Determining Affect Intensity on a Continuous Range

Ravikiran Parameshwara
Human-Centered Technology Research Centre, University of Canberra, Australia

I. INTRODUCTION

There has been substantial progress in the automated analysis of the affect (emotions, mood) over the past two decades, owing to its applications in various computing technologies like medical imaging, ambient intelligence etc. Although multiple studies focus towards the categorical classification of affect, there is a compelling need for analysing affective states in a continuous space as discrete emotions are not representative of the complex emotions displayed by humans on a daily basis. Typically, continuous emotional states are represented on the two-dimensional space of *valence* (how negative or positive the emotional display is) and *arousal* (how calming or exciting the emotional display looks like) [4].

While determining the affective state in a continuous space, it is crucial to consider the temporal information of the emotional state [3]. This dynamic parameter reveals additional information about the affective state, as compared to the observations at a specific instance (static). Studies show that information from the temporal dimension may have psychological relevance to the genuineness and other behaviours.

Despite the existence of multiple affective video databases annotated with continuous/categorical emotional information, they are limited by the amounts of annotated data. This limitation can be attributed to the tedious process of in-the-wild data collection, and cumbersome process of accurate annotation by experts.

The aim of this PhD project is to analyse human affect on a continuous time scale. This project has numerous applications in psychology and psychiatric studies, for example, in understanding impaired social functioning, such as schizophrenia, depression, and autism spectrum disorder. *In-the-wild* video data will be considered in this research as temporal data collected under uncontrolled settings manifests real-world situations. Specifically, this research aims at addressing the following research questions (RQ):

(RQ1) **What is the influence of limited video data on emotion inference?**
While studies have established the existence of *signature* facial expressions corresponding to the basic categorical emotions [1], individual differences in emoting facial expressions nevertheless exist; factoring out these idiosyncrasies is critical for effective emotion inference, especially when the data is limited.
In this research, we explore continuous human affect inference using AFEW-VA [2], an 'in-the-wild' video dataset with limited data, employing *subject-independent* (SI) and *subject-dependent* (SD) settings. To-date, a novel dynamically-weighted loss function is employed with a Convolutional Neural Network (CNN)-Long Short-Term Memory (LSTM) architecture to optimize dynamic affect prediction. As per the results obtained, superior prediction is achieved in the SD setting, as compared to the SI counterpart.

(RQ2) **How can emotions be inferred using limited data?**
Although deep learning approaches have shown impressive successes across domains, their success often depends on the the large amount of quality data. To deal with this limitation, in recent years, researchers have introduced *few-shot learning* algorithm, mainly for the classification task, while limited techniques exist for regression task. Furthermore, these algorithms are less explored for video data.
In this research, we exploit few-shot learning paradigm for continuous affect inference to deal with limited video data.

(RQ3) **Can fusing the multiple modalities like facial expression, acoustic features, bio-signals, etc. result in better affect inference?**
While (RQ1) and (RQ2) are focused towards affect inference from unimodal (video) data, (RQ3) examines the importance of multimodal data. Since humans express affect using multiple modalities like facial expressions, acoustic cues, gesture and body movements, physiological signals, we aim to explore how unimodal (single modality) or multimodal (combination of difference modalities) approaches influence affect inference.

II. TIMELINE

Since the current results indicate that factoring the idiosyncrasies is critical for affect inference in limited data (RQ1), the ongoing direction of research is employing few-shot learning techniques (RQ2) which will be completed by December 2022. By mid-2023, the experiments of multimodal learning (RQ3) will be executed. The expected thesis submission date is January 2024.

ACKNOWLEDGEMENTS

I thank my supervisors Dr. Ibrahim Radwan, Prof. Roland Goecke, and Dr. Ramanathan Subramanian for their valuable support and guidance. I would also like to thank Dr. Akshay Asthana and Dr. Iman Abbasnejad for fruitful discussions. This work is funded by the Australian Research Council Discovery Project ARC DP190101294.

REFERENCES

[1] L. F. Barrett, R. Adolphs, S. Marsella, A. M. Martinez, and S. D. Pollak. Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychological science in the public interest*, 20(1):1–68, 2019.
[2] J. Kossaifi, G. Tzimiropoulos, S. Todorovic, and M. Pantic. Afew-va database for valence and arousal estimation in-the-wild. *Image and Vision Computing*, 65:23–36, 2017.
[3] A. Metallinou and S. Narayanan. Annotation and processing of continuous emotional attributes: Challenges and opportunities. In *2013 10th IEEE international conference and workshops on automatic face and gesture recognition (FG)*, pages 1–8. IEEE, 2013.
[4] R. W. Picard. *Affective computing*. MIT press, 2000.