

IDTW を用いた変化点検知とメタラベリングを用いた GBDT による株式変動予測

東京工業大学 *田崎誠 TASAKI Makoto
01405430 東京工業大学 中田和秀 NAKATA Kazuhide

1. はじめに

近年の日本において、個人の資産運用の必要性が高まっている。特に2019年には「老後2000万円問題」が取り沙汰され、国民自身で老後に向けた資産運用を行う必要性が認知されてきている。さらに2024年からNISAの恒久化及び年間投資枠が年360万円に拡充されるなど、個人の資産運用を後押しする政策も行われている。しかし一方で既存の単純な投資戦略を使用するだけでは市場リターンを上回ることができなくなっていることから、個人の資産運用は難しくなっている。

そこで本研究では、トレンド変化点検知と機械学習を組み合わせた手法を提案する。この提案手法は月単位かつ長期的なトレンドに基づいて取引をするため、ポジションの変更が少なく済むことから時間的・金銭的コストがかかりにくく、個人の資産運用に使用しやすいものである。提案手法では既存のモメンタム戦略と変化点検知を改良した新しいトレンド変化点検知手法による予測を行う。さらにその予測を改善するための新たな工夫を施した機械学習を用いた予測方法も提案する。これらの手法によって月次予測で既存手法を超える高い予測精度と、予測に基づいた高リターン・低リスクな取引を実現している。

2. 関連研究

変化点検知、DTW、機械学習による株式変動予測と本研究との関係について述べる。

変化点検知の主要な手法の1つにICSSアルゴリズムアプローチがある。この手法は[1]によって提案され、CUSUM検定とICSSアルゴリズムを用いた変化点検知手法が提案されている。また最新手法として、クラスカル=ウォリス検定を用いたKW-ICSS[2]も提案されている。しかし既存の変化点検知手法には時系列の形状まで考慮したものはない。

DTWは時系列データ間の類似度を測る手法である。ユークリッド距離や相関係数よりも直感的な類似度になることが知られている。また長さの異なる時系列データの類似度も測ることが可能である。金融時系列データの類似度をより投資家の認識に近くなるように改良した手法にIDTW[3]がある。本研究ではこのIDTWを使用した変化点検知手法であるIDTW-CPDを提案する。

機械学習による株式変動予測にはニューラルネットワーク系か決定木系の手法が用いられることが多い。本研究ではデータ数が少ないため決定木系の手法を用いる。[4]はLightGBMを用いた株式変動予測を提案している。機械学習による株式変動予測では日単位や分単位のような頻度の高い予測は多く研究されているが、月次予測に関してはそれら程多く行われていない。

3. 提案手法

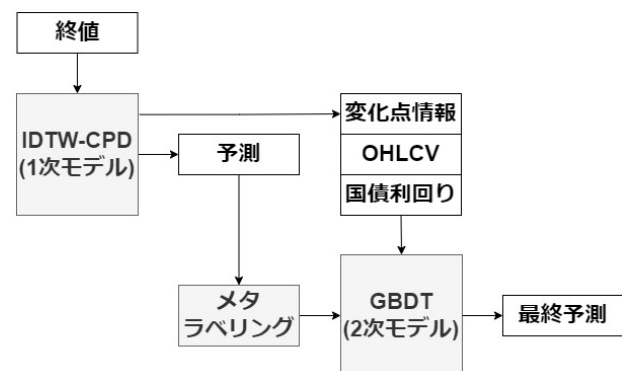


図1: 提案手法の全体像

図1は提案手法の全体像である。提案手法の1つ目がIDTW-CPDを用いて予測を出力する部分であり、2つ目がIDTW-CPDの予測結果をメタラベリングし、GBDTを用いてその予測を修正した最終的な予測結果を出力する部分である。

提案手法の1つ目の部分はIDTWを用いたトレンド変化点検知手法である。図2は図の上側が終

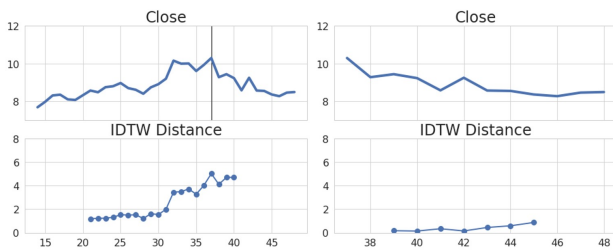


図 2: IDTW を用いた変化点検知

値で、下側がその時点で時系列を分割し、その分割された2つの時系列データ間でIDTWを測った値である。左の図のようにトレンド変化点が存在する時系列データであれば、変化点の時点でIDTWは大きくなる。しかし右の図のようにトレンド変化点がない場合はIDTWは大きくなる。この特性を利用してトレンド変化点を検知するのが、IDTW-CPDである。この手法によって既存の変化点検知の形状を考慮できないという問題点を解決できる。

しかしIDTW-CPDの予測にも、変化点の検知には数か月経ってからでないと検知できないという問題が存在する。提案手法の2つ目の部分ではその問題点を解決するため、IDTW-CPDの予測が正解するかをメタラベリングしてから、LightGBMを用いてIDTW-CPDの予測の正解確率を推定し、その確率に基づいてIDTW-CPDの予測を修正する。さらに特徴量に景気の先行指標となる国債利回りに関する情報を入れることで、LightGBMが景気の変化とトレンドの変化の関係を学習できるようにする。こうすることで、IDTW-CPDでは予測が外れやすい変化点直後の予測を修正でき、予測精度を改善させることができる。

4. 実験

データはS&P500の1960年1月から2022年9月までのものと、米国国債利回りの1962年1月から2022年9月までのものを使用している。実験はウォークフォワード法を用いて行い、1962年から1990年を最初の学習期間、1990年から2000年までを検証期間、2000年から2005年までをテスト期間とし、その後は5年ずつ期間をずらしながら予測を行っていき、最終的なテスト期間は2000年1月から2022年9月までとなる。この22年9か月のテスト期間での性能を既存手法と比較する。

表 1: パフォーマンスの比較

手法	年率	年換算	精度
	リターン	ボラティリティ	
Buy&Hold	4.44%	15.37%	61.76%
時系列モメンタム [5]	3.11%	15.41%	57.35%
IDTW+k*-NN[3]	6.03%	15.33%	61.76%
KW-ICSS[2]	8.23%	15.26%	59.34%
IDTW-CPD	8.48%	15.23%	62.87%
IDTW-CPD&LightGBM メタラベリングなし	5.21%	15.35%	61.40%
IDTW-CPD&LightGBM メタラベリングあり	11.46%	15.09%	64.71%

表1がバックテストによって得られた各手法ごとの性能である。IDTW-CPD&LightGBMメタラベリングありが本研究の提案手法であり、年率リターン、年換算ボラティリティ、予測精度において既存の手法をすべて上回っていることがわかる。また、IDTW-CPDのみを使用した場合でも、既存の変化点検知手法を含めた既存の手法よりも優れた結果を得られている。

参考文献

- [1] C. Inclán and G. C. Tiao. Use of cumulative sums of squares for retrospective detection of changes of variance. *J. Amer. Stat. Assoc.*, vol. 89, pp. 913–923, 1994.
- [2] M. Lee K. Kim, J. H. Park and J. W. Song. Unsupervised change point detection and trend prediction for financial time-series using a new cusum-based approach. *IEEE Access*, Vol. 10, pp. 34690–34705, 2022.
- [3] M. Imamura K. Nakagawa and K. Yoshida. Stock price prediction with fluctuation patterns using indexing dynamic time warping and k*-nearest neighbors. *JSAI-isAI 2017: New Frontiers in Artificial Intelligence*, pp. 97–111, 2017.
- [4] M. Shangchen. Predicting the sp500 index trend based on gbdt and lightgbm methods. *International Conference on Energy Big Data and Low-carbon Development Management*, Vol. 214, 2020.
- [5] Y. H. Ooi T. J. Moskowitz and L. H. Pedersen. Time series momentum. *Journal of Financial Economics*, Vol. 104, Issue 2, pp. 228–250, 2012.