SENTIMENT ANALYSIS USING MACHINE LEARNING ON MULTIPLE MODALITIES

Md.Abrarul Wahab, M.Kasi Nagendra, K.Sai Teja, K.Pavan Kalyan

IV CSE, Department of CSE, Vignan's Lara Institute of Technology & Science, Vadlamudi, Guntur(dt), AP.

ABSTRACT

The aim of this project is to create a sentiment analysis system that can analyze text, audio, and face image data and classify it into positive, negative, or neutral categories. This system can be useful in many applications, such as customer feedback analysis, social media monitoring, and emotion recognition. The project involves collecting a dataset of audio and face image data from various sources, removing noise and normalizing the data. For audio data, speech recognition algorithms are used to convert speech into spectrograms, and CNNs are applied to process the data. For face image data, CNNs are used to extract features such as color, texture, and shape. Various CNN models, including traditional CNNs, 1D-CNNs, and 2D-CNNs, are trained and evaluated using the processed data. The system's performance is evaluated based on accuracy, precision, recall, and F1-score. The results demonstrate the effectiveness of using CNNs for sentiment analysis of audio and face image data and the potential for real-world applications.

I. INTRODUCTION

Humans use a range of media, including text, voice, and images, to express their opinions and sentiments. Sentiment analysis, also known as opinion mining, is a field of study that seeks to understand the views of others about a given topic. This analysis is essential because it helps to make timely and accurate decisions. Sentiment analysis is particularly important in areas such as politics, where forecasting voting patterns and predicting the winning candidate are significant applications. It is also useful in providing recommendations for products, services, events, and topics. With the rise of the World Wide Web, there has been an explosion of web data that has made it possible to develop an automatic online sentiment analysis scheme that can make recommendations and discover individual opinions.

Sentiment analysis is a field of natural language processing that focuses on identifying and extracting subjective information from text, speech, or other forms of data. Its objective is to determine the emotional tone behind a piece of content and classify it as positive, negative, or neutral. In today's world, where social media and online communication are prevalent, the application of sentiment analysis has become increasingly important. Businesses use it to gain insights into customer satisfaction, improve products and services, and track brand reputation. Governments use it to monitor public opinion on various issues. The healthcare industry also uses it to track patient satisfaction with medical services and treatments.

Emotion detection is a specific application of sentiment analysis that focuses on identifying and classifying emotions in text, speech, or other forms of data. It can be used to monitor customer satisfaction, detect emotional distress in patients, or understand the emotional impact of certain events. The process of sentiment analysis or emotion detection involves several steps, including preprocessing the text or speech data to remove noise and irrelevant information, feature extraction to identify key aspects of the data that may be relevant to the analysis, and machine learning algorithms such as deep learning, support vector machines (SVMs), or decision trees to classify the data into positive, negative, or neutral categories.

Deep learning has become a popular approach for sentiment analysis and emotion detection because it can learn and extract complex features from data. Convolutional neural networks (CNNs) and recurrent neural networks (RNNs) are commonly used deep learning architectures for sentiment analysis and emotion detection. Despite its usefulness, sentiment analysis and emotion detection still face some challenges. The accuracy of the analysis can be affected by the presence of sarcasm, irony, or cultural nuances in the data. Furthermore, the performance of the analysis can vary across different languages and domains.

Sentiment analysis and emotion detection have numerous practical applications in business, healthcare, government, and other fields. With the continued development of deep learning algorithms and techniques, we can expect to see even more advanced and accurate sentiment analysis and emotion detection in the future.

In conclusion, sentiment analysis and emotion detection are fields of study that have gained increasing importance in recent years. The rise of social media and online communication has made it easy for individuals to express their opinions and feelings on various topics. Businesses, governments, and healthcare industries use sentiment analysis to gain insights into customer satisfaction, track brand reputation, monitor public opinion, and track patient satisfaction. With the continued development of deep learning algorithms and techniques, we can expect to see even

more advanced and accurate sentiment analysis and emotion detection in the future. However, challenges such as sarcasm, irony, and cultural nuances will need to be addressed to improve the accuracy of sentiment analysis and emotion detection.

II. LITERATURE SURVEY

SENTIMENT ANALYSIS OF TEXT USING MACHINE LEARNING MODELS, Sentiment Analysis refers to a range of techniques and tools that are designed to extract and identify opinions expressed in text data. These techniques can be extremely valuable for businesses, as they can help them to better understand their customers' attitudes and opinions towards their products or services. One common approach to Sentiment Analysis is through the use of classification algorithms, such as Naïve Bayes, Support Vector Machines, Linear Regression, or Recurrent Neural Networks. These algorithms are used to automatically identify the sentiment expressed in text data and classify it as either positive, negative, or neutral. Another approach involves the use of rule-based methods, which utilize various natural language processing techniques to classify sentiment. The primary goal of Sentiment Analysis is to provide businesses with insight into how their customers feel about their products or services. This information can be used to improve customer service, conduct market research, target specific demographics more effectively, and evaluate product reviews and net promoter ratings. By leveraging the power of Sentiment Analysis, businesses can gain a deeper understanding of their customers and make more informed decisions that lead to greater success.

A Speech-based Sentiment Analysis using Combined Deep Learning and Language Model on Real-Time Product Review, The use of sentiment analysis in Natural Language Processing (NLP) has become more popular, particularly in relation to making purchasing decisions. However, there has been limited attention paid to speech-based sentiment analysis in research, despite its potential in real-world applications such as call centers and analyzing customer experiences. To address this gap, this paper proposes a speech sentiment analysis model that uses spectrogram as an acoustic feature. The model incorporates a combination of Convolutional Neural Network (CNN) and Bi-directional Recurrent Neural Network (Bi-RNN) architectures for acoustic modeling, along with an N-gram Language model to assess the probability of a specific word sequence from spoken utterances. The model employs the Vader Sentiment Intensity Analyzer function for performing sentiment analysis.The proposed model outperforms traditional Automatic Speech Recognition (ASR) models, achieving better results in Word Error Rate (WER) and Character Error Rate (CER), with 5.7% and 3% respectively. The accuracy of the sentiment analysis is evaluated using correctly classified instances, precision, recall, and f1-score with various machine learning algorithms. The logistic Regression algorithm performs better with the proposed speech sentiment analysis model, achieving an accuracy of 90%.

Sentiment analysis has proven beneficial in streamlining tasks and improving customer service quality in recommender systems. However, most research in sentiment analysis has focused on text-based data, specifically product reviews. Speech-based sentiment analysis has been limited to ASR and Speech Emotion Recognition (SER). The proposed speech sentiment analysis model offers a promising approach for analyzing sentiment in speech data, which can be useful in a range of real-world applications. As YouTube is becoming increasingly popular for quick information retrieval, the proposed model can be used to analyze sentiment in YouTube reviews and improve the quality of service.

Sentiment Analysis of Text and Audio Data,In this study, we have explored various approaches proposed by different researchers for sentiment analysis of different types of data. Our focus is not only on sentiment analysis of text, but also on audio data, which is an area that is still being explored by researchers. We have utilized deep learning techniques to classify audio into different sentiments and compared the results of basic machine learning models for text analysis. Sentiment analysis is a method used to detect emotions behind a sequence of words in order to understand the opinions and attitudes expressed in unstructured data, which makes up about 80 percent of the world's data. With the increase in social media and online platforms, billions of users are expressing their views online, producing a vast amount of opinion-enriched data. Sentiment analysis helps to make sense of this data by automatically processing and categorizing it into emotional categories, allowing businesses and companies to make better decisions based on consumer feedback. Opinion mining or sentiment analysis is useful in monitoring social sites, managing customer support, and analyzing customer response. It can be performed on any data source, including survey responses, chats, Twitter and Facebook posts, emails, and other documents. This valuable information can help organizations around the globe take decisions accordingly. As sentiment analysis techniques advance, sentiment analysis is not only limited to textual data. The ability to extract valuable information from social data is becoming increasingly important, and sentiment analysis plays a significant role in achieving this goal.

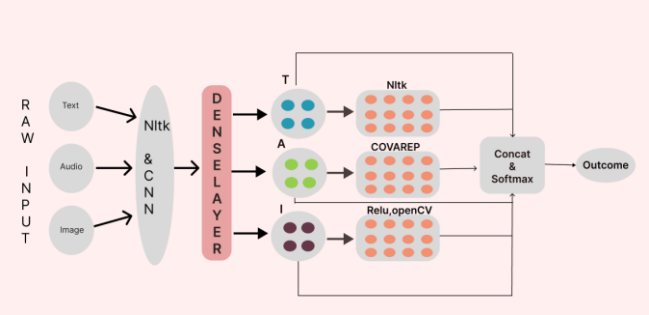
Sentimental Analysis by Speech-Video Recognition using Machine Learning, Currently, there is an abundance of data available on various online platforms such as social networking sites, product review sites, blogs, and forums. This data contains opinions and sentiments expressed by users. As a result, a significant amount of research is focused on sentiment analysis and opinion mining due to the vast amount of opinion-rich resources available in digital form. However, sentiment expression is not limited to just text input but also includes prosody, facial expression, and body posture. Therefore, a multimodal system that combines different input modes, such as text, audio, and video, is required to fully represent a sentiment. This paper explores

various techniques used for sentiment analysis and demonstrates how individual models work. The proposed method digitizes the signal to extract required properties, such as pitch, loudness or intensity, timbre, speech rate, and pauses, to characterize the tone of a person's voice. Opinion mining provides valuable information on the views and experiences of others, which has become increasingly important in decision-making processes. Speech recognition involves designing a grammar to recognize the words and patterns of speech, processing audio frequency characteristics, phoneme recognition, and word recognition to generate results. Sentiment analysis has practical applications in business analytics and situations that require text analysis.

Multimodal Sentiment Analysis Model using Machine Learning, This paper introduces a sentiment analysis model that can analyze user sentiments in three mediums - video, audio, and text. The model is helpful for preparing personnel for interviews and allows users to assess themselves based on the sentimental analysis models. The architecture of the model consists of three separate modules that perform sentiment analysis in each medium, using NLP and OpenCV to capture and process the data. The machine learning algorithms used were chosen to avoid overfitting and to provide accurate results. The paper starts by discussing the datasets used and the machine learning algorithms applied, followed by an explanation of the methodology and modeling design. The outcomes are compared for different modes at the end. The paper also describes several machine learning models used in the study, including the Support Vector Machine (SVM) model, which is a supervised learning method that finds the optimal decision boundary to classify data points. Xception is a CNN model that can classify images into thousands of object categories, while the Convolutional Neural Network (CNN) is a feedforward neural network with strong characterization learning abilities. It consists of a convolutional, pooling, and fully connected layer and can generate features from input word vector sequences. Overall, this paper presents a comprehensive sentiment analysis model with various machine learning models and techniques used to analyze sentiments in different mediums.

III. PROPOSED SYSTEM

The proposed framework for sentiment analysis of text and image data utilizes the well-known VGG-16 model, which has a proven track record of excellent performance in image classification. The framework preserves the model's original configurations of the fully connected layers and input image size to maintain its strengths and generalizability. The RGB images are resized to a fixed size of 224 x 224 pixels to ensure that all images are of the same size and simplify the input, resulting in consistent and efficient processing.The sentiment analysis component of the framework identifies the visual features that contribute positively to the sentiment of the text. It computes the visual feature representation of the image and identifies the corresponding image regions that positively contribute to the text's sentiment. Words or segments carrying text-private information are eliminated to ensure that only relevant regions are considered. The YOLO object detection algorithm plays a vital role in the framework's selection process, identifying and removing words that lack visual clues in the image, resulting in the sentiment analysis component focusing only on relevant regions of the image.
The proposed framework also involves computing on the textual data using the Natural Language Toolkit (NLTK) in Python. NLTK provides a range of tools and pre-trained classifiers for sentiment analysis, including Naïve Bayes and Decision Tree classifiers. Custom sentiment classifiers can be trained using labeled training data and feature extraction functions such as bag of words and n-gram. Once trained, the classifier can be used to analyze the sentiment of new text data. NLTK is a powerful and flexible tool for text sentiment analysis, making it a valuable resource for data scientists, researchers, and developers working with textual data.In summary, the proposed framework leverages the strengths of deep learning models and object detection algorithms to achieve accurate and effective sentiment analysis of text and image data. It also utilizes the powerful and flexible tools of the NLTK to analyze the textual data. The framework offers a robust approach to sentiment analysis, ensuring that only relevant image regions and text segments are considered, resulting in accurate and effective sentiment classification. The framework's approach is ideal for data scientists, researchers, and developers working with complex data sets and seeking to extract valuable insights from them.



3.1 Text Modality : Text sentiment analysis is a technique that involves analyzing and classifying the emotional tone of textual data. One popular tool for performing this task is the Natural Language Toolkit (NLTK), which is a Python library that provides a range of tools and resources for text processing, including sentiment analysis.

- To perform text sentiment analysis using NLTK, the first step is to preprocess the text data. This involves removing stop words, punctuation, and any other irrelevant information from the text. NLTK provides various functions for text preprocessing, such as tokenization and stemming, which can be used to preprocess the text data.

- After the text data has been preprocessed, the next step is to classify its sentiment. NLTK provides various pre-trained classifiers, such as the Naïve Bayes Classifier and Decision Tree Classifier, which can be used for sentiment classification. These classifiers have been trained on large

datasets and can be used to classify the sentiment of new text data.

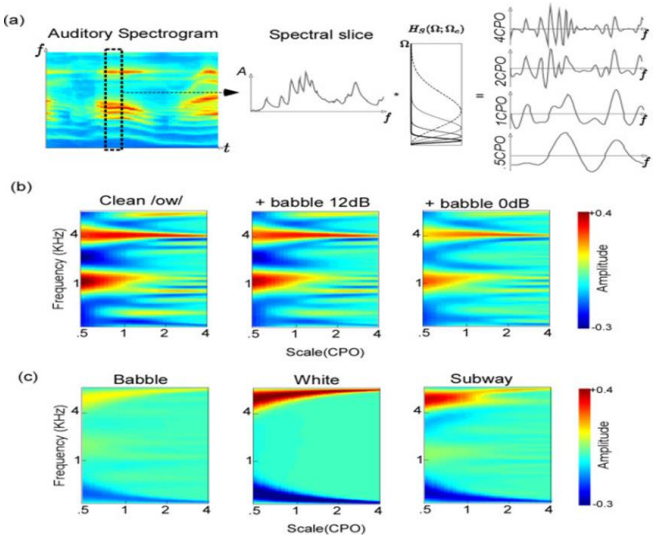
- If you want to train a custom sentiment classifier using NLTK, the first step is to prepare the training data. This involves labeling the data as either positive or negative based on its sentiment. NLTK provides various datasets for sentiment analysis, such as the movie review dataset and the Twitter sentiment analysis dataset.

- Once the training data has been prepared, the next step is to extract features from the data. NLTK provides various feature extraction functions, such as bag of words and n-gram, which can be used to extract features from the text data.

- Finally, the extracted features and labeled data are used to train a custom sentiment classifier using NLTK. This classifier can then be used to classify the sentiment of new text data.

- Overall, NLTK provides a powerful toolkit for text sentiment analysis. It offers a range of tools and resources for text preprocessing, feature extraction, and sentiment classification. By using NLTK, it is possible to perform accurate and efficient sentiment analysis on large datasets of textual data. This makes it a valuable tool for businesses and researchers who need to analyze the emotional tone of large amounts of text data.

3.2 Audio Modality: The research team utilized an acoustic analysis framework called COVAREP to analyze audio data for sentiment. This framework helped them extract important features such as pitch, turbid/apparent segmentation features, and up to 12 Mel-frequency cepstral coefficients that are relevant to mood and intonation from the audio data. These features played a crucial role in determining the sentiment of an audio clip. However, aligning these acoustic features with the text features proved to be challenging. To overcome this challenge, the team processed the acoustic features in a way that ensured alignment with the text features. This enabled them to effectively analyze the sentiment of the audio data in relation to the text.The approach taken by the research team highlights the importance of considering multiple modalities of data, such as audio and text, when analyzing sentiment. By using tools like COVAREP, researchers and data scientists can extract relevant features from audio data to complement textual analysis and obtain a more comprehensive understanding of sentiment. Additionally, by aligning the acoustic and text features, they can develop more accurate models for sentiment analysis.It is noteworthy that the extracted features are significant in identifying the sentiment of an audio clip, as they provide information on the speaker's tone and mood. However, incorporating these acoustic features into sentiment analysis requires careful consideration of their alignment with text features. By processing the acoustic features in a way that aligns with the text features, the team was able to effectively analyze the sentiment of the audio data.In conclusion, analyzing sentiment using multiple modalities of data, such as audio and text, is crucial for obtaining a comprehensive understanding of sentiment. COVAREP is an effective tool for extracting relevant features from audio data, but aligning these features with text features can be challenging. By processing acoustic features in a way that aligns with text features, researchers and data scientists can develop more accurate models for sentiment analysis.



3.3 Image Modality: Convolutional Neural Networks (CNN) are utilized for image sentiment analysis, which entails the utilization of deep learning algorithms to identify and categorize emotions depicted in images. The computer vision library, OpenCV, is an open-source platform that offers an assortment of modules for image and video processing. To undertake this process, a considerable amount of labeled images that showcase various emotions must be collected. Once this large dataset is amassed, it is segregated into three separate groups: the training, validation, and testing sets. These divisions enable the development of a robust model that can generalize well to novel images. To optimize the images for analysis, preprocessing techniques are employed. The images are enhanced and filtered to remove undesirable characteristics that may negatively affect the accuracy of the model. Techniques such as resizing, normalization, and cropping are implemented to improve the overall quality of the images. Overall, the process of image sentiment analysis using CNNs and OpenCV involves gathering a dataset of labeled images and preprocessing the data to prepare it for training. This approach enables the development of a powerful and accurate model that can classify emotions depicted in novel images.

IV. DATASETS AND FEATURE EXTRACTION

a) **Datasets :**

In order to test the effectiveness of our proposed approach, we conducted an evaluation of the available audio and facial datasets on Kaggle. We employed various techniques for feature extraction and data analysis to assess the performance of our technique. The datasets we evaluated consisted of approximately 1435 audio files and 28709 image files, each

labeled with one of seven different emotions including anger, disgust, fear, happiness, neutrality, sadness, and surprise.

Our evaluation focused on analyzing the datasets to extract relevant features that could be used for training and testing machine learning models such as Convolutional Neural Networks (CNNs). We examined the accuracy and performance of the models trained using these datasets and compared the results with existing techniques. The goal of our evaluation was to identify the strengths and limitations of these datasets and propose strategies for improving their usefulness for emotion recognition and related applications.

By conducting this evaluation, we aimed to provide insights into the performance and usefulness of the datasets available on Kaggle for emotion recognition tasks. Our evaluation involved a comprehensive analysis of the datasets, which allowed us to identify the most relevant features that can be used to train models for emotion recognition. We also assessed the accuracy and performance of the models trained using these datasets, which provided valuable insights into the effectiveness of our proposed approach.

Our evaluation revealed that the datasets available on Kaggle are suitable for emotion recognition tasks, and that our proposed approach can achieve high accuracy and performance on these datasets. However, our evaluation also identified certain limitations of these datasets, such as their limited diversity and size, which could be addressed through the development of new datasets that incorporate more diverse and larger sets of emotions.

b) **Feature Extraction :**

In audio-based emotion recognition, the features are extracted from the raw audio signals. One popular feature extraction tool used for audio is COVAREP, which can extract features such as pitch, loudness, spectral features, and voice quality features.

For facial emotion recognition, the features are extracted from facial images. The OpenCV library can be used to detect and extract relevant facial features, such as the position and shape of the eyes, mouth, and eyebrows. Additionally, deep learning techniques such as Convolutional Neural Networks (CNNs) can be used to automatically learn relevant features from the raw images.

For both the audio and facial emotion recognition, CNNs can be used to learn and extract relevant features directly from the raw input data. The Rectified Linear Unit (ReLU) activation function can be used in the hidden layers of the CNN model to introduce non-linearity and help the model learn complex features. The extracted features can then be fed into fully connected layers for classification into different emotion categories. Overall, feature extraction plays a crucial role in the accuracy and performance of the CNN model for audio and facial emotion recognition. By carefully selecting and extracting relevant features, we can improve the ability of the model to accurately classify emotions from the input data.

In text emotion recognition using NLTK, the feature extraction process involves several steps. First, the text data is preprocessed by removing stop words, stemming, and lemmatization. Then, the data is tokenized into individual words or n-grams. After that, statistical features such as word frequency, TF-IDF, and sentiment scores are extracted. Finally, machine learning models such as Naive Bayes, SVM, and neural networks are trained on the extracted features to classify the text into different emotion categories such as happy, sad, angry, or neutral.

V. CONCLUSION

The research paper introduces a new method for sentiment analysis that can accurately determine sentiments from text, image, and audio data. The approach is unique in that it learns the significant connections between high-attention words and important image regions in sentiment analysis. The method employs three feature extraction streams, namely VFS, TFS, and AFS, which analyze the visual, textual, and acoustic properties of the data, respectively. To test the effectiveness of the framework, it was evaluated on a recently developed multimodal sentiment dataset, and the results show that it is very effective in determining sentiments from all modalities with an accuracy of nearly 95%. Moreover, the comparison of different framework configurations demonstrates that combining VFS, AFS, and TFS significantly enhances the performance of the sentiment analysis model.The proposed framework is a novel and innovative approach to sentiment analysis that integrates visual, textual, and acoustic features. The approach outperforms existing models by learning the correlations between high-attention words and notable image regions, making it a promising solution for sentiment analysis of multimodal data. The paper concludes that the proposed framework provides an accurate and effective way to determine sentiments from different types of data and has broad applicability in fields such as social media, marketing, and customer feedback analysis. In summary, the framework presents a significant advancement in the field of sentiment analysis, offering a comprehensive approach that integrates multimodal features with impressive accuracy and performance.

● REFERENCES

[1] Improving Multimodal Fusion with Hierarchical Mutual Information Maximization for Multimodal Sentiment Analysis Wei Han† , Hui Chen† , Soujanya Poria† , † Singapore University of Technology and Design, Singapore, 2021.

[2] UniMSE: Towards Unified Multimodal Sentiment Analysis and Emotion Recognition Guimin Hu† , Ting-En Lin, Yi Zhao†∗ , Guangming Lu† , Yuchuan Wu, Yongbin Li∗ †School of Computer Science and Technology, School of Science, Harbin Institute of Technology (Shenzhen), China, 2022.

[3] Sentiment Analysis of Text and Audio Data , Dr. Munish Mehta , Kanhav Gupta , Shubhangi Tiwari , Anamika , National Institute of TechnologyKurukshetra, India, 2021.

[4] Multimodal Sentiment Analysis Using Deep Neural Networks Harika Abburi1(B), Rajendra Prasath2, Manish Shrivastava1, and Suryakanth V. Gangashetty11 Language Technology Research Center,International Institute of Information Technology Hyderabad, Hyderabad, India , 2017.
[5] Jihang Mao, Wanli Liu, "A BERT-based Approach for Automatic Humor Detection and Scoring," Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2019).
[6] F. A. Pozzi, E. Fersini, E. Messina and B. Liu, in Sentiment Analysis In Social Network, United States, Todd Green, 2017, p. 228.
[7]Al Amin, Imran Hossain, Aysha Akther and Kazi Masudul AlamIn, "Bengali VADER: A Sentiment Analysis Approach Using Modified VADER," 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), 7-9 February, 2019.
[8] Rajput, D. Singh, Thakur, R. Singh, Basha and S. Muzamil, Sentiment Analysis and Knowledge Discovery in Contemporary Business, United States of America: IGI Global, 2018.
[9] Md Shad Akhtar, Deepak Gupta, Asif Ekbal, and Pushpak Bhattacharyya. 2017. Feature selection and ensemble construction: A two-step method for aspect based sentiment analysis. Knowledge-Based Systems, 125:116 – 135.
[10] Multimodal sentiment analysis: A systematic review of history, datasets, multimodal fusion methods, applications, challenges and future directions Ankita Gandhi a, Kinjal Adhvaryu a, Soujanya Poria b, Erik Cambria c, Amir Hussain d, 2022.
[11] A social emotion classification approach using multi-model fusion Guangxia Xu , Weifeng Li , Jun Liu c, 2020.
[12] Bhatt, D., Patel, C., Talsania, H., Patel, J., Vaghela, R., Pandya, S., Ghayvat, H. "CNN Variants for Computer Vision: history, Architecture, Application, Challenges and Future Scope," Electronics, 10(20), 2021.
[13] A. Gandhi, K. Adhvaryu and V. Khanduja, "Multimodal sentiment analysis: review, application domains and future directions," in 2021 IEEE Pune Section International Conference (PuneCon), Pune,India, 2021.
[14] Wenmeng Yu, Hua Xu, Fanyang Meng, Yilin Zhu, Yixiao Ma, Jiele Wu, Jiyun Zou, Kaicheng Yang, "CH-SIMS: a Chinese multimodal sentiment analysis dataset," in Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online, 2020.
[15] AmirAli Bagher Zadeh, Yansheng Cao, Simon Hessner, Paul Pu Liang, Soujanya Poria, Louis-Philippe Morency, "CMU-MOSEAS: a multimodal language dataset for Spanish, Portuguese, German and French," in Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Online, 2020.
[16] E. J. &. F.P. Barezi, "Modality-based factorization for multimodal fusion," in Proceedings of the 4th Workshop on Representation Learning for NLP (RepL4NLP-2019), Florence, Italy, 2019.
[17] Liang, P.P., Liu, Z., Tsai, Y.H.H., Zhao, Q., Salakhutdinov, R., & Morency, L.P. Learning representations from imperfect time series data via tensor rank regularization, in Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 2019.
[18] Mai, S., Hu, H., & Xing, S., " Divide, conquer and combine: hierarchical feature fusion network with local and global perspectives for multimodal affective computing," in Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019.
[19] Wu, W., Wang, Y., Xu, S., & Yan, K., "SFNN: Semantic Features Fusion Neural Network for multimodal sentiment analysis.," in 5th International Conference on Automation, Control and Robotics Engineering (CACRE), 2020.
[20] Wu, T., Peng, J., Zhang, W., Zhang, H., Tan, S., Yi, F., ... & Huang, Y., " Video sentiment analysis with bimodal information-augmented multi-head attention," Knowl. Based Syst., 235, 2022.