

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^{*1} 山田健太^{*1*2} 和泉潔^{*1*3}
 Atsuki Nakayama Kenta Yamada Kiyoshi Izumi

^{*1}東京大学大学院工学系研究科 ^{*2}科学技術振興機構 PRESTO
 Graduate School of Engineering, The University of Tokyo PRESTO, JST

^{*3}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。また、価格や注文時間は離散的なため、サポートベクターマシンやニューラルネットワークなどの既存のパターン抽出手法はそのままでは適用できない。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提案する。

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を組み合わせて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

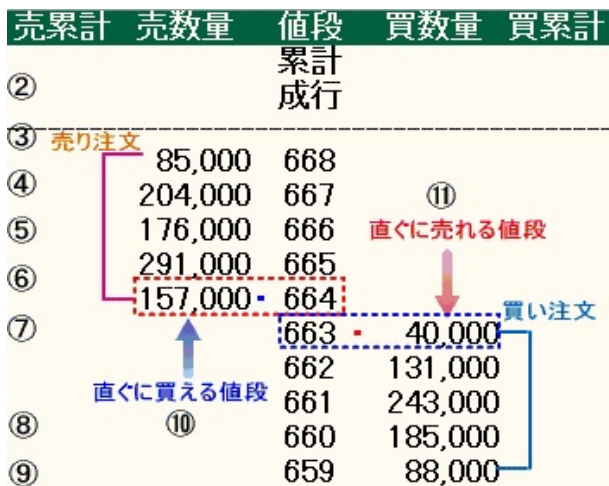


図 1: 板情報概念図 [板情報概念図]

本研究は、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提案する。

株式市場には板と呼ばれる仕組みが存在する。これは、100円で買いたい注文が累計5000株あり、101円で売りたい注文が累計10000株ある、といった情報を価格ごとにまとめたものである。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れ、結果としてある注文が1時間前に行われていた場合と、2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と言える。

今回提案するのは、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などの手法を適用す

連絡先: 中山敦貴, atsukinakayama@gmail.com

る事が、応用例として考えられる。

3. 板情報の分析手法

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、この論文では主に入力からの特徴量の抽出について述べる。

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10本値までの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、その注文量は0とする。つまり、10本値だからといって、10ティック上までしか見ないとはいえない。表示する上限は、学習のために実用上では、基準時点の最良気配値の上下30ティック、などと一定の範囲に決めるのが良いと思われる。だが今回は、株価の変化の視認性を高めるために、基準時点を含む1日に登場する全ての価格を含むようにした。このベクトルは、10本値より先の部分は注文量0で埋められる。

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。例えばある1秒で101円で売り注文が2000株あった後、同じ1秒の間に101円で500株の売り注文キャンセルがあった場合、1秒間の合計注文量は1500となる。これが買いであった場合、注文量はマイナス1500となる。

次に、1秒前の最良気配値を中心として、上下10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する。この操作を60秒間分繰り返して、60×価格数の画像を生成する。

学習に用いる際は、この画像を、10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、といった手法が考えられる。

上述の操作において注意が必要な点として、ある価格の位置が例えば第9本値から第11本値に変化した場合が挙げられる。今回は、注文量が見えなくなったということで、一律その後の合計注文量は0として扱った。また、データに欠損があった場合も、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。

4. 実験

本論文では、実験で作成した画像の中から、株価が大きく変化している様子を例として提示する。具体的には、銘柄コード7733(オリンパス)の2011年10月14日前場を選択した。選択した理由として、ニュース報道により大きな株価の変化が起こっている点、有名な銘柄であり流動性も高い点が挙げられる。画像

化は、元となる板情報を元に、上述の手法で行った。その後の実際の学習過程として想定出来る手順としては、畳み込みニューラルネットワークにラベル付けした画像のうち1000枚を入力、200枚を検証用に、100枚をテストに使い、その正答率などを評価する事が考えられる。

5. 板情報の画像化の結果と考察

以下のような画像を、1分につき1枚得る事が出来た。ここで、青が濃いほど売りが優勢である事を示し、色が薄く白に近いほど買いが優勢である事を示す。横軸が時間であり、縦軸が注文価格を表す。



図2: 横軸は2011年10月14日9時32分から32分59秒まで、縦軸は2460円から2490円までを表すよう拡大してある

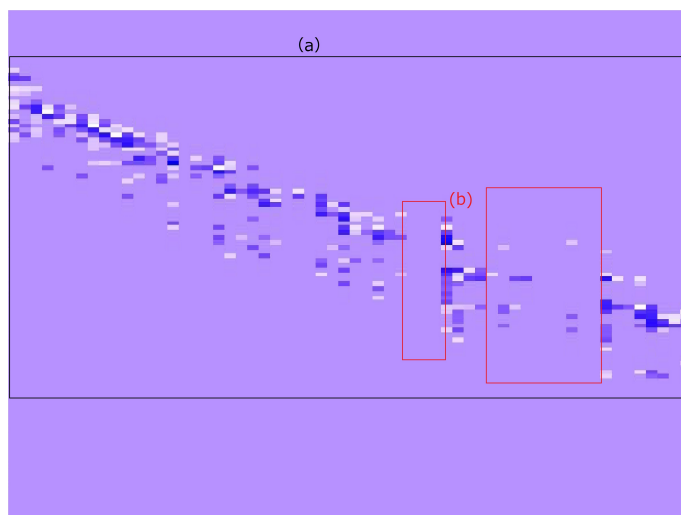


図3: 横軸は2011年10月14日9時33分から33分59秒まで、縦軸は2380円から2490円までを表すよう拡大してある

図2から図4を見ると、図3と図4の左上から右下にかけての黒色四角(a)内の価格が下がっている最中では、全体的に売りが優勢となっており、特に10本値の上端で青が濃くなっている。価格の下落が始まり、売り注文や買い注文のキャンセルが大量に入っている様子がここから分かる。また下がっている

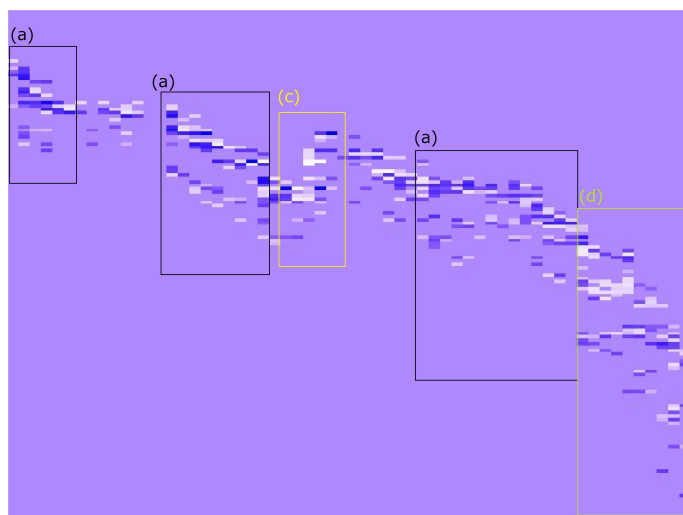


図 4: 横軸は 2011 年 10 月 14 日 9 時 34 分から 34 分 59 秒まで、縦軸は 2440 円から 2290 円までを表すよう拡大してある

最中でも図 3 右下の赤色四角 (b) 内のように、一時的に注文量が殆ど無い時間帯 (あるいは注文量と注文のキャンセル量がほぼ等しい時間帯) があったりしており、注文の多さには秒単位で大きなばらつきがある事が分かる。そして、図 4 中央上側あたりの黄色四角 (c) の価格が下げ止まっている部分では一時的に急速に買いが優勢になって、それから価格も上昇している様子が良く分かる。また、図 4 右下の黄緑色四角 (d) 内のように、急落中は板が薄くなり、広い価格に 10 本値が分布する傾向も分かる。一枚の画像でこのように市場の状態をある程度推察する事が出来る事は、一定の価値があると言う事が出来るだろう。

今後の展望としては、二通りの路線が考えられる。一つは、まずこれらの画像に画像処理の手法を適用する事だ。そしてもう一つは、これら画像を 1 枚ではなく複数枚セットにする事で情報量を増やす事である。例えば今回は 1 秒間の合計注文量を基本単位にしたが、その 1 秒間での平均累積注文量を表すもう一枚の画像と組にすることで、より正確に市場の状態を捉えられると考えられる。これは 2 色のデータから成る画像を処理する際の方法論として語る事が出来るだろう。

具体的な学習プロセスとしては、まずは今回のような 1 枚の画像を 1300 枚程度用意した上で、例えば 10 秒後に価格が上がっていれば 1、上がっていなければ 0 といったラベル付けを行う。その内 1000 枚を畳み込みニューラルネットワークに入力し、それにディープラーニングを繋げ、最終段においてロジスティック回帰を行うようなシステムの学習を進める。validation を 200 枚の画像で時々行いながら学習を進め、十分良いモデルを構築し終えた段階で、残り 100 枚でテストを行い、その正答率などを評価する事が考えられる。

参考文献

[板情報概念図] <http://www.kabudream.com/hajimete/hajimete2.5.html>

[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)

[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)