

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、「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、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。

しかし、本研究ではデータベースから抽出したデータに対してそのまま分析を行いたいため Python からの操作を行う

```
import mysql.connector

db=mysql.connector.connect(host="localhost", user="trump_tw", password="trump")

cursor=db.cursor()

cursor.execute("USE trump_tweet")
db.commit()
```

図 2: Python での操作

取得したデータをデータベースにインポートする際に、TwitterAPI から得られるタイムスタンプの型が MySQL の DATETIME 型に合わなかったため変換を行う必要があった。

```
date = 'Sun Jul 29 20:12:15 +0000 2018'
cursor.execute("SELECT str_to_date(%s, '%a %b %d %T %f %Y') % date)
#cursor.execute("SELECT str_to_date('Sun Jul 29 20:12:15 +0000 2018', '%a %b %d %T %f %Y')")
rows = cursor.fetchall()
```

図 3: 型の変換

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は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。

```
[mysql> select User,Host from mysql.user;
```

User	Host
twitter	%
user	%
mysql.infoschema	localhost
mysql.session	localhost
mysql.sys	localhost
root	localhost
test_user	localhost

```
7 rows in set (0.03 sec)
```

図 1: terminal での操作