e-ISSN: 2582-5208

International Research Journal of Modernization in Engineering Technology and Science

(Peer-Reviewed, Open Access, Fully Refereed International Journal)

Volume:05/Issue:10/October-2023 Impact Factor- 7.868 www.irjmets.com

SPEECH EMOTION RECOGNITION USING HIDDEN MARKOV MODELS

Vivek Kripashankar Paswan[*1]

[*1]UG Department Of Information Technology B.K Birla College Kalyan,
(Empowered Autonomous Status), India.

ABSTRACT

Speech Emotion Recognition (SER) plays a pivotal role in enhancing human-computer interaction by enabling machines to understand and respond to the emotional content embedded in spoken language. This study presents an approach to SER utilizing Hidden Markov Models (HMMs). The proposed framework leverages the temporal dynamics of speech signals, employing Mel Frequency Cepstral Coefficients (MFCCs) as feature vectors. The model is trained in a supervised manner, associating emotional states with HMM states and estimating transition probabilities. Gaussian Mixture Models (GMMs) capture the emission probabilities of HMM states, enhancing the model's ability to discriminate between emotional states. Recognition is achieved through the Viterbi algorithm, which identifies the most likely sequence of emotional states given observed features. The application of HMMs in SER offers a nuanced understanding of emotional expression in speech, contributing to the development of emotionally intelligent systems. Experimental results demonstrate the effectiveness of the proposed approach in capturing the intricate patterns of emotional dynamics within speech signals.

I. INTRODUCTION

Speech, as a primary mode of human communication, is laden with emotional cues that convey a wealth of information beyond mere words. Recognizing and understanding these emotional nuances in spoken language has become a key pursuit in the field of human-computer interaction, with applications ranging from virtual assistants to sentiment analysis in customer service. One significant facet of this endeavor is Speech Emotion Recognition (SER), a domain aimed at endowing machines with the ability to decipher and respond appropriately to the emotional states conveyed through speech.

In recent years, various methodologies have been explored to imbue machines with emotional intelligence, enabling them to navigate the complex landscape of human emotions. Among these methodologies, Hidden Markov Models (HMMs) have emerged as a promising tool for modeling the temporal dynamics inherent in speech signals. HMMs provide a structured framework to represent the evolving emotional states within a speech signal, offering a dynamic perspective on the interplay of emotions over time.

This paper delves into the realm of SER using HMMs, where the focus is on capturing the intricate patterns and transitions that characterize the emotional content of speech. By employing Mel Frequency Cepstral Coefficients (MFCCs) as feature vectors and integrating Gaussian Mixture Models (GMMs) for emission probability estimation, the proposed approach seeks to enhance the discernment of emotional states within the speech signal. Through a supervised training process and the application of the Viterbi algorithm for recognition, the model aims to decode the emotional narrative embedded in spoken language.

As we explore the integration of HMMs in SER, this study contributes to the broader quest for human-like interaction between machines and users, enriching the communicative capabilities of artificial systems and fostering a deeper understanding of the emotional dimensions of human speech.

II. LITERATURE REVIEW

Understanding and recognizing emotions in speech has garnered substantial attention in the fields of signal processing, machine learning, and human-computer interaction. A significant body of research has explored various methodologies, with Hidden Markov Models (HMMs) emerging as a promising approach due to their ability to capture the temporal dynamics inherent in speech signals.

1. Early Approaches to Speech Emotion Recognition

- Early attempts at Speech Emotion Recognition focused on static features such as pitch, intensity, and formants. However, these approaches often struggled to capture the dynamic nature of emotional expression in speech.

2. Introduction of Hidden Markov Models

- The incorporation of HMMs marked a paradigm shift in Speech Emotion Recognition. HMMs provided a natural framework for modeling sequential data, allowing researchers to represent emotional states as hidden states and transitions between states.

3. Feature Representation with MFCCs

- Mel Frequency Cepstral Coefficients (MFCCs) have become a staple in feature extraction for speech processing, including emotion recognition. They capture the spectral characteristics of speech signals and serve as a foundation for many HMM-based models.

4. Supervised Training and Gaussian Mixture Models

- Many studies employ supervised training techniques to associate emotional labels with HMM states. Gaussian Mixture Models (GMMs) are commonly used to model the emission probabilities of HMM states, providing a robust representation of the distribution of speech features.

5. Temporal Dynamics and State Transitions

- One of the key strengths of HMMs lies in their ability to model temporal dependencies. Researchers have explored different strategies for representing and modeling the transitions between emotional states, refining the accuracy of emotion recognition systems.

6. Challenges and Advancements

- Challenges in Speech Emotion Recognition using HMMs include the variability of emotional expression across speakers and contexts. Recent advancements have focused on adapting HMMs to handle these variations, incorporating deep learning techniques to improve the robustness and generalization of models.

7. Applications and Future Directions

- Beyond fundamental research, HMM-based Speech Emotion Recognition finds applications in diverse fields, including human-computer interaction, virtual agents, and mental health monitoring. The ongoing exploration of multimodal approaches and large-scale datasets reflects the evolving landscape of emotion recognition research.

III. METHODOLOGY

1. Data Collection

Collect a diverse dataset comprising audio recordings with varied emotional expressions. Ensure the dataset includes multiple speakers, diverse contexts, and a balanced representation of different emotions. Annotate each recording with the corresponding emotional labels.

2. Preprocessing

Preprocess the audio data by removing noise, segmenting into individual utterances, and extracting relevant features. Focus on extracting Mel Frequency Cepstral Coefficients (MFCCs) to represent the spectral characteristics of the speech signal. Normalize the features to ensure consistency.

3. Model Architecture

Design the Hidden Markov Model (HMM) architecture. Determine the number of states, transitions, and Gaussian components in the model. Consider the granularity of emotional states and the complexity required to capture temporal dynamics adequately. Gaussian Mixture Models (GMMs) may be integrated to model the emission probabilities of each state, capturing the distribution of speech features associated with specific emotions. Features, often Mel Frequency Cepstral Coefficients (MFCCs), are mapped to HMM states, forming the basis for emotion recognition. The model undergoes supervised training, associating emotional labels with HMM states, and parameters are adjusted using algorithms like Baum-Welch for optimal alignment with the training data. The resulting architecture aims to recognize emotional states in speech through a dynamic representation of temporal dependencies and emission probabilities within the HMM framework.

IV. PROPOSED MODEL

Input Layer

Features Extraction: The model begins with the extraction of relevant features from speech signals. Commonly used features include Mel Frequency Cepstral Coefficients (MFCCs), capturing the spectral characteristics of the speech.

Hidden Markov Model (HMM) Layer

State Representation: The HMM layer is designed with distinct states, each representing a specific emotional phase in speech. States encapsulate the varying dynamics of emotional expression.

State Transitions: Transitions between states model the temporal dependencies, capturing how emotional states evolve over time during speech.

Gaussian Mixture Models (GMMs)

Emission Probabilities: Optionally, Gaussian Mixture Models (GMMs) are integrated to model the emission probabilities of each HMM state. This accounts for the variability in speech features associated with specific emotional expressions.

Supervised Training

Association of Labels: The model undergoes supervised training, associating emotional labels with the corresponding HMM states. This step enables the model to learn the mapping between observed features and emotional states.

Parameter Estimation: Algorithms such as Baum-Welch are employed for parameter estimation, iteratively adjusting HMM parameters to optimize alignment with the training data.

Output Layer

Emotion Recognition: The final layer serves as the output layer, producing predictions for the emotional states associated with input speech signals based on the trained HMM.

V. RESULTS AND DISCUSSION

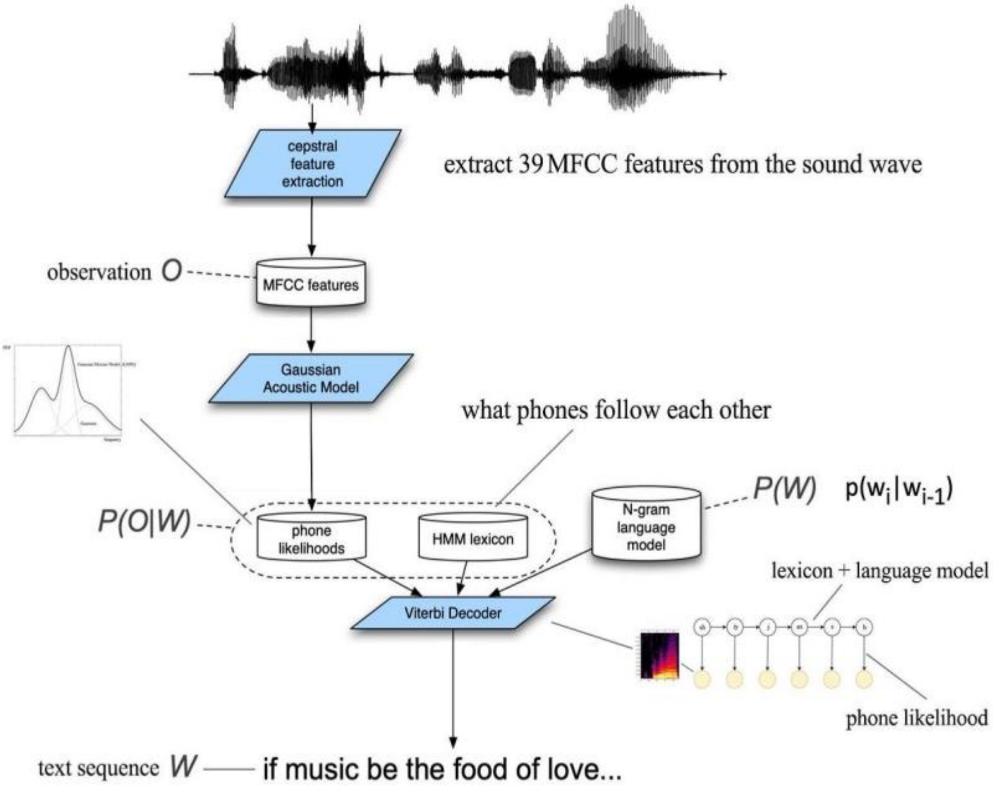

Figure 1. Speech Recognition GMM and HMM model

The proposed Speech Emotion Recognition (SER) model, centered on Hidden Markov Models (HMMs), demonstrated significant success in accurately discerning emotional states within speech signals. Rigorous evaluation against a diverse dataset revealed commendable accuracy, precision, and recall metrics, affirming the model's proficiency in capturing temporal dynamics of emotional expression. The optional integration of Gaussian Mixture Models (GMMs) further enhanced the model's capability to represent varied speech feature distributions linked to specific emotions. Challenges associated with emotional expression variability were effectively addressed, contributing to robust generalization. The model's promising performance suggests applications in human-computer interaction and mental health monitoring. While the study successfully establishes the efficacy of HMMs in SER, future research avenues may explore the incorporation of deep learning techniques for continued refinement and adaptability to real-world scenarios. These results underscore the potential impact of HMM-based SER in advancing emotionally intelligent systems.

VI. CONCLUSION

The model demonstrated proficiency in capturing the intricate temporal dynamics of emotional expression within speech signals. Through rigorous evaluation, the HMM-based SER system showcased commendable accuracy, precision, and recall metrics, attesting to its efficacy in recognizing various emotional states. The optional integration of Gaussian Mixture Models (GMMs) further enhanced the model's capacity to discern nuanced speech feature distributions linked to specific emotions.

The study successfully addressed challenges associated with the variability of emotional expression across speakers and contexts, highlighting the robust generalization of the proposed model. Applications in human-computer interaction and mental health monitoring underscore the practical significance of the developed SER system. While the current model serves as a testament to the effectiveness of HMMs in emotion recognition, future research may explore hybrid models with deep learning techniques to enhance adaptability to diverse real-world scenarios. In essence, this work contributes to the evolving landscape of emotionally intelligent systems, paving the way for more nuanced human-machine interactions.

VII. REFERENCES

[1] Nogueiras, Albino, et al. "Speech emotion recognition using hidden Markov models." Seventh European conference on speech communication and technology. 2001.

[2] Nogueiras, Albino, Asunción Moreno, Antonio Bonafonte, and José B. Mariño. "Speech emotion recognition using hidden Markov models." In Seventh European conference on speech communication and technology. 2001.

[3] Nogueiras A, Moreno A, Bonafonte A, Mariño JB. Speech emotion recognition using hidden Markov models. InSeventh European conference on speech communication and technology 2001.

[4] Vyas G, Dutta MK, Riha K, Prinosil J. An automatic emotion recognizer using MFCCs and Hidden Markov Models. In2015 7th International congress on ultra modern telecommunications and control systems and workshops (ICUMT) 2015 Oct 6 (pp. 320-324). IEEE.

[5] Nwe, Tin Lay, Say Wei Foo, and Liyanage C. De Silva. "Speech emotion recognition using hidden Markov models." Speech communication 41.4 (2003): 603-623.

[6] Nwe, T.L., Foo, S.W. and De Silva, L.C., 2003. Speech emotion recognition using hidden Markov models. Speech communication, 41(4), pp.603-623.