ARTICLE IN PRESS

Neurocomputing xx (xxxx) xxxx–xxxx

Contents lists available at ScienceDirect

Neurocomputing

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

Reservoir computing for emotion valence discrimination from EEG signals

Lachezar Bozhkov[a], Petia Koprinkova-Hristova[b], Petia Georgieva[c,*]

[a] Technical University of Sofia, Bulgaria
[b] Institute of Information and Communication Technologies, Bulgarian Academy of Sciences, Sofia, Bulgaria
[c] Department of Electronics Telecommunications and Informatics/IEETA, University of Aveiro, Portugal

ARTICLE INFO

Keywords:
Reservoir Computing
Echo State Networks
Intrinsic Plasticity
Deep Neural Encoders
Feature dimensionality reduction
Affective computing

ABSTRACT

In this paper we propose a new approach for feature dimensionality reduction based on Reservoir Computing (Echo State Networks). The method is validated with EEG data to identify the common neural signatures based on which the positive and negative valence of human emotions across multiple subjects can be reliably discriminated. The key step in the proposed approach is the Intrinsic Plasticity (IP) adaptation of the reservoir states. Learning Echo State Networks (ESN) with IP maximizes the entropy of the distribution of reservoir vectors given static data as a fixed input, which is supposed to follow Gaussian distribution. The equilibrium reservoir vector is extracted for each static input vector by iterating updates of the reservoir vector until it converges. Standard classification and clustering models provided with selected combinations of reservoir neurons are ranked based on their discriminate performance. The IP tuned ESNs is more powerful technique to map the high dimensional input feature vector into a low dimensional representation and improve the emotion valence discrimination compared to classical ESNs and Deep Neural Encoders.

1. Introduction

Affective Computing (AC) is a research field that aims to automatically detect and quantify human emotions. Major techniques for affect detection are voice or facial expressions, text, body posture or language [1,2]. Affective Neuroscience emerged recently as a new AC approach that attempts to find the neural correlated between human emotions and registered brain activity.

Electroencephalography (EEG) is one of the most exploited brain imaging techniques due to its accessibility in several research labs and clinical environments. Comprehensive review of EEG-based emotion recognition systems is provided in [3]. It is difficult to compare the results, because of factors such as data acquisition scenarios, number of participants, emotion models, type of stimulus and extracted features, temporal windows, etc. However, most of the solutions build a recognition model for each participant (subject dependent model) [4–7]. Building models that cover multiple subjects (i.e. subject independent models) is a challenging problem due to the intrinsic variability of the registered brain activity among different participants [8]. For example in [3] based on the Power Spectral Density (PSD) in different frequency bands of the EEG signal and Support Vector Machine (SVM) classifier, happy/unhappy emotions are discriminated with accuracy of 75.62% for the subject dependent setting and only 62.12% for the generalized subject independent setting. In [9], emotion clustering with 84% inter-subjects accuracy is presented based on similarities between phase portraits constructed from the time-response of a Duffing oscillator. The results are obtained with simulated data with low signal to noise ratio (< 10 dB). In [10] exhaustive feature reduction improves the inter-subject accuracy up to 80%. More challenging problem of four emotions (pleasure, joy, sadness, anger) is considered in [11] with 82% inter-subject accuracy.

The emotion recognition systems across multiple subjects have to come up with the problem of choosing the most relevant features for all participants and the usually high dimensional feature space. Working on this problem in [12] we proposed the Echo State Network (ESN) as a mechanism for low dimensional feature space projection in the EEG-based affective computing. The ESNs are dynamical structures designed to facilitate learning in Recurrent Neural Networks (RNN), normally applied for time series modeling. Using ESNs for static data classification or clustering gained less attention. In [13] we proposed a novel approach to extract the equilibrium reservoir states that maximize the model discrimination capacity after Intrinsic Plasticity (IP) adaptation of the reservoir states. Based on IP-trained ESN and after search for the optimal combination of equilibrium reservoir states we obtained subject independent classification models of human emotion valence that outperformed the above referenced studies. The experience gained suggested that a reliable inter-subject neural decoding is feasible only if the features are carefully reduced to the most

* Corresponding author.
E-mail address: petia@ua.pt (P. Georgieva).

http://dx.doi.org/10.1016/j.neucom.2016.03.108
Received 15 November 2015; Received in revised form 3 March 2016; Accepted 31 March 2016
Available online xxxx
0925-2312/

discriminative ones.

The goal of the present work is twofold. First, we explore the limits of the ESNs low dimensional representations and define the optimal reservoir states combination that can serve as the most discriminative features particularly in inter-subject affective computing setting.

Secondly, we make a broad comparison between the IP-trained ESNs with classical ESNs (with randomly generated reservoir) and the state of the art Deep Neural Autoencoders (DNA). DNAs are based on the concept of Deep Neural Networks (DNN) [14] where the subsequent hidden layers, called auto-encoders, map the input data into representation features with different level of abstraction. We argue that the extraction of optimal low dimensional combinations of IP-trained reservoir equilibrium states is more promising mechanism than the auto-encoders for data representation and feature dimensionality reduction in the framework of the EEG-based affective computing.

The paper is organized as follows. In Section 2 the IP-adaptation in Reservoir Computing (RC) is outlined and our approach for RC-based feature extraction is presented. In Section 3 the experimental scenario and the EEG data acquisition are described. In Section 4 a clustering approach for the ESN-based emotion valence discrimination from EEG signals is proposed, and in Section 5 the same problem is solved as a classification task. IP-tuned ESN is compared with Deep Neural Encoders in Section 6. Discussion of the results and conclusions are summarized in Section 7.

2. Reservoir adaptation with Intrinsic Plasticity

2.1. Reservoir Computing

Recurrent Neural Networks (RNN) represent a large class of computational models particularly suitable for nonlinear time series processing. However, despite their successful application in academic and practical problems, the RNN training is only feasible for relatively small networks. A promising approach to overcome the RNN shortcomings was proposed independently by Wolfgang Maass under the name *Liquid State Machines* [15] and by Herbert Yaeger under the name of *Echo State Networks* [16]. This approach is often referred to as *Reservoir Computing* (RC), [17]. In this work the Echo State Network (ESN) formalism of RC is used. The basic structure of the ESN, presented on Fig. 1, consists of one hidden layer, called dynamic reservoir, of randomly connected sigmoid neurons, usually tanh nonlinear functions f^{res}

$$r(k) = f^{res}(W^{in}in(k) + W^{res}r(k-1)) \tag{1}$$

and a linear function f^{out} at the output (readout):

$$out(k) = f^{out}(W^{out}[in(k) \quad r(k)]) \tag{2}$$

Here k denotes discrete time instant, *in(k)* is the vector of ESN inputs with dimension n_{in}, *r(k)* is the vector of reservoir neurons states with dimension n_r and *out(k)* is the vector of ESN outputs with dimension n_{out}. W^{in} and W^{res} are randomly generated matrices of the input and the reservoir weights that remain unchanged during network training. Only the output weights (matrix W^{out}) are trained. In some applications direct input-output connection is not supposed.

In contrast to the classical RNN where the reservoir and the readout weights are trained in the same way, the ESN treats them differently. The purpose of the reservoir is to expand the input history into a rich enough reservoir state space *r(k)* while the readout combines the neuron signals *r(k)* into the desired output signal $out_{target}(k)$. The weights of the reservoir neurons are randomly generated from a uniform distribution symmetric around zero. An important ESN property is that the reservoir should possess the so called *echo state property*. This means that the effect of the current state *r(k)* and the current input *in(k)* on a future state *r(k+n)* should decrease as time passes ($n \to \infty$). There have been various attempts to define rules that guarantee the echo state property and optimize the reservoir. Many of them are based on experience, heuristics or exhaustive search over the parameter space. A promising biologically motivated mechanism called intrinsic plasticity (IP) has recently attracted a wide attention in the reservoir computing community [18].

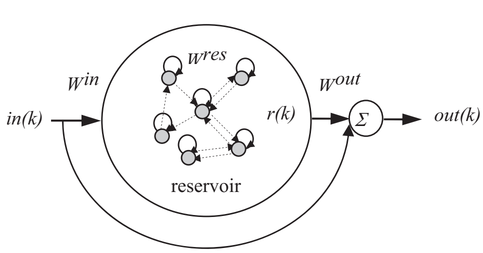

Fig. 1. Echo state network basic structure.

2.2. Intrinsic Plasticity

The IP adaptation rule aims at maximization of information transmission trough the ESN that is equivalent to its output entropy maximization. Motivation of this approach is related to known biological mechanisms that change neural excitability according to the distribution of the input stimuli. In [19] a gradient method for adjusting the biases is proposed and an additional gain term to achieve the desired distribution of outputs by minimizing the Kullback-Leibler divergence:

$$D_{KL}(p(r), p_d(r)) = \int p(r)\log\left(\frac{p(r)}{p_d(r)}\right)dr \tag{3}$$

D_{KL} measures the difference between the actual *p(r)* and the desired $p_d(r)$ probability distribution of the reservoir states r. In [18] the IP formalism for hyperbolic tangent neuron transfer functions and Gaussian output distribution is derived

$$p_d(r) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(r-\mu)^2}{2\sigma^2}\right) \tag{4}$$

Now, the expression (3) can be rearranged as follows:

$$D_{KL}(p(r), p_d(r)) = -H(r) + \frac{1}{2\sigma^2}E((r-\mu)^2) + \log\frac{1}{\sigma\sqrt{2\pi}}, \tag{5}$$

where H denotes entropy. The last term in (5) is constant and the second term determines the deviation of the reservoir state r from the desired mean value. The minimization of (5) is a compromise between the entropy maximization and the minimization of the distance between μ and r. In order to achieve this effect two additional reservoir parameters – the gain a and the bias b (both vectors with size n_r) - are introduced as follows:

$$r(k) = f^{res}(diag(a)W^{in}in(k) + diag(a)W^{res}r(k-1) + b) \tag{6}$$

The IP adaptation adjusts a and b using iterative gradient descent:

$$a_{it} = a_{it-1} + \Delta a$$

$$b_{it} = b_{it-1} + \Delta b$$

where it denotes iteration number and the gradients $\Delta a$and Δb are derivatives of Kullback-Leibler divergence with respect to vectors a and b. In the case of hyperbolic tangent reservoir neurons (f^{res}=$tan\ h$) the gradients are:

$$\Delta b = -\eta\left(-\frac{\mu}{\sigma^2} + \frac{r}{\sigma^2}(2\sigma^2 + 1 - r^2 + \mu r)\right)$$

$$\Delta a = \frac{\eta}{a} + \Delta b net_{input}$$

$$net_{input} = W^{in}in + W^{res}r$$

Here η is the learning rate. The IP adaptation tends to concentrate