Proceedings of the 4th International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/54/20241506

Research on emotional analysis and integration of multimodal data report

Xiaodi Jin

Artificial Intelligence Application, Shanghai Urban Construction Vocational College, Shanghai, China

Licorice-king@outlook.com

Abstract. In recent years, emotional analysis has remained a hot topic, permeating people's daily lives and playing an indispensable role in various environments. Researchers can improve products and services and forecast user behavior using sentiment analysis to understand consumers' emotional inclinations and reactions. This paper's comprehensive review of emotional analysis is conducted through various research data. This study integrates emotional analysis and multimodal data research, elaborating on emotional analysis's characteristics and pros and cons from aspects such as facial expressions, voice information, and textual data. This paper points out the current applications of emotional analysis and recent research advancements, selecting various features and methods for emotion tendency classification, and utilizing existing technological tools for analysis. Lastly, this work combines multimodal data with various studies to make emotional analysis more widely and aptly applied in everyday life. There are greater opportunities for sentiment analysis in various domains thanks to the advancements in multimodal sentiment analysis research and modal fusion technologies.

Keywords: Emotional analysis, voice, facial expressions, multimodal data.

1. Introduction

Emotional analysis is a task designed to utilize computers to automatically analyze and comprehend people's emotional expressions. It is pivotal in human-computer interaction and criminal investigations [1]. Through emotional analysis, researchers can gain insights into users' emotional inclinations and reactions, enhancing products and services and predicting user behavior. Researchers employ three primary modal data types: facial expression information, voice information, and textual data, for conducting multimodal information analysis, thereby enabling emotional analysis to better assist individuals in their daily lives.

This paper employs a comprehensive integrative approach, researching facial, voice, and textual emotional analysis, providing an overview of various methods and models in emotional analysis. The aim is to integrate diverse modal data and different methods, offering researchers multiple avenues for emotional analysis and sharing some research findings.

© 2024 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/54/20241506

2. Emotional analysis of facial expressions

Facial expression information plays a significant role in emotional analysis, as extracting features from facial expressions can help researchers understand and analyze people's emotions [2]. Researchers employ various feature extraction methods to conduct emotional analysis of facial expressions.

One commonly used feature extraction method involves image processing techniques to extract features from facial expressions. This includes using face detection and facial keypoint localization algorithms to extract facial regions, and then representing emotional states by extracting specific facial features such as lip shape and eye wrinkles.

Another common approach is to utilize deep learning techniques, such as Convolutional Neural Networks (CNN). CNNs can learn to extract abstract features related to facial expressions from raw images, for example, by using pre-trained convolutional neural networks to extract key features from facial expressions [3].

Researchers have made significant progress in emotional analysis based on facial expression information. They are not only focused on associating facial expressions with discrete emotion categories but have also started to pay attention to the continuity and subtle variations in expressions [4].

Some studies have explored facial expression emotional analysis based on video sequences. By analyzing the continuous changes in facial expressions in video, a better understanding of the dynamic changes in people's emotional states can be achieved. Furthermore, some research combines facial expression information with other modal data, such as voice and textual data, to enhance the accuracy and robustness of emotional analysis.

While researchers have progressed in the emotional analysis of facial expressions, they still face some challenges, such as handling facial expressions under different lighting conditions and capturing subtle changes. Future research can further explore these issues and develop more precise and robust methods for emotional analysis based on facial expressions.

3. Emotional analysis based on voice

3.1. Feature extraction methods for voice information

In emotional analysis based on voice, researchers employ various methods to extract features from voice information. One commonly used approach is to extract spectral features from the audio signal, for example, by using Short-Time Fourier Transform (STFT) to convert the audio signal into a spectrogram and then extract features such as frequency and energy from the spectrogram. Additionally, there are methods based on Mel-Frequency Cepstral Coefficients (MFCC), which involve transforming the spectrogram into a Mel-scaled spectrogram and calculating Mel-frequency cepstral coefficients as features. Furthermore, methods based on cepstral features, pitch period features, and others can extract emotion-related information from voice signals.)

3.2. Advances in emotional analysis based on voice information

In research on emotional analysis based on voice, researchers primarily focus on utilizing voice information to identify and classify emotional states [5]. They achieve this by constructing classification models, training, and predicting using machine and deep learning methods. Regarding feature extraction, researchers extract features such as spectral features and Mel-frequency cepstral coefficients from voice signals to transform them into representations suitable for the models to learn from. Subsequently, they use these features to train classifiers to predict emotional states. In recent years, deep learning methods such as Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) have shown promising results in emotional analysis based on voice [6]. These methods can learn higher-level feature representations from voice signals, enhancing the accuracy of emotion classification.

Overall, research in emotional analysis based on voice continues to advance, with improvements in feature extraction methods and classification models enabling researchers to more accurately identify and classify emotional states. However, this technology still faces some challenges, such as handling long time series and issues related to information loss. Future research can explore more effective feature

DOI: 10.54254/2755-2721/54/20241506

extraction methods and model architectures to address these challenges and apply emotional analysis based on voice to more practical scenarios.

4. Feature extraction methods for text-based emotional analysis

In text-based emotional analysis, feature extraction is a crucial step. Feature extraction transforms text into a form that computers can process to further analyze and understand emotions. Common methods for text feature extraction include:

4.1. Bag-of-words model (BoW)

The Bag-of-Words model treats text as a collection of words, ignoring word order and grammar rules, and considering word frequency as features. It generates a vector representation of the text by counting the occurrence of each word or using methods like TF-IDF (Term Frequency-Inverse Document Frequency).

4.2. N-gram model

The N-gram model considers individual words as features and examines combinations of adjacent words. It extracts consecutive n words as features, capturing associations between words [7, 8].

4.3. Word embedding

Word embedding maps each word to a fixed-length vector space, preserving the semantic relationships between words. Common word embedding methods include Word2Vec and BERT, among others.

5. Advances in text-based emotional analysis

In the field of text-based emotional analysis, researchers have made significant progress. They have employed machine learning and deep learning algorithms for text-based sentiment classification and emotion prediction. Some common methods include:

5.1. Traditional machine learning methods

Traditional machine learning techniques like Support Vector Machines (SVM) and Naive Bayes extract text features and train classifiers for sentiment discrimination.

5.2. CNN

CNNs have achieved success in text information processing. They employ one-dimensional convolutional layers to capture local features and utilize pooling layers for feature extraction and dimension reduction.

5.3. RNN

RNNs, with the introduction of a recurrent structure, can capture contextual relationships between text sequences. Variants of RNNs, such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU), help address the issue of information loss in long text sequences.

5.4. Transformer models

Based on self-attention mechanisms, transformer models can capture global semantic relationships and have achieved significant success in natural language processing. BERT (Bidirectional Encoder Representations from Transformers) is one of the pre-training methods based on the Transformer model and has shown excellent performance in sentiment analysis [9].

In summary, research in text-based emotional analysis has made significant progress through continuous improvements in feature extraction methods and the introduction of deep learning models. These methods provide a crucial foundation for multimodal emotional analysis, allowing for integration with results from other modalities, further enhancing the accuracy and comprehensiveness of emotional analysis.

DOI: 10.54254/2755-2721/54/20241506

6. Integration of multimodal data in emotional analysis

In emotional analysis, integrating various modalities of data to obtain more accurate results is an important research direction. Modality fusion techniques are widely applied in multimodal emotional analysis to combine information from different modalities effectively.

Modality fusion techniques combine features or representations from multiple modalities to extract more comprehensive and accurate emotional information. Common modality fusion methods include attention-based methods and tensor-based methods. Attention mechanisms can identify the optimal weights to integrate features from different modalities, highlighting important information and suppressing noise. Tensor-based methods can project features from all modalities into a shared space, facilitating the computation of interactions between modalities.

Significant research achievements have been made in emotion analysis based on multimodal data. By leveraging a variety of modalities, such as facial expression, textual, and voice information, researchers have improved the accuracy and effectiveness of emotion analysis [10]. Innovative feature extraction methods and modality fusion strategies have been proposed, such as methods based on multi-level fusion, which can enhance the quality of single-modal feature vectors. However, it's important to note that excessive modality fusion can lead to overfitting issues in small-sample situations.

Neural network-based multimodal emotional analysis has also made substantial progress, and researchers are delving deeper into research data [11].

7. Experimental results and analysis

In the context of existing research, such as the multimodal feature fusion and word embedding-driven 3D retrieval method developed by Guan Ripeng and his research team, they devised a suitable application method. They employed point cloud and view modality branches to extract and process data, retrieving the 20 most closely matched 3D models, thereby improving accuracy. They optimized the feature fusion scheme of 3D models, further enhancing retrieval precision [12]. This paper summarizes and analyzes the research.

By reviewing different modalities' feature extraction methods and modality fusion techniques, researchers can conclude that multimodal emotional analysis has significantly improved accuracy and overall effectiveness. By combining various modalities such as facial expression, textual, and voice information, researchers can obtain more comprehensive and accurate emotional information.

However, some challenges and issues need to be addressed. For instance, when dealing with long time series, the problem of information loss may arise, necessitating the introduction of suitable techniques to mitigate it. Additionally, attention mechanisms play a crucial role in finding the optimal weight in modality fusion and require further research and improvement.

8. Discussion and outlook

Emotional analysis still faces several challenges and questions. For instance, better utilizing multimodal data for emotional analysis remains a research hotspot. Additionally, the practical applications of emotional analysis are diverse, necessitating further exploration of emotional analysis techniques in various application domains.

Based on the current research progress, future directions may include the following:

1. Enhancing the effectiveness and stability of modality fusion techniques to better leverage multimodal data.
2. Research emotion analysis methods and models tailored to specific application scenarios.
3. Addressing long time-series issues in emotional analysis and proposing more effective methods to resolve the problem of information loss.

In summary, the research progress in multimodal emotional analysis and the development of modality fusion techniques provide researchers with more possibilities to apply emotional analysis in different fields. Further research and innovation will drive the application of emotional analysis in areas such as human-computer interaction and criminal investigations, and provide strong support for better understanding and explaining emotional behavior.

DOI: 10.54254/2755-2721/54/20241506

9. Conclusion

This paper provided a detailed overview of the current research progress in emotional analysis and multimodal data integration. We discussed the advancements in emotional analysis based on facial expression, textual, and voice information and the application of modality fusion techniques in multimodal emotional analysis. Despite the existing challenges and issues, multimodal emotional analysis demonstrates significant potential and has made substantial progress in improving accuracy and effectiveness. Future research can further explore and enhance modality fusion techniques, address emotion analysis issues in specific application scenarios, and resolve challenges related to information loss in handling long time-series data. This will provide valuable support and guidance for researchers to gain a more comprehensive and in-depth understanding of emotional behavior.

References

[1] Jie Y, Qiang Z 2012 Research on emotion analysis and calculation based on human-computer interaction system *Computer Application Research* **29** 7 2763-2765
[2] Xiaojiang P, Yu Q 2020 Progress and challenges in facial expression analysis
[3] Xiao S, Ting P, Fuji R 2016 Facial expression recognition based on ROI-KNN convolutional neural network *Journal of Automation* **42** 6 883-891
[4] Jiahui L, Shumei Z, Junli Z 2020 Research on facial expression recognition based on deep learning *Application Research of Computers/Jisuanji Yingyong Yanjiu* **37** 4
[5] Wenjing H, Haifeng L, Huabin R, Lin M 2013 Review of research progress in speech emotion recognition *Journal of Software* **25** 1 37-50
[6] Fan S, Dan Q, Wenlin Z, Lili Z, Wu G 2017 Low-resource speech recognition method using long short-term memory network *Journal of Xi'an Jiaotong University* 10 120-127
[7] Jifeng L, Qun L 2004 High-speed Chinese character encoding recognition system based on N-Gram model (Doctoral dissertation)
[8] Qiong W, Wenzhen K, Li X 2021 Text error checking algorithm based on improved N-gram model and knowledge base *Computer Applications and Software* **38** 10 310-315
[9] Wenting L, Xinming L 2022 Research progress on Transformer based on computer vision *Journal of Computer Engineering & Applications* **58** 6
[10] Xiaoming Z, Yijiao Y, Shiqing Z 2022 Research progress on multi-modal emotion recognition based on deep learning *Computer Science and Exploration* **16** 7 1479
[11] Minhong L, Zuqiang M 2020 Multimodal sentiment analysis based on attention neural network *Computer Science* **47** S2 508-514
[12] Ripeng G, Liqun K, Shichao J, Fengguang X, Xie H 2023 Three-dimensional retrieval method driven by multi-modal feature fusion and word embedding *Computer Engineering* **49** 4 101-107 113