Electroencephalogram Signal Clustering with Convex Cooperative Games

Chenglong Dai, Jia Wu, Dechang Pi, Lin Cui, Blake Johnson, and Stefanie I. Becker

***Abstract*—Currently, electroencephalogram (EEG) is mostly analyzed in a supervised way, which requires EEG labels (e.g., EEG classification). With the ever-increasing amount of unlabeled/mislabeled EEG in neuropsychiatric disorder diagnosis, BCI, and rehabilitation, manually labeling of EEG data is a labor intensive and time-consuming process, and few labs have developed algorithms to analyze EEG in an unsupervised manner (i.e., EEG clustering). In this paper, we propose a cooperative game inspired approach to cluster multi-trial EEG data. The idea is to map multi-trial EEG clustering to the coalition formation in a cooperative game, and then identify cluster center (the EEG trial with highest Shapley value) and assign EEG trials into proper clusters based on their cross correlation-transformed Shapley values. We demonstrate the mapped EEG cooperative game is convex, and it leads to an algorithm for multi-trial EEG clustering named CoGEEGc. The CoGEEGc yields high-quality multi-trial EEG clustering with respect to intra-cluster compactness and inter-cluster scatter. We show that CoGEEGc outperforms 15 state-of-the-art EEG or time series clustering approaches through detailed experimentation on real-world multi-trial EEG datasets. Comparison against 15 methods with four theoretical properties of clustering further illustrates the superiority of CoGEEGc, as it satisfies two properties while other approaches only satisfy one.**

***Index Terms*—EEG clustering, convex cooperative game, Shapley value, modified cross correlation.**

I. Introduction

ELECTROENCEPHALOGRAM (EEG) is one specific type of weak, complex, non-stationary, and low signal-to-noise ratio electrical potentials generated by cerebral cortex. As it can reflect human brain functions and body status, EEG is widely analyzed in a supervised way to help diagnose neuropsychiatric disorders such as Alzheimer's disease (AD) [1], stroke [2], epileptic seizures [3], [4] and amyotrophic lateral sclerosis (ALS) [5]. EEG is also applied to improve the quality of life of disabled people, by supporting rehabilitation or using motor imagery EEG signals [6] to control wheelchairs [7] or robot arms [8] with brain-computer interfaces (BCIs) [9], [10]. Moreover, these applications are mainly based on labeled EEG classification. Although EEG research has made significant progress in supervised applications, the majority of studies have focused on single-trial [11], [12], single-channel [2], or multi-channel EEG [13], [14], and seldom consider correlations among EEG trials. Multi-trial EEG [15], as its name indicates, consists of many EEG trials based on multiple channels that are recorded across several sessions during several days, which contains more complete information with potential varieties even for the same cerebral activity of same subject. Further, with the increasing number of multi-trial EEG studies, lack of labels or mislabeling in practical applications has become a serious limitation for supervised technique-based solutions like classifications [12], [13], [16], [17] that require EEG labels. To address the brand new problem in EEG study, we investigate unlabeled multi-trial EEG clustering and propose a novel solution to this problem in the paper.

A. Motivation

With the ever-increasing amount of unlabeled multi-trial EEG data in neuropsychiatric disorder diagnosis, control, and rehabilitation, the traditional supervised analysis methods, such as classification, are not applicable, and manually labeling EEG is a time-consuming and labor intensive process. Therefore, EEG clustering becomes a promising approach to analyze unlabeled EEG signals. With clustering, interesting patterns and correlations among unlabeled EEG trials can be identified and then help to pre-diagnose cerebral diseases through correlation comparison between new unlabeled EEG trials and those in clusters. Unfortunately, multi-trial EEG clustering is rarely reported in existing studies [18]. EEG can be regarded as one special type of time series and many time series-based clustering methods have been proposed and performed well on traditional time series [19], but these methods may be inapplicable for multi-trial EEG clustering due to EEG's specific characteristics of high dimension, complexity, non-stationary, and low signal-to-noise ratio.

As Michalski [20] pointed out, most current clustering algorithms just consider the distance of data to a hypothetical centroid, which does not completely reflect the intrinsic notion of clustering and underrates the importance of other data in the same cluster. This kind of approaches can characterize the distance between object and its closest cluster center, but they fail to represent the correlations between object and others. In contrast, clustering is not only based on the object-to-center (i.e., EEG-to-center in the paper) relationship, but also on object-to-object (i.e., EEG-to-EEG) correlations. Therefore, a novel clustering approach that can reflect intrinsic notion of clustering and consider the collective behavior of EEG data should be proposed, to simultaneously capture the EEG-to-center and EEG-to-EEG correlations. To address this problem

C. Dai and D. Pi are with the College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China (e-mail: chenglongdai@nuaa.edu.cn; dc.pi@nuaa.edu.cn).

J. Wu is with Department of Computing, Macquarie University, NSW 2109, Australia (e-mail: jia.wu@mq.edu.au).

L. Cui is with the Intelligent Information Processing Laboratory, Suzhou University, Suzhou 234000, China (e-mail: jsjxcuilin@126.com).

B. Johnson is with the Department of Cognitive Science at Macquarie University, Sydney, Australia. E-mail: blake.johnson@mq.edu.au.

Stefanie I. Becker is with the School of Psychology, University of Queensland, St Lucia, QLD 4072, Australia (e-mail: s.becker@psy.uq.edu.au).

for multi-trial EEG clustering, a cooperative game-inspired method is introduced in the paper, based on cross correlation-transformed Shapley values that are used to assign EEG trials with strong correlations (i.e., with nearly equal Shapley values) into same clusters to form compact clusters while to scatter weak-correlation ones in different clusters.

The second weakness of existing centroid-based clustering methods is that they rely on an optimization function that localizes the centroid through a time-consuming iteration process that may also converge on a local minimum (rather than finding the optimal centroid). Further, these approaches are also sensitive to cluster center initialization or the order presence of data, resulting in inconstant clusters with several runs and precluding the satisfiability of order independence property [21] (we will discuss it in Section VI). To address this problem, we use Shapley value based on marginal contribution of each EEG trial to efficiently initialize cluster centers, instead of a time-consuming iteration strategy. In detail, the EEG trial holds highest Shapley value is defined as a cluster center, then with similarity threshold, the EEG trials with closest Shapley values to the center will be assigned into the corresponding cluster. Emphatically, this strategy guarantees the clusters with intra-cluster compactness and inter-cluster scatter.

In summary, we propose a novel approach for multi-trial EEG clustering based on cooperative game theory, which is mapped to the collaboration of EEG trials in cooperative games. Cooperative game is a game that the players collaborate with each other to build a collectively and coalitionally rational relationship with strong constraint force and compulsorily executive agreement that aims to achieve the optimal coalitions that maximizes the value of the game [22]. Further, the goal of clustering is to group similar objects into same cluster while different clusters separated as far as possible, and this process is similar to the coalition formation in a cooperative game. In fact, like clustering, there are many potential coalitions (i.e., clusters) of players (i.e., objects in cluster or EEG trials in this paper) in a cooperative game and most of them are meaningless, and the core of a cooperative game is an optimal solution for coalition searching. In detail, the core is the collection of allocations that are collectively (intra-cluster compact) and coalitionally (inter-cluster scattered) rational and the allocations in the core are stable. That is, with the core of cooperative game, we can cluster EEG trials into optimal groups by simultaneously considering intra-cluster compactness and inter-cluster scatter. Furthermore, different players have different contributions to the total value of the cooperative game, and the Shapley value can achieve fair allocations for players based on the marginal contributions [22], with which the cluster centers (with highest Shapley values) and EEG trials (with nearly equal Shapley values) assigned into which coalition (i.e., cluster) can be also determined efficiently in the end, to form intra-cluster compacted and inter-cluster scattered clusters.

B. Contributions and Outline

Multi-trial EEG clustering is a non-trivial task for the brand new type of unlabeled EEG data in applications. However, studies on it have been rarely reported so far. In the paper, we explored the issue with cooperative game theory and addressed the challenges of (1) convexity discussion of EEG clustering-mapped cooperative game, (2) Shapley value-based EEG cluster center selection, and (3) simultaneously considering intra-cluster compactness and inter-cluster scatter of EEG clusters. In summary, the contributions of this paper are listed below.

- We solved the problem of multi-trial EEG clustering by mapping it to coalition formation in a cooperative game. We also showed that the EEG clustering-mapped cooperative game is convex. To the best of our knowledge, this is the first attempt to apply cooperative game strategy to cluster unlabeled multi-trial EEG signals.
- We proposed a novel approach named CoGEEGc for multi-trial EEG clustering based on Shapley value and similarity threshold. The cross-correlation-measured Shapley value characterizes the importance of every EEG trial to cluster and efficiently determines cluster centers for CoGEEGc, while similarity threshold determines EEG trials with nearly equal Shapley values assigned into corresponding clusters. In the end, CoGEEGc achieves a high-quality multi-trial EEG clustering with respect to intra-cluster compactness and inter-cluster scatter.
- We demonstrated the efficacy of CoGEEGc with detailed experimentation on 15 real-world multi-trial EEG datasets, and comparisons with 15 state-of-the-art clustering approaches also clearly indicate the superiority of CoGEEGc on multi-trial EEG clustering.
- We also discussed the satisfiability of four clustering theoretical properties for CoGEEGc, which shows CoGEEGc outperforms baseline methods, since it satisfies two properties of richness and order independence while 15 baselines only satisfy one.

The remainder of this paper is presented as follows: Related works on EEG or time series clustering are summarized in Section II. Background of cross correlation for EEG similarities and cooperative game theory is introduced in Section III. The cooperative game inspired multi-trial EEG clustering method, i.e., CoGEEGc, is presented in Section IV, including its convexity discussion and Shapley value computation. Detailed experimentation on real-world multi-trial EEG datasets is described in Section V. We then discuss the satisfiability of four desirable clustering properties for CoGEEGc in Section VI. In Section VII, we present our conclusions and list some directions for future work.

II. Related Works

Most of existing reports on EEG analysis focus on supervised approaches, especially EEG classification. Only a few studies are reported on multi-trial EEG clustering to date. Dai et al. [18] proposed a centroid-based EEG clustering method with cross correlation, which achieved good results compared to several state-of-the-art time series clustering methods. However, its performance relies on an optimization function that required many iterations and time for convergence and is probably subject to the problem of converging on a local minimum. Besides, it also requires to preset cluster number,

which prevents it from achieving all desired partitions of EEG trials. Wahlberg and Lantz [14] introduced two multichannel-based methods for epileptic EEG spikes clustering: fuzzy C-means (FCM) and graph-theoretic clustering (GTC). Like k-means, FCM requires the initialization of cluster number and cluster centers, which probably results in inconstant clusters with multiple runs. It also demands the convergence of its coefficients to a upper bound, which likely falls into local convergence. The GTC requires an initialization of component number to cluster (i.e., k-*components*), which may not produce expected number of clusters, and it also likely groups objects that belong to different clusters into the same cluster due to the constraint of k-*components*. Further, this study mainly focused on multichannel epileptic EEG, rather than multi-trial ones.

Although only several studies investigated EEG clustering [14], [18], promising time series-based approaches may provide potential solutions for multi-trial EEG clustering, since EEG can be regarded as time series with specific characteristics. Recently, various time series clustering methods are emerged out and they can be roughly categorized into five types.

K-means-type methods, such as k-means++ [23], robust and sparse fuzzy k-means (RSFKM) [24] and k-multiple-means (KMM) [25], exploit the strategy of k-means to partition time series into k clusters. The most popular and classical k-means clusters data by randomly initializing cluster centers and assigning remaining candidates to closest centers according to their Euclidean distance or relative distance to centers, so it has high efficiency but low accuracy. k-means++, an variant of k-means, initializes cluster center based on probability distribution, which improves clustering performance to some extent, but its first center is still randomly initialized. Further like k-means, k-means++ is very sensitive to outliers. To handle outliers, a fuzzy k-means clustering method termed RSFKM is proposed by considering the robustness and sparseness of object-cluster memberships. In addition, KMM extends k-means that only considers one center to model each class by considering multiple sub-cluster means (i.e., sub-centers) that is modeled into a bipartite graph partitioning problem with the constrained Laplacian rank [26]. Although these k-means-type methods made great improvements in conventional time series clustering through k-means strategy, they may inherit main weaknesses of k-means that requires a build-in number of k-specified clusters, and they may be not robust to (sub-)center/prototype initializations, which likely results in inconstant clusters during multiple runs. Additionally, a dynamic clustering approach for time series is recently introduced in [27], which dynamically monitors the cluster membership switch of time series over time, but it applies k-means++ to initialize clusters, thus it also inherits weakness of k-means.

Density-based methods identify clusters as the regions with high densities. In detail, they first identify dense regions with density estimators, and then link neighbouring dense regions to cluster [28], such as the most famous and classical DBSCAN [29], OPTICS [30], and SNN [31]. Although they are widely applied in clustering, they involve some inherent weaknesses. DBSCAN is empirically sensitive to the parameters of ϵ (the radius of neighborhood) and MinPts (the minimum number of objects in the neighborhood), and it probably fails to find clusters with differing densities. OPTICS may not exactly separate clusters with its reachability ordering, since the ordering just follows the nearest neighboring reachability distance. SNN uses SNN similarity [32] to identify core points and build clusters, which is based on the overlap between their kNN lists. Therefore, it relies on the use of kNN strategy and is highly sensitive to the k parameter. Generally, the density-based methods, as stated in [28], probably cannot identify all clusters with differing densities. Recently, to address low-density clustering problem in traditional density-based methods, the density ratio-based strategy is proposed in [28] through replacing density estimator with density-ratio estimator. However, the performance of reconditioned density ratio-based strategy also depends on the separation of cluster modes (i.e., points of the highest estimated density) and it is also inherently sensitive to η-neighborhood density.

Feature selection-based methods, such as regularized discriminative feature selection (UDFS) [33], nonnegative spectral feature selection (NDFS) [34], robust unsupervised feature selection (RUFS) [35], and robust spectral feature selection (RSFS) [36], enhance or improve clustering performance by selecting discriminative features of time series. However, they cluster time series embedded with k-means, such that they also require initialization of cluster number in advance, which precludes them getting all desired clusters. Although they select discriminative features from time series, some important information is also lost in the process. These feature selection-based methods are also affected by many parameters that needs tuning, such as feature subset dimension, scatter separability, or maximum likelihood [37]. Further, these methods seem sensitive to a prior setting number of features. Moreover, some other feature selection-based methods such as cooperative learning-based [38] and local learning-based [39] clustering may not require the setup of the feature number or cluster number, but they demand the process of optimizing an objective function to converge and then to cluster, which seems to result in unfixed clusters during different runs due to the inconstant convergence.

Distance-based methods cluster time series data based on distance measures, such as Euclidean distance (ED), Dynamic time warping (DTW) and its variant cDTW [40], relative distance (e.g., K-SC [41]), and cross correlation [18]. A proper distance measure probably achieves good clusters. However, it is unclear which is the most applicable distance measure for EEG trials. Particularly, ED requires that sequences have the same length and it cannot capture sequences of variable lengths which is typical in EEG trials. DTW concentrates very much on minimizing the accumulation of all local distances between adjacent points [42]. As such, DTW does not deal well with one of two sequences sampled less frequently [43] and it is affected by the warping window width. Paparrizos et al. [44] showed that, with respect to time series clustering, cross correlation outperforms related distance, ED, DTW, and competitive to cDTW. In the newest study on EEG clustering [18], it also demonstrated the superiority of cross correlation on measuring distances among EEG trials. However, these distance-based methods that consider all data points of time series are sensitive to outliers and noise [45]. Besides, they also

exploit an optimization function to search centroid sequences of clusters, which is time-consuming because of the mass of iterations for optimization. Further, they cannot achieve consistent clusters due to the convergence of the optimization function for centroid search.

Shape/shapelet-based approaches [44], [46], [47] learn a series of discriminative shapes/shapelets to represent original time series, and then cluster based on their correlations/similarities to these shapes/shapelets. The most promising shape/shapelet-based time series clustering approaches, k-shape [44] and USSL [47], perform well on traditional time series, but they may not be suitable for multi-trial EEG clustering. k-shape searches for representative shapes by iteratively running a shape extraction function, which also likely results in inconsistent clusters due to the inconstant solutions of objective function after multiple runs. USSL is influenced by many parameters that require empirical tuning, and thus it cannot guarantee consistency in the number of clusters. Further, both k-shape and USSL require initialization of the number of clusters prior to EEG clustering. Besides, these methods are very time-consuming, as it takes much time to search the distinctive shapes/shapelets. Recently, another sequence-to-sequence representation method named Deep Temporal Clustering Representation (DTCR) [45] is proposed, and it generates cluster-specific representations of time series through k-means process and then clusters, which improves the cluster structures and obtains non-linear cluster-specific temporal representations, but it is still affected by the ability of encoder and may be degraded by missing values.

III. Preliminaries

In this section, we introduce the background of cross correlation for EEG's Shapley value computation and cooperative game theory. First of all, to facilitate reading, the main symbols used in this paper are briefly summarized in Table I.

A. Cross Correlation

Cross correlation [18], known as a sliding or shifting point-to-point inner product, is widely used in signal processing. It is also a scaling and shift invariant measure [44] which means it can deal with sequences with different phase (local or global alignment), length or scale. Namely, cross correlation is a similarity measure for two signals to evaluate displacement of one relative to the other. Given two signals $\boldsymbol{x}$ and $\boldsymbol{y}$, their cross correlation is defined as

$$CC_\tau(\boldsymbol{x},\boldsymbol{y}) = R_\tau(\boldsymbol{x},\boldsymbol{y}) = \sum_{i=1}^{m-|\tau|} (x_{i+\tau}\cdot y_i) \tag{1}$$

where $\tau \in \{-m,-(m-1),\cdots,0,1,\cdots,(m-1),m\}$ denotes the shift locations, m is the length of signals. Importantly, cross correlation can measure similarities among signals with different lengths, but we just utilize it to measure signals (EEG trials) with same length in this paper. Therefore, the resultant cross correlation sequence of $\boldsymbol{x}$ and $\boldsymbol{y}$ contains $2m-1$ samples.

Considering all the possible shift sequences, the cross correlation sequences $CC_w(\boldsymbol{x},\boldsymbol{y}) = (c_1,\cdots,c_w)$ with the length of $2m-1$ is defined as

$$CC_w(\boldsymbol{x},\boldsymbol{y}) = R_{w-m}(\boldsymbol{x},\boldsymbol{y}) = R_\tau(\boldsymbol{x},\boldsymbol{y}) \tag{2}$$

TABLE I
Notations and Descriptions

Notation	Description
$\boldsymbol{N}$	Set of players in a cooperative game
$\boldsymbol{S}$	Subset of $\boldsymbol{N}$
$\boldsymbol{\Omega}$	Orders of permutation
$\boldsymbol{\Theta}$	Set of all permutations of $\boldsymbol{N}$
$\boldsymbol{C},\boldsymbol{D}$	Coalition of players in a cooperative game
$\boldsymbol{E}$	Set of multi-trial EEG
$\boldsymbol{O},\boldsymbol{Q}$	Coalition of EEG trials
$\boldsymbol{\phi}$	Set of Shapley values of EEG coalition
$\boldsymbol{e}$	A single EEG trial
CC	Traditional cross correlation of EEG trials
$\boldsymbol{\zeta}$	A contribution segment for cross correlation
v	Value function of a given cooperative game
τ	Shift of an EEG trial to another
θ	Permutation of players in a cooperative game
δ	Similarity threshold
α	Weight to normalized cross correlation
β	Weight to local tendency
n	Number of multi-trial EEG
m	Length of EEG trials

where $w \in \{1,2,\cdots,2m-1\}$, $CC_w(\boldsymbol{x},\boldsymbol{y}) \in [-1,1]$ and $R_{w-m}(\boldsymbol{x},\boldsymbol{y})$ is calculated as follows

$$R_\tau(\boldsymbol{x},\boldsymbol{y}) = \begin{cases} \sum_{k=1}^{m-\tau} x_{k+\tau}\cdot y_k, & \tau \geq 0 \\ R_{-\tau}(\boldsymbol{y},\boldsymbol{x}), & \tau < 0 \end{cases} \tag{3}$$

where $\tau = w-m$. The goal of cross correlation between $\boldsymbol{x}$ and $\boldsymbol{y}$ is to search the position of w that maximizes $CC_w(\boldsymbol{x},\boldsymbol{y})$, as also illustrated in Fig.1(a).

B. Cooperative Game

Cooperative game is a game, in which groups of players enforce cooperative behavior that is a competition between coalitions of players, not between individual players [48]. A cooperative game is simply defined as the formation of $(\boldsymbol{N},v)$, where $\boldsymbol{N} = \{1,2,\cdots,n\}$ denotes players in the cooperative game and $v: 2^{\boldsymbol{N}} \to \mathbb{R}, v(\varnothing) = 0$ is called value function. In a cooperative game $(\boldsymbol{N},v)$, a payoff allocation $\boldsymbol{x} = \{x_1,x_2,\cdots,x_n\}$ such that $x_i \geq v(\{i\}), \sum_{i=1}^n x_i = v(\boldsymbol{N})$ denotes a vector of contribution of players to the overall gain of cooperative game, where $i \in \boldsymbol{N}(\boldsymbol{N} = \{1,2,\cdots,n\})$. The gain of every player x_i in a payoff allocation must be larger than that without cooperation, $v(\{i\})$. That is, a payoff allocation must be coalitionally rational which satisfies $x_i \geq v(\{i\})$. Obviously, there are many payoff allocations for a cooperative game, but most of them are unacceptable. Subsequently, we introduce two important solution concepts for cooperative game: the core and the Shapley value.

1) The Core: The core of a game $(\boldsymbol{N},v)$ is the coalitionally and collectively rational collection of all payoff allocations [49], as $\boldsymbol{C}o(v)$ denotes, where $\boldsymbol{C}o(v) \subseteq \boldsymbol{A}(v)$ and $\boldsymbol{A}(v)$ is the set of all allocations. The core is the set of imputations, under which no coalition has a value greater than the sum of its elements' payoffs. Therefore, no payoff coalition has incentive to leave the grand coalition $\boldsymbol{N}$ and receive a larger payoff. Namely, the payoff allocation in the core is stable since no

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2021.3060742, IEEE Transactions on Knowledge and Data Engineering

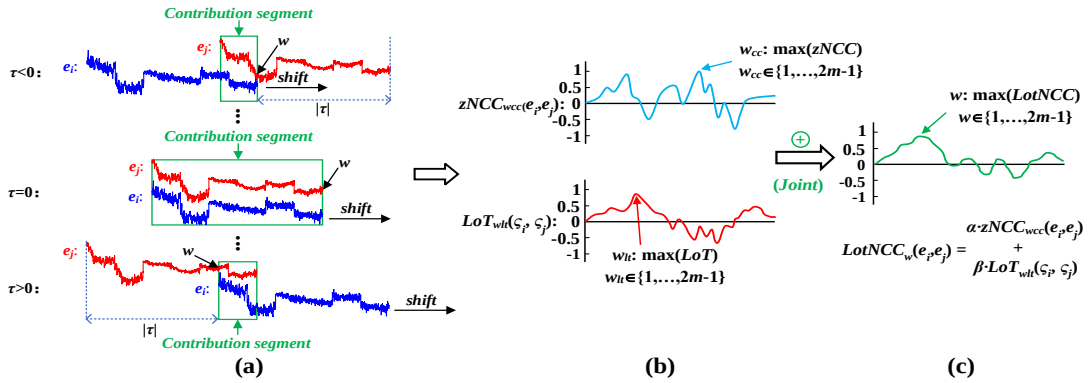


Fig. 1. An example of computing $LotNCC_w$ by jointing $zNCC_{w_{cc}}$ and $LoT_{w_{lt}}$, where w denotes the final optimal shifted location such that $\max LotNCC$ when jointing $zNCC$ and LoT, w_{cc} denotes the optimal location for traditional cross correlation such that $\max zNCC$ along with w, and w_{lt} denotes the optimal location for contribution segments such that $\max LoT$ along with w. (a) To compute shift of $\boldsymbol{e}_x$ and $\boldsymbol{e}_y$, keep $\boldsymbol{e}_y$ static and move $\boldsymbol{e}_x$ along with $\boldsymbol{e}_y$, then it will obtain $2m-1$ shift locations $w \in \{1, \cdots, 2m-1\}$ and $2m-1$ possible lengths for contribution segments corresponding to w, namely $w_{lt} \in \{1, \cdots, 2m-1\}$; (b) to simply calculate $LotNCC_w$, first compute $zNCC_{w_{cc}}$ (Blue line), $LoT_{w_{lt}}$ (Red line), respectively based on their corresponding w_{cc} and w_{lt} ($w_{cc}, w_{lt} \in \{1, \cdots, 2m-1\}$); (c) subsequently, joint $zNCC_{w_{cc}}$ and $LoT_{w_{lt}}$ with α and β to obtain the final $LotNCC_w$ (Green line) along with w.

any player can benefit by unilaterally deviating from a payoff allocation in the core.

2) Shapley Value: The Shapley value [50] is a solution concept of cooperative games, which assigns a unique distribution of a total surplus generated by the coalition of all players. Namely, some players in the cooperative game contribute more to the coalition than others, so these players intuitively deserve higher payoff, and the Shapley value fairly measures the gains to every player. Mathematically, the Shapley value of cooperative game $(\boldsymbol{N}, v)$ is defined as

$$\boldsymbol{\phi}(\boldsymbol{N}, v) = (\phi_1(\boldsymbol{N}, v), \phi_2(\boldsymbol{N}, v), \cdots, \phi_n(\boldsymbol{N}, v)) \tag{4}$$

where $\phi_i(\boldsymbol{N}, v)$ is the Shapley value of player i in the cooperative game $(\boldsymbol{N}, v)$. In detail, the fairly optimal payoff to the player i is calculated by

$$\phi_i(\boldsymbol{N}, v) = \sum_{\boldsymbol{S} \subseteq \boldsymbol{N} \setminus \{i\}} \frac{|\boldsymbol{S}|!(n - |\boldsymbol{S}| - 1)!}{n!} \Big\{ v(\boldsymbol{S} \cup \{i\}) - v(\boldsymbol{S}) \Big\} \tag{5}$$

where n is total number of players in cooperative game, and the sum extends over all subsets $\boldsymbol{S}$ of $\boldsymbol{N}$ without player i.

3) Convex Cooperative Games: According to Shapley's conclusion [51], $(\boldsymbol{N}, v)$ is convex when the marginal contribution of each player i in a larger coalition is correspondingly larger than that in a smaller coalition. Namely,

$$v(\boldsymbol{C} \cup \{i\}) - v(\boldsymbol{C}) \leq v(\boldsymbol{D} \cup \{i\}) - v(\boldsymbol{D}) \tag{6}$$

where $\forall \boldsymbol{C} \subseteq \boldsymbol{D} \subseteq \boldsymbol{N} \setminus \{i\}$. The value of $v(\boldsymbol{D} \cup \{i\}) - v(\boldsymbol{D})$ is defined as the marginal contribution to the coalition $\boldsymbol{D}$.

4) Shapley Value in Convex Games: As introduced in [49], for possible $|\boldsymbol{N}|!$ permutations θ of players in cooperative game $(\boldsymbol{N}, v)$, the orders of permutation is given by

$$\boldsymbol{\Omega}_{\theta,s} = \{i \in \boldsymbol{N} : \theta(i) \leq s\}, s \in \{1, 2, \cdots, |\boldsymbol{N}|\} \tag{7}$$

where $\boldsymbol{\Omega}_{\theta,0} = \varnothing$, $\boldsymbol{\Omega}_{\theta,|\boldsymbol{N}|} = \boldsymbol{N}$, and $\theta(i)$ represents the location of player i in the permutation of θ. To determine the core for the specific θ, Eq.7 can be solved by

$$x_i^{\theta}(\boldsymbol{\Omega}_{\theta,s}) = v(\boldsymbol{\Omega}_{\theta,s}) \tag{8}$$

The solution of Eq.8 defines the payoff vector x^{θ} with elements defined as

$$x_i^{\theta} = v(\boldsymbol{\Omega}_{\theta,\theta(i)}) - v(\boldsymbol{\Omega}_{\theta,\theta(i)-1}), \forall i = 1, 2, \cdots, |\boldsymbol{N}| \tag{9}$$

The Shapley value of convex games, according to [51], is the center of payoff vector x^{θ}, which can be calculated by

$$\phi_i = \frac{1}{|\boldsymbol{N}|!} \sum_{\theta \in \boldsymbol{\Theta}} x_i^{\theta} \tag{10}$$

where $\boldsymbol{\Theta}$ denotes the set of all permutations of $\boldsymbol{N}$.

IV. CoGEEGc: Cooperative Game Inspired Multi-Trial EEG Clustering

In the section, we introduce the idea of cooperative game inspired multi-trial EEG clustering in detail.

A. Cross Correlation-Based Multi-trial EEG Similarity

To measure similarities among EEG trials, a modified cross correlation is adopted in the paper, as introduced in Dai et al's work [18]. In detail, the modified cross correlation between two EEG trials $\boldsymbol{e}_i$ and $\boldsymbol{e}_j$ with length of m at optimal shifted location w is defined as Eq.11 that contains two items: z-normalized cross correlation $zNCC$ and local tendency LoT. Further, $LotNCC$ can be transformed to separately compute $zNCC$ and LoT at corresponding optimal shift locations w_{cc} and w_{lt}, and then search their jointly optimal shifted location w such that $\max LotNCC_w$ from $2m-1$ values to get the final $LotNCC_w$, see Fig.1(b)(c).

$$LotNCC_w(\boldsymbol{e}_i, \boldsymbol{e}_j) = \alpha \cdot zNCC_{w_{cc}}(\boldsymbol{e}_i, \boldsymbol{e}_j) + \beta \cdot LoT_{w_{lt}}(\boldsymbol{\zeta}_i, \boldsymbol{\zeta}_j) \tag{11}$$

where $LotNCC_w \in [-1, 1]$, $\alpha, \beta \in [0, 1]$ $(\alpha + \beta = 1)$ denote the weights of traditional z-normalized cross correlation and local tendency, and $\boldsymbol{\zeta}_i \subseteq \boldsymbol{e}_i$ and $\boldsymbol{\zeta}_j \subseteq \boldsymbol{e}_j$ are the exact sequences (*i.e, contribution sequences* in [18]) used to do inner product when computing cross correlation between $\boldsymbol{e}_i$ and $\boldsymbol{e}_j$.

Especially, the z-normalized cross correlation $zNCC(\boldsymbol{e}_i, \boldsymbol{e}_j)$ is defined as

$$zNCC(\boldsymbol{e}_i, \boldsymbol{e}_j) = \frac{CC(\boldsymbol{e}_i, \boldsymbol{e}_j)}{\sqrt{R_0(\boldsymbol{e}_i, \boldsymbol{e}_i) \cdot R_0(\boldsymbol{e}_j, \boldsymbol{e}_j)}} \tag{12}$$

Besides, the local tendency between $\boldsymbol{e}_i$ and $\boldsymbol{e}_j$ to modify $zNCC$ with contribution segments is defined as

$$LoT(\boldsymbol{\zeta}_i, \boldsymbol{\zeta}_j) = \frac{\sum_{k=1}^{p-q}(e_{i_{k+q}} - e_{i_k})(e_{j_{k+q}} - e_{j_k})}{\sqrt{\sum_{k=1}^{p-q}(e_{i_{k+q}} - e_{i_k})^2 \sum_{k=1}^{p-q}(e_{j_{k+q}} - e_{j_k})^2}} \quad (13)$$

where $\boldsymbol{\zeta}_i \subseteq \boldsymbol{e}_i$ and $\boldsymbol{\zeta}_j \subseteq \boldsymbol{e}_j$ such that $|\boldsymbol{\zeta}_i| = |\boldsymbol{\zeta}_j|$ are the contribution segments [18] to do inner product when computing cross correlation between $\boldsymbol{e}_i$ and $\boldsymbol{e}_j$; $q(1 \leq q < p)$ denotes the interval length, commonly $q = 1$; $p = l + w - m$ (m denotes the length of EEG trials and l, p are the index of beginning and ending point of $\boldsymbol{\zeta}$); $LoT(\boldsymbol{\zeta}_i, \boldsymbol{\zeta}_j) \in [-1, 1]$, and positive $LoT(\boldsymbol{\zeta}_i, \boldsymbol{\zeta}_j)$ indicates most of neighboring points in contribution segments have same trend; otherwise, opposite trend would be present.

Clearly, there are $2m - 1$ values of $LotNCC_w$ along with $2m - 1$ possible locations of w, and the maximum (i.e., $\max LotNCC_w$) is the final one that we are searching for. To the end, the distance $dist(\boldsymbol{e}_i, \boldsymbol{e}_j)$ of two EEG trials $\boldsymbol{e}_x$ and $\boldsymbol{e}_j$ in the paper is measured by

$$dist(\boldsymbol{e}_i, \boldsymbol{e}_j) = 1 - \max LotNCC_w(\boldsymbol{e}_i, \boldsymbol{e}_j) \quad (14)$$

where $dist \in [0, 2]$, and the smaller the $dist$, more similar the two EEG trials.

B. Multi-Trial EEG Clustering Mapped Cooperative Games

Given a multi-trial EEG dataset $\boldsymbol{E}_{n\times m} = \{\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_n\}$ of n trials with length of m, we redefined a similarity function based on Eq.14: $f(dist) : [0, 2] \rightarrow (0, 1]$, where $dist : \boldsymbol{E} \times \boldsymbol{E} \rightarrow [0, 2]$, and $dist(\boldsymbol{e}_i, \boldsymbol{e}_j), \forall \boldsymbol{e}_i, \boldsymbol{e}_j \in E$ denotes the distance between $\boldsymbol{e}_i$ and $\boldsymbol{e}_j$, satisfying $dist(\boldsymbol{e}_i, \boldsymbol{e}_i) = 0$. Further, $f : [0, 2] \rightarrow (0, 1]$ is set as the monotonically non-increasing similarity mapping function on $dist$ and particularly $f(0) = 1$, since a smaller distance between EEG signals literally corresponds to a higher similarity between them. Correspondingly, define a dissimilarity mapping function $f' : [0, 2] \rightarrow [0, 1)$, such that $f'(dist) = 1 - f(dist)$. The goal is to group together those EEG trials that are similar with each other based on f.

We transformed multi-trial EEG clustering into coalition formation in a cooperative game $(\boldsymbol{N}, v)$ in such a way that every EEG trial $\boldsymbol{e}_i$ in $\boldsymbol{E}_{n\times m}$ is mapped to a player i in the cooperative game, and the number of players is that of EEG trials: $|\boldsymbol{N}| = n$. Players cooperate with each other and aim to build a coalition (i.e., cluster) while maximizing the value of the coalition indicated by v. In this instance, the single player is not a member of any coalition; therefore, we set the value of the single player i to $v(\{i\}) = 0$. Subsequently, we defined the value of a coalition $\boldsymbol{C}$ as

$$v(\boldsymbol{C}) = \frac{1}{2} \sum_{\substack{\boldsymbol{e}_i, \boldsymbol{e}_j \in \boldsymbol{C} \\ \boldsymbol{e}_i \neq \boldsymbol{e}_j}} f\Big(dist(\boldsymbol{e}_i, \boldsymbol{e}_j)\Big) \quad (15)$$

where $v(\cdot)$ is the value function for a coalition in this formulation. Since the distance between two EEG trials is symmetric, that is $dist(\boldsymbol{e}_i, \boldsymbol{e}_j) = dist(\boldsymbol{e}_j, \boldsymbol{e}_i)$, the sum of pairwise similarities between EEG trials is $\frac{1}{2}$. Finally, a scale-invariant similarity function f is defined as

$$f\Big(dist(\boldsymbol{e}_i, \boldsymbol{e}_j)\Big) = 1 - \frac{dist(\boldsymbol{e}_i, \boldsymbol{e}_j)}{dist_{\max} + 1} \quad (16)$$

where $dist_{\max}$ denotes the maximum distance between any two EEG trials. Correspondingly, $f'\Big(dist(\boldsymbol{e}_i, \boldsymbol{e}_j)\Big) = 1 - f\Big(dist(\boldsymbol{e}_i, \boldsymbol{e}_j)\Big) = \frac{dist(\boldsymbol{e}_i, \boldsymbol{e}_j)}{dist_{\max}+1}$.

C. Convexity of Multi-Trial EEG Cooperative Games

As introduced in Section III-B, the convex cooperative game should satisfy the marginal contribution of player i to the coalition $\boldsymbol{D}$, namely, $v(\boldsymbol{C} \cup \{i\}) - v(\boldsymbol{C}) \leq v(\boldsymbol{D} \cup \{i\}) - v(\boldsymbol{D}), \forall \boldsymbol{C} \subseteq \boldsymbol{D} \subseteq \boldsymbol{E}$. With the value function of Eq.15, the multi-trial EEG clustering mapped cooperative game is a convex game.

Theorem 1. *The cooperative game of multi-trial EEG clustering with the value function* $v(\boldsymbol{C}) = \frac{1}{2} \sum_{\boldsymbol{e}_i, \boldsymbol{e}_j \in \boldsymbol{C}, \boldsymbol{e}_i \neq \boldsymbol{e}_j} f\Big(dist(\boldsymbol{e}_i, \boldsymbol{e}_j)\Big)$ *is a convex game, where* $v(\{\boldsymbol{e}_i\}) = 0, i \in \{1, 2, \cdots, n\}$ *and* $\boldsymbol{C} \subseteq \boldsymbol{E}$ *is a coalition with* k *EEG trials.* (The proof is shown in APPENDIX A.)

D. Shapley Value of Multi-Trial EEG Cooperative Games

In the convex cooperative game $(\boldsymbol{N}, v)$, the Shapley value of each EEG trial $\boldsymbol{e}_i$ can be computed as Theorem 2 introduced.

Theorem 2. *The Shapley value of an EEG trial* $\boldsymbol{e}_i$ *in the convex cooperative game* $(\boldsymbol{N}, v)$ *can be computed by* $\phi_i = \frac{1}{2} \sum_{\boldsymbol{e}_j \in \boldsymbol{E}, j \neq i} f\Big(dist(\boldsymbol{e}_i, \boldsymbol{e}_j)\Big)$. (The proof is shown in APPENDIX B.)

In addition, data points (players in a given cooperative game) located near each other have approximately equal Shapley values [49]. Similarly, EEG trials that are near each other have approximately equal Shapley values as well.

Theorem 3. *For two EEG trials* $\boldsymbol{e}_i$ *and* $\boldsymbol{e}_j$ *such that* $dist(\boldsymbol{e}_i, \boldsymbol{e}_j) \leq \xi, \xi \rightarrow 0$ *in the convex game have nearly equal Shapley values.* (The proof is shown in APPENDIX C.)

E. CoGEEGc: Multi-Trial EEG Clustering Algorithm

Algorithm 1 outlines the proposed approach for multi-trial EEG clustering. First, it computes the Shapley value for every EEG trial, see lines 1-6. Subsequently, it updates the cluster center $\boldsymbol{e}_c$ by selecting an EEG trial with maximum Shapley value from the un-clustered EEG trials as shown in line 9, while assigning EEG trials whose similarity to $\boldsymbol{e}_c$ is larger than δ to the same cluster (see lines 11-12). With the threshold δ, the algorithm can assign those EEG trials with nearly equal Shapley values into the same cluster (see Section IV-D).

F. Time Complexity of CoGEEGc

CoGEEGc clusters EEG trials using the Shapley value and modified cross correlation. Moreover, cross correlation-based EEG similarity computation is prior to Shapley value computation. We, therefore, discussed their time costs respectively.

1) Time Consumption of Modified Cross Correlation: As Dai et al introduced in [18], the modified cross correlation $LotNCC_w$ can be transformed to compute $zNCC_{w_{cc}}$ and $LoT_{w_{lt}}$, separately, and then search the modified cross correlation such that $\max LotNCC_w$ from $2m-1$ joint values of $zNCC_{w_{cc}}$ and $LoT_{w_{lt}}$. As analyzed in [18], computing $zNCC_{w_{cc}}$ and $LoT_{w_{lt}}$ along with $2m-1$ possible locations of w between any two EEG trials requires $O(m\log m)$ and $O(m^2)$, respectively. Particularly, $zNCC_{w_{cc}}$ is computed by Discrete Fourier Transform (DFT) and Inverse Discrete Fourier Transform (IDFT), which requires $O(m^2)$ time. But with the Fast Fourier Transform (FFT) algorithm proposed by Cooley and Tukey in [52], its computation time can be reduced to $O(m\log m)$. Consequently, the time consumption of $LotNCC_w$ for n EEG trials requires $O\Big(\max\{n^2 m\log m, n^2m^2\}\Big) = O(n^2m^2)$ (since, $n \ll m$ commonly), where m denotes the length of the EEG trial.

2) Time Consumption of Shapley Value: As Theorem 2 introduced, the Shapley value of every EEG trial $\boldsymbol{e}_i$ in the EEG convex cooperative game can be computed with a closed-form formula $\phi_i = \frac{(n-2)!}{n!}\Big[\Big(1+\cdots+(n-1)\Big)\sum\limits_{\boldsymbol{e}_j\in E, j\neq i} f\Big(dist(\boldsymbol{e}_i,\boldsymbol{e}_j)\Big)\Big]$. Consequently, it requires exactly $O(n^2)$ to compute Shapley values for n EEG trials, which is also shown in Algorithm 1.

To the end, the exact time complexity of CoGEEGc for n-trial EEG clustering is $O\Big(\max\{n^2m^2, n^2\}\Big)$. In practice, given that, usually, $n \ll m$, the time consumption for CoGEEGc equates to $O(n^2m^2)$, which indicates that distance computation costs much more time than Shapley value computation.

V. Experimental Results

This section reported the experimental results of CoGEEGc comparing with 15 state-of-the-art EEG/time series clustering approaches on 15 real-world multi-trial EEG datasets.

Algorithm 1 CoGEEGc.

Input:
$\boldsymbol{E}_{n\times m} = \{\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_n\}$: n EEG trials with length of m;
$\delta \in (0, 1]$: Similarity threshold;
Output:
Clusters of multi-trial EEG;
for $i = 1$ to $n-1$ **do**
  for $j = i$ to n **do**
    $dist(\boldsymbol{e}_i, \boldsymbol{e}_j) \leftarrow 1 - \max LotNCC(\boldsymbol{e}_i, \boldsymbol{e}_j)$;
    $\phi_i = \frac{1}{2}\sum\limits_{\boldsymbol{e}_j\in\boldsymbol{E}, j\neq i} f(dist(\boldsymbol{e}_i, \boldsymbol{e}_j))$;
  end for
end for
Initialize $\boldsymbol{T} = \boldsymbol{E}, \boldsymbol{C} = \varnothing$;
repeat
  $c = \arg\max\limits_{\boldsymbol{e}_i\in\boldsymbol{T}} \phi_i$;
  $\boldsymbol{C} = \boldsymbol{C} \cup \{\boldsymbol{e}_c\}$;
  $\boldsymbol{O}_c = \{\boldsymbol{e}_i \in \boldsymbol{T} : f(dist(\boldsymbol{e}_c, \boldsymbol{e}_i)) \geq \delta\}$;
  $\boldsymbol{T} = \boldsymbol{T} \backslash \boldsymbol{O}_c$;
until $\boldsymbol{T} = \varnothing$;

A. Multi-trial EEG Datasets

As Table II shows, the multi-trial EEG data we used include (1) slow cortical potentials (SCPs [1], one type of EEG), i.e., #1–#5. The SCPs are recorded from two subjects, one is healthy (Ia) and the other is an amyotrophic lateral sclerosis (ALS) patient (Ib); A cross-healthy-patient EEG dataset is also constructed with datasets Ia and Ib: we firstly deleted one EEG channel (vEOG) of Ib that is used to detect vertical eye movements, then down-sampled EEG trials of remaining 6 channels of Ib into 5377 to mix with dataset Ia, which finally leads to EEG datastes #3–#5; (2) Three-class mental imagery EEG with precomputed features [2] from 3 healthy subjects, i.e., #6–#8; (3) Four-class motor imagery EEG [3] from 3 healthy subjects, i.e., #9–#11; (4) Two-class simple motor imagery EEG datasets [3] from 2 healthy subjects, i.e., #12–#13; (5) Four-class hand movement EEG datasets [3], i.e., #14–#15. Importantly, as the multi-trial EEG were all originally labeled, we deleted these labels prior to applying clustering algorithms. Besides, all the EEG data are z-normalized before used to cluster with those algorithms in the paper.

B. Evaluation Criteria

We evaluated our method with 6 criteria that are from "view of quality" such as *intra-cluster compactness*, *inter-cluster scatter*, *integrated ratio*, and from "view of accuracy" such as *adjusted rand index*, *F-score*, and *Fleiss' kappa*.

• ***Intra-cluster Compactness*** [18] measures the compactness of multi-trial EEG in the same cluster. Mathematically, $S_{In} = \frac{1}{|\boldsymbol{C}|}\sum_{C_i\in\boldsymbol{C}}(\frac{1}{|C_i|}||\boldsymbol{dist}_i - \mu_i||_2)$, where $C_i \in \boldsymbol{C}$ denotes the i^{th} EEG cluster, μ_i is the mean similarity of C_i, and $\boldsymbol{dist}_i$ denotes the local tendency modified cross-correlation similarities matrix of C_i. **A smaller S_{In} indicates higher holistic similarity among EEG trials in the same cluster.**

• ***Inter-cluster Scatter*** [18] evaluates the scatters of EEG trials from different clusters. Mathematically, $S_{Be} = \frac{1}{|\boldsymbol{C}|}\sum_{C_i,C_j\in\boldsymbol{C}}\Big(\frac{1}{\sqrt{|C_i||C_j|}}\sum_{i\neq j}(\boldsymbol{dist}_{ij}-\mu_i)(\boldsymbol{dist}_{ij}-\mu_j)^T\Big)^{\frac{1}{2}}$. **A larger S_{Be} demonstrates higher scatter between clusters.**

• ***Integrated Ratio (γ)*** [18] is an integrated ratio (or tradeoff) that simultaneously considers the intra- and inter-cluster evaluations, i.e., $\gamma = \frac{S_{Be}}{S_{In}}$. **A higher γ indicates higher quality and better clustering achieved by the method.**

• ***Adjusted Rand Index (ARI)*** [53], as an improvement of Rand Index ($RI = \frac{TP+TN}{TP+FP+TN+FN}$), measures the agreement between two partitions: one given by the cluster process and the other defined by external criteria. Namely, $ARI = \dfrac{\sum\limits_{ij}\binom{n_{ij}}{2} - \Big[\sum\limits_{i}\binom{a_i}{2}\sum\limits_{j}\binom{b_j}{2}\Big]/\binom{n}{2}}{\frac{1}{2}\Big[\sum\limits_{i}\binom{a_i}{2}+\sum\limits_{j}\binom{b_j}{2}\Big] - \Big[\sum\limits_{i}\binom{a_i}{2}\sum\limits_{j}\binom{b_j}{2}\Big]/\binom{n}{2}}$, where n_{ij}: number of objects in both class u_i and cluster v_i; a_i: number of objects in class u_i and b_j: number of objects in cluster v_j; n: total number of objects.

[1]The SCP datasets are available publicly online and can be downloaded from http://www.bbci.de/competition/ii/.

[2]The mental imagery EEG datasets are also available online at http://www.bbci.de/competition/iii/.

[3]The motor imagery EEG datasets are available as online archives at http://www.bbci.de/competition/iv/.

TABLE II
MULTI-TRIAL EEG DATASETS

Dataset	Title	Descriptions	# of EEG Trials	Length of EEG Trial (m)	# of EEG Channels	# of Clusters
# 1	Ia (Traindata_0 & Traindata_1)	SCPs from a healthy subject	268	5377	6	2
# 2	Ib (Traindata_0 & Traindata_1)	SCPs from an ALS patient	200	8065	7	2
# 3	Traindata_0, Ia & Traindata_0, Ib	Mixed SCPs labeled "class 0" of a healthy subject and an ALS patient	235	5377	6	2
# 4	Traindata_1, Ia & Traindata_1, Ib	Mixed SCPs labeled "class 1" of a healthy subject and an ALS patient	233	5377	6	2
# 5	Ia & Ib	Mixed SCPs of dataset Ia (from a healthy subject) and Ib (from an ALS patient)	468	5377	6	4
# 6	III_V_s1	Mental imagery EEG with precomputed features of left hand, right hand, and word association from 3 healthy subjects: s1, s2, and s3	3488	97	8	3
# 7	III_V_s2		3472			
# 8	III_V_s3		3424			
# 9	IV_2a_s1	Motor imagery EEG of left hand, right hand, both feet, and tongue from 3 healthy subjects: s1, s2, and S3	288	6887	22	4
# 10	IV_2a_s2					
# 11	IV_2a_s3					
# 12	IV_2b_s1	Simple motor imagery EEG of left hand and right hand from 2 healthy subjects: s1 and s2	120	940	3	2
# 13	IV_2b_s2					
# 14	IV_3_s1	Hand movement EEG in directions of left, right, forward, and backward from 2 healthy subjects: s1 and s2	160	4001	10	4
# 15	IV_3_s2					

• ***F-score*** [54] unequally weighs the false positives FP and false negatives FN in ***RI*** with a scale parameter $\beta \geq 0$ on *recall*. Mathematically, $F-score = \frac{(1+\beta^2)pr}{\beta^2 p+r}$, where *precision* $p = \frac{TP}{TP+FP}$, *recall* $r = \frac{TP}{TP+FN}$, TP is true positives and commonly $\beta = 1$.

• ***Fleiss' kappa*** (κ) [55] is a statistical measure for assessing the coherence of decision ratings among classes. In detail, $\kappa = \frac{\overline{P}-\overline{P_e}}{1-\overline{P_e}}$, where $\overline{P} = \frac{1}{Nn(n-1)}(\sum_{i=1}^{N}\sum_{j=1}^{k} n_{ij}^2 - Nn)$ and $\overline{P_e} = \sum_{j=1}^{k}(\frac{1}{Nn}\sum_{i=1}^{N} n_{ij})^2$.

Higher *ARI*, *F-score*, and κ correspond to higher "accuracy" of multi-trial EEG clusters.

C. Baselines

We compared CoGEEGc with 15 state-of-the-art EEG/time series clustering algorithms that are introduced in detail in Section II, including the extensions of classical k-means: k-means++ [23] and KMM [25]; density ratio-based clustering: ReCon_DBSCAN [28], ReCon_SNN [28], and ReCon_OPTICS [28]; feature selection-based clustering: UDFS [33], NDFS [34], RUFS [35], and RSFS [36]; distance-based clustering: DBA [40], K-SC [41], and MTEEGC [18]; shape/shapelet-based clustering: k-shape [44] and USSL [47]; and dynamic clustering: DMDLM [27]. To render the clustering results more reliable, all the baseline methods are performed 10 times and the results are averaged. The number of clusters for the baselines are set according to the number of classes in original EEG datasets, and parameter values in baseline methods are set as same as the original paper (For details, please refer to the original paper). All algorithms were conducted with Matlab R2014b, on Windows 10 machines with 4∗3.30 GHz CPUs and 16GB memory.

D. Multi-trial EEG Clustering Results

The results from "view of quality" with respect to intra-cluster compactness (S_{In}), inter-cluster scatter (S_{Be}), and integrated ratio (γ) are shown in Table III. Generally, CoGEEGc achieved the highest S_{In} and S_{Be} on most of multi-trial EEG datasets, and it yielded the highest integrated ratio $\gamma = \frac{S_{Be}}{S_{In}}$ for all datasets. Further introduced in Table III, CoGEEGc obtains the best average ranks of S_{In} (i.e., 1.9333), S_{Be} (i.e., 1.3333) and γ (i.e., 1.0). As a consequence, taking intra-cluster compactness and inter-cluster scatter into account simultaneously, CoGEEGc achieved better clustering quality than the 15 state-of-the-art methods. Most of these 15 methods focused on EEG-to-center (or EEG-to-subcenter, EEG-to-means) relationship or just considered density ratio rather than EEG-to-EEG relationships; yet, they underrated the correlations between any two EEG trials in the same cluster or dissimilarities among different clusters. That means these 15 methods did not consider the coalitional compactness among all EEG trials, as they grouped dissimilar EEG trials into the same cluster, which leads to relatively poor intra-cluster compactness and inter-cluster scatter. CoGEEGc is also a centroid-based method, but it is based on the Shapley value that the EEG trial with highest Shapley value is selected as the center, and then with similarity threshold δ, those EEG trials with mostly equal Shapley value to the center will be assigned into corresponding clusters. In the process, it considers both correlations among EEG trials in the same cluster and those in different clusters. Again, similarity threshold also helps assign similar EEG with mostly equal Shapley values into same clusters and separate dissimilar EEG into different clusters.

To further illustrate the superiority of CoGEEGc, we also analyzed it from "view of accuracy" using *ARI*, *F-score*, and κ. The results are in detail shown in Table IV, which clearly indicates that CoGEEGc outperforms the 15 state-of-the-art methods, since CoGEEGc obtains the best performance in terms of the lowest average ranks of *ARI* (i.e., 1.1333), *F-score* (i.e., 1.1333) and κ (i.e., 1.2). EEG trials in the same cluster share similar patterns while different EEG have different ones.

TABLE III
MULTI-TRIAL EEG CLUSTERING WITH 15 MULTI-TRIAL EEG DATASETS WITH RESPECT TO S_{In}, S_{Be} AND γ .
(**SMALLER S_{In}, OR LARGER S_{Be}, OR LARGER γ INDICATES A HIGHER "QUALITY" OF EEG CLUSTERS.**)

Dataset	Measure	k-means++	ReCon_ DBSCAN	ReCon_ SNN	ReCon_ OPTICS	UDFS	NDFS	RUFS	RSFS	K-SC	DBA	k-shape	USSL	KMM	DMDLM	MTEEGC	CoGEEGc
#1	S_{In}	0.5901	0.5978	0.5893	0.5871	0.5801	0.5801	0.5798	0.5811	0.5828	0.5816	0.5919	0.5845	0.5836	0.5821	**0.5732**	0.5734
	S_{Be}	0.6149	0.6154	0.6171	0.6169	0.6144	0.622	0.6212	0.6176	0.6184	0.6192	0.6308	0.6258	0.6233	0.6207	0.6385	**0.6413**
	γ	1.042	1.0294	1.0472	1.0508	1.0591	1.0722	1.0714	1.0628	1.0611	1.0646	1.0657	1.0707	1.0680	1.0663	1.1139	**1.1184**
#2	S_{In}	0.5665	0.5798	0.5773	0.5783	0.562	0.5826	0.5771	0.5702	0.5743	0.5691	0.5701	0.5736	0.5774	0.5745	0.5628	**0.5602**
	S_{Be}	0.5889	0.5891	0.5921	0.6011	0.602	0.6266	0.6185	0.6264	0.6291	0.6262	0.6344	0.6314	0.6298	0.6247	0.6358	**0.6382**
	γ	1.0395	1.0160	1.0256	1.0394	1.0712	1.0755	1.0717	1.0986	1.0954	1.1003	1.1128	1.1008	1.0908	1.0874	1.1297	**1.1392**
#3	S_{In}	0.5768	0.5686	0.5662	0.5649	0.5703	0.569	0.5723	0.5615	0.5595	0.5556	0.5468	0.5649	0.5623	0.5652	0.5436	**0.5402**
	S_{Be}	0.5662	0.5643	0.5615	0.5632	0.5647	0.568	0.5724	0.5684	0.5571	0.5553	0.5528	0.5689	**0.5743**	0.5577	0.5741	0.5732
	γ	0.9816	0.9924	0.9917	0.9970	0.9902	0.9982	1.0002	1.0123	0.9957	0.9995	1.011	0.9881	1.0213	0.9867	1.0561	**1.0611**
#4	S_{In}	0.5885	0.5918	0.5902	0.5891	0.5871	0.5911	0.5892	0.5825	0.5927	0.5888	0.5799	0.5862	0.5753	0.5775	**0.5684**	0.5712
	S_{Be}	0.5799	0.5768	0.5761	0.5788	0.5801	0.5823	0.5789	0.5837	0.5891	0.5827	0.5811	0.5806	0.5868	0.5855	0.5841	**0.5923**
	γ	0.9854	0.9747	0.9761	0.9825	0.9881	0.9851	0.9825	1.0021	0.9939	0.9896	1.0021	0.9904	1.0200	1.0139	1.0276	**1.0369**
#5	S_{In}	0.5733	0.5721	0.5692	0.5698	0.5712	0.5789	0.5688	0.5714	0.5775	0.5651	0.5667	0.5698	0.5640	0.5691	0.5642	**0.5628**
	S_{Be}	0.8674	0.8688	0.8706	0.8717	0.8732	0.8791	0.8716	0.8772	0.8779	0.8596	0.8762	0.8731	0.8721	0.8713	0.8782	**0.8834**
	γ	1.513	1.5186	1.5295	1.5298	1.5287	1.5186	1.5323	1.5352	1.5202	1.5211	1.5461	1.5323	1.5463	1.5310	1.5565	**1.5697**
#6	S_{In}	0.5583	0.5632	0.5598	0.5584	0.5471	0.5413	0.5509	0.5388	0.5468	0.5527	0.5457	0.5621	0.5473	0.5496	0.5401	**0.5312**
	S_{Be}	0.6192	0.6255	0.6348	0.6379	0.6395	0.6360	0.6202	0.6416	0.6532	0.6469	0.6487	0.6337	0.6443	0.6417	0.6656	**0.6662**
	γ	1.1091	1.1106	1.1340	1.1424	1.1689	1.1749	1.1258	1.1908	1.1946	1.1704	1.1887	1.1274	1.1772	1.1676	1.2324	**1.2541**
#7	S_{In}	0.5628	0.5583	0.5572	0.5548	0.5528	0.5512	0.5495	0.5431	0.5478	0.5507	0.5446	0.5672	0.5435	0.5574	0.5422	**0.5398**
	S_{Be}	0.6473	0.6489	0.6511	0.6503	0.6572	0.6534	0.6487	0.6592	0.6581	0.6519	0.6606	0.6415	0.6686	0.6515	0.6657	**0.6711**
	γ	1.1501	1.1623	1.1685	1.1721	1.1889	1.1854	1.1805	1.2138	1.2014	1.1838	1.2130	1.1310	1.2302	1.1688	1.2278	**1.2432**
#8	S_{In}	0.5688	0.5688	0.5693	0.5672	0.5571	0.5622	0.5584	0.5603	**0.5523**	0.5572	0.5618	0.5675	0.5541	0.5613	0.5538	0.5544
	S_{Be}	0.6417	0.6454	0.6468	0.6487	0.6538	0.6569	0.6582	0.6594	0.6613	0.6558	0.6513	0.6502	0.6627	0.6568	0.6641	**0.6659**
	γ	1.1282	1.1347	1.1361	1.1437	1.1736	1.1684	1.1787	1.1769	1.1974	1.1770	1.1593	1.1457	1.1960	1.1701	1.1992	**1.2011**
#9	S_{In}	0.5835	0.5711	0.5678	0.5688	0.5674	0.5722	0.5697	**0.5615**	0.5668	0.5682	0.5634	0.5721	0.5703	0.5776	0.5618	0.5636
	S_{Be}	0.6352	0.6136	0.6132	0.6155	0.6117	0.6233	0.6172	0.6087	0.6216	0.6203	0.6168	0.6221	0.6218	0.6278	0.6305	**0.6358**
	γ	1.0886	1.0744	1.0800	1.0821	1.0781	1.0893	1.0834	1.0841	1.0967	1.0917	1.0948	1.0874	1.0903	1.0869	1.1223	**1.1281**
#10	S_{In}	0.5713	0.5708	0.5697	0.5683	0.5679	0.5628	0.5589	0.5688	0.5618	0.5611	0.5674	0.5652	**0.5558**	0.5624	0.5582	0.5562
	S_{Be}	0.5887	0.5895	0.5921	0.5943	0.5936	0.6137	0.6184	0.6203	0.6116	0.5988	0.6012	0.5931	0.6018	0.6045	0.6191	**0.6296**
	γ	1.0305	1.0328	1.0393	1.0458	1.0453	1.0904	1.1065	1.0905	1.0886	1.0672	1.0596	1.0494	1.0828	1.0749	1.1091	**1.1320**
#11	S_{In}	0.5688	0.5682	0.5634	0.5645	0.5598	0.5654	0.5610	0.5584	0.5572	0.5621	0.5587	0.5711	0.5612	0.5603	0.5551	**0.5533**
	S_{Be}	0.5976	0.5983	0.6016	0.6047	0.6125	0.6088	0.6143	0.6132	0.6083	0.6112	0.6157	0.6028	0.6107	0.6023	0.6129	**0.6162**
	γ	1.0506	1.0530	1.0678	1.0712	1.0941	1.0768	1.0950	1.0981	1.0917	1.0874	1.1020	1.0555	1.0882	1.0750	1.1041	**1.1137**
#12	S_{In}	0.5589	0.5572	0.5596	0.5578	0.5596	0.5608	0.5572	0.5563	0.5512	**0.5469**	0.5487	0.5534	0.5528	0.5576	0.5513	0.5478
	S_{Be}	0.5722	0.5729	0.5719	0.5736	0.5818	0.5782	0.5756	0.5735	0.5867	0.5812	0.5788	0.5716	0.5873	0.5833	0.5862	**0.5876**
	γ	1.0238	1.0282	1.0220	1.0283	1.0397	1.0310	1.0330	1.0309	1.0644	1.0627	1.0549	1.0329	1.0624	1.0461	1.0633	**1.0727**
#13	S_{In}	0.5576	0.5624	0.5582	0.5568	0.5527	0.5618	0.5588	0.5603	0.5571	0.5523	0.5489	0.5598	0.5537	0.5583	0.5509	**0.5485**
	S_{Be}	0.5723	0.5735	0.5712	0.5743	0.5776	0.5745	0.5782	0.5769	0.5821	0.5830	0.5815	0.5784	0.5805	0.5761	**0.5839**	0.5819
	γ	1.0264	1.0197	1.0233	1.0314	1.0451	1.0226	1.0347	1.0296	1.0449	1.0556	1.0594	1.0332	1.0484	1.0319	1.0599	**1.0610**
#14	S_{In}	0.5886	0.5893	0.5862	0.5876	0.5724	**0.5698**	0.5722	0.5711	0.5866	0.5837	0.5845	0.5823	0.5758	0.5812	0.5706	0.5721
	S_{Be}	0.8628	0.8552	0.8539	0.8547	0.8715	0.8687	0.8738	0.8692	0.8636	0.8576	0.8588	0.8607	0.8587	0.8642	0.8704	**0.8759**
	γ	1.4659	1.4512	1.4567	1.4546	1.5225	1.5246	1.5271	1.5220	1.4722	1.4692	1.4693	1.4781	1.4913	1.4869	1.5254	**1.5310**
#15	S_{In}	0.5792	0.5814	0.5828	0.5833	0.5753	**0.5676**	0.5721	0.5694	0.5777	0.5812	0.5764	0.5789	0.5711	0.5736	0.5686	0.5682
	S_{Be}	0.8710	0.8583	0.8618	0.8639	0.8716	0.8698	0.8723	0.8745	0.8637	0.8654	0.8622	0.8576	0.8725	0.8694	0.8723	**0.8791**
	γ	1.5038	1.4763	1.4787	1.4811	1.5150	1.5324	1.5247	1.5358	1.4951	1.4890	1.4958	1.4814	1.5278	1.5157	1.5341	**1.5472**
Avg S_{In} Rank		12.9333	14.0667	12.1333	11.4667	7.8	10.3333	8.3333	6.4667	7.7333	7.0667	6.4667	11.0	6.3333	9.0667	2.466	**1.9333**
# Best S_{In}		0	0	0	0	0	2	0	1	1	1	0	0	1	0	2	**7**
Avg S_{Be} Rank		12.8	13.9333	13.8	12.0	9.2	7.3333	8.1333	7.0	6.6	9.1333	7.6667	10.1333	5.3333	8.5333	3.0	**1.3333**
# Best S_{Be}		0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	**13**
Avg γ Rank		13.8667	15.0	13.6667	12.2	9.5333	8.5333	7.6	6.4	7.0667	8.0	6.2	10.2	5.2	9.0	2.2667	**1.0**
# Best γ		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**15**

Similar to the preceding analyses, with Shapley value and similarity threshold, CoGEEGc grouped similar EEG with similar pattern (i.e., EEG trials with mostly equal Shapley values) into same clusters and separated dissimilar EEG trials with dissimilar patterns into different clusters. Most of these 15 methods clustered EEG trials mainly based on EEG-to-center (or EEG-to-prototype, EEG-to-means) relationship or local density ratio, which only considered similarities between EEG trials and centers (or sub-cluster prototypes, cluster means) or high-density EEG clusters, but ignored similarities to other EEG trials in the same cluster or in low-density EEG clusters, so it degraded their clustering "accuracy".

E. Sensitivity Analysis

1) Impact of δ: CoGEEGc assigns EEG trials to a cluster based on Shapley values and similarity threshold δ. When the Shapley value of an EEG trial is close to that of cluster center C (the EEG trial with highest Shapley value) and its similarity to the cluster center is larger than δ, it will be assigned into the corresponding cluster. Here we analyzed the impact of δ on CoGEEGc for multi-trial EEG clustering. First of all, we introduced how to select proper values of δ for CoGEEGc based on the discrete similarity distribution $d(\delta)$ of EEG trials, that is defined by Eq.17.

$$d(\delta) = \frac{\sum\limits_{s\in[\delta_i,\delta_i+\Delta]} t_s}{T^2} \tag{17}$$

where T, t_s denote the total number of modified cross correlation (MCC)-based EEG similarities and number of MCC similarities located in the interval of $[\delta_i, \delta_i + \Delta]$, respectively. Besides, the similarity distributions of δ on 15 EEG datasets are illustrated in Fig.S1 of Appendix D, which shows that MCC-based EEG similarities generally locate in the range

TABLE IV
MULTI-TRIAL EEG CLUSTERING COMPARISONS ON 15 MULTI-TRIAL EEG DATASETS WITH RESPECT TO *ARI*, *F-score*, AND κ. (LARGER *ARI*, *F-score*, OR κ INDICATES A HIGHER "ACCURACY" OF EEG CLUSTERS.)

Dataset	Measure	k-means++	ReCon_ DBSCAN	ReCon_ SNN	ReCon_ OPTICS	UDFS	NDFS	RUFS	RSFS	K-SC	DBA	k-shape	USSL	KMM	DMDLM	MTEEGC	CoGEEGc
	ARI	0.0145	0.0031	0.0033	0.0036	0.0572	0.0648	0.0518	0.0008	0.001	0.0189	0.0258	0.0201	0.0559	0.0475	0.1147	**0.1422**
#1	*F-score*	0.4028	0.4867	0.4984	0.2494	0.433	0.5118	0.5604	0.5316	0.4233	0.5614	0.5691	0.4852	0.5547	0.5278	0.5728	**0.5882**
	κ	0.2018	0.0089	0.0238	0.0238	0.1878	0.2	0.1198	0.1239	0.074	0.0663	0.1711	0.1028	0.1834	0.1122	0.2158	**0.2213**
	ARI	0.0017	0.0001	0.0002	0.0002	0.0036	0.0027	0.005	0.0038	0.0018	0.0008	0.0081	0.0032	0.0068	0.0024	0.0103	**0.0162**
#2	*F-score*	0.5514	0.3321	0.3495	0.3744	0.4352	0.5514	0.5652	0.5688	0.5381	0.4724	0.5395	0.4718	0.5615	0.5569	0.5732	**0.5872**
	κ	0.04	0.02	0.02	0.02	0.09	0.04	0.02	0.06	0.03	0.05	0.01	0.02	0.04	0.06	0.1	**0.12**
	ARI	0.8053	0.6234	0.6450	0.6522	0.5151	0.5308	0.5017	0.5857	0.0874	0.1551	0.2887	0.2211	0.6821	0.8114	0.6642	**0.8532**
#3	*F-score*	0.597	0.6525	0.6670	0.6714	0.6055	0.6958	0.598	0.6998	0.6459	0.676	0.7035	0.6088	0.7102	0.6811	0.7152	**0.7278**
	κ	0.1441	0.3827	0.3971	0.4131	0.4025	0.4511	0.377	0.4611	0.2922	0.408	0.4694	0.2581	0.4122	0.3638	0.4771	**0.4907**
	ARI	0.8829	0.5079	0.5238	0.5419	0.8193	0.8349	0.8774	0.8829	0.0267	0.7428	0.1766	0.6484	0.8942	**0.9015**	0.8911	0.9012
#4	*F-score*	0.5838	0.4867	0.6228	0.8124	0.8372	0.8488	0.851	0.8581	0.534	0.8459	0.8596	0.5823	0.8744	0.8753	0.8732	**0.8820**
	κ	0.2189	0.1089	0.2763	0.3466	0.699	0.7012	0.7105	0.7212	0.1653	0.6924	0.6691	0.2829	0.7411	0.7476	0.7435	**0.7515**
	ARI	0.2947	0.2065	0.2248	0.2264	0.2119	0.2351	0.2286	0.2811	0.0344	0.0712	0.2332	0.2377	0.2253	0.2328	0.3217	**0.3568**
#5	*F-score*	0.5304	0.4198	0.4278	0.6467	0.676	0.681	0.6862	0.6814	0.5731	0.5923	0.6855	0.5912	0.6618	0.6412	0.6987	**0.7201**
	κ	0.201	0.2115	0.2185	0.2315	0.311	0.3122	0.3155	0.31	0.2762	0.2658	0.3154	0.2837	0.3085	0.3134	0.3716	**0.3898**
	ARI	0.5314	0.3460	0.3494	0.4126	0.5696	0.5773	0.5583	0.5843	0.5688	0.5268	0.5712	0.5238	0.5762	0.5494	0.5821	**0.5985**
#6	*F-score*	0.3720	0.3389	0.2304	0.3042	0.5411	0.5358	0.5325	0.5527	0.5314	0.3642	0.5290	0.3118	0.5558	0.5152	0.5543	**0.5563**
	κ	0.3305	0.1067	0.0102	0.0108	0.3179	0.3177	0.3067	0.3258	0.3112	0.2812	0.3017	0.3206	0.3168	0.2988	**0.3346**	0.3328
	ARI	0.5247	0.3487	0.3513	0.4081	0.5643	0.5449	0.5209	0.5627	0.5332	0.5124	0.5436	0.5013	0.5714	0.5277	0.5845	**0.6021**
#7	*F-score*	0.7165	0.2087	0.1809	0.3271	0.7213	0.6855	0.5828	0.7528	0.7120	0.6046	0.7199	0.3234	0.7489	0.7197	**0.7545**	0.7532
	κ	0.1541	0.0077	0.007	0.007	0.2755	0.2310	0.1569	0.3082	0.1967	0.1645	0.2320	0.3015	0.2788	0.2249	0.3119	**0.3134**
	ARI	0.5211	0.3334	0.3368	0.3709	0.5634	0.5596	0.5721	0.5687	0.5469	0.5423	0.5133	0.4989	**0.5888**	0.5379	0.5744	0.5886
#8	*F-score*	0.4169	0.1901	0.1603	0.3036	0.4814	0.4882	0.4835	0.4836	0.4235	0.4355	0.3395	0.3007	0.4894	0.4211	0.4884	**0.4912**
	κ	0.1364	0.0137	0.0124	0.0126	0.2155	0.2112	0.2172	0.2219	0.1435	0.1683	0.0565	0.2187	0.2269	0.1586	0.2245	**0.2297**
	ARI	0.4322	0.2474	0.3097	0.5926	0.6735	0.6844	0.6284	0.7050	0.7025	0.6659	0.7008	0.6669	0.6817	0.5875	0.7272	**0.7447**
#9	*F-score*	0.2692	0.0863	0.1385	0.1006	0.4355	0.4686	0.4305	0.4613	0.4101	0.4007	0.4128	0.3813	0.4679	0.4478	0.4672	**0.4738**
	κ	0.1019	0.0486	0.0602	0.0602	0.1902	0.1857	0.1925	0.1958	0.1787	0.1532	0.1794	0.1122	0.1885	0.1653	0.1984	**0.2018**
	ARI	0.3528	0.2474	0.2785	0.4323	0.6415	0.6578	0.6494	0.5897	0.7102	0.6611	0.7011	0.6917	0.6667	0.5634	0.7236	**0.7372**
#10	*F-score*	0.3241	0.3020	0.3271	0.3481	0.4601	0.4562	0.4588	0.4522	0.4254	0.4237	0.4217	0.3665	0.4451	0.3985	0.4577	**0.4636**
	κ	0.1272	0.1020	0.1039	0.1231	0.2012	0.1989	0.2033	0.2006	0.1602	0.1573	0.1648	0.1382	0.1947	0.1853	0.2038	**0.2121**
	ARI	0.2513	0.2474	0.2920	0.4104	0.5867	0.5919	0.5504	0.5785	0.5612	0.5578	0.5674	0.5594	0.5864	0.4932	0.6104	**0.6265**
#11	*F-score*	0.3155	0.3178	0.3814	0.4003	0.4482	0.4537	0.4456	0.4511	0.4350	0.4273	0.4264	0.3852	0.4429	0.4142	0.4471	**0.4589**
	κ	0.1804	0.0816	0.1003	0.1087	0.2735	0.2711	0.2812	0.2574	0.2372	0.2357	0.2448	0.1831	0.2668	0.2712	0.2822	**0.2872**
	ARI	0.5311	0.4935	0.4958	0.4971	0.5732	0.5638	0.5696	0.5735	0.5686	0.5894	0.6114	0.5782	0.5804	0.5496	0.6028	**0.6246**
#12	*F-score*	0.4416	0.3556	0.3568	0.4584	0.4933	0.4752	0.4985	0.4979	0.4888	0.5042	0.5667	0.4812	0.5108	0.4783	0.5697	**0.5722**
	κ	0.1	0.05	0.05	0.0536	0.0163	0.0287	0.0318	0.1087	0.0167	0.0833	0.1333	0.0204	0.1118	0.1123	**0.1371**	0.1349
	ARI	0.5212	0.4385	0.4648	0.4958	0.5434	0.5343	0.5488	0.5545	0.5812	0.5719	0.5682	0.5789	0.5634	0.5435	0.5936	**0.6188**
#13	*F-score*	0.4565	0.2835	0.3835	0.4664	0.4899	0.4863	0.4902	0.4896	0.5489	0.5249	0.5172	0.4757	0.5136	0.4812	**0.5633**	0.5624
	κ	0.0333	0.0257	0.0231	0.0269	0.0403	0.0433	0.0482	0.0487	0.1	0.0667	0.0833	0.0369	0.0536	0.0346	0.1086	**0.1237**
	ARI	0.5215	0.5458	0.5653	0.5546	0.7746	0.8104	0.7947	0.7812	0.7713	0.5967	0.6123	0.5534	0.8074	0.6164	0.8305	**0.8528**
#14	*F-score*	0.3479	0.2452	0.2935	0.3051	0.4177	0.4213	0.4256	0.4158	0.3934	0.3842	0.3990	0.3473	0.4113	0.3768	0.4141	**0.4341**
	κ	0.1238	0.1135	0.1187	0.1303	0.1612	0.1758	0.1735	0.1707	0.1556	0.1428	0.1603	0.1372	0.1645	0.1586	0.1765	**0.1788**
	ARI	0.4972	0.4453	0.5367	0.5381	0.6635	0.6542	0.6683	0.6143	0.5687	0.5721	0.6004	0.5023	0.6372	0.6214	0.6812	**0.7036**
#15	*F-score*	0.3202	0.1962	0.2753	0.2836	0.4154	0.4125	0.4236	0.4113	0.3801	0.3816	0.3869	0.3312	0.4175	0.3812	0.4167	**0.4254**
	κ	0.1422	0.0883	0.1036	0.1183	0.1725	0.1688	0.1758	0.1702	0.1683	0.1585	0.1623	0.1338	**0.1796**	0.1579	0.1752	0.1792
Avg *ARI* Rank		11.0667	15.0	13.5333	12.4667	7.5333	6.7333	8.0	6.7333	9.6667	10.2	7.6	9.8	5.0667	8.8	2.5333	**1.1333**
# Best *ARI*		0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	**13**
Avg *F-score* Rank		12.8	14.9333	14.1333	12.9333	7.4667	6.4	6.2	5.3333	9.8667	8.7333	7.0	12.6	4.6667	8.7333	3.0	**1.1333**
# Best *F-score*		0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	**13**
Avg κ Rank		11.2667	14.4	13.9333	12.8	6.8	6.8667	7.0667	5.3333	10.4667	9.8667	8.0667	10.6	5.7333	8.2667	2.1333	**1.2**
# Best κ		0	0	0	0	0	0	0	0	0	0	0	0	1	0	2	**12**

of $[0.25, 0.9]$. Moreover, to obtain balanced EEG clusters for 15 EEG datasets, δ located in $[0.25, 0.9]$ is generally chosen in the paper. Afterwards, we discussed CoGEEGc's clustering "accuracy" on the δ in $[0.25, 0.9]$, and the results are illustrated in Fig.S2 of Appendix D. As Fig.S2 indicates, its clustering "accuracy" increased and then decreased along with the increase of δ, and CoGEEGc achieved the best clustering with a moderate δ, such as $\delta \in \{0.6, 0.65, 0.7\}$. Presumably, a large δ would lead CoGEEGc to cluster only strictly similar EEG trials with extremely equal Shapley value into the same coalition, which resulted in too many clusters. On the contrary, a small δ would group more dissimilar EEG trials into the same cluster, returning fewer clusters than the realistic amount. Further, as Fig.S2 shows, EEG similarities of 15 EEG datasets are intensively located in the range of $\delta \in \{0.6, 0.65, 0.7\}$, so it may be the potential reason why CoGEEGc achieved the best clustering "accuracy" with similarity threshold $\delta \in \{0.6, 0.65, 0.7\}$.

2) Impact of distance measures: We analyzed the impact of distance measures on CoGEEGc by comparing the modified cross correlation (MCC) with Euclidean distance (ED), dynamic time warping (DTW), and the traditional cross correlation (TCC). The results are shown in Fig.S3 of Appendix E, which clearly demonstrates that the CoGEEGc with MCC achieved the highest *ARI*, *F-score*, and κ compared to ED, DTW, and TCC, which is also in accord with the conclusion in [18]. Practically, ED is quite sensitive to outliers. DTW concentrates too much on minimizing the accumulation of all local distances between adjacent points of EEG trials and it is also affected by warping window width. TCC focuses on the global similarity of EEG trials, without consideration of local similarity. MCC considers both global and local similarity of

EEG trials, as a result, MCC assists CoGEEGc to achieve better clustering results than ED, DTW, and TCC.

3) Impact of α and β ($\alpha + \beta = 1$): A modified cross correlation is applied to measure similarity among EEG trials in our method, using local tendency (Eq.13) to modify the normalized cross correlation (Eq.12). In Eq.11, α and β ($\alpha + \beta = 1$) weight the normalized cross correlation and local tendency, and finally together determines similarities between EEG trials. Here, we discussed the impact of combination weights of α and β on CoGEEGc for EEG clustering, and the results are illustrated in Fig.S4 of Appendix F, which demonstrates that a moderate combination weight (α, β) (i.e., $\alpha \in [0.4, 0.6]$, correspondingly $\beta \in [0.4, 0.6]$) yields better clusters, especially $(\alpha = 0.5, \beta = 0.5)$. A large α (correspondingly, small β) or large β (correspondingly, small α) sub-optimally assigns too much weight to the global similarity (traditional cross correlation) or the local tendency. As a result, it cannot properly measure similarities among EEG trials from the global and local view. In an extreme case, when only considering the normalized cross correlation (i.e., $(\alpha = 1, \beta = 0)$) or local tendency (i.e., $(\alpha = 0, \beta = 1)$), CoGEEGc does not perform as well as with moderate (α, β).

4) Execution Time: As introduced earlier, the exact time complexity of CoGEEGc is $O(n^2m^2)$, but in practice, its execution time is mainly on the computation of modified cross correlations for EEG trials, rather than Shapley values. Once the cross-correlation similarities are computed, CoGEEGc is efficient to cluster EEG trials, since it initializes cluster centers based on Shapley values rather than an optimization objective function and assigns EEG trials into corresponding clusters also based on Shapley values as well as similarity threshold. The time consumptions on 15 EEG datasets are displayed in Fig.S5 of Appendix G, which clearly shows that CoGEEGc is more efficient than most of the 15 algorithms. Admittedly, k-means++, ReCon_DBSCAN, ReCon_SNN, ReCon_OPTICS, and KMM are the most efficient algorithms, and the potential reason is that they use efficient distance measure (e.g., ED) or density estimator or k-means-alike strategy (i.e., KMM) to cluster. Meanwhile, UDFS, NDFS, RUFS, RSFS, K-SC, DBA, k-shape, USSL, and MTEEGC require convergence of an objective function for feature selection or centroid extraction, which costs much time. Further, DMDLM dynamically analyzes cluster membership switch for each data point over time, which is computationally expensive to cluster those long-length EEG trails that overlap with different clusters.

VI. Clustering Properties Satisfied by CoGEEGc

We discussed the clustering properties for CoGEEGc, including Kleinberg's *scale invariance*, *richness*, *consistency* [56], and Ackerman's *order independence* [21]. In detail, (1) scale invariance emphasizes that clustering algorithms do not build in "length scale"; (2) richness requires that all potential/desired clusters should be achieved by clustering algorithms; (3) consistency demands that clustering results do not change when shrinking or expanding the intra-cluster or inter-cluster distances, and (4) order independence imposes that clustering algorithms are not sensitive to the presentation order of forthcoming data. Namely, the clustering algorithm achieves the same clusters during multiple runs without being influenced by the presentation order of data.

1) Scale invariance: CoGEEGc clusters EEG trials with a scale invariant similarity function $f(dist) = 1 - \frac{dist}{dist_{\max}+1}$, but it builds in a similarity threshold δ in the process, which precludes it from satisfying scale invariance. Although ReCon_DBSCAN, ReCon_SNN, and ReCon_OPTICS cluster EEG trials based on their density ratio with corresponding density estimator, they embed a given core radius/distance (i.e., ϵ-neighborhood) or/and a minimum number of EEG trials (i.e., MinPts) to identify "core EEG" and then form EEG clusters. Namely, in these algorithms, the cardinality of the neighborhood should exceed a threshold, which denotes they also build in a scale, so they do not satisfy the scale invariance. k-means++, UDFS, NDFS, RUFS, RSFS, K-SC, DBA, k-shape, USSL, MTEEGC, KMM, and DMDLM satisfy scale invariance property, since they cluster EEG trials based on the pre-set number of clusters k, instead of built-in length scale.

2) Richness: CoGEEGc clusters EEG trials based on their Shapley values and similarity threshold. In other words, the number of clusters is determined by Shapley values and similarity threshold δ. Therefore, it can adjust δ to generate all the desired partitions (i.e., clusters). With this, CoGEEGc satisfies the richness property. In contrast, k-means++, UDFS, NDFS, RUFS, RSFS, K-SC, DBA, k-shape, USSL, MTEEGC, KMM, and DMDLM do not satisfy this property, since they cluster EEG trials based on a prior setting of cluster number k, which prevents them from meeting the requirement of richness property. Obviously, in ReCon_DBSCAN, ReCon_SNN, and ReCon_OPTICS, MinPts, the pre-set parameter, determines the "core EEG" and constrains the minimum number of a sub-cluster, so it precludes these three algorithms from getting any desired clusters with expected numbers of EEG trials (e.g., the clusters with less than MinPts EEG trials). Therefore, they do not satisfy the richness property.

3) Consistency: CoGEEGc, k-means++, UDFS, NDFS, RUFS, RSFS, K-SC, DBA, k-shape, USSL, and MTEEGC are all center/centroid-based clustering methods. According to Kleinberg's conclusion [56], no center/centroid-based clustering methods satisfy consistency property. Therefore, none of them satisfies the consistency property. As introduced before, ReCon_DBSCAN, ReCon_SNN, and ReCon_OPTICS cluster EEG trials based on density-ratio strategy with their corresponding density estimator that is embedded ϵ-neighborhood (radius of the neighborhood). When shrinking or expanding the intra-cluster or inter-cluster distances, it probably changes the EEG trials in the ϵ-neighborhood radius, then the density ratio has to change accordingly, and the clusters are changed in the end, especially for clusters with significantly different densities. Consequently, being embedded ϵ, the three density ratio-based algorithms violate the consistency. KMM considers the distances between data and prototypes. When expanding intra-cluster distances or/and shrinking inter-cluster distances, the distances and probabilities of prototypes assigning to data are correspondingly changed, finally resulting in inconsistent clusters. In addition, the prototypes in KMM also can be regarded as the centers of sub-clusters. When the probabilities

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2021.3060742, IEEE Transactions on Knowledge and Data Engineering

of prototypes assigned to data are updated, the prototypes are also updated, until the multiple-mean neighbor assignment is no longer changed. As a result, KMM can be defined as a (sub)center-based clustering algorithm, which does not satisfy the consistency property according to Kleinberg's conclusion [56]. Although DMDLM is not a distance/similarity-based clustering algorithm, it utilizes k-means++ to initialize clusters which can be affected by the shrinkage or expansion of inter-cluster or intra-cluster distances, thus it, similar to k-means++, does not satisfy consistency property.

4) order independence: It is easy to see that CoGEEGc is order independent. CoGEEGc chooses cluster centers based on Shapley values of EEG trials rather than by random selection. Subsequently, EEG trials are assigned into clusters based on similarity threshold as well as their Shapley values. In this assignment, for an incoming EEG trial $\boldsymbol{e}_{i+1}$, the coalition of $i+1$ EEG trials is defined as $(C_{\theta,i}, \boldsymbol{e}_{i+1}) = C_{\theta,i} + \sum_{\theta(j)\leq\theta(i)} f\Big(dist(\boldsymbol{e}_{i+1}, \boldsymbol{e}_j)\Big)$, where $C_{\theta,i} = \sum_{\theta(j)<\theta(m)\leq\theta(i)} f\Big(dist(\boldsymbol{e}_m, \boldsymbol{e}_j)\Big)$ denotes the coalition for i EEG trials whose permutation orders satisfy $\theta(j) < \theta(m) \leq \theta(i)$. That is to say, the final coalition is not changed along with EEG trials' presentation order. Therefore, CoGEEGc satisfies order independence. In fact, ReCon_DBSCAN, ReCon_SNN, and ReCon_OPTICS use a density ratio-based strategy, modified from the classical density-based strategy, to capture "core EEG" and clusters. In such a strategy, the order of core EEG capture just influences the local sub-clusters but never changes the links between EEG trials and "core EEG" in the final clusters, since their density ratios always meet the threshold within the ϵ-neighborhood radius even though with different initializations of "core EEG", which guarantees to achieve same clusters during different runs. Hence, ReCon_DBSCAN, ReCon_SNN, and ReCon_OPTICS are order independent. k-means++, UDFS, NDFS, RUFS, RSFS, K-SC, DBA, k-shape, USSL, and MTEEGC do not satisfy order independence, since they are not robust to cluster center/centroid initialization that probably results in inconstant clusters during multiple runs. Particularly, their optimization function for searching cluster centers/centroid or features cannot guarantee a global convergence, thus they also likely get different clusters after several runs, which violates the requirement of order independence. Similarly, KMM is actually an extension of k-means-type clustering algorithm, and its results vary on different initialization, such as the initialization of m prototypes. In other words, KMM would get different clusters with multiple runs, which precludes from satisfying order independence. Although DMDLM applies a locally constant model called Evolutional Dirichlet Process [57] to dynamically monitor the cluster membership switch of data that leads to a locally constant cluster, it embeds the cluster initialization through k-means++ process that inherits the order uncertainty of cluster initialization in k-means++, so it cannot guarantee the same clustering results during multiple runs. Consequently, DMDLM is not order independent.

As summarized in Fig.2, CoGEEGc stands out as it satisfies two of four theoretical clustering properties, while those 15 state-of-the-art baseline methods only satisfy one.

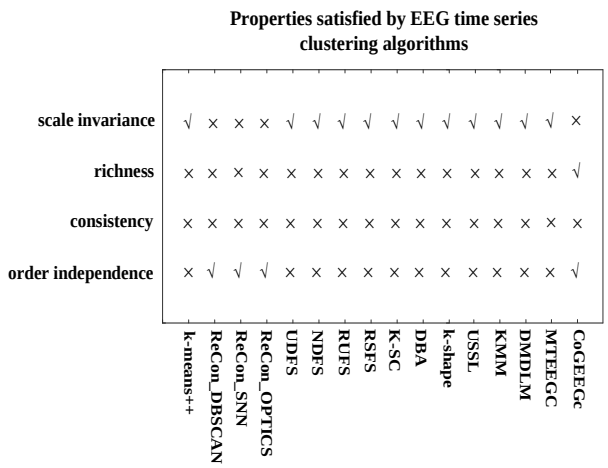


Fig. 2. Key properties satisfied by multi-trial EEG/time series clustering methods. $\surd$ and $\times$, respectively, denote the "satisfied" and "dissatisfied" property.

VII. Conclusions and Future Work

This paper investigated multi-trial EEG clustering and proposed a novel approach, inspired by cooperative game theory, called CoGEEGc. CoGEEGc transformed multi-trial EEG clustering into a convex cooperative game, utilized modified cross correlation-transformed Shapley values to determine cluster centers, and clustered EEG trials based on Shapley values and similarity threshold. CoGEEGc considered the intrinsic properties of clustering, since it, like cooperative game, considered the basis of distances between EEG trials to center and correlations to other EEG trials in same clusters. CoGEEGc achieved high-quality multi-trial EEG clusters with respect to intra-cluster compactness and inter-cluster scatter, and it satisfies two of four desired clustering properties: richness and order independence. Furthermore, comparisons with 15 state-of-the-art EEG/time series clustering approaches using standard cluster validity criteria on 15 multi-trial EEG datasets further demonstrated the efficacy and superiority of CoGEEGc for multi-trial EEG clustering.

In future, it could assess the impact of other similarity/distance metrics on CoGEEGc with additional real-world multi-trial EEG datasets, such as cross-subject EEG datasets, or discuss parameter setting in automatic (not empirically manual) ways on more general cases. In addition, another future work could explore further options to reduce time consumption for CoGEEGc.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (Grant No. U1433116 and 61702355) and Fundamental Research Funds for the Central Universities (Grant No. NP2017208).

References

[1] C. Babilon, C. Del Percio, R. Lizio, et al., "Abnormalities of cortical neural synchronization mechanisms in patients with dementia due to Alzheimer's and Lewy body diseases: an EEG study," *Neurobiol. of aging*, vol. 55, pp. 143–158, 2017.

[2] S. Khatun, B. Morshed, and G. M. Bidelman, "A single-channel EEG-based approach to detect mild cognitive impairment via speech-evoked brain responses," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 5, pp. 1063–1070, 2019.

[3] S. Wang, W. A. Chaovalitwongse, and S. Wong, "Online seizure prediction using an adaptive learning approach," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 12, pp. 2854–2866, 2013.

[4] M. Fan and C. Chou, "Detecting Abnormal Pattern of Epileptic Seizures via Temporal Synchronization of EEG Signals," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 3, pp. 601–608, 2019.

[5] Y. Shahriari, M. Vaughan, L. M. McCane, et al., "An exploration of BCI performance variations in people with amyotrophic lateral sclerosis using longitudinal EEG data," *J. Neural Eng.*, vol. 16, no. 5, article no. 056031, 2019.

[6] K. K. Ang and C. Guan, "EEG-based strategies to detect motor imagery for control and rehabilitation," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 4, pp. 392–401, 2017.

[7] M. Tariq, P. M. Trivailo, and M. Simic, "EEG-based BCI control schemes for lower-limb assistive-robots," *Front. Hum. Neurosci.*, vol. 12, article no. 312, 2018.

[8] Y. Ekawa, T. Nishida, and N. Yamawaki, "An improvement of robot arm for EEG-based brain-computer interface system," *Neurosci. Research*, vol. 68, sup. 1, pp. e215–e216, 2010.

[9] H. Zeng and A. Song, "Optimizing Single-Trial EEG Classification by Stationary Matrix Logistic Regression in Brain-Computer Interface," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 11, pp. 2301–2313, 2016.

[10] Y. Mishchenko, M. Kaya, E. Ozbay, et al., "Developing a three- to six-state EEG-based Brain-Computer Interface for a virtual robotic manipulator control," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 4, pp. 977–987, 2019.

[11] F. Qi, Y. Li, and W. Wu, "RSTFC: A Novel Algorithm for Spatio-Temporal Filtering and Classification of Single-Trial EEG," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 12, pp. 3070–3082, 2015.

[12] I. Razzak, M. Blumenstein, and G. Xu, "Multiclass support matrix machines by maximizing the inter-class margin for single trial EEG classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 6, pp. 1117–1127, 2019.

[13] Y. Wang and K. C. Veluvolu, "Evolutionary algorithm based feature optimization for multi-channel EEG classification," *Frontiers Neurosci.*, vol. 11, article no. 28, 2017.

[14] P. Wahlberg and G. Lantz, "Methods for robust clustering of Epileptic EEG spikes," *IEEE Trans. Biomed. Eng.*, vol. 47, no. 7, pp. 857–868, 2000.

[15] J. Metsomaa, J. Sarva, and R. J. Ilmoniemi, "Multi-trial evoked EEG and independent component analysis," *J. Neurosci. Meth.*, vol. 228, pp. 15–26, 2014.

[16] C. Dai, D. Pi, S. I. Becker, et al., "Valid EEG selection for classification," *ACM Trans. Knowl. Discov. Data*, vol. 14, no. 2, article no. 18, 2020.

[17] C. Dai, J. Wu, D. Pi, et al., "Brain EEG time series selection: A novel graph-based approach for classification," in *SDM*, 2018, pp. 558–566.

[18] C. Dai, D. Pi, L. Cui, et al., "MTEEGC: A novel approach for multi-trial EEG clustering," *Appl. Soft Comput.*, vol. 71, pp. 255–267, 2018.

[19] T. W. Liao, "Clustering of time series data-a survey," *Pattern Recogn.*, vol. 38, no. 11, pp. 1857–1874, 2005.

[20] R. S. Michalski, "Knowledge acquisition through conceptual clustering: A theoretical framework and an algorithm for partitioning data into conjunctive concepts," *J. Policy Analysis and Information systems*, vol. 4, no. 3, pp. 219–244, 1980.

[21] M. Ackerman and S. Ben-David, "Measures of clustering quality: A working set of axioms of clustering," in *NIPS*, 2008, pp. 121–128.

[22] R. B. Myerson, *Game Theory: Analysis of Conflict*. Harvard Univ. Press, 1997.

[23] D. Arthur and S. Vassilvitskii, "k-means++: The advantages of careful seeding," in *SODA*, 2007, pp. 1027–1035.

[24] J. Xu, J. Han, K. Xiong, et al., "Robust and sparse fuzzy k-means clustering," in *IJCAI*, 2016, pp. 2224–2230.

[25] F. Nie, C. Wang, and X. Li, "K-multiple-means: A multiple-means clustering method with specified k clusters," in *KDD*, 2019, pp. 959–967.

[26] F. Nie, X. Wang, M. I. Jordan, et al., "The constrained Laplacian rank algorithm for graph-based clustering," in *AAAI*, 2016, pp. 1969–1976.

[27] Victhor S. Sartório and Thaís C. O. Fonseca, "Dynamic clustering of time series data," 2020. [Online]. Available: arXiv:2002.01890v1.

[28] Y. Zhu, K. M. Ting, and M. J. Carman, "Density-ratio based clustering for discovering clusters with varying densities," *Pattern Recogn.*, vol. 60, pp. 983–997, 2016.

[29] M. Ester, H. P. Kriegel, J. Sander, et al., "A density-based algorithm for discovering clusters in large spatial databases with noise," in *KDD*, 1996, pp. 226–231.

[30] M. Ankerst, M. M. Breunig, H. P. Kriegel, et al., "OPTICS: ordering points to identify the clustering structure," *ACM SIMOD Record*, vol. 28, no. 2, pp. 49–60, 1999.

[31] L. Ertöz, M. Steinbach, V. Kumar, et al., "Finding clusters of different sizes, shapes, and densities in noisy, high dimensional data," in *SDM*, 2003, pp. 47–58.

[32] R. A. Jarvis and E. A. Patrick, "Clustering using a similarity measure based on shared near neighbors," *IEEE Trans. Comput.*, vol. 100, no. 11, pp. 1025–1034, 1973.

[33] Y. Yang, H. T. Shen, Z. Ma, et al., "$l_{2,1}$-norm regularized discriminative feature selection for unsupervised learning," in *IJCAI*, 2011, pp. 1589–1594.

[34] Z. Li, Y. Yang, J. Liu, et al., "Unsupervised feature selection using nonnegative spectral analysis," in *AAAI*, 2012, pp. 1026–1032.

[35] M. Qian and C. Zhai, "Robust unsupervised feature selection," in *IJCAI*, 2013, pp. 1621–1627.

[36] L. Shi, L. Du, and Y. D. Shen, "Robust spectral learning for unsupervised feature selection," in *ICDM*, 2014, pp. 977–982.

[37] J. G. Dy and C. E. Brodley, "Feature selection for unsupervised learning," *J. Mach. Learn. Res.*, vol. 5, pp. 845–889, 2004.

[38] Y. Cheung, M. Li, Q. Peng, et al., "A cooperative learning-based clustering appraoch to lip segmentation without knowing segment number," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 1, pp. 80–93, 2017.

[39] H. Zeng and Y. Cheung, "Feature selection and kernel learning for local learning-based clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 8, pp. 1532–1547, 2011.

[40] F. Petitjean, A. Ketterlin, and P. Gancarski, "A global averaging method for dynamic time warping, with applications to clustering," *Pattern Recogn.*, vol. 44, no. 3, pp. 678–693, 2011.

[41] J. Yang and J. Leskovec, "Patterns of temporal variation in online media," in *WSDM*, 2011, pp. 177–186.

[42] P. Chen, K. Xu, G. Li, et al., "Local Fréchet distance in specific emitter identification," in *Proc. 2017 Int. Conf. Commun. Softw. Netw.*, 2017, pp. 842–845.

[43] A. Driemel, A. Krivosija, and C. Sohler, "Clustering time series under the Fréchet distance," in *SODA)*, 2016, pp. 766–785.

[44] J. Paparrizos and L. Gravano, "k-shape: Efficient and accurate clustering of time series," in *ACM SIGMOD*, 2015, pp. 1855–1870.

[45] Q. Ma, J. Zheng, S. Li, et al., "Learning representations for time series clustering," in *NeurIPS*, 2019, pp. 3776–3786.

[46] J. Zakaria, A. Mueen, and E. Keogh, "Clustering time series using unsupervised-shapelets," in *ICDM*, 2012, pp. 785–794.

[47] Q. Zhang, J. Wu, P. Zhang, et al., "Salient subsequence learning for time series clustering," *IEEE Trans. Pattern. Anal. Mach. Intell.*, vol. 41, no. 9, pp. 2193–2207,2019.

[48] C. Dai and D. Pi, "Parameter auto-selection for hemispherical resonator gyroscope's long-term prediction model based on cooperative game theory," *Knowl.-based Syst.*, vol. 134, pp. 105–115, 2017.

[49] V. K. Garg, Y. Narahari, and M. N. Murty, "Novel biobjective clustering (BiGC) based on cooperative game theory," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 5, pp. 1070–1082, 2013.

[50] L. S. Shapley, "A value for n-Person games," in *Proc. Annu. of Math. Contributions to the Theory of Games*, 1953, pp. 307–317.

[51] L. S. Shapley, "Cores of convex games," *Int. J. Game Theory*, vol. 1, no. 1, pp. 11–26, 1971.

[52] J. W. Cooley and J. W. Tukey, "An algorithm for the machine calculation of complex Fourier series," *Math. Comput.*, vol. 19, no. 90, pp. 297–301, 1965.

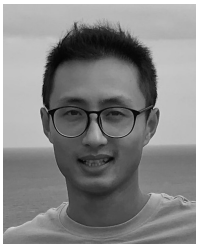
[53] L. Hubert and P. Arabie, "Comparing partitions," *J. Classif.*, vol. 2, no. 1, pp. 193–218, 1985.

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

[55] J. L. Fleiss, "Measuring nominal scale agreement among many raters," *Psychol. Bull.*, 1971, vol. 76, no. 5, pp. 378–382.

[56] J. Kleinberg, "An impossibility theorem for clustering," in *NIPS*, 2002, pp. 463–470.

[57] T. C. Fonseca and M. A. Ferreira, "Dynamic multiscale spatiotemporal models for poisson data," *J. Am. Stat. Assoc.*, 2017, vol. 112, no. 517, pp. 215–234.



Chenglong Dai received the master's degree in computer science and technology from Nanjing University of Aeronautics and Astronautics (NUAA), Nanjing, China, in 2014. He is working toward the PhD degree in Computer Science and Technology in NUAA. His research interests include EEG processing, EEG analyzing, and data mining. He has published several related high-quality papers in journals like IEEE TCYB, ACM TKDD and ACM TIST, and top conferences like SDM (awarded the Best Paper Award in Data Science Track). He also has served as a reviewer for IJCNN'18, IJCNN'19, and IJCNN'20.

Jia Wu (M'16) received the Ph.D. degree in computer science from the University of Technology Sydney, Ultimo, NSW, Australia. He is currently a Lecturer in the Department of Computing, Macquarie University, Sydney and a Chair Professor in School of Computer Science, Wuhan University, China. His current research interests include data mining and machine learning. Since 2009, he has published 100+ refereed journal and conference papers, including TPAMI, TKDE, TNNLS, ACM TKDD, IJCAI, AAAI, ICDM, and SIAM SDM.

Dr Wu was the recipient of SDM'18 Best Paper Award in Data Science Track, IJCNN'17 Best Student Paper Award, and ICDM'14 Best Paper Candidate Award. He is the Associate Editor of the ACM TKDD, *Journal of Network and Computer Applications* and *Neural Networks*.

Dechang Pi is a professor in the School of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics (NUAA), China. He has published 7 research books and more than 100 papers in journals and conferences, won over 20 grants including the National Science Foundation of China, Technology Foundation of National Defence, Aviation Science Foundation of China, and so on. His main research interests lie in the area of big data including data mining, EEG data preprocessing.

Lin Cui received the master's degree from the School of Computer Science and Information Engineering, Hefei University of Technology (HFUT), China, in 2008. She is currently a professor at Intelligent Information Processing Laboratory, Suzhou University, Anhui, China. Her research interests include data mining, POI recommendation, and social network analysis.

Blake Johnson is associate professor in the Department of Cognitive Science at Macquarie University, Sydney, Australia, a chief investigator in the Australian Research Council Centre of Excellence for Cognition and its Disorders (CCD) and a project leader in the Australian Government's Hearing Cooperative Research Centre (Hearing CRC). His neuroimaging research investigates typical and atypical human brain function using a variety of brain measurement techniques including EEG and MEG (magnetoencephalography). Current work on the development of speech motor control in children is funded by the Australian Research Council.

Stefanie I. Becker is an Associate Professor and ARC Future Fellow at the University of Queensland, Brisbane, Australia. She is an internationally recognised expert on visual attention, and received several awards for her work involving eye tracking, EEG, and fMRI. Since her PhD in 2007, she has authored or co-authored over 60 papers in high-ranking journals, and served as an associate editor for the journal JEP-HPP.