MULTIMODAL FUSION MODEL FOR EMOTION RECOGNITION

Nurul Nadhrah Kamaruzaman[1], Nor Azura Husin[1], Norwati Mustapha[1], Razali Yaakob[1], Siti Khadijah Ali[1], Masrina Mustaffa[1]
[1]Universiti Putra Malaysia

ABSTRACT

Emotion recognition gains its relevance in diverse areas especially in the field of healthcare. As time goes on, research in emotion recognition analysis has developed from the traditional single modal to complex multimodal analysis. Extraction of data from various modalities such as text, video or audio are required in order to gain a meaningful pattern, thus increase the accuracy of the recognition. In line with that, this study addresses several issues relating to computer understanding of human emotions involving extraction of salient features and contextual learning. Addressing such issues, this study aims to implement several deep learning networks and explore novel fusion strategies with the exploitation of multimodal data including text, visual and audio. As a result, this study is expected to produce an accurate and efficient multimodal fusion model that gives the best performance for emotion recognition and classification.

INTRODUCTION

Emotion recognition is a challenging task as it involves predicting an indefinite and subjective emotional state from multimodal input data. Extraction of salient features from multimodal data is the first important step for emotion recognition as human does not have a uniform standard in expressing emotions. This process is essential to identify the features in which are able to distinguish human different emotional state. Besides that, the role of context in emotion perception is very important as context can infer more emotional states of human. It is necessary to analyse the contextual information thoroughly in order to identify the emotional states precisely.

OBJECTIVE

This research aims to propose an accurate multimodal fusion model for emotion recognition by exploring the best way on how to configure the deep learning networks, in order to capture the salient features of human emotions and maximize the contextual learning from three modalities (audio, visual and text)

APPROACH

Feature Extraction

Audio	Visual	Text
Python package for audio analysis	3DConvolutionalNetwork = Capable in identifying gesture and motion in video data	Convolutional Network

Contextual Learning

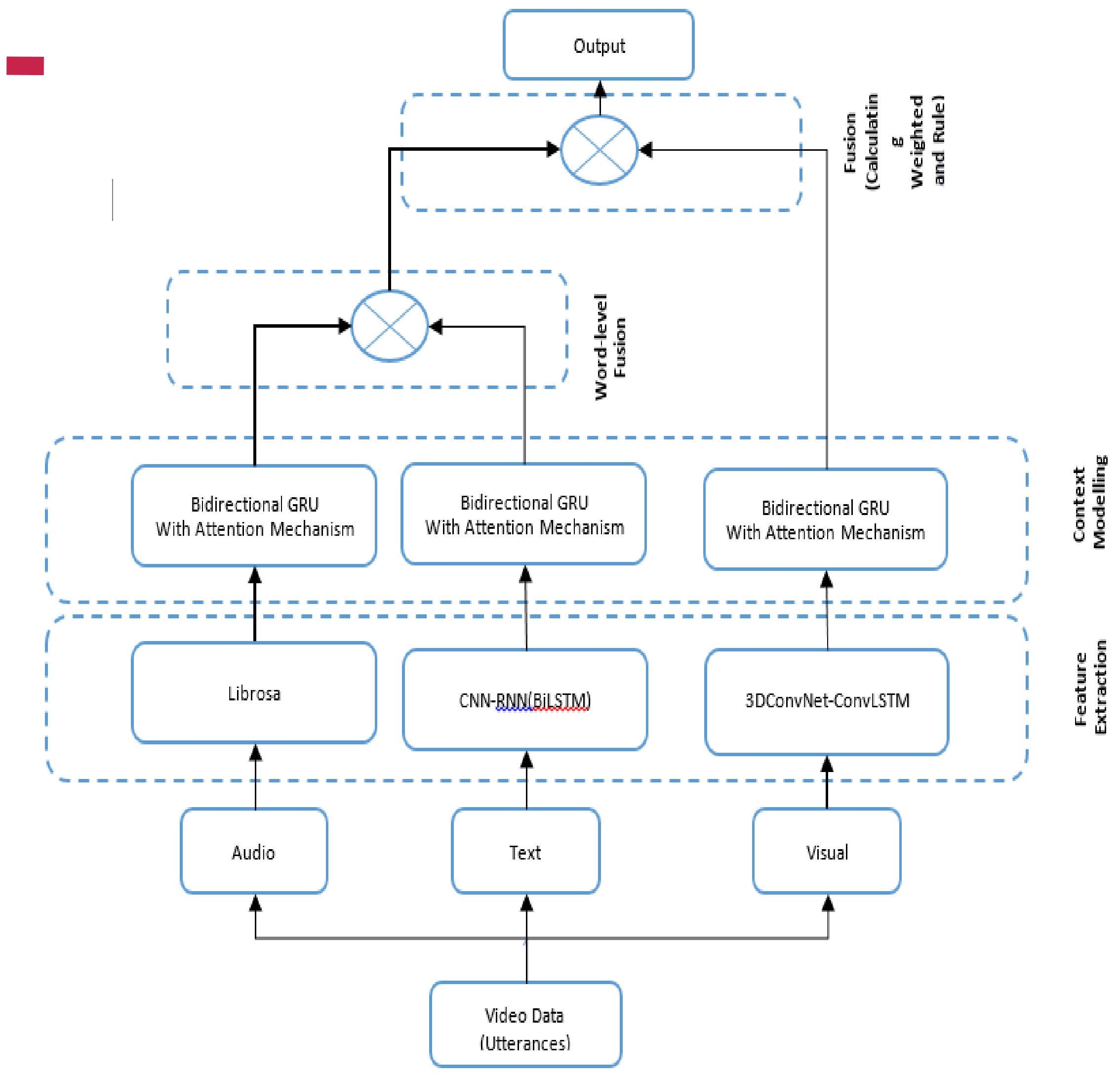
Modality based attention: The idea is to prioritize one of the modalities at the world level, which is when any of the features of the modalities is more relevant to capture emotion, the model should prioritize that particular feature and vice versa. In order to achieve this, bidirectional GRU with attention mechanism will be incorporated as it is able to capture contextual information between words effectively

Context based attention: Model implement attention mechanism over the entire utterance. This idea will focus on mass probability over the words that capture emotional states along the sequence

Fusion Strategy

Model implement word-level fusion of associating the text and audio at each word. This study will then fuse the visual modality by calculating the weighted rules.

DESIGN


Output
Fusion (Calculatin g Weighted and Rule)
Word-level Fusion
Bidirectional GRU With Attention Mechanism
Bidirectional GRU With Attention Mechanism
Bidirectional GRU With Attention Mechanism
Context Modelling
Librosa
CNN-RNN(BiLSTM)
3DConvNet-ConvLSTM
Feature Extraction
Audio
Text
Visual
Video Data (Utterances)

CONCLUSION

The advancement in emotion recognition technology brings such a huge positive impact in various areas especially in healthcare. With the current implementation and innovation, emotion recognition can greatly facilitate the healthcare professional in analysing and detecting a range of mental state that are particularly relevant to mental health care. The result of analysis may provide an early intervention for people with potential mental health problem and save them from suffering.

CASH 2020
INTERNATIONAL CONFERENCE ON COMPUTER ASSISTED SYSTEM IN HEALTH, EDUCATION AND SUSTAINABLE DEVELOPMENT

Contact:
nurul.nadhrah0111@gmail.com / n_azura@upm.edu.my
norwati@upm.edu.my/ razaliy@upm.edu.my /
ctkhadijah@upm.edu.my / masrina@upm.edu.my