



1. はじめに
2. 分析手法

# Twitter mood predicts the stock market.

Development of Financial Market Forecasting Method  
Considering Alternative Data on Economy

Itaru Aso

Graduate School of Information Engineering, Toyama Prefectural University  
t855001@st.pu-toyama.ac.jp

Toyama Prefectural Univ.



# 1.1 本研究の背景

## 背景

株式市場の予測が学門やビジネスから注目を集めている。

⇒ 株式市場の予測は本当に可能であるのか？

## Efficient Market Hypothesis (EMH)

株価の変動は現在や過去の価格ではなく、新しいニュースや情報によって引き起こされる

EMH … 効率的市場仮説



## 1.2 先行研究

- 1. はじめに
- 2. 分析手法

### ニュースについて

ニュースは予測不可能なものではなく、オンラインソーシャルメディア（ブログや Twitter など）からいち早く指標を見つけることで様々な経済的および商業的な変化の予測が可能であるのではないかと示唆している

### SNS の応用例

- 1 オンラインチャットアクティビティによる書籍販売の予測
- 2 ブログの感情分析による映画の売上予測
- 3 Twitter による映画に関する市民の感情をもとに興行収入を予測

ニュースが確かに株価に最も影響を与えるが、市民の雰囲気や感情も同様に影響を与えるかもしれない

⇒ 株式市場の予測に適応できるのではないか



## 1.2 本研究の目的

- 1. はじめに
- 2. 分析手法

### 目的

Twitter から得られた人々のツイートの感情を分析することによって  
ダウ平均を予測する

感情を分析する際には、OpinionFinder と Google Profile of Mood States(GPOMS) を用いる

- OpinionFinder

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

- Google Profile of Mood States

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



## 2.1 前処理

- 1. はじめに
- 2. 分析手法

### サンプリング

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)

### 前処理

- 1 ありふれ過ぎている語や句読点を除去
- 2 投稿者の感情が表れているツイートのみを抽出 ( “ i feel ” , “ i am feeling ” , “ i ’ m feeling ” , “ i dont feel ” , “ l ’ m ” , “ lm ” , “ l am ” , and “ makes me ” )
- 3 スпамをはじくために “ http: ” や “ www. ” を含む物は除去



## 2.1 感情分析

- 1. はじめに
- 2. 分析手法

### OpinionFinder

OF 辞書は 2718 の positive 単語、4912 の negative 単語からなる辞書

### Google-Profile of Mood States (GPOMS)

72 に及ぶ POMS の調査項目を参考に作成された 964 語からなる辞書

OpinionFinder と GPOMS を比較するためにスケールを合わせる

$$Z_{X_t} = \frac{X_t - \bar{x}(X_{t \pm k})}{\sigma(X_{t \pm k})} \quad (1)$$

$X_t$ :感情値,  $k$ :日付の幅



## 2.2 ケーススタディ

1. はじめに
2. 分析手法

2008 年 11 月 4 日のアメリカ大統領選挙と 2008 年 11 月 27 日の感謝祭に行われている

### アメリカ大統領選挙

- 1 前日は Calm が大きく落ち込んでいる
- 2 当日は Calm, Sure, Happy の上昇している

### 感謝祭

OF と Happy が特徴を捉えている

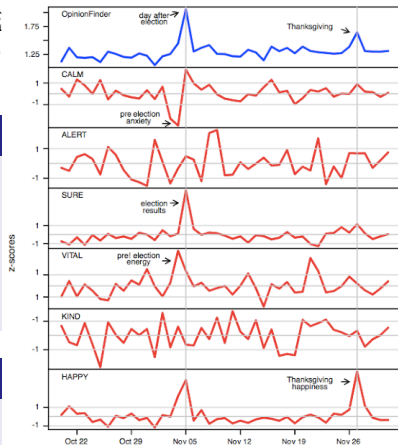


Figure: 1 z-score のグラフ



## 2.3 相関

### OF と GPOMS の相関

- 1 Sure, Vital, Happy が相関が高い
- 2 Calm と Alert, Kind はそこまで相関が高くない

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: \*)

Figure: 2 感情値とダウ平均の相関





## 2.3 相関

1. はじめに
2. 分析手法

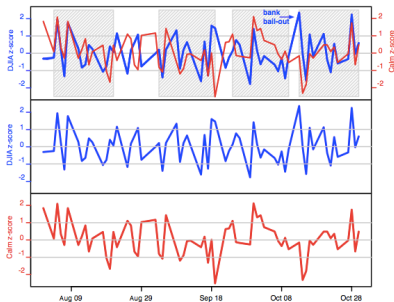


Figure: 2 ダウ平均と Calm の相関

3日後にずらした Calm の時系列データとダウ平均の時系列データをグラフ化

実際に、重回帰分析を行なったところ 87.6 %の精度で予測が可能であった