EEG-BASED EMOTION CLASSIFICATION USING GRAPH SIGNAL PROCESSING

Seyed Saman Saboksayr⋆, Gonzalo Mateos⋆, and Mujdat Cetin⋆†

⋆Department of Electrical and Computer Engineering, University of Rochester, Rochester, NY, USA
†Goergen Institute for Data Science, University of Rochester, Rochester, NY, USA

ABSTRACT

The key role of emotions in human life is undeniable. The question of whether there exists a brain pattern associated with a specific emotion is the theme of many affective neuroscience studies. In this work, we bring to bear graph signal processing (GSP) techniques to tackle the problem of automatic emotion recognition using brain signals. GSP is an extension of classical signal processing methods to complex networks where there exists an inherent relation graph. With the help of GSP, we propose a new framework for learning class-specific discriminative graphs. To that end, firstly we assume for each class of observations there exists a latent underlying graph representation. Secondly, we consider the observations are smooth on their corresponding class-specific sough graph while they are non-smooth on other classes' graphs. The learned class-specific graph-based representations can act as sub-dictionaries and be utilized for the task of emotion classification. Applying the proposed method on an electroencephalogram (EEG) emotion recognition dataset indicates the superiority of our framework over other state-of-the-art methods.

***Index Terms*—** Graph learning, graph signal processing, discriminative transform learning, emotion recognition, EEG.

1. INTRODUCTION

Over the years, many real-world problems have been tackled via traditional signal processing (SP) approaches. *Graph signal processing* (GSP) has been introduced to extend the knowledge of classical SP to complex networks wherein the data rely on an underlying graph different from the chain or grid structures assumed by traditional SP [1]. The underlying graph captures the connections between entities of the network. In some networks, the inherent topology is known *a priori* (e.g., sensor networks) or directly observable (e.g., social networks) [2]. However, this is not the case in many networks, e.g., brain networks. Therefore, it is of interest to learn the underlying graph using the data associated with the network. Furthermore, the prosperity of GPS depends on how accurate the learned graph represents the underlying relations in the network.

The so-called task of *graph learning* or *topology inference* aims to estimate the underlying relationships of the *graph signals*. It is known that the graph learning based on observed signals is an *NP-hard combinatorial* problem. Hence, solving graph learning requires applying some constraints to the problem. Such constraints encode information about the structure of the network. Based on the literature, common constraints that have been used include *smoothness* and *sparsity* [3,4]. Smoothness constraint implies that the graph signals that are related to each other behave similarly. Recently, *bandlimitedness* assumption has been taken into account in several studies [5,6]. Bandlimited graph signals admit a sparse representation in the graph spectral domain. By considering graph signals as random vectors drawn from a Gaussian Markov random field distribution, the graph learning problem becomes the estimation of the inverse covariance matrix [7,8].

An important aspect of human life is emotion since emotions play an important role in human decision makings and communications [9]. Moreover, investigating whether there exists a specific pattern of brain connectivity related to a particular emotion has become a key challenge in affective neuroscience. As mentioned before, the human brain is recognized as a complex network and the underlying graphs for particular cognitive, perceptual, or emotional processes are unknown. However, one can try to exploit GSP tools to capture brain patterns from the brain signals and classify human emotions such as valence and arousal using these patterns. Electroencephalogram (EEG) signals are recorded in a non-invasive manner and are more affordable. Therefore, the vast majority of studies in the field of emotion recognition use EEG signals [10–12].

Relation to prior work and contributions. While studies such as [3,4] focus on learning the underlying graph that is efficient with respect to the signal representation; other studies focus on applications involving classification and try to learn a representation from the signal to improve the classification result. Typically the goal is to define a class-specific sub-dictionary by using the signals of the corresponding class [8]. Previously, studies such as [13] suggested defining a common graph for all classes and learn class-specific graph transforms based on the signals in each class which play the role of sub-dictionaries. The lack of discrimination between the class-specific graph transforms, which comes from using only one graph for all classes, may reduce the classification performance [8]. The distinctive goal here is that of discriminative transform learning, to effectively tackle a classification problem involving network data. The aim of our proposed topology inference method is to take into account the discriminability while maintaining efficient signal representation. Different from [3], our formulation enforces that each classes' signals to be smooth on their class-specific learned graph while they are non-smooth on other classes' graphs. The discriminative graph learning problem studied here was first formulated in [8] but under the lens of probabilistic graphical model selection. Therein, graph signals are viewed as random vectors adhering to a Gaussian Markov random field distribution, where the unknown class-specific precision matrices typically play the role of graph Laplacians. However, the discriminative graphical lasso estimator in [8] is not guaranteed to return a valid graph Laplacian for each class, since the search is performed over the whole positive semi-definite cone. Incorporating Laplacian constraints may challenge the block coordinate-descent algorithm in [8]. Accordingly, one misses on the GSP insights offered here in terms of signal smoothness and bandlimitedness in the graph spectral domain.

Work in this paper was supported by the NSF awards CCF-1750428, CCF-1934962 and ECCS-1809356.

2. PRELIMINARIES

We define the graph signal $\mathbf{x} = [x_1, \ldots, x_N]^\top \in \mathbb{R}^N$ over a weighted, undirected graph $\mathcal{G}(\mathcal{V}, \mathcal{E}, \mathbf{W})$, where $\mathcal{V} = \{1, \ldots, N\}$ represents the node set of cardinality N and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ the set of edges. In the context of EEG signal processing, graph signals correspond to the EEG signals and the graph represents the brain functional connectivity. Accordingly nodes stand for the EEG electrodes. Following this notation, $x_i \in \mathbb{R}$ and $\mathbf{W} \in \mathbb{R}_+^{N \times N}$ denote the signal value at node $i \in \mathcal{V}$ and the adjacency matrix of edge weights, respectively. The symmetric and non-negative coefficients $W_{ij} = W_{ji} \in \mathbb{R}_+$ indicate the strength of the connection (or similarity) between electrode i and electrode j. In the absence of connection [i.e., $(i,j) \notin \mathcal{E}$] one has $W_{ij} = 0$. Moreover, we assume that $\mathcal{G}$ does not include any self-loops which implies $W_{ii} = 0$, $\forall i \in \mathcal{V}$. The adjacency matrix $\mathbf{W}$ encodes the brain functional connectivity. Beyond the adjacency matrix $\mathbf{W}$, results in spectral graph theory often motivate choosing the combinatorial graph Laplacian $\mathbf{L} := \text{diag}(\mathbf{W}\mathbf{1}) - \mathbf{W}$, where $\mathbf{1}$ refers to the all-one vector. In particular, $\mathbf{L}$ plays a central role in defining a useful and intuitive graph Fourier transform (GFT) as described next.

Graph Fourier transform and signal smoothness. In order to introduce the network's spectral basis and define the GFT, we decompose the (symmetric and positive semi-definite) combinatorial graph Laplacian as $\mathbf{L} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^\top$, where $\mathbf{\Lambda} := \text{diag}(\lambda_1, \ldots, \lambda_N)$ denotes the diagonal matrix of non-negative eigenvalues and $\mathbf{V} := [\mathbf{v}_1, \ldots, \mathbf{v}_N]$ the orthonormal matrix of eigenvectors. The GFT of $\mathbf{x}$ with respect to $\mathbf{L}$ is the signal $\tilde{\mathbf{x}} := \mathbf{V}^\top \mathbf{x}$. The inverse (i)GFT of $\tilde{\mathbf{x}} := [\tilde{x}_1, \ldots, \tilde{x}_N]^\top$ is given by $\mathbf{x} = \mathbf{V}\tilde{\mathbf{x}} = \sum_{k=1}^N \tilde{x}_k \mathbf{v}_k$, which is a proper inverse due to the orthonormality of $\mathbf{V}$. The GFT encodes a notion of signal variability over $\mathcal{G}$ (akin to frequency in Fourier analysis of temporal signals) by synthesizing $\mathbf{x}$ as a sum of orthogonal frequency components $\mathbf{v}_k$. The GFT coefficient $\tilde{x}_k$ is the contribution of $\mathbf{v}_k$ to the graph signal $\mathbf{x}$.

To elaborate on the notion of frequency for graph signals, consider the total variation (or *Dirichlet energy*) of $\mathbf{x}$ with respect to the combinatorial graph Laplacian $\mathbf{L}$ defined as

$$\text{TV}(\mathbf{x}) := \mathbf{x}^\top \mathbf{L}\mathbf{x} = \sum_{i \neq j} W_{ij} (x_i - x_j)^2 = \sum_{k=1}^N \lambda_k \tilde{x}_k^2. \quad (1)$$

The quadratic form (1) acts as a smoothness measure, because it effectively quantifies how much the graph signal $\mathbf{x}$ changes with respect to $\mathcal{G}$'s topology [14]. If we evaluate the total variation of eigenvector $\mathbf{v}_k$ (itself a graph signal), one immediately obtains $\text{TV}(\mathbf{v}_k) = \lambda_k$. Accordingly, the Laplacian eigenvalues $0 = \lambda_1 < \lambda_2 \leq \cdots \leq \lambda_N$ can be viewed as graph frequencies indicating how the eigenvectors (i.e., frequency components) vary with respect to $\mathcal{G}$. Smoothness is a cardinal property of many real-world network processes. The last equality in (1) suggests that smooth (or bandlimited) signals admit a sparse representation in the graph spectral domain. Intuitively, they tend to be spanned by a few Laplacian eigenvectors associated with small eigenvalues.

Learning graphs from observations of smooth signals. Consider the following network topology identification problem. Given a set $\mathcal{X} := \{\mathbf{x}_p\}_{p=1}^P$ of possibly noisy graph signal observations from P trials, the goal is to learn an undirected graph $\mathcal{G}(\mathcal{V}, \mathcal{E}, \mathbf{W})$ with $|\mathcal{V}| = N$ nodes such that the observations in $\mathcal{X}$ are smooth on $\mathcal{G}$. In this section, we review the solution proposed in [3], that we build on in the rest of the paper.

Given $\mathcal{X}$, one can form the data matrix $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_P] \in \mathbb{R}^{N \times P}$, and let $\bar{\mathbf{x}}_i^\top \in \mathbb{R}^{1 \times P}$ denote its i-th row collecting those P measurements at vertex i. The key idea in [3] is to establish a link between smoothness and sparsity, namely

$$\sum_{p=1}^P \text{TV}(\mathbf{x}_p) = \text{trace}(\mathbf{X}^\top \mathbf{L}\mathbf{X}) = \frac{1}{2}\|\mathbf{W} \circ \mathbf{Z}\|_1, \quad (2)$$

where $\circ$ denotes the Hadamard (element-wise) product and the Euclidean-distance matrix $\mathbf{Z} \in \mathbb{R}_+^{N \times N}$ has entries $Z_{ij} := \|\bar{\mathbf{x}}_i - \bar{\mathbf{x}}_j\|^2$, $i, j \in \mathcal{V}$. The intuition is that when the given distances in $\mathbf{Z}$ come from a smooth manifold, the corresponding graph has a sparse edge set, with preference given to edges (i,j) associated with smaller distances Z_{ij}.

In this context, a general purpose framework for learning graphs under a smoothness prior is advocated in [3], which entails solving

$$\begin{aligned} &\min_{\mathbf{W}} \ \|\mathbf{W} \circ \mathbf{Z}\|_1 + f(\mathbf{W}) \quad (3) \\ &\text{s. t.} \quad \text{diag}(\mathbf{W}) = \mathbf{0},\ W_{ij} = W_{ji} \geq 0,\ i \neq j. \end{aligned}$$

The convex objective function $f(\mathbf{W})$ augments the smoothness criterion $\|\mathbf{W} \circ \mathbf{Z}\|_1$, and several choices have been proposed to, e.g., recover common graph constructions based on the Gaussian kernel [15], accommodate time-varying graphs [16], or to scale other related graph learning algorithms [4]. Identity (2) offers a favorable way of formulating the inverse problem (3), because the space of adjacency matrices can be described via simpler (meaning entry-wise decoupled) constraints relative to its Laplacian counterpart. As a result, the convex optimization problem (3) can be solved efficiently with complexity $\mathcal{O}(N^2)$ per iteration, by leveraging provably-convergent primal-dual solvers amenable to parallelization [17].

3. DISCRIMINATIVE GRAPH LEARNING

In an effort to address a classification problem involving network data, we bring to bear GSP insights to learn a discriminative graph-based representation of the signals; see e.g., [8, 18]. To this end, a new formulation for the graph learning problem is proposed which can be solved using primal-dual algorithms. After this training phase, the GFTs of the optimum graphs can be used to extract discriminative features of the test signals we wish to classify.

Problem statement. Consider a dataset $\mathcal{X} = \bigcup_{c=1}^C \mathcal{X}_c$ comprising labeled graph signals $\mathcal{X}_c := \{\mathbf{x}_p^{(c)}\}_{p=1}^{P_c}$ from C different classes e.g., high and low valence. The signals in each class possess a very distinctive structure, namely they are assumed to be smooth (or bandlimited) with respect to unknown class-specific graphs $\mathcal{G}_c = (\mathcal{V}, \mathcal{E}_c, \mathbf{W}_c)$, $c = 1, \ldots, C$. This notion is analogous to a multiple linear subspace model in which graph signals of each class are assumed to be spanned by a few vectors (namely, the basis of the corresponding low-dimensional subspace) [19]. In fact, the discussion in Section 2 implies one can equivalently restate the class-conditional signal model assumptions as follows: there is a network $\mathcal{G}_c$ such that graph signals in $\mathcal{X}_c$ are spanned by a few Laplacian eigenvectors (associated with small eigenvalues), for $c = 1, \ldots, C$. Similar to [8], given $\mathcal{X}$ the goal is to learn the class-specific adjacency matrices $\mathbf{W}_c$ under signal smoothness priors, so that the obtained GFT bases can be subsequently used to classify unseen (and unlabeled) graph signals effectively.

Our approach blends elements and ideas from the graph learning framework in [3] [cf. (3)] along with the discriminative graphical lasso estimator [8]. Indeed, the algorithm in [3] optimizes network topology recovery under smoothness assumptions, but is otherwise

agnostic to the performance of a potential downstream (say, classification) task the learned graph may be integral to. Inspired by [8], we seek graph representations that capture the underlying network topology (i.e., the class structure), but at the same time are discriminative to boost classification performance. To this end, we propose to learn a graph representation $\mathcal{G}_c$ per class by solving the following convex optimization problems [cf. (3)]

$$\min_{\mathbf{W}_c \in \mathcal{W}_m} \|\mathbf{W}_c \circ \mathbf{Z}_c\|_1 - \alpha \mathbf{1}^\top \log(\mathbf{W}_c \mathbf{1}) + \beta \|\mathbf{W}_c\|_F^2 - \gamma \sum_{k \neq c}^{C} \|\mathbf{W}_c \circ \mathbf{Z}_k\|_1, \quad (4)$$

where $\mathbf{W}_c$ is the adjacency matrix of $\mathcal{G}_c$, constrained to the set $\mathcal{W}_m = \{\mathbf{W} \in \mathbb{R}_+^{N \times N} : \mathbf{W} = \mathbf{W}^\top, \text{diag}(\mathbf{W}) = 0\}$. Moreover, $\mathbf{Z}_c$ is the distance matrix constructed from class c signals $\mathbf{X}_c := [\mathbf{x}_1^{(c)}, \ldots, \mathbf{x}_{P_c}^{(c)}] \in \mathbb{R}^{N \times P_c}$, while α, β, and γ are positive regularization parameters.

Taking a closer look at the objective function, minimizing $\|\mathbf{W}_c \circ \mathbf{Z}_c\|_1$ encourages a graph $\mathbf{W}_c$ over which the signals in $\mathcal{X}_c$ are smooth. At the same time, the last term enforces non-smoothness of the signals in the other $C-1$ classes. This composite criterion will thus induce a GFT with better discrimination ability than the $\gamma = 0$ case – the energy of class c signals will be predominantly concentrated in lower frequencies, while the spectral content of the other classes is pushed towards high-pass regions of the spectrum. Under Gaussianity assumptions, this can be interpreted as a Fisher discrimination criterion used in Linear Discriminant Analysis (LDA) [15, Ch. 4.3], which entails minimization of the inter-class scatter as well as maximizing the intra-class scatter; see [8].

The logarithmic barrier on the nodal degree sequence $\mathbf{W}_c \mathbf{1}$ precludes the trivial all-zero solution. Moreover, it ensures the estimated graph is devoid of isolated vertices. The Frobenius-norm regularization on the adjacency matrix $\mathbf{W}_c$ controls the graph's edge sparsity pattern by penalizing larger edge weights (the sparsest graph is obtained for $\beta = 0$). Overall, this combination forces degrees to be positive but does not prevent most individual edge weights from becoming zero [3].

Primal-dual graph learning algorithm. Following [3], one can solve (4) based on primal-dual techniques; see [17] for a tutorial treatment on primal-dual methods and their applications. To make (4) amenable to this optimization method, recall that the adjacency matrix $\mathbf{W}_c \in \mathcal{W}$ is symmetric with diagonal elements equal to zero. Therefore, the independent decision variables are effectively the upper-triangular elements $[\mathbf{W}_c]_{ij}$, $j > i$, which we collect in the vector $\mathbf{w}_c \in \mathbb{R}_+^{\frac{N(N-1)}{2}}$. Given all these considerations one can solve (4) via Algorithm 1, where $\mathbf{z}_c$ is a vector containing the upper-triangular entries of $\mathbf{Z}_c$ and $\mathbf{S} \in \{0,1\}^{N \times \frac{N(N-1)}{2}}$ is a matrix that satisfies $\mathbf{d}_c = \mathbf{W}_c \mathbf{1} = \mathbf{S}\mathbf{w}_c$. Algorithm 1 is guaranteed to converge when the step size is $\mu \in \left(0, 1 + 4\beta + \sqrt{2(N-1)}\right)$ [17].

Classification via low-pass graph filtering. During the training phase of the classification task, the goal is to learn C class-specific graphs $\mathcal{G}_c$ from labeled graph signals $\mathcal{X}_c := \{\mathbf{x}_p^{(c)}\}_{p=1}^{P_c}$. This can be accomplished by running C parallel instances of Algorithm 1. Let $\hat{\mathbf{W}}_c$ denote the estimated adjacency matrix of the graph representing class C, and likewise let $\hat{\mathbf{L}}_c = \text{diag}(\hat{\mathbf{W}}_c \mathbf{1}) - \hat{\mathbf{W}}_c$ be the combinatorial graph Laplacian. Finally, let $\hat{\mathbf{V}}_c$ denote the orthonormal GFT basis of Laplacian eigenvectors for class C; see Section 2.

In the operational or test phase, we are presented with an unseen and unlabeled graph signal $\mathbf{x}$ which we wish to classify into one of the C classes. To that end, we will process $\mathbf{x}$ with a filter-bank comprising C graph filters. The c-th branch yields the graph-frequency domain output $\tilde{\mathbf{x}}_{F,c} = \text{diag}(\tilde{\mathbf{h}})\tilde{\mathbf{x}}_c = \tilde{\mathbf{h}} \circ \tilde{\mathbf{x}}_c$, where $\tilde{\mathbf{x}}_c$ are the GFT coefficients of $\mathbf{x}$ with respect to graph $\mathcal{G}_c$, and $\tilde{\mathbf{h}} = [\tilde{h}_1, \ldots, \tilde{h}_N]^\top$ is the frequency response of an ideal low-pass filter with bandwidth $w \in \{1, 2, \ldots, N\}$, i.e., $\tilde{h}_i := \mathbb{I}\{i \leq w\}$. Typically, one chooses the tunable parameter w to be $N/2$ or smaller in order to implement a low-pass filter. Notice that while the frequency response $\tilde{\mathbf{h}}$ is the same for all C branches, the graph filters $\mathbf{H}_c = \mathbf{V}_c \text{diag}(\tilde{\mathbf{h}}) \mathbf{V}_c^\top$ differ because the learned graphs (hence the GFT transforms) vary across classes. From the definition of $\tilde{\mathbf{h}}$, it immediately follows that $\tilde{\mathbf{x}}_{F,c}$ is nothing else than the projection of $\mathbf{x}$ onto the eigenvectors of $\hat{\mathbf{L}}_c$ corresponding to the smallest w eigenvalues.

If $\mathbf{x}$ belongs to class $c^\star$, say, then this graph signal should be smoothest with respect to $\mathcal{G}_{c^\star}$. Equivalently, for fixed (appropriately low) bandwidth w we expect the signal power to be largest when projected onto the GFT basis constructed from $\hat{\mathbf{L}}_c$. Accordingly, the adopted classification rule is simply $\hat{c} = \text{argmax}_c \{\|\tilde{\mathbf{x}}_{F,c}\|_2^2\}$. A classification error occurs whenever $\hat{c} \neq c^\star$.

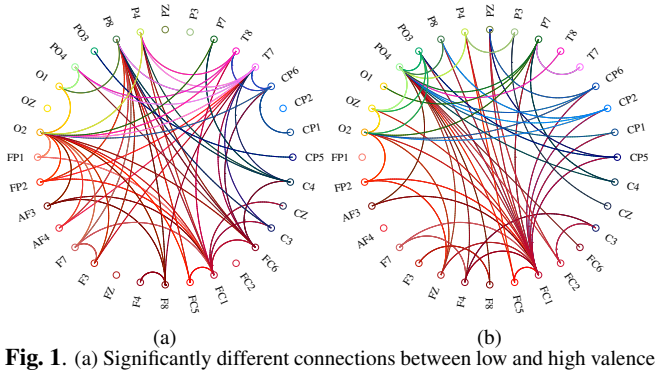
Algorithm 1: Primal-dual for graph learning, class c

Input parameters $\alpha, \beta, \gamma, \mu$, data $\{\mathbf{z}_c\}_{c=1}^C$, initial $\mathbf{w}_c, \mathbf{d}_c$.
Set $\bar{\boldsymbol{\mu}} = 2\mu \left(\mathbf{z}_c - \gamma \sum_{k \neq c}^{C} \mathbf{z}_k\right)$.
while *not converged* **do**
  $\mathbf{y}_1 = \mathbf{w}_c - \mu(2\beta \mathbf{w}_c + \mathbf{S}^\top \mathbf{d}_c)$.
  $\mathbf{y}_2 = \mathbf{d}_c + \mu \mathbf{S}\mathbf{w}_c$.
  $\mathbf{p}_1 = \max(0, \mathbf{y}_1 - \bar{\boldsymbol{\mu}})$.
  $\mathbf{p}_2 = (\mathbf{y}_2 + \sqrt{\mathbf{y}_2^2 + 4\alpha\mu\mathbf{1}})/2$.
  $\mathbf{q}_1 = \mathbf{p}_1 - \mu(2\beta \mathbf{p}_1 + \mathbf{S}^\top \mathbf{p}_2)$.
  $\mathbf{q}_2 = \mathbf{p}_2 + \mu \mathbf{S}\mathbf{p}_1$.
  $\mathbf{w}_c \leftarrow \mathbf{w}_c - \mathbf{y}_1 + \mathbf{q}_1$.
  $\mathbf{d}_c \leftarrow \mathbf{d}_c - \mathbf{y}_2 + \mathbf{q}_2$.
end

4. EEG EMOTION RECOGNITION

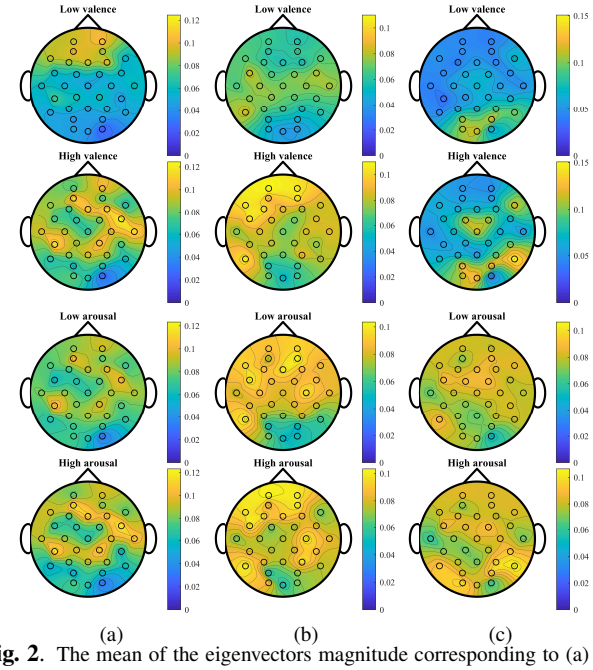
In this section, we study the application of the proposed discriminative graph learning algorithm in the context of emotion recognition using EEG signals. To this end, we apply our method to a widely used and publicly available EEG data-set called DEAP [20]. The DEAP data-set contains EEG and peripheral physiological signals of 32 participants. The data were recorded while subjects were watching one-minute long music videos. Each participant has 40 trials (music videos) and rates each video in terms of the levels of valence, arousal, like/dislike, dominance, and familiarity [20]. In this study, we focus on the valence and arousal classification. Ratings are decimal numbers between 1 and 9 and in order to make this a binary classification task we divide the ratings into two classes of *low* when the ratings are smaller than 5, and *high* when ratings are larger than or equal to 5. We exploit the pre-processed version of the data-set which contains 32 EEG channels with 128 Hz sampling rate and the 3 second pre-trial baseline is discarded.

We perform the classification task in leave-one-trial-out scheme where for each subject we use 39 trials as the training set and test on the one remaining trial. We repeat this 40 times and report the mean accuracy. This is a subject dependent procedure which means that we do the classification (training and testing) for each subject separately. Classification follows the procedure described in Section 3



(a) (b)

Fig. 1. (a) Significantly different connections between low and high valence with $p \leq 0.002$. (b) Significantly different connections between low and high arousal with $p \leq 0.03$. These results suggest the learned graphs for low and high emotional responses show significant difference with each other.



(a) (b) (c)

Fig. 2. The mean of the eigenvectors magnitude corresponding to (a) low, (b) mid, and (c) high frequency for valence (top two rows) and arousal (bottom two rows). These results demonstrate the different patter between low and high emotions, where most of these differences are aligned with the literature.

and we project the test signals on the first $1/4$ of the GFT basis ($w = N/4$). Note that we did a grid search to find the best regularization parameter and the results of the paper are the best parameters that we find. The discriminative graph is learned for the training set and then normalized to have unit Frobenius norm. The average classification accuracy over all the trials and all the participants is reported in Table 1. The results of the other state-of-the-art methods are also reported for the comparison. Results of the discriminative graphical lasso method in [8] are omitted since the provided code did not converge for some of the DEAP dataset trials. As it is shown in Table 1, the proposed discriminative graph learning approach outperforms state-of-the-art methods in the classification of emotions. Moreover, we can see 6 and 5 percent improvement in the classification of valence and arousal, respectively with respect to the method in [3] which is the foundation of our model. That improvement is because of the added discriminative term.

Now that we established the superior performance of our proposed framework, it is of interest to investigate whether there is any useful pattern in the underlying learned graph. To this end, we first discard the trials that have ratings in $(4.5, 5.5)$ where the participant is not confident enough in rating the trials. Then we study the connections that are significantly different between classes. Finally, we illustrate the decomposed signals and the eigenvectors with respect to the low, mid, and high frequency in order to investigate whether our findings are aligned with the literature.

Table 1. The classification accuracy averaged over all participants.

Study	Valence	Arousal
Proposed method	**92.73**	**93.44**
Kalofolias [3]	86.56	88.91
Chao and Liu [11]	77.02	76.13
Rozgić *et al.* [21]	76.90	69.10
Chen *et al.* [22]	76.17	73.59
Tripathi *et al.* [12]	81.40	73.36

Accordingly, we learn two graphs corresponding to low and high emotions per person, using the parameters that gave us the best results in the classification. Interestingly, it seems that the edge weights are related to the intensity of the emotions, i.e., valence and arousal. In Fig. 1, we show the significantly different connections between low and high emotional responses. The statistical way to check whether the learned representations are discriminative is through analyzing the significantly different connections. Here, we apply a non-parametric test of the null hypothesis called Wilcoxon rank-sum test [23]. Fig. 1(a,b) show connections that are significantly different between low and high valence with $p \leq 0.002$ and between low and high arousal with $p \leq 0.03$, respectively. The original paper of the DEAP data-set indicates the significant channels [20, Table 4] which most of them are part of the significantly different connections that we capture. Frequency analysis is the common theme in the EEG signal processing literature. Therefore, we conduct a similar analysis with the help of GFT and the underlying learned graph. Fig. 2(a,b,c) demonstrate the average of the magnitude of the eigenvector sets associated with low (first $1/4$ components), mid, and high (last $1/4$ components) frequencies, respectively. The asymmetrical pattern of the frontal EEG activity can be seen in Fig. 2(a) which is consistent with the findings in [24]. Also, as it is noted in Fig 1(a) most of the connections are related to the frontal lobe (red-shaded colors). Moreover, the authors in [25] discovered that for classifying positive from negative emotions, the features are generally in the right occipital lobe and parietal lobe for the alpha band, the central lobe for the beta band, and the left frontal and right temporal lobes for the gamma band. Since we do the classification via low-frequency components, the employed features are mainly in the low-frequency representation Fig. 2(a) and evidently there are different patterns in the mentioned areas such as left frontal, right temporal, central, and parietal lobes between high and low valence/arousal. Moreover, in high frequencies, we observe different patterns in the right parietal and left occipital lobes.

5. CONCLUSION

In this work, we proposed a novel graph learning framework from smooth signals by proposing an objective function that contains a discriminative term. Results of the EEG emotion recognition experiments show that the proposed approach outperforms state-of-the-art algorithms in emotion classification, while it recovers interpretable graphs offering insights into the structure of the data classes.

6. REFERENCES

[1] A. Ortega, P. Frossard, J. Kovačević, J. M. F. Moura, and P. Vandergheynst, "Graph signal processing: Overview, challenges, and applications," *Proc. IEEE*, vol. 106, no. 5, pp. 808–828, 2018.

[2] G. Mateos, S. Segarra, A. G. Marques, and A. Ribeiro, "Connecting the dots: Identifying network structure via graph signal processing," *IEEE Signal Process. Mag.*, vol. 36, no. 3, pp. 16–43, 2019.

[3] V. Kalofolias, "How to learn a graph from smooth signals," in *Artif. Intel. and Stat. (AISTATS)*, 2016, pp. 920–929.

[4] X. Dong, D. Thanou, P. Frossard, and P. Vandergheynst, "Learning Laplacian matrix in smooth graph signal representations," *IEEE Trans. Signal Process.*, vol. 64, no. 23, pp. 6160–6173, 2016.

[5] S. Sardellitti, S. Barbarossa, and P. Di Lorenzo, "Graph topology inference based on sparsifying transform learning," *IEEE Trans. Signal Process.*, vol. 67, no. 7, pp. 1712–1727, 2019.

[6] B. Le Bars, P. Humbert, L. Oudre, and A. Kalogeratos, "Learning laplacian matrix from bandlimited graph signals," in *IEEE Intl. Conf. Acoust., Speech and Signal Process. (ICASSP)*, 2019, pp. 2937–2941.

[7] J. Friedman, T. Hastie, and R. Tibshirani, "Sparse inverse covariance estimation with the graphical lasso," *Biostatistics*, vol. 9, no. 3, pp. 432–441, 12 2007.

[8] J. Kao, D. Tian, H. Mansour, A. Ortega, and A. Vetro, "Disc-GLasso: Discriminative graph learning with sparsity regularization," in *IEEE Intl. Conf. Acoust., Speech and Signal Process. (ICASSP)*, 2017, pp. 2956–2960.

[9] W. A. Cunningham and T. Kirkland, "Emotion, cognition, and the classical elements of mind," *Emotion Review*, vol. 4, no. 4, pp. 369–370, 2012.

[10] W. Zheng, J. Zhu, and B. Lu, "Identifying stable patterns over time for emotion recognition from EEG," *IEEE Trans. Affect. Comput.*, vol. 10, no. 3, pp. 417–429, 2019.

[11] H. Chao and Y. Liu, "Emotion recognition from multi-channel EEG signals by exploiting the deep belief-conditional random field framework," *IEEE Access*, vol. 8, pp. 33002–33012, 2020.

[12] S. Tripathi, S. Acharya, R. D. Sharma, S. Mittal, and S. Bhattacharya, "Using deep and convolutional neural networks for accurate emotion classification on DEAP dataset," in *Proc. of the Thirty-First AAAI Conference on Artificial Intelligence*, 2017, pp. 4746–4752.

[13] D. Thanou, D. I. Shuman, and P. Frossard, "Learning parametric dictionaries for signals on graphs," *IEEE Trans. Signal Process.*, vol. 62, no. 15, pp. 3849–3862, 2014.

[14] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 83–98, 2013.

[15] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, Springer, New York, 2 edition, 2009.

[16] V. Kalofolias, A. Loukas, D. Thanou, and P. Frossard, "Learning time varying graphs," in *IEEE Intl. Conf. Acoust., Speech and Signal Process. (ICASSP)*, 2017, pp. 2826–2830.

[17] N. Komodakis and J. Pesquet, "Playing with duality: An overview of recent primal–dual approaches for solving large-scale optimization problems," *IEEE Signal Process. Mag.*, vol. 32, no. 6, pp. 31–54, 2015.

[18] D. Thanou, D. I. Shuman, and P. Frossard, "Parametric dictionary learning for graph signals," in *IEEE Global Conf. Signal and Info. Process. (GlobalSIP)*, 2013, pp. 487–490.

[19] R. Vidal, "Subspace clustering," *IEEE Signal Process. Mag.*, vol. 28, no. 2, pp. 52–68, 2011.

[20] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "DEAP: A database for emotion analysis using physiological signals," *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 18–31, 2011.

[21] V. Rozgić, S. N. Vitaladevuni, and R. Prasad, "Robust EEG emotion classification using segment level decision fusion," in *IEEE Intl. Conf. Acoust., Speech and Signal Process. (ICASSP)*, 2013, pp. 1286–1290.

[22] M. Chen, J. Han, L. Guo, J. Wang, and I. Patras, "Identifying valence and arousal levels via connectivity between EEG channels," in *Intl. Conf. Affect. Comput. and Intel. Interaction (ACII)*, 2015, pp. 63–69.

[23] M. P. Fay and M. A. Proschan, "Wilcoxon-Mann-Whitney or t-test? on assumptions for hypothesis tests and multiple interpretations of decision rules," *Statistics surveys*, vol. 4, pp. 1, 2010.

[24] L. A. Schmidt and L. J. Trainor, "Frontal brain electrical activity (EEG) distinguishes valence and intensity of musical emotions," *Cognition & Emotion*, vol. 15, no. 4, pp. 487–500, 2001.

[25] D. Nie, X. Wang, L. Shi, and B. Lu, "EEG-based emotion recognition during watching movies," in *Intl. IEEE/EMBS Conf. Neural Eng.*, 2011, pp. 667–670.