

# Twitterを用いた株価予測

---

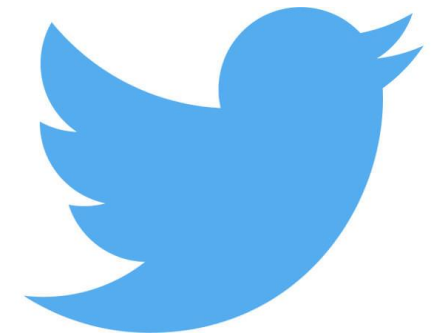
情報工学基盤講座 麻生到

# はじめに

---

Twitterを含めたSNS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、その情報量を膨大かつ増え続けている

➡ そのSNSの情報を利用して、**経済動向**の分析が行われている



# 既存研究

---

Johan BollenらがTwitterから感情分析を用いてダウ平均株価を予測している [1]

- ツイート

2008年2月28日～12月19日に取得されたツイート  
9,853,498ツイート(~270万ユーザー)

- 感情分析

Opinion Finder (OF) :

OFは辞書を使ってpositive/negativeに振り分けるフリーソフト

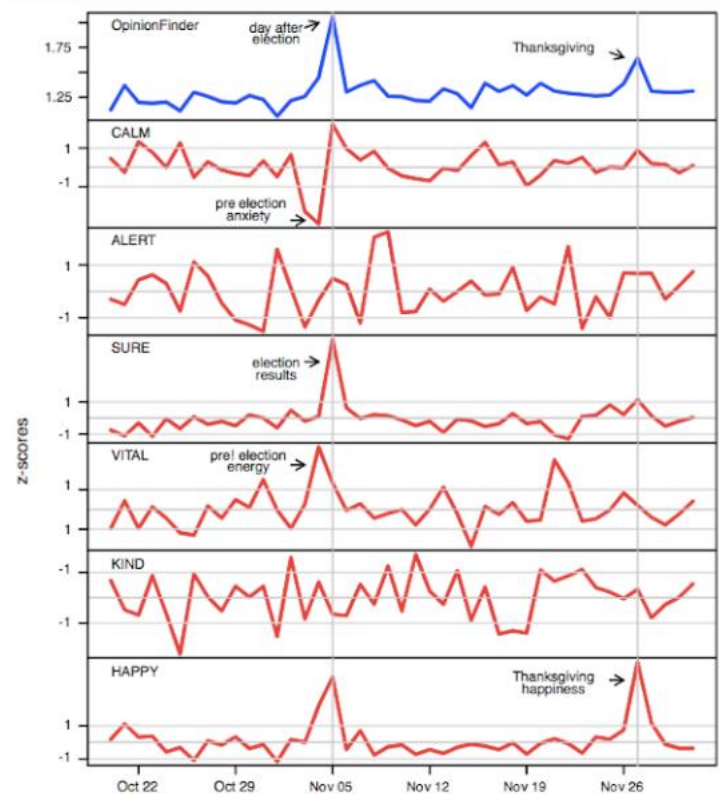
Google-Profile of Mood States (GPOMS) :

6種類の感情因子 (Calm, Alert, Sure, Vital, Kind, and Happy) に分解する辞書

[1] Johan Bollen, Huina Mao, Xiao-Jun Zeng, “Twitter mood predicts the stock market.”, arXiv:1010.3003v1 [cs.CE], 14, Oct, 2010

Johan Bollenらの研究では、“I feel”、“I am”といった心理的表現のみのツイートを取得して分析を行っている

Calm(平穏)とダウ平均株価の相関が高く、CalmのScoreを用いて予測を行っている



Lag	OF	Calm	Alert	Sure	Vital	Kind	Happy
1 day	<b>0.085*</b>	0.272	0.952	0.648	0.120	0.848	0.388
2 days	0.268	<b>0.013**</b>	0.973	0.811	0.369	0.991	0.7061
3 days	0.436	<b>0.022**</b>	0.981	0.349	0.418	0.991	0.723
4 days	0.218	<b>0.030**</b>	0.998	0.415	0.475	0.989	0.750
5 days	0.300	<b>0.036**</b>	0.989	0.544	0.553	0.996	0.173
6 days	0.446	<b>0.065*</b>	0.996	0.691	0.682	0.994	<b>0.081*</b>
7 days	0.620	0.157	0.999	0.381	0.713	0.999	0.150

(p-value < 0.05: \*\*, p-value < 0.1: \*)

Lag : 何日ずらしたか

ダウ平均株価との相関

# データマイニング

## Twitter API

```
Wed Jul 04 04:27:03 +0000 2018, フシだいぶ初めて任天堂の株主総会行った時、予想に反してマリオのコスプレとかしてる人一人もおらんかった。ほとんどはお爺さんお婆さん。質問もゲームに関してはほぼなし。", 2, 6
Wed Jul 04 04:26:49 +0000 2018, 任天堂このまま25000円まで行くの期待, 0, 0
Wed Jul 04 04:25:57 +0000 2018, 残念だけど、あたしの出演作とか生まれた会社とか、少なくとも任天堂のゲームに関心が持てない人にはフォローは返せないわ, 0, 0
Wed Jul 04 04:25:42 +0000 2018, "これ、スプラトゥーン3とか出たら
イカとタコの2個ソフト出るやつやでえ...😂😂😂任天堂の商法は小学校の時から知ってるんだからな!!😂
&lt;RT", 0, 17
Wed Jul 04 04:25:38 +0000 2018, "ただ買うのに350万円いるからな。
任天堂は株式分割したら、全国のゲーマーがファンクラブ会員証のような感覚で買うんちゃうやらか。
いや、変なコスプレした人とかに総会来てほしくないからわざとやってないのか。", 2, 7
```

いいね数

投稿日

ツイート

リツイート数

### 制限情報

- 1週間前までのツイートまでしか取得できない
- アクセス回数: 15分間に180回
- 1回のアクセスに200件程度のツイートしか取得できない

# 感情辞書

## PN Table

東工大の高村教授が作って公開している辞書 [2]

良い: よい: 形容詞: 0.999995  
喜ぶ: よろこぶ: 動詞: 0.999979  
褒める: ほめる: 動詞: 0.999979  
めでたい: めでたい: 形容詞: 0.999645  
賢い: かしこい: 形容詞: 0.999486  
善い: いい: 形容詞: 0.999314  
適す: てきす: 動詞: 0.999295  
天晴: あっばれ: 名詞: 0.999267  
祝う: いわう: 動詞: 0.999122  
功績: こうせき: 名詞: 0.999104  
賞: しょう: 名詞: 0.998943  
嬉しい: うれしい: 形容詞: 0.998871  
喜び: よろこび: 名詞: 0.998861  
才知: さいち: 名詞: 0.998771  
徳: とく: 名詞: 0.998745  
才能: さいのう: 名詞: 0.998699  
素晴らしい: すばらしい: 形容詞: 0.998617  
芳しい: かんばしい: 形容詞: 0.998578  
称える: たたえる: 動詞: 0.998576  
適切: てきせつ: 名詞: 0.998406  
崇める: あがめる: 動詞: 0.998357  
助ける: たすける: 動詞: 0.998356

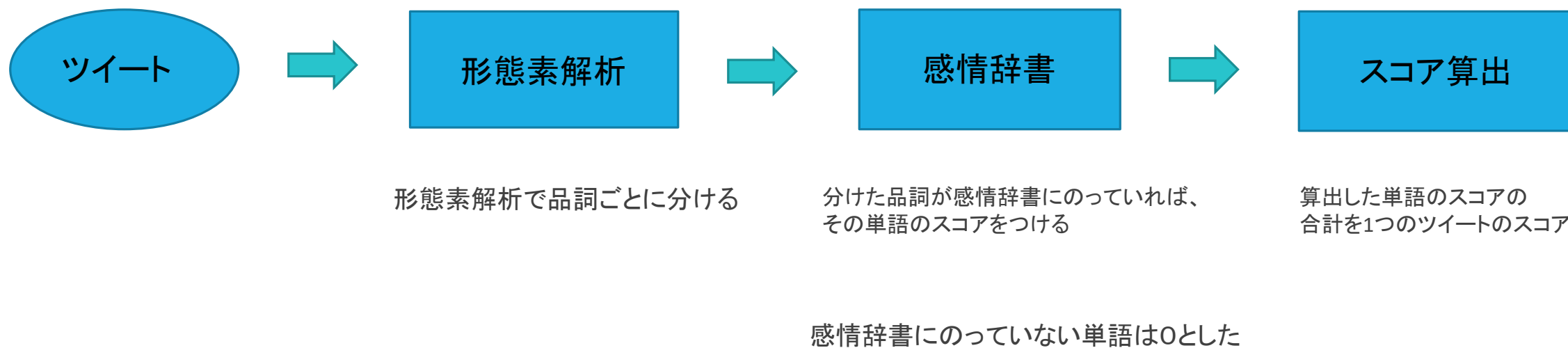
1つの単語に対して、+1~-1のスコアがつけられている

43748単語収録されている

[2] 高村 大也, 乾 孝司, 奥村 学, “スピンモデルによる単語の感情極性抽出, 情報学会論文誌, Vol. 47, No. 2, Feb. 2006

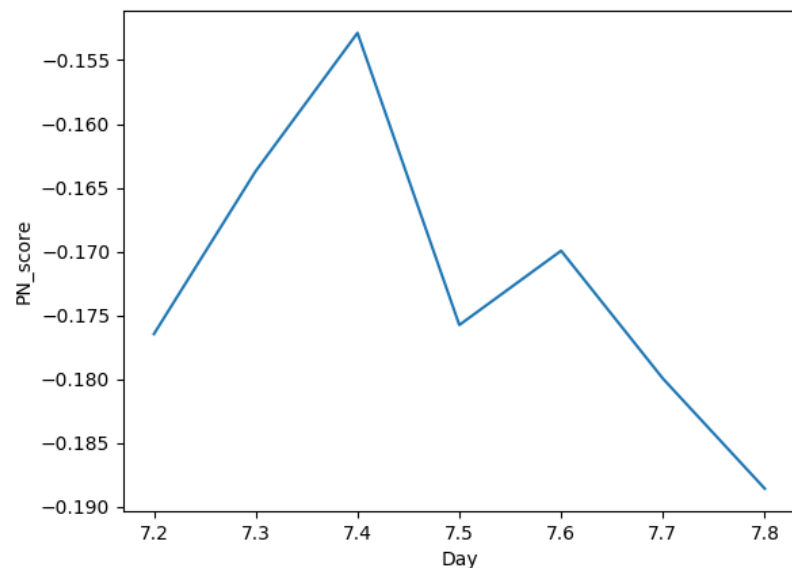
# 適用手法

---



# 実験

7月2日～7月8日までのツイートを分析対象とした(20000ツイート)



キーワード: 任天堂の感情スコア

$$\text{1日のスコア} = \frac{\text{1日の合計スコア}}{\text{1日のツイート数}}$$

この1週間の変動はあまりみられなかった



# これからの課題

---

- ・ 日経平均株価を予測するにあたり、ツイートの取得条件が定まっていない  
(キーワードで検索をかけるか、アカウントで検索をかけるか)
- ・ 感情分析を行うにあたり、ツイートの中に感情辞書に含まれていない単語も多くスコアをつけることがうまくいっていない
- ・ 株価予測モデルがまだ決まっていない