

高頻度注文情報を用いたデータの層化と 多段階事前学習による株価動向予測

Stock Price Prediction Using Limit Order Book Data with Data Stratification and Multi-Phase Pre-training

松原冬樹^{1*} 和泉潔¹ 坂地泰紀¹
Fuyuki Matsubara¹ Kiyoshi Izumi¹ Hiroki Sakaji¹

¹ 東京大学大学院工学系研究科

¹ Graduate School of Engineering, The University of Tokyo

Abstract: Predicting the movement of stock price is an important issue for market participants. Recently, there have been many attempts applying machine learning techniques in financial time series prediction. However, overfitting presents a huge challenge when machine learning approaches are used in financial time series prediction. In this paper, we propose a stock price prediction method utilizing limit order book data from stocks other than target stocks by stratifying the data and holding a multi-phase pre-training considering market liquidity. Experimental results shows that the proposed approach enhances prediction performance.

1 はじめに

金融商品の価格は市場参加者の行動によって変化し続けている。市場参加者は金融商品を購入価格より高い価格で売却、もしくは売却価格より低い価格で購入することによって収益を上げることができる。金融商品の将来における価格変動を過去の情報を用いることによって事前に予測することが可能であるならば、予測を基に取引をしリスクに見合う以上の収益を追求することができると考えられる。経済学、物理学などの学問分野の学者や金融市場に身を置く実務家によって「金融市場における将来の価格予測が可能であるか」という問いに関して数多くの理論研究、実証分析が行われてきた。

[Fama 70] は、「いかなる時も利用可能なすべての情報が証券価格に完全に反映される」とする効率的市場仮説を提唱した。実際に取引が行われる取引価格に対して、すべての市場参加者が情報を共有し情報の非対称性がない場合の均衡価格を本源的価値とすると、本源的価値はすべての市場参加者が価格付けの方法に関して合意している場合の証券価格となる。

一方で、効率的市場仮説への反証をあげる研究、事例も存在する [Cont 01] は経験的に知られている市場の定型化された事実 (stylized facts) として証券のリターンが正規分布ではなくべき乗分布に従うことやボラティ

リティ(資産のリターン標準偏差)に正の自己相関が存在するボラティリティクラスタリングといった現象等について言及している。これらの市場の経験則は効率的市場仮説とは相容れない性質である。また、2010年5月6日には米国市場においてダウ工業株30種平均が数分間で9% (約1000ドル) も下落したフラッシュクラッシュは、市場が効率的であるとする仮説とは非整合的な現象であると考えられる。

これらのことから、市場は必ずしも効率的であるとは言えないと考えられる。金融商品の価格変動に一定の傾向を見出すことができるとすると、そのパターンを利用して金融市場の動向を予測できる可能性はあると言える。近年では計算機性能の向上や取引の電子化に伴い入手可能な金融データが増えたことで、人工知能分野の技術を用いた金融市場の動向予測に関する研究も増えてきている。中でも、機械学習手法は大規模なデータの中から隠れたパターンを認識することに長けているため、金融市場予測において同手法を用いることの妥当性は高いと言える。

1.1 金融時系列情報への機械学習の応用に関する研究

高頻度注文情報には市場参加者の様々な行動が反映されており、価格のみに関する時系列情報と比較して市場参加者の行動に関する情報が豊富に織り込まれていると考えられる。近年、高頻度注文情報を機械学習

*連絡先：東京大学大学院工学系研究科
〒113-8656 東京都文京区本郷 7-3-1
Email:m2019fmatsubara@socsim.org

手法によってを分析し、証券価格動向を予測する数々の研究がなされてきた。

[Dixon 18] は仲値から上下五本の価格、注文数量に加え、成行注文の比率をRNN(Recurrent Neural Network)に入力し、仲値の動向予測を行った。RNNによる予測はロジスティック回帰やカルマンフィルタによる予測精度を上回り、次の仲値変動と核の注文系列、流動性の偏りとの時間依存性を考慮に入れた非線形な関係を捉えることに適していると論じた。また、ニューラルネットワークを金融時系列予測へ適用するには過学習への陥りやすいことを述べている。[Zhang 19] はCNN(Convolutional Neural Network)とLSTM(Long Short-Term Memory)を組み合わせて、注文情報の空間的な構造と時間依存性の2側面を捉えるDeepLOBを提案した。DeepLOBは価格と注文数量を入力としたときの仲値動向予測において、既存手法の精度を上回ると述べた。また、データ量が少ない場合は過学習を起こし、十分な汎化性能をモデルが出来ない可能性について指摘している。金融時系列予測においてモデルが訓練データに過剰適合をする場合、その予測を基に取引を行う主体は損失を出す危険性が高まると考えられる。したがって、機械学習手法を金融時系列予測に用いる際、過学習を抑制するは主要な課題の一つであると言える。

1.2 銘柄を超えた学習による株価動向予測に関する研究

[Sirignano 19b] は米国株式の高頻度注文情報から得られる価格、注文数量を入力とし、LSTMを用いて次に仲値が変動する際の方向の予測を行った。その際、予測対象の銘柄のみでモデルを学習させたケースよりも予測対象以外の銘柄のデータも併せて学習させた場合の方が予測精度が高くなる結果を得ている。

1.3 投資家行動と価格形成に関する研究

マーケットマイクロストラクチャー研究は投資家行動と価格形成の関係に焦点を当て、在庫モデルや戦略的トレーダーモデルを扱う理論研究や、注文情報を様々な角度から分析、解釈する実証分析が存在する。[Maureen 15] はこうした市場環境下において、資産の価格形成に市場の流動性がより一層大きな役割を担っていると述べ、マーケットマイクロストラクチャーに着目し現実市場の理解を深めることの重要性について論じた。[David 11] は、2010年5月6日のフラッシュクラッシュに関する研究を行った。特定の取引日の出来高から計測される売買の偏りから算出され、情報の非対称性を表すVPIN(Volume-Synchronized Probability of

Informed Trading)の累積密度関数が急激な価格変動に先行する可能性を指摘した。また、[David 12]ではLFT(Low Frequency Trading)戦略を用いる取引主体がTWAP(Time Weighted Average Price)といった執行アルゴリズム等を用いて物理時間を基に取引を行うことにより記録されるデータの特徴からHFT(High Frequency Trader)に取引意図を推察され、利用されている可能性について言及した。また、[Cao 09]は豪州株式市場における注文情報の分析を行い、需給を表す注文の偏り(Order Imbalance)と証券の短期リターンの間に有意な関係性が存在すると述べている。これらの研究からマーケットマイクロストラクチャーを考慮に入れることが金融市場の動向予測に役立つ可能性が示唆され、データドリブンなアプローチによる金融時系列予測を行う際にもマーケットマイクロストラクチャーを反映させることで予測精度、解釈性の向上に繋がると考えられる。

金融商品の価格動向を予測することができれば、投資家にとってリターンの追求やリスクの低減が可能となる。効率的市場仮説は金融商品の動向の予測は不可能であるとするが、数々の実証研究が示唆するように、現実の市場は完全に効率的とは必ずしも言えない事例が多く存在する。金融市場予測を行うにあたって、高頻度注文情報は様々な市場参加者の注文行動が記録されているため、金融商品の価格予測に適していると考えられる。また、高頻度注文情報のデータ量は膨大であるため金融時系列の動向予測をする際に機械学習手法を用いることの妥当性は高いが、過学習の抑制は最も重要な課題の一つである。機械学習手法を用いた金融時系列予測における過学習の主な原因として、金融時系列の実データは時間軸方向の数に限りがあるため訓練データが不足してしまうことが挙げられる。

本研究では、機械学習手法を用いた株価動向予測を行うにあたり、予測対象銘柄のデータに加えて他銘柄のデータも併せて学習に用いることでデータ数の不足にアプローチを行う。予測対象銘柄以外の銘柄の注文情報を用いる際、市場参加者や価格形成の挙動といった性質の類似度が高い銘柄と低い銘柄が混在しており、予測対象銘柄と類似度が低い銘柄の情報で学習を行うと予測精度が低下すると考えられる。そのため、予測精度を向上させるためには予測対象銘柄と類似度が低い銘柄の割合を下げるのが妥当であると考えられる。したがって、銘柄の類似度を基に高頻度注文情報のサブグループングを行い、多段階の学習フェーズを設けることとする。この手続きにより、他銘柄のデータと併せて学習を行うことによって訓練データの数を補いつつ、学習データにおける予測対象銘柄と類似度が低い銘柄のデータの割合を段階的に下げていくことができると考える。また、この手法を用いる際データが層状に分割されるため、以降高頻度注文情報のサブグループ

ピングを行い、多段階の学習フェーズを設ける一連の手続きを「データの層化」と呼ぶ。データの層化手法の詳細は第2章で述べる。また、1.3節で見たように短期市場では特に投資家行動、価格形成にマーケットマイクロストラクチャーが影響していると考えられるため、本研究ではマーケットマイクロストラクチャーを考慮に入れた高頻度注文情報のサブグルーピングを提案する。高頻度注文情報を用いた機械学習手法による金融市場予測タスクにおいて、マーケットマイクロストラクチャーを考慮に入れた多段階事前学習を行う研究の例はなく、本研究における新規性となる。本目的の達成は投資家にとっての運用成績向上につながるだけでなく、金融市場という市場参加者の行動が複雑に集積する系における深層学習の有効性を示すことに貢献できると考える。

2 データの層化と多段階事前学習

本章では、高頻度注文情報の層化を行い多段階の学習フェーズを設ける提案手法の枠組みについて述べる。2.1節では本研究で用いるデータの概要について説明し、2.2節では枠組みの詳細を述べる。

2.1 使用データ

東京証券取引所は市場の売買取引に関する注文、約定、売買高、といった情報をリアルタイムで提供し、蓄積されたものはヒストリカルデータとして取得可能にしている。中でも、FLEX Standardは普通債・TOKYO PRO-BOND Market 銘柄を除く全上場銘柄に関する四本値、売買高、売買代金、寄前/複数気配値段・数量、成行注文数量といった時系列情報を提供している。

本研究では、このFLEX Standardのうち、東証株価指数(TOPIX)ニューインデックスシリーズのうちTOPIX Core30とTOPIX Large70を合わせた計100銘柄(TOPIX100)に関する高頻度注文情報(2015年1月4日から2015年9月18日まで)を用いる。TOPIXニューインデックスシリーズは年に1,2回の頻度で銘柄の規模の変化に伴って定期的に区分の選定が行われる。本研究で用いる銘柄は2014年10月31日に実施された定期選定結果に従った。なお、2014年10月31日以降は2015年10月30日までニューインデックスシリーズの入れ替えは行われていない。

2.2 手法

本節では、深層学習モデルに多段階の事前学習をさせる際の高頻度注文情報の層化の枠組みについて述べる。1.3.3節で言及したように、市場参加者の取引は注文を

介するため高頻度注文情報には予測という観点で有用な情報が含まれている可能性がある。また、流動性投資家、利益追求投資家の需要と供給によって価格形成がなされることは株式市場において銘柄を問わず共通している。このことから、予測対象以外の銘柄の高頻度注文情報を用いることによって株価動向予測の精度向上が期待できる。1.3.3節で述べたように、[Sirignano 19b]は、米国株式を対象に、銘柄を超えた学習による精度向上を示した。一方で、銘柄をセクターやティックサイズの大小によってデータ分割を行い学習させる場合、使用可能な全情報を用いて仲値変動の予測をする場合と比較して精度向上にはつながらなかった。これはデータの分割を行うことによって学習に用いられるデータの量が減少したことが原因の一つであると考えられる。したがって、用いることのできる全銘柄のデータで学習させた後、予測対象銘柄のデータを用いてファインチューニングをすることによって全銘柄の情報をしつつ、予測対象の銘柄の性質に即した学習を行うことができると考えられる。

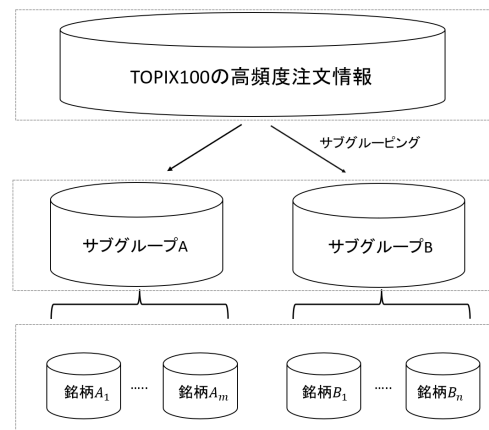


図1: サブグルーピングによるデータの層化

図1では、サブグループを2種類(サブグループA、サブグループB)とし、サブグループAの銘柄数が m 銘柄、サブグループBの銘柄数が n 銘柄(ただし $m+n=100$)とした場合の高頻度注文情報の層化の概要を示した。

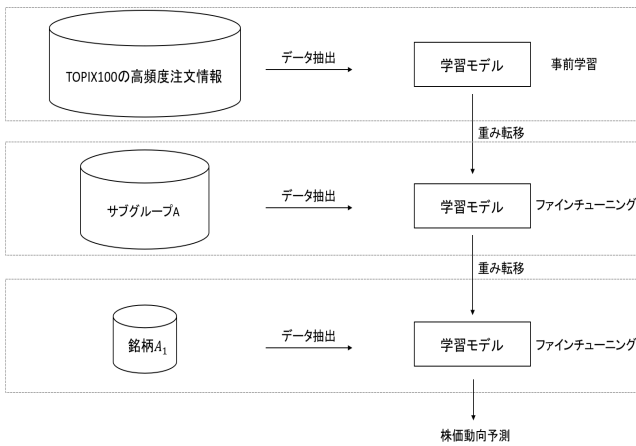


図 2: データ層化による多段階事前学習の概要図

図 2 では高頻度注文情報を用いて TOPIX100 の構成銘柄のサブグループを形成し、多段階の事前学習の枠組みの概要を示した。銘柄 A_1 を予測対象銘柄とするとき、TOPIX100 をグルーピングした際に銘柄 A_1 がサブグループ 1 に属する場合の学習手順の例である。

予測対象銘柄の比較した際に、注文情報や市場参加者の挙動及び性質が類似している銘柄のデータと合わせて学習させることが類似度の低い銘柄のデータと合わせて学習させるよりも予測精度向上に有効であると仮定すると、過去の注文情報及び市場参加者の性質が類似している銘柄同士でグルーピングを行いデータの層化を行った後、多段階の学習フェーズを設けることで予測精度の向上に繋がると考える。今回用いるデータセットの TOPIX100 は TOPIX100 ニューインデックス区分の TOPIX Core30 と TOPIX70 から構成されており、サブグルーピングを行うにあたってニューインデックス区分に従った 2 サブグループに分割させるサブグルーピングは自然な発想であると言える。次章では FLEX Standard から計測した指標を基に TOPIX100 のサブグルーピングを行うが、サブグループの数はニューインデックス区分を基に分割されるグループ数に合わせて 2 つのサブグループに分割する。分割手法の詳細は次章で述べる。また、データセットの銘柄数が N 銘柄から成る場合データは最大で N 層にまで層化することが可能であるが本研究では 3 層までとする。

3 マーケットマイクロストラクチャーを考慮に入れたサブグルーピング

第 1 章で述べた通り、近年金融市場環境は急速に変化している。特に短期で売買を行う高頻度取引、そして電子取引を活用した大口取引の執行アルゴリズムの存在感は年々増ってきている。こうした取引戦略や執

行戦略は [Jarnecic 14]、[Kissell 06] で言及されているように収益機会の向上や取引コストの削減のため、市場の流動性に影響を受けると考えられる。したがって、市場の短期価格変動の分析をする際に流動性を考慮に入れることで現実に即した予測アプローチを取ることができると考える。2.1 節で述べた通り TOPIX ニューインデックスシリーズは時価総額、流動性を基準に銘柄の選定を行っている。米国市場を襲ったフラッシュクラッシュは流動性の枯渇によって引き起こされたと考察されるが、発生直後に出来高自体は急増していること ([David 11]) を考慮に入れると売買代金のみで一元的に流動性を算出することは流動性の側面を捉える際に制限が大きいと言える。したがって、本研究では流動性の性質より詳細に捉えるため、流動性の概念を表す複数の側面に分解して考えサブグループを形成することとする。流動性指標の計測方法については 3.1 節で述べる。

3.1 流動性指標の計測

本節では [太田 11] を参考に流動性の概念を導入し、本研究で用いる流動性の計測方法について述べる。[太田 11] は流動性に 3 つの側面があるとし、それぞれ

1. 取引したいタイミングで取引できるか
2. より低い取引費用で取引できるか
3. 必要な数量を取引できるか

とした。それぞれの側面は順に即時性、タイトネス、デプスと呼ばれ、本研究では FLEX Standard を用いて各流動性を計測した。即時性を表す指標としては約定間隔 (Transaction Interval) を、タイトネスを表す指標として気配スプレッド率 (Quoted Spread Rate) [太田 11]、デプスを表す指標として仲値を中心とする上下 8 本の売り注文、買い注文の総量 (Order Volume) を選定した。1 取引日当たりの各指標を以下のように表す。

$$Transaction\ Interval = \frac{1}{n} \sum_{i=1}^n (Transaction\ Time_{i+1} - Transaction\ Time_i) \quad (1)$$

$$Quoted\ Spread\ Rate = \frac{1}{N} \sum_{i=1}^N \left(\frac{BidAsk\ Spread_i}{MidPrice_i} \right) \quad (2)$$

$$Order\ Volume = \frac{1}{N} \sum_{i=1}^N \left\{ \sum_{j=1}^8 (Ask\ Price_j^i \times Ask\ Volume_j^i + Bid\ Price_j^i \times Bid\ Volume_j^i) \right\} \quad (3)$$

3.2 クラスタリング

3.1 節で述べた方法により計測された流動性指標によって TOPIX100 の計 100 銘柄のクラスタリングを行いサブグループを形成する。ここで、クラスター数は TOPIX ニューインデックスシリーズにより TOPIX100 が Core30 と Large70 の 2 グループから構成されていることを考慮し、2 クラスターとする。本研究では k-mean 法を用いる。なお、各々の流動性指標に関してはヒストグラムに偏りが見られたため、自然対数を取るスケールングを行ったのち標準化を行った。その後、100 銘柄をユークリッド距離を距離関数とする k-means 法によって 2 クラスターに分類する。

3.3 結果

本節では、3.1 節および 3.2 節で述べた流動性指標の計測方法にしたがって TOPIX100 の銘柄をクラスタリングした結果を示す。

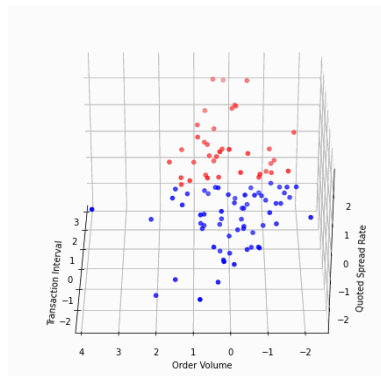


図 3: 流動性指標を用いた TOPIX100 のクラスタリング結果-その 1-

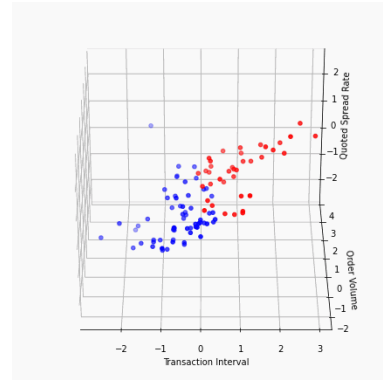


図 4: 流動性指標を用いた TOPIX100 のクラスタリング結果-その 2-

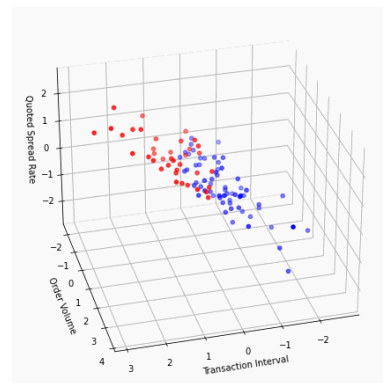


図 5: 流動性指標を用いた TOPIX100 のクラスタリング結果-その 3-

図 3、図 4、図 5 における青色に示された銘柄群をグループ 0、赤色に示された銘柄群をグループ 1 とするとグループ 0 は即時性、タイトネスの観点で流動性が平均的に高く、グループ 1 は即時性、タイトネスの観点で流動性が平均的に低い結果となった。

4 深層学習による株価動向予測

本章では、深層学習による株価動向予測を行い、予測精度を比較することにより、前章までに述べた多段階事前学習の枠組みの有効性を検証する。

4.1 実験

本研究では下記の 5 通りで TOPIX100 に対して次の仲値変動の方向の予測を行う。

1. 予測対象銘柄のデータのみを用いて学習
→ 予測

2. TOPIX100 の全銘柄のデータを用いて学習
→予測
3. TOPIX100 の全銘柄のデータを用いて事前学習
→予測対象銘柄のデータを用いてファインチューニング
→予測
4. TOPIX100 の全銘柄のデータを用いて事前学習
→予測対象銘柄がCore30に属する場合はCore30、Large70に属する場合はLarge70のデータを用いてファインチューニング
→予測対象銘柄のデータを用いてファインチューニング
→予測
5. TOPIX100 の全銘柄のデータを用いて事前学習
→流動性を基準に分割したサブグループのデータを用いてファインチューニング
→予測対象銘柄のデータを用いてファインチューニング
→予測

本研究では以後上記の5手法を Target only、TOPIX100 only、TOPIX100+Target、TOPIX100+TOPIX New Index+Target、TOPIX100+Liquidity+Target と表記することとする。

本実験では、株価動向予測を行うにあたり FLEX Standard のうち、TOPIX100 に関する高頻度注文情報 (2015 年 1 月 4 日から 2015 年 9 月 18 日まで) を用いる。2015 年 1 月 4 日から 2015 年 3 月 31 日までを訓練用データ、2015 年 4 月 1 日から 2015 年 4 月 30 日までを検証用データ、2015 年 5 月 1 日から 2015 年 9 月 18 日までを評価用データとした。本研究における実験では、13 時 00 分 00 秒から 13 時 01 分 00 秒までの高頻度注文情報をサンプリングし、13 時 01 分 00 秒より後に初めて仲値が変動したときにそれが上昇であるか下落であるかの 2 クラス分類を行った。仲値変動が上昇である場合には教師ラベル 0 を、下落である場合には教師ラベル 1 を与えた。評価指標には F1 値を用いた。入力には、[Cao 09]、[Gould 16]、[Sirignano 19a]、[Stoikov 16] を参考に Order Book Imbalance を用いた。Order Book Imbalance は (4) 式のように表される。

$$OrderBookImbalance = \frac{V_k^b - V_k^a}{V_k^a + V_k^b} \quad (4)$$

これは売買注文の偏りを表し、 V_k^a を仲値から k 番目の売り注文数量 ($k = 1$ のとき V_k^a は最良売り気配)、 V_k^b を仲値から k 番目の買い注文数量 ($k = 1$ のとき V_k^b は最良買い気配) とすると以下のように表される。本研究で用いた高頻度注文情報は仲値を中心とした上下 8 本までの売り注文および買い注文が記録されているため、

入力の次元数は 8 次元となる。また、学習モデルには 1 層の LSTM を用い、LSTM の最終隠れ層の出力はユニット数 2 の全結合層へ入力され、ソフトマックス関数により得られた確率による分類を行った。

4.2 結果

各手法ごとの F1 値の TOPIX100 全体に関する平均値は以下ようになった。

表 1: 手法別 F1 値の TOPIX100 に関する平均値

Method	F1 Score
Target only	0.5115
TOPIX100 only	0.5713
TOPIX100 + Target	0.5819
TOPIX100 + TOPIX New Index + Target	0.5884
TOPIX100 + Liquidity + Target	0.5975

4.3 考察

4.2 節から、データを層化させ多段階の学習フェーズを設ける手法によって予測精度が平均的に向上していると言える。一方で、予測対象銘柄のデータのみで学習した場合など、データの層数が少ない方が予測精度が高い銘柄も存在した。他銘柄のデータと併せて学習を行うことにより予測精度の低下が起きる原因として、以下のことが考えられる。

まず、予測対象銘柄の市場参加者や価格形成の挙動といった性質が他銘柄のそれと大きく異なる場合が考えられる。この場合、他銘柄の注文情報と併せて学習する際にデータ数の増大による予測精度向上比べ、予測対象銘柄とは類似度の低いデータによる学習の割合が高いことによる予測精度の低下の度合いが大きくなる可能性があり、最終的には予測対象銘柄のデータのみで学習した場合のほうが予測精度が高くなることが考えられる。次に、訓練用データと評価用データでの市場参加者や価格形成の挙動といった性質の時間的変化を本実験では考慮していないことがこの結果につながったと考えられる。訓練用データと評価用データにおいて市場参加者や価格形成の挙動といった性質が変化した場合、訓練用データにおいて類似度が高い銘柄群が評価用データにおいては必ずしも類似度が高いとは言えず、評価用データにおける予測精度の低下を招いたと考えられる。

4.4 投資シミュレーション

本節では、投資シミュレーションの設定、結果について述べる。4.2 節で得られた結果の中で最も F1 値が

高い銘柄数が多かった TOPIX100+Liquidity+Target モデルによる投資シミュレーションのパフォーマンスをロングオンリー戦略、ショートオンリー戦略の2手法をベースラインとして比較する実験を行う。現実の市場において仲値での取引はできないが、同じ条件下でベースラインのパフォーマンスを上回るのであればモデルに予測力が存在することを示唆していると考えられる。ここでは、13時00分00秒から13時01分00秒までの高頻度注文情報をサンプリングし、13時01分00秒より後に初めて仲値が変動したときにそれが上昇とモデルが予測した場合には買いを、下落であると予測した場合には売りをサンプリング期間の直後に入れ、仲値が変動したときに反対売買を行う戦略に基づくパフォーマンスを比較する。なお、ロングオンリー戦略はモデルの予測結果に関わらずサンプリング期間の直後に買いを、ショートオンリー戦略は売りを、仲値が変動した直後に反対売買を行う。また、各手法の取引は2015年5月1日から2015年9月18日まで一営業日に一度行い、初期状態の資金を1として複利で運用した。

TOPIX100に関する投資シミュレーションの結果、100銘柄中64銘柄において、提案手法によるパフォーマンスが最も高く、ロングオンリー戦略、ショートオンリー戦略はそれぞれ13銘柄、23銘柄においてパフォーマンスが最も高い結果となった。したがって、提案手法の取引パフォーマンスはベースラインを多くの場合上回っていると言える。このことから、提案手法は概ね実務への応用につながる予測力を有していると考えられる。

5 今後の課題

今後の課題としては、今回計測した流動性指標では説明できない予測精度の向上、低下も見られたことから、流動性指標計測の精緻化が考えられる。本研究では即時性、タイトネス、デプスをそれぞれ約定時間、気配スプレッド率、注文総量で代表させ日次で平均をとる処理を行ったが、各値の分散も考慮に入れることでより解釈性向上につながると考えられる。次に、市場参加者や価格形成の挙動といった市場構造の時間的変化への対応が挙げられる。具体的には、学習データを期間ごとに分割しデータの層化を行い古いデータを段階的に減少させるよう多段階の学習フェーズを設ける学習手法や生成ネットワークを用いて生成した注文情報によるデータ拡張といった手法が挙げられる。最後に、本研究ではマーケットマイクロストラクチャーを考慮に入れるべく流動性指標を計測し、それを基に銘柄のクラスターリングを行ったが、自己符号化器などを用いて高頻度注文情報を直接的に入力したクラスター

リングを行うことができれば人手を介する処理を省くことにつながる。人手を介することによる恣意性を排除できれば、バイアスを抑制でき予測精度の向上につながると思われる。

参考文献

- [Cao 09] Cao, C., Hansch, O., and Wang, X.: The information content of an open limit-order book, *Journal of Futures Markets: Futures, Options, and Other Derivative Products*, Vol. 29, No. 1, pp. 16–41 (2009)
- [Cont 01] Cont, R.: Empirical properties of asset returns: stylized facts and statistical issues (2001)
- [David 11] Easley, D., De Prado, M. M. L., Maureen O’Hara: The microstructure of the “flash crash”: flow toxicity, liquidity crashes, and the probability of informed trading, *The Journal of Portfolio Management*, Vol. 37, No. 2, pp. 118–128 (2011)
- [David 12] Easley, D., Prado, de M. M. L., Maureen O’Hara: The volume clock: Insights into the high-frequency paradigm, *The Journal of Portfolio Management*, Vol. 39, No. 1, pp. 19–29 (2012)
- [Dixon 18] Dixon, M.: Sequence classification of the limit order book using recurrent neural networks, *Journal of computational science*, Vol. 24, pp. 277–286 (2018)
- [Fama 70] Fama, E. F.: Efficient capital markets: A review of theory and empirical work, *The Journal of Finance*, Vol. 25, No. 2, pp. 383–417 (1970)
- [Gould 16] Gould, M. D. and Bonart, J.: Queue imbalance as a one-tick-ahead price predictor in a limit order book, *Market Microstructure and Liquidity*, Vol. 2, No. 02, p. 1650006 (2016)
- [Jarnecic 14] Jarnecic, E. and Snape, M.: The provision of liquidity by high-frequency participants, *Financial Review*, Vol. 49, No. 2, pp. 371–394 (2014)
- [Kissell 06] Kissell, R.: The expanded implementation shortfall: Understanding transaction cost components, *The Journal of Trading*, Vol. 1, No. 3, pp. 6–16 (2006)
- [Maureen 15] Maureen O’Hara: High frequency market microstructure, *Journal of Financial Economics*, Vol. 116, No. 2, pp. 257–270 (2015)

- [Sirignano 19a] Sirignano, J. A.: Deep learning for limit order books, *Quantitative Finance*, Vol. 19, No. 4, pp. 549–570 (2019)
- [Sirignano 19b] Sirignano, J. and Cont, R.: Universal features of price formation in financial markets: perspectives from deep learning, *Quantitative Finance*, Vol. 19, No. 9, pp. 1449–1459 (2019)
- [Stoikov 16] Stoikov, S. and Waeber, R.: Reducing transaction costs with low-latency trading algorithms, *Quantitative Finance*, Vol. 16, No. 9, pp. 1445–1451 (2016)
- [Zhang 19] Zhang, Z., Zohren, S., and Roberts, S.: Deeplob: Deep convolutional neural networks for limit order books, *IEEE Transactions on Signal Processing*, Vol. 67, No. 11, pp. 3001–3012 (2019)
- [太田 11] 太田亘, 宇野淳, 竹原均 F 株式市場の流動性と投資家行動, 中央経済社 (2011)