Speech Emotion Recognition with Bi-GRU and Self-Attention based Feature Representation

Authors : Bubai Maji, Monorama Swain

Abstract : Speech is considered an essential and most natural medium for the interaction between machines and humans. However, extracting effective features for speech emotion recognition (SER) is remains challenging. The present studies show that the temporal information captured but high-level temporal-feature learning is yet to be investigated. In this paper, we present an efficient novel method using the Self-attention (SA) mechanism in a combination of Convolutional Neural Network (CNN) and Bi-directional Gated Recurrent Unit (Bi-GRU) network to learn high-level temporal-feature. In order to further enhance the representation of the high-level temporal-feature, we integrate a Bi-GRU output with learnable weights features by SA, and improve the performance. We evaluate our proposed method on our created SITB-OSED and IEMOCAP databases. We report that the experimental results of our proposed method achieve state-of-the-art performance on both databases.
Keywords : Bi-GRU, 1D-CNNs, self-attention, speech emotion recognition
Conference Title : ICCPS 2022 : International Conference on Cyber-Physical Systems
Conference Location : Tokyo, Japan
Conference Dates : April 25-26, 2022