WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Emotion Recognition in Deep Learning Across Diverse Speech and Text Corpora

Makendran R[1], Balaji K[2], Barath Raj S[3], Boobesh S[4],Gowtham K[5]

[1]Assistant Professor,[2345]student, 8th semester B.E. Computer Science and Engineering, Dhirajlal Gandhi College of Technology

ABSTRACT

The proposed a deep learning model of convolutional neural networks (CNNs) and Long Short Term Memory(LSTMs) to capture both local and temporal context information. We collect and annotate diverse speech and text corpora from several sources, including public databases and social media platforms. Experimental results demonstrate the effectiveness of the proposed model in emotion recognition tasks on the collected corpora. The study also compares the performance of different deep learning architectures and analyzes the impact of feature selection and data augmentation techniques. The findings highlight the potential of deep learning for emotion recognition across multiple modalities and emphasize the importance of diverse corpora for training robust models. Our research sheds light on the nuanced emotional cues present indiverse linguistic expressions

INTRODUCTION

Emotions are important in human communication because they influence our decisions, interactions, and overall well-being. As deep learning continues to be integrated into diverse applications, the significance of comprehending human emotions has grown, ensuring the development of technologies that are empathetic and responsive. Many studies concentrate solely on developing and testing models using specific corpus datasets, overlooking the drawbacks associated with this approach. One significant limitation is the lack of diversity inherent in a single dataset. In real-world situations, emotions exhibit a wide range of variations influenced by cultural, regional, and individual differences. Unfortunately, a solitary dataset often fails to capture this complexity, limiting the model's understanding of diverse motional expressions.

Factors in Emotion Recognition with Deep Learning Models Using Speech and Text on Multiple Corpora

In this study, the speech encoder employs a hybrid approach, integrating Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and a series of fully connected layers augmented by an attention mechanism. On the other hand, the text encoder follows contemporary transfer learning techniques, utilizing a pre trained BERT model coupled with fine-tuning strategies. The effectiveness of these models is initially assessed using the IEMOCAP corpus, enabling direct comparisons with methods described in existing literature.

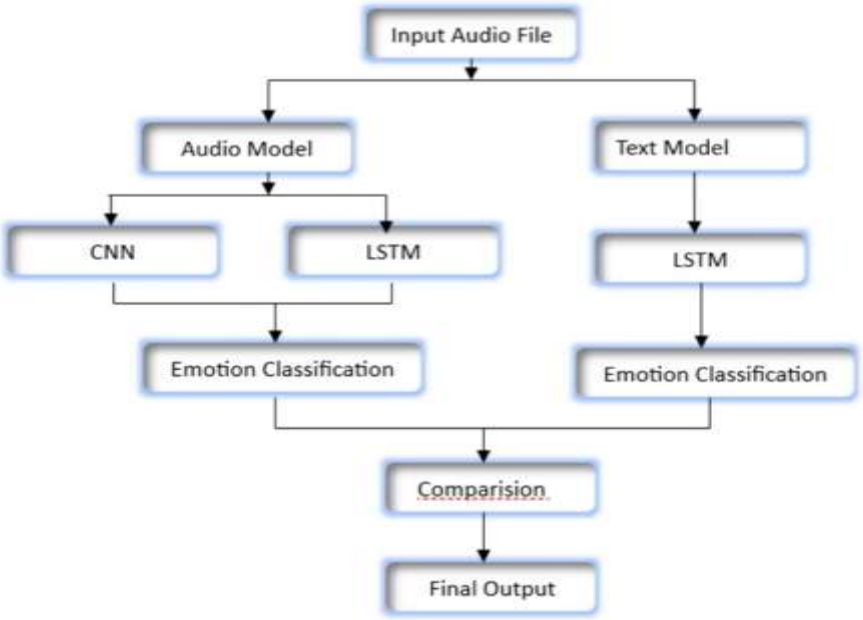
Deep Learning-Based Speech Emotion Recognition Using Multi-Level Fusion of Concurrent Features

Detecting and categorizing emotional states in speech requires analyzing both audio signals and textual transcriptions. Examining the extracted features over various time interval sis crucial, as these features exhibit intricate relationships that are vital for inferring emotion sex pressed in speech. These relationships can be categorized as spatial, temporal, and semantic tendencies. Besides the emotional attributes present in each modality, the textual modern compasses semantic and grammatical inclinations within the spoken sentences.

Speech Emotion Recognition Based on Transfer Emotion-Discriminative Features Subspace Learning

Initially, acoustic features are extracted using Open SMILE from both the source and target datasets. These features are then processed through a CNN+BLSTM architecture to capture higher-level global features and time series patterns. To bridge the differences between the source and target data, we employ Linear Discriminant Analysis (LDA) to find a common low-dimensional subspace. Additionally, Maximum Mean Discrepancy (MMD) and Graph Embedding (GE) techniques are utilized to further align the datasets.

SYSTEM ARCHITECTURE



METHODOLOGY

EXISTING SYSTEM

The existing approach often fails to recognize complex and subtle emotions. Due to their diversity and multidimensionality, deep learning algorithms are currently unable to fully represent and classify the complexity of emotions. This limitation might lead to inaccurate classification of emotions or make it challenging to recognize emotions that do not fit into the predetermined categories, which would reduce the overall accuracy and reliability of the system. The current method could be vulnerable to noise and variations in speech and text corpora

PROPOSED SYSTEM

The main goal of the proposed research is to use cutting-edge deep learning algorithm son a variety of voice and text datasets to address the challenging job of emotion recognizing. The capacity of a computer system to accurately identify and understand human emotions from a variety of input sources, including speech and text, is known as emotion recognition, and it is still a difficult task. Deep learning is a branch of machine learning that draws inspiration from the neural networks found in the human brain. It has shown promise in a number of fields ,such as voice and picture recognition. However, because emotions are subjective and there is no standardized vocabulary for emotions, the field of emotion identification presents particular difficulties.

RESULT

The primary objective of the deep learning-based emotion detection system is to increase the accuracy of emotion identification across many sources over a range of speech and text corpora. This system aims to surpass existing models by addressing problems caused by linguistic variations, cultural disparities, and environmental subtleties impacting emotion identification. The system uses deep learning architectures like Long Short-Term Memory (LSTM) networks and convolutional neural networks (CNNs) to efficiently extract complicated patterns from emotional input. The ability to adjust to many languages, dialects, and emotional expressions is guaranteed by extensive and varied voice and text training datasets. The quality of input data may be increased by applying preprocessing methods including feature extraction, standardization, and dimensionality reduction. The application of information from adjacent areas to enhance generalization and accuracy the area of emotion recognition is made possible by transfer learning and multi-task learning frameworks.

CONCLUSION

To sum up, the deep learning system created to identify emotions in a variety of voice and text datasets has great potential for developing the area of emotion analysis. Through the use of advanced deep learning algorithms, the system is able to extract and analyze emotional aspects from a variety of

sources, including text and audio. This skill overcomes the drawbacks of unimodal techniques and advances a more thorough knowledge of human emotions.

REFERENCE

[1] Norbert Braunschweiler , Rama Doddipatla , Member, IEEE, Simon Keizer, Member, IEEE,and Svetlana Stoyanchev, Member, IEEE"Factors in Emotion Recognition With Deep LearningModels Using Speech and Text on Multiple Corpora" , IEEE signal processsing letters, Vol.29, 2022.

[2] Pansy Nandwani1 · Rupali Verma1"A review on sentiment analysis and emotion detectionfrom text",IEEE,28 Aug 2021.

[3] Samuel Kakuba 1,2, Alwin Poulose 3, Dong Seog Han 4,"Deep Learning-Based SpeechEmotion Recognition Using Multi-Level Fusion of Concurrent Features",IEEE Vol.10, 2022.

[4] Lim-Min Zhang 1,2, Giap Weng Ng 2, Yu-Beng Leau 2, (Senior Member, IEEE), and HaoYan 1"A Parallel-Model Speech Emotion Recognition Network Based on Feature Clustering"IEEE Vol 11,2023.

[5] Zhang Kexin and Liu Yunxiang "Speech Emotion Recognition Based on Transfer Emotion Discriminative Features Subspace Learning" IEEE Vol 11,2023.