A Survey of Sentimental Analysis Methods on COVID-19 Research

Areeba Umair[1], Elio Masciari[1,2]

[1]Department of Electrical Engineering and Information Technologies, University of Naples, Federico II
[2]Institute for High Performance Computing and Networking (ICAR), National Research Council, Naples, Italy

Abstract

In this era of social media, people share anything they feel or experience on social media in the form of posts or comments. These posts, comments or reviews of the people can be analyzed using sentimental analysis, which is emerging field in text mining. COVID-19 has people's life all over the globe and thus has declared as pandemic. Due to COVID, people are feeling panic, anxiety, rage, sorrow, misery, stress and other issues. In this review, we have presented the sentimental analysis data sources, approaches, scenarios, methods and tools by comparing thirty studies. The results illustrated that most researchers have used SVM and Naive Bayes for sentimental analysis on COVID research. We also concluded that most of the researchers work on the sentiments of students, reopening sentiments, vaccine sentiments etc.

Keywords

Social media Big Data, Sentiments related to COVID, Social Media Reviews, Data analytic, COVID-19

1. Introduction

Now-a-days, many people use social networks to express their opinion, thoughts or feedback about anything, [1]. In this era of technology, almost all the industries provide their customers with the ability to buy product online and also share their reviews or feedback on their website of social media pages [2]. This feedback can be positive or negative which can help other customer in making decision and help the industry to improve the product according to the customer need [3]. Such kind of review data on internet can be used in extraction of sentiments from the raw data that can be used for well-being of the society as well as business or organization [4], [5]. Sentimental analysis is the natural language processing tasks in which text is classified into positive, negative or neutral sentiments based on their meanings in the sentence. There are three types of sentimental analysis i.e. document level, sentence level and aspect level sentimental analysis. In order to gain the fine grain sentimental expression, aspect level sentimental analysis is used [1]. Let's take an example to understand aspect level sentimental analysis. "The food is very tasty but its quality is low". In this example, "very tasty and "low" show two different sentiments i.e. positive and negative respectively. The traditional sentimental analysis methods have been eliminated due to advancement in artificial intelligence [3].

SEBD 2022: The 30th Italian Symposium on Advanced Database Systems, June 19-22, 2022, Tirrenia (PI), Italy
areeba.umair@unina.it (A. Umair); elio.masciari@unina.it (E. Masciari)
https://www.docenti.unina.it/elio.masciari (E. Masciari)

The whole world is facing the biggest challenge in the form of COVID, which has destroyed the economy of many under-developed countries [6]. Corona-virus was discovered in Wuhan, China in the month of December 2019 and it has started spreading in the world and thus declared as pandemic. According to John Hopkins University, $435,427,191$ people have been affected due to COVID, thus causing $5,966,417$ number of deaths till 27 February 2022. People are facing different psychological problems due to COVID such as anger, depression, fear, and many others.

The traditional machine learning methods and deep learning methods are available to resolve the sentimental classification problems [7], [8]. The traditional ML (machine learning) classifiers for sentimental classification are Support Vector Machine (SVM) and Naive Bayes however, deep learning methods for sentimental classification are Recurrent Neural Network (RNN) and Convolutional Neural Network (CNN). These methods extract meaningful features automatically [1]. RNN has a recurrent nature due to which it suffers from gradient vanishing problem and CNN has short-comings for sequential dependencies. Thus, the literature shows the different issues and limitations in the exiting approaches such as low accuracy and performance and high complexity [3]. The inconsistent sentimental polarity in the sentence causes the word dependency to be weaken. In such scenario, attention mechanism can be fruitful for sentimental classification tasks.

In this research, we have collected thirty primary studies related to sentimental analysis with respect to COVID-19 and performed the survey. The purpose of the survey was to identify the main data sources which are providing COVID-19 related data and the widely used applications that have been applied on such data. This survey also identifies the applications or topics on which research is being processed with respect to COVID-19 sentimental analysis. At the end the future implications of COVID-research have also been presented in this survey.

2. Methodology

Thirty primary studies were selected for the comparison and their review was performed. Table 1 has six columns, first columns shows the references while in second column, data sources used in COVID-19 research have been mentioned. The purpose of mentioning the data sources or data sets is to assist the new researchers in collecting the similar datasets for their research. We have also mentioned the volume of the dataset used in the primary studies in column 3. The methods and approaches frequently used for sentimental analysis of COVID-19 have been specified in column 4. The column 5 illustrated the application scenarios for COVID sentimental analysis research, it can give new directions for the future research. The future research directions have been shown in column 6.

2.1. COVID-19 Data-sets

During the COVID pandemic, many people experience different mental issues which caused their emotions to change. The people used social media to express these emotions. Therefore, the social media provide huge amount of data to understand the peoples feelings and reactions to the situations they faced during pandemic. The data sources for COVID-19 research have been shown in Table 1. It illustrates that the main data source during COVID was twitter.

Twenty four out of thirty primary studies used twitter as a source in their research. However, the remaining data sources were WeChat account, Yelp, Reddit, and other Media forums.

Twitter: Twitter has been used worldwide for sharing the thoughts and opinions. It is most popular app having 81.47 million users [9]. People post their feelings in the form of "tweets". According to research in [10], almost 200 billion tweets are published in one year.

2.2. Sentiments classification methods

With the increase of social media platforms and social media data, more powerful analytical tools should be developed. Different approaches were adopted for COVID-19 research to perform the sentimental analysis. They can be divided into three different types i.e. machine learning, lexicon based and hybrid.

2.2.1. ML and Deep Learning (DL) Methods:

The ML methods which can be used for sentimental analysis are supervised learning approaches and unsupervised approaches.

The supervised learning works on the labelled data . Different researchers used different sentimental analysis methods on COVID data as seen in Table 1. In and [11], Naive Bayes algorithm was used as a supervised learning method for sentimental classification. Naive Bayes uses the Bayesian theorem given in equation 1.

$$P(H|X) = P(X|H)P(H)/P(X) \tag{1}$$

Support vector machine works by finding the hyper-plane in the whole data by creating high dimensional feature from he feature space. [12], [13] and [14] used SVM in their research for sentimental analysis. Decision tree found different decision rules from the entire dataset and used them to train its model. Random forest also chooses random features and instance from the entire dataset. It has been used by [15], [13] and [14]. Many other researchers used other sentimental analysis techniques such as KNN Linear Regression [16], Logistic Regression [11], [14], LSTM [17], [13], RNN [18] and BERT model [10], [9] etc.

The unsupervised learning ML methods uses unlabeled data. There are different methods that have been applied on sentimental analysis during COVID. The researchers used K-means clustering in . However, many other studies such as [19], [20], [21], [14], [22], [23] have used Latent Dirichlet Allocation (LDA).

2.3. Application scenarios on COVID-19 data

COVID has effected people's life and thus they are facing different psychological issues. Many researchers pursued their research to analysis the people's sentiments during COVID-19.

2.3.1. Mental health analysis of students during the lockdown

To control the spread of COVID, social distancing was applied which reduced the human-to-human interactions. Many countries imposed lockdown, and closed the airspace, educational and other institutes. Due to lockdown, people specially students had to stay far away from their homes, stuck in their hostels, and had to quit their educational activities, which causes anxiety and stress in students. Students express their sentiments using social platforms and researchers tried to explore their sentiments [20], [24] and students [25]. In [18], [26], [13], [11], [27], [28], [21], [22], [19], [29], [30], twitter data was used to understand the people's sentiments during lockdown.

2.3.2. Reopening after COVID-19:

Coronavirus has effected the lives of billion of peoples directly or indirectly. It has caused economical crisis all over the world which is a hurdle towards reopening [31]. The long-term closing of economy is a threat for any country to survive. Due to these reasons, people are forcing to reopen the businesses and going to normal life [32]. Hence, in [32] and [31], the researchers put their efforts on the discovering what are people thinking about re-opening after COVID-19.

2.3.3. Restaurant reviews

In today's digital era, the customers can share their opinion and feedback about quality of product or services they use from different organizations. These reviews help other customers to make decisions when they are about to use the service and product. The online reviews are associated with the star rating which effect the revenue of the restaurant. During COVID, special SOPs were announced for the restaurants and people were very concerned about the COVID-spread. Therefore, many restaurants got negative reviews for cold outdoor area and slow service. Researchers analyzed the people' feedback about restaurants which helped the management of restaurants to maintain a quality food and ambience [15].

2.3.4. Vaccine sentiments and racial sentiments

The development of COVID vaccine can be useful to control the spread of COVID. Therefore, it many industries put their efforts and develop different kind of vaccines. But, to control the COVID with vaccines, the acceptance and receiving of vaccines is the main requirement [33]. If people are not willing to get themselves, it will be a clear hurdle in the control of COVID [34]. Researchers analyzed the public sentiments about vaccines in [9]. COVID also caused the feelings of discrimination across the boarders and therefore people became more racists [12].

3. Comparison of Studies

Table 1 shows the summary of the comparison of the primary studies used in this survey.

Table 1: Comparison of methods and scenarios used for analyzing people's sentiments related to COVID-19

Ref	Data Source/set	Volume of Data	Methods	Application	Future Direction
[15]	Yelp	112,412	GBDT, RF, LSTM, SWEM	Analyze restaurant reviews	Restaurant locations.
[28]	Tweets	20,325,929 tweets	CrystalFeel	Trends of fear, anger, sadness, and joy	Include other media platforms.
[35]	Tweets	500,000 tweets	TextBlob	Finding tweets polarity	Explore other social media
[9]	Tweets	57.5M	BERT	Vaccine sentiments	Real-time social media monitoring
[18]	Tweets	N/A	NLP, RNN	Analyze sentiments	Visualization, clustering and classification
[12]	Tweets	3,377,295	SVM	Racial sentiment changes	Temporal changes in racial attitudes
[36]	Tweets	840,000	TextBlob, LDA,	identification of Anxiety, stress, and trauma	Perception changes for different biographies
[27]	Tweets	57 454	NLP	Analyse the characteristics of polish COVID-19	N/A
[26]	Tweets	370	WordCloud	Sentimental analysis	N/A
[32]	Tweets	293,597	Binary logit model	Reopening sentiment	Socioeconomic and household information
[13]	Tweets	7528	TextBlob, CNN-LSTM, RF, SVC, ETC, DT,	Perform sentiment analysis	Use deep learning approaches
[11]	Tweets	900000	NB, LR,	Public sentiment associated with the progress of Coronavirus	Include news articles and personal communications data.
[31]	Tweets	293,597	N-gram, R packages Syuzhet and sentimentr	Reopen Sentiments	Replicate on other social media data
[29]	Tweets	16 million	TClustVID	Investigate Topics and Sentiment	Explore other data repositories.
[21]	RateMDs	55,612 PORs	TF–IDF, LDA	Patients views	Trend in high death and recovery rate
[22]	Tweets	4 million	LDA, NLP	COVID-19–related sentiments	Explore public trust
[30]	Tweets	13 million	Dynamic Topic Models	Detecting Topic	More specific topics
[23]	Qingbo	N/A	LDA	Emotional change	Precise location information

4. Conclusion

Twitter based sentimental classification is a new paradigm in the social media research. A review of almost thirty primary studies was performed in our research. The comparison of data sources used, volume of data used, approaches, and application scenarios with respect to COVID-19 was established. This survey presents its contribution in the field of sentimental analysis and open doors for the new researchers. This survey paper shows that twitter is the most popular data source for sentiments analysis and Naive Bayes and SVM are the algorithms which researchers used for sentimental analysis during COVID. During COVID-19, various researchers worked on the different dimensions such as mental health of students, reopening sentiments, restaurants reviews and vaccine sentiments. Thus, the use of advanced methods of machine learning and deep learning along with the social media data can explore more interesting topics in future.

References

[1] Z. Zhou, F. Liu, Q. Wang, R-Transformer network based on position and self-attention mechanism for aspect-level sentiment classification, IEEE Access 7 (2019) 127754–127764. doi:10.1109/ACCESS.2019.2938854.

[2] M. Ceci, R. Corizzo, F. Fumarola, M. Ianni, D. Malerba, G. Maria, E. Masciari, M. Oliverio, A. Rashkovska, Big data techniques for supporting accurate predictions of energy production from renewable sources, volume 0, 2015, p. 62 – 71. doi:10.1145/2790755.2790762, cited by: 15.

[3] X. Wang, Y. Tong, Application of an emotional classification model in e-commerce text based on an improved transformer model, PLoS One 16 (2021) 1–16. URL: http://dx.doi.org/10.1371/journal.pone.0247984. doi:10.1371/journal.pone.0247984.

[4] A. Umair, E. Masciari, Sentimental and spatial analysis of covid-19 vaccines tweets, Journal of Intelligent Information Systems (2022) 1–21.

[5] B. Fazzinga, S. Flesca, F. Furfaro, E. Masciari, Rfid-data compression for supporting aggregate queries, ACM Transactions on Database Systems 38 (2013) 1 – 45. doi:10.1145/2487259.2487263, cited by: 11.

[6] M. Ceci, R. Corizzo, F. Fumarola, M. Ianni, D. Malerba, G. Maria, E. Masciari, M. Oliverio, A. Rashkovska, Big data techniques for supporting accurate predictions of energy production from renewable sources, in: B. C. Desai, M. Toyama (Eds.), Proceedings of the 19th International Database Engineering & Applications Symposium, Yokohama, Japan, July 13-15, 2015, ACM, 2015, pp. 62–71. URL: https://doi.org/10.1145/2790755.2790762. doi:10.1145/2790755.2790762.

[7] S. Flesca, E. Masciari, Efficient and effective web change detection, Data Knowl. Eng. 46 (2003) 203–224. URL: https://doi.org/10.1016/S0169-023X(02)00210-0. doi:10.1016/S0169-023X(02)00210-0.

[8] S. Flesca, F. Furfaro, E. Masciari, On the minimization of xpath queries, J. ACM 55 (2008) 2:1–2:46. URL: https://doi.org/10.1145/1326554.1326556. doi:10.1145/1326554.1326556.

[9] M. Müller, M. Salathé, Addressing machine learning concept drift reveals declining vaccine sentiment during the COVID-19 pandemic, arXiv (2020) 1–12. arXiv:2012.02197.

[10] N. Chintalapudi, G. Battineni, F. Amenta, Sentimental Analysis of COVID-19 Tweets Using Deep Learning Models, Infect. Dis. Rep. 13 (2021) 329–339. doi:10.3390/idr13020032.
[11] J. Samuel, M. M. Rahman, G. G. N. Ali, Y. Samuel, A. Pelaez, P. H. J. Chong, M. Yakubov, Feeling Positive about Reopening? New Normal Scenarios from COVID-19 US Reopen Sentiment Analytics, IEEE Access 8 (2020) 142173–142190. doi:10.1109/ACCESS.2020.3013933.
[12] T. T. Nguyen, S. Criss, P. Dwivedi, D. Huang, J. Keralis, E. Hsu, L. Phan, L. H. Nguyen, I. Yardi, M. M. Glymour, A. M. Allen, D. H. Chae, G. C. Gee, Q. C. Nguyen, Exploring U.S. shifts in anti-Asian sentiment with the emergence of COVID-19, Int. J. Environ. Res. Public Health 17 (2020) 1–13. doi:10.3390/ijerph17197032.
[13] F. Rustam, M. Khalid, W. Aslam, V. Rupapara, A. Mehmood, G. S. Choi, A performance comparison of supervised machine learning models for Covid-19 tweets sentiment analysis, PLoS One 16 (2021) 1–23. URL: http://dx.doi.org/10.1371/journal.pone.0245909. doi:10.1371/journal.pone.0245909.
[14] X. Xiang, X. Lu, A. Halavanau, J. Xue, Y. Sun, P. H. L. Lai, Z. Wu, Modern Senicide in the Face of a Pandemic: An Examination of Public Discourse and Sentiment About Older Adults and COVID-19 Using Machine Learning, J. Gerontol. B. Psychol. Sci. Soc. Sci. 76 (2021) e190–e200. doi:10.1093/geronb/gbaa128.
[15] Y. Luo, X. Xu, Comparative study of deep learning models for analyzing online restaurant reviews in the era of the COVID-19 pandemic, Int. J. Hosp. Manag. 94 (2021) 102849. URL: https://doi.org/10.1016/j.ijhm.2020.102849. doi:10.1016/j.ijhm.2020.102849.
[16] H. Adamu, S. L. Lutfi, N. H. A. H. Malim, R. Hassan, A. Di Vaio, A. S. A. Mohamed, Framing twitter public sentiment on Nigerian government COVID-19 palliatives distribution using machine learning, Sustain. 13 (2021). doi:10.3390/su13063497.
[17] H. Jelodar, Y. Wang, R. Orji, H. Huang, Deep sentiment classification and topic discovery on novel coronavirus or COVID-19 online discussions: NLP using LSTM recurrent neural network approach, arXiv 24 (2020) 2733–2742.
[18] L. Nemes, A. Kiss, Social media sentiment analysis based on COVID-19, J. Inf. Telecommun. 5 (2021) 1–15. URL: https://doi.org/10.1080/24751839.2020.1790793. doi:10.1080/24751839.2020.1790793.
[19] M. Hung, E. Lauren, E. S. Hon, W. C. Birmingham, J. Xu, S. Su, S. D. Hon, J. Park, P. Dang, M. S. Lipsky, Social network analysis of COVID-19 sentiments: Application of artificial intelligence, J. Med. Internet Res. 22 (2020) 1–13. doi:10.2196/22590.
[20] S. V. Praveen, R. Ittamalla, G. Deepak, Analyzing Indian general public's perspective on anxiety, stress and trauma during Covid-19 - A machine learning study of 840,000 tweets, Diabetes Metab. Syndr. Clin. Res. Rev. 15 (2021) 667–671. URL: https://doi.org/10.1016/j.dsx.2021.03.016. doi:10.1016/j.dsx.2021.03.016.
[21] A. M. Shah, X. Yan, A. Qayyum, R. A. Naqvi, S. J. Shah, Mining topic and sentiment dynamics in physician rating websites during the early wave of the COVID-19 pandemic: Machine learning approach, Int. J. Med. Inform. 149 (2021). doi:10.1016/j.ijmedinf.2021.104434.
[22] J. Xue, J. Chen, C. Chen, C. Zheng, S. Li, T. Zhu, Public discourse and sentiment during the COVID 19 pandemic: Using latent dirichlet allocation for topic modeling on twitter, PLoS One 15 (2020) 1–12. URL: http://dx.doi.org/10.1371/journal.pone.0239441. doi:10.1371/

journal.pone.0239441. arXiv:2005.08817.
[23] B. Zhu, X. Zheng, H. Liu, J. Li, P. Wang, Analysis of spatiotemporal characteristics of big data on social media sentiment with COVID-19 epidemic topics, Chaos, Solitons and Fractals 140 (2020) 110123. URL: https://doi.org/10.1016/j.chaos.2020.110123. doi:10.1016/j.chaos.2020.110123.
[24] V. Ajantha Devi, A. Nayyar, Evaluation of Geotagging Twitter Data Using Sentiment Analysis During COVID-19, volume 166, Springer Singapore, 2021. URL: http://dx.doi.org/10.1007/978-981-15-9689-6{_}65. doi:10.1007/978-981-15-9689-6_65.
[25] A. Agarwal, B. Agarwal, P. Harjule, A. Agarwal, Mental Health Analysis of Students in Major Cities of India During COVID-19, Springer Singapore, 2021. URL: http://dx.doi.org/10.1007/978-981-33-4236-1{_}4. doi:10.1007/978-981-33-4236-1_4.
[26] S. Raheja, A. Asthana, Sentimental analysis of twitter comments on COVID-19, Proc. Conflu. 2021 11th Int. Conf. Cloud Comput. Data Sci. Eng. (2021) 704–708. doi:10.1109/Confluence51648.2021.9377048.
[27] E. Probierz, A. Gałuszka, T. Dzida, Twitter text data from #Covid-19: Analysis of changes in time using exploratory sentiment analysis, J. Phys. Conf. Ser. 1828 (2021). doi:10.1088/1742-6596/1828/1/012138.
[28] M. O. Lwin, J. Lu, A. Sheldenkar, P. J. Schulz, W. Shin, R. Gupta, Y. Yang, Global sentiments surrounding the COVID-19 pandemic on Twitter: Analysis of Twitter trends, JMIR Public Heal. Surveill. 6 (2020) 1–4. doi:10.2196/19447.
[29] M. S. Satu, M. I. Khan, M. Mahmud, S. Uddin, M. A. Summers, J. M. Quinn, M. A. Moni, TClustVID: A novel machine learning classification model to investigate topics and sentiment in COVID-19 tweets, medRxiv (2020). doi:10.1101/2020.08.04.20167973.
[30] H. Yin, S. Yang, J. Li, Detecting Topic and Sentiment Dynamics Due to COVID-19 Pandemic Using Social Media, volume 12447 LNAI, Springer International Publishing, 2020. doi:10.1007/978-3-030-65390-3_46. arXiv:2007.02304.
[31] J. Samuel, G. G. N. Ali, M. M. Rahman, E. Esawi, Y. Samuel, COVID-19 public sentiment insights and machine learning for tweets classification, Inf. 11 (2020) 1–22. doi:10.3390/info11060314. arXiv:2005.10898.
[32] M. Mokhlesur Rahman, G. G. Nawaz Ali, X. J. Li, K. C. Paul, P. H. Chong, Twitter and Census Data Analytics to Explore Socioeconomic Factors for Post-COVID-19 Reopening Sentiment, arXiv (2020). doi:10.2139/ssrn.3639551.
[33] H. Seale, A. E. Heywood, J. Leask, M. Sheel, D. N. Durrheim, K. Bolsewicz, R. Kaur, Examining Australian public perceptions and behaviors towards a future COVID-19 vaccine, medRxiv (2020) 1–9. doi:10.1101/2020.09.29.20204396.
[34] M. S. Green, R. Abdullah, S. Vered, D. Nitzan, A study of ethnic, gender and educational differences in attitudes toward COVID-19 vaccines in Israel – implications for vaccination implementation policies, Isr. J. Health Policy Res. 10 (2021) 1–12. doi:10.1186/s13584-021-00458-w.
[35] K. H. Manguri, R. N. Ramadhan, P. R. Mohammed Amin, Twitter Sentiment Analysis on Worldwide COVID-19 Outbreaks, Kurdistan J. Appl. Res. (2020) 54–65. doi:10.24017/covid.8.
[36] M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, L. Zettlemoyer, Improving Language Understanding by, OpenAI (2018) 1–10.