Classifier Fusion Method Based Emotion Recognition for Mobile Phone Users

Luobing Dong[1,2](✉), Yueshen Xu[1], Ping Wang[2], and Shijun He[1]

[1] Xidian University, Xi'an 710071, Shaanxi, China
{lbdong,ysxu}@xidian.edu.cn, 18700197078@163.com
[2] National Key Laboratory of Science and Technology on ATR,
College of Electronic Science National University of Defense Technology,
Changsha 410073, Hunan, China
760521407@qq.com

Abstract. With the development of modern society, people are paying more and more attention to their mental situation. An emotion is an external reaction of people's psychological state. Therefore, emotion recognition has attached widespread attention and become a hot research topic. Currently, researchers identify people's emotion mainly based on their facial expression, human behavior, physiological signals, etc. These traditional methods usually require some additional ancillary equipment to obtain information. This always inevitably makes trouble for users. At the same time, ordinary smart-phones are equipped with a lot of sensor devices nowadays. This enables researchers to collect emotion-related information of mobile users just using their mobile phones. In this paper, we track daily behavior data of 50 student volunteers using sensors on their smart-phones. Then a machine learning based classifier pool is constructed with considering diversity and complementary. Base classifiers with high inconsistent are combined using a dynamic adaptive fusion strategy. The weights of base classifiers are learned based on their prior probabilities and class-conditional probabilities. Finally, the emotion status of mobile phone users are predicted.

Keywords: Classifier fusion method · Emotion recognition · Dynamic adaptive fusion strategy · Machine learning

1 Introduction

Nowadays, people have greater material wealth than previous generations. We can access to an abundance of material resources. However, life has become more pressured and challenging than ever before. These pressures come from a wide range of sources, including our workplaces, the education system, our family and friends, etc. The constant pressure that they exert on us can have a profound effect on our mental health. This makes people pay more and more attention

Supported by the Fundamental Research Funds for Central Universities (JB161004).

Q. Li et al. (Eds.): BROADNETS 2019, LNICST 303, pp. 216–226, 2019.
https://doi.org/10.1007/978-3-030-36442-7_14

to their mental health which is fundamentally linked with their physical health. Researchers state that emotion regulation is an essential feature of mental health [1]. Therefore, real-time emotion recognition is very important for people's timely emotional regulation and the clinical diagnosis of mental illness. In recent years, emotion recognition has been studied as a promising technology in a broad range of mental health monitor [2–4].

Emotion recognition has also been an important research topic in the fields of artificial intelligence owing to its significant academic and commercial potential [5]. By 2019, AI-based emotion detection has become a 20 billion dollars industry [6]. Scientists typically identify human emotion based on physiological signals including facial or verbal expressions [7], skin temperature [8], blood volume pulse [9], electromyographic signal [10], etc. However, these signals always need special bio-sensors or equipment to get. This makes trouble to its application. People and their doctors often hope to monitor their emotion changing process to help them do some judgments and decisions. For example, some doctors want to establish a connection between some clinical symptoms and patients' emotion status. To do that, patients always need to rent or buy some special carry-on equipment. This is inconvenient and may cause negative emotions. We know that life is filled with ups and downs. It is common for people to have shifts in emotions. Therefore, it is very important that the emotions of the observed person are unconsciously recorded and recognized in time.

Over recent decades the pace of the Internet has gradually changed from personal computer (**PC**) storm to smart phone competition. Not only are mobile phones more affordable, but mobile phones also are becoming more powerful in functions. People spend more and more time on mobile phones everyday. And most people carry their mobile phones with them all the day. Furthermore, sensor devices are no longer the special configuration of high-end mobile phones. Most mid-end and low-end mobile phones are equipped with dozens of sensor devices, such as gravity, gyroscope, pressure, temperature, lighting, etc. These make unconsciously record and recognize uses' emotion possible [11]. The strength of the finger on the screen, the speed of walking, the volume of the sound, etc. indicate different emotion status of users.

The human emotion status is an important factor in creating good mental or physical health. Therefore, identifying factors that directly influence the emotion of individuals and using them to predict human emotional state in real-time will have enormous societal benefits. In this paper, we develop a novel method to identify the emotion status of mobile users by collecting environmental and physiological features using sensors of their mobile phones. We track daily behavior data of 50 student volunteers using sensors on their smart-phones. Then a machine learning based classifier pool is constructed with considering diversity and complementary. Base classifiers with high inconsistent are combined using a dynamic adaptive fusion strategy. The weights of base classifiers are learned based on their prior probabilities and class-conditional probabilities. Finally, the emotion status of mobile phone users are predicted.

The rest of this paper is organized as follows. The related work is discussed in Sect. 2. Then, the overview of the proposed method is presented in Sect. 3. In Sect. 4, we discuss the methodology. The experiment evaluation is presented in Sect. 5. Finally, we conclude the paper in Sect. 6.

2 Related Work

Emotion recognition methods can be easily divided into three categories: text analysis based emotion recognition, facial and verbal expression analysis based emotion recognition, and physiological signal analysis based emotion recognition. In this section, we will introduce this three categories separately.

There are two main types of text analysis based emotion recognition methods: word classification based method and semantic analysis based method. The former requires a special dictionary which stores as many emotional scores of emotionally salient words as possible. With this dictionary, the emotional scores of all emotionally salient words which are contained in a piece of text can be obtained. Then the emotion that this text wants to express can be identified by aggregating these scores. Identifying and collecting as many emotionally salient words as possible is critical to the effectiveness of this method. Therefore, Many studies focus on how to find emotionally salient words and score them [12,13]. The latter performs the emotion recognition of the corresponding text through the semantic network. This method depends heavily on the richness of the semantic knowledge base which it uses. Baccianella presents SentiwordNet which is devised for supporting sentiment classification and opinion mining applications [14]. Cambria combines the largest existing public knowledge classification with a natural language-based common sense knowledge semantic network [15]. The multi-dimensional extension of the generated knowledge base is used for sentiment analysis.

The rapid development and rise of the deep learning neural network make the analysis of images, text, and speech more effective. This makes emotion recognition based on facial and verbal expression a hot topic. Different expressions often reflect different emotions. So we can judge people's emotions by analyzing people's facial expressions and accompanying facial muscle movements. For example, when the corner of the mouth curves up and wrinkles radiate out from the corners of the eyes, we can usually judge the emotion status of someone as happy. When someone's eyes pop out and his brows are puckered in a frown. Obviously, his emotional status at this time is anger. Facial expression based emotion recognition may be either based on local features or based on overall features. The former is based on the fact that people in different situations have different shapes, sizes, and relative positions of facial features. Considering the difference of the overall facial features under different emotions, the latter is based on the overall facial feature. And the range of extracted features is the entire face. Researchers have made many achievements in this field. For example, Affectiva creates a facial expression dataset with a large number of facial expression pictures of different races, ages, and genders [16]. Then Affectiva design artificial intelligence algorithms to identify people's emotions by observing all the changes of facial expression such as changes of textures and wrinkles of the face, changes in the shape of the facial features, etc.

Many people always wear a poker face in their normal life. They like emotionless expressions that give no indication of their thoughts or intentions. This limit the effective of facial and verbal expression based emotion recognition technology. Therefore, researchers try to identify people's emotion by analyzing their physiological signals [17–21]. Liu et al. propose a real-time emotion recognition system based on **EEG** signals. The advantage of this system is that it is combined with the clip database [22]. Wang et al. study various emotional characteristics of EEG signals, track the changes of EEG signals. They establish the connection between EEG characteristics and emotion status and finally identify the classification of emotions [23]. Machine learning has a good effect on the fusion of multimodal information, such as the fusion of various physiological features and EEG signals. This makes machine learning based algorithms have good performance on separating signals with high spatial and frequency dimensions [24,25].

3 Overview of the Method

In this section, we introduce the main idea of the proposed method. Figure 1 shows its flow chart. We choose 50 student volunteers from Xidian University. They install our data acquisition application EAmobile on their Android mobile phones. EAmobile can collect their daily emotion related information through the gravity sensor, pressure sensor, gyroscope sensor, speed sensor and voice sensor. This collection process lasts for half a year. The volunteers help us label their information with their real emotion status. Their labeled information are transferred to our remote server. After preprocessing and feature extraction, the data is stored in Mysql database. This data is used to train our fusion classifier. The trained model is used to predict the real-time emotion status of other mobile phone users based on their instantaneous mobile phone sensing information.

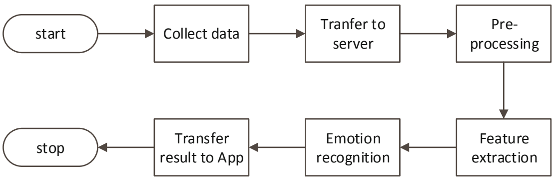

Fig. 1. The flow chart of the proposed method

Inspired by [26], we propose an emotional model considering the unsuitability of discrete variables for linear analysis and transformation. It is shown in Fig. 2. This model is consist of two dimensions: activity level and pleasure level. The pleasure level is used to measure the positive or negative degree of the user. The activity level is used to measure the degree of proactiveness or passiveness of the user. The coordinate system is established with the level of pleasure as the x-axis and the level of activity as the y-axis. Then the corresponding emotion status can be positioned in the coordinate system.

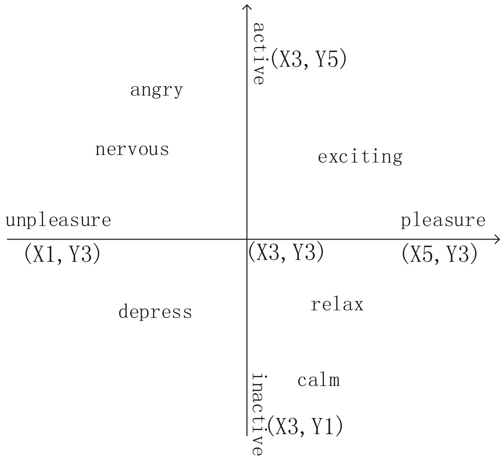

Fig. 2. The emotional model

We divide the pleasure level and activity level into five levels separately. We set $(X3, Y3)$ as the origin. Then $(X1, Y1)$ represents very unpleasant and very inactive. $(X5, Y5)$ represents very pleasant and very active. The statistics result based on the collected information shows that most emotion status of the mobile users stay between the third and fourth levels. This means that the value of their emotion status (Xi, Yi) satisfies: $X3 \leq Xi \leq X4$ and $Y3 \leq Yi \leq Y4$. The result indicates that users always are in calm status at most time. And the probability of being in a happy status is greater than the probability of being in a sad status. This is obviously reasonable. Table 1 shows the users' emotion status distribution which is calculated based on our statistics.

Table 1. The distribution (%) of users' emotion status.

Pleasure level \ Activity level	$X1$	$X2$	$X3$	$X4$	$X5$
$Y1$	0.25	0.23	0.21	0	0
$Y2$	0	6.79	2.18	0.32	0.41
$Y3$	0.57	5.62	42.06	1.32	0
$Y4$	0.78	1.56	20.05	12.57	0.78
$Y5$	0	0	0.58	1.95	1.77

4 Methodology

This section focus on the detailed classifier fusion method used in this paper. We first describe how to eliminate the useless data contained in the collected daily behavior data. Then, we explain the extracted features in detail, followed by description to the used classifier fusion method.

4.1 Data Pre-processing

Data obtained directly through mobile phone sensors cannot be used directly. Because the collected raw data has many problems such as noise, incomplete, inconsistency, etc. There are many reasons for this useless. For example, the application may not be used correctly. Data collection process may be disrupted or abort by device failures or interference. So we should pre-process the data so that all data can be used. In this paper, the following methods are used for data pre-processing:

- Filtering illegal data. We filter illegal data in the system artificially to avoid importing them into the emotion recognition process and confusing classifiers.
- Eliminating incomplete data. If a large number of data items are missing in a certain sample, the system will reject the sample. If there are only very few items miss in a sample, we just fill it with the default value.
- Medium filtering. We perform median filtering on the acquired waveform data. This method can make the waveform bumpier. Then most key features will be more clear.

4.2 Feature Extraction

In this paper, we collect mobile users' daily behavior data through a lot of mobile phone sensors. Data from different sensors have different features which are related the emotion status of users. Table 2 shows the details of these features. Due to the length limitation of this paper, we just give the feature detail of light sensor as an example.

Data which is collected through the light sensor has notable and distinctive values. When we put the phone in a confined space such as a pocket or a bag, the output value of light sensor is small. It is basically in $[0, 100]$. When we put it under normal light, the values are generally concentrated in $[100, 1000]$. When we are outdoors, the values are mostly in $[1000, 2000]$. Therefore, we can easily extract the phone's current status of use by the output value of the light sensor.

4.3 Classifier Fusion Method

The classifier fusion method we use can be divided into three phases: generation phase, selection phase, fusion phase. We will describe these three phases separately.

Table 2. Features need to be extracted from each types of sensor.

Sensor	Number of features	Description
Gravity	3	Time of portrait orientation, time of landscape orientation, times of exchanging between these two orientations
Gyroscope	2	Rotation time, average angular velocity
Accelerator	2	Total shaking time, severity of shaking, times of shaking
Light	1	One of following states: no use, outdoor, indoor
Pedometer	2	Step count, difference between the average speed and the largest speed
Global Positioning System (**GPS**)	1	Location entropy
Network speed	3	Fluctuation of speed, speed, strength of signals

Generation Phase: At this stage, a base classifier pool is constructed considering the diversity of base classifiers. We choose most of the classification algorithms based on neural networks and decision trees as the base classifiers. They are trained on different data sets.

Selection Phase: At this stage, classifiers which are good for the data set to be classified are selected from the classifier pool for the next phase. Breiman suggested that the more unstable the base classifier the better the combined classifier will achieve [27]. This instability means that even a very subtle change in the training sample may result in a very different final decision classification result. In other words, the more sensitive the base classifier algorithm is to the sample during the learning and training process, the better the final combination result.

Therefore, in order to improve the accuracy and generalization ability of the final decision classifier, it is necessary to select suitable classifiers from the base classifier pool. The selection method generally considers the accuracy and difference of the base classifier. In terms of accuracy, we set a threshold. The base classifier whose accuracy is above the threshold can participate in the selection phase. In terms of difference, we use the inconsistent measurement [28].

For two classifiers C_i and C_j, we define their difference as formula 1.

$$\Delta_{i,j} = \frac{S_{\bar{c_i},c_j} + S_{c_i,\bar{c_j}}}{S_{\bar{c_i},\bar{c_j}} + S_{\bar{c_i},c_j} + S_{c_i,\bar{c_j}} + S_{c_i,c_j}} \tag{1}$$

Here, $S_{\bar{c_i},\bar{c_j}}$ represents the number of the samples that the classifier C_i and C_j both make the wrong prediction. $S_{\bar{c_i},c_j}$ represents the number of samples that the classifier C_i makes the wrong prediction and C_j makes the right prediction. $S_{c_i,\bar{c_j}}$ represents the number of samples that the classifier C_i makes the wrong prediction and C_j makes the right prediction. S_{c_i,c_j} represents the number of samples that the classifier C_i and C_j both make the right prediction. The difference between the two classifiers $\varDelta_{i,j}$ is in the range [0, 1]. For all samples, if the two classifiers have the same recognition effect, the difference is 0. A larger value indicates a greater difference between two classifiers.

We assume that there are n classifiers in the base classifier pool. Let $\varGamma_i$ denote the average of the difference between the base classifier C_i and all other base classifiers. It can be defined by formula 2.

$$\varGamma_i = \frac{\sum_{j=1}^{n} \varDelta_{i,j}}{n-1} \tag{2}$$

Then the average difference of the base classifier pool can be calculated by formula 3.

$$Avg = \frac{\sum_{i=1}^{n} \varGamma_i}{n} \tag{3}$$

When $\varGamma_i \geq Avg$, it indicates that the base classifier C_i is good. Then C_i will be chosen as a member of the final fusion classifier.

Fusion Phase: There are many fusion strategies such as average, majority vote, weight, etc. The average method and majority vote method are not stable in performance and easily affected by extreme values. The weighting method gives different weights to different base classifiers. For example, the reference weight can be given based on the accuracy of the base classifier. The more accurate the base classifier, the greater its weight.

We use a dynamic adaptive weight strategy for classifier fusion. Inspired by Bayesian theory, we use the historical decision of the training data as the prior probability and use the classification confidence of the current input test sample as its class-conditional probability to dynamically compute the weight of each base classifier. In the training phase, we train each base classifier on the training data set and build up the confusion matrix for each classifier. The confusion matrix is be used to measure the apriori behavior of each base classifier. Given m class labels, the confusion matrix of classifier C_i can be calculate as formula 4.

$$M^i = \begin{bmatrix} e_{11}^i & e_{12}^i & \cdots & e_{1m}^i \\ e_{21}^i & e_{22}^i & \cdots & e_{2m}^i \\ \cdots & \cdots & \cdots & \cdots \\ e_{m1}^i & e_{m2}^i & \cdots & e_{mm}^i \end{bmatrix} \tag{4}$$

Here, e_{lk}^i indicates the probability that a sample which belongs to C_l is identified as a member of C_k by the ith classifier. We set the confusion matrix which is generated by the training data as the prior probability. In the recognition process, we set the confidence $P_i(D_k|X)$ which is the output of one classifier C_i as the class-conditional probability of class D_k. We define the inverse reliability of C_i at D_k according to D_l as formula 5.

$$\phi_{kl}^{i} = exp(-|P_i(D_k|X) - P_i(D_l|X)|) \tag{5}$$

Then, we combine confusion matrix which is based on the prior probability with the inverse reliability which is based on the class-conditional probability to get the fusion weight of C_i. Formula 6 gives the definition of the fusion weight of C_i.

$$w_{ki} = \frac{e_{kk}^{i}}{\sum_{l=1,l\neq k}^{m} e_{kl}^{i}\phi_{kl}^{i} + \sum_{l=1,l\neq k}^{m} e_{lk}^{i}\phi_{lk}^{i}} \tag{6}$$

The confidence of the fusion classifier for class D_i can be got by formula 7.

$$P(D_i|X) = \sum_{i=1}^{n} w_{ki} P^{i}(C_k|X) \tag{7}$$

The class with the largest confidence of the fusion classifier is the final classification decision.

5 Evaluation

The evaluation results of the proposed system will be described in this section. We choose 8 base classifiers to construct the classifier pool. The original accuracy of these base classifiers on our sample are shown in Table 3.

Table 3. Original accuracy.

Algorithm	Accuracy (%)	Algorithm	Accuracy (%)
LAD	66.43	Part Rule	61.60
CART	64.87	MLP	58.54
REP	63.76	SVM	53.27
DTNB	62.31	LR	51.04

We compare three different fusion strategies: majority vote method, accuracy based method and our method. The accuracy of them on our samples are shown in Table 4. All experiments are repeated 1000 times. And the result is the average result of all experiments. We can see that our method can get the largest accuracy. We create two base classifier pools with two different groups of base classifiers. The first one is consist of LAD, REP and CART. These three classifiers have the largest accuracy and average difference. So the accuracy strategy has better performance than the vote strategy. When we change the strategy to the dynamic adaptive strategy with our parameters, the accuracy is significantly improved. In the second group, we replace REP which has the worst accuracy percentage with DTNB. The accuracy of the system is improved under all strategies.

Table 4. Experiment results.

Base classifier1	Base classifier2	Base classifier3	Fusion strategy	accuracy
LAD	REP	CART	majority vote	64.85
LAD	REP	CART	accuracy	65.49
LAD	REP	CART	dynamic adaptive strategy	67.53
LAD	CART	DTNB	majority vote	66.78
LAD	CART	DTNB	accuracy	68.39
LAD	CART	DTNB	dynamic adaptive strategy	71.67

6 Conclusion

In this paper, we present an emotion recognition system for mobile phone users. By collecting behavior data of 50 student volunteers through their mobile phone sensors, we train the dynamic adaptive weight fusion classifier. The system achieve an average accuracy of 71.67%. In the future, we intend to extend our work to combine our system with android smart-phones to manage users' emotions in a real-time environment.

References

1. Gross, J.J., Muñoz, R.F.: Emotion regulation and mental health. Clin. Psychol.: Sci. Pract. **2**(2), 151–164 (1995)
2. Valstar, M., et al.: Avec 2016: depression, mood, and emotion recognition workshop and challenge. In: Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge, pp. 3–10. ACM (2016)
3. Trigeorgis, G., et al.: Adieu features? End-to-end speech emotion recognition using a deep convolutional recurrent network. In: 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 5200–5204. IEEE (2016)
4. Zhao, M., Adib, F., Katabi, D.: Emotion recognition using wireless signals. In: Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking, pp. 95–108. ACM (2016)
5. Ko, B.: A brief review of facial emotion recognition based on visual information. Sensors **18**(2), 401 (2018)
6. Baur, D.: AI-based emotion detection has become a 20B industry (2019). https://www.theguardian.com/technology/2019/mar/06/facial-recognition-software-emotional-science/. Accessed 01 July 2019
7. Li, M., et al.: Facial expression recognition with identity and emotion joint learning. IEEE Trans. Affect. Comput. (2018)
8. Greco, A., et al.: Skin admittance measurement for emotion recognition: a study over frequency sweep. Electronics **5**(3), 46 (2016)
9. Zhao, B., et al.: EmotionSense: emotion recognition based on wearable wristband. In: 2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI), pp. 346–355. IEEE (2018)

10. Vijayan, A.E., Sen, D., Sudheer, A.P.: EEG-based emotion recognition using statistical measures and auto-regressive modeling. In: 2015 IEEE International Conference on Computational Intelligence & Communication Technology, pp. 587–591. IEEE (2015)
11. Farhan, A.A.: Modeling Human Behavior using Machine Learning Algorithms (2016)
12. Deng, Z.-H., Luo, K.-H., Yu, H.-L.: A study of supervised term weighting scheme for sentiment analysis. Expert Syst. Appl. **41**(7), 3506–3513 (2014)
13. Khan, F.H., Qamar, U., Bashir, S.: Lexicon based semantic detection of sentiments using expected likelihood estimate smoothed odds ratio. Artif. Intell. Rev. **48**(1), 113–138 (2017)
14. Baccianella, S., Esuli, A., Sebastiani, F.: SentiWordNet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In: Lrec, vol. 10, pp. 2200–2204 (2010)
15. Cambria, E., et al.: New avenues in opinion mining and sentiment analysis. IEEE Intell. Syst. **28**(2), 15–21 (2013)
16. McDu, D., et al.: Affectiva-mit facial expression dataset (AM-FED): naturalistic and spontaneous facial expressions collected. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 881–888 (2013)
17. Martini, N., et al.: The dynamics of EEG gamma responses to unpleasant visual stimuli: from local activity to functional connectivity. NeuroImage **60**(2), 922–932 (2012)
18. Frantzidis, C.A., et al.: Toward emotion aware computing: an integrated approach using multichannel neurophysiological recordings and affective visual stimuli. IEEE Trans. Inf. Technol. Biomed. **14**(3), 589–597 (2010)
19. Balconi, M., Mazza, G.: Brain oscillations and BIS/BAS (behavioral inhibition/activation system) effects on processing masked emotional cues: ERS/ERD and coherence measures of alpha band. Int. J. Psychophysiol. **74**(2), 158–165 (2009)
20. Jenke, R., Peer, A., Buss, M.: Feature extraction and selection for emotion recognition from EEG. IEEE Trans. Affect. Comput. **5**(3), 327–339 (2014)
21. Khezri, M., Firoozabadi, M., Sharafat, A.R.: Reliable emotion recognition system based on dynamic adaptive fusion of forehead biopotentials and physiological signals. Comput. Methods Programs Biomed. **122**(2), 149–164 (2015)
22. Liu, Y.-J., et al.: Real-time movie-induced discrete emotion recognition from EEG signals. IEEE Trans. Affect. Comput. **9**(4), 550–562 (2017)
23. Wang, X.-W., Nie, D., Lu, B.-L.: Emotional state classification from EEG data using machine learning approach. Neurocomputing **129**, 94–106 (2014)
24. Balconi, M., Lucchiari, C.: EEG correlates (event-related desynchronization) of emotional face elaboration: a temporal analysis. Neurosci. Lett. **392**(1–2), 118–123 (2006)
25. Iacoviello, D., et al.: A real-time classification algorithm for EEGbased BCI driven by self-induced emotions. Comput. Methods Programs Biomed. **122**(3), 293–303 (2015)
26. Yang, R., Xi, C., Xi, S.: J. Front. Comput. Sci. Technol. **10**(6), 751–760 (2016)
27. Breiman, L.: Bias, variance, and arcing classifiers. Technical report, 460, Statistics Department, University of California, Berkeley (1996)
28. Bowes, D., Randall, D., Hall, T.: The inconsistent measurement of message chains. In: 2013 4th International Workshop on Emerging Trends in Software Metrics (WETSoM), pp. 62–68. IEEE (2013)