© 2025 IJNRD | Volume 10, Issue 2 February 2025 | ISSN: 2456-4184 | IJNRD.ORG

Review EmoFusion: Integrating Multi-Modal Data for Comprehensive Human Emotion Detection

Bhushan Deshmukh
Department Of Information technology
Tulsiramji Gaikwad Patil College of Engineering and Tech
Nagpur, INDIA

Dr. Prof. Mukul Pande
Department Of Information technology
Tulsiramji Gaikwad Patil College of Engineering and Tech
Nagpur, INDIA

Abstract

Emotion recognition methodologies from physio logical signals are increasingly becoming person alized, due to the subjective responses of different subjects to physical stimuli. Existing works mainly focused on modelling the involved physi- ological corpus of each subject, without consider- ing the psychological factors. The latent correlation among different subjects has also been rarely examined. We propose to investigate the influence of personality on emotional behavior in a hyper graph 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 assign each of them with weights. The emotion relevance learned on the vertex-weighted multi-modal multi-task hypergraphs is employed for emotion recognition. We carry out extensive experiments on the ASCERTAIN dataset and the results demonstrate the superiority of the proposed method.

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 [D'mello and Kory, 2015]. 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 [Shu and Wang, 2017]. 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 and personality.

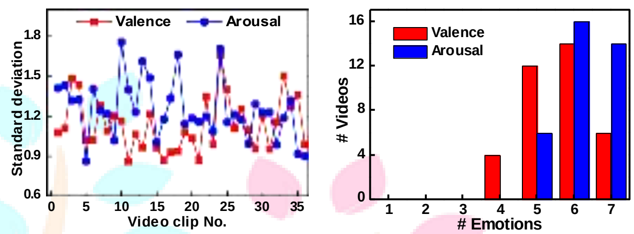


Figure 1: Left: the valence and arousal standard deviations of the 58 subjects on the 36 video clips. Right: the video distribution with different annotated emotion numbers (7-scale) in the ASCERTAIN dataset, where "# Emotions" and "# Videos" represent the numbers of annotated emotions and videos, respectively. 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.

In this paper, we focus on personalized emotion recognition (PER) from physiological signals, which enables wide user-centric applications, ranging from character analysis to personalized recommender systems. The emotion we aim to recognize here is perceived emotion. For the difference between expressed, perceived and induced emotions, please refer to [Juslin and Laukka, 2004]. 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 [D'mello and Kory, 2015], such as Electroencephalogram (EEG), Electrocardiogram (ECG), Galvanic Skin Response (GSR), and temperature, *etc*. Different subjects may have different physiological responses of the same emotion on the same modality signal. Further, the importance of various physiological signals to emotions differs from each other. Combining the complementary multi-modal data would obtain better results.

Multi-factor influence. Besides the physical stimuli, there are many other factors that may influence the emotion perceptions. For example, personal interest and personality may directly influence the emotion perceptions [Kehoe *et al.*, 2012]; viewers' emotions are often influenced by their recent past emotions [Frijda, 1986] and by their friends on social networks [Yang *et al.*, 2014].

Incomplete data. Due to the influence of many normal factors in data collection, such as electrode contact noise, and sensor device failure [Shu and Wang, 2017], physiological signals may be sometimes corrupted, which results in a common problem - data missing, i.e. physiological data from some modalities are not available [Wagner *et al.*, 2011].

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, 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 multimodal physiological signals, trying to solve the incomplete data issue simultaneously.

Specifically, we propose to employ the hypergraph structure to formulate the relationship among physiological signals and personality. Recently, hypergraph learning [Zhou *et al.*, 2006] has shown superior performances in various vision and multimedia tasks, such as music recommenda- tion [Bu *et al.*, 2010], object retrieval [Gao *et al.*, 2012; Su *et al.*, 2017], social event detection [Zhao *et al.*, 2017b] and clustering [Purkait *et al.*, 2017]. However, tradition- al hypergraph structure treats different vertices, hyperedges, and modalities equally, which is obviously unreasonable. 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 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 semi-supervised learning is conducted and the estimated factors, referred as emotion relevance, are used for emotion recognition. The emotions of multiple subjects can be recognized simultaneously. We evaluate the proposed method on the ASCERTAIN dataset [Subramanian *et al.*, 2016].

The contributions of this paper are three-fold:

1. We propose to computationally study the influence of personality on personalized emotion recognition from physiological signals.
2. We present a novel hypergraph learning algorithm, i.e. 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 significantly outperforms the state-of-the-art and can easily handle the challenge of data incompleteness.

2 Related Work

Emotion recognition from physiological signals. Due to the complex expression nature of human emotions, many ER methods employ a multimodal framework by consider- ing multiple physiological signals [D'mello and Kory, 2015]. Lisetti and Nasoz [2004] employed GSR, heart rate, and temperature signals to recognize human emotions elicited by movie clips and mathematics questions. Muscle movements, heart rate, skin conductivity, and respiration changes are used to recognize emotions induced by music clips [Kim and Andre´, 2008]. Koelstra *et al.* [2012] analyzed the mapping between blood volume pressure, respiration rate, skin temperature, Electrooculogram (EOG) and emotions induced by 40 music videos. Soleymani *et al.* [2012] constructed 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 [Subramanian *et al.*, 2014]. The mappings from Magnetoencephalogram (MEG), Electromyogram (EMG), EOG and ECG to emotions are studied for both music and movie clips [Abadi *et al.*, 2015b]. Subramanian *et al.* [2016] investigated binary emotion recognition from physiological features, including GSR, EEG, ECG and facial landmark trajectories (EMO), on their collected ASCERTAIN dataset. Besides the psychological signals, a playgame context is also considered to estimate the player experience or emotion [Tognetti *et al.*, 2010; Martinez *et al.*, 2013; Camilleri *et al.*, 2017]. However, all these methods do not consider any psychological factor besides physiological signals and contextual interaction. In this paper, we employ GSR, EEG, ECG, and EMO for emotion recognition by considering the influence of personality.

Among the above ER approaches, both categorical emotion states (CES) [Lisetti and Nasoz, 2004; Soleymani *et al.*, 2012] and dimensional emotion space (DES) [Koelstra *et al.*, 2012; Subramanian *et al.*, 2014; Abadi *et al.*, 2015b; Subramanian *et al.*, 2016] are used to represent emotions. Similar to [Subramanian *et al.*, 2016], we represent emotions using the discretized VA model.

One close work is personalized emotion prediction of social images by considering visual content, social context, temporal evolution, and location [Zhao *et al.*, 2016]. 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 (ENACO) [Costa and MacCrae, 1992]. A comprehensive survey of personality computing is presented in [Vinciarelli and Mohammadi, 2014]. As for the personality and emotion relationship, Winter and Kuiper [1997] extensively examined it in social psychology. Van Lankveld *et al.* [2011] proposed to estimate personality via a player's game behaviors in a video game. Abadi *et al.* [2015a] and Subramanian *et al.* [2016] recognized personality and emotion separately using physiological signals without considering their intrinsic correlation and influence. Though personality is believed to affect emotions [Kehoe *et al.*, 2012], personality and emotion relationship from physiological signals has not yet been studied comprehensively in a computational setting, due to various problems such as invasiveness of sensing equipment, subject preparation time and the paucity of reliable annotators [Subramanian *et al.*, 2016].

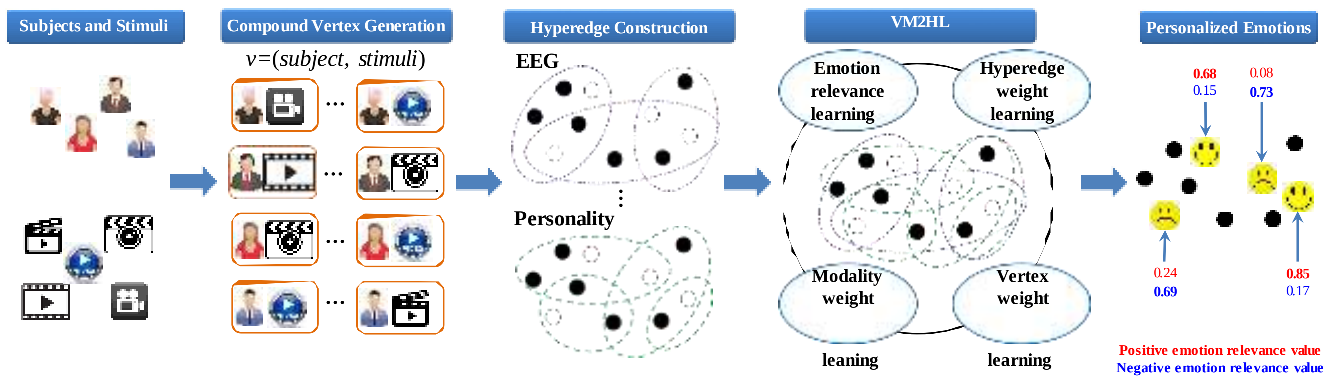


Figure 2: The framework of the proposed method for personality-aware personalized emotion recognition from physiological signals by jointly learning the emotion relevance, hyperedge weight, vertex weight, and modality weight. Each circle represents a compound vertex (subject, stimuli). The filled ones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

Besides physiological signals, we investigate the influence of personality on emotions computationally.

Multi-modal learning. We might have multi-modal data to describe a target [Atrey *et al.*, 2010], either from different sources [D'mello and Kory, 2015] or with multiple features (also called multi-view learning) [Gao *et al.*, 2012; Zhao *et al.*, 2017c; 2017a]. Typically different modal data can represent different aspects of the target. Jointly combining them together to explore the complementation may promisingly improve the performance [Atrey *et al.*, 2010; D'mello and Kory, 2015]. Besides the traditional early fusion and late fusion [Wang *et al.*, 2009], there are many other multi-modal fusion strategies, such as hypergraph learning [Zhou *et al.*, 2006], multigraph learning [Wang *et al.*, 2009] and multimodal deep learning [Ngiam *et al.*, 2011]. By jointly exploring the different weights of vertices, hyperedges, and modalities, we present VM2HL to make full use of personality and physiological signals for PER.

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 its high-order correlation modelling. Considering the fact that the importance of different vertices, hyperedges and modalities in a hypergraph is different, we propose a novel method, named Vertex-weighted Multi-modal Multi-task Hypergraph Learning (VM2HL), for PER. The framework is shown in Figure 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 [Costa and MacCrae, 1992], personality is labelled using the big five model in the ASCERTAIN dataset [Subramanian *et al.*, 2016], 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

$$s_{PER}(u_i, u_j) = \frac{\langle p_i, p_j \rangle}{\|p_i\| \cdot \|p_j\|}, \qquad (1)$$

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 [D'mello and Kory, 2015]. As in [Subramanian *et al.*, 2016], we extract different features from 4 kinds of physiological signals: ECG, GSR, EEG, and EMO, 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 dimensions are 32, 31, 88 and 72, respectively. Please refer to the Table 3 in [Subramanian *et al.*, 2016] for feature extraction details. Similar to Eq. (1), Cosine function is used to measure the pairwise similarity of each modality feature extracted from physiological signals. 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 with the emotion relevance learned from 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,

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

Different from traditional hypergraph learning method, which simply regards all the vertices equally, we assign different weights to 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 $\mathbf{D}^v_m$ and $\mathbf{D}^e_m$ as $\mathbf{D}^v_m(i, i) = d_m(v_i)$ and $\mathbf{D}^e_m(i, i) = \delta_m(e_i)$.

3.2 VM2HL

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 $\mathbf{y}_{1c} = [y^c_{11}, \cdots, y^c_{1n_1}]^{\mathrm{T}}, \ldots, \mathbf{y}_{Nc} = [y^c_{N1}, \cdots, y^c_{Nn_N}]^{\mathrm{T}}$ $(c = 1, \cdots, n_e)$, and the to-be-estimated emotion relevance values of all stimuli related to the specified users of the cth emotion category are $\mathbf{r}_{1c} = [r^c_{11}, \cdots, r^c_{1n_1}]^{\mathrm{T}}, \ldots, \mathbf{r}_{Nc} = [r^c_{N1}, \cdots, r^c_{Nn_N}]^{\mathrm{T}}$. We denote $\mathbf{y}_c$ and $\mathbf{r}_c$ as

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

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

The proposed VM2HL is conducted as a semi-supervised learning 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

$$\arg\min \{\Gamma(\mathbf{R}) + \lambda\Psi(\mathbf{R}, \mathbf{W}, \mathbf{U}, \boldsymbol{\alpha}) + \eta \mathrm{R}(\mathbf{W}, \mathbf{U}, \boldsymbol{\alpha})\}, \quad (4)$$

where $\boldsymbol{\alpha}$ represents the weights of different hypergraphs to evaluate the importance of different modality features, which satisfies $\sum_{m=1}^{M} \alpha_m = 1$, and

$$\Theta_m = (\mathbf{D}^v_m)^{-\frac{1}{2}}\mathbf{U}_m\mathbf{H}_m\mathbf{W}_m(\mathbf{D}^e_m)^{-1}\mathbf{H}^{\mathrm{T}}_m\mathbf{U}_m(\mathbf{D}^v_m)^{-\frac{1}{2}}. \quad (7)$$

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

R is the regularizer on the weights of modalities, vertices and hyperedges and one simple version is adopted by

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

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

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

where $\lambda > 0$. According to [Zhou *et al.*, 2006], $\mathbf{R}$ can be solved by

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

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

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

where $\mathbf{D}^v_m(v, v) = \sum_{e \in \mathcal{E}_m} \mathbf{W}_m(e)\mathbf{H}_m(v, e)$, $\eta > 0$, 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.

$\mathbf{R}, \mathbf{W}, \mathbf{U}, \boldsymbol{\alpha}$

where λ and μ 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

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

Ψ is the regularizer on the hypergraph structure

$$\Psi(\mathbf{R}, \mathbf{W}, \mathbf{U}, \boldsymbol{\alpha}) = \frac{1}{2} \sum_{c=1}^{n_e} \sum_{m=1}^{M} \alpha_m \sum_{e \in E_m} \sum_{\mu, v \in V_m} \frac{\mathbf{W}_m(e)\mathbf{U}_m(\mu)\mathbf{H}_m(\mu, e)\mathbf{U}_m(v)\mathbf{H}_m(v, e)}{\delta(e)} \left(\frac{\mathbf{r}_c(\mu)}{\sqrt{\mathbf{D}_m^v(\mu, \mu)}} - \frac{\mathbf{r}_c(v)}{\sqrt{\mathbf{D}_m^v(v, v)}} \right)^2 = \sum_{c=1}^{n_e} \mathbf{r}_c^{\top} \sum_{m=1}^{M} \alpha_m (\mathbf{U}_m - \Theta_m)\mathbf{r}_c, \quad (6)$$

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

$$\arg\min_{\boldsymbol{\alpha}} \{\lambda \sum_{c=1}^{n_e} \mathbf{y}_c^{\top} \alpha_m (\mathbf{U}_m - \Theta_m)\mathbf{y}_c + \eta M \operatorname{tr}(\boldsymbol{\alpha}^{\top}\boldsymbol{\alpha})\}, \quad (12)$$

where $\sum_{m=1}^{M} \alpha_m = 1$ and $\eta > 0$. Similar to [Gao *et al.*, 2012], we employ the Lagrange multiplier to solve the optimization problem and can derive

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

The above optimization procedure is repeated until convergence. Since each of the steps above decreases the objective function which has a lower bound 0, the convergence of the alternating optimization can be guaranteed.

	SVM_L	SVM_R	NB	HL	HL_E	VM2HL
V	58.89	56.60	60.52	63.44	65.10	**74.34**
A	64.68	62.18	66.80	69.02	70.92	**79.46**

Table 1: Performance comparison between the proposed method and the state-of-the-art approaches in terms of recognition accuracy (%).

	SVM_L	SVM_R	NB	HL	HL_E
V	3.24	4.83	2.65	3.47	4.16
A	5.31	6.46	4.15	4.13	6.25

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

4 Experiment Setup

4.1 Dataset

To the best of our knowledge, ASCERTAIN [Subramanian *et al.*, 2016] 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 from [Abadi *et al.*, 2015b] , which are 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 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 [Perugini and Di Blas, 2002]. 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 13th, 15th, 27th, and 34th GSR signals of the 3rd student are missing.

4.2 Baselines

To compare with the state-of-the-art for PER, we select the following methods as baselines: (1) Support Vector Machine with linear kernel (SVM_L) [Subramanian *et al.*, 2016] and with radial basis function kernel (SVM_R), (2) Naive Bayes (NB) [Subramanian *et al.*, 2016], (3) hypergraph learning (H-L) [Zhou *et al.*, 2006], and (4) hypergraph learning with hyperedge weight update (HL_E) [Gao *et al.*, 2013]. Late fusion for SVM and NB is implemented as in [Subramanian *et al.*, 2016] to deal with multi-modal physiological signals, which are connected in one hypergraph in HL and HL_E.

4.3 Implementation Details

Similar to [Subramanian *et al.*, 2016], 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) [Subramanian *et al.*, 2016] as the evaluation metric. $0 \leq Acc \leq 1$ 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 [Subramanian *et al.*, 2016]. 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 a superior and stable performance with a wide range of parameter values. For a fair comparison, we carefully tune the parameters of the baselines and report the best results. Further, we perform 10 runs and report the average results to remove the influence of any randomness.

5 Results and Analysis

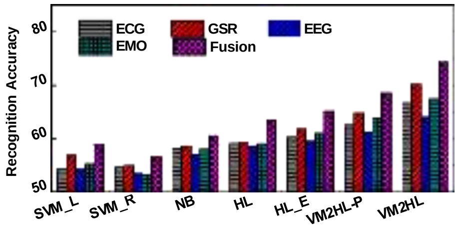
5.1 Comparison with the State-of-the-art

First, we compare the performance of the proposed method with the state-of-the-art approaches for personalized emotion recognition. The result measured by recognition accuracy is shown in Table 1, where the best methods are highlighted in bold. The Mann-Whitney-Wilcoxon test results are given in Table 2. From the results, we observe that: (1) the pro- posed method significantly outperforms the baselines on both valence and arousal under 95% confidence interval; (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, which is probably because that the standard deviation of arousal is larger in most cases, as shown in Figure 1, which may lead to larger interclass difference. Specifical- ly, the performance gains of VM2HL over SVM L, SVM R, NB, HL and HL E are 26.25%, 31.35%, 22.84%, 17.19%, 14.20% on valence, and 22.86%, 27.78%, 18.95%, 15.12%, 12.03% on arousal, respectively.

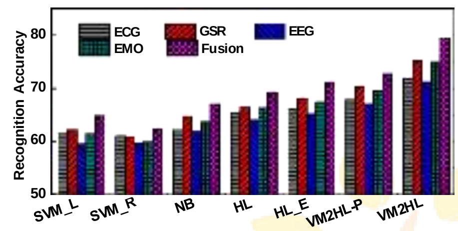
The better performance of the proposed method can be attributed to the following reasons. 1. The hypergraph structure is able to explore the complex high-order relationship among multi-modal features, which leads to the superior performance of hypergraph learning families over other models. 2. We take personality into account, which connects different subjects with similar personality values. 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 of vertices, hyperedges, and modalities are jointly learned, which can accordingly generate a better correlation.

5.2 On Different Physiological Signals

Second, we compare the performance of different uni-modal physiological signals. The results on valence and arousal are



(a) Valence



(b) Arousal

Figure 3: Performance comparison between different single physiological signal and the fusion strategy of different methods in terms of recognition accuracy (%).

reported in Figure 3(a) and Figure 3(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, the performances of different methods follow the similar order to the above Subsection.

5.3 On Personality

Third, we evaluate the influence of personality on the recognition performance by removing the personality hyperedge in VM2HL. The comparison between with and without personality in the proposed method is shown in Table 3. 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 subject- s, the latent information is extensively explored, which has a similar impact as increasing the number of training samples and thus improves the recognition performance.

5.4 On Vertex, Hyperedge, and Modality Weights

Fourth, 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 Table 4. 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

	VM2HL-P	VM2HL
Valence	68.53	74.34
Arousal	72.54	79.46

Table 3: Personalized emotion recognition results with and without personality in terms of recognition accuracy (%), where “-P” indicates without personality.

	VM2HL-V	VM2HL-E	VM2HL-M	VM2HL
Valence	71.56	73.29	72.92	74.34
Arousal	76.42	77.97	77.61	79.46

Table 4: Personalized emotion recognition results with and without optimizing vertex, hyperedge, and modality weights in terms of recognition accuracy (%), where “-V”, “-E”, and “-M” indicate without optimizing vertex weights, hyperedge weights, and modality weights, respectively.

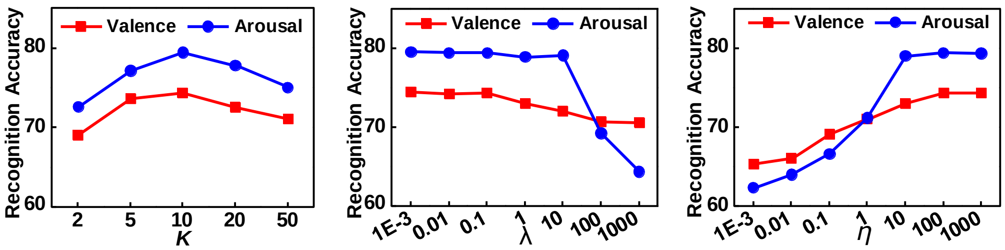
is similar to the multi-task version of the hypergraph learning method with hyperedge and vertex weights update [Su *et al.*, 2017]. Generally, vertex weights give more contribution to the overall performance, following by modality weights and hyperedge weights. We can conclude that jointly optimizing the weights of vertices, hyperedges, and modalities would generate more discriminative hypergraph structure and produce better emotion recognition performance.

5.5 On Hyperedge Generation

Fifth, 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 4(a), 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.

5.6 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 on the regularizer of the hypergraph structure and η on the weights of vertices, hyperedges, and modalities. To validate the influences of λ and η, we first fix η as 100 and vary λ, and then fix λ as 0.1 and vary η, with results shown in Figure 4(b) and Figure 4(c), 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. We can conclude that selecting proper λ and η can indeed improve the performance



(a) K (b) λ (c) η

Figure 4: The influence of (a) hyperedge generation parameter K, (b) regularization parameter λ, and (c) regularization parameter η on the emotion recognition performance of the proposed method in terms of recognition accuracy (%).

of emotion recognition, which indicates the significance of the joint exploration of different regularizers.

5.7 Limitation Discussion

The tested dataset is relatively small. As the only available dataset that connects personality and emotional states via physiological responses, ASCERTAIN [Subramanian *et al.*, 2016] 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 remain our future work.

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 [Yao *et al.*, 2016] and hierarchical hypergraph learning strategy [Wen *et al.*, 2014].

Dichotomizing ordinal VA values turns out to yield split criterion biases. The reason behind is similar to [Subramanian *et al.*, 2016], 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 [Yannakakis *et al.*, 2017], 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, to the best of our knowledge, the first comprehensive computational study about the influence of personality on emotion. We presented Vertex-weighted Multimodal 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. The importance of different vertices, hyperedges, and modalities is effectively explored by learning the optimal weights. Further, the proposed method can easily handle the data incompleteness issue. Experimental results on the ASCERTAIN dataset demonstrated the superiority 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. How to improve the computational efficiency of hypergraph learning to deal with large-scale data is also worth studying.

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 (No. 2017M610897), the Royal Society Newton Mobility Grant (No. IE150997), the National Key R&D Program of China (No. 2017YFC011300), and the Berkeley Deep Drive.

References

[Abadi *et al.*, 2015a] Mojtaba Khomami Abadi, Juan Abdo´n Miranda Correa, Julia Wache, Heng Yang, Ioannis Patras, and Nicu Sebe. Inference of personality traits and affect schedule by analysis of spontaneous reactions to affective videos. In *IEEE FGR*, volume 1, pages 1–8, 2015.

[Abadi *et al.*, 2015b] Mojtaba Khomami Abadi, Ramanathan Subramanian, Seyed Mostafa Kia, Paolo Avesani, Ioannis Patras, and Nicu Sebe. Decaf: Meg-based multimodal database for decoding affective physiological responses. *IEEE TAFFC*, 6(3):209–222, 2015.

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

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

[Camilleri *et al.*, 2017] Elizabeth Camilleri, Georgios N Yannakakis, and Antonios Liapis. Towards general models of player affect. In *ACII*, pages 333–339, 2017.

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

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

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

[Gao *et al.*, 2012] Yue Gao, Meng Wang, Dacheng Tao, Rongrong Ji, and Qionghai Dai. 3-d object retrieval and recognition with hypergraph analysis. *IEEE TIP*, 21(9):4290–4303, 2012.

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

[Juslin and Laukka, 2004] Patrik N Juslin and Petri Laukka. Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening. *Journal of New Music Research*, 33(3):217–238, 2004.

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

[Kim and Andre´, 2008] Jonghwa Kim and Elisabeth Andre´. Emotion recognition based on physiological changes in music listening. *IEEE TPAMI*, 30(12):2067–2083, 2008.

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

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

[Martinez *et al.*, 2013] Hector P Martinez, Yoshua Bengio, and Georgios N Yannakakis. Learning deep physiological models of affect. *IEEE CIM*, 8(2):20–33, 2013.

[Ngiam *et al.*, 2011] Jiquan Ngiam, Aditya Khosla, Mingyu Kim, Juhan Nam, Honglak Lee, and Andrew Y Ng. Multimodal deep learning. In *ICML*, pages 689–696, 2011.

[Perugini and Di Blas, 2002] Marco Perugini and Lisa Di Blas. Analyzing personality related adjectives from an eticemic perspective: the big five marker scales (bfms) and the italian ab5c taxonomy. *Big Five Assessment*, pages 281–304, 2002.

[Purkait *et al.*, 2017] Pulak Purkait, Tat-Jun Chin, Alireza Sadri, and David Suter. Clustering with hypergraphs: the case for large hyperedges. *IEEE TPAMI*, 39(9):1697–1711, 2017.

[Shu and Wang, 2017] Yangyang Shu and Shangfei Wang. Emotion recognition through integrating eeg and peripheral signals. In *ICASSP*, pages 2871–2875, 2017.

[Soleymani *et al.*, 2012] Mohammad Soleymani, Jeroen Lichtenauer, Thierry Pun, and Maja Pantic. A multimodal database for affect recognition and implicit tagging. *IEEE TAFFC*, 3(1):42–55, 2012.

[Su *et al.*, 2017] Lifan Su, Yue Gao, Xibin Zhao, Hai Wan, Ming Gu, and Jiaguang Sun. Vertex-weighted hypergraph learning for multi-view object classification. In *IJCAI*, pages 2779–2785, 2017.

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

[Subramanian *et al.*, 2016] Ramanathan Subramanian, Julia Wache, Mojtaba Abadi, Radu Vieriu, Stefan Winkler, and Nicu Sebe. Ascertain: Emotion and personality recognition using commercial sensors. *IEEE TAFFC*, 2016.

[Tognetti *et al.*, 2010] Simone Tognetti, Maurizio Garbarino, Andrea Bonarini, and Matteo Matteucci. Modeling enjoyment preference from physiological responses in a car racing game. In *IEEE CIG*, pages 321–328, 2010.

[Van Lankveld *et al.*, 2011] Giel Van Lankveld, Pieter Spronck, Jaap Van den Herik, and Arnoud Arntz. Games as personality profiling tools. In *IEEE CIG*, pages 197–202, 2011.

[Vinciarelli and Mohammadi, 2014] Alessandro Vinciarelli and Gelareh Mohammadi. A survey of personality computing. *IEEE TAFFC*, 5(3):273–291, 2014.

[Wagner *et al.*, 2011] Johannes Wagner, Elisabeth Andre, Florian Lingenfelser, and Jonghwa Kim. Exploring fusion methods for multimodal emotion recognition with missing data. *IEEE TAFFC*, 2(4):206–218, 2011.

[Wang *et al.*, 2009] Meng Wang, Xian-Sheng Hua, Richang Hong, Jinhui Tang, Guo-Jun Qi, and Yan Song. Unified video annotation via multigraph learning. *IEEE TCSVT*, 19(5):733–746, 2009.

[Wen *et al.*, 2014] Longyin Wen, Wenbo Li, Junjie Yan, Zhen Lei, Dong Yi, and Stan Z Li. Multiple target tracking based on undirected hierarchical relation hypergraph. In *CVPR*, pages 1282–1289, 2014.

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

[Yang *et al.*, 2014] Yang Yang, Jia Jia, Shumei Zhang, Boya Wu, Qicong Chen, Juanzi Li, Chunxiao Xing, and Jie Tang. How do your friends on social media disclose your emotions? In *AAAI*, pages 306–312, 2014.

[Yannakakis *et al.*, 2017] Georgios N Yannakakis, Roddy Cowie, and Carlos Busso. The ordinal nature of emotions. In *ACII*, pages 248–255, 2017.

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

[Zhao *et al.*, 2016] Sicheng Zhao, Hongxun Yao, Yue Gao, Rongrong Ji, Wenlong Xie, Xiaolei Jiang, and Tat-Seng Chua. Predicting personalized emotion perceptions of social images. In *ACM MM*, pages 1385–1394, 2016.

[Zhao *et al.*, 2017a] Sicheng Zhao, Guiguang Ding, Yue Gao, and Jungong Han. Approximating discrete probability distribution of image emotions by multi-modal features fusion. In *IJCAI*, pages 466–4675, 2017.

[Zhao *et al.*, 2017b] Sicheng Zhao, Yue Gao, Guiguang Ding, and Tat-Seng Chua. Real-time multimedia social event detection in microblog. *IEEE TCYB*, 2017.

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

[Zhou *et al.*, 2006] Dengyong Zhou, Jiayuan Huang, and Bernhard Scho¨lkopf. Learning with hypergraphs: Clustering, classification, and embedding. In *NIPS*, pages 1601–1608, 2006.