



**HAL**  
open science

# Entropy-Assisted Emotion Recognition of Valence and Arousal Using XGBoost Classifier

Sheng-Hui Wang, Huai-Ting Li, En-Jui Chang, An-Yeu (andy) Wu

► **To cite this version:**

Sheng-Hui Wang, Huai-Ting Li, En-Jui Chang, An-Yeu (andy) Wu. Entropy-Assisted Emotion Recognition of Valence and Arousal Using XGBoost Classifier. 14th IFIP International Conference on Artificial Intelligence Applications and Innovations (AIAI), May 2018, Rhodes, Greece. pp.249-260, 10.1007/978-3-319-92007-8\_22 . hal-01821082

**HAL Id: hal-01821082**

**<https://inria.hal.science/hal-01821082v1>**

Submitted on 22 Jun 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Entropy-Assisted Emotion Recognition of Valence and Arousal Using XGBoost Classifier

Sheng-Hui Wang<sup>(✉)</sup>, Huai-Ting Li, En-Jui Chang, and An-Yeu (Andy) Wu

Graduate Institute of Electronics Engineering, National Taiwan University,  
Taipei, Taiwan  
{harry, wesli, enjui, andywu} @access.ee.ntu.edu.tw

**Abstract.** Emotion recognition is an essential function to realize human-machine interaction devices. Physiological signals which can be collected easily and continuously by wearable sensors are good inputs for emotion analysis. How to effectively process physiological signals, extract critical features, and choose machine learning model for emotion classification has been a big challenge. In this paper, an entropy-based processing scheme for emotion recognition framework is proposed, which includes entropy domain feature extraction and prediction by XGBoost classifier. We experiment on AMIGOS database and the experimental results show that the proposed scheme for multi-modal analysis outperforms conventional processing approaches. It achieves approximately 80% and 68% accuracy of prediction for two affect dimensions, valence and arousal. For one modality case, we found that galvanic skin response (GSR) channel is the most potential modality for prediction, which leads to best performances.

**Keywords:** Affective computing · Emotion recognition · Physiological signal processing · Entropy domain features · XGBoost

## 1 Introduction

Affective computing is a key technology for human–computer interaction (HCI) in the future era of Internet of thing (IoT) [1], which makes it possible for machines and computers to realize human’s emotion and mentality in real time. Furthermore, it can give appropriate responses and services based-on human’s current mental status. Related databases have been developed for related researches on affective computing. Among the open-source databases available on the Internet, FEEDTUM [2] and Berlin Database of Emotional Speech (EmoDB) [3], released in 2005, used facial video and speech for emotion recognition, respectively. They predicted basic emotion of people, such as anger and happy, by single type of signal.

As intelligent IoT develops, more and more wearable devices are equipped with different kinds of sensors. We can effectively get various signals sensed from subjects and have the access for continuously monitoring. For this scenario,

physiological signals can be good inputs for affective computing framework, since we do not express our emotions on faces and make sounds every time. Integrating multiple-channel of physiological signals for emotion recognition and aggregating machine learning classifiers for ensemble learning have gain attention in recent years. DEAP database [4] proposed in 2012 has been experimented in many works [5]-[6], it contains several types of physiological signals which were recorded when subjects were watching music videos. In DEAP, emotions perceived by subjects are self-assessed on two affect dimensions, valence and arousal. Valence tells how positive or negative the emotion is and arousal indicates the intensity of emotion. In 2016, ACERTAIN database [7] was proposed that first used commercial wearable sensor to collect data and added personality into experiments. AMIGOS database [8] proposed in 2017 contains stable data, several types of videos, and elements for other related research, such as social context and mood. These all indicate that multi-modality analysis of physiological signals is a trend of affective computing. Therefore, how to effectively process different physiological signals and further extract critical features from different channels for emotion classification have been a big challenge.

The aim of this paper is to establish reliable emotion recognition framework based on physiological signals. To improve the accuracy of emotion recognition, we firstly extract the entropy domain features to quantify the regularity and randomness of signal, which has high potential to represent the different levels of emotion. Next, we apply the XGBoost [9] to enhance the performance of classification which has high scalability and efficiency for training and has the ability to learn from high dimension data without using high complexity feature selection algorithms.

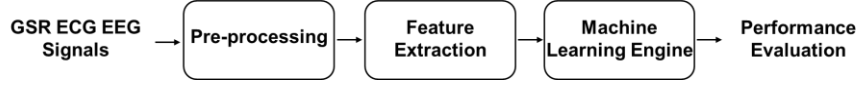
The rest of the paper is organized as follows. Section 2 describes common methods for emotion recognition task. Section 3 illustrates the enhanced methods for feature extraction and machine learning engine. Section 4 shows the experimental results and Section 5 concludes.

## **2 Common Approaches of Emotion Recognition Framework**

The flow chart of the emotion recognition framework in this study is shown in Fig. 1. Three kinds of physiological signals are inputs and the output is the prediction for high or low of affect dimensions. Three other blocks including pre-processing, feature extraction, and machine learning engine will be introduced as below.

### **2.1 Pre-Processing**

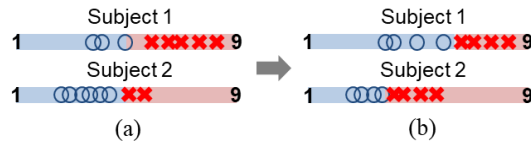
We use the processed data provided by the official website of AMIGOS dataset. In AMIGOS database, each subject watched both long and short videos as stimulus for emotion. Because long-length video data contain multiple values for valence



**Fig. 1.** Flow chart of general emotion recognition framework.

and arousal, only short-length video data are used in this work to reduce the uncertainty. There are total 40 subjects and 16 short-length videos in database. Each video of each subject is seen as one data. Each of data has 14 channels for Electroencephalography (EEG), 2 channels for Electrocardiography (ECG), and 1 channel for galvanic skin response (GSR). 7 subjects whose physiological signals of videos contain some missing values are removed (ID number: 9, 12, 21, 22, 23, 24, 33). Therefore, there are totally  $(40-7)$  (subjects)  $\times$  16 (videos) = 528 data. We further checked the data stability. First Channel of ECG signals suffer from fewer noises compared to second channel and are used for further analyzing. In addition, GSR signals are filtered by the low-pass filter with cut-off frequency 3Hz to remove abnormal high frequency noises.

In this work, we do binary classification for valence and arousal. Self-assessments of each subject on two affect dimensions provided in AMIGOS database is used for labels. They are originally values between [1, 9] and needed to be transformed to either positive class or negative class label, which represents high and low, respectively. Fig. 2 shows two different ways to process. Directly cut by threshold 5 is intuitive, but hard to distinguish one's relative high and low emotion. In this way, values may highly accord to personal tendency to rate high or low scores, which results in imbalance of labels and decrease in performances. Our method is to define labels by subject-dependent mean value among all 16 videos of each person. Using mean value instead of median value can avoid ambiguity whether the median is labeled positive class or negative class.



**Fig. 2.** (a) Define labels by comparing with threshold score 5. (b) Define labels by comparing with subject-dependent mean among all videos.

## 2.2 Feature Extraction

Feature extraction is implemented on all three kinds of signals with different ways to reduce the input dimension before using machine learning methods. Features illustrated in this Section are slightly different from the feature set in AMIGOS database. Features extracted among physiological signals are shown in Table 1.

**Table 1.** Features extracted among physiological signals

Modality	Features Extracted
GSR signal (32 features)	<p><b>SR features (18):</b> mean, mean of derivative, mean of negative derivative, proportion of negatives in derivative, number of local minimum, mean of each rise time series, spectral power 0-2.4 Hz (12, per 0.2Hz).</p> <p><b>SC features (4):</b> mean, standard deviation, mean of derivative, mean of second derivative.</p> <p><b>SCSR features (4):</b> mean, standard deviation, mean of derivative, mean of second derivative.</p> <p><b>SCSR &amp; SCVSR features (6):</b> zero crossing rate of SCSR, zero crossing rate of SCVSR, average magnitude of SCSR occurrences, average magnitude of SCVSR occurrences, average magnitude of all SCSR and SCVSR occurrences, ratio of number of SCSR occurrences to SCVSR.</p>
ECG signal (77 features)	<p><b>Features from raw ECG (60):</b> spectral power 0-6 Hz (per 0.1Hz)</p> <p><b>HRV features (10):</b> mean, standard deviation, skewness, kurtosis, root mean square, root mean square of the successive differences, percentage of points higher than mean + standard deviation, percentage of points lower than mean - standard deviation, low frequency spectral power (0.01-0.08 Hz), middle frequency spectral power (0.08-0.15 Hz), high frequency spectral power (0.15-0.50 Hz).</p> <p><b>HR features (6):</b> mean, standard deviation, skewness, kurtosis, percentage of points higher than mean + standard deviation, percentage of points lower than mean - standard deviation</p>
EEG signal (105 features)	<p><b>PSD features (70):</b> average 5 bands (theta, slow alpha, alpha, beta, and gamma) PSD in 14 channels.</p> <p><b>Asymmetry features (35):</b> difference of 5 bands PSD between 7 pairs of channel.</p>

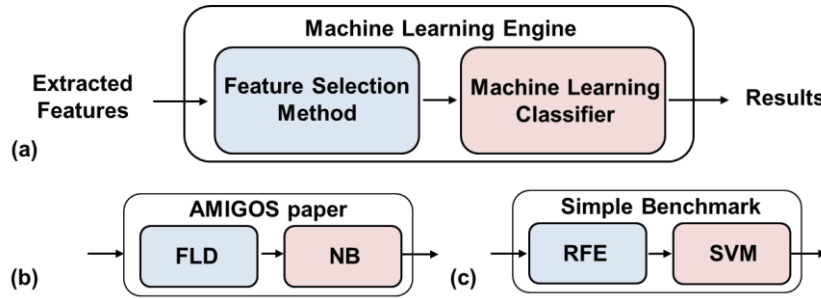
**GSR signal:** Partly follow methods in [8] and [10], skin response (SR) related features (18) are first extracted and skin conductance (SC) signal is got by computing SR signal's reciprocal. SC signal is then normalized and used to extract features (4). Next, skin conductance slow response (SCSR) and skin conductance very slow response (SCVSR) is got by low-pass filtering normalized SC signal at 0.2 and 0.08 Hz, respectively. Afterwards, some SCSR related features are extracted (4). Last, after de-trending SCSR and SCVSR by using empirical mode decomposition (EMD) methods [11] to remove the last half intrinsic mode functions (IMFs), some related features are extracted.

**ECG signal:** Partly follow methods in [8] and [10], 0-6 Hz spectral power features are first extracted (60, per 0.1). Then, R-R interval (RRI) series is calculated after detecting R peaks in ECG. Heart rate variability (HRV) and heart rate (HR) time series can be computed by using RRI. Finally, HRV related features (11) and HR related features (6) are extracted.

**EEG signal:** Follow the methods in [4], average power spectral density (PSD) of theta band (4-7 Hz), slow alpha band (8-10 Hz), alpha band (8-13 Hz), beta band (14-30 Hz), gamma band (31-47 Hz) of each EEG channel are extracted ( $5 \times 14 = 70$ ). Also, asymmetry of PSD of 5 bands between 7 pairs of EEG channel are extracted ( $5 \times 7 = 35$ ).

### 2.3 Machine Learning Engine

This block consists of a feature selection method and a machine learning classifier, as shown in Fig. 3a. Feature space which contains 214 features is comparatively big while we only have 528 (16 subjects \* 33 videos) data in AMIGOS database. Thus, feature selection is applied to eliminate redundant features and reduce the model complexity of machine learning algorithm in next stage. This is helpful for enhancing overall performance and sparing lots of computation resources.



**Fig. 3. (a) Overview of machine learning engine block. (b) Machine learning engine block in AMIGOS paper. (c) Machine learning engine block of simple benchmark.**

In Fig. 3b, overview of machine engine block in AMIGOS database is shown. For feature selection method in original paper of Amigos database [8], Fisher's linear discriminant (FLD)  $J$  [12] is calculated over each feature, which is defined as

$$J(f) = \frac{|\mu_1 - \mu_0|}{\sigma_1^2 + \sigma_0^2}. \quad (1)$$

Afterward, user can decide how much discriminant features they would like to select based on validation sets and pick up features with the highest  $J$  values. For machine learning classifier, Gaussian Naïve Bayes (GaussianNB) classifier is used. Assuming features are independent and Gaussian distributed, GaussianNB is given as

$$G(f_1, \dots, f_n) = \operatorname{argmax}_c p(C = c) \prod_{i=1}^n p(F_i = f_i / C = c), \quad (2)$$

and calculate the probability of each sample for each class, where  $F$  is feature set and  $C$  is the classes set. These two methods are not powerful and lead to relatively low f1-score shown in [8].

In this work, we first implement support vector machine (SVM) [13], [14] based approach as the benchmark, as shown in Fig. 3c. SVM is a famous tool in the field of machine leaning and it has good performance on classification and prediction over many applications. Compared with other classifier, SVM can deal with small-size datasets well, since it uses only support vectors to construct hyperplane. When it comes to feature selection, recursive feature elimination (RFE) algorithm [15] can efficiently remove irrelevant features and be suit for SVM. RFE-SVM eliminates features with smallest weight in SVM model on a sequential backward selection (SBS) based process. The progress continues until one feature is left and the features combined with SVM model with the highest performance is outputted.

### 3 Proposed Framework with Entropy Domain Features and XGBoost Classifier

#### 3.1 Entropy Domain Features

Non-linear entropy domain features, such as sample entropy, permutation entropy  $\tau$  based features, are widely used on physiological signals. Extracted entropy value can help quantify the regularity of signal and thus be applied on medical diagnosis. For this emotion recognition work, three types of entropy domain features, including refined composite multiscale entropy (RCMSE), turning point ratio (TRP), and Shannon entropy, are applied to measure the complexity of physiological signals. The details are shown as follows.

**Refine Composite Multiscale Entropy (RCMSE) [16]:** Multiscale entropy (MSE) [17] has been used widely to evaluate physiological control mechanisms, such as atrial fibrillation and Alzheimer’s disease [18]. RCMSE is an improved version of MSE. It reduces the possibility of undefined value problem when the signal length is short and has a better accuracy for entropy estimation. The concept of RCMSE is using different scales of local matching pattern to compute the regularity of the signal. There are two steps of RCMSE. First step: time series of signal is coarse-grained into multiscale series. For each scale factor  $\tau$ ,  $\tau$  series are generated by average  $\tau$  points in non-overlapping windows and each of series overlap  $\tau-1$  points with neighbor series. The  $j$ -th point of  $k$ -th coarse-

grained series,  $\mathbf{y}_k^{(\tau)} = \{y_{k,1}^{(\tau)} y_{k,2}^{(\tau)} y_{k,3}^{(\tau)} \dots y_{k,p}^{(\tau)}\}$ , with scale factor  $\tau$  of signal  $x$  is defined:

$$\mathbf{y}_{k,j}^{(\tau)} = \frac{1}{\tau} \sum_{i=(j-1)\tau+k}^{j\tau+k-1} x_i, \quad 1 \leq j \leq \frac{N}{\tau}, \quad 1 \leq k \leq \tau. \quad (3)$$

In the conventional MSE algorithm, output of coarse graining is only first series of each scale factor. Second step: averaged sample entropy of each  $\tau$  is calculated as below:

$$RCMSE(x, \tau, m, r) = -\ln\left(\frac{\sum_{k=1}^{\tau} n_{k,\tau}^{m+1}}{\sum_{k=1}^{\tau} n_{k,\tau}^m}\right), \quad (4)$$

where  $m$  is the matching pattern length and  $r$  is the similarity criterion. In (4),  $n_{k,\tau}^m$  is the number of two sets of simultaneous data points of length  $m$  in  $k$ th series have the difference  $< r$ . When the ratio of  $\sum_{k=1}^{\tau} n_{k,\tau}^{m+1}$  and  $\sum_{k=1}^{\tau} n_{k,\tau}^m$  is small, sample entropy would be large, which means high complexity of signal. Value of RCMSE is undefined only when  $\sum_{k=1}^{\tau} n_{k,\tau}^{m+1} = 0$ , which means  $n_{k,\tau}^m$  of  $k$  series are all zero. By summation of matching pattern of multiple series, RCMSE has less probability to be undefined than conventional MSE.

In this work, RCMSE is applied to RRI series of ECG signals. Because the length of ECG signals are relatively short (each video 55s-155s), the scale factor  $\tau$  is set up to be 3. In addition, we set the matching pattern  $m$  up to be three and the similarity criterion  $r$  is set to be 0.2 of standard deviation.

**Turning Points Ratio (TPR):** TPR is proposed on the basis of nonparametric ‘‘Runs Test’’ to evaluate the randomness in a time-series [19] and the idea was used in RRI of ECG signal [20]. The concept of TPR is to measure the complexity of the signal by number of turning points compared to total points. Turning point is found by comparing each point with left and right neighbor points and TPR is calculated as follows:

$$TPR = \frac{\sum_{i=2}^{N-1} [(x_i - x_{i-1})(x_i - x_{i+1}) > 0]}{N - 2}, \quad (5)$$

where  $N$  is the length of the signal  $x$ . Besides original TPR, we extracted modified TPR (MTPR) on signals. The procedure of MTPR is to use EMD methods to extract trend of the signal first, in order to remove trivial peaks of the signal. Next, compute TPR as (5) on the extracted trend.

In this work, TPR and MTPR are calculated on RRI of ECG signals and GSR signals. EEG signal has less information on time series and is skipped.

**Shannon Entropy:** Shannon entropy is commonly calculated in the domain of information theory. It is defined for a given discrete probability distribution, using probability of each symbol to measure the uncertainty or randomness of the data.



However, almost every point has different values. We can't see each of different points as a new symbol, otherwise almost the same value of Shannon entropy we would get from different signals. In other way, we classify every points into one group of the group set. First, outliers that have larger differences than three standard deviations with mean are removed. Second, we sort rest data points into N groups, which are equally divided between max and min value. Last,  $p_i$  of each  $i$ th group is calculated and Shannon entropy we can get by:

$$\text{Shannon Entropy} = -\sum_{i=1}^N p_i \log(p_i). \quad (6)$$

The concept of Shannon entropy is to observe the complexity of the signal by overall distribution. Each group is seen as a symbol.

In this work, Shannon entropy is calculated on RRI of ECG signals and GSR signals. Since the optimal group number is data dependent, we apply total 4, 8, 16, 32, 64 groups for simulation.

For these three enhanced feature extraction methods, total 26 features are added into original feature space. The details are listed in Table 2.

**Table 2.** Enhanced features extracted among physiological signals

Modality	Features Extracted
GSR signal (7 features)	<b>TPR (2):</b> TPR, MTPR. <b>Shannon Entropy (5):</b> number of groups: 4, 8, 16, 32, 64
ECG signal (19 features)	<b>RCMSE (12):</b> m = 0: $\tau = \{1, 2, 3\}$ , sum of $\tau = \{1, 2, 3\}$ m = 1: $\tau = \{1, 2, 3\}$ , sum of $\tau = \{1, 2, 3\}$ m = 2: $\tau = \{1, 2, 3\}$ , sum of $\tau = \{1, 2, 3\}$ <b>TPR (2):</b> TPR, MTPR. <b>Shannon Entropy (5):</b> number of groups: 4, 8, 16, 32, 64

### 3.2 Extreme Gradient Boosting (XGBoost)

In this part, we would like to change the content of the machine learning blocks mentioned before in Section 2.3, where RFE based SVM approach was used. Among all the machine learning algorithms, gradient boosting tree based model [21] has shown in many applications in different domains. XGBoost [9] is an efficient and scalable gradient boosting machine, which has won lots of machine competitions in recent years [22], [23]. It is an ensemble model consisting of sets of classification and regression tree (CART). While XGB is used for supervised learning problems and we use training data  $x_i$  to predict a target variable  $y_i$ , the model can be described in the form:

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), f_k \in F, \quad (7)$$

where  $K$  is the total number of trees,  $f_k$  for  $k$ th tree is a function in the functional space  $F$ , and  $F$  is the set of all possible CARTs. In the training, each of new-trained CART will try to complement the so-far residual. Objective function been optimized at  $(t+1)$ th CART is described:

$$obj = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t)}) + \sum_{i=1}^t \Omega(f_i), \quad (8)$$

Where  $l(\cdot)$  denotes the training loss function,  $y_i$  the is ground truth, and  $\hat{y}_i^{(t)}$  is the prediction value at step  $t$ .  $\Omega(\cdot)$  given by

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \quad (9)$$

is the regularization term, where  $T$  are the number of leaves and  $w_j$  is the score on  $j$ th leaf. When (9) is optimized, Taylor's expansion is used so that gradient descent can be used for different loss functions. Furthermore, feature selection is no need when we use XGBoost approach. During training period of XGBoost, good features would be chosen as node in trees, which means features not used are abandoned.

In this work, we use the scikit-learn API for XGBoost classification. The inputs of XGBoost are total 240 features (214 traditional features + 26 entropy domain feature) and the outputs are prediction results for valence or arousal. For loss function in (8), logistic loss function is set. The details of used features and performances are shown in next Section.

## 4 Experimental Settings and Results

Single trial classification for two affect dimensions, arousal and valence, is experimented and the flow is shown in Fig. 1. After signal is pre-processed and transformed into features, all features are normalized to [1 -1]. Leave-one-subject-one approach is conducted to evaluate performance. That is, every time one of the 33 subjects is leaved as test set and machine learning engine is trained using remaining 32 subjects' features. When the procedure is repeated 33 times, final performance is calculated by averaging 33 values in each procedure.

There are 3 processing schemes that are compared.

- Scheme 1: Using the feature set and the method described in original AMIGOS database [8], where FLD and GaussianNB were used for machine learning engine block.

- Scheme\_2: Using the feature set illustrated in Section 2.2 and RFE-SVM for machine learning engine block.
- Scheme\_3: Using new feature set with entropy domain features and XGBoost for machine learning engine block.

Table 3 shows the comparison of f1-score and accuracy over three schemes, F1-score is the harmonic mean of precision and recall and here we average f1-scores of positive and negative class as final value. Four scenarios were experimented: using only GSR, ECG, EEG, and using all modalities. Table 4 shows the features used in trees of XGBoost classifier for classification of valence and arousal.

**Table 3.** Performance of scheme\_1- scheme\_3 on emotion recognition framework. (F1-score is mean for positive and negative class. Red value indicates the highest accuracy or f1-score in each affect dimension. Scheme\_1: Amigos feature set + FLD + GaussianNB; Scheme\_2: commom feature set + RFE-SVM; Scheme\_3: new feature set + XGBoost)

		Accuracy				F1-score			
		GSR	ECG	EEG	All	GSR	ECG	EEG	All
V	Scheme 1	-	-	-	-	0.531	0.535	0.576	0.570
	Scheme 2	0.640	0.615	0.589	0.680	0.610	0.588	0.556	0.666
	Scheme 3	0.776	0.634	0.581	<b>0.801</b>	0.774	0.630	0.577	<b>0.800</b>
A	Scheme 1	-	-	-	-	0.548	0.550	0.592	0.585
	Scheme 2	0.644	0.623	0.566	0.663	0.643	0.612	0.557	0.660
	Scheme 3	0.682	0.542	0.579	<b>0.684</b>	<b>0.707</b>	0.566	0.604	0.698

**Table 4.** Features selected in XGBoost classifier for classification of two affect dimensions. (Entropy domain features are marked in red color)

Dimension	Features
Valence	<b>GSR:</b> spectral power 0-0.2 0.4-0.6 0.6-0.8 0.8-1.0 1.4-1.6, SC_mean of derivative, SCSR_mean, <b>SR_MTPR</b> <b>ECG:</b> spectral power 1.2-1.3 2.5-2.6, HRV_root mean square, HRV_low frequency spectral power, HRV_high frequency spectral power, HRV_percentage of points lower than mean - standard deviation, <b>TPR</b> , <b>Shannon_group4</b> , <b>RCMSE_m=2 sum</b> <b>EEG:</b> T8_gamma, O2_gamma, T7_gamma, O2_beta, T7-T8_slow alpha, T7-T8_beta, T7-T8_theta, O1-O2_beta
Arousal	<b>GSR:</b> spectral power 0-0.2 1.6-1.8, SC_mean of derivative, SC_mean of second derivative, SCSR_number of local minimum <b>ECG:</b> spectral power 1.8-1.9 1.9-2.0 3.0-3.1, HRV_percentage of points lower than mean - standard deviation, HRV_low frequency spectral power, <b>RCMSE_m=1</b> , <b><math>\tau = 1</math></b> <b>EEG:</b> FC6_alpha

F1-score for Scheme\_1 is directly obtained from [8]. For dimension valence and arousal, best accuracy is about 80% and 68%, respectively. Scheme\_3 outperforms Scheme\_1 and Scheme\_2 in almost every scenario on both affect

dimensions, especially in valence dimension, outperforms by more than 10% of accuracy. Using all-modalities for prediction can have highest performances except for f1-score in arousal. In our experiments (Scheme\_2 and Scheme\_3), using GSR channel to predict can have the best performance for both valence and arousal dimensions, when only one channel is used. However, the features extracted in GSR channel is the least among three modalities, which indicates that the quality but not the quantity of features lead to better prediction ability.

## 5 Conclusion

In this paper, we proposed the scheme including both entropy domain features and XGBoost. We enhance feature extraction methods which are entropy domain and helpful for evaluating the complexity of physiological signals. On the other hand, XGBoost classifier which gains popularity in recent years is used for learning and prediction. The proposed scheme can reach the performance of approximately 80% and 68% accuracy on valence and arousal dimension, respectively. It outperforms the processing scheme in original paper of AMIGOS database and the scheme that contains common features and traditional SVM model.

**Acknowledgements.** This work was supported by the Ministry of Science and Technology of Taiwan (MOST 106-2221-E-002-205-MY3 and MOST 106-2622-8-002-013-TA), National Taiwan University and Pixart Imaging Inc.

## References

1. R. Picard, "Affective Computing for HCI," in International Conference on Human-Computer Interaction Munich Germany Lawrence Erlbaum Associates Inc. pp. 829 - 833. (1999)
2. F. Wallhoff, Facial expressions and emotion database, available at: [http://www.mmk.ei.tum.de/\\_waf/fgnet/feedtum.html](http://www.mmk.ei.tum.de/_waf/fgnet/feedtum.html) (2005)
3. F. Burkhardt, A. Paeschke, M. Rolfes, W. Sendlmeier, B. Weiss, "A Database of German Emotional Speech", in Proceedings of Interspeech, pp 1517-1520 (2005)
4. S. Koelstra et al., "DEAP: A Database for Emotion Analysis; Using Physiological Signals," in *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 18-31, Jan.-March (2012)
5. M. Chen, J. Han, L. Guo, J. Wang and I. Patras, "Identifying valence and arousal levels via connectivity between EEG channels," *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*, Xi'an, pp. 63-69 (2015).
6. S. Wu, X. Xu, L. Shu and B. Hu, "Estimation of valence of emotion using two frontal EEG channels," *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Kansas City, MO, pp. 1127-1130 (2017)

7. R. Subramanian; J. Wache; M. Abadi; R. Vieriu; S. Winkler; N. Sebe, "ASCERTAIN: Emotion and Personality Recognition using Commercial Sensors," in IEEE Transactions on Affective Computing, vol. PP, no.99, pp.1-1 (2016)
8. J. Abdon Miranda-Correa, M. Khomami Abadi, N. Sebe, and I. Patras, "AMIGOS: A Dataset for Mood, Personality and Affect Research on Individuals and Groups," ArXiv e-prints (2017)
9. T. Chen and C. Guestrin, "XGBoost," Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. - KDD '16, pp. 785–794 (2016)
10. J. Kim and E. Andr, "Emotion recognition based on physiological changes in music listening," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 30, no. 12, pp. 2067–2083 (2008)
11. N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis," in Proceedings of the Royal Society of London A: mathematical, physical and engineering sciences, vol. 454, no. 1971. The Royal Society, pp. 903–995 (1998)
12. F. Song, D. Mei, and H. L., "Feature Selection Based on Linear Discriminant Analysis," in Intelligent System Design and Engineering Application (ISDEA), 2010 Int'l. Conf., vol. 1, pp. 746–749 (2010)
13. Cristianini N, Shawe-Taylor J. *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge, UK: Cambridge University Press (2000)
14. Luts J, Ojeda F, van de Plas Raf R, de Moor B, van Huffel S, Suykens JAK. A tutorial on support vector machine-based methods for classification problems in chemometrics. *Analytica Chimica Acta*. (2010)
15. Guyon I, Weston J, Barnhill S, Vapnik V: Gene selection for cancer classification using support vector machines. *Machine Learning*, 46(1):389–422. (2002)
16. S. Wu, C. Wu, S. Lin, K. Lee, C.K. Peng, "Analysis of complex time series using refined composite multiscale entropy," *Phys. Lett. A*, 378 (20), pp. 1369-1374 (2014)
17. M. Costa, A. L. Goldberger, and C. K. Peng, "Multiscale entropy analysis of biological signals," *Phys. Rev. E*, vol. 71, pp. 1–17 (2005)
18. P. H. Tsai, C. Lin, J. Tsao, "Empirical mode decomposition based detrended sample entropy in electroencephalography for Alzheimer's disease," *Journal of Neuroscience Methods*, vol. 210, pp. 230-237, Sep (2012)
19. W. A. Wallis and G. H. Moore, "A Significance Test for Time Series Analysis." *Journal of the American Statistical Association*, vol. 36, pp. 401- 409 (1946)
20. S. Dash, E. Raeder, S. Merchant and K. Chon, "A statistical approach for accurate detection of atrial fibrillation and flutter," in Proc, Annual Computers in Cardiology Conference (CinC), pp. 137-140 (2009)
21. J. Friedman. Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, 29(5):1189–1232 (2001)
22. Adam-Bourdarios, C.; Cowan, G.; Germain-Renaud, C.; Guyon, I.; Kégl, B.; Rousseau, D. The Higgs Machine Learning Challenge. *J. Phys. Conf. Ser* (2015)
23. Phoboo, A.E. Machine Learning wins the Higgs Challenge. *CERN Bull* (2014)