IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON DATA-ENABLED INTELLIGENCE FOR DIGITAL HEALTH

Received May 30, 2019, accepted June 27, 2019, date of publication July 15, 2019, date of current version July 30, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2928691

Interpretable Emotion Recognition Using EEG Signals

CHUNMEI QING[1], RUI QIAO[1], XIANGMIN XU[1], AND YONGQIANG CHENG[2]

[1]School of Electronic and Information Engineering, South China University of Technology, Guangzhou 510641, China
[2]Department of Computer Science and Technology, University of Hull, Hull HU6 7RX, U.K.

Corresponding author: Xiangmin Xu (xmxu@scut.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant U180120050, Grant 61702192, and Grant U1636218.

ABSTRACT Electroencephalogram (EEG) signal-based emotion recognition has attracted wide interests in recent years and has been broadly adopted in medical, affective computing, and other relevant fields. However, the majority of the research reported in this field tends to focus on the accuracy of classification whilst neglecting the interpretability of emotion progression. In this paper, we propose a new interpretable emotion recognition approach with the activation mechanism by using machine learning and EEG signals. This paper innovatively proposes the emotional activation curve to demonstrate the activation process of emotions. The algorithm first extracts features from EEG signals and classifies emotions using machine learning techniques, in which different parts of a trial are used to train the proposed model and assess its impact on emotion recognition results. Second, novel activation curves of emotions are constructed based on the classification results, and two emotion coefficients, i.e., the correlation coefficients and entropy coefficients. The activation curve can not only classify emotions but also reveals to a certain extent the emotional activation mechanism. Finally, a weight coefficient is obtained from the two coefficients to improve the accuracy of emotion recognition. To validate the proposed method, experiments have been carried out on the DEAP and SEED dataset. The results support the point that emotions are progressively activated throughout the experiment, and the weighting coefficients based on the correlation coefficient and the entropy coefficient can effectively improve the EEG-based emotion recognition accuracy.

INDEX TERMS EEG, emotion activation, emotion recognition, machine learning.

I. INTRODUCTION

With the rapid development of computer and human-computer interaction technology, there is a high demand to build a more intelligent and humanized human-machine interface (HMI) in the field of human-computer interaction (HCI) [1], [2]. The original intention and goal of HCI are to better help users to achieve the intended interactive purpose. However it is worth noting that in the process of HCI, the user's interactive behavior is only an external behavior, and the nature of that behavior is driven by the user's perception. Contemporary cognitive scientists believe that in addition to traditional cognitive processes such as perception, learning, memory, and speech, emotion is also an important cognitive process. Compared to machines, humans naturally have complex emotional systems, and a person's behavior is often influenced by emotions. If the machine has the ability to accurately recognizing human emotions, it is significant to build a more intelligent and humanized human-computer interaction system. In this context, affective computing emerges as required, and it is being studied as a hot spot [3]–[7]. Affective computing is attracting more and more attention.

Affective computing is the study and development of systems and devices that can recognize, interpret, process, and simulate human affects [8]. In affective computing, researchers use various sensors to collect physiological and behavioral signals triggered by emotions and use computer technology to analyze these signals to obtain emotional models [9]. Based on the obtained emotional model, the HCI system can perceive, recognize and understand human emotions, and make targeted responses to different emotional states of the users, making the whole HCI system more intelligent and humanized.

The associate editor coordinating the review of this manuscript and approving it for publication was Yonghong Peng.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see http://creativecommons.org/licenses/by/4.0/

On the other hand, significant amount of research has been reported to use computer technology to analyze human emotions by both biologists and computer scientists. Russell [10] proposed a valence-arousal emotion model which mapped each emotional state to an area in the two-dimensional space. Among them, the horizontal axis represents the valence of the emotion, to some extent reflects the degree of positive or negative emotion and the vertical axis represents the arousal of emotions, reflecting the level of neurophysiological activation of emotions. There is also a three-dimensional emotional model called PAD (Pleasure - Arousal - Dominance) model, proposed by Russell and Mehrabian in 1974 [11]. The degree of pleasure and arousal is consistent with the definition in the valence-arousal emotion model, while the dominance indicates the individual's state of control over the situation and others. In this way, computer-based emotional state analysis has a uniform standard that makes computer-based emotion recognition possible.

Although the measurement of emotional state can be quantified using Russell's model, recognizing the emotional state is still a challenging task. Emotion assessment methods can be broadly divided into subjective and objective ones [12], [13]. Subjective measures use self-rating instruments, such as Self-Assessment Manikin (SAM) [12] and questionnaires, while objective measures can be acquired from physiological cues derived from the physiology theories of emotion [14]. In the physiological signals used to assess emotions, electroencephalogram (EEG) signal has attracted more and more attention in recent years and has been widely used in medical, Affective Computing and other fields. Abeer et al., used power spectral density (PSD) features which were extracted from EEG in combination with deep neural network (DNN) to categorize emotions [14]. Samarth et al., combined the statistical features extracted from EEG with deep learning models such as the CNN and DNN to identify emotions [15]. Raja et al., optimized the extracted EEG feature set by using p-values, and combined it with ensembles methods to recognize emotions. All the above studies have achieved excellent recognition accuracy.

Although much of the recent work has achieved excellent recognition accuracy, most of the existing EEG-based methods for emotion recognition will cut the instance into a series of a fixed duration of segments (e.g., 2s or 4s). Chen et al. used 4-second sliding and 2-second overlapping time windows to divide the EEG signal into 29 segments [16]. Zhuang et al. used 2-second sliding and 1-second overlapping time windows to divide the EEG signal into 49 segments [17]. In their methods, the labels of all segments are the same, consistent with the emotional labels of the instance. However, emotions cannot always be in the same state during a trial, hence the applicability of this type of marking methods are limited. In order to solve this problem, some studies only use the second half of a trial to train their models. These studies believe that the emotions will be more apparent in the second half of the experiment, and the results confirm that the classification accuracy is actually better [18]–[20]. However, this is only confirmed from an experimental point of view. Without a theoretical basis, we still do not know how and when emotions are stimulated in an experiment.

As can be seen from the introduction, most of the traditional studies have not formed a clear understanding of the activation process of emotions. In addition, the conclusions of relevant researches are rarely supported by psychology. Aiming at the problem of insufficient awareness of emotional stimulation mechanism in most traditional studies, we propose an innovative method of emotional activation mechanism based on machine learning and EEG signals in this paper. The contributions can be summarized as below:

1) We innovatively constructed correction coefficients and entropy coefficients of emotions by extracting features from the EEG signals. Based on these two coefficients, a novel activation curve of emotions is constructed. This activation curve carries information of the emotional stimulation mechanism.

2) We use the obtained correlation coefficients and entropy coefficients to construct weight coefficients to improve the emotion recognition accuracy.

The rest of this paper is organized as follows: Section II introduced the data used in this paper. The proposed method is presented in Section III. Section IV contains the results and their discussion. Finally, the conclusion is presented in Section V.

II. DATA PRERARATION

The experiments in this paper were carried out on the DEAP dataset [21] and the SEED dataset [22] which are commonly used benchmark datasets by many researchers [23]–[27]. The DEAP dataset and the SEED dataset are briefly introduced below.

A. DEAP

The Database for Emotion Analysis using Physiological Signals (DEAP) is a benchmark affective EEG database for the analysis of emotions. It was acquired in a controlled laboratory setting. The DEAP contains 32-channel electroencephalogram (EEG) and 8-channel peripheral physiological signals from 32 subjects among which 22 subjects have additional positive videos recorded. Figure 1 shows the electrode placements for the EEG. The various emotions of subjects were stimulated through 40 1-minute music-videos, corresponding to different emotional state. For each subject, 40 videos were presented in 40 trials following their ratings (1-9) of Arousal, Valence, Like/Dislike, Dominance and Familiarity of this trial recorded through SAM. After obtaining the score, the emotional state is defined according to the valence-arousal emotion model. With a threshold of 5, the two dimensional emotional space can be divided into four regions, namely high valence-high arousal (HVHA), high valence-low arousal (HVLA), low valence high arousal (LVHA) and low valence-low arousal (LVLA). The scatter plot of the DEAP dataset is shown in Figure 2.

The DEAP dataset contains two versions of physiological signal data, which are raw data and pre-processed data. For

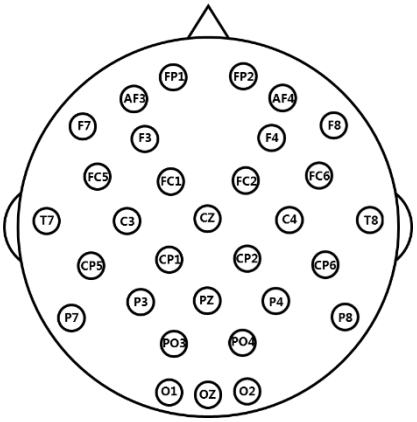


FIGURE 1. Location of EEG electrodes in DEAP.

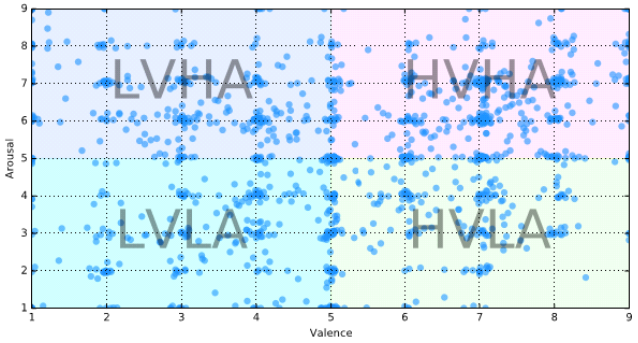


FIGURE 2. The scatter plot of DEAP, based on valence-arousal model.

the raw data, due to the pre-processing process such as noise reduction, different results may be obtained due to different methods. Therefore, in order to ensure consistency, the pre-processed data is used in this paper. The pre-processed data in the DEAP dataset includes 32 channels of EEG signals (128Hz) and 8 channels of peripheral physiological signals.

B. SEED

The SJTU Emotion EEG Dataset (SEED) is a free and publicly available EEG dataset for emotional analysis provided by Shanghai Jiao Tong University in 2015 [22]. The SEED dataset contained 62 channels of EEG signals from 15 subjects for 15 experiments. In each experiment, every subject firstly watched 15 emotional film clips, then the subject had 45 seconds to self-assess and 15 seconds to calm down. Throughout the experiment, Zheng et al. used film clips to elicit three emotions of the subject: positive, neutral, and negative.

The SEED dataset contains 62-channel electroencephalogram (EEG) which was recorded using an ESI NeuroScan System at a sampling rate of 1000 Hz from 62-channel electrode cap according to the international 10-20 system [22]. In this paper, down sampled the raw data to 200Hz, and applied a 0-75Hz bandpass filter to filter out the unwanted signals. The electrode placements for the EEG is illustrated in Figure 3.

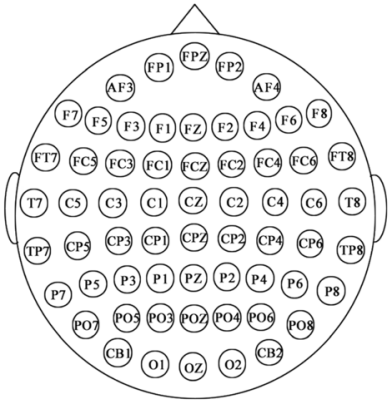


FIGURE 3. Location of EEG electrodes in SEED [22].

III. METHODOLOGY

The main challenge of this paper is how to visualize the activation process of emotions and verify that the results we get are reasonable. In the response to the problems above, this paper proposed a research method of emotional activation mechanism based on EEG signals and machine learning. The flowchart of the method proposed in this paper is shown in Figure 4.

A. PREPROCESSING AND FEATURE EXTRACTION

1) PREPROCESSING

For the DEAP dataset, the duration of each EEG signal is 63s, the first three seconds of the signal is the pre-trial baseline signal and it should be removed. In this study, we use 2-second sliding and 1-second overlapping time windows to cut the 60 seconds EEG signal into 59 segments. Figure 5 shows the segmentation process.

For the SEED dataset, the duration of each experiment is different. In order to unify the standard, we choose 185s which is the shortest duration of all experiments as the standard experimental duration. For the experiments which duration is longer than 185s, the last 185s were selected. In addition, since the experiments in SEED lasts for a long time (approximately 4 minutes), in order to avoid the possible interference or the emotions that have not been elicited at the beginning of the experiment, we have removed the first 30 seconds of EEG data, that is, for the SEED dataset, only 155 seconds of data were used in this paper. The data addition, since the experiments in SEED lasts for a long time (approximately 4 minutes), in order to avoid the possible interference or the emotions that have not been elicited at the beginning of the experiment, we have removed the first 30 seconds of EEG data, that is, for the SEED dataset, only 155 seconds of data were used in this paper. The data segmentation method is based on the recommended setting in [22], the time window's duration is set to 1s without overlapping.

2) FEATURE EXTRACTION

The feature extraction is performed firstly on individual EEG channel signal for each subject, then a feature tensor is formed

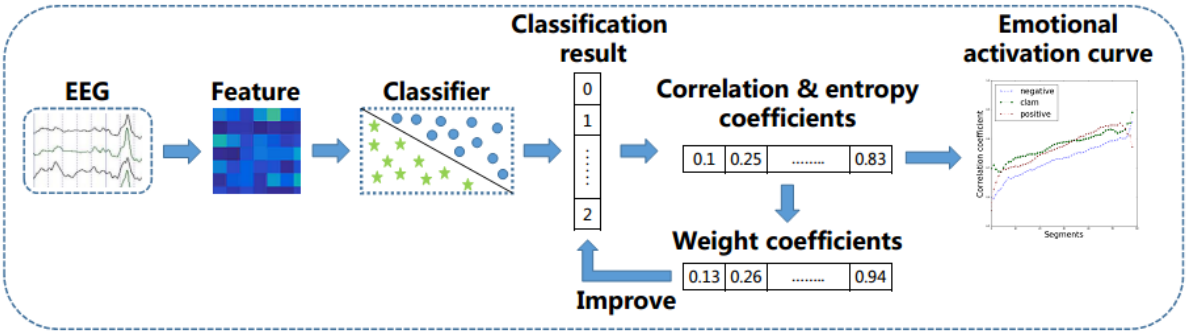


FIGURE 4. Flowchart of the method proposed in this paper.

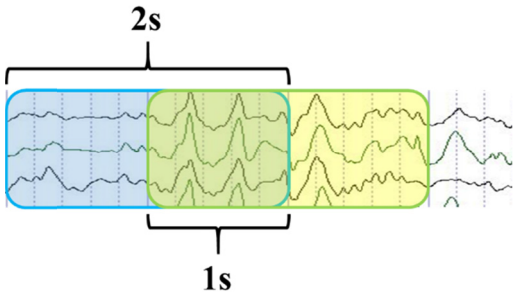


FIGURE 5. EEG segmentation process in the DEAP dataset.

by ensemble of these features. The following takes a certain channel of EEG as an example to describe the feature extraction method and process on DEAP and SEED.

- DEAP

For the DEAP dataset, the features selected here are 1st and 2nd order differential features, which are two statistical features widely used in the field of emotion recognition based on EEG signals. The features extraction from EEG signals in this paper are based on the definition of 1st and 2nd order difference in [28]. The extracted features are described below.

The 1st order derivative is given by,

$$\delta = \frac{1}{T-1}\sum_{t=1}^{T-1}|x(t-1)-x(t)| \tag{1}$$

where T is the duration of the signal x.

The 2nd order derivative is given by,

$$\gamma = \frac{1}{T-2}\sum_{t=1}^{T-2}|x(t-2)-x(t)| \tag{2}$$

For additional features, we calculated the normalized 1st order derivative and the normalized 2nd order derivative. Normalized 1st order difference is given by,

$$\delta' = \frac{\delta}{\sigma} \tag{3}$$

where σ is the standard deviation of the signal x. Similarly, we can get the normalized 2nd difference. In this way, we extract 4 features for each EEG channel.

As an example, after the feature extraction process, a $T \times S \times C \times F$ tensor can be obtained, where T is the number of experiment for each subject which is 40; S represents the number of segments, here is 59; C represents the number of 32 EEG channels; F represents the feature dimension, and here is 4. Here, the features of the 32 channels are connected to form a feature vector. So for a subject, a 40 $\times$59 $\times$ 128 tensor can be obtained.

- SEED

For the SEED dataset, the differential entropy features provided by the SEED dataset are used in this paper. We calculate differential entropy for each sample over the five bands of each EEG channel (delta, theta, alpha, beta and gamma) [22]. Finally, on a channel of a sample, a feature vector of length 5 can be extracted.

For a sample, after the feature extraction process, with the DEAP dataset, a T $\times$ S $\times$ C $\times$ F tensor can be obtained, where T is the number of the experiment for each subject of 15; S represents the number of segments, here is 155; C is 62, representing 62 EEG channels, F represents the feature dimension, here is 5. Here, the features of the 62 channels are connected to form a feature vector, so, for a subject, a 15 $\times$155 $\times$ 310 tensor can be obtained.

3) AUTOENCODER

In this paper, since the difference feature extracted from the DEAP dataset is a relatively simple statistical feature, in order to improve the discriminative power of the feature, we use the autoencoder to further process the differential feature. The differential entropy feature is more complex and discriminative than the differential feature. Therefore, we only use the autoencoder to abstract the difference feature extracted from the DEAP dataset. The following is a brief introduction to the autoencoder.

Autoencoder is a commonly used neural network model and widely used in EEG-based emotion recognition [29], [30]. The goal of an autoencoder is to learn encoding of input data and the encoding can be used as an abstraction

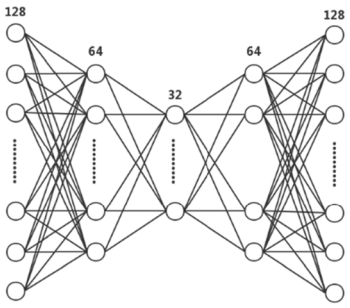


FIGURE 6. Network structure of stacked autoencoder.

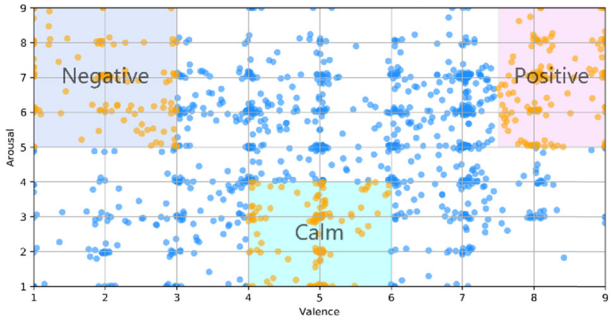


FIGURE 7. The division of emotional subsets in the DEAP dataset.

of input data. In general, this kind of abstract feature is often more discriminative. An autoencoder is generally formed by a feed-forward neural network, consisting of an input layer, an output layer and one or more hidden layers between the input layer and the output layer. On the whole, an autoencoder consists of an encoder and a decoder. The encoder abstracts the input data and the decoder produces the reconstruction of the corresponding input data. If the difference between the data reconstructed by the decoder and the input data is small, then we have reason to believe that the encoding given by the encoder to the original input data is a good representation of the input data.

Single-layer autoencoder often have limited abstraction capabilities. If we want to improve representation or modeling capacity, we usually use the structure of stacked autoencoder. Stacked autoencoder is a neural network consisting of multiple layers of autoencoder in which the outputs of each layer are wired to the inputs of the successive layer.

In this paper, we use the stacked autoencoder to extract abstract features based on difference features. As mentioned before, the feature dimension is 128, so the dimension of the input and output layers of the autoencoder are 128, and the output dimension of each fully connected layer is 64, 32 and 64 respectively. The activation function of each layer is Rectifier Liner Unit (ReLU) and we choose MSE as the loss function. In addition, batch size is set to 32 and adadelta was selected as optimizer. The network structure is shown in Figure 6.

B. THE CHOICE OF TARGET EMOTION

Since there are only three emotional labels (positive, negative and neutral) in the SEED dataset, in order to be consistent, this paper extracts three emotional subsets from the DEAP dataset: positive, negative and calm. The three subsets were selected by extending our previous works [18] and made some extensions.

In [18], the calm subset was composed of the samples with an arousal level lower than 4 and a valence level between 4 and 6. Similarly, the negative subset was consisted of the samples with an arousal level higher than 5 and a valence level lower than 3. In the end, a total of 279 samples were selected, of which 146 were calm emotion samples and 133 were negative emotion samples. In this paper, we added a positive group which was consisted of the samples with an arousal level higher than 5 and a valence level higher than 7.5. The reason for choosing the valence score of 7.5 as the threshold is to balance the number of samples in each subset. The total number of analyzed instances was 439, i.e., 146 from calm subset, 133 from negative subset and 160 from positive subset. The division of the three subsets in the DEAP dataset is shown in Figure 7.

C. CLASSIFIER

We choose the soft voting strategy to build the classifier. Soft voting strategies can combine the advantages of multiple classifiers to predict the label of the samples, making it more robust. Soft voting strategies rely on a series of independent classifiers and predicts the class label based on the argmax of the sums of the predicted probabilities. Here we choose Decision Tree, KNN and Random Forest as the base classifier. After getting the labels of each segment of a trial, in order to get the label of this trial, we introduced a classification strategy. Here, we define $label(y_i)$ as the label corresponding to y_i, so the classification strategy can be expressed as

$$Y_{predFinal} = \arg\max_{l} \sum_{i=0}^{N-1} I(label(y_i), l), \tag{4}$$

where l is positive, negative or calm and $\mathrm{I}(label(y_i), l)$ is the indicator function defined as,

$$\mathrm{I}(label(y_i), l) = \begin{cases} 1, & label(y_i) = l \\ 0, & label(y_i) \neq l \end{cases}. \tag{5}$$

The classification strategy is shown in Figure 8. For example, as mentioned before, in DEAP dataset, we use 2 second sliding and 1-second overlapping time windows to divide the 60 seconds EEG signal into 59 segments and 59 segments will be fed to a classifier simultaneously. For the 59 samples from the same trial, we chose the category with the largest number of occurrences in the classification result as the overall label for the 59 samples.

In addition, in the model evaluation step, we aim to use the K-Fold cross validation to assess the predictive

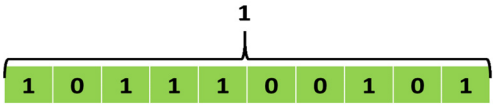


FIGURE 8. The classification strategy.

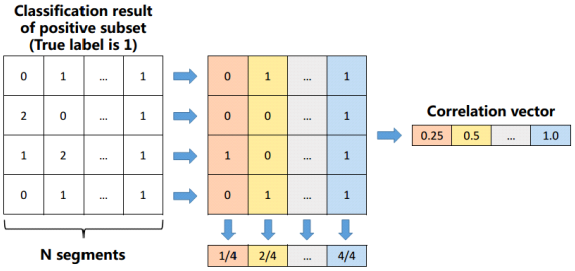


FIGURE 9. The calculation process of the correlation coefficient.

performance of our proposed method. Here, we choose 10 as the value of K. For example, in DEAP dataset, the total number of analyzed instances was 439, these instances will be divided into 10 subsets. For each fold, we choose one of the subsets as the test set and the other as the training set. Finally, we calculated the mean of the accuracy as the final result.

D. EMOTIONAL ACTIVATION CURVE

Here we will introduce the correlation coefficient and the entropy coefficient, which are based on the classification results of each segment obtained in the previous step.

1) THE CORRELATION COEFFICIENT

The correlation coefficient is defined below. Assuming that the classification result of N (N is 59 in DEAP and 155 in SEED) samples belonging to one trial is $Y_{pred} = \{y_1, y_2, \cdots, y_N\}$ and the true label of the trial is y_{true}. The ith element c_i of the correlation coefficient C can be expressed as:

$$c_i = \mathrm{I}(y_i, y_{true}), \quad 1 \le i \le N, \tag{6}$$

where I( ) is the same as the indicator function mentioned in Equation 5.

In this way, we can get the correlation coefficient C. For all trials in the test set, we calculate the correlation coefficient C and after the 10-fold cross-validation, we can obtain the correlation coefficients for all trials. Then we calculate the average of the correlation coefficients for all samples that belong to the same emotion category. Finally, we can obtain three correlation coefficient, C_{Calm}, $C_{Negative}$ and $C_{Positive}$, corresponding to the three emotions. The calculation process of the correlation coefficient is shown in Figure 9. Next, we can map these three coefficients in a two-dimensional coordinate system and smooth them with the mean filter to get the correlation curve.

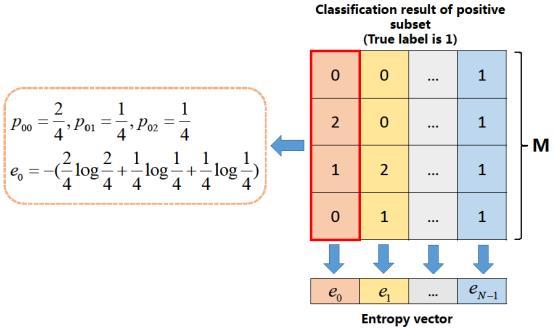


FIGURE 10. The calculation process of the entropy coefficient.

2) THE ENTROPY COEFFICIENT

As mentioned before, after the 10-fold cross-validation, the prediction results are obtained for all trials. We divided the trials into three subsets according to the emotion label and we will get three prediction result arrays and the shape of arrays is M × N, where M is the number of trials in the subsets. Then, the ith element e_i of the entropy coefficient E can be calculated as follows,

$$e_i = -\sum_{j=1}^{3} p_{ij} \log(p_{ij}), \quad 0 \le i \le N-1 \tag{7}$$

where p_{ij} is defined as,

$$p_{ij} = \frac{m_{ij}}{M}. \tag{8}$$

Here m_{ij} is the number of label j in segmenting i. and we define that if $p_{ij} = 0$, then

$$p_{ij} \log(p_{ij}) = 0 \tag{9}$$

The calculation process of the entropy coefficient is shown in Figure 10. In this way, we can get three entropy coefficient E_{Clam}, $E_{Positive}$ and $E_{Negative}$, corresponding to three emotions. Next, we can map these three coefficients in a two-dimensional coordinate system and smooth them with the mean filter to get the entropy curve.

3) THE WEIGHT COEFFICIENT

After obtaining the correlation coefficient and the entropy coefficient, we can get the weight coefficient based on these two coefficients. Here, we define $label(y_i)$ as the label corresponding to y_i, the weight coefficients W_{Clam}, $W_{Positive}$ and $W_{Negative}$ can be expressed as:

$$W_l = \sum_{i=0}^{N-1} \mathrm{I}(label(y_i), l) \frac{(C_l[i] + 1 - E_l[i])}{2}, \tag{10}$$

where l is positive, negative or calm.

The weight coefficients can be used to get the final emotional label of a trial. Unlike the classification strategy mentioned before, for a trial, the value of W_{Clam}, $W_{Positive}$ and $W_{Negative}$ can be obtained, and then we can select the label

corresponding to the maximum value as the final label of this trial. The weight coefficients and the classification strategy mentioned before will be compared in the next section.

IV. RESULTS AND DISCUSSION

A series of experiments have been designed and conducted on both DEAP and SEED datasets to validate the effectiveness of our method including emotion activation curves, the interpretability and accuracy improvement. The time period comparison experiment shows the impact of using different section of EEG signals to train classification models on classification accuracy. Based on the classification results, we can get the correlation curve and entropy curve. The correlation curve obtained from our method reflects the correlation between target emotion and emotion at different time points, which is, the activation process of emotion. The entropy curve focuses on the uncertainty of emotional state at different time points and shows hints of interpreting human emotional activities during the experiments. Furthermore, we showed that the weight coefficients based on the correlation coefficients and entropy coefficients have improved classification accuracy compared to current benchmark algorithms. It is worth mentioning that misclassification samples have been removed from the calculations. The misclassification samples refers to samples that are misclassified by the classifier

A. EXPERIMENT ENVIRONMENT

The experimental environment was built on a PC running Windows operating system with CORE i5 CPU and 8G memory. The computing environment was Python 3.6.

B. THE TIME PERIOD COMPARISON EXPERIMENT

We have compared the classification accuracy using different sections of the EEG signals. The results are shown in Table 1. The highest classification accuracy in each group is bolded.

It can be seen from Table 1 that the classification accuracy obtained in DEAP dataset has significant improvements when the model is trained using the second half of the EEG data. In particular, when the model was trained using the last 34 seconds of the EEG data, the highest classification accuracy of 62.63% was obtained. Similarly improvements have also been achieved in the SEED dataset when the model is trained using the second half of the EEG data. When the model was trained using the last 75 seconds EEG data, the highest classification accuracy of 74.85% was obtained.

Compared to the classification accuracy of 72% achieved in [22] on the SEED dataset, the best accuracy of ours on the same dataset is 74.85%. This shows the classification result in our paper is credible. The two-category classification accuracy in [18] has achieved a close to 70% accuracy on the DEAP dataset and the best result of ours achieved a comparable 62.63% on three categories. As we all know, when a model transforms from a two-category task to a three-category task, its classification accuracy tends to be greatly reduced. So we have reason to believe that our model has similar discriminatory power compared to that of [18], which means, the classification result is also credible.

In summary, the results in Table 1 confirm that when we train the model with the second half of the EEG data, the classification accuracy will be higher. This is consistent with the fact that training the model with the latter half of the trial has the potential to improve the results as reported in [18]–[20]. In other words, the EEG signal in the second half of the whole trial have a stronger discriminating ability for recognizing the target emotions.

C. THE CORRELATION CURVE AND EMOTION ACTIVATION PROCESS

Figure 11 show the correlation curves of the negative, neutral and positive emotions in the DEAP dataset (a-d) and SEED dataset (e-h), respectively, where the blue dashed lines represent negative emotion, yellow ones are calm emotions and pink ones represent positive emotions. To eliminate the bias introduced by classifiers, we have used four different classifiers to construct the correlation curves, namely the ensemble model defined in section III(C), Decision Tree, KNN and Random Forest.

In Figure 11(a-d), there are two things are worth noted. Firstly, we found that the correlation coefficients of calm emotions tend to increase from beginning. We believe that this is related to the fact that human emotions are usually in a relatively calm state, which is also very consistent with our intuitive understanding of emotions. Secondly, the correlation curve of positive emotions starts to show a descending trend after the peak, while the negative emotion correlation curve remains as ascending trend. Reference [31] shows that positive emotion reached peak value earlier than negative in heart rate response and our correlation curve from EEG shows the same patterns. We think this is in line with the point made in literature [32] that people are more likely to get used to and adapt to positive emotions. This seems to be more consistent with people's daily life experience, that is, people are easy to adapt to and ignore positive emotional experience, but difficult to forget negative one.

As one can see from Figure 11(e-h), the correlation curves of the three emotions obtained on the SEED dataset show a similar gradual ascending trend. Although the overall trend of the correlation curve from the SEED dataset is roughly the same as that from the DEAP dataset, there are still some differences. Compared with the DEAP dataset, the correlation curve obtained from the SEED dataset has a higher position, indicating that the emotional correlation is generally higher than that in the DEAP dataset. In other words, the emotional expression in the

SEED data set may be more intense, which is also consistent with the fact that the classification accuracy based on the SEED data set is higher than the accuracy achieved on the DEAP dataset.

In addition, in the SEED dataset, the peak of positive emotions appears later than that of negative emotions, which is inconsistent with the conclusions in [31], but is consistent

TABLE 1. Results of training models using different period of EEG signals.

DEAP		SEED	
Period	**Accuracy**	**Period**	**Accuracy**
All	57.15%	All	71.41%
Last 30 seconds	60.13%	Last 75 seconds	**74.85%**
Last 34 seconds	**62.63%**	Last 95 seconds	72.27%

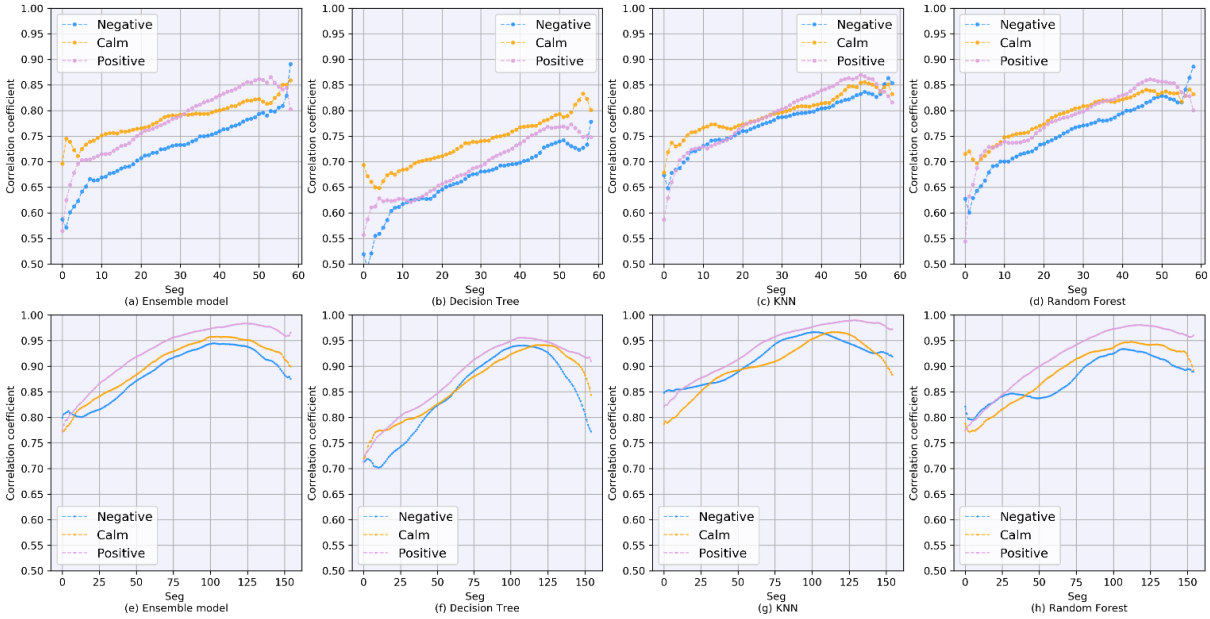


FIGURE 11. The correlation curve of DEAP(a-d) and SEED(e-h).

with the theory of negativity bias in psychology. Negativity bias theory refers to a priority attention mechanism for negative emotions. The negativity bias, is the notion that, even when of equal intensity, things of a more negative nature (e.g. unpleasant thoughts, emotions, or social interactions; harmful/traumatic events) have a greater effect on one's psychological state and processes than neutral (calm) or positive things [33]–[36]. More in-depth research may be needed on this point.

In general, although there are some differences in the emotional activation curves obtained based on DEAP and SEED, both of them show an ascending trend over time, supporting the theory of progressive activation process of emotions.

Furthermore, the correlation curves of the three emotions all show an ascending trend over time which indicates that as the trials progress, the correlation between the subject's emotions and the target emotions increases. In other words, the emotional correlation curve reflects the progressive activation process of emotions. This result seems suggest that emotions are gradually activated during the trial.

D. THE ENTROPY CURVE AND INTERPRETABILITY

Figure 12 shows the entropy curves of the negative, calm and positive emotions obtained from the DEAP dataset (a-d) and SEED dataset (e-h), respectively. The blue lines represent negative emotion; yellow lines represent calm emotions and pink one represent positive emotions. Four different classifiers have been applied to construct the entropy curves, i.e. the ensemble model, Decision Tree, KNN and Random Forest from left to right. As seen from Figure 12, the entropy curves of all three emotions demonstrate a descending trend over time, indicating that the uncertainty of the three emotions gradually decreases in their respective trial subsets.

Two points worth to be made from Figure 12(a-d). Firstly, the entropy value of calm emotions is relatively low at the beginning which indicates that the calm emotions are relatively less uncertain. This result is consistent with the ones measured by correlation curves in Figure 11(a-d). Secondly, the entropy curve of positive emotions began to rise after reaching the lowest point, while the entropy curve

of negative emotions still shows a descending trend, which is consistent with the results obtained from Figure 11(a-d).

As can be seen from Figure 12(e-h), the entropy curves of the three emotions obtained from the SEED dataset also show a descending trend. As with the correlation curve, the entropy curve obtained on the SEED dataset is somewhat different from that obtained on the DEAP dataset, i.e. the smaller entropy values and the valley of negative emotions comes earlier than positive emotions. This represents the same pattern as explained earlier in the correlation curves.

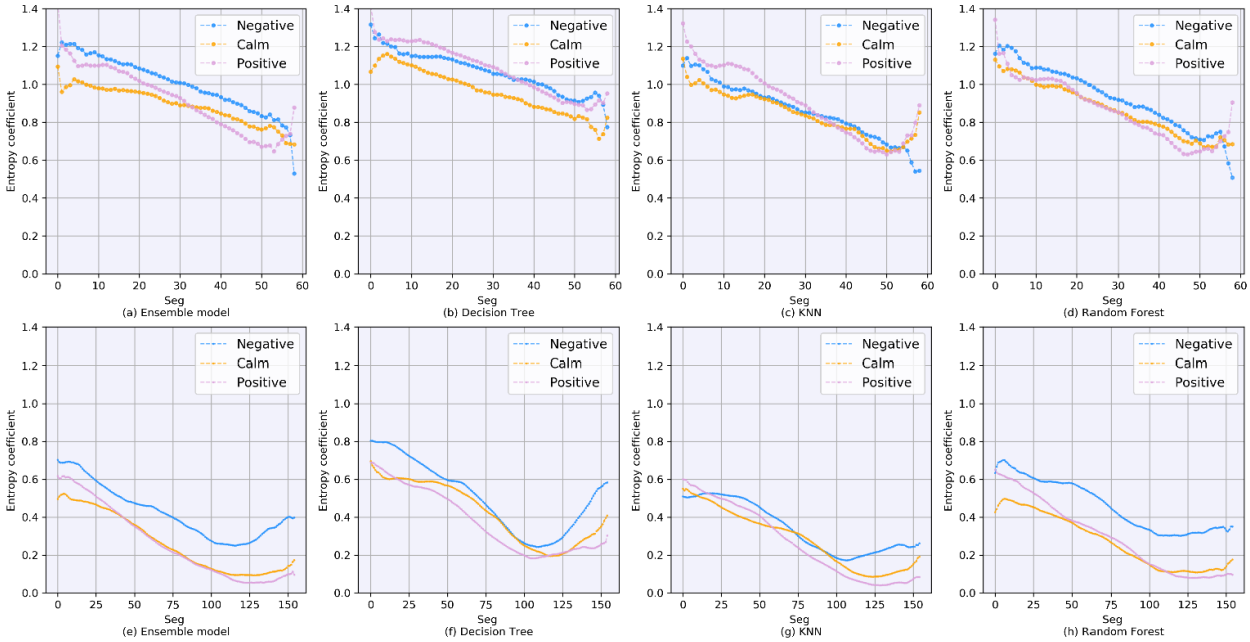


FIGURE 12. The entropy curve of DEAP(a-d) and SEED(e-h), corresponding to ensemble model, Decision Tree, KNN and Random Forest from left to right.

TABLE 2. Results of the weight coefficients.

DEAP		SEED	
Model	**Accuracy**	**Model**	**Accuracy**
Raw	62.63%	Raw	74.85%
With weight coefficients	**63.09%**	With weight coefficients	**75%**

Both correlation curves and entropy curves have demonstrated the interpretability of our method. The trends showed in the curves are compliant with the progressive activation process of emotion reported in the literatures.

E. THE WEIGHT COEFFICIENTS EXPERIMENT

The weight coefficients impact is explored on classification accuracy in this section. The results are shown in Table 2. The highest classification accuracy in each group are bolded. As seen in Table 2 , the classification accuracy of the weight coefficients is higher than that of the original classification strategy mentioned in Section III.C on both the DEAP dataset and the SEED data set, demonstrating the validity of the proposed weighting coefficients. Compared with the simple voting-based classification strategy, the weighting coefficient-based classification strategy proposed in this paper considers the influence of emotions at different time points on target emotions. Since the weight coefficient is based on the correlation coefficient and entropy coefficient, the validity of the weight coefficient also indicates the validity of the proposed correlation coefficient and entropy coefficient from another aspect, which indicates that the proposed theory of progressive activation of emotions is reasonable.

V. CONCLUSION

In this paper, aiming at the challenge of insufficient awareness of emotional stimulation mechanism in most traditional studies, we proposed a coefficients-based method based on machine learning using EEG signals. This method not only outperformed the benchmark algorithms in terms of accuracy but also interpret the progress of emotion activation. Firstly, we extracted features from EEG signals and classified emotions using machine learning techniques. We further found that the latter stage of EEG signals have better correlations with emotions, hence better classifier performance can be achieved if the second half of the trial is used for training. Secondly, based on the classification results, the correlation curves and entropy curves of emotions are constructed, which to a certain extent indicate the emotional activation progression. It is found that emotion was progressively activated. The proposed method has provided a quantitative tool to theoretically explain emotional activation mechanism such as why the second half of a trial leads to better classification results. Finally, the obtained correlation coefficients and entropy coefficients are used to construct weight coefficients to improve the classification accuracy compared to current benchmark algorithms. Since the weight coefficient is based on the correlation coefficient and entropy coefficient,

the validity of the weight coefficient also indicates the validity of the proposed correlation coefficient and entropy coefficient from another aspect, which indicates that the proposed theory of progressive activation of emotions is reasonable.

REFERENCES

[1] A. Kappas and N. C. Krämer, *Face-to-Face Communication Over the Internet: Emotions in a Web of Culture, Language, and Technology*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[2] M. S. Hossain, "Patient state recognition system for healthcare using speech and facial expressions," *J. Med. Syst.*, vol. 40, no. 12, p. 272, Dec. 2016.

[3] R. W. Picard, *Affective Computing*. Cambridge, MA, USA: MIT Press, 2000.

[4] C. Kleine-Cosack, "Recognition and simulation of emotions," May 2008.

[5] J. H. Janssen, E. L. van den Broek, and J. H. D. M. Westerink, "Tune in to your emotions: A robust personalized affective music player," *User Model. User-Adapted Interact.*, vol. 22, no. 3, pp. 255–279, 2012.

[6] R. Yonck, *Heart of the Machine: Our Future in a World of Artificial Emotional Intelligence*. New York, NY, USA: Skyhorse, 2017.

[7] S. Asteriadis, P. Tzouveli, K. Karpouzis, and S. Kollias, "Estimation of behavioral user state based on eye gaze and head pose–application in an e-learning environment," *Multimedia Tools Appl.*, vol. 41, no. 3, pp. 469–493, 2009.

[8] (2019). *Affective Computing*. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Affective_computing&oldid=882501304

[9] N. Garay-Vitoria, I. Cearreta, J. M. López, and I. Fajardo, "Assistive technology and affective mediation," *Hum. Technol.*, vol. 2, no. 1, pp. 55–83, 2006.

[10] J. A. Russell, "A circumplex model of affect," *J. Personality Social Psychol.*, vol. 39, no. 6, pp. 1161–1178, 1980.

[11] J. A. Russell and A. Mehrabian, "Evidence for a three-factor theory of emotions," *J. Res. Pers.*, vol. 11, no. 3, pp. 273–294, 1977.

[12] M. M. Bradley and P. J. Lang, "Measuring emotion: The self-assessment manikin and the semantic differential," *J. Behav. Therapy Experim. Psychiatry*, vol. 25, no. 1, pp. 49–59, 1994.

[13] J. J. Allen, J. A. Coan, and M. Nazarian, "Issues and assumptions on the road from raw signals to metrics of frontal EEG asymmetry in emotion," *Biol. Psychol.*, vol. 67, nos. 1–2, pp. 183–218, 2004.

[14] A. Al-Nafjan, M. Hosny, Y. Al-Ohali, and A. Al-Wabil, "Recognition of affective states via electroencephalogram analysis and classification," in *Proc. Int. Conf. Intell. Hum. Syst. Integr.* Cham, Switzerland: Springer, 2018, pp. 242–248.

[15] S. Tripathi, S. Acharya, R. D. Sharma, S. Mittal, and S. Bhattacharya, "Using deep and convolutional neural networks for accurate emotion classification on DEAP dataset," in *Proc. 29th IAAI Conf.*, 2017, pp. 4746–4752.

[16] J. Chen, B. Hu, Y. Wang, Y. Dai, Y. Yao, and S. Zhao, "A three-stage decision framework for multi-subject emotion recognition using physiological signals," in *Proc. IEEE Int. Conf. Bioinformatics Biomed. (BIBM)*, Dec. 2016, pp. 470–474.

[17] N. Zhuang, Y. Zeng, K. Yang, C. Zhang, L. Tong, and B. Yan, "Investigating patterns for self-induced emotion recognition from EEG signals," *Sensors*, vol. 18, no. 3, p. 841, 2018.

[18] B. García-Martínez, A. Martínez-Rodrigo, R. Z. Cantabrana, J. M. P. García, and R. Alcaraz, "Application of entropy-based metrics to identify emotional distress from electroencephalographic recordings," *Entropy*, vol. 18, no. 6, p. 221, 2016.

[19] N. Kumar, K. Khaund, and S. M. Hazarika, "Bispectral analysis of EEG for emotion recognition," *Procedia Comput. Sci.*, vol. 84, pp. 31–35, May 2016.

[20] Y. Zhang, S. Zhang, and X. Ji, "EEG-based classification of emotions using empirical mode decomposition and autoregressive model," *Multimedia Tools Appl.*, vol. 77, no. 20, pp. 26697–26710, 2018.

[21] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "DEAP: A database for emotion analysis ;using physiological signals," *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 18–31, Jan./Mar. 2012.

[22] W.-L. Zheng and B.-L. Lu, "Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks," *IEEE Trans. Auton. Mental Develop.*, vol. 7, no. 3, pp. 162–175, Sep. 2015.

[23] X. Li, D. Song, P. Zhang, Y. Zhang, Y. Hou, and B. Hu, "Exploring EEG features in cross-subject emotion recognition," *Frontiers Neurosci.*, vol. 12, p. 162, May 2018.

[24] Y. Luo, S.-Y. Zhang, W.-L. Zheng, and B.-L. Lu, "WGAN domain adaptation for EEG-based emotion recognition," in *Proc. Int. Conf. Neural Inf. Process.* Cham, Switzerland: Springer, 2018, pp. 275–286.

[25] Z. Wen, R. Xu, and J. Du, "A novel convolutional neural networks for emotion recognition based on EEG signal," in *Proc. Int. Conf. Secur., Pattern Anal., Cybern.*, Dec. 2017, pp. 672–677.

[26] Z. Lan, O. Sourina, L. Wang, R. Scherer, and G. R. Müller-Putz, "Domain adaptation techniques for eeg-based emotion recognition: A comparative study on two public datasets," *IEEE Trans. Cogn. Develop. Syst.*, vol. 11, no. 1, pp. 85–94, Mar. 2019.

[27] Y. Luo and B.-L. Lu, "EEG data augmentation for emotion recognition using a conditional Wasserstein GAN," in *Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2018, pp. 2535–2538.

[28] D. E. Hernández, L. Trujillo, E. Z-Flores, O. M. Villanueva, and O. Romo-Fewell, "Detecting epilepsy in EEG signals using time, frequency and time-frequency domain features," in *Computer Science and Engineering-Theory and Applications*. Cham, Switzerland: Springer, 2018, pp. 167–182.

[29] D. Ayata, Y. Yaslan, and M. Kamasak, "Multi channel brain EEG signals based emotional arousal classification with unsupervised feature learning using autoencoders," in *Proc. 25th Signal Process. Commun. Appl. Conf. (SIU)*, May 2017, pp. 1–4.

[30] Q. Zhang, X. Chen, Q. Zhan, T. Yang, and S. Xia, "Respiration-based emotion recognition with deep learning," *Comput. Ind.*, vols. 92–93, pp. 84–90, Nov. 2017.

[31] P. Yisheng, "A parametric study of the dynamics of emotional response time," M.S. thesis, Capital Normal Univ., Beijing, China, 2012.

[32] N. H. Frijda, "The laws of emotion," *Amer. Psychol.*, vol. 43, no. 5, p. 349, 1988.

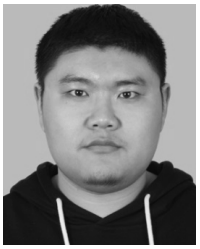
[33] P. Rozin and E. B. Royzman, "Negativity bias, negativity dominance, and contagion," *Personality Social Psychol. Rev.*, vol. 5, no. 4, pp. 296–320, 2001.

[34] M. Lewicka, J. Czapinski, and G. Peeters, "Positive-negative asymmetry or 'when the heart needs a reason,'" *Eur. J. Social Psychol.*, vol. 22, no. 5, pp. 425–434, 1992.

[35] R. F. Baumeister, E. Bratslavsky, C. Finkenauer, and K. D. Vohs, "Bad is stronger than good," *Rev. Gen. Psychol.*, vol. 5, no. 4, pp. 323–370, 2001.

[36] D. E. Kanouse and L. R. Hanson, Jr., "Negativity in evaluations," Lawrence Erlbaum Associates, Univ. California, Los Angeles, CA, USA, 1987.

CHUNMEI QING received the B.Sc. degree in information and computation science from Sun Yat-sen University, China, in 2003, and the Ph.D. degree in electronic imaging and media communications from the University of Bradford, U.K., in 2009. She was a Postdoctoral Researcher with the University of Lincoln, U.K. Since 2013, she has been an Associate Professor with the School of Electronic and Information Engineering, South China University of Technology (SCUT), Guangzhou, China. Her main research interests include image/video processing, computer vision, pattern recognition, and machine learning.



RUI QIAO received the B.Eng. degree in information engineering from the South China University of Technology, Guangzhou, China, in 2016, where he is currently pursuing the M.Eng. degree with the School of Electronic and Information Engineering. His current research interests include emotion recognition and machine learning.

XIANGMIN XU received the Ph.D. degree from the South China University of Technology, Guangzhou, China, where he is currently a Full Professor with the School of Electronic and Information Engineering. His current research interests include image/video processing, human–computer interaction, computer vision, and machine learning.

YONGQIANG CHENG is currently a Senior Lecturer with the Department of Computer Science and Technology, University of Hull, U.K. His research interests include digital healthcare technologies, AI, control theory and applications, embedded systems, secure communication, and data mining.