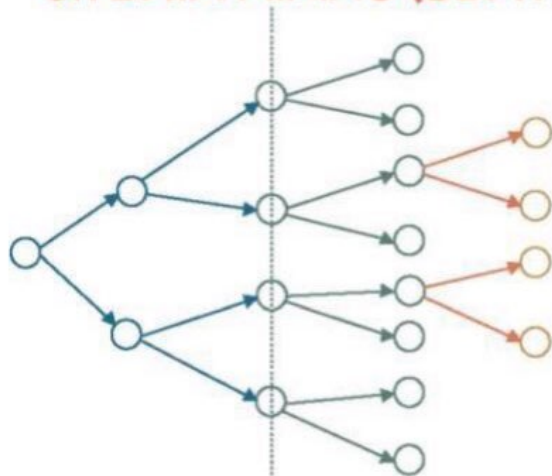## Society for Industrial and Applied Mathematics

**SIAM International Conference on DATA MINING (SDM18)**

May 3-5, 2018

San Diego Marriott Mission Valley, San Diego, California, USA

# Best Applied Data Science Paper Award

Presented to

Chenglong Dai, Jia Wu, Dechang Pi, Lin Cui

For

Brain EEG Time Series Selection: A Novel Graph-based Approach for Classification

*Martin Ester and Dino Pedreschi*

Program Chairs

# Brain EEG Time Series Selection: A Novel Graph-Based Approach for Classification

Chenglong Dai*        Jia Wu†        Dechang Pi*        Lin Cui *,‡

## Abstract

Brain Electroencephalography (EEG) classification is widely applied to analyze cerebral diseases in recent years. Unfortunately, invalid/noisy EEGs degrade the diagnosis performance and most previously developed methods ignore the necessity of EEG selection for classification. To this end, this paper proposes a novel maximum weight clique-based EEG selection approach, named mwcEEGs, to map EEG selection to searching maximum similarity-weighted cliques from an improved Fréchet distance-weighted undirected EEG graph simultaneously considering edge weights and vertex weights. Our mwcEEGs improves the classification performance by selecting intra-clique pairwise similar and inter-clique discriminative EEGs with similarity threshold $\delta$. Experimental results demonstrate the algorithm effectiveness compared with the state-of-the-art time series selection algorithms on real-world EEG datasets.

## 1 Introduction

In a noninvasive way, Brain Electroencephalography (EEG) classification is widely used to diagnose Alzheimer's disease [3], epileptic seizure [22], stroke [9], and so on. But in such cerebral disease diagnosis, invalid/noisy EEGs significantly affect the diagnosis accuracy, since the invalid/noisy EEGs degrade the distinction of target features. Invalid/noisy EEG is stimulated by the non-target brain activities, whose contour shape is dissimilar with those of target ones stimulated by the specific brain activities. Invalid EEGs are mainly from (1) the environmental noises which are always ignored when analyzed and (2) the non-target bioelectrical potentials. Actually, more invalid EEGs mix in raw EEGs of those patients suffering with epilepsy, Alzheimer's disease, stroke, amyotrophic lateral sclerosis (ALS), etc., due to the uncontrolled neural actions in their brain. To improve the diagnosis accuracy, independent component analysis (ICA) [11], principal component analysis (PCA) [10], common spatial pattern (CSP)[26], blind source separation (BSS) [25], and wavelet transform (WT) [17] mainly consider the artifact removal and they improve the accuracy in some extent, but they ignore the impact of invalid/noisy EEGs to the follow-
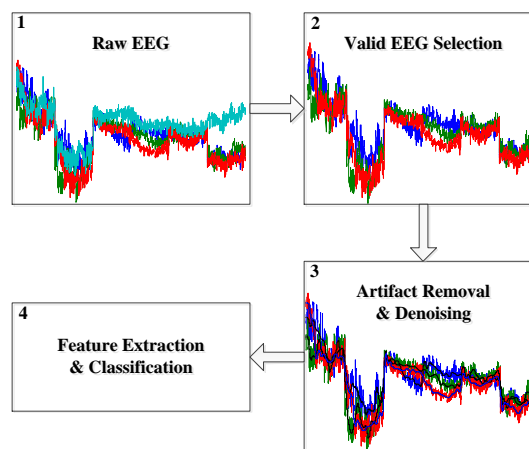


Figure 1: A simplified framework of EEG analysis. 1 illustrates 4 raw EEGs including one invalid EEG (light blue); 2 shows the EEG selection that the light blue one is selected out; 3 indicates artifact removal or denoising with selected EEGs, in which the black lines demonstrate the processed results; 4 displays the final EEG classification.

up analyses such as EEG artifact removal, denoising, feature extraction and classification. In another word, EEG selection is the most advance process for EEG analyses, especially for EEG classification, as Figure 1 illustrates. Furthermore, EEG selection aims to reduce the invalid/noisy EEGs stimulated by non-target brain activities.

EEG selection is a source control to reduce the degradation from the invalid ones. As far as we know, none of existing previous researches focused on EEG selection, and they jumped this step to artifact removal, feature extraction and classification, seeing Figure 1. We study EEG selection based on maximum weight clique in the work, providing more target EEGs for EEG classification. To the best of our knowledge, this is the first try to select EEGs with maximum weight clique for its further classification.

This paper presents a novel EEG selection. It aims to map EEG selection to searching maximum weight cliques in a similarity-weighted EEG graph in such a way that EEGs in the same clique are more similar to each other than to those in different cliques. This method simultaneously considers the weights of vertices and edges in the weighted EEG graph. Meanwhile, the proposed method focuses on the correlation between pairwise EEGs in the same class and scatter among different classes. Our contributions can be summarised as

---

*Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China. {chenglongdai, dc.pi, jsjxcuilin}@nuaa.edu.cn

†Department of Computing, Macquarie University, NSW 2109, Australia. wujiawb@gmail.com

‡Intelligent Information Processing Laboratory, Suzhou University, Suzhou 234000, China

follows:

• We present a novel method mwcEEGs for EEG selection. It maps EEG selection to searching maximum weight cliques from an similarity-weighted EEG graph, simultaneously considering the edge weights and the vertex weights.

• We demonstrate the superiority of mwcEEGs, with several popular and newest classifiers, over the state-of-the-art time series selection approaches through a detailed experimentation using standard classification validity criteria on the real-world EEG datasets.

The structure of this paper is as follows: In Section 2, we provide some backgrounds into similarity measure and maximum weight clique applied in this paper. In Section 3, we describe the proposed method including the EEG selection algorithm mwcEEGs and its detail description. In Section 4, we outline the datasets, criteria, and baselines to compare. The results and discussion are also presented in this section. Finally, we conclude the work in Section 5.

## 2 Preliminaries

This section introduces the backgrounds of similarity measure: Fréchet distance and maximum weight clique problem, which are the main two parts of the proposed method.

**2.1 Fréchet Distance** The Fréchet distance (FD), Hausdorff distance (HD), and dynamic time warping (DTW) are the most widely used similarity measures. FD takes into account the location and ordering of the points along the curves, which makes it a better similarity measure for EEG than HD and DTW. Since HD regards the EEG as arbitrary point sets, it ignores the order of points along the EEG. It is possible for two EEGs to have small HD but large FD [1]. DTW measures the distance between curves by warping the sequences in time dimension which ignores timing orders of points and degrades the synchronism of two curves. It sometimes generates unintuitive alignments and results in inferior results [13], since the DTW similarity measure is not essentially positive definite. Hence, the DTW does not reflect exact similarity of two EEGs because of its time warping [20]. Therefore, we applied Fréchet distance to be the similarity measure in our work.

Mathematically, for two EEGs $f, g \subseteq \mathbb{R}^2$ with continuous mapping $f, g : [0,1] \to \mathbb{R}^2$, the Fréchet distance $\delta_F(f,g)$ between $f$ and $g$ is defined as

$$(2.1) \qquad \delta_F(f,g) = \inf_{\alpha,\beta} \sup_{t \in [0,1]} \| f(\alpha(t)) - g(\beta(t)) \|$$

where $|| \cdot ||$ is the underlying norm, and $\alpha, \beta : [0,1] \to [0,1]$.

**2.2 Maximum weight Clique** Maximum weight clique problem (MWCP) is to search a complete subgraph (any two vertices are connected by an edge) with the maximum weights of vertices or edges from a weighted graph. Mathematically, given a weighted undirected graph $G =$

$(V, E, \eta, \mu)$, where $V$ and $E$ respectively denote vertex and edge of the graph; $\eta : V \to \{0\} \cup \mathbb{R}^+$ and $\mu : E \to \{0\} \cup \mathbb{R}^+$ are respectively the weights of them. $\sum_{v \in V} \eta_v + \sum_{e \in E} \mu_e$ is the weight of $G$. Define $\mathbb{N}_n = \{1, \cdots, n\}, n = |V|$ and $e_{ij} = \{i, j\} \in E$, the aim of MWCP is to search a clique $C$ with maximum weight from $G$, see (2.2).

$$(2.2) \qquad \sum_{i=1}^{n} \eta_i v_i + \sum_{1 \le i < j \le n} \mu_{ij} e_{ij} \to \max_{v_i, e_{ij} \in \{0,1\}}$$

Especially, when $\{i, j\} \in E_C, e_{ij} = 1$, otherwise $e_{ij} = 0$, the MWCP is transformed to maximum clique problem (MCP) which aims to search a complete subgraph of maximum cardinality. In the case, the MWCP is to maximize $\sum_{e_{ij} \in \{0,1\}, i,j \in V_C} \mu_e e_{ij}$.

## 3 The proposed method

The proposed method selects EEGs through searching maximum weight cliques in an improved Fréchet distance weighted EEG graph simultaneously considering the weights of vertices and edges.

**3.1 Weights of Edges** In this work, EEGs are regarded as pairwise connected vertices in an undirected weighted complete graph. The Fréchet distance-based similarities of EEGs are the weights on the edge, also called edge weight, that determines which edge is cut when partitioning complete subgraphs. However, the conventional Fréchet distance (CFD) ignores the temporal structure and it is sensitive to global trends [6]. To improve the CFD, local tendency is brought to evaluate the trend of EEGs. Mathematically, for two EEGs $tr_i = (a_1, a_2, \cdots, a_m)$ and $tr_j = (b_1, b_2, \cdots, b_n)$, $tr_i, tr_j \in \textbf{Trial}$, $\textbf{Trial}$ is the set of EEGs, the local trend of $tr_i, tr_j$ is evaluated by (3.3).

$$(3.3) \quad LocT(tr_i, tr_j) = \frac{\sum\limits_{t=1}^{p-q} (a_{t+q} - a_t)(b_{t+q} - b_t)}{\sqrt{\sum\limits_{t=1}^{p-q} (a_{t+q} - a_t)^2 \sum\limits_{t=1}^{p-q} (b_{t+q} - b_t)^2}}$$

where $1 \le q < p = \min\{m, n\}$. $q$ denotes the length of segment EEG. As $a_{t+q} - a_t$ indicates, a larger $q$ probably ignores more local tendencies than those of shorter segments whose length $\le q$. Hence, commonly $q = 1$ or downsampling $\max\{m, n\}$ into $p = \min\{m, n\}$; $LocT(tr_i, tr_j) \in [-1, 1]$ estimates the local tendency observed simultaneously on EEGs. This index indicates the synchronization of two EEGs in temporal structure.

With local and global measurements, the improved Fréchet distance (IFD) is calculated with (3.4) and (3.5).

$$(3.4) \quad \begin{aligned} s_{ij} = &\lambda \cdot norm(\delta_F(tr_i, tr_j)) \\ &+ (1-\lambda) \frac{1 - LocT(tr_i, tr_j)}{2} \end{aligned}$$

where

$$\delta_F(tr_i, tr_j) =$$
(3.5)
$$\inf_{\alpha,\beta:[0,1]\to[0,1]} \sup_{t\in[0,1]} S\Big(tr_i\big(\alpha(t)\big), tr_j\big(\beta(t)\big)\Big)$$

and $norm(\delta_F(tr_i, tr_j)) \in [0,1]$ is the normalized value of $\delta_F(tr_i, tr_j)$; $\lambda \in [0,1]$ is the weight for normalized global Fréchet distance while $1 - \lambda$ for local tendency.

All the similarities of $n$ pairwise connected EEGs calculated by IFD with local and global measurements construct the edge weight of the undirected weighted complete graph. Let $\boldsymbol{Trial}$ denote a $n \times n$ EEG matrix $\boldsymbol{Trial}^{n\times n} = (tr_1, tr_2, \cdots, tr_n)$ and the diagonal normalized similarity symmetric matrix $\boldsymbol{S}^{n\times n}$, the edge weights $\boldsymbol{\mu}$, is formed as:

(3.6)
$$\boldsymbol{\mu} = \boldsymbol{S}^{n\times n} = (\boldsymbol{s_1}, \boldsymbol{s_2}, \cdots, \boldsymbol{s_n})^T$$

where $s_{ji} = s_{ij}$; $s_{ij}$, $s_{ji} \in \boldsymbol{s_i}$, and $i$, $j$ of $s_{ij}$ denote $tr_i, tr_j \in \boldsymbol{Trial}, |\boldsymbol{Trial} = n|$.

**3.2 Weights of Vertices** Vertex weight of $G = (V, E, \eta, \mu)$ indicates the importance of vertex to the potential maximum similarity-weighted clique, i.e., it measures the importance of the vertex to the potential clique and also determines which vertices can be partitioned together into a same clique. For a set of EEGs $\boldsymbol{Trial}$ in $G = (V, E, \eta, \mu)$, their importance to the potential clique can be measured by the similarity partially ordered matrix $\boldsymbol{\eta}$, also called the vertex weights, such that $< \boldsymbol{\eta}^{n\times 1}, \preccurlyeq_\eta >= \{< \eta_i, \eta_j >| \eta_i \geq \eta_j; i \neq j; i, j \in \boldsymbol{Trial}\}$, where $\eta_i$ is computed by (3.7).

(3.7)
$$\eta_i = \frac{1}{|\boldsymbol{Trial}| - 1} \sum_{j\in \boldsymbol{Trial}\setminus i} s_{ij}$$

Specifically, $|\boldsymbol{Trial}| = 1$, $\eta = 0$. $\eta_i$ also represents the similarity rank of objective EEG $tr_i$ to the rest of ones $tr_j \in \boldsymbol{Trial} \setminus tr_i$. When $tr_i$ is partitioned into clique $C$, $tr_j \in \{tr_1, \cdots, tr_k\}$ with high rank similarity based on vertex weights $\boldsymbol{\eta}$ to $tr_i$ are correspondingly highly likely grouped into $C$. In detail, the vertex weight matrix $\boldsymbol{\eta}$ is formed as

(3.8)
$$\boldsymbol{\eta}^{n\times 1} = (\eta_1, \ldots, \eta_k, \ldots, \eta_n)^T$$

Simultaneously considering the edge weights $\boldsymbol{\mu}^{n\times n}$ and vertex weights $\boldsymbol{\eta}^{n\times 1}$, the pairwise high-weight EEGs with same label $L_i \in \boldsymbol{L}^{n\times 1}$ grouped into the same clique $C_i \in \boldsymbol{\mathcal{C}}$ with respect to similarity threshold $\delta$ can be represented by (3.9), with which the graph partition (vertices selection) can group most similar vertices with same labels into the same clique with a minimum weight loss of edge cut. In other word, based on (3.9), the proposed method repeats selecting the vertex with highest value of $\boldsymbol{C}_i = \boldsymbol{L}_i\boldsymbol{\eta}_i^T\boldsymbol{\mu}_i$ such that $\boldsymbol{\mu}_i \geq \delta_i$ in class $L_i$ as the one adding into the clique $C_i$ to form a new clique with larger weight, rather than randomly selecting one, which insures vertices with high importance can be grouped together into the same clique.

(3.9)
$$\boldsymbol{\mathcal{C}} = \boldsymbol{L}\boldsymbol{\eta}^T\boldsymbol{\mu}_{\geq\delta} = \cup_{i=1}^k \boldsymbol{C}_i = \cup_{i=1}^k \boldsymbol{L}_i\boldsymbol{\eta}_i^T\boldsymbol{\mu}_{i\geq\delta_i}$$

**3.3 The mwcEEGs** In this paper, EEG selection is mapped to multi-searching cliques with maximum weight based on $\boldsymbol{C} = \boldsymbol{L}\boldsymbol{\eta}^T\boldsymbol{\mu}_{\geq\delta}$ and similarity threshold $\delta$, which is named mwcEEGs. In detail, the mwcEEGs selects most similar EEGs with same label into the same clique and separates discriminative ones into different cliques with respect to $\delta$ and in this selecting the invalid/noisy EEGs are removed. That is to say, mwcEEGs not only selects most intra-clique similar EEGs as well as inter-clique discriminative ones, but also reduces the influence of noisy EEGs on the classifier.

Given a labeled weighted EEG graph $G_L = (V_L, E, \eta, \mu)$ with label matrix $\boldsymbol{L}^{n\times 1}$, where $\eta : V_L \to \{0\} \cup \mathbb{R}^+$ and $\mu : E \to \{0\} \cup \mathbb{R}^+$, and positive integers $N_1, \cdots, N_m$ such that $\sum_{i=1}^m N_i \leq n, n \leq |V_L|$, the mwcEEGs aims to select a family $\mathcal{C} = \{C_1, \cdots, C_m\}_{\geq\delta}$ of $m_{\geq 2}$ disjoint labeled cliques with maximum weight: $\sum_{C_k \geq \delta_k \in \mathcal{C}} w_{C_k}$ based on $\boldsymbol{C} = \boldsymbol{L}\boldsymbol{\eta}^T\boldsymbol{\mu}_{\geq\delta}$ with respect to the similarity threshold $\delta$. Given any edge $(i,j) \in E(C_k)$, $k \in \mathbb{N}_m$, and the weight of edge in $C_k$ such that $\mu_{ij} \in \{s_{ij}|s_{ij} \geq \delta_{C_k}\}$, the weight function simultaneously considering the weights of edges and vertices is modified as (3.10) defines.

(3.10)
$$w_{ij} = \begin{cases} \frac{\eta_i+\eta_j}{N_k-1} + \mu_{ij}, & \text{if } \mu_{ij} \geq \delta_{C_k}, i \neq j; \\ 0, & \text{otherwise.} \end{cases}$$

Where $N_k$ denotes the number of vertices of clique $C_k$. $N_k$ increases along with a new vertex $v_t$ such that $\forall v_i \in C_k, \mu_{ti} \geq \delta_{C_k}$ joins into $C_k$ and then the modified weighted $w_{ij}$ is correspondingly updated. Moreover, the similarity thresholds $\delta$ are crucial to EEG selection. Simply, similar vertices whose edge weights larger than corresponding $\delta_k$ are likely grouped into the same clique $C_k$. Namely, $\delta$ influences the EEG selection results. Furthermore, for searching $m$ cliques, it needs $m-1$ thresholds, since $m-1$ thresholds achieve $m-1$ cliques while remaining vertices are naturally regarded as a clique, then $m$ cliques are finally achieved.

PROPOSITION 3.1. *The mwcEEGs with the modified weight function (3.10) can be written equivalently as*

$$\mathcal{F}(\mathcal{C}) = \sum_{k=1}^m F(C_k) = \sum_{k=1}^m \sum_{\substack{e\in E(C_k), \delta_0=1 \\ \delta_k \leq \mu_e < \delta_{k-1}}} w_e \to \max_{\mathcal{C}\in\mathscr{C}}$$

*or*

$$\mathcal{F}(\mathcal{C}) = \frac{1}{2}\sum_{k=1}^m \sum_{i\in C_k} \sum_{\substack{j\in C_k, \delta_0=1 \\ \delta_k \leq \mu_{ij} < \delta_{k-1}}} w_{ij} \to \max_{\mathcal{C}\in\mathscr{C}}$$

*where* $\mathscr{C} = \Big\{\mathcal{C} = \{C_1, \cdots, C_m\} : C_i \cap C_j = \varnothing; |C_i| = N_i, |C_j| = N_j; \{i,j\} \in \mathbb{N}_m\Big\}$, *k denotes the EEG class index.*

*Proof.* Set $A_k = A(C_k) = \sum_{i \in C_k} \eta_i$, then

$$\mathcal{F}(\mathcal{C}) = \sum_{k=1}^{m} \sum_{\substack{e \in E(C_k), \delta_0=1 \\ \delta_k \le \mu_e < \delta_{k-1}}} w_e = \frac{1}{2} \sum_{k=1}^{m} \sum_{i \in C_k} \sum_{\substack{j \in C_k \setminus i, \delta_0=1 \\ \delta_k \le \mu_{ij} < \delta_{k-1}}} w_{ij}$$

$$= \frac{1}{2} \sum_{k=1}^{m} \sum_{i \in C_k} \sum_{\substack{j \in C_k \setminus i, \delta_0=1 \\ \delta_k \le \mu_{ij} < \delta_{k-1}}} \left( \frac{\eta_i + \eta_j}{N_k - 1} + \mu_{ij} \right)$$

$$= \frac{1}{2} \sum_{k=1}^{m} \sum_{i \in C_k} \Big( \sum_{j \in C_k \setminus i} \frac{\eta_i}{N_k - 1} + \sum_{j \in C_k \setminus i} \frac{\eta_j}{N_k - 1} + \sum_{\substack{j \in C_k \setminus i, \delta_0=1 \\ \delta_k \le \mu_{ij} < \delta_{k-1}}} \mu_{ij} \Big)$$

$$= \frac{1}{2} \sum_{k=1}^{m} \sum_{i \in C_k} \Big( \eta_i + \frac{A_k - \eta_i}{N_k - 1} + \sum_{\substack{j \in C_k \setminus i, \delta_0=1 \\ \delta_k \le \mu_{ij} < \delta_{k-1}}} \mu_{ij} \Big)$$

$$= \sum_{k=1}^{m} \Big( \frac{(N_k - 2)\eta_i + A_k}{2(N_k - 1)} + \frac{1}{2} \sum_{i \in C_k} \sum_{\substack{j \in C_k \setminus i, \delta_0=1 \\ \delta_k \le \mu_{ij} < \delta_{k-1}}} \mu_{ij} \Big)$$

$$= \sum_{k=1}^{m} \Big( \sum_{i \in C_k} \eta_i + \sum_{\substack{e \in E(C_k), \delta_0=1 \\ \delta_k \le \mu_e < \delta_{k-1}}} \mu_e \Big)$$

Proposition 3.1 indicates that searching the maximum similarity-weighted cliques based on the modified weight function is to maximize the total weight of vertices and edges satisfying the similarity thresholds. With the modified weight function simultaneously considering edge weights and vertex weights, the pseudo-code of mwcEEGs for EEG selection is shown in Algorithm 1. The mwcEEGs firstly sets $m - 1$ similarity thresholds, vertex weight matrix $\boldsymbol{\eta}$ with (3.4,3.5,3.6) and edge weight matrix $\boldsymbol{\mu}$ with (3.7,3.8) for initializing the labeled EEG graph $G_L = (V_L, E, \boldsymbol{\eta}, \boldsymbol{\mu})$, seeing line 1. Then it calculates the value of $\boldsymbol{C}_k$ and selects the vertex with the maximum value into the $k^{th}$ clique without randomly selecting one, see lines 4-5. Subsequently, mwcEEGs compares the total weight of the new clique $C_k \cup \{v_t\}$ with the old one $C_k$ to determine the new vertex $v_t$ joining to the clique or not, as lines 6-13 indicate. Additionally with lines 9-13, the matrix $\boldsymbol{C}_k$, $\boldsymbol{\eta}_k$ and $\boldsymbol{\mu}_k$ are updated based on the vertex adding. When the $k^{th}$ clique $C_k$ whose vertex set is $V_{C_k}$ is searched out, the $\boldsymbol{\eta}_k$ and $\boldsymbol{\mu}_k$ are modified correspondingly in lines 9-13, to calculate the weight matrix of the remaining vertices $V - V_{C_k}$ and their edges with same label $l_k$, and then the vertex with $l_k$ holding the largest value in $\boldsymbol{C}_k$ achieved by $\boldsymbol{\eta}_k \boldsymbol{\mu}_k$ is most likely chosen as the next vertex into the clique, to form a new clique $C_k$ with larger total weight.

As Algorithm 1 and Proposition 3.1 demonstrate, the algorithm selects labeled EEGs $v_i$ such that $\forall v_j \in C_k, \mu_{ij} \ge \delta_k$ and then trains the classifier model with such selected labeled EEGs. In other words, this process with respect to $\delta_k$ not only chooses the most distinguished labeled EEGs with

---

**Algorithm 1**    Graph-based EEG Selection.

**Input:**
>  $\boldsymbol{\delta}$: Set of $m_{\ge 1}$ similarity thresholds $\boldsymbol{\delta} = \{\delta_k | 0 \le \delta_k < \delta_{k-1}; \delta_0 = 1, k = 1, \cdots, m\}$;
>  $\boldsymbol{L}^{n \times 1}$: label matrix of $n$ EEGs;
>  $\boldsymbol{\eta}^{n \times 1}$, vertex weight matrix of $n$ EEGs;
>  $\boldsymbol{\mu}^{n \times n}$, edge weight matrix of $n$ EEGs;

**Output:**
>  $\mathcal{C}$: set of $m$ labeled cliques such that $\max_{\mathcal{C} \in \mathscr{C}} \mathcal{F}(\mathcal{C})$;

1: Initialize labeled EEG graph $G_L = (V_L, E, \boldsymbol{\eta}, \boldsymbol{\mu})$ and $\mathcal{C} = \varnothing$;
2: **repeat**
3:     $\delta_k \in \boldsymbol{\delta}, \boldsymbol{L}_k = \boldsymbol{L}$;
4:     $\boldsymbol{C}_k = \boldsymbol{L}_k \boldsymbol{\eta}_k^T \boldsymbol{\mu}_k$;
5:     $\{v_t$ labeled $l_k$ with maximum value in $\boldsymbol{C}_k\} \in V_L$;
6:     **if** $\forall \{v_n\} \in V_{C_k}, \delta_k \le \mu_{v_t v_n} < \delta_{k-1}$ **then**
7:         $W_{C_k \cup \{v_t\}} = \sum_{i,j \in V_{C_k} \cup \{v_t\}} w_{ij}$;
8:     **end if**
9:     **if** $W_{C_k \cup \{v_t\}} \ge W_{C_k}$ **then**
10:         $V_{C_k} = V_{C_k} \cup \{v_t\}$;
11:         $V_L = V_L \setminus \{v_t\}$;
12:         $W_{C_k} = W_{C_k \cup \{v_t\}}$;
13:     **end if**
14:     $\mathcal{C} = \mathcal{C} \cup C_k$;
15:     $\boldsymbol{L} = \boldsymbol{L} \setminus \boldsymbol{L}_k, \boldsymbol{\delta} = \boldsymbol{\delta} \setminus \{\delta_k\}$;
16:     $k = k + 1$;
17: **until** $\boldsymbol{\delta} = \varnothing$ or $\boldsymbol{L} = \varnothing$;

---

high similarity to train the classifier model, but also reduces the influence of invalid/noisy EEGs.

In the algorithm, a vertex $v_t$ labeled $l_k$ joining the clique $C_k$ if it simultaneously satisfies 2 conditions: (1) $\forall v_j \in C_k$, $\mu_{tj} \ge \delta_k$; (2) $\sum_{v_i, v_j \in C_k \cup \{v_t\}} w_{ij} \ge \sum_{v_i, v_j \in C_k} w_{ij}$. Actually, once $\delta_k$ is set, adding $v_t$ into $C_k$ just needs satisfying (1).

THEOREM 3.1. *A vertex $v_t$ labeled $l_k$ joins the clique $C_k$ to construct a new larger-weight clique $C_k^*$ if and only if $\forall v_i \in C_k, \mu_{it} \ge \delta_k$.*

*Proof.* Since $v_t \in C_k^*$, obviously $\forall v_i \in C_k, \mu_{it} \ge \delta_k$, and $l_k = l_{C_k}$. For a vertex $v_t$ labeled $l_k$ such that $\forall v_j \in C_k, \mu_{tj} \ge \delta_k$, according to Proposition 3.1: $\sum_{e \in E(C_k)} w_e = \sum_{v_i \in C_k} \eta_i + \sum_{e \in E(C_k)} \mu_e$, then

$$\sum_{e \in E(C_k \cup \{v_t\})} w_e - \sum_{e \in E(C_k)} w_e$$
$$= \sum_{v_i \in C_k \cup \{v_t\}} \eta_i + \sum_{e \in E(C_k \cup \{v_t\})} \mu_e - \sum_{v_i \in C_k} \eta_i - \sum_{e \in E(C_k)} \mu_e$$
$$= \sum_{v_i \in C_k} \eta_i + \eta_t + \sum_{e \in E(C_k)} \mu_e + \sum_{v_i \in C_k} \mu_{it}$$
$$- \sum_{v_i \in C_k} \eta_i - \sum_{e \in E(C_k)} \mu_e = \eta_t + \sum_{v_i \in C_k} \mu_{it} > 0$$

Namely, $v_t$ joining $C_k$ increases the weight, therefore $v_t$ and $\forall v_i \in C_k$ construct a new clique $C_k^* = C_k \cup \{v_t\}$ with a

larger weight than that of $C_k$.

Based on Theorem 3.1, once $\delta_{k=1,\cdots,m}$ is set, searching $m$ labeled cliques such that $\max_{\mathcal{C}\in\mathscr{C}}$ can be transformed to $m$-search labeled maximum similarity-weighted cliques, namely $m$-repeating mwcEEGs with $\delta_k$, $k = 1,\cdots,m$.

PROPOSITION 3.2. *The mwcEEGs for $m$ labeled cliques selection can be equivalently transformed to $m$-time repeating mwcEEGs with $0 \leq \delta_k < \delta_{k-1}, \delta_0 = 1$.*

*Proof.* Recall Proposition 3.1. The mwcEEGs, can be written as $\mathcal{F}(\mathcal{C}) = \sum\limits_{k=1}^{m} \sum\limits_{\substack{e\in E(C_k),\delta_0=1 \\ \delta_k \leq \mu_e < \delta_{k-1}}} w_e \to \max\limits_{\mathcal{C}\in\mathscr{C}}$, then

$$\mathcal{F}(\mathcal{C}) = \sum_{k=1}^{m} \sum_{\substack{e\in E(C_k),\delta_0=1 \\ \delta_k \leq \mu_e < \delta_{k-1}}} w_e \to \max_{\mathcal{C}\in\mathscr{C}}$$

$$= \sum_{k=1}^{m} \Big( \sum_{i\in C_k} \eta_i + \sum_{\substack{e\in E(C_k),\delta_0=1 \\ \delta_k \leq \mu_e < \delta_{k-1}}} \mu_e \Big) \to \max_{\mathcal{C}\in\mathscr{C}}$$

$$= \Big( \big(\sum_{i\in C_1} \eta_i + \sum_{e\in E(C_1),\delta_1\leq\mu_e<1} \mu_e\big) + \cdots$$
$$+ \big( \sum_{i\in\mathcal{C}\backslash C_1\backslash\cdots\backslash C_{m-1}} \eta_i + \sum_{\substack{e\in E(\mathcal{C}\backslash C_1\backslash\cdots\backslash C_{m-1}) \\ \delta_m \leq \mu_e < \delta_{m-1}}} \mu_e\big) \Big) \to \max_{\mathcal{C}\in\mathscr{C}}$$

$$\because e \in E(C_k), 0 \leq \delta_k \leq e < \delta_{k-1}, \delta_0 = 1; k = 1,\cdots,m$$

$$= \Big( \big(\sum_{i\in C_1} \eta_i + \sum_{e\in E(C_1),\delta_1\leq\mu_e<1} \mu_e\big) \to \max_{C_1\in\mathcal{C}} \Big) + \cdots$$
$$+ \Big( \big( \sum_{i\in\mathcal{C}\backslash C_1\backslash\cdots\backslash C_{m-1}} \eta_i + \sum_{\substack{e\in E(\mathcal{C}\backslash C_1\backslash\cdots\backslash C_{m-1}) \\ \delta_m \leq \mu_e < \delta_{m-1}}} \mu_e\big) \to \max_{C_m\in\mathcal{C}} \Big)$$

$$= \sum_{k=1}^{m} \Big\{ \big(\sum_{i\in C_k} \eta_i + \sum_{\substack{e\in E(C_k),\delta_0=1 \\ \delta_k \leq \mu_e < \delta_{k-1}}} \mu_e\big) \to \max_{C_m\in\mathcal{C}} \Big\}$$

According to Proposition 3.2 and Theorem 3.1, the mwcEEGs is transformed by $m$-searching labeled maximum cliques, as Algorithm 2 shows, in which any algorithm for maximum clique problem (MCP) [27] can be applied to search the cliques with maximum weight in the given graph. Importantly, in every iteration to search the maximum clique, the vertex weights ($\eta_v \in \boldsymbol{\eta}_{V_k}$) of $v \in V_k$ will be ranked in a descending order and the weight matrix ($\boldsymbol{C}_k = \boldsymbol{L}_k\boldsymbol{\eta}_k^T\boldsymbol{\mu}_k$) is calculated in line 8, so that the algorithm can choose vertex with highest weight into the potential clique $C_k$. This procedure contributes high-quality selection and fast searching maximum clique. A higher $\eta_i \in \boldsymbol{\eta}$ indicates the vertex $v_i$ has higher similarity with all the other vertices. That is, vertices with higher $\eta_i$ are likely grouped into the same clique. Meanwhile, selecting the vertex with largest value in $\boldsymbol{C}_k$ as the new vertex adding to the potential clique also reduces the time consumption compared with the conventional methods that randomly select vertices.

---

**Algorithm 2**   mwcEEGs via Maximum Weight Cliques

**Input:**
  $\boldsymbol{\delta}$: Set of $m_{\geq 1}$ similarity thresholds $\boldsymbol{\delta} = \{\delta_k | 0 \leq \delta_k < \delta_{k-1}; \delta_0 = 1, k = 1, \cdots, m\}$;
  $\boldsymbol{L}^{n\times 1}$: label matrix of $n$ EEG;
  $\boldsymbol{\eta}^{n\times 1}$, vertex weight matrix of $n$ EEG;
  $\boldsymbol{\mu}^{n\times n}$, edge weight matrix of $n$ EEG;
**Output:**
  $\mathcal{C}$: set of $m$ labeled cliques such that $\max\limits_{\mathcal{C}\in\mathscr{C}} \mathcal{F}(\mathcal{C})$;
1: Initialize labeled EEG graph $_LG = (V_L, E, \eta, \mu)$ and $\mathcal{C} = \varnothing$;
2: **repeat**
3:   $G_k = G: V_k = V_L, E_k = E$;
4:   $\delta_k \in \boldsymbol{\delta}, \boldsymbol{L}_k = \boldsymbol{L}$;
5:   **if** $\mu_{ij} < \delta_k$ **then**
6:     $E_k = E_k \setminus \{e_{ij}\}$;
7:   **end if**
8:   Search $C_k = (V_C, E_C)$ such that $\max |V_C|$ from $G_k = (V_k, E_k, \eta, \mu)$ with any MCP algorithm, improved with descending sorting ($\eta_v \in \boldsymbol{\eta}_{V_k}$) of $v \in V_k$ and $C_k = \boldsymbol{L}_k\boldsymbol{\eta}_k^T\boldsymbol{\mu}_k$;
9:   $\mathcal{C} = \mathcal{C} \cup C_k$;
10:  $G_M = (V_M, E_M, \eta, \mu)$:
       $V_M = V_L - V_C = \{i \mid i \in V_L; i \notin V_C\}$,
       $E_M = \{e_{ij} \mid i, j \in V_L; i, j \notin V_C; i \neq j\}$;
11:  $G = G_M$;
12:  $\boldsymbol{L} = \boldsymbol{L} \setminus \boldsymbol{L}_k, \boldsymbol{\delta} = \boldsymbol{\delta} \setminus \{\delta_k\}$;
13:  $k = k + 1$;
14: **until** $\boldsymbol{\delta} = \varnothing$ or $\boldsymbol{L} = \varnothing$;

---

Table 1: Datasets

| EEG Cases | Datasets | Training:Testing | # of Classes |
|---|---|---|---|
| EEG Case 1 | Traindata_0, Ia + Traindata_1, Ia | 180 : 88 (268) | 2 |
| EEG Case 2 | Traindata_0, Ib + Traindata_1, Ib | 134 : 66 (200) | 2 |
| EEG Case 3 | Traindata_0 + Traindata_1, Ia Traindata_0 + Traindata_1, Ib | 313 : 156 (468) | 4 |

## 4  Experiments

**4.1  Datasets**  The EEG data we experiment with are slow cortical potentials (SCPs [1]) provided by Institute of Medical Psychology and Behavioral Neurobiology from University of Tübingen. In detail, Dataset Ia with 135 EEG trials labeled '0' (Traindata_0, Ia) and 133 labeled '1' (Traindata_1, Ia) are taken from a healthy subject (HS). Dateset Ib with 100 EEG trials labeled '0' (Traindata_0, Ib) and 100 labeled '1' (Traindata_1, Ib) are taken from an amyotrophic lateral sclerosis subject (ALS). 3 cases of experiments are set up in Table 1. In this paper, we apply Hold-out strategy [14] to evaluate methods. The data sets are divided into two parts: training data and testing data with the proportion of 2:1. Furthermore, the Hold-out strategy is applied 3 times to produce 3 groups of training and testing data for the methods.

---

[1]The data set is publicly available as online archives at http://www.bbci.de/competition/ii/.

**4.2 Evaluation Methodology** The following criteria, which have been well used in data mining area [5, 8, 19], are selected to evaluate the proposed method.

• **Rand index** (*RI*) [21] estimates the quality of classification with respect to the right classes of the data. It measures the percentage of right decisions made by the method. In detail, $RI = \frac{TP+TN}{TP+TN+FP+FN}$, where *TP*, *FP*, *TN*, and *FN* respectively denotes the number of true positives, false positives, true negatives, and false negatives.

• **F-score** [24] weighs FP and FN in RI unequally through weighting a parameter $\beta \geq 0$ on *recall*, commonly $\beta = 1$. Mathematically, *F-score*$= \frac{(1+\beta^2)pr}{\beta^2 p+r}$, where *precision*: $p = \frac{TP}{TP+FP}$ and *recall*: $r = \frac{TP}{TP+FN}$.

• **Fleiss' kappa** (*κ*) [7] is a statistical measure for assessing the coherence of decision ratings among classes. Mathematically, $\kappa = \frac{\overline{P}-\overline{P_e}}{1-\overline{P_e}}$, where $\overline{P} - \overline{P_e}$ denotes the degree of agreement actually achieved over chance and $1 - \overline{P_e}$ denotes the degree of agreement attainable above chance. Meanwhile, $\overline{P} = \frac{1}{Nn(n-1)}(\sum_{i=1}^{N}\sum_{j=1}^{k} n_{ij}^2 - Nn)$, $\overline{P_e} = \sum_{j=1}^{k}(\frac{1}{Nn}\sum_{i=1}^{N} n_{ij})^2$, and $N$ denotes the number of subjects, $n$ the number of ratings per subject, $k$ the number of classes into which assignment are made.

**4.3 Baselines** We compared mwcEEGs with the state-of-the-art EEG time series selection methods, as follows.

• **lwEEGs** [28]: Local weighted EEG time series selection computes time series centroid in each class and selects $k$ nearest [12] time series from the same labeles to the corresponding centroid as the training time series for classifiers.

• **gwEEGs** [28]: Global weighted EEG time series selection computes the centroid of all labeled time series and selects $k$ closest ones from all classes to the centroid time series as the training ones for classifiers.

• **lrtEEGs** [29]: Local recursion testing EEG time series selection recursively selects $m$ nearest time series to every testing one with same label and chooses the $k \leq m$ most nearest ones from $m$ selected time series in each class as the training data for classifiers, which focuses on the local correlation between time series to the testing ones.

• **grtEEGs** [29]: Global recursion testing EEG time series selection recursively selects $m$ nearest time series to each testing time series without considering the class labels and then chooses the $k \leq m$ most similar ones as the training time series for classifiers. It considers the global similarity between all the time series to the entire testing time series.

Meanwhile, in order to evaluate all the methods to select EEGs, we apply most popular and newest classifiers to classify EEGs with EEG selection methods. The applied classifiers are introduced below, which mainly include most widely applied SVM, shapelet-based, ensemble-based, and structure-based classifiers.

• **SVM**: We apply LIBSVM [4] in this section as one of the baselines to classify EEG data. With LIBSVM, the kernel width and $C$ of SVM are respectively tuned as $\gamma \in \{10^{-3}, 10^{-2}, \ldots, 10^{1}\}$ and $C \in \{10^{-2}, 10^{-1}, \ldots, 10^{1}, 10^{2}\}$.

• **st-TSC** [16]: Shapelet transform-based method time series classifier extracts $k$ discriminative subsequences that best distinguish time series in different classes and uses an optimization formulation to search for fixed length time series subsequences that best predict the target variable by calculating the distances from a series to each shapelet.

• **RPCD** [23]: Recurrence patterns compression distance time series classifier uses recurrence plots as representation domain for time series classification via applying Campana-Keogh distance to estimate similarity.

• **COTE** [2]: An ensemble-based classifier classifies time series by applying a heterogeneous ensemble onto transformed representations. A flat collective of transform based ensembles (COTE) fuses various classifiers into a single one, which includes whole time series classifiers, shapelet classifiers, and spectral classifiers.

• **SAX-SEQL** [18]: An efficient linear classifier learns long discrete discriminative subsequences from time series by exploiting all-subsequences space based on symbolic aggregate approximation (SAX) [15] which smooths and compresses time series into discrete representations.

**4.4 Parameter setting** For mwcEEGs, $\delta$ determines the selection of EEGs and affects the classification performance. The mwcEEGs with a smaller $\delta$ selects less discriminative but more general EEGs while mwcEEGs with a larger $\delta$ selects more discriminative but less general EEGs. Then the classification with classifiers are influenced by the selected EEGs, namely, the final classification results are affected by $\delta$. And more distinguished EEGs with a large $\delta$ degrades the classification performance of classifiers because of low generality of selected discriminative EEGs which cannot represent the general EEG data. As a consequence, the mwcEEGs with a moderate or optimal $\delta$ selects a better amount of EEGs which balances the discrimination and generality, and seems to achieve better classification performance. Here we select the optimal $\delta$ based on the discrete probability distribution of similarities among EEG. Mathematically, the discrete probability distribution $d(\delta)$ is defined as

$$(4.11) \qquad d(\delta) = \frac{n_{s\in[\delta_i,\delta_i+Interval]}}{N^2}$$

where $N$ denotes the number of vertices (EEGs) in the similarity-weighted graph $G$, $n$ denotes the number of similarities $s \in \mathbf{S}$ locate in $[\delta_i, \delta_i + Interval]$.

With (4.11), $\delta$ is chosen based on the (4.12)

$$(4.12) \quad \delta = f^{-1}(D) \in \left\{ \delta | f(\delta) : \sum_{i\in[0,\frac{1}{Interval}-1]}^{\frac{1}{Interval}} d_i(\delta) = A \right\}$$

where $A \in (0,1)$ and a larger $A$ achieves a smaller $\delta$.

Similarity threshold $\delta_{k=1,\cdots,m,m\geq1}$ in the work for EEG selection with 3 datasets is respectively set based on the

discrete probability distribution shown in (a) – (e) of Figure 2. (a) – (e) respectively shows that most EEGs from the same dataset are similar with each other and the EEG similarity is displayed in (f).

To illustrate the influence of $\delta$ on mwcEEGs, we set $\delta \in \{0.3, 0.35, 0.4, 0.45, 0.5, 0.55\}$ for 3 EEG Cases. The number of selected EEGs for lwEEGs, gwEEGs, lrtEEGs, and grtEEGs is set as same as mwcEEGs. For classifiers SVM, st-TSC, RPCD, COTE, and SAX-SEQL, we set the optimal parameters as same as the references set. As introduced before, data set in 3 cases is divided into 3 groups of training and testing data with the proportion of 2:1 based on the Hold-out strategy. All the methods are run on 3 groups of EEG data for each case and the results are averaged.

**4.5 Experimental Results and Discussion** In this paper, we proposed the maximum weight clique inspired method mwcEEGs to select EEGs. To firmly establish the efficacy of our method, we compared the mwcEEGs with the state-of-the-art time series selection methods on several popular and newest time series classifiers for EEG classification. The experimental results with 3 cases (3 datasets) are shown in Figure 3 – 5 respectively. We can see from the experimental results that with selected EEGs by mwcEEGs, the classification performance is improved compared that without selected EEGs. Moreover, a small or a moderate $\delta$ yields a better classification performance than a larger $\delta$. In other word, a small or moderate $\delta$ calculated with (4.12) achieves a high-quality EEG classification. The reason is that the larger the $\delta$ is, more discriminative EEGs are selected by mwcEEGs. That is to say, the selected discriminative EEGs with a larger $\delta$ reduce more discriminative features of EEGs and cannot represent the general EEGs, so its classification results are probably lower than that with a smaller $\delta$ achieves or even lower than that without selected ones. As a consequence, mwcEEGs with small or moderate $\delta$ yields best classification results with respect to RI, F-score, and *kappa*, which indicates that mwcEEGs is superior over the state-of-the-art time series selection methods for EEG classification on several promising classifiers.

## 5 Conclusion

This paper explores Brain EEG selection. Raw EEGs without selection contains many invalid/noisy data which degrades the corresponding learning performance. Since EEG is weak, complicated, fluctuated and with low signal-to-noise, conventional time series selection methods are not applicable for EEG selection. To address this issue, a novel approach (called mwcEEGs) based on maximum weight clique is proposed to select valid EEGs. The main idea of mwcEEGs is to map EEG selection to searching a family of cliques with maximum weights simultaneously combining edge weights and vertex weights in an improved Fréchet
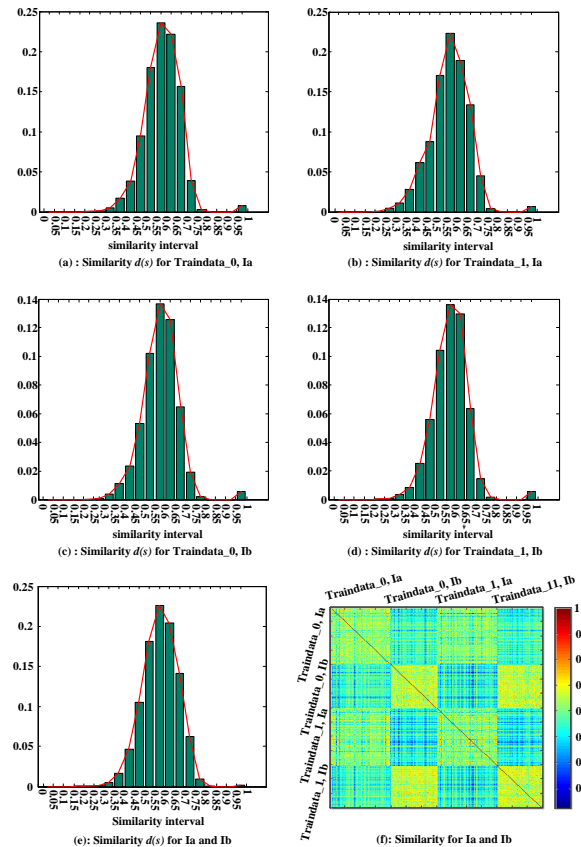


Figure 2: Similarity distribution $d(s)$ of 3 EEG datasets.

distance-weighted EEG graph while reducing the influence of invalid/noisy EEGs according to similarity thresholds $\delta$. The experimental comparisons with the state-of-the-art time series selection methods based on different evaluation criteria on real-world EEG data demonstrate the effectiveness of the mwcEEGs for EEG selection.

## Acknowledgments

## References

[1] P. K. Agarwal, R. B. Avraham, H. Kaplan, and M. Sharir, *Computing the discrete Fréchet distance in subquadratic time*, SIAM journal on computing, 43 (2014), pp. 429–449.

[2] A. Bagnall, J. Lines, J. Hills, and A. Bostrom, *Time-series classification with COTE: The collective of transformation-based ensembles*, IEEE trans. knowledge and data engineering, 27 (2015), pp. 2522–2535.
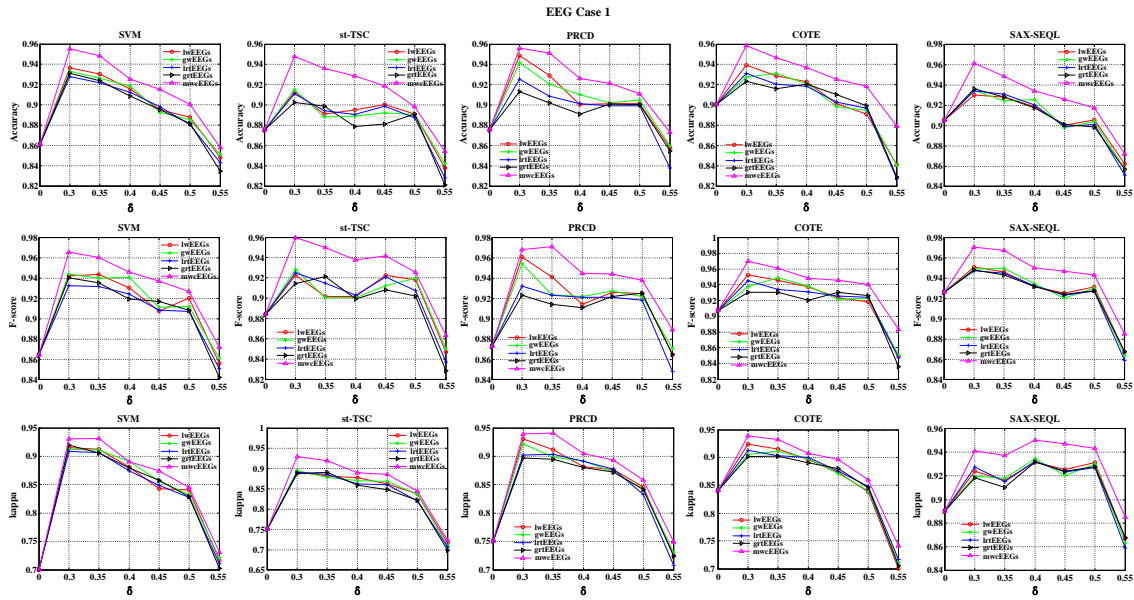
Figure 3: Classification results of 5 classifiers with selected EEGs by different EEG selection methods on EEG Case 1.
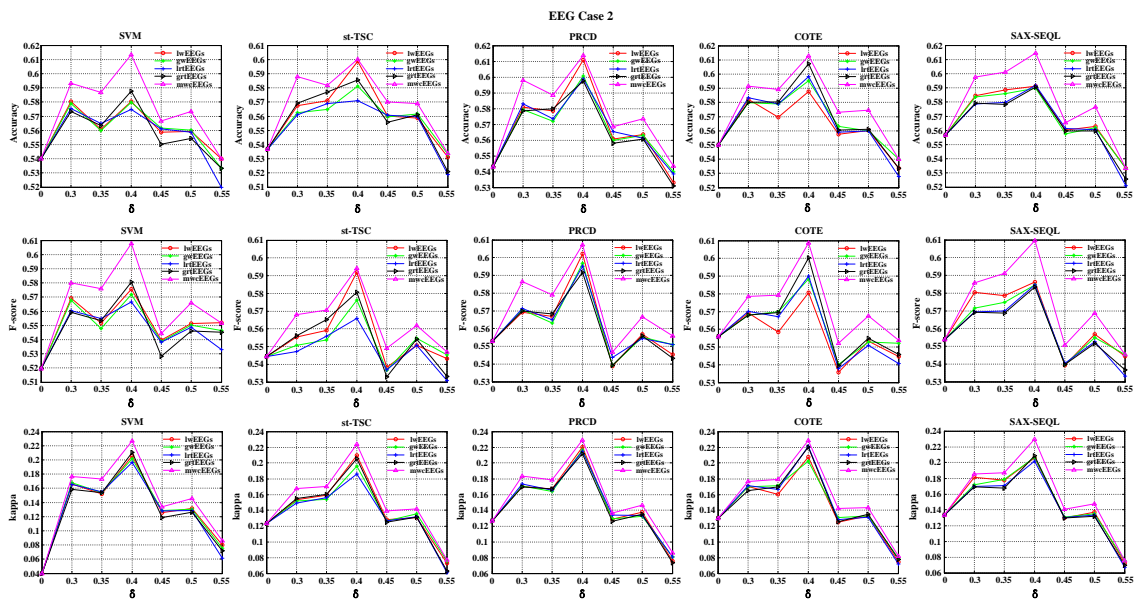


Figure 4: Classification results of 5 classifiers with selected EEGs by different EEG selection methods on EEG Case 2.

[3] E. Barzegaran, B. van Damme, R. A. Meuli, and M. G. Knyazeva, *Perception-related EEG is more sensitive to Alzheimer's disease effects than resting EEG*, Neurobiology of aging, 43 (2016), pp. 129-139.

[4] C. C. Chang and C. J. Lin, *LIBSVM: A library for support vector machines*, ACM trans. intelligent systems and technology, 2 (2011), pp. 1–27.

[5] J. Wu, Z. Cai and X. Zhu, *Self-adaptive probability estimation for Naive Bayes classification*, In 2013 International Joint Conference on Neural Networks (IJCNN), Dallas, TX, 2013, pp. 1–8.

[6] A. Chouakria-Douzal and P. N. Nagabhushan, *Improved Fréchet distance for time series*. In Proceedings of 10th Jubilee Conference of the International Federation of Classification Societies, (2006), pp. 13–20.

[7] J. L. Fleiss, *Measuring nominal scale agreement among many raters*, Psychological bulletin, 76 (1971), pp. 378–382.

[8] J. Wu, S. Pan, X. Zhu, C. Zhang and P. S. Yu, *Multiple Structure-View Learning for Graph Classification*, IEEE Transactions on Neural Networks and Learning Systems, PP (2017), pp. 1-16.

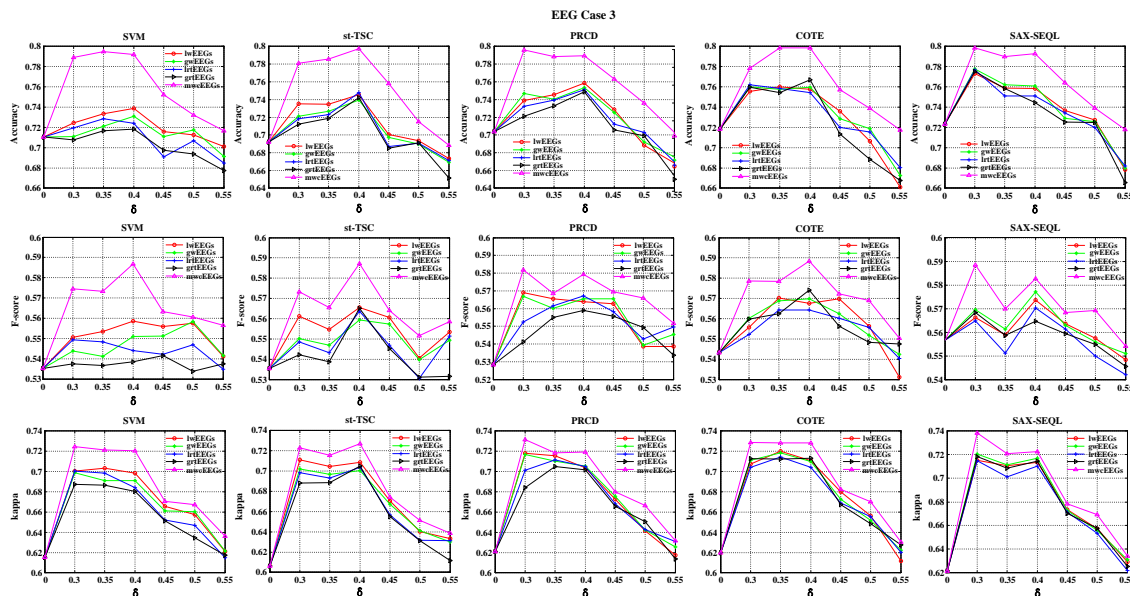[9] B. Hordacre, N. C. Rogasch, and M. R. Goldsworthy, *Com-*

Figure 5: Classification results of 5 classifiers with selected EEGs by different EEG selection methods on EEG Case 3.

*mentary: utility of EEG measures of brain function in patients with acute stroke*, Frontiers in human neuroscience, 10 (2016), article no. 621.

[10] R. Jiang, H. L. Fei, and J. Huan, *A family of joints sparse PCA algorithms for anomaly localization in network data streams*, IEEE trans. knowledge and data engineering, 25 (2013), pp. 2421–2433.

[11] Y. Jonmohamadi, G. Poudel, C. Innes, and R. Jones, *Source-space ICA for EEG source separation, localization, and time-course reconstruction*, Neuroimage, 101 (2014), pp. 720–737.

[12] J. Wu, Z. Cai and Z. Gao, *Dynamic K-Nearest-Neighbor with Distance and attribute weighted for classification*, In 2010 International Conference on Electronics and Information Engineering, 2010, pp. 356–60.

[13] E. Keogh and M. J. Pazzani, *Derivative dynamic time warping*, In SDM, 2001, pp. 1–11.

[14] J. Kim, *Estimating classification error rate: Repeated cross-validation, repeated hold-out and bootstrap*, Computational statistics and data analysis, 53 (2009), pp. 3735–3745.

[15] J. Lin, E. Keogh, L. Wei, and S. Lonardi, *Experiencing sax: a novel symbolic representation of time series*, Data mining and knowledge discovery, 15 (2007), pp. 107–144.

[16] J. Lines, L. M. Davis, J. Hills, and A. Bagnall, *A shapelet transform for time series classification*, In ACM SIGKDD, 2012, pp. 289–297.

[17] Y. Liu, W. Zhou, Q. Yuan, and S. Chen, *Automatic seizure detection using wavelet transform and SVM in long-term intracranial EEG*, IEEE trans. neural systems and rehabilitation engineering, 20 (2012), pp. 749–755.

[18] T. L. Nguyen, S. Gsponer, and G. Ifrim, *Time series classification by sequence learning in all-subsequence space*, In ICDE, 2017, pp. 947–958.

[19] J. Wu, X. Zhu, C. Zhang, and Z. Cai, *Multi-instance multi-graph dual embedding learning*, In ICDM, 2013, pp. 827–836.

[20] F. Petitjean, A. Ketterlin, and P. Gancarski, *A global averaging method for dynamic time warping, with applications to clustering*, Pattern Recognition, 44 (2011), pp. 678–693.

[21] W. Rand, *Objective criteria for the evaluation of clustering methods*, Journal of the American statistical association, 66 (1971), pp. 846–850.

[22] K. Samiee, P. Kovacs, and M. Gabbouj, *Epileptic seizure classification of EEG time-series using rational discrete short-time Fourier transform*, IEEE transactions on biomedical engineering, 62 (2015), pp. 541–552.

[23] D. F. Silva, V. M. A. de Souza, and G. E. A. P. A. Batista, *Time series classification using compression distance of recurrence plots*, In ICDM, 2013, pp. 687–696.

[24] C. J. Van Rijsbergen, *Information Retrieval*, 2nd ed., London: Butterworths, 1979.

[25] R. R. Vázquez, H. Vélez-Pérez, R. Ranta, V. L. Dorr, D. Maquin, and L. Maillard, *Blind source separation, wavelet denoising and discriminant analysis for EEG artefacts and noise cancelling*, Biomedical signal processing and control, 7 (2012), pp. 389–400.

[26] W. Wu, Z. Chen, X. Gao, Y. Li, E. N. Brown, and S. Gao, *Probabilistic common spatial patterns for multichannel EEG analysis*, IEEE trans. pattern analysis and machine intelligence, 37 (2015), pp. 639–653.

[27] Q. H. Wu and J. Hao, *A review on algorithms for maximum clique problems*, European journal of operational research, 242 (2015), pp. 693–709.

[28] J. Wu, S. Pan, X. Zhu, and C. Zhang, *Dual instance and attribute weighting for naive bayes classification*, In IJCNN, 2014, pp. 1675–1679.

[29] L. Jiang, *Learning Instance Weighted Naive Bayes from labeled and unlabeled data*, Journal of Intelligent Information Systems, 38 (2012), pp. 257-268.