

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、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、ツイートが為替に影響しているのかの検証、為替の価格のみによる予測、為替の価格にインジケータを追加した予測、為替の価格とインジケータ、ツイートの情報を追加した予測を行なった。

2.1.1 目的

SNS などの情報が為替に影響していることを検証し、社会的要因を即時に考慮することが可能な為替予測手法の提案

2.1.2 時系列データのクラスタリング

本研究では、K-shape を用いる。この手法の基本的な仕組みは、K-mean 法に基づいている。距離尺度に基づき、適用するデータの重心を計算し各クラスタの分割を繰り返すを行う。この手法では、Shape-based distance(SBD) と呼ばれる距離尺度を用いる。

2.1.3 エルボー法

まず、クラスタリングを行う前にエルボー法によりクラスタ数を決定する。エルボー法とは、各クラスタ中心との距離の二乗の合計をクラスタ内誤差平方和 (SSE) として計算することにより最適なクラスタを求める方法。

エルボー法では、肘のように曲がっている点が最適なクラスタであると示唆されるので図1よりクラスタ数は4か5が適していると考えられる。

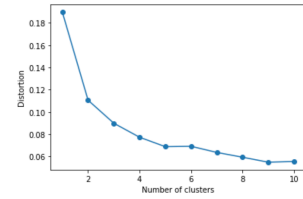


図 1: 1 エルボー法の結果

2.1.4 クラスタリング結果

トランプ大統領のツイート後から 20 分後の為替とランダムな日時の 20 分間の為替の価格それぞれのデータセットを用意してクラスタリングを行った。

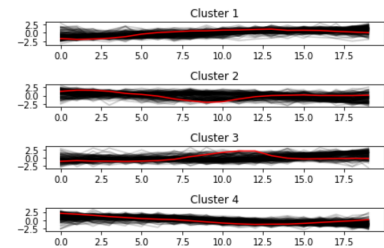


図 2: 2 K-shape による分析結果

Cluster2 と Cluster3 にツイート後の為替のデータセットが多く分類された。

2.1.5 考察

実際の価格の変動においても、ツイートの 7 分後から 10 分後辺りから為替が変動していることがわかる。そのため、ツイートによる為替への影響はあると考えられる。

2.2 為替予測

時系列予測が得意な LSTM を用いる。特徴量を図2のように 10 分間のデータを 1 セットとして予測を行なっていく。

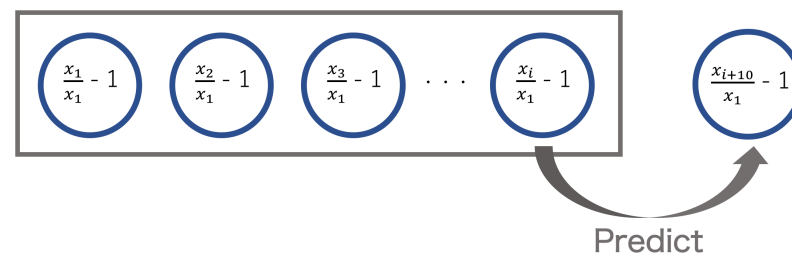


図 3: 2 データセット

2.2.1 検証 1

ここでは、為替の価格のみで予測を行う。学習データとして 2018 年 9 月の 1ヶ月分の為替データを用いた。

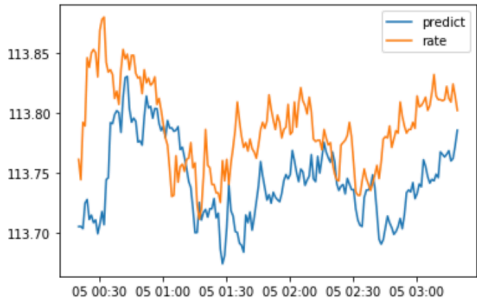


図 4: 3 検証 1 の分析結果

図 3 に分析結果を示す。決定係数は、0.683 であった。図 3 より予測と実測値の動きが全く逆の場合が多々見られた。10 分後の予測は終値だけでは難しい可能性が大きい

2.2.2 検証 2

検証 2 では、検証 1 の特徴量にインジケータを追加することにより、精度の向上を図る。用いる特徴量は、終値と 36 種類のインジケータである。

| ema_26 | ema_9 | bb_up | bb_mid | ... | apo | ppo | cmo | ht_dcperiod | ht_dcphase | ht_p_inphase | ht_ |
|------------|------------|------------|-----------|-----|-----------|-----------|------------|-------------|------------|--------------|-----|
| 111.075654 | 111.046402 | 111.148167 | 111.08480 | ... | -0.029788 | -0.026814 | -50.573348 | 19.598470 | -1.067944 | -0.014234 | |
| 111.071292 | 111.041989 | 111.145145 | 111.08020 | ... | -0.031538 | -0.028390 | -56.418381 | 21.458003 | -0.743397 | -0.016602 | |
| 111.068694 | 111.042247 | 111.139150 | 111.07625 | ... | -0.032301 | -0.029078 | -31.971764 | 23.536594 | -4.567837 | -0.018716 | |
| 111.067035 | 111.044687 | 111.133171 | 111.07305 | ... | -0.032333 | -0.029107 | -20.557873 | 24.324707 | 1.647142 | -0.026141 | |
| 111.065768 | 111.047364 | 111.124396 | 111.06960 | ... | -0.031353 | -0.028225 | -16.927649 | 25.708262 | 2.564275 | -0.033605 | |

図 5: 4 インジケータの一部

図 4 に用いるインジケータの一部を示す。

3 おわりに

今後の展望としては、検証 2 の実装と検証 3 において、感情スコアをどのように学習データに追加するかを検討する