

Research on Sentiment Analysis Based on Short Video Data

Jiayi She

Guangdong Polytechnic Normal University, Guangzhou, Guangdong, 510450, China

ABSTRACT

Comments on short videos are predominantly everyday spoken language. In different contexts, the same expression carries distinct meanings, which should be interpreted according to contextual information. Furthermore, internet slang emerges frequently, and ironic expressions are widely prevalent. These characteristics significantly raise the difficulty of sentiment analysis. This paper systematically reviews the core literature in this field from 2020 to 2025. Integrating literature induction with content analysis, it establishes a three-dimensional research framework involving technology, application scenarios and theory. Furthermore, it summarizes the technological evolution trajectories, research hotspots, and practical application challenges. The results show that the evolution of sentiment analysis for short video comments can be divided into the following stages: traditional machine learning, deep learning, pre-trained models, lightweight models, and multimodal fusion. Mainstream models have achieved remarkable improvements in accuracy. According to existing studies, the state-of-the-art accuracy in Chinese irony recognition is 78%, and lightweight models can reduce the size of the original BERT model by approximately 70%. Nevertheless, several challenges remain, including inadequate cross-modal fusion, the absence of dynamic adaptation mechanisms for new internet vocabulary, underdeveloped lightweight deployment solutions, and significant shortcomings in data ethics and compliance. This study suggests future research directions including cross-cultural adaptation, internet neologism handling, causal inference, and engineering deployment, providing technical references for content moderation, personalized recommendations, and brand public opinion monitoring on short video platforms.

KEYWORDS

Sentiment analysis; Short video comments; Natural language processing; Pre-trained models; Multimodal fusion

1 Introduction

1.1 Research Background

In recent years, short video platforms have achieved rapid development and profoundly transformed the way users worldwide acquire information^[1]. As of 2024, mainstream platforms including Douyin and Kuaishou have accumulated massive user bases and interactive data. Comments on short videos can timely reflect users' real emotions with linguistic features highly consistent with daily spoken language^[2-3].

Sentiment analysis, also known as opinion mining, refers to the computational method of identifying public opinions, attitudes and emotions toward objects, figures, topics or events^[4]. Most traditional sentiment analysis studies focus on product reviews, news reports and formal posts on social media^[5]. In contrast, short video comments differ greatly from conventional texts. They are characterized by a high degree of colloquial expression, frequent emergence of internet neologisms, strong reliance on context for semantic comprehension, and widespread use of irony. These distinctive features collectively increase the difficulty of conducting effective sentiment analysis^[6-7].

1.2 Research Problems and Objectives

Current research suffers from three major limitations. First, research findings across diverse disciplines are fragmented. The combination of natural language processing technologies and real-world short video application scenarios remains insufficient, and a unified integrated framework is still lacking^[1,8]. Second, most technical studies focus on algorithm optimization while ignoring the actual operation environment of online platforms and lightweight deployment demands, resulting in a scarcity of feasible solutions for practical industrial use^[9]. Third, short videos are featured by colloquial language, sarcastic expressions and multimodal attributes. Nevertheless, systematic collation and summary of targeted technical solutions for the above features are still absent^[6,10]. This paper puts forward three research objectives. First, it systematically sorts out the technical evolution of sentiment analysis for short video comments, and constructs a three-dimensional analysis framework covering technology, scenario and theory. Second, based on data comparison, this paper reveals the performance differences among various models in short video comment scenarios, so as to clarify the applicable scope of each model^[4,11]. Third, it summarizes the core technical challenges in current research, and proposes practical research directions and application suggestions for real-world implementation^[9,12].

2 Literature Review

This paper retrieves and organizes relevant literature following standardized screening procedures. With "short video", "multimodality", "sentiment analysis" and "audience perception" as keywords, literature is collected from multiple databases at the initial stage. After excluding conference abstracts, book reviews, letters and non-academic documents, the remaining literatures are further reviewed and filtered. Finally, valid literature is determined to ensure the research quality and its correlation with the research topic^[1,8].

2.1 Current Research Status of Methods related to Short Video Sentiment Analysis

2.1.1 Traditional Machine Learning Methods

In the early stage of sentiment analysis, methods such as Naive Bayes, Support Vector Machine and Random Forest were widely adopted. Such approaches rely on manual extraction of text features, including phrase combination, part-of-speech tagging and sentiment dictionaries. Although they can achieve satisfactory results on standardized texts, their performance degrades significantly when applied to short video comments, accompanied by weak generalization ability^[13-14].

2.1.2 Deep Learning Methods

Deep learning has brought profound changes to sentiment analysis. For instance, models such as LSTM are good at capturing long-distance contextual dependencies, while CNN is more suitable for extracting local semantic features. Both of them can learn text features automatically^[15]. More importantly, Transformer-based BERT series models have become the mainstream at present. With bidirectional pre-training, they enhance the ability of contextual semantic understanding, and can effectively handle abbreviations, ambiguous expressions and highly context-dependent content^[5,15]. In addition, lightweight improved models tailored for short video scenarios present prominent advantages, with faster operating speed and convenient terminal deployment^[6-7].

2.1.3 Multimodal Sentiment Analysis

Multimodal fusion has become a prevailing research hotspot. By integrating multiple types of information including texts, emojis and video frames, it effectively improves the recognition of irony and implicit emotions^[11,16]. Nevertheless, this field still confronts practical challenges. It is difficult to realize cross-modal feature alignment, and the redundancy of multimodal information remains prominent. Besides, its adaptability in cross-cultural scenarios is far from satisfactory. These issues have not been thoroughly resolved yet^[7,10,17].

2.2 Research on Application Scenarios

2.2.1 Content Creation Optimization

Scholars have explored the auxiliary role of sentiment analysis in content creation, and analyzed how it helps creators capture audience feedback^[18]. Relevant studies have compared the emotional expression patterns of different content types such as entertainment, life skills and marketing. The results indicate obvious differences in emotional presentation among various categories, which requires differentiated analytical methods^[18-19].

2.2.2 Intelligent Recommendation of Platforms

By embedding the sentiment analysis module into the recommendation system, short video platforms tend to prioritize content that triggers positive interactions, so as to optimize user experience. Existing studies have explored the correlation between comment sentiment tendency and content communication popularity^[12,20].

2.2.3 Brand Marketing Evaluation

Brand marketing evaluation serves as a vital application direction in this field. Researchers have constructed a monitoring framework based on comment sentiment analysis to measure the actual effectiveness of marketing campaigns. At present, most studies focus on large enterprises and MCN institutions, while paying insufficient attention to the demands of small and medium-sized creators.

2.3 Comparison of Research Status at Home and Abroad

Research focuses differ between domestic and foreign studies. Foreign researches lay more emphasis on technical optimization and practical application. They mainly commit to improving pre-trained models for better processing of colloquial and fragmented texts, and attach great importance to public opinion evolution as well as emotional polarization mechanisms^[1,15]. In contrast, domestic research keeps pace with international technological development. Massive adaptive improvements have been made for Chinese context. Certain progress has been achieved in the processing of internet neologisms and irony recognition. Furthermore, domestic studies concentrate on user behaviors and emotional expression on short video platforms, and conduct audience-related research combined with local cultural characteristics^[2-3,18].

3 Research Methods

3.1 Research Design

This paper comprehensively adopts three research methods, including literature induction and analysis, content analysis and comparative research. The whole research is carried out in the process of retrieval, screening, coding, induction and suggestion putting forward, so as to ensure the standardization and traceability of the research process.

3.2 Data Sources and Literature Screening

The data of this study are collected from databases such as CNKI and Wanfang Database. The retrieval keywords cover short video comments, sentiment analysis, multimodality and other related terms, with the time span limited from

January 2020 to June 2025. The included literature is restricted to core journal papers, dissertations and high-level conference papers. All selected studies focus on short video sentiment analysis and provide clear methodological explanations, model descriptions or application conclusions. After systematic screening, a core sample consisting of 24 valid documents is finally determined. Based on the thematic induction and content analysis of core literatures, existing studies are classified into three main research directions: bimodal feature fusion of audio and text, visual emotion representation and scene understanding, as well as audience emotional response and communication effect modeling^[10-11,16].



Figure 1 Clustering Diagram of Research Directions in Sentiment Analysis of Short Video Comments

3.3 Analysis Process

This study uses Excel to statistically analyze the publication years, journal sources, and the proportion of research methods of the collected literature. Meanwhile, a three-dimensional coding system covering four dimensions, including technical routes, model performance, application modes and research limitations, is constructed to conduct content analysis^[1,8].

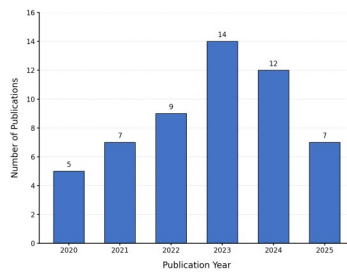


Figure 2 Annual Publication Statistics of Short Video Sentiment Analysis Literature (2020–2025)

4 Research Results

4.1 Research Evolution and Hotspots

A review of the relevant literature reveals that the number of studies in the field of short video sentiment analysis has increased significantly since 2020, with the growth trend accelerating further after 2022. Three main research directions are identified through induction: the adaptive optimization of pre-trained models, multimodal fusion technology, and application research oriented to specific platforms^[1,8]. The stages of technological evolution are clearly defined: from 2020 to 2021, traditional machine learning dominated; from 2021 to 2022, the focus shifted to deep learning; from 2022 to 2023, pre-trained models became mainstream; from 2023 to 2024, lightweight models received increasing attention; and from 2024 to 2025, the field entered the multimodal fusion stage^[6-7,15].

Table 1 Comparison of Technical Routes for Short Video Comment Sentiment Analysis

Technical Route	Representative Models	Core Advantages	Main Limitations	Application Scenarios
Traditional Machine Learning	SVM, Naive Bayes, Random Forest	Strong interpretability, low computational requirements	Relies on manual feature engineering, poor generalization	Small-scale corpora, formal texts
Deep Learning	CNN-LSTM, BiLSTM	Automatic feature extraction, end-to-end learning	Weak modeling of long-distance semantic dependencies	Short text sequences, sentiment polarity classification
Pre-trained Models	BERT, RoBERTa, ERNIE	Strong context understanding, good at sarcasm/ambiguity	High computing power and storage requirements	Colloquial texts, sarcasm recognition
Lightweight Models	ALBERT, DistilBERT, TinyBERT	Suitable for mobile devices and real-time inference	Slightly lower accuracy compared to BERT	Low-computing-power devices, online analysis
Multimodal Fusion	Attention Fusion, Collaborative Learning, Contrastive Learning	Integrates emoticon/image/text information	Difficult cross-modal alignment, scarcity of labeled data	Comments with emoticons, video-based comments

4.2 Analysis of Technical Methods

4.2.1 Data Collection Analysis

After statistically analyzing the included literature, this study identifies two prominent problems regarding data in this research field. First, most studies adopt self-built datasets while unified benchmark datasets are lacking. More than 60% of existing researches construct datasets independently by crawling user comments from short video platforms such as Douyin and Kuaishou^[3,9], with data volumes ranging from several thousand to tens of thousands of entries. Differences in data sources, annotation rules and emotion classification standards across studies hinder the horizontal comparison of model performance. Only a small number of studies adopt public datasets, which are not specially designed for short video comments and thus lack sufficient coverage of colloquial expressions and ironic samples^[6].

Second, the quality of data annotation varies greatly. Some studies only label positive and negative emotions, ignoring complex categories such as irony, neutrality and mixed emotions^[6-7]. Meanwhile, few researches report the number of annotators and consistency tests, which reduces data credibility^[4]. Therefore, establishing an open and unified benchmark dataset for short video comment sentiment analysis that covers irony and emerging internet buzzwords is a fundamental task to promote the further development of this field.

4.2.2 Model Architecture and Performance Comparison

Existing studies have shown that sentiment analysis models integrating contextual semantics and emoji information can achieve an accuracy of over 75% in irony recognition tasks, with a significant improvement in the recognition of ambiguous emotions^[6-7]. According to statistical results, BERT-series models (including BERT-MLFF) are the most widely adopted. Due to their outstanding performance in contextual semantic comprehension, they have become the mainstream basic models in the current research on text sentiment analysis of short videos.

The CNN-LSTM hybrid model ranks second in application frequency. It combines the advantages of CNN in local feature extraction and the capability of LSTM in temporal dependency modeling, and is widely applied in traditional deep learning-based sentiment analysis tasks. In addition, models such as ELMO-TextCNN-Reformer, emotion lexicon fusion models, MODA attention-based multimodal models, and FE-ALBERT-AR models are mainly applied in specific segmented scenarios. They correspond to specialized research directions including in-depth semantic enhancement, emoji sentiment fusion, cross-modal information interaction and lightweight model optimization respectively.

Table 2 Performance Comparison of Models for Short Video Comment Sentiment Analysis

Model Category	Representative Models	Accuracy Range	F1-Score Range	Sarcasm Recognition Accuracy
Traditional Machine Learning	SVM	65%–72%	0.63–0.70	—
CNN-LSTM Hybrid	CNN-LSTM, BiLSTM-Attention	74%–80%	0.72–0.78	~72%
BERT Series	BERT, RoBERTa, BERT-MLFF	80%–86%	0.78–0.84	78%
Lightweight Models	ALBERT, DistilBERT	76%–82%	0.74–0.80	~75%
Multimodal Models	MODA, MISA, MMIM	78%–84%	0.76–0.82	~76%

The overall distribution indicates that pre-trained language models have become the mainstream choice in this field. Traditional deep learning models still maintain steady application value, while multimodal and lightweight specialized models serve as supplementary solutions to meet diverse and segmented research demands.

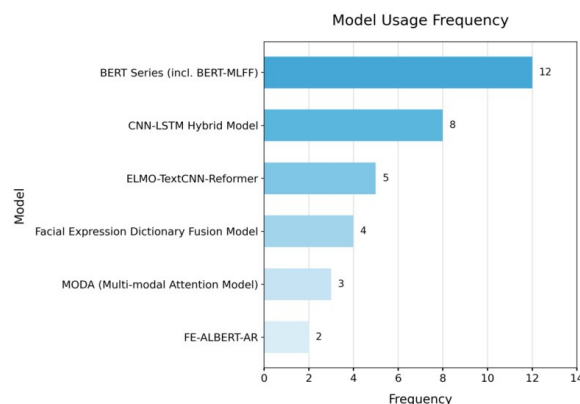


Figure 3 Usage Frequency Statistics of Mainstream Sentiment Analysis Models

4.3 Application Scenario Analysis

In terms of application coverage, short video sentiment analysis models involve a wide range of fields, which can adapt to general scenarios and achieve refined performance in segmented scenarios. For universal scenarios including short video content, commodity reviews and public opinion analysis, such models present the highest adaptability. It proves their cross-domain and full-scenario analytical capabilities, which can be applied to content ecosystem governance, e-

commerce consumption research and public opinion monitoring simultaneously.

In other segmented scenarios, the models maintain balanced performance and comprehensive coverage, which are specified as follows. The lightweight short-text scenario focuses on real-time and rapid analysis of bullet comments, titles and other brief content. The multimodal scenario processes images, texts and emojis collaboratively, making it suitable for mining mixed emotional signals in social contexts. The text unimodal scenario judges emotional polarity merely through plain texts, and fits for ordinary comment texts on Weibo and short video platforms. The fine-grained analysis and few-sample scenario can decompose multi-dimensional sentiments such as logistics and packaging in commodity reviews, and complete analytical tasks with limited datasets.

The above scenarios have formed a complete application matrix centered on three major directions. First, in content creation optimization, beyond simple traffic statistics, the technology helps creators perceive the actual emotional feedback of audiences. It provides targeted revision suggestions by identifying dull segments and weak emotional expression, so as to enhance the attractiveness of works^[18-19]. Second, in intelligent platform recommendation, emotion is introduced as an essential adjustment dimension instead of relying merely on personalized preference. It filters negative content that users resist, diversifies emotional distribution in recommendation streams, and prevents users from falling into extreme emotions, thereby optimizing overall user experience^[2,20]. Third, in brand marketing evaluation, sentiment analysis quantifies the emotional impact of brands. It distinguishes public trust and preference toward brands, compares emotional characteristics with competitors, and captures abnormal emotional changes in advance to realize early warning and emergency response before public opinion crises break out.

In this way, sentiment analysis technology is no longer limited to basic emotional polarity recognition. It can provide substantial technical support for core links including content production, platform recommendation and brand operation.

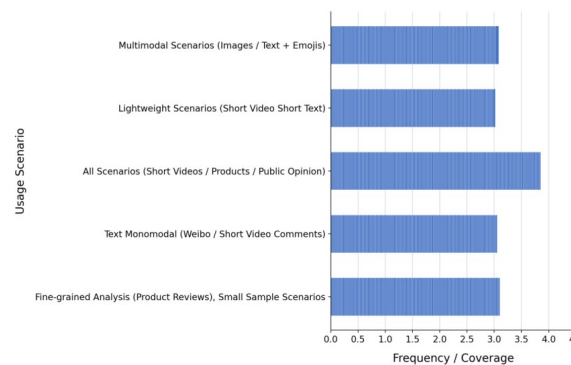


Figure 4 Application Scenario Adaptation of Short Video Sentiment Analysis Models

4.4 Identification of Research Gaps

Through a comprehensive review of the selected literature, several obvious deficiencies in the current research on short video sentiment analysis are summarized. In terms of cross-modal fusion, most existing studies simply splice multimodal features or adopt weighted methods to realize shallow fusion^[10-11,16]. Only a few researches adopt the attention mechanism^[6], and there is still a lack of end-to-end multimodal collaborative learning frameworks specially tailored for short video comments.

In terms of dynamic language adaptation, internet buzzwords update rapidly, while most current models adopt static training patterns without effective online updating mechanisms. In terms of lightweight deployment, although relevant researches on lightweight models have been carried out, engineering implementation and practical deployment for mobile terminals and edge devices remain insufficient^[6-7,9].

Furthermore, cross-cultural research is nearly blank. Existing literature mainly focuses on single-language scenarios, with few comparative studies on emotional expression between Chinese and English short video comments^[1,8]. In addition, ethical research is severely inadequate. The included studies rarely involve data privacy protection, algorithmic fairness and bias assessment^[9,12].

5 Conclusion and Suggestions

5.1 Main Conclusions

The technology of short video comment sentiment analysis updates rapidly. Fortunately, a complete development path has been formed in recent years: from traditional machine learning to deep learning, then to pre-trained models, lightweight models, and finally to multimodal fusion, with the overall accuracy continuously improving^[11,15]. However, the emergence of problems such as colloquialism, dense new words, irony, and strong contextual dependence remains the core technical challenges^[3,6]. Moreover, most applications are concentrated in fields such as content creation, recommendation systems, and brand marketing. A more common issue is that there is an obvious technical gap between large institutions and small-and-medium-sized creators, the degree of interdisciplinary integration is insufficient, and the

combination of natural language processing technology and social science theories is not in-depth enough ^[1,8].

5.2 Future Research Directions

In the future, focus can be placed on promoting the construction of cross-linguistic and cross-cultural standard datasets and evaluation systems for short video sentiment analysis ^[11-12] to improve the accuracy and integration of datasets. In addition, research on how models can automatically adapt to internet buzzwords and adopt lightweight update methods to improve corresponding accuracy; explore the causal relationship between emotional expression and communication behavior ^[9]; establish an ethical evaluation framework for data compliance and algorithmic fairness; and accelerate the research and development of lightweight multimodal sentiment analysis systems that can be used collaboratively on terminals, edges, and clouds ^[16].

5.3 Practical Suggestions

Practical suggestions include the following aspects: For researchers, we should strengthen interdisciplinary cooperation, build open datasets and unified evaluation standards, and strive for win-win cooperation ^[1,8]; for platform operators, we should optimize API opening and privacy protection, and strive to improve algorithm transparency ^[2,9]; for brands and MCN institutions, we should establish a full-link emotional analysis and early warning mechanism to better meet market and user needs; for ordinary creators, we should provide more convenient and lightweight technical tools to reduce the threshold of using sentiment analysis technology and promote the popularization and application of related technologies.

References

- [1] Wang, X.X. (2025). Research on the current status of domestic and foreign short video research based on bibliometric analysis.
- [2] Zhong, Z. H., Xiao, J. H., Wu, Y., & Wang, X. C. (2021). Research on popularity modeling of online short videos based on Douyin platform. *J. Electron. Sci. Tech. Univ.*, 50(5): 774–781.
- [3] Wu, J.Y., & Zhang, P.P. (2022). Research on emotional turn in comments of vaccination short videos. *News Knowl.*, (11): 78–88.
- [4] Liu, M.P. (n.d.). Research on key technologies of video emotion analysis. Doctoral dissertation, Southeast University.
- [5] Li, L. (2024). Research on sentiment analysis of online consumer reviews based on BERT: A case study of mobile phone products. Doctoral dissertation, Hangzhou Dianzi University.
- [6] Wei, S.X. (n.d.). Research on short video sentiment analysis and irony recognition based on deep learning. Doctoral dissertation, Lanzhou Univ. Technol.
- [7] Hong, J.J. (n.d.). Multimodal sentiment analysis based on audio and video enhanced text. Doctoral dissertation, Yangtze University.
- [8] Kang, C.Y., & Pei, M.F. (2025). Visual analysis of research context and hotspots of short video emotional communication. *Commun. Copyr.*, (6).
- [9] Chen, J.H., Wang, Y.F., & Nie, H.Z. (n.d.). Research on analysis model of public opinion evolution of short videos in sudden events. *J. Inf. Resour. Manage.*
- [10] Liu, T.L., Xiao, Y.H., & Dai, X.B. (2022). Method and system for annotation associated short video emotion recognition based on vision and language. China Patent No. CN114882412A.
- [11] Yang, J.T. (n.d.). Research on key technologies of sentiment analysis based on multimodal fusion. Doctoral dissertation, University Name.
- [12] Wang, X.B. (n.d.). Text sentiment analysis of online public opinion events based on multi-source data: A case study of China-US trade war. Doctoral dissertation, Xiangtan University.
- [13] Xie, F. (2023). Research on short text sentiment analysis of social network platforms based on deep learning. Doctoral dissertation, Nanchang University.
- [14] Zhang, Y.M., & Li, Y. (2026). Fine-grained sentiment analysis of microblog texts based on BERT model. *Mod. Inf. Technol.*, 10(4): 112–115.
- [15] Chan, Y.J. (2025). Survey of sentiment analysis based on deep learning. *Inf. Comput.*, (21): 78–80.
- [16] Liu, X.W. (n.d.). Research on video sentiment analysis based on multimodal fusion. Doctoral dissertation, Nanchang University.
- [17] Zhou, J., Ye, J.M., & Li, C. (2021). Multimodal learning emotional computing: Motivation, framework, and suggestions. *e-Educ. Res.*, 42(7): 8.
- [18] Shi, J.Y. (n.d.). Study on influencing factors of emotional communication effect of short videos of mainstream media: Taking Tiktok short videos of the Central Committee of the Communist Youth League as an example. Doctoral dissertation, Yangzhou University.
- [19] Zhao, X.T., & Luo, N. (2024). Analysis of user demand characteristics of cultural short videos: Based on KANO model. *J. News Res.*, 15(6): 17–20.
- [20] Song, X. (2022). Research on emotional turn in short video content production of mainstream media. Doctoral dissertation, Hunan University.