# Research on the Application of Social Media Data Mining Based on Sentiment Analysis

Yanning Hu<sup>1,a,\*</sup>

<sup>1</sup>School of Information and Communication Engineering, Hainan University, Haikou, 570228, China a. 20233006303@hainanu.edu.cn \*corresponding author

Abstract: As social software becomes increasingly prevalent, social media platforms have evolved into a major avenue for individuals to express their emotions. On these platforms, users convey their emotions through text, images, and videos. Their expressions not only reveal their personal emotional state but also correlate with subsequent behavioral patterns. As an invaluable tool for understanding human psychology, it is able to identify emotional tendencies and make correlation predictions by analyzing vast quantities of user data on social media. Recently, sentiment analysis methods and applications have evolved significantly, spanning fields like marketing, public relations, and social research. However, it continues to encounter significant challenges in complex contexts and cross-cultural communication. To gain a deeper understanding of the applications of sentiment analysis in social media, the paper surveys the principal methods and domains of sentiment analysis in recent years by literature investigation, with particular attention to multidisciplinary case studies. The findings indicate that sentiment analysis techniques are effective in some areas like real-time monitoring and trend prediction, which are vital for business decision-making and planning. Meanwhile, sentiment analysis is increasingly being applied to areas such as mental health monitoring as well as social opinion analysis. Nevertheless, current methods still struggle with complex contexts and cross-cultural sentiment expressions, necessitating further research and improvement.

*Keywords:* Sentiment Analysis, Real-time Monitoring, Social Media, Business Applications, User Data.

# 1. Introduction

As a key branch of natural language processing, sentiment analysis has made significant strides in understanding and predicting human emotional expressions in recent years. The rapid development of social media platforms has generated a large amount of user data, which includes multimodal content such as text messages, images and videos, providing a rich resource for sentiment analysis. It lays foundation for fine-grained sentiment recognition and multidimensional analysis. At present, the primary technological advances in sentiment analysis are reflected in the following aspects. The optimization of pre-trained models greatly improves the performance of sentiment analysis on various datasets. Also, the integration of multimodal analysis techniques has prompted researchers to extract sentiment features from multiple data sources, enhancing the depth and accuracy of analysis. And the

<sup>@</sup> 2025 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

research on the interpretability of models has guaranteed the trustworthiness and utility of the results of sentiment analysis. Despite the fruitful results of many studies, there are still several important research gaps in the field of sentiment analysis. These gaps include support for low-resource languages, the in-depth analysis of complex emotions, research on dynamic sentiment expression, and the exploration of cross-cultural adaptation challenges. This paper aims to explore the detailed application of sentiment analysis in the area of dynamic affective expressions. To this end, relevant literature in recent years has been systematically organized, focusing on the main methods and application areas of sentiment analysis. With the aim of exploring the current status and future development direction of the application of sentiment analysis in multiple domains, the paper has mined massive data of social media platforms, such as Facebook, Twitter, Weibo, and WeChat. In addition, it seeks to assess the contribution in promoting user experience and service quality, thus highlighting the current challenges and research opportunities.

## 2. Basic Methods of Sentiment Analysis

## 2.1. Sentiment Dictionary and Rule-based Methods

Rule-based Sentiment Analysis is an approach that identifies and extracts sentiment tendencies in text using predefined rules, typically relying on linguistics, semantics and pattern recognition, with Natural Language Processing (NLP) playing a crucial role. To analyze sentiment in text effectively, the process involves text pre-processing, feature extraction, as well as model evaluation. Feature extraction primarily relies on NLP-based word vectorization and embedding techniques, such as Word2Vec, GloVe, TF-IDF, and BERT. Different domains use different sentiment lexicons such as SentiWordNet, AFINN, HowNet, and LIWC. selecting and constructing the right sentiment lexicon based on the data can significantly improve the accuracy of sentiment analysis. For instance, in 2020, Zhang et al. used a web crawler program to analyze the Chinese review data on e-commerce platforms [1]. They used Word2Vec for word vector construction, combined the TF-IDF algorithm and lexical methods to extract keywords, and built a baseline sentiment lexicon based on lexical positions and fundamental sentiment words. By using the SO-PMI algorithm to determine sentiment tendencies, better results were achieved with the HowNet sentiment dictionary. Ultimately, the sentiment analysis accuracy using the self-constructed dictionary reached 89.2% and 86.7% on datasets from the computer and mobile phone domains, respectively. In addition, Zhang et al. tackled microblog sentiment classification by creating six sentiment dictionaries, incorporating modifiers and sentiment words, and using weighted sentiment polarity classification for constituent and whole sentences. This approach enhances analysis accuracy, aiding public opinion monitoring and detecting disguised sentiment [2]. However, rule-based approaches, while customizable, become difficult to maintain and adapt as data complexity increases, requiring frequent updates to the sentiment lexicon to cope with new expressions. Combining rule-based approaches with machine learning models can more effectively improve generalization in complex contexts.

#### 2.2. Traditional Machine Learning-Based Methods

Machine learning-based sentiment analysis is a method that uses machine learning algorithms to identify and classify sentiment tendencies in text. Such type of analysis is typically used for processing large amounts of data, such as customer reviews, social media posts, product reviews, etc. Therefore it can automatically determine whether the sentiment in it is positive, negative, or neutral. Machine learning techniques include supervised, unsupervised, and semi-supervised learning methods. Common methods include Support Vector Machine(SVM), Random Forest, and Plain Bayes. Machine learning excels at processing large-scale datasets by automatically learning features, minimizing the need for manual feature engineering. Its performance improves with more data,

making it popular in business intelligence, customer service, and market analysis. Liu et al utilized the TF-IDF algorithm and SVM to distribute the weights of sentiment words and neutral words. Retain the overall information of the text while highlight the importance of sentiment words [3]. The results indicate that the accuracy of text sentiment analysis based on this method was as high as 82.1%, which is 13.9% higher than the rule-based sentiment dictionary method and 7.7% higher than the TF-IDF weighting method. Aftab et al. compared the performance of different machine learning models for sentiment analysis and found that SVMs usually perform better in terms of accuracy and efficiency [4]. It was noted that although Naive Bayes performs well in some cases, its independence assumptions may not always be valid in real-world applications. Singh et al. utilized four machine learning classifiers, Naïve Bayes, J48, BFTree, and OneR, to perform sentiment analysis using three manually annotated datasets, and averaged over 29 experimental periods to show that OneR's classification was more accurate, while Naïve Bayes showed a faster learning rate [5]. Despite the variety of machine learning algorithms, choosing diverse algorithms for different data types can better optimize model learning. However, they still rely on high-quality training data and effective feature engineering. Also, the interpretability of machine learning models is usually worse than rule-based ones. Developing efficient algorithms that minimize reliance on large labeled datasets has become a key focus in machine learning research, aiming to enhance model interpretability and ensure more transparent decision-making.

# 2.3. Deep Learning-Based Methods

Deep learning-based sentiment analysis is an approach that uses deep learning models to identify and extract sentiment tendencies in text. It can automatically learn sentiment-related features from large amounts of data without designing feature extraction rules manually. These models typically use multi-layer neural networks, such as Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Long Short-Term Memory Networks (LSTM), and Transformers. They are able to efficiently capture local and global features. Aslan et al. extracted a large amount of social media text by designing an API and used a CNN model for feature extraction [6]. The results indicated that the TSA-CNN-AOA (KNN) model achieved the highest text categorization accuracy at 95.1%. Huang et al. proposed an AEC-LSTM model that greatly improves the emotional intelligence and attentional mechanisms by enhancing emotion classification performance [7]. It solved the negative impact of high-level abstraction on text feature learning. Ren et al. introduced multi-granularity semantic features for Chinese text sentiment analysis using a bidirectional LSTM and attention mechanism, achieving an F1 score of 84.80%, significantly improving upon existing methods[8]. These studies demonstrate that integrating new models and features enhances sentiment analysis accuracy, with deep learning offering improved expressiveness and generalization [9].

# 3. Applications of Sentiment Analysis in Social Media

# 3.1. Brand Management and Consumer Sentiment

Brands increasingly leverage social media platforms to monitor and analyze consumer behavior and sentiment, thereby facilitating more effective management of brand image and enhancing consumer satisfaction. By analyzing emotional expressions and sentiment, brands can gain deeper insights into consumer emotions and preferences, thus enabling the personalization of communication strategies and product marketing efforts. This approach enhances consumer engagement while enabling more targeted and adaptive brand positioning. Linear equations can model product prices and ratings from online reviews, suggesting that future sales can be predicted by analyzing consumers' preferences for various product features. Also, sentiment-based latent semantic analysis models, can be used to predict sales through an auto-regressive sentiment perception approach. For movie reviews and box

office data, sentiment quality and online reviews heavily impact box office predictions, highlighting the effectiveness of analyzing consumer feedback and sentiment data to predict market trends. Besids, by collecting and analyzing consumer feedback, brands can establish communities on social media, fostering greater interaction among consumers and enhancing brand loyalty, thus gaining a competitive edge in a crowded market. In response to negative events, brands can also manage consumer sentiment effectively, reducing the impact of negative publicity.

# 3.2. Public Opinion Monitoring

Opinion dynamics detection refers to the real-time monitoring and analysis of public statements, emotions, and attitudes on the internet through technical methods, with the aim of understanding and assessing public perceptions and reactions to specific events, brands, products, or services. And this process usually involves collecting feedback on social media, using sentiment analysis techniques to determine the emotional color of the remarks, and monitoring the trends and hot spots of public opinion, which is important for crisis management and market analysis. For example, on February 27, 2019, Fuling Zhacai Group released its earnings report, showing a revenue of 1.99 billion yuan, a slight increase of 3.93% year-on-year, but a net profit decline of over 8%. In August of the same year, someone on a TV program from Taiwan China mistakenly speculated that the drop in Fuling Zhacai's stock price was related to consumers' purchasing power, which quickly sparked widespread discussion. In response, Fuling Zhacai capitalized on this incident by launching a promotional lottery campaign, cleverly addressing the controversy. The campaign went viral, with the number of shares on social media exceeding 100,000, setting a new record. They successfully transformed negative public opinion into positive publicity through proactive social media marketing, boosting brand awareness and consumer favorability. By engaging with consumers, The company showcased the flexibility and appeal of its brand, effectively maintaining its brand image in a timely manner. Public opinion monitoring revealed that its stock price and turnover rate both increased during the incident, indicating that trending social media events had a significant impact on listed companies, helping to mitigate negative effects and enhance market competitiveness.

#### 3.3. Crisis Response and Intervention

Sentiment analysis plays an important role in crisis response and intervention, helping organizations, governments, and relief teams understand the needs and emotions of the public during a crisis. By analyzing sentiment in social media, it is possible to monitor the public's reaction to crisis events in real time and use sentiment analysis techniques to develop early warning systems that predict early signs of a potential crisis in advance. This process not only reveals public sentiment, but also aids decision makers in developing targeted response strategies. For example, on April 9, 2017, United Airlines Flight 3411 sparked widespread discussion after an overbooking incident resulted in passengers being forced off the flight. A study that sentiment analyzed 323 tweets found that 39.3% of tweets expressing emotion reflected disgust, 26% reflected anger, and 9.6% reflected pride [10]. These findings reflect a public outcry and a demand for justice. However, United Airlines failed to interpret these signals effectively. Their initial response, a brief apology, was insufficient to address the gravity of the situation, aggravating public anger. The CEO's dismissive stance and continued support for the staff further alienated the public, exacerbating the crisis. Although a formal apology and an internal review were issued later, the damage to the company's reputation was substantial, as evidenced by a 6% decline in stock value, resulting in a loss of over \$800 million in market capitalization. This case demonstrates the effectiveness of sentiment analysis in identifying mood swings and trends in crisis management, helping to make more timely and appropriate responses and enhancing the effectiveness of interventions. Conversely, ignoring public sentiment may lead to an escalation of the crisis and make management more difficult.

## 3.4. Multidomain Applications of Sentiment Analysis

Sentiment analysis has significant potential in a number of fields, including business, healthcare, and education, enabling businesses to gain insights into user emotions and needs through social media data, while also facilitating the rapid identification of potential issues or opportunities.

In business, analyzing social media data enables companies to extract consumer sentiment through model training on annotated datasets, helping them understand demand, optimize services, refine strategies, and enhance user experience. Monitoring social media discussions enables the timely identification of public sentiment, assessment of public emotions, and rapid response to negative information, protecting brand reputation. Besides, analyzing trending topics helps brands capitalize on current events to enhance their image and boost sales [11].

In the healthcare sector, online social media has become an important source of health-related information for both healthcare professionals and the public [11]. Text categorization and sentiment analysis are widely employed in healthcare to assess emotional tendencies. By classifying text into categories like positive, negative, or neutral, and applying sentiment analysis in complex contexts, emotional states expressed by patients in online forums, social media posts, or medical surveys can be effectively identified. Monitoring user statements helps detect negative emotional signals, like anxiety or depression, enabling early warning systems for timely intervention in at-risk individuals. Furthermore, analyzing emotional states can lead to more personalized healthcare services, better tailored to meet patients' needs.

In education, the rise of online learning, including MOOCs, enriches extracurricular learning and provides real-time feedback opportunities. Analyzing student feedback on courses and instructors can enhance teaching quality and aids in course selection and career planning, and monitoring emotional expressions can help identify psychological and academic issues early, enabling timely support and fostering a positive online learning environment. For instance, a sudden drop in social interaction or negative posts may signal psychological issues. Sentiment analysis can detect these concerns, enabling early intervention and support for students, and can also monitor public opinion, reduce negative content, and promote a positive online environment.

#### 4. Challenges and Future Trends in Sentiment Analysis Technology

Despite the strides made in sentiment analysis, a number of challenges remain. In particular, analyzing polysemous words requires determining the sentiment value in context, a complex task in itself. Improving the accuracy of this type of analysis usually requires training with large amounts of labeled data or dynamic learning models. In addition, complex context processing is another major difficulty in sentiment analysis, and the challenge lies in capturing implicit messages in text, such as irony, humor, and puns. These linguistic phenomena are difficult to recognize by simple rules or dictionaries, so deep learning models are needed to capture long-distance dependencies in text. Cross-language sentiment recognition further complicates the task, making the development of generalized sentiment analysis models adaptable to different languages and domains a key issue in the context of cultural exchange and globalization. Meanwhile, the quality and consistency issues of data annotation are also a major challenge. Most sentiment analysis models rely on dataset training and autonomous learning, especially when analyzing and processing large-scale data streams in real-time, where removing noisy data is critical. In addition, the interpretability and transparency of sentiment analysis models, and the issues of user privacy and data security, require urgent attention.

Neverthelesss, sentiment analysis technology show great potential in future trends. The use of deep learning and transfer learning paves the way for solutions to more complex tasks. The development of multimodal sentiment analysis, which will combine multiple data sources such as text, images, and sound, will provide a more comprehensive analytical perspective. Personalized and user-customized sentiment analysis models will better meet the specific needs of users. Fine-grained sentiment analysis, which seeks to identify more nuanced sentiments in text, such as specific aspects of a brand, product, or service, will be a central focus of future research. In addition, cross-platform and cross-domain applications, such as health monitoring and political analysis, will further expand the scope of applications for sentiment analysis. Real-time monitoring and early warning systems will enable sentiment analysis to offer timely alerts for potential negative events. The widespread adoption and standardization of sentiment analysis tools will enhance the accuracy and reliability of sentiment analysis. Meanwhile, integration with other AI technologies, such as image recognition and speech recognition, will provide a more comprehensive perspective on sentiment understanding. As sentiment analysis techniques address current challenges, they are evolving toward more intelligent, precise, and widespread applications.

## 5. Conclusion

This paper provides an overview of the key sentiment analysis methods developed in recent years and examines their applications in social media data mining. Through comparative analysis, it can be seen that rule-based, machine learning, and deep learning-based sentiment analysis methods have become relatively well-established. However, they are difficult to achieve optimal results when dealing with complex situations or dynamic changes. Meanwhile, sentiment analysis is widely used in the fields of public opinion monitoring, crisis response, marketing planning and public opinion control. As existing methods are not yet sufficient to cope with the rapidly evolving social media environment, there is an urgent need to explore new approaches to expand the application of sentiment analysis techniques and to develop tools capable of handling real-time data streams. Future research could focus on cross-modal sentiment analysis, enhancing interpretability, and developing industry-specific models to improve the effectiveness and adaptability of sentiment analysis in real-world applications.

#### References

- [1] Zhang, Y., et al. (2020) Sentiment Analysis of E-commerce Text Reviews Based on Sentiment Dictionary. ICAICA, 1346-1350.
- [2] Zhang, S., et al. (2018) Sentiment analysis of Chinese micro-blog text based on extended sentiment dictionary. Future Gener. Comput. Syst., 81: 395-403.
- [3] Liu, H., Chen, X. and Liu, X. (2022) A Study of the Application of Weight Distributing Method Combining Sentiment Dictionary and TF-IDF for Text Sentiment Analysis. IEEE Access, 10: 32280-32289.
- [4] Malviya, S., et al. (2020) Machine Learning Techniques for Sentiment Analysis: A Review. SAMRIDDHI : A Journal of Physical Sciences, Engineering and Technology, 12: 72-78.
- [5] Singh, J., Singh, G. and Singh, R. (2017) Optimization of sentiment analysis using machine learning classifiers. Hum. Cent. Comput. Inf. Sci. 7: 32.
- [6] Aslan, S., Kızıloluk, S. and Sert, E. (2023) TSA-CNN-AOA: Twitter sentiment analysis using CNN optimized via arithmetic optimization algorithm. Neural Computing & Application 35: 10311-10328
- [7] Huang, F., et al. (2021) Attention-Emotion-Enhanced Convolutional LSTM for Sentiment Analysis. IEEE Transactions on Neural Networks and Learning Systems, 33: 4332-4345.
- [8] Ren, J. and Liu, Z. (2023) Integrating multi-granularity semantic features into the Chinese sentiment analysis method. Journal of East China Normal University (Natural Science), 44: 3754-3760
- [9] Jiawa, Z., et al. (2021) Review of Methods and Applications of Text Sentiment Analysis, "Data Anal. Knowl. Discov., 54(6): 1-13.
- [10] Qiu, W. (2018) Emotional Trends in Social Media: The United Airlines Passenger Removal Incident. http://media.people.com.cn/n1/2018/0126/c416773-29789118-2.html

[11] Nandwani, P. and Verma, R. (2021) A review on sentiment analysis and emotion detection from text. Soc. Netw. Anal. Min. 11: 81.