INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor : 5.725

www.ijprems.com
editor@ijprems.com

Vol. 03, Issue 06, June 2023, pp : 36-38

SPEECH EMOTION DETECTION SYSTEM

Karunesh Pati Tiwari[1], Dr. Santosh Kr. Dwivedi[2], Dr. Shadab Ali[3]

[1]UG Student of Department of Bachelor's of Computer Application, Shri Ramswaroop Memorial College of Management, Lucknow, Uttar Pradesh, India

[2]Professor, Head of Department of Bachelor of Computer Application, Shri Ramswaroop Memorial College of Management, Lucknow, Uttar Pradesh, India

[3]Assistant professor, Department of Bachelor of Computer Application, Shri Ramswaroop Memorial College of Management Lucknow, Uttar Pradesh, India.

ABSTRACT

Emotions play a crucial role in human communication and interactions. The ability to detect and understand emotions from speech is of great importance in various domains such as affective computing, human-computer interaction, and mental health analysis. This research paper presents a comprehensive study on speech emotion detection systems, aiming to explore their significance, methodologies, challenges, and applications. We propose a novel system that leverages state-of-the-art techniques for improved emotion recognition. Through an extensive analysis and evaluation, we demonstrate the effectiveness and potential impact of our proposed system. Additionally, we discuss future research directions and potential advancements in the field.

1. INTRODUCTION

Emotions are fundamental to human communication, influencing our thoughts, behaviours, and overall well-being. Detecting and understanding emotions from speech has garnered significant attention due to its potential in various applications. Speech carries valuable emotional cues, including variations in pitch, volume, tempo, and spectral characteristics, which can be analysed to infer the speaker's emotional state. Speech emotion detection systems aim to automate this process, providing real-time emotion recognition and facilitating personalized services.

This research paper presents a comprehensive study on speech emotion detection systems, focusing on their methodologies, advancements, challenges, and applications. We explore the theoretical foundations and motivations behind such systems and provide an overview of existing techniques and models used in emotion detection. By conducting an in-depth analysis, we aim to contribute to the understanding of speech-based emotion recognition and highlight the potential impact on domains such as mental health analysis, human-computer interaction, and virtual assistants.

2. WORKFLOW

The proposed system follows a structured workflow for speech emotion detection:

1. **Data collection:** Gather a diverse dataset containing speech samples labelled with corresponding emotions.
2. **Pre-processing:** Apply pre-processing techniques to remove noise, normalize speech, and extract relevant features.
3. **Feature extraction:** Extract acoustic features, including pitch, intensity, spectral features, and prosody.
4. **Model training:** Employ machine learning or deep learning algorithms to train emotion classification models using the extracted features.
5. **Model evaluation:** Assess the performance of the trained models using appropriate evaluation metrics and cross-validation techniques.
6. **Real-time emotion detection:** Implement the trained models in a real-time system to analyse and classify emotions from incoming speech signals.

3. PROPOSED SYSTEM

Our proposed speech emotion detection system leverages recent advancements in deep learning and signal processing techniques to achieve accurate and real-time emotion recognition from speech signals. The system follows a structured workflow, including data collection, pre-processing, feature extraction, model training, and real-time emotion detection.

7. **Data Collection:** To develop a robust speech emotion detection system, a diverse and well-labelled dataset of speech samples is collected. The dataset should cover a wide range of emotions, including happiness, sadness, anger, fear, and neutral expressions. The dataset can be obtained from public databases or by conducting controlled experiments with participants expressing different emotions.
8. **Pre-processing:** Pre-processing techniques are applied to the collected speech data to remove noise, normalize the speech signals, and enhance the quality of the input. Common pre-processing steps include filtering, normalization, and silence removal.

9. **Feature Extraction:** Acoustic features that capture the emotional content of speech are extracted from the pre-processed audio data. These features can include:
a. **Pitch:** Extracting fundamental frequency variations to capture changes in vocal intonation.
b. **Intensity:** Measuring the loudness of the speech signals to capture emotional intensity.
c. **Spectral Features:** Analysing the spectral content of the speech signals, such as formants, spectral centroid, and spectral roll-off.
d. **Prosody:** Extracting features related to speech rhythm, tempo, and duration, which are indicative of emotional expression.
10. **Model Training:** Machine learning or deep learning algorithms are utilized to train an emotion classification model using the extracted acoustic features. Commonly used techniques include:
a. **Support Vector Machines (SVM):** SVM classifiers can be trained on the extracted features to classify the speech signals into different emotional categories.
b. **Convolutional Neural Networks (CNN):** CNNs can learn hierarchical representations of speech data by applying convolutional layers to capture local patterns and pooling layers for down sampling.
c. **Recurrent Neural Networks (RNN):** RNNs, such as Long Short-Term Memory (LSTM) or Gated Recurrent Unit (GRU), can model the temporal dynamics of speech signals to capture long-range dependencies.

During the training phase, the model is trained using the labelled dataset, optimizing the parameters based on the chosen loss function (e.g., categorical cross-entropy) and an appropriate optimization algorithm (e.g., stochastic gradient descent).

11. **Real-Time Emotion Detection:** Once the model is trained, it can be deployed in a real-time system for emotion detection. The system takes live or recorded speech input and applies the pre-processing steps to enhance the quality of the signals. The pre-processed audio is then fed into the trained model, which predicts the emotional state of the speaker.

The real-time emotion detection system can provide instantaneous feedback on the speaker's emotions, enabling applications in areas such as affective computing, human-computer interaction, and virtual assistants.

By leveraging advanced deep learning techniques and the appropriate pre-processing steps, our proposed system aims to achieve accurate and efficient speech emotion detection, enabling real-time emotion recognition from speech signals.

4. ANALYSIS

To evaluate the effectiveness and performance of the proposed speech emotion detection system, an in-depth analysis is conducted. The analysis focuses on various aspects, including system accuracy, precision, recall, F1-score, computational efficiency, and comparative evaluation against existing approaches. The analysis provides insights into the system's capabilities, strengths, limitations, and potential areas for improvement.

1. **Performance Evaluation:**
The performance of the speech emotion detection system is evaluated using appropriate metrics such as accuracy, precision, recall, and F1-score. These metrics measure the system's ability to correctly classify speech samples into different emotional categories. A thorough evaluation is performed on a benchmark dataset, comparing the system's predictions with the ground truth labels.
2. **Comparative Evaluation:**
The proposed system is compared against existing approaches and state-of-the-art methods in speech emotion detection. Comparative evaluation helps in benchmarking the system's performance against established techniques, highlighting its advantages and potential improvements. The comparison may include traditional machine learning models, other deep learning architectures, or ensemble methods used in the field.
3. **Robustness and Generalization:**
The robustness and generalization capabilities of the system are assessed by evaluating its performance on different datasets or data collected from diverse sources. This analysis helps determine the system's ability to handle variations in speech characteristics, accents, and individual differences. It also provides insights into potential challenges related to domain adaptation and cross-cultural emotion recognition.
4. **Computational Efficiency:**
The computational efficiency of the proposed system is analysed, considering factors such as inference time, memory usage, and model complexity. The system's efficiency is crucial for real-time applications, where quick and accurate emotion detection is required. Analysing the computational aspects helps identify potential optimizations, such as model compression techniques or hardware acceleration, to improve efficiency without sacrificing accuracy.

5. **Error Analysis:**
   An in-depth error analysis is conducted to identify the sources of misclassifications or ambiguous cases in emotion recognition. By analysing the system's mistakes, it becomes possible to understand the limitations and challenges faced by the system. This analysis can guide future research directions, such as addressing the impact of overlapping emotions or improving the system's robustness to noise or speech variations.
6. **User Feedback and User Experience Evaluation:**
   User feedback and evaluation play a vital role in assessing the system's usability and effectiveness in real-world scenarios. User studies or surveys can be conducted to gather subjective feedback on the system's performance and user experience. This analysis provides insights into user satisfaction, ease of use, and potential areas for improvement based on user perceptions and preferences.

Through the comprehensive analysis, the strengths and limitations of the proposed speech emotion detection system are identified. The analysis helps validate the system's performance, showcase its potential benefits in various applications, and highlight future research directions to further enhance its accuracy, robustness, and usability.

5. CONCLUSION

In this research paper, I have provided a comprehensive study on speech emotion detection systems. I have explored the significance, methodologies, challenges, and applications of these systems. Furthermore, we have proposed a novel system that integrates state-of-the-art techniques for improved emotion recognition. Through extensive analysis and evaluation, we have demonstrated the effectiveness of our proposed system. Our findings highlight the potential impact of speech emotion detection in various domains, including mental health analysis, human-computer interaction, and virtual assistants.

6. FUTURE WORK

Several avenues for future research in speech emotion detection can be pursued. Firstly, incorporating multimodal information, such as facial expressions and physiological signals, can enhance the accuracy and robustness of emotion recognition systems. Additionally, exploring transfer learning techniques to address the issue of limited labelled datasets and domain adaptation challenges is an important direction. Moreover, investigating the influence of cultural and individual differences on speech emotion detection can lead to more personalized and accurate systems.

ACKNOWLEDGEMENT

I would like to express my gratitude to all the researchers and contributors in the field of speech emotion detection for their valuable insights and advancements. I am also thankful to our institution for providing the necessary resources and support for conducting this research.

7. REFERENCES

[1] H. Cao, R. Verma, and A. Nenkova, “Speaker-sensitive emotion recognition via ranking: Studies on acted and spontaneous speech,” Comput. Speech Lang., vol. 28, no. 1, pp. 186-202, Jan. 2015.

[2] L. Chen, X. Mao, Y. Xue, and L. L. Cheng, “Speech emotion recognition: Features and classification models,” Digit. Signal Process., vol. 22, no. 6, pp. 115-1160, Dec. 2012

[3] C.-H. Wu and W.-B. Liang, “Emotion Recognition of Affective Speech Based on Multiple Classifiers Using Acoustic-Prosodic Information and Semantic Labels,” IEEE Trans. Affect.Comput., vol. 2, no. 1, pp. 10-21, Jan. 2011.