Advances in Computer, Signals and Systems (2025)
Clausius Scientific Press, Canada

DOI: 10.23977/acss.2025.090213
ISSN 2371-8838 Vol. 9 Num. 2

Review of Public Opinion Sentiment Recognition Technology

Sijie Cao[1,a,*]

[1]*Department of Economics and Management, Nanjing University of Science and Technology, Nanjing, China*

[a]*1037891226@qq.com*

**Corresponding author*

Keywords: Public Opinion Sentiment Recognition; Text Recognition; Single-Modal Analysis; Multi-Modal Fusion

Abstract: With the rapid development of the internet and social media, public opinion sentiment recognition has become a crucial research topic. This paper first introduces the background and significance of sentiment recognition technology, followed by a comprehensive classification and analysis of single-modal and multi-modal approaches in public opinion sentiment recognition. Single-modal research primarily focuses on textual data analysis, employing techniques such as sentiment lexicons, machine learning, and deep learning for emotion classification. Multi-modal research enhances the accuracy and robustness of sentiment recognition by integrating information from diverse modalities, including text, audio, and visual data. Although existing technologies have achieved notable progress, challenges remain, such as difficulties in processing complex emotions, limited model interpretability, and insufficient generalization capabilities. Future research should prioritize multi-modal data fusion, fine-grained sentiment classification, and improvements in real-time processing and model interpretability to better support enterprise decision-making and social governance.

1. Introduction

Since the 21st century, with the popularization of the Internet and the booming development of social media, online public opinion has become an important tool for gauging public opinions and sentiments. Public opinion not only reflects the attitudes of the community towards major events, policies and public topics, but also has a profound impact on social governance, corporate brand management and market behavior. Against this background, how to efficiently and accurately identify and analyze the emotional tendencies in massive public opinion information has become one of the key issues in public opinion research.

Sentiment recognition techniques, a core task in Natural Language Processing (NLP), aim to automatically identify sentiment polarity from text, including positive, negative, and neutral sentiment categories. By analyzing the sentiment of news, comments, and social media content on different platforms, researchers and policy makers are able to better understand changes in public sentiment and the drivers behind them. In recent years, with the development of machine learning and deep learning technologies, sentiment recognition methods have evolved from early rule- and dictionary-based approaches to statistical learning and neural network-based models. These methods are able to capture the implicit emotional information in language more deeply, improving the accuracy and robustness of opinion sentiment recognition. Meanwhile, research in the directions of cross-modal sentiment recognition and multilingual sentiment analysis also provides a broad prospect for the future development of this field.

2. Classification of Opinion Sentiment Recognition Studies

Public Opinion Sentiment Recognition is the emotional identification of social public opinion, which is the foundation and basis for unfolding sentiment analysis and evolution. The direction of public opinion sentiment recognition is divided into traditional single-modal and multimodal fusion. Single-modal means analyzing a certain kind of data, such as text (mainly), pictures, audio or video, in order to identify the public's emotional expression of a specific topic or event. Multimodal is to combine multimodal information in multimedia data, and comprehensively use sentiment dictionary, machine learning, deep learning and other methods to mine more accurate and rich sentiment information.

2.1. Single-modal Opinion Sentiment Recognition Research

Opinion sentiment recognition is the core foundation of social opinion sentiment analysis, which can be divided into two technical directions: unimodal and multimodal. Unimodal technology realizes sentiment recognition by analyzing a single data type (text/image/audio/video), of which text modality is the main research target due to the easy accessibility of data. Unimodal emotion recognition mainly uses emotion lexicon, machine learning and integrates deep learning methods to realize more accurate emotion mining.

Most of the unimodal sentiment recognition techniques focus on the deep mining of text keywords. For example, Min Li [1]et al. extracted keywords for each topic by collecting topic data from Sina Weibo in the post-epidemic era, analyzing the temporal changes in sentiment using SnowNLP, and clustering the data into topics by combining TF-IDF and LDA topic models. Early stage phase research mainly relies on artificial feature engineering by constructing sentiment lexicon with statistical learning model. With the development of the Internet and computer technology, unimodal opinion sentiment recognition techniques began to involve more deep learning. Yuan Qiongfang[2]et al. constructed a framework for classifying public opinion on public health emergencies based on TextCNN model, and for the first time verified the advantage of convolutional neural network in long text sentiment classification. Zhang, Haitao[3]et al. further optimized the structure of the CNN model, and the experimental results show that its classification accuracy is 12.3% higher than that of traditional SVM methods, especially when dealing with complex semantics such as metaphors and ironies in microblog texts. Shisong Tang's ([4])team innovatively integrates the BERT and Bi-LSTM architectures, and significantly improves the model's ability to capture low-frequency emotion words through the near-synonym replacement data enhancement strategy. To address the problem of irony semantic recognition, Hongpeng Pan [5]develops a collaborative bidirectional encoding model, which improves the accuracy of irony text recognition by 15.8% on the test set, filling the technical gap in this area; Lan You [6]constructs a BERT-BiGRU multimode integration model, which effectively integrates the surface features of the text with the deeper semantic characterization, and improves the generalization ability by 9.2%; Wei Meng ([7]) proposes amulti-feature fusion framework, and the BERT-BiGRU model is a new approach to the recognition of ironic words.)team proposed a multi-feature fusion framework, which innovatively integrates word embedding, syntactic structure and sentiment lexicon features to realize a richer semantic space representation. These studies show that technology fusion has become a key path to break through the bottleneck of unimodal recognition.

2.2. Research on Public Opinion Sentiment Recognition with Multimodal Fusion

Multimodal public opinion sentiment recognition technology is an artificial intelligence technology that integrates multidimensional unimodal data such as text, image, audio, etc., and analyzes and mines the implicit emotional tendency of social public opinion through algorithms. Its core advantage lies in breaking through the limitations of single modality, such as combining text semantics, image expression, audio tone and other multi-source information for comprehensive judgment, which significantly improves the comprehensiveness and accuracy of emotion recognition. Domestic scholars have carried out a series of innovative research in this field.

Fan Tao[8]et al. (2024) proposed a multimodal joint attention mechanism-based sentiment analysis model (MCSAM), capturing graphical and textual modal interaction features through word-guided and graph-guided bi-directional attention mechanisms, and verified the effectiveness of the model on multimodal datasets such as "New Crown Pneumonia", which improves the accuracy by 3-5% compared with the traditional methods. Jinmin Zhang [9]et al. (2024) constructed a multimodal analysis framework based on BiGRU, combined word2vec and CNN to extract text and image features respectively, and used linear fusion strategy to realize the sentiment classification, and achieved an accuracy of 73.6% on the dataset of "Li-Ning's sky-high priced shoes", which is about 4 percentage points higher than the baseline model. Chen Jie[10]et al. proposed the DR-Transformer model, which fuses DenseNet image features and RoBERTa text features, and realizes dynamic weight assignment through Transformer Encoder, achieving an accuracy of 79.84% on the microblogging dataset, which is an improvement of 4.74% compared with the unimodal method, and verifies the advantages of multimodal fusion. Wen Peiyu [11]et al. constructed a three-modal fusion network, combining text, audio and video features, and realized multi-dimensional feature extraction through BiGRU and attention mechanism, and the F1 value reaches 63.4% on IEMOCAP and other datasets, which is about 8 percentage points higher than that of the unimodal method. Li Hui [12]et al. proposed the TIsA model, which fuses STFT audio features, BERT text features, and CNN image features, and achieves multimodal feature complementarity through the structure of multi-branch network, and achieves an accuracy of 82.3% on the CH-SIMS dataset, which demonstrates the effectiveness of multimodal fusion. Yang Ruyun [13]et al. proposed a knowledge-enhanced Res-ViT model, combined with an external knowledge base to semantically enhance the text, and used residual networks and visual transformer to extract image features, with an F1 value of 85.2% in the multimodal sentiment analysis task, which is 3.1 percentage points higher than that of the baseline model. information of different modalities, significantly improving the accuracy of sentiment recognition. Wu Peng [14]et al. proposed a bi-directional long and short-term memory model-based classification method (EBiLSTM) for negative emotions of Internet users, which realizes multi-classification recognition by fusing emotionally semantically-enhanced word vectors and bi-directional long and short-term memory networks. The results of the study show that the model can effectively recognize the three negative emotions of anger, sadness, and fear of Internet users through the enhancement of textual emotion semantic by the expression word vectors combined with the context-capturing ability of bi-directional long and short-term memory networks. The results show that the model can effectively recognize the three negative emotions of anger, sadness and fear. Domestic scholars have formed characteristic technical routes in graphic fusion, multimodal modeling, and attention mechanism optimization, which provide important methodological support for public opinion analysis. These studies show that multimodal fusion significantly improves the accuracy of emotion recognition by complementing the feature information of different modalities.

Domestic scholars have formed characteristic technical routes in graphic fusion, multimodal modeling, and attention mechanism optimization, which provide important methodological support for public opinion analysis.

Through the bibliometric method, the research status of domestic online opinion sentiment analysis was sorted out, and the domestic sentiment recognition literature in the China Knowledge Network (CNKI) database was visualized and studied to get the main sentiment recognition models and techniques in the field of intelligence and library intelligence, especially based on sentiment lexicon, machine learning and deep learning, as shown in the table 1 below.

Table 1: Common Models and Techniques for Unimodal and Multimodal Direction Emotion Recognition

Machine Learning	Support Vector Machine (SVM); Random Forest (RF)
Deep Learning	Convolutional Neural Network (CNN); Bidirectional Long Short-Term Memory Network (BiLSTM); Short-Time Fourier Transform (STFT); BERT; Bidirectional Gated Recurrent Unit (BiGRU); BERT combined with Bidirectional Long Short-Term Memory (Bi-LSTM) model ; Deep Auto-Encoder (AE); ResNet34; BERT-BiGRU; VGG19; Xception; DR-Transformer; T5; Res-ViT

3. Research shortcomings and reflections

Existing research on opinion sentiment recognition has limited ability to handle complex sentiment in unimodal studies, and many models can only handle simple binary categorized sentiments (positive and negative), ignoring complex sentiment categories such as neutrality, sarcasm, irony, and so on. Most studies focus on textual data and fail to effectively combine multimodal data such as images and videos for sentiment analysis, limiting the ability of models to capture richer sentiment information. Rule-based methods require a lot of manual labeling, which is time-consuming and costly. The model's ability to quickly react to changes in online public opinion is limited. In a rapidly changing social media environment, this can lead to lagging analysis results. Many deep learning models, while performing well on specific tasks, lack sufficient interpretability, making their decision-making process less transparent. Existing models tend to be optimized for specific domains or datasets and lack adaptability to sentiment expression in different domains and cultural contexts. In the research field of multimodal emotion recognition, models have a high demand for computational resources and a significant increase in training time when dealing with high-resolution images and long-duration audio, which affects the usefulness of the models. Many studies focus on the fusion of text and images, ignoring the emotion recognition of other modalities such as audio and video, limiting the comprehensiveness and accuracy of the models. Some models have been validated only on Chinese datasets, and their adaptability and generalizability on multilingual and cross-cultural datasets still need to be further tested. Whether unimodal or multimodal, their models lack generalization ability.

4. Summary and Outlook

As technology advances, future sentiment analysis will employ more multimodal data, such as the combination of text, images, audio, and video, to capture and analyze human emotional expressions more comprehensively. Researchers will explore more efficient data fusion strategies, such as attentional mechanisms and deep learning models, to optimize feature integration across modalities. Current sentiment analysis models, such as BERT and BiLSTM, have achieved remarkable results but still suffer from high computational complexity and long training time.

In areas such as opinion monitoring and customer service, there are increasing demands for real-time and interpretability of sentiment analysis. Future research will be devoted to improving the real-time processing capability of the model and exploring the visualization of the model's decision-making process to enhance the credibility of the model and the understanding of the user. High-quality multimodal datasets are the foundation of sentiment analysis research. Future work will include the construction of larger and more diverse datasets, as well as the development of finer-grained labeling schemes, such as multidimensional sentiment labeling and sentiment evolution labeling. Future research will be devoted to developing more lightweight models, reducing resource consumption through techniques such as model pruning and quantization, and improving the utility of models.

References

[1] Li Min. Research on governance of public health events in urban communities based on resilience perspective [D]. Fujian Normal University,2023.

[2] YUAN Qiongfang. Research on emotion recognition of online public opinion on public emergencies based on deep learning[J]. Computer Knowledge and Technology,2023,19(07):42-44+48.

[3] ZHANG Haitao, WANG Dan, XU Hailing,et al. Research on microblog opinion sentiment classification based on convolutional neural network[J]. Journal of Intelligence,2018,37(07):695-702.

[4] TONG Shisong, LIAN Liping, HE Chenglong, et al. A kind of sentiment recognition of public opinion information based on Bert and Bi-LSTM[J]. Network Security Technology and Application,2021,(07):57-59.

[5] PAN Hongpeng, WANG Dong, LIU Zhongyi, et al. A study on collaborative two-way coding opinion comment sentiment analysis considering irony semantic recognition[J]. Journal of Intelligence,2022,41(05):99-105+111.

[6] You Lan, Zeng Han, Han Fanyu, et al. Emotion semantic recognition based on BERT-BiGRU integrated learning[J]. Computer Technology and Development,2023,33(05):159-166.

[7] MENG Wei, ZHANG Li. Research on the strategy of semantic sentiment analysis in online public opinion content based on artificial intelligence technology[J]. Science and Technology Communication,2020,12(14):165-167.

[8] FAN Tao, WU Peng,CAO Qi. Research on multimodal fusion netizen sentiment recognition based on deep learning[J]. Journal of Information Resources Management,2020,10(01):39-48.

[9] ZHANG Jinmin, LI Xufang, FAN Dijun. Sentiment analysis of multimodal online public opinion based on BiGRU model[J]. Intelligent Computer and Application,2024,14(01):191-193+199.

[10] Chen J, Ma J,Li X, et al. Research on multimodal emotion recognition based on DR-Transformer model[J]. Intelligence Science,2022,40(03):117-125.

[11] WEN Peiyu, NIE Guohao, WANG Xingmei, et al. Research and application of multimodality in emotion recognition[J]. Applied Science and Technology,2024,51(01):51-58+97.

[12] Li Hui, Pang Jingwei. Research on multimodal Internet user sentiment recognition based on text-graph-audio fusion[J/OL]. Data Analysis and Knowledge Discovery,1-17[2024-09-19].

[13] Ruyun Yang, Jing Ma. A feature-enhanced multimodal sentiment recognition model incorporating knowledge and Res-ViT[J]. Data Analysis and Knowledge Discovery,2023,7(11):14-25.

[14] WU Peng, YING Yang, SHEN Si. Research on the classification of netizens' negative emotions based on bidirectional long and short-term memory model[J]. Journal of Intelligence,2018,37(08):845-853.