



1. 本研究について
2. ツイートが為替に影響しているかの検証
3. 為替予測

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

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 取り組み

1. 本研究について

2. ツイートが為替に影響しているかの検証

3. 為替予測

研究の流れ

- 1 ツイートが為替に影響しているのかの検証
- 2 為替の価格のみによる予測
- 3 為替の価格にインジケータを追加した予測
- 4 為替の価格とインジケータ、ツイートの情報を追加した予測

研究の目的

SNS などの情報が為替に影響していることを検証し、社会的要因を即時に考慮することが可能な為替予測手法の提案



2.1 検証方法

- 1. 本研究について
- 2. ツイートが為替に影響しているかの検証
- 3. 為替予測

目的

ツイート後の為替の価格とランダムな日時の為替の価格をクラスタリングすることによって影響があるかの検証を行う

時系列データのクラスタリングを行う際には、主に以下の2つ手法が適用されることが多い。

時系列データのクラスタリング

- 1 特徴ベース (Feature-based)
→ 時系列データを低次元の特徴ベクトルに変換することによって、従来のアルゴリズムを用いることができる
- 2 形状ベース (Shape-based)
→ 時系列データの形状を抽出することによってクラスタリングを行う。



2.2 時系列データのクラスタリング

本研究で適用する手法

現状の段階では、形状ベースのアルゴリズムである K-shape という手法を適用する。

K-shape

基本的な仕組みは、K-mean 法に基づいている。
距離尺度に基づき、適用するデータの重心を計算し各クラスタの分割を繰り返すを行う。

この手法では、Shape-based distance(SBD) と呼ばれる距離尺度を用いる。

$$SBD(\vec{x}, \vec{y}) = 1 - \max_w \left(\frac{CC_w(\vec{x}, \vec{y})}{\sqrt{R_0(\vec{x}, \vec{x}) \cdot R_0(\vec{y}, \vec{y})}} \right)$$

ただし、

$$CC_w(\vec{x}, \vec{y}) = R_{w-m}(\vec{x}, \vec{y}), \quad w \in \{1, 2, \dots, 2m-1\}$$

$$R_k(\vec{x}, \vec{y}) = \begin{cases} \sum_{l=1}^{m-k} x_{l+k} \cdot y_l, & k \geq 0 \\ R_{-k}(\vec{y}, \vec{x}), & k < 0 \end{cases}$$

1. 本研究について
2. ツイートが為替に影響しているかの検証
3. 為替予測



2.2 時系列データのクラスタ数

エルボー法

各クラスタ中心との距離の二乗の合計をクラスタ内誤差平方和 (SSE) として計算することにより最適なクラスタを求める方法.

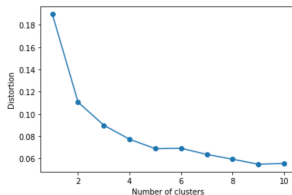


Figure: 1 エルボー法の結果

エルボー法では、肘のように曲がっている点が最適なクラスタであると示唆されるので図よりクラスタ数は4か5が適していると考えられる.

本研究では、クラスタ数は4で分類を行った.



2.3 時系列データのクラスタリングの結果

トランプ大統領のツイート後から 20 分後の為替とランダムな日時の 20 分間の為替の価格それぞれのデータセットを用意してクラスタリングを行った。

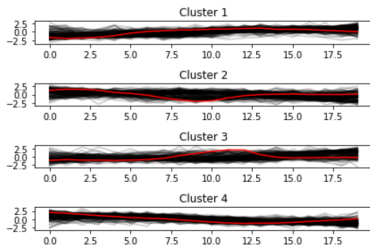


Figure: 2 K-shape による分析結果

分析結果

Cluster2 と Cluster3 にツイート後の為替のデータセットが多く分類された。

1. 本研究について
2. ツイートが為替に影響しているかの検証
3. 為替予測



2.3 時系列データのクラスタリングの考察

1. 本研究について
2. ツイートが為替に影響しているかの検証
3. 為替予測

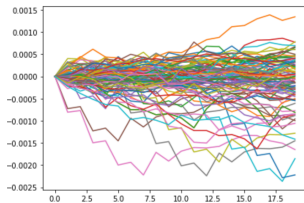


Figure: 3 ツイート後の為替の変動

考察

実際の価格の変動においても、ツイートの7分後から10分後辺りから為替が変動していることがわかる。
そのため、ツイートによる為替への影響はあると考えられる。



3.1 分析手法

分析手法

時系列予測が得意な LSTM を用いる

用いる特徴量

10 分間のデータを 1 セットとして予測を行なっていく

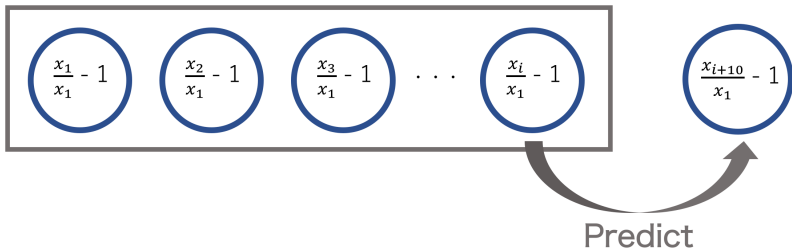


Figure: 4 データセット



3.1 検証 1(為替の価格のみ)

学習データ

2018 年 9 月の 1 ヶ月分の為替データ

	Date	Close	Open	High	Low	Volume
867	2018-09-04 00:00:00	111.042	111.033	111.049	111.031	21
868	2018-09-04 00:01:00	111.031	111.044	111.049	111.029	29
869	2018-09-04 00:02:00	111.047	111.034	111.052	111.034	16
870	2018-09-04 00:03:00	111.056	111.045	111.059	111.045	12
871	2018-09-04 00:04:00	111.059	111.059	111.062	111.057	5

Figure: 5 学習データの 5 日分

本研究では、終値のデータのみを用いる

1. 本研究について
2. ツイートが為替に影響しているかの検証
3. 為替予測



3.2 検証 1 の分析結果

1. 本研究について
2. ツイートが為替に影響しているかの検証
3. 為替予測

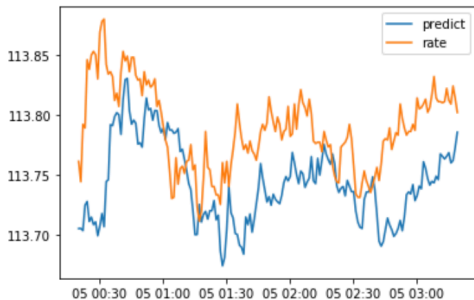


Figure: 6 検証 1 の分析結果

精度

決定係数 : 0.683640



3.3 検証1の考察

- 1. 本研究について
- 2. ツイートが為替に影響しているかの検証
- 3. 為替予測

考察

- 1 予測と実測値の動きが全く逆の場合が多々見られた
- 2 10分後の予測は終値だけでは難しい可能性が大きい
→ 1分後の値の予測は、決定係数が0.85で精度が高めであった



3.4 検証 2(為替とインジケーター)

検証 2 では、検証 1 の特徴量にインジケーターを追加することにより、精度の向上を図る。

精度

用いる特徴量は、終値と 36 種類のインジケータである。

ema_26	ema_9	bb_up	bb_mid	...	apo	ppo	cmo	ht_dcperiod	ht_dcphase	ht_p_inphase	ht_
111.075654	111.046402	111.148167	111.08480	...	-0.029788	-0.026814	-50.573348	19.598470	-1.067944	-0.014234	
111.071292	111.041989	111.145145	111.08020	...	-0.031538	-0.028390	-56.418381	21.458003	-0.743397	-0.016602	
111.068694	111.042247	111.139150	111.07625	...	-0.032301	-0.029078	-31.971764	23.536594	-4.567837	-0.018716	
111.067035	111.044687	111.133171	111.07305	...	-0.032333	-0.029107	-20.557873	24.324707	1.647142	-0.026141	
111.065768	111.047364	111.124396	111.06960	...	-0.031353	-0.028225	-16.927649	25.708262	2.564275	-0.033605	

Figure: 7 インジケーターの一部

1. 本研究について
2. ツイートが為替に影響しているかの検証
3. 為替予測



3.4 今後の課題

- 1. 本研究について
- 2. ツイートが為替に影響しているかの検証
- 3. 為替予測

今後の課題

- 1 検証 2 の実装
- 2 検証 3 において、感情スコアをどのように学習データに追加するかを検討する