

高頻度注文情報の時系列性考慮による短期市場動向予測

Short-time market trend prediction by considering time series of high frequency order

前田 巖^{1*} 松島 裕康¹ 坂地 泰紀¹ 和泉 潔¹
ディグロー デビット² 富岡 博和² 加藤 惇雄³
Iwao Maeda¹ Hiroyasu Matsushima¹ Hiroki Sakaji¹ Kiyoshi Izumi¹
David deGraw² Hirokazu Tomioka² Atsuo Kato³

¹ 東京大学

¹ The University of Tokyo

² 大和証券株式会社

² Daiwa Securities Co. Ltd.

³ 株式会社大和総研

³ Daiwa Institute of Research Ltd.

Abstract: 高頻度取引 (HFT) により得られる高頻度注文情報を用いた短期の市場動向予測に注文が集まっている。高頻度注文情報は同時多発的な注文の系列であるため、隣接するデータの順序情報にノイズが大きく、より広いデータ範囲におけるロバストなパターン認識が必要である。本研究では注文情報のパラメータ化および畳み込みニューラルネットワーク (CNN) の活用を通し、この問題の解決を図った。注文情報に対する適切な情報抽出を行い、CNN を用いて予測モデルを構築することで、注文パターンに対しロバストな予測が可能になる手法を提案する。実験の結果、FLEX_FULL 注文データに対する解析で、本手法が高い精度で未来の株価動向を予測できることを確認した。

1 序論

金融市場の電子化・高速化に伴い、アルゴリズム取引や高頻度取引 (HFT) といった機械的な取引が台頭している。同時に、機械的な取引により生成される注文情報、取引情報も高頻度化、膨大化が進行し、金融市場の動向予測、および市場のメカニズム解明に対し有効な利用が期待されている。特に、近年発展を続ける機械学習・深層学習技術の適用により、高頻度注文情報からデータ背後の関係性を学習し、特徴抽出や予測を行う研究が行われている [1]。

高頻度注文情報は非常に高頻度で獲得できる注文情報であり、ミリ秒単位、注文単位の非常に短期間における市場動向の予測に活用できると考えられる [2]。高頻度注文情報を用いた短期の市場動向予測においては、高頻度注文情報の時系列性を適切に考慮し、特徴抽出を行うことが必要である。ミリ秒単位の同時多発的な注文の系列である高頻度注文情報では、隣接する注文の順序情報にノイズが大きい。単純に注文系列のパター

ンを学習しただけでは、市場の動向に有用な情報を抽出できるとは考えづらい。より大きな時間領域の注文パターンをロバストに学習することにより、有用な予測が可能になると考えられる。

本研究ではロバストなパターン認識を可能にするため、注文情報のパラメータ化および畳み込みニューラルネットワーク (CNN) の活用を通しこの問題の解決を図った。

2 データ

FLEX_FULL ヒストリカルデータを用いた。各銘柄について、2013年7月1日から100営業日分を使用した。予測対象は、予測時点から20注文後時点での仲値の上昇・下降とし、予測時点の仲値から1 tick分をしきい値として、上昇、下降、変化なしの3クラス分類で解析を行った。入力には、予測時点から過去100件の注文情報を用いた。

*連絡先：東京大学大学院工学系研究科システム創成学専攻和泉研究室

3 手法

3.1 注文情報のパラメータ化

注文情報を入力データとして深層学習モデルを構築するため、注文情報のパラメータ化を行った。本研究では、注文の種類（指値注文、成行注文、キャンセル注文）、売り買い、価格、注文量、および注文時の仲値を考慮し、以下のように各注文について3次元の数値データ ($x_{\text{price}}, x_{\text{amount}}, x_{\text{midprice}}$) への変換を行うことで、注文からの特徴抽出を行った。3つの変数はそれぞれ、注文価格、注文量、注文時の仲値を表している。

$$x_{\text{price}} = P - MP \quad (1)$$

$$x_{\text{amount}} = \log(A) \quad (2)$$

$$x_{\text{midprice}} = MP - MP_- \quad (3)$$

ここで、P, MP, A, MP_- はそれぞれ、注文価格、注文時の仲値、注文量、直前注文の仲値を表す。注文価格、および仲値について差分を取ることで、データの正規化を行っている。また、買い注文では $x_{\text{price}} > 0$ 、売り注文では $x_{\text{price}} < 0$ 、指値注文および成行注文では $x_{\text{amount}} > 0$ キャンセル注文では $x_{\text{amount}} < 0$ となり、注文の種類が区別可能となる。

以上のように変換した3次元データを過去100注文分結合し、 100×3 次元の特徴量を各時点に対する説明変数とした。

3.2 深層学習モデル

表1のようなCNNモデルを作成し、学習を行った。Dropout層の25%はdropoutを行う割合を表している。1層目の畳み込み層では注文方向および注文パラメータ方向に畳み込みを行い、2層目以降では注文方向に畳み込みを行っている。活性化関数にはReLU関数を用いた。学習時には、サンプル数の比に応じクラスごとに重みを設定し損失関数の計算を行っている。

4 結果

使用した100営業日のデータのうち、前半50日をモデル構築用データ、後半50日をモデル検証用データとして解析を行った。

企業コード3407に対する結果を以下に示す。提案手法の精度評価指標を表2に示す。Trainingがモデル構築用データに対する精度指標を、Testがモデル検証用データに対する精度指標を表している。比較として、ロジスティック回帰で予測モデルを作成した場合の結果を表3に示す。各データに対するAccuracyおよびF1

表 1: 提案手法のネットワーク構成

type	patch size	stride	output size
convolution	3×3	1	$98 \times 1 \times 64$
convolution	3×1	1	$98 \times 1 \times 64$
max pool	3×1	2	$49 \times 1 \times 64$
dropout(25%)			$49 \times 1 \times 64$
convolution	3×1	1	$49 \times 1 \times 64$
convolution	3×1	1	$49 \times 1 \times 64$
max pool	3×1	2	$24 \times 1 \times 64$
dropout(25%)			$24 \times 1 \times 64$
convolution	3×1	1	$24 \times 1 \times 64$
convolution	3×1	1	$24 \times 1 \times 64$
max pool	3×1	2	$12 \times 1 \times 64$
dropout(25%)			$12 \times 1 \times 64$
linear			3
softmax			3

scoreは、上昇・下降・変化なし3クラスそれぞれに対するスコアの、各クラスのデータ数を用いた重み付き平均で計算されている。AccuracyおよびF1 scoreの値より、高い精度で短期の仲値動向を予測できていることがわかる。

提案手法の優れている点として、予測の正解率 (precision) が高く、実際の動向と逆方向の予測を出すことが非常に少ないことが挙げられる。表4および表5に、提案手法、ロジスティック回帰のモデル検証用データに対する混同行列、および予測に対する割合を示す。

予測ラベルおよび実際のラベルは、Downが下降、Stayが変化なし、Upが上昇を表している。表4より提案手法ではDownと予想した場合に実際に実際はUpおよびDownである割合がそれぞれ23.8%, 1.1%となっており、実際と逆方向の変動を予測する可能性が、実際の変動を正しく予測する可能性に比べ非常に低いことがわかる。対してロジスティック回帰では、同様の割合がそれぞれ10.6%, 1.5%となっており、提案手法に比べ正解率および実際と逆方向の予測を行う確率が高くなっている。予測の正解率および実際と逆方向の予測を行う確率は特に金融商品の取引において重要であり、その点で提案指標がより適しているといえる。

表 2: 提案手法の精度指標

	Accuracy	F1 Score
Training	0.880	0.866
Test	0.875	0.855

表 3: ロジスティック回帰の精度指標

	Accuracy	F1 Score
Training	0.356	0.488
Test	0.353	0.497

表 4: 提案手法のモデル検証用データに対する混同行列および予測に対する割合

	Predicted:		
	Down	Stay	Up
Actual:			
Down	7823(23.8%)	51345(4.7%)	219(1.3%)
Stay	24676(75.1%)	976845(90.3%)	11452(66.8%)
Up	368(1.1%)	53894(5.0%)	5464(31.9%)

表 5: ロジスティック回帰のモデル検証用データに対する混同行列および予測に対する割合

	Predicted:		
	Down	Stay	Up
Actual:			
Down	14110(10.6%)	42656(5.1%)	2621(1.7%)
Stay	116261(87.8%)	759223(90.2%)	137489(86.9%)
Up	2003(1.5%)	39561(4.7%)	18162(11.5%)

5 まとめ

高頻度注文情報を用いた短期の株価動向予測を行った。注文情報のパラメータ化および CNN の活用により、高頻度注文情報が持つ系列性を適切にモデリングでき、高い精度で価格動向を予測することができた。

6 免責事項

本稿は、著者の個人見解を表すものであり、大和証券株式会社および株式会社大和総研の公式見解を表すものではありません。

参考文献

- [1] E. Chong, C. Han, and F. C. Park,: Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies, *Expert Syst. Appl.*, Vol. 83, pp. 187–205 (2017)
- [2] 田代大悟: 平成 29 年度 修士論文 (2018)