

# 板情報を利用した指値執行システムの提案

## Proposal of limit order execution system using order book information

大矢 倫靖<sup>1\*</sup> 村岡 洋一<sup>2</sup>  
Tomoyasu Oya<sup>1</sup> Yoichi Muraoka<sup>2</sup>

<sup>1</sup> 早稲田大学基幹理工学研究科

<sup>1</sup> Department of Computer Science and Engineering, Waseda University

<sup>2</sup> 早稲田大学

<sup>2</sup> Waseda University

**Abstract:** We proposed an idea of order book analysis for predicting the direction of Nikkei 225 index in short terms using neural network. Not only the prediction of the market was proposed, but also the probability of execution was described as the first passage time of the Brownian motion in our model. Our system worked well for predicting the market behavior to some extent.

## 1 はじめに

本研究では、板 (Order Book) を用いたデータの分析を基にアルゴリズムトレーディングに使えるモデルを構築する。株式の売買に用いられる板は、指値と成行注文に基づき、価格が決定される。近年においては、長期的な株価変動や日足ベースのデータを用いた研究だけではなく、1分足のデータやティックデータなど、分析粒度のさらに小さい短い期間のマーケットの性質が注目されている [1][2][3]。さらに、そのティックデータが出来上がる根本となっているメカニズムが、板のようなミクロな要素であり、その積み重ねが長期的な株価変動を生み出していると考えられる。板を分析して、モデリングしようという研究は盛んになっている [4]。人間が株式等をトレーディングする場合においても板を見ている場合が多いと考えられ、板を見ることで有用な情報が手に入れられる場合も多いと考えられる。実際に日本の株式市場において Rama[5] らが行った、板の値動きを待ち行列の理論を用いて説明しようという試みがある。本研究においては、板や過去のトレンドを見ながら、どこに指値を指せばよいのかという意思決定を行えるようなモデルを構築することを目的とする。

### 1.1 板の構成要素

株式市場で使われる板とは、売買の注文が表示される場所であり、売りと買いの需要と供給がマッチしたところで価格が決定される。主に指値注文と成行注文があり、指値注文は「何円になったら買い」のように値段を指定して注文を出すことであり、成行注文は、「値段はいくらでもよいからすぐに買いたい」という注文である。成行注文は指値注文よりも優先される。表 1 に板の様子 (日経 225 先物) を示す。

表 1: 板の様子

	売り気配	価格	買い気配
(売り気配 5)	250	10030	
(売り気配 4)	190	10020	
(売り気配 3)	305	10010	
(売り気配 2)	263	10000	
(売り気配 1)	35	9990	
		9980	25 (買い気配 1)
		9970	350 (買い気配 2)
		9960	270 (買い気配 3)
		9950	390 (買い気配 4)
		9940	125 (買い気配 5)

\*連絡先: 早稲田大学基幹理工学研究科  
〒169-8555 東京都新宿区大久保 3-4-1  
E-mail: ooya-oh-yeah@toki.waseda.jp

## 2 提案方法

提案システムの概要は次の図1の通りである。まず、板情報、トレンドなどの情報を利用してニューラルネットワークの入力とする。ニューラルネットワーク内のパラメータはあらかじめ学習されたものを用い、出力は予測に基づき、上がる、下がる、変化なしの3パターンに分類する。その情報を元に、指値を上に出すべきか、下に指すべきかの意思決定を行う。

実際に指値がどれくらいの確率で約定するかを計算するために、ブラウン運動の初到達時刻分布の性質を用いて、累積分布関数から何分後に約定するかといった確率を計算することでより良い約定を目指す、ということが本システムの手法である。

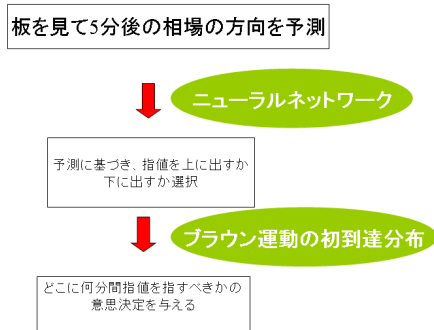


図 1: システム概要図

## 3 ニューラルネットワークによる株価予測

### 3.1 バックプロパゲーション

本研究においては、ニューラルネットワークの学習としてバックプロパゲーションを用いている。バックプロパゲーションは最急降下法を用いた手法の一つであり、学習データに対して逐次重みをアップデートすることで教師信号との誤差を最小にすることで学習を行う。学習サンプルの  $k$  番目の教師信号を  $t_k$  とし、出力信号を  $O_k$  とする。誤差

$$E = \frac{1}{2} \sum_{i=1}^n (t_i - O_i)^2 \quad (1)$$

を最小化するような重みを決定することを考える。 $E$  を  $W_{kj}$  で偏微分すればよく、中間層-出力層に関しては、学習係数を  $\eta$ 、中間層の出力を  $H$  とすると

$$\Delta W_{kj} = -\eta \frac{\partial E}{\partial W_{kj}} = -\eta \frac{\partial E}{\partial O_k} \frac{\partial O_k}{\partial U_k} \frac{\partial U_k}{\partial W_{kj}} \quad (2)$$

となるような修正量にすればよく、

$$\Delta W_{kj} = -\eta \{ -(t_k - O_k) O_k (1 - O_k) H_j \} \quad (3)$$

のようにして、出力層の結合荷重を求められる。入力層-中間層に関しても同様に、

$$\Delta W_{ji} = -\eta \frac{\partial E}{\partial W_{ji}} \quad (4)$$

$$= -\eta \left( \sum_{i=1}^n \frac{\partial E_i}{\partial O_i} \frac{\partial O_i}{\partial U_k} \frac{\partial U_k}{\partial H_j} \right) \frac{\partial H_j}{\partial T_j} \frac{\partial T_j}{\partial W_{ji}} \quad (5)$$

$$= \eta H_j (1 - H_j) X_i \sum_{i=1}^n W_{kj} (t_k - O_k) O_k (1 - O_k)$$

のようにして重みをアップデートして、誤差が一定値以下となるように学習をさせる。本研究においては教師データとして、実際に5分後に株価が、上がる、下がる、変わらない、のどのパターンに当てはまったかを学習させている。バックプロパゲーションの様子を図2に示す。

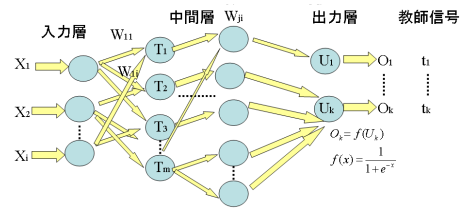


図 2: バックプロパゲーション様子

### 3.2 板情報の扱い方

まず、板情報を正規化する。板を各値段の高いほうから  $1, 2, \dots, 16$  とし、枚数のベクトル  $v = (v_1, v_2, \dots, v_{16})$  (8本値) のように表した上で、各枚数を正規化する。

正規化にあたってはニューラルネットワークの入力が  $-1$  から  $+1$  の範囲に収まっている必要があるため、以下のような手順で行う。正規化済み買い(売り)板に並んでいる数を  $p_{adj}$  とすると、

$$p_{adj} = \frac{p - \mu}{\sigma} \quad (6)$$

$p$  は買い(売り)板に並んでいる数、 $\mu$  はその時間における、買い(売り)板に並んでいる平均、 $\sigma$  はその時間における、買い(売り)板に並んでいる枚数の標準偏差である。 $-1$  から  $+1$  の範囲に抑えるために、 $-1$  以下の数は  $-1$  に、 $+1$  以上の数は  $+1$  に変換した。ニューラルネットワークの入力としては、このうち、価格形成に大きく影響を与えるであろう、買い気配 1-4、売り気配 1-4 の直近の価格に近い部分を採用した。

### 3.3 トレンド情報の扱い方

次に、板情報とは別に、トレンドの情報を入力に与える。 $n$ 分前から $n+1$ 分前 ( $n = 0, 1, 2, 3, 4$ ) の価格変化を見て、上昇していれば  $+1$  を、変化しなければ  $0$ 、下落している場合には  $-1$  をとしている。

### 3.4 寄付からの時間

株価の変動に関して、一日という時間の中においても、売買が集中する時間帯が存在したり、特定のパターンが生じるなど、何らかの特徴があると考えられる [6]。寄付付近は売買が集中するなど、傾向が見られると考え、寄付からの時間を  $-1$  から  $+1$  の範囲で表現し、パラメータの一つに加えた。実際の日中の株価と出来高の関係を図 3 に示す。

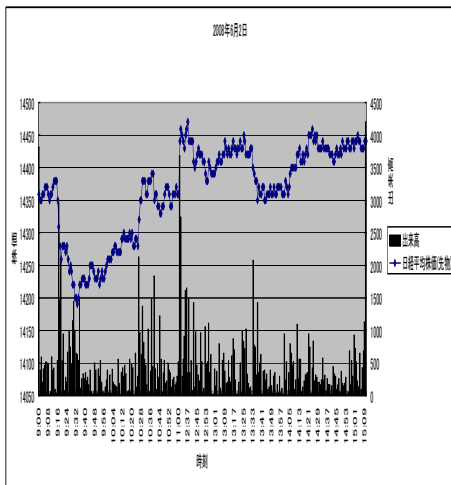


図 3: 日中の価格変動と出来高 (2008/06/02)

### 3.5 モデルの入出力

上述した変数を用いて、5 分後の株価指数を予測することを考える。入力としては表 2 を使用した。これらの値  $x = (x_1, x_2, \dots, x_{14})$  を入力にして、

$$y = f(x) \quad (7)$$

$y$  を出力とする ( $-1 \leq y \leq 1$ )。ただし、 $f$  はニューラルネットワークによる関数である。入力に対して、実際の正解 (5 分後に株価が上がるかどうか) データを出力にしている。5 分後に株価が 2 ティック以上上昇すれば、 $+1$  を 2 ティック以上下降すれば  $-1$  を、それ以外の場合には  $0$  としている。2 ティックとしているのは、1 ティックの上下は、直前に約定した値段によって生じてしまう可能性があるためである。ニューラルネット

ワークの中間層は 1 層とし、ニューロン数は 50 とした。ニューラルネットワークの実装としては、高速なライブラリである FANN[8] を用いた。

表 2: 各係数

係数	意味
$x_1$	売り気配 4
$x_2$	売り気配 3
$x_3$	売り気配 2
$x_4$	売り気配 1
$x_5$	買い気配 1
$x_6$	買い気配 2
$x_7$	買い気配 3
$x_8$	買い気配 4
$x_9$	0 分前から 1 分前の価格変化
$x_{10}$	1 分前から 2 分前の価格変化
$x_{11}$	2 分前から 3 分前の価格変化
$x_{12}$	3 分前から 4 分前の価格変化
$x_{13}$	4 分前から 5 分前の価格変化
$x_{14}$	寄付からの経過時間

### 3.6 トレーニング期間

ニューラルネットワークモデルのトレーニング期間を 2008 年 7 月から 9 月までの 3ヶ月間とした。5 分おきに、板のスナップショットを取っており、1 日当たり約 60 サンプルのデータ数が存在する。3ヶ月で 3717 サンプルを学習させた。

## 4 予測結果

表 3: トレーニング結果

教師信号	学習結果
全パターン総数	全パターン正解数
3717	3326(89.5%)
上がる	上がる正解数
479	264(55.1%)
変化なし	変化なし正解数
3067	3005(98.0%)
下がる	下がる正解数
171	57(33.3%)

表 4: テスト (未知データ予測) 結果

正解信号	予測
データ総数 1298	データ正解数 620(47.8%)
上がる 378	上がる正解数 47(12.4%)
変化なし 677	変化なし正解数 562(83.0%)
下がる 172	下がる正解数 13(7.6%)
正解信号	予測した個数 (不正解を含む)
上がる 378	個数 168
変化なし 677	変化なし正解数 1088
下がる 172	下がる正解数 43

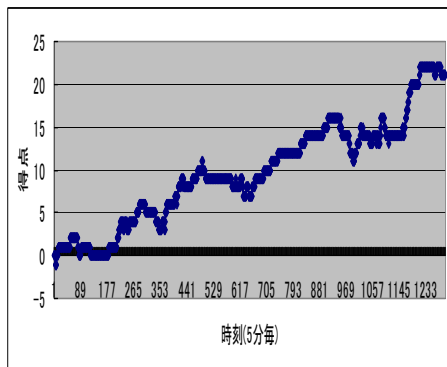


図 4: ニューラルネットワークの評価

実際にこのモデルを使用し、2008年10月の1ヶ月間の相場を予測した。結果は表3,4のようになった。この結果から、ニューラルネットワークモデルの予想は「変化なし」という予想を多く出したことが分かる。テストデータ(未知データ)に関して、正答率がそれほど高くないようにも感じるが、実際にこのシグナルを用いて売買を行ったときに有益な結果となっているのかを調べるために、簡単なテストを行ってみた。

シミュレーションとして、ニューラルネットワークが「上がる」と予想した場合に、実際に上がれば+1点を、変化なしならば0点を、下がってしまった場合には-1点とする。また、ニューラルネットワークが「変化なし」と予想した場合に、実際に上がった場合、変化なしの場合、下がってしまった場合全てに関して0点を、ニューラルネットワークが「下がる」と予想し

た場合に、実際に下がれば+1点を、変化なしならば0点を、上がってしまった場合には-1点とする。

この得点を時系列で累積させれば、ある程度、実際にこのシステムを使った場合に収益が上がりやすいかの目安になると考えられる。この方法に基づき、時間と点数の関係をプロットした結果が図4である。この結果、ニューラルネットワークモデルの予測力はある程度機能していることが確認できる。

## 5 ブラウン運動の初到達時刻分布

指値の約定する確率を計算するのに、ブラウン運動の性質を用いる。

### 5.1 ブラウン運動

ブラウン運動は、 $(\Omega, \mathcal{F}, P)$  を確率空間とし、各  $w \in \Omega$  に対して  $W(0) = 0$  を満たし、 $w$  に依存する連続関数  $W(t), t \geq 0$  が存在すると仮定する。このとき、 $W(t), t \geq 0$  がブラウン運動であるとは、全ての  $0 = t_0 < t_1 < \dots < t_m$  に対して、増分

$$W(t_1) - W(t_0), \dots, W(t_m) - W(t_{m-1}) \quad (8)$$

が独立で、これらの増分それぞれが正規分布で、以下の条件を満たす場合に言う。

$$E[W(t_{i+1}) - W(t_i)] = 0 \quad (9)$$

$$Var[W(t_{i+1}) - W(t_i)] = t_{i+1} - t_i \quad (10)$$

### 5.2 ブラウン運動の到達時刻

ブラウン運動に基づく確率過程

$$S(t) - S(0) = \sigma W(t) (\sigma \text{ はボラティリティ}) \quad (11)$$

とする。正のレベル(株価の変動で達するバリアの量)を  $m$  と時刻  $t$  を考える。時刻  $t$  以前にレベル  $m$  に到達するブラウン運動の経路(すなわち、レベル  $m$  への到達時刻  $\tau_m$  が  $t$  以下である経路)を考える。

これは、ブラウン運動の鏡像原理 [7] を用いると、確率変数  $\tau_m$  は累積分布関数

$$\begin{aligned} P\{\tau_m \leq t\} &= P\{\tau_m \leq t, S(t) \leq m\} \\ &+ P\{\tau_m \leq t, S(t) \geq m\} \\ &= 2P\{S(t) \geq m\} \\ &= \frac{2}{\sqrt{2\pi}} \int_{\frac{m}{\sigma\sqrt{t}}}^{\infty} e^{-\frac{x^2}{2}} dx \quad (12) \end{aligned}$$

に従い、確率密度関数は

$$f_{\tau_m}(t) = \frac{d}{dt} P\{\tau_m \leq t\} \quad (13)$$

$$= \frac{|m|}{\sigma t \sqrt{2\pi t}} e^{-\frac{m^2}{2\sigma^2 t}} \quad (14)$$

となる。この性質を利用して、指値が約定するという現象を、 $t$  分間における株価の変化  $S(t)$  として、レベル  $m$  に達するまでの確率として捉えることが本研究の手法であり、この確率は  $\frac{m}{\sigma}$  を与えれば一意に決定されるので、このパラメータを実測の株価データから推定することで求められる。この性質を元にした指値が約定する確率分布関数と累積分布関数をそれぞれ図 5,6 に示す。

### 5.3 ブラウン運動による初到達分布の確率密度関数と累積分布関数

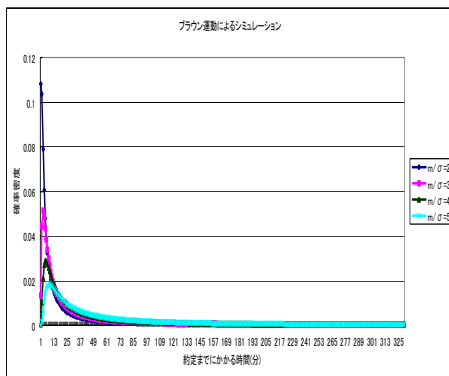


図 5: 確率密度関数

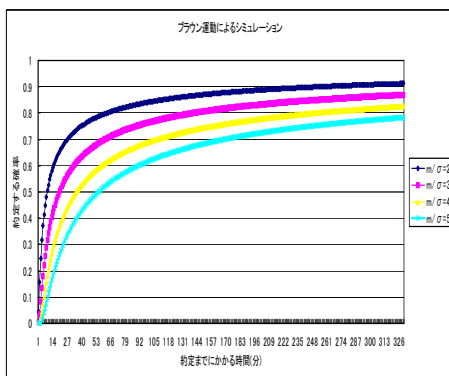


図 6: 累積分布関数

### 5.4 実測の株価データ

実際の板のデータを用いて、約定にかかる時間の分布を計測した。データは 2008 年 6 月から 2008 年 8 月までの毎日の 9:00 と 10:00 からスタートして、200 枚の板が食われるまでの時間を計測した。板は各気配ごと (買い気配 1,2,3,4,...) のように売り買い両方それぞれの板について計測している。図 7 は買い気配 3 における確率密度関数のグラフを、図 8 には、買い気配 1,2,3,4,5 の各注文が約定する確率の累積分布関数を与えたものである。330 分以降が非連続的なのは、一日のザラ場の時間が限られているからである。

### 5.5 理論と実測の比較

結果のグラフを見ると、形状はおおむね理論曲線と実際の曲線が似ていることが分かる。 $m/\sigma$  の値をどのように推測するかということが、課題ではあるが、このデータを元に、ユーザは 80% の確率で約定させるには何円のところに指値を何分間置けばよいかという意思決定が出来ると考えられる。

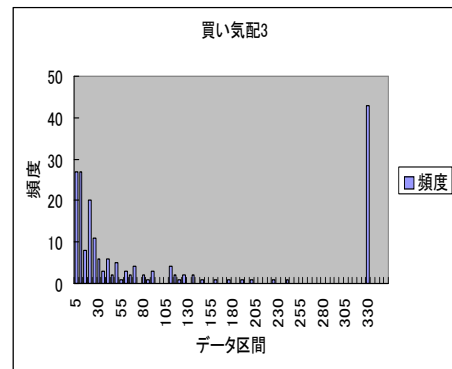


図 7: 買い気配 3

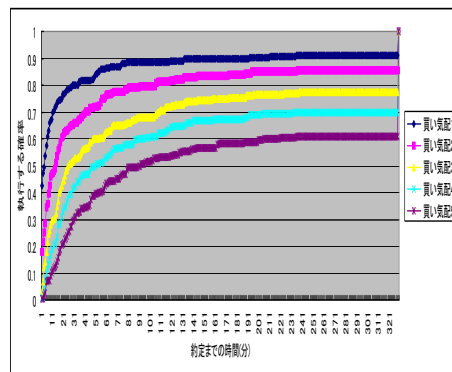


図 8: 累積分布関数 (実測買い気配)

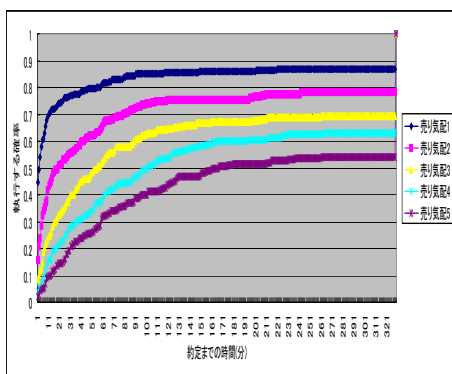


図 9: 累積分布関数 (実測売り気配)

## 6 むすび

本研究においては、ニューラルネットワークによる相場予測、ならびに指値の約定確率についての分析を行った。ニューラルネットワークによる相場予測に関して、ある程度予測力があるという結果が得られ、指値の執行確率を「何分でどれだけ確率で約定するか」という視点で数理モデルで表現した。

ニューラルネットワークの予測に関して、ある程度相場の予測能力があると考えられる。また、指値の約定確率に関して、パラメータをうまく推定できれば、実測のデータと理論曲線は非常に近いものになると考えられる。

機械学習や数理モデルを組み合わせることで、より市場の性質を深く知ることが出来、精度の高い予測も可能になっていくであろうと考えられる。近年のようなマーケット環境がより高速になっている中で、指値をどこに何分指せば良いかという問題に対してはこれからの関心が高まりつつあると考えられ、さらなる研究が必要である。

## 参考文献

- [1] 宇野淳, 価格はなぜ動くのか, 金融マーケットの謎を解き明かす, 日経 BP 社, 2008
- [2] 見並 良治, 久米川 昌弘, 尹 熙元 株式市場における売買注文状況と日中価格変動の解析, The 22nd Annual Conference of the Japanese Society for Artificial Intelligence, 2008
- [3] Yacine Ait-Sahalia, Per A. Mykland, Lan Zhang, Ultra High Frequency Volatility Estimation with Dependent Microstructure Noise. Working Paper, Princeton University, University of Chicago, and University of Illinois, 2006

- [4] Foucault, T., O. Kadan, E. Kandel. Limit order book as a market for liquidity. *Review of Financial Studies* 18(4) 1171-1217, 2005
- [5] Rama Cont, Sasha Stoikov, Rishi Talreja, A Stochastic Model for Order Book Dynamics, Working Paper, SSRN, 2008
- [6] Ryan Garvey, Fei Wu, Intraday time and order execution quality dimensions, Volume 12, Issue 2, Pages 203-228, 2009
- [7] S.E. シュリーヴ (著), 今井 達也 (翻訳), 河野 祐一 (翻訳), 田中 久充 (翻訳), 長山 いづみ (翻訳), ファイナンスのための確率解析, シュプリンガー
- [8] Fast Artificial Neural Network Library (FANN) <http://leenissen.dk/fann/>