Proc. Blockchain in Kyoto 2022 (BCK22)
JPS Conf. Proc. **40**, 011006 (2023)
https://doi.org/10.7566/JPSCP.40.011006

Acquiring Semantic Mode signal from Tweets of Cryptoassets

Hiroshi Uehara[1], Wataru Souma[1] and Yuichi Ikeda[2]

[1] *Faculty of Data Science, Rissho University, 1700 Magechi, Kumagaya, Saitama 360-0194, Japan*
[2] *Graduate School of Advanced Integrated Studies in Human Survivability, Kyoto University, Kyoto 606-8306, Japan*

E-mail: uehara@ris.ac.jp,souma@ris.ac.jp,ikeda.yuichi.2w@kyoto-u.ac.jp

(Received November 14, 2022)

This study proposes a method for detecting collective motions, the time series situation where dispersed information becomes inclined to a unique direction. Our proposal, Semantic mode signal, is distinctive among related methods in providing the situation with semantic contexts extracted from time series texts. Furthermore, the proposal is characterized by its applicability to high-dimensional word space with sparsity, such as numerous tweets. We applied the method to tweets concerning 19 cryptoassets. The empirical results indicated that the method appropriately detected the collective motions representing the contexts semantically coincident with the news events and the price trends, supporting the efficiency of our proposal.

KEYWORDS: Mode signal, Economic anomaly, Cryptoasset, Topic model, Text analysis, Bayesian inference, Geopolitical risk

1. Introduction

In this era of uncertainty, much attention has focused on unpredictable economic anomalies. This is especially the case with cryptoassets which have highly unpredictable price movements because of their weak linkage to the real side of the economy. Therefore, detecting market anomaly latent signals has been a crucial research topic. Although several studies have successfully detected anomalies [1, 2], these results are limited to specifying the time and degree of anomalous events without semantic information to interpret the events.

The social media platform, Twitter, has been widely analyzed for grasping social trends and concerns wherein counting words in tweets by each time slice enables the extraction of information representing trends. However, counting words in tweets is only sometimes effective for capturing semantics regarding anomalous events. In general, market anomaly signals are represented by the degree of movements inclined toward a specific direction. The situation of these inclined movements is called collective motions. In the case of cryptoassets, collective motions are price movements of all the related cryptoassets toward one direction. Counting words in tweets, however, does not capture such collective motions effectively wherein detected words with high frequencies represent scattered interests but not unique semantic contexts inclined in one direction.

In this study, a method is proposed to provide semantic information for the interpretation of cryptoasset market anomalous events by applying mode signal in tweets concerning cryptoassets . Mode signal is an effective method for detecting collective motions [8]. The study uses the mode signal method of detecting latent situations dominated by the first eigenvector. Nevertheless, some issues still exist in applying mode signal to tweets. In each tweet, the tweet space is an extremely high dimension with sparsity because it is comprised of diversified vocabularies and sporadic frequency. The proposed semantic mode signal simultaneously enables dimensional reduction with transforming tweet vocabulary space to semantically summarized space called topic space. Once the summarized low-dimensional space is acquired, principal directions in the space become semantically interpretable in terms of collective motions. As such, this study aims at enabling the semantic interpretation of the cryptoasset market collective motions by semantic mode signal, which is a homogeneous method to mode signal, and is applicable to extremely high-dimensional space with sparsity.

The rest of this paper is as follows: Section 2 reviews studies on detecting market anomalies and tweets analysis; Section 3 explains the proposed method; Section 4 introduces data for examination; Section 5 discusses the empirical settings; Section 6 presents the results; and Section 7 concludes the study.

2. Related Work

Twitter is one of the enormous data sources for interpreting social trends of interest and could provide semantic information to understand anomalies in the cryptoasset market. Generally, in the case of financial market research, tweets are classified into time series sentiment (i.e. positive/negative) by applying natural language algorithms wherein the inferred sentiment is then used as features for price predictions [10–13]. All these research works are characterized by the supervised learning methods for acquiring accurate inferences of the sentiment and precise prediction of prices based on the sentiment. Furthermore, apart from sentiment, a few studies have adopted other attributes of tweets to predict prices [14, 15]. Otabek and Choi (2022) used four attributes: the number of followers of the tweet poster, the number of comments on the tweet, the number of likes, and the number of retweets [14].

A topic model [16] is an effective clustering method for semantic clustering and dimensional reduction. It is a generative statistical model characterized by the simultaneous generative procedure as if each word is generated from both the words' distributions (i.e. topics) and documents (i.e. tweets). Therefore, inference outcomes include topics as a form of word distribution while the documents represented by topic distributions comprise much smaller dimensions than the original vocabulary space. In the study, topic model has been used to classify social interests from tweets. For example, Gangwar and Singh [12] have analyzed social trends of interest concerning covid19 from tweets [17]. Similarly, Hai et al. (2015) have inferred topics related to stock prices thereby extracting sentiment specific to stock prices that improved predictive performance [18]. While these studies have used a topic model for clustering topics of interest, this study uses it not only for clustering but also for dimensional reduction.

Mode signal captures anomalous events as collective motions. It derives principal components from a space comprised of multiple time series data (e.g. prices of cryptoassets) and detects latent situations in the series where the first eigenvector becomes dominant. Applying mode signal derives prominent peaks representing collective motions from complicated fluctuations of the multiple time series data [3–7, 9].

This study applies unsupervised learning to tweets using the topic model to infer clusters of words' distributions, each representing semantic coherence as if generated from a latent unique context. The topic model transforms the high-dimensional vocabulary space of tweets into low-dimensional semantic space to try and detect collective motions from the space based on a method similar to mode signal which is semantic mode signal. This method can acquire peaks of collective motions as semantically interpretable forms.

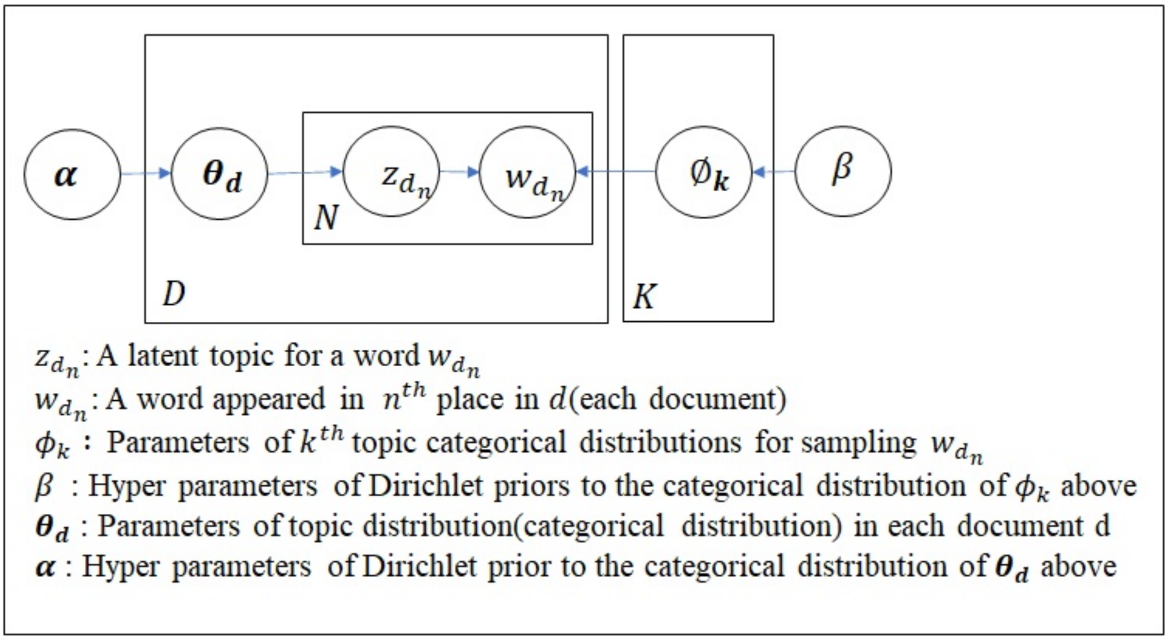


Fig. 1. A graphical model of Topic model Generative process of Topic model represents the latent topics z_{d_n} are simultaneously generated both from topic distribution θ_d and from word distribution ϕ_k

3. Semantic Mode Signal Method

The proposed semantic mode signal is comprised of three procedures. First, is to acquire clusters of contexts as topics and represent each document (i.e. tweets) as distributions of the topics by adopting the topic model. Second, is to detect situations where collective motions become distinctive. Third, is to provide semantic interpretation with the collective motions. The following elaborates on the method.

3.1 Topic model

3.1.1 Generative modeling

In the topic model, each word in the documents is assumed to be generated through two simultaneous processes as shown in **Figure** 1; on the right side is the generative process from topics while the left side is the one from documents where ϕ_k represents the parameters of topics subject to multinomial distributions of words and K is the given number of topics. Since the topic model is a Bayesian statistical model, ϕ_k has the conjugate prior distribution subject to Dirichlet distribution represented by β. θ_d which are the parameters of documents subject to multinomial distribution of topics. Similar to topics, it has the conjugate prior distribution subject to Dirichlet distribution represented by α. Equation (1) corresponds to the generative model given in **Figure** 1(refer to **Table** I for the description of each variable in detail).

$$p\left(\mathbf{W},\mathbf{Z},\theta,\phi\right)=\prod_{d}\mathrm{Dir}\left(\theta_{\mathbf{d}}|\alpha\right)\prod_{k}\mathrm{Dir}\left(\phi_{\mathbf{k}}|\beta\right)\prod_{n}Mult\left(z_{d_n}|\theta_d\right)Mult(w_{d_n}|\emptyset_{\mathbf{k}},z_{d_n}) \tag{1}$$

Table I. Descriptions of variables

Variables	Description
Dir	Dirichlet distribution
$Mult$	Multinomial distribution
$\phi_{\mathbf{k}}$	Parameters of word distribution in each topic, k, with hyperparameters, β
θ_d	Parameters of topic distribution in d(each document)
α	Hyperparameters of θ_d
z_{dn}	A latent topic for a word, w_{dn}, appeared in n^{th} place in d
N_{dk}	The number of words belonging to k in d
N_k	The number of words belonging to k
N_d	The number of words in each document
N_{kw}	The number of each word,w, belonging to k
N_{dk}^{-dn}	The number of words belonging to k in d except dn. (Hereafter, "-" of the upper suffix represents similar exceptions)
ϕ_{kv}	The component in ϕ_k corresponding to v(vocabulary)
θ_{dk}	The component in θ_d corresponding to k(topic)
ω_i^τ	Weight of semantic information of collective motions for v(word) in τ(time slice)
n_{t_v}	The number of v(word) appeared in t(documents in τ)
ϕ_{kv}	Word2topic probability of v(word) in k(topic)
g_k^τ	Time series topic weights of k(topic) in τ (i.e., component value in 1st eigenvector in each time slice)
T	The number of documents in τ
K	The number of topics

Then, the algorithm of the generative model is as follows:

(1) For $k \in 1, \cdots, K$:

  a. $\phi_{\mathbf{k}} \sim Dir(\beta)$

(2) For $d \in 1, \cdots, D$:

  a. Draw $\theta_d \sim Dir(\alpha)$
  b. For $n \in 1, \cdots, N_d$:
    i. Draw $z_{dn} \sim Mult(\theta_d)$
    ii. Draw $w_{dn} \sim Mult(\phi_{\mathbf{z_{dn}}})$

3.1.2 Inference

The study adopts the collapsed Gibbs sampling, one of the Markov chain Monte Carlo methods, for inference. Based on the method, z_{dn} are sampled according to formula (2). Then, both ϕ_k and θ_d are computed based on z_{dn} by formula (3). These procedures are iterated until the sampled values are converged. The convergence is assessed by perplexity (4) which approaches asymptotically to a constant value at the convergence (refer to **Table** I for the description of each variable in detail).

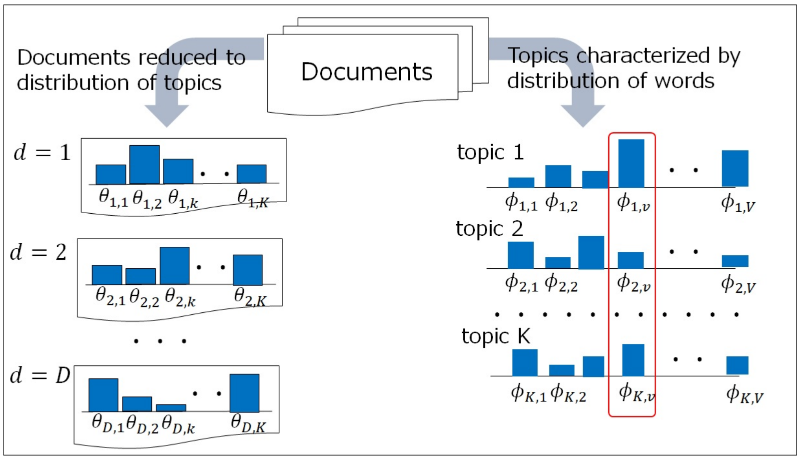


Fig. 2. Topic model inferred parameters of distributions : Both topic distribution θ_d and word distribution ϕ_k are simultaneously estimated by the inference. The original dimension of the documents are reduced to θ_d, and the context of each topic is represented by ϕ_k. Data within the red rectangular border represent the semantic weights of a word v in each topic.

$$p\left(z_{dn} = k \middle| \mathbf{W}, \mathbf{Z}^{-\mathbf{dn}}, \alpha, \beta\right) \quad \propto (N_{dk}^{-dn} + \alpha) \frac{N_{k,w_{dn}}^{-dn} + \beta}{N_k^{-dn} + \beta \mathbf{V}} \tag{2}$$

$$\begin{aligned} \phi_{kv} &= \frac{N_{kv} + \beta}{N_k + \beta V} \\ \theta_{dk} &= \frac{N_{dk}}{N_d + \sum \alpha} \end{aligned} \tag{3}$$

$$ppl = \exp\left(\frac{\sum_{d=1}^{D} \sum_{n=1}^{N_d} \log \sum_{k=1}^{K} \theta_{dk} \phi_{kw_{dn}}}{N}\right) \tag{4}$$

Figure 2 illustrates the results of the inference. The bar charts on the right side of **Figure** 2 annotated by topic number are the distributions with each parameter ϕ_{vk} representing a word's probability. Moreover, the words with high probabilities in each distribution represent semantic coherence as if it has a unique context that characterizes each topic. Also, the number of bar charts is the document space dimension.

Each document is transformed to the distribution of topics with the parameters θ_{dk} as illustrated on the left side of **Figure** 2. As a result, the document space dimension is reduced from words' space to topic space (hereafter, the transformed documents in the topic space are referred to as transformed documents).

3.2 Detecting collective motions

Mode signal derives principal components from time series data and the first eigenvector is assumed to be latent collective motions. Then, the first eigenvector is projected onto a vector in each time slice in the space. The time slices with large projected values indicate the situation where collective motions are distinctive. Following the method, principal components from the time series document space are derived. While the mode signal vector space consists of a unique vector allocated to each time slice, the vector space of the documents

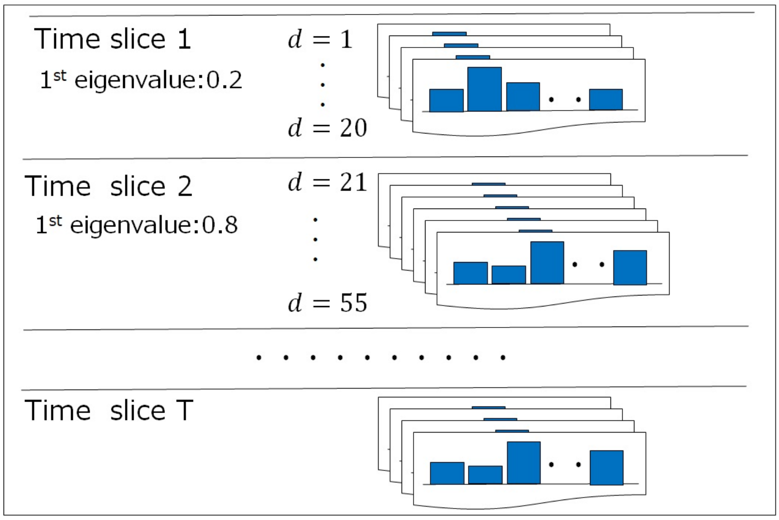


Fig. 3. Deriving 1st eigenvectors from topic vector space by each time slice: Dimensionally reduced documents in Fig.2 are partitioned by time slice, enabling repetitive calculation of eigenvalues with non-sparse topic vector spaces.

could consist of multiple vectors in each time slice because the number of vectors is according to the number of documents (e.g. tweets). Thus, principal components are recalculated by each time slice and time slices with large first eigenvalues are specified, which correspond to the semantic collective motions. **Figure** 3 illustrates the method. The number of transformed documents varies with each time slice. The principal components are derived from each time slice: 1,2, and so on. In this case, time slice 2 indicates the collective motion because of its high first eigenvalue.

3.3 Providing semantic information for the collective motions

The first eigenvector in each time slice forms the principal direction in the topic space of the transformed documents. Moreover, the first eigenvector components are the weights to determine the direction in the topic space. Namely, the components represent the degree of interest in each topic which represents fluctuating interests in the time series. With this respect, the components are referred to as time series topic weights hereafter.

Because documents are transformed into topics, each word in the documents probabilistically belongs to topics (hereafter referred to as word2topic probability). The word2topic probabilities $\phi_{kv}s$ of a word, v, are indicated by data in the red rectangular border vertically across the distributions as shown in **Figure** 2. The word2topic probabilities are constant to maintain semantic coherence throughout the time series. While the word2topic probabilities are constant, time series topic weights and the frequency of each word fluctuate depending on each time slice, reflecting fluctuating interests. Thus, the collective motion semantic information in each time slice is represented by weighted word frequency by the word2topic probability and time series topic weights as given in formula (5), (refer to **Table** I for the detailed description of each variable).

$$\omega_i^\tau = \sum_{t=1}^{T} \sum_{k=1}^{K} n_{t_i} \phi_{kv} g_k^\tau \tag{5}$$

4. Data

Two kinds of data were collected concerning 19 kinds of cryptoassets: time series prices and tweets. The period of collection is from 00:00 to 23:59 UTC on Feb.24.2022, the date of the Russian invasion of Ukraine. Furthermore, the data is sliced by one-minute so that the time series is partitioned into 1440 datasets.

4.1 Price data of cryptoassets

The one-minute price data on February 24, 2020 was downloaded for 19 cryptoassets (BTC, ETH, XRP, ADA, USDT, DOGE, XLM, DOT, UNI, LINK, USDC, BCH, LTC, GRT, ETC, FIL, AAVE, ALGO, and EOS) by using CoinDesk API. Of these cryptoassets, USDT and USDC are stable coins whose values are backed by an external asset such as the US Dollar. The time series data of n-th component at time t is denoted as $x_{n,t}$. Therefore, the logarithmic change of time series is defined as follows:

$$r_{n,t} = \log(x_{n,t}) - \log(x_{n,t-1}) \ , \tag{6}$$

where $n = 1, \cdots, 19$ cryptoassets and $t = 1, \cdots, 1440$ minutes.

4.2 Tweets of cryptoassets

All tweets in English hashtagged by the name of the 19 cryptoassets was collected by the study. The number of tweets amounts to approximately 20,000. Then, morphological analysis is applied and the tweets are tokenized to words limited to nouns, verbs, adverbs, and adjectives.

Next, dictionary and vocabulary lists based on the words were constructed. The vocabulary number forms the document (i.e. tweets) space dimension. Meanwhile, the dictionary is referenced to count the frequency of each word in each tweet. Each tweet is then transformed into a word frequency vector. Thus, according to the dictionary, all the tweets are vectorized to the same dimension. From the 20,000 tweets were acquired 1,120 vocabularies so that every word frequency vector becomes very sparse because each tweet generally comprises around ten words. Furthermore, the vectors are the topic model dataset described in the previous section.

5. Empirical Settings

5.1 Threshold for computing eigenvectors

Time slices with a smaller number of tweets than the number of topics virtually have a smaller dimension than the topic space. This could result in more frequent dominance of the first eigenvalues than in the case of the topic space, leading to nominal collective motions. To avoid this situation, the threshold is set to compute the first eigenvalues so that the values are zero in case the time slices have many tweets less than one of the topics. Therefore, the larger number of topics results in many more zeros. In this case, the number of topics is set to 15, equivalent to the average number of tweets in each time slice.

5.2 The number of iterations

Given the 15 topics, the topic model parameters are inferred by the iterative procedure described in the method section. **Figure** 4 shows the state of convergence where perplexity (4) becomes flat after 50 iterations.

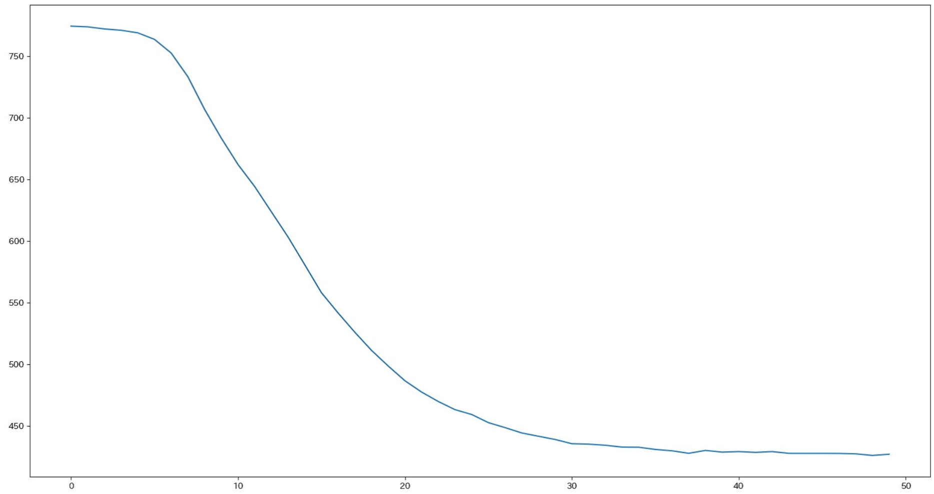


Fig. 4. **Perplexity Convergence** : 50 iterations result in sufficient convergence of the inference of the parameters, ϕ_{kv} and θ_{dk}.

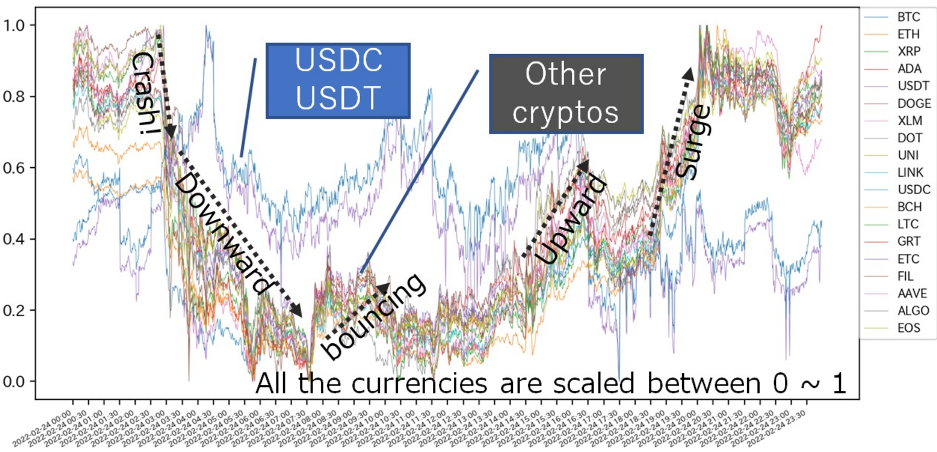


Fig. 5. **Price time series of 19 cryptoassets as of Feb. 24, 2022**: USDC and USDT indicate fluctuating patterns contrary to those of other cryptoassets.

6. Empirical Results

Given the aforementioned data, the study examines whether semantic mode signal adequately captures collective motions in terms of both the peaks and semantic information. The following elaborates on the results.

6.1 Price data time series

Figure 5 illustrates the price data time series fluctuation scaled in the range of 0 to 1. Most currencies show steep crashes followed by a moderate decline in the earlier period of the day. After repetitive up/down movement in the middle term, the trend gradually changes to a clear upward direction toward the end. USDT and USDC show opposing movements against the other cryptoassets probably because these have linkages to USD, the real currency.

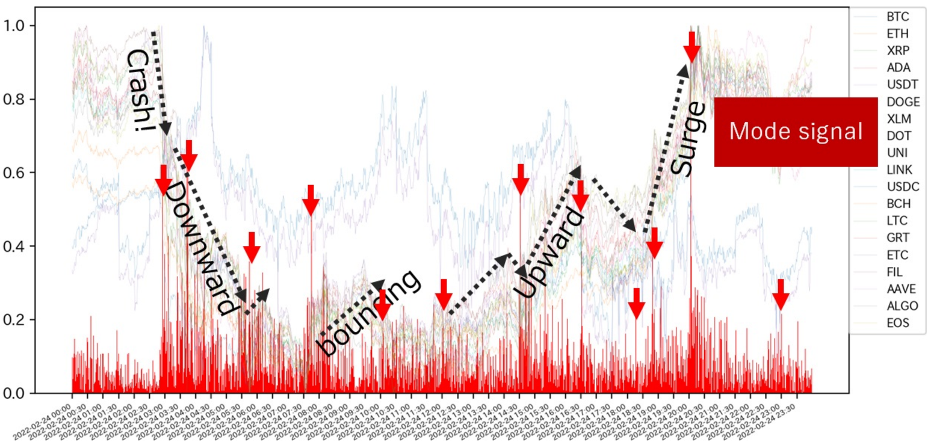


Fig. 6. Collective motions by mode signal: Red arrows indicate distinctive collective motions, each of which locates the changing point of price trends.

6.2 Applying mode signal

The red bar charts in **Figure** 6 are the results of applying mode signal to the price data. Firm peaks coincide with the changing point of trends. The most substantial peak is around 20:00, coinciding with the surging peak of prices which implies important events.

6.3 Applying semantic mode signal

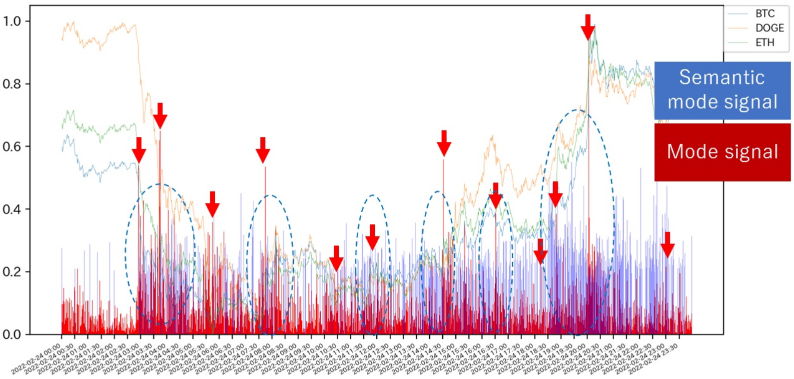
Blue bar charts in **Figure** 7 (a) are the results of applying semantic mode signal to tweets. The height of each bar chart is the first eigenvalue in each time slice based on subsection 3.2 Detecting collective motions. Peaks tend to populate around those of the red bar charts and mode signal with the peaks becoming dense toward the most substantial mode signal peak. For comparison, **Figure** 7 (b) shows the result of counting the number of tweets in each time slice (green bar charts). While semantic mode signal peaks tend to coincide with the ones of mode signal, the number of tweets shows unrelated peaks to mode signal. An example is indicated by the strong peak of the green bar at around 22:00.

6.4 Semantic interpretation of collective motions

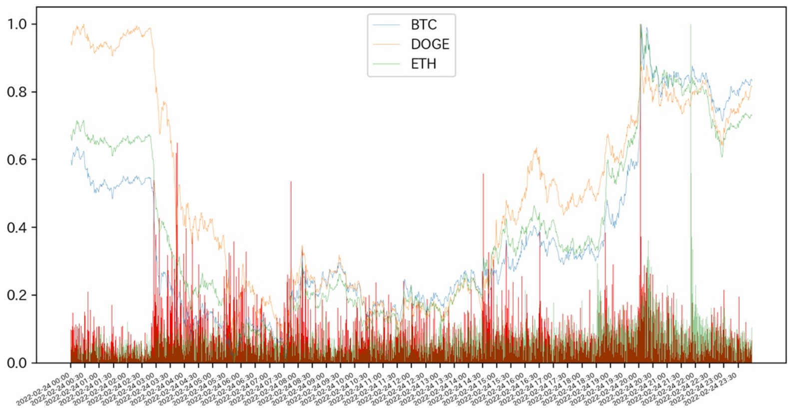
Word clouds in **Figure** 8 show the semantic interpretations for peaks of blue bar charts. The font sizes of the words reflect weighted frequencies defined by formula (5). Each word cloud is annotated by the time of the corresponding bar, the first eigenvalue (i.e. the height of the bar), and the number of tweets in the time slice. Likewise, each bar has its word cloud by one-minute. So, it is impossible to view all the word clouds simultaneously. Therefore, an interactive user interface that enables showing the corresponding word cloud by clicking each bar was developed.

The right side word cloud at 2:58 shows Putin, Russian, and Ukrainian which imply the invasion. Coincidentally, a web news article posted at 3:46 reports that Putin announced the invasion as described by the right side below. At 3:31, another word cloud shows drop, safe, asset, and gold, indicating large fonts. Both the word clouds at 2:58 and 3:46 are the ones during the sharp downward trend of cryptoasset, implying negative responses to the invasion. Market participants attempt to avoid the risk and shift to safe assets like gold.

At 14:29, Putin, the news, and the military remain to be closed up. However, word clouds are indicating a change in people's responses at 18:53 and 19:52. At 18:53, the word cloud is symbolized by human, right, stand, criminalized, and attack. Likewise at 19:52, the word



(a) **Semantic mode signal collective motions** : Semantic mode signal distinctive collective motions (i.e. blue bars) are populated around the ones by mode signal.



(b) **The number of tweets in each time slice**: Distinctive peaks of green bars are not coincident with the ones of mode signal.

Fig. 7. (a) Semantic mode signal collective motions compared with mode signal and (b) the number of tweets

cloud coinciding with the highest peak of mode signal shows similar words. Also, the word "community" is distinctive to the word cloud at 19:52. Since both of the word clouds appear during the upward trend of prices, these imply positive responses as opposed to those in earlier downward trends. In other words, people try to stand against the criminalized attacks and protect human rights, and community is the crypto community trying to donate cryptoassets to Ukraine. Furthermore, these responses coincide with the web news, "Hundreds of people detained in anti-war protests around Russia," posted at 18:26 which is described at the right side below in **Figure** 8

In summary, the semantic mode signal detected a negative response to the invasion in terms of avoiding risk in the early stage. Meanwhile, it detected a positive response to the invasion in terms of strong protest against the event in the latter stage. These semantic features coincide with the price trends as well as peaks of mode signal. For comparison, **Figure** 9 shows the word cloud at each bar corresponding to **Figure** 7 (b). The words with large fonts do not indicate semantic coherence and do not correspond with price trends.

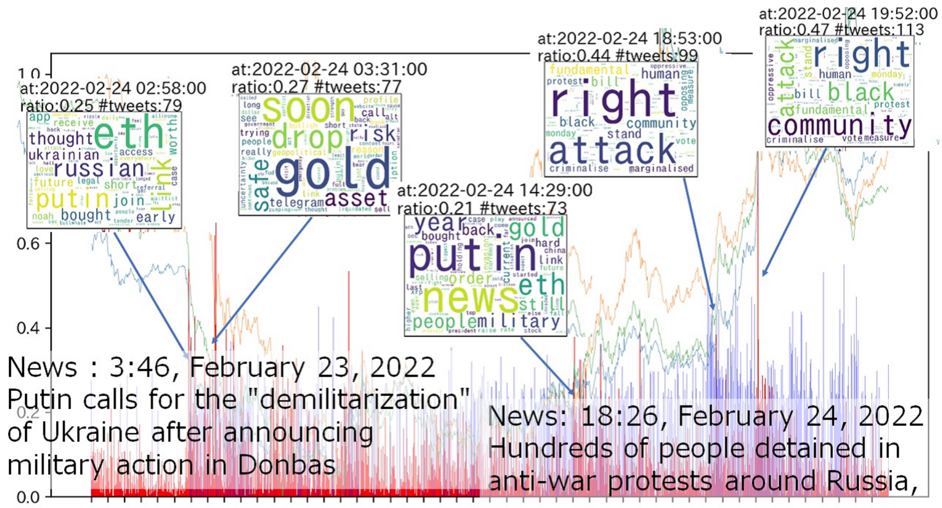


Fig. 8. Semantic interpretation of collective motions by semantic mode signal: Words with larger fonts represent principal semantics for each word cloud. Each word cloud indicates coherent context and coincides with posted news announcements described at the bottom.

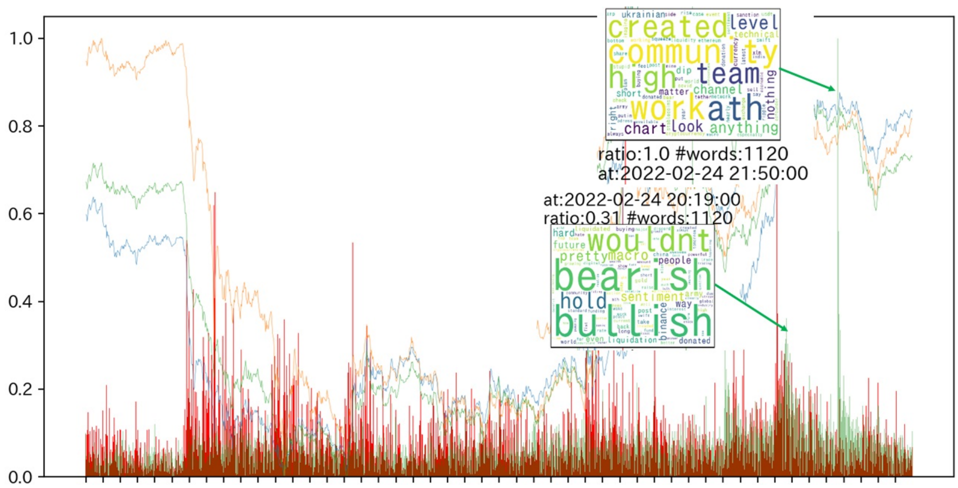


Fig. 9. Semantics of the peaks based on the number of tweets in each time slice: Words with larger fonts are the more frequent ones in each time slice. Each word cloud does not represent a coherent context.

7. Conclusion

This study presented semantic mode signal, a method that detects collective motions as peaks in the time series and provides the semantic interpretation of motions. Semantic mode signal is proposed to overcome the issue of applying mode signal to the vector space with large dimensions with sparsity by implementing the topic model which is a Bayesian model for transforming document space from word vectors to topic vectors. The empirical study of tweets concerning 19 cryptoassets supports the efficiency of the proposed method in terms of the following results:

(1) Semantic mode signal peaks were populated around the ones of mode signal.

(2) The weighted words for each of the major peaks represented semantic coherence coincident with the contents of web news.

(3) Also, the weighted words represented emotions coincident with price trends.

Because the empirical data are limited to short-term tweets focused on cryptoassets, the proposed method needs to be examined for wider varieties of long-term text data to verify its versatility. While this study assumed topics to be constant throughout the period, changes in topics should also be taken into consideration for long-term analysis. Moreover, in the case of long-term analysis, the time slice interval could be wider, like one hour or one day, so that the number of tweets in any time slice might satisfy the threshold for computing eigenvectors described in subsection 5.1. Such time series eigenvalues without forced zero enable quantitative comparison with mode signal using the cross-correlation function. All these issues will be addressed in a future work.

References

[1] Yu, R and Qiu, H and Wen, Z and Lin, CY and Liu, Y, "A survey on social media anomaly detection. ACM SIGKDD Explor. Newsl. 18 (1), 1-14,2016,

[2] Hassan, Muneeb Ul and Rehmani, Mubashir Husain and Chen, Jinjun, "Anomaly detection in blockchain networks: A comprehensive survey," IEEE Communications Surveys & Tutorials, 2022.

[3] Iyetomi, Hiroshi and Nakayama, Yasuhiro and Aoyama, Hideaki and Fujiwara, Yoshi and Ikeda, Yuichi and Souma, Wataru, " Fluctuation-dissipation theory of input-output interindustrial relations," Physical Review E, Statistical, Nonlinear, and Soft Matter Physics,Vol.83(1), p.016103,2011.

[4] Iyetomi, Hiroshi and Nakayama, Yasuhiro and Yoshikawa, Hiroshi and Aoyama, Hideaki and Fujiwara, Yoshi and Ikeda, Yuichi and Souma, Wataru, "What causes business cycles? Analysis of the Japanese industrial production data," Vol.25(3), pp.246-272, 2011.

[5] Arai, Yuta and Yoshikawa, Takeo and Iyetomi, Hiroshi, "Complex principal component analysis of dynamic correlations in financial markets," Intelligent Decision Technologies, pp.111-119, 2013.

[6] Kichikawa, Yuichi and Arai, Yuta and Iyetomi, Hiroshi, "Complex principle component analysis on dynamic correlation structure in price index data," Procedia Computer Science, Vol.60, pp.1836-1845,2015.

[7] Vodenska, I. AND Aoyama, H. AND Fujiwara, Y. AND Iyetomi, H. AND Arai, Y., "Interdependencies and causalities in coupled financial networks," PLoS ONE, Vol.11(3),http://dx.doi.org/10.1371%2Fjournal.pone.0150994,2016.

[8] Aoyama, Hideaki and Fujiwara, Yoshi and Ikeda, Yuichi and Souma, Wataru, "Macroeconophysics: new studies on economic networks and synchronization," Cambridge University Press, 2017.

[9] Souma, Wataru, "Characteristics of Principal Components in Stock Price Correlation, " Frontiers in Physics, Vol.9, https://www.frontiersin.org/article/10.3389/fphy.2021.602944,2021.

[10] Pant, Dibakar Raj and Neupane, Prasanga and Poudel, Anuj and Pokhrel, Anup Kumar and Lama, Bishnu Kumar, "Recurrent neural network based bitcoin price prediction by twitter sentiment analysis," IEEE 3rd International Conference on Computing, Communication and Security (ICCCS), pp.128-132, 2018.

[11] Huang, Teng-Chieh and Zaeem, Razieh Nokhbeh and Barber, K Suzanne, "It is an equal failing to trust everybody and to trust nobody: Stock price prediction using trust filters and enhanced user sentiment on Twitter," ACM Transactions on Internet Technology (TOIT), Vol.19(4), pp.1-20, 2019.

[12] Swathi, T and Kasiviswanath, N and Rao, A Ananda, "An optimal deep learning-based LSTM for stock price prediction using twitter sentiment analysis," Applied Intelligence, pp.1-14, 2022.

[13] Alostad, Hana and Davulcu, Hasan, " Directional prediction of stock prices using breaking news on twitter," IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), Vol.1, pp.523-530, 2015.

[14] Otabek, Sattarov and Choi, Jaeyoung, "Twitter Attribute Classification With Q-Learning on Bitcoin Price Prediction," IEEE Access, Vol.10, pp.96136-96148, 2022.
[15] Al Guindy, Mohamed, "Cryptocurrency price volatility and investor attention," International Review of Economics & Finance, Vol.76, pp.556-570, 2021.
[16] Blei, David M and Ng, Andrew Y and Jordan, Michael I, "Latent dirichlet allocation," Journal of machine Learning research, Vol.3, pp.993-1022, 2003.
[17] Gangwar, Pradeep and Singh, Vrijendra, "Time-Based aggregation for Bi-term Topic Model to Analyze CoVID-19 Twitter Data," IEEE 8th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON), pp.1-5, 2021.
[18] Nguyen, Thien Hai and Shirai, Kiyoaki and Velcin, Julien, "Sentiment analysis on social media for stock movement prediction," Expert Systems with Applications, Vol.42(24), pp.9603-9611, 2015.