Paper Title: Pattern recognition and features selection for speech emotion recognition model using deep learning

Paper link :https://link.springer.com/article/10.1007/s10772-020-09690-2

1.Summary

1.1 Motivation

The study presents a unique deep learning-based speech recognition model that focuses on identifying emotions in human speech. By recognizing emotional states in spoken language, it seeks to enhance human-machine communication. This technology has applications in emergency services, customer service, and other domains.

1.2 Contribution

The development of a deep learning model that performs better than current techniques at identifying emotions in speech is the paper's primary contribution. Using several speech variables, the suggested model shows great accuracy in emotion identification, reaching as high as 94.21%.

1.3 Methodology

The procedure includes taking audio recordings of people speaking and extracting characteristics using a pre-trained Convolutional Neural Network (CNN) ResNet34 model. To enhance the accuracy of the model, approaches such as data augmentation, optimal learning rate finding, stochastic gradient descent with restarts, and fine-tuning are implemented. The performance of the model is also assessed in the research using a variety of speech aspects, such as prosodic, MFCC, LPC, and LSP features.

1.4 Conclusion

The deep learning model that is being presented shows significant advancements in speech emotion identification, outperforming more conventional techniques like kNN and GMM and attaining high accuracy rates. The model's hopeful implications for practical applications stem from its ability to accurately recognize emotional states in human speech.

2.Limitation

2.1 First Limitation

One limitation of the paper is the lack of discussion regarding the computational resources required to train and deploy the deep learning model. Deep learning models can be computationally intensive, and this aspect should be addressed, especially for real-time applications.

2.2 Second Limitation

The performance of the proposed framework is the primary emphasis of the research; however, possible challenges, such as managing accents, loud or non-standard speech, or changes in speech patterns, are not thoroughly examined. Resolving these issues is essential to the model's adaptability in practical situations.

3. Synthesis

The paper's conclusions have important consequences for a variety of applications where it's critical to comprehend the speaker's emotional state, including emergency response and customer service. The capability of the deep learning model to accurately recognize emotions in speech is a possibility for enhancing human-computer interaction. Further studies may concentrate on resolving the detected limitations and increasing the model's utilization in real-world scenarios.