Understanding User Engagement in Online Communities during COVID-19 Pandemic: Evidence from Sentiment and Semantic Analysis on YouTube

Chenyang Li
Beijing Normal University-Hong Kong Baptist University United International College, Zhuhai, China, n830021020@mail.uic.edu.cn

Mingze Xie
Beijing Normal University-Hong Kong Baptist University United International College, Zhuhai, China, n830021039@mail.uic.edu.cn

May Wang
Beijing Normal University-Hong Kong Baptist University United International College, Zhuhai, China, ywang@uic.edu.cn

Li, C., Xie, M. & Wang, M. (2021). Understanding user engagement in online communities during covid-19 pandemic: Evidence from sentiment and semantic analysis on YouTube. In *Proceedings of The International Conference on Electronic Business, Volume 21* (pp. 651-656). ICEB'21, Nanjing, China, December 3-7, 2021.

Understanding User Engagement in Online Communities during COVID-19 Pandemic: Evidence from Sentiment and Semantic Analysis on YouTube

(Work in Progress)

Chenyang Li [1,2]
Mingze Xie [1,3]
May Wang [1,4,*]

*Corresponding author
[1] Beijing Normal University-Hong Kong Baptist University United International College, Zhuhai, China,
[2] Undergraduate Student (Year 4), n830021020@mail.uic.edu.cn
[3] Undergraduate Student (Year 4), n830021039@mail.uic.edu.cn
[4] Associate Professor, ywang@uic.edu.cn

ABSTRACT

Since the outbreak of COVID-19, the pandemic has changed the lives of many people and brought dramatic motional experiences. Among many social media platforms, YouTube saw the most significant growth of any social media app among American users during the pandemic, according to the Pew Research Center on 7th April 2021. Exposure to COVID-19 related news can have a significant impact on user engagement on social networks. Different news may trigger different emotions (i.e., anger, anticipation, disgust, fear, joy, sadness, surprise, or trust), and a user may engage differently in response to the news. On YouTube, user engagement is manifested through actions such as liking, disliking, commenting, or sharing videos. During the pandemic, many users provide constructive comments that are encouraging, respectful, and informative to support each other. We applied sentiment analysis in the study to investigate different emotions and applied semantic analysis to investigate positive appraisal (i.e., encouraging, respectful, and informative) to identify salient factors that can motivate user engagement. The findings of the work shed light on how social network platforms could encourage constructive comments to help people provide emotional support to each other during pandemics through using positive appraisal in online news comments.

The first research objective is to study the impact of sentiment valence of different emotions on people's liking of news comments. News about COVID-19 on social networks may provide valuable information but also bring about public panic. In response to this COVID-19 related news, reviewers expressed their feelings by clicking the like, dislike buttons to the video and comments, or writing some comments under the video on YouTube. Some positive news was followed by comments expressing their anticipation, joy, and trust, while negative news might trigger sadness, fear, disgust, or anger. Our research focuses on sentiment analysis of news titles and the comments following each video. News title provides important information about the video, showing the summary of the video and allowing people to get a first glimpse of the content of the video. Through sentiment analysis of title and comments, correlations could be found between title/comments sentiment and user engagement.

The second research objective is to investigate the impact of comments' positive appraisal (i.e., encouraging, respectful, and informative content) on user engagement. The informative comments under the negative news have significant implications for the audience. They can be considered as a complement or judgment of the video content. Encouraging and respectful comments also help people build good conversations online. Our research focuses on semantic analysis of news titles and comments based on the three dimensions of positive appraisal and analyzes their impacts on user engagement to like the corresponding comment. We discuss the correlation between video title sentiment and the positive appraisal followed in the comments of the video to provide good conversations on the platform.

A group of 38,085 online comments was collected from more than 400 different publishers from January 1st to January 30th, 2021, on YouTube. The dataset contains the most-viewed videos that were related to at least one of the following search queries: coronavirus, COVID-19, pandemic, or vaccine. NRC lexicon is adopted in the sentiment analysis to identify different emotions in titles and comments of the video. We adopt the topic modeling method and build a classifier from the Yahoo News Annotated Comments Corpus to identify constructive online comments for specific topics. We also measure inter-annotator agreements and compare the reliability of manual annotation and the classifier. We find that longer titles and sad emotions can obtain more likes on the comments of COVID-19 related news. During the pandemic, people tend to show their support when they find others are quite sad. We also expect to see correlations between some positive appraisals and user engagement.

Keywords: Sentiment analysis, user engagement, COVID-19, YouTube

INTRODUCTION

Since the outbreak of the COVID-19, many people have had to stay at home during the lockdown, spending more time on the internet. YouTube is one of the biggest video platforms in the world for people to obtain information about the pandemic. Many media companies also have their official accounts on YouTube to enhance user engagement in the online platform. User engagement can be manifested through actions such as liking, disliking, commenting, or sharing videos. While semantic analysis on polarities, such as positive, negative, or neutral emotions on social media, were investigated (Saragih & Girsang,2017; Schumaker *et al*., 2012; Salehan & Kim, 2016), the majority of the literature did not look into the impacts of specific emotions in comments (Chen *et al.* 2017, Khan 2017) on user engagement. During the pandemic, news videos cause various emotional responses during the early few weeks after the outbreak of the COVID-19 (Haroon & Rizvi 2020). Therefore, the first research questions of this study are:

RQ1: How will specific emotions lead to more readership of video/comments on YouTube?

During a pandemic, users may provide constructive comments that are encouraging, respectful, and informative to support each other on a social network platform. The second research question is:

RQ2: How will positive appraisals lead to more readership of video/comments on YouTube?

HYPOTHESIS AND RESEARCH MODEL

In the model, we will consider different emotions of sentiment in the text of titles and reviews besides general sentiment (i.e., positive, negative, or neutral). Polarity refers to sentiment direction, which can be positive, negative, and neutral in the results. The research model is illustrated as the following (see Figure 1):

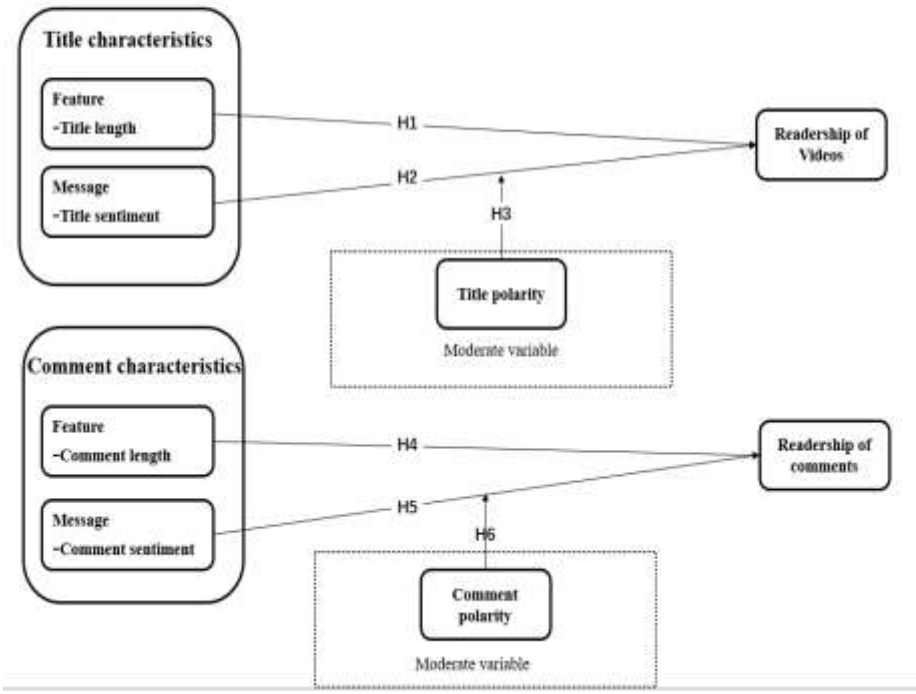

Figure 1: Research model

Hypotheses Development

Effect of emotions on readership of videos

A short title usually does not contain enough information. Some examples are “Will COVID-19 vaccine passports work?” or “Cahn Health Hour COVID 19 Vaccine Panel”. Compared with short sentences, long sentences provide more information to attract the audience to read. For example, “Health Official Warns 'Wait-And-See Approach' On Covid Vaccine Could Be Deadly.” Therefore, we hypothesize that:

H1. The length of the title of a review has a positive effect on the readership of the video.

According to the selective attention perspective, people only have limited attention when they are processing information. The text with stronger emotions can gain more attention and feedback compared to the text with neutral emotions. Therefore, we hypothesize that:

H2. The larger the total amount of sentiment the title of a video exhibits, the more readership it may receive.

According to the selective perception theory, people will filter the information because of the people's mental structure. People are more willing to absorb the information that is fixed with their mental structure instead of choosing to believe information that is not meet their mental structure. People are like to read and click likes to the opinion in the news that they are agreed. Therefore, we hypothesize that:
H3. Title polarity moderates the effect of title sentiment on the readership of the video. The effect will be larger for a positive title than negative and neutral ones.

Effect of emotions on readership of comments
The longer comment is more persuasive than the short comment. Longer comments may contain more information, and it is easily empathizing with others, so it also will be considered it is more helpful than short content. For example, short comments only include a general idea about the news, such as "nice" or "Good news, sir." For longer comments, for example, "Pfizer's figure of 95 % is fake too, figures were manipulated or omitted in 3rd phase trial, the efficacy is roughly 30 %.", it gives keyword information like "Pfizer" "fake" and many data to support the idea. Thus, we hypothesize that:
H4. The length of the comment has a positive effect on the readership of the comments.

Following similar logic in *H2* and *H3*, we propose that:
H5. The larger the total amount of sentiment the comment exhibits, the more readership it may receive.
H6. Comment polarity moderates the effect of title sentiment on the readership of the video. The effect will be larger for a positive title than negative and neutral ones.

METHOD

A group of 38,085 online comments from more than 400 different publishers was collected from YouTube. The videos in this dataset are between January 1st, 2021, and January 30th, 2021. The dataset contains the most-viewed videos that were related to at least one of the following search queries: coronavirus, covid-19, pandemic, vaccine. We selected videos whose total comments likes were no less than 1 to ensure there is a minimum number of likes accumulated for the comments. The final sample consisted of 10,860 comments with 201 videos. Semantic analysis is applied to investigate eight emotions contained in each video title and its comment based on NRC Emotion Lexicon. Sentiment polarity as well sentiment are calculated

Polarity = positive sentiment + negative sentiment.

The positive sentiment has a range from 1 to 5, and the negative sentiment has a range from -1 to -5, so the polarity will range from -4 to 4. The other method to assess the sentiment of the information is to calculate the total amount of sentiment in each information. The negative sentiment output is a negative number. We calculate the absolute value of it and then use the number to add up with the positive number. The following is our formula of the readership, the equation 2:

Sentiment = (positive sentiment – negative sentiment) – 2

The positive sentiment has a range from 1 to 5, minus the negative sentiment, which has a from -1 to -5, so the total sentiment will have a range of 2 to 10. To normalize it, we minus two from (positive sentiment－negative sentiment). Finally, it will have a range from 0 to 8 (Salehan and Kim 2016).

The length of the title and comments was obtained by calculating the characters of the information. Longevity was measured by counting the number of days since the review was created. The readership of a video's title is measured by the total number of likes and dislikes in a video. The agreement of a video's comment was measured by dividing the number of likes by the total number of comments' likes under each video to reduce the problem of a low number of likes per comment due to a low number of total video comments.

DATA ANALYSIS AND RESULTS METHOD

We used negative binomial regression to test our first model because our dependent variable follows a negative binomial distribution. Descriptive statistics, correlation analysis, and regression coefficients were provided in Appendix.

Table 1: Hypotheses Testing Result

1. H1: Title Length → Readership	-0.164***	Supported
2. H2: Title Sentiment → Readership	-0.163***	Supported
3. H3: Title polarity moderates the effect of title sentiment.	-0.126***	Supported
4. H4: Comment Length → Readership	0.079***	Supported
5. H5: Comment Sentiment → Readership	-0.12	Not supported
6. H6: Comment polarity moderates the effect of comment sentiment	0.016	Not supported

*** $p < 0.001$ ** $p < 0.01$ * $p < 0.5$

From the descriptive statistics in Table A.1, we can find that the most salient emotions in title sentiment are anticipation (mean=0.26), trust (mean=0.24), and fear (mean=0.17). The most salient emotions in content sentiment are trust (mean=0.57), anticipation (mean=0.47), and fear (mean=0.36). As shown in Table 1, H1-H4 is supported. However, H5 on general comment sentiment and H6 is not supported. Table A.4 shows that among all specific emotions, sadness was found to be a significant predictor of the number of likes for comments among all eight emotions ($b = -0.047$, $p < 0.05$). We could also find that the moderation effect of polarity.

DISCUSSION AND CONTRIBUTIONS

We made contributions through investigating more specific emotions instead of general sentiment, which can only address positive, negative, and neutral polarity. The study provides the value to investigate different dimensions of sentiment in a different context. Since the pandemic brings a dramatic change to people's lives, it may change the distribution of different emotions on the internet. People tend to be more supportive of comments which are sad and try to support each other during this difficult time. Other emotions are not as salient as sadness in the context of COVID-19 at its very beginning. The moderating effect of polarity of emotion also indicates that people tend to provide sympathetic responses to others and to seek resonance.

LIMITATIONS AND FUTURE WORK

There are several limitations of the study. First, the sample of the video includes only one month's data in January 2021. Second, there isn't any information about the replies to each comment thread in the data. Therefore, the data analysis does not consider the interactions of sequential comments. Third, the content of the video is not considered at this stage. In the future, more comprehensive data with different comment threads will be collected, and panel data analysis will be adopted. The content of the video should also be taken into consideration in future research.

To answer the second research question, we adopt the topic modeling method and build a classifier from the Yahoo News Annotated Comments Corpus to identify constructive online comments for specific topics. We also measure inter-annotator agreements and compare the reliability of manual annotation and the classifier. In a pretest, we found that positive appraisal may affect user engagement in a different way on different topics, i.e., COVID-19 and vaccine.

In conclusion, we confirm previous findings that longer titles and comments could attract more likes. The general sentiment (being positive, negative, or neutral) in the video title is salient to attract users. However, general sentiment for comments is not significantly correlated with likes. Sadness emotion in the comments can trigger more likes to the comments. During the pandemic, people tend to show their support when they find others are quite sad. We also expect to see correlations between some positive appraisals and user engagement in different topics.

ACKNOWLEDGMENT

This work is supported by the college student research grant (Grant No. A1067).

REFERENCE:

Chen, Y.-L., Chang, C.-L. & Yeh, C.-S. (2017). Emotion classification of YouTube videos. *Decision Support Systems, 101*, 40-50. doi: 10.1016/j.dss.2017.05.014

Haroon, O. & Rizvi, S. A. R. (2020). COVID-19: Media coverage and financial markets behavior—A sectoral inquiry. *Journal of Behavioral and Experimental Finance, 27*, 100343. doi: 10.1016/j.jbef.2020.100343.

Khan, M. L. (2017). Social media engagement: What motivates user participation and consumption on YouTube?. *Computers in Human Behavior, 66*, 236–247. doi: 10.1016/j.chb.2016.09.024.

Lee, J. (2013). What makes people read an online review? The relative effects of posting time and helpfulness on review readership.*Cyberpsychology, Behavior And Social Networking, 16(7)*, 529-535. doi: 10.1089/cyber.2012.0417.

Mudambi, S. & Schuff, D. (2010). What makes a helpful online review? A study of customer reviews on Amazon.com. *MIS Quarterly, 34(1)*, 185-200. doi: 10.2307/20721420.

Salehan, M. & Kim, D. J. (2015). Predicting the performance of online consumer reviews: A sentiment mining approach to big data analytics. *Decision Support Systems. 81*, 30-40. doi: 10.1016/j.dss.2015.10.006.

Saragih, M. H. & Girsang, A. S. (2017). Sentiment analysis of customer engagement on social media in transport online. In *Proceedings of 2017 International Conference on Sustainable Information Engineering and Technology* (pp. 24-29). SIET 2017, Malang, Indonesia, November 24-25.

Schulz., S., Schulz, S., Baak, M., Stahl, G., & Adams, B. (2021). Restorative practices for preventing/countering violent extremism: an affective-discursive examination of extreme emotional incidents. *British Journal of Sociology of Education, 42(8)*, 1227-1245. doi: 10.1080/01425692.2021.1993789

Schumaker, R. P., Zhang Y. L., Huang, C. N., & Chen, H. (2012). Evaluating sentiment in financial news articles. *Decision Support Systems, 53(3)*, 458-464. doi: https://doi.org/10.1016/j.dss.2012.03.001

APPENDIX A:

Table A.1: Descriptive Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
title_length	10838.00	11.49	3.12	5.00	19.00
title_positive	10838.00	1.14	0.40	1.00	3.00
title_negative	10838.00	-1.43	0.71	-4.00	-1.00
title_polary	10838.00	-0.29	0.87	-3.00	2.00
title_positive_dummy	10838.00	0.11	0.31	0.00	1.00
title_sentiment	10838.00	0.58	0.76	0.00	3.00
likes	10838.00	2524.00	4193.84	13.00	25716.00
dislikes	10838.00	627.53	1977.39	4.00	20509.00
totle_likes	10838.00	3151.53	4924.82	78.00	29447.00
comment_count	10838.00	986.67	1465.14	16.00	11679.00
comment_positive	10838.00	1.59	0.76	1.00	5.00
comment_negative	10838.00	-1.62	0.91	-5.00	-1.00
comment_sentiment	10838.00	1.21	1.18	0.00	7.00
comment_polary	10838.00	-0.02	1.20	-4.00	4.00
review_neutral	10838.00	0.41	0.49	0.00	1.00
comment_length	10838.00	17.38	18.59	1.00	175.00
comment_likes	10838.00	20.70	98.61	0.00	3473.00
title_totle_like	10838.00	1971.86	2923.55	41.00	14799.00
comment_anger	10838.00	0.23	0.54	0.00	6.00
comment_anticipation	10838.00	0.47	0.85	0.00	9.00
comment_disgust	10838.00	0.18	0.47	0.00	6.00
comment_fear	10838.00	0.36	0.79	0.00	20.00
comment_joy	10838.00	0.33	0.72	0.00	9.00
comment_sadness	10838.00	0.27	0.61	0.00	6.00
comment_surprise	10838.00	0.23	0.54	0.00	7.00
comment_trust	10838.00	0.57	1.01	0.00	11.00
title_anger	10838.00	0.15	0.40	0.00	2.00
title_anticipation	10838.00	0.26	0.49	0.00	2.00
title_disgust	10838.00	0.09	0.28	0.00	1.00
title_fear	10838.00	0.17	0.40	0.00	3.00
title_joy	10838.00	0.09	0.29	0.00	1.00
title_sadness	10838.00	0.12	0.35	0.00	2.00
title_surprise	10838.00	0.11	0.32	0.00	2.00
title_trust	10838.00	0.24	0.46	0.00	2.00

Table A.2: Correlations

	likes_total_lil	anger	anticipation	disgust	fear	joy	sadness	surprise	trust	log_length	sentiment_	log_longev
likes_total_likes	1.00											
anger	0.02	1.00										
anticipation	0.01	0.17	1.00									
disgust	0.02	0.55	0.12	1.00								
fear	0.01	0.52	0.26	0.40	1.00							
joy	0.00	0.09	0.55	0.05	0.15	1.00						
sadness	0.01	0.57	0.14	0.51	0.61	0.08	1.00					
surprise	0.02	0.35	0.46	0.20	0.28	0.43	0.26	1.00				
trust	0.01	0.10	0.43	0.08	0.17	0.58	0.09	0.32	1.00			
review_neutral	0.00	-0.16	-0.14	-0.15	-0.17	-0.18	-0.16	-0.14	-0.15			
log_length	0.05	0.26	0.35	0.19	0.31	0.25	0.29	0.24	0.33	1.00		
sentiment_neutral	0.02	0.15	0.11	0.13	0.14	0.13	0.14	0.11	0.11	0.23	1.00	
log_longevity	-0.20	0.03	0.04	0.01	0.10	0.03	0.05	0.02	0.04	0.08	0.03	1.00

Table A.3: Readership of Video

----------------------------	------------	------------	--------	----------	----------------------
log_total_likes |	Coef.	Std. Err.	t	P>|t|	Beta
---------------------------+	------------	------------	--------	----------	---------------------
avg_comment_disgust	-1.66	0.12	-14.24	0.00	-0.12
avg_comment_surprise	-0.91	0.11	-8.19	0.00	-0.08
log_avg_comment_length	-0.30	0.05	-6.01	0.00	-0.07
avg_sentiment_neutral	-0.01	0.01	-0.83	0.41	0.00
log_avg_comment_longevity	0.21	0.01	14.75	0.00	0.15
title_sentiment	-0.21	0.01	-14.65	0.00	-0.12
title_length	-0.06	0.00	-13.31	0.00	-0.13
sentiment_positive	0.12	0.02	4.62	0.00	0.03
log_title_longevity	0.07	0.01	5.92	0.00	0.04

Table A.4: Readership of Comments

----------------------------------	------------	-----------	---------	-------	--------------------------
comment_likes_total_likes	Coef.	Std. Err.	t	P>|t|	Beta
---------------------------------	------------	-----------	---------	-------	-------------------------
comment_anger_polarity	0.00	0.00	0.72	0.47	0.02
comment_anticipation_polarity	0.00	0.00	0.92	0.36	0.02
comment_disgust_polarity	0.00	0.00	1.63	0.10	0.04
comment_fear_polarity	0.00	0.00	1.50	0.13	0.03
comment_joy_polarity	0.00	0.00	-1.44	0.15	-0.03
comment_sadness_polarity	0.00	0.00	-2.20	0.03	-0.05
comment_surprise_polarity	0.00	0.00	-0.18	0.86	0.00
comment_trust_polarity	0.00	0.00	-0.05	0.96	0.00
comment_anger_anger_polarity	0.00	0.00	-1.19	0.23	-0.03
comment_anticipation_anticipatio	0.00	0.00	-1.76	0.08	-0.04
comment_disgust_disgust_polarity	0.00	0.00	-1.39	0.17	-0.03
comment_fear_fear_polarity	0.00	0.00	-1.45	0.15	-0.03
comment_joy_joy_polarity	0.00	0.00	1.46	0.15	0.03
comment_sadness_sadness_polarity	0.00	0.00	2.59	0.01	0.06
comment_surprise_surprise_polarity	0.00	0.00	0.91	0.37	0.02
comment_trust_trust_polarity	0.00	0.00	-0.97	0.33	-0.02
review_neutral	0.00	0.00	-0.24	0.81	0.00
log_comment_length	0.00	0.00	7.55	0.00	0.09
sentiment_neutral	0.00	0.00	1.09	0.28	0.01
log_comment_longevity	-0.01	0.00	-25.53	0.00	-0.30
title_sentiment	0.00	0.00	0.56	0.58	0.01
title_length	0.00	0.00	-2.60	0.01	-0.05
sentiment_positive	0.00	0.00	3.90	0.00	0.05
log_title_longevity	0.00	0.00	-0.91	0.36	-0.01