KNU KYUNGPOOK NATIONAL UNIVERSITY

Does Twitter Affect Stock Market Decisions?

Financial Sentiment Analysis During Pandemics
: Comparative Study the H1N1 and the COVID-19 Periods

Valle-Cruz et al. (2022)

Presenter : Jonghyun Ha,
Ph.D. student in Economics at KNU

Index

1 Executive Summary

- The relationship between Twitter posts and stock market during the pandemic period.

1. Lexicon-based approach on financial Twitter accounts to detect polarity in financial news on Twitter.
2. The shifted correlation analysis to study relationship between the polarity of Twitter posts and financial indices.
3. The markets reacted faster to Twitter posts during the COVID-19 period than during the H1N1 period.
4. The most influential Twitter accounts were NYTimes, Bloomberg, CNN, and Investing.com.

2 About the Authors

- Professors in different fields(Engineering, Finance, Political science) At UAEMEX

David Valle-Cruz (https://orcid.org/ 0000-0002-5204-8095)

FOLLOW

Universidad Autónoma del Estado de México
Verified email at uaemex.mx - Homepage

Applied Artificial Intelligence Sentiment Analysis Emerging Technologies Data Science E-Government

Vanessa Fernandez-Cortez

Autonomous Mexico State University
Verified email at uaemex.mx

Finance budgeting artificial intelligence

Asdrúbal López Chau

Universidad Autónoma del Estado de México
Verified email at uaemex.mx - Homepage

Data mining Machine learning Embedded Systems Image processing Education

Rodrigo Sandoval-Almazan

Universidad Autonoma del Estado de Mexico
Verified email at uaemex.mx - Homepage

e-government open government social media

3 Keyword : Lexicon-based Approach

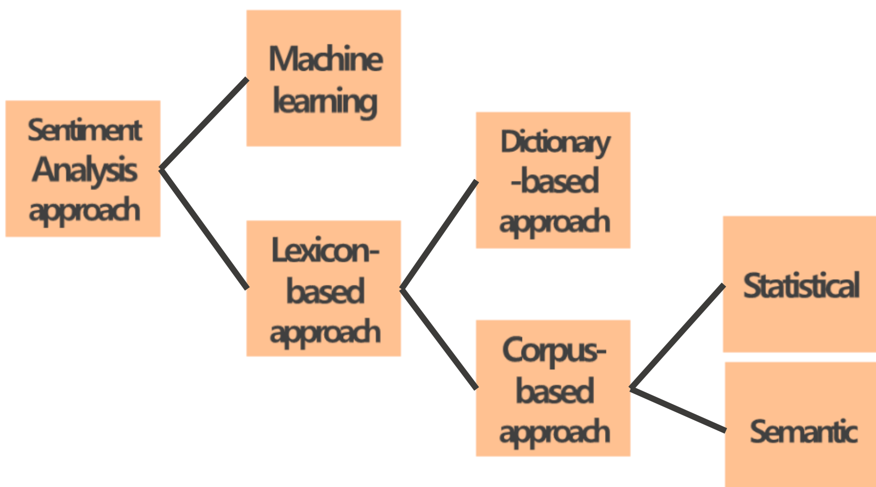
- To determine sentiment orientation, use a lexicon-based approach.

[Lexicon] : The total inventory of morphemes in a given language.

- This technique calculates the sentiment orientations of the whole document or set of sentence(s) from semantic orientation of lexicons.
- Semantic orientation can be positive, negative, or neutral.

Relating to the meanings of words

Types of Sentiment Analysis



Processing

Sentence Splitting ❯ Tokenize ❯ Computing

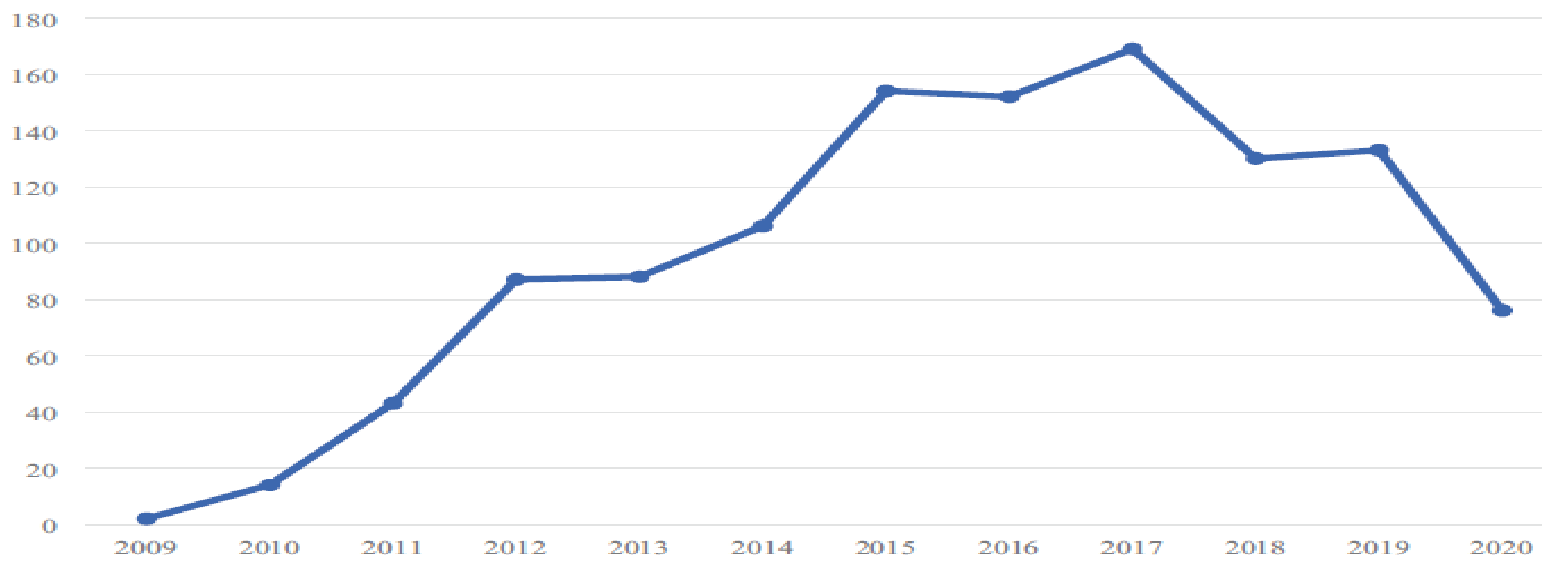
4 State of the Art: Sentic Computing

- The Application of sentic computing in finance has focused on sentiment analysis generated by news related to the stock markets.

[Sentics] : natural language concepts exploited for tasks such as emotion recognition from text/speech or sentiment analysis

- The main goal of sentic computing is to create emotion detection models supported by different disciplines.

Papers related to sentic computing per year



5 Literature Review

1

Investors bet on the credibility of the media outlets that publish rumours despite knowing that the information is not always reliable.

- Oberlechner et al. (2004) and Cruz et al. (2012)

2

However, this trend is growing with the use of online social media networks such as Twitter, Facebook, Instagram and YouTube.

- Doerr et al. (2012)

3

The information shared on social media generates sentiments and reactions in investors that influence their decisions to buy or sell financial assets, especially on the stock markets.

- Kearney et al. (2014)

4

News shared on social media, whether true or not, cause changes in the trends of international stock market indices.

- Jiao et al. (2020)

This study continues this path in exploring pandemic periods

: To study how long an event will affect the financial market

6 Data Extraction(1)

1. For the H1N1 pandemic, the period for data collection was from June to July 2009. For the COVID-19 pandemic, the period for data collection was from January to May 2020.

2. The above periods were selected based on the time when the first stock market index peaked and began to fall, i.e., the point at which the pandemics started to influence financial indices.

3. This study selected adjusted closing prices, where each of the financial indices in the world reached their maximum and minimum(Mexico, USA, UK, Brazil, France, Germany, Japan, China).

4. Twitter accounts selected for sentiment analysis consisted of personal, company, or organizational accounts, as well as news broadcasts that were influential or important for finance

6 Data Extraction(2)

Table 4 Maximum and minimum prices of the stock exchange indices

Index	Maximum price date	Maximum price	Minimum price date	Minimum price	Percentage loss (%)
During the H1N1 pandemic					
IPC	11/June/2009	25,372.84	07/July/2009	23,359.93	− 7.93
NASDAQ 100	11/June/2009	1862.36	07/July/2009	1746.17	− 6.23
Dow Jones	11/June/2009	8770.91	10/July/2009	8146.52	− 7.11
S&P 500	11/June/2009	944.89	10/July/2009	879.13	− 6.95
FTSE 100	10/June/2009	4436.80	10/July/2009	4127.20	− 6.97
BOVESPA	12/June/2009	53,558.00	14/July/2009	48,873.00	− 8.74
CAC 40	11/June/2009	3334.93	10/July/2009	2983.10	− 10.55
DAX	11/June/2009	5107.25	10/July/2009	4576.31	− 10.39
Hang Seng	11/June/2009	18,791.02	13/July/2009	17,254.63	− 8.17
Nikkei 225	12/June/2009	10,135.82	13/July/2009	9050.33	− 10.70
SSE Composite	10/June/2009	2816.25	12/June/2009	2743.76	− 2.57
During the COVID-19 pandemic					
IPC	12/February/2020	45,338.37	23/March/2020	32,964.21	− 27.29
NASDAQ 100	19/February/2020	9,817.17	20/March/2020	6,879.52	− 29.92
Dow Jones	12/February/2020	29,551.41	23/March/2020	18,591.92	− 37.08
S&P 500	19/February/2020	3386.14	23/March/2020	2237.39	− 33.92
FTSE 100	19/February/2020	7457.02	20/March/2020	5190.78	− 30.39
BOVESPA	19/March/2020	68,332.00	20/March/2020	67,069.00	− 1.84
CAC 40	19/February/2020	6111.24	18/March/2020	3754.84	− 38.55
DAX	19/February/2020	13,789.00	18/March/2020	8441.70	− 38.77
Hang Seng	17/February/2020	27,959.59	23/March/2020	21,696.13	− 22.40
Nikkei 225	12/February/2020	23,861.21	19/March/2020	16,552.83	− 30.62
SSE Composite	13/January/2020	3115.57	23/March/2020	2660.17	− 14.61

6 Data Extraction(3)

- It consisted of personal, company or organizational accounts as well as news broadcasts.

Table 5 Analysed Twitter accounts

Twitter account	Total downloaded tweets	
	2009	2020
@business	50	551
@Carl_C_Icahn	NA*	2
@CNNBusiness	747	1725
@ecb	NA*	298
@Investingcom	NA*	279
@InvestOfficeAD	NA*	120
@jpmorgan	NA*	86
@JPMorganAM	NA*	30
@lloydblankfein	NA*	8
@nytimes	490	244
@SEC_Enforcement	NA*	2
@UBS_CEO	NA*	9
@USTreasury	NA*	78
@WarrenBuffett	NA*	0

*Not applicable because the user did not have a Twitter account in 2009

6 Data Extraction(4)

- Preprocessing of Twitter data

Algorithm 1 Preprocessing of Data

Input: Tweets downloaded

Output: Data cleaned.

For each tweet **do**

Step 1: Split text

Step 2: Remove punctuation symbols, interrogation and exclamation symbols

Step 3: Remove special elements, such as hashtags and html symbols

Step 4: Remove stop words

Step 5: Compute the stem of remaining words

Step 6: Join the words to form a string

6 Lexicon-based Analysis(1)

- Collecting polarity intensity from different lexicons

Algorithm 2 Calculating Polarity Intensity

Input: Text **T**, Lexicon **L**

Output: Polarity intensity of **T**

Step 1: Split **T** into a list ***W*** of words

For each $\boldsymbol{w} \in \boldsymbol{W}$ **do**

Step 2: Recover$polarity\ intensity\ of\ \boldsymbol{w}$: $value\ of\ term\ \boldsymbol{t} \mid \boldsymbol{t} \in \boldsymbol{L}, \boldsymbol{t} = \boldsymbol{w}$

Step 3: Convert *polarity* to a real number (see Table 6)

Step 4: If polarity is positive, then assign $\boldsymbol{w_{pi}} \leftarrow polarity$

Step 5: Else assign $\boldsymbol{w_{ni}} \leftarrow polarity$

Step 6: Compute positivity and negativity of **T** (see equations (1) and (2), respectively)

6 Lexicon-based Analysis(2)

- Polarity values were transformed into numerical values.

Table 6 Transformation of values of polarity into real numbers

Lexicon	Rule
Bing Liu	Value "positive" is transformed into + 1 Value "negative" is transformed into − 1
Sentiment 140	Value 0 is transformed into − 1 Value 2 is transformed into 0 Value 4 is transformed into + 1
NRC	Sentiments (+) are transformed into + 1 Sentiments (-) are transformed into − 1
Affin	Values are divided by 5
SenticNet	polarity_intense () method

6 Lexicon-based Analysis(3)

-For each tweet, the strength of polarity was computed and normalized according to Euclidean norm.

$$\text{positivity} = \frac{\sum w_{pi}}{\sqrt{(\sum w_{pi})^2 + (\sum w_{ni})^2}} \qquad (1)$$

$$\text{negativity} = \frac{\sum pw_{ni}}{\sqrt{(\sum w_{pi})^2 + (\sum w_{ni})^2}} \qquad (2)$$

where

w_{pi}: positive weight of ith word in text **T**.

w_{ni}: negative weight of ith word in text **T**.

Table 7 Hypothetical example of a matrix $\boldsymbol{M}_{pricesentic}$ of dates, adjusted prices and polarities of posts

Date	AdjPrice	S140+	S140-	Bing+	Bing-	NRC+	NRC-	Affin+	Affin-	SN+	SN-
12-March-2020	2923.49	− 0.56	0.24	0.33	− 0.14	0.12	− 0.16	0.58	− 0.06	0.05	0.57
13-March-2020	2887.43	0.10	0.31	0.34	− 0.18	0.50	− 0.20	0.44	− 0.41	0.70	− 0.52
16-March-2020	2789.25	0.50	0.28	− 0.07	− 0.59	0.14	0.33	0.52	0.52	− 0.03	0.44

6 Date-Shifted Correlations

- They searched for the date to find the value of the offset that produced the most significant correlation in an absolute value .

Algorithm 3 Detecting correlations with date shift

Input: ***Index***, ***date-shift*** value

Output: Maximum value of correlation $\boxed{\left(\max_{shift-date} |r(Mpricesentic)|\right)}$

Step 1: Apply date shift to the dates of posts.

Step 2: Compute $M_{pricesentic}$ corresponding to *index*

Step 3: Compute the correlation between column **AdjPrice** and polarities

Step 4: Search for the maximum absolute value of correlations (see equation (4))

Table 8 Example of polarity vector with a date shift $= -7$

Date	S140	Bing+	Bing-	NRC+	NRC-	Affin	SenticNet	Adj.Close
2020/May/12	0.000	0.29	0.25	0.48	0.30	– 0.02	– 0.58	2891.56
2020/May/13	0.000	0.23	0.33	0.51	0.38	– 0.03	0.59	2898.05

7 Results

1. They identified the best correlations to understand financial market behavior and sentiment analysis during H1N1 and COVID-19 periods.

2. They found that their method is better at detecting highly shifted correlations by using SenticNet compared with other lexicons.

Table 10 Most significant correlations in absolute value for each data set

Data set	H1N1 (2009)			COVID-19 (2020)		
	Shift	$\max_{shift-date} \|r\|$	lexicon	Shift	$\max_{shift-date} \|r\|$	lexicon
CNNBusiness	− 2	0.46	SenticNet	− 6	0.70	SenticNet
Bloomberg	+2	0.54	SenticNet	− 7	1.00	SenticNet
NYTimes	− 1	0.39	SenticNet	− 13	0.97	SenticNet
Investing.com	NA	NA	SenticNet	4	0.93	SenticNet

8 Discussion

1. The findings show that some Twitter posts had a more significant effect in the COVID-19 era than in the H1N1 period, because the use of Twitter was more widespread in 2020, and more accounts published information regarding COVID-19 and its effects on finances.

2. It took 0 to 10 days for markets to react to information shared and disseminated on Twitter during the COVID-19 pandemic. During the H1N1 pandemic, this period was from 0 to 15 days.

3. It indicates that the behavior of the stock market affects the reactions of Twitter users. In the case of H1N1, this took 1 to 11 days, and in the case of COVID-19, this took 1 to 6 days.

4. The findings show that the number of followers on Twitter influences the performance of financial indices, regardless of whether the publication is accurate or not.

9 Conclusion (1)

1. The combination of a lexicon-based approach is enhanced by a shifted correlation analysis, as latent or hidden correlations can be found in data.

2. There is an important effect of sentiments on Twitter on financial indices, and this effect is observed a few days after the information is posted on Twitter.

3. They confirmed that SenticNet performed better than other lexicons.

4. They proposed a correlation matrix that contains the polarity of Twitter posts and the behavior of financial markets to study their relationship when the date is shifted.

9 Conclusion (2)

5 The data show some important effects days after and days before the posts were made.

6 They also confirm that social media propagation—more Twitter accounts—over this period of 11 years has a direct impact on the indices' behavior.

7 They presented the state of the art on sentic computing and finance.

8 The findings confirm that the drop in stock prices during the COVID-19 era, compared with the H1N1 period, was more dramatic because there was more speculation, rumors and negative news.

10 Presenter Comments

1. It is insufficient to explain a causal relationship as to whether Twitter posts influenced the financial indices. In the paper, the causal relationship is determined by whether the date shift has a negative or positive value. However, with the same date shift, both positive and negative values exit.

2. Since they used indices in bearish market after the pandemic, it is difficult to be generalized that the correlation between Twitter posts and stock market indices in general markets, which can be bullish or bearish.

3. In this paper, COVID-19 period H1N1 period are compared, after the emergence of Twitter, but it is hard to compare directly because of the asymmetric number of tweets and the number of accounts in other periods.

KYUNGPOOK NATIONAL UNIVERSITY

경북대학교