



- 1. はじめに
- 2. 市場と投資判断
- 3. 金融テキストマイニング
- 4. 提案手法
- 5. 数値実験
- 6. おわりに

# オルタナティブデータを考慮した 変数選択と数法則発見式を用いた 金融マーケット予測手法の開発

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

L205, AM 9:00-9:25, Friday, December 8, 2018,  
Toyama Prefectural Univ.



# 1.1 本研究の背景

- 1. はじめに
- 2. 市場と投資判断
- 3 金融テキストマ  
イニング
- 4. 提案手法
- 5. 数値実験
- 6. おわりに

## 本研究の背景

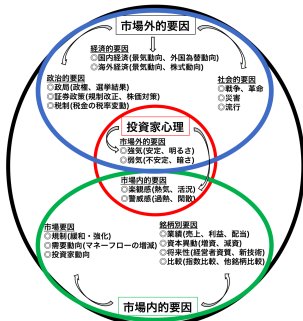
- 1 経済活動の活発化に伴い、金融市場の規模は大きくなっている
- 2 計算機科学の発展により、ビッグデータの蓄積や蓄積したデータを機械学習を用いて分析することが可能
- 3 ビッグデータを用いたデータ分析は、金融経済現象にも応用
- 4 オルタナティブ・データを活用することで新たな金融工学の地平が切り開かれている



## 1.2 本研究の概要

### 金融市場

- 1 金融市場は複雑に様々な要因が絡み合っていることから全容を明らかにすることは困難
- 2 金融市場をモデル化するためには、市場内外に起こる事象を関連づけることが重要である





## 1.3 本研究の目的

- 1. はじめに
- 2. 市場と投資判断
- 3 金融テキストマ  
イニング
- 4. 提案手法
- 5. 数値実験
- 6. おわりに

### 本研究の方針

金融市場の動向をモデル化するためには市場外的要因と市場内的要因を含んだ分析をすることが必要であると考え

### 本研究の目的

- 1 市場内外内の要因を考慮することにより、金融マーケットに関する幅広い情報をカバーした予測手法の提案
- 2 変数選択と数法則発見式を用いて金融マーケットに関する情報の中影響力のある情報の特定



## 2.1 市場

### 市場原理

- 1 株式や為替などが取引される際にその価格と数量は需要と供給によって決定される
- 2 市場参加者は利益を追求するので、市場全体が結果的にバランスを保つように機能されることが多い。
- 3 それらを考慮して投資家は利益最大化のために振る舞う

### 投資

- 一般的に参考とする情報として過去の価格の値動きから得られる指標が用いられる
- 効率的市場仮説により過去のデータのみの情報だけでは市場の動向を予測する事が難しい  
→ 過去のデータのみだけでなく市場に影響すると考えられるニュースや新聞などを考慮することが必要

1. はじめに
2. 市場と投資判断
3. 金融テキストマ  
イニング
4. 提案手法
5. 数値実験
6. おわりに



## 2.2 投資

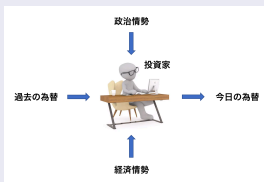
### 投資判断

- 1 投資家の行動によって値動きが発生するため、最も重要となる要因は投資家心理である。
- 2 投資家心理に影響を与えるているのが、市場内的要因と市場外的要因である。

### 投資家心理

$$S(t) = f(S(t-1), P, E, \dots) \quad (1)$$

$S(t)$ :現在の市場価格  $t$ : 時間  $P$ : 市場内的要因,  $E$ : 市場外的要因





## 2.3 投資判断

### 投資の分析手法

一般的に投資を行う際の分析手法としてファンダメンタル分析とテクニカル分析の2つの手法が用いられる。

**1** ファンダメンタル分析

→ 国際的な経済の動きや個別の企業の財務情報など市場外的要因を考慮する手法

**2** テクニカル分析

→ 現在の市場のトレンドを把握する方法であり、テクニカル指標を用いて市場内的要因を考慮する手法

- 1. はじめに
- 2. 市場と投資判断
- 3. 金融テキストマ  
イニング
- 4. 提案手法
- 5. 数値実験
- 6. おわりに



## 2.4 ファンダメンタル分析

### 分析対象

社会的要因や政治的要因となることから、新聞記事やネットニュース、SNS からテキストデータを取得して分析することが効果的であると考えられる。

### 従来研究

- 1 Twitter のテキストを用いて翌日のダウ・ジョーンズ工業株価平均の変動を 87.6 % の精度で予測する研究 [1]
- 2 日本の Yahoo! 株式掲示板における投稿数が多い東証一部 50 社を対象に、インターネット株式掲示板と株式指標の関係を検証した研究 [2]
- 3 日本銀行の金融経済月報のテキストデータを用いて日本国債市場の月末の価格を予測する研究 [3]

[1] Johan Bollen, Huina Mao, Xiao-Jun Zeng, Twitter mood predicts the stock market, CoRR, Vol.abs/1010.3003(2010)

[2] 丸山健, 梅原英一, 諏訪博彦, 太田敏澄, インターネット株式掲示板の投稿内容と株式市場の関連性, 証券アナリストジャーナル, Vol.46, No.11・12, pp110-127(Mar. 2008)

[3] 和泉潔, 後藤卓, 松井藤五郎, 経済テキスト情報を用いた長期的な市場動向推定, 情報処理学会論文誌, Vol.52, No.12, pp3309-3315(Dec. 2011)





## 2.5 テクニカル分析

- 1. はじめに
- 2. 市場と投資判断
- 3. 金融テキストマイニング
- 4. 提案手法
- 5. 数値実験
- 6. おわりに

### 分析対象

市場内要因や銘柄別要因であることから過去のデータ

### テクニカル指標

- 1 現在の相場のトレンド傾向
- 2 現在のトレンドの強さ
- 3 相場が上昇や下降時の転換点
- 4 値頃感や相場の変動幅

### 従来研究

- テクニカル指標の最適な組み合わせを学習することによって  
10tick 先の価格変動の方向を自動予測する研究 [4]

→ テクニカル指標を組み合わせることで予測精度が向上すると考えられている

[4] 徳岡聖二, 田中三栄子, 進化計算による tick 価格変動のトレンド予測, 情報処理学会論文誌, Vol.48, No.SIG19(TOM19), pp68-74(Dec. 2007)



## 2.6 ファンダメンタル分析とテクニカル分析

### 問題点

#### ■ ファンダメンタル分析の問題点

ファンダメンタル分析のみでスキャルピングのような高頻度の取引は難しく、デイトレードが限界である

#### ■ テクニカル分析の問題点

急なニュースや決算の発表や業績の発表による値動きの大きな変化に対応できない

### 解決策

- 1 ファンダメンタル分析とテクニカル分析は相補的な関係にある  
[5]
- 2 市場内的要因と市場外的要因を考慮した投資判断の手法が効果的である

[5] Bettman, Jenni L, Sault, Stephen J, Schultz, Emma L, Fundamental and technical analysis: substitutes or complement?, Accounting and Finance, Volume 49, Number 1, March 2009, pp. 21-36



## 3.1 金融テキストマイニングの分析対象

### 分析対象

1. はじめに
2. 市場と投資判断
- 3 金融テキストマイニング
4. 提案手法
5. 数値実験
6. おわりに

#### 1 SNS

- ・最も多様であり、書き手の多いテキスト
- ・日常的な内容なものが多く統一性のないもの
- ・リアルタイム性に優れている
- ・要人の発言や社会情勢の情報などがいち早く獲得可能

#### 2 オンライン上の経済情報

- ・SNS 上のテキストよりもより専門的な情報を獲得可能
- ・ロイターや Bloomberg のオンライン上の経済ニュース
- ・書き手が経済や社会情勢の知識がある記者

#### 3 経済レポート

- ・金融機関の発行するテキストであり、最も専門的なテキスト
- ・投資家や他の専門家の人々に向けて市況を解説したテキスト
- ・内容だけでなく言葉遣いやフォーマットも統一性がある
- ・発行される頻度が少ない



## 3.2 金融テキストによる市場の影響

### 本研究の分析対象

本研究では、市場内外の要因を考慮して予測手法を提案する.

→ 最も更新頻度の高い金融テキストを用いる必要があるので  
Twitter のテキストを分析対象とする.

### ツイートの為替への影響

- 1 ツイートがどのように為替へどのように影響しているのか検証する
- 2 ツイート後の為替の値動きとランダムな日時の為替の値動きの比較
- 3 時系列データのクラスタリング手法の k-Shape によって検証する

1. はじめに
2. 市場と投資判断
- 3 金融テキストマ  
イニング
4. 提案手法
5. 数値実験
6. おわりに



### 3.3 ツイートによる為替の影響の検証

#### 検証対象のデータ

トランプ大統領の 2019.9.1 から 2019.10.31 の約 2ヶ月間のツイート  
クラスタリングによる検証を行う前に、クラスタ数をエルボー法により決定する

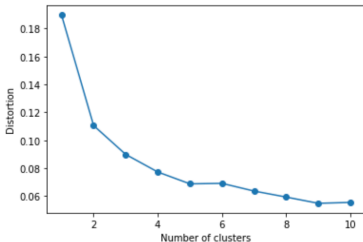


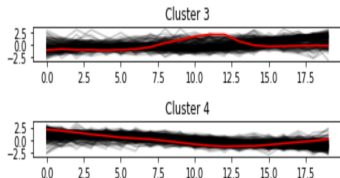
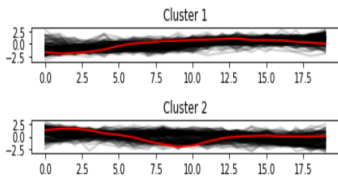
Figure: 2. エルボー法の分析結果

エルボー法は肘のように SSE 値が曲がった点が適しているクラスタ数である。以上の図より、クラスタ数を 4 に決定した。



### 3.4 k-Shape によるクラスタリング

1. はじめに
2. 市場と投資判断
3. 金融テキストマイニング
4. 提案手法
5. 数値実験
6. おわりに



#### 考察

- 1 Cluster2 と Cluster3 は 7.5 分を過ぎたあたりからそれぞれ上下に大きく変動している
- 2 Cluster2 と Cluster3 にはトランプ大統領のツイート直後の為替の価格の変動が多くみられた



## 3.5 センチメント分析

### センチメント分析

- 1 Web 上のユーザー生成コンテンツの量は、主に SNS やブログなど自分の個人コンテンツを共有することを可能にする無数のプラットフォームの出現により、ますます急速に増加している。
- 2 ユーザー生成コンテンツは意見や感情が豊富であり、株式市場の変動の予測であったり、マーケティングの戦略の参考にされることが多い。
- 3 オンライン上の人々の情報を自動的に分析する手法が重要となってきた。

### 本研究のセンチメント分析

本研究では、ツイートを分析対象としており、提案手法の入力データとして用いる。

→ センチメント分析によりテキストデータのベクトル化を図る。

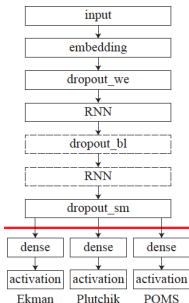
1. はじめに
2. 市場と投資判断
- 3 金融テキストマ  
イニング
4. 提案手法
5. 数値実験
6. おわりに



## 3.6 感情スコアの導出

### センチメント分析

感情スコアを Twitter での感情認識を参考にシステムを作成した [6]



[6]NikoColneric, JanezDemsar, Emotion Recognition on Twitter: Comparative Study and Training a Unison Model, IEEE Transactions on Affective Computing, PP(99):1-1(Feb 2018)





## 3.7 センチメント分析による感情スコアの算出

### 感情スコア

センチメント分析によって、Ekman の感情の感情分類に基づいてスコアを算出する

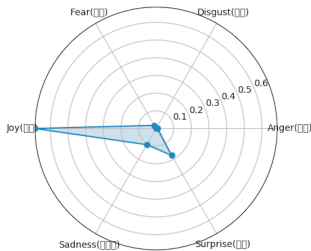


Figure: 2. Pultick の感情の環に基づいた感情スコア

「A years ago, A star was Born, and here we are 6 times pink platium」



## 4.1 金融マーケットの予測手法

### 予測手法

金融マーケット予測手法として主成分回帰分析を用いた GMDH を適用する

### GMDH

- 1 少ない入出力データからシステムに関する先験的な知識を必要とすることなくモデリングできる
- 2 階層型ニューラルネットワークと同じく、多層構造と並列処理を特徴としたアルゴリズム

### 主成分回帰分析を用いる GMDH のメリット

- 1 最適な変数の組み合わせを自己選択する.
- 2 層の積み重ねを自動停止する.

1. はじめに
2. 市場と投資判断
3. 金融テキストマイニング
4. 提案手法
5. 数値実験
6. おわりに



## 4.2 簡単な同定問題への適用

### 適用する式

$$\Phi = (1.0 + 1.1x_1 + 1.2x_2 + 1.3x_3)^4 + \varepsilon \quad (2)$$

ここで,  $\varepsilon$  は  $N(0, 25^2)$  の正規性白色雑音を示す.

入力変数:  $x_1 \sim x_4$  (4 変数)

→ 適用する式において  $x_4$  は含まれていないが, 入力に不必要な変数が含まれている場合においてもシステムの同定が可能であることを示すために  $x_4$  を入力に含めている.

同定に用いる入出力データ: 20 個

モデルの評価に用いる入出力データ: 20 個

1. はじめに
2. 市場と投資判断
3. 金融テキストマイニング
4. 提案手法
5. 数値実験
6. おわりに



## 4.3 同定されたモデル

### 同定されたモデル

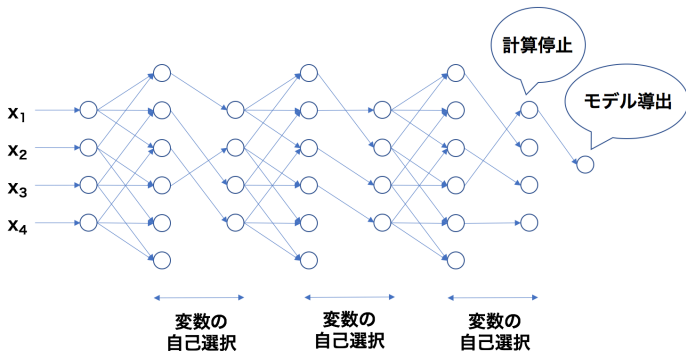


Figure: 同定されたモデル



## 4.4 同定結果

### AIC と固有値のグラフ

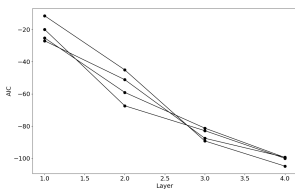


Figure: AIC

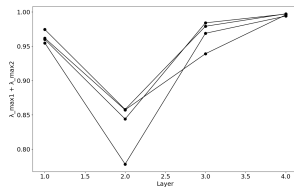


Figure: 固有値

1. はじめに
2. 市場と投資判断
3. 金融テキストマイニング
4. 提案手法
5. 数値実験
6. おわりに



## 4.4 同定結果

### 予測精度

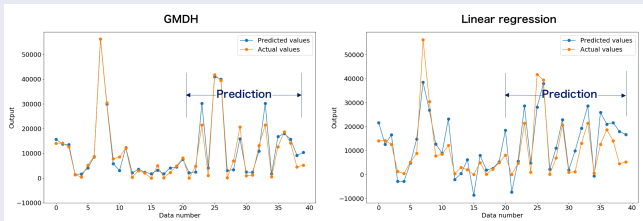


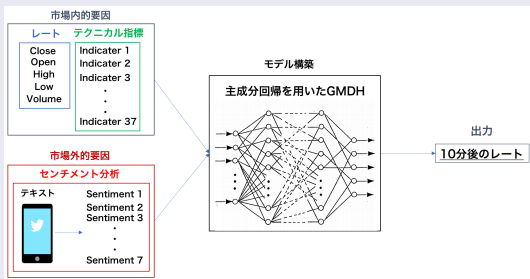
Figure: GMDH と線形回帰の予測精度

1. はじめに
2. 市場と投資判断
3. 金融テキストマイニング
4. 提案手法
5. 数値実験
6. おわりに



## 4.5 提案手法

### 提案手法



### 特徴量

- 1 市場内要因  
→ 終値と始値, 高値, 安値, インジケータ (37 種類)
- 2 市場外的要因  
→ Twitter のツイート



## 4.6 特徴量

- 1. はじめに
- 2. 市場と投資判断
- 3. 金融テキストマ  
イニング
- 4. 提案手法
- 5. 数値実験
- 5. おわりに

### 市場内的要因

- 1 投資家本人がチャートと指標を見て判断を行うため体系的な予測方法が確立されていない
- 2 モデルを構築するごとに選択するインジケータだったり組み合わせが異なるため柔軟なストラテジーの作成が可能となる

### 市場外的要因

- 1 金融レポートや新聞記事やニュースでは発信される頻度が時間帯が決まっており市場の動向についていけない
- 2 投資家らは、それらの媒体が発信される前に動いているため情報の鮮度が落ちてしまう
- 3 より速く情報が発信される SNS を用いることで情報の鮮度が確保できると考えられる





## 4.7 モデル構築

### GMDH によるモデルの構築

- 1 特徴量の組み合わせごとにノードを作成
- 2 ノードに対して AIC をもとに特徴量の自己選択  
→ 多重共線性を引き起こす中間変数の取り込みを防いでいる
- 3 層の積み重ねの停止を固有値をもとに自動で停止
- 4 ハイパーパラメータが少ないことから先験的な知識を必要とせずにモデルを構築可能
- 5 モデルの構造を数式として算出可能  
→ モデル式から特徴量における重要度や有用な組み合わせが明確になる

1. はじめに
2. 市場と投資判断
3. 金融テキストマイニング
4. 提案手法
5. 数値実験
6. おわりに



## 5.1 実験の概要

### 概要

- 1 外国為替のドル円を取引を行う
- 2 提案手法の有意性を示すため2つの売買ルールによって検証を行う.
- 3 取引時間 : 9時 ~ 6時
- 4 取引通貨量 : 1000
- 5 手数料 : 2pips

### 売買ルール

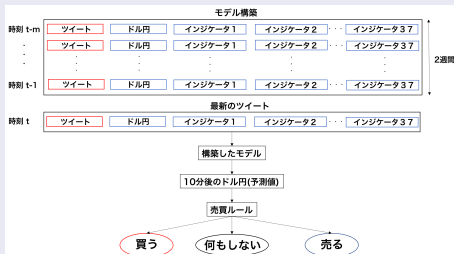
- 1 ルール 1  
→ 市場外的要因と市場内的要因を考慮した売買ルール
- 2 ルール 2  
→ 市場内的要因のみを考慮して作成した売買ルール

1. はじめに
2. 市場と投資判断
- 3 金融テキストマ  
イニング
4. 提案手法
5. 数値実験
6. おわりに



## 5.2 ルール 1

### ルール 1



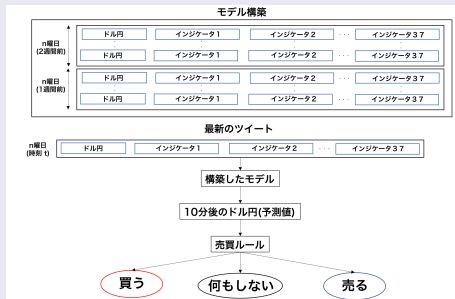
ドル円の終値:  $x_t$ , 予測値:  $x_{pred}$

- 1  $x_{pred} - x_t > 0.2$  の時  
→ 買いと判断
- 2  $x_{pred} - x_t < -0.2$  の時  
→ 売りと判断
- 3  $0.2 \geq x_{pred} - x_t \geq -0.2$  の時  
→ 何もしないと判断



## 5.2 ルール 2

### ルール 2



- 1  $x_{pred} - x_t > 0.2$  の時  
→ 買いと判断
- 2  $x_{pred} - x_t < -0.2$  の時  
→ 売りと判断
- 3  $0.2 \geq x_{pred} - x_t \geq -0.2$  の時  
→ 何もしないと判断



## 5.3 実験結果

1. はじめに
2. 市場と投資判断
3. 金融テキストマ  
イニング
4. 提案手法
5. 数値実験
6. おわりに

日付	収益 (pips)
9/16	-15400
9/17	-160000
9/18	+13100
9/19	+78000
9/20	+1400
9/23	-59000
9/24	-3900
9/25	+145500
9/26	+2100
9/27	-26200
9/30	-77200
10/1	+113600
10/2	+151700
10/3	+17100
10/4	+33200
10/7	+29090
10/8	+106400
10/9	+4000
10/10	+43200
10/11	-8300
10/14	-20100
10/15	-12800
10/16	-87100
10/17	-41600
10/18	-39790
10/21	-34200
10/22	+51200
10/23	-21100
10/24	-3090
10/25	-7200

日付	収益 (pips)
9/16	-10900
9/17	-43590
9/18	+15900
9/19	-28290
9/20	-1100
9/23	-2400
9/24	-4700
9/25	-29700
9/26	-1100
9/27	+19400
9/30	+30200
10/1	+29400
10/2	+66000
10/3	+23890
10/4	+11400
10/7	+4000
10/8	+51200
10/9	+0
10/10	-5300
10/11	+2000
10/14	-3600
10/15	-11200
10/16	-11200
10/17	-8300
10/18	-3900
10/21	+2200
10/22	+9000
10/23	-6000
10/24	+1500
10/25	+3100

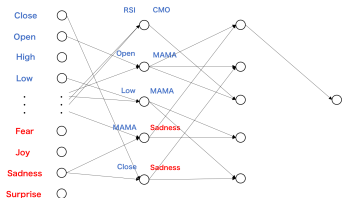
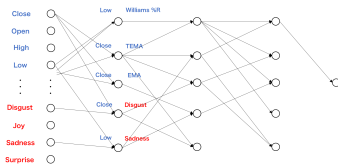
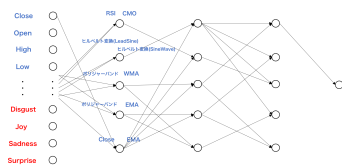
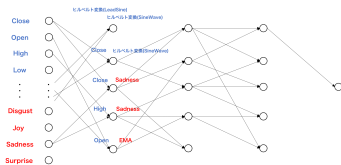
### 収益

- 1 ルール 1 : 236800pips
- 2 ルール 2 : 94200pips



## 5.4 モデル式

1. はじめに
2. 市場と投資判断
3. 金融テキストマ  
イニング
4. 提案手法
5. 数値実験
6. おわりに





## 6.1 おわりに

### 今後の課題

- 1 市場外の情報を増やす
- 2 センチメント分析の日本語への対応
- 3 リアルタイム取引への対応

1. はじめに
2. 市場と投資判断
3. 金融テキストマ  
イニング
4. 提案手法
5. 数値実験
6. おわりに



### 3.3 ツイートの為替の影響

トランプ大統領がツイートした直後の為替の価格とランダムな日時の為替の価格の変動をクラスタごとに色分けしたグラフを以下に示す

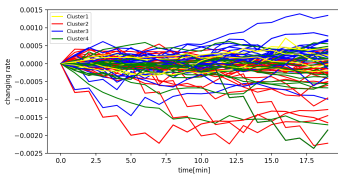


Figure: ツイート後のレート

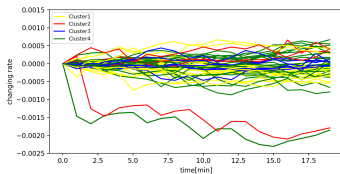


Figure: ランダムのレート

1. はじめに
2. 市場と投資判断
3. 金融テキストマイニング
4. 提案手法
5. 数値実験
6. おわりに