Analyzing sentiment and its application in deep learning: Consistent behavior across multiple occasions

Yanxiong Xiang

School of Computer Science, South China Business College Guangdong University of Foreign Studies, Guangzhou, 510545, China

1840606237@e.gwng.edu.cn

Abstract. This article offers a systematic review of the evolution in sentiment analysis techniques, moving from unimodal to multimodal to multi-occasion methodologies, with an emphasis on the integration and application of deep learning in sentiment analysis. Firstly, the paper presents the theoretical foundation of sentiment analysis, including the definition and classification of affect and emotion. It then delves into the pivotal technologies used in unimodal sentiment analysis, specifically within the domains of text, speech, and image analysis, examining feature extraction, representation, and classification models. Subsequently, the focus shifts to multimodal sentiment analysis. The paper offers a survey of widely utilized multimodal sentiment datasets, feature representation and fusion techniques, as well as deep learning-based multimodal sentiment analysis models such as attention networks and graph neural networks. It further addresses the application of these multimodal sentiment analysis techniques in social media, product reviews, and public opinion monitoring. Lastly, the paper underscores that challenges persist in the area of multimodal sentiment fusion, including data imbalance and disparities in feature expression. It calls for further research into cross-modal feature expression, dataset augmentation, and explainable modeling to enhance the performance of complex sentiment analysis across multiple occasions.

Keywords: artificial intelligence, sentiment analysis, deep learning, application scenarios.

1. Introduction

In our modern era of digitization and artificial intelligence, sentiment analysis stands as a paramount area of research in AI. It holds a key significance in understanding and applying emotional roles within interpersonal communications. As the popularity of social media and big data technologies rise, a tremendous amount of sentiment data, including text, images, voice, and video, is generated in our everyday lives. Sentiment analysis aims to precisely extract, comprehend, and express the emotional content from these diverse datasets, thus serving as a vital aid in areas like social interactions, marketing, public opinion monitoring, emotional support, and intelligent human-computer interactions.

Traditionally, sentiment analysis has relied primarily on rule-based systems, statistical machine learning, and shallow feature engineering [1]. However, these methods have often struggled to effectively grasp the intricate relationship between semantics and sentiment, leading to significant challenges in sentiment analysis for real-world situations. In recent years, deep learning, a robust method for representation learning, has made impressive strides in natural language processing, computer vision, and speech processing via the construction of deep neural networks. Its application in sentiment analysis has garnered considerable attention from both academia and industry. Despite these advances, the bulk of research is focused on unimodal sentiment analysis, such as text or speech-based sentiment analysis [2]. Yet, human emotional expressions in real-world scenarios are often multimodal, deriving information from various sensors and data sources. Consequently, multimodal sentiment analysis has emerged as a fresh area of investigation.

This study also addresses the task of analyzing sentiment of identical behaviors across multiple contexts. People might display different sentiments for the same behavior depending on the situation. For instance, a song might evoke pleasure in some settings, while inciting sadness or anger in others. The exploration of sentiment analysis for the same behavior across multiple contexts offers a broader understanding of sentiment diversity and complexity. This paper aims to systematically review recent advances in deep learning for sentiment analysis, particularly focusing on multimodal sentiment analysis and same-behavior sentiment analysis across multiple contexts. It begins by introducing the fundamentals of sentiment analysis, its significance, and applications in social and business domains. It then concentrates on existing unimodal sentiment analysis methods and essential techniques, such as preprocessing and feature fusion methodologies. Subsequently, the paper explores the evolution of multimodal sentiment analysis, its datasets, and representations, alongside deep learning-based multimodal sentiment analysis models. The study will then dive into the significance and challenges of analyzing the sentiment of the same behavior across various occasions.

This work also sheds light on the application of sentiment analysis in social media, product reviews, and public opinion monitoring. Current problems and challenges such as model migration and data scarcity are also summarized. Lastly, the paper discusses the challenges and future outlook of multimodal sentiment fusion. Despite being a core issue in multimodal sentiment analysis, the fusion process has limitations. The paper outlines potential future research directions, including enhancements in model migration, data augmentation techniques, and more efficient sentiment fusion methods. The author hopes that readers will gain an understanding of the recent research progress in deep learning for sentiment analysis, recognize the significance of multimodal sentiment analysis, and gain insights for future research in multimodal sentiment fusion.

2. Concepts of sentiment analysis

Affect and emotion are two important aspects of human emotional experience, and they play a key role in sentiment analysis. In this section, the definitions, differences, and common ways of categorizing will be explored.

Affect is the subjective experience that an individual has of things or events, and is a psychological response that covers a wide range of feelings, such as joy, anger, sadness, fear, disgust, and so on. Emotions tend to be more complex and long-lasting, and they can be influenced by the external environment and internal factors, as well as by an individual's thinking and behavior. Emotions are a form of expression of feelings, which are short-lived and relatively strong emotional responses. Emotions tend to be sudden, they can be caused by specific events or stimuli, and usually manifest themselves in physiological changes and emotional experiences.

Although emotions are distinguished from affect, they are closely related and are often intertwined in actual experience. In sentiment analysis, researchers usually focus on the recognition and classification of sentiment, aiming to understand the sentimental content contained in data such as text, speech, and images [3]. The complexity of affects and emotions allows them to be categorized in many different ways. In the field of sentiment analysis, common classification methods include basic sentiment classification, dimensional sentiment classification, and complex sentiment classification. Basic sentiment categorization: basic sentiment categorization is the division of emotions into a set of basic, discrete sentiment categories. Common basic sentiment categories include anger, joy, sadness, fear, disgust, and so on. This classification is relatively simple and intuitive, and is suitable for rough description and analysis of emotions. Dimensional sentiment categorization: dimensional sentiment categorization views emotions as varying continuously across multiple dimensions. Common dimensions include pleasure and arousal. Pleasure indicates the positive or negative tendency of the sentiment and arousal indicates the intensity of the sentiment. By categorizing sentiment in dimensions, it is possible to describe the emotional experience in a more refined way. Complex sentiment categorization: complex sentiment categorization is the division of sentiments into more and more fine-grained categories. For example, in addition to the basic sentiment categories, sentiments such as satisfaction, embarrassment, envy, etc. may be included. Complex sentiment categorization can express the diversity and complexity of sentiment more comprehensively. In the practical application of sentiment analysis, depending on the tasks and needs, appropriate affective and emotional categorization methods can be selected to accurately capture the emotional content.

By discussing the definition and categorization of affect and emotion, this chapter lays the foundation for subsequent in-depth discussions of sentiment analysis methods and applications. In the following chapters, this paper will continue to explore the key techniques and application cases of unimodal and multimodal sentiment analysis.

3. Unimodal sentiment analysis

Text Sentiment Analysis serves as a crucial research direction within the sphere of Sentiment Analysis. Its purpose is to analyze and comprehend the sentiment information encapsulated within text data [4]. This section zeroes in on the fundamental tasks and primary techniques associated with text sentiment analysis, encompassing preprocessing, feature representation, and sentiment classification models.

3.1. Sentiment analysis of texts

Sentiment analysis of text usually consists of two main tasks: Sentiment categorization: Sentiment categorization is the task of classifying text data into different sentiment categories. Common sentiment categories include positive, negative, neutral, etc. Sentiment categorization aims to determine the polarity of the sentiment expressed in the text, i.e. whether it is positive, negative or neutral. Sentiment intensity analysis: Sentiment intensity analysis is the task of measuring the intensity or degree of sentiment of a text. Emotional intensity can be categorized into several levels such as mild, moderate and strong. Such tasks aim to further refine the expression of emotion and quantify the degree of sentiment. The key technologies of text sentiment analysis include preprocessing, feature representation and sentiment classification models: Preprocessing: It is the first step of text sentiment analysis, which includes text cleaning, word segmentation, and removal of stop words. Text cleaning is aimed at removing noise and useless information in the text, such as HTML tags, special characters and so on. Segmentation cuts the text into individual words for subsequent processing. Deactivation is to filter out common words that have no practical significance in sentiment analysis, such as "的", "是 ", "在", etc. Feature Representation: It is to transform text data into a vector form that can be understood by the computer. Common feature representation methods include Bag-of-Words and Word Embedding. Bag-of-Words model represents text as a vector, where each dimension corresponds to a word, and the value in the vector indicates the frequency or importance of the word in the text. And Word Embedding is a representation that maps words into a real vector space, which captures semantic relationships between words and contextual information. Sentiment classification model: a sentiment classification model is a model that realizes sentiment classification by learning the mapping relationship from feature representations to sentiment categories. Common sentiment classification models include plain Bayesian classifier, support vector machine (SVM), logistic regression, and deep learning models such as recurrent neural network (RNN) and convolutional neural network (CNN). Deep learning models have achieved remarkable results in text sentiment analysis, which can better capture semantic and contextual information in text, thus improving the accuracy of sentiment classification.

In summary, text sentiment analysis is an important research direction in the field of sentiment analysis. Through key techniques such as preprocessing, feature representation and sentiment classification models, researchers are able to accurately recognize and understand the sentiment information embedded in text data.

3.2. Sentiment analysis of voice

The sentiment analysis of voice is an important research direction in the field of Sentiment Analysis, which aims to analyze and understand the sentiment information embedded in speech signals. This section concentrates on the tasks and principal techniques involved in speech sentiment analysis, incorporating acoustic feature extraction, sentiment recognition models, and the significance of datasets. Speech-based emotion classification stands as a vital aspect within a variety of artificial intelligence applications, particularly those concerning human-computer interfaces. Numerous studies focusing on speech-based emotion classification have been put forward to date. These investigations predominantly fall into two categories: methods grounded in deep learning and those utilizing traditional signal processing and statistical techniques. Deep learning techniques are being used to recognize speech-based emotions in recent times [5]. The sentiment analysis of voice usually involves the following two main tasks: Sentiment categorization: sentiment categorization is the task of classifying speech signals into different sentiment categories such as anger, joy, sadness, neutral, etc. This task aims at determining the polarity of emotions expressed by the speaker in speech and thus understanding his or her emotional state. Sentiment Intensity Analysis: Sentiment intensity analysis is a task that measures the intensity or degree of emotion in speech. Sentiment intensity can be categorized into several levels such as mild, moderate and strong. This task aims to provide insight into the degree of emotional expression and the intensity of emotional experience of the speaker.

The key techniques for the sentiment analysis of voice include acoustic feature extraction and sentiment recognition modeling. Acoustic feature extraction: acoustic features are numerical representations extracted from speech signals to describe the acoustic properties and audio characteristics of speech. Common acoustic features include fundamental frequency, spectrogram, Mel-frequency cepstral coefficients (MFCCs) and so on. These acoustic features can capture the audio frequency, energy and harmonic information in the speech signal, providing an important basis for subsequent emotion classification. Sentiment recognition model: a sentiment recognition model is a model that realizes sentiment classification by learning the mapping relationship from acoustic features to sentiment categories. Traditional emotion recognition models include Gaussian Mixture Model (GMM) and Support Vector Machine (SVM). However, in recent years, deep learning models, especially Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs), have achieved significant performance improvements in speech sentiment analysis. Deep learning models are able to learn more abstract and complex feature representations from speech signals to better capture emotional states. In the process of analyzing sentiment of voice, the selection and construction of datasets are critical to model performance. Constructing large-scale and diverse speech sentiment datasets can help deep learning models to better generalize and adapt to different sentiment expressions. When constructing a dataset, it is necessary to consider the balance of different sentiment categories, and also to ensure that the dataset contains a variety of speech types, speakers, and sentiment scenarios in order to fully reflect real-world speech sentiment.

In summary, the sentiment analysis of voice, as an important branch in the field of sentiment analysis, is able to accurately recognize and understand the sentimental state of the speaker by means of key technologies such as acoustic feature extraction and sentimental recognition models [6]. The application of deep learning models in speech sentiment analysis provides humans with more powerful tools to help researchers better mine the sentimental information embedded in speech data. In Sentiment Analysis of Voice, sentiment classification modeling is a key step, which achieves sentiment classification by learning mapping relationships from acoustic features to sentiment categories. In recent years, deep learning models have performed well in speech sentiment classification, especially Recurrent Neural Networks (RNNs) and Convolutional Neural Networks. Several commonly used speech emotion classification models are described below:

Recurrent Neural Networks: RNN is a classical sequence model for processing time series data, such as speech signals. In speech emotion classification, RNNs are able to model the temporal information of the speech signal and capture the temporal evolution of the speaker's emotion. Common RNN structures are Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU). By passing hidden states over multiple time steps, RNNs are able to categorize the sentiment of an entire speech sequence.

Convolutional Neural Networks: CNN is a deep learning model mainly used in image recognition, but it also performs well in speech emotion classification. With one-dimensional convolutional layers, CNNs are able to capture local features of the speech signal. These local features can be combined in subsequent layers to obtain higher level abstract features for sentiment classification. The advantages of CNN in speech sentiment classification are simple model structure and high computational efficiency.

Hybrid Models: Hybrid models combine the advantages of multiple deep learning models, a common combination is to combine CNN and RNN. In a hybrid model, CNN is used to extract the local information of acoustic features, and RNN is used to capture the dependencies of these features on the time series. Such a hybrid model can represent the speech signal more comprehensively and improve the accuracy of sentiment classification.

Pretrained Models: Pretrained models are deep learning models that are pre-trained on large-scale speech data and then fine-tuned on specific tasks. Pretrained models have a more powerful representation and can better utilize large-scale data to extract useful features. In speech sentiment classification, pre-trained models have achieved significant performance gains [7].

In summary, classification models of analyzing the sentiment from voice play a crucial role in unimodal sentiment analysis. Deep learning models such as RNN, CNN, hybrid models and pre-trained models can accurately recognize and understand the speaker's emotional state by learning the mapping relationship between acoustic features and emotion categories. With the continuous development of deep learning technology, better performance and wider application in speech emotion classification can be expected in the future.

3.3. Sentiment analysis of image

Sentiment analysis of images is an emerging field that utilizes deep learning and computer vision to interpret emotions and sentiments conveyed through visual content. By analyzing facial expressions, objects, colors, and scenes, this technique provides valuable insights into user reactions to products, advertisements, and social media images. The fusion of computer vision and deep learning empowers us to understand and interact with images in a more emotionally intelligent way, revolutionizing our perception and interpretation of visual data.

4. Multimodal sentiment analysis

Due to its many applications, which include the diagnosis of mental illnesses, the comprehension of human behavior. Human machine, robot interaction, and autonomous driving systems, multi-modal emotion recognition has recently grown to be a significant research topic in the affective computing community [8].

4.1. Multimodal affective data sets and representation methods

One of the most difficult and significant study areas, with many applications, is emotion recognition based on multimodal data (e.g., audio, video, text, etc.). To determine the best multimodal model for emotion identification integrating audio and visual modalities, this research effort has rigorously explored model-level fusion. For audio and video data, different unique feature extractor networks are suggested [9].

Multimodal sentiment analysis involves multiple types of data, such as text, speech, and images, for comprehensive analysis and understanding of sentiment expressions. In this section the focus will be on dataset selection and representation in multimodal sentiment analysis.

A multimodal sentiment dataset is the basis for performing multimodal sentiment analysis, which consists of multiple types of data and contains samples of different sentiment states. Constructing high-quality multimodal sentiment datasets is essential for training and evaluating multimodal sentiment analysis models. Common multimodal sentiment datasets include: IEMOCAP (Interactive Emotional Dyadic Motion Capture): the IEMOCAP dataset is a multimodal sentiment dataset containing video, audio and text data. The dataset is generated from simulated dialogues and performances and contains real sentimental expressions. The IEMOCAP dataset is widely used in sentiment recognition and sentiment intensity analysis tasks. MOSI (Multimodal Corpus of Sentiment Intensity): the MOSI dataset contains speech, text and visual information from YouTube videos. It is mainly used for sentiment intensity analysis task, i.e., to measure the sentiment intensity of the speaker in the video. CMU-MOSEI (CMU Multimodal Opinion Sentiment and Emotion Intensity): the CMU-MOSEI dataset is a large-scale multimodal sentiment dataset containing video, speech and text data from YouTube. It is suitable for sentiment classification and sentiment intensity analysis tasks. SEMAINE (Sustained Emotionally colored Machine-human Interaction using Nonverbal Expression): the SEMAINE dataset contains multimodal emotion data from virtual characters and real humans. It covers different emotional states and interaction scenarios and is used to study the application of emotion expression in computer interaction.

4.2. Feature fusion methods

When performing multimodal sentiment analysis, different types of data need to be converted into a unified representation to facilitate model processing and learning [10]. Commonly used multimodal sentiment data representation methods include: Early Fusion: early fusion is the fusion of different types of data at the input level. For example, text features, speech features and image features are spliced or averaged at the input level and then fed into a model for training. The advantage of early fusion is that it is simple and intuitive, but it may face the problems of inconsistent feature dimensions and information redundancy. Late Fusion: Late fusion involves inputting different types of data into the corresponding models for feature extraction and learning respectively, and then fusing the outputs of the models. For example, the outputs of text model, speech model and image model are weighted average or spliced, and then input into the sentiment classification model. Late fusion is more flexible compared to early fusion and can better capture the characteristics of each type of data. Cross-modal Attention: cross-modal attention is a widely used method in multimodal sentiment analysis. It assigns different weights to each data type by learning the degree of association between the different types of data. Cross-modal attention helps the model to automatically focus on data types that contribute significantly to sentiment analysis. Graph-based Fusion: graph neural networks, a class of deep learning methods for processing graph-structured data, can also be applied for feature fusion in multimodal sentiment analysis. Multimodal features can be effectively fused by constructing features of different modalities into graph structures and performing information transfer and feature aggregation on the graph. The graph neural network can capture the complex relationship between different modal features and further improve the performance of sentiment analysis.

In summary, the selection and representation of multimodal sentiment datasets is the key to constructing an effective multimodal sentiment analysis model. By selecting a suitable multimodal sentiment dataset and adopting an appropriate data representation method, the correlation between different types of data can be better utilized to improve the performance of multimodal sentiment analysis.

4.3. Multimodal sentiment analysis model based on deep learning

Deep learning models serve as a linchpin in multimodal sentiment analysis, thanks to their ability to effectively handle intricate multimodal data and discern interaction information between modalities [11]. A number of deep learning based multimodal sentiment analysis models are expounded below:

Multimodal Fusion Network (MFN): This network adopts a feature-level fusion method, merging features from different modalities and channeling them into a unified classifier for sentiment classification. Depending on the task requirements and data characteristics, early fusion or late fusion may be deployed in this model. MFN, with its simplicity and intuitive nature, is suitable for handling sentiment data with fewer modalities. Cross-modal Attention Network (CAMN): By introducing an attention mechanism, CAMN facilitates adaptive learning of modal weights across different modal data, enabling a more effective fusion of information from various modalities. It establishes dynamic connections between different modal data, capturing nuanced variations in emotional expressions with greater efficacy. CAMN is well-suited for handling more complex multimodal emotional data, such as the fusion of video, audio, and text. Graph Neural Network (GNN): GNN is a deep learning model designed for handling graph-structured data, which can also be utilized for feature fusion in multimodal sentiment analysis. By constructing features of different modalities into graph structures, performing information transfer, and aggregating features on the graph, multimodal features can be effectively integrated. GNN captures complex relationships between different modal features, which significantly enhances sentiment analysis performance. Two-Stream Fusion Network (TSFN): TSFN uses two separate models to process different modal data, and subsequently fuses their outputs. For instance, a text model and a speech model can independently process text and speech data, with their outputs later merged into a unified classifier for sentiment classification. TSFN offers a more flexible approach to handling data of different modalities, making it suitable for more complex sentiment analysis tasks.

In a nutshell, deep learning-based multimodal sentiment analysis models showcase excellent performance in multimodal feature fusion. These models leverage multimodal information to its fullest to enhance the performance of sentiment classification and sentiment intensity analysis. Nonetheless, different tasks and data scenarios might impose distinct requirements on model selection and design. Therefore, in practical applications, the choice of the most suitable model and fusion strategy must align with the specific situation at hand.

5. Application scenarios of deep learning in sentiment analysis

Deep learning has indeed transformed sentiment analysis, revolutionizing the interpretation and analysis of human emotions and opinions expressed in text. An array of application scenarios has surfaced, highlighting the versatility and efficacy of deep learning in this domain.

5.1. Sentiment analysis in social media

Social media has become a platform for people to express their emotions widely, with huge amounts of text, images and videos being shared every day. Deep learning sentiment analysis models can help automate the identification of users' emotional expressions on social media to understand their attitudes towards topics, events and products. This is especially important for companies and brands to understand users' feedback and sentiment tendencies in a timely manner, so as to optimize products and services and improve user satisfaction. Sentiment analysis of social media evaluations is thought to be useful for smart cities. In smart cities, social media evaluations can be useful for a variety of reasons [12].

5.2. Analysis of product reviews and user feedback

Assessing customer and user sentiments has always played a pivotal role in decision-making across a range of sectors, most notably within the marketing industry. Sentiment Analysis (SA) encapsulates a variety of methodologies and strategies aimed at evaluating the feelings, emotions, and views expressed by users in text and other forms of media. An aspiration to gain a deeper understanding of these perspectives has sparked the development of advanced techniques that focus on analyzing sentiments related to specific product attributes. This progress has resulted in the evolution of the field known as Aspect-Based Sentiment Analysis (ABSA) [13]. On e-commerce platforms, users often provide reviews and feedback on the products they purchase. Through deep learning sentiment analysis, these reviews can be sentiment classified and sentiment intensity analyzed to understand the user's preferences and dissatisfaction with the product. This helps companies discover the advantages and improvement points of their products in a timely manner, so as to improve product quality and user experience.

5.3. The role of sentiment analysis in public opinion monitoring

Public opinion monitoring is the real-time tracking and analysis of the public's sentiment tendency towards an event, topic or brand. Through deep learning sentiment analysis models, it is possible to automate the sentiment analysis of massive news reports, social media content and forum posts to understand the public's attitudes and emotional tendencies towards a specific event or brand. This is of great significance for governments, enterprises and public organizations to help them identify and respond to the public's sentiment needs and concerns in a timely manner.

6. Challenges and expectations of multimodal sentimental fusion

The concept of multimodal sentiment fusion serves as an instrumental method of integrating sentiment information derived from a variety of modalities. A superior approach compared to single-modal sentiment analysis, multimodal sentiment fusion boasts several advantages. Firstly, it offers an abundance of information, as diverse modalities harbor complementary data. When fused, a more comprehensive and in-depth sentiment analysis can be realized. Secondly, multimodal fusion works to eradicate noise and ambiguity present in singular modalities, thereby enhancing the accuracy and reliability of sentiment analysis. Thirdly, this approach procures a more global context, assisting in understanding the backdrop and situation of the sentiment expression more effectively. Lastly, the fusion of multimodal data allows sentiment analysis to better adapt to the individual sentiment expressions of various users, providing a personalized sentiment analysis service.

However, the execution of efficient multimodal sentiment fusion encounters a few challenges. These include data imbalance, where unequal data across modalities may lead to the marginalization or underestimation of some information, thus affecting the fusion output. Modality mismatch, where discrepancies between modalities might affect fusion outcomes, is another issue. Additionally, the extraction of effective features from diverse modal data presents a challenge due to varied representations and feature expressions, making efficient feature fusion an essential topic. Finally, the selection of the appropriate fusion strategies, such as early or late fusion, including the correct attention mechanisms and weight allocation methods, is a key aspect of the process. Looking ahead, multimodal sentiment fusion continues to be a primary focus and challenge within the sentiment analysis realm [14]. For the optimization of multimodal sentiment fusion, in-depth exploration is required in several areas. These areas include cross-modal representation learning to ascertain effective methods to convert varying modal data into a unified feature space for improved feature fusion and sentiment analysis. Optimization of fusion strategies also remains a critical area of focus, studying the strengths and weaknesses of various fusion strategies, and choosing suitable methods for different tasks and data scenarios. In the face of data imbalance, the exploration of multimodal data enhancement techniques can help broaden the dataset and enhance the model's ability to generalize. Further, the examination of the correlation and conversion laws of sentiment expression across different modalities will aid in understanding and fusing multimodal sentiment information more effectively. Lastly, enhancing model interpretability is of the essence to make the prediction and fusion processes of the models easier to comprehend and interpret.

Despite these challenges, sentiment analysis seeks to identify and extract subjective information from text through natural language processing [15]. Addressing these hurdles and constantly advancing the development of multimodal sentiment fusion technology will help deal with the demand for sentiment analysis in multiple contexts, increase the precision and applicability of sentiment analysis, and deliver more valuable information and insights to society and industry.

7. Conclusion

In conclusion, this paper provides a comprehensive review and exploration of the application and development of deep learning in sentiment analysis, with a distinct focus on two trajectories:

multimodal sentiment analysis and multi-context same-behavior sentiment analysis. Initially, it lays a theoretical foundation by defining and classifying emotions and sentiments. Following this, an in-depth examination of the three aspects of unimodal sentiment analysis - text, speech, and image - is conducted. It introduces their respective task definitions, key technologies, and sentiment classification models. Subsequently, the paper places an emphasis on multimodal sentiment analysis, presenting an overview of multimodal datasets, feature representation and fusion methods, and a multimodal analysis model based on deep learning. The significance of multimodal sentiment analysis is then discussed in specific application scenarios, such as social media, product reviews, and public opinion monitoring. Finally, it highlights the key challenges of multimodal sentiment fusion, including data imbalance and feature expression inconsistency, and suggests potential directions for future technical development.

Through this review, it's evident that deep learning has led to significant advancements in the field of sentiment analysis, facilitating complex sentiment analysis across multiple modalities and contexts. However, key issues remain to be addressed, such as the establishment of a sentiment analysis model that can transfer across contexts, the creation of high-quality multimodal sentiment datasets, and the design of improved feature representation and fusion mechanisms. It is anticipated that with the ongoing development of deep learning technology, further breakthroughs and innovations will be realized in the field of sentiment analysis. This will open up more possibilities for understanding and applying sentiment, and cater to a wide range of application scenarios, such as social interaction, human-computer interaction, and business analysis.

References

[1] Hu, M., & Chen, S. (2019). One-Pass Incomplete Multi-View Clustering. Proceedings of the AAAI Conference on Artificial Intelligence, 33(01), 3838-3845.

[2] Abdullah, T., & Ahmet, A. (2023). Deep Learning in Sentiment Analysis: Recent Architectures. ACM Computing Surveys, 55(8), Article 159.

[3] Chan, J.YL., Bea, K.T., Leow, S.M.H. et al. (2023). State of the art: a review of sentiment analysis based on sequential transfer learning. Artificial Intelligence Review, 56, 749–780.

[4] WANG, Y., ZHU, J., WANG, Z., BAI, F., & GONG, J. (2022). Review of applications of natural language processing in text sentiment analysis. Journal of Computer Applications, 42(4), 1011-1020.

[5] Atila, O., Şengür, A. (2021). Attention guided 3D CNN-LSTM model for accurate speech based emotion recognition. Applied Acoustics, 182, 108260.

[6] Kumar, V. S., Pareek, P. K., Costa de Albuquerque, V. H., Khanna, A., Gupta, D., & S, D. R. (2022). Multimodal Sentiment Analysis using Speech Signals with Machine Learning Techniques. 2022 IEEE 2nd Mysore Sub Section International Conference (MysuruCon), 1-8.

[7] Abdaoui, A., Pradel, C., & Sigel, G. (2020). Load What You Need: Smaller Versions of Multilingual BERT. Proceedings of SustaiNLP: Workshop on Simple and Efficient Natural Language Processing, 119–123.

[8] Mocanu, B., Tapu, R., & Zaharia, T. (2023). Multimodal emotion recognition using cross modal audio-video fusion with attention and deep metric learning. Image and Vision Computing, 133, 104676.

[9] Middya, A. I., Nag, B., & Roy, S. (2022). Deep learning based multimodal emotion recognition using model-level fusion of audio–visual modalities. Knowledge-Based Systems, 244, 108580.

[10] Song, T., Zhang, X., Ding, M., Rodriguez-Paton, A., Wang, S., & Wang, G. (2022). DeepFusion: A deep learning based multi-scale feature fusion method for predicting drug-target interactions. Methods, 204, 269-277.

[11] Moshayedi, A. J., Roy, A. S., Kolahdooz, A., & Shuxin, Y. (2022). Deep Learning Application Pros And Cons Over Algorithm. AIRO, EAI.

[12] Jain, P.K., Quamer, W., Saravanan, V. et al. (2023). Employing BERT-DCNN with sentic knowledge base for social media sentiment analysis. Journal of Ambient Intelligence and Humanized Computing, 14, 10417–10429.

[13] D'Aniello, G., Gaeta, M., & La Rocca, I. (2022). KnowMIS-ABSA: an overview and a reference model for applications of sentiment analysis and aspect-based sentiment analysis. Artificial Intelligence Review, 55, 5543–5574.

[14] Gandhi, A., Adhvaryu, K., Poria, S., Cambria, E., & Hussain, A. (2023). Multimodal sentiment analysis: A systematic review of history, datasets, multimodal fusion methods, applications, challenges and future directions. Information Fusion, 91, 424-444.

[15] Wankhade, M., Rao, A.C.S., & Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. Artificial Intelligence Review, 55, 5731–578.