Data and Information Management, 2021; 5(1): 110–118

2020 ASIS&T Asia-Pacific Regional Conference (Virtual Conference), December 12-13, 2020, Wuhan, China

Open Access

Yi Zhao, Haixu Xi, Chengzhi Zhang*

Exploring Occupation Differences in Reactions to COVID-19 Pandemic on Twitter

https://doi.org/10.2478/dim-2020-0032
received August 15, 2020; accepted September 15, 2020.

Abstract: Coronavirus disease 2019 (COVID-19) pandemic-related information are flooded on social media, and analyzing this information from an occupational perspective can help us to understand the social implications of this unprecedented disruption. In this study, using a COVID-19-related dataset collected with the Twitter IDs, we conduct topic and sentiment analysis from the perspective of occupation, by leveraging Latent Dirichlet Allocation (LDA) topic modeling and Valence Aware Dictionary and sEntiment Reasoning (VADER) model, respectively. The experimental results indicate that there are significant topic preference differences between Twitter users with different occupations. However, occupation-linked affective differences are only partly demonstrated in our study; Twitter users with different income levels have nothing to do with sentiment expression on covid-19-related topics.

Keywords: occupational differences, COVID-19, Twitter, topic discovery, sentiment analysis

1 Introduction

The outbreak of coronavirus Disease 2019 (COVID-19) is rapidly spreading worldwide and causing a profound effect on various aspects of society. At the time of writing, Johns Hopkins University reported more than 21,056,850 confirmed cases of COVID-19 globally, including 762,293 deaths and 13,100,902 recovered.[1] To slow down the spread of virus, social distancing and self-isolation have been implemented globally, and social media has become an important channel for people to post their opinions and attitudes. The outbreak of the COVID-19 pandemic caused a rapid increase in COVID-19-related information on social media platforms, including YouTube, Facebook, Twitter, etc. (Abd-Alrazaq, Alhuwail, Househ, Hamdi, & Shah., 2020). Extensive research has shown that social media is a popular source of data in understanding public concerns and attitudes, and it is an important way to support crisis communication between the public and the government (Jordan et al., 2018; Shah & Dunn, 2019).

Occupation is one of the noteworthy demographic variables of Twitter users (Sloan, Morgan, Burnap, & Williams, 2015). In a recent study, Kern, McCarthy, Chakrabarty, and Rizoiu (2019) automatically assessed the personality of different occupations based on the tweets and found that personality was related to distinctive occupations. In addition, several studies have found an association between personality and perceptions of the COVID-19 situation (Carvalho, Pianowski, & Goncalves, 2020; Zajenkowski, Jonason, Leniarska, & Kozakiewicz, 2020). The above research may imply that Twitter users engaged in different occupations may have different topic preferences and sentiment expressions, when faced with COVID-19 pandemic. This motivates us to provide a comprehensive analysis based on Twitter data from the perspective of occupation differences, which helps policymakers understand the fine-grained public concerns.

Previous research has been focused on gender-linked affective or topic differences in social media platforms (Thelwall & Thelwall, 2020; Thelwall & Vis, 2017; Vegt & Kleinberg, 2020). Few works have been devoted to occupational differences in response to COVID-19, and

***Corresponding author: Chengzhi Zhang,** Department of Information Management, School of Economics and Management, Nanjing University of Science and Technology, Nanjing, China, Email: zhangcz@njust.edu.cn
Yi Zhao, Haixu Xi, Department of Information Management, School of Economics and Management, Nanjing University of Science and Technology, Nanjing, China

1 https://coronavirus.jhu.edu/map.html.

the research mainly focuses on a particular occupation, including world leaders (Rufai & Bunce, 2020) and college students (Duong,Luo, Pham, Yang, & Wang, 2020). Additionally, the income level may be another factor that affects topic preference and sentiment expression. The purpose of this study is to demonstrate that Twitter users with different occupation respond differently to COVID-19 and different income levels affect topic preferences and sentiment expressions.

2 Related Works

When it comes to discovering topics from Twitter, a majority of studies utilize the LDA technique and its improved versions. Abd-Alrazaq et al. (2020) utilized LDA to identify the main topic in the language of COVID-19 tweets. Guo, Vargo, Pan, Ding, and Ishwar (2016) conducted empirical research on two Twitter datasets for the 2012 presidential election, and they concluded that the LDA generated meaningful results. A lot of extended version of the LDA was used for topic analysis on social media platforms. N-gram LDA technique was used by E. H.-J. Kim, Jeong, Kim, Kang, and Song. (2016) to investigate the topic coverage of Twitter and news publications about the Ebola virus. Yan, Guo, Lan, and Cheng (2013) proposed a bi-term topic model to capture the word co-occurrence directly and enhance the topic quality. Sasaki, Yoshikawa, and Furuhashi(2014) implemented a Twitter–TTM model based on the topic tracking model that is competent in online inference, and the model can capture the dynamic topic trends on Twitter. To improve the interpretation of the topic model, especially for short text documents, Alkhodair, Fung, Rahman, and Hung (2018) developed a new model that combined Twitter–LDA, WordNet, and hashtags and assessed the effectiveness of the model against two real Twitter datasets.

Sentiment analysis for Twitter textual data is also a hot topic. Two main methods have been utilized to classify the sentiment polarity: supervised machine learning method and rule-based method (Hutto & Gilbert, 2014; Kim & Hovy, 2014). Tang et al. (2014) developed a new method that learned sentiment from specific word embedding and outperformed the previous top-performing system. Ren, Wang, and Ji (2016) used SVM for sentiment classification, and LDA was adopted for the improvement of word embedding. Recently, a large pretrained model was adopted for the classification task. BERT and RoBERTa were used by Duong et al. (2020) to examine the Twitter user's sentiment expression. Supervised machine learning methods are powerful for sentiment classification, but large numbers of annotation data are needed for model training, which is costly and time consuming. Rule-based methods, including sentiwordnet (Esuli & Sebastiani, 2006) and Valence Aware Dictionary and sEntiment Reasoning (VADER) (Hutto & Gilbert, 2014), are also widely used on Twitter. Pandarachalil, Sendhilkumar, and Mahalakshmi(2014) proposed an unsupervised method that incorporated SenticNet, SentiWordNet, and SentislangNet, and it performed well on large scales. Chaithra (2019) presented a hybrid approach combining VADER and Naive Bayes to use the comments of the mobile unboxing videos to predict the sentiment, and the result concluded that VADER has improved the performance of the Naive Bayes classifier in predicting the sentiment of the comments.

To our knowledge, no previous study that adopted topic modeling and sentiment analysis to investigate the occupation distinctions in reactions to COVID-19 pandemic, which will be addressed in this study. As a result of a trade-off between cost and accuracy, LDA and VADER methods are applied in the study.

3 Material and Methods

3.1 Data Collection

Using the Twitter IDs provided by Chen, Lerman, and Ferrara (2020) on the repository,[2] we collected 8716,289 tweets from May 1 to 15, 2020. In this study, we focus only on the English-language tweets and Twitter user populations of those who can identify occupational information from their self-description field. Since the self-description field is an open text in which Twitter users can choose to write whatever they like, it's essential to design an appropriate method for occupation extraction. Inspired by Kern et al. (2019), regular expression matching was used to extract occupation information and we obtained 15,984 job titles from their research. The difference is that the job titles are aligned into major occupation groups according to the O*net[3] alternative titles data.[4] For example, the Assignment Editor and Morning News Producer are two job titles that belong to the Producers and Directors group.

2 https://github.com/echen102/COVID-19-TweetIDs.

3 The Occupational Information Network (O*NET) is the primary source of occupational information in the United States, and the O*NET database is the central of the network, which contains hundreds of standardized and occupation-specific descriptors on nearly 1,000 occupations covering the entire U.S. economy.

4 https://www.onetcenter.org/dictionary/20.3/excel/alternate_titles.html.

Table 1
The Distribution of Occupations of Twitter Users

Income Level	Occupation	Occupation Abbreviations	Number of tweets
High	Computer and Information Research Scientists	CIRS	1,652
	Marketing Managers	MM	1,700
	Dentists, General	DEN	1,835
Medium	Management Analysts	MA	1,909
	Business Teachers, Postsecondary	BTP	1,736
	Financial Analysts	FA	1,841
Low	Farmworkers, Farm, Ranch, and Aquacultural Animals	FFRAA	1,876
	Production Workers, All Other	PW	1,750
	Landscaping and Groundskeeping Workers	LGW	1,988

Furthermore, only the major occupation for each Twitter user was reserved during the process of occupation extraction. In our subsequent studies, we only considered the major occupation group as a research object. After removing retweets and filtering out corporate Twitter users with more than 50 tweets, we acquired 622,687 unique COVID-19-related tweets, belonging to 373,773 users, representing 800 occupations.

Due to the diversity of occupational categories and the restrictions of space, we selected nine occupations from different income levels. The classification criteria on occupation type and income levels are lacking, so we incorporate the salary data from the Bureau of Labor Statistics[5] and information from Wall Street News (Suneson, 2019) to generate the following classification standard: Occupations in the high-income level group with incomes over $100,000 per year; occupations in the medium-income level group with incomes $30,000 to $100,000 per year; occupations in the low-income level group with incomes less than $30,000 per year. The nine occupations encompass a variety of social occupations, including technical occupations, managerial occupations, service occupations, etc. Of these nine occupations, the smallest one included 1,121 unique Twitter users. For balance, we randomly sampled 1,121 unique Twitter users from each occupation. The selected occupations are shown in Table 1.

3.2 Method

Many researchers have utilized Latent Dirichlet Allocation (LDA) to identify topics from social media data (Alkhodair et al., 2018; Asghari, Sierra-Sosa, & Elmaghraby, 2020; Giannetti, 2018). LDA is a three-level hierarchical Bayesian model (Blei, Ng, & Jordan, 2003), it is an unsupervised machine learning technique used to create a representation of documents by topic, where each topic consisted of a set of words. In this study, we employed an LDA algorithm from the Python Gensim library.[6]

To obtain a clean corpus, we conducted data preprocessing at first. Then, we used regular expressions to remove URLs, HTML tags, and Twitter user mentions. Next, we also removed punctuation, stop words, and nonprintable characters from tweets. Finally, all tweets are lower-cased, tokenized, and lemmatized. It is well known that phrases are more meaningful than individual tokens. Hence bigrams and trigrams are created and added to the corpus. Before the corpus was fed into the LDA model (Blei et al., 2003), we used two kinds of document representation methods, bag of words (BOW) and term frequency–inverse document frequency (TF-IDF) (Salton & Yu, 1975), to represent tweets. The number of topics that need to be determined before running the LDA and it is a hyperparameter, so CV coherence measure is used to fine-tune our LDA topic model to obtain the optimal topic number (Röder, Both, & Hinneburg, 2015).

Sentiment analysis was also performed in our study. Sentiment analysis was conducted on the cleaned tweets using the VADER, a lexicon- and rule-based sentiment analysis model for social media text (Hutto & Gilbert, 2014). Hutto and Gilbert (2014) compared VADER with multiple methods, including Affective Norms for English Words (ANEW),[7] Linguistic Inquiry and Word Count

5 https://www.bls.gov/ooh/home.htm.

6 https://radimrehurek.com/gensim.

7 http://csea.phhp.ufl.edu/media/anewmessage.html.

Figure 1. Coherence measurement for bag of words (BOW) and term frequency–inverse document frequency (TF-IDF).

(LIWC),[8] SentiWordNet (Esuli & Sebastiani, 2006), the General Inquirer,[9] and machine learning techniques based on SVM, Naïve Bayes, Maximum Entropy, and the results show that the VADER model that is superior to these methods. One significant advantage of using VADER is that it can quickly and accurately obtain the sentiment of each tweet on such a large scale. The sentiment scores ranged from –1 to +1, with -1 as the most negative sentiment and 1 as the most positive sentiment; furthermore, when the score is between −0.05 and +0.05, it means neutral sentiment. The sentiment of each tweet was calculated using the library “vaderSentiment” in Python,[10] which is the implementation of the python version of VADER model.

4 Experiment and Result

4.1 Selection of the Optimal Number of Topics

All 622,687 tweets were fed into the LDA model, we repeated the experiment on a different number of topics (ranging from 2 to 32) and reported the coherence value for two document representation methods. As shown in Figure 1, BOW performed better than TF-IDF, and the coherence value selected 14 as the optimal number of topics. Since the words of Topics 3, 5, 7, 8, and 13 are difficult to assign a specific theme, the authors reached a consensus on selecting nine highly relevant topics as the final result, as shown in Table 2.

As shown in Table 2, we assigned the potential themes to nine topics based on the top 15 most relevant words. Topic 0 discussed the preparation for reopening, with words such as “social distancing,” “reopen,” “school,” and “guideline” indicating this overarching theme. Topic 1 was primarily about U.S. President Donald Trump’s lies about the COVID -19. Topic 2 was about the coronavirus new cases and deaths, which were identified in tweets that mentioned the rapid growth in the number of confirmed cases. Topic 4 talked about free online support, and Twitter users on the topic mainly discussed joining a local free online support team to provide useful pandemic-related information. A potential theme identified in Topic 6 was protests against the stay-at-home order, with the words “fuck” and “shit” showing the feelings for the order. Topic 9 primarily referred to the risk caused by COVID-19, which can affect business, jobs and food. Topic 10 was measures to slow the spread of COVID-19, including wearing masks and staying at home. The theme of research on the vaccine and treatment was identified in topic 11. Topic 12 was related to virus misinformation and fake news, and Twitter users debated whether the novel coronavirus originated in Wuhan’s laboratory and spread around the world.

8 www.liwc.net.
9 http://www.wjh.harvard.edu/~inquirer.
10 https://github.com/cjhutto/vaderSentiment/tree/0150f59077ad3b8d899eff5d4c9670747c2d54c2#introduction.

Table 2
Topics in COVID-19-Associated Tweets

Topic	Themes	Top 15 most relevant words	Number of tweets
0	Preparation for reopening	*social_distancing, reopen, open, school, take, people, back, life, place, economy, say, child, close, measure, guideline*	35,336
1	President's lies about COVID-19	*trump, american, say, president, response, cdc, america, death, lie, white_house, claim, government, administration, medium, covid*	40,838
2	Coronavirus new cases and deaths	*case, death, test, new, report, number, positive, total, covid, update, day, state, rate, high, data*	50,863
3	*	*get, see, come, work, covid, back, hard, time, like, hit, interesting, one, since, bit, well*	35,531
4	Free online support	*help, support, need, crisis, student, work, time, learn, online, community, free, join, impact, new, team*	49,828
5	*	*people, know, get, think, covid, like, make, would, good, thing, one, say, die, want, bad*	107,280
6	Protests against the stay-at-home order	*state, fuck, vote, governor, house, party, game, get, man, right, play, bbc_news, stay-at-home_order, shit, protest*	25,435
7	*	*day, love, read, great, time, watch, video, one, today, story, thank, good, new, share, stayhome*	62,833
8	*	*may, week, month, next, plan, ease, due, year, end, last, start, two, restriction, thread, day*	29,162
9	The risk caused by COVID-19	*health, business, worker, pay, care, government, public, work, need, risk, job, service, food, staff,* help	54,488
10	Measures to slow the spread of COVID-19	*home, stay, people, work, safe, mask, keep, get, wear_mask, order, face, family, back, protect, life*	48,237
11	Research on the vaccine and treatment	*vaccine, world, virus, patient, covid, use, disease, could, doctor, study, find, spread, cure, treatment, people*	39,989
12	Virus misinformation and fake news	*corona, virus, china, india, fight, wuhan, chinese, world, spread, lab, come, country, war, outbreak, govt*	21,210
13	*	*like, look, question, year, well, say, answer, ask, science, london, wow, would, could, break, first*	21,657

Note: * indicates that it is difficult to assign a specific theme to the topic and will not be analyzed in a subsequent section.

4.2 Distribution of Topics among Different Occupations

Figure 2 shows the topic distribution of Twitter users engaged in different occupations. Topics 3, 5, 7, 8, and 13 are grouped into a separate group (other Topics).

Overall, there is a significant difference in topic concern between Twitter users with different occupations, while Twitter users at different income levels showed a slight tendency toward some topics. For high-income level groups, Topic 2 is more concerned with Computer and Information Research Scientists (CIRS) and Dentists, General (DEN) and Marketing Managers (MM) cares more about Topics 4 and 9, and it indicates the topic preference for Twitter users with difference occupations. In addition, Topic 4 also attracts more attention to high-income and medium-income occupations than to low-income occupations. Twitter users in low-income occupations are more interested in Topic 1 than Twitter users in high-income and medium-income occupations. Compared with other topics, all of Twitter users who engaged in these nine occupations showed less curiosity about the virus misinformation and fake news (Topic 12).

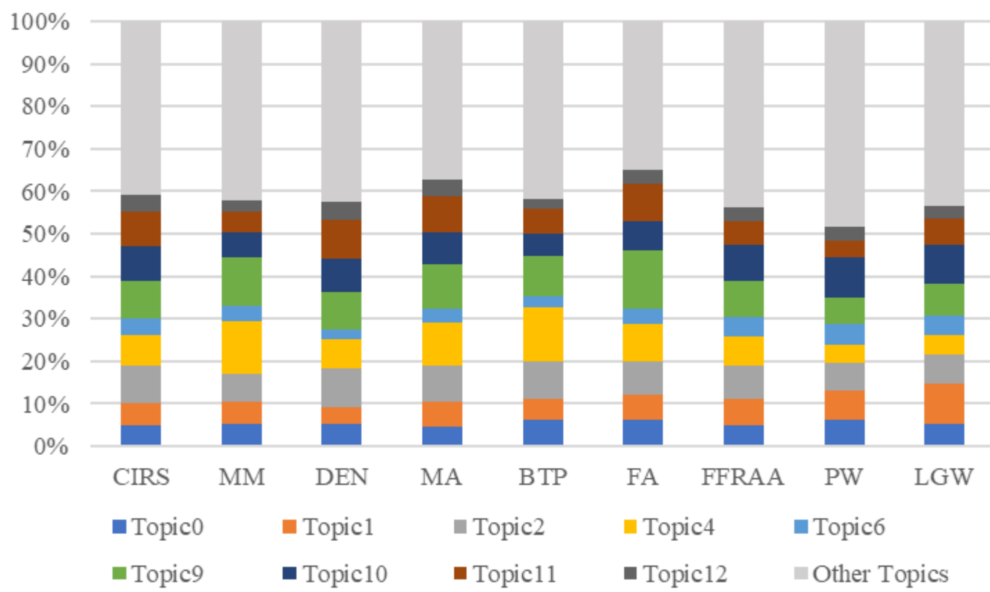

Figure 2. Topic distribution of Twitter users engaged in different occupations.

4.3 Topic-based Sentiment Analysis among Different Occupations

To understand occupation differences in sentimental responses to COVID-19, we analyze the sentiment distribution toward the different topics in Figures 3 and 4.

Overall, Twitter users in different occupations express different sentiments on different topics, but the sentiment expressed by Twitter users of different occupations toward a topic seems to have nothing to do with their income levels. Overall sentiment trends are almost identical for Twitter users with different occupations for most of the topics. For example, regardless of the Twitter user's occupation, negative sentiment ratio is greater than positive sentiment ratio toward Topic 1(President's lies about the COVID-19), that is, U.S. President Donald Trump admitted to journalist Bob Woodward that he played down the severity of the pandemic[11]; the number of negative tweets posted on Twitter on this topic is very alarming. Moreover, online support groups offered medical support to the public during the COVID-19 pandemic, enhancing the public's ability to self-protection (Gong, Xu, Cai, Chen, & Wang, 2020), and thus about 58% of tweets positively responded to the Topic 4 (free online support). At the same time, differences exist in emotional expression toward Topic 2, that is, CIRS, MM, and DEN respond more positively sentiment toward the coronavirus new cases and deaths, whereas, Financial Analysts (FA), Farmworkers, Farm, Ranch, and Aquacultural Animals (FFRAA), Production Workers (PW), and Landscaping and Groundskeeping Workers (LGW) are more likely to express negative feelings on this topic. Additionally, for Topic 6, the positive sentiment ratio of CIRS and LGW is a little higher than that of other occupations, and MA tends to be subject to less positive sentiment on the topic of protesting against the stay-at-home order.

5 Conclusion and Future Works

In this study, we collected 622,687 tweets from 800 occupations and selected nine occupations with different income levels as topic for the research. We found that there was a significant difference in topic concern between Twitter users with different occupations, but sentiment expression differences only existed toward Topic 2, and income level seems to have nothing to do with emotional expression. These findings only partly demonstrate the hypothesis that Twitter users engaged in different occupations have different sentiment expressions. Furthermore, our study finds significant occupation differences in topic preference.

As a short paper, there are also a few limitations. First, we only compare nine occupations, and since the sample size is not big enough, it is necessary to investigate all the occupations and may provide more convincing results. Second, supervised machine learning can apply to topic

11 https://www.cnn.com/2020/09/10/politics/trump-woodward-lies-about-lying-coronavirus-fact-check/index.html.

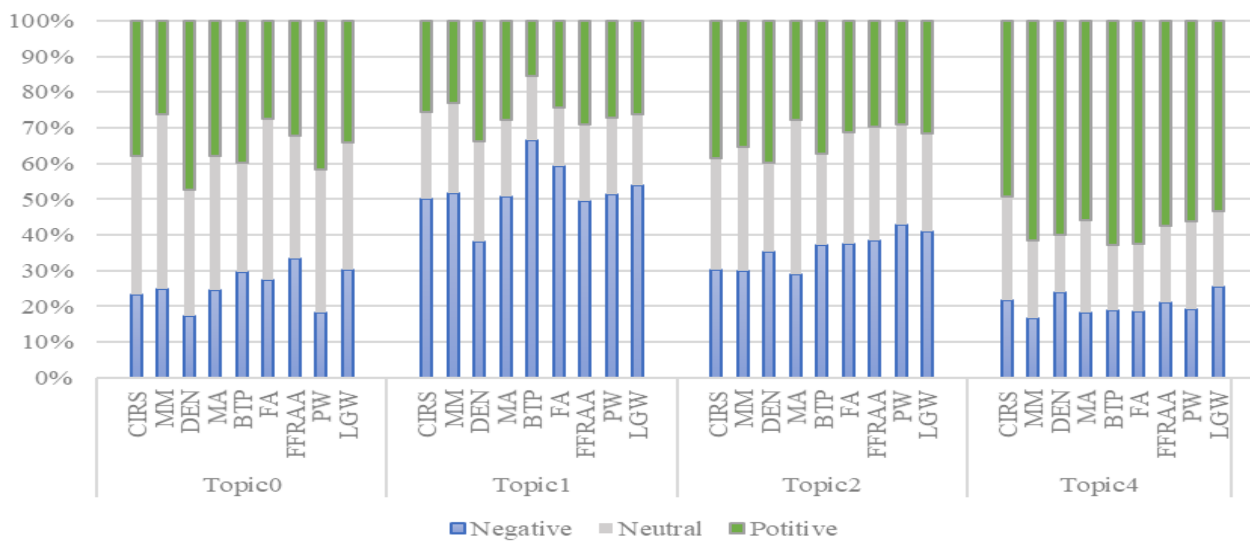

Figure 3. Sentiment distribution on the topics 0, 1, 2, and 4.

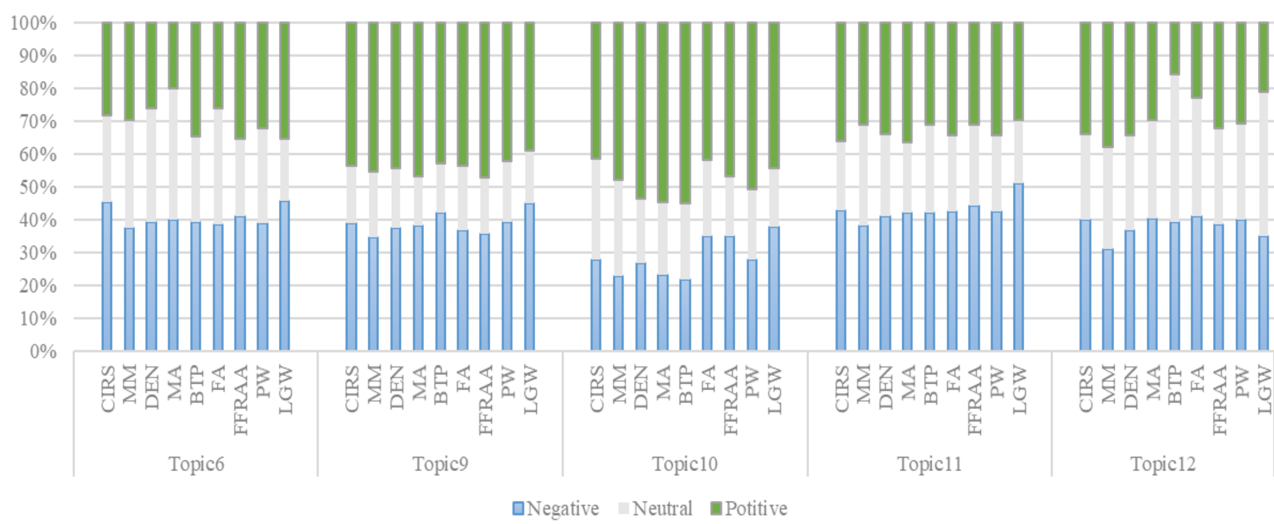

Figure 4. Sentiment distribution on the topics 6, 9, 10, 11, and 12.

discovery and sentiment analysis to improve the topic quality and classification accuracy, it may provide a better analytical result.

There are some related studies that can be implemented in the future. First, this article mainly focuses on sentiment differences toward topics among groups of Twitter users clustered by occupations; a more nuanced exploration should be conducted on emotional difference, which could provide us with an in-depth understanding of people's actions. Second, the topics posted on Twitter have been continuously changing with the development of the COVID-19 pandemic, so merely examining the topic preferences of Twitter users for a given period of time may be insufficient. Future researchers can therefore conduct the research throughout the lifetime of the pandemic and explore that whether the phenomena of topic preferences of Twitter users with different occupations still exist.

Acknowledgements: This work is supported by Jiangsu Social Science Fund (grant no. 20TQA001), Humanities and Social Science Research Fund of the Ministry of Education in China (grant no. 18YJC840045), and Postgraduate Research & Practice Innovation Program of Jiangsu Province (grant no. KYCX20_0347). The authors are grateful to all the anonymous reviewers for their precious comments and suggestions.

References

Abd-Alrazaq, A., Alhuwail, D., Househ, M., Hamdi, M., & Shah, Z. (2020). Top concerns of Tweeters during the COVID-19 pandemic: Infoveillance study. *Journal of Medical Internet Research, 22*(4), e19016.

Alkhodair, S. A., Fung, B. C. M., Rahman, O., & Hung, P. C. K. (2018). Improving interpretations of topic modeling in microblogs. *Journal of the Association for Information Science and Technology, 69*(4), 528–540. doi:10.1002/asi

Asghari, M., Sierra-Sosa, D., & Elmaghraby, A. S. (2020). A topic modeling framework for spatio-temporal information management. *Information Processing & Management*. doi:10.1016/j.ipm.2020.102340

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research, 3*(1), 993–1022.

Carvalho, L. F., Pianowski, G., & Goncalves, A. P. (2020). Personality differences and COVID-19: are extroversion and conscientiousness personality traits associated with engagement with containment measures? *Trends in Psychiatry and Psychotherapy, 42*(2), 179–184. doi:10.1590/2237-6089-2020-0029

Chaithra, V. D. (2019). Hybrid approach: Naive bayes and sentiment VADER for analyzing sentiment of mobile unboxing video comments. *Iranian Journal of Electrical and Computer Engineering, 9*(5), 4452–4459.

Chen, E., Lerman, K., & Ferrara, E. (2020). Tracking Social Media Discourse About the COVID-19 Pandemic: Development of a Public Coronavirus Twitter Data Set. *JMIR Public Health and Surveillance, 6*(2), e19273. doi:10.2196/19273

Duong, V., Luo, J., Pham, P., Yang, T., & Wang, Y. (2020). *The ivory tower lost: How college students respond differently than the general public to the COVID-19 pandemic*. Retrieved from https://arxiv.org/pdf/2004.09968.pdf

Esuli, A., & Sebastiani, F. (2006, May). *SENTIWORDNET: A publicly available lexical resource for opinion mining*. Paper presented at the 5th International Conference on Language Resources and Evaluation (LREC'06), Genoa, Italy. Retrieved from http://www.lrec-conf.org/proceedings/lrec2006/pdf/384_pdf.pdf

Giannetti, F. (2018). A twitter case study for assessing digital sound. *Journal of the Association for Information Science and Technology, 69*(5), 687–699. doi:10.1002/asi.23990

Gong, K., Xu, Z., Cai, Z., Chen, Y., & Wang, Z. (2020). Internet Hospitals Help Prevent and Control the Epidemic of COVID-19 in China: Multicenter User Profiling Study. *Journal of Medical Internet Research, 22*(4), e18908. doi:10.2196/18908

Guo, L., Vargo, C. J., Pan, Z., Ding, W., & Ishwar, P. (2016). Big social data analytics in journalism and mass communication. *Journalism & Mass Communication Quarterly, 93*(2), 332–359. doi:10.1177/1077699016639231

Hutto, C. J., & Gilbert, E. (2014). *VADER:A parsimonious rule-based model for sentiment analysis of social media text*. Paper presented at the Eighth International Conference on Weblogs and Social Media (ICWSM-14), Ann Arbor, MI.

Jordan, S., Hovet, S., Fung, I., Liang, H., Fu, K.-W., & Tse, Z. (2018). Using Twitter for public health surveillance from monitoring and prediction to public response. *Data, 4*(1), 1–20. doi:10.3390/data4010006

Kern, M. L., McCarthy, P. X., Chakrabarty, D., & Rizoiu, M. A. (2019). Social media-predicted personality traits and values can help match people to their ideal jobs. *Proceedings of the National Academy of Sciencecs, 116*(52), 26459–26464. doi:10.1073/pnas.1917942116

Kim, E. H.-J., Jeong, Y. K., Kim, Y., Kang, K. Y., & Song, M. (2016). Topic-based content and sentiment analysis of Ebola virus on Twitter and in the news. *Journal of Information Science, 42*(6), 763–781. doi:10.1177/0165551515608733

Kim, S. -M., & Hovy, E. (2014, August). *Determining the sentiment of opinions*. Paper presented at the 20th International Conference on Computational Linguistics, Geneva, Switzerland. doi:10.3115/1220355.1220555

Pandarachalil, R., Sendhilkumar, S., & Mahalakshmi, G. S. (2014). Twitter sentiment analysis for large-scale data: An unsupervised approach. *Cognitive Computation, 7*(2), 254–262. doi:10.1007/s12559-014-9310-z

Ren, Y., Wang, R., & Ji, D. (2016). A topic-enhanced word embedding for Twitter sentiment classification. *Information Sciences, 369*, 188–198. doi: 10.1016/j.ins.2016.06.040

Röder, M., Both, A., & Hinneburg, A. (2015). Exploring the space of topic coherence measures. *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining - WSDM' 15*, 399–408. doi: 10.1145/2684822.2685324

Rufai, S. R., & Bunce, C. (2020). World leaders' usage of Twitter in response to the COVID-19 pandemic: A content analysis. *Journal of Public Health (Oxford, England), 42*(3), 510–516. doi:10.1093/pubmed/fdaa049

Salton, G., & Yu, C. T. (1975). On the construction of effective vocabularies for information retrieval. *ACM SIGPLAN Notices, 10*(1), 48–60.

Sasaki, K., Yoshikawa, T., & Furuhashi, T. (2014). Online topic model for Twitter considering dynamics of user interests and topic trends. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1977–1985. doi: 10.3115/v1/D14-1212

Shah, Z., & Dunn, A. G. (2019). Event detection on Twitter by mapping unexpected changes in streaming data into a spatiotemporal lattice. *IEEE Transactions on Big Data*, 1–1. doi:10.1109/TBDATA.2019.2948594

Sloan, L., Morgan, J., Burnap, P., & Williams, M. (2015). Who tweets? Deriving the demographic characteristics of age, occupation and social class from Twitter user meta-data. *PLoS One, 10*(3), e0115545. doi:10.1371/journal.pone.0115545

Suneson, G. (2019). What are the 25 lowest paying jobs in the US? Women usually hold them. *Wall Street*. Retrieved from https://www.usatoday.com/story/money/2019/04/04/25-lowest-paying-jobs-in-us-2019-includes-cooking-cleaning/39264277/

Tang, D., Wei, F., Yang, N., Zhou, M., Liu, T., & Qin, B. (2014). Learning sentiment-specific word embedding for Twitter sentiment classification. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, 1555–1565. doi: 10.3115/v1/P14-1146

Thelwall, M., & Thelwall, S. (2020). Covid-19 tweeting in English: Gender differences. *El Profesional de la Información, 29*(3). doi:10.3145/epi.2020.may.01

Thelwall, M., & Vis, F. (2017). Gender and image sharing on Facebook, Twitter, Instagram, Snapchat and WhatsApp in the UK. *Aslib Journal of Information Management, 69*(6), 702–720. doi:10.1108/AJIM-04-2017-0098

Vegt, I. v. d., & Kleinberg, B. (2020). Women worry about family, men about the economy: Gender differences in emotional responses to COVID-19. *Proceedings of the 12th International Conference on Social Informatics*, 397–409. doi: 10.1007/978-3-030-60975-7_29

Yan, X., Guo, J., Lan, Y., & Cheng, X. (2013). A biterm topic model for short texts. *Proceedings of the 22nd International Conference on World Wide Web*, 1445–1456. doi: 10.1145/2488388.2488514

Zajenkowski, M., Jonason, P. K., Leniarska, M., & Kozakiewicz, Z. (2020). Who complies with the restrictions to reduce the spread of COVID-19? Personality and perceptions of the COVID-19 situation. *Personality and Individual Differences, 166*, 110199, 1–6. doi:10.1016/j.paid.2020.110199