

Affective Recognition Using EEG Signal in Human-robot Interaction

Chen Qian, Tingting Hou, and Shan Fu

School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai, 200240, P.R. China
{qian_chen}@sjtu.edu.cn

Abstract. TODO : this part is waiting to be rewritten.

Keywords: Affective Computing; Arm-Robot Operation; Time Domain Features; Domain Feat

1 Introduction

Hitherto, mechanical arm, as a vital product in industry field, has been broadly used in medical, exploration, rescue field etc. However, the mistaken operation caused by human error, which should have been avoided, is still one of the dominant reasons causing accidents. As we all know, the performance of human has a close link to the cognitive state of human and to some degree emotion is the main reason causing the change of the cognitive state. Therefore, one of the critical ways to avoid huamn error in mechanical arm operation is recognizing the emotion during the opreration process through detecting the physiological signal. Since R. W. Picard has defined *Affective Computing (AC)*^[1] in 1995, affective computing has been a critical field in human-computer interaction area. There are numerous physiological signals which could reflect the cognitive state of human, such as Blood Volume Pressure (BVP), Skin Conductance Response (SCR), Respiration (RESP), Electrocardiogram (ECG), Electromyogram (EMG), Electrocorticogram (ECoG), Electroencephalogram (EEG), Heart Rate (HR), Oxygen Saturation (SaO₂) and Surface Temperature (ST)[2]. In all physiological signals, EEG is no doubt the most capable signal directly reflecting the brain activity. Therefore this paper chooses to use 32 dry electrodes EEG signal acquisition equipment to obtain EEG signal data during mechanical arm operation.

There are lots of models which have been proposed to describe the emotion, such as six basic emotions model proposed by Ekman et al.[3], eight basic emotions model proposed by Plutchik[4] and the valence-arousal scale proposed by Russell[5]. For simplifing this problem, this paper chooses to use the valence in the valence-arousal model of Russell to evaluate the emotion of the subjects. And the valence reflecting the positive or negative aspect of the subjects is enough to describe the cognitive state of the subjects.

Besides the model of emotion, the way to obtain the ground truth of the subjects cognitive state is also crucial. Almost all research in Affective Computing field use the self-assessment scores to estimate the true cognitive state of the subjects. However, even the subjects themselves could hardly to retell the exact emotion state in the mechanical arm operation and using one single scores to estimate the cognitive state during the entire operation process is obviously not reasonable in detail. So this paper proposes to use objective and real time indexes to represent the cognitive state of the subjects. In this paper, we record the track of the end point of the mechanical arm and extract the features of the track to represent the fluctuation of the cognitive state of the subjects. Meanwhile, we assume that the workload and the time pressure could stimulate the emotion change of the subjects, so we give a basic score, which reflects the emotion state the subjects should be, and add the weighted features scores to the basic score to reflect the fluctuation of the emotion state.

In our experiment, we designed three levels of operating tasks in different difficulty to stimulate the emotion state change. And the level of difficulty is determined by the workload and the time pressure. To eliminate the influence of unfamiliarity, one minute of free exploration is added before these three tasks and to eliminate the interaction of different tasks, a 30 seconds reset time interval is added between the different tasks.

During the data process part, because of the low signal-to-noise ratio (SNR), the disturbance of EMG signal and the electromagnetic interference, the raw data have firstly been filter to the 1-64Hz frequency band[6]. After normalization process, different scales sliding window are induced in extracting features. Because of the real-time label we obtain from the track mentioned above, it allows us to consider the data in single sliding window as one sample.

The feature extraction methods are detailed summarized in paper[6], we choose three time domain features and one frequency feature to represent the raw data according to the value of the weighted relative occurrence. And at the feature selection process, we apply Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) to select the extracted features above. And at the classification design process, we apply Support Vector Machine (SVM), which is an effective classification discriminator, to predict the emotion state.

Here is the remainder organization of this paper: Section 2 introduce the apparatus used in the experiment and the detail experiment protocol. Section 3 describes the data preprocess procedure, feature extraction, feature selection classification design methods. Section 4 focus on showing the data process results. Section 5 discusses the results. Section 6 gives the conclusion of our experiment.

2 Experiment Setup

2.1 Apparatus

There are 3 key equipments we use in our experiment: EEG Signal Acquisition Equipment, Mechanical Arm and Joystick. And there are 3 personal computers

for communicating with EEG equipment, controling movement of mechanical arm through joystick and showing the end point of the mechanical arm.

EEG Signal Acquisition Equipment: The EEG Signal Acquisition Equipment we apply is the *Cognionics HD-72 Dry EEG Headset*^[7] (see in Fig. 1). Considering the difficulty of wearing EEG equipment and the comfort of the subjects, we choose 32 dry electrodes, according to the international 10-20 system, to obtain the EEG signal, which are showed in Fig. 2. The sampling rate is 500Hz which is sufficient for obtaining EEG signal.



Fig. 1. Cognionics HD-72 Dry EEG Headset

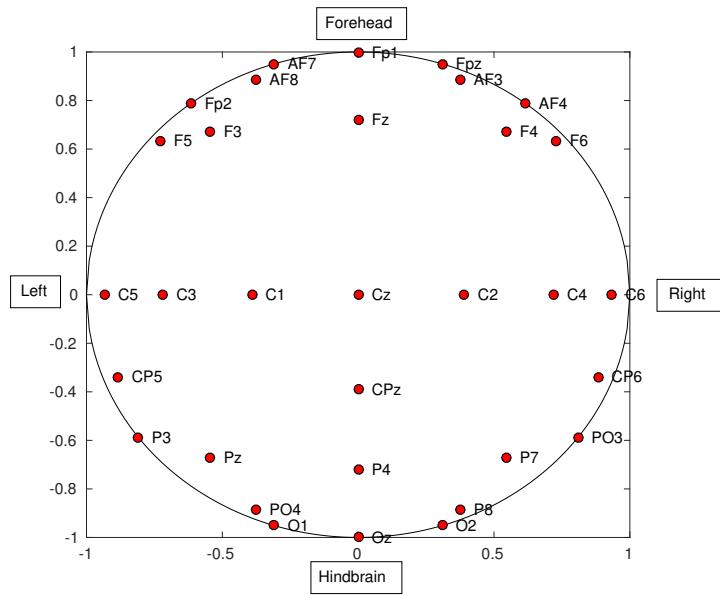


Fig. 2. 32 dry electrodes sensor location

Mechanical Arm: The mechanical arm we apply is the *Dobot Magician mechanical arm*^[8] (see in Fig. 3), because it supports reprogramming according to the need of users and the control precision (0.2mm) is adequate for our experiment.



Fig. 3. Dobot Magician mechanical arm

Joystick: The joystick we use is the flying joystick named *Extreme 3D Pro*^[9] produced by Logitech (see in Fig. 4). The perfect ergonomic design with a custom twist-handle rudder relies its one-handed control resulting in a smaller device footprint. There are six programmable buttons on the base. Each programmable button can be configured to execute simple single commands or intricate macros involving multiple keystrokes, mouse events, and more. In our experiment, we only operate the rocker in six basic direction movement, which are left-right direction, front-behind direction and left-right rotation.



Fig. 4. Logitech Extreme 3D Pro joystick

In our experiment, we use Robot Operating System (ROS) to receive the control signal from joystick, release the control signal to mechanical arm, receive the position information of the end point of the mechanical arm and record the track information with real time mark point in a sampling rate of 10Hz (code url: https://github.com/chenyuxuan/EEG_Human_Robot_Interaction).

<https://github.com/QCH1993/DobotMagician>). Meanwhile, real time EEG signal are recorded with mark information which could be used in subsequent time correcting process.

2.2 Experiment Protocol

Five healthy participants (two females, three males), aged between 23 and 25, participated in the experiment. Before all experiments, all participants were asked to have adequate sleepiness.

In our experiment, firstly, participants were asked to read the experiment procedures and notes and the experimenter would read out the procedures and notes for a second reminder. And the experimenter was also present there to answer any questions. At the beginning of each experiment, one minute of free exploration were given to remove the interference caused by unfamiliarity. And then the subjects were asked to operate the mechanical arm to touch the different color point on the desktop according to a certain order. We designed three levels of operating tasks in different difficulty (easy, medium and hard mode) according to the number and position of points and the time constraint. In easy mode, the subjects were asked to touch three colored points without time pressure. In medium mode, the subjects were asked to touch five colored points without time pressure. In hard mode, the subjects were asked to touch five colored points in 90 seconds. To eliminate the interplay between the three tasks, a 30 seconds break time for resetting was added after each task.

The experiment environment is showed in Fig. 5. The Fig. 5(a) is the overview of the entire experiment environment. The subject operation platform (see in Fig. 5(b)) is insulated from the mechanical arm platform and all the information helping the subjects to move the mechanical arm was from three camera set around the mechanical arm and on the end point of the mechanical arm (see in Fig. 5(c)). The screen interface presenting the information from cameras are showed in Fig. 5(d).

3 Data Preprocess Methods

In data science field, data preprocess is the key procedure before any data analysis procedure. The quality of data preprocess directly affects the final accuracy of recognition. In EEG experiment, because of the low signal noise rate (SNR) and various interference, filtering and amplifying EEG signal is a crucial step before EEG data analysis. And in our experiment, we need to do some label data preprocess to obtain more objective real time label of EEG data because of the use of sliding window which would be mentioned in Section 4 and the need of combining self-assessment with objective data.

3.1 EEG Data Preprocess

For one subject experiment, the raw data are drawn in Fig. 6(a). And then according to the track mark and EEG signal mark, the EEG signal and the

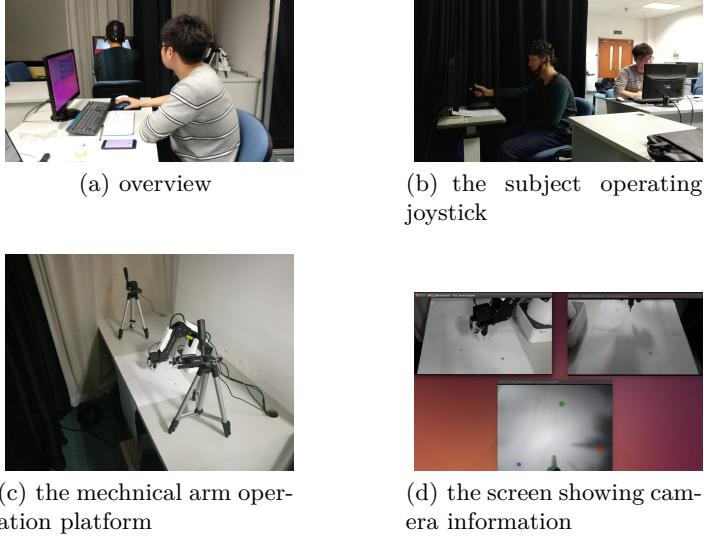


Fig. 5. The experiment environment

track were aligned in time. And we divided the entire EEG signal to easy mode, medium mode and hard mode procedure according to the mark information. The divided easy mode raw data, as an example, are showed in Fig. 6(b). Next, to remove the disturbance of EMG signal and the interference of various electromagnetic signal which are generally high frequency signals, we use band filter to filter out 1-64Hz signal, which are the dominant frequency band of EEG signal(see in Fig. 6(c)). Because the EEG signal has extra low SNR, the filtered signal at the beginning and end part is unstable. Therefore we selected to remove the first and last 5 seconds data (see in Fig. 6(d)). And for the computing convenience, the selected data are normalized to [-1,1] area according to all 32 channels signal by min-max normalization.

3.2 Label Data Preprocess

Label data includes three parts in total: the self-assessment of subjects, the difficulty level experimenters design and the performance the track reflects. The self-assessment is the subjective label data and the difficulty level and the performance are the objective data.

For the self-assessment, after the entire experiment were finished, each subjects were asked to fill a table about their emotion state during operating different tasks. There were seven level the subjects could choose. "1" means the most positive emotion state, "4" means neutral emotion state and "7" means the most negative emotion state. The greater the number, the worse the emotion state. And then we used each score to weighted each task.

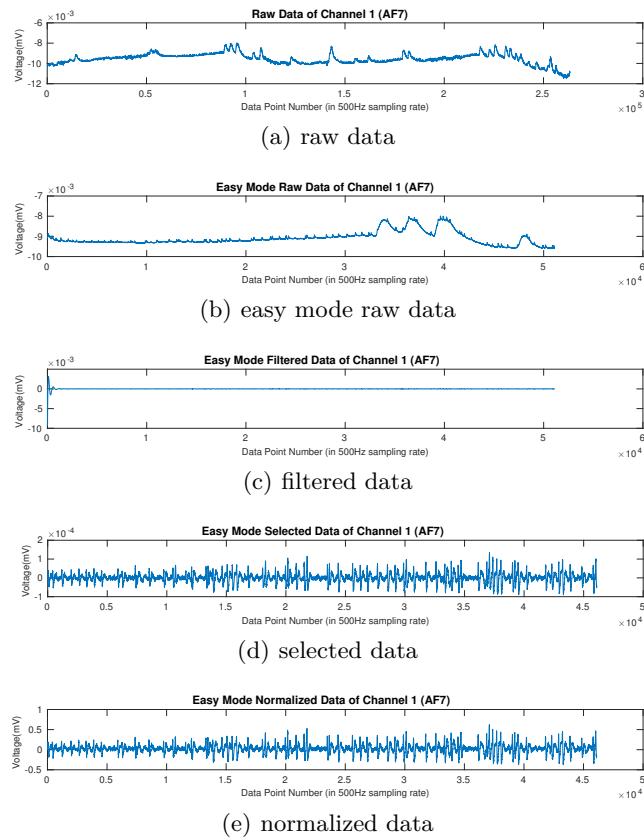


Fig. 6. Data preprocess wave charts

For the the difficulty level experimenters desgin, we simply gave the scores of difficulty according to the workload and the time pressure. In easy task, we asked the subjects to touch three points without time constraint, which means that we assumed the most positive emotion state could be stimulated by this task. On the contrary, in hard task, we asked the subjects to touch five points with a ninety seconds constraint, which means that we assumed the most negtive emotion state could be stimulated by this task. We used "2.5", "4" and "5.5" to represent the emotion states different difficulty levels we assumed could stimulate.

For the performace, we recorded the tip trajectory of mechanical arm. As showed in Fig. 7, in easy mode, the subjects were asked to touch orange point, purple point and pink point in order from a random start point. We assumed that the subjects would feel positive when they moved the mechanical arm smoothly, so we counted the numbers of direction change, in selected sliding window mentioned in Section 4, as the estimation of the emotion state of the subjects.

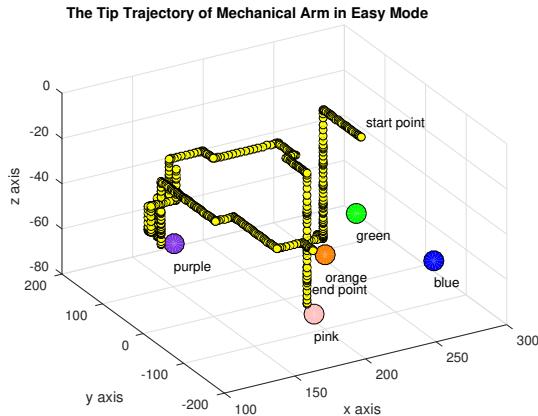


Fig. 7. The tip trajectory of mechanical arm in easy mode

We use the formula below to caculate the final label.

$$L = \frac{1}{3}(L_{self-assessment}) + \frac{1}{3}(L_{difficulty-level}) + \frac{1}{3}(L_{direction-change}) \quad (1)$$

where $L_{self-assessment}$ means the self-assessment score, $L_{difficulty-level}$ means the difficulty level score, $L_{direction-change} = 3 \frac{N_{direction-change} - N_{min}}{N_{max} - N_{min}} - 1.5$ means the direction change score, among them, $N_{direction-change}$ means the number of direction changes, N_{min} means the minium of $N_{direction-change}$ and N_{max} means the maxium of $N_{direction-change}$.

4 Affective Recognition Methods

Traditional pattern recognition process includes 4 steps: data preprocess, feature extraction, feature selection and classification design. In this paper, we use several classic methods in feature extraction, feature selection and classification design procedure to explore their performance in our experiment data.

4.1 Multi-scale Feature Extraction

The multi-scale feature extraction procedure are divided in two steps: 1) sliding window selection and 2) feature extraction.

Sliding Window Selection Because uncertain relation between time interval and emotion state change, we chose 2s, 4s, 6s, 8s, 10s as the length of sliding window and 0.2s as the length of stride to extract multi-scale features.

Feature Extraction According to the paper[6], we chose 18 dimensions time domain features, which are statistical features, higher order crossings (HOC) features and fractal dimension feature, and 53 dimensions frequency domain features, which are band power, bin power and the ratio of mean band powers β/α .

Time Domain Features

In this paper, we chose seven *statistical features* to represent the raw EEG time series. These are:

- a) Mean: $\mu_{\eta} = \frac{1}{T} \sum_{t=1}^T \eta(t)$
- b) Power: $P_{\eta} = \frac{1}{T} \sum_{t=1}^T |\eta(t)|^2$
- c) Standard deviation: $\sigma_{\eta} = \sqrt{\frac{1}{T} \sum_{t=1}^{T-1} (\eta(t) - \mu_{\eta})^2}$
- d) 1st difference: $\delta_{\eta} = \frac{1}{T-1} \sum_{t=1}^{T-1} |\eta(t+1) - \eta(t)|$
- e) Normalized 1st difference: $\bar{\delta}_{\eta} = \frac{\delta_{\eta}}{\sigma_{\eta}}$
- f) 2nd difference: $\gamma_{\eta} = \frac{1}{T-2} \sum_{t=1}^{T-2} |\eta(t+2) - \eta(t)|$
- h) Normalized 2nd difference: $\bar{\gamma}_{\eta} = \frac{\gamma_{\eta}}{\sigma_{\eta}}$

To find more robust features and pattern of EEG, Petrantonakis and Hadjileontiadis proposed *higher order crosssections (HOC)* features[10]. Therefore we applied HOC to represent the raw EEG time series. The higher order series could be caculated by the formula below:

$$H_k\{\eta(t)\} = \nabla^{k-1} \eta(t) \quad (2)$$

where ∇ means $\eta(t) - \eta(t-1)$. Therefore, the features of HOC could be caculated by counting the sign changes or equation below:

$$F_k = \sum_{t=1}^{T-k} \psi(H_k\{\eta(t)\} H_k\{\eta(t+1)\}), k = 1, 2, \dots, 10 \quad (3)$$

where $\psi(t)$ is a section function which is defined below:

$$\psi(x) = \begin{cases} 0 & \text{if } x \geq 0 \\ 1 & \text{if } x < 0 \end{cases} \quad (4)$$

The *fractal dimention (FD)* as a feature measuring the complexity is widely used. There are many mehtods computing the FD feature. In this paper, we chose to use the Higuchi algorithm[11] to caculate the FD feature. To compute the FD feature, the EEG series is rewritten as:

$$\{\boldsymbol{\eta}(p), \boldsymbol{\eta}(p+q), \boldsymbol{\eta}(p+2q), \dots, \boldsymbol{\eta}(q + [\frac{T-q}{q}]q)\}, p = 1, 2, \dots, q \quad (5)$$

where $[x]$ means the largest integer less than x . Then we could define the series $M_p(q)$ as below:

$$M_p(q) = \frac{T-1}{[\frac{T-p}{q}]q^2} \sum_{k=1}^{[\frac{T-p}{q}]} |\boldsymbol{\eta}(p+kq) - \boldsymbol{\eta}(p+(k-1)q)| \quad (6)$$

Therefore we could further define the average:

$$F(q) = \frac{1}{N_p} \sum_{p=1}^{N_p} M_p(q) \quad (7)$$

where N_p means the maximum of p . According to paper[11], we knew that $F(p)$ is proportional to $p^{-F_{FD}}$ (where F_{FD} means the feature of FD). Therefore we could assumed:

$$F(q) = r \cdot p^{-F_{FD}} \quad (8)$$

when we log the equation, we could obtain:

$$\log(F(q)) = \log(r) - F_{FD} \log(p) \quad (9)$$

Therefore we could obtain FD feature F_{FD} by the slope of $\log(F(q))$ to $\log(p)$.

Frequency Domain Features

For time series signal, spectrum analysis is an important method extracting features. Through short-time fourier transform (STFT)[12], which is a more robust method, we could get the power value as a function of time and frequency: $p(t, f)$. Further, the relation of the power and the frequency could be written as below:

$$P(f) = \sqrt{\sum_{t=0}^T [p(t, f)]^2} \quad (10)$$

As we konw, effective EEG signals exists in low frequency band. Further, this valid low frequency band could be divided to five smaller parts: δ (1-4 Hz) band, θ (4-8 Hz) band, α (8-12 Hz) band, β (12-30 Hz) band, γ (30-64 Hz) band.

Therefore, for each band on $[f_a, f_b]$, *STFT band power* could be caculated by the formula below. Additionally, the ratio of mean band powers $\frac{\beta}{\alpha}$ was computed.

- a) Mean: $\mu_P = \frac{1}{f_b - f_a} \sum_{f=f_a}^{f_b} P(f)$
- b) Minium: $P_{max} = \max(P(f)), f \in [f_a, f_b]$
- c) Maximum: $P_{min} = \min(P(f)), f \in [f_a, f_b]$
- d) Varience: $var(P) = \frac{1}{f_b - f_a} \sum_{f=f_a}^{f_b} |P(f)|^2$

Similarly, we could divide the frequency band in a higher resolution to obtain *STFT bin power* features. We divided the 1-64 Hz band into 32 subband, which means $32 \Delta f = 2\text{Hz}$ bands are extracted for further processing. And then, for each Δf band, we could caculate the STFT bin power feature by the formula above.

As showed in Fig. 8, We concatenate the time domain features and the frequency domain features as the representation of original raw data. Therefore, the features number of one channel of one sample is 71.

7 Statistical Features	10 HOC Features	1 FD Features	21 STFT Band Power Features	32 STFT Bin Power Features
Time domain features			Frequency domain features	

Fig. 8. Feature vector composition

4.2 Feature Selection

Principal component analysis (PCA), proposed by Pearson[13], is a effective and classic method to reduce dimensions of origianl matrix through keeping the lower order principal compoenets. Therefore we use PCA to select the original features.

After the feature extraction steps, we could concatenate 32 channels feature to a long features vector and use all samples to constitute a feature matrix $\mathbf{M} = (\phi_1, \phi_2, \dots, \phi_N)$, where ϕ_i means one sample features vector, N means the number of samples.

Then we use the samples to estimate the mean and the covariance matrix:

$$\bar{\phi} = \frac{1}{N} \sum_{i=1}^N \phi_i \quad (11)$$

$$\Sigma = \frac{1}{N} \sum_{i=1}^N (\phi_i - \bar{\phi})(\phi_i - \bar{\phi})^T = \frac{1}{N} \mathbf{X} \mathbf{X}^T \quad (12)$$

Then we could caculate the eigenvectors (μ_i) and the eigenvalue (λ_i) of Σ . The value λ_i means the importance of the corresponding eigenvector μ_i . We

sorted these eigenvalues in descending order to obtain the descending eigenvalue series: $\lambda'_1, \lambda'_2, \dots, \lambda'_i, \dots, \lambda'_L$, where L is the number of features. Correspondingly, the eigenvectors could be rearranged: $\mu'_1, \mu'_2, \dots, \mu'_i, \dots, \mu'_L$. Then we define the information integrity I by the formula below:

$$I = \frac{\sum_{i=1}^k \lambda'_i}{\sum_{i=1}^L \lambda'_i} \quad (13)$$

In this paper, we chose

$$K = \arg \max_k I(k) > 0.99 \quad (14)$$

Then K is the feature dimension of feature matrix after dimensionality reduction. And the mapping matrix is $M_{PCA} = (\mu'_1, \mu'_2, \dots, \mu'_K)$

4.3 Classification

Support vector machine (SVM) is a effective classification proposed by *Cortes and Vapnik*^[14]. Because the advantages of SVM in non-linear and high-dimensional pattern recognition, we applied it in our experiment as the classification.

When we classify data, each sample data is composed of a feature vector and a label value: $D_i = (\mathbf{x}_i, y_i)$, where \mathbf{x}_i is the feature vector in high dimension, y_i is the category the sample belongs. In particular, the classification problem of two category is taken as an example, and the multi-classification problem could be solved as the combination of multiple two classification problems. Specially, in two classification problem, we define the distance between one sample and a hyperplane is η_i , and $\eta_i = y_i(\omega \mathbf{x}_i + b)$. And to normalize the distance, ω and b are replaced by $\frac{\omega}{\|\omega\|}$ and $\frac{b}{\|\omega\|}$. The normalized distance could be rewritten as

$$\delta_i = \frac{|g(\mathbf{x}_i)|}{\|\omega\|} \quad (15)$$

where $g(\mathbf{x}_i) = \omega \mathbf{x}_i + b$. Because the realtion between δ and misclassification number N is

$$N \leq \left(\frac{2R}{\delta} \right)^2 \quad (16)$$

where $R = \max_i \|\mathbf{x}_i\|$, $i = 1, 2, \dots, n$, the larger the distance δ is, the smaller the N is. Therefore, the optimizing equation is

$$\begin{aligned} & \max \delta \\ & s.t. \quad y_i(\omega \mathbf{x}_i + b) - 1 \geq 0, i = 1, 2, \dots, n \end{aligned} \quad (17)$$

Further, according to equation 15, the optimizing target could be rewritten as below:

$$\begin{aligned} & \min \frac{1}{2} \|\omega\|^2 \\ & s.t. \quad y_i(\omega \mathbf{x}_i + b) - 1 \geq 0, i = 1, 2, \dots, n \end{aligned} \quad (18)$$

Because ω determined by sample data, ω could be assumed as:

$$\omega = \sum_{i=1}^n a_i y_i \mathbf{x}_i^T \quad (19)$$

a_i would be unequal to zero only when the sample is on the closest hyperplanes. In other word, hyperplanes are supported by the sample points which are close to these hyperplanes. Therefore, discriminant function could be written as:

$$g(\mathbf{x}) = \omega \mathbf{x} + b = \sum_{i=1}^n a_i y_i \mathbf{x}_i^T \mathbf{x} + b \quad (20)$$

Since the samples on hyperplanes are known, b could be caculated by $y_i(\omega \mathbf{x}_i + b) - 1 = 0$. When test sample \mathbf{x}_{test} need to be classify, the value of $g(\mathbf{x}_{test})$ decides the category \mathbf{x}_{test} belongs.

5 Results

6 Discussion

7 Conclusion

References

1. Picard R W. Affective computing[M]. MIT Press, 1997.
2. Khosrowabadi R, Wahab A, Ang K K, et al. Affective computation on EEG correlates of emotion from musical and vocal stimuli[C] //Neural Networks, 2009. IJCNN 2009. International Joint Conference on. IEEE, 2009: 1590-1594.
3. Ekman P, Friesen W V, O'sullivan M, et al. Universals and cultural differences in the judgments of facial expressions of emotion[J]. Journal of personality and social psychology, 1987, 53(4): 712.
4. Plutchik R. Emotions: A general psychoevolutionary theory[J]. Approaches to emotion, 1984, 1984: 197-219.
5. Russell J A. A circumplex model of affect[J]. Journal of personality and social psychology, 1980, 39(6): 1161.
6. Jenke R, Peer A, Buss M. Feature extraction and selection for emotion recognition from EEG[J]. IEEE Transactions on Affective Computing, 2014, 5(3): 327-339.
7. Documents of The Cognionics HD-72 Dry EEG, <http://www.cognionics.com/images/docs/HD72.pdf>
8. Specification of The Dobot Magician mechanical arm, <https://www.dobot.cc/dobot-magician/specification.html>
9. Features of The Extreme 3D Pro joystick, <https://www.logitechg.com/en-us/product/extreme-3d-pro-joystick#featuresAnchor>
10. Petrantonakis P C, Hadjileontiadis L J. Emotion recognition from EEG using higher order crossings[J]. IEEE Transactions on Information Technology in Biomedicine, 2010, 14(2): 186-197.

11. Liu Y, Sourina O. Real-time fractal-based valence level recognition from EEG[M] //Transactions on Computational Science XVIII. Springer, Berlin, Heidelberg, 2013: 101-120.
12. Lin Y P, Wang C H, Jung T P, et al. EEG-based emotion recognition in music listening[J]. IEEE Transactions on Biomedical Engineering, 2010, 57(7): 1798-1806.
13. Pearson K. LIII. On lines and planes of closest fit to systems of points in space[J]. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 1901, 2(11): 559-572.
14. Cortes C, Vapnik V. Support-vector networks[J]. Machine learning, 1995, 20(3): 273-297.