Neural Computing and Applications
https://doi.org/10.1007/s00521-018-3664-1

S.I. : EMERGENCE IN HUMAN-LIKE INTELLIGENCE TOWARDS CYBER-PHYSICAL SYSTEMS

Emotion recognition based on physiological signals using brain asymmetry index and echo state network

Fuji Ren[1,2] · Yindong Dong[1] · Wei Wang[1]

Received: 29 April 2018 / Accepted: 24 July 2018

Abstract
This paper proposes a method to evaluate the degree of emotion being motivated in continuous music videos based on asymmetry index (AsI). By collecting two groups of electroencephalogram (EEG) signals from 6 channels (Fp1, Fp2, Fz and AF3, AF4, Fz) in the left and right hemispheres, multidimensional directed information is used to measure the mutual information shared between two frontal lobes, and then, we get AsI to estimate the degree of emotional induction. In order to evaluate the effect of AsI processing on physiological emotion recognition, 32-channel EEG signals, 2-channel EEG signals and 2-channel EMG signals are selected for each subject from the DEAP dataset, and different sub-bands are extracted using wavelet packet transform. *k*-means algorithm is used to cluster the wavelet packet coefficients of each sub-band, and the probability distribution of the coefficients under each cluster is calculated. Finally, the probability distribution value of each sample is sent as the original features into echo state network for unsupervised intrinsic plasticity training; the reservoir state nodes are selected as the final feature vector and fed into the support vector machine. The experimental results show that the proposed algorithm can achieve an average recognition rate of 70.5% when the subjects are independent. Compared with the case without AsI, the recognition rate is increased by 8.73%. On the other hand, the ESN is adopted for the original physiological feature refinement which can significantly reduce feature dimensions and be more beneficial to the emotion classification. Therefore, this study can effectively improve the performance of human–machine interface systems based on emotion recognition.

Keywords Emotion recognition · Physiological signals · Brain asymmetry index · Echo state network

1 Introduction

In recent years, human–computer interaction has received a lot of attention from researchers and scholars and has influenced us all in aspects of our lives, such as human–human interaction, job search and entertainment, etc. [1, 2]. If the machine can understand the human emotion state, human–machine interface (HMI) can be more intuitive, smooth and effective to adapt to a variety of human–machine applications, and we call it affective computing (AC) [3, 4]. Affective computing is a field of research that allows the machine to perceive and express some emotions; it can recognize, interpret and process human emotions through related system and intelligent equipment; therefore, emotion recognition is the primary research hot spot for affective computing. Facial expressions [5, 6], speech [7], posture [8], text [9, 10] and physiological signals from the

✉ Yindong Dong
dongyindong66@mail.hfut.edu.cn

Fuji Ren
ren@is.tokushima-u.ac.jp

Wei Wang
wangwei_hfut@hfut.edu.cn

[1] School of Computer and Information, Anhui Province Key Laboratory of Affective Computing and Advanced Intelligent Machine, Hefei University of Technology, Hefei 230601, China

[2] Graduate School of Advanced Technology and Science, University of Tokushima, Tokushima 7708502, Japan

Published online: 02 August 2018

human spontaneous nervous system are used to express and identify emotional cognition [11, 12, 34]. Due to the high temporal resolution of EEG signals and the electrical signals collected directly from the central nervous system, it is possible to reflect truly and intuitively the emotional changes in the brain. In recent years, electroencephalogram-based emotion recognition (EEG-ER) has received extensive attention in medical care, emotional companionship and long-range teaching [33, 35].

Emotional induction is mostly a physiological response that an individual stimulates after being stimulated by related emotionally impressive pictures, videos, music or spontaneous emotional imaginations [4, 16]. In the process of emotion recognition based on physiological signals, most researchers are concerned about the study of feature extraction and classification accuracy for each physiological signal trial under all subjects. There is a question: does all subjects have the same level of emotional incentives in the same emotionally evoked medium? Different individuals are often affected by social relationships, family, weather or physical conditions [3, 9]. Therefore, under the same emotion-induced factor, the degree of emotional stimulation may vary greatly, and some of them cannot be induced by effective motivating emotion. Individuals who collect physiological signals cannot carry a large amount of corresponding emotional information, which will affect the overall emotional recognition effect and may even lead to wrong decisions in the MCI. In addition, based on the diversity and complexity of physiological signals, we will get too high feature dimensions, which can easily lead to feature disasters, and then the phenomenon of overfitting will occur so as to affect the classification accuracy.

In view of the above problems, we have introduced an index (AsI) for measuring the degree of emotional induction in EEG signals [11]. In terms of spatial dimensions, the asymmetry of the left and right hemispheres of the brain is an important role for the initiation of human emotions. Davidson [13, 14] found that EEG signals exhibited emotional asymmetry in the left and right hemispheres of the forehead of the alpha band, mainly because activity in the left forehead area was stimulated by positive emotions, and brain electrical activity in the right forehead area was stimulated by negative emotions. Zheng [15] extracted 27 pairs of asymmetrical differential entropy feature information in the 62-channel EEG system and used deep neural network for emotion recognition. The results showed that the forehead asymmetry has significant difference in the emotion recognition in the beta and gamma frequency bands [32]; Daimi [16] selected the frontal energy asymmetry of different brain regions in each frequency in the 4 dimensions of valence, arousal, dominance, and liking, respectively, and found that the classification accuracy in valence and arousal was higher than in liking and arousal. Therefore, the paper selects the important influence of the asymmetry of the frontal lobe on the emotion in the brain to quantify the degree of EEG emotion induced. The idea can be used to extract the physiological signals related to emotional stimulation. The multidimensional directed information (MDI) analysis method was used to quantify the amount of interactive information between the left and right hemispheres [17], and then, the asymmetry index (AsI) was calculated based on this quantification.

In this paper, firstly, we can get the asymmetric index (AsI) of the frontal and right regions by the asymmetric principle of the brain [14, 18] and then screen out the strong emotion-induced window signals according to the index value. In order to more fully express the information features, we selected 36-channel total physiological signals of EEG, EMG and EOG for each subject to carry out wavelet packet transform, and then, the wavelet packet coefficients were extracted in their respective sub-bands, and the clustering analysis was performed for the wavelet packet coefficients in each channel; finally, the probability distribution for each cluster in each sub-band of the wavelet was obtained. In view of the high dimension of the original feature and the large redundancy between the features, the direct use of these features for classification will lead to inaccuracy. Therefore, we propose to integrate the original feature and extract the state node as the new feature space by using the training principle of ESN's reservoir.

2 Related work

2.1 Multidimensional direct information MDI

There are many ways to characterize causal relationships between two time series, such as directed coherent analysis [19], directed information analysis (DTF) [20], partially directed coherence [21], Granger causality [22] and multidimensional directed information MDI [17]. MDI analysis is a method of recognizing causal information sharing between two time series in consideration of a conditional time series. Compared with other methods, MDI can express the amount of information propagation in terms of absolute bits. The brief description of MDI is as follows:

We take the time series X, Y and Z of the three EEG signals. The length of the sequence is denoted as L. To express the causal relationship between the time series in a more accurate manner, use $L=A+1+B$ for describing the length of the learning sequence before and after one of the time point in the sequences. For example, for the EEG sequence X:

$$X = x_{k-A}x_{k-A+1}\ldots x_{k-1}x_k x_{k+1}x_{k+2}\ldots x_{k+B} = X^A x_k X^B \quad (1)$$

In this formula, $X^A = x_{k-A}x_{k-A+1}\ldots x_{k-1}$, $X^B = x_{k+1}x_{k+2}\ldots x_{k+B}$.

In the literature [17], the information generated at the k-point of the X-sequence is passed to the sequence Y through the delay of b-time. Here, the time series Z is considered as a conditional sequence of mutual information, and information is transmitted to X and Y, respectively. The directed mutual information is expressed as in the following formula (2):

$$I\left(x_k \to y_{k+b}|X^A Y^A Z^A y_k z_k\right) = \frac{1}{2}\log\frac{\left|R(X^A Y^A Z^A x_k y_k z_k)\right| \cdot \left|R(X^A Y^A Z^A y_k z_{k+b})\right|}{\left|R(X^A Y^A Z^A y_k z_k)\right| \cdot \left|R(X^A Y^A Z^A x_k y_k z_{k+b})\right|} \quad (2)$$

Here, the covariance matrix $x_1, x_2, \ldots, x_n$ of the random variable is expressed as $R(x_1 x_2 \ldots x_n)$. Using Eq. (2), we can calculate the total information amount of X passes to Y under consideration of the information Z, as shown in Eq. (3):

$$S^{XY} = I\left(x_k \to y^B|X^A Y^A Z^A y_k z_k\right) = \sum_{b=1}^{B}\frac{1}{2}\log\frac{\left|R(X^A Y^A Z^A x_k y_k z_k)\right| \cdot \left|R(X^A Y^A Z^A y_k z_{k+b})\right|}{\left|R(X^A Y^A Z^A y_k z_k)\right| \cdot \left|R(X^A Y^A Z^A x_k y_k z_{k+b})\right|} \quad (3)$$

It should be noted that the calculated result in the traditional directional analysis method is X to Y directed mutual information, and no consideration is given to the third-party sequence Z. In this paper, in the process of calculating directed mutual information X, Y under the condition of sequence Z, where Z contains the common components of the X and Y information, it can be seen as the process that the information of Z flows to X and Y, instead of the process that the information directly flows between X and Y, which avoids the ambiguity of information transfer. Through the analysis of MDI, we can assume that X and Y are the EEG signals of the left and right hemispheres of the brain, so that we can explore the asymmetry of the brain's forehead through the estimation of S^{XY} to define the asymmetric index AsI.

2.2 Asymmetry index

When the subjects produce negative emotions, the activities in the right forehead of the brain increase; when the subject produces positive emotions, the activities in the left forehead of the brain are inspired accordingly. According to this asymmetry, we assume that when the individual is calm, the left and right sides of the brain appear symmetrical and have the maximum amount of mutual information; when the individual is emotionally inspired, the forehead is asymmetrical on both sides of the brain, and the mutual information is in minimum. Based on this asymmetric principle, the following formula (4–5) is defined:

$$S_c = S_c^{XY} + S_c^{YX} \quad (4)$$

$$S_a = S_a^{XY} + S_a^{YX} \quad (5)$$

Among them, $S_c^{XY}, S_c^{YX}, S_a^{XY}$ and S_a^{YX} is calculated by Eq. (3). In it, S_c indicates the total amount of two-way information interaction in which the subjects are in the peace state in the baseline signal, and S_a represents the total amount of two-way information interaction in the emotional incentive process of the subjects. Through the above formula (4–5), we can get the effective index AsI on emotional incentives, as in the following formula (6):

$$\mathrm{AsI} = \left|(S_c - S_a) \times \frac{\sqrt{2}}{2}\right| \quad (6)$$

In this paper, AsI can be used to eliminate some signals with low emotional relevance and individuals who are not easily stimulated by emotions before emotion recognition.

2.3 AsI implementation rules

We used MDI to calculate $\mathrm{AsI}_{\mathrm{Fp1-Fp2}}$, $\mathrm{AsI}_{\mathrm{AF3-AF4}}$, respectively, in the two groups of electrodes (Fp1, Fp2, Fz and AF3, AF4, Fz). As shown in Fig. 1, all AsI values are normalized to the range of [0:1], in order to more fully retain the relevant information related to emotional motivation. We formulate rules to give $\mathrm{AsI}_{\mathrm{Fp1-Fp2}}$ higher priority. The specific rules are as follows:

(a) When $\mathrm{AsI}_{\mathrm{Fp1-Fp2}} \in [0.5, 1]$, $\mathrm{AsI}_{\mathrm{AF3-AF4}}$ takes any value, then related signals are reserved.

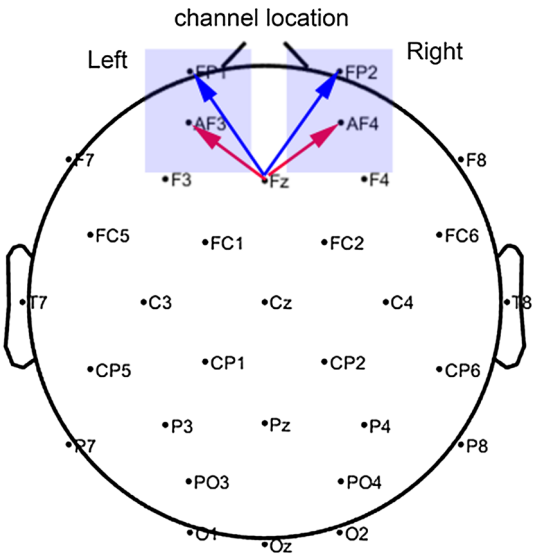

Fig. 1 Two groups of channels of 32 electrode location shown for AsI

(b) When $\mathrm{AsI}_{\mathrm{Fp1-Fp2}} \in [0.4, 0.5)$, $\mathrm{AsI}_{\mathrm{AF3-AF4}} \in [0.15, 0.6)$, then related signals are retained; otherwise, $\mathrm{AsI}_{\mathrm{AF3-AF4}} \in [0, 0.15)$, then related signals are removed.

(c) When $\mathrm{AsI}_{\mathrm{Fp1-Fp2}} \in [0, 0.4)$, $\mathrm{AsI}_{\mathrm{AF3-AF4}} \in [0.6, 1]$, the related signals are retained; otherwise, $\mathrm{AsI}_{\mathrm{AF3-AF4}} \in [0, 0.6)$, the related signals are rejected.

Here, we have discarded the entire 60 s signal satisfying the number of signal windows with low emotional incentives greater than 15 in the above rules, and we believe that these subjects did not perform inspired emotions well in watching this music video. In Fig. 2, the situation in which 40 s emotional incentive signal of No. 7 subject watching music the 9th video is given. In the red box, the brain electrical signals that have been removed after AsI are eliminated in accordance with the rules as above. Through experiments, the compliant windows for retaining emotional stimuli are four, so we can see that No. 7 subject failed to stimulate emotions effectively. We discarded the entire 60 s computer signal and did not participate in the final emotion classification and recognition.

3 Feature extraction and compression

3.1 Features extraction

After AsI processing, we used the wavelet packet transform to divide frequency band into four sub-bands each subject: theta: 4–7 Hz, alpha: 8–13 Hz, beta: 14–30 Hz and gamma: 31–45 Hz, and used the sliding window technology in each of 4 sub-bands; then, we will gain wavelet packet coefficients of each window in each sub-band; this study uses k-means clustering algorithm to cluster these wavelet packet coefficients [29–31]; the k-means algorithm mainly includes the following steps:

1. Randomly select different k wavelet packet coefficients as the initial cluster center.
2. Calculate the membership matrix U.

$$u_{ij} = \begin{cases} 1, & \text{if } \|x_k - c_i\|^2 \le \|x_k - c_t\|^2, \forall t \ne i \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

3. Calculate the objective function.

$$J = \sum_{i=1}^{k} \left(\sum_{k} \|x_k - c_i\|^2 \right) \tag{8}$$

4. Update the cluster center.

$$c_i = \frac{\sum_j^N u_{ij} x_j}{\sum_j^N u_{ij}} \tag{9}$$

5. Return to step 2 to find that the cluster center no longer changes and the algorithm ends.

Finally, the probability distribution of each wavelet coefficient under each cluster can be obtained. $|S_{ij}|$ is the number of wavelet packet coefficients belonging to cluster i in window j, $|S_j|$ is the total number of wavelet packet coefficients in window j, and n is the number of sliding windows.

$$P_{ij} = \frac{|S_{ij}|}{|S_j|}, \quad i = 1, 2, \ldots k \text{ and } j = 1, 2, \ldots, n \tag{10}$$

In order to fully probe into the effect of the distribution of physiological signal features in the emotion classification, we selected the number of clusters $k \in [2{:}6]$ in the process of feature extraction. Therefore, for the EEG signal, a $32 \times 4 \times k$ dimensional feature vector will be generated in the four bands (theta, alpha, beta, and gamma). In

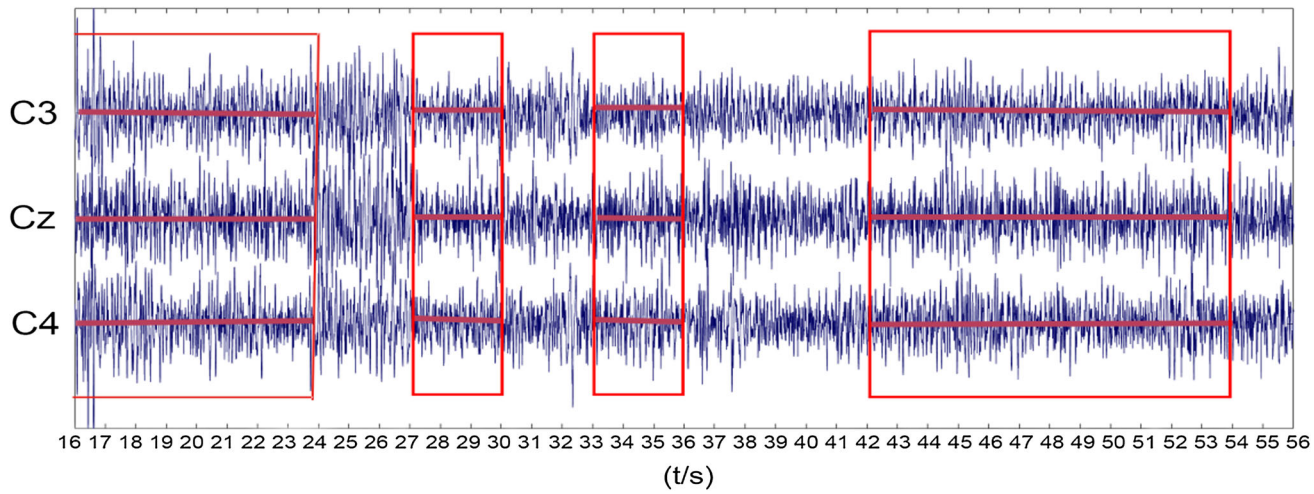

Fig. 2 Emotional motivation situation that No. 7 subject views video the 9th in the central three channels (C3, Cz, C4)

addition, electrooculogram (EOG) and electromyographic (EMG) signals are selected as complementary features. The effective band ranges of the power spectrum energy of EOG and EMG signals are 4–30 Hz and 4–40 Hz, respectively. Therefore, the wavelet packet coefficients of theta, alpha and beta bands are extracted separately for the 2-channel EOG signals and 2-channel EMG signals; the other processing is the same as EEG; this produces a feature vector of $4\times3\times k$ dimensions in the three bands; therefore, a total of $140\times k$ dimensions for a feature vector of is obtained in each time window. High-dimensional feature vectors will bring dimension disasters and lower the classification accuracy. Therefore, this paper uses ESN to further compress and integrate the initial EEG feature signals.

3.2 Compression and selection of features

3.2.1 Echo state network

In this paper, we choose a recurrent neural network–echo state network (ESN) suitable for multi-time series research [23–25]. A standard ESN consists of three parts: the input layer, the reservoir and the input layer, as shown in the following Fig. 3.

In Fig. 3, an input sequence $u(t)=\{u(t), i=1,2,\ldots,k\}$, k is the number of input nodes, X is a state vector of length N, and W is a linkage matrix of size $[N\times N]$. In the modeling process, an offset b is added to the linkage matrix. Therefore, the input matrix w^{in} is $[N\times K]$, the output is $y(t)$, the output node is L, so the output matrix w^{out} is $L\times N$. Then the activation unit of the internal state is updated at time $t+1$ to the following Eq. (11):

$$x(t+1)=f(W\times x(t)+W_{\text{in}}\times u(t)+b) \tag{11}$$

In this equation, $x(t+1)=\{x_i(n+1:i=1,2,\ldots N)\}$, the transfer function we use is tanh, so the range of the attribute values of the state node is $[-1, 1]$. w_{in}, w and b are all randomly initialized when the network mode is initially set up and are fixed in training. Neural network input is accomplished by a linear transformation of the following formula:

$$y(t)=w^{\text{out}}*x(t) \tag{12}$$

In the entire ESN training process, w_{out} is the only parameter that needs to be trained, so the training speed is quite fast. However, in this article, intrinsic plasticity (IP) training process is an unsupervised parameter tuning algorithm that does not involve the training of the target tag data. In order to simplify the calculation, the output dimension is set to 1.

3.2.2 Steady-state training for dimensionality reduction in ESN reservoir

Researchers have confirmed that the appearance of any stable state will be accompanied by a local maximum entropy [26]. Steil proposed a stable intrinsic plasticity (IP) algorithm to improve the performance of the reservoir. This algorithm can dynamically adjust steady state of inside neurons in ESN for the input node, so that the output entropy in the reservoir can reach the maximum value [27].

$$D_{KL}(p(x),p_d(x))=\int p(x)\log\left[\frac{p(x)}{p_d(x)}\right] \tag{13}$$

In (13), we assume $p_d(x)$ is the state output distribution of an reservoir, and this distribution follows a Gaussian distribution with mean μ and variance σ. Here, we implement IP algorithm by measuring the difference between true distributions $p(x)$ and $p_d(x)$.

$$p_d(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x-u)^2}{2\sigma^2}\right) \tag{14}$$

Substituting Eq. (14) into Eq. (13), we can get

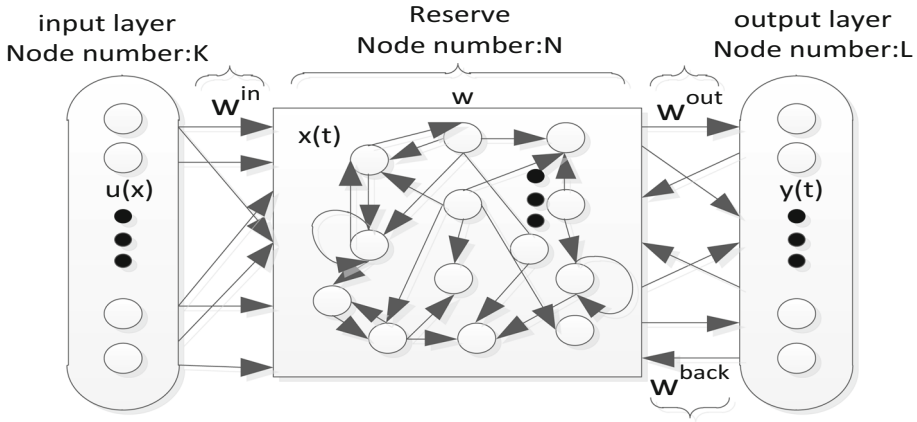

Fig. 3 A standard echo state network structure

$$D_{KL}(p(x), p_d(x)) = -H(x) + \frac{1}{2\sigma^2}E\left((x-u)^2\right) + \log\frac{1}{\sigma\sqrt{2\pi}} \tag{15}$$

Here, $H(x)$ represents an entropy value. The second term is the deviation between the true output x of the reservoir and the Gaussian mean u. Therefore, we can see that the increase in $H(x)$ and the decrease in the difference between u and x can minimize D_{KL}. In order to achieve the optimization of minimum distance, we introduce the gain parameter a and the offset b. IP training is a process of constantly adjusting parameters a and b by using gradient descent in the process of minimization D_{KL}.

$$x_e = \tanh(\mathrm{diag}(a)w_{\mathrm{in}}u_c(t) + \mathrm{diag}(a)wx_e + b) \tag{16}$$

In (16), for any input constants $u(c)$, we can get the final steady state x_e of the reservoir through training. Due to a given value $u_c(t)$ for each input node, we can combine it with the offset b to obtain (17).

$$\begin{aligned} x_e &= \tanh\left(\mathrm{diag}(a)wx_e + b_{\mathrm{in}(c)}\right) \\ b_{\mathrm{in}(c)} &= \mathrm{diag}(a)wu_c(t) + b \end{aligned} \tag{17}$$

In the paper [36], the key parameter setting is $\mu = 0, \sigma = 0.1$, the learning rate η in the gradient descent process is 0.0005, the initial state of the adjustable parameter (a, b) can be set to be $a = 1, b = 0$. When the state variable does not change significantly, the training ends. Because the IP algorithm is an unsupervised training process, the state vector in the entire reservoir is an intrinsic reflection of each input sample structure, so there is no occurrence of overfitting problems in training.

3.2.3 State node selection in reservoir

The literature [26] points out that the sample vectors whose input spaces are close to each other are sent to the reservoir; then, the state nodes are also close to each other. Therefore, this paper makes use of such a discriminative effect on state nodes in the reservoir as low-dimensional features to perform emotion recognition. The specific process is as follows:

Step 1 Randomize the initial reservoir based on the dimensions of the initial sample vector.

Step 2 Use the IP algorithm to train each input sample.

Step 3 Collect the steady-state vectors of the reservoir in each time period, and then, calculate the probability density distribution of each state node.

Step 4 Select the first c state node neurons with higher number of local maxima in their probability density distribution in the reservoir; then, combine these neurons into c-dimensional feature vector and send them to support vector machine (SVM) for classification and identification.

4 Experiments results and analysis

4.1 DEAP emotion databases

In the paper, we do emotional analysis in the DEAP database. Please refer to the literature [28] for the database collection process. Thirty-two subjects with physical and mental health (50% male to female, mean age 26.9 years) were invited to participate in the collection of experimental data. A total of 40 channel signals were recorded, of which the EEG signals was 32 channels. Eight channels are peripheral physiological signals. The data were recorded at a sampling frequency of 512 Hz; then, it was down-sampled to 128 Hz and finally filtered by the filter to 4–45 Hz. The artifacts of the EEG were removed by independent component analysis. The 32 channels were fixed on the scalp in accordance with the International 10–20 system standard. Thirty-two participants were scheduled to record physiological signals and face recordings in 40 different emotional music videos. Each video lasted 1 min. Before each video, the subjects will have a 3-s baseline preparation time, and they will be in a calm state in the baseline time. For each music video, we get 20 different AsI, and then, we filter out strong emotional motivation signals for emotion recognition according to the rules in Sect. 2.3. After each video is played, participants will have sufficient time to evaluate emotional degree from 1 to 9 in the 3D continuous model (valence, arousal, dominance) of the video just played. In this paper, we choose two emotional dimensions: valence and arousal, and we can classify four emotional categories, including HVHA (high valence high arousal), LVHA (low valence high arousal), HVLA (high valence low arousal), and LVLA (low valence low arousal), corresponding to the four main emotions which are happy/joy, anger/disgust, relaxation/calm and depression/sadness.'

Support vector machine (SVM) is a supervised classification model that can be used to predict the target value of unlabeled data. In this paper, we use Libsvm toolbox to classify emotional physiological signals. The kernel function chosen in this paper is Gaussian radial basis function (RBF), $K_{\mathrm{RBF}}(X_i, X_j) = \exp\left(-\frac{\|X_i - X_j\|}{2\sigma^2}\right)$, where σ is the

width parameter of the function; in addition, the parameters g and the penalty factor c in the training process are optimized using particle swarm optimization (PSO).

4.2 Influence of different window sizes on classification accuracy

To explore the sensitivity of physiological signal time window size to classification accuracy in emotion recognition, each window is set to 2, 3, 4, 6, 8 s, respectively, and the number of sliding steps is incremented by the step length of 0.5 s. Table 1 is based on the classification effect of the SVM classifier under different classification windows, where the reserve pool size is set to 500. The classification accuracy takes the average value of different clustering number k.

From Table 1, it can be clearly seen that when the window length is 1–4 s, the classification accuracy increases roughly with the increase in time; when window length is more than 4 s, the classification precision begins to decrease with the increase in time. When window length is 8 s, the accuracy decreases significantly. In addition, when the sliding step length is near the center value of the window length, the higher precision of the given window length can often be achieved. In this paper, when window length is 4 s and the sliding step is 2.5 s, the average classification precision reaches the maximum value of 70.5%. Therefore, we believe that in the identification process of physiological emotion signal, we should compromise the overall representation and detailed description of the information characteristic of biological signal and also take into account the coherent expression of the information of pre- and post-signal windows.

Table 1 The average classification accuracy (%) of different window lengths under their respective sliding steps

Sliding steps (s)	Window length (s)					
	1	2	3	4	6	8
0.5	61.4	63.2	62.0	59.9	62.8	58.3
1	62.1	65.0	66.1	61.2	60.3	61.1
1.5		66.7	69.7	65.1	65.4	60.4
2		64.9	63.2	68.5	67.2	63.5
2.5			66.7	70.5	63.3	62.0
3			65.3	68.0	64.1	64.3
4				62.9	61.5	61.7
6					62.9	61.4
8						63.2

4.3 Comparison of AsI and unAsI in classification performance

In this paper, we used the physiological signals after AsI and those without AsI for comparative analysis, respectively. In order to more fully analyze the effectiveness of AsI in emotion classification, in the training process, we also used the SID (subject independent) and SD (subject dependent) to compare the recognition results, as shown in Table 2, corresponding to Fig. 4. The proposed method is abbreviated to PM, as well as original features represented as org. SD refers to training samples and test samples from the same subjects and then takes the average of all subjects tested; SID refers to training samples and test samples, which may be separately from different subjects, and then take the average of the results of all trials; in the sense of research, SID focuses on the common expression of emotion analysis, so the research on SID can be better promoted in the BCI system and can meet certain real-time requirements. In this experiment, leave-one-out cross-validation and 13-fold cross-validation method are adopted, respectively. Among them, the leave-one-out cross-validation method is to use $n-1$ of the selected n subjects as the training set and the remaining one as the verification. For 13-fold cross-validation, we randomly divided the entire data set into 13 and repeated 120 times. Here we take the average of the two as the result.

As can be seen from Fig. 4, recognition accuracy of the physiological signals processed using AsI obtains a significant improvement, and when the number of clusters is 4, a higher classification recognition rate is obtained, which may be related to brain functional area cooperation of each other of each sub-band. In the case of SID (Fig. 5a), the average accuracy on the three kinds of feature expression is 65.7, 65.5 and 70.5%, respectively, which is 9.9, 6.4 and 8.7% higher than unAsI; and for SD (Fig. 5b), the average accuracy of the three characteristics was 69.2, 65.4 and 71.9%, respectively, which was 5.5, 4.2 and 6.2% higher than that of unAsI; thus, it can be clearly demonstrated that AsI can effectively choose some physiological signals with strong emotional stimulation, and these signals can successfully capture some information features related to emotions, which have greatly promoted the classification accuracy. In addition, for SID, the improvement rate of accuracy rate is higher than that of SD, which shows that SID has stronger fault tolerance with stronger emotion signals. At the same time, for AsI (in Fig. 5c), the accuracy rate with SD is slightly greater than SID in the case of using the reservoir state features selected, but the overall accuracy rate is not significantly different; for unAsI (Fig. 5d), the accuracy rate of SD versus SID has undergone a significant improvement. This is likely because SID

Table 2 Recognition rates (%) of the SVM trained for AsI and unAsI in SID and SD (numerical values associated with Fig. 4)

Cluster number	AsI						unAsI					
	SID			SD			SID			SD		
	Org	All ESN	PM	Org	All ESN	PM	Org	All ESN	PM	Org	All ESN	PM
K=2	59.2	62.4	69.9	62.3	61.2	68.4	51.2	54.8	56.2	62.2	59.7	65.1
K=3	62.3	65.8	67.2	**72.1**	68.3	71.9	53.3	57.7	58.0	65.5	62.4	68.9
K=4	69.8	**67.0**	**73.8**	71.3	**69.7**	75.3	**60.4**	60.6	64.3	**66.2**	62.1	**69.4**
K=5	**71.7**	65.2	71.8	70.9	68.2	71.3	56.3	59.2	61.5	64.3	**65.4**	64.0
K=6	68.3	66.9	70.2	71.3	69.2	**76.1**	55.0	61.1	63.4	62.7	61.1	65.4
K=7	63.2	64.8	72.3	69.1	62.5	70.3	54.6	**61.4**	**61.8**	63.4	57.1	62.0
K=8	65.7	65.9	68.0	67.5	58.4	69.7	56.9	59.0	61.2	61.3	60.7	65.2
On average	65.7	65.5	70.5	69.2	65.4	71.9	55.4	59.1	60.8	63.7	61.2	65.7

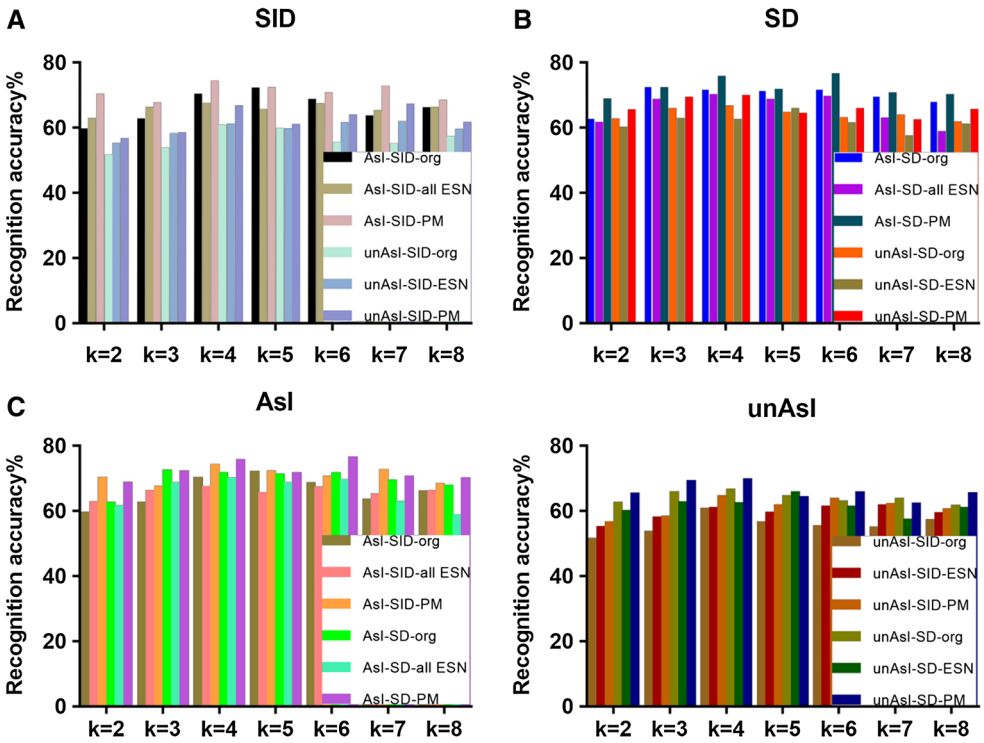

Fig. 4 Recognition rates of the SVM trained for AsI and unAsI in SID and SD

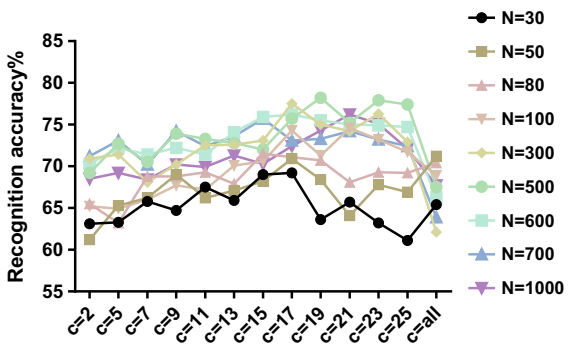

Fig. 5 Recognition rates (%) of the SVM trained by ESN reservoir state nodes extraction in different reservoir sizes

considers the addition of different participants' signals during the training process which is equivalent to carrying out emotional commonality training from different individual, and this means that the AsI obtained by preprocessing can effectively select the physiological signals related to emotional stimulation. In addition, it can be seen from Fig. 4 that whether we have an AsI process or not, the classification effect of the reservoir state features selected is higher than the classification accuracy under the corresponding original features.

4.4 Comparison of classification recognition rate under different reservoir sizes *N*

In order to facilitate the analysis of the effect of reservoir size on the classification accuracy, we consider the case of the SID after the AsI processing, the K in the fixed feature clustering process is 4 and the number N of reservoir nodes is 30, 50, 80,100, 300, 500, 600,700 and 1000. Through the selective extraction of state node information in the reserve pool, the feature can be further compressed and integrated to reduce the feature dimension to the maximum and reduce the redundancy between the features. c range is set to [2, 25], the step size is 2 and the experimental results are shown in Table 3, corresponding to Fig. 5.

From Fig. 5, when the number of nodes selected is about 19, emotion classification can achieve good results. In order to further explore the effect of the state node features in the reservoir on the classification accuracy, this paper selects all the state nodes under the ESN model for classification, and it is found that when the classification precision reaches the maximum in the selected node, the classification accuracy decreases with the increase in the number of selection nodes in the reservoir. This is probably due to the increase in the number of nodes in the reservoir, and the degree of nonlinearity between the nodes in the reservoir is too complex, which leads to more clustering between nodes and the occurrence of overfitting. In addition, we can see that the classification accuracy is up to 78.2% when the scale is N=500. The increase in the size of the reservoir will lead to a significant decrease in the computational efficiency, so it is not recommended to use large-scale ESN for modeling in the application of ER-MCI.

4.5 Accuracy comparison under different classification method

In this article, we make a comparison of the classification performance from the features and the classification methods. From Table 4, the following abbreviations are used: SPF: spectral power features, AI: asymmetry index based power, WPCD: wavelet packet coefficient distribution; we can see that SPF- and AI-based power [16, 18], energy ratio, Hjorth parameter, C0 complexity are used to compare with WPCD; in the case of choosing the appropriate window length and the clustering coefficient k, WPCD can achieve a relatively good classification performance. In the selection of the classification methods, we choose SVM, KNN, ESN [37], KNN+RF [38] and the proposed method. In the case of unAsI or AsI, the proposed scheme can achieve the highest classification accuracy of 72.3 and 78.2%.

5 Conclusions

First, due to subjective or objective reasons, some individuals cannot be inspired by music video to stimulate the effective emotion; we put forward a pre-processing scheme before the emotion recognition; this scheme makes full use of 3 s baseline signals of DEAP database. After the calculation of the AsI, the signal segments with low emotional motivation under the music video or the individual is filtered, and effective emotional stimulus signals are reserved for feature extraction and emotion classification. Second, we choose the 32-channel EEG signals and 4-channel peripheral physiological signals (EOG and

Table 3 Recognition rates (%) of the SVM trained by ESN reservoir states extraction algorithm in different reservoir sizes (numerical values associated with Fig. 5)

State node c	Number of nodes in reservoir N								
	30	50	80	100	300	500	600	700	1000
c=2	63.1	61.2	65.5	65.2	70.9	69.2	70.3	71.2	68.5
c=5	63.3	65.3	63.2	64.9	71.4	72.7	72.2	73.1	69.2
c=7	65.8	66.2	68.8	66.0	68.1	70.5	71.4	70.2	68.4
c=9	64.7	69.0	68.8	67.7	70.2	73.9	72.2	74.2	70.2
c=11	67.5	66.2	69.3	67.2	72.5	73.3	71.4	72.4	69.9
c=13	65.9	67.1	67.9	70.1	72.6	72.8	74.1	73.6	71.3
c=15	69.0	68.2	**71.5**	70.6	73.1	72.0	75.9	**75.7**	70.3
c=17	**69.2**	70.9	71.1	74.3	**77.5**	75.7	**76.2**	73.1	72.4
c=19	63.6	**68.4**	70.7	71.3	75.0	**78.2**	75.5	73.3	74.3
c=21	65.7	64.1	68.1	**74.6**	74.2	75.2	75.0	74.2	**76.2**
c=23	63.2	67.8	69.3	73.2	76.3	77.9	74.9	73.2	75.1
c=25	61.1	66.9	69.2	71.7	72.9	77.4	74.7	72.4	72.1
c=all	65.4	71.2	70.4	68.9	62.1	67.5	66.3	63.9	67.7
平均	65.1	66.8	68.6	69.7	72.9	74.1	73.7	73.0	71.3

Table 4 The classification accuracy (%) with different features and classification schemes

Feature	Method	unAsI	AsI
SPF/AI [16, 28]	SVM	52.2	57.8
SPF/AI	KNN	49.5	55.2
WPCD	SVM	55.1	59.7
SPF/AI	ESN [37]	61.6	65.9
WPCD	ESN	64.9	68.2
Energy ratio, Hjorth parameter, C0 complexity, variance and spectral entropy [38]	KNN+RF	70.0	
WPCB	The proposed method	72.3	78.2

EMG) and use the wavelet packet transform to gain wavelet packet coefficients, and the coefficients are clustered, and the probability distribution values in each subband are obtained as the original feature, and then, these original features are sent to the ESN for training, respectively. Finally, the appropriate state nodes are selected in the reservoir as the final features to be sent to the SVM.

Through the comparison and analysis of the experimental results, it is found that (a) after AsI processing, the accuracy of the emotional classification is 8.73% higher than that without AsI, which is an effective scheme for preprocessing of EEG emotion classification. (b) When the clustering number is 4, the effect of emotion classification can reach the best. (c) By training and selecting the state nodes in the reservoir, we can use these node features to achieve 4.8% performance improvement compared with original features in emotion classification. The feature processing method comes down to a feature perception compression process based on the black box in ESN. It can not only reduce the feature dimension, but also re-integrate the feature vector. The experimental results show that the experimental model can enhance the robustness of MCI emotional decision making and can play a very good auxiliary role in psychological medical and emotional care. In the next step, we will explore the features interaction of each brain region at different frequencies and the influence of different classification models on EEG emotion classification.

Acknowledgements This research has been partially supported by National Natural Science Foundation of China (Grant No. 61432004), NSFC-Shenzhen Joint Foundation (Key Project) (Grant No. U1613217).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Vilar P (2014) Designing the user interface: strategies for effective human–computer interaction (5th edition). Inf Process Manage 61(5):1073–1074
2. Andreasson R, Alenljung B, Billing E et al (2018) Affective touch in human–robot interaction: conveying emotion to the nao robot. Int J Social Robot 3:1–19
3. Zhang Z, Tanaka E (2017) Affective computing using clustering method for mapping human's emotion. In: IEEE international conference on advanced intelligent mechatronics. IEEE, pp 235–240
4. Fragopanagos N, Taylor JG (2005) Emotion recognition in human–computer interaction. Neural Netw 18(4):389
5. Hu M, Zheng Y, Ren F et al (2015) Age estimation and gender classification of facial images based on Local Directional Pattern. In: IEEE international conference on cloud computing and intelligence systems. IEEE, pp 103–107
6. Ren F, Huang Z (2016) Automatic facial expression learning method based on humanoid robot XIN-REN. IEEE Trans Hum Mach Syst 46(6):810–821
7. Wang K, An N, Li BN et al (2017) Speech emotion recognition using Fourier parameters. IEEE Trans Affect Comput 6(1):69–75
8. Camurri A, Camurri A, Camurri A (2016) Adaptive body gesture representation for automatic emotion recognition. ACM Trans Interact Intell Syst 6(1):6
9. Ren F (2009) Affective information processing and recognizing human emotion. Elsevier Science Publishers B.V., Amsterdam
10. Ren F, Wang L (2017) Sentiment analysis of text based on three-way decisions. J Intell Fuzzy Syst 33(1):245–254
11. Petrantonakis PC, Hadjileontiadis LJ (2012) Adaptive emotional information retrieval from EEG signals in the time–frequency domain. IEEE Trans Signal Process 60(5):2604–2616
12. Yoon HJ, Chung SY (2013) EEG-based emotion estimation using Bayesian weighted-log-posterior function and perceptron convergence algorithm. Comput Biol Med 43(12):2230–2237
13. Davidson RJ, Ekman P, Saron CD et al (1990) Approach-withdrawal and cerebral asymmetry: emotional expression and brain physiology: I. J Pers Soc Psychol 58(2):330
14. Davidson RJ, Schwartz GE, Saron C, Bennett J, Goleman DJ (1979) Frontal versus parietal EEG asymmetry during positive and negative affect. Psychophysiology 16:202–203
15. Zheng WL, Lu BL (2015) Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. IEEE Trans Auton Ment Dev 7(3):162–175
16. Daimi SN, Saha G (2014) Classification of emotions induced by music videos and correlation with participants' rating. Expert Syst Appl 41(13):6057–6065

17. Sakata O, Shiina T, Saito Y (2002) Multidimensional directed information and its application. Electron Commun Jpn 85(4):45–55
18. Petrantonakis PC, Hadjileontiadis LJ (2011) A novel emotion elicitation index using frontal brain asymmetry for enhanced EEG-based emotion recognition. IEEE Trans Inf Technol Biomed 15(5):737–746
19. Sakata O, Shiina T, Satake T et al (2006) Short-time multidimensional directed coherence for EEG analysis. IEEJ Trans Electr Electron Eng 1(4):408–416
20. Deshpande G, Laconte S, Peltier S et al (2006) Directed transfer function analysis of fMRI data to investigate network dynamics. In: International conference of the IEEE engineering in medicine and biology society, p 671
21. Xu X, Ye Z, Peng J (2007) Method of direction-of-arrival estimation for uncorrelated, partially correlated and coherent sources. Microw Antennas Propag IET 1(4):949–954
22. Roebroeck A, Formisano E, Goebel R (2005) Mapping directed influence over the brain using Granger causality and fMRI. Neuroimage 25(1):230–242
23. Jaeger H (2001) The "echo state" approach to analysing and training recurrent neural networks. Technical report GMD Report 148. German National Research Center for Information Technology
24. Jaeger H (2002) Tutorial on training recurrent neural networks, covering BPPT, RTRL, EKF and the echo state network approach. GMD-Forschungszentrum Informationstechnik, Bonn
25. Han M, Xu M (2018) Subspace echo state network for multivariate time series prediction. IEEE Trans Neural Netw Learn Syst 29(1):238–244
26. Koprinkova Hristova P, Tontchev N (2012) Echo state networks for multi-dimensional data clustering. In: International conference on artificial neural networks and machine learning. Springer-Verlag, pp 571–578
27. Fourati R, Ammar B, Aouiti C et al (2017) Optimized echo state network with intrinsic plasticity for EEG-based emotion recognition. In: International conference on neural information processing. Springer, Cham, pp 718–727
28. Koelstra S, Muhl C, Soleymani M et al (2012) DEAP: a database for emotion analysis; using physiological signals. IEEE Trans Affect Comput 3(1):18–31
29. Kanungo T, Mount DM, Netanyahu NS et al (2002) An efficient *k*-means clustering algorithm: analysis and implementation. IEEE Trans Pattern Anal Mach Intell 24(7):881–892
30. Hartigan JA (1979) A *k*-means clustering algorithm. Appl Stat 28 (1):100–108
31. Huang Z (1998) Extensions to the *k*-means algorithm for clustering large data sets with categorical values. Data Min Knowl Disc 2(3):283–304
32. Zheng WL, Zhu JY, Lu BL et al (2016) Identifying stable patterns over time for emotion recognition from EEG. IEEE Trans Affect Comput 99(1949):1. https://doi.org/10.1109/TAFFC.2017.2712143
33. Lin YP, Wang CH, Jung TP et al (2010) EEG-based emotion recognition in music listening. IEEE Trans Biomed Eng 57 (7):1798–1806
34. Jenke R, Peer A, Buss M (2017) Feature extraction and selection for emotion recognition from EEG. IEEE Trans Affect Comput 5 (3):327–339
35. Yin Z, Wang Y, Liu L et al (2017) Cross-subject EEG feature selection for emotion recognition using transfer recursive feature elimination. Front Neurorobot 11:19
36. Schrauwen B, Wardermann M, Verstraeten D et al (2008) Improving reservoirs using intrinsic plasticity. Neurocomputing 71(7–9):1159–1171
37. Skowronski MD, Harris JG (2007) Special issue: automatic speech recognition using a predictive echo state network classifier. Elsevier Science Ltd, Amsterdam
38. Chen J, Hu B, Wang Y et al (2017) A three-stage decision framework for multi-subject emotion recognition using physiological signals. In: IEEE international conference on bioinformatics and biomedicine. IEEE, pp 470–474