4L-5

Twitterを用いた株式市場の変動予測

長尾　将宏†　　長尾　智晴†
†横浜国立大学　大学院環境情報学府

1　まえがき

近年テキストマイニングの技術を用いて株価を予測する研究が盛んに行われている．従来研究では，新聞の記事，金融経済月報，Twitterなどの情報を用いるものが提案されている．

金融経済月報を用いた研究[1]では，含まれる単語の出現頻度や共起関係を元に，重回帰分析を用いて変動の要因となる特徴量を分析している．金融経済月報は月ごとに公表されるため，翌日の株価を予測するといった短期的な予測には向いていないが，景気の長期的な分析に有効であることを示した．

Twitterを用いた研究[2]では，Twitter上の感情表現を含む英文ツイートから作成した感情モデルが社会的な出来事と関連性があることが示している．気分に関係する語彙として，POMSから得られたそれぞれ「Calm」，「Alert」，「Sure」，「Vital」，「Kind」，「Happy」に属する72の語彙と，POMSと共起した一般のWebページから抽出された4,5-gramの語彙を使用した．予測する際には，気分に関係する語彙と感情表現（“I feel”など）を含むツイート数の推移と，ダウ平均株価を自己組織化ファジィニューラルネットワークに入力した．15日間のダウ平均株価の変動を予測した結果，ダウ平均株価の値のみを用いた場合は73.3％の精度だったのに対し，ダウ平均株価の値と「Calm」の語彙を利用した場合は86.7％の精度で予測することができた．同様にして，Twitter上のツイート情報を用いて短期的に株価を予測することができるとする研究がいくつか存在する．[3][4]

本研究では，Twitter上の投資家によるツイートを集めることで株価を予測することを目標とする．

2　提案手法

Twitterから取得できるツイートにおいて，株価予測に対してより影響力があるのは投資家によるツイートであると考えられる．その中でも特に重要なのは，これからの値動きを予想するようなツイートであり，本稿では予想ツイートと呼ぶ．予想ツイートは「上昇（下降）関係の単語」および「予想関係の単語」を含むと仮定し，抽出を行った．また，本稿における投資家とは,「日経平均」を含むツイートをしたユーザである．各ユーザの過去のツイートにおける予想ツイート数により，ユーザごとの信頼度を算出し，この影響を調べた．

2.1　抽出方法

Step1:「上昇（下降）関係の単語」の抽出

それぞれのツイートを形態素解析し，一日ごとにそれぞれの単語の頻度を求めた．その頻度と日経平均株価のデータ列N日分について，相関係数を出した．相関係数0.2以上ならば上昇関係の単語（表1）とし，相関係数-0.2以下ならば下降関係の単語（表2）とした．

上昇関係の単語	相関係数
上がる	0.589
銭高	0.530
上げ	0.475
上げ幅	0.446
あがる	0.426

表1: 上昇関係の単語（上位5単語）

下降関係の単語	相関係数
下げる	-0.684
下げ幅	-0.677
下落	-0.600
下がる	-0.573
下げ	-0.531

表2: 下降関係の単語（上位5単語）

Step2:「予想関係の語彙」の抽出

それぞれのツイートを形態素解析した結果から，出現頻度の高かった予想関係の語彙，計43語彙（表3）を抽出した．

Step3: 信頼度の算出

各ユーザの過去のツイートにおける予想ツイート

Using Twitter To Predict Stock Market Movements
†Masahiro Nagao　†Tomoharu Nagao
†Graduate School of Environment and Information Sciences, Yokohama National University

期待感	らしい
でしょう	ようです
してくる	だろう
あり得る	なんとなく
予感	考えられる
思う	かな

表 3: 予想関係の語彙（一部）

数を出し，これを信頼度とする．あるユーザ i の信頼度 w_i を次式で定義する．

$$w_i = 1 + Count_{predict}$$

$Count_{predict}$: 予想 $Tweet$ の数

3 実験

「日経平均」を含むツイートを取得し，そこから予想ツイートを抽出した．ツイートの取得期間は 2013/01/01 から 2013/11/26 である．4〜9 日遅らせた予想ツイートの割合と，日経平均株価変化率の間に相関があるかを，グレンジャー因果性テストにより調べた．第 i 日目の予想ツイートの割合 r_i を次式で定義する．

$$r_i = \frac{\sum R_{up} + \sum R_{down}}{Count_{tweet}}$$

$R_{up}(> 0)$: $Tweet$ 中の上昇単語の相関係数

$R_{down}(< 0)$: $Tweet$ 中の下降単語の相関係数

$Count_{tweet}$: 予想 $Tweet$ 数

また，予測が可能であるかを方向的中率により評価した．表 4,5 は信頼度を考慮しない場合と，信頼度を考慮した場合のグレンジャー因果性テストと方向的中率である．これより，特に 5〜6 日遅らせた予想 Tweet の割合と日経平均株価変化率の間に，関係性がないという帰無仮説は有意水準 1 %で棄却された．すなわち，5〜6 日遅らせた予想ツイートの割合と日経平均株価変化率の間には，語彙の抽出期間について相関があることが示された．方向的中率については，実験期間について，60 %程度で 5〜6 日後の日経平均株価の方向性を的中させることができた．また，信頼度を考慮する場合の方が，良好な方向的中率を示すことがわかった．

遅らせた日数	グレンジャー因果	方向的中率
4	$9.18*10^{-2}$ （p ＜.01)	51.97 %
5	$1.32*10^{-4}$ （p ＜.01)	57.89 %
6	$6.38*10^{-5}$ （p ＜.01)	64.47 %
7	$1.92*10^{-2}$ （p ＜.05)	65.79 %
8	$7.96*10^{-1}$	63.16 %
9	$4.76*10^{-1}$	59.21 %

表 4: 信頼度を考慮しない場合のグレンジャー因果性テストと方向的中率の結果

遅らせた日数	グレンジャー因果	方向的中率
4	$1.33*10^{-2}$ （p ＜.05)	51.32 %
5	$1.95*10^{-4}$ （p ＜.01)	57.23 %
6	$7.74*10^{-5}$ （p ＜.01)	65.13 %
7	$2.79*10^{-2}$ （p ＜.05)	66.44 %
8	$5.62*10^{-1}$	63.82 %
9	$5.90*10^{-2}$	59.87 %

表 5: 信頼度を考慮した場合のグレンジャー因果性テストと方向的中率の結果

4 まとめ

本発表では，予想ツイートの割合と日経平均株価変化率に相関があることを示した．また，ユーザの過去のツイートから得られた信頼度を考慮することで，わずかではあるが精度を向上できることが確認できた．しかしながらまだ優位性があると言える結果ではないので，今後さらに精度を上げるべく実験中である．

参考文献

[1] 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 25(3), 383-387, 2010

[2] Johan Bollen, Huina Maoa, Xiaojun Zengb.“Twitter mood predicts the stock market”. Journal of Computational Science, Volume 2, Issue 1, March 2011, Pages 18

[3] 迫村光秋, 和泉潔, セーヨー・サンティ.“Twitter のテキストとネットワークの解析による経済動向分析”. 第 10 回 人工知能学会研究会資料.

[4] 桃井 達明, 須鎗 弘樹.“Twitter から生成した感情モデルと社会経済的現象との相関”. 情報科学技術フォーラム講演論文集 11(1), 127-130, 2012-09-04.