Proceedings of CONF-CDS 2026 Symposium: Computer Vision-Based Multimodal Cognitive Load Estimation for Adaptive Media Communication
DOI: 10.54254/2755-2721/2026.GL35343

The Technological Evolution of Sentiment Analysis —A Comparative Study from SVM to Large Language Models

Yaxuan Wang

College of Computer Science and Electronic Engineering, Hunan University, Changsha, China
3052829590@qq.com

Abstract. As the ecosystem encompassing social media and product reviews grows ever more intricate, emotional expression presents prominent traits including subtlety, sarcasm, fragmentation and multimodal fusion. Traditional machine learning models (e.g., SVM), which rely on manual feature engineering, encounter bottlenecks in recognition accuracy when dealing with irony, metaphor, and long-distance sentiment dependencies. Sentiment analysis of reviews is thus trapped in the dual predicament of "semantic noise" and "shallow understanding." This paper focuses on the advantages and accuracy verification of Large Language Models (LLMs) in tackling high-difficulty sentiment analysis of reviews. This study abandons the mere enumeration of single accuracy values and instead delves into the cognitive breakthroughs of LLMs across three key dimensions: context-aware ambiguous meaning resolution, which deeply interprets ironic and euphemistic connotations; fine-grained sentiment element extraction, accurately identifying the polarity of praise or criticism toward specific product attributes; and closed-loop verification through sentiment generation and explanation, providing traceable justifications via chain-of-thought mechanisms.

Keywords: Large Language Models, Sentiment Analysis, Semantic Ambiguity Resolution, Irony Detection

1. Introduction

With the profound integration of Web3.0, social media and e-commerce, user-generated text reviews have become core data for enterprises to perceive user sentiment and advance product iteration. Sentiment analysis, as a core task in natural language processing, aims to identify the emotional polarity, attitude tendency, and subjective information embedded in text. However, current review data increasingly exhibit significant semantic complexity: user expressions are becoming more subtle, with irony, exaggeration, niche jargon, and fragmented language prevalent. Dictionary-based approaches and shallow statistical learning confront a twofold predicament when analyzing such texts—insufficient semantic penetration and limited domain generalization ability.

In recent years, large language models (LLMs) based on the Transformer architecture have exhibited robust language comprehension and generation competence, owing to their hundreds of billions of parameters and context-aware mechanisms. Unlike traditional machine learning models, LLMs possess the ability to reason about implicit semantics and adapt to unseen contexts. However,

DOI: 10.54254/2755-2721/2026.GL35343

in the vertical application scenario of sentiment analysis, how to verify LLMs' superior accuracy compared with conventional algorithms over traditional methods in complex review scenarios, and how to reveal their internal mechanisms in addressing specific pain points such as irony detection, remain cutting-edge issues awaiting in-depth exploration.

Against the above background, the core research question of this paper is: In scenarios where the emotional expression of user reviews is becoming increasingly obscure and the recognition accuracy of traditional models has reached a bottleneck, how can large language models break through the limitations of shallow semantic understanding to achieve higher-precision sentiment discrimination in complex review texts. Through comparative analysis, this paper examines the extent and internal logic of the accuracy improvement achieved by large language models when dealing with complex samples such as ironic expressions, implicit sentiment interpretation, and domain-specific terminology.

The primary objective of this study is to validate and explain the accuracy advantages of large language models in the task of sentiment analysis of review texts. Different from existing generalized review studies, this paper constructs a closed-loop argumentation path that proceeds from "analysis of technical pain points" to "analysis of model architecture" and finally to "validation of actual performance."

This paper clarifies the theoretical differences in the field of sentiment analysis between "shallow statistical methods" and "deep semantic understanding methods" by analyzing the fundamental differences in feature extraction mechanisms between traditional approaches such as SVM and large language models. Furthermore, by revealing the performance of large models in real-world review environments, this paper helps enterprises more accurately capture user dissatisfaction and latent needs, thereby reducing the risk of decision-making errors caused by misinterpreting ironic reviews.

2. Literature review

2.1. The research lineage of sentiment analysis technology

As a core task in the field of natural language processing, sentiment analysis has undergone a technical evolution from rule-driven to data-driven approaches, and from shallow learning to deep semantic understanding. Tracing this lineage helps to accurately position this study within the technological system of sentiment analysis.

Traditional Machine Learning Era:In the early stages of sentiment analysis research, lexicon- and rule-based methods dominated. These approaches constructed sentiment vocabulary databases and designed combinatorial rules to determine the sentiment orientation of texts, showing a certain degree of applicability in specific domains. However, sentiment lexicons have limited coverage, rule design heavily relies on expert knowledge, and cross-domain generalization ability is weak. Subsequently, traditional machine learning methods represented by Support Vector Machines (SVM) and Naive Bayes gradually became the mainstream paradigm. These approaches transform sentiment analysis into a text classification problem, extract statistical features such as N-grams and TF-IDF through feature engineering, and then use classifiers for polarity judgment. Research has shown that traditional machine learning-based sentiment analysis methods perform well in scenarios with explicit emotional expressions, but limitations such as feature sparsity and insufficient semantic representation capability make it difficult for them to handle complex contexts.

Deep Learning Revolution Period:With the development of word embedding techniques and neural networks, deep models such as LSTM, Bi-LSTM, and CNN have been widely applied to sentiment analysis tasks. Compared to traditional methods, deep learning models can automatically

DOI: 10.54254/2755-2721/2026.GL35343

learn distributed representations of text and capture sequential semantic information, significantly improving the accuracy of sentiment classification. Zhao Kaibo et al. [1] pointed out that the introduction of BERT marked an important milestone in the field of natural language processing (NLP), as it significantly enhanced the performance of many NLP tasks through innovations such as the bidirectional Transformer encoder and the Masked Language Model (MLM). Subsequently, pre-trained language models represented by BERT further revolutionized the research paradigm: by learning general language knowledge through large-scale pre-training on corpora and then fine-tuning for downstream tasks, they have achieved breakthrough progress in multiple sentiment analysis benchmarks.

Rise of Large Language Models Period: In recent years, large language models (LLMs), represented by the GPT series, ChatGLM, and others, have pushed sentiment analysis into a new stage of generative intelligence. Liu Ziwei [2] and colleagues proposed that LLMs, endowed with strong contextual reasoning and semantic association modeling capabilities, open up new research avenues for fine-grained sentiment analysis. Their context modeling mechanism can effectively capture implicit aspect–sentiment correlations in review texts, while the few-shot learning property helps extract key features related to sentiment expression from limited information. These characteristics provide new technical pathways for addressing sentiment semantic sparsity and noise interference in multimodal reviews. Chen Liming [3] and colleagues developed SentiGLM, a sentiment analysis model specifically designed for social media text based on ChatGLM3. Experimental results demonstrate that it achieves superior performance over classic models including BERT and LSTM in Weibo sentiment classification tasks, while also offering better cost-effectiveness and timeliness. Liu Yiding [4] and colleagues, targeting aspect-level sentiment understanding in dialogue scenarios, proposed a multi-agent consistency reflection method that effectively mitigates the hallucination problem of large language models, outperforming existing mainstream approaches.

2.2. The specificity and research progress of Chinese sentiment analysis

Chinese sentiment analysis faces unique linguistic challenges. First, there are no explicit spaces between Chinese words,hence segmentation errors will spread to subsequent processing links and impair sentiment classification accuracy. Second, Chinese emotional expression is more subtle and euphemistic, with frequent use of rhetorical devices such as irony, metaphor, and exaggeration, often resulting in a discrepancy between literal meaning and true sentiment. Furthermore, constant updates of online slang and specialized terms further raise barriers to semantic comprehension.

In the field of Chinese sentiment analysis, domestic scholars have conducted systematic research. Zhong Jiawa [5] and colleagues carried out a bibliometric analysis of text sentiment analysis methods from 2011 to 2020, summarizing the applicable scenarios, advantages, and disadvantages of three types of approaches—lexicon-based, traditional machine learning, and deep learning, and indicated that multi-strategy hybrid algorithms have emerged as a key optimization trend. Tao Jianhua [6] and colleagues reviewed the current state of emotion recognition and understanding from a multimodal perspective, noting that large model-based transfer learning recognition schemes will become a major development trend, and that the application value of affective computing in detecting emotional disorders such as depression and anxiety is becoming increasingly prominent. Xin Qinying [7] tested the accuracy of Chinese sentiment analysis between ChatGPT and SnowNLP using complex emotional sentences containing "zhenshi", and found that ChatGPT demonstrated higher accuracy when processing words with dual emotional connotations and complex contextual

DOI: 10.54254/2755-2721/2026.GL35343

meanings, which is attributed to its deep language understanding ability and rich knowledge background.

2.3. Real-world challenges of sentiment analysis on review texts

Sentiment analysis of review texts holds significant value in industrial applications, but faces three real-world challenges:

Semantic Obscurity: User reviews on e-commerce platforms and social media are becoming increasingly subtle, where direct emotional vocabulary is often replaced by euphemisms, exaggerated rhetoric, or niche jargon. For example, ironic expressions such as "This phone's heat dissipation is really 'good'" are easily misjudged by traditional keyword-based methods. Tao Jianhua [6] and colleagues pointed out that sentiment recognition relying merely on single modalities and shallow features will suffer sharp accuracy degradation under interference and fail to conform to real human cognitive patterns.

Fine-grained Requirement Upgrade: Practical application scenarios can no longer meet demands for coarse-grained polarity judgment at the document or sentence level. Users need to know which specific attribute of a product a review expresses what kind of sentiment — that is, aspect-level sentiment analysis. Cao Tianya [8] pointed out that Aspect-Based Sentiment Analysis (ABSA) has gained continuous research attention because it can simultaneously capture sentiment polarities toward multiple aspects. Yu Chuanming [9] and colleagues also state that as an important branch of sentiment analysis, ABSA focuses on identifying sentiment information related to specific aspects in text. Traditional aspect-based sentiment analysis generally uses machine learning methods, but such methods cannot effectively process large amounts of data and are inadequate for the task.

Cross-domain Generalization Challenge: Emotional expression patterns differ across diverse domains, and a model trained on restaurant reviews tends to undergo substantial performance drops when applied to electronic product reviews. Although deep models have brought some improvements, the domain dependence of annotated data remains a constraining factor.

2.4. Related research status

Although large language models have demonstrated significant advantages in sentiment analysis, several issues remain in existing research. First, there is a lack of systematic evaluation on reviews: most comparative studies use standard public datasets, which primarily consist of explicit sentiment expressions and fail to sufficiently cover challenging samples such as irony, implicit sentiment, and cross-sentence coreference. Consequently, the capability boundaries of LLMs in real-world difficult scenarios have yet to be clearly characterized. Second, there is a shortage of empirical research on LLMs for fine-grained sentiment analysis: existing studies mostly focus on sentence-level polarity classification, and lack systematic validation of LLM performance in the typical scenario where a single review expresses different sentiments toward different attributes. Finally, there is a lack of multi-dimensional perspectives in technology selection: current research emphasizes accuracy comparisons while giving little attention to practical factors such as deployment cost and inference efficiency; a comprehensive analytical framework of "efficiency–cost–applicable scenario" has not yet been established.

DOI: 10.54254/2755-2721/2026.GL35343

2.5. Future research directions

Based on the findings of this study and reflection on its limitations, the following expandable research directions are proposed.

Direction 1: Probing the Capability Boundaries of Large Language Models in Cross-Cultural Sentiment Expression

The data in this study are primarily from the Chinese review context. However, across different cultural contexts, there are differences in the degree of subtlety in emotional expression, the frequency and forms of irony, and the social acceptability of specific rhetorical strategies. Li Xiangming [10] points out that current large language models have deficiencies in handling text generation characteristics across different cultural dimensions, especially in understanding and generating complex text content related to specific cultures, including key areas such as precise sentiment expression and in-depth contextual analysis. Future research could construct multilingual, cross-cultural parallel test sets of difficult samples, aligning typical expressions of the same sentiment intention across different languages and cultural backgrounds through annotation, and systematically explore the cultural transferability of large language models' sentiment analysis capabilities.

Direction 2: Constructing a Three-Dimensional Trade-off Model of Accuracy, Cost, and Explainability

While this study introduces the concept of multi-dimensional evaluation at a methodological level, it has not yet developed a comprehensive assessment model. Future research could focus on building a decision-making framework that balances three dimensions — accuracy, cost, and explainability. For the accuracy dimension, the F1 score for difficult samples can be adopted as a core metric; for the cost dimension, API invocation fees and inference latency should be included; and for the explainability dimension, standardized quality scoring rules need to be developed. This direction holds practical value for bridging the gap in methodological selection between academic research and industrial applications.

Direction 3: Interdisciplinary Integration of Sentiment Analysis with Social Affective Computing and Mental Health Monitoring

Tao Jianhua et al. [6] point out that the application value of sentiment analysis in detecting emotional disorders such as depression and anxiety is becoming increasingly prominent. While this study focuses on the consumer review scenario, its methodological framework—particularly the focus on difficult samples and the analysis of cognitive dimensions can be transferred to the fields of social affective analysis and mental health monitoring. Future research could explore the application of LLMs' fine-grained sentiment extraction and reasoning capabilities to the early identification of mental health risks in social media texts, and construct a specialized sentiment analysis evaluation system for mental health by integrating psychological frameworks of emotion classification.

3. Conclusion

This study takes sentiment analysis of review texts as the task scenario and systematically examines the technological evolution from SVM to large language models. A comprehensive analysis shows that the core breakthrough of large language models in sentiment analysis does not lie in fully replacing traditional methods in conventional scenarios, but rather in deep understanding of complex semantics. Leveraging contextual reasoning capabilities, they effectively eliminate pragmatic ambiguity in irony and implicit expressions, and achieve fine-grained extraction of sentiment elements from reviews through large-scale pre-training. This capability shifts sentiment analysis

DOI: 10.54254/2755-2721/2026.GL35343

from matching keywords to understanding the speaker's true intention, which constitutes the unique value of generative intelligence in discriminative tasks. Currently, this field still faces challenges such as the lack of a three-dimensional trade-off framework, insufficient cross-cultural sentiment understanding, and non-standardized cognitive dimension evaluation systems. Future research can construct a technology selection framework that balances accuracy, cost, and explainability, conduct systematic evaluations in multilingual scenarios, and extend sentiment analysis to social application domains such as mental health monitoring, thus achieving an organic integration of technical efficiency and humanistic care.

References

[1] Zhao, K., Li, Y., Peng, C., Li, J., & Bao, L. (2026). A review of large model development: Technological evolution from BERT to DeepSeek. *Microelectronics & Computer*. 1000-7180.

[2] Liu, Z., Ni, L., Ni, Z., Zhu, X., & Zhang, W. (2026). A multimodal conditional fusion sentiment analysis method enhanced by large models. *Data Analysis and Knowledge Discovery*. ISSN 2096-3467, CN 10-1478/G2.

[3] Chen, L., Su, Y., Wang, X., Jiang, J., & Zhang, Z. (2025). Re-measurement of residents' subjective well-being: An analysis based on social media text and the large language model SentiGLM. *Journal of Econometrics, 5*(1), 267-292.

[4] Liu, Y., Wang, J., Luo, J., & Zhou, G. (2025). Multi-agent consistency reflection for aspect-level sentiment understanding in LLM dialogues. *Journal of Software, 36*(10), 4753-4767.

[5] Zhong, J., Liu, W., Wang, S., & Yang, H. (2021). A review of text sentiment analysis methods and applications. *Data Analysis and Knowledge Discovery, 5*(6), 1-13.

[6] Tao, J., Fan, C., Lian, Z., Lyu, Z., Shen, Y., & Liang, S. (2024). Current status and trends of multimodal emotion recognition and understanding. *Journal of Image and Graphics, 29*(6), 1–20.

[7] Xin, Q. (2024). A comparative study on the accuracy of sentiment analysis between ChatGPT and SnowNLP. *Journal of Chinese Language and Literature*(119), 51–68.

[8] Cao, T., Wang, Z., Shi, P., Li, K., & Li, S. (2026). Aspect-based sentiment analysis based on KGE and multi-channel attention. *Computer Engineering*. ISSN 1000-3428, CN 31-1289/TP.

[9] Yu, C., Jiang, Z., & Sun, Z. (2026). Multimodal aspect-based sentiment analysis model enhanced by contrastive learning and large language models. *Modern Information, 46*(2), 77–90.

[10] Li, X. (2024). Analysis of text generation and emotional expression capabilities of large language models in cross-cultural contexts. *Technology Enhanced Foreign Language Education*(4).