

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

清水 豪士

富山県立大学 情報基盤工学講座
t715038@st.pu-toyama.ac.jp

June 24, 2021

はじめに

2/17

背景

- ソーシャルメディア上の情報は、社会を観測するセンサと考えることができ、それをソーシャルセンサと呼ぶ。
- ソーシャルセンサは実世界の観測と予測の可能性を広げた。
- 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。

目的

- 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。
- 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分析を行う。

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。

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

- 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 に関する投資行動を行い、金融市場が変動する。

定式化したものを end-to-end の NN である Recurrent Neural Network(RNN) を用いてモデル化する。

提案手法

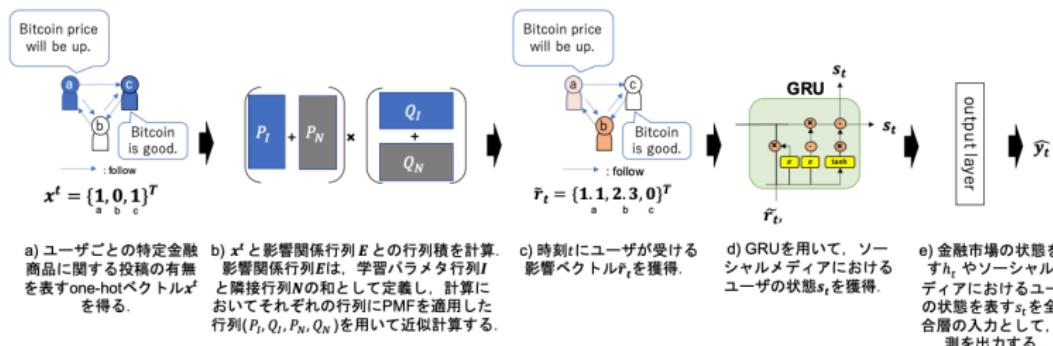


図 1: 提案方式

- 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 にユーザが受ける影響ベクトル 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场の予測を出力する。

提案手法

閲覧などによって引き起こされるユーザ u_i がユーザ u_j から受ける影響 $s_{j,i}$ を要素にもつ影響関係行列 \mathbf{E} を end-to-end ネットワークの学習パラメタとして定義する。

影響関係行列 \mathbf{E} による影響関係の表現と抽出

ソーシャルメディアのユーザ M 人のうち、ユーザ u_i がユーザ 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 \mathbf{N} で表現し、ユーザ u_i がユーザ u_j の投稿を閲覧した時に受ける影響度合を、学習パラメタ \mathbf{I} によって表す。

このとき影響関係行列 \mathbf{E} を以下のように定義する。

$$\mathbf{E} = \mathbf{N} + \mathbf{I} \quad (1)$$

そしてある時刻 t において K 人のユーザが、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 \mathbf{x}_t と影響関係行列 \mathbf{E} の行列積によって、ある時刻 t にユーザが受ける影響ベクトル \mathbf{r}_t を獲得する。

$$\mathbf{r}_t = \mathbf{E}\mathbf{x}_t \quad (2)$$

影響関係行列 \mathbf{E} を構成する隣接行列 \mathbf{N} に対して PMF を行い、行列 \mathbf{P}_N と行列 \mathbf{Q}_N に分解し、 \mathbf{I} に基づく学習パラメタ \mathbf{P}_I と学習パラメタ行列 \mathbf{Q}_I を定義することで、式(1)の影響関係行列 \mathbf{E} を以下のように近似する。

$$\mathbf{E} \sim (\mathbf{P}_I + \mathbf{P}_N)(\mathbf{Q}_I + \mathbf{Q}_N)$$

このとき、式(2)は以下のように近似される。

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

PMF : Probabilistic Matrix Factorization
高次元なグラフネットワークを低次元で表現する手法

提案手法

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。

認知・心理の時系列内部状態の獲得

式(3)で獲得された影響ベクトル r_t に対して、処理を行い、埋め込みベクトル \tilde{r}_t を得る。

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

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 s_t を GRU を用いて獲得する。

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

獲得された 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 m_t に対し、GRU を適用し、金融市場の状態を表す特徴量 h_t を獲得する。

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

W_{in} および b_{in} は学習パラメタ

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、 s_t および h_t を 1 次元の値に変換し、正規化したのち最終層に入力として予測を行う。

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

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

$$\hat{y}_t = \tanh(\mathbf{W}_{out} concat(s_t, h_t) + b_{out}) \quad (6)$$

\hat{y}_t は時刻 t に出力される予測値。 $concat(s_t, h_t)$ は s_t と h_t の連結ベクトルそれ以外は、学習パラメタである。

実験設定

9/17

実験対象

- 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。
- 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。

データセット

- 3つの暗号資産の価格データ、Twitter REST API 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。

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 m_t として利用する。

データセット

- ① $ChangePrice_{\Delta t}$: 時刻 t から $\Delta t = 1, 3, 5, 10, 25$ 日前からの終値変化率
- ② $DeviationRate_{\Delta t}$: 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 $Volatility_{\Delta t}$: 時刻 t における $\Delta t = 3, 5, 10, 25$ 日間のボラティリティ

ツイッターのデータはある投資家をフォローしている約 200000 人のユーザのうち、2017 年 1 月から 4 月までに 5 回以上ツイートがある 9900 人のユーザのアカウントを対象とし、9900 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獲得した。

評価方法

暗号資産市場の将来のボラティリティの予測において、RMSE を用いる。
 \hat{y} を予測値、 y を実際の値とすると、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。

比較モデル

12/17

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。

評価方法

- 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、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。

実験結果

13/17

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を図2に示す。

	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	0.0215	0.0220	0.0267	0.0414

	1日後	3日後	5日後	10日後
Market	0.0146	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	0.0184	0.0226	0.0237

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

図2: 上:ビットコイン、左:イーサリアム、右:リップル のボラティリティ予測における提案手法のRMSE比較

ほとんどの比較検証において、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えている。

提案手法は、ソーシャルメディアの投稿の閲覧に伴う影響関係や、ユーザの認知や心理状態を抽出・表現するため、ノイズに対して頑強で、より安定して高い精度を出すことができると考えられる。

分析

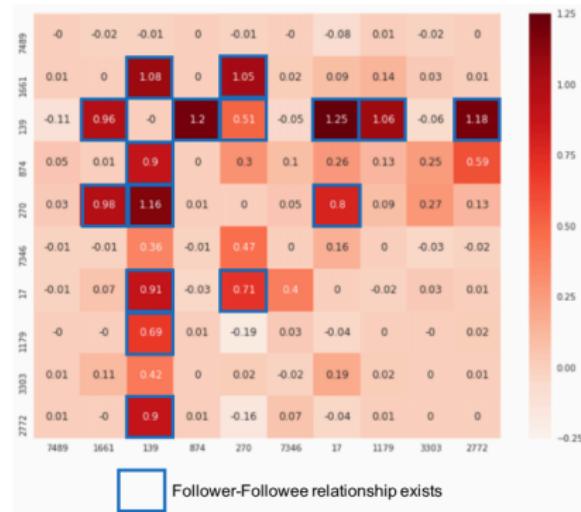


図 3: ビットコインの 3 日後のボラティリティ予測の検証で獲得された影響関係を表す行列 E においてサンプリングされた 10 アカウントの影響関係に対応するパラメタをヒートマップ

行は影響を与えるユーザ，列は影響を受けるユーザを表している。

青四角で囲まれたユーザ関係は，ツイッターにおいて列のユーザが行のユーザをフォローしていることを表している。

値が大きいほど影響が大きいことを示している。

終わりに

15/17

まとめ

- 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。
- 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。

今後の課題・改良点

- NN の構造や学習方法そのものは、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。

実験結果

実験結果