Emotion Recognition in Social Networks Using Large Language Models

Ming Gao

Department of Mathematics and Applied Mathematics, Nanchang University, Jiangxi, China

2DA6047F6AF500BD@outlook.com

Abstract. With the rapid development of the internet, people have become increasingly reliant on social media platforms, generating vast amounts of social network data. Among these, sentiment data stands out as a particularly valuable category. By analyzing sentiment data, it is possible to predict the evolution of social network opinions and promptly monitor users emotional changes for early intervention. This paper discusses various large language models from the perspective of reducing sentiment classification categories, evaluating their performance through core metrics including accuracy, precision, and recall. Through seven different approaches—such as adding output linear layers, enhancing text recognition capabilities, implementing three-level joint prompt structures, and integrating COMET knowledge reasoning—this paper fine-tuned large language models to achieve more precise identification of emotional data in social networks. The optimized models demonstrated improved overall performance. This comprehensive review of large language models in social network sentiment analysis provides scholars with actionable insights for better fine-tuning strategies in this field.

Keywords: Social network, emotion recognition, large language model.

1. Introduction

In recent years, with the rapid development of the internet, social networks have grown significantly. A social network is a complex system containing a wealth of information, and understanding the emotional information within it can help better grasp the dynamics of public opinion. To analyze emotions in social networks, various models have been proposed: the Graph Convolutional Network (IGCN) model [1], the Topological Label-Rotational Graph Convolutional Network (TL-RGCN) model [2], the Emotion-Rotational Graph Convolutional Network (Emotion-RGCNET) model [3], the ROBERT Large model [4], deep learning-based emotion analysis models [5], and Bert-based models combined with hypergraph dual attention networks [6]. However, accurately distinguishing similar emotional categories and ignoring differences in linguistic morphological structures remain challenges. To address this, researchers have explored various large language models. For instance, Shah et al. achieved 96.4% accuracy in identifying depressive content by fine-tuning GPT-3.5 Turbo 1106 and LLaMA2-7B models [7]. Kwon et al. improved accuracy by combining large language models with linear output layers for English nuclear energy-related text analysis [8]. Polat et al. developed a DialogueRNN architecture integrated with fine-tuned large language models to capture Turkey's unique linguistic features [9]. Zhang et al. designed a three-stage joint prompt inference method achieving 77.93% and 82.31% accuracy respectively [10]. Lin et al. proposed a large language model-based social media text stance detection method that improved F1 scores in stance detection tasks [11]. Long et al. proposed a multi-task generation restructured dialogue emotion recognition model (M-GERC) based on large language models (LLMs), which demonstrated favorable W-F1 performance [12]. Maazallahi et al. fine-tuned the GPT-3.5 Turbo 1106 and LLaMA2-7B models to develop a robust framework for accurate emotion classification [13]. With the rapid advancement of large language models, a comprehensive review of emotion recognition in social networks has become essential.

This paper aims to systematically different large language models from the perspective of their categories of emotion classification, analyzing their advantages and limitations.

2. The Application of LLMs for Emotion Analysis in Social Network

2.1. Datasets

With the advancement of large language models, numerous datasets have emerged, each with its own focus in terms of data collection and variety. For the training of large language models, different models utilize distinct datasets. For instance: The core dataset for ChatGPT is the publicly available Common Crawl, All Gemini models are trained on multimodal and multilingual datasets, Both LLaMA and LLaMA 2 are pre-trained on a publicly available dataset of 2 trillion tokens, sourced from web content scraped by Common Crawl, Wikipedia articles in 20 languages, and LaTeX source code of scientific papers from ArXiv [14].

2.2. Evaluation Metrics

In social networks, accuracy, precision, recall, and F1-score are core metrics for emotion recognition [15]. Additionally, metrics such as AUC and ROC curves, relative error, Kappa statistic, ranking loss, mean absolute error, and minimum absolute error are also used to evaluate emotion recognition from various perspectives.

2.3. Related Work

To provide clearer insights, this study systematically evaluates major language models by categorizing their ability to classify sentiment types across various datasets, from the most limited to the most comprehensive, while analyzing their respective advantages and limitations.

Depression differs from simple positive/negative emotional categorizations, making accurate identification challenging. Shah et al. achieved 96.4% precision in detecting depressive content through fine-tuning GPT-3.5 Turbo 1106 and LLaMA2-7B models [7]. To fully capture the complexity of depressive expressions and analyze nuanced emotional nuances and contextual layers in social media texts, targeted fine-tuning of large language models enhanced their ability to accurately identify depression-related functional indicators and improve text comprehension. This approach enables these models to better handle diverse linguistic expressions while maintaining high accuracy in emotion classification for social media texts. Notably, traditional machine learning methods heavily rely on predefined emotional features and require extensive manual feature extraction from text data. In contrast, large language models can dynamically recognize various linguistic patterns through extensive training with diverse datasets, demonstrating superior adaptability to social media's linguistic environment.

To better understand public sentiment toward nuclear energy on U.S. social media platforms, Kwon et al. employed three large language models—BERT, GPT, and Llama—augmented with output linear layers to generate logical value pairs for analyzing English texts related to nuclear energy on platforms like Twitter [8]. To avoid biases in emotional labeling standards, the researchers utilized seven automated sentiment annotation tools from a multi-emotion annotation toolkit, applying majority voting to categorize data into positive, negative, and neutral sentiments. Comparative analysis between large language models, traditional machine learning models, and deep neural network models revealed that large language models performed better, with BERT and GPT achieving accuracy rates of 81% and 80% respectively. However, Llama2 showed poor performance due to memory reduction and quantization processing. To accurately assess public sentiment regarding nuclear energy on U.S. social media, Kwon et al. conducted training and testing on high-credibility and high-confidence tweets respectively. Results demonstrated higher accuracy rates for experiments using high-confidence tweets.

Current training datasets for emotion recognition predominantly consist of English data. While English remains the most widely used language in practical applications, other languages still hold unique value in sentiment analysis. Different languages exhibit distinct grammatical structures and morphological features. Mainstream large language models demonstrate significant biases when processing ambiguous and nuanced text across languages, with these differences varying depending on the language and model type. To enhance the adaptability of large language models in diverse linguistic contexts, Polat et al. conducted contextual multi-dimensional analysis using fine-tuned models like BERT, GPT-3.5 Turbo, Llama 2, and GPT-4 integrated with DialogueRNN architecture. They analyzed manually annotated Turkish marital dialogues, adapting the models to address Turkish's agglutinative grammar characteristics and complex morphological structures. This approach improved sentiment analysis capabilities, enabling rapid capture of unique linguistic features across languages and facilitating cross-linguistic model utilization [9]. The proposed model resolves limitations inherent in English-only trained models, effectively handling Turkish's structural complexity while paving the way for sentiment analysis in other languages including Chinese.

Building upon prompt-based fine-tuning methodologies, Zhang et al. developed a three-stage joint prompt reasoning framework [10]. The tri-level implicit sentiment analysis model consists of three components: first-level prompts for extracting emotional elements, second-level prompts for sentiment polarity analysis using pre-trained models, and third-level prompts that synthesize semantic information from the first two levels through sentiment mapping to determine polarity. To enhance performance, the researchers constructed distinct semantic label sets with varying granularity levels of synonyms for positive, neutral, and negative sentiment polarities. Each polarity category contains multiple labels expressing related emotional tendencies. They categorized all instances from the benchmark SemEval14Laptop and Restaurant datasets containing target aspect annotations into explicit and implicit sentiments. Experimental results demonstrated that the TPISA model achieved 77.93% accuracy on the SemEval14Laptop dataset and 82.31% accuracy on the Restaurant dataset for implicit sentiment analysis tasks.

Lin et al. proposed a stance detection method for social media texts based on large language models [11]. The approach incorporates three key components: a stance detection prompt template, an emotion-stance multi-task prompt, and a low-rank adapter fine-tuning mechanism. To avoid experimental bias, the researchers applied this method to various model architectures—including Llama3-8B, Llama2-13B, and Yi-9B—with differing complexities, computational efficiencies, and performance characteristics. They further fine-tuned these models using QLoRa technology, enabling effective adaptation to social media text stance detection tasks. Experimental results from SemEval-2016 Task 6A (a dataset categorizing target stances into five types—atheism, climate change as a real issue, feminist movement, Hillary Clinton, and abortion legalization—and classifying stance labels as support, opposition, or neutral with additional positive/negative/neutral sentiment tags) showed an F1 score of 85.49% in the multi-task analysis. Notably, fine-tuning Llama2-13B, Yi-9B, Falcon Mamba-7B, and Jamba-1.5-mini models improved their F1 scores accordingly. Through ensemble voting, the average F1 score for stance detection reached 86.28%.

Long et al. proposed a multi-task generative reconstruction dialogue emotion recognition model called M-GERC [12], which is based on large language models (LLMs). The model incorporates two auxiliary tasks—speaker identification and topic prediction—along with a retrieval module. By integrating the COMET common sense knowledge reasoning model, it supplements top-k retrieved statements with contextual knowledge to generate rich semantic information, providing clues about causality, intent, and responses. Finally, pre-trained LLMs are employed for sentiment analysis. Experimental results were obtained across three datasets: DailyDialog (divided into seven emotions: anger, disgust, fear, happiness, sadness, surprise, and neutral), MELD (classified into seven emotions: happiness, anger, fear, disgust, sadness, surprise, and neutral), and EmoryNLP (categorized into seven emotions: neutral, happiness, calm, firm, fear, madness, and sadness). The findings revealed that M-GERC demonstrated superior W-F1 scores across all three datasets, particularly excelling on MELD and EmoryNLP. Compared with benchmark models like DialogueGCN, KET, and COSMIC, M-GERC also showed improved W-F1 performance.

The model's ability to accurately identify emotions is determined by its training on datasets, meaning the accuracy of emotional categorization in the training set directly impacts the model's performance. Maazallahi et al. fine-tuned GPT-3.5 Turbo 1106 and Llama 2-7B models to propose a robust emotion classification framework [13], which coordinates differences in emotional labels and ensures accurate categorization. The model constructs compliant training sets using annotated instance datasets to avoid potential ambiguities, then enhances classification accuracy through heterogeneous graph neural networks based on compliant data. Validation on three diverse datasets—GoEmotion, Friends, and TEC—demonstrated significant improvements over RoBERTa, DistilRoBERTa, and DistilRoBERTa-v2 models in accuracy, weighted F1 score, and recall rate. However, it's noteworthy that the model's precision remains slightly lower and heavily relies on datasets with rich contextual information.

3. Discussion

3.1. Challenges

Compared with traditional machine models, fine-tuned large language models have made progress in the direction of emotion recognition in social networks, but there are still challenges as follows:

1) Large language models primarily rely on textual features for emotion recognition on social media. However, texts on social networks are more colloquial compared to other types of text, often characterized by short sentences, missing subjects, frequent typos, and grammatical errors. In particular, people often use identical phrases to express entirely different emotions. These characteristics of social media texts pose objective difficulties for emotion recognition by large language models.

2) In existing research, multiple emotion labels have been used to annotate data, yet it remains impossible to cover all emotions. Moreover, in real-world social media data, a single instance often contains multiple emotions rather than just one. This complexity of emotions presents additional challenges for large language models in recognizing emotions in social networks.

3) Not only on social media but also in daily communication, the same utterance may be interpreted with different emotions—or even misunderstood—by different people. This issue also arises during emotional annotation of data. Incorrect emotional labeling can significantly impact the accuracy of emotion recognition in social networks. Therefore, non-compliant emotional annotations pose considerable difficulties for large language models in recognizing emotions in social networks.

3.2. Future Prospects

1) In order to alleviate the influence of stylistic features in social networks on emotion recognition of large language models, future efforts should be made

Information is collected from various aspects and angles, such as the overall emotional tendency of the society, the emotional expression mode of users personal habits, and the influence of users emotional changes.

2) In order to alleviate the impact of the complexity of emotion itself on the emotion recognition of large language models, it is necessary to better express complex emotions from the perspective of combining basic emotion combinations and digital annotated emotion degrees in the future.

3) In order to alleviate the impact of non-compliant emotional labeling on the emotion recognition of large language models, it is necessary to select most labels for emotional labeling and make a more detailed coordination and balance between emotional labels and emotions from multiple perspectives in the future, so as to make the fuzzy data classified clearly.

4. Conclusion

This paper explores seven optimization strategies for social media sentiment analysis in large language models, evaluating their performance through metrics including tag classification accuracy and emotional precision. The study reveals three key findings: 1) Large language models demonstrate superior accuracy in sentiment classification compared to traditional machine learning models and deep neural network architectures. 2) Enhanced classification precision of sentiment labels within databases significantly improves overall analytical accuracy. 3) Targeted fine-tuning of large language models for specific linguistic pairs can substantially boost their capability in detecting particular linguistic sentiments.

References

[1] Khemani B, Patil S, Malave S, Gupta J. Improved graph convolutional network for emotion analysis in social media text. MethodsX. 2025; 14: 103325.

[2] Xie J, Wang Y, Meng T, Tai J, Zheng Y, Varatnitski YI. Multimodal emotion recognition method based on domain generalization and graph neural networks. Electronics. 2025; 14 (5): 885.

[3] Yan J, Pu P, Jiang L. Emotion-RGC net: A novel approach for emotion recognition in social media using RoBERTa and graph neural networks. PLoS One. 2025; 20 (3): e0318524.

[4] Hossain I, Puppala S, Alam MJ, Talukder S, Talukder Z. A visual approach to tracking emotional sentiment dynamics in social network commentaries. In: Proc Int AAAI Conf Web Soc Media. 2024; 18: 596-609.

[5] Li Y, Xie M, Yi Y. Fine-grained sentiment analysis on social network platforms based on deep learning. Appl Res Comput. 2017; 34 (3): 743-7.

[6] Xu G, Liu L, Wang J, Chen Z. Text sentiment analysis based on BERT and hypergraph dual attention network. Appl Res Comput. 2024; 41 (3): 786-93.

[7] Shah SM, Gillani SA, Baig MSA, Saleem MA, Siddiqui MH. Advancing depression detection on social media platforms through fine-tuned large language models. Online Soc Netw Media. 2025; 46: 100311.

[8] Kwon OH, Vu K, Bhargava N, Radaideh MI, Cooper J, Joynt V, et al. Sentiment analysis of the United States public support of nuclear power on social media using large language models. Renew Sustain Energy Rev. 2024; 200: 114570.

[9] Polat EN, Yildiz OT, Demiroğlu C, Kafescioğlu N. Decoding emotional dynamics: A comparative analysis of contextual and non-contextual models in sentiment analysis of Turkish couple dialogues. IEEE Access. 2024.

[10] Zhang X, Yan Z. A three-level joint prompt implicit sentiment analysis method incorporating large language models. Appl Res Comput. 2024; 41 (10): 2900-5.

[11] Li J, Shi L, Ding M, Lei Y, Zhao D, Chen L. Stance detection in social media texts based on large language models. J Front Comput Sci Technol. 2025; 19 (5): 1302-12.

[12] Long Y, Gou Z, Chen Y, Qin L. Multi-task generative reconstruction for conversational emotion recognition based on large language models. Appl Res Comput. 2025; 42 (7): 1964-71.

[13] Maazallahi A, Asadpour M, Bazmi P. Advancing emotion recognition in social media: A novel integration of heterogeneous neural networks with fine-tuned language models. Inf Process Manag. 2025; 62 (2): 103974.

[14] Buscemi A, Proverbio D. ChatGPT vs. Gemini vs. LLaMA on multilingual sentiment analysis. arXiv preprint arXiv: 2402. 01715. 2024.

[15] Saxena A, Reddy H, Saxena P. Introduction to sentiment analysis covering basics, tools, evaluation metrics, challenges, and applications. In: Principles of Social Networking: The New Horizon and Emerging Challenges. 2021. p. 249-77.