An Evaluation of Feature Extraction in EEG-Based Emotion Prediction with Support Vector Machines

Itsara Wichakam and Peerapon Vateekul
Department of Computer Engineering
Chulalongkorn University
Bangkok, Thailand
Itsara.W@Student.Chula.ac.th, Peerapon.V@Chula.ac.th

Abstract*—*Electroencephalograph (EEG) data is a recording of brain electrical activities, which is commonly used in emotion prediction. To obtain promising accuracy, it is important to perform a suitable data preprocessing; however, different works employed different procedures and features. In this paper, we aim to investigate various feature extraction techniques forEEG signals. To obtain the best choice, there are fourfactors investigatedin the experiment: (i) the number of channels, (ii) signal transformation methods, (iii) feature representations, and (iv) feature transformation techniques. Support Vector Machine (SVM) is chosen to be our baseline classifier due to its promising performance. The experimentswere conducted on the DEAP benchmark dataset. The results showed that the prediction on EEG signals from 10 channels represented by the bandpower one-minute features gave the best accuracy and F1.

Keywords—EEG; emotion; feature extraction; classification; prediction

I. INTRODUCTION

Electroencephalogram (EEG) is a measurement method to record the neural activity in our brain. It has applied in many researches, particularly in emotion prediction. There are five major brain waves: delta (0.5-4 Hz), theta (4-7.5 Hz), alpha (8-13 Hz), beta (14-26 Hz), and gamma (30+ Hz) [1]. Note that each work may define these brain wave ranges differently [2]. The effective bandwidth for EEG signals is approximately up to 100 Hz. However, it is possible to receive EEG signals with the range of bandwidth between 200 and 300 Hz. In the emotion prediction task, researchersusually consider onlysignals with the frequency range between 0.5 and 50 Hz. Fig. 1 shows the electrode positions to probe the brain wave (called 10-20position system), which is recommended by the International Federation of Societies for Electroencepha-lography and Clinical Neurophysiology. In many experiments, the number of probes can be smaller than the standard one.

In the domain of emotion prediction, there are two measurement spaces of emotion states [2]: continuous and discrete. The first space, continuous, is mostly used to indicate a rating for stimulus, such as picture, music, etc., provided during the experiment. The example continuous ratings are valence, arousal, and liking in the range of 1 to 9. The second space, discrete, is employed when there is a standard definition for each emotion state and stimulus in the experiment. The example discrete ratings are positive or negative, happy or unhappy, calm or excited, etc. In this domain, there are many benchmark databases, e.g.,International Affective Picture System (IAPS) and International Affective Digital Sound (IADS) [3]. Many recent trials showed that human emotion can be induced by different kinds of stimulus (picture, music, and video) [2, 4]. One of the most widely used benchmark is a database for analysis using physiological signals (DEAP) [5]. In this dataset, participants evaluated their emotion states using continuous scales.

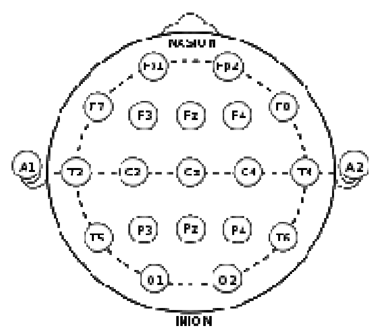

Fig. 1. The 10-20 electrode position system [6].

There were many prior trials in the emotion prediction using EEG signals, where each of them performed different inputs and procedures. It is surprising that there is no standard procedure for EEG data preprocessing. Some of them used 32 probes, while others suggested that 10 probes are sufficient. Also, there are many choices of EGG signal transformations: bandpower and Power Spectral Density (PSD) wavelet[4]. Since EGG signals are continuous,there are many levels of samplings, e.g., seconds, minutes, and statistics of minutes. Sometimes, feature transformation techniques, e.g., PCA and LDA, are appliedin order to enhance the prediction performance. Hence, it should be very helpful if there is a standard procedure of data preprocessing for the emotion prediction task that can ensure to achieve promising accuracy.

In this paper, we focus on finding the best procedure to extract from raw EGG signals to feature vectors that are suitable for the emotion prediction. To achieve the goal, four comparisons are conducted including (*i*) 32 vs. 10 channels, (*ii*) bandpower vs. PSD wavelets, (*iii*) minute data vs. statistics based

978-1-4799-5822-1/14/$31.00 ©2014 IEEE

data, and (*iv*) with or without feature transformation (PCA). All of the comparisons are performedona standard benchmark, DEAP and evaluated by two measures: accuracy and F1. To test statistical significance, there are two types of cross validation: leave-one-trail-out (LOTO) and leave-one-subject-out (LOSO).

In Section II, we present related worksand details of the DEAP dataset. The factors included in the experiment are shown in Section III. The results are discussed in Section IV.

II. RELATED WORKS

In this section, we present details of all related works: experimental dataset, a brief overview of SVM, and a short review of related works.

A. Dataset

The experiment in this paper is conducted based on the dataset DEAP: A Database for Emotion Analysis Using Physiological Signals [5]. There are 32 participants who watched one minute segment videos from YouTube. There are 40 videos in the dataset selected from 120 total videos. During the experiment, 32 electrode positions were recorded during the experiment. After end of each video, participants were rating their tastes level in continuous scale value of 1-9 point called "self-assessment (SAM)". For valence, each level is represented from unhappy or sad (1) to happy or joyful (9). For arousal, each level is represented from calm or bored (1) to stimulated or excited (9). For dominance and liking, the rating score is also between 1 to 9. Moreover, participants have to evaluate their familiarity of the video using a discrete rate: 1 is"Never heard it before the experiment" and 5is"Knew the song very well". Fig. 2 show images used for self-assessment. Table I summarizes a statistics the of DEAP dataset.

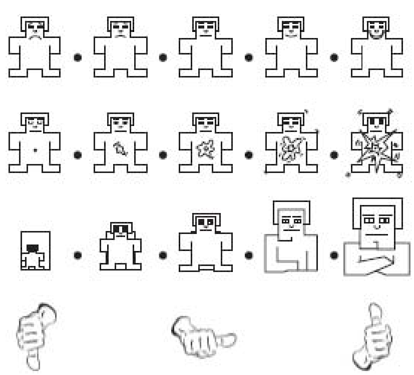

Fig. 2. Images used for self-assessment (SAM): valence (top row), arousal, dominance, and liking (bottom row) [5].

TABLE I. DEAP DATASET SUMMARY

Participants	32
Videos	40
Rating scales	Arousal, valence, dominance, and liking are 1-9 continuous values, and familiarity is 1-5 discrete values
Recorded signals	32 electrode positions, and 512 Hz sampling rate

B. Support Vector Machines

Support vector machines (SVM) is one of the most famous supervised classification technique that is used in variety research domains. The idea of the SVM was proposed in the 1979 byVapnik [1]. The main concept of SVM is a binary classification that has a decision boundary called a separation hyperplane in the form of $\vec{h} = \vec{w}\vec{x} + b$, where w is an orientation vector and b is a bias. The result of a prediction is a SVM score, whose sign is used as a prediction result either positive or negative.

SVM can model complex, real-world problems such as text classification, image classification and bioinformatics analysis. Furthermore, it performs very well on handling the problem domains where the number of features exceeds the number of training examples.

C. Previous Works in EEG-Based Emotion Prediction

*1) Discovering Emotion-Inducing Music Features Using EEG Signals [4]:*In these works, the data collection is divided into two phases. In the first phase, participants listened to a set of songs and, then, they gave each song a rating score between 1-5 for joyful, sad, relaxing, and stressful. In the second phase, EEG signals were monitored during the participants were listening to a different set of songs.The data is gathered from 10 electrode positions, i.e., Fp1, Fp2, F3, F4, T3, T4, P3, P4, O1, and O2. Finally, each EEG signal wasextracted into a set of feature vectors representing different fequency bands: theta (5-8 Hz), alpha (8-13 Hz), and beta (13-20 Hz) filtered byfast Fourier transform (FFT).

*2) Real-Time EEG-Based Happiness Detection System [2]:*In this work, 10 participantswere stimulatedby pictures from Geneva Affective Picture Database (GAPED) [7] and sounds from classical emotion elicitation. The brain waves were collected from 14 channels, i.e., AF3, AF4, F3, F4, F7, F8, FC5, FC6, P7, P8, T7, T8, O1, and O2. Each EEG signal was represented by 5 frequency vectors: delta (0-4 Hz), theta (4-8 Hz), alpha (8-16 Hz), beta (16-32 Hz), and gamma (32-64 Hz). Finally, Gaussian SVM was employed as a classifier to predict participants' emotion on different stimulus (pictures and musics).

III. PROPOSED FACTORS FOR FEATURE EXTRACTION

This paper aims to find the best feature extraction for EEG signals. This section demonstrates all variations of feature sets and how to create them.

A. The number of electrode positions

There are two candidate numbers of electrode positions: 32 and 10 channels.

*1) 32channels:*it is the full data provided in the DEAP dataset.

*2) 10 channels:*it is a subset of channels proposed in [4] including Fp1, Fp2, F3, F4, T7, T8, P3, P4, O1, and O2.

B. Feature Extraction

In the DEAP dataset, there are 32 participants (subjects), 40 videos (trials) for each subject,32 channels recorded for each trial on 60 seconds length[5]. Therefore, the overall DEAP dataset (*D*) comprises of 32-subjects×40-trials × 32-channels × 60-seconds. Moreover, each EEG signal is transformed into four frequency bands: theta (4-8 Hz), alpha (8-12 Hz), beta (13-30 Hz), and gamma (30-45 Hz). These bands follow the ranges used in the DEAP dataset [5]. Also, the delta band is removed following prior works [2, 5, 4, 8] since it has proved that the low frequency band is not necessary for the emotion prediction task.

Using the DEAP dataset, many variations datasets were generated and tested in order to find the best EGG representation that can give the highest prediction performance.There are two major EEG signal transformations: bandpower and PSD wavelets.

*1) Bandpower:*there are four variation datasets.

a) The dataset"D_10_1": It is constructed using EEG signals from 10 channels. The signal is represented in minutes, which is converted from the EEG signal in seconds via an average function.Therefore, each participant is illustrated in a vector of 40 features (4frequency bands × 10channels × 1 minute feature) as follow.

$$\text{Instance } X : \langle x_1, ..., x_{40} \rangle \quad (1)$$

b) The dataset"D_10_5": The signal is collected using 10 channels and represented in minutes along with a set of 5 statistics values: average, max, min, range, and standard deviation. Hence, each observation is represented in a vector of 200 features (4frequency bands × 10channels × 5statistics features)as follow.

$$\text{Instance } X : \langle x_1, ..., x_{200} \rangle \quad (2)$$

c) The dataset "D_32_1": The wave is gathered using all of the 32 channels and represented it in minutes using an average function; thus, each example is illustrated in a vector of 128 features (4frequency bands × 32 channels × 1minute feature)as in the following formula.

$$\text{Instance } X : \langle x_1, ..., x_{128} \rangle \quad (3)$$

d) The dataset "D_32_5": It is built around the signal collected from all of the 32 channels. Each signal is shown in minutes along with 5 statistics values: average, max, min, range, and standard deviation, so there are 640 features for each instance(4frequency bands × 32 channels × 5statistics features)as follow.

$$\text{Instance } X : \langle x_1, ..., x_{640} \rangle \quad (4)$$

*2) PSD by Wavelet Transform:*there are two variation datasets.

*a) The dataset "D_10_1_WT":*The data gathering is processed on 10 channels. It is similar to the dataset "*D_10_1*" except the brain wave is transformed using PSD. Each example is represented by a vector of 40 features as follow.

$$\text{Instance } X : \langle x_1, ..., x_{40} \rangle \quad (5)$$

*b) The dataset"D_32_1_WT":*The data collection is processed on 32 channels. It is similar to the dataset "*D_32_1*" except the brain wave is transformed using PSD. Each example is represented by a vector of 128 features as follow.

$$\text{Instance } X : \langle x_1, ..., x_{128} \rangle \quad (6)$$

All of the datasets comprises of 1,280 cases (40trials × 32subjects) as in Equation 7. Moreover, each feature value is normalized into the range of [0, 1] using Equation 8.

$$\text{Dataset } D : \langle X_1, ..., X_{1280} \rangle \quad (7)$$

$$\forall x_i \in X_{Subj.j} \left(Norm. = \frac{x_i - \min_{Subj.j}}{MAX_{Subj.j} - \min_{Subj.j}} \right) \quad (8)$$

C. Binary Class Conversion

There are three emotion states including valence, arousal, and liking. Each emotion state is originally rated by participants using a continuous scale (1-9) in the DEAP dataset. In this paper, this continuous scale is converted into a binary scale, either positive (+1) or negative (-1); therefore, each emotion state is represented as a binary class and there are total three binary classes (emotions).

The conversion is based on a cutoff value, where any scores less than the cutoff are negatives, and the others are positives. The cutoff is an average score of all participants.

TABLE II showsthe class distribution of each emotion state. The positive class outnumbers the negative one; this circumstance is called "imbalance" [5].

TABLE II. THE NUMBER OF POSITIVE / NEGATIVE CLASS

	Positive	Negative
Valence	724	556
Arousal	754	526
Liking	857	423

IV. EXPERIMENTS

This section shows details of the experiments including data distribution, the measurement, and the results along with their analysis.

LIBSVM [9] is our choice of SVM since it is publicly available and widely used. Radial Basis Function (RBF) is a selected kernel function and all related parameters (c and gamma) are computed by a grid search method with 3 fold-cross validation.

A. Experimental Setup

1) Data Distribution: As in Section 3 (B), there are 6 experimental datasets (4 datasets filtered by bandpower and 2 datasts transformed by PSD). Each dataset has 3 binary classes: valence, arousal, and liking. To evaluate the result, there are two strategies. First, leave-one-trial-out (LOTO): leave one trial for test and the remaining trails for train. Second, leave-one-subject-out (LOSO): leave one subject for test and the remaining subjects for train. Therefore, there are 40 folds in LOTO and 32 folds in LOSO.

2) Experimental Measurements: Accuracy (Acc) and F-measure (F1) are our main performance criteria. Given*TP*: the number of true positive, *TN*: the number of true negative, *FP*: the number of false positive, and *FN*: the number of false negative.Equation 9 shows the accuracy, which is an average of *k*-fold cross validation accuracy.

$$Accuracy = \frac{\sum_{n=1}^{k-fold} TP_{fold_n} + FN_{fold_n}}{k \times \text{number of instances}} \quad (9)$$

For the F-score, it is computed based on two other measures call Precision (Pr) and Recall (Rc). For the positive class, Pr^+, Rc^+, and $F1^+$ are shown in Equations 10-12. To evaluate F1 on both classes, an average function is employed as shown in Equation 13.

$$\Pr^{+} = \frac{TP}{TP + FP} \quad (10)$$

$$Rc^{+} = \frac{TP}{TP + FN} \quad (11)$$

$$F1^{+} = \frac{2 \times \Pr^{+} \times Rc^{+}}{\Pr^{+} + Rc^{+}} \quad (12)$$

$$F1^{*} = \frac{F1^{+} + F1^{-}}{2} \quad (13)$$

B. Experimental Results.

a) 1 Minute feature vs 5 statistic features: The first experimental, we compared between 1 minute feature training sets and 5 statistic features training sets (D_10_1, D_10_5, D_32_1, and D_32_5) for both 10 and 32 channels for all class labels. The result (in the TABLE III) show that 1 minute feature training sets is absolutely better than 5 statistic features training sets. Thus, we can conclude that the EEG signals might not be represented with the statistic values features for classification.

b) 10 Channels vs 32 Channels: The second experimental, we compared between 10 channels training sets and 32 channels training sets (D_10_1, D_10_5, D_32_1, and D_32_5) for all class labels. The result (in the TABLE III) show that 10 channels features training sets is better than 32 channels features training sets for 2 class labels, valence, and arousal except liking that show us the 32 channels is slightly better than 10 channels. So, we used the T-test method to test the accuracy and the F-score for analyzed how is difference between two training sets (D_10_1, and D_32_1). The result of T-test are p-values. The first is a p-value of accuracy is 0.313, and the second is a p-value of F-score is 0.001. Thus, we can conclude that the number of channel to used for feature extraction method is 10 channels is enough for classify. That may help us to reduce the number of resource which consistent to many works e.g., [4], and [8].

c) Normal features vs Principal component analysis (PCA) features: The third experimental, we selected the training sets that gave us the best accuracy (from the first fourth rows in the TABLE III) for all class labels. Then, we computed the PCA of the training sets that we selected (D_10_1, and D_10_5), and then, we selected the first 95% of eigen values from the eigen vectors to the features vector, then we compared those. The result show that the normal training sets is absolutely better than the PCA taining sets.

TABLE III. COMPARISION ON OUR DIFFERENCE TRAINING SETS

	Valence		Arousal		Liking	
	Acc	**F1**	**Acc**	**F1**	**Acc**	**F1**
D_10_1	**0.649**	0.698	**0.649**	0.721	0.668	0.798
D_10_5	0.547	0.684	0.605	0.722	0.670	0.802
D_32_1	0.623	0.680	0.629	0.701	**0.673**	0.799
D_32_5	0.548	0.633	0.623	0.708	0.669	0.801
PCA(D_10_1)	0.629	0.687	0.629	0.708	0.670	0.800
PCA(D_10_5)	0.640	0.690	0.640	0.704	0.666	0.794

d) Bandpower vs PSD by Wavelet transform: The fourth experimental, We selected the training sets from the best accuracy for all labels from the fourth rows in the TABLE V i.e. D_10_1, and D_32_1 where as a difference channels that has been extracted the features vector by the bandpower method to compared with the D_10_1_WT, and D_32_1_WT training sets that has been extracted the features vector by PSD by wavelet transform method. The result (in the TABLE IV) show that the bandpower method is absolutely better than the wavelet transform method for valence label for both 10 and 32 channels, arousal label for 10 channels, and liking label for 32 channels while the PSD by wavelet transform method is slightly better than the bandpower method for arousa label for 32 channels, and liking label for 10 channels. Thus, we can conclude that the bandpower method is absolutely better than the PSD by wavelet transform method, while the PSD by wavelet transform method is non significantly better than the bandpower method for all class labels.

TABLE IV. COMPARISION BETWEEN BANDPOWER AND PSD

	Valence		Arousal		Liking	
	Acc	**F1**	**Acc**	**F1**	**Acc**	**F1**
D_10_1	**0.649**	0.698	**0.649**	0.721	0.668	0.798
D_10_1_WT	0.564	0.654	0.630	0.719	0.671	0.796
D_32_1	0.623	0.680	0.629	0.701	**0.673**	0.799
D_32_1_WT	0.600	0.678	0.634	0.708	0.670	0.797

e) LOTO vs LOSO: The last experimental, We compared between LOTO and LOSO method for all class labels using training set D_10_1. The result (in the TABLE V) show that LOTO method is better than LOSO method for valence and liking label. Thus, we can conclude that the accuracy of emotion classigication from DEAP EEG dataset has been depends on trial more than on subject.

TABLE V. COMPARISON BETWEEN LOTO AND LOSO

	Valence		Arousal		Liking	
	Acc	**F1**	**Acc**	**F1**	**Acc**	**F1**
LOTO	**0.649**	0.698	**0.649**	0.721	0.668	0.798
LOSO	0.511	0.590	0.529	0.623	**0.670**	0.802

In the TABLE VI, we compared our experiment result with DEAP [5] and the random technique called "all positive". The DEAP was used a Gaussian naïve Bayes technique with Fisher's linear discriminantfeature selection to classify [5]. All positive is a technique that calculated in term of assumes all predict classes to positive for all instances. The result shows that our feature extraction method gets better accuracy than DEAP. Thus, we can conclude that our feature extraction method can be good enough to build the model for classify a DEAP EEG signals.

TABLE VI. OUR EXPERIMENT RESULT COMPARED TO DEAP

	Valence		**Arousal**		**Liking**	
	Acc	**F1**	**Acc**	**F1**	**Acc**	**F1**
Our best selected	**0.649**	0.5140*	**0.65**	0.508*	**0.668**	0.416*
DEAP	0.567	0.5633*	0.62	0.583*	0.554	0.502*
All positive	0.566	0.7226	0.59	0.741	0.670	0.820

* F1 that calculated from the average for both classes in Equation 16.

V. CONCLUSION

The goal of this paper is to suggest the best data preprocessing (feature extraction) for the EGG data to predict the emotion states, such as happy vs. unhappy, bored vs. excited, etc.It should be very helpful as a standard procedure for the future research. The experiment was conducted on the DEAP benchmark based on SVM. The result suggests that the best feature extraction is "one-minute EEG data using bandpower filter from 10-channel probes." There is no performance improvement from using the whole 32 channels, adding any additional features, or performingthe feature transformation, PCA. Moreover, our proposed framework outperforms the original system presented in DEAP in terms of accuracy and F1.

In the future research, we plan to vary and select a suitable set of positions specifically for each emotion state.

REFERENCES

[1] S. Sanei and J.A. Chambers, EEG Signal Processing. 2007.

[2] N. Jatupaiboon, S. Pan-ngum, and P. Israsena, "Real-time EEG-based happiness detection system," The Scientific World Journal, vol. 2013.

[3] University of Florida. "The center for the study of emotion and attention." Internet: csea.phhp.ufl.edu/media.html, [Jan. 30, 2014].

[4] R. Cabredo, R. Legaspi, P. S. Inventado, and M. Numao, "Discovering emorion-inducing music features using EEG signals," Journal of Advanced Computational Intelligence and Intelligent Informatics, vol. 17, no. 3, 2013.

[5] S Koelstra, C. Muhl, M. Soleymani, J.S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, and A. Nijholt, "DEAP: a databased for emotion analysis using physiological signals", IEEE Trans. On Affective Computing, vol. 3, no. 1, pp. 18-31, Jan-Mar 2012.

[6] Wikipedia. "10-20 system (EEG)." Internet: en.wikipedia.org/wiki/10-20_system_(EEG), [Jan. 30, 2014].

[7] Affective Science. "Research material." Internet: www.affective-sciences.org/researchmaterial, [Jan. 30, 2014].

[8] N. Jatupaiboon, S. Pan-ngum, and P. Israsena, "Emotion classification using minimal EEG channels and frequency bands," in Proc. the 10th International Joint Conference on Computer Science and Software Engineering (JCSSE2013), pp. 21-24, 2013.

[9] C. C. Chang, and C. J. Lin, "LIBSVM: a library for support vector machines," ACM Trans. On Intelligent Systems and Technology, 2011. Software available at www.csie.ntu.edu.tw/~cjlin/libsvm.