Personalized Emotion Recognition by Personality-aware High-order Learning of Physiological Signals*

SICHENG ZHAO, Tsinghua University, China and University of California Berkeley, USA
AMIR GHOLAMINEJAD, University of California Berkeley, USA
GUIGUANG DING, Tsinghua University, China
YUE GAO, Tsinghua University, China
JUNGONG HAN, Lancaster University, UK
KURT KEUTZER, University of California Berkeley, USA

Due to the subjective responses of different subjects to physical stimuli, emotion recognition methodologies from physiological signals are increasingly becoming personalized. Existing works mainly focused on modelling the involved physiological corpus of each subject, without considering the psychological factors, such as interest and personality. The latent correlation among different subjects has also been rarely examined. In this paper, we propose to investigate the influence of personality on emotional behavior in a hypergraph learning framework. Assuming that each vertex is a compound tuple (subject, stimuli), multi-modal hypergraphs can be constructed based on the personality correlation among different subjects and on the physiological correlation among corresponding stimuli. To reveal the different importance of vertices, hyperedges and modalities, we learn the weights for each of them. As the hypergraphs connect different subjects on the compound vertices, the emotions of multiple subjects can be simultaneously recognized. In this way, the constructed hypergraphs are vertex-weighted multi-modal multi-task ones. The estimated factors, referred to as emotion relevance, are employed for emotion recognition. We carry out extensive experiments on the ASCERTAIN dataset and the results demonstrate the superiority of the proposed method, as compared to the state-of-the-art emotion recognition approaches.

CCS Concepts: • **Human-centered computing** → *Human computer interaction (HCI)*; • **Computing methodologies** → *Supervised learning by classification*; • **Applied computing** → *Psychology*;

Additional Key Words and Phrases: Personalized emotion recognition, personality-sensitive learning, physiological signal analysis, multi-modal fusion, hypergraph learning

*Corresponding authors: Guiguang Ding, Yue Gao

Authors' addresses: Sicheng Zhao, Tsinghua University, Beijing, 100084, China, University of California Berkeley, Berkeley, 94720, USA, schzhao@gmail.com; Amir Gholaminejad, University of California Berkeley, Berkeley, 94720, USA, amirgh@berkeley.edu; Guiguang Ding, Tsinghua University, Beijing, 100084, China, dinggg@tsinghua.edu.cn; Yue Gao, Tsinghua University, Beijing, 100084, China, gaoyue@tsinghua.edu.cn; Jungong Han, Lancaster University, Lancaster, LA1 4YW, UK, jungonghan77@gmail.com; Kurt Keutzer, University of California Berkeley, Berkeley, 94720, USA, keutzer@berkeley.edu.

1551-6857/2018/1-ART1 $15.00
https://doi.org/10.1145/3233184

ACM Reference Format:
Sicheng Zhao, Amir Gholaminejad, Guiguang Ding, Yue Gao, Jungong Han, and Kurt Keutzer. 2018. Personalized Emotion Recognition by Personality-aware High-order Learning of Physiological Signals. *ACM Trans. Multimedia Comput. Commun. Appl.* 1, 1, Article 1 (January 2018), 19 pages. https://doi.org/10.1145/3233184

1 INTRODUCTION

Emotion recognition (ER) plays an important role in both interpersonal and human-computer interaction. Though being studied for years, ER still remains an open problem, which has to face the fact that human emotions are not expressed exclusively but through multiple channels, such as speech, gesture, facial expression and physiological signals [10]. Unlike other signals that can be adopted voluntarily or involuntarily, physiological signals are controlled by the sympathetic nervous systems, which are generally independent of humans' will and cannot be easily suppressed or masked. Therefore, physiological signals may provide more reliable information for emotions compared to visual cues and audio cues [36]. Meanwhile, human emotions are a highly subjective phenomenon, as shown in Figure 1, which can be influenced by a number of contextual and psychological factors, such as interest, personality and temporal evolution.

In this paper, we focus on personalized emotion recognition (PER) from physiological signals, which enables wide user-centric applications, ranging from personalized recommender systems to intelligent diagnosis. For example, comparing the difference of physiological responses towards specified emotion between normal people and people with depression or autism can help recognize the patients automatically and contribute to clinical diagnosis. The emotion we aim to recognize here is perceived emotion. For the difference between expressed, perceived and induced emotions, please refer to [23]. However, PER is still a non-trivial problem because of the following challenges:

Multi-modal data. Emotions can be expressed through physiological signals from different modalities [10], such as Electroencephalogram (EEG), Electrocardiogram (ECG), Galvanic Skin Response (GSR), respiration and temperature, *etc.* Different subjects may have different physiological responses of the same emotion on the same modality signal. Furthermore, the importance of various physiological signals to emotions differs from each other. Combination of complementary multi-modal data by fusing strategies would obtain better results.

Multi-factor influence. Besides the physical stimuli, there are many other factors that may influence the emotion perceptions. Personal interest and personality may directly influence the emotion perceptions [24, 39]. Viewers' emotions are often temporally influenced by their recent past emotions [12]. How a viewer's emotion is influenced by their friends on social networks is quantitatively studied in [51].

Incomplete data. Due to the influence of many normal factors in data collection, such as electrode contact noise, power line interference and sensor device failure [36], physiological signals may be sometimes corrupted, which results in a common problem - data missing, i.e. physiological data from some modalities are not available [46].

Existing methods on PER mainly worked on the first challenge by designing effective fusion strategies, based on the assumption that the signals from all modalities are always available [10, 34], which is often unrealistic in practice. In this paper, we make the first attempt at estimating the influence of one psychological factor, i.e. personality, on PER from multi-modal physiological signals, trying to solve the incomplete data issue simultaneously.

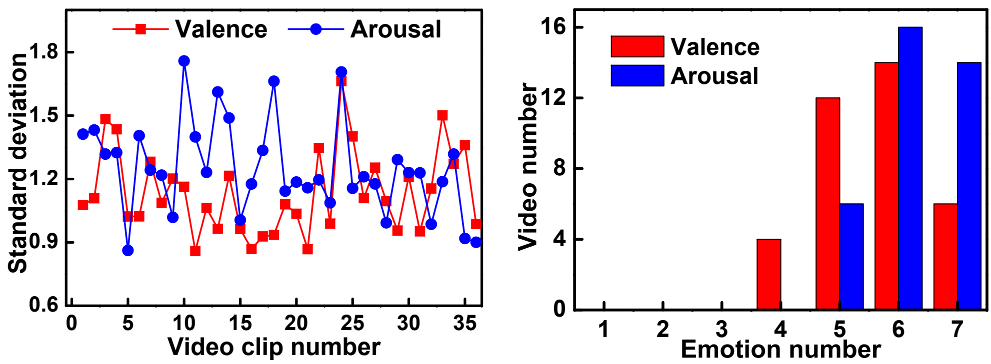


Fig. 1. Left: the valence and arousal standard deviations (STD) of the 58 subjects on the 36 video clips. Right: the video distribution with different annotated emotion numbers (7-scale) in the ASCERTAIN dataset, where "# Emotions" and "# Videos" represent the numbers of annotated emotions and videos, respectively. These two figures clearly show the emotion's subjectiveness in this context: the left figure shows that the valence and arousal STD of most videos are larger than 1, while the right one indicates that all the videos are labeled with at least 4 emotions by different subjects.

Specifically, we propose to employ the hypergraph structure to formulate the relationship among physiological signals and personality. A hypergraph[1] is a generalization of a graph in which an edge can join any number of vertices. A hypergraph is often composed of a set of vertices, and a set of non-empty subsets of vertices called hyperedges. Recently, hypergraph learning [66] has shown superior performances in various vision and multimedia tasks, such as image retrieval [20], music recommendation [7], object retrieval [13, 40], social event detection [60] and clustering [35]. However, traditional hypergraph structure treats different vertices, hyperedges and modalities equally [66], which is obviously unreasonable, since the importance is actually different. For example, different vertices have varied representation abilities and their importance varies during the learning process. To this end, we propose a Vertex-weighted Multi-modal Multi-task Hypergraph Learning (VM2HL) for PER, which introduces an updated hypergraph structure considering the vertex weights, hyperedge weights and modality weights. In our method, each vertex is a compound tuple (subject, stimuli). The personality correlation among different subjects and the physiological correlation among corresponding stimuli are formulated in a hypergraph structure. The weights of different hypergraphs and both the vertices and hyperedges of each hypergraph are automatically learned. The vertex weights and hypergraph weights are used to define the influence of different samples and modalities on the learning process, respectively, while the hyperedge weights are used to generate the optimal representation. The learning process is conducted on the vertex-weighted multi-modal multi-task hypergraphs and the estimated factors, referred as emotion relevance, are used for emotion recognition. As the vertices are compound ones, which include the information of different subjects, VM2HL can recognize the emotions of multiple subjects simultaneously. We evaluate the proposed method on the ASCERTAIN dataset that is labeled with personality and emotion information.

In summary, the contributions of this paper are three-fold:

[1] https://en.wikipedia.org/wiki/Hypergraph

1. To the best of our knowledge, this is the first comprehensive computational study about the influence of personality on personalized emotion recognition from physiological signals.
2. We propose a novel hypergraph learning algorithm, i.e. vertex-weighted multi-modal multi-task hypergraph learning (VM2HL), to jointly model the physiological signals and personality by considering the weighted importance of vertices, hyperedges and modalities.
3. Extensive experiments are conducted on the ASCERTAIN dataset with the conclusion that the proposed VM2HL obtains significant performance gains over the state-of-the-arts and can easily handle the challenge of data incompleteness.

One preliminary conference version on investigating the influence of personality on personalized emotion recognition was first introduced in our previous work [59]. The new improvement compared with the conference version lies in the following three aspects: (1) we perform a more comprehensive survey of related works; (2) we provide the motivation more clearly and detail the algorithm analysis as well as the experimental settings; and (3) we conduct more comparative experiments and enrich the analysis of the results.

The rest of this paper is organized as follows. Section 2 reviews the related work. Section 3 presents the proposed method in detail, including vertex-weighted multi-modal multi-task hypergraph construction and corresponding learning procedure. Section 4 describes the experimental setup, including the datasets, extracted features, baselines and implementation details. Experimental results and analysis are reported in Section 5, followed by the conclusion and future work in Section 6.

2 RELATED WORK

In this section, we briefly review related work on emotion recognition from physiological signals, personality and emotion relationship, and multi-modal learning.

Emotion recognition from physiological signals. As an active research topic for several years, ER from physiological signals has attracted some attention from both the academic and industrial communities. Due to the complex expression nature of human emotions, many ER methods employ a multimodal framework by considering multiple physiological signals [10]. A survey of the state-of-the-art emotion recognition methods can be found in [3, 10, 34].

Lisetti and Nasoz [29] employed GSR, heart rate, and temperature signals to recognize human emotions elicited by movie clips and difficult mathematics questions. Douglas-Cowie et al. [11] provided 3 naturalistic and 6 induced affective databases from 8-125 participants in the HUMAINE project. Besides speech, gesture and face descriptors, four physiological signals are captured: ECG, skin conductance, respiration and skin temperature. Heart rate, muscle movements, skin conductivity, and respiration changes are used to recognize emotions induced by music clips [25]. Koelstra et al. [27] analyzed the mapping between blood volume pressure, respiration rate, skin temperature, Electrooculogram (EOG) and emotions induced by 40 music videos on the popular DEAP dataset. Soleymani et al. [38] constructed MAHNOB-HCI, a multimodal dataset with synchronized face video, speech, eye-gaze and physiological recordings, including ECG, GSR, respiration amplitude, and skin temperature. User responses are correlated with eye movement patterns to analyze the impact of emotions on visual attention and memory [41]. The utility of various eye fixation and saccade-based features is examined for valence recognition [17]. Li et al. [28] employed the temporal correlation between continuous arousal peaks combined with GSR to detect induced emotions when watching 30 movies from the LIRIS-ACCEDE database [6]. The mappings from Magnetoencephalogram (MEG), Electromyogram (EMG), EOG and ECG to emotions are studied for both music and movie clips on the DECAF dataset with the conclusion that emotions are better elicited with movie clips [2]. More recently, Subramanian et al. [42] investigated binary emotion recognition from physiological features, including GSR, EEG, ECG and facial landmark trajectories (EMO), on their collected ASCERTAIN dataset. Miranda-Correa et al. [31] presented a novel dataset, AMIGOS, for multimodal research of affect, personality traits and mood from neuro-physiological signals. EEG, ECG, GSR, audio, visual, and depth modalities are fused to recognize affect from two social contexts, one with individual viewers and the other with groups of viewers. Besides the psychological signals, a playgame context is also considered to estimate the player experience or emotion [8, 30, 43].

Among the above mentioned methods, both categorical emotion states (CES) and dimensional emotion space (DES) are used to represent emotions. Being straightforward for users to understand and label, CES methods directly map emotions to one of a few basic categories [29, 38]. More descriptive DES methods employ a 3-D or 2-D space to represent emotions, such as valence-arousal-dominance (VAD) [27], and valence-arousal (VA) [2, 11, 17, 38, 41, 42]. Some works also discretize the DES into a few typical scales to combine the advantages [25, 27, 38, 42]. The authors also represent emotions using the discretized VA model as in [42].

Affective analysis has also been widely studied in different types of multimedia data, which are used to evoke human emotions, such as text [15], image [4, 22, 61, 63], music [52], speech [26], and video [48]. One close work is personalized emotion perception prediction of social images by considering visual content, social context, temporal evolution, and location information [62, 64]. Differently, our work aims to recognize personalized emotions from physiological signals by modelling personality.

Personality and emotion relationship. Human personality can be described by the big-five or five-factor model in terms of five dimensions - Extraversion, Neuroticism, Agreeableness, Conscientiousness and Openness [9]. A comprehensive survey of personality computing is presented in [45]. As for the personality and emotion relationship, Winter and Kuiper [50] extensively examined it in social psychology. Van Lankveld et al. [44] proposed to estimate personality via a player's game behaviors in a video game. By inserting the relative score of the Myers-Briggs types, Henriques et al. [19] showed that psychological traits can increase the emotion recognition performance. Henriques and Paiva [18] defined seven principles, based on empirical results, for recognizing and describing emotions during affective interactions from physiological signals. Expressive signal representations are proposed to correct individual differences and to account for subtle variations, and the integration of sequential and feature-based models. Abadi et al. [1] and Subramanian et al. [42] recognized personality and emotion separately using physiological signals without considering their intrinsic correlation and influence.

Multi-modal learning. In real-world applications, we might have multi-modal data to describe a target [5], either from different sources [10, 55] or with multiple features (also called multi-view learning) [13, 56–58, 65]. Typically different modal data can represent different aspects of the target. Jointly combining them together to explore the complementation may promisingly improve the performance [5, 10]. Besides the traditional early fusion and late fusion [16, 37, 47], there are many other multi-modal fusion strategies, such as hypergraph learning [66], multigraph learning [47] and multimodal deep learning [32].

Motivation of the proposed method. All the above-mentioned ER methods do not consider any psychological factor besides physiological signals and contextual interaction. Though personality is believed to affect emotions [24], the interleaved connection between personality and emotion has not yet been studied comprehensively in a computational

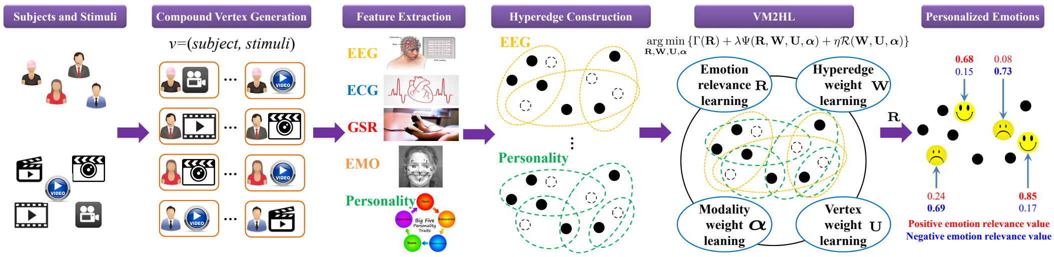


Fig. 2. The framework of the proposed method for personality-aware personalized emotion recognition from physiological signals by jointly learning the emotion relevance, hyperedge weight, vertex weight and modality weight. Each circle represents a compound vertex (subject, stimuli). The filled ones indicate training samples, while the empty ones are testing samples.

setting. On one hand, this is due to various problems such as invasiveness of sensing equipment, subject preparation time and the paucity of reliable annotators [42]. On the other hand, previous works on hypergraph learning treat the hyperedge weight and vertex weight equally [66], update hyperedge weight [14], or update hyperedge and vertex weights [40], without jointly learning the optimal weights of vertices, hyperedges and modalities. In this paper, we employ GSR, EEG, ECG, and EMO for emotion recognition and investigate the influence of personality on emotions computationally. Specifically, we present Vertex-weighted Multi-modal Multi-task Hypergraph Learning to make full use of personality and physiological signals for personalized emotion recognition.

3 THE PROPOSED METHOD

Our goal is to recognize personalized emotions from physiological signals considering personality and dealing with missing data. We employ a hypergraph structure to formulate the relationship among physiological signals and personality, taking advantage of the hypergraph on high-order correlation modelling. Considering the fact that the importance of different vertices, hyperedges and modalities in a hypergraph is different, i.e. the contribution of different elements to the learning process varies, we propose a novel method, named Vertex-weighted Multi-modal Multi-task Hypergraph Learning (VM2HL), for PER.

The framework of the proposed method is shown in Fig. 2. First, given the subjects and stimuli that are used to evoke emotions in subjects, we generate the compound tuple vertex (subject, stimuli). Second, we construct the multi-modal hyperedges to formulate the personality correlation among different subjects and the physiological correlation among corresponding stimuli. Finally, we obtain the PER results after the joint learning of the vertex-weighted multi-modal multi-task hypergraphs.

3.1 Hypergraph Construction

As stated above, the vertex in the proposed method is a compound one, including the subject and involved stimuli. We can construct different hyperedges based on the features of each element of the vertex.

Similar to [9], personality is labelled using the big five model in the ASCERTAIN dataset [42], i.e. personality is represented by a 5-dimension vector. We employ Cosine function to measure the pairwise personality similarity between two users u_i and u_j as follows

$$s_{PER}(u_i, u_j) = \frac{< p_i, p_j >}{\|p_i\| \cdot \|p_j\|}, \tag{1}$$

where p_i is the personality vector of user u_i.

A specific emotion perceived in humans usually leads to corresponding changes in different physiological signals [10]. As in [42], we extract different features from 4 kinds of physiological signals: ECG, GSR, EEG and EMO. Please refer to Section 4.2 for the detailed extraction process. Similar to Eq. (1), Cosine function is used to measure the pairwise similarity of each modality feature extracted from physiological signals. Please note that other similarity or distance measures can also be used here.

Given the pairwise similarities above, we can formulate the relationship among different samples in a hypergraph structure. Each time one vertex is selected as the centroid, and one hyperedge is constructed to connect the centroid and its K nearest neighbors in the available feature space. Please note that we construct personality hyperedges from both inter-subject and intra-subject perspectives. All the vertices from the same subject are connected by one hyperedge. Further, for each subject, we select the nearest K subjects based on personality similarity and connect all the vertices of these subjects by constructing another hyperedge.

Suppose the constructed hypergraphs are $\mathcal{G}_m = (\mathcal{V}_m, \mathcal{E}_m, \mathbf{W}_m)$, where $\mathcal{V}_m$ is the vertex set, $\mathcal{E}_m$ is the hyperedge set, and $\mathbf{W}_m$ is the diagonal matrix of hyperedge weight for the mth hypergraph ($m = 1, 2, \cdots, M$, $M = 5$ in this paper, including 4 hypergraphs based on physiological signals and 1 hypergraph based on personality). We can easily tackle the missing data challenge by removing the hyperedges of corresponding vertices. For example, if the EEG is missing for one subject, we just simply do not construct hyperedges based on EEG for this subject. This still works because the model can learn the emotion relevance by ECG, GSR, EMO, and personality.

Given the constructed hypergraph $\mathcal{G}_m$, we can obtain the incidence matrix $\mathbf{H}_m$ by computing each entry as,

$$\mathbf{H}_m(v, e) = \begin{cases} 1, & \text{if } v \in e, \\ 0, & \text{if } v \notin e. \end{cases} \tag{2}$$

Different from traditional hypergraph learning method, which simply regards all the vertices equally, we learn different weights of the vertices to measure their importance and contribution to the learning process. Suppose $\mathbf{U}_m$ is the diagonal matrix of vertex weight. The vertex degree of vertex $v \in \mathcal{V}_m$ and the edge degree of hyperedge $e \in \mathcal{E}_m$ are defined as $d_m(v) = \sum_{e \in \mathcal{E}_m} \mathbf{W}_m(e)\mathbf{H}_m(v, e)$ and $\delta(e) = \sum_{v \in \mathcal{V}_m} \mathbf{U}_m(v)\mathbf{H}_m(v, e)$. According to $d_m(v)$ and $\delta_m(e)$, we define two diagonal matrices $\mathbf{D}_m^v$ and $\mathbf{D}_m^e$ as $\mathbf{D}_m^v(i, i) = d_m(v_i)$ and $\mathbf{D}_m^e(i, i) = \delta_m(e_i)$.

3.2 Vertex-weighted Multi-modal Multi-task Hypergraph Learning

Given N subjects $u_1, \ldots, u_N$ and the involved stimuli s_{ij} ($j = 1, \cdots, n_i$) for u_i, our objective is to jointly explore the correlations among all involved physiological signals and the personality relations among different subjects. Suppose the compound vertices and corresponding labels of the cth emotion category are $\{(u_1, s_{1j})\}_{j=1}^{n_1}, \cdots, \{(u_N, s_{Nj})\}_{j=1}^{n_N}$ and $\mathbf{y}_{1c} = [y_{11}^c, \cdots, y_{1n_1}^c]^{\mathrm{T}}, \ldots, \mathbf{y}_{Nc} = [y_{N1}^c, \cdots, y_{Nn_N}^c]^{\mathrm{T}}$, where $c = 1, \cdots, n_e$, n_e is the number of emotion categories, and the to-be-estimated values of all stimuli related to the specified users of the cth emotion category, referred to as emotion relevance, are $\mathbf{r}_{1c} = [r_{11}^c, \cdots, r_{1n_1}^c]^{\mathrm{T}}, \ldots, \mathbf{r}_{Nc} = [r_{N1}^c, \cdots, r_{Nn_N}^c]^{\mathrm{T}}$. We denote $\mathbf{y}_c$ and $\mathbf{r}_c$ as

$$\mathbf{y}_c = [\mathbf{y}_{1c}^{\mathrm{T}}, \cdots, \mathbf{y}_{Nc}^{\mathrm{T}}]^{\mathrm{T}}, \mathbf{r}_c = [\mathbf{r}_{1c}^{\mathrm{T}}, \cdots, \mathbf{r}_{Nc}^{\mathrm{T}}]^{\mathrm{T}}. \tag{3}$$

Let $\mathbf{Y} = [\mathbf{y}_1, \cdots, \mathbf{y}_c, \cdots, \mathbf{y}_{n_e}], \mathbf{R} = [\mathbf{r}_1, \cdots, \mathbf{r}_c, \cdots, \mathbf{r}_{n_e}]$.

Similar to the regularization framework in [13, 40, 66], the relevance matrix $\mathbf{R}$ is learned from a joint optimization process to minimize the empirical loss and the regularizer on the hypergraph structure as well as on the weights of vertices, hyperedges and modalities simultaneously by

$$\arg\min_{\mathbf{R},\mathbf{W},\mathbf{U},\boldsymbol{\alpha}} \{\Gamma(\mathbf{R}) + \lambda\Psi(\mathbf{R},\mathbf{W},\mathbf{U},\boldsymbol{\alpha}) + \eta\mathcal{R}(\mathbf{W},\mathbf{U},\boldsymbol{\alpha})\}, \tag{4}$$

where λ and η are two trade-off parameters, $\mathbf{W} = \{\mathbf{W}_1, \cdots, \mathbf{W}_M\}$, $\mathbf{U} = \{\mathbf{U}_1, \cdots, \mathbf{U}_M\}$ and the three components are defined as follows:

Γ is the empirical loss:

$$\Gamma(\mathbf{R}) = \sum_{c=1}^{n_e} ||\mathbf{r}_c - \mathbf{y}_c||^2. \tag{5}$$

Ψ is the regularizer on the hypergraph structure:

$$\begin{aligned}
&\Psi(\mathbf{R},\mathbf{W},\mathbf{U},\boldsymbol{\alpha}) = \frac{1}{2}\sum_{c=1}^{n_e}\sum_{m=1}^{M}\alpha_m \sum_{e\in\mathcal{E}_m}\sum_{\mu,\nu\in\mathcal{V}_m} \\
&\frac{\mathbf{W}_m(e)\mathbf{U}_m(\mu)\mathbf{H}_m(\mu,e)\mathbf{U}_m(\nu)\mathbf{H}_m(\nu,e)}{\delta(e)}\left(\frac{\mathbf{r}_c(\mu)}{\sqrt{\mathbf{D}_m^v(\mu,\mu)}} - \frac{\mathbf{r}_c(\nu)}{\sqrt{\mathbf{D}_m^v(\nu,\nu)}}\right)^2 \\
&= \sum_{c=1}^{n_e}\mathbf{r}_c^{\mathrm{T}}\sum_{m=1}^{M}\alpha_m(\mathbf{U}_m - \Theta_m)\mathbf{r}_c,
\end{aligned} \tag{6}$$

where $\sum_{m=1}^{M}\alpha_m = 1$ and

$$\Theta_m = (\mathbf{D}_m^v)^{-\frac{1}{2}}\mathbf{U}_m\mathbf{H}_m\mathbf{W}_m(\mathbf{D}_m^e)^{-1}\mathbf{H}_m^{\mathrm{T}}\mathbf{U}_m(\mathbf{D}_m^v)^{-\frac{1}{2}}. \tag{7}$$

$\Delta = \sum_{m=1}^{M}\alpha_m(\mathbf{U}_m - \Theta_m)$ can be viewed as a vertex-weighted fused hypergraph Laplacian.

$\mathcal{R}$ is the regularizer on the weights of modalities, vertices and hyperedges and one simple version is adopted by

$$\mathcal{R}(\mathbf{W},\mathbf{U},\boldsymbol{\alpha}) = \sum_{m=1}^{M}(\mathrm{tr}(\mathbf{W}_m^{\mathrm{T}}\mathbf{W}_m) + \mathrm{tr}(\mathbf{U}_m^{\mathrm{T}}\mathbf{U}_m) + \mathrm{tr}(\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\alpha})), \tag{8}$$

where tr() is the trace of a matrix.

Solution

To solve the optimization task of Eq. (4), we employ an alternative strategy. First, we fix $\mathbf{W},\mathbf{U},\boldsymbol{\alpha}$, and optimize $\mathbf{R}$. The objective function of Eq. (4) turns to

$$\arg\min_{\mathbf{R}}\{\sum_{c=1}^{n_e}||\mathbf{R}(:,c) - \mathbf{Y}(:,c)||^2 + \lambda\mathbf{R}^{\mathrm{T}}\Delta\mathbf{R}\}, \tag{9}$$

where $\lambda > 0$. According to [66], $\mathbf{R}$ can be solved by

$$\mathbf{R} = \left(\mathbf{I} + \frac{1}{\lambda}\Delta\right)^{-1}\mathbf{Y}. \tag{10}$$

Second, we fix $\mathbf{R}, \mathbf{U}, \boldsymbol{\alpha}$, and optimize $\mathbf{W}$. Since each $\mathbf{W}_m$ is independent from each other, the objective function can be rewritten as

$$\arg\min_{\mathbf{W}_m}\{\lambda \sum_{c=1}^{n_e} \mathbf{y}_c^{\mathrm{T}} \alpha_m (\mathbf{U}_m - \Theta_m)\mathbf{y}_c + \eta \mathrm{tr}(\mathbf{W}_m^{\mathrm{T}} \mathbf{W}_m)\}, \tag{11}$$

where $\mathbf{D}_m^v(v,v) = \sum_{e \in \mathcal{E}_m} \mathbf{W}_m(e)\mathbf{H}_m(v,e)$, $\eta > 0$, and $\mathbf{W}_m(e) \geq 0$. Replacing Θ_m with Eq. (7), the above optimization task is convex on $\mathbf{W}_m$ and can be easily solved via off-the-shelf quadratic programming methods.

Third, we fix $\mathbf{R}, \mathbf{W}, \boldsymbol{\alpha}$, and optimize $\mathbf{U}$. Since each $\mathbf{U}_m$ is independent from each other, the optimization of $\mathbf{U}$ is similar to the optimization of $\mathbf{W}$.

Finally, we fix $\mathbf{R}, \mathbf{W}, \mathbf{U}$, and optimize $\boldsymbol{\alpha}$. The objective function of Eq. (4) reduces to

$$\begin{aligned} &\arg\min_{\boldsymbol{\alpha}}\{\lambda \sum_{c=1}^{n_e} \mathbf{y}_c^{\mathrm{T}} \alpha_m (\mathbf{U}_m - \Theta_m)\mathbf{y}_c + \eta M \mathrm{tr}(\boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{\alpha})\}, \\ &\qquad \text{s.t.} \quad \sum_{m=1}^{M} \alpha_m = 1, \eta > 0. \end{aligned} \tag{12}$$

Similar to [13], we employ the Lagrange multiplier to solve the optimization problem and can derive:

$$\alpha_m = \frac{1}{M} + \frac{\sum_{c=1}^{n_e} \mathbf{y}_c^{\mathrm{T}} \sum_{m=1}^{M} (\mathbf{U}_m - \Theta_m)\mathbf{y}_c}{2\eta M^2} - \frac{\sum_{c=1}^{n_e} \mathbf{y}_c^{\mathrm{T}} (\mathbf{U}_m - \Theta_m)\mathbf{y}_c}{2\eta M}. \tag{13}$$

The above optimization procedure is repeated until convergence. Intuitively, all the three components in Eq. (4) are greater than or equal to 0, so the objective function has a lower bound 0. When updating each of the steps above, the corresponding objective turns to a quadratic optimization problem, for which an optimal solution can be computed [13, 40, 66], and thus decreases the overall objective function Eq. (4). Therefore, the convergence of the alternating optimization is guaranteed.

The computational cost is computed as follows. The complexity of hypergraph construction is $O(\sum_{m=1}^{M} d_m (\sum_{i=1}^{N} n_i)^3 \log \sum_{i=1}^{N} n_i)$. The complexity of emotion recognition is $O(T_a (\sum_{i=1}^{N} n_i)^2 n_e T_b)$, where T_a and T_b are the iteration number for the alternating optimization process and the iteration number of the iterative process (here we assume that the iterations of optimizing $\mathbf{R}, \mathbf{W}, \mathbf{U}$ are the same), respectively. Please note that the computational cost can be further reduced by data downsampling [54] and hierarchical hypergraph learning strategy [49], which remains our future work.

4 EXPERIMENT SETUP

In this section, we introduce the detailed experimental settings, including the ASCERTAIN dataset that contains both personality and emotion information with physiological signals, compared baselines and implementation details.

4.1 Dataset

To the best of our knowledge, ASCERTAIN [42] is the only published and released dataset to date that connects personality and emotional states via physiological responses. 58 university students (21 female, mean age = 30) were invited to watch 36 movie clips used in [2] between 51-127s long to evoke emotions. All the subjects were fluent in English and were habitual Hollywood movie watchers. The movie clips are shown to be uniformly distributed (9 clips

Table 1. Extracted features for each modality [42], where "#" indicates the dimension of each feature, and "Statistics" denote mean, standard deviation (std), skewness, kurtosis of the raw feature over time, and % of times the feature value is abovebelow mean±std.

Modality	#	Extracted features
ECG	32	Ten low frequency ([0-2.4] Hz) power spectral densities (PSDs), four very slow response ([0-0.04] Hz) PSDs, IBI, HR and HRV statistics.
EEG	88	Average of first derivative, proportion of negative differential samples, mean number of peaks, mean derivative of the inverse channel signal, average number of peaks in the inverse signal, statistics over each of the 8 signal channels provided by the Neurosky software.
GSR	31	Mean skin resistance and mean of derivative, mean differential for negative values only (mean decrease rate during decay time), proportion of negative derivative samples, number of local minima in the GSR signal, average rising time of the GSR signal, spectral power in the [0-2.4] Hz band, zero crossing rate of skin conductance slow response ([0-0.2] Hz), zero crossing rate of skin conductance very slow response ([0-0.08] Hz), mean SCSR and SCVSR peak magnitude.
EMO	72	Statistics concerning horizontal and vertical movement of 12 motion units (MUs) specified in [21].

per quadrant) over the VA space. During watching the clips, several sensors were used to record the physiological signals. After watching each clip, the participators were requested to label the VA ratings reflecting their affective impression with a 7-point scale, i.e. -3 (very negative) to 3 (very positive) scale for V, and 0 (very boring) to 6 (very exciting) scale for A. Personality measures for the big-five dimensions were also compiled using a big-five marker scale questionnaire [33]. The standard deviations of ENACO are 1.0783, 0.7653, 0.7751, 0.9176, and 0.6479, respectively. Please note that the dataset is incomplete with missing data. For example, the 13rd, 15th, 27th, and 34th GSR signals of the 3rd student are missing.

4.2 Extracted Features

Following [42], different features are extracted for the 4 kinds of physiological signals: ECG, GSR, EEG, and EMO. GSR measures the transpiration rate of the skin, EEG measures the small changes in the skull's electrical field produced by neural activity, ECG evaluates the heart rate characteristics, and EMO calculates the statistical measures of different landmarks. These features are extracted over the final 50 seconds of stimulus presentation, owing to (1) the clips are more emotional towards the end, and (2) some employed features are nonlinear functions of the input signal length. The detailed features are summarized in Table 1. Please note that in the ASCERTAIN dataset, due to the influence of normal factors in data collection, one or more modality features are missing for some specified subjects.

4.3 Baselines

To compare with the state-of-the-art approaches for PER, we select the following methods as baselines: (1) Support Vector Machine with linear kernel (SVM_L) [42] and with radial basis function kernel (SVM_R), (2) Naive Bayes (NB) [42], (3) hypergraph learning (HL) [66], and (4) hypergraph learning with hyperedge weight update (HL_E) [14]. As our goal is to recognize personalized emotions, we need to train one model for each subject using the baselines, such as SVM and NB. We indeed can take the personality of each subject or the subject

Table 2. Mann-Whitney-Wilcoxon test of the proposed VM2HL and VM2HL-P with the baselines measured by p-value ($\times 10^{-3}$).

		SVM_L	SVM_R	NB	HL	HL_E	VM2HL-P
VM2HL	Valence	3.24	4.83	2.65	3.47	4.16	1.32
	Arousal	5.31	6.46	4.15	4.13	6.25	2.82
VM2HL-P	Valence	4.35	6.62	3.59	4.61	5.38	-
	Arousal	5.62	7.28	5.63	6.75	8.54	-

correlations as input, but it makes no sense or cannot contribute to the model. This is because for each subject, the personality is always the same, which means that the personality feature for the training and test samples are identical. Therefore, we do not consider personality features for these baselines. Late fusion for SVM and NB is implemented as in [42] to deal with multi-modal physiological signals, which are connected in one hypergraph in HL and HL_E. SVM_L, SVM_R, and NB are state-of-the-art methods for emotion recognition [31, 42]. HL and HL_E are traditional hypergraph learning methods [14, 66].

4.4 Implementation Details

Similar to [42], we dichotomize the valence and arousal affective ratings based on the median values for binary emotion recognition, since the number of movie clips each subject watched and labelled is relatively small for fine-grained emotion recognition. We employ the recognition accuracy (*Acc*) [42] as evaluation metric. For each subject, *Acc* is defined as the fraction of correctly recognized emotions among the total test emotions. For each test run, the overall *Acc* is the average *Acc* of all subjects. $0 \leq Acc \leq 1$ and a larger *Acc* value indicates better performance. 50% of stimuli and corresponding physiological signals and emotions of each subject are randomly selected as the training set and the rest constitute the testing set. The parameters of the baselines are selected by 10-fold cross validation on the training set. For example, the gamma and C parameters of SVM are selected via grid search, similar to [42]. Unless otherwise specified, parameter K in hyperedge generation is set to 10, and regularizer parameters $\lambda = 0.1$ and $\eta = 100$ are adopted in experiment. Empirical analysis on parameter sensitivity is also conducted, which demonstrates that the proposed VM2HL has superior and stable performance with a wide range of parameter values. The weights of vertices, hyperedges and modalities are initialized to 1 and optimized by the proposed method. For fair comparison, we carefully tune the parameters of the baselines and report the best results. Further, we perform 10 runs and show the average results together with the standard deviations to remove the influence of any randomness.

5 RESULTS AND ANALYSIS

In this section, we report the results on comparison with the state-of-the-art approaches and on the influence of different factors in the proposed method.

5.1 Comparison with the State-of-the-art

First, we conduct experiments to compare the performance of the proposed method with the state-of-the-art approaches for personalized emotion recognition. The results measured by recognition accuracy are shown in Fig. 3, while the Mann-Whitney-Wilcoxon test results are given in Table 2. “VM2HL-P” indicates the proposed method without modelling personality. Please note that the five baselines and VM2HL-P are based on the fusion of all physiological signals, while the proposed VM2HL jointly models physiological signals and personality.

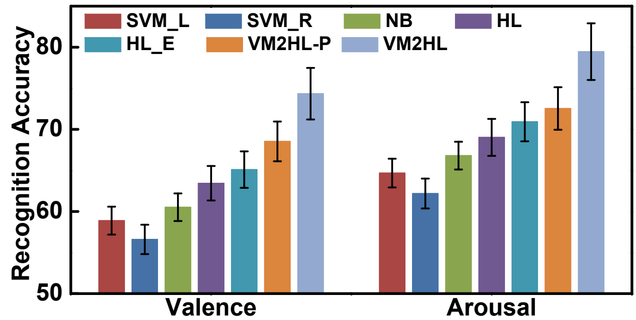


Fig. 3. Performance comparison between the proposed method and the state-of-the-art approaches in terms of recognition accuracy and the standard deviation (%), where "-P" indicates without personality.

From the results, we have the following observations: (1) the proposed method (both VM2HL-P and VM2HL) significantly outperforms the baselines on both valence and arousal; (2) the hypergraph learning families achieve better results than traditional SVM and NB classifiers; (3) NB performs slightly better than SVM; though simple, the linear kernel of SVM is superior to the RBF kernel; (4) all the methods achieve above-chance (50%) emotion recognition performance with physiological features; (5) the performance on arousal is better than valence.

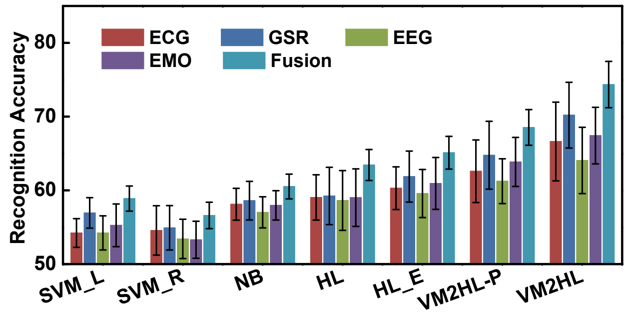
The better performance of the proposed method can be attributed to the following three reasons. 1. The hypergraph structure is able to explore the complex high-order relationship among multi-modal features, which leads to the superior performance of hypergraph learning families over other models. 2. We take personality into account, which connects different subjects with similar personality values. In this way, the recognition process turns to a multi-task learning problem for multiple subjects. The latent correlations among different subjects are effectively explored, which can be deemed as a way to enlarge the training set for each subject. 3. The different importance or contribution of vertices, hyperedges and modalities are jointly learned, which can accordingly generate a better correlation.

Comparing the results of VM2HL-P and VM2HL, it is clear that after removing personality, the performance decreases significantly. Comparing with VM2HL-P, VM2HL achieves 8.48% and 9.54% performance gains on valence and arousal, respectively. This is reasonable because personality is the only element that connects different subjects and corresponding physiological signals. By changing from single-task learning for each subject to multi-task learning for multiple subjects, the latent information is extensively explored, which has a similar impact as increasing the number of training samples and thus improves the recognition performance.

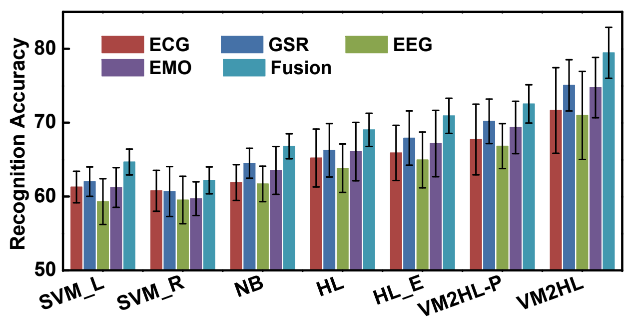
5.2 On Different Physiological Signals

Second, we compare the performance of different uni-modal physiological signals for personalized emotion recognition. The results on valence and arousal are reported in Figure 4(a) and Figure 4(b), respectively.

Comparing the results, we can observe that: (1) fusing multi-modal physiological signals can obtain better recognition performance than most uni-modal ones for all the methods; (2) generally, GSR features produce the best performance for both valence and arousal, while ECG and EEG features are less discriminative; (3) for most physiological signals,



(a) Valence



(b) Arousal

Fig. 4. Performance comparison between different single physiological signal and the fusion strategy of different methods in terms of recognition accuracy and the standard deviation (%).

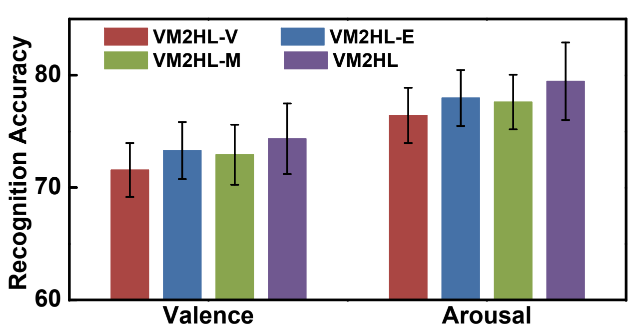
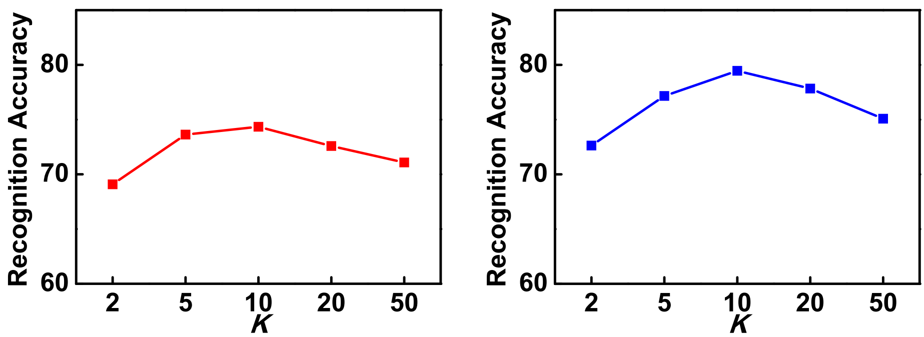


Fig. 5. Personalized emotion recognition results with and without optimizing vertex, hyperedge and modality weights in terms of recognition accuracy and the standard deviation (%), where "-V", "-E" and "-M" indicate without optimizing vertex weights, hyperedge weights and modality weights, respectively.

the performance comparison of different methods follows the similar order to the above Subsection. Please note that in these figures, VM2HL considers personality besides different uni-modal physiological signals.



(a) Valence (b) Arousal

Fig. 6. The influence of K in the hyperedge generation stage on the emotion recognition performance of the proposed method in terms of recognition accuracy (%).

5.3 On Vertex, Hyperedge and Modality Weights

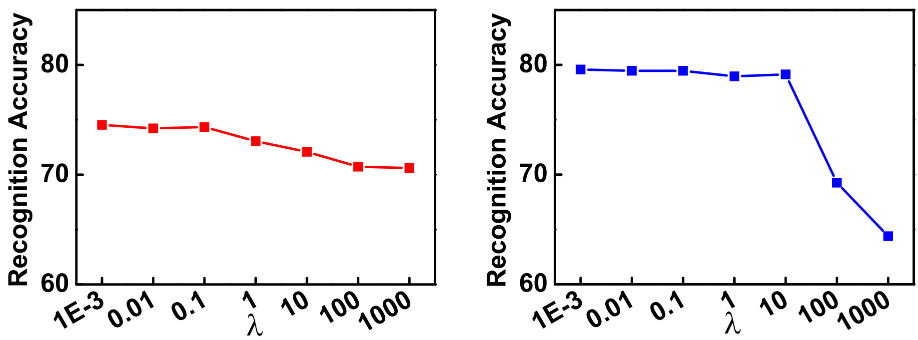
Third, we investigate the influence of optimal vertex, hyperedge and modality weights by removing the optimization of just one kind of weight. The results are shown in Fig. 5. We can see that all the three kinds of weights indeed contribute to the performance of the proposed method. The performance gains of VM2HL over VM2HL-V, VM2HL-E, and VM2HL-M are 3.88%, 1.43%, 1.95% on valence, and 3.98%, 1.91%, 2.38% on arousal, respectively. Please note that VM2HL-M is similar to the multi-task version of the hypergraph learning method with hyperedge and vertex weights update [40]. Generally, vertex weights give more contribution to the overall performance, following by modality weights and hyperedge weights. We can conclude that jointly optimizing the weights of vertices, hyperedges and modalities would generate more discriminative hypergraph structure and produce better emotion recognition performance.

5.4 On Hyperedge Generation

Fourth, we evaluate the influence of the selected neighbor number K in hyperedge generation on the performance of the proposed method. The result is shown in Figure 6, with K varying from 2 to 50. It is clear that the performance is relatively steady with a wide range. When K becomes too small or too large, the performance turns to be slightly worse. When K is too small, such as $K = 2$, too few vertices are connected in each hyperedge, which cannot fully explore the high-order relationship among different vertices. However, when K is too large, such as $K = 50$, too many vertices are connected in each hyperedge, which could also limit the discriminative ability of the hypergraph structure. We can conclude that both too small and too large K values will degenerate the representation ability and thus degrade the performance.

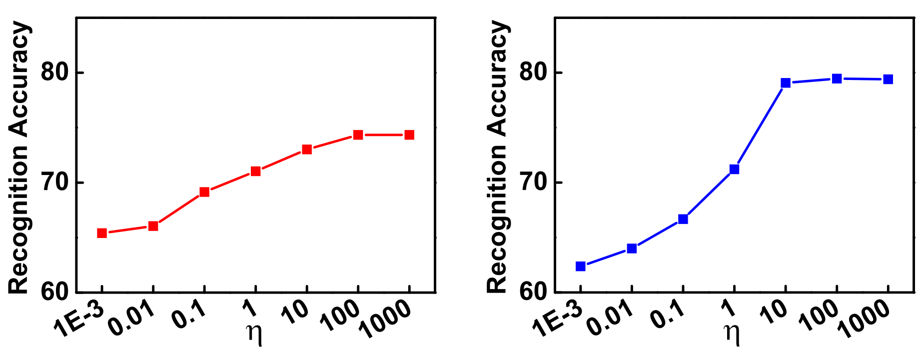
5.5 On Parameter Sensitivity

There are two regularization parameters in the proposed method that control the relative importance of different regularizers in the objective function, i.e. λ which is the regularizer for the hypergraph structure and η which is the regularizer for the weights, verticies, hyperedges and modalities. To validate the influences of λ and η, we first fix η as 100 and vary λ, and



(a) Valence (b) Arousal

Fig. 7. The influence of regularization parameter λ on the emotion recognition performance of the proposed method in terms of recognition accuracy (%).



(a) Valence (b) Arousal

Fig. 8. The influence of regularization parameter η on the emotion recognition performance of the proposed method in terms of recognition accuracy (%).

then fix λ as 0.1 and vary η, with results shown in Figure 7 and Figure 8, respectively. From these results, we can observe that (1) the proposed method can achieve steady performances when λ and η vary in a large range; (2) with the increase of λ, the performance tends to be stable when $\lambda \leq 10$, and then turns worse; (3) with the increase of η, the performance tends to be better and becomes stable when $\eta \geq 100$. Too large or too small values would either dominate the objective function or have quite little influence on the results, which is expected. We can conclude that selecting proper λ and η can indeed improve the performance of emotion recognition, which indicates the significance of the joint exploration of different regularizers.

5.6 Limitation Discussion

The tested dataset is relatively small. As the only available dataset that connects personality and emotional states via physiological responses, ASCERTAIN [42] only includes 58 subjects and 36 movie clips. Constructing a large-scale dataset with personality and physiological signals, and testing the proposed method on large-scale data would make more sense.

The computational efficiency of hypergraph learning would greatly increase when dealing with large-scale data. To reduce the computational cost, there are two possible solutions: data downsampling [54] and hierarchical hypergraph learning strategy [49].

Dichotomizing ordinal VA values turns out to yield split criterion biases. The reason behind is similar to [42], i.e. the number of movie clips each subject watched and labelled is relatively small. Our method can be easily extended to fine-grained emotion classification if large-scale data is available. Like other hypergraph learning methods, the proposed method can only be used for emotion classification, without supporting emotion regression. As shown in [53], the ordinal labels are a more suitable way to represent emotions. Currently, the proposed method cannot tackle the ordinal emotions.

6 CONCLUSION

In this paper, we proposed to recognize personalized emotions by jointly modelling personality and physiological signals, which is the first comprehensive computational study about the influence of personality on emotion. We presented Vertex-weighted Multi-modal Multi-task Hypergraph Learning as the learning model, where (subject, stimuli) forms the vertices, and the relationship among personality and physiological signals is formulated as hyperedges. By introducing the vertex weights, hyperedge weights and modality weights, our method is able to jointly explore the importance of different vertices, hyperedges and modalities. The learning process on a hypergraph is thus more optimal for personalized emotion recognition. Further, the proposed method can easily handle the data incompleteness issue by constructing the corresponding hyperedge or not. Experimental results on the ASCERTAIN dataset demonstrated the effectiveness of the proposed PER method, which can generalize to new subjects if the personality or physiological signals are known.

For further studies, we plan to combine the multimedia content employed to evoke emotions and the physiological signals for PER. In addition, we will predict emotion and personality simultaneously in a joint framework to further explore the latent correlation. Constructing a reliable large-scale dataset with personality and physiological signals would greatly promote the research of PER. Recognizing group emotions to balance personalized emotions and dominant emotions is an interesting and worthwhile topic. How to improve the computational efficiency of hypergraph learning to deal with large-scale data remains to be discussed.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (Nos. 61701273, 61571269, 61671267), the Project Funded by China Postdoctoral Science Foundation (Nos. 2018T110100, 2017M610897), the Royal Society Newton Mobility Grant (No. IE150997), the National Key R&D Program of China (Grant No. 2017YFC011300), and the Berkeley Deep Drive. The authors would also like to thank the Handling Guest Editor X. Alameda-Pineda and the anonymous reviewers for their insightful comments to help us improve the paper.

REFERENCES

[1] Mojtaba Khomami Abadi, Juan Abdón Miranda Correa, Julia Wache, Heng Yang, Ioannis Patras, and Nicu Sebe. 2015. Inference of personality traits and affect schedule by analysis of spontaneous reactions to affective videos. In *IEEE International Conference and Workshops on Automatic Face and Gesture Recognition*, Vol. 1. 1–8.

[2] Mojtaba Khomami Abadi, Ramanathan Subramanian, Seyed Mostafa Kia, Paolo Avesani, Ioannis Patras, and Nicu Sebe. 2015. DECAF: MEG-based multimodal database for decoding affective physiological responses. *IEEE Transactions on Affective Computing* 6, 3 (2015), 209–222.

[3] Hussein Al Osman and Tiago H Falk. 2017. Multimodal Affect Recognition: Current Approaches and Challenges. In *Emotion and Attention Recognition Based on Biological Signals and Images*. InTech.

[4] Xavier Alameda-Pineda, Elisa Ricci, Yan Yan, and Nicu Sebe. 2016. Recognizing emotions from abstract paintings using non-linear matrix completion. In *IEEE Conference on Computer Vision and Pattern Recognition*. 5240–5248.

[5] Pradeep K Atrey, M Anwar Hossain, Abdulmotaleb El Saddik, and Mohan S Kankanhalli. 2010. Multimodal fusion for multimedia analysis: a survey. *Multimedia Systems* 16, 6 (2010), 345–379.

[6] Yoann Baveye, Emmanuel Dellandrea, Christel Chamaret, and Liming Chen. 2015. Liris-accede: A video database for affective content analysis. *IEEE Transactions on Affective Computing* 6, 1 (2015), 43–55.

[7] Jiajun Bu, Shulong Tan, Chun Chen, Can Wang, Hao Wu, Lijun Zhang, and Xiaofei He. 2010. Music recommendation by unified hypergraph: combining social media information and music content. In *ACM International Conference on Multimedia*. 391–400.

[8] Elizabeth Camilleri, Georgios N Yannakakis, and Antonios Liapis. 2017. Towards General Models of Player Affect. In *International Conference on Affective Computing and Intelligent Interaction*. 333–339.

[9] Paul T Costa and Robert R MacCrae. 1992. *Revised NEO personality inventory (NEO PI-R) and NEO five-factor inventory (NEO-FFI): Professional manual.* Psychological Assessment Resources, Incorporated.

[10] Sidney K D'mello and Jacqueline Kory. 2015. A review and meta-analysis of multimodal affect detection systems. *Comput. Surveys* 47, 3 (2015), 43.

[11] Ellen Douglas-Cowie, Roddy Cowie, Ian Sneddon, Cate Cox, Orla Lowry, Margaret Mcrorie, Jean-Claude Martin, Laurence Devillers, Sarkis Abrilian, Anton Batliner, et al. 2007. The HUMAINE database: addressing the collection and annotation of naturalistic and induced emotional data. In *International Conference on Affective Computing and Intelligent Interaction*. 488–500.

[12] Nico H Frijda. 1986. *The emotions*. Cambridge University Press.

[13] Yue Gao, Meng Wang, Dacheng Tao, Rongrong Ji, and Qionghai Dai. 2012. 3-d object retrieval and recognition with hypergraph analysis. *IEEE Transactions on Image Processing* 21, 9 (2012), 4290–4303.

[14] Yue Gao, Meng Wang, Zheng-Jun Zha, Jialie Shen, Xuelong Li, and Xindong Wu. 2013. Visual-textual joint relevance learning for tag-based social image search. *IEEE Transactions on Image Processing* 22, 1 (2013), 363–376.

[15] Anastasia Giachanou and Fabio Crestani. 2016. Like it or not: A survey of twitter sentiment analysis methods. *Comput. Surveys* 49, 2 (2016), 28.

[16] Hatice Gunes and Massimo Piccardi. 2005. Affect recognition from face and body: early fusion vs. late fusion. In *IEEE International Conference on Systems, Man and Cybernetics*, Vol. 4. 3437–3443.

[17] R Hamed, Adham Atyabi, Antti Rantanen, Seppo J Laukka, Samia Nefti-Meziani, Janne Heikkilä, et al. 2015. Predicting the valence of a scene from observers eye movements. *PloS One* 10, 9 (2015), e0138198.

[18] Rui Henriques and Ana Paiva. 2014. Seven Principles to Mine Flexible Behavior from Physiological Signals for Effective Emotion Recognition and Description in Affective Interactions.. In *International Conference on Physiological Computing Systems*. 75–82.

[19] Rui Henriques, Ana Paiva, and Claudia Antunes. 2013. Accessing emotion patterns from affective interactions using electrodermal activity. In *Humaine Association Conference on Affective Computing and Intelligent Interaction*. 43–48.

[20] Yuchi Huang, Qingshan Liu, Shaoting Zhang, and Dimitris Metaxas. 2010. Image retrieval via probabilistic hypergraph ranking. In *IEEE Conference on Computer Vision and Pattern Recognition*. 3376–3383.

[21] Hideo Joho, Jacopo Staiano, Nicu Sebe, and Joemon M Jose. 2011. Looking at the viewer: analysing facial activity to detect personal highlights of multimedia contents. *Multimedia Tools and Applications* 51, 2 (2011), 505–523.

[22] Dhiraj Joshi, Ritendra Datta, Elena Fedorovskaya, Quang-Tuan Luong, James Z Wang, Jia Li, and Jiebo Luo. 2011. Aesthetics and emotions in images. *IEEE Signal Processing Magazine* 28, 5 (2011), 94–115.

[23] Patrik N Juslin and Petri Laukka. 2004. Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening. *Journal of New Music Research* 33, 3 (2004),

217–238.

[24] Elizabeth G Kehoe, John M Toomey, Joshua H Balsters, and Arun LW Bokde. 2012. Personality modulates the effects of emotional arousal and valence on brain activation. *Social Cognitive and Affective Neuroscience* 7, 7 (2012), 858–870.

[25] Jonghwa Kim and Elisabeth André. 2008. Emotion recognition based on physiological changes in music listening. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30, 12 (2008), 2067–2083.

[26] Yelin Kim and Emily Mower Provost. 2015. Emotion recognition during speech using dynamics of multiple regions of the face. *ACM Transactions on Multimedia Computing, Communications, and Applications* 12, 1s (2015), 25.

[27] Sander Koelstra, Christian Muhl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and Ioannis Patras. 2012. DEAP: A database for emotion analysis; using physiological signals. *IEEE Transactions on Affective Computing* 3, 1 (2012), 18–31.

[28] Ting Li, Yoann Baveye, Christel Chamaret, Emmanuel Dellandréa, and Liming Chen. 2015. Continuous arousal self-assessments validation using real-time physiological responses. In *ACM International Workshop on Affect & Sentiment in Multimedia*. ACM, 39–44.

[29] Christine Lætitia Lisetti and Fatma Nasoz. 2004. Using noninvasive wearable computers to recognize human emotions from physiological signals. *EURASIP Journal on Advances in Signal Processing* 2004, 11 (2004), 929414.

[30] Hector P Martinez, Yoshua Bengio, and Georgios N Yannakakis. 2013. Learning deep physiological models of affect. *IEEE Computational Intelligence Magazine* 8, 2 (2013), 20–33.

[31] Juan Abdon Miranda-Correa, Mojtaba Khomami Abadi, Nicu Sebe, and Ioannis Patras. 2017. AMIGOS: A dataset for Mood, personality and affect research on Individuals and GrOupS. *arXiv preprint arXiv:1702.02510* (2017).

[32] Jiquan Ngiam, Aditya Khosla, Mingyu Kim, Juhan Nam, Honglak Lee, and Andrew Y Ng. 2011. Multimodal deep learning. In *International Conference on Machine Learning*. 689–696.

[33] Marco Perugini and Lisa Di Blas. 2002. Analyzing personality related adjectives from an eticemic perspective: the big five marker scales (BFMS) and the Italian AB5C taxonomy. *Big Five Assessment* (2002), 281–304.

[34] Soujanya Poria, Erik Cambria, Rajiv Bajpai, and Amir Hussain. 2017. A review of affective computing: From unimodal analysis to multimodal fusion. *Information Fusion* 37 (2017), 98–125.

[35] Pulak Purkait, Tat-Jun Chin, Alireza Sadri, and David Suter. 2017. Clustering with hypergraphs: the case for large hyperedges. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39, 9 (2017), 1697–1711.

[36] Yangyang Shu and Shangfei Wang. 2017. Emotion recognition through integrating EEG and peripheral signals. In *IEEE International Conference on Acoustics, Speech and Signal Processing*. 2871–2875.

[37] Cees GM Snoek, Marcel Worring, and Arnold WM Smeulders. 2005. Early versus late fusion in semantic video analysis. In *ACM International Conference on Multimedia*. 399–402.

[38] Mohammad Soleymani, Jeroen Lichtenauer, Thierry Pun, and Maja Pantic. 2012. A multimodal database for affect recognition and implicit tagging. *IEEE Transactions on Affective Computing* 3, 1 (2012), 42–55.

[39] Robert C Solomon. 1993. *The passions: Emotions and the meaning of life*. Hackett Publishing.

[40] Lifan Su, Yue Gao, Xibin Zhao, Hai Wan, Ming Gu, and Jiaguang Sun. 2017. Vertex-Weighted Hypergraph Learning for Multi-View Object Classification. In *International Joint Conferences on Artificial Intelligence*. 2779–2785.

[41] Ramanathan Subramanian, Divya Shankar, Nicu Sebe, and David Melcher. 2014. Emotion modulates eye movement patterns and subsequent memory for the gist and details of movie scenes. *Journal of Vision* 14, 3 (2014), 31:1–31:18.

[42] Ramanathan Subramanian, Julia Wache, Mojtaba Abadi, Radu Vieriu, Stefan Winkler, and Nicu Sebe. 2016. ASCERTAIN: Emotion and personality recognition using commercial sensors. *IEEE Transactions on Affective Computing* (2016).

[43] Simone Tognetti, Maurizio Garbarino, Andrea Bonarini, and Matteo Matteucci. 2010. Modeling enjoyment preference from physiological responses in a car racing game. In *IEEE Conference on Computational Intelligence and Games*. 321–328.

[44] Giel Van Lankveld, Pieter Spronck, Jaap Van den Herik, and Arnoud Arntz. 2011. Games as personality profiling tools. In *IEEE Conference on Computational Intelligence and Games*. 197–202.

[45] Alessandro Vinciarelli and Gelareh Mohammadi. 2014. A survey of personality computing. *IEEE Transactions on Affective Computing* 5, 3 (2014), 273–291.

[46] Johannes Wagner, Elisabeth Andre, Florian Lingenfelser, and Jonghwa Kim. 2011. Exploring fusion methods for multimodal emotion recognition with missing data. *IEEE Transactions on Affective Computing* 2, 4 (2011), 206–218.

[47] Meng Wang, Xian-Sheng Hua, Richang Hong, Jinhui Tang, Guo-Jun Qi, and Yan Song. 2009. Unified video annotation via multigraph learning. *IEEE Transactions on Circuits and Systems for Video Technology* 19, 5 (2009), 733–746.

[48] Shangfei Wang and Qiang Ji. 2015. Video affective content analysis: a survey of state of the art methods. *IEEE Transactions on Affective Computing* 6, 4 (2015), 410–430.

[49] Longyin Wen, Wenbo Li, Junjie Yan, Zhen Lei, Dong Yi, and Stan Z Li. 2014. Multiple target tracking based on undirected hierarchical relation hypergraph. In *IEEE Conference on Computer Vision and Pattern Recognition*. 1282–1289.

[50] Kathy A Winter and Nicholas A Kuiper. 1997. Individual differences in the experience of emotions. *Clinical Psychology Review* 17, 7 (1997), 791–821.

[51] Yang Yang, Jia Jia, Shumei Zhang, Boya Wu, Qicong Chen, Juanzi Li, Chunxiao Xing, and Jie Tang. 2014. How Do Your Friends on Social Media Disclose Your Emotions?. In *AAAI Conference on Artificial Intelligence*. 306–312.

[52] Yi-Hsuan Yang and Homer H Chen. 2012. Machine recognition of music emotion: A review. *ACM Transactions on Intelligent Systems and Technology* 3, 3 (2012), 40.

[53] Georgios N Yannakakis, Roddy Cowie, and Carlos Busso. 2017. The Ordinal Nature of Emotions. In *International Conference on Affective Computing and Intelligent Interaction*. 248–255.

[54] Chao Yao, Jimin Xiao, Tammam Tillo, Yao Zhao, Chunyu Lin, and Huihui Bai. 2016. Depth map down-sampling and coding based on synthesized view distortion. *IEEE Transactions on Multimedia* 18, 10 (2016), 2015–2022.

[55] Quanzeng You, Liangliang Cao, Hailin Jin, and Jiebo Luo. 2016. Robust Visual-Textual Sentiment Analysis: When Attention meets Tree-structured Recursive Neural Networks. In *ACM International Conference on Multimedia*. 1008–1017.

[56] Sicheng Zhao, Guiguang Ding, Yue Gao, and Jungong Han. 2017. Approximating Discrete Probability Distribution of Image Emotions by Multi-Modal Features Fusion. In *International Joint Conference on Artificial Intelligence*. 466–4675.

[57] Sicheng Zhao, Guiguang Ding, Yue Gao, and Jungong Han. 2017. Learning Visual Emotion Distributions via Multi-Modal Features Fusion. In *ACM International Conference on Multimedia*. 369–377.

[58] Sicheng Zhao, Guiguang Ding, Yue Gao, Xin Zhao, Youbao Tang, Jungong Han, Hongxun Yao, and Qingming Huang. 2018. Discrete Probability Distribution Prediction of Image Emotions With Shared Sparse Learning. *IEEE Transactions on Affective Computing* (2018).

[59] Sicheng Zhao, Guiguang Ding, Jungong Han, and Yue Gao. 2018. Personality-Aware Personalized Emotion Recognition from Physiological Signals. In *International Joint Conferences on Artificial Intelligence*.

[60] Sicheng Zhao, Yue Gao, Guiguang Ding, and Tat-Seng Chua. 2017. Real-Time Multimedia Social Event Detection in Microblog. *IEEE Transactions on Cybernetics* (2017).

[61] Sicheng Zhao, Yue Gao, Xiaolei Jiang, Hongxun Yao, Tat-Seng Chua, and Xiaoshuai Sun. 2014. Exploring principles-of-art features for image emotion recognition. In *ACM International Conference on Multimedia*. 47–56.

[62] Sicheng Zhao, Hongxun Yao, Yue Gao, Guiguang Ding, and Tat-Seng Chua. 2018. Predicting personalized image emotion perceptions in social networks. *IEEE Transactions on Affective Computing* (2018).

[63] Sicheng Zhao, Hongxun Yao, Yue Gao, Rongrong Ji, and Guiguang Ding. 2017. Continuous probability distribution prediction of image emotions via multitask shared sparse regression. *IEEE Transactions on Multimedia* 19, 3 (2017), 632–645.

[64] Sicheng Zhao, Hongxun Yao, Yue Gao, Rongrong Ji, Wenlong Xie, Xiaolei Jiang, and Tat-Seng Chua. 2016. Predicting personalized emotion perceptions of social images. In *ACM International Conference on Multimedia*. 1385–1394.

[65] Sicheng Zhao, Hongxun Yao, You Yang, and Yanhao Zhang. 2014. Affective image retrieval via multi-graph learning. In *ACM International Conference on Multimedia*. 1025–1028.

[66] Dengyong Zhou, Jiayuan Huang, and Bernhard Schölkopf. 2006. Learning with Hypergraphs: Clustering, Classification, and Embedding. In *Advances in Neural Information Processing Systems*. 1601–1608.

Received September 2017; revised March 2018 and April 2018; accepted June 2018