

進化型ニューラルネットワークとフル板情報による株価変動の分析

熊添 勇人[†] 藤田 悟[†]法政大学 情報科学部[†]

1. まえがき

株式市場のデータとして板情報と呼ばれる各気配値に対する注文情報を表形式にまとめたデータが提供されている。以前は、最良気配値から上下 8 本の範囲にある株の売買数量のみしか確認できなかったが、フル板情報が提供され始めてからは、全ての気配値に対する注文情報が確認できるようになった。板情報を分析した関連研究には、過去の板情報が短期的に将来の株価へ影響を及ぼすことが示されている[2]。さらに、フル板情報を利用した関連研究では、投資家の注文行動や市場特性の抽出などが行われており、フル板情報には有用な情報が存在する可能性が示されている[1]。

本稿では板情報より多くの情報を含むフル板情報から抽出した特徴を用いて株価変動をニューラルネットワークと進化型ニューラルネットワークで予測し、過去のフル板情報から将来の株価変動へどのような影響があるかを調査した。

2. フル板情報の特徴抽出

本研究ではフル板情報が将来に及ぼす短期的影響を調べるために情報が更新される度に特徴抽出を行う。そして、大量に得られた特徴量から無作為に一定量を選択して予測に利用する。

フル板情報から特徴抽出を行うにあたり、板情報の各要素をニューラルネットワークの汎化性能を高めるために正規化する。本稿では、フル板情報の売り板の数量や件数、買い板の数量や件数のそれぞれの要素の総和が 1 に正規化している。これは、売りと買いの比率を用いることで、フル板情報を利用して投資家が取引する際に注目するポイントの 1 つである板の厚みの特徴とできると考えたからである。以上の議論を踏まえて、現在の株価 x_1 ・総売買数量の比率 x_2 ・総売買件数の比率 x_3 ・最良気配から 9 本目以降の総売数量 x_4 ・最良気配から 8 本分の各売数量 $x_{5\sim 12}$ ・最良気配から 8 本分の各買数量 $x_{13\sim 20}$ ・最良気配から 9 本目以降の総買数量 x_{21} ・最良気配から 9 本目以降の総売件数 x_{22} ・最良気配から 8 本分の各売件数 $x_{23\sim 30}$ ・最良気配から 8 本分の各買件数 $x_{31\sim 38}$ ・最良気配から 9 本目以降の総買件数 x_{39} を特徴としてフル板情報から抽出する。この特徴量では、

9 本目以降は大きくひとくくりにまとめ、最良気配に近い部分は変動への関連が強いと考え、1 本毎に利用している。

現在の株価 x_1 は式(1)で得た値である。これは、株価の変動に制限が付いていることを利用して、現在の株価をその日に変動できる下限ならば-1、上限ならば1となるようにしたものである。

$$x_1 = \frac{\text{price-base}}{|\text{cap}|} \quad (1)$$

この時、*price*は現在の株価であり、*base*は基準値段（前日の終値）を表し、*cap*は制限値幅を表す。

総売買数量の比率 x_2 は式(2)で与えられる。*amount_{bid}*はこの時の全気配値に対する売数量の総和を、*amount_{ask}*は同様に買数量の総和を表す。

$$x_2 = \begin{cases} -\frac{\text{amount}_{bid}}{\text{amount}_{ask}} & (\text{amount}_{bid} < \text{amount}_{ask}) \\ \frac{\text{amount}_{ask}}{\text{amount}_{bid}} & (\text{amount}_{bid} \geq \text{amount}_{ask}) \end{cases} \quad (2)$$

総売買件数の比率 x_3 は式(2)の売買数量の総和を売買件数の総和に変更した式で得ることができるため、省略とする。 x_2, x_3 の値は、より厚い板を基準としてもう一方がどれだけの厚さであるかを表す。さらに、売り板の方が薄い場合は負の値として表す要素である。

次に、売数量の特徴 $x_{4\sim 12}$ について述べる。最良気配から 9 本目以降の総売数量 x_4 は式(3)で与えられる。

$$x_4 = -\frac{\sum_{i=9}^N \text{amount}_{bid}^i}{\text{amount}_{bid}} \quad (3)$$

*amount_{bid}ⁱ*は最良気配から*i*本目の売数量を意味する。この要素では、総売数量に対して 9 本目以降に存在する売数量の総和がどれほどの比率かを示す。また、負の値になっているのは、買いを正の値に売りを負の値にすることで明確に区別するためである。

最良気配から 8 本分の各売数量 $x_{5\sim 12}$ は式(4)で算出する。式(4)において、*amount_{bid}^N*は最良気配値から*N*本目の売数量を表している。

$$x_{5\sim 12} = -\frac{\text{amount}_{bid}^N}{\text{amount}_{bid}} \quad (4)$$

次に、買数量の特徴 $x_{13\sim 21}$ の特徴について述べる。最良気配から 8 本分の各買数量 $x_{13\sim 20}$ は式(4)の売数量を買数量に変更し、符号を正にして算出する。同様に、最良気配から 9 本目以降の総買数量 x_{21} は式(3)

The Analysis of stock price fluctuation with evolving neural network

[†] Yuto Kumazoe, Satoru Fujita

[†] Faculty of C.I.S, Hosei University

の売数量を買数量に変更して符号を正にすることで求める。

売買件数に関する特徴は、其々の「件数」部分を「数量」に読み替えた特徴と対応している。そして、式(3)(4)の数量を件数に変更して算出を行う。

以上の方法で得た値をまとめて、39次元の特徴を構成した。

3. 予測実験

本章では、2章で述べた特徴抽出手法を用いて得た特徴ベクトルを入力とし、NNと進化型ニューラルネットワークの1つであるNEAT[3]で予測実験を行った。また、時系列の影響を調べるために、時刻tの特徴・時刻tとt-1の特徴を結合したもの・時刻t-t-2の特徴を結合したものの3つを入力として実験を行なった。教師データには、次状態の板がどう変化するかという次の3種を利用した。

- 最良気配値の境界：上昇・下落・不変
 - 株価の変動方向：高騰・下落・不変
 - 注文位置
 - ▶ 最良売気配値以上の気配値に対する売注文
 - ▶ 最良売気配値未満の気配値に対する売注文
 - ▶ 最良買気配値以下の気配値に対する買注文
 - ▶ 最良買気配値を超過する気配値に対する買注文
- 本実験は、銘柄番号6502,6701,6702の3つを対象として、2010年1月5日~2011年6月27日のデータを教師データの要素毎に10000件ずつ無作為に選択して特徴抽出に利用した。NNの実験において、層構造(数値はノード数)は、入力層・512・1024・1024・256・出力層のような6層で全結合のニューラルネットワークを利用した。また、NNのエポック数は3000、NEATの世代数は5000とした。

表1. 教師データ「最良気配の境界」の予測精度(%)

銘柄	特徴結合数	NN TRAIN	NN TEST	NEAT TRAIN	NEAT TEST
6502	1	98.97	63.72	64.48	65.76
	2	99.63	67.79	64.32	63.92
	3	99.76	71.39	54.08	53.84
6701	1	98.17	70.75	69.40	69.04
	2	99.16	73.87	67.33	67.17
	3	99.27	75.47	53.17	52.08
6702	1	99.81	56.59	60.54	60.56
	2	99.97	61.04	57.16	56.85
	3	99.94	63.35	58.30	58.53

表2. 教師データ「株価の変動方向」の予測精度(%)

銘柄	特徴結合数	NN TRAIN	NN TEST	NEAT TRAIN	NEAT TEST
6502	1	83.64	38.71	39.19	37.77
	2	99.03	40.68	38.04	37.37
	3	99.21	39.91	37.10	35.31
6701	1	91.78	53.41	41.52	40.55
	2	96.10	53.48	39.05	39.28
	3	97.32	51.45	38.93	38.95
6702	1	94.36	37.75	39.55	39.35
	2	98.06	36.59	38.36	37.52
	3	99.31	37.12	35.43	35.75

表3. 教師データ「注文位置」の予測精度(%)

銘柄	特徴結合数	NN TRAIN	NN TEST	NEAT TRAIN	NEAT TEST
6502	1	92.07	30.77	33.01	32.96
	2	97.03	29.91	33.49	32.42
	3	98.96	29.71	29.65	29.91
6701	1	91.02	30.23	32.16	32.14
	2	95.13	31.27	31.51	31.24
	3	95.98	31.53	31.69	31.11
6702	1	99.35	30.24	31.30	31.73
	2	99.63	30.94	32.50	32.03
	3	99.83	30.85	33.50	33.20

以上の条件で行った実験結果を表1, 表2, 表3に示す。

4. 考察

各実験結果を見ると、3種類の教師データに対するテストデータの予測精度は、最良気配の境界>株価の変動方向>注文位置となった。株価の変動方向というのは直近の約定価格に基づくため、板の状態があまり変わらずに株価の上昇・下降判定が行われる。一方、最良気配の境界が上昇・下降するときには最良気配値が変化している。この板の状態変化が識別結果に影響を与えたと推測している。さらに、時系列方向に特徴量を結合した場合、最良気配の境界の予測精度は向上したが、株価の変動方向では著しい変化はなかった。これも先に述べた板の変化と同様の影響と考えている。

注文位置の予測精度が最も低い原因は、投資家が戦略的に売買したり、板を見ないで取引したりする場合もあるからではないかと考えている。しかし、4値分類を無作為に行った時の25%より5%程、分類精度が良くなっているのは、投資家の取引戦略を傾向として捉えたからであると予測している。

5. むすび

本稿では、板の厚みに着目してフル板情報から特徴抽出を行った。そして、抽出した特徴と3種類の教師データを用いて予測実験を行なった。3種類の教師データに対する精度は、最良気配の境界>株価>注文位置となった。また、フル板情報を用いることで、株価の変動や投資家の戦略的行動が抽出できる可能性を示した。

文献

- [1] 石山 武人, 藤田 悟, “クラスタリング手法を用いた株式市場におけるフル板情報の分析”, 情報処理学会第74回全国大会, 4M-2, 2012年.
- [2] 見並 良治, 久米川 昌弘, 尹 熙元, “株式市場に置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, 1E2-2, 2008.
- [3] Kenneth O.Stanley, “Evolving Neural Networks through Augmenting Topologies”, Evolutionary Computation, Vol. 10, No. 2, pp. 99-127, 2002.