

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

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。

また、計算機科学の発展により、ビッグデータの蓄積や蓄積したデータを機械学習を用いて分析することが可能となっている。その分析は金融経済現象へ応用されている。現在では、オルタナティブ・データを活用することで新たな金融工学の地平が切り開かれている。

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、Twitter の情報を取得し、その情報を活用することで金融マーケットの状況を予測する。行動経済学では、感情が個人の判断に大きく影響することが判明している。そこで、Twitter から感情に関わる情報を引き出す事で、マーケットの状況を予測出来るという仮定に基づき為替予測を行うシステムを開発する。

2 分析手法

本研究では、Twitter からの情報を収集し、収集した情報をセンチメント分析・クラスタリングを経て為替取引を行う。

2.1 今回の取り組み

今回は、Python による自動売買システムの実行とトランプ大統領のツイートをを用いた為替予測を行った。

2.1.1 自動売買システムの実行

自動売買システムの実行の様子を図1に示す。

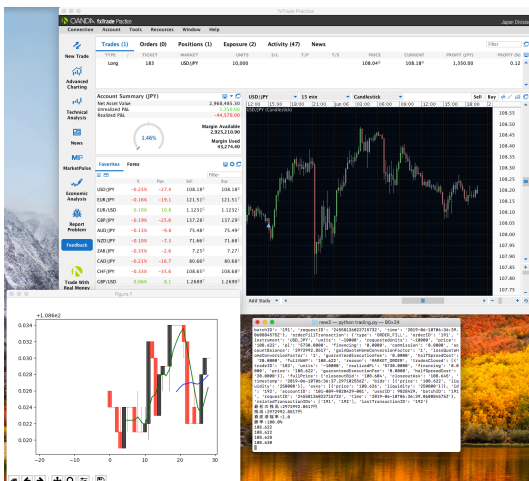


図 1: 1 自動売買システムの実行画面

設定したルールに従って売買が実行できると同時に価格変動のグラフもリアルタイムで描画できる。

2.2 トランプ大統領のツイートをを用いた為替予測

為替予測に用いるシステムの流れを図2に示す。

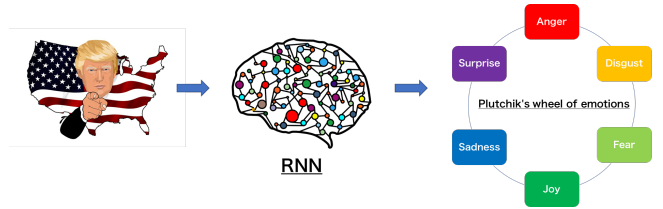


図 2: 2 システムの流れ

取得したツイートに対して、感情分析を行い為替の予測を行っていく。本研究では、RNN を用いた感情分析を採用している。また、予測に用いたデータフレームを図3に示す。

	Tweet	Anger	Disgust	Fear	Joy	Sadness	Surprise	t	c	target
0	If the Fed backs off and starts talking a little more Dovish I think were going to be right back to our 2800 to 2900 target range that weve had fo...	0.066892	0.025094	0.161228	0.241742	0.293027	0.210017	2018-10-29 22:53:00	112.492	1
1	The Stock Market is up massively since the Election but is now taking a little pause - people want to see what happens with the Midterms. If you w...	0.142287	0.040632	0.541550	0.064098	0.013292	0.198140	2018-10-29 22:34:00	112.445	1
2	Congressman Kevin Brady of Texas is so popular in his District and far beyond that he doesnt need any help - but I am giving it to him anyway. He ...	0.058110	0.001772	0.127102	0.784119	0.018262	0.010635	2018-10-29 22:25:00	112.458	1

図 3: 3 分析に用いたデータフレーム (上位3件)

Tweet は、トランプ大統領のツイート内容、Anger・Disgust・Fear・Joy・Sadness・Surprise は、感情スコア、t は、ツイートした日時、c はツイートしてから 10 分後の為替価格、target は教師ラベルである。

2.2.1 分析結果

テストデータ 100 ツイートに対して行った予測の精度は、約 63 % であった。前回の BoW を用いた分析の結果は約 52 % であったため、前回の結果は上回った。

今回も単純なナイーブベイズを用いた分析であったためその点において、まだまだ改善の余地があると考えられる。

3 おわりに

今後の展望としては、予測手法の改善や変更であったり、対象がトランプ大統領だけであったので分析対象を増やす。また、今回は 10 分後の為替の価格を対象としていたのでその点を考慮する必要がある。