

# ソーシャルメディアにおける影響関係から金融市場に対する作用のモデル化と分析

## Modeling and Analysis of Effects on Financial Markets from Influence Relationships in Social Media.

田村 浩一郎

Koichiro Tamura

東京大学 大学院 工学系研究科技術経営戦略学専攻

Department of Technology Management for Innovation, School of Engineering The University of Tokyo

koichiro.tamura@weblab.t.u-tokyo.ac.jp

松尾 豊

Yutaka Matsuo

(同 上)

matsuo@weblab.t.u-tokyo.ac.jp

**keywords:** social media, deep learning, social sensor, social actuator, financial market

### Summary

The data of social media has received much attention to observe and predict real-world events. For example, It is used to predict financial markets, products demand, and voter turnout. While these works regards social media as a sensor of real world, as social media become more popular, it become more natural to think social media significantly effects on real worlds events. The canonical example might be cryptocurrencies, where supply and demand are more susceptible to investor sentiment and therefore interactions within social media cause significant effects on the price of them. On the hypothesis that social media actuate real-world events, we propose a neural network based model to predict the price fluctuations of financial assets, including cryptocurrencies. We model the effect of social media which cannot be directly observed, using an end-to-end neural network, Recurrent Neural Network. By simulating the effect within the social media, we show that the method that models the effect of social media on financial markets can observe and predict the price fluctuations of cryptocurrencies more precisely and stably. By analyzing the model, we suggest that networks within social media can be influential relationships throughout time, even if they are not directly connected, and that the intensity of the influence from social media on financial markets varies depending on the nature of the financial assets.

### 1. はじめに

金融市場の観測や予測をする際、ソーシャルメディアを実世界の射影と捉え、観測・予測に有用な情報源として用いる手法が提案されている [Bollen 11, Kristoufek 13, Nofer 15, Zhang 11]. ソーシャルメディア上の情報は、実世界で起こっていること、それに対する人々の反応を示す情報であり、全体として社会を観測するセンサと考えることができる [松尾 13]. こうした機能は、ソーシャルセンサ [Sakaki 10] と呼ばれ、実世界における社会での事象やトレンドの検知を可能にし、我々の実世界の観測と予測の可能性を広げた. 投資家が何を認知し、どのような感情を抱いているかを反映するセンサとしてソーシャルメディアを用いることで、金融市場の観測や予測をより精緻に行なうことが可能になっている [Bollen 11].

一方で、近年、ソーシャルメディア内の相互作用が、投資家の心理状態の変化を引き起こし、金融市場の価格変動の決定要因の一つになっている可能性がある [Oda 18]. ソーシャルメディアにおいて他者の投稿を閲覧し影響を受けることで、特定の金融資産に対する認知や心理状態

を変化させ、投資行動を行っている投資家が存在する. [Oda 18] では、ある特定のツイッターユーザのツイートによって株価が変動している可能性が言及されている. 本研究は、ソーシャルメディアが、金融市場の価格変動に働きかけるアクチュエータとしても機能しているという仮説に基づく.

暗号資産のような需要と供給がより投資家心理に左右されやすい金融資産の登場に伴い、ソーシャルメディア内の相互作用と心理状態の変化が金融市場に作用する事象をモデル化し、分析する重要性は高まっている. 暗号資産には、ビットコイン、イーサ(イーサリアムの内部通貨、以下イーサリアムと呼ぶ)、リップルなど様々な暗号資産があり、ビットコインは2017年初頭では\$900ほどであったのに対して、2017年12月には\$20,000近くまで価格が上昇し、注目を集めた. 暗号資産はマクロ経済に基づくファンダメンタルズ要素がないとされ [Kristoufek 13]、将来のキャッシュフローモデル [Gordon 62]、購買力平価 [Krugman 08, Reinert 09]、および金利パリティ [Feenstra 08, Levi 07] のような既存の経済理論や金融理論によって、暗号資産の

価格変動を説明することは困難である。暗号資産の需要と供給は、相対的に投資家心理により左右されると推察される。

本研究は、ソーシャルメディアにおける影響関係から金融市場に対する作用をモデル化し、暗号資産をはじめとする金融資産の価格変動をより精緻にかつ安定して予測することを目的とする。そして、モデルを用いた解析によって、ソーシャルメディアにおける影響関係から金融市場に対する作用の分析を行う。

ソーシャルメディアにおけるユーザの影響関係とその金融市場に対する作用は、直接観測することができず、時間を通して相互に複雑に関連しているため、モデル化し分析することは困難であった。金融市場に対して作用をもたらすソーシャルメディア内の影響関係は、投稿やいいね、リツイートといった観測可能な情報拡散を取り扱う information diffusion[Zafarani 14] の研究と異なる。他のユーザの投稿の閲覧に伴う影響関係を含めた、直接観測できないユーザの認知や心理状態の変化を考慮する必要がある。また、他のユーザから繰り返し影響を受けたユーザが、時間を経て他者に影響を与えるという可能性もあり、ソーシャルメディアにおける相互作用とユーザの状態、及び金融市場への影響は、時間を通して複雑に関連している。

そこで本研究は、ソーシャルメディアにおける影響関係から金融市場に対する作用を定式化し、end-to-end のニューラルネットワークである Recurrent Neural Network(以下 RNN と呼ぶ)を用いてモデル化する手法(図 1)を提案する。金融市場で観測される値を予測対象とした損失関数を持つ end-to-end の RNN モデルを学習させることで、直接観測できず、実際の金融資産の価格変動に寄与する相互作用や、時間を通して複雑に変化するユーザの認知や心理状態を抽出・分析ができるようになると期待される。

本稿では、金融市場の予測のタスクに対して、提案手法のモデルと、ソーシャルメディアにおける影響関係から金融市場に対する作用を反映していない複数のモデルを比較検証した。検証実験では、サンプリングされた投資家のツイッターアカウント 9,900 人の 1,133,801 個の有向エッジを持つグラフネットワーク及び 2017 年 1 月 1 日から 2017 年 12 月 31 日までの 9,842,699 件のツイートを集積し、暗号資産市場と日本の株市場の 2 つの金融市場の価格データを用いた。その後、提案手法のモデルのパラメタを解析することによって、ソーシャルメディアにおける影響関係から金融市場への作用に関する分析を行った。

本研究の貢献は、以下の通りである。

- ソーシャルメディアの影響関係から金融市場への作用をモデル化する手法を提案し、既存の価格決定要因では説明することが難しい暗号資産の価格変動をより精緻にかつ安定して予測可能であることを示したこと。

たこと。

- 提案手法のモデルによって、ソーシャルメディアの影響関係から金融市場への作用の分析可能性を高めたこと。その分析の中で、ソーシャルメディア内のネットワークが直接接続されていなくとも、時間を通して影響を与える関係になることや、金融資産の性質によってソーシャルメディアから金融市場への影響の強さが異なることを示唆したこと。

本論文は、2 章でソーシャルメディアを用いた実世界の観測や予測を行うソーシャルセンサ [Sakaki 10] に関する研究や、information diffusion[Zafarani 14] の研究を俯瞰し、本研究の学術的見地を明確にする。3 章で提案手法について説明し、4 章では、検証実験について述べる。5 章では、検証実験における結果を踏まえてソーシャルメディアの影響関係から金融市場への作用の分析をする。最後に、6 章で結論を述べる。

## 2. 関連研究

ソーシャルメディアにおける情報は、実世界で何が起こっているのか、そしてそれに対する人々の反応を示す情報であり、社会で起こりつつあるイベントやトレンド、現象を検出するソーシャルセンサとして機能している [松尾 13]。ソーシャルセンサを通して得られた出力から、統計処理などによって実世界の観測や予測を行うことができ、ソーシャルセンサを用いた実世界の観測や予測に関する研究がなされてきた。

例えば、ソーシャルセンサによって、地震が起こった時に、ツイッターにおける地震に関する投稿から地震を検知することができる [Sakaki 10]。将来の映画の収益をツイッターから得られる情報から予測したり [Asur 10]、インスタグラムの投稿からウェブ上で人気になったり関心を集める事象を予測することができる [Ferrara 14]。選挙における投票では、ツイッターから得られる情報を用いることで当選者を予測する研究 [Tumasjan 10, Wang 12] が行われている。経済指標をツイッターから得られるセンチメント情報によって予測するモデルも提案されており [Bollen 11, Nofer 15, Zhang 11]、Google トレンドとビットコインのマーケットの相関の存在なども示されている [Kristoufek 13]。

近年はソーシャルメディアを用いた観測と予測を、現実のタスクにおいてより精緻に行うための研究が続けられている。[Brambilla 17] では、分析対象となるコンテンツの新規性の高さゆえ、観測頻度が低く、抽出が難しいコンテンツを抽出する手法を提案している。[Shao 17] では、ソーシャルメディアにおいて特定の領域の分析を行うには、サブグラフを特定する必要があるが、ソーシャルメディアのような動的で多次元的なネットワークにおいてそうしたサブグラフを効率的に抽出する手法が研究されている。また、ユーザの人気、social rank を考慮し

た network embedding の手法 [Gu 18] は、大規模なネットワーク構造を持つソーシャルメディアの活用可能性を高めた。

ソーシャルセンサに関する既存研究は、実世界の観測と予測の可能性を広げた一方で、ソーシャルメディアを単なる実世界の射影や、実世界を観測・予測したりするための情報源としてしかみなしておらず、ソーシャルメディアに形成される社会ネットワーク内の相互作用が、実世界に働きかける可能性について考慮した研究は行われてこなかった。実世界におけるイベントがソーシャルメディアのネットワーク構造や内部のコミュニケーションに影響を与えることを分析した研究 [Romero 16] や、ソーシャルメディアのユーザの時系列状態を隠れマルコフモデル(以下 HMM と呼ぶ)を用いて表現し、ユーザのソーシャルメディア内における行動を予測する研究 [De Choudhury 09] はなされている。しかし、ソーシャルメディアに形成される社会ネットワークにおいて、投稿の閲覧に伴う影響関係や、ユーザの認知・心理状態の変化が、実世界に対して働きかけていることを考慮し、実世界の観測と予測を行う研究はなされていない。

また、ソーシャルメディアにおけるユーザの影響関係をモデル化する際、直接観測することのできない影響関係をモデル化することは難しく、こうした研究は少ない。ソーシャルネットワークにおける影響関係をモデル化する既存手法として、Linear Threshold Model(以下 LTM と呼ぶ)[Kempe 03] がある。しかし、LTM では、ユーザ  $u_i$  がユーザ  $u_j$  から受けた影響度を直接知ることができないため、影響度をネットワークグラフの指標などで代替して表現する場合が多かった [Zafarani 14]。金融市場に対して影響を及ぼすソーシャルメディアの相互作用をモデル化する場合、金融資産の性質やユーザの興味関心といった性質を反映し、直接観測することのできないユーザの認知や心理状態の変化をより精緻に考慮する必要がある。

ソーシャルメディアのネットワークにおける Information cascades の研究 [Bourigault 16, Cao 17, Li 17, Wang 17] では、Gated Recurrent Unit[Hochreiter 97](以下、GRU と呼ぶ)を用いて、情報拡散するネットワークの状態を end-to-end で表現・獲得することを試みている。しかし、これらの研究は直接観測される情報拡散を観測・予測対象としており、問題設定が本研究と異なるため、投稿に対するいいねやリツイート、リプライといった直接観測可能な information flow のみしか取り扱っておらず、ソーシャルメディアにおいてアクションがないユーザに関しては考慮していない。ソーシャルメディアに形成される社会ネットワークから金融市場への作用を考慮し、金融市場の観測・予測を行う場合、直接観測することができないような、他者の投稿を閲覧することによるユーザ間の影響関係や、時間を通して変化・蓄積されるユーザの認知・心理状態をモデル化することが必要となる。

### 3. 提案手法

#### 3.1 ソーシャルメディアの影響関係から金融市場への作用の定式化

ソーシャルメディアの一部のユーザは、他者の投稿を閲覧し、特定の金融資産に対する認知や心理状態を変化させ、意思決定を行っている。こうして新たな投資行動を引き起こし、結果として金融市場が変動していると推察できる。本研究では、モデル化するソーシャルメディアの影響関係から金融市場への作用を以下のステップに定式化する。

1. 観測と投稿 ユーザ  $u_j$  が金融市場に関わる事象 X を観測し、事象 X に関連する投稿をソーシャルメディアに行う。
2. 影響 ユーザ  $u_j$  の投稿を閲覧したユーザ  $u_i$  は、事象 X に対する認知や心理状態に影響を受け、場合によっては事象 X に関連する発信を行い、ソーシャルメディアにおいて情報拡散が起こる。
3. 認知と心理変化 ユーザ  $u_i$  や、ユーザ  $u_i$  の投稿を閲覧した他のユーザが、さらに事象 X に対する認知を高めたり、事象 X に対する心理状態を変化させる。
4. 行動 ソーシャルメディアから認知や心理状態に対して影響を受けたユーザが、実際に事象 X に関連する投資行動を行い、金融市場が変動する。

#### 3.2 end-to-end のネットワークを用いたモデル化

次に、3.1 節で定式化したソーシャルメディアの影響関係から金融市場に対する作用を、end-to-end のニューラルネットワークである RNN を用いてモデル化する手法(図 1)を説明する。

金融市場の観測値を予測対象とした損失関数を持つ end-to-end の RNN モデルを学習させることで、実際の金融資産の価格変動に寄与する相互作用や、時間を通して複雑に変化するユーザの認知や心理状態を抽出・分析できると期待される。

モデルの概要は図 1 の通り。まず、ソーシャルメディアから特定の金融資産に関連するユーザの投稿の有無を表す one-hot ベクトルを入力とし、ユーザの影響関係を表す行列  $E$  との行列積を計算することで、時刻  $t$  にユーザが受けた影響ベクトル  $r_t$  を獲得する。その後 RNN の一種である GRU を用いて時系列におけるユーザの認知・心理状態を表す  $s_t$  を獲得し、最終層を経て金融市場の予測を出力する。

なお、本研究では、ソーシャルメディアにおける影響関係から金融市場に対する作用をモデル化することで、金融資産の価格変動をより精緻にかつ安定して予測することを目的としている。そのため、モデルの詳細は次節以降で説明するが、ニューラルネットワークの構造や具体的な学習手法は、その性能や分析可能性を考慮し、異なるものを適用しても良い。

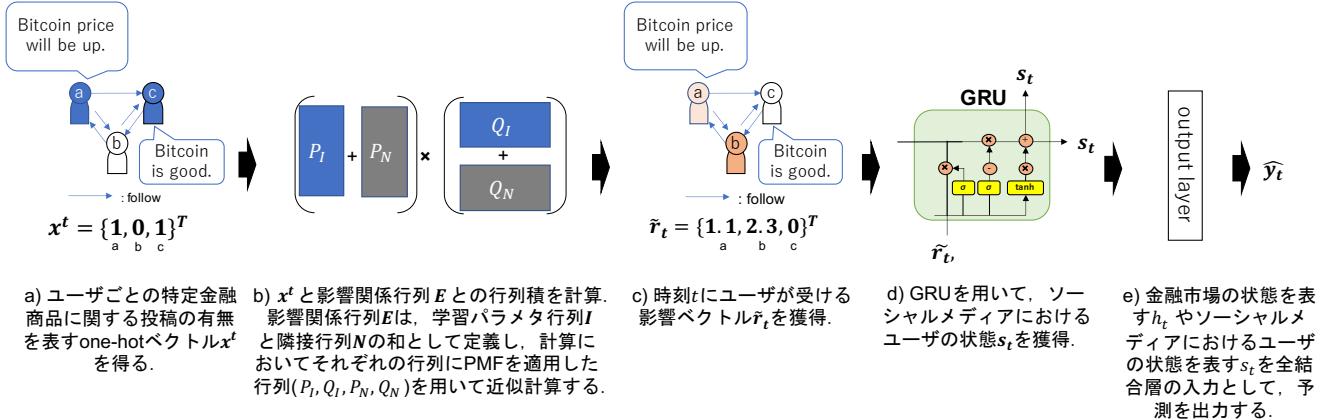


図1 end-to-end のネットワークを用いたモデルの概要

### 3.3 影響関係行列 $E$ による影響関係の表現と抽出

ソーシャルメディアを用いて金融市場の予測を行う場合、リプライやリツイートなどの直接観測されるユーザーだけでなく、閲覧によって引き起こされる直接観測できないユーザーの影響関係を考慮することが望ましい。

LTM[Kempe 03]では、ユーザー $u_j$ からユーザー $u_i$ に対する影響度 $e_{j,i}$ は直接観測できないので、影響度 $e_{j,i}$ をフォロワー数などのネットワーク指標で代用するが、本研究のように時間を通したユーザーの影響関係を取り扱う際に、ネットワーク指標のみを影響関係指標とすることには、以下の2つの課題がある。

- (1) ユーザ $u_i$ がユーザー $u_j$ をフォローし投稿を閲覧する場合でも、ユーザー $u_i$ がユーザー $u_j$ を信用していない、またはユーザー $u_j$ が金融市場に関連のない投稿をしている場合は、金融市場に作用をもたらすほど影響関係は実際には存在せず、相互作用の関係を過大評価する可能性がある。
- (2) ユーザ $u_i$ がユーザー $u_j$ をフォローしておらず、直接投稿を閲覧しない場合でも、ユーザー $u_i$ の興味関心がユーザー $u_j$ の投稿する情報と類似している場合、他のユーザーを介した影響の連鎖によって、時間を通して影響関係にある可能性がある。

そこで本研究では、閲覧などによって引き起こされるユーザー $u_i$ がユーザー $u_j$ から受ける影響 $s_{j,i}$ を要素にもつ影響関係行列 $E$ をend-to-endネットワークの学習パラメタとして定義し、金融市場で観測される値を予測対象とした損失関数を用いた学習過程で表現獲得することを試みる。

また、ユーザー $u_i$ がユーザー $u_j$ をどのように評価し、影響を受けるかという $e_{j,i}$ を推定するタスクと、それに関連する上述した課題は、未知なる2者間の評価関係を推定するという枠組みにおいて、レコメンドアルゴリズムにおけるユーザー $u_i$ のアイテム $i_j$ に対する評価を推定するタスクおよびその課題に類推できる。[Liang 16]の研究では、ユーザー $u_i$ がアイテム $i_j$ を閲覧することと、ユーザー $u_i$ がアイテム $i_j$ に対して実際に影響を受けることを

別々の行列で表すことで、上述した二つの課題に類似する問題の解決を試みている。そこで本研究も、対象とするソーシャルメディアのユーザー $M$ 人のうち、ユーザー $u_i$ がユーザー $u_j$ の投稿を閲覧することを、ソーシャルメディアのフォロー関係を表す隣接行列 $N$ によって表現し、ユーザー $u_i$ がユーザー $u_j$ の投稿を閲覧した時に受ける影響度を、学習パラメタ行列 $I$ によって表す。そして、隣接行列 $N$ と学習パラメタ行列 $I$ の和によって、影響関係行列 $E \in \mathbf{R}^{M \times M}$ を定義する。

$$E = N + I \quad (1)$$

ただし、行列 $N$ の要素 $n_{i,j}$ はユーザー $u_i$ がユーザー $u_j$ をフォローしているか否かを表すバイナリ値であり、学習パラメタ行列 $I$ の要素 $i_{i,j}$ は閲覧などに伴いユーザー $u_i$ がユーザー $u_j$ から受ける影響を表現する値である。

そして、ある時刻 $t$ において、 $K$ 人のユーザーが、対象としている金融資産に関する投稿をしたかどうかを表したone-hotベクトル $x_t \in \mathbf{R}^M$ と影響関係行列 $E$ の行列積によって、ある時刻 $t$ にユーザーが受ける影響ベクトル $r_t \in \mathbf{R}^M$ を獲得する。

$$r_t = Ex_t \quad (2)$$

ソーシャルメディアのような高次元のグラフを取り扱う時、対象とする $K$ 人のユーザーのすべての影響関係を学習することは計算量が膨大である上、ユーザーの投稿の有無はスパースなデータであり、学習過程において過学習をする恐れもある。高次元なグラフネットワークを低次元で表現する手法として、Probabilistic Matrix Factorization[Mnih 08](以下PMF)を用いる。影響関係行列 $E$ を構成する隣接行列 $N$ に対してPMFを行い、行列 $P_N \in \mathbf{R}^{M \times K}$ と行列 $Q_N \in \mathbf{R}^{K \times M}$ に分解し、 $I$ に基づく学習パラメタ行列 $P_I \in \mathbf{R}^{M \times K}$ と学習パラメタ行列 $Q_I \in \mathbf{R}^{K \times M}$ を定義することで、式(1)の影響関係行列 $E$ を以下のように近似する。

$$E \sim (P_I + P_N)(Q_I + Q_N)$$

よって、式(2)は、以下のように近似され表される。

$$\mathbf{r}_t \simeq (\mathbf{P}_I + \mathbf{P}_N)(\mathbf{Q}_I + \mathbf{Q}_N)\mathbf{x}_t \quad (3)$$

### 3.4 認知・心理の時系列内部状態の獲得

ソーシャルメディアのユーザは、他のユーザの影響を受けてからすぐに関連する投稿をしたり、投資行動を起こすとは限らない。例えば、ユーザ  $u_i$  は他のユーザ  $u_j$  の金融資産  $X$  に対する投稿を繰り返し閲覧することによって、その金融資産  $X$  に関して認知を深め、興味関心や心理状態が変化することで、投資を行う場合がある。この時、時間を通した影響関係やソーシャルメディアのユーザの認知や心理状態を考慮する必要がある。

ネットワークの状態は、直接観測することができないため、HMM によって表現する手法 [De Choudhury 09] を用いることが考えられる。しかし、ユーザの影響関係と、認知・心理状態の時間変化は相互に密接に関連すると考えられ、本研究では end-to-end のアルゴリズムであり、時系列モデルを取り扱う RNN の一種である GRU [Hochreiter 97] を用いる。

まず、式(3)で獲得された影響ベクトル  $\mathbf{r}_t$  に対して、次の処理を行い、埋め込みベクトル  $\tilde{\mathbf{r}}_t \in \mathbf{R}^{hidden}$  を得る。

$$\tilde{\mathbf{r}}_t = \text{dropout}(\tanh(\mathbf{W}_{in} \frac{\mathbf{r}_t}{\sqrt{M}} + \mathbf{b}_{in}))$$

ただし、*dropout* は学習時に用いる Dropout[Srivastava 14] 層、 $M$  は対象とするソーシャルメディアのユーザ数、 $\mathbf{W}_{in}$  および  $\mathbf{b}_{in}$  は学習パラメタである。

そして、時刻  $t$  において計算できるソーシャルメディアのユーザの認知・心理状態を表す  $\mathbf{s}_t$  を、GRU を用いて獲得する。

$$\mathbf{s}_t = \text{GRU}(\tilde{\mathbf{r}}_t) \quad (4)$$

一方で、獲得された  $\mathbf{s}_t$  と、タイムステップ  $t$  において計算できる金融市場の価格変動を表す特徴量  $\mathbf{m}_t$  に対し、GRU を適用し、金融市場の状態を表す特徴量  $\mathbf{h}_t$  を獲得する。

$$\mathbf{h}_t = \text{GRU}(\tilde{\mathbf{m}}_t) \quad (5)$$

金融資産の特性に応じて、ソーシャルメディアの重要性が異なる可能性があり、ソーシャルメディアが金融市場の価格変動の予測にどれだけ寄与しているのかという分析の可能性を高めるため、本研究では  $\mathbf{s}_t$  および  $\mathbf{h}_t$  を 1 次元の値に変換し、Batch Normalization[Ioffe 15] を用いて正規化したのち最終層に入力して予測を行う。

$$s_t = \tanh(\mathbf{W}_{s-out}\mathbf{s}_t + \mathbf{b}_{s-out})$$

$$h_t = \tanh(\mathbf{W}_{h-out}\mathbf{h}_t + \mathbf{b}_{h-out})$$

$$\hat{y}_t = \tanh(\mathbf{W}_{out} \text{concat}(\mathbf{s}_t, \mathbf{h}_t) + \mathbf{b}_{out}) \quad (6)$$

ただし、*concat*( $\mathbf{s}_t, \mathbf{h}_t$ ) は  $\mathbf{s}_t$  と  $\mathbf{h}_t$  の連結ベクトルであり、 $\mathbf{W}_{s-out}$ ,  $\mathbf{b}_{s-out}$ ,  $\mathbf{W}_{h-out}$ ,  $\mathbf{b}_{h-out}$ ,  $\mathbf{W}_{out}$ ,  $\mathbf{b}_{out}$  は学習パラメタ、 $\hat{y}_t$  は時刻  $t$  に出力される予測値である。

## 4. 実験

### 4.1 実験設定

#### §1 実験対象

本研究ではまず、ツイッターのデータを用いて、暗号資産の将来の価格変動のボラティリティを予測する実験を行い、提案手法の学習モデルの精度を評価する。ソーシャルメディアの影響関係から金融市場に対する作用をモデル化する提案手法が、既存の価格決定要因では説明することが難しい暗号資産の価格変動をより精緻に予測できることを検証する。本実験では、ビットコイン、イーサリアム、リップルの暗号資産市場で時価総額が大きい 3 つの暗号資産について、将来のボラティリティの予測を行う。予測するボラティリティは、予測する時点から特定タイムステップ経過後までの期間の、1 タイムステップ前からの価格変化率の標準偏差とする。

#### §2 データセット

実験には、ビットコイン、イーサリアム、リップルの暗号資産市場で時価総額が大きい 3 つの暗号資産の価格データ、Twitter REST API<sup>\*1</sup> から取得した投資家のツイートおよびフォロー/フォロワー関係のデータを用いる。

暗号資産市場のデータは、2017 年 1 月 1 日から 2017 年 12 月 31 日までの 1 年間の価格の日足データを用いる。なお、本実験では 1 タイムステップの  $\Delta t = 1$  は 1 日とする。取得したデータから、以下の要素を計算し、時刻  $t$  における暗号資産の価格変動を表す特徴量  $\mathbf{m}_t$  として利用する。

- (1) *ChangePrice<sub>Δt</sub>*: 時刻  $t$  から  $\Delta t = 1, 3, 5, 10, 25$  日前からの終値変化率
- (2) *DeviationRate<sub>Δt</sub>*: 時刻  $t$  における  $\Delta t = 3, 5, 10, 25$  日間の終値移動平均からの乖離率
- (3) *Volatility<sub>Δt</sub>*: 時刻  $t$  における  $\Delta t = 3, 5, 10, 25$  日間のボラティリティ(過去  $\Delta t$  日間の、1 日前からの終値変化率 *ChangePrice<sub>1</sub>* の標準偏差)

また、時刻  $t$  から 1, 3, 5, 10 日後までの期間の将来のボラティリティを計算し、予測対象とする。

ツイッターのデータは、[Oda 18] で言及されているある投資家ツイッターアカウントをフォローしている約 200,000 人(2018 年 1 月時点)のユーザのうち、2017 年 1 月から 4 月までの 3 ヶ月間で 5 回以上ツイートがあるアクティブユーザで、かつフォロワーが 100 人以上いる 9,900 人のユーザのアカウントを対象とする。そして、対

\*1 <https://dev.twitter.com/twitterkit/android/access-rest-api>

象アカウント 9,900 人の全てのツイート、フォローしているユーザ、フォロワーのユーザを独自に獲得した。全部で、2017年1月1日から2017年12月31日までの9,842,699件のツイートおよび、1,133,801個の有向エッジを持つ9,900人の対象ユーザのグラフネットワークを分析対象とした。

### §3 評価方法

評価方法として、暗号資産市場の将来のボラティリティの予測においては、RMSEを用いる。 $\hat{y}$ を予測値、 $y$ を実際の値とすると、RMSEは以下の式で得られる。

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2}.$$

暗号資産市場のデータは、1月1日から10月31日までの約300レコードを学習データとし、11月1日から12月31日までの60レコードをテストデータとして用いて評価を行う。

### §4 ハイパーパラメタ

モデルの学習には、Adamを用いて、学習率を0.001とする。トレーニングステップは最大1000として、100ステップごとにバリデーションスコアを確認し、バリデーションスコアが前回の値を下回った場合に学習を終了させる。バリデーションのためのデータセットは、学習データの時系列の最後から32ステップとした。ドロップアウト率を0.3、影響ベクトル $r_t$ の次元を $2^6$ 、隠れ層ノードの数を $2^4$ とする。提案モデルのRNNにおいて、本検証実験では時系列長を5タイムステップとした。

## 4.2 比較モデル

実験の目的は、ソーシャルメディアの影響関係から金融市場に対する作用を、end-to-endのニューラルネットワークであるRNNを用いてモデル化する手法が、既存の価格決定要因では説明することが難しい暗号資産の価格変動をより精緻に予測できることを検証することである。

提案手法の評価の正当性を保つために、比較手法は、提案手法のモデルをベースに、既存研究との差分の構造を変えた以下の3つのモデルとする。

- (1) **Market:** 提案手法のモデルにおいて、ソーシャルメディアから得られる情報を用いず、金融市場のデータのみを扱う時系列モデル
- (2) **Social Sensor:** 提案手法のモデルにおいて、式(4)にソーシャルメディアにおける投稿に関する統計量を用いた、ソーシャルセンサの時系列モデル
- (3) **LTM:** 提案手法のモデルにおいて、式(4)にソーシャルメディアの影響関係としてLTMを利用した時系列モデル

### §1 Market: ソーシャルメディアから得られる情報を用いず、金融市場のデータのみを扱う時系列予測モデル

ソーシャルメディアから取得可能なデータを用いずに、上述した時刻 $t$ における暗号資産の価格変動を表す特徴量 $\mathbf{m}_t$ のみを入力として、GRUを用いた予測を行う。出力として、以下を得る。

$$\tilde{\mathbf{h}}_t = GRU(\mathbf{m}_t)$$

$$\hat{y} = \mathbf{W}_{out} \tilde{\mathbf{h}}_t + b_{out},$$

ただし、3章と同様に、 $\mathbf{W}_{out}, b_{out}$ は学習パラメタである。

### §2 Social Sensor: ソーシャルメディアにおける投稿に関する統計量を用いた、ソーシャルセンサの時系列モデル

次に、金融市場のデータ $\mathbf{m}_t$ に加えて、ソーシャルメディアにおいて対象とする $M$ 人のユーザのうち、時刻 $t-1$ から $t$ までの間に特定の金融資産についての言及を含む投稿したユーザの割合 $num_t$ を用いて、金融市場の予測を行う。

$$\tilde{\mathbf{h}}_t = GRU(concat(\mathbf{m}_t, num_t))$$

$$\hat{y} = \mathbf{W}_{out} \tilde{\mathbf{h}}_t + b_{out},$$

ただし、 $concat(\mathbf{m}_t, num_t)$ は $num_t$ と $\mathbf{m}_t$ を連結したベクトルであり、 $\mathbf{W}_{out}, b_{out}$ は学習パラメタである。

暗号資産市場に関する投稿は、ビットコイン、イーサリアム、リップルのそれぞれの暗号資産において以下のワードが含まれている投稿を対象とした。

ビットコイン: ビットコイン, Bitcoin, BTC

イーサリアム: イーサリアム, Ethereum, ETH

リップル: リップル, Ripple, XPR

### §3 LTM: ソーシャルメディアにおける閲覧による影響関係を考慮した、LTMを利用した時系列モデル

ソーシャルメディアにおける閲覧に伴う影響関係を考慮するモデルとしてLTMを用いることが考えられる。LTMでは、ユーザ $u_j$ からユーザ $u_i$ に対する影響度 $s_{j,i}$ を用いるが、影響度 $s_{j,i}$ を直接観測することができないので、影響度をネットワーク指標などで代用して表す必要がある。[Zafarani 14]において影響度をフォロワー数で表現するモデルが示されており、またinformation cascadesのようなソーシャルメディアにおけるネットワーク構造の将来予測に関しても、ソーシャルネットワークの構造を表すネットワーク指標を入力として用いることの有用性が示されている。

そこで、Social Sensorモデルのように、単にユーザの投稿割合を入力にするのではなく、その投稿を行ったユーザのネットワーク指標によって重み付けした値 $num_t$ を入力として用いる。

$$num_t = \frac{1}{\sqrt{M}} \sum_{i=1}^n degree(user_i)$$

表1 提案手法と比較手法のデータ利用・構造の違い

	Market	Social Sensor	LTM	Proposed Model
金融市場データの利用	○	○	○	○
ソーシャルメディアの投稿に関する統計量の利用	×	○	○	○
ネットワークにおける閲覧を伴う影響関係の利用	×	×	○	○
金融資産に対する時間を通した影響関係と心理状態の利用	×	×	×	○

ただし,  $M$  は検証実験で対象としているユーザ数,  $n$  は時刻  $t$  において特定の金融資産に関する投稿を行ったユーザの数, そして  $degree(user_i)$  は  $user_i$  のソーシャルメディアにおける対象としているユーザのネットワーク指標(本研究では, フォロワー数とする)を表す.

出力として, 以下を得る.

$$\tilde{h}_t = GRU(concat(\mathbf{m}_t, num_t))$$

$$\hat{y} = \mathbf{W}_{out}\tilde{h}_t + b_{out},$$

ただし,  $concat(\mathbf{m}_t, num_t)$  は  $num_t$  と  $\mathbf{m}_t$  を連結したベクトルであり,  $\mathbf{W}_{out}, b_{out}$  は学習パラメタである.

### 4.3 実験結果

ビットコイン, イーサリアム, リップルの将来の価格変動ボラティリティ予測の結果は, それぞれ表2, 表3, 表4の通り.

表2 ビットコインの将来のボラティリティ予測における提案手法のRMSE比較

	1日後	3日後	5日後	10日後
Market	0.0230	0.0309	0.0323	0.0420
Social Sensor	0.0394	0.0416	0.0307	0.0459
LTM	0.0317	0.0323	0.0392	0.0431
Proposed model	<b>0.0215</b>	<b>0.0220</b>	<b>0.0267</b>	<b>0.0414</b>

表3 イーサリアムの将来のボラティリティ予測における提案手法のRMSE比較

	1日後	3日後	5日後	10日後
Market	<b>0.0146</b>	0.0200	0.0245	0.0269
Social Sensor	0.0176	0.0218	0.0244	0.0239
LTM	0.0160	0.0196	0.0253	0.0243
Proposed model	0.0184	<b>0.0184</b>	<b>0.0226</b>	<b>0.0237</b>

表4 リップルの将来のボラティリティ予測における提案手法のRMSE比較

	1日後	3日後	5日後	10日後
Market	0.0654	0.0813	0.0931	0.1042
Social Sensor	<b>0.0572</b>	0.0730	0.0957	0.0993
LTM	0.0617	0.0728	0.0900	0.0994
Proposed model	0.0627	<b>0.0655</b>	<b>0.0790</b>	<b>0.0938</b>

表2, 表3, 表4 からわかるように, ほとんどの比較検証実験の場合において, 提案手法が他の手法よりも精緻に将来のボラティリティの予測を行っている. ソーシャルメディアにおける影響関係から金融市場への作用をモデル化する提案手法が, 既存の価格決定要因では説明することが難しい暗号資産の価格変動をより精緻に予測できることを示唆する.

一方で, 表3, 表4において, 予測時点から1日後のイーサリアムおよびリップルのボラティリティの予測では, 提案手法はMarketやSocial Sensorの手法を上回ることができなかった. 予測時点直後のボラティリティに関しては, ソーシャルメディアにおけるユーザの状態が反映されるよりも, 現在のボラティリティの傾向が継続しやすく, Marketのような比較的単純な手法の方が, より精緻に予測ができていたのだと推察できる.

また, Social SensorやLTMの手法は, ソーシャルメディアの情報を用いないMarketに対して, 安定的に優れた予測精度を出すことができていない. 3種類の暗号資産の, 1, 3, 5, 10日後の予測実験の計12パターンの実験において, Social Sensorの手法は, 5パターンがMarketの精度を上回ることができず, LTMの手法は, 6パターンがMarketの精度を上回ることができなかった. 特にビットコインの将来ボラティリティ予測では, Social Sensorの手法は, 5日後の予測のみMarketの精度を上回り, LTMの手法においては, Marketの精度を上回ることができていない.

Social SensorやLTMの手法では, ソーシャルメディアの投稿の閲覧に伴う影響関係や, 時間を通して変化・蓄積されるユーザの認知や心理状態を抽出・表現することができおらず, 将来の予測という枠組みでは, 将来に対する影響が少なく, ただ価格変動に対して反応したユーザの投稿が, かえってノイズとなってしまっている可能性がある. 一方で, 提案手法は, 複雑でかつ直接観測されないソーシャルメディアの投稿の閲覧に伴う影響関係や, 時間を通して変化・蓄積されるユーザの認知や心理状態を抽出・表現するため, そうしたノイズに対して頑強で, 表2, 表3, 表4で示されるように, より安定して高い精度を出すことができていると考えられる.

## 5. 分析

### 5.1 ソーシャルメディアの社会ネットワークにおける影響関係の抽出と分析

ソーシャルメディアにおいて、リツイートなどの情報拡散だけでなく、閲覧を伴う直接観測されない情報拡散も存在するはずである。また、フォロー/フォロワーという直接的な関係だからと言って、どれだけ影響を受けるかどうかはわからないし、時間を通してユーザーの認知や心理状態が変化することで、フォロー/フォロワーという直接的な関係がなくても、影響関係にある可能性がある。

そこで、提案手法の学習モデルのユーザー間の影響度を表すパラメタ行列  $E$  を分析することで、ソーシャルメディアにおける影響関係から金融市場への作用を考慮した、ソーシャルメディア内のユーザー同士の影響関係を分析する。

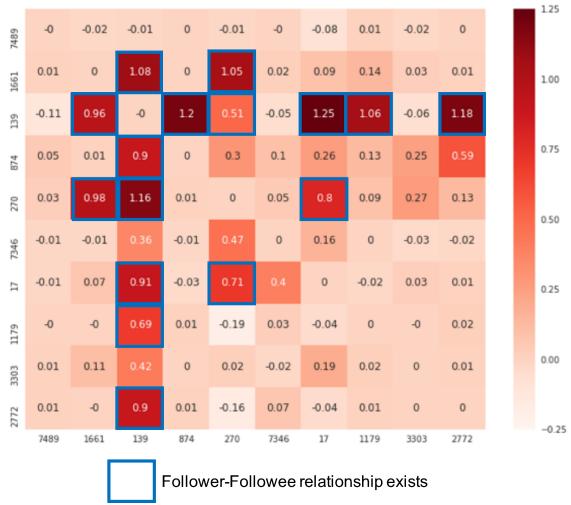


図2 ビットコインの3日後のボラティリティ予測の検証実験で獲得された影響関係を表す行列  $E$  において、サンプリングされた10アカウントの影響関係のヒートマップ。値が大きいほど影響が大きいことを示している。

図2は、ビットコインの3日後のボラティリティ予測の検証実験で獲得された影響関係を表す行列  $E$  において、サンプリングされた10アカウントの影響関係に対応するパラメタをヒートマップで表したものであり、値が大きいほど影響度が高い。行は影響を与えるユーザーを表しており、列は影響を受けるユーザーを表している。また、青四角で囲まれたユーザー関係は、ツイッターにおいて列のユーザーが行のユーザーをフォローしていることを表している。

提案手法を用いて、複雑でかつ直接観測されないソーシャルメディアの投稿の閲覧に伴う影響関係を陽に表現する手法であり、図2のように、誰が誰にどれだけ影響を与えていたかを獲得し、可視化することが可能である。

図2を見ると、青四角で囲まれておらず、フォローしていないが影響度が高く抽出されている関係があること

がわかる。例えば、あるユーザー  $u_{2772}$  がユーザー  $u_{874}$  から与えられる影響度は、 $E_{(874,2772)} = 0.59$  であり、フォロー関係になかったにもかかわらず強い影響関係が見て取れる。ユーザー  $u_{2772}$  は、実際に、暗号資産市場に関する多くのつぶやきを投稿していたことが確認され、ユーザー  $u_{2772}$  はビットコインに対して強い関心を持ち、実際にビットコインに対して投資をしていると推察される。一方、ユーザー  $u_{874}$  はツイッターで有名な暗号資産ディーラーであり、ビットコインや他の暗号資産について何度も投稿し、情報発信を行っていた。

この結果は、ソーシャルメディアのネットワークにおける直接的な関係がなくとも、お互いの興味関心と情報発信の内容が類似しているユーザーは、時間を通して影響関係になる可能性を示唆している。

### 5.2 金融資産の性質に対するソーシャルメディア特徴量の予測に対する重要度の分析

実験では、暗号通貨市場を対象に検証を行ったが、ソーシャルメディアにおける影響関係が金融市場に作用するその度合いは、金融資産の性質にも依存するという仮説が成り立つ。そこで、需要と供給がより投資家の心理に左右されると推察される暗号資産だけでなく、日本の株式市場において提案手法を適用し、事業内容や市場特性ごとに最終層のパラメタ  $W_{out}$  を可視化することによって、 $s_t$  と  $h_t$  の予測に対する重要度を獲得し、各金融資産の性質に応じてソーシャルメディアにおける事象がどれだけ予測に寄与しているかを相対的に分析・考察する。

ツイッターデータは4章と同様のものを用いて、日本の株式市場のデータは、Yahoo Finance<sup>\*2</sup> から得られる2017年1月1日から2017年12月31日までの1年間のデータを利用する。価格変動を表す特徴量  $m_t$  は、4章と同様に計算する。関連するソーシャルメディアへの投稿は、日本株の証券コードおよび企業名を含む投稿を対象とした。なお、式(6)の入力変数は、共に独立であることが証明できず、かつ標準化されていないため、パラメタの絶対値から予測に対する重要度を議論することはできないことに注意する。 $W_{out}$  におけるパラメタの初期値を全て0.5にし、日本株式の各金融資産ごとに5日後の価格変動のボラティリティ予測のための学習を行い、学習後の  $W_{out}$  のパラメタを測定することによって、ソーシャルメディアにおける事象が金融市場の予測に相対的に寄与しやすい特性を分析する。

付録Aに記述する事業、市場ごとに  $W_{out}$  の分析を行った結果は以下の通り。

各事業・市場ごとに、学習過程における  $W_{out}$  の推移を見てみると、自動車関連事業および銀行関連事業はツイッターにおける事象を表す変数  $s_t$  に対する重みよりも、金融市場における価格変動を表す変数  $h_t$  に対する重

\*2 <https://finance.yahoo.co.jp/>

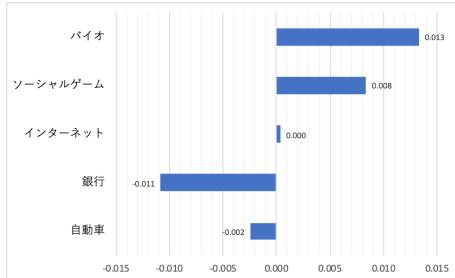


図3 付録Aに記述する事業ごとの  $W_{out}$  の分析結果. 金融市場における価格変動を表す変数  $h_t$  への重みに対して, ソーシャルメディアにおける事象を表す変数  $s_t$  への重みの差分を表している. 値が大きいほどソーシャルメディアが予測に寄与する比重が大きい. バイオ事業やソーシャルゲーム事業において, よりソーシャルメディアにおける事象を表す変数  $s_t$  が予測に寄与していることがわかる.

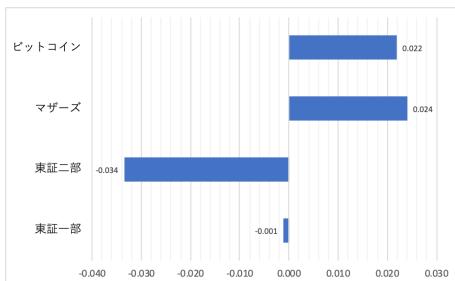


図4 付録Aに記述する市場ごとの  $W_{out}$  の分析結果. 金融市場における価格変動を表す変数  $h_t$  への重みに対して, ソーシャルメディアにおける事象を表す変数  $s_t$  への重みの差分を表している. ビットコイン市場やマザーズ市場において, よりソーシャルメディアにおける事象を表す変数  $s_t$  が予測に寄与していることがわかる.

みが大きくなっていることがわかる. 一方で, ソーシャルゲームやバイオ関連事業は, ツイッターにおける事象を表す変数  $s_t$  に対する重みが, 金融市場における価格変動を表す変数  $h_t$  に対する重みよりも大きくなっていることがわかる. インターネット関連事業は, それぞれの重みはそれっぽ等しかった. また, 東証1部市場の銘柄は, より金融市場における価格変動を表す変数  $h_t$  に対する重みが大きかったのに対して, マザーズ市場やビットコインは, よりツイッターにおける事象を表す変数  $s_t$  に対する重みが大きかった.

自動車や銀行関連銘柄, 東証一部の銘柄などの事業規模も大きい成熟産業銘柄よりも, ソーシャルゲームやバイオ関連事業, またマザーズ市場やビットコインといった, 時価総額が小さく, プレスリリースなどの1つのニュースでも価格変動が比較的簡単に起こる金融資産は, ソーシャルメディアのユーザである個人投資家が好む金融資産である. 金融資産・市場の性質によって, ソーシャルメディアから金融市場への影響の強さが異なることが示唆される.

## 6. 終わりに

ソーシャルメディアは, 金融市場において投資家が何を認知し, どのような感情を抱いているかを獲得することができるセンサとして機能しているだけでなく, 金融市場の価格変動に働きかけるアクチュエータとしても機能している可能性がある. そこで本研究は, 直接観測することができず, 時間を通して相互に複雑に関連している, ソーシャルメディアにおける影響関係から金融市場への作用を, RNNを用いてモデル化する手法を提案した.

実験を通して, ソーシャルメディアにおける影響関係から金融市場への作用をモデル化する提案手法が, 既存の価格決定要因では説明することが難しい暗号資産の価格変動をより精緻にかつ安定して予測できることを示した. また, 提案手法のモデルは, ソーシャルメディアの影響関係から金融市場への作用の分析可能性を高めた. 分析の中で, ソーシャルメディアのネットワークが直接接続されていなくとも, 時間を通して影響関係になる可能性があり, さらに金融資産・市場の性質によってソーシャルメディアから金融市場への影響度が異なることを示した.

なお, 本研究では, ソーシャルメディアにおける影響関係から金融市場に対する作用をモデル化することで, 金融資産の価格変動をより精緻にかつ安定して予測することを目的としている. ニューラルネットワークの構造や学習手法そのものは, その性能や分析可能性を考慮し, 異なる手法を適用できるという点で拡張性を持ち, 改良の余地がある. また, 本実験のデータ期間として2017年を対象としたが, 暗号資産市場は上昇局面であった. 金融資産・市場の性質によってソーシャルメディアから金融市場への影響度が異なるため, 様々な市場に本研究を適用するためには, 更なる実験が必要となる可能性がある.

ソーシャルメディアから金融市場への因果関係そのものを証明することは困難である. しかし, 本研究がソーシャルメディアから金融市場に対する作用を示唆し, 後の研究と金融市場の成熟に貢献することを願っている.

## 謝辞

本研究は JSPS 科研費 JP25700032, JP15H05327, JP16H06562 の助成を受けたものです.

## ◇ 参考文献 ◇

- [Asur 10] Asur, S. and Huberman, B. A.: Predicting the future with social media, in *Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, Volume 01, pp. 492–499, IEEE Computer Society (2010)
- [Bollen 11] Bollen, J., Mao, H., and Zeng, X.: Twitter mood predicts the stock market, *Journal of Computational Science*, Vol. 2, No. 1, pp. 1–8 (2011)
- [Bourigault 16] Bourigault, S., Lamprier, S., and Gallinari, P.: Representation learning for information diffusion through social networks: An embedded cascade model, in *Proceedings of the 9th ACM Inter-*

- [national Conference on Web Search and Data Mining, pp. 573–582, ACM (2016)]
- [Brambilla 17] Brambilla, M., Ceri, S., Della Valle, E., Volonterio, R., and Acero Salazar, F. X.: Extracting emerging knowledge from social media, in *Proceedings of the 26th International Conference on World Wide Web*, pp. 795–804, International World Wide Web Conferences Steering Committee (2017)
- [Cao 17] Cao, Q., Shen, H., Cen, K., Ouyang, W., and Cheng, X.: DeepHawkes: Bridging the gap between prediction and understanding of information cascades, in *Proceedings of the 2017 ACM Conference on Information and Knowledge Management*, pp. 1149–1158, ACM (2017)
- [De Choudhury 09] De Choudhury, M.: Modeling and predicting group activity over time in online social media, in *Proceedings of the 20th ACM Conference on Hypertext and Hypermedia*, pp. 349–350, ACM (2009)
- [Feeenstra 08] Feeenstra, R. C. and Taylor, A. M.: *International Macroeconomics*, Worth Publishers (2008)
- [Ferrara 14] Ferrara, E., Interdonato, R., and Tagarelli, A.: Online popularity and topical interests through the lens of instagram, in *Proceedings of the 25th ACM Conference on Hypertext and Social Media*, pp. 24–34, ACM (2014)
- [Gordon 62] Gordon, M. J.: *The Investment, Financing, and Valuation of the Corporation*, RD Irwin (1962)
- [Gu 18] Gu, Y., Sun, Y., Li, Y., and Yang, Y.: RaRE: Social rank regulated large-scale network embedding, in *Proceedings of the 2018 World Wide Web Conference on World Wide Web*, pp. 359–368, International World Wide Web Conferences Steering Committee (2018)
- [Hochreiter 97] Hochreiter, S. and Schmidhuber, J.: Long short-term memory, *Neural Computation*, Vol. 9, No. 8, pp. 1735–1780 (1997)
- [Ioffe 15] Ioffe, S. and Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift, *arXiv preprint arXiv:1502.03167* (2015)
- [Kempe 03] Kempe, D., Kleinberg, J., and Tardos, É.: Maximizing the spread of influence through a social network, in *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 137–146, ACM (2003)
- [Kristoufek 13] Kristoufek, L.: Bitcoin meets Google Trends and Wikipedia: Quantifying the relationship between phenomena of the Internet era, *Scientific Reports*, Vol. 3, p. 3415 (2013)
- [Krugman 08] Krugman, P. R.: *International Economics: Theory and Policy*, 8/E, Pearson Education India (2008)
- [Levi 07] Levi, M. D.: *International Finance: Contemporary Issues*, Routledge (2007)
- [Li 17] Li, C., Ma, J., Guo, X., and Mei, Q.: DeepCas: An end-to-end predictor of information cascades, in *Proceedings of the 26th International Conference on World Wide Web*, pp. 577–586, International World Wide Web Conferences Steering Committee (2017)
- [Liang 16] Liang, D., Charlin, L., McInerney, J., and Blei, D. M.: Modeling user exposure in recommendation, in *Proceedings of the 25th International Conference on World Wide Web*, pp. 951–961, International World Wide Web Conferences Steering Committee (2016)
- [Mnih 08] Mnih, A. and Salakhutdinov, R. R.: Probabilistic matrix factorization, in *Advances in Neural Information Processing Systems*, pp. 1257–1264 (2008)
- [Nofer 15] Nofer, M. and Hinz, O.: Using twitter to predict the stock market, *Business & Information Systems Engineering*, Vol. 57, No. 4, pp. 229–242 (2015)
- [Oda 18] Oda, S.: The Mysterious Twitter User Drawing a Swarm of Japan Traders, <https://www.bloomberg.com/news/articles/2018-01-14/the-mysterious-twitter-user-attracting-a-swarm-of-japan-traders> (2018)
- [Reinert 09] Reinert, K. A., Rajan, R. S., Glass, A. J., and Davis, L. S.: *The Princeton Encyclopedia of the World Economy (Two volume set)*, Vol. 1, Princeton University Press (2009)
- [Romero 16] Romero, D. M., Uzzi, B., and Kleinberg, J.: Social networks under stress, in *Proceedings of the 25th International Conference on World Wide Web*, pp. 9–20, International World Wide Web Conferences Steering Committee (2016)
- [Sakaki 10] Sakaki, T., Okazaki, M., and Matsuo, Y.: Earthquake shakes Twitter users: Real-time event detection by social sensors, in *Proceedings of the 19th International Conference on World Wide Web*, pp. 851–860, ACM (2010)
- [Shao 17] Shao, M., Li, J., Chen, F., Huang, H., Zhang, S., and Chen, X.: An efficient approach to event detection and forecasting in dynamic multivariate social media networks, in *Proceedings of the 26th International Conference on World Wide Web*, pp. 1631–1639, International World Wide Web Conferences Steering Committee (2017)
- [Srivastava 14] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting, *The Journal of Machine Learning Research*, Vol. 15, No. 1, pp. 1929–1958 (2014)
- [Tumasjan 10] Tumasjan, A., Sprenger, T. O., Sandner, P. G., and Welpe, I. M.: Predicting elections with twitter: What 140 characters reveal about political sentiment, *The International Conference on Weblogs and Social Media*, Vol. 10, No. 1, pp. 178–185 (2010)
- [Wang 12] Wang, H., Can, D., Kazemzadeh, A., Bar, F., and Narayanan, S.: A system for real-time twitter sentiment analysis of 2012 us presidential election cycle, in *Proceedings of the ACL 2012 System Demonstrations*, pp. 115–120, Association for Computational Linguistics (2012)
- [Wang 17] Wang, J., Zheng, V. W., Liu, Z., and Chang, K. C.-C.: Topological recurrent neural network for diffusion prediction, in *2017 IEEE International Conference on Data Mining (ICDM)*, pp. 475–484, IEEE (2017)
- [Zafarani 14] Zafarani, R., Abbasi, M. A., and Liu, H.: *Social Media Mining: An Introduction*, Cambridge University Press (2014)
- [Zhang 11] Zhang, X., Fuehres, H., and Gloor, P. A.: Predicting stock market indicators through twitter “I hope it is not as bad as I fear”, *Procedia-Social and Behavioral Sciences*, Vol. 26, pp. 55–62 (2011)
- [松尾 13] 松尾豊：ウェブからの実世界の観測と予測, 電子情報通信学会論文誌 B, Vol. 96, No. 12, pp. 1309–1315 (2013)

〔担当委員：高野 雅典〕

2020年6月1日 受理

## ◇ 付 錄 ◇

### A. 分析した事業・市場において選択した銘柄

各事業・市場において、以下に示される2017年時点で上場しておりかつ時価総額上位5銘柄を選択し、分析対象とした。

- バイオ事業: サンバイオ (4592), そーせい (4565), ユーグレナ (2391), アンジェス (4563), ペプチドリーム (4587)
- ソーシャルゲーム事業: ガンホー・オンライン・エンターテイメント (3765), Klab(3656), イグニス (3689), ドリコム (3793), アカツキ (3932)
- インターネット事業: ソフトバンクグループ (9984), 楽天 (4755), ヤフー (4689), サイバーエージェント (4751), ディー・エヌ・エー (2432)
- 自動車事業: トヨタ自動車 (7203), 本田自動車 (7267), 日産自動車 (7201), スズキ (7269), SUBARU(7270)
- 銀行事業: 三菱UFJ ファイナンシャルグループ (8306), ゆうちょ銀行 (7182), 三井住友フィナンシャルグループ (8316), みずほフィナンシャルグループ (8411)
- 東証一部市場: トヨタ自動車 (7203), NTT ドコモ (9437), 日本電信電話 (9432), ソフトバンクグループ (9984), 三菱UFJ ファイナンシャルグループ (8306)
- 東証二部市場: 東芝 (6502), アトム (7412), 三谷商事 (8066), 帝国ホテル (9708), 日本精機 (7287)
- マザーズ市場: サンバイオ (4592), ミクシィ(2121), ジャパンインベストメントアドバイザー (7272), ティーケーピー (3479), UUUM (3990)

---

著者紹介

---

田村 浩一郎

東京大学大学院工学系研究科 技術経営戦略学専攻 博士 2 年。



松尾 豊(正会員)

1997 年東京大学工学部卒業。2002 年同大学院博士課程修了。博士(工学)。産業技術総合研究所、スタンフォード大学を経て、2007 年より、東京大学大学院工学系研究科技術経営戦略学専攻 准教授。2019 年より同大学院人工物工学研究センター／技術経営戦略学専攻 教授。2014 年より 2018 年まで人工知能学会倫理委員長。2017 年より日本ディープラーニング協会理事長。人工知能学会論文賞、情報処理学会長尾真記念特別賞、ドコモモバイルサイエンス賞など受賞。専門は、人工知能、深層学習、Web 工学。

