Biomedical Signal Processing and Control 18 (2015) 118–126

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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

Affective detection based on an imbalanced fuzzy support vector machine

Jing Cheng[a,*], Guang-Yuan Liu[b]

[a] College of Computer and Information Science, Southwest University, Chongqing 400715, China
[b] School of Electronic and Information Engineering, Southwest University, Chongqing 400715, China

ARTICLE INFO

Article history:
Received 16 July 2014
Received in revised form
18 December 2014
Accepted 19 December 2014
Available online 4 February 2015

Keywords:
Imbalanced classification
IBFSVM
Affective detection
Galvanic skin response

ABSTRACT

The interpretation of physiological signals is an important subject in affective computing. In this paper, we report an experiment to collect affective galvanic skin response signals (GRS), and describe a new imbalanced fuzzy support vector machine (IBFSVM) for their classification. IBFSVM introduces denoising factors and class compensation factors, thus defining a new fuzzy membership. The effectiveness of IBFSVM is verified on various real and artificial datasets. We define an appropriate evaluation criterion (*g_mean*) that combines the classification accuracy of positive and negative samples, and show that IBFSVM outperforms traditional support vector machines on imbalanced datasets. By running the IBFSVM for the datasets in our experiment, we can find that the *g_mean* of happiness, sadness, angry and fear is 85.17%, 86.6%, 87.4%, and 81.53% respectively. So IBFSVM is an effective and feasible solution for imbalanced learning in our experiment.

1. Introduction

Affective detection represents the foundation of harmonious human–machine interaction, and is thus one of the key problems in affective computing. Human emotional states are inherently linked to psychological and physiological changes, which are manifested by behavioral, facial, and physical signals. By analyzing these observable signals, a human's affective state can be inferred. Compared with facial expressions, speech, and body posture, physiological signals better reflect the activity pattern of the autonomic nervous system. Hence, affective detection based on physiological signals has been studied by many researchers [1–3]. Support vector machines (SVMs) are widely used for pattern recognition, and have been employed for affective detection based on physiological signals [4,5].

Although SVMs perform well with balanced datasets, their capability is greatly reduced when the dataset is imbalanced [6,7]. SVM classifiers can produce models that are biased toward the majority class and perform poorly with the minority class. In the real world, imbalanced datasets are common, so improving the performance of imbalanced learning is a very important for SVMs.

SVMs are very sensitive to noise and outliers, which affect the slope of the decision hyperplane, and thus the performance of the SVM. Lin and Wang [8] proposed a fuzzy SVM (FSVM) to overcome the influence of outliers, but FSVM cannot deal with imbalanced samples; the accuracy in the minority class is still poor. Based on the above analysis, we propose an imbalanced FSVM (IBFSVM) that not only deals with outliers, but can also handle imbalanced datasets. In this paper, IBFSVM is applied to imbalanced learning in affective detection based on galvanic skin response (GSR) signals.

The remainder of this paper is arranged as follows. In Section 2, we discuss some related work on imbalanced learning with SVMs, and Section 3 describes the IBFSVM algorithm in detail, and compares the results from applying IBFSVM to artificial datasets and UCI datasets. Section 4 introduces the experimental process for affective detection and presents the classification results of IBFSVM for affective GSR datasets. Finally, Section 5 summarizes the results and conclusions from this study.

2. Related work

In the field of imbalanced learning, we often regard the minority class as the positive class and the majority class as the negative class. In general, there are two methods for dealing with imbalanced classification. The first preprocesses the datasets to balance them before training the classifier. Different resampling methods, such as random and focused oversampling and undersampling, fall

* Corresponding author. Tel.: +86 13983390890.
E-mail address: cjcat@swu.edu.cn (J. Cheng).

http://dx.doi.org/10.1016/j.bspc.2014.12.006
1746-8094/

into this category. An oversampling method has been proposed that can increase the number of positive classes to balance the training data before classification [9]. However, this method simply duplicates the positive samples, which will augment the system overhead and lead to over-fitting. Chawla et al. [10] described the synthetic minority oversampling technique algorithm (SMOTE), which generates new positive data by linear interpolation in the neighborhood of existing positive samples, rather than directly duplicating them. This method is widely used to treat imbalanced datasets and enhance imbalanced learning. Rehan et al. [11] studied an undersampling method in which the number of samples in the majority class is decreased to balance the training data. This method loses some useful information from the negative samples.

The second solution to imbalanced learning is to modify existing SVMs to construct cost-sensitive classifiers that are less sensitive to imbalanced classes. For example, Veropoulos et al. [12] proposed a method called DEC (the Different Error Cost) SVM, which modifies the objective function to assign two misclassification cost values C^+ and C^- as follows:

$$\begin{aligned} &Min \left(\frac{1}{2} w \cdot w + C^+ \sum_{[i|y_i=1]}^{l} \varepsilon_i + C^- \sum_{[i|y_i=-1]}^{l} \varepsilon_i \right) \\ &s.t. \quad y_i(w \cdot \phi(x_i) + b) \geq 1 - \varepsilon_i \\ &\varepsilon_i \geq 0 \end{aligned} \tag{1}$$

In Eq. (1), C^+ is the misclassification cost for the positive-class examples, and C^- is the misclassification cost for the negative-class examples. By assigning a higher misclassification cost for the positive (minority) samples than for the negative (majority) samples, i.e., $C^+ > C^-$, the influence of any imbalance between the classes can be significantly reduced. Imam et al. [13] proposed another modified SVM, named zSVM, for the learning of imbalanced datasets. First, consider the SVM decision function:

$$f(x) = sign\left(\sum_{i=1}^{l} \alpha_i y_i K(x_i, x) + b \right) = sign\left(\sum_{i}^{l_1} \alpha_i^+ y_i K(x_i, x) + \sum_{i}^{l_2} \alpha_i^- y_i K(x_i, x) + b \right) \tag{2}$$

where α_i^+ are the coefficients of the positive-support vectors, α_i^- are the coefficients of the negative-support vectors, and l_1, l_2 represent the sample size of positive- and negative-training examples, respectively. In the zSVM method, the magnitude of α_i^+ is increased by multiplying the positive-support vectors by some small, positive value z. Then, the modified SVM decision function can be represented as follows:

$$f(x) = sign\left(z \sum_{i}^{l_1} \alpha_i^+ y_i K(x_i, x) + \sum_{i}^{l_2} \alpha_i^- y_i K(x_i, x) + b \right) \tag{3}$$

This modification of α_i^+ increases the weight of the positive-support vectors in the decision function, thus decreasing the bias toward the majority class. In addition to these methods, several kernel-modification techniques for training SVMs with imbalanced datasets have been proposed in [14].

The FSVM proposed by Lin and Wang [8] assigns different fuzzy membership values s_i to different examples to reflect the importance to their own class. More important examples are assigned higher membership values, and less important ones, such as outliers and noise, are assigned lower membership values. The mathematical model of FSVM is as follows:

$$\begin{aligned} &\min \quad \frac{1}{2}||w||^2 + C\sum_{i=1}^{n} s_i \xi_i \\ &s.t. \quad y_i[w \cdot \varphi(x_i) + b] - 1 + \xi_i \\ &\quad \xi_i \geq 0, \quad i = 1, 2, 3, \ldots, n \end{aligned} \tag{4}$$

To solve the FSVM optimization problem, Eq. (4) is transformed into the following dual Lagrangian:

$$\begin{aligned} &\max \quad W(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j K(x_i, x_j) \\ &s.t. \quad \sum_{i=1}^{n} \alpha_i y_i = 0, 0 \leq \alpha_i \leq s_i C, \quad i = 1, 2, 3, \ldots, n \end{aligned} \tag{5}$$

In Eq. (4), the membership s_i of sample x_i is incorporated into the objective function. Therefore, smaller s_i will reduce the effect of parameter ξ_i in the objective function, meaning the corresponding x_i is considered to be less important. The parameter C is effectively the cost assigned to a misclassification. In Eq. (4), each sample is assigned a different misclassification cost value $s_i C$, such that more important samples are assigned higher costs, and less important ones are assigned lower costs. For imbalanced classification, $s_i C$ becomes larger as the value of s_i is increased in the minority-class samples, which will augment the influence of positive samples when training the SVM. Literature [15] has proposed to use fuzzy support vector machine to solve unbalanced classification. In this article Batuwita and Palade defined the membership function as follows:

$$m_i^+ = f(x_i^+) r^+ \tag{6}$$

$$m_i^- = f(x_i^-) r^- \tag{7}$$

In formula (6) and (7), they assign $r^+ = 1$, and $r^- = r$, where r is the minority-to-majority class ratio. And function $f(x_i)$ gives the within-class importance of a training example. In this article we adopt similar principles of design fuzzy membership functions. We design two fuzzy factor, one is used to get rid of noise, another is to realize class compensation, then we use adjustable parameter a to control the two factor's weight. Experimental results show that our method can achieve better performance.

3. Algorithm for the imbalanced fuzzy support vector machine

Enhancing the classification accuracy of the positive class is the key problem in imbalanced learning. This involves improving the design of the membership function s_i in traditional FSVM. In this section, we discuss two factors that reduce the classification accuracy of the positive class and cause the SVM hyperplane to deviate from its optimal position. The membership function s_i of IBFSVM is then designed according to these two factors.

3.1. Factors that skew the separating hyperplane

Following a series of experiments and related analysis, we have determined two factors that cause the hyperplane to become skewed in imbalanced learning. The first factor is a large difference between the number of positive and negative samples, such as in the detection of cancer patients in a massive normal crowd. The second factor concerns the distribution of the two classes. Although the number of positive and negative samples may be well balanced, the distribution can be quite different. For example, in one class the samples may be very close, whereas in the other they are

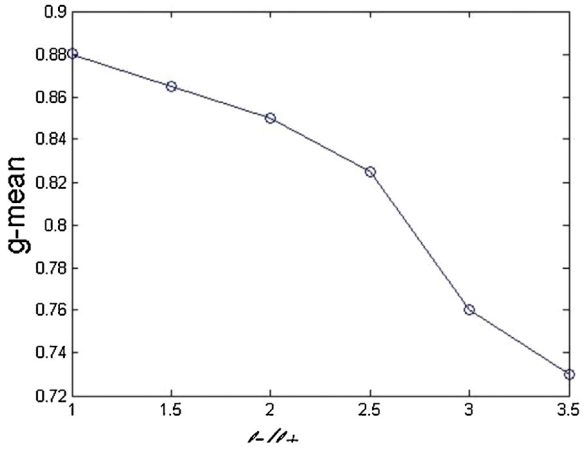

Fig. 1. Influence of sample size variance in the two classes.

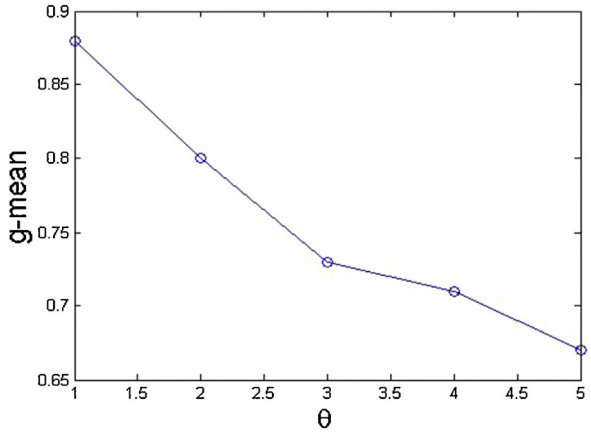

Fig. 2. Influence of different distributions in the two classes.

very sparse. The following data and tests illustrate that an imbalance in sample size or sample distribution will lead to a skewed hyperplane.

Experiment 1. An imbalance in the positive and negative sample sizes will reduce the capacity of standard SVM. First, positive and negative data are generated artificially. Their mean vectors are $\mu^+ = [3,3]$ and $\mu^- = [6,6]$, respectively, and the covariance matrices are $\Sigma^+ = \Sigma^- = [1, 0; 01]$. The sample size of the positive class is $l^+ = 200$, and the sample size of the negative class is varied to be $l^- = 300$, 400, 500, 600, and 700. We use a standard SVM, known as C-SVM, to train the classifier, and take the kernel to be a radial basis function (RBF). Kernel options of $\gamma = 2^{-8}$, $C = 100$ are applied. The C-SVM algorithm is run 10 times for each value of l^-. In every test, the first 100 positive data are placed in the training set, and the last 100 positive data are placed in the test set. The last 100 negative data are put in test set, while the remaining negative data are placed in the training set. We use $g\_mean = \sqrt{sp \times se}$ to evaluate the algorithm's performance. *sp* represents the classification accuracy of the positive class, and *se* represents the classification accuracy in the negative class. The trend of *g_mean* with respect to the ratio l^-/l^+ is shown in Fig. 1. This shows that, as the ratio between l^- and l^+ increases, the value of *g_mean* decreases, which means with the increasing of the ratio between l^- and l^+, the performance of the classifier will be worsened.

Experiment 2. A large difference in the distribution of the two classes will reduce the performance of the SVM. Again, positive and negative data are artificially generated to have mean vectors of $\mu^+ = [3,3]$ and $\mu^- = [6,6]$, respectively. In this experiment, we use the covariance matrices Σ^+ and Σ^- to describe the different distributions of the two classes, the covariance matrices satisfy $\Sigma^+ = [1, 0; 01]$ and $\Sigma^- = \theta\Sigma^+$, where the value of θ is either 2, 3, 4, or 5. There are 300 samples in both the positive class and the negative class. We use C-SVM to train the classifier, take RBF as the kernel, and set kernel options of $\gamma = 2^{-8}$, $C = 100$. The algorithm is run 10 times for each value of θ. In every test, we take the first 200 data as the training set and the last 100 data as the test set. We again use the *g_mean* criterion to evaluate the performance of C-SVM. The trend of *g_mean* with respect to θ is presented in Fig. 2. From Fig. 2, we can see that, for equivalent sample sizes, an increase in the distribution difference between the two classes worsens the performance of the SVM.

From the results of the above experiments, we can conclude that the hyperplane will be skewed toward the majority class when there is a large disparity in sample size. If the two classes are of equal size, the hyperplane will be skewed toward that which has a sparser distribution.

3.2. Design of a fuzzy factor for class compensation

The above analysis shows that differences in sample size and sample distribution affect the performance of the SVM. Thus, we aim to combine these two factors to produce a fuzzy membership function s_i that can reflect the importance of imbalanced samples and enhance the accuracy of the minority class.

First, based on the main idea of cost-sensitive SVMs, different costs of C^+ and C^- will be applied to the positive and negative classes, respectively. Generally, $C^+ > C^-$, which means that the cost of misclassifying positive samples is greater than that of negative samples. The relationship between C^+ and C^- is discussed in [16]. Setting $C+$ and C^- to satisfy Eq. (8) will reduce, to some extent, the influence of having different numbers of samples in the two classes.

$$\frac{C^+}{C^-} \propto \frac{l^-}{l^+} \tag{8}$$

It is well known that the covariance matrix reflects the degree to which samples deviate from the mean, and that the projected standard deviation of the covariance matrix, which is projected on the normal vector of the decision hyperplane, can represent the distribution of the samples. The projected standard deviation of the covariance matrix for positive and negative samples is given by:

$$n^+ = \sqrt{w^T \Sigma^+ w}, \quad n^- = \sqrt{w^T \Sigma^- w}, \tag{9}$$

where Σ^+ and Σ^- represent the covariance matrices for the positive and negative classes, respectively. Eq. (9) gives theoretical expressions that cannot be used in the actual calculation. To calculate n^+ and n^-, the following steps are needed:

(1) Train the preliminary decision hyperplane by standard SVM, and the normal vector ω_1 of the hyperplane is given by:

$$\omega_1 = \sum_{i=1}^{n} \alpha_i y_i \varphi(x_i) \tag{10}$$

(2) Calculate the projected value of sample $\varphi(x_j)$, which is projected in the normal vector ω_1 of the decision hyperplane. This is given by:

$$\omega_1 \cdot \varphi(x_j) \sum_{i=1}^{n} \alpha_i y_i \varphi(x_i)\varphi(x_j) = \sum_{i=1}^{n} \alpha_i y_i k(x_i, x_j), \quad j = 1, 2, \ldots, n. \tag{11}$$

(3) Calculate the projected standard deviation of the covariance matrix for the positive and negative classes as:

$$n^+ = \sqrt{\frac{1}{l^+ - 1}\sum_{j=1}^{l^+}\left[\sum_{i=1}^{l}\alpha_i y_i k(x_i, x_j) - \frac{1}{n^+}\sum_{j=1}^{l^+}\sum_{i=1}^{l}\alpha_i y_i k(x_i, x_j)\right]^2} \quad (12)$$

$$n^- = \sqrt{\frac{1}{l^- - 1}\sum_{j=1}^{l^-}\left[\sum_{i=1}^{l}\alpha_i y_i k(x_i, x_j) - \frac{1}{n^-}\sum_{j=1}^{l^-}\sum_{i=1}^{l}\alpha_i y_i k(x_i, x_j)\right]^2} \quad (13)$$

To reduce the influence of differences in the distribution of the positive and negative classes, according to Ref. [17], the costs of C^+ and C^- will be satisfied formula (14):

$$\frac{C^+}{C^-} \propto \frac{n^+}{n^-}. \quad (14)$$

According to Eqs. (8) and (14), the costs C^+ and C^- should satisfy:

$$\frac{C^+}{C^-} \propto \frac{n^+}{l^+}\Big/\frac{n^-}{l^-}. \quad (15)$$

Based on Eq. (15), the fuzzy membership factors s_{1i} for class compensation can be designed as follows:

$$s_{1i} = \begin{cases} s_{1i}^+ = \dfrac{n^+/l^+}{n^+/l^+ + n^-/l^-}, & y_i = +1 \\ s_{1i}^- = \dfrac{n^-/l^-}{n^+/l^+ + n^-/l^-}, & y_i = -1 \end{cases} \quad (16)$$

3.3. Design of the denoising fuzzy factors

To reduce the influence of noise on the decision hyperplane, FSVM [8] introduced a fuzzy membership s_i based on the distance from each sample to the class center. Using an appropriate kernel function, SVMs can adopt a transformation $\Phi : R^n \to F$ to project each sample from the input space I to a high-dimensional feature space F. The distance from each sample to the class center in I can also be projected to the feature space by the kernel function. In this paper, we improve the classic FSVM by introducing denoising fuzzy factors based on the projected distances in the feature space F.

According to the transformation Φ, samples x_i and x_i are projected as $\Phi(x_i)$ and $\Phi(x_j)$. The projected distance between $\Phi(x_i)$ and $\Phi(x_j)$ is:

$$d_{ij} = ||\Phi(x_i) - \Phi(x_j)|| = \sqrt{k(x_i, x_i) - 2k(x_i, x_j) + k(x_j, x_j)} \quad (17)$$

The projected mean vectors of the minority and majority classes are $\Phi^+ = \frac{1}{l^+}\sum_{i=1}^{l^+}\Phi(x_i)$ and $\Phi^- = \frac{1}{l^-}\sum_{i=1}^{l^-}\Phi(x_i)$, respectively. The distance D_{i+} (D_{i-}) from minority (majority) sample x_i to Φ^+ (Φ^-) can then be calculated as:

$$D_i^+ = \sqrt{k(x_i, x_i) - \frac{2}{l^+}\sum_{j=1}^{l^+}k(x_i, x_j) + \frac{1}{(l^+)^2}\sum_{j=1}^{l^+}\sum_{k=1}^{l^+}k(x_j, x_k)}, \quad (18)$$

$$D_i^- = \sqrt{k(x_i, x_i) - \frac{2}{l^-}\sum_{j=1}^{l^-}k(x_i, x_j) + \frac{1}{(l^-)^2}\sum_{j=1}^{l^-}\sum_{k=1}^{l^-}k(x_j, x_k)} \quad (19)$$

Let R^+ and R^- denote the radii of the minority and majority datasets, respectively. R^+ and R^- satisfy the following equations: $R^+ = \max D_i^+$, $R^- = \max D_i^-$.

According to [18], the denoising fuzzy factor s_{2i} can be designed according to the projected distances D_i^+ and $D_i{}^-$ in feature space:

$$s_{2i} = \begin{cases} s_{2i}^+ = \dfrac{D_i^+}{R^+ + \delta}, & y_i = +1 \\ s_{2i}^- = \dfrac{D_i^-}{R^- + \delta}, & y_i = -1 \end{cases}. \quad (20)$$

According to Eqs. (16) and (20), the final fuzzy membership of IBFSVM is given by:

$$s_i = \begin{cases} s_i^+ = a \cdot s_{1i}^+ + (1-a)\cdot s_{2i}^+, & y_i = +1 \\ s_i^- = a \cdot s_{1i}^+ + (1-a)\cdot s_{2i}^+, & y_i = -1 \end{cases}, \quad (21)$$

where a is an adjustable parameter that balances the class compensation and denoising fuzzy factors.

3.4. Testing IBFSVM

To test the performance of IBFSVM, we used three artificial datasets and six UCI datasets [18]. The artificial datasets (Test1, Test2, and Test3) had ten noisy data points intentionally added. Detailed information about the artificial datasets is listed in Table 1.

Six further datasets were taken from the UCI repository. Detailed information about these datasets is listed in Table 2.

Before testing, we normalized the datasets. Some 60% of the samples were selected at random as training data, and the remaining 40% formed the test data. The RBF kernel $k(x_i, x_j) = \exp(-\gamma \cdot ||x_i - x_j||^2)$ was selected, and the kernel parameter γ and penalty parameter C were determined by a grid search technique. We use $g\_mean = \sqrt{se \times sp}$ to evaluate the imbalanced classification performance, where sp means classification accuracy for positive class and se means classification accuracy for negative class. g_mean can only become large when both se and sp are high. Thus, g_mean not only reflects the overall classification performance, but also the accuracy of the positive sample classification.

The C_SVM, FSVM, and IBFSVM algorithms were used to classify the above datasets. The values of sp, se, and g_mean are listed in Table 3.

From Table 3, we can see that IBFSVM outperforms C-SVM and FSVM for most datasets. If the sample size of the positive and negative classes diverges greatly, such as for Abanole, whose imbalance ratio is 39.55, the classification accuracy of positive samples given by C_SVM and FSVM becomes very poor. For Abanole, se values of C_SVM and FSVM are 5.16% and 5.79% respectively, and g_mean values are 22.53% and 23.91%, respectively. However, IBFSVM clearly improves these results, producing the se value of 78.25% and g_mean of 74.35% for Abanole. With the Banana dataset, which has an imbalance ratio of 5.5, IBFSVM exhibits slightly better performance than both C_SVM and FSVM. With balanced datasets, such as Test1, Ionosphere, and Wdbc, IBFSVM does not produce any clear advantage; indeed, with Wdbc, C_SVM gives a slightly bigger value of g_mean than both FSVM and IBFSVM.

As shown in Eq. (21), the parameter a can be used to adjust the weight between the class compensation factor s_{1i} and the denoising factor s_{2i}. Different values of a will produce different classification performance in IBFSVM. Fig. 3 shows the change in g_mean as a varies from 0.0 to 1.0. From Fig. 3, we can see that the optimal value of a changes with the dataset. If the positive and negative class sizes are significantly different, the choice of a will dramatically influence the classification performance. For example, in Abanole, when $a = 0$, the fuzzy membership function is mainly determined by the denoising factor, this produces the worst performance of IBFSVM, giving a g_mean value of just 58.7%. When $a = 0.9$, where class compensation factor plays a more major role in fuzzy membership, IBFSVM gives the best performance with this dataset, producing a g_mean value of 74.35%. For datasets with little variation in sample size, such as Banana, Pima-Indians, Breast, Ionosphere, and Wdbc, changing the value of a has little impact on the classification performance of IBFSVM. Most of the UCI datasets are better classified when $a > 0.5$, suggesting better recognition accuracy when the class compensation factor plays a more pronounced role. However, for the Pima-Indians and Ionoshpere datasets, the classifier performs

Table 1
Artificial datasets and related properties.

Datasets	Mean of positive class	Standard deviation of positive class	Mean of negative class	Standard deviation of negative class	Sample size of positive class	Sample size of negative class	Imbalance ratio
Test1	[3,3]	[1,0; 0, 1]	[3,3]	[5,0; 0, 5]	205	305	1.487
Test2	[3,3]	[1,0; 0, 1]	[3,3]	[1,0; 0, 1]	205	705	3.439
Test3	[3,3]	[1,0; 0, 1]	[3,3]	[1,0; 0, 1]	205	1005	4.9

Table 2
UCI datasets and related properties.

Datasets	Sample size in positive class	Sample size in negative class	Total number	Imbalance ratio (IR)
Abanole	103	4074	4177	39.55
Banana	200	1100	1300	5.5
Pima-Indians	268	500	768	1.87
Breast	239	444	683	1.85
Ionosphere	126	225	351	1.78
Wdbc	212	357	569	1.68

better when $a < 0.5$, which may be because these two datasets contain more noise.

4. Affective detection based on physiological signals

The fundamental purpose of IBFSVM is to enable the imbalanced classification of affective detection based on physiological signals. We now describe the experimental process of affective detection.

4.1. Experimental schema

To study the physiological signals of emotional states in an experimental environment, we must firstly devise an experimental schema that induces four kinds of basic emotions, that is, happiness, sadness, anger and fear. At beginning we should choose the materials which can elicit the emotion in the lab environment. Audios [19], pictures [20], and videos [21] were once to be used as inspiration materials. And for it has two channels such as auditory and visual channels, videos can produce a better inspirational effect than other materials. Therefore, we use videos to elicit the subjects' emotion.

Then we recruited subjects from the freshmen in Southwestern University in China. The ages of subjects are from 18 to 22, and male and female are 50% respectively. Every subject must sign an informed consent form. And finally we collected physiological signals from 112 people, and the valid data is happiness 96, sadness 87, fear 59, and anger 36. The number of anger sample is minimal, because the material elicited anger often inspires several kinds of complex emotion, so a lot of angry samples are invalid.

Table 3
Comparison of C-SVM, FSVM, and IBFSVM.

Datasets	Algorithm	*se* [%]	*sp* [%]	*g_mean*
Test1	C_SVM	75	83.8	79.04
	FSVM	83.75	89.17	86.41
	IBFSVM ($a = 1.0$)	92.5	87.5	**89.96**
Test2	C_SVM	80	94.4	86.92
	FSVM	82.5	89.28	85.82
	IBFSVM ($a = 0.5$)	80	95	**87.18**
Test3	C_SVM	60	96.5	76.09
	FSVM	76.25	93.75	84.54
	IBFSVM ($a = 0.8$)	88.75	86.5	**87.62**
Abanole	C_SVM	5.16	98.35	22.53
	FSVM	5.79	98.77	23.91
	IBFSVM ($a = 0.9$)	78.25	70.65	**74.35**
Banana	C_SVM	77	93	84.62
	FSVM	80	80.5	80.42
	IBFSVM ($a = 0.8$)	90	94	**91.97**
Pima-Indians	C_SVM	50	94.5	68.74
	FSVM	70.37	75.5	72.89
	IBFSVM ($a = 0.3$)	82.4	74.5	**78.35**
Breast	C_SVM	89.5	91.87	90.67
	FSVM	94.79	100	97.36
	IBFSVM ($a = 0.8$)	97.92	98.87	**98.39**
Ionosphere	C_SVM	78	93.3	85.32
	FSVM	88	93.33	90.63
	IBFSVM ($a = 0.2$)	88	94.4	**91.16**
Wdbc	C_SVM	95.29	97.8	**96.53**
	FSVM	95.29	92.95	94.11
	IBFSVM ($a = 0.8$)	95.29	96.47	95.88

The bold value means the maximum accuracy of classifier.

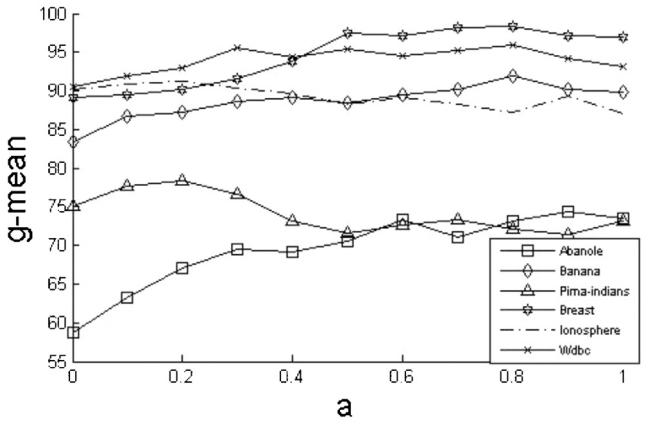

Fig. 3. Change in *g_mean* with respect to *a*.

In the process of collecting physiological signals, we devise a recall experimental paradigm, which can help us to obtain the subjects' self-evaluation for their subjective emotional experience. That means, the movie of one kind of basic emotions will be played twice and only one kind of emotion will be elicited.

When the video is played in the first time, subjects are asked to experience the target emotion carefully, while a BIOPAC MP150 is used to collect their physiological signals synchronously. When the video is played again, subjects are asked to recall their subjective emotional experience in the first presentation, and are asked to press a button when they experience the target emotion. If subjects have experienced the target emotion, they press the button, and the character '1' will be recorded in a self-evaluation file. However if subjects do not taste the target emotion, they will not press the button, and the character '0' will be recorded in this file. In this way, we can collect both physiological signals and the subjects' self-evaluation of their emotional inspiration. Fig. 4 shows the arrangement of the video clips. Fig. 5 displays the self-evaluation files of subjects. In this file, the string '1' means subjects experience the target emotion for a period. And based on the self-evaluation files, we can calculate the start time and end time when subjects taste the target emotion.

4.2. Collection and preprocessing of data

During the first presentation of the movie clip, the BIOPAC MP150 records various physiological signals, including galvanic skin response (GSR), electrocardiograph (ECG), pulse rate (PR) and oxygen saturation of blood (OS). In this study, only the GSR signals were processed and used for subsequent analysis.

The original GSR includes many electrical noises, and we eliminate these noises by bior 5.5 wavelet. By bior 5.5 the GSR will be decomposed into five layers. Using stein unbiased likelihood estimation principle, the threshold in every layer is selected. And then using wdencmp function wavelet denoising is realized. After denoising GSR, we normalize the GSR. Eq. (22) is the formula about normalization. Fig. 6 displays the original GSR and the pretreated GSR.

$$GSR_{normalized} = \frac{GSR - GSR_{\min}}{GSR_{\max} - GSR_{\min}} \tag{22}$$

Most segments of the overall GSR signals are not relevant to the changing of the target emotion. Thus, it is vital to intercept the affective segments from GSR. For this purpose, the subjects' self-evaluation files were employed for statistical analysis, the results of which are presented in Fig. 7. From Fig. 7(a), we can see that most subjects pressed the button at 280 s, 338 s, 378 s, and 398 s in the 'Happiness' video, which implies that the happy emotion was elicited at these points in the movie. Fig. 7(b) suggests that sadness was generally elicited in the latter half of the 'Grief' video. That means, in the time scale from 640 s to 800 s the subjects' sadness is elicited dramatically. As shown in Fig. 7(c), most subjects experienced anger at 290 s, 410 s, and 460–500 s, and Fig. 7(d) indicates that fear was experienced from 320 to 370 s and 410 to 460 s.

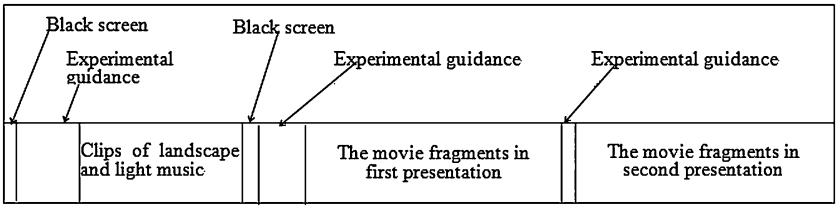

Fig. 4. Arrangement of video clips.

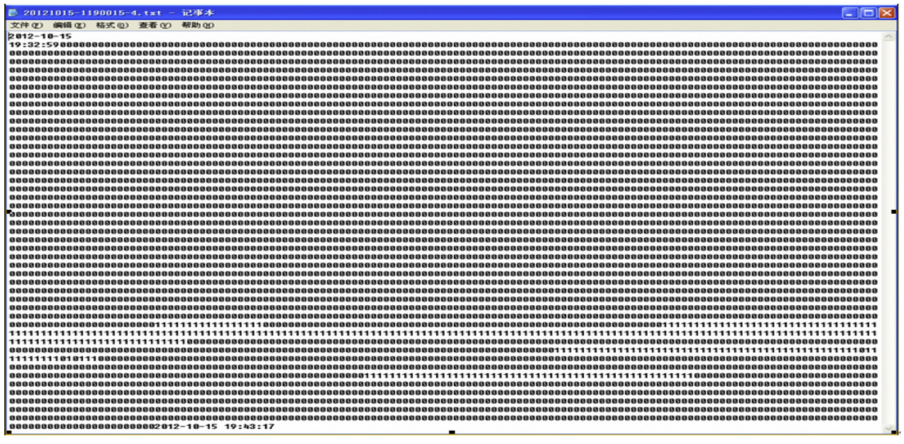

Fig. 5. The self-evaluation file of subjects.

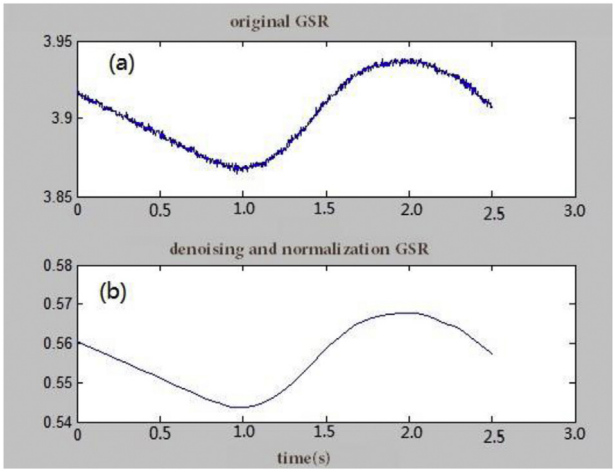

Fig. 6. Preprocessing GSR.

Using this analysis, we can intercept the affective segments of GSR. Thus, starting from 280 s, 640 s, 460 s, and 320 s, respectively, we intercepted 50 s GSR segments to represent happiness, sadness, anger, and fear signals. Finally, we obtained sample sets for the four emotions: 96 happy samples, 87 sad samples, 36 anger samples, and 59 fear samples.

4.3. Extraction of nonlinear features

We extracted 24 nonlinear features of GSR [22] to detect the target emotion. These features include the embedding dimension [23], delay time [23], largest Lyapunov exponent [24], correlation dimension [25], approximate entropy [26], sample entropy [26], and some relative indices based on recurrence quantification analysis (RQA) and the Hurst exponent calculated by multi-fractal detrended fluctuation analysis (MF-DFA). There we will simply introduce the methods of RQA and MF_DFA.

4.4. Recurrence plots and recurrence quantification analysis

Recurrence plots (RP) is put forward by Eckmann [27] in 1987, which can display the status of phase space in higher dimension (more than 3 dimension) by two-dimensional graphics. The mathematical definition of Recurrence plots is as follows:

$$\boldsymbol{R}_{i,j}^{\boldsymbol{m}} := \Theta(\varepsilon - ||x_i - x_j||), \qquad x_i \in \boldsymbol{R}^{\boldsymbol{m}}, \quad i, j = 1, \ldots, N \tag{23}$$

In the above formula, x_i and x_j represent two states in m-dimensional phase space; m says the embedding dimension after the phase space reconstruction; N is the number of system state after the phase space reconstruction; ε is the distance threshold; $||\cdots||$ is the Euclidean norm; and Θ is Heaviside function, that is, $\Theta(x) = \begin{cases} 1, & when \ \ x \geq 0 \\ 0, & when \ \ x \leq 0 \end{cases}$.

In order to analyze recurrence plots more effectively, Zbilut and Webber et al. [28] developed a new methods named as recurrence quantification analysis (RQA), where several parameters reflected the characteristics of RP can be calculated. And in the paper we use RQA to calculate these parameters as follows, and these parameters will be used as features of GSR.

(1) Recurrence rate (RR), the percentage of recursive point in recurrence plots, and

$$\mathrm{RR} = \frac{1}{N^2} \sum_{i,j=1}^{N} R_{i,j} \tag{24}$$

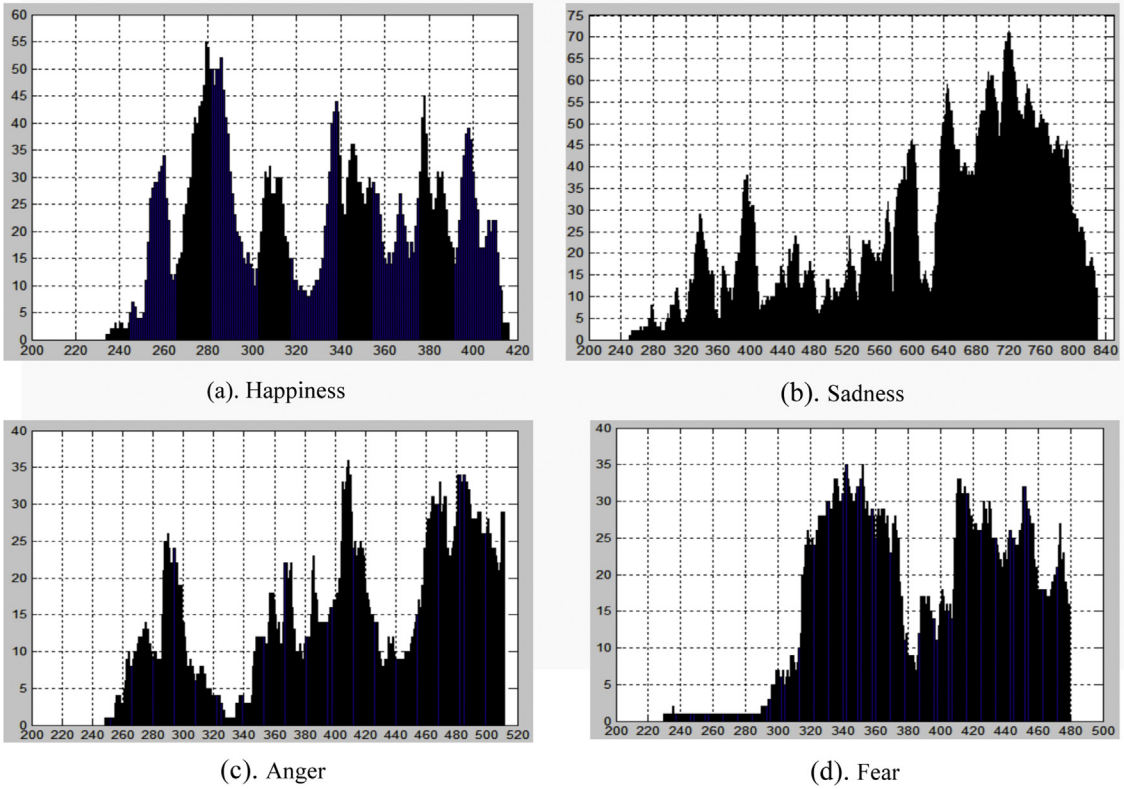

Fig. 7. Distribution of emotional responses based on button presses. Horizontal axis represents the time scale of the video clips, vertical axis denotes number of subjects who have pressed the corresponding button.

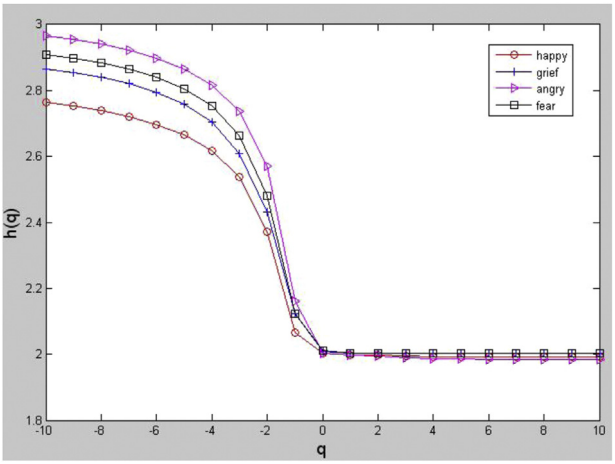

Fig. 8. Relational graph of $h(q)$ and q in affective GSR.

(2) Determinism (DET), the percentage of recursive point in first diagonal (lines paralleled to the main diagonal), and

$$\mathrm{DET} = \frac{\sum_{l=l\min}^{N} lp(l)}{\sum_{i,j}^{N} R_{i,j}} \tag{25}$$

(3) Laminarity (LAM), the percentage of recursive point in vertical or horizontal line, and

$$\mathrm{LAM} = \frac{\sum_{l=l\min}^{N} vp(v)}{\sum_{i,j}^{N} R_{i,j}} \tag{26}$$

(4) The ratio of the DET and RR, that is,

$$\mathrm{RATIO} = N^2 * \frac{\sum_{l=l\min}^{N} lp(l)}{\left(\sum_{i,j}^{N} R_{i,j}\right)^2} \tag{27}$$

4.4.1. Multi-fractal detrended fluctuation analysis

Multi-fractal detrended fluctuation analysis (MF-DFA) is the most important and reliable tool in the research of long-rang correlation and multi-fractal characterization of non-stationary time series. Compared with DFA [29], MF-DFA can not only calculate the scale exponents, but also analyze the multi-fractal spectrum.

The algorithm of MF-DFA is specified in Ref. [30], and at the end of the algorithm MF-DFA will figure out the fluctuation function $F_q(s)$ showed in Eq. (28):

$$F_q(s) = \left\{ \frac{1}{2N_s} \sum_{v=1}^{2N_s} [F^2(s,v)]^{q/2} \right\}^{1/q}, \tag{28}$$

where the time series $\{x_i\}$ will be segmented into $2N_s$ intervals, and the length of every interval is s, q is the order of multi-fractal, $F^2(s,v)$ represent the mean value of each interval removed trend. So $F_q(s)$ is the function of s and q, that is, $F_q(s) = As^{h(q)}$.

In order to discuss the correlation between q and s, we fix q firstly, then draw the bi-logarithmic graph of $\ln[F_q(s)] \sim \ln s$, and the slope of the curve is the Hurst exponent $h(q)$. Based on the Hurst exponent $h(q)$ the relational graph of $h(q)$ and q is drawn, shown in Fig. 8. If $h(q)$ is a constant, the time series $\{x_i\}$ is the single fractal; if $h(q)$ is the function of q, the time series $\{x_i\}$ is the multi-fractal. We can draw the relational graph of $h(q)$ and q in four kinds of basic emotion, shown as Fig. 8. We can find that the four kinds of basic emotion all have multifractal characteristics.

Table 4
Classification performance of C-SVM, FSVM, and IBFSVM for the four emotions.

Target emotion	Classifier	*se* [%]	*sp* [%]	*g_mean*
Happiness	C-SVM	75.4	70.12	72.7
	FSVM	82.14	79.23	80.67
	IBFSVM	89.28	81.25	**85.17**
Sadness	C-SVM	67.43	86.51	76.37
	FSVM	71.52	95	82.43
	IBFSVM	83.33	90	**86.6**
Fear	C-SVM	65.35	82	73.20
	FSVM	71.14	86.5	78.44
	IBFSVM	85.2	89.6	**87.4**
Anger	C-SVM	58.33	80	68.31
	FSVM	69.74	87.5	78.11
	IBFSVM	75.2	88.4	**81.53**

The bold value means the maximum accuracy of classifier.

Based on the Hurst exponent $h(q)$, we can calculate the singular exponent α and multi-fractal spectrum $f(\alpha)$:

$$\alpha = h(q) + q * h(q)' = h(q) + q * \frac{dh(q)}{dq} \tag{29}$$

$$f(\alpha) = q * [\alpha - h(q)] + 1 \tag{30}$$

Based on Eqs. (29) and (30), we will extract the four parameters in the multi-fractal spectrum $f(\alpha)$, where are:

(1) $\alpha_o(f_{\max} = f(\alpha_0))$, $\alpha_0 \in [\alpha_{miin}, \alpha_{\max}]$;
(2) $\Delta\alpha = \alpha_{\max} \times \alpha_{\min}$, which means the width of singular spectrum.
(3) $\Delta f = f(\alpha_{\max}) - f(\alpha_{\min})$, which means the percentage of peaks and troughs in the multi-fractal spectrum $f(\alpha)$.
(4) Symmetrical parameters B of multi-fractal spectrum, which can be got by least-square fitting as formula (29):

$$f(\alpha) = A(\alpha - \alpha_0)^2 + B(\alpha - \alpha_0) + C \tag{31}$$

We use the above four parameters as multifractal features of GSR.

4.5. Detection of target emotion by IBFSVM

After extracting the nonlinear features, the three classifiers were employed to detect the four target emotions. This is a typical multi-classification problem. The "One-Versus-Rest" strategy is applied, which convert a four-classification problem into four binary classification problems. For every binary classifier, the samples of one kind of emotion will be regarded as positive examples, and the samples of the other emotions will be regarded as negative examples. For instance, if happy samples are taken as positive examples, then sadness, anger, and fear samples would be set as negative examples. Clearly, the One-Versus-Rest strategy will produce far fewer positive samples than negative samples, i.e., we have an imbalanced problem.

We employ five-fold cross-validation and an RBF kernel with $\gamma = 2^{-8}$, and set the penalty parameter $C = 100$. The classification performance of C-SVM, FSVM, and IBFSVM is listed in Table 4. We can see that IBFSVM outperformed both C-SVM and FSVM for all four emotions. The best recognition performance from IBFSVM was for the fear emotion, and the worst was for anger, giving *g_mean* values of 87.4% and 81.53%, respectively.

5. Summary

As our experiments considered the four target emotions of happiness, sadness, anger, and fear, the GSR-based affective detection represented a typical multi-classification problem. The strategy

of One-Versus-Rest was employed to resolve this problem, transforming the classification into four binary classifiers. An inherent shortcoming of One-Versus-Rest is the imbalance in the number of positive-class and negative-class samples. The standard SVM and classic FSVM do not deal with imbalanced learning very well, sacrificing the classification accuracy of the positive class to accomplish good overall classification accuracy. Therefore, we proposed IBFSVM, which deals with imbalanced classification very well. We introduced a denoising fuzzy factor to reduce the effects of noise, and also considered a class compensation mechanism to reduce the effects of an imbalanced number and distribution of samples in each class. Using three artificial datasets and six UCI datasets, we verified the effectiveness of IBFSVM. When employed to classify affective datasets extracted from GSR signals, IBFSVM outperformed both C-SVM and FSVM. The proposed method attained a maximum *g-mean* for the fear emotion, 87.4%, whereas the lowest *g-mean* value of 81.53% was for anger. Happiness and sadness gave *g_mean* values of 85.17% and 86.6%, respectively.

In future work, we intend to: (1) apply IBFSVM to oversampling and undersampling to further enhance the accuracy of imbalanced learning; (2) use other physiological signals, such as electrocardiograms and heart rate, to detect emotion; (3) collect more physiological signals and enlarge the sample size.

Acknowledgments

This paper is supported by Key Grant Project of Chinese Ministry of Education (No. 311032), and the Fundamental Research Funds for the Central Universities (No. XDJK2015C023).

References

[1] O. AlZoubi, S.K. D'Mello, R.A. Calvo, Detecting naturalistic expressions of nonbasic affect using physiological signals, IEEE Trans. Affect. Comput. 3 (3) (2012) 298–310.
[2] G. Valenza, A. Lanata, E.P. Scilingo, The role of nonlinear dynamics in affective valence and arousal recognition, IEEE Trans. Affect. Comput. 3 (2) (2011) 237–249.
[3] J. Kim, E. Andre, Emotion recognition based on physiological changes in music listening, IEEE Trans. Pattern Anal. Mach. Intell. 30 (12) (2008) 2067–2083.
[4] G.E. Sakr, I.H. Elhajj, H.A.-S. Huijer, Support vector machines to define and detect agitation transition, IEEE Trans. Affect. Comput. 1 (2) (2010) 98–108.
[5] G.E. Sakr, I.H. Elhajj, U.C. Wejinya, Multi level SVM for subject independent agitation detection, in: Proceedings in IEEE/ASME International Conference on Advanced Intelligent Mechatronics, 2009, pp. 538–543.
[6] R. Barandela, J.S. Sanchez, V. Garcia, E. Rangel, Strategies for learning in class imbalance problems, Pattern Recogn. 36 (3) (2003) 849–851.
[7] Y. Lin, Y. Lee, G. Wahba, Support vector machines for classification in nonstandard situations, Mach. Learn. 46 (1–3) (2002) 191–202.
[8] C.F. Lin, S.D. Wang, Fuzzy support vector machines, IEEE Trans. Neural Netw. 13 (2) (2002) 464–471.
[9] H. Haibo, E.A. Garcia, Learning from imbalanced data, IEEE Trans. Knowl. Data Eng. 21 (9) (2009) 1263–1284.
[10] N. Chawla, K. Bowyer, P. Kegelmeyer, SMOTE: synthetic minority oversampling technique, J. Artif. Intell. Res. 16 (2002) 321–357.
[11] R. Barandela, J.S. Sanchez, V. Garcia, et al., Strategies for learning in class imbalance problems, Pattern Recogn. 36 (3) (2003) 849–851.
[12] K. Veropoulos, C. Campbell, N. Cristianini, Controlling the sensitivity of support vector machines, in: Proceedings of the International Joint Conference on AI, 1999, pp. 55–60.
[13] T. Imam, K. Ting, J. Kamruzzaman, z-SVM: an SVM for improved classification of imbalanced data, Lecture Notes Comput. Sci. 4304 (2006) 264–273.
[14] G. Wu, E.Y. Chang, KBA: kernel boundary alignment considering imbalanced data distribution, IEEE Trans. Knowl. Data Eng. 17 (6) (2005) 786–795.
[15] R. Batuwita, V. Palade, FSVM-CIL: fuzzy support vector machines for class imbalance learning, IEEE Trans. Fuzzy Syst. 18 (3) (2010) 558–571.
[16] M.A. YD, Z. Du, S.Y. Liu, A new noise-immune fuzzy SVM algorithm for unbalanced data, J. Xi'an Technol. Univ. 28 (3) (2008) 297–300.
[17] C.D. Qin, S.Y. Liu, S.F. Zhang, Balanced fuzzy support vector machines based on imbalanced data set, Comput. Sci. 39 (6) (2012) 188–190.
[18] http://archive.ics.uci.edu/ml/
[19] E. Eich, T.W. Joycelin, D. Macaulay, A.D. Percy, et al., Combining music with thought to change mood, in: Handbook of Emotion Elicitation and Assessment, Series in Affective Science, 2007, pp. 124–136.
[20] P.C. Petrantonakis, L.J. Hadjileontiadis, Emotion recognition from brain signals using hybrid adaptive filtering and higher order crossings analysis, IEEE Trans. Affect. Comput. 1 (2) (2010) 81–97.
[21] J.J. Gross, R.W. Levenson, Emotion elicitation using films, Cognit. Emot. 9 (1) (1995) 87–108.
[22] J. Cheng, G.L. Liu, X.W. Lai, Calculation of nonlinear features of SC for emotion recognition, J. Comput. Inf. Syst. 10 (6) (2014) 2331–2339.
[23] Z.B. Lu, Z.M. Cai, K.Y. Jiang, Determination of embedding parameters for phase space reconstruction based on improved C–C method, J. Syst. Simul. 19 (11) (2007) 2527–2538.
[24] A. Wolf, J.B. Switf, H.L. Swinney, et al., Determining Lyapunov exponents from a time series, Physica D 16 (3) (1985) 285–317.
[25] P. Grassbegrer, I. Procaccia, Characterization of strange attractors, Phys. Rev. Lett. 50 (5) (1983) 346–349.
[26] J.S. Richman, J.R. Moorman, Physiological time series analysis using approximate entropy and sample entropy, Am. J. Physiol. Heart Circ. Physiol. 278 (6) (2000) 2039–2049.
[27] J.P. Eckmmn, S.O. Kamphorst, D. Ruelle, Recurrence plots of dynamical systems, Europhys. Lett. (1987) 441–445.
[28] J.P. Zbilut, C.L. Webber, Embeddings and delays as derived from quantification of recurrence plots, Phys. Lett. A 171 (3–4) (1992) 199–203.
[29] C.K. Peng, S.V. Buldyrev, S. Havlin, et al., Mosaic organization of DNA nucleotides, Phys. Rev. E 49 (2) (1994) 1685–1689.
[30] J.W. Kantelhardt, S.A. Zschiegner, E. Koscielny-Bunde, et al., Multifractal detrended fluctuation analysis of nonstationary time series, Physica A 316 (1–4) (2002) 87–114.