

“© 2015 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.”

Investigation of Window Size in Classification of EEG-Emotion Signal with Wavelet Entropy and Support Vector Machine

Henry Candra*, Mitchell Yuwono, Rifai Chai, Ardi Handojoseno, Irraivan Elamvazuthi, Hung T. Nguyen, and Steven Su

Abstract—When dealing with patients with psychological or emotional symptoms, medical practitioners are often faced with the problem of objectively recognizing their patients' emotional state. In this paper, we approach this problem using a computer program that automatically extracts emotions from EEG signals. We extend the finding of Koelstra et. al [IEEE trans. affective comput., vol. 3, no. 1, pp. 18–31, 2012] using the same dataset (i.e. the DEAP: dataset for emotion analysis using electroencephalogram, physiological and video signals), where we observed that the accuracy can be further improved using wavelet features extracted from shorter time segments. More precisely, we achieved accuracy of 65% for both valence and arousal using the wavelet entropy of 3 to 12 seconds signal segments. This improvement in accuracy entails an important discovery that information on emotions contained in the EEG signal may be better described in term of wavelets and in shorter time segments.

I. INTRODUCTION

Research in human emotion recognition system has been conducted with many varieties of physiological signals including the brain wave or electroencephalography (EEG). A number of investigations attempted to map the EEG signal to the conceptual arousal and valence emotional dimensions as described in Russel's circumplex model [1]. According to Russel, arousal and valence represented engagement/disengagement and pleasant/unpleasant feeling, respectively [1].

Various algorithm has been proposed to identify arousal and valence. Classification results varied from low to high depending on the method used. An example of such successful research was achieved by investigating the heart rate and skin impedance as shown in a previous paper by Swagner and Kaber [2].

A substantial factor for determining the reliability of an emotion recognition systems in general is the feature selection step [3]. Particularly for features obtained using spectral/wavelet analysis, there are a number of parameters that needs to be tweaked including the size of the time window. The window size is measured in time with unit

of second. An overly wide window can lead to information overload which causes the feature to be mixed up with other information. Similarly the information about emotion might not be adequately extracted if the time window is too short.

The discrete wavelet transform (DWT) coefficient has recently gained popularity for extracting time-frequency domain features in EEG signals [4], [5]. It has been argued that the DWT is advantageous due to: 1. the relative wavelet energy can be naturally associated to the EEG frequency bands, and; 2. The relative entropy of its wavelet power can serve as a reliable estimate of the degree of similarity between different segments of the signal [6].

EEG signals are generally subdivided into several bandwidths known as delta, theta, alpha, beta, and gamma. These bands need to be properly selected as not all of them carry the relevant information for emotion recognition. For example, Zhu et. al. discovered that EEG patterns in the Beta and Gamma bands are generally stable across emotions and subjects. [7]. This means that proper selected frequency bands will improve classification result.

Our contribution in this paper are as follows. We investigate the effective window size to be used for the segmentation of EEG-emotion signals using a publicly available Dataset for Emotion Analysis using electroencephalogram, Physiological and Video Signals (DEAP) [8]. We extracted the features with DWT with varying window widths and calculated the entropy of the detail coefficients corresponding to the alpha, beta, and gamma bands. These features fed into the support vector machine (SVM) classification algorithm [9] for classifying between high/low arousal and high/low valence, respectively. Our investigation revealed that arousal can be classified up to 65.33% accuracy using the window length of 3–10 seconds; while valence can be classified up to 65.13% accuracy using the window length of 3–12 seconds. This result implies that the information regarding these emotions may be appropriately localized at 3–12 seconds time segments.

II. MATERIALS AND METHOD OVERVIEW

A. Materials

The experiment is conducted with DEAP dataset by first carried out the pre-processing step following the procedures used in [8] to remove artifacts from the original EEG signal such as eye blinking utilizing 4 to 45 Hz band pass filter for each channel; averaged the channels to a common reference; and then down sampled the signal to 128 sample/second. The dataset contains 32 channels EEG signals of 32 participants

*Henry Candra, Mitchell Yuwono, Rifai Chai, Ardi Handojoseno, Steven W. Su, and Hung. T. Nguyen are with the Centre for Health Technologies, Faculty of Engineering and Information Technology, University of Technology, Sydney, New South Wales, Australia (Henry.Candra@student.uts.edu.au, mitchellyuwono@gmail.com, Rifai.Chai@uts.edu.au, AluysiusMariaArdi.Handojoseno@student.uts.edu.au), (Steven.Su, Hung.Nguyen)@uts.edu.au) Irraivan Elamvazuthi, is with Dept. of Electrical and Electronic Engineering Universiti Teknologi PETRONAS, Tronoh, Malaysia (Email: irraivan.elamvazuthi@petronas.com.my)

while watching 40 music videos with one minute duration for each video. The participants rated each video with continuous scale 1 to 9 for arousal and valence emotions. Using the rating, two emotions arousal and valence are labeled with High and Low label, respectively. The middle value of the rating is used as the baseline. For the training and testing purpose, a group of 5 participants is created.

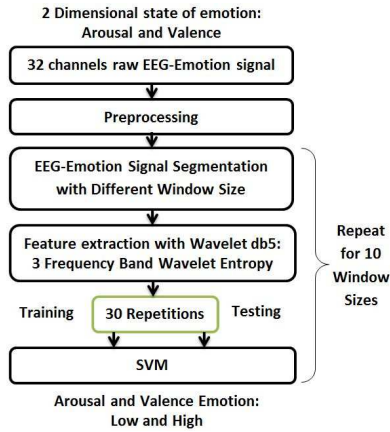


Fig. 1. Block Diagram of Proposed Window Size Investigation Process

B. Method Overview

The proposed method is illustrated in Fig.1. which can be divided into 3 steps: 1. EEG signal segmentation with various window size; 2. Wavelet Features extraction; and; 3. Emotion classification using SVM. The whole process is repeated 10 times for each arousal and valence emotion accommodating the variation of window size. The experiment generated total 20 segmented datasets for EEG-emotion signal classification.

III. METHOD

A. EEG Signals Segmentation with Various Window Size

Before the feature extraction process, the preprocessed EEG signals were segmented into various window sizes. The first window is the largest as much as the size of the EEG signal provides in DEAP dataset that is 60 second. The subsequent steps include reducing the window size by half and continuously taking half of the previous window size until getting the smallest windows size of 1 second is obtained (the smallest window size of 1 second is due to the sampling rate 128 sample/second). To refine the result of the investigation, additional window sizes were inserted between the existing sizes, which in total gave 10 window sizes which are: 60; 30; 20; 15; 12; 10; 8; 6; 3; and 1 second. The block diagram of window segmentation process is illustrated in Fig. 2. For each window size, the signals were then stored as a new dataset. This process resulted in the multiplication of data size in accordance with the number of segment used. For example, taking window size 30 second will result in doubling the data size of 60 second segment, while using 1 second window will multiply the data size 60 times.

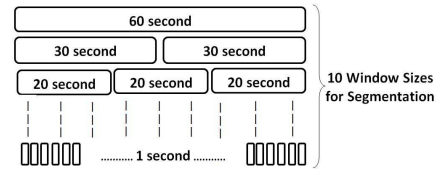


Fig. 2. Block Diagram of Window Segmentation

B. Wavelet Features Extraction

This process can be divided into more details as follows: Wavelet Coefficients calculation; Wavelet Entropy features formation of 3 frequency bands; Array creation of 32 channels Wavelet Entropy resulted in 96 units of array vector.

1) *Wavelet Coefficients calculation*: Discrete wavelet transforms (DWT) based on dyadic scales and positions are utilized for feature extraction, which defined as follows [5]:

$$DWT(j, k) = \int_{-\infty}^{\infty} x(t) \frac{1}{\sqrt{2^j}} \psi \left(\frac{t - 2^j k}{2^j} \right) dt \quad (1)$$

where $2^j k$ and 2^j are the time localization and scale respectively, while $\psi(t)$ denotes the mother wavelet function.

For multi resolution signal decomposition of DWT, the signal $y(t)$ is processed through low pass filter (LPF) and high pass filter (HPF) before down sampled by 2. The LPF output gives the approximation A_j within band frequency $f_m/2$ to f_m , while the HPF output provides the detail D_j within the band frequency $f_m/2$ to f_m . The frequency subband of the sampling frequency related to the original signal is f_s where $f_m = f_s/2^{(1+l)}$ with 1 is the decomposition level. For a given EEG signal in a channel $y(t)$, the coefficient of wavelet decomposition with DWT is:

$$y(t) = \sum_{k=-\infty}^{\infty} A(k) \psi_k(t) + \sum_{j=0}^{\infty} \sum_{k=-\infty}^{\infty} D(j, k) \psi_{j,k}(t) \quad (2)$$

2) *Wavelet Entropy features formation of 3 frequency band*: For this purpose, Daubechies (db5) wavelet is selected as the wavelet function, dividing decomposition EEG signal with sampling rate 128 sample/second provide the results into 5 bands: delta (3-4Hz); Theta (5-8Hz); Alpha (9-16Hz); Beta (17-32Hz); and Gamma (33-64Hz). The energy of related frequency bands is then extracted. From this 5 band, only 3 bands are utilized, namely: alpha, beta, and gamma. The Energy of wavelet coefficients both approximation and detail can be represented as:

$$E_k = \sum_{k=1}^N |C_{j,k}|^2, j = 1, \dots, l \quad (3)$$

where $C_{j,k}$ is the detail or approximation, and N is the number of wavelet coefficients at each decomposition level.

To normalize the energy of the wavelet packet, energy of detail is compared to the total energy of the signal with:

$$|E_k| = \frac{E_k}{E_t} \quad (4)$$

Where E_t is total energy of the signal as in:

$$E_t = \sum_k E_k \quad (5)$$

The processed is continued to generate the relative wavelet entropy. According to [6] wavelet entropy calculate the degree of similarity between different segments of the EEG signal. Total wavelet entropy can be calculated as follows:

$$G(x) = - \sum_j p_{j,k} \log p_{j,k} \quad (6)$$

3) *Array creation of 32 channels of Entropy*: All the normalize entropy of wavelet are then arranged in an array of 32 channels with each array related to one segment of EEG-emotion signal, respectively.

C. SVM training and testing

For the classification process, SVM with a Radial Basis Function (RBF) kernel is utilized. The kernel radius RSVM is inferred from the training data using Ensemble Rapid Centroid Estimation (ERCE) [10], [11]. The SVM radius estimation step is as follows,

- 1) Let $\mathbf{p}_j : p \in \{0,1\}$ be a vector of probabilities describing the wavelet entropy features in the training set,
- 2) Execute ERCE [10], [11] to cluster $\mathbb{P} = \{\mathbf{p}_1, \mathbf{p}_2, \dots\}$ to an arbitrary number of cluster based on Jensen-Shannon distance.
- 3) Aggregate the ensemble clustering results using average linkage to get the final clustered sets $\{\mathbb{C}_1 \cup \dots \cup \mathbb{C}_K\}$, where K is determined automatically by ERCE at ensemble aggregation. The corresponding centroid vectors $\{\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_K\}$ are computed as follows,

$$\boldsymbol{\mu}_k = \frac{1}{|\mathbb{C}_k|} \sum_{\mathbf{p} \in \mathbb{C}_k} \mathbf{p} \quad (7)$$

- 4) The SVM radius R_{SVM} is taken as the average cluster radius in terms of euclidean distance as follows,

$$R_{\text{SVM}} = \frac{1}{K} \sum_{k=1}^K \sum_{\mathbf{p} \in \mathbb{C}_k} \frac{\|\mathbf{p} - \boldsymbol{\mu}_k\|}{|\mathbb{C}_k|}. \quad (8)$$

The SVM is then trained using Sequential Minimal Optimization (SMO) algorithm.

IV. RESULTS AND DISCUSSION

The training and testing processes are repeated for the 10 different sizes of window segments. The training utilized 50% of the dataset, while the rest 50% were used for testing. The classification result from 30 repetitions of training and testing of arousal and valence emotions with 10 various window sizes are listed in Table I and Table II, respectively.

Table I shows that the classification result can be broadly summarized into 3 regions:

- 1) *20–60 seconds*: The window size was reduced dramatically from 60 to 20 second. A gradual change of accuracy about 1.57% were observed. The accuracy in this region is between 60.60% and 62.17%.

TABLE I
CLASSIFICATION RESULT OF AROUSAL WITH 10 WINDOW SIZES

| Window Size | Accuracy | Sensitivity | Specificity |
|---------------|---------------|----------------|----------------|
| 60 sec | 60.60% ± 4.2% | 44.19% ± 16.0% | 74.43% ± 12.9% |
| 30 sec | 61.32% ± 2.8% | 45.08% ± 17.4% | 74.36% ± 12.7% |
| 20 sec | 62.17% ± 2.3% | 42.56% ± 11.3% | 77.61% ± 10.0% |
| 15 sec | 63.24% ± 2.4% | 48.56% ± 9.8% | 74.57% ± 7.4% |
| 12 sec | 63.36% ± 2.4% | 49.94% ± 10.3% | 73.68% ± 7.6% |
| 10 sec | 64.32% ± 1.7% | 53.44% ± 7.5% | 72.61% ± 6.2% |
| 8 sec | 64.76% ± 1.7% | 51.47% ± 8.8% | 75.03% ± 6.7% |
| 6 sec | 64.80% ± 1.2% | 50.90% ± 5.9% | 75.69% ± 4.5% |
| 3 sec | 65.33% ± 1.1% | 55.99% ± 5.9% | 72.37% ± 4.5% |
| 1 sec | 63.67% ± 1.0% | 48.35% ± 5.1% | 75.31% ± 3.8% |

- 2) *3–20 seconds*: Accuracy increased by 2.15% when the window size is reduced from 20 seconds to 10 seconds, and 3.72% when the window size is reduced from 60 seconds to 10 seconds. This region was characterized by small variation of accuracy. The highest accuracy in this region was 65.33% (3 seconds window) which about 4.73% increase compare to the lowest accuracy in the region.
- 3) *1–3 seconds* The accuracy in this region reduces immensely from 65.33% to 63.67%.

TABLE II
CLASSIFICATION RESULT OF VALENCE WITH 10 WINDOW SIZES

| Window Size | Accuracy | Sensitivity | Specificity |
|---------------|---------------|----------------|----------------|
| 60 sec | 59.63% ± 6.5% | 39.08% ± 20.6% | 76.83% ± 15.8% |
| 30 sec | 60.63% ± 4.2% | 40.56% ± 15.1% | 77.72% ± 13.6% |
| 20 sec | 61.47% ± 2.5% | 39.28% ± 10.1% | 80.72% ± 7.7% |
| 15 sec | 63.62% ± 2.2% | 56.28% ± 6.8% | 69.85% ± 4.7% |
| 12 sec | 65.05% ± 1.6% | 59.79% ± 6.9% | 69.45% ± 5.4% |
| 10 sec | 64.34% ± 1.3% | 58.67% ± 6.0% | 69.17% ± 5.2% |
| 8 sec | 64.66% ± 1.3% | 59.54% ± 5.1% | 68.98% ± 5.0% |
| 6 sec | 65.13% ± 1.2% | 61.23% ± 4.9% | 68.39% ± 3.7% |
| 3 sec | 65.03% ± 1.0% | 56.37% ± 6.3% | 72.24% ± 4.4% |
| 1 sec | 63.36% ± 0.6% | 49.20% ± 3.6% | 75.37% ± 3.0% |

Similar to the first table, the classification result for valence emotion in Table II can also be divided into 3 corresponding regions:

- 1) *20–60 seconds*: Accuracy rose from 59.63% to 61.47% (about 1.84% slight rise) for the large reduction of window size from 60 to 20 second.
- 2) *3–20 seconds*: Accuracy rose about 5.42% compared to the lowest accuracy in the previous region. The accuracy varies between 61.47% and 65.13%.
- 3) *1–3 seconds*: Again, 1.67% reduction of accuracy were observed for only 2 second reduction of window size. Comparison between arousal and valence emotions indicated that classification of arousal is 0.20% higher than valence emotion which is relatively insubstantial. Furthermore, sensitivity and specificity of valence classification followed closely those of arousal.

Further investigation was conducted with graphical representation of Table I and II which illustrated in Fig. 3. The

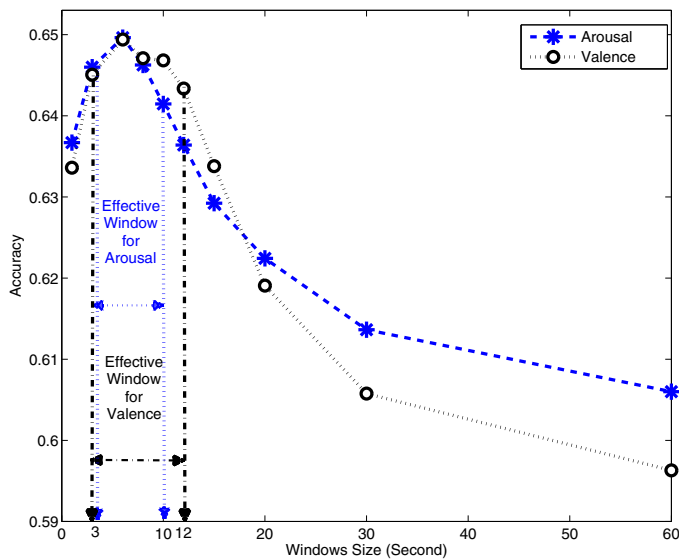


Fig. 3. Graphical Representation of Arousal and Valence Classification

first region for both arousal and valence were dominated by wider range of window size with less accuracy. The accuracy peaked at the second region and fell in the third region. It can be observed that the effective window size for arousal and valence emotions were both located in the second region. The effective window size for arousal was between 3–10 seconds while the effective window size valence was between 3–12 seconds. Both emotions were correctly classified with accuracies of higher than 64%. This finding suggests that the window size for arousal and valence emotion might be located in a slightly overlapping, yet differing range. To our knowledge, this subtle difference has never been considered as the two emotions were generally given the same treatment. From the results obtained in this experiment, it can be observed that the information regarding emotions in the EEG signal may be appropriately localized at around 3–12 seconds time segments.

V. CONCLUSION

Our investigation on the effect of window size suggests that information about valence and arousal emotions were localized in specific time segments in the EEG signal. Our experimental results reveal that the effective window size for arousal and valence is between 3–10 and 3–12 second, respectively. Within these regions, the accuracy of higher than 64% can be achieved for both emotions. Using the appropriate time window, a reliable emotion detector with higher accuracy, sensitivity and specificity can be trained. Future work will include further investigation for other emotions, features, and classification methods.

REFERENCES

- [1] J. A. Russell, "A circumplex model of affect." *Journal of Personality and Social Psychology*, vol. 39, no. 6, pp. 1161–1178, 1980.
- [2] M. Swangnetr and D. B. Kaber, "Emotional state classification in patient–robot interaction using wavelet analysis and statistics-based feature selection," *Human-Machine Systems, IEEE Transactions on*, vol. 43, no. 1, pp. 63–75, 2013.
- [3] R. Jenke, A. Peer, and M. Buss, "Feature extraction and selection for emotion recognition from eeg," *IEEE Trans. Affective Comput.*, vol. 5, no. 3, pp. 327–339, 2014.
- [4] Rizon, "Discrete wavelet transform based classification of human emotions using electroencephalogram signals," *American Journal of Applied Sciences*, vol. 7, no. 7, pp. 878–885, 2010.
- [5] A. Handojoseno, J. Shine, T. Nguyen, Y. Tran, S. Lewis, and H. Nguyen, "The detection of freezing of gait in parkinson's disease patients using eeg signals based on wavelet decomposition," in *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, Aug 2012, pp. 69–72.
- [6] O. A. Rosso, S. Blanco, J. Yordanova, V. Kolev, A. Figliola, M. Schürmann, and E. Başar, "Wavelet entropy: a new tool for analysis of short duration brain electrical signals," *Journal of Neuroscience Methods*, vol. 105, no. 1, pp. 65–75, 2001.
- [7] J.-Y. Zhu, W.-L. Zheng, Y. Peng, R.-N. Duan, and B.-L. Lu, "Eeg-based emotion recognition using discriminative graph regularized extreme learning machine," in *Neural Networks (IJCNN), 2014 International Joint Conference on*, July 2014, pp. 525–532.
- [8] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "Deap: A database for emotion analysis using physiological signals," *IEEE Trans. Affective Comput.*, vol. 3, no. 1, pp. 18–31, 2012.
- [9] M. I. Schlesinger and V. Hlavac, *Ten lectures on statistical and structural pattern recognition*. Kluwer Academic, 2002.
- [10] M. Yuwono, S. Su, B. Moulton, and H. Nguyen, "Data clustering using variants of rapid centroid estimation," *Evolutionary Computation, IEEE Transactions on*, vol. 18, no. 3, pp. 366–377, June 2014.
- [11] M. Yuwono, S. Su, B. Moulton, Y. Guo, and H. Nguyen, "An algorithm for scalable clustering: Ensemble rapid centroid estimation," in *Evolutionary Computation (CEC), 2014 IEEE Congress on*, July 2014, pp. 1250–1257.