Entropy and Complexity Assisted EEG-based Mental Workload Assessment System

Graduate Institute of Electronics Engineering, National Taiwan University, Taipei, Taiwan
b03901062@ntu.edu.tw, {kane,cys,edan,andywu}@access.ee.ntu.edu.tw

Abstract—As the era of Brain-Computer Interfacing (BCI) arrives, computationally measuring human mental workload via Electroencephalography (EEG) signal has become a crucial research field. Conventionally, mental workload assessment studies are mainly based on time-statistics, frequency, and wavelet domain features. In this paper, we present a mental workload assessment system in discriminating high and low mental workload by extracting EEG features from two new domains: time-complexity and entropy domains features. According to statistical analysis, the result demonstrates that the Frontal and Frontal-Central are two dominating regions. In addition, by fusing the traditional and new features, we boosted the classification performance from 69% to 88%. It indicates time-complexity and entropy domain features are able to extract some non-linear characteristics of EEG, which could not be achieved by traditional approaches. We conclude that the new features are feasible to assess human mental workload, and could provide complementary information to traditional features.

Index Terms—Mental Workload, EEG, Complexity, Entropy.

I. INTRODUCTION

Mental workload (MW) is an indicator of the brain's effort. It reflects the information processing capability that used to perform a task [1]. Due to its good representation of the human cognitive and mental state, it possesses numerous applications, ranging from safety to smart technology, including driver awareness, mental health monitoring and Brain-Computer Interfacing (BCI). Hence, Mental Workload Assessment (MWA) has been an important issue for real-world applications.

Recently, MWA has been studied in Electroencephalography (EEG) widely. [2] developed a classification model with Event-Related Potential (ERP), power frequency features. The model achieves high accuracy in estimating MW. [3] study the behavior of alpha rhythm (8-12Hz), they found the suppression happened when condition changes from rest to task. Besides, [4] investigates the performance of several physiological signals while EEG showed the highest performance over all modalities.

Conventionally, EEG is commonly analyzed in time, frequency and wavelet domains. For example, the power of different frequency bands (θ,α,β) is always treated as powerful features, especially for the alpha band (8-12Hz) [2], [4]. Besides, wavelet features are also commonly used [5]. They reveal both time and frequency information, similar

performance can also be achieved by Wavelet features. However, EEG is a multidimensional, non-stationary time series, so using only time, frequency and wavelet domain features might not able to capture all the EEG characteristics, especially for those non-linear characteristics. Hence, some non-linear methods to extract useful information is necessary for MWA.

Recently, non-linear EEG features, such as complexity and entropy measurements, were commonly used in disease detection. [6] employed permutation entropy to perform complexity analysis of alzheimer's disease. [7] used wavelet entropy in discriminating Attention-Deficit Hyperactivity Disorder patient, 96% accuracy is achieved in this task. Fractal Dimension [8], considered as a complexity features, is also treated as a biomarker for Dementia. From the evidence of these related works, complexity and entropy features do reveal some important characteristics of EEG, which may also be informative in assessing mental workload. Hence, we investigate the feasibility of complexity and entropy features in this study.

Our research has two remarks: 1) We find that complexity and entropy domain features are highly correlated to MW. 2) Complexity or entropy domain features do have complementary information to time and frequency domain features. By fusing them with traditional time and frequency domain features, classification result has shown a huge improvement (+18%).

II. BACKGROUND

A. Task Selection and Dataset

Our study is based on an open-access brain-imaging dataset [9], which consists of 28-channels EEG according to the international 10-5 system. 26 subjects performed N-backs task (0-, 2- and 3-back) experiment, which requires subjects to memorize the alphabet serially, and response when the letter shows up same with the N-th times before. By tuning N, different levels of MW can be obtained. Our goal is to determine MW (0-back v.s. 3-back) based on [9].

B. Processing Flow

Our processing flow includes data pre-processing, feature extraction, selection, and classification. First, bandpass filter and rejection methods were used to remove baseline noise and eye blink artifacts respectively. Next, feature extraction was performed on processed data. In addition to extracting

*These two authors contributed equally.
time-statistics, frequency, wavelet domain features, we also extracted time-complexity and entropy domain features, which mentioned in section III. The extracted features would undergo Recursive Feature Elimination (RFE) to filter out redundant features. Last, selected features were fed into Support Vector Machine (SVM) to perform binary classification of MW (0-back vs. 3-back).

III. METHOD

A. Time-Statistics, Frequency and Wavelet Domain Features

To make comparisons, we first implemented traditional features commonly used in MWA, including time-statistics [10], frequency and wavelet domain features [11]. Corresponding extracted features are shown in Table I.

B. Time-Complexity Domain Features

1) Higher-Order Crossing (HOC)

HOC [12] is an efficient and robust method to capture the oscillatory pattern of EEG. As shown in Equ. 1, in order to extract the pattern under different scale, this method first applies iteratively difference operator, \( \nabla^k \), to the zero-mean time series \( Z(X) \). The oscillatory characteristic under different scale can be obtained by this operator. Lastly, those sequence's zero crossing number is considered as HOC. In this study, we set \( k \) to 1,2,3,...,10.

\[
HOC = \text{ZeroCount}(\nabla^{k-1} Z(X))
\]  

2) Higuchi Fractal Dimension (HFD)

HFD [13] measures the change of curve length according to the sampling frequency \( k \). We first generate \( k \) new template vectors from time series \( x \) with length \( N \):

\[
x^k = \{x(m), x(m+k), x(m+2k),...x(m+[\frac{N-m}{k}]} \times k)\}
\]  

where \( \lfloor . \rfloor \) is Gauss' operator, \( k \) is the time interval between sample points, and \( m = 1,2,...,k \). The curve length \( L_m(k) \) for each new template vector is further derived as

\[
L_m(k) = \frac{\sum_{i=1}^{\lfloor \frac{N-m}{k} \rfloor} ||x(m+i\times k) - x(m + (i-1) \times k)||}{\lfloor \frac{N-m}{k} \rfloor \times k}
\]  

Next, \( L(k) \), the average curve length for the time interval \( k \), is derived as:

\[
L(k) = \frac{1}{k} \sum_{m=1}^{K} L_m(k)
\]  

Lastly, we plot \( L(k) \) against \( 1/k \) on a double logarithmic scale and perform linear regression, the slope is HFD.

3) Pattern Spectrum (PS)

Morphological operations are commonly used in image processing, while they are applied on analyzing one-dimensional data such as physiological signal recently [14]. In one-dimensional time series, the idea of Dilation \( D_g(X) \) and Erosion \( E_g(X) \) could be viewed as max and min operation by structure element \( g \) respectively:

\[
D_g(X) = y_1, y_2...y_n, \ where \ y_i = \max\{x_1-g,...,x_i+g\},
E_g(X) = y_1, y_2...y_n, \ where \ x_i = \min\{x_1-g,...,x_i+g\}.
\]  

Next, we can adapt the Opening and Closing operation:

\[
(f \circ g) = E_g(D_g(X)) \ , \ (f \bullet g) = D_g(E_g(X)).
\]

Lastly, Opening, Closing Pattern Spectrum are derived as:

\[
OPS^A_g = A(f \circ (\lambda - 1)g) - A(f \circ g),
OPS^A_g = A(f \bullet (\lambda - 1)g) - A(f \bullet g),
\]

where \( A(\cdot) \) denotes the area between signal and the zero axis. \( OPS^A_g \) is the Opening Pattern Spectrum, which reflects the change of area after doing \( \lambda \) times and \( \lambda - 1 \) times scaling operations with structure element \( g \). Likewise, the Closing Pattern Spectrum \( CPS^A_g \) could be derived in the similar way. In our study, we set the structure element \( g \) to 1,2,3,4,5, and \( \lambda \) to 1.

C. Entropy Domain Features

We extracted entropy features from two approaches: RCMPE (signal-channel), and MMPE (multi-channel).

1) Refined Composite Multi-Scale Permutation Entropy (RCMPE)

Permutation Entropy (PE) is conducive to evaluating the entropy of the time series [15]. The definition of PE is:

\[
PE(x) = - \sum_{j=1}^{m} p(p_j)\log p(p_j),
p(p_j) = \frac{\#\{i|0 \leq i \leq m, (x_i+1,...,x_i+m) has type p_j\}}{N-m+1}.
\]

Given an embedding dimension \( m \), \( \{p_1,...,p_j,...,p_m\} \) are \( m! \) permutation patterns, \( j \) is the pattern index and \( p(p_j) \) is the relative frequency of pattern \( p_j \). \( p_j \) represents different kinds of amplitude variation of the time series. The PE of a monotonic series would be 0. For a random series where the frequency of all patterns are the same, the value of PE would be \( \log m! \).

Adapted from PE, RCMPE [16] solves several issues in PE and makes more accurate measurement of signal permutation. The first step of RCMPE is coarse graining, after defining scaling factor \( \tau \), \( \tau \) coarse-grained series are generated from the original length \( N \) time series \( x \). We define the \( j \)-th index of \( k \)-th new series \( y_k^{(t)} = \{y_k^{(t)}(1), y_k^{(t)}(2),...,y_k^{(t)}(N)\} \) as:

\[
y_{k,j}^{(t)} = \frac{1}{\tau} \sum_{i=(j-1)\tau+k}^{j\tau+k-1} x_i, \ 1 \leq j \leq \frac{N}{\tau}, \ 1 \leq k \leq \tau.
\]
Eq. 9 is effective to reduce the time series and observe the signal in the larger scale. Next, RCMPE is computed as:

$$RCMPE(x, \tau, m) = - \sum_{q=1}^{m!} \bar{p}^q(\tau) \ln \bar{p}^q(\tau),$$

(10)

where $q$ is the permutation index. Considering all of the coarse-grained series $y_{k}^{r}$ together, $\bar{p}^q(\tau)$ represents the average relative frequency of the permutation pattern $\tau$. The averaging calculation recovers the statistical reliability which decreases in Eqn. 9. We set $\tau = 1, 3, 5, ..., 19$ and $m = 2, 3, 4, 5, 6$ respectively.

2) Multivariate Multi-Scale Permutation Entropy (MMPE)

Since there exist correlations between different EEG channels, we use MMPE [17] to analyze $p$-variate series. First we also conduct coarse-graining to generate new series $y_{i,j}^{r}$ from the selected EEG channels:

$$y_{i,j}^{r} = \frac{1}{\tau} \sum_{t=(j-1)\tau+1}^{j\tau} x_{i,t}, \quad 1 \leq i \leq p, \quad 1 \leq j \leq N / \tau,$$

(11)

where $i$ is channel index and $j$ is the index of new coarse-grained series. Second, we compute MMPE as:

$$MMPE(x, \tau, m) = - \sum_{q=1}^{m!} \bar{p}^q(\tau) \ln \bar{p}^q(\tau),$$

(12)

where $q$ is permutation pattern index. Likewise, in view of all the coarse-grained series $y_{i,j}^{r}$, $\bar{p}^q(\tau)$ represents the average relative frequency of the permutation pattern $\tau$. We selected 7 EEG channel groups: (AF5, AF6, AFz), (F1, F2), (FC1, FC2), (C3, C4, Cz), (CP1, CP2), (P1, P2), (O1, O2) from 7 scalp regions to calculate their MMPE respectively. We assigned $\tau = 1, 3, 5, ..., 19$ and $m = 2, 3, 4, 5, 6$.

D. Feature Selection and Classification

Due to the high dimension of the extracted features, we integrated Recursive Feature Elimination (RFE) into processing flow to eliminate feature redundancy. RFE is a recursive method which eliminates unimportant feature in each iteration based on a kernel estimator, and finds the optimal feature number to achieve the best result. Finally, Linear SVM was employed to identify different MW. We used leave-one-subject-out cross-validation in evaluation.

IV. EXPERIMENTAL RESULT

A. Statistical Analysis

We employed Analysis of Variance (ANOVA) to conduct statistical analysis for time-complexity and entropy features. Significant threshold was set to 0.05.

Significant time-complexity domain features are listed in Table II. In HOC, significant features distribute from order 0 to 9, suggesting that low and high order could capture complexity characteristics from different perspectives. For HFD, 6 out of 28 channels are significant, proving that fractal geometry is a good indicator of MW. As for PS, we deduce that when structure element $g$ is small, the effect of morphological operations would be too inappreciable, while too large $g$ might be too destructive for the original time series. Hence, the features are significant only when $g=2,3,4$.

Significant entropy domain features are listed in Table III. Both RCMPE and MMPE become significant when the scaling factor $\tau$ is greater than 15. This result implies that EEG permutation dynamic exists in larger temporal scales, and the coarse graining is necessary to extract EEG characteristics. The metrics to use "Multi-Scale" measurement mentioned in [16] [17] could be verified here.

Considering channel locations, we observe that the significant features in both complexity and entropy domain are generally located in F1, FC1, FC2, CP1, CP2, and most of them are symmetry. We also obtain significant MMPE features in (F1, F2), and (FC1, FC2) pairs, indicating that Frontal and Central might be the most informative regions for MWA, in agreement with previous researches [18]. Fig. 1 visualizes the number of significant features in each channel.

B. Classification Results

To measure the utility of time-complexity and entropy domain features, we conducted the experiment with three schemes based on the feature selection and classification procedure mentioned in section III-D. Traditional Scheme aggregated traditional features including time-statistic, frequency, and wavelet domains, while New Scheme consisted of time-complexity and entropy domain features. Lastly, we fused traditional features with time-complexity and entropy domain features in the Fusion Scheme.
V. Conclusion

This study evaluates the feasibility of time-complexity and entropy domain EEG features for MWA. Statistical analysis manifests the behavior of the features in terms of different parameters and channel locations. Classification results demonstrate that time-complexity and entropy features are capable to assess MW, complement traditional feature domains, and boost the performance.

REFERENCES


