

# 経済に関するオルタナティブ・データを考慮した金融マーケット予測手法の開発

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標を抽出し、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、トランプ大統領のツイートを取得した。取得したツイートの例を図 1 に示す。

```
Sun Jul 29 19:28:39 +0000 2018 "...Also, why is Mueller only appointing Angry Docs, some of whom have worked for Crooked Hillary, others, including: https://t.co/093207091d", 18486, 6267
Sun Jul 29 19:35:15 +0000 2018 "There is No Collusion! The Robert Mueller Rigged Witch Hunt, headed now by 17 (increased from 13, including an Obam. https://t.co/mtUqvuvu2D", 23319, 83236
Sun Jul 29 19:39:19 +0000 2018 "...and the Amazon Washington Post do nothing but write bad stories even on very positive achievements – and they will never change!", 11189, 4732
Sun Jul 29 19:40:18 +0000 2018 "...and the Amazon Washington Post do nothing but write bad stories even on very positive achievements – and they will never change!", 11189, 48385
Sun Jul 29 19:49:18 +0000 2018 "..., unfortunately, 40% of media coverage of my Administration is negative, despite the tremendously positive results we...", 16261, 29369
Sun Jul 29 19:59:17 +0000 2018 "when the media – driven insane by their Trump Derangement Syndrome – reveals internal deliberations of our government. https://t.co/0rad50hWek", 16533, 64139
```

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。

本研究では、為替の変動に影響しているツイートを抽出することで為替変動への効果的なアプローチを試みる。例として図 2 と図 3 に為替に影響していると考えられるツイートと影響していないと考えられるツイートを示す。

## 3 センチメント分析

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。本研究では、為替の情報を扱うため日本語の感情辞書だけでは不十分と考えられる。そのため、英語版の感情辞書の作成も行った。

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。



図 2: トランプ大統領の tweet の採用例



図 3: トランプ大統領の tweet の不採用例

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。

そして、為替の変動に影響していると思われるツイートは経済専門用語などが含まれていると想定されるため辞書の中にそのような用語がないと未知語として処理される。よって、辞書の作成の際には経済関連の文章を用いる。

また、ポジネガのみでは文章の評価が制限されてしまいうまく感情の評価を行な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図 4 のように「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。

本研究でも、「喜び」、「悲しみ」、「嫌悪」、「憧れ」、「怒り」、「恐怖」、「驚き」、「期待」の 8 つの感情を採用し、評価した。



Fig. 4: Plutick の感情の環

## 4 クラスタリング

為替（ドル円）が上がるか下がるか変化なしといったラベル付けし、3 クラスのクラスタリングを行う。クラスタリングを行う際の入力としては、Plutick の感情の環に基づいた 8 つの感情スコアを使用する。

ある感情  $c$  の 1 ツイートあたりの感情スコア  $s_i^{(c)}$  を以下に示す。

$$s^{(c)} = \frac{\sum_{i=0}^L x_i^{(c)}}{L} \quad (1)$$

$x_i^{(c)}$ : 感情  $i$  のスコア,  $L$ : 分析対象となる 1 ツイートあたりの単語数

そして、1日あたりのスコア  $S_i$  を以下に示す

$$S^{(c)} = \frac{\sum_{n=0}^N S_n^{(c)}}{N} \quad (2)$$

$N$ :1日あたりのツイート数

また、

$$S = \{S^{(c)} | c = 0, 1, \dots, 8\} \quad (3)$$

$M$  日間ツイートを取得した場合のデータセットは、

$$S_{(z)}^{(c)} \equiv \{S_m^{(c)} \in \mathbb{R}^D | (z_m) = (z), m = 0, 1, \dots, M-1\} \quad (4)$$

ここで、 $z$  は為替の変化状態を表す

為替の変化状態  $z$  が与えられたとき、感情スコア  $\{s^{(c)}\}$  が算出されたことを支持する度合いは以下のように示すことができる。

$$\begin{aligned} P(z|S, \{s^{(c)}\}) &\propto \prod_{c=0}^{C-1} \int \int \mathcal{N}_D(s^{(c)} | \mu_{(z)}^{(c)}, (\Lambda_{(z)}^{(c)})^{-1}) \times Q(\mu_{(z)}^{(c)}, \Lambda_{(z)}^{(c)} | X^{(c)}) d\mu_{(z)}^{(c)} d\Lambda_{(z)}^{(c)} \\ &= S_D(s^{(c)} | m^*, Q^*, f^*) \end{aligned} \quad (5)$$

$$m^* = m_{(e,z)}^{(c)} \quad (6)$$

$$Q^* = (\nu_{(e,z)}^{(c)} - D + 1) \frac{\beta_{(e,z)}^{(c)}}{\beta_{(e,z)}^{(c)} + 1} W_{(e,z)}^{(c)} \quad (7)$$

$$f^* = \nu_{(e,z)}^{(c)} - D + 1 \quad (8)$$

(9)

ここで、 $\mathcal{N}_D(s^{(c)} | \mu_{(z)}^{(c)}, (\Lambda_{(z)}^{(c)})^{-1})$  は  $s^{(c)}$  の平均ベクトル  $\mu$  と分散共分散行列  $\Lambda^{-1}$  の  $D$  変量ガウス分布であり、  $Q(\mu_{(z)}^{(c)}, \Lambda_{(z)}^{(c)} | X^{(c)})$  は事後分布である。

自動売買システムにおいて買う・売る、動かないの判断は式 (5) を用いて以下のように求める

$$z_1^* = \arg \max_{z \in \mathbb{T}_M} P(z|S, \{s^{(c)}\}) \quad (10)$$

$z_1^*$  を投資判断に活用する

## 5 為替予測

抽出したツイートから辞書を用いた求めたスコアを入力として為替を予測する。従来研究では1日単位での長期予測を行っていた。

長期予測では多くのツイートを対象として分析できるため、予測を行いやすい。本研究では、Twitter を用いているためリアルタイム性を活用して短期予測を行えればいいと考えている。しかし、トランプ大統領など多数のユーザを対象とするならば1日単位で考慮する必要がある

実際に、Fig. 5 のような流れで為替予測を行う。

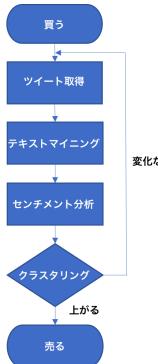


Fig. 5: 自動売買の流れ

## 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.