

Speech Emotion Recognition Using Spectrogram Based Neural Structured Learning

Dawn Sivan^{1,2*}, *Haripriya P. H.*³ and *Rajan Jose*²

¹Advanced Intelligent Materials, Universiti Malaysia Pahang, 26300 Gambang, Pahang, Malaysia

²Faculty of Industrial Sciences and Technology, Universiti Malaysia Pahang, 26300 Gambang, Pahang, Malaysia ³Logiprompt Techno Solutions India Private Limited, Thiruvananthapuram, Kerala, India

*Corresponding author: dawnsivan91@gmail.com

Abstract

Human emotions are extremely crucial in our daily life. Emotion analysis based solely on auditory data is difficult due to the lack of visible visual information on human faces. Thus, a unique emotion recognition system based on robust characteristics and machine learning from the audio speech is reported in this paper. Audio details are used as input to the person-independent emotion recognition system, from which the spectrogram values are extracted as features. The generated features are then used to train and understand the emotions via Neural Structured Learning (NSL), a fast and accurate deep learning approach. During studies on an emotion dataset of audio speeches, the proposed approach of integrating spectrogram and NSL produced improved recognition rates compared to other known models. The system can be used in smart environments like homes or clinics to provide effective healthcare, music recommendations, customer support, and marketing, among several other things. As a result, rather than processing data and making judgments from far distant data sources, the decision-making could be made closer to where the data lives. The Toronto Emotional Speech Set (TESS) dataset that contains 7 emotions has been used for this research. The algorithm is successfully tested with the dataset with an accuracy of ~97%.

Keywords: Deep learning; Human computer interface; Neural structured learning spectrogram; Speech emotion recognition.