

Twitter を用いた為替予測

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(O と Google-Profile of Mood States を用いて、「calmess」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。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定さ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。

実際に、取得したツイートの例を図1に示す。

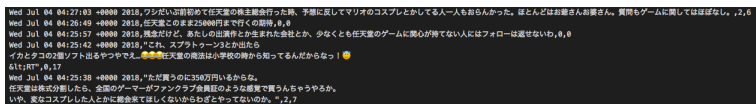


図 1: 取得したツイートの例

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 感情辞書

感情分析を行う多くの場合、感情辞書を用いることが多い。既存の感情辞書として、東工大の高村教授が公開している“PN Table”や東北大の乾・岡崎研究室が公開している“日本語評価極性辞書”，Yahoo!JAPAN研究所の鍛冶伸裕さんという方が作られた“Polar Phrase Dictionary”がある。

これらの既存の感情辞書では、単語ごとにスコアリングがしており、+1に近いほどその単語はポジティブであり、-1に近い単語ほどネガティブな単語であるといった性質を持つ。極性のスコアの計算は、重み $\sum t_w$ と単語の頻度 $\sum t_{freq}$ を用いて以下の式を用いて行う。

$$score = \sum t_w \times \log(t_{freq}) \quad (1)$$

また、辞書を作成するための基となる文章は本や新聞などのきっちりとした文章を用いている。

3 感情分析

一般的な感情分析では、文章に対して既存の感情辞書を用いて評価する。分析では、文章を形態素解析で品詞ごとに分けた後、品詞ごとに感情辞書と照らし合わせてスコアリングしていく。

しかし、これらの感情辞書は極性辞書なのでポジ・ネガの1つの感情軸しか持っていない。そのため、本研究で扱う Twitter のような話し言葉の分析を行う場合、感情軸が1つだけであると文章を評価が不十分になると考えられる。

また、辞書を作成するための基となる文章は本や新聞などのきっちりとした文章を用いていることから既存の感情辞書は話し言葉の分析を想定していないと示唆される。

本研究では、感情軸を増やしツイートのような話し言葉の分析を想定した感情辞書を作成した。Word2Vecを用いて文章から辞書モデルを作成し、感情辞書とした。話し言葉を想定しているため、辞書を作るための基の文章は楽天とAmazonのレビューを用いた。

また、ポジネガのみでは文章の評価が制限されてしまうく感情の評価を行うことが困難であると考えられるため感情軸を増やした。先行研究として、心理学者 Robert Plutick が考案した感情の環を元に感情を選択した研究がある [1]。Robert Plutick が考案した感情の環は、「喜び」、「悲しみ」、「嫌悪」、「憧れ」、「怒り」、「恐怖」、「驚き」、「期待」の8の感情軸から構成されている。しかし、先行研究では「喜び」、「悲しみ」、「怒り」、「恐怖」、「驚き」、「期待」の6つ感情を採用している。「嫌悪」と「信頼」の軸は隣接する感情に似たような性質を持っているため排除されていた。

本研究でも、「喜び」、「悲しみ」、「怒り」、「恐怖」、「驚き」、「期待」の6つ感情を採用し、評価した。

4 実験

実際に作成した辞書を用いて以下の取得したツイートの評価を行った。

Case1：好きな曲を自分で弾けるのは楽しいですよ～♪

Case2： そういや任天堂からメール全然来ない

それぞれのツイートの評価を示した図を図1と図2に示す。

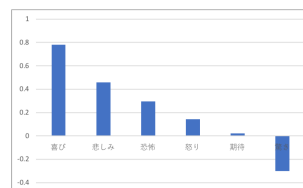


Fig. 1: Case1

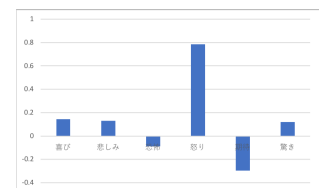


Fig. 2: Case2

否定語が反映されているかを確認するため、以下の文章の評価を図3と図4に示す。

Case3：野球は楽しい

Case4：野球は楽しい

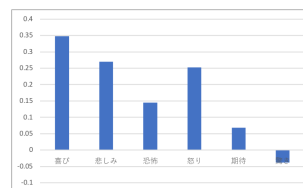


Fig. 3: Case3

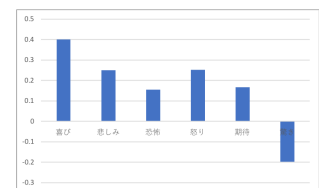


Fig. 4: Case4

評価結果より、否定語が全く反映されていないことがわかった。そのため、否定語を含んだ場合、スコアを逆転させる特別なルールを追加する必要がある。

参考文献

- [1] 三井健史, 伊藤智也, 中西勇人, 濱川礼, SNS の投稿を用いた感情記録ライフログシステム, 情報処理学会研究報告, Vol. 2014-EC-32 No. 1, pp.1-6, 2014.6.6.