

# 高頻度板情報の時空間パターン分析による株価動向推定

Stock Price Estimation by Spatio-temporal Pattern Analysis of High Frequency Tick Data

中山敦貴<sup>\*1</sup>      山田健太<sup>\*1\*2</sup>      和泉潔<sup>\*1\*3</sup>  
Atsuki Nakayama      Kenta Yamada      Kiyoshi Izumi

<sup>\*1</sup>東京大学大学院工学系研究科      <sup>\*2</sup>科学技術振興機構 PRESTO  
Graduate School of Engineering, The University of Tokyo      PRESTO, JST

<sup>\*3</sup>科学技術振興機構 CREST  
CREST, JST

金融市場における情報技術の発展により、市場価格だけでなく買い注文と売り注文の集合である板情報の解析が可能となった。板情報は、株価変動の背後にある注文情報を含むので、価格情報より網羅的に市場動向を解析できると考えられ、暴落の予兆発見などの応用が期待される。しかし、情報量が市場価格に比べて桁違いに大きくなるためその扱いは困難である。そこで本研究では、板情報より各価格での注文量の増減を時間と価格を軸に持つ2次元平面上にマッピングした画像データを作成し、これを用いて板情報の時空間パターンを抽出する方法を用いる。そして実際に複数銘柄の板情報を画像化したものから、ある時点から10秒後の株価が上昇するかどうかという枠組みでの実験を行う。手法としては、ロジスティック回帰と畳み込みニューラルネットワークという2種類の手法を用いた学習の結果を、画像サイズも変化させながら比較した。またその際、今回は板情報の変化を注文、キャンセル、約定の3種類の情報に分解し、それぞれから画像を作成し、この3通りの結果を比較した。

## 1. はじめに

近年の情報通信技術の発達と金融取引の電子化に伴い、金融に関する膨大なデータが日々生成されている。そして、それら膨大なデータを元に株価などを予測したいという需要が高まりつつある。その中で、機械学習と呼ばれる分野では、金融データマイニング、金融テキストマイニングなどといった名前で知られている研究が行われている。金融データマイニングの分野では、GA(遺伝的アルゴリズム)を用いてテクニカル指標に依ったシステムトレードを行った研究や[新井 13]、ニューラルネットワークとGAを組み合わせる取引による利益向上を計った研究[山口 03]などが存在する。また、金融テキストマイニングと呼ばれる分野では、新聞記事や日銀の月報を使って予測を行う研究[和泉 10]などが存在している。

## 2. 関連研究と本研究の目的

売数量	値段	買数量
2000	1931	
4000	1930	
3000	1929	
1500	1928	
	1927	2400
	1926	4000
	1925	2000
	1924	3000

図 1: 板情報概念図

本研究では、[中山 15]で提案した、板情報と呼ばれる注文時刻と注文株数がペアになった高頻度な数値情報を用いて、株価変動の予兆発見のために空間情報として取り扱って時空間パターンを抽出する手法を用いる。

株式市場には板と呼ばれる仕組みが存在する。これは、1927円で買いたい注文が累計2400株あり、1928円で売りたい注文が累計1500株ある、といった情報を価格ごとにまとめた物である。板の動きの結果として株価が変動しているという因果関係があるために、この板に関する情報(板情報と呼ばれる)を活用する事は、株価などを予測する際に有用だと考えられる。板情報を研究に用いた取り組みとしては、株式の板情報から抽出した特徴ベクトルを用いて市場の状態変化を見た研究[梅岡 12]、板情報に対して混合ガウス分布と呼ばれる確率モデルを用いて、市場の状態変化を検出した研究[宮崎 13]などが存在する。

[梅岡 12]や[宮崎 13]では、1分単位で板情報を分割し、最良気配値を中心とした各価格に入る合計注文量の対数を取ったものを、1分間にその位置に入った回数で割って、8次元のベクトルを1分間に1つ得ている。そしてこのベクトルを長時間に渡って連結したものを、一つの特徴量として用いて、クラスタリングなどを行っている。

ここで用いられている特徴量の問題点として挙げられるのは、例えばある注文が1分前に行われていただけで、全く違う位置のベクトル成分に変化が起こり、結果としてある注文が1時間前に行われていた場合と1分前に行われていた場合で、この1時間前にずれたベクトルと1分前にずれたベクトルの距離の差が同じになってしまう事である。同様に、ある注文がある価格の1円上に移動した場合と、10円上に移動した場合でも、2つのベクトルの距離の変化は全く同じ物になる。まとめると、従来手法は株価の時間方向や価格方向の位置変化に対する脆弱性がある。

上述の問題を解決するために、本研究で用いるのは、板情報を用いた株価の予測の前段階として、板情報を一枚の画像のような時空間パターンとして表現する手法である。画像処理の分野は、機械学習の中でも古くから研究が進んでいる分野の一つ

連絡先: 中山敦貴, atsukinakayama@gmail.com

で、学習に画像処理の手法を用いる事を可能にする事は意義が大きいと考えられる。例えば、上述の位置変化に対する耐性（位置不変性があると言われる）を持つ手法である畳み込みニューラルネットワーク [ビショップ 08] や、これを組み込んだディープラーニングを用いる事は、位置関係が保存された画像データに対しての操作として有効な手法なのではないかと考えられる。そこで本研究では、畳み込みニューラルネットワークとロジスティック回帰という2手法を実際に株価の予測のための学習に使い、結果を比較、評価する事を目的とした。またその際、今回は板情報の変化を注文とキャンセルと約定の3種類に分解し、それぞれに対して画像を作成し、3通りの結果を得た。これにより、どの情報が株価予測のために効果的なのかを評価する事も目的とした。

### 3. 手法

今回の論文で想定するフレームワーク、「入力」→「特徴量抽出」→「学習」→「検証」の4段階のうち、この章ではまず「入力」である板情報からの「特徴量抽出」、つまり画像化について述べる。次に、「学習」手法として用いるロジスティック回帰、畳み込みニューラルネットワークについて述べるが、こちらについては新規性のある手法では無いため、参考論文を紹介するにとどめた。最後に「検証」手法について、実際の実験の流れを述べる。

#### 3.1 板情報の画像化

この節では、板情報を画像にまとめるための具体的な手法について述べる。まず、ある時点での最良気配値を中心として、上下30ティックまでの1秒間の合計注文量の絶対値の対数を取り、その後買いの注文量にはマイナスをつけたベクトルを作成する。この際、ティックサイズ的に存在するはずの価格が飛んでいる場合、例えば100円と102円に注文があるが101円には無い場合、101円を飛ばすのではなく、101円に対応する注文量を0とする。つまり、上下30ティックだからといって、上下30本値まで見るとは限らない。

1秒間の合計注文量を見るには、1秒間の全ての板情報を順番に見ていき、注文量の変化があるたびにそれをカウントして保存し、その和を取る。例えば基本となる取引単位が1000株な銘柄で、ある1秒で101円に売り注文が10000株あった後、同じ1秒の間に101円で5000株の売り注文があった場合、1秒間の合計注文量は、15000を基本単位である1000株で割った上でその対数を取った物となる。これが買いであった場合、注文量は $15000 \div 1000$ の対数を取った物にマイナスを掛けた物となる。

次に、1秒前の最良気配値を中心として、上下30本値までの1秒間の合計注文量を同様にベクトル化し、先に作った最初のベクトルの横につなげて行列としていく。この操作を60秒間分繰り返して、60秒×60ティックの画像を生成する。

学習に用いる際は、この画像を、60秒ごとに1つ作り、画像の次の10秒間に値上がりしたのか値下がりまたは価格変化無しかったのかの2クラスで画像をラベル付ける、という手法を取る。

上述の操作において注意が必要な点として、データに欠損があった場合や、一日の価格の下端や上端を画像が超える場合、その間の合計注文量は0とした。

上述のように画像化するメリットとして、まず板情報の視認性が良くなる事が挙げられる。時間を横軸に、価格を縦軸にとっているので、板の状態変化をアニメーションなどを用いず一枚の画像だけで理解出来る事だけでも、画像化の意義はあると考

えられる。次に、単なる注文量の羅列ではなく、注文の位置や距離が時間や価格ときちんと対応する事が挙げられる。これにより、1円のズレと10円のズレはきちんと異なる物として扱われ、また適切な手法と組み合わせる事で位置不変性などを持った価格予測アルゴリズムの開発などが可能になる、といった展望が開けるようになる。図2、図3、図4に、具体的に今回用いる注文、キャンセル、約定の情報を元に作成した画像の例を示した。

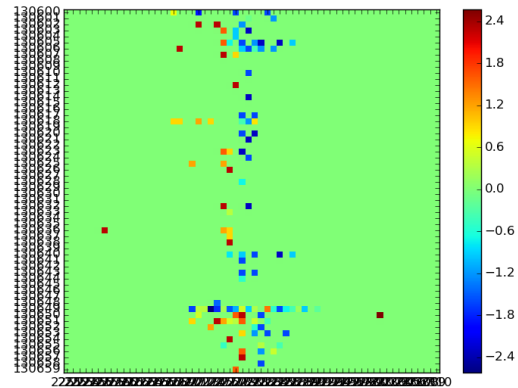


図2: 作成した画像の例.60×60で、縦軸が時刻、横軸は価格を表す。この図は注文のみを表示している画像であり、2013年7月1日後場の画像セット内の1枚である

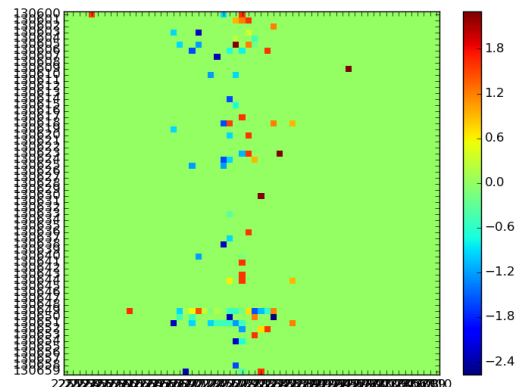


図3: この図はキャンセルのみを表示している画像であり、図2と同時刻の1枚である

#### 3.2 ロジスティック回帰

ロジスティック回帰は、ソフトマックス関数と呼ばれる関数の係数とバイアス項を教師あり学習していく事で、教師データに対し適応していく機械学習の手法の一つである。詳しくは、[deeplearning.net1]が実装まで分かり易く解説されている。

#### 3.3 畳み込みニューラルネットワーク

畳み込みニューラルネットワークは、与えられた2次元のデータに対して、何種類ものフィルタによるフィルタリングを行い、その出力結果を縦横に繋げた物に対して maxpooling と呼ばれる処理を行った上で、場合によってはもう一度同様の操作を行

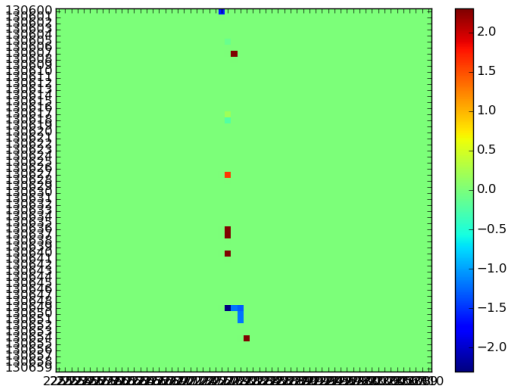


図 4: 同様に、この図は約定のみを表示している画像であり、図 2 と同時刻の 1 枚である

うニューラルネットワークである。この結果を通常のニューラルネットワークあるいはディープラーニングに入力して、最終的な結果を得る、そしてフィルタ係数やニューラルネットワーク部分の係数やバイアスの学習を行うタイプの畳み込みニューラルネットワークは深層学習の一種であり、機械学習の手法の一つである [deeplearning.net2]。

#### 4. 実験

本論文では、先の章で述べた手順での画像作成を、注文データ、キャンセルデータ、約定データのそれぞれについて行い、さらにこれら 3 種類の画像から、画像の下側中央  $40 \times 40$  の切り出しもしくは下側中央  $20 \times 20$  の切り出しを行った計 6 通りの画像について実験を行った。銘柄は今回選択した 5 銘柄 (銘柄コード 1925, 4188, 5401, 5411, 6301) とした。手法は、先に述べたロジスティック回帰および畳み込みニューラルネットワークの 2 通りを用いて、結果を比較した。プログラムの作成は、[deeplearning.net1], [deeplearning.net2] を参考にしつつ、改変を加えて行った。

実験手順は、以下の通りである。まず、期間を 2013 年 7 月から 2014 年 6 月と定め、この 1 年間の全ての板情報のうち、前場と後場から 1 日 1 時間ずつ画像化を行い、正解としてその画像の 10 秒後に株価が上昇しているを 1 とし、株価が上昇していないを 0 とし各画像にラベル付けを行った。

そうして出来た数万枚の画像のうち前半 7 割を学習用セット、その後の 2 割を検証用セット、最後の 1 割をテスト用セットとして分けた。そしてそれぞれのセットの正解 1 と正解 0 の数が揃っていない問題を解決するために、それぞれのセットに対して 1 と 0 の数が同じになるように画像と正解のコピーを行った。これにより、完全にランダムに判定する学習器の結果は 50% に非常に近くなるようになり、異なる銘柄や手法、画像の種類の比較を正解 1 と正解 0 の数の比にかかわらず偏りなく行う事が出来るようになるメリットがある。

そうして作成した画像データのセットを用いて、ロジスティック回帰、および畳み込みニューラルネットワークで学習を行う。その際、ある一定数の画像の学習を行うたびに、検証用セットを用いて結果の検証を行う。この結果がこれまでの最高値だった場合のみ、さらにテスト用セットを用いて結果のテストを行う。そうして得られた結果を、以下の章でテスト、あるいはテストの結果と表記する。

#### 5. 結果

実験結果を分析しやすいよう、各実験のテストにおける正答率の平均値を取って表の形にまとめた物が表 1, 表 2, 表 3, 表 4, 表 5, 表 6 になる。表 1, 表 2, 表 3, 表 4 では銘柄ごとの比較、及び各銘柄の平均値を、画像サイズが  $20 \times 20$  か  $40 \times 40$  なのか、およびロジスティック回帰と畳み込みニューラルネットワークについてまとめている。表 5, 表 6 では、画像の種類に着目し、注文のみ、キャンセルのみ、約定のみの場合のそれぞれについて、画像サイズごと、手法ごとの平均値をまとめている。

#### 6. 考察

まず、表 1, 表 2, 表 3, 表 4 について述べる。銘柄ごとの比較では、3 画像を平均した正答率の傾向は手法、画像サイズを問わず銘柄ごとに一定の類似性があると見る事もできるが、画像サイズの変化により値が上昇する銘柄も存在していれば下落する銘柄も存在しており、一定のランダム性を伴うものであると見る事が出来る。どの銘柄も、基本的な「学習しやすさ」のような物が存在している上で、手法や画像サイズにより相性の良いものとそうでないものが存在しているのではないかと考えられる。5 銘柄の平均値は、ロジスティック回帰の  $40 \times 40$  の画像が一番良いという結果が得られた。

次に、表 5, 表 6 について述べる。まず全体の傾向として、この中で一番良い学習器になるのは、約定画像のみを使った場合であるという事が言える。それに続いて注文画像のみ、キャンセル画像のみ、と続く場合が 3 通り存在する中で、 $40 \times 40$  の畳み込みニューラルネットワークでは、キャンセルが注文を逆転しているが、これは例外的で、全体の平均値で考えると、注文画像のみはキャンセル画像のみよりも良い学習器を作ると言う事が出来るだろう。ここから示唆されるのは、約定という行為が持つ情報が株価を予想する上でもっとも大きなファクターであり、次に注文が来て、最後のキャンセルはあまり重要でない情報であるという結論である。

また、総評として、正答率が高い銘柄では 60% を超えており、機械学習の力で株価上昇を予想しようという試みは、一定の結果を示したと言えるだろう。一方で、銘柄 1925 は他銘柄に比べるとあまり良くない結果が出ており、この原因として、流動性の足りなさというのがファクターとして挙げられる可能性があると考え調べてみたものの、結果、注引量自体は他の銘柄と比べても少ないとはいえず、原因究明は今後の課題と言える。

#### 7. 結論

今回は、板情報の画像化に関して、学習手法の導入、注文、キャンセル、約定情報それぞれによる画像の生成を行い、結果として一定の成果を残せる学習器を作ることが出来た。また、注文、キャンセル、約定の情報を比べた場合、今回の正解率の結果を踏まえると、約定、注文、キャンセルの順で重要なのではないかと考えられる。

#### 参考文献

[ビショップ 08] C・M・ビショップ著 (元田浩, 栗田多喜夫, 樋口知之, 松本裕治, 村田昇訳), 『パターン認識と機械学習- ベイズ理論による統計的予測 上』, シュプリンガー・ジャパン (2008)

[deeplearning.net1] <http://deeplearning.net/tutorial/logreg.html>

[deeplearning.net2] <http://deeplearning.net/tutorial/lenet.html>

[新井 13] 新井佑弥, 折原良平, 中川博之, 清雄一, 田原康之, 大須賀昭彦「GA を用いた適合度関数と相場変化に着目したシステムトレード」, 第 10 回 SIG-FIN, 2013

[山口 03] 山口和孝, 坂井修一, 「ニューラルネットと遺伝的アルゴリズムを用いた株式売買支援システム」, 東京大学大学院情報理工学系研究科電子情報学専攻修士論文 (2003)

[和泉 10] 和泉 潔, 後藤 卓, 松井 藤五郎, 「テキスト情報による金融市場変動の要因分析」, 人工知能学会論文誌, Vol. 25, No. 3, pp. 383387 (2010)

[梅岡 12] 梅岡利光, 鳥海不二夫, 平山高嗣, 榎堀優, 石井健一郎, 間瀬健二, 「板情報を用いた株式市場の状態変化の分析」, 第 37 回 JAFEE 大会 (2012)

[宮崎 13] 宮崎文吾, 和泉潔, 鳥海不二夫, 橋諒, 「混合ガウスモデルを用いた市場注文状況の変化の検出」, JPX ワーキングペーパー, Vol. 3, (2013)

[中山 15] 中山敦貴, 山田健太, 和泉潔「高頻度板情報の時空間パターン分析による株価動向推定」, JSAI2015

表 1: 畳み込みニューラルネットワーク, 20 × 20 の画像を使用した結果

銘柄コード	1925	5411	6301	5401	4188
外挿予測精度 (%)	52.27	53.75	57.03	55.02	55.04

表 2: 畳み込みニューラルネットワーク, 40 × 40 の画像を使用した結果

銘柄コード	1925	5411	6301	5401	4188
外挿予測精度 (%)	52.94	53.88	57.09	58.76	53.42

表 3: ロジスティック回帰, 20 × 20 の画像を使用した結果

銘柄コード	1925	5411	6301	5401	4188
外挿予測精度 (%)	50.59	54.98	59.45	56.08	57.76

表 4: ロジスティック回帰, 40 × 40 の画像を使用した結果

銘柄コード	1925	5411	6301	5401	4188
外挿予測精度 (%)	50.86	57.61	61.14	58.58	60.24

表 5: 畳み込みニューラルネットワークを使用した結果

畳み込み	20*20 平均	40*40 平均
注文情報 (%)	55.34	54.23
キャンセル情報 (%)	52.21	54.70
約定情報 (%)	56.30	56.71

表 6: ロジスティック回帰を使用した結果

畳み込み	20*20 平均	40*40 平均
注文情報 (%)	53.93	55.03
キャンセル情報 (%)	52.19	52.51
約定情報 (%)	55.77	57.68