EMOTION RECOGNITION THROUGH INTEGRATING EEG AND PERIPHERAL SIGNALS

*Yangyang Shu and Shangfei Wang**

Key Lab of Computing and Communication Software of Anhui Province,
School of Computer Science and Technology, University of Science and Technology of China, Hefei, Anhui, P.R.China, 230027

ABSTRACT

The inherent dependencies among multiple physiological signals are crucial for multimodal emotion recognition, but have not been thoroughly exploited yet. This paper propose to use restricted Boltzmann machine (RBM) to model such dependencies.Specifically, the visible nodes of RBM represent EEG and peripheral physiological signals, and thus the connections between visible nodes and hidden nodes capture the intrinsic relations among multiple physiological signals. The RBM generates new representation from multiple physiological signals. Then, a support vector machine is adopted to recognize users' emotion states from the generated features. Furthermore, we extend the proposed fusion method for incomplete datas, since physiological signals are often corrupted due to artifacts. Specifically, we pre-train the RBM using all the complete data, then we update missing values and RBM parameters to minimize free energy of visible vectors using both complete and incomplete data. Experiments on two benchmark databases demonstrate the effectiveness of the proposed methods.

Index Terms— multimodal emotion recognition; RBM; EEG; peripheral physiological signal; missing data;

1. INTRODUCTION

Due to the importance of emotions in human's daily life, emotion recognition is essential to human computer interaction. Emotions can be detected from multiple channels, i.e. visual cues, audio cues, and physiological signals[1]. Physiological signals reflect unconscious body changes, and are controlled by the sympathetic nervous system, while visual cues and audio cues can be adopted voluntarily or involuntarily. Thus, physiological signals may provide more reliable information for emotions compared to visual cues and audio cues. The paper focuses on emotion recognition from multiple physiological signals.

Current research on emotion recognition from multiple physiological signals can be categorized into feature-level fusion and decision-level fusion[2]. Feature level fusion approaches integrate multiple physiological signals for emotion recognition through concatenating features from multiple physiological signals into one feature vector, while decision-level fusion approaches combine emotion classifiers from each modality through decision strategies, such as major vote or weighted combination. Although both feature-level fusion and decision-level fusion exploit multiple physiological signals for facilitating emotion recognition to some extend, the inherent dependencies among multiple physiological signals can not be effectively captured through simply concatenating multiple features or combining the recognized emotions from multiple physiological signals. Therefore, in this paper, we propose a new fusion method to model the high-order dependencies among multiple physiological signals. Specifically, we employ a Restricted Boltzmann Machine (RBM), whose visible nodes represent multiple physiological signals, i.e. EEG signals and peripheral physiological signals. Through introducing hidden nodes, the connections between visible nodes and hidden nodes capture the intrinsic relations among EEG signals and peripheral physiological signals, and thus the RBM can generate new representation for multiple physiological signals. Then, the support vector machine is used to recognize users' emotion states from the generated features.

Curren research on emotion recognition from multiple physiological signals assumes that all channels of data are always available. However, data missing is a common problem encountered when investigating physiological signals. Physiological signals may be corrupted by power line interference, motion artifacts, electrode contact noise, or sensor device failure. To the best of our knowledge, there is little work of emotion recognition from multiple physiological signals considering missing physiological data. Only recently, researchers begin to realized that it is too optimistic to assume that all data from all modalities is available at all times. Wagner et al.[3] maybe the first to explore fusion methods for multimodal emotion recognition with missing data. Other than discarding all the data instances containing invalid modalities, which results in a substantial amount of unusable data, they propose handling missing data at the decision-level fusion by integrating all the available modalities. In this paper, we propose to treat the missing data in the same way as the other parameters. Specifically, we first pre-train the RBM using all the complete data, then we initialize the missing data randomly. After that, we update the missing values each time when we update the weights to minimize the free energy of visible vectors using both complete and incomplete data.

2. METHOD

The framework of the proposed method is shown in Fig. 1. We first extract features from peripheral physiological signals and EEG signals. Then, we adopt the RBM to model the high-order relations among multiple physiological features, and generate a new feature representation. After that, we use a SVM classifier to map the generated features to emotion labels. For the missing data, we first train RBM model using the complete data to initialize the parameters of our model. After that, we iteratively generate the missing data as well as update the model parameters in the whole data set.

*This is the corresponding author

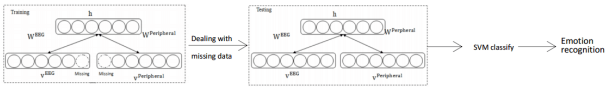

Fig. 1. The Framework of Our Method

2.1. Feature Extraction

2.1.1. EEG features

We preprocessed EEG signals by adopting a band-pass filter with a lower cutoff frequency of 0.3Hz and a higher cutoff frequency of 45Hz to mitigate the noise. Then we extract the power spectrum (PS) features from EEG signals. The PS of five frequency sub-bands, i.e. delta(0-4Hz), theta(4-8Hz), alpha(8-13Hz), beta(13-30Hz) and gamma(30-45Hz), from 32 electrodes are extracted. Furthermore, we calculate the PS asymmetry between 14 pairs of electrodes from alpha, beta, theta and gamma sub-bands. That is used as the EEG features [4].

2.1.2. Peripheral features

Peripheral physiological signals include electro-oculogram (EOG), electromyograms (EMG) of Zygomaticus and Trapezius muscles, electrocardiograph, galvanic skin response (GSR), respiration amplitude (RESP), skin temperature (TEMP) and blood volume by plethysmograph signals (PLET). Before features extraction, these signals are preprocessed using band-pass filters to restrain the noise. Then, several commonly used features are adopted.

For EOG and EMG signals, 0.4Hz and 1Hz low-pass filters are adopted respectively, and then energy, mean and variance are used as features. For ECG signals, 1Hz low-pass filters are used. Heart rate variability (HRV), root mean square of the mean squared difference of successive beats, standard deviation of beat interval change per respiratory cycle, 14 spectral power in the bands from 0-1.5Hz, low frequency 0.01-0.08Hz, medium frequency 0.08-0.15Hz and high frequency 0.15-0.5Hz components of HRV power spectrum and poincare analysis features (2 features) [5] are extracted as features. For GSR signals, mean, mean of the derivative, mean of the positive derivatives, proportion of negatives in the derivative, number of local minima and 10 spectral powers within 0-2.4Hz are extracted as features after using 3Hz low-pass filters. For RSP signals, 3Hz low-pass filters are adopted. Band energy ratio, average respiration signal, mean of the derivative, standard derivation, range of greatest breath, 10 spectral powers within 0-2.4Hz, average and median peak to peak time are extracted as features. For TEMP signals, mean, mean of the derivative, spectral powers in 0-0.1 Hz and 0.1-0.2 Hz are as features after 0.45Hz low-pass filters. For PLET signals, 0.5Hz low-pass filters are adopted. Average and standard derivation of HRV and inter-beat intervals, energy ratio between 0.04-0.15 Hz and 0.15-0.5 Hz, spectral power in 0.1-0.2 Hz, 0.2-0.3 Hz, 0.3-0.4 Hz, 0.01-0.08 Hz, 0.08-0.15 Hz and 0.15-0.5 Hz components of HRV are as features. The more details can be found in [6][7][5][4].

2.2. Capturing relations between EEG signals and peripheral physiological signals by RBM for complete data

The RBM model can learn the joint probability distribution over its visible nodes throught the hidden nodes. Therefore, we adopt the RBM to capture the high-order dependencies between EEG and peripheral physiological features. We refer to the work of Gao et al. [8]

As shown in Fig 1, the visible nodes of RBM can be divided into two parts, one part for EEG features and the other for the peripheral physiological features. Because the EEG features and peripheral physiological features are continuous, we adopt the Gaussian units for each visible node. $V^E \in R^{D^E}$ represents EEG features and $V^P \in R^{D^P}$ represents peripheral features where D^E and D^P represent the dimensions of EEG and peripheral features respectively, and $H \in \{0,1\}^{nhidden}$ are binary stochastic hidden units, where $nhidden$ represents the number of hidden nodes. The energy of the state V^E, V^P and h of our Gaussian-Bernoulli RBM [9] is defined as follows [10]:

$$E(V^E, V^P, h|\theta) = \sum_{i=1}^{D^E} \frac{(v_i^E - b_i^E)^2}{2(\sigma_i^E)^2} + \sum_{i=1}^{D^P} \frac{(v_i^P - b_i^P)^2}{2(\sigma_i^P)^2} - \sum_{i=1}^{D^E}\sum_{j=1}^{nhidden} \frac{v_i^E}{\sigma_i^E} W_{ij}^E h_j - \sum_{i=1}^{D^P}\sum_{j=1}^{nhidden} \frac{v_i^P}{\sigma_i^P} W_{ij}^P h_j - \sum_{j=1}^{nhidden} b_j^h h_j \quad (1)$$

where $\theta = \{\mathbf{b}, \sigma, \mathbf{W^E}, \mathbf{W^P}\}$ are the parameters of our model. The joint distribution over visible units is shown as follows:

$$P(V^E, V^P|\theta) = \frac{1}{Z(\theta)} \sum_H exp(-E(V^E, V^P, h|\theta)), \quad (2)$$

where $Z(\theta) = \int_{v^E}\int_{v^P}\sum_H exp(-E(V^E, V^P, h|\theta))dv^P dv^E$

Given a set of observations $\{V_n^E, V_n^P\}_{n=1}^N$, the derivative of the log-likelihood respect to $\mathbf{W^E}$ and $\mathbf{W^P}$ are shown in Eq 3 and Eq 4 respectively .

$$\frac{1}{N}\sum_{n=1}^{N} \frac{\partial logP(V_n^E, V_n^P; \theta)}{\partial W_{ij}^E} = E_{P_{data}}\left[\frac{v_i^E}{\sigma_i^E} h_j\right] - E_{P_{model}}\left[\frac{v_i^E}{\sigma_i^E} h_j\right] \quad (3)$$

$$\frac{1}{N}\sum_{n=1}^{N} \frac{\partial logP(V_n^E, V_n^P; \theta)}{\partial W_{ij}^P} = E_{P_{data}}\left[\frac{v_i^P}{\sigma_i^P} h_j\right] - E_{P_{model}}\left[\frac{v_i^P}{\sigma_i^P} h_j\right] \quad (4)$$

The Contrastive Divergence (CD) [10] algorithm is adopted to learn the parameters θ of our model. After training, the RBM model can capture the high-order relations among EEG and peripheral physiological signals, and generate new representations from multiple physiological signals. .

2.3. Capturing relations between EEG signals and peripheral physiological signals by RBM for incomplete data

For the incomplete data, our RBM model can generate the missing data as well as update the model parameters. After learning the RBM with the complete data, we use the incomplete data as well as the complete data to fine tune our RBM model. The missing part of the data is initialized randomly. Then, the missing values are treated as the same way as the model parameters, and are updated according to Eq. 5 from [10]:

$$v_i^t = v_i^{t-1} + \Delta v_i^t = v_i^{t-1} + \epsilon(\frac{\partial F}{\partial v_i^{t-1}} - \frac{\partial F}{\partial v_i^{t-1}}) \quad (5)$$

where t means the tth time iteration, ϵ represents the learning rate, v_i means the value of the ith visible node which is missing,$\hat{v}_i$ means the condition probability of the missing value given the hidden nodes and F satisfies Eq. 6.

$$e^{-F(V^E,V^P|\theta)} = \sum_h e^{-E(V^E,V^P,h|\theta)} \quad (6)$$

The details of training RBM with incomplete data are shown in Algorithm 1.

Algorithm 1 Training RBM with incomplete data

Require: training data (v^E, v^P), learning rate λ
Ensure: the parameters $\theta = \{\mathbf{b}, \sigma, \mathbf{W^E}, \mathbf{W^P}\}$.
  Initialize the parameters θ with complete data
  Initialize the missing value randomly
  repeat
    for each training instance (v^E, v^P) **do**
      $\hat{h}_j \leftarrow g(\sum_{i=1}^{D^E} W_{ij}^E \frac{v_i^E}{\sigma_i^E} + \sum_{i=1}^{D^P} W_{ij}^{(P)} \frac{v_i^P}{\sigma_i^P} + b_j^h)$
      $h_j \sim g(\sum_{i=1}^{D^E} W_{ij}^E \frac{v_i^E}{\sigma_i^E} + \sum_{i=1}^{D^P} W_{ij}^{(P)} \frac{v_i^P}{\sigma_i^P} + b_j^h)$
      $\hat{v}^E \leftarrow P(v^E|h)$
      $\hat{v}^P \leftarrow P(v^P|h)$
    end for
    update θ with Eq.3
    for each missing value v_i **do**
      $v_i^t = v_i^{t-1} + \Delta v_i^t = v_i^{t-1} + \epsilon(\frac{\partial F}{\partial \hat{v}_i^{t-1}} - \frac{\partial F}{\partial v_i^{t-1}})$
    end for
  until Convergence

3. EXPERIMENTAL RESULTS AND ANALYSIS

3.1. Experimental Conditions

We conduct the experiments on two benchmark databases to evaluate the performance of our method, the DEAP database [6] and the MAHNOB-HCI database [7].

The DEAP database is a multimodal dataset for analysis of human affective states. It records eight kinds of physiological signals, i.e. EEG, EOG, EMG, ECG, GSR, RSP, TEMP and PLET, from 32 participants during their watching music videos. Each segment is labeled by self-reported feelings using 9-scale ratings (1-9). We divide them into two classes, positive (rating 6-9) and negative (rating 1-5). Specifically, For valence, there are 672 positive and 544 negative segments, for arousal, there are 726 high and 490 low segments in the DEAP database.

The MAHNOB-HCI database is a multimodal database for affect recognition and implicit tagging. It contains five kinds of physiological signals, i.e. EEG, ECG, GSR, RESP and TEMP, from 27 participants watching 20 emotional videos. Each segment is labeled by self-reported feelings using 9-scale ratings (1-9). We divide them into two classes, positive (rating 6-9) and negative (rating 1-5). Totally, in the valence space, 289 samples are positive and the rest are negative. In the arousal space, there are 268 positive and 265 negative samples.

To validate the effectiveness of our model, we conduct two groups of experiments on both database: emotion recognition with complete data and emotion recognition with incomplete data. For the complete data, we compare our model with merely using peripheral features and merely using EEG features by RBM. Furthermore, we compare our model with feature-level fusion and decision-level fusion using SVM. For feature-level fusion, the EEG features and peripheral physiological features are concatenated as the input of SVM. For decision-level fusion, a weighted strategy is adopted to combine the recognition results from two SVMs, which are trained from EEG features and peripheral physiological features respectively. For the incomplete data, we randomly miss EEG features or peripheral features at the rate of 5%, 10%, 20% and 40% respectively. We compared our method with the method discarding the whole data that have missing part. The compared method also use RBM to generate new features, and SVM as the classifier. We adopted leave-one-video out cross-validations in our experiments and used three commonly-used metrics accuracy, i.e.F1-score and kappa to verify the effectiveness of our method.

3.2. Experimental results and analyses of complete data

The results of emotional recognition from EEG and peripheral signal on the DEAP database and the MAHNOB-HCI database are shown in Table 1 and Table 2 respectively.

From the above two tables, we can obtain the following observations:

First, compared with peripheral physiological signals, EEG signals may be better for emotion recognition, since emotion recognition from EEG signals outperforms that from peripheral physiological signals using either RBM+SVM or SVM on both databases. Second,among the three fusion methods , the proposed RBM fusion method performs best. This demonstrates that the proposed fusion method can successfully capture the dependencies among multiple physiological signals, and result in good performance on multimodal emotion recognition. Third, for emotion recognition from EEG signals or peripheral physiological signals, the RBM+SVM method outperforms the SVM method in most cases, further suggesting the good representation of RBMs.

3.3. Experimental results and analyses of incomplete data

Fig. 2 and Fig. 3 show the emotion recognition results from incomplete data on the DEAP database and MAHNOB-HCI database respectively. From the two figures, we find that the proposed method outperforms the method which discards the whole data with missing part, demonstrating the proposed method successfully exploit all available data for emotion recognition. With the increase of the missing rate, the emotion recognition performance become worse and worse. It is reasonable, since less EEG signals and peripheral physiological signals provide less information for emotion recognition, and thus results in decreased performance. Compared Fig.2(c), Fig.2(d), Fig.3(c), Fig.3(d) with Fig.2(a), Fig.2(b), Fig.3(a) Fig.3(b)respectively, we can find that missing the EEG features will heavily influence the emotion recognition performance. it may indicate that EEG signals contain more information for emotion recognition.

3.4. Comparision with related works

Few works recognize emotions from EEG and peripheral physiological signals on the DEAP database and MAHNOB-HCI database.

Table 1. Emotion recognition results on the DEAP database with complete data

	Valence							Arousal						
	RBM +SVM EEG	RBM +SVM Peripheral	Our model	SVM EEG	SVM Peripheral	SVM feature-level fusion	SVM decision-level fusion	RBM +SVM EEG	RBM +SVM Peripheral	Our model	SVM EEG	SVM Peripheral	SVM feature-level fusion	SVM decision-level fusion
Accuracy	59.5%	56.3%	**60.7%**	58.0%	51.6%	58.9%	58.0%	60.3%	54.9%	**64.6%**	61.7%	58.1%	62.8%	61.7%
F1 score	0.535	0.510	**0.541**	0.522	0.464	0.527	0.522	**0.532**	0.508	0.512	0.511	0.480	0.521	0.511
Kappa	0.177	0.114	**0.199**	0.147	0.024	0.192	0.147	0.216	0.103	**0.240**	0.196	0.130	0.218	0.196

Table 2. Emotion recognition results on the MAHNOB-HCI database

	Valence							Arousal						
	RBM +SVM EEG	RBM +SVM Peripheral	Our model	SVM EEG	SVM Peripheral	SVM feature-level fusion	SVM decision-level fusion	RBM +SVM EEG	RBM +SVM Peripheral	Our model	SVM EEG	SVM Peripheral	SVM feature-level fusion	SVM decision-level fusion
Accuracy	58.3%	51.8%	**59.1%**	52.9%	46.5%	57.4%	52.9%	65.3%	58.5%	**65.9%**	58.9%	56.3%	64.0%	58.9%
F1 score	0.539	0.505	0.542	**0.569**	0.504	0.608	**0.569**	0.646	0.588	**0.654**	0.588	0.574	0.642	0.588
Kappa	0.159	0.038	**0.173**	0.050	0.076	0.142	0.050	0.306	0.171	**0.317**	0.178	0.125	0.280	0.178

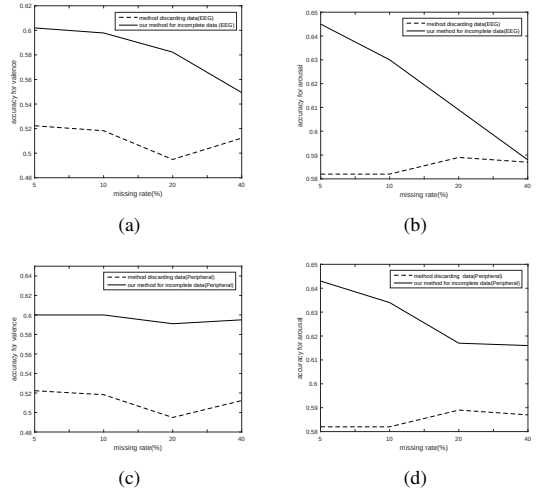

Fig. 2. Experimental results for incomplete data on the DEAP database

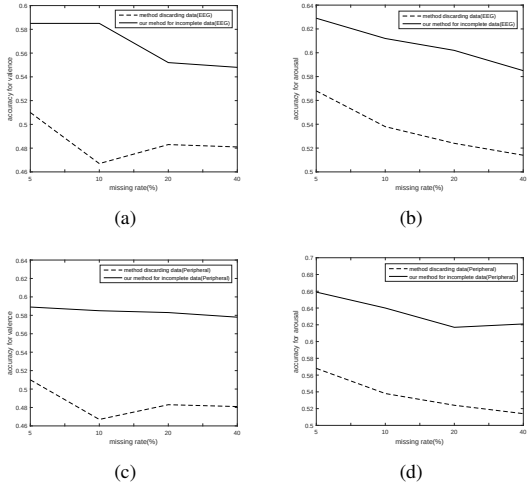

Fig. 3. Experimental results for incomplete data on the MAHNOB-HCI database

Table 3. Comparision with related works

Database	DEAP				MAHNOB-HCI			
Models	Valence		Arousal		Valence		Arousal	
	Acc.	$F1^*$	Acc.	$F1^*$	Acc.	$F1^*$	Acc.	$F1^*$
Chen's experiment	58.1%	0.575	62.7%	0.602	57.4%	0.551	61.3%	0.613
Ours Fusion	60.7%	0.599	64.6%	0.618	59.1%	0.586	65.9%	0.654

Therefore, we compare our method with feature-level fusion and decision-level fusion using SVM as shown in Section 3.2. Furthermore, we compare our work with Chen et al.'s [4] , which adopt the CCA to model the relations between EEG signals and peripheral physiological signals for emotion recognition, and conduct experiments on the DEAP database and the MAHNOB-HCI database. The results are listed in Table3, where $F1^*$ means the average F1 score for two classes which is adopted in [4]. From Table 3, we can find that our method outperforms theirs, further demonstrating that our method successfully captures the high-order dependence between EEG and peripheral physiological signals, and constructs a good feature space for emotion recognition. Since Chen et al.'s used both EEG and peripheral physiological signals for training, but only peripheral physiological signals for testing. The above comparison is only for reference.

4. CONCLUSION

In this paper, we propose the RBM model to capture the relations between EEG and peripheral physiological signals for a better feature representation. The proposed model handles the missing value of the training data explicitly. The results on the MAHNOB-HCI and the DEAP database demonstrate that with complete data, our model can combine EEG and peripheral physiological signals to construct a better feature space for emotion recognition, with incomplete data, our model can exploit the rest data to achieve better performance.

5. ACKNOWLEDGMENT

This work has been supported by the National Science Foundation of China(Grant No. 61473270, 61228304, 61175037), and the project from Anhui Science and Technology Agency (1508085SMF223).

6. REFERENCES

[1] Sidney K D'mello and Jacqueline Kory, "A review and meta-analysis of multimodal affect detection systems," *ACM Computing Surveys (CSUR)*, vol. 47, no. 3, pp. 43, 2015.

[2] C Vinola and K Vimaladevi, "A survey on human emotion recognition approaches, databases and applications," *ELCVIA Electronic Letters on Computer Vision and Image Analysis*, vol. 14, no. 2, pp. 24–44, 2015.

[3] Johannes Wagner, Elisabeth Andre, Florian Lingenfelser, and Jonghwa Kim, "Exploring fusion methods for multimodal emotion recognition with missing data," *IEEE Transactions on Affective Computing*, vol. 2, no. 4, pp. 206–218, 2011.

[4] Shiyu Chen, Zhen Gao, and Shangfei Wang, "Emotion recognition from peripheral physiological signals enhanced by eeg," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2016, pp. 2827–2831.

[5] Jonghwa Kim and Elisabeth André, "Emotion recognition based on physiological changes in music listening," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 30, no. 12, pp. 2067–2083, 2008.

[6] 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," *Affective Computing, IEEE Transactions on*, vol. 3, no. 1, pp. 18–31, 2012.

[7] Mohammad Soleymani, Jeroen Lichtenauer, Thierry Pun, and Maja Pantic, "A multimodal database for affect recognition and implicit tagging," *Affective Computing, IEEE Transactions on*, vol. 3, no. 1, pp. 42–55, 2012.

[8] Zhen Gao and Shangfei Wang, "Emotion recognition from eeg signals by leveraging stimulus videos," in *Pacific Rim Conference on Multimedia*. Springer, 2015, pp. 118–127.

[9] Geoffrey E Hinton and Ruslan R Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[10] Geoffrey Hinton, "A practical guide to training restricted boltzmann machines," *Momentum*, vol. 9, no. 1, pp. 926, 2010.