Biomedical Signal Processing and Control 47 (2019) 324–333

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Audio-video emotional response mapping based upon Electrodermal Activity

Vivek Sharma [a,*], Neelam R. Prakash [b], Parveen Kalra [c]

[a] Centre of Excellence, Industrial and Product Design PEC University and Technology, Chandigarh, India
[b] Department of Electronics & Communication Engineering PEC University and Technology, Chandigarh, India
[c] Department of Production & Industrial Engineering PEC University and Technology, Chandigarh, India

ARTICLE INFO

Article history:
Received 18 October 2017
Received in revised form 21 July 2018
Accepted 20 August 2018

Keywords:
Affective computing
Multilayer neural networks
Electrodermal activity
Music videos

ABSTRACT

In this paper, a machine learning algorithm is proposed for emotional pattern recognition during audio-visual stimuli (music videos) using Electrodermal Activity (EDA). For emotion prediction apart from conventional time domain features of EDA signal, various features in different signal representation i.e. frequency and wavelet were analysed. The comparative result indicated that the wavelet features subset outperformed the conventional time domain features in term of classification accuracy. For identification of optimal network configuration, various combination of optimization algorithms (i.e. backpropagation algorithms) and error function were explored. The best performance of 79% for arousal, 69.8% for valence and 71.2% for dominance were obtained for emotion recognition respectively.

1. Introduction

Emotional response are reflection of automatic nervous system (ANS) activation which are trigger by brain response to an emotional stimulus [1,2]. The emotional response of an individual is mainly elicited in state and trait factors. The state factors are expressed in voice, facial expressions and body expressions of an individual. The trait factors are reflected in psychophysiological signals such as Electromyography (EMG), Electrocardiogram (ECG), Photoplethysmogram (PPG), Pupil Size and Electrodermal Activity (EDA) etc. [3]. In some controversial cases subjects usually hide/conceal their state response due to societal and cultural issues. The trait factors are hard to conceal than state factors and can provide accurate information about the emotional status of an individual [3].One of such widely used psychophysiological signal for emotion recognition is skin conductance (SC) measured via Electrodermal Activity (EDA) of the skin, which is also referred as Galvanic Skin Resistance (GSR).It can be easily recorded via cheap electrical circuitry via noninvasive methods [4]. It is widely considered as a reflection of sympathetic arousal, as sweat glands activities are linked with the sympathetic nervous system branch of ANS [4]

EDA signal mainly consists of two components, namely tonic and phasic. Tonic components are reflection of a slow moving changes in skin conductance which are also referred as a baseline level of the skin conductance (SCL). Phasic component or Skin Conductance Response (SCR) consist of dynamic changes in skin conductance which can be attributed to a particular stimulus [4]. Conventionally EDA is decomposed into phasic and tonic components using certain threshold criteria for peak detection in predefined response window [4]. Duration of response window ranges from 1 to 5 s after the onset of a stimulus and threshold varies from 0.001 μs to 0.05 μs [5,6]. These methods suffer from drawbacks such as a) underestimation or overestimation of SCR amplitude due to subjective estimation b) overlapping of SCR if inter-stimulus interval (ISI) i.e. Interval between two stimuli, is less than recovery time of stimulus c) SCR slope distortion due to inference of recovery slope from the previous SCR response [7]. These drawbacks can result in signal attenuation and improper measurement of SCR parameter's which are indicative of sympathetic arousal/emotional reaction. To overcome these drawbacks various mathematical models for EDA signal decomposition have been proposed in the last two decades. These model can provide accurate decomposition of EDA signal into phasic and tonic component compared to conventional peak method [8] which are describe Section 2.

Previous researchers such as [9–11] have proposed machine learning algorithm using EDA signal for emotion prediction since it is free from the parasympathetic division of ANS compared to other signals [4]. Most of these studies were based upon conventional techniques such as peak to peak for decomposition of EDA

* Corresponding author.
E-mail address: sharma.v0516@gmail.com (V. Sharma).

https://doi.org/10.1016/j.bspc.2018.08.024
1746-8094/

signal which suffers from various drawbacks, as mentioned earlier. Further, only time domain features of EDA such as rise time, half time, max SCR, the number of peaks, average peak height, etc. were explored for emotional mapping without taking into account other signal representation features such as frequency and joint time-frequency which can provide more useful information. Some researchers, such as Ghaderyan and Abbasi [12] have used features from various signal representations i.e. time, frequency and wavelet for estimation of mental stress using EDA signal. Most of these studies, including Ghaderyan and Abbasi [12] were based upon stimulus of shorter duration such as driving tasks, self-induced, images, arithmetic problems and audio etc. which can provide a low emotional response [13–15].The longer duration audio-video films are one of the widely used stimulus for eliciting emotional response with application ranging in TV commercial for product marketing, political campaigns, and music videos etc. Also, during these stimulus emotional response continuously fluctuate and overlap because of which accurate emotional state is difficult to predict by using conventional window techniques.

To address these issues, objective of the present work are listed as following a) To develop a machine learning algorithm for predicting emotion state elicited by audio-visual stimuli (music videos) using EDA signal b) To explore different linear and nonlinear features of EDA signal in different signal representation i.e. Time Domain, Frequency Domain and Wavelet Domain c) To identify optimal feature subset and network configuration. The remaining paper is organized as data and methods (Section 2), then the results (Section 3), Discussion and Conclusion (Sections 4 and 5).

2. Data and methods

2.1. DEAP database

In this study, physiological data, i.e. EDA signal from publicly and freely available database DEAP i.e. “A Database for Emotion Analysis” was utilized. This dataset developed by Koelstra et al. [16]. Permission to download and under fair usage was taken from database administrator via email. Physiological Data from only those participants who have given consent to have their physiological used in publications was utilized. This dataset consists of physiological signal and self-emotional rating (using self-assessment manikin (SAM) scale) from 32 individuals during viewing of 40 one-minute music videos. The physiological signal consists of Electroencephalogram (EEG), Blood Volume Pulse (BVP), Electrodermal (EDA), Electrooculography (EOG), Electromyography (EMG), Skin temperature and Respiration respectively. As per the description given in their manual, inform consent from each participant was taken. Since this study focuses only on EDA signal, the same was extracted from the database using the description given on their site (www.eecs.qmul.ac.uk/mmv/datasets/deap).

As per database description, EDA signal was recorded at 512 Hz on non-dominant hand of all participants. Further, EDA data were down-sampled to 128 Hz and 3 s baseline was subtracted to normalize the data. The participant also rated their emotional experience of each trial on SAM on arousal, valence, dominance, and liking on a scale of 1-9. Rating on familiarly on a scale of 1–5 was also obtained from participants. The down-sampled and normalized EDA data along with self-assessed emotional rating via SAM questionnaire of all participants were downloaded from their site.

2.2. Pre-processing

The purpose of this stage was to remove any artefacts in EDA signal that may be present due to power lines interferences, movement, faulty connection etc. As per the literature, EDA signal frequency range lies in 0.05–2 Hz in which tonic component (SCL) ranges from 0.05 to 0.5 and phasic component (SCR) from 0.5 to 2 Hz [17]. Hence, EDA signal was filtered by using low pass buttworth filter with cut-off frequency of 2 Hz.

Table 1
Inter correlation between Classes Arousal, Valence and Dominance.

R(Correlation Coefficient)				
		A	V	D
Spearman’s rho	A	1.000	0.176	0.531
	V	0.176	1.000	0.300
	D	0.531	0.300	1.000

* A = Arousal D = Dominance and V = Valence.

To remove subject based bias in emotional rating. The SAM scores of emotional dimensions (arousal, valence and dominance) from each individuals (i.e. during all 40 trials) were normalized by using z scores. To test of any association between classes i.e. arousal, valence and dominance correlation analysis using spearman’s rho correlation (Table 1) was performed.

As shown in Table 1 correlation coefficient(r) is lowest among arousal and valence with 0.176 and highest among arousal and dominance with 0.531 while arousal and valence moderately correlated with r = .300.None of classes is significantly correlated as r<.7(threshold value for strong correlation) in all the cases. The k-mean clustering algorithm was used for categorized arousal and dominance in neutral and active classes. The valence dimension was categorized into positive and negative classes.

2.3. EDA decomposition

The basic objective of model based EDA decomposition is to determine the underlying neurological process that results in SCR generation. For example, SA-> SCR where SA is sympathetic arousal. It is a kind of forward model that is turned backward for estimation of SA referred by term “model inversion” [8].This model was further expanded by Bach [8] into SA->SMNA->SCR where SMNA is sudomotor nerve activity which is refer as one of the main cause of SCR generation. The various mathematical models in the last two decades have been proposed for its estimation some of these models are describe in this section.

Lim et al. [18] proposed a four-parameter model which can be extended upto 8 parameters to describe SCL and SCR activities. Their decomposed technique was based on curve fitting using 10 s windows of EDA data. Using this technique, they reported 15% improvement in SCL amplitude and 140 ms in peak latency compared to conventional analysis. Optimization parameters used in their study require visual inspection which can be prone to operator based errors.

Alexander et al. [19] proposed linear time invariant (LTI) model for estimation of SMNA using deconvolution approach. In their model, SC is represented as a convolution between SMNA and Bateman function i.e. biexponential impulse response function (IRF) as describe in Eqs. (1)–(4).

$$IRF\,(t) = (e^{-\frac{t}{t1}} - e^{-\frac{t}{t2}}) \cdot u(t) \tag{1}$$

$$SC = SMNA * IRF;\ *represent\ convolution \tag{2}$$

$$SMNA = Driver_{tonic} + Driver_{phasic} \tag{3}$$

$$Driver_{tonic} + Driver_{phasic} = SC ** (IRF)^{-1}\ \ **represent\ deconvolution \tag{4}$$

Bach et al. [7] also proposed their model based on deconvolution. In their model, a single canonical IRF was used to represent