

DTW を用いた株価指数変動予測モデルの分析とその運用

長倉大輔研究会 経済学部 3 年 吉田光佑・森谷太智

要旨

本研究は、過去の株価変動パターンを参考にし、翌月の株価の騰落を予想する機械学習モデルをテーマにしている。通常は恣意性が含まれるテクニカル分析を機械的に行い、予測精度の向上への試みに興味を持ち、本研究に至る。今回は、音声認識の分野で使われる動的時間収縮法 DTW(Dynamic Time Wrapping) を用いて類似を計算し、過去の変動パターンとの類似度の計測を行ない、k-NN と組み合わせて、翌月の株価の騰落を予想した。12 ヶ月の変動を学習して、対象とする 12 ヶ月の変動から翌月の株価の騰落を予想するモデルを組んだ。

目次

1 概要

1.1 研究の背景と目的

1.2 先行研究と研究の方向性

2 モデル説明

2.1 DTW

2.2 IDTW

2.3 k-NN

3 実証分析

3.1 分析手法

3.2 データ

3.3 分析結果

3.4 追加分析

4 結論

4.1 まとめ

4.2 課題と今後の展望

5 参考文献

1 概要

1.1 研究の背景と目的

現在「貯蓄から投資へ」という政策のもと、日本社会で個人投資家による投資が活発化してきている。株価投資を行う際には、主に株価変動を予測することになるが、分析には2種類ある。テクニカル分析とファンダメンタル分析がある。テクニカル分析とは、過去の値動きをチャートで表して、そこからトレンドやパターンなどを把握し、予測するものである。ファンダメンタル分析とは、相場の大きな方向性を掴むために景気動向、金融政策、財政政策等の変化が市場にどのような影響を及ぼすのかを分析する。テクニカル分析は、従来の方法では恣意性が強いと言われている。この分析を機械的に行うことで予測精度を上げることができないかと考えて本研究に至る。本研究では、投資における予想の機械化、及び、予想モデルを用いた運用方法を提案することが目的である。

1.2 先行研究と研究の方向性

類似度を DTW で測定する先行研究として、中川(2019)の研究がある。これは、株価の日次データによって構成される月毎の過去の変動を学習して、ある月の変動から翌月の変動を予測するというものである。類似度を表す指標として DTW とそれから派生した DDTW、IDTW の3つを用いた。DDTW は時系列データの変分に対して DTW を適用する Derivative DTW である。IDTW とは指数化し比較しやすくした Indexing DTW

を指す。また、類似度が近いものを選択し騰落予想する時のアルゴリズムとして k -NN と k^* -NN の二つを用いていた。 k^* -NN は、 k -NN のハイパーパラメータを過去のデータから決めるものだ。結果としては、指標としては IDTW、アルゴリズムとしては k^* -NN の精度が高く、IDTW と k^* -NN の組み合わせが最も精度が良かったとしている。

本研究では、この先行研究を受けて、IDTW と k -NN の組み合わせを用いてテクニカル分析を行おうと考えた。月次データによって構成される年毎の過去の変動を学習して、ある 1 年間(12 ヶ月)の変動から翌月の変動を予測するという先行研究よりも長期の目線で分析することにした。

2 モデル説明

2.1 DTW

DTW は中川(2019)による定義を利用している。DTW は時系列データ間の類似度を計算するための手法である。単純な類似度計測手法としてユークリッド距離があるが、時間軸方向に少しでもずれがあると類似度が減少する。図 1 でユークリッド距離と DTW 法に基づく時系列データ類似度比較の例を示す。アルゴリズムは図 2 のようになっている。

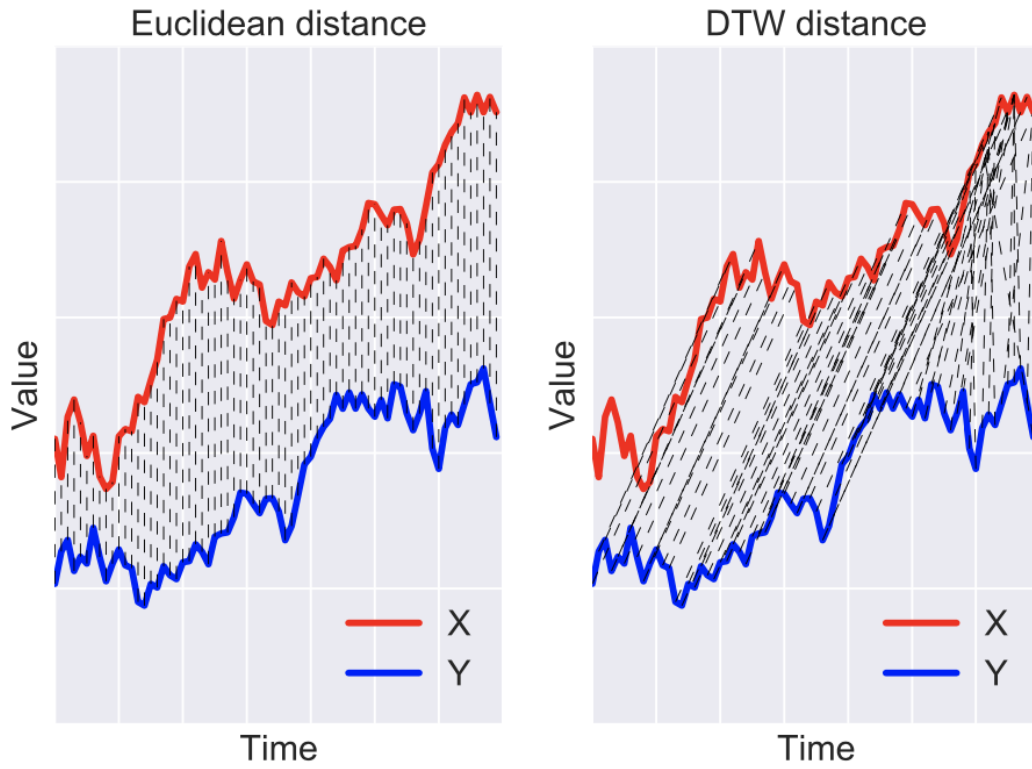


図 1 : ユークリッド距離と DTW の比較の例

(中川(2019)より抜粋)

Algorithm 1 DTW distance

```

1: procedure DTW( $x, y$ )
    ▷ Initialize matrix D

2:   Var  $D[N + 1, M + 1]$ 
3:    $D[1, 1] = 0$ 
4:   for  $i = 2$  to  $N + 1$  do
5:     for  $j = 2$  to  $M + 1$  do
6:        $D[i, j] = \infty$ 
7:     end for
8:   end for

    ▷ Calculate DTW distance

9:   for  $i = 2$  to  $N + 1$  do
10:    for  $j = 2$  to  $M + 1$  do
11:       $D[i, j] = d(x[i - 1], y[j - 1])$ 
         $+ \min(D[i, j - 1], D[i - 1, j], D[i - 1, j - 1])$ 
12:    end for
13:  end for
14:  return  $D[N + 1, M + 1]$ 
15: end procedure

```

図 2 : DTW のアルゴリズム(中川(2019)より抜粋)

2.2 IDTW

変動の最初の値を基準として指数化を行い、それによって表される系列をもとに DTW を算出する。これにより、比較する 2 つの変動の最初の値が大きく離れていても距離に影響が出ず、類似度がより正確に算出される。

2.3 k-NN

分類問題において、 k 個の最近傍の平均を取って予測を行う。この際に距離を定義する必要があり、今回の予測では DTW と IDTW を用いる。

3 実証分析

3.1 分析手法

今回の分析の概要は、過去の変動を学習し、DTW と IDTW を用いて対象の変動と学習した変動との類似度を計算する。k-NN を用いて類似度が高い過去の変動を k 個選び、その翌月のリターンを k 個の平均を取る。その値が正ならば翌月の株価を上昇予想し、負ならば下降予想する。騰落予想の正答率とリターンを計算する。上昇予測の際には翌月 TOPIX の値と参照する変動の最後の月の TOPIX の値の差分をリターンとする。下落予想の際には空売りを想定し、翌月 TOPIX の値と参照する変動の最後の月の TOPIX の値の差分をリターンとして計算し累積していく。先行研究の中川(2019)では月の変動を学習し、1 月分の変動を対象として翌月の騰落を予測していた。本研究では 12 ヶ月

分のデータを一つの変動として学習し、12 ヶ月分の変動を対象として翌月の騰落を予測する。評価指標としては、騰落予想の正答率とリターンを計算する。上昇予測の際には翌月 TOPIX の値と参照する変動の最後の月の TOPIX の値の差分をリターンとする。下落予想の際には空売りを想定し、翌月 TOPIX の値と参照する変動の最後の月の TOPIX の値の差分をリターンとして計算し累積していく。

3.2 データ

データ全体は 1955 年 1 月から 2021 年の 12 月の月次の TOPIX データである。訓練期間を 1955 年 1 月から 2013 年 12 月とし、2014 年 1 月から 2021 年 12 月までをテスト期間としている。学習データは 12 ヶ月分を変動として学習するが、学習する変動は 1 ヶ月ごとずらして学習している。予測する際に参照する変動も 1 ヶ月ずつずらして予想をする。

3.3 分析結果

分析結果の評価軸として、騰落予想の精度、およびリターンで評価する。本研究では k-NN のハイパーパラメータの値である k を 1 から 5 に設定を変更し、最も正答率の高い k の値が 3 である場合の結果を分析している。DTW を用いた分析結果は表 1 のようになった。正答率は予想する月の数のうち何回予想が的中したのかということである。リターンは前述した方法で計算している。利益率は、投資開始日から最終日まで保有し

ていた時のリターン、つまり、投資終了日の TOPIX の値と投資開始日の TOPIX の値の差を分母にし、累積リターンを分子にしたときの百分率の値である。結果を見ると、全体の正答率は 60%を超え、利益率は 140.3%となっている。

正答率			リターン	
上昇予測(%)	下落予測(%)	全体(%)	累積リターン	利益率(%)
61.2	60.0	60.7	761.6	140.3

表 1：分析手法に DTW を用いた際の騰落予想の正答率と運用のリターン

空売りを行った結果、下落予測が失敗した場合には、上昇予測の失敗と同様に損失を生み出すことになるため、空売りをしないという選択を続けた方が良いのではないかという疑問が出てくる。そこで、空売りの効果を検証するため、下落予想時には空売りではなく、株式投資をしないという選択を行った場合との比較を行った。その方針で分析した結果が表 2 である。累積リターンが減少していることから、空売りというリスクのある運用方法にある程度の効果があることが分かる。

正答率			リターン	
上昇予測(%)	下落予測(%)	全体(%)	累積リターン	利益率(%)
61.2	60.0	60.7	663.5	122.2

表 2：下落予想時に投資をしないという運用のリターン

また、IDTW を用いた分析結果は表 3 のようになった。DTW の結果と比較すると、全体の正答率は 55%、下落予測では 50%を下回るものの、利益率は 168.7%と高水準であ

る。また、下落予想時に投資をしないというリスクを抑えた運用を行った場合のリターンを上回っていることから、上昇予測精度の改善が相殺している、若しくは、大きな下落の予測に関しては外さないのではないかと考えられる。

正答率			リターン(空売り)		リターン(投資なし)	
上昇予測(%)	下落予測(%)	全体(%)	累積リターン	利益率(%)	累積リターン	利益率(%)
63.3	42.9	54.8	915.9	168.7	740.6	136.4

表 3：分析手法に IDTW を用いた際の騰落予想の正答率と運用のリターン

ここまで、リターンの予測を行うにあたって、最近傍の単純平均を用いていたが、過去のデータとの類似度が高いケースの値を優先的に反映させれば、より適切に類似度を用いた分析ができるはずである。そこで、騰落予測の計算に加重平均を用いて予測値を求める分析を行った。下落予測時の行動は空売りで、近傍数ごとの重みは表 4 のとおりとした。計算の結果、モデルが DTW であるときの最適な近傍数は 2、IDTW であるときの最適な近傍数は 4 であることが分かった。分析結果は表 5 のとおりである。IDTW の予測精度が若干改善され、また、両モデルについての利益率が 160%を上回るという結果が得られた。

近傍数	重み(左から類似度が高いもの)				
1	1				
2	0.7	0.3			
3	0.5	0.3	0.2		
4	0.4	0.3	0.2	0.1	
5	0.3	0.3	0.2	0.1	0.1

表 4：近傍数ごとの重みの設定

モデル	近傍数	正答率			リターン	
		上昇予測(%)	下落予測(%)	全体(%)	累積リターン	利益率(%)
DTW	2	55.1	65.7	59.5	992.2	182.8
IDTW	4	61.2	48.6	56.0	880.9	162.3

表 5：騰落予測に加重平均を導入し運用した場合の分析結果

3.4 追加分析

ここまで、最適な近傍数を決定するために、全ての近傍数の場合を計算し、最も結果が良い近傍数を設定していた。しかし、株価が不規則に変化中、設定された近傍数が常に最適であるとは限らず、時間が経つにつれて最適な近傍数は変化し続けるはずである。そこで、直近の予測精度によって最適だと思われる近傍数を変化させ、予測をするという分析を行った。近傍数の範囲を 1 から 5 に設定し、直近の 12 か月、36 か月の予測精度が最も良かった近傍数の値を利用、下落予測時の行動は空売りとし、予測値の計算については先述の加重平均を用いてそれぞれ分析を行う。分析結果は表 6 のようになった。

モデル	参照期間	正答率			リターン	
		上昇予測(%)	下落予測(%)	全体(%)	累積リターン	利益率(%)
DTW	12か月	51.0	77.1	61.9	1137.9	209.6
	36か月	44.9	82.9	60.7	1093.8	201.5
IDTW	12か月	63.3	51.4	58.3	1152.8	212.3
	36か月	65.3	51.4	59.5	1243.4	229.0

表 6：利用する近傍数を変化させて運用した場合の分析結果

モデルに DTW を用いた際には、上昇予測の精度が落ちたものの、下落予測については大幅に精度が良くなり、利益率も 200%を上回っている。IDTW についても、上昇予測精度はそのままに、下落予測精度が改善され、リターンが 210%を上回る結果となった。

4 結論

4.1 まとめ

今回の研究では、12 ヶ月の変動から翌月の株価の騰落を予想するモデルを組んだ。このモデルに基づいて株を運用すると、投資開始日から保有した場合と比較して 150%程度の利益が、近傍数を変化させた場合には 200%程度の利益が出るという結果となった。

近傍数を変化させる手法による騰落予想の精度はおよそ 60%程度であるが、上昇予測と下落予測のどちらが良いかについてはモデルによって分かれる。本研究の先行研究である中川(2019)の日次データを月ごとに区切って予想する手法と比較して、正答率はモデルが IDTW である場合に大きな変化は見られず、DTW である場合には改善されるという結果が得られた。また、いずれの手法を取った場合でも、利益率は 100%を上回り、

投資開始日から株を保有していた場合よりも利益をあげることができるという結果となった。また、これらモデルに基づいて投資を行う際に、下落予想時には投資しないというリスクが小さい運用方法についても同時に分析を行ったが、空売り時と比較して利益率が落ちており、空売りの効果が表れていると読み取れる。

4.2 課題と今後の展望

本研究の課題としては、2点ある。1点目は、騰落予測に加重平均を用いるケースにおいて、ハイパーパラメータである重みを恣意的に決定してしまったことである。先行研究の中川(2019)では、 k^* -NN のハイパーパラメータである近傍数 k と重み λ を予想の度に決定するアルゴリズムを用いて予測精度を向上させている。ハイパーパラメータのチューニングを行えば、より精度の高いモデルを構築することができると感じた。2点目は、1年という中期のトレンドを訓練し、予測を行うのが1か月後である、短期トレンドという設定である。

今後の展望としては、日時データを用いた研究を行うことや、変動を区切る幅を変更しての分析、騰落予測に k -NN 以外のモデルで行うことを考えている。

5 参考文献

<https://info.monex.co.jp/technical-analysis/column/001.html>

<https://www.ibm.com/jp-ja/topics/knn>

価格変動パターンを用いた株価予想手法の実証研究

2019 年 中川慧 筑波大学大学院 ビジネス科学研究科 企業科学専攻