G Model
ARTMED-1569; No. of Pages 8
ARTICLE IN PRESS
Artificial Intelligence in Medicine xxx (2018) xxx–xxx
Contents lists available at ScienceDirect
Artificial Intelligence in Medicine
journal homepage: www.elsevier.com/locate/aiim

Anytime multipurpose emotion recognition from EEG data using a Liquid State Machine based framework

Obada Al Zoubi[a,b,c], Mariette Awad[a,*], Nikola K. Kasabov[d]

[a] Department of Electrical and Computer Engineering, American University of Beirut, Lebanon
[b] School of Electrical and Computer Engineering, University of Oklahoma, USA
[c] Laureate Institute for Brain Research, OK, USA
[d] Auckland University of Technology, New Zealand

ARTICLE INFO

Article history:
Received 30 October 2017
Received in revised form
29 December 2017
Accepted 3 January 2018

Keywords:
Emotion recognition
EEG
Liquid State Machine
Machine learning
Pattern recognition
Feature extraction

ABSTRACT

Recent technological advances in machine learning offer the possibility of decoding complex datasets and discern latent patterns. In this study, we adopt Liquid State Machines (LSM) to recognize the emotional state of an individual based on EEG data. LSM were applied to a previously validated EEG dataset where subjects view a battery of emotional film clips and then rate their degree of emotion during each film based on valence, arousal, and liking levels. We introduce LSM as a model for an automatic feature extraction and prediction from raw EEG with potential extension to a wider range of applications. We also elaborate on how to exploit the separation property in LSM to build a multipurpose and anytime recognition framework, where we used one trained model to predict valence, arousal and liking levels at different durations of the input. Our simulations showed that the LSM-based framework achieve outstanding results in comparison with other works using different emotion prediction scenarios with cross validation.

1. Introduction

The affective states are psycho-physiological components that can be measured using two main principle dimensions: valence and arousal. Valence varies from negative to positive, and measures emotion's consequences, emotion eliciting circumstances or subjective feeling and attitudes. Arousal measures the activation of the sympathetic nervous system and ranges in intensity from not-at-all to extreme. A couple of studies proposed different models to explain the affective state such as the six basic emotions model [1], dimensional scale of emotions model [2], the tree structure of emotions model [3] and the valence-arousal scale model [4]. In this work we rely on the valence-arousal scale model, due to its simplicity. The model explains emotion variation in a 2D plane, where emotion is affiliated with the corresponding valence and arousal levels. Fig. 1 shows the valence-arousal scale proposed by Russell in which emotions are described in a 2D plane; the horizontal axis represents the valence while the vertical one represents the arousal. More specifically, Russell's model is divided into four regions: Low Valence–Low Arousal (LVLH), Low Valence–High Arousal (LVHA), High Valence–Low Arousal (HVLA) and High Valence–High Arousal (HVHA). Thus, the problem of identifying the emotional state is converted in most of the cases into determining valence and arousal levels.

There are different resources to infer the emotional state in humans such as facial expression, speech, and physiological signals like skin temperature, galvanic resistance, ECG, fMRI and EEG.

This work uses EEG signals for emotion recognition. EEG signals are brainwaves that are produced by population action potential of brain's neurons during activities. Hence, they may be one of the most reliable sources of emotion due to their high temporal resolutions. Moreover, EEG signals are relatively easy to acquire due to the recent advancement in building wireless and wearable EEG sensors [5,6]. To identify and study the emotional state from EEG, several machine learning (ML) techniques have been applied such as deep learning (DL) [7–9], support vector machine (SVM) [10], *k*-nearest neighbors (KNN) [11], and Artificial Neural Networks (ANN) [10].

This works applies a novel framework based on Liquid State Machine (LSM) [12–14] approach for emotion recognition. LSM is a temporal pattern recognition paradigm, and hence it is apt to handle the temporal nature of EEG signals. LSM has been applied successfully to many problems that include spatio/spectro temporal properties like speech recognition [15–17], facial expression

* Corresponding author.
E-mail addresses: obada.alzoubi@ou.edu (O. Al Zoubi), ma162@aub.edu.lb (M. Awad), nkasabov@aut.ac.nz (N.K. Kasabov).

https://doi.org/10.1016/j.artmed.2018.01.001
0933-3657/

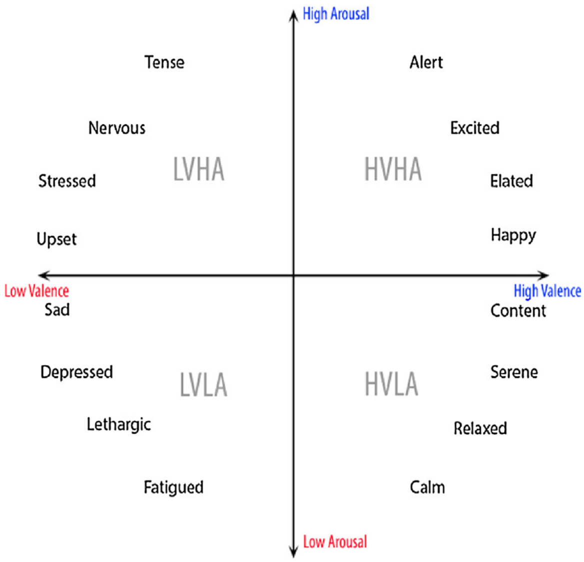

Fig. 1. Russell's model for emotion representation.

recognition [18], robots arm motion prediction [19], real time imitation learning [20], movement prediction from videos [21] and stochastic behavior modeling [22]. Furthermore, several efforts have been made to build hardware-inspired LSM [23–26].

Most of the work done on EEG emotion recognition have suffered from finding informative features from EEG data. In addition, they suffer from converting channel responses into global responses induced by single stimulus. Our work proposes an LSM-based framework for an automatic feature extraction and information consolidation from different channels. This is done by exploiting the temporal unsupervised learning in LSM, where each input to LSM produces a resilient activation patterns inside the LSM that are then converted into features. The same concept has been applied successfully in DL and showed that the extracted features by DL are more informative than the traditional feature extraction approaches [27]. In addition, we reveal how LSM can be adopted to perform multipurpose emotion recognition task in an anytime fashion. By multipurpose, we mean that one trained LSM will be used to predict valence, arousal and liking. And anytime means that the processed signal is not constrained to be of a specific size to capture the sustained emotion. In other words, the framework can conduct the temporal pattern recognition task from a variable length of an input. To evaluate the LSM-based framework, we conducted several experiments to test for performance, linearity and scenario-based emotion prediction accuracies at different lengths of the input. The obtained results from our work show that the framework is capable to surpass other ML approaches used by other research.

The reminder of this work is organized as follows: Section 2 describes the dataset used to validate the framework and surveys the related work. Section 3 introduces LSM and its properties, and then Section 4 elaborates on the proposed LSM framework for EEG emotion recognition. In Section 5, the work tests the proposed framework and discusses the reported results. Finally, the work concludes with a summary and future work in Section 6.

2. Literature review

This section is divided into two parts: in the first part, we introduce the dataset. Part two provides a survey for related work.

Table 1
DEAP dataset description.

Feature	Description
Number of subjects	32
Number of videos/stimuli	40
Number of EEG channels	32
Labels	Valence, arousal and liking
Sampling rate	128 Hz

2.1. DEAP dataset for EEG emotion recognition

We chose the DEAP dataset [28] to validate and test the proposed framework for emotion recognition because the DEAP dataset was recently introduced and used by various EEG emotion recognition research. The next part of literature review surveys the important works that used DEAP dataset. DEAP dataset consists of EEG data recordings from 32 subjects, while watching 40 musical videos. The 40 videos were chosen among 120 initial YouTube videos. Half of the 120 were drawn manually, while the remaining were selected semi-automatically. After that, a 1-min highlight part was determined from each of the 120 initial videos and then was presented to a subjective assessment experiment. The top 40 consistently ranked videos were chosen to be presented to the 32 subjects. Subjects were 50% female, aged from 19 to 37 years with an average of 26.9. Each video was presented to a subject and then she/he was asked to fill a self-assessment for her/his valence, arousal, liking and dominance. Valence scales from 1 to 9 (1 represents sad and 9 represents happy). Arousal scales from 1 to 9 (1 represents calm and 9 represents excited). Liking measures whether a subject likes the video or not, and it corresponds to a number from 1 to 9 (1 means that a subject did not like the video, while 9 means that a subject strongly liked it). The EEG data were recorded according to the 10–20 international system using a 32-channel array at the rate of 512 Hz. Afterwards the data was preprocessed to remove outliers, and then downsampled to 128 Hz. Table 1 shows a summary of DEAP dataset.

2.2. Related work

Several studies have tried to use video clips to study emotions. For example, [11] used five subjects to record 62 channels EEG data in four emotional states: joy, relax, sad and fear stimulated by watching pre-chosen elicitation clips. The work extracted features from time and frequency domains. The testing showed that frequency domain features are more informative than time domain features with the best reported accuracy of 66.51% using SVM classifier. Similarly, [10] used 30 pictures from the International Affective Picture System (IAPS) as an elicitation for 20 subjects. EEG data were recorded for 5 s for each picture using six channels. The best reported result was using time domain features with 56.1% accuracy achieved by SVM classifier.

Other studies used DEAP dataset to evaluate their work. The remaining part of the literature review focuses on these works. In work [7], the authors applied DL with a stack of three autoencoders, two softmax layers and 50 neurons in each hidden layer on DEAP dataset. The work used the power spectral of five frequency bands of EEG: delta, theta, alpha, beta and gamma as an input. The dataset was labeled according into three valence states (Negative, Neutral and Positive) and three arousal states (Negative, Neutral and Positive). The best reported results were 53.42% for valence and 52.03% for arousal when using Principle Component Analysis (PCA) with Covariate Shift Adaptation (CSA) transformation at the input of DL network.

Another study [29] proposed a method to fuse features from segment level into response level. Each problem is considered