

# Twitterを用いた 株価予測

富山県立大学情報システム工学専攻  
1855001 麻生 到

指導教員：奥原浩之

## 1 はじめに

現在、ツイッターなどのマイクロブログには、様々なニュースやそれに対する人々の反応が書かれており、その情報量は膨大かつ、増加し続けている。この膨大な情報を実世界の動きを観測するためのソーシャルセンサーとして利用する研究の数は増加しており、観測する対象を予め設定し、それについて詳細な分析を行ったものが多く見られる。特に、経済動向を分析対象としたものとして、ツイッターからキーワードを用いて株式に関する情報を収集し、株価動向との関連の分析に取り組んだ事例があるなど、ツイッター情報は経済動向の分析に大いに用いられている。

既存研究としては、Bollenらが、ツイートを対象にOpinionFinder(OF)とGoogle-Profile of Mood Statesを用いて、「calmness」などの6つの心的状態を表す指数を抽出し、ダウ平均株価の予測を行った[1]。しかし、分析対象となるツイートは“I feel”, “I’m”といった心的状態を明言したものに限られていることに加えて、ツイート情報はダウ平均株価の過去の数値データによる予測を補うものとして用いられている。

本研究では、日本語のツイートに対して感情分析を行い株価の予測を行う。

## 2 データマイニング

### 2.1 Twitter API

本研究では、Twitterが提供しているAPIを用いてツイートの取得を行った。Twitter APIから取得できる情報は、投稿日やツイート、お気に入り数、リツイート数などを得ることができる。また、取得できるツイートの期限が決まっており1週間前までのツイートしか取得することができない。

実際に、“任天堂”をキーワードとして取得したツイートの例を図1に示す。

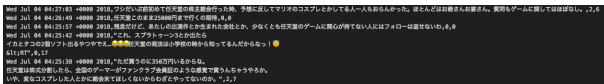


図 1: ツイートの例

Twitter APIの制限情報として、1週間前までのツイートしか取得できないことやアクセス回数が15分間に180回までであり、1回のアクセスにつき200件までしか取得することができないことが挙げられる。

### 2.2 感情分析

本研究では、東工大の高村教授が作って公開している辞書であるPN Tableを使用した。この感情辞書では、各電子のスピンは上向きと下向きのどちらかの値をとり、隣り合ったスピンは同じ値をとりやすいといったことを応用した。各単語を電子と見なし、単語の感情極性をスピンの向きと見なし、関連する単語ペアを連結することにより語彙ネットワークを構築している[2]。PN Tableの一部の例を図2に示す。

各単語には、-1 1のスコアが付けられている。ポジティブに近い単語ほどスコアが大きくなり、ネガティブに近い単語ほどスコアが小さくなるように定義されている。

### 2.3 分析

図3のように1つのツイートに対して、形態素解析を行い品詞ごとに分割する。そして、分割した単語をPN Tableと照らし合わせてスコア付けを行っていく。この時、辞書に含まれていない単語のスコアは0とした。

1日分のツイートのスコアは、その日のすべてのツイートのスコアの合計をその日のツイート数で割った値とした。

## 4 実験



良い:よい:形容詞:0.999995
喜ぶ:よるこぶ:動詞:0.999979
喜める:ほめる:動詞:0.999979
めでたい:めでたい:形容詞:0.999645
實い:かしこい:形容詞:0.999486
善い:いい:形容詞:0.999314
適す:てきす:動詞:0.999295
天晴:あっぱれ:名詞:0.999267
祝う:いわう:動詞:0.999122
功績:こうせき:名詞:0.999104
實:しよ:名詞:0.998943
嬉しい:うれしい:形容詞:0.998871
喜び:よるこび:名詞:0.998861
才知:さいち:名詞:0.998771
徳:とく:名詞:0.998745
才能:さいのう:名詞:0.998699
素晴らしい:すばらしい:形容詞:0.998617
芳しい:かんばしい:形容詞:0.998578
称える:たたえる:動詞:0.998576
適切:てきせつ:名詞:0.998406
喜める:あがめる:動詞:0.998357
助ける:たすける:動詞:0.998356

図 2: PN Table の例



図 3: 分析の流れ

キーワードを任天堂とし、7月2日から7月8日の合計2万ツイートを分析対象とした。取得したデータに対してそれぞれ日ごとに感情分析を行った。その結果を図4に示す。

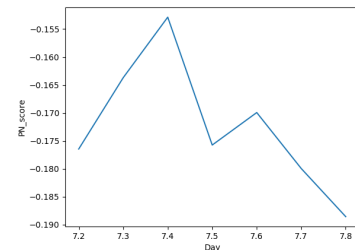


図 4: 感情分析の結果

結果として、この1週間では大きな変動は見られなかった。期間を延ばして考察する必要があると考えられる。

## 5 おわりに

日経平均株価を予測するにあたり、ツイートを取得する条件を定めなければならない。今回は任天堂で分析を行ったが日経平均株価の予測には適していないので早急に決める必要がある。また、感情分析を行う際ツイートの中に辞書に含まれていない単語が多々あったため検討する必要がある。

## 参考文献

- [1] Johan Bollen, Huina Mao, Xiao-Jun Zeng, “Twitter mood predicts the stock market.”, arXiv. 1010.3003v1. cs.CE, 14, Oct, 2010.
- [2] 高村 大也, 乾 孝司, 奥村 学, “スピンモデルによる単語の感情極性抽出, 情報学会論文誌, Vol. 47, No. 2, Feb. 2006.