

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 指導教員：奥原浩之
1855001 麻生 到

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、「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。

```
# OANDAのデモ口座へのAPI接続
api = API(access_token=access_token, environment="practice")

# unitsの+は買い、-は売り
data_buy = {
    "order": {
        "instrument": "USD_JPY",
        "units": "+10000",
        "type": "MARKET",
        "positionFill": "DEFAULT"
    }
}

data_sell = {
    "order": {
        "instrument": "USD_JPY",
        "units": "-10000",
        "type": "MARKET",
        "positionFill": "DEFAULT"
    }
}

#API経由で成行注文を実行
r = orders.OrderCreate(accountID, data=data_buy)
```

図 2: 取引を行うソースコード

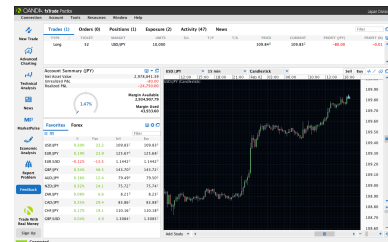


図 3: oanda の GUI の画面

2 データマイニング

2.1 cron を用いたツイート自動取得

本研究では、cron を用いてツイートを定期的に自動取得する。cron は、ジョブを自動実行するためのデモンプロセスである。

図 1 に cron の処理画面を示す。このような処理を行うことで 1 週間に 1 度自動的にツイートを取得することができる。

```
mas_code — vi • crontab — 80x24
PYTHONIOENCODING = 'utf-8'
LANG=ja_JP.UTF-8
59 10 * * 2 /Users/beta/.pyenv/shims/python /Users/beta/Documents/mas_code/get_tweet_acc.py
~
~
```

図 1: cron の処理画面

3 為替取引

3.1 OandaAPI を用いて為替取引

Oanda API では、相場情報の取得とトレーディングが可能である。相場の情報は、90 以上の通貨ペアのリアルタイム為替レートを取得し、24 時間を替相場の状況をリアルタイムに監視が可能。また、過去レートを 10 年間以上遡ってアクセスすることができる。また、トレーディングではアカウント情報、残高、注文、約定などの情報を取得可能である。

本研究では、OandaAPI の中でも REST API のテスト用環境で取引を行う

図 2 のようにパラメータを変更することによって簡単に取引を行うことが可能である。

また、Oanda では GUI も用意されており図 3 に本研究で使用しているアカウントの画面を示す。

4 クラスタリングにおける学習データ

クラスタリングに用いる学習データは、1 日あたりのツイートの 8 つの感情スコアから構成される。

まず、学習データとして使用できるツイートを抽出する。使用できるツイートとは、為替の価格は土日休日はつけられていないため、それに対応するツイートである。

次に、抽出したツイートから 1 ツイートあたりの感情スコアを算出し、その後、それらのスコアを使用し、1 日単位の感情スコアを算出する必要がある。1 日あたりのスコアを算出する際には、その日にツイートされたツイート数の感情スコアの平均から算出しようと考えている。算出したスコアに対して、価格変動に関するラベルを紐付ける。

しかし、取得しているツイートのデータの容量が 10GB ほどあるため単純な処理だと 1 回の処理だと 3~4 日かかってしまっている

このままデータ量が多くなると膨大な時間を要することになる。改善点として、ソースコードを改良して処理速度を上げるかデータベースを構築するなど工夫しないといけない

現在は単純な処理を実行しているため、処理を分割するなど工夫して処理を行う必要がある。