

Emotion Recognition from Speech Signal Using Deep Learning

Intelligent Data Communication Technologies and Internet of Things pp 471–481 | Cite as

Conference paper

First Online: 13 February 2021

- 142 Downloads

Part of the [Lecture Notes on Data Engineering and Communications Technologies](#) book series (LNDECT, volume 57)

Abstract

Emotions play a vital role in a human's mental life. Speech is a medium through which expression of perspective and identification of one's mental state is possible. Recognizing the feelings that others are trying to convey through speech is essential. There are various parameters of the speech signal that define the feelings of a person. Thus, speech emotion recognition (SER) from the speech signal is a challenging task. This paper proposed an SER system based on the features extracted and obtained by Mel-frequency cepstral coefficient (MFCC) spectrograms. As for a complex model, the audio features obtained by MFCC play an important role in emotion recognition. The 1D-CNN model architecture is implemented in this paper. The work is performed on "The Ryerson Audio-Visual Database of Emotional Speech and Song" (RAVDESS) dataset. Six emotions based on their gender are classified (i.e., 6×2 emotions) with 82.3% accuracy. The emotions classified are happy, sad, angry, calm, fear, and nervous.

Keywords

Speech emotion recognition (SER) Mel-frequency cepstral coefficient (MFCC)

Deep learning Convolution neural network (CNN)

This is a preview of subscription content, [log in](#) to check access.

References

1. Ayadi ME, Kamel MS, Karray F (2011) Survey on speech emotion recognition. In: Features, classification schemes, and databases. Pattern Recog 44(3):572–587. <https://doi.org/10.1016/j.patcog.2010.09.020> (<https://doi.org/10.1016/j.patcog.2010.09.020>) (2011)