

Data Mining and Big Data in Social Media Public Opinion Monitoring

Dandan Zhang^{1,a,*}

¹*Department of Public Opinion Imaging and Advertising, Donmyung University, Busan, Busan 612-022, Korea*

^a*dd420138027@gmail.com*

^{*}*Corresponding author*

Keywords: Data Mining, Big Data Analysis, Social Media Public Opinion, Real-Time Monitoring

Abstract: With the popularization of the Internet, social media has become one of the most common communication methods, and a large amount of information is generated on social media at every moment. These pieces of information contain a large amount of public opinion data, making social media an important source of public opinion. In the application of social media public opinion monitoring, utilizing Big Data (BD) technology to process and mine massive data from social media has become an important means of public opinion monitoring. At present, there are mature technologies and tools in this field, but further research is needed for multilingual and multicultural Data Mining (DM), as well as BD analysis and linkage prediction of micro and macro public opinion. This article uses DM and BD analysis to monitor public opinion on social media, and verifies it by conducting keyword mining on corresponding web pages in the medical, transportation, education, public security, and military industries. It can be concluded that the number of keywords captured by DM and BD analysis methods in healthcare, transportation, education, public security, and military is 980, 830, 789, 445, and 657, respectively. The number of keywords captured by traditional methods is 670, 545, 489, 245, and 557. From this, it can be seen that through DM and BD analysis techniques, valuable public opinion information can be extracted from social media data, achieving the goal of real-time tracking of social public opinion and predicting the development of micro and macro public opinion.

1. Introduction

With the development of network technology, social media has become an integral part of people's lives and work. The large amount of data generated from this can provide guidance for people's lives and work after processing and analysis. By analyzing the BD environment, it can be seen that DM technology and BD analysis technology are facing challenges, while also bringing opportunities for social media public opinion monitoring. Text DM and BD analysis in social media public opinion monitoring have become an important research direction. It can explore the value and impact of applying BD analysis and DM techniques to social media public opinion monitoring.

In order to effectively monitor and analyze public opinion on social media, this article adopts DM and BD analysis methods for public opinion monitoring on social media. BD analysis can statistically and analyze large-scale real-time data on social media, achieving real-time tracking of social public opinion. DM can partition texts and identify themes or hotspots of public opinion through key information such as keywords and association relationships. It can be deeply mined to obtain trends in predicting the development of public opinion.

2. Related Work

In recent years, with the development of network technology, the popularization of intelligent mobile terminals and social media, social media has become one of the most common communication methods, and a large amount of information is generated on social media every moment. Therefore, social media has become an important source of public opinion, and more and more research related to BD analysis, DM, and social media public opinion monitoring has emerged [1-2]. Based on Sina Weibo data released by tourists from 2017 to 2019, Feng Zeqi used the BERT (Bidirectional Encoder Representations from Transformers) model for text analysis of Weibo data. He explored the spatiotemporal distribution patterns of tourist emotions and the emotional characteristics of tourists under different themes. He found that the emotions of tourists on Weibo showed rhythmic changes, and there were differences in emotional response intensity and rhythm between tourists of different genders, and they were prone to strong emotions towards the themes of "weather" and "dining". He proposed to explore the emotional characteristics of tourists from multiple dimensions and levels, providing reference for the monitoring and early warning system of public opinion in tourist destinations [3]. SHI Feng proposed an intelligent DM analysis algorithm for identifying social media opinion manipulators and their manipulation strategies outside the domain, and conducted empirical analysis. He analyzed public opinion theme planning through social media insight systems and intelligent social media DM algorithms, and identified organizations behind the scenes that manipulate negative public opinion. The current relatively effective response measure is to use technological means for real-time monitoring and early warning [4]. Chen Qinghua used the practical concept of BD analysis in the teaching evaluation system as a starting point to explain the drawbacks of traditional teaching evaluation systems. He demonstrated the practical advantages of BD analysis in the teaching evaluation system and further explored the practical solutions of BD analysis in the teaching evaluation system [5].

However, traditional social media public opinion monitoring has some shortcomings in its application, such as limited data processing capabilities, slow processing speed, insufficient breadth and depth of monitoring scope, and lack of real-time updates. Therefore, this article adopts BD analysis and DM methods to monitor public opinion on social media. It can not only effectively monitor and analyze public opinion on social media, identify public opinion hotspots, and analyze and predict the direction of public opinion. This also greatly improves the efficiency and accuracy of public opinion monitoring [6-7].

3. Social Media Public Opinion Monitoring

3.1 Importance

Social media public opinion monitoring can help companies timely understand consumer evaluations and feedback, as well as deepen their understanding of consumer needs and preferences [8-9]. Analyzing user comments on social media can help them better grasp market trends and consumer psychology [10-11]. Social media public opinion monitoring can help enterprises detect potential crises in a timely manner, prevent the spread of negative public opinion, and quickly

implement corresponding measures by monitoring public opinion on social media in real time, thereby effectively controlling crisis situations [12-13]. By analyzing public opinion on social media, the government can better formulate policies and measures, and enhance the public's trust in the government [14-15]. At the same time, social media public opinion monitoring can help government departments conduct public opinion warnings, detect important information that affects the government's image in advance, and formulate response strategies.

3.2 Current Situation

Due to the massive, heterogeneous, and dynamic data generated by social media, traditional public opinion monitoring methods have certain limitations in screening, analyzing, and summarizing information. It mainly relies on keyword search, making it difficult to obtain and process relevant public opinion information in depth. At the same time, it is difficult to effectively identify and process implicit information such as user satire and suggestion. Therefore, this article adopts BD analysis and DM methods to monitor public opinion on social media, which can effectively monitor and analyze public opinion on social media and analyze and predict the direction of public opinion [16-17].

4. DM, BD, and Social Media Public Opinion Monitoring

4.1 DM and BD

The large amount of data generated by social media is filled with people's emotions, attitudes, and behavioral patterns. Analyzing this data can help people better understand the direction of social media public opinion. The role played by DM and BD analysis techniques cannot be ignored [18].

BD analysis provides the possibility for social media public opinion monitoring, allowing people to track and analyze various social dynamics, structured and unstructured data in real time, such as text, audio, video, etc. The application of BD analysis technology can quickly statistics and analyze large-scale real-time data, understand people's behavior and thoughts, discover the heat and direction of public opinion events, and achieve real-time tracking of social public opinion [19].

Mining data can quantify and structure inherently unquantifiable text information through techniques such as information extraction, DM, and machine learning. It is possible to identify key themes and emotional trends in information, analyze context, and even identify emotional tendencies, thereby achieving monitoring of social media public opinion. On social media platforms, the interactive relationships formed between different users constitute a vast social network. By using DM techniques, network relationships can be analyzed to identify nodes with significant influence and fast dissemination, to identify the source of public opinion and analyze the propagation path and trend of public opinion [20].

In summary, it can be concluded that BD analysis and DM techniques play a significant role in social media public opinion monitoring. It plays a greater role in grasping social public opinion, warning social risks, and optimizing social management.

4.2 Role of DM and BD in Social Media Public Opinion Monitoring

Social media public opinion monitoring includes determining analysis objects, information classification models, analysis methods, and model for analyzing data structure expression. The information in traditional social media public opinion monitoring usually does not have a type of information displayed in the form of analysis and processing, which is often intuitive and

superficial. This article uses BD analysis and DM methods to monitor public opinion on social media, which can effectively monitor and analyze public opinion on social media, and identify hot topics of public opinion. Its analysis and prediction of public opinion trends can greatly improve the efficiency and accuracy of social media public opinion monitoring. This article uses two decomposition methods based on the dimensions and scope of social media public opinion monitoring analysis, and applies DM and BD analysis to derive specific quantifiable indicators for public opinion monitoring. Specifically, assuming that A_n is the total amount of public opinion on a specific topic in n time periods, A_{n-1} is the total amount of public opinion on a specific topic in the previous time period in n time periods, and a_n is the growth rate of public opinion in n time periods, the relationship between them is shown in formula (1):

$$a_n = \frac{A_n - A_{n-1}}{A_{n-1}} \quad (1)$$

$A(X_{(i)})_n$ is used to represent the total amount of public opinion in dimension $X_{(i)}$ during the n time period, and $A(X_{(i)} - Y_{(j)})_n$ to represent the total amount of public opinion in dimension $X_{(i)}$ and dimension $Y_{(j)}$ during the n time period. The relationship between them is shown in formulas (2) and (3):

$$A_n = \sum_{i=1}^{len(x)} A(X_{(i)})_n \quad (2)$$

$$A(X_{(i)})_n = \sum_{j=1}^{len(y)} A(X_{(i)} - Y_{(j)})_n \quad (3)$$

The driving force of public opinion representing dimension $X_{(i)}$ on the total amount of public opinion is shown in formula (4):

$$a_n = \sum_{i=1}^{len(x)} L(X_{(i)})_n \quad (4)$$

Through the above formulas, data mining and big data analysis have decomposed the growth rate of the main node of public opinion into the calculation of driving forces from various dimensions of the nodes.

5. Testing of DM and BD in Social Media Public Opinion Monitoring

5.1 Accuracy Testing of DM and BD in Social Media Public Opinion Prediction

In order to test the accuracy of DM and BD analysis in predicting social media public opinion, this article compared the actual passenger flow of a month at the World Expo with the predicted passenger flow of two types of social media public opinion monitoring. The following is a comparison of the prediction and actual passenger flow of traditional media and new media based on DM and BD analysis in this article, as shown in Figure 1.

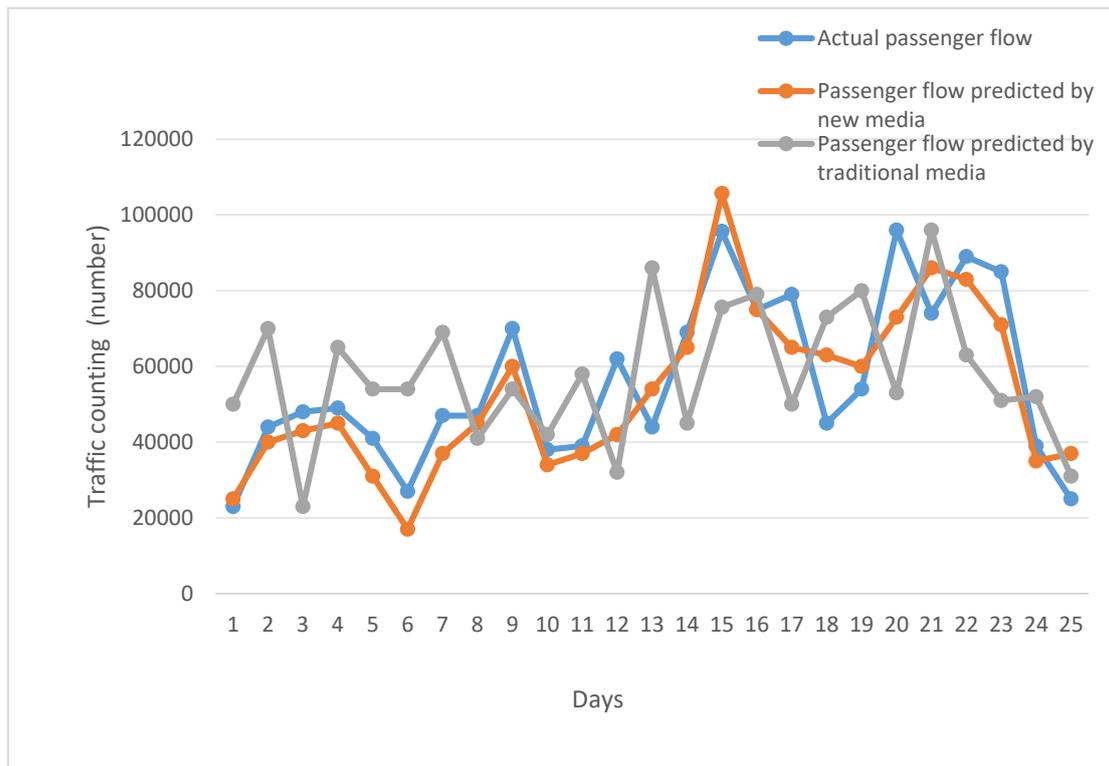


Figure 1: The actual passenger flow of the World Expo and the predicted passenger flow of social media public opinion

5.2 Results of DM and BD Analysis in Predicting the Accuracy of Social Media

From Figure 1, it can be seen that the new media prediction based on DM and BD analysis proposed in this article has better performance than traditional media prediction. This article calculates the average relative error, accuracy, and standard rate of two types of social media public opinion predictions, as shown in Table 1. From Table 1, it can be seen that the media public opinion prediction based on DM and BD analysis proposed in this article has an average relative error of 7.1% lower than traditional media public opinion prediction, and has higher accuracy and standard rate, greatly improving the prediction effect.

Table 1: Average relative error, accuracy, and standard rate

Forecast methods	Traditional media forecasts (%)	New media forecasts (%)
Mean relative error	15.6	8.5
Accuracy	84.4	91.5
Standard rate	75.4	88.3

Through experimental verification, the media public opinion prediction based on DM and BD analysis in this article has shown good results, with improved prediction accuracy compared to traditional media prediction methods. This is because the new media public opinion prediction based on DM and BD analysis proposed in this article is better than traditional media public opinion prediction. Therefore, social media public opinion monitoring based on DM and BD analysis can improve the accuracy of predictions and achieve ideal prediction results.

5.3 Testing the Mining Degree of DM and BD in Social Media Public Opinion Monitoring

In order to compare two monitoring methods of traditional media public opinion monitoring and the new media public opinion monitoring based on DM and BD analysis in this article, 400 web pages were randomly selected for medical, transportation, education, public security, and military data, and the pages of these web pages were divided and statistically analyzed. In order to verify the ability to process web data and mine keyword information, keyword mining statistics were conducted for the corresponding categories of medical, transportation, education, public security, and military industries. The comparison results are shown in Figure 2. At the same time, in order to verify the completeness of webpage data and the ability to filter and analyze information, this article also compared the recall rates of medical, transportation, education, public security, and military industries. The comparison results are shown in Figure 3.

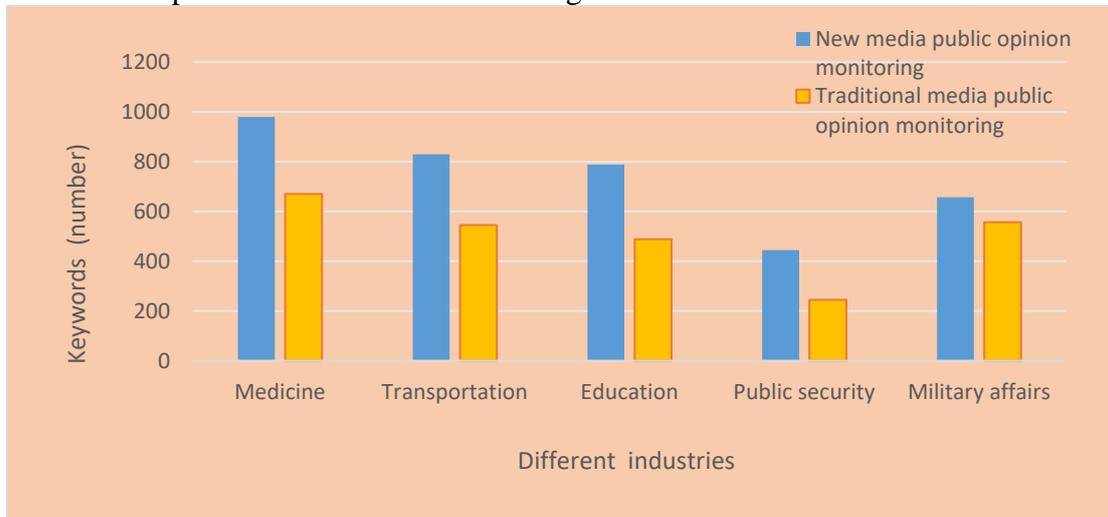


Figure 2: Media public opinion monitoring for keyword mining in different industries

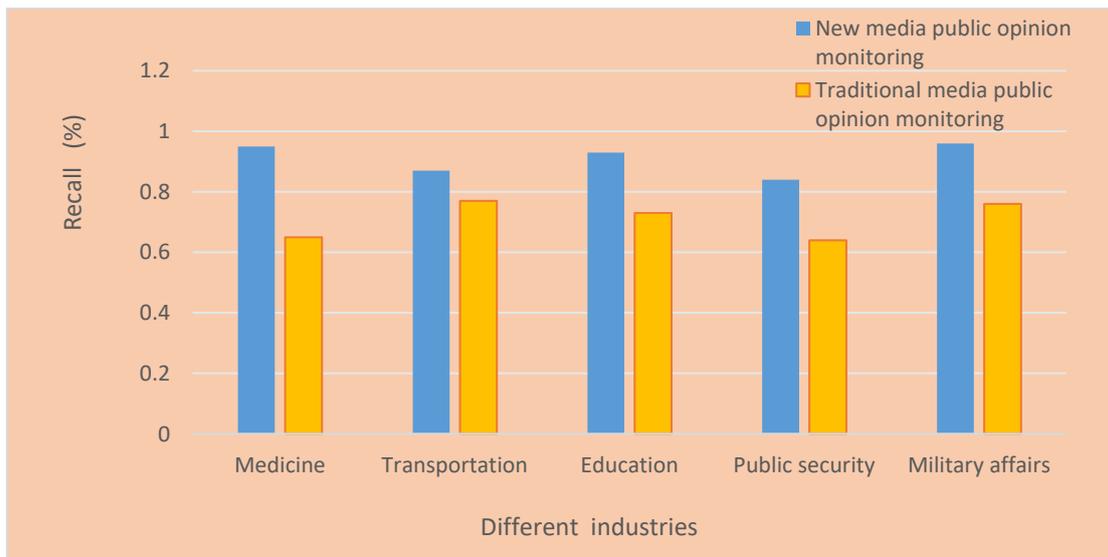


Figure 3: The recall rate of media public opinion monitoring for different industries

5.4 Results of DM and BD in Social Media Public Opinion Monitoring and Mining

From Figure 2, it can be seen that among the randomly crawled keywords on the same webpage,

social media public opinion monitoring based on DM and BD analysis has mined 980 keywords in the medical industry, while traditional media public opinion monitoring has only 670 keywords. This is the case in other different industries. From this, it can be seen that the deep mining of medical, transportation, education, public security, and military industries in social media public opinion monitoring based on DM and BD analysis is significantly better than the mining depth of traditional media public opinion monitoring. It has stronger capabilities in web data processing and mining, and can mine more keyword information.

From Figure 3, it can be seen that DM and BD analysis have significantly better recall rates in social media public opinion monitoring for medical, transportation, education, public security, and military industries than traditional media public opinion monitoring. The recall rate of social media public opinion monitoring based on DM and BD analysis on the same webpage is over 80%, with the highest recall rate of 96% for the military industry. The highest recall rate of traditional media public opinion monitoring is only 77%. From this, it can be seen that social media public opinion monitoring based on DM and BD analysis has a stronger ability to comprehensively search for web data, and more comprehensive and in-depth screening and analysis of information.

6. Conclusions

This article applies DM and BD analysis to the field of social media public opinion monitoring. By analyzing and mining social media data, it is possible to timely detect changes in public opinion, grasp the emotional tendencies of the masses, identify hot topics, and better respond to future predictions and changes in events. At the same time, it can also provide important decision support and strategic insights for enterprises and governments. With the continuous development of social media and the expansion of the number of people, the application of DM and BD analysis in social media public opinion monitoring would also become more important and widespread. At the same time, it would also continue to expand with the continuous progress of technology and application scenarios. These technologies would provide more accurate, comprehensive, and intelligent social media public opinion monitoring and analysis services for enterprises, governments, and organizations.

References

- [1] Rao A S, Krishna B V, Saravanan D, et al. Supervision calamity of public opinion actions based on field programmable gate array and machine learning. *International Journal of Nonlinear Analysis and Applications*, 2021, 12(2): 1187-1198.
- [2] Zhong Z. Internet public opinion evolution in the COVID-19 event and coping strategies. *Disaster Medicine and Public Health Preparedness*, 2021, 15(6): e27-e33
- [3] Feng Zeqi, Peng Xia, Wu Yachao. Tourists' emotional perception based on social media data mining. *Geography and Geo-Information Science*, 2022, 38(1):31-36.
- [4] Shi Feng, Li Yanze, Shao Lei. Data Mining Algorithm and Empirical Analysis of Social Media Public Opinion Manipulator Identification: A Case Study of International Large-scale Events. *Journal of Intelligence*, 2023, 42(6):147-153.
- [5] Chen Qinghua. Suggestions and practice of big data analysis in teaching evaluation system. *Journal of Yunyang Normal College*, 2020, 040(003):91-94.
- [6] Zhong Y, Chen L, Dan C, et al. A systematic survey of data mining and big data analysis in internet of things. *The Journal of Supercomputing*, 2022, 78(17): 18405-18453.
- [7] Jayasri N P, Aruna R. Big data analytics in health care by data mining and classification techniques. *ICT Express*, 2022, 8(2): 250-257.
- [8] Haoxiang W, Smys S. Big data analysis and perturbation using data mining algorithm. *Journal of Soft Computing Paradigm (JSCP)*, 2021, 3(01): 19-28.
- [9] Ageed Z S, Zeebaree S R M, Sadeeq M M, et al. Comprehensive survey of big data mining approaches in cloud systems. *Qubahan Academic Journal*, 2021, 1(2): 29-38.

- [10] Rao Z, Yuan J. *Data mining and statistics issues of precision and intelligent agriculture based on big data analysis. Acta Agriculturae Scandinavica, Section B—Soil & Plant Science*, 2021, 71(9): 870-883.
- [11] Nti I K, Quarcoo J A, Aning J, et al. *A mini-review of machine learning in big data analytics: Applications, challenges, and prospects. Big Data Mining and Analytics*, 2022, 5(2): 81-97.
- [12] Mahmud M S, Huang J Z, Salloum S, et al. *A survey of data partitioning and sampling methods to support big data analysis. Big Data Mining and Analytics*, 2020, 3(2): 85-101.
- [13] Niham D, Elle L, Yuriah A, et al. *Utilization of Big Data in Libraries by Using Data Mining. International Journal of Cyber and IT Service Management*, 2023, 3(2): 79-85.
- [14] Sufi F K, Razzak I, Khalil I. *Tracking anti-vax social movement using AI-based social media monitoring. IEEE Transactions on Technology and Society*, 2022, 3(4): 290-299.
- [15] Tran M, Draeger C, Wang X, et al. *Monitoring the well-being of vulnerable transit riders using machine learning based sentiment analysis and social media: Lessons from COVID-19. Environment and Planning B: Urban Analytics and City Science*, 2023, 50(1): 60-75.
- [16] He W, Xia D, Liu J, et al. *Research on the dynamic monitoring system model of university network public opinion under the big data environment. Mobile Networks and Applications*, 2022, 27(6): 2352-2363.
- [17] Li J, Xu X. *A study of Big Data-based employees' public opinion system construction. Journal of Industrial Integration and Management*, 2020, 5(02): 225-233.
- [18] Liu P. *Information dissemination mechanism based on cloud computing cross-media public opinion network environment. International Journal of Information Technologies and Systems Approach (IJITSA)*, 2021, 14(2): 70-83.
- [19] Xu B, Liu Y. *The role of big data in network public opinion within the colleges and universities. Soft Computing*, 2022, 26(20): 10853-10862.
- [20] Wei W, Wang L, Li X, et al. *Design and implementation of public opinion monitoring system based on cloud platform. Journal of Internet Technology*, 2021, 22(3): 569-581