

Twitter から仮想通貨市場に対する作用の分析

Twitter Actuates the Cryptocurrency Market

田村浩一郎

Koichiro Tamura

大澤昇平

Shohei Ohsawa

松尾豊

Yutaka Matsuo

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, The University of Tokyo

Social media is not only a sensor but also an actuator. A *social sensor* can detect various social phenomena or trends by using the data obtained from social media. Though the information from social sensors has vastly enhanced our potential to observe and predict real-world phenomena, the causality from the social media itself to the real world has not been studied to date. Meanwhile, recent trends in SNS-driven consumer behavior, such as taking beautiful photos for Instagram, so-called "Instagenic," have further highlighted the importance of studying the causality from social media to the real world. This paper demonstrates a new concept of *social actuator*. We introduce internal states, which represent the states of the social media users influenced by others, and show how to address the confounding structure in the inference of causality from social media to the real world. Using the results of our experiments on Twitter data and cryptocurrency-market data, we show that our proposed method can detect the influences between the users on social media, and describe the causation from Twitter to the cryptocurrency market. Finally, we discuss the effectiveness of the proposed method for different datasets and suggest that we all have the potential to impact the real world through social media either intentionally or unintentionally.

1. 序論

ソーシャルメディアは、ソーシャルセンサ [Sakaki 10] としての機能を持つだけでなく、アクチュエーターとしての機能も持っている。ソーシャルセンサは様々な社会現象やトレンドなどを、ソーシャルメディアにおける投稿やお気に入りといった情報を用いて検知する [Sakaki 10]。その一方で、近年、インスタグラムに投稿するための見栄えの良い写真、所謂「インスタ映え」する写真を取るために消費行動が観測されている。また、投資家は他の投資家のツイートを元に投資判断を行い、株価が特定のユーザーのツイートに連動している可能性も述べられている [Oda 18]。この研究は、ソーシャルメディアが実世界に作用するソーシャルアクチュエーターとして機能しているという仮説を実証することを目的とする。

実世界の観測や予測を行う際に、social sensor によって観測されたデータの有用性や、実世界のデータとの相関が議論されていた [Sakaki 10, Tumasjan 10, Asur 10, Bollen 11, Kristoufek 13] が、social media から実世界の現象に対する因果については十分に議論されて来なかった。social media と対象としている実世界の事象の双方に介入する交絡因子が存在する場合、social media と対象とする実世界の事象の関係は疑似相関である可能性がある。例えば [Kristoufek 13] では、google trend と bitcoin market の値の双方向の関係性が存在していることが示されているが、Bitcoin について google で検索したから価格が変動したのではなく、Google trend が Bitcoin market に影響を与える実世界の trend を反映しているにすぎず、google trend から Bitcoin に対する因果関係については疑似相関である可能性が考えられる。

そこで本研究は、Twitter から現実世界に対する因果関係を検証するための手法を提案する。そして 9401 人の twitter users を対象に、2017-01-01 から 2017-12-31 までの tweet および follow/followee 関係と、cryptocurrency market の値を

連絡先: 田村浩一郎, 東京大学大学院工学系研究科,
koichi@weblab.t.u-tokyo.ac.jp

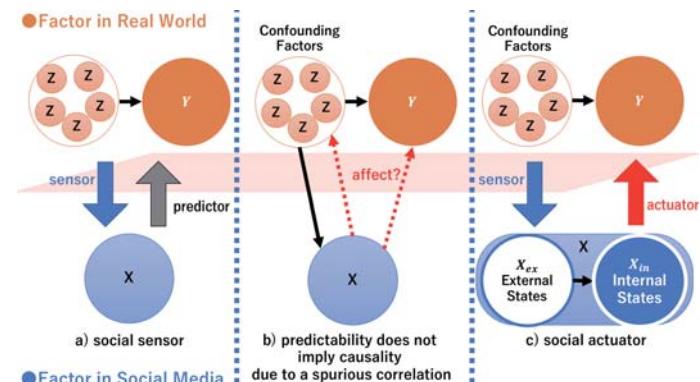


図 1: A new concept of social actuator

用いて実験を行い、提案手法を用いた将来の cryptocurrency market value の予測精度がいくつかのベースラインを上回る精度であることを示した。そしてその結果より、twitter における internal states が、cryptocurrency market に対して働きかけていることを説明した。

本研究の貢献は以下の通り。

- 提案手法が、twitter 内におけるリツイートなどの直接観測される interaction だけでなく、直接観測することができないユーザー間の影響を考慮し、学習の枠組みの中でユーザー間の影響度を獲得できるものであることを示した。
- Twitter における他者の post から影響を受けて蓄積された実世界に働きかける social motivation が、cryptocurrency market に働きかけていることを示唆し、social media を通して誰もが実世界にインパクトを与える可能性を提示した。

2. 関連研究

2.1 Social Sensor

2000年代後半、TwitterやFacebookと行ったソーシャルメディアが登場した。スマートフォンの普及と共に、「人々は、実世界に情報端末をもち出し、そこで起きた出来事や感じたことをソーシャルメディアに投稿するようになったため、社会や実世界で起きていることが計算機に記録されるようになった。このことは、社会現象の実態やその発生メカニズムを解明しようとしてきた社会学において、貴重なデータ資源になることになった」[Hijikata 16]。ソーシャルメディアにおけるユーザーの投稿や、投稿に対する「いいね」などの反応によって、ユーザーの興味や関心を観測することが可能になり、「社会で起こりつあるイベントやトレンド、現象などを検出する」[Hijikata 16]「ソーシャルセンサ」[Sakaki 10]としての役割を果たしている。ソーシャルセンサを用いて実世界の現象の観測や予測を行う研究は数多くなされていた [Sakaki 10, Tumasjan 10, Asur 10, Bollen 11, Kristoufek 13, Zhang 11]。しかしこれらの既存研究は、ソーシャルメディアが実世界に影響を与えるという因果関係を議論していない。

2.2 Information Diffusion on Social Network

ソーシャルメディアは、任意のユーザーがそこでの出来事や感じたことを投稿するという特徴に加えて、ユーザー間でネットワークを構成し、社会相互しているという特徴がある。「TwitterやFacebookなどのソーシャルメディアが普及し、そこで構築された友人関係のネットワークは、本物の人間のつながりを表したものと認識される」[Hijikata 16]ようになり、実世界では直接観測することが難しい友人関係のネットワークを分析することが可能になったことから、ソーシャルメディアで構築されるネットワークに関する研究も多くなされ、情報拡散 (information diffusion) を行うメディアとしての強力を示すこととなった」[Hijikata 16]。ソーシャルネットワークにおける情報拡散の一つに、information cascades[Zafarani 14] があげられる。information cascades は”An information cascade describes the process of diffusion when individuals merely observe their immediate neighbors” [Zafarani 14] と定義され、information cascades を予測する研究 [Li 17, Cheng 14, Budak 11] がなされている。しかしこれらの研究は、ソーシャルメディア内部での情報拡散を対象としたものであり、実世界との関係を分析した研究は筆者の知る限り存在していない。

3. 提案手法

本研究は、以下の手法を提案し、Twitterから実世界に対する因果関係を実証することを試みる。

- figure 1 で表されるように、social media の要素を以下の 2 つの因子に分解し表現する

external states social sensor によって観測されるユーザーの投稿などを表す因子 X_{ex}

internal states 他のユーザーから受ける影響が蓄積されたユーザーの内部状態を表す因子 X_{in}

- 対象としている将来の実世界の事象 Y の予測において、以下の 2 つの手法を比較し、Granger causality の存在の有無を確かめる

- 対象としている実世界の事象 Y と X_{ex} を入力とする回帰モデル

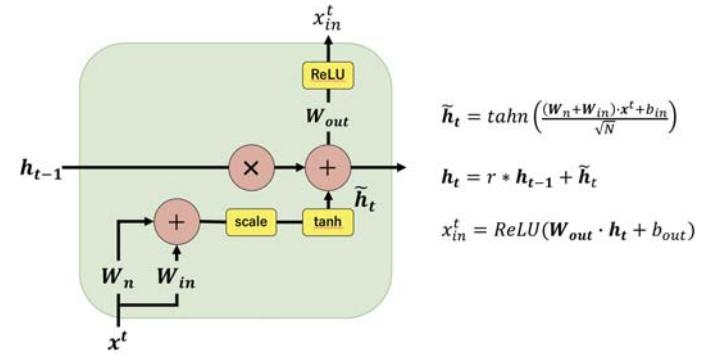


図 2: Social Internal States Module

- 対象としている実世界の事象 Y と X_{ex} および X_{in} を入力とする回帰モデル

この時、social media に対する external states は実世界の要因 Y や交絡要因 Z に影響を受けるのに対して、短期的には social media 内での internal states は、ユーザーの投稿などの external states にのみ介入を受けると仮定できる時、因果の逆転や疑似相関の可能性を否定し、Granger causality から、internal states から対象としている将来の実世界の事象 Y に対する因果関係を示唆することができる。

3.1 Social Internal States Module

ソーシャルメディアにおけるユーザーが他のユーザーとの相互作用によって影響を受け、蓄積されたユーザーの内部状態、internal states を獲得するために、Social Internal States Module(Figure 2) を提案する。

まず、 N 人のユーザーのクラスタをサンプリングし、サンプリングされたユーザー同士の follow/followee 関係を分析する。そして、follow/followee 関係に基づき、隣接行列 $\mathbf{W}_n \in \mathbf{R}^{N \times N}$ を作成する。なお、 \mathbf{W}_n の要素 $w_{ij} = 1$ である場合、user_i が user_j を follow していることを意味する。時刻 t において実験対象について言及する投稿を行なったユーザーを one-hot vector \mathbf{x}^t で表現する。例えば、Bitcoin の将来の出来高を予測する場合、user_j が Bitcoin について言及した投稿を行なったとき、 $\mathbf{x}_t[j] = 1$ とする。そして、以下のようにユーザーが他のユーザーから受ける影響を表すベクトル \tilde{h}_t を計算する。

$$\tilde{h}_t = \tanh\left(\frac{(\mathbf{W}_n + \mathbf{W}_{in}) \cdot \mathbf{x}^t + b_{in}}{\sqrt{N}}\right),$$

なお $\mathbf{W}_{in} \in \mathbf{R}^{N \times N}$ および $b_{in} \in \mathbf{R}^N$ は学習パラメタである一方、 \mathbf{W}_n は定数パラメタである。そして、再帰的アルゴリズムによってユーザーに蓄積する内部状態を以下に計算する。

$$h_t = r * h_{t-1} + \tilde{h}_t,$$

$r \in \mathbf{R}^1$ は学習パラメタで、初期値を 0.1 とする。最後に Social Internal States Module の出力として以下を計算する。

$$x_{in} = \text{ReLU}(\mathbf{W}_{out} \cdot h^t + b_{out}),$$

$\mathbf{W}_{out} \in \mathbf{R}^N$ および $b_{out} \in \mathbf{R}^1$ は学習パラメタである。

3.2 Comparing Regression Models

ソーシャルメディアの内部状態から現実世界への因果関係の存在を検証するために、まず、以下の 2 つの方法を比較し Granger 因果関係の存在を確認する。

Social sensor model

実世界の事象 Y を予測するために、ソーシャルセンサによって観測されるソーシャルメディアにおける事象 X_{ex} を用いた回帰モデル。

$$\hat{y} = \mathbf{W}_{ss} \cdot concat(\mathbf{x}_y, \mathbf{x}_{ex}) + b_{ss},$$

\mathbf{x}_y は事象 Y を表す特徴ベクトル、 \mathbf{x}_{ex} は X_{ex} を表す特徴ベクトルであり、 \mathbf{W}_{ss} and b_{ss} は学習パラメタである。

Social actuator model

実世界の事象 Y を予測するために、ソーシャルセンサによって観測されるソーシャルメディアにおける事象 X_{ex} およびソーシャルメディアのユーザーの内部状態 X_{in} を用いた回帰モデル。

$$\hat{y} = \mathbf{W}_{sa} \cdot concat(\mathbf{x}_y, \mathbf{x}_{ex}, \mathbf{x}_{in}) + b_{sa}, \quad (1)$$

\mathbf{x}_{in} は social internal states module から取得した X_{in} を表す特徴ベクトル、 \mathbf{W}_{sa} and b_{sa} は学習パラメタである。

social actuator model の予測が social sensor model の予測よりも正確であるとき、ソーシャルメディアの内部状態から現実世界へのグレンジャー因果関係の存在を推論することができる。

しかし、グレンジャー因果は通常の因果の必要条件にすぎず、擬似相関の可能性がある。本研究では、Figure 1 で示されるように、ソーシャルセンサによって得られた external states の要素 X_{ex} は現実世界の対象要素 Y 以外の現実世界の交絡因子 Z に介入されるが、ソーシャルメディアの内部状態 X_{in} は短期的には external states の要素 X_{ex} のみに介入されると仮定する。これは、follow/followee や相互作用といったソーシャルメディアにおける関係性は、予測を行うような短期的な期間においては実世界の事象と独立であると考えることができるからだ。この場合、擬似相関の問題を仮想的に解決し、internal states から実世界の現象 Y に対する因果関係を述べることができる。

4. 実験

近年、ブロックチェーンという技術に基づいた仮想通貨の存在が話題になっている。現在、仮想通貨の使い道は法定通貨に対して限られており、仮想通貨のなす経済圏の成長があるわけでもなく、また保有することによる金利なども存在していないため、現在の仮想通貨の市場はほとんどすべて需要と供給によって決定している [Kristoufek 13]。よって、ソーシャルメディアにおける影響が存在するならば、他の要因が比較的少ないため検証しやすいと考え、今回、実験対象として仮想通貨市場を取り扱う。

実験には、CoinGecko^{*1} から取得した Bitcoin, Ethereum, Ripple の 3 つの仮想通貨の価格および出来高と、Twitter REST API^{*2} から取得した投資家のツイートおよび follow 関係のデータを用いる。仮想通貨のデータは、2017 年 1 月 1

*1 <https://www.coingecko.com/en>

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

日から 2017 年 12 月 31 日までの一年間の価格および出来高の日足データを利用する。取得したデータから、1,3,5 日前比の出来高変化率、および 1,3,5 日前比の価格変化率を作成し、仮想通貨の情報を表す入力特徴量とする。そして、1,3,5,30 日後の出来高変化率を作成し、予測対象とする。1 月 1 日から 10 月 31 日までの約 300 データを学習データとし、11 月 1 日から 11 月 30 日までの 30 データをテストデータとして用いる。Twitter のデータは、Okasanman, one of the famous twitter accounts of investors in Japan を follow しているユーザーのうち、2016 年 9 月から 12 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ーでかつ follower が 100 人以上いる 9401users のアカウントを対象として、取得していた全てのツイートおよび follow/followee 関係を用いる。

そして、以下の 4 つのモデルの mean square error(MSE) を比較検証する。

BASE: predict 0 all

常に 0 と予測を行うモデルをベースラインとする。このベースラインの MSE よりも悪い場合、そのモデルの予測は意味をなさないということになる。

$$\hat{y} = 0.$$

VAR: only cryptocurrency market values

まず、仮想通貨市場の値のみを用いた自己回帰モデル (VAE) [Toda 94] を検証する。出力として、以下を得る。

$$\hat{y} = \mathbf{W}_{var} \cdot \mathbf{x}_y + b_{var},$$

\mathbf{x}_y は仮想通貨市場の値の入力ベクトル、 $\mathbf{W}_{var}, b_{var}$ は学習パラメタである。

Social Sensor: with external states of Twitter

次に、ソーシャルセンサによって獲得される external states を用いた回帰モデルを検証する。予測対象となる仮想通貨に対するツイートは、それぞれ以下のワードを含むものを検出する。

Bitcoin: Bitcoin, BTC, ビットコイン

Ethereum: Ethereum, ETH, イーサリアム

Ripple: Ripple, XPR, リップル

モデルの出力は以下の通り。

$$\hat{y} = \mathbf{W}_{ss} \cdot concat(\mathbf{x}_y, \mathbf{x}_{ex}) + b_{ss},$$

\mathbf{x}_{ex} はツイート数などソーシャルセンサから得られた入力ベクトル、 \mathbf{W}_{ss}, b_{ss} は学習パラメタである。

Information Cascades: with DeepCas pipeline

DeepCas [Li 17] は cascades graph の構造をノードをサンプリングすることで表現し、End-to-end の枠組みで cascades を予測する手法である。この手法は、ネットワークにおける関係性を表現している点で上述した internal states を表現していると考えることもできる。そこで、social internal states module の代わりに、DeepCas pipeline を用いる。

得られた結果は表 1 の通り。ほとんどの場合で提案手法である Social Actuator のモデルが他のモデルよりも優れた予測を行なっている。

表 1: Performance measured by MSE

	Bitcoin				Etheruem				Ripple			
	volume_1	volume_3	volume_5	volume_30	volume_1	volume_3	volume_5	volume_30	volume_1	volume_3	volume_5	volume_30
BASE	0.2335	0.4507	0.5316	0.4126	0.1865	0.5913	0.5530	1.1061	0.9345	2.0870	1.9373	0.6639
VAR	0.2150	0.3122	0.3225	0.3184	0.1639	0.3687	0.3890	1.1070	0.6279	0.7111	1.1307	1.2217
Social Sensor	0.2287	0.3551	0.3141	0.2576	0.1636	0.3679	0.3755	1.1195	0.6142	0.5660	0.9268	1.3226
Information Cascades	0.2569	1.0195	1.1131	1.4950	0.4321	0.4458	0.4088	2.8378	0.6220	1.6445	1.1410	2.4280
Social Actuator	0.2236	0.3087	0.3076	0.2117	0.1489	0.3640	0.3226	0.9291	0.5570	0.5298	0.9156	1.2951

5. 考察

実験では、VARによる手法、Social sensorを用いて投稿数に着目した手法、information diffusionを考慮したDeepCasを用いた手法に対して、提案手法の精度が高いという結果を得た。既存のSocial sensorを用いた予測手法は、実世界を観測したユーザーがソーシャルメディアに投稿することを利用して実世界の現象を間接的に観測・予測するものであって、誰が、どれくらいの数のユーザーが投稿するかということは、実世界の現象によるものだとされていた。ソーシャルセンサからの情報に加えて、Twitterの投稿者と情報の受信者の間の影響という、ソーシャルメディア内の社会相互を入力に加えた提案手法の精度が高いということは、Twitterにおける社会相互から仮想通貨の市場へのグレンジャー因果性が存在していることを意味する。

グレンジャー因果は通常の因果性が存在する必要条件ではあるが、十分条件ではなく、擬似相関である可能性がある。今回の実験の場合、仮想通貨の市場とTwitterにおける現象に対して、共通の交絡要因がある場合である。例えば、仮想通貨のマイニングに必要な電力に対する料金の変化や、新たに仮想通貨に対する規制など、ファンダメンタルズな要因が仮想通貨の市場とTwitterのユーザー双方に影響を及ぼす可能性があり、両者の交絡要因であると考えられる。本研究の提案手法の入力として追加したTwitterの投稿者と情報の受信者の間の関係において与える影響は小さいと仮定できる。なぜなら、比較的短期的な期間におけるファンダメンタルズ要因によって、Twitterにおけるfollow/followee関係が変化する可能性は低く、user同士の影響度が変わることは考えにくいかだ。もしTwitterにおける社会相互が実世界の現象と独立である場合、Twitterにおける社会相互から仮想通貨の市場へのグレンジャー因果性から、Twitterにおける社会相互と仮想通貨の市場への因果関係を推論することができる。

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. 492–499IEEE 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)
- [Budak 11] Budak, C., Agrawal, D., and El Abbadi, A.: Structural trend analysis for online social networks, *Proceedings of the VLDB Endowment*, Vol. 4, No. 10, pp. 646–656 (2011)
- [Cheng 14] Cheng, J., Adamic, L., Dow, P. A., Kleinberg, J. M., and Leskovec, J.: Can cascades be predicted?, in *Proceedings of the 23rd international conference on World wide web*, pp. 925–936ACM (2014)
- [Hijikata 16] Hijikata, : 社会学とAI—社会と人間を映すソーシャルメディア, *人工知能学会誌*, Vol. 32, No. 3 (2016)
- [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)
- [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. 577–586International World Wide Web Conferences Steering Committee (2017)
- [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)
- [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. 851–860ACM (2010)
- [Toda 94] Toda, H. Y. and Phillips, P. C.: Vector autoregression and causality: a theoretical overview and simulation study, *Econometric reviews*, Vol. 13, No. 2, pp. 259–285 (1994)
- [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., *Icwsrm*, Vol. 10, No. 1, pp. 178–185 (2010)
- [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)