

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。

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、「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、Web上の景気に関する情報やSNSからの情報のテキストからセンチメント分析を行い、金融マーケットの状況の予測を行う。

2 取り組んだ内容

Twitterから取得したデータを扱う上で、集計や抽出を行う際前回までの状況では処理時間が遅いという問題が発生した。今後もデータ量も増えることを考慮した場合、データベースを用意してデータを管理する必要がある。そこで、本研究ではMySQLを利用してデータの管理を行う。

今回の研究会まで取り組んだ内容としては、MySQLのインストールから基本操作、Twitterのデータはcsvファイルで管理してあるためデータベースへのインポートまで実行した。

2.1 MySQL

MySQLは全世界で最も利用されているリレーショナルデータベースの1つである。主な特徴としては、オープンソースなため基本的に無料で利用できることや大量のデータを高速で扱うことができる、MyAdminなどをはじめとした便利なツールが充実していることがあげられる。MySQLの操作は一般的には、terminalなどから操作する。

id	date	week	tweet
0	2019-01-07 11:43:09	Monday	チームDのMC-U。カーディガンの赤は血染め。顔のシミはタトゥー。
1	2018-12-30 14:43:49	Sunday	まず、お箸の練習しろ。
2	2019-01-06 03:38:58	Sunday	本日ラジオ。
3	2019-01-01 14:12:03	Tuesday	あけー。ことでーす！
4	2018-12-31 07:56:34	Monday	良いお年を。
5	2018-12-30 07:31:22	Sunday	本日ラジオ。
6	2018-12-28 06:54:00	Friday	年末ラストは歌合戦。
7	2018-12-25 16:01:53	Tuesday	あと一回だけスプラトゥーンやって寝よう。。。
8	2018-12-25 16:00:09	Tuesday	ねるー！
9	2018-12-25 15:51:49	Tuesday	おまえはいつも恭社感を漂わせるな。

図2: idカラムを追加したテーブル

idカラムを追加する以前では、重複した場合1つだけ残すといった処理ができず重複したレコードを全て削除する処理しか実行できなかった。実際に、重複したデータを削除する処理内容としてはidが最も小さいレコードのデータのみを残して、それ以外のレコードは削除する。

3 今後の取り組み

MySQLへのデータの格納は完了したため、次からは実際にデータをMySQLから取り出して感情分析を行う処理を実行する。
そして、算出した感情スコアをもとに為替予測を行ってみる。

図1: 以前までのテーブル

MySQLにデータを格納する際に、重複した内容のデータが存在しているといった問題が発生した。そこで、その解決策としてidを格納するカラムを設けることにより重複したデータを削除する。