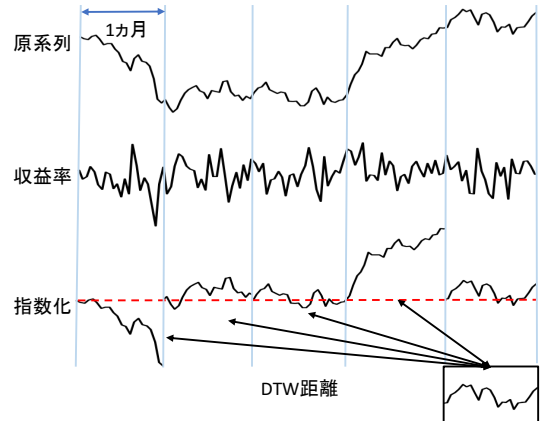


図 2: IDTW による株価変動の比較

て比較する方が自然である。この点を考慮して、(2) については、月初の値を 1 とし、月内の日毎の株価推移を、月初との比を使って表した。

以上、各月毎に抽出した日毎の株価推移のデータ系列間で DTW 距離を計算し、距離が近いほど株価変動が似ていると判断した (図 2)。

### 3.3 提案手法の特徴

前述のように株価の変動パターンを用いた研究として収益率、ボラティリティ及び出来高を使って類似度を計算した [3] や [11, 14] 等があるが、本研究と比して決定すべきパラメータが多い、結果が安定していない、等検討課題が多く残されている。

また、次章で述べるような高い収益率と株価の動向予測の正解率が提案手法の特徴であるが、これらの特徴が及ぼす上記の手順 (月内の日次の株価変動を DTW 距離計算の単位とした事と、月初の値を 1 とし月内の株価推移の比を用いた事) の影響については [10] で議論し、本論文では結果の概要を報告する。

## 4. 実証分析

### 4.1 分析 1: TOPIX を用いた分析

IDTW によって抽出された過去の株価変動が予測に有効な特徴量であることを次の通り確認する。分析にあたっては TOPIX 指数の日次の配当込指数を使用する。指数データは情報端末の Bloomberg から取得した。データ期間は 1989 年 1 月から 2016 年 12 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 2016 年 12 月までの 10 年間とした。分析にあたって検証期間と似た変動パターンをした月の株価変動を、IDTW を用いて学習期間から探索する。その後、株価変動が似ていると判断された月の翌月の株価を予測値として利用する。これを 1 か月毎に繰り返し予測を行った。ただし学習期間の始点は 1989 年 1 月に固定とした。具体的な手順は以下の通り。

1.  $t$  月初の株価を 1 に標準化し、各月ごとの月間の価格変動を作成する (指数化)
2.  $t-1$  月までの月間の価格変動と  $t$  月の月間の価格変動の DTW 距離を計算する

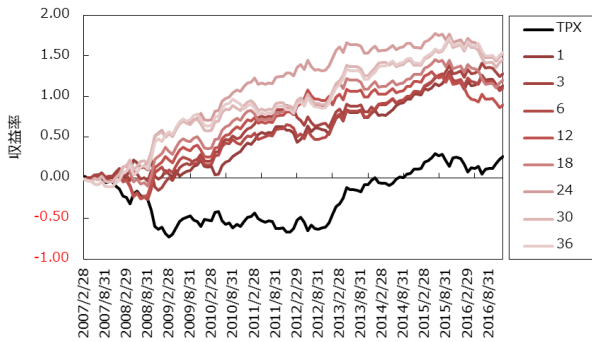


図 3: 各月数毎の累積収益率の推移

	TPX	1	3	6	12
年率リターン [%]	2.64	11.33	12.83	11.14	9.03
年率リスク [%]	19.39	19.12	19.04	19.13	19.23
シャープレシオ [倍]	0.14	0.59	0.67	0.58	0.47
合計収益率 [%]	26.43	113.28	128.27	111.41	90.30
正答率 [%]	55.00	60.83	64.17	60.83	58.33

	TPX	18	24	30	36
年率リターン [%]	2.64	12.03	15.28	14.56	15.44
年率リスク [%]	19.39	19.09	18.89	18.94	18.88
シャープレシオ [倍]	0.14	0.63	0.81	0.77	0.82
合計収益率 [%]	26.43	120.29	152.80	145.59	154.41
正答率 [%]	55.00	60.83	63.33	60.83	63.33

表 1: 各月数毎の結果

- DTW 距離に基づき過去の株価変動パターンが近い月を順に  $n$  個選ぶ (パラメータ)
- 選ばれた  $n$  個の月の翌月リターンの平均値がプラスなら  $t$  月末の価格で買い、マイナスなら売りとして  $t+1$  月の収益を計算する
- $t = t + 1$  に進めて 1. に戻る

手順 3. の平均を計算する月数毎の累積収益率と TOPIX の推移が図 3 である。検証期間である 2007 年 1 月から 2016 年 12 月まで TOPIX の累積収益率はほぼ横這いであるのに対して、提案手法で抽出した場合、月数  $1 \leq n \leq 36$  の全てにおいて IDTW は TOPIX を上回った。

表 1 が各月数および TOPIX の年率換算後のリターン、リスク、シャープレシオ、合計収益率、正答率のサマリーである。表に示した通り、特徴量として利用する  $n$  個の月数に依存せず、累積収益率は 100%程度 TOPIX 指数に対して上回り、正答率もおおむね 60%と良好な結果であった。これは分析期間は異なるものの、本研究と同じく 10 年間の TOPIX の月次予測を、テキスト情報をもとにおこなった [12] と同程度の予測力である。

各年度ごとの正答率の一覧が表 2 である。各月数および TOPIX との相関はすべて弱い逆相関 (0 から -0.2 程度) であった。

#### 4.2 分析 2: モメンタムとの比較

本研究で提案した IDTW による過去の株価変動パターンを用いた株価予測と、ファイナンスの分野での過去の株価変動を用いた特徴量であるモメンタムとの比較を図 4、表 3、4 に示す。

	TPX	1	3	6	12
2007 年	45.45	63.64	72.73	45.45	36.36
2008 年	25.00	50.00	41.67	58.33	66.67
2009 年	58.33	66.67	75.00	50.00	50.00
2010 年	50.00	83.33	75.00	66.67	83.33
2011 年	41.67	58.33	75.00	66.67	58.33
2012 年	66.67	50.00	41.67	58.33	50.00
2013 年	66.67	66.67	58.33	75.00	75.00
2014 年	58.33	75.00	75.00	75.00	66.67
2015 年	66.67	50.00	75.00	75.00	58.33
2016 年	66.67	50.00	50.00	33.33	33.33

	TPX	18	24	30	36
2007 年	45.45	45.45	36.36	45.45	36.36
2008 年	25.00	66.67	91.67	83.33	75.00
2009 年	58.33	50.00	58.33	50.00	66.67
2010 年	50.00	83.33	83.33	66.67	66.67
2011 年	41.67	58.33	66.67	50.00	50.00
2012 年	66.67	58.33	58.33	58.33	58.33
2013 年	66.67	66.67	66.67	75.00	75.00
2014 年	58.33	66.67	58.33	66.67	75.00
2015 年	66.67	66.67	66.67	75.00	75.00
2016 年	66.67	41.67	41.67	33.33	50.00

表 2: 各年度毎の正答率 [%]

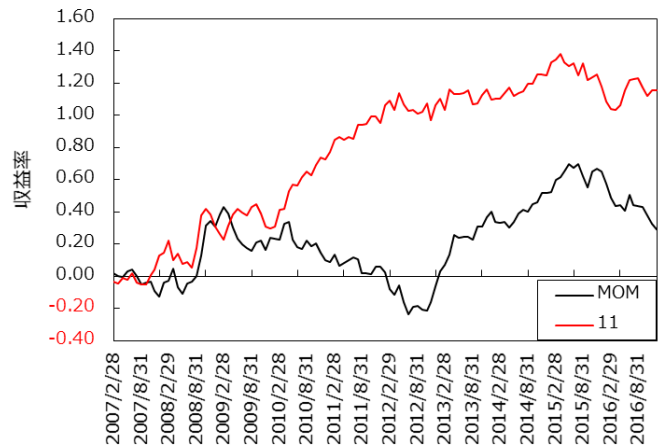


図 4: モメンタム戦略との累積収益率の推移の比較

モメンタムの計測は [1] にならない、過去 12 か月の累積リターンのうち直近 1 か月を除いた 11 か月の累積リターン値で計測している。モメンタムがプラスなら月末の価格で買い、マイナスなら売りとしてモメンタム戦略の収益を計算した。過去の株価変動パターン数はモメンタムと同じく、最も類似度の高い月を除いた 11 個の平均値を使用する。その他の条件は分析 1 と同様である。

両戦略の累積収益率の推移が図 4、結果のサマリーが表 3、各年度ごとの正答率の一覧が表 4 である。両戦略の相関係数は 0.15 であった。モメンタム戦略に対して、提案手法の累積収益率は 100%程上回った。

#### 5. まとめ

本研究では IDTW を用いて抽出した現在の株価変動に近い、過去の株価変動パターンが予測に有効な特徴量であることを実証した。分析 1: TOPIX を用いた分析では、以下のような良好な結果を得た。

- おおむね 60%近い正答率

	MOM	11
年率リターン [%]	2.92	11.68
年率リスク [%]	19.08	19.21
シャープレシオ [倍]	0.15	0.61
合計収益率 [%]	28.98	115.83
正答率 [%]	51.26	60.50

表 3: モメンタム戦略との比較

	MOM	11
2007 年	45.45	54.55
2008 年	75.00	58.33
2009 年	41.67	41.67
2010 年	33.33	75.00
2011 年	41.67	66.67
2012 年	41.67	58.33
2013 年	83.33	66.67
2014 年	66.67	66.67
2015 年	66.67	58.33
2016 年	16.67	58.33

表 4: 各年度毎の正答率の比較 [%]

- 合計収益率は 100%程度 TOPIX 指数に対して上回った
- 特徴量として利用する  $n$  箇の月数に依存しない

この結果は当月の変動が過去のある月の変動と似ていた場合、その先も似た変動をする可能性が高いことを示している。すなわちファイナンスの分野で一般に受け入れられている効率的市場仮説が示唆する弱度の効率性<sup>\*1</sup>が TOPIX 指数では成立していない可能性を示唆する。

一方で過去の株価変動を用いた特徴量としてモメンタムがあるが、終値の差により変化を捉えるモメンタムに対して、提案手法は株価の変化パターンに着目する点異なる。分析 2: モメンタムとの比較では、一般に日本市場では有効性が認められないモメンタムに対して、合計収益率、正答率ともに大幅に上回った。

今後の展望として、本研究では日本市場、とくに株式市場のみを対象に分析を行ったが、外国株式市場や為替など金融市場の様々なマーケットに対して本研究の提案手法の有効性を検証することが挙げられる。

## 参考文献

- [1] Asness, C. S., Moskowitz, T. J., and Pedersen, L. H.: Value and momentum everywhere. The Journal of Finance, 68(3), pp. 929-985. (2013).
- [2] Bouman, S., and Jacobsen, B.: "The Halloween indicator," Sell in May and go away": Another puzzle. The American Economic Review, 92(5), 1618-1635. (2002).
- [3] Coelho, M. S.: Patterns in financial markets: Dynamic time warping. PhD Thesis NSBE-UNL. (2012).
- [4] Fama, E. F.: Efficient capital markets: A review of theory and empirical work. The journal of Finance, 25(2), 383-417. (1970).

- [5] Fama, E. F. and K. R. French.: Common risk factors in the returns on stock and bonds. Journal of Financial Economics, Vol. 33, pp. 356. (1993).
- [6] Hamilton, J. D.: Time series analysis (Vol. 2). Princeton: Princeton university press. (1994).
- [7] Itakura, F.: Minimum prediction residual principle applied to speech recognition. IEEE Trans Acoustics Speech Signal Process. (1975).
- [8] Keogh E. J. and M. J. Pazzani: Scaling up Dynamic Time Warping for Datamining Application. Proc. Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 285-289. (2000).
- [9] Takahashi, S., Takahashi, H., Tsuda, K. and Terano, T.: Analyzing Asset Management Knowledge from Analyst's Reports through Text Mining. International IPSI-2004. (2004).
- [10] 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. 第 31 回人工知能学会全国大会. (2017).
- [11] 小沢育実, 関和広: 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. (2016).
- [12] 藏本貴久, 和泉潔, 吉村忍, 石田智也, 中嶋啓浩, 松井藤五郎, 中川裕志: 新聞記事のテキストマイニングによる長期市場動向の分析. 人工知能学会論文誌, 28(3), pp. 291-296. (2013).
- [13] 五島圭一, 高橋大志, 寺野隆雄: ニュースのテキスト情報から株価を予測する. 第 29 回人工知能学会全国大会. (2015).
- [14] 白浜公章: 経済データに対する値と形状に基づく時系列類似尺度の比較. 神戸大学紀要. (2016).

\*1 過去の価格の持っている全ての情報は現在の価格に十分に反映されており、過去の時系列データを分析することで市場平均を凌駕するような利益を上げられないような状態を指す [4].