EEG Based Emotion Identification Using Unsupervised Deep Feature Learning

Xiang Li[1], Peng Zhang[1], Dawei Song*[1,2], Guangliang Yu[1], Yuexian Hou[1], Bin Hu[3]
[1]Tianjin Laboratory of Cognitive Computing and Application, Tianjin University, China
[2]The Computing Department, The Open University, United Kingdom
[3]The Ubiquitous Awareness and Intelligent Solutions Lab, Lanzhou University, China
{xli0924, pzhang, dwsong, glyu, yxh}@tju.edu.cn, bh@lzu.edu.cn

ABSTRACT
Capturing user's emotional state is an emerging way for implicit relevance feedback in information retrieval (IR). Recently, EEG-based emotion recognition has drawn increasing attention. However, a key challenge is effective learning of useful features from EEG signals. In this paper, we present our on-going work on using Deep Belief Network (DBN) to automatically extract high-level features from raw EEG signals. Our preliminary experiment on the DEAP dataset shows that the learned features perform comparably to the use of manually generated features for emotion recognition.

Categories and Subject Descriptors
H.3.3 [**Information Storage and Retrieval**] Information Search and Retrieval

Keywords
Emotion Recognition, EEG, Deep Feature Learning

1. INTRODUCTION
Implicit relevance feedback (IRF) has been extensively studied in IR, for relevance prediction based on user interaction and behaviors. A recently emerging line of research in IRF is to capture the users' affective or emotional states. A typical approach is based on the user's facial expressions, from which various basic emotional states can be detected and used as implicit relevance indicators [1]. While encouraging results have been achieved, these methods do not directly reveal the users' actual internal cognitive states.

Compared with facial expression, electroencephalogram (EEG) is a reliable approach to probe the internal cognitive and emotional changes of users. It has attracted increasing attention in the fields of Human Computer Interaction (HCI), Brain Computer Interface (BCI) and Cognition Science. Its high temporal resolution indeed provides us with a large amount of useful information in analyzing the users' real emotional state directly. A key problem is how to extract useful features from raw EEG signals for accurate emotion prediction.

A commonly used way is to manually extract various time and frequency domain features that are potentially useful for EEG-based emotional state recognition tasks [3]. However, the manual feature extraction process is time consuming and requires extensive domain-dependent signal processing strategies, but it is often not obvious which features will be relevant for a given task.

Therefore, our research question is: for a given emotional state recognition task, can we automatically learn higher level features directly from the raw EEG signals, which better represent the underlying data than manually engineered signal-processing features?

For this purpose, we propose to use deep learning technique for automatic feature extraction, given the deep architecture's proven unsupervised feature learning ability. We will show in the experiment that these learned features generate a good emotion recognition performance that is comparable to the use of frequency features extracted by signal-processing technique.

NeuroIR'15, August 13, 2015, Santiago, Chile.

2. METHODOLOGY
We adopt a two-hidden-layer Deep Belief Networks (DBN) as the deep architecture for unsupervised feature learning and reduction [2]. DBN is a probabilistic generative model, which is composed of stacked Restricted Boltzmann Machines (RBM) and can be trained in a greedy manner layer-by-layer with unlabeled data. The whole architecture of our work is demonstrated as follows in Figure 1.

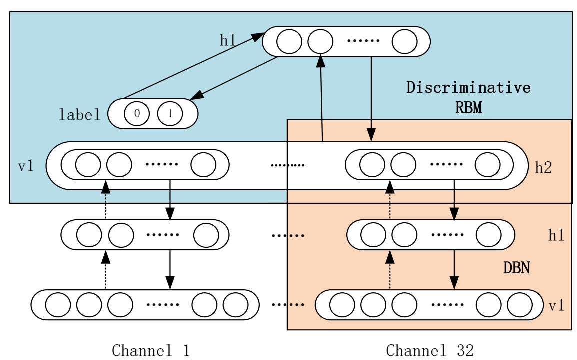

Fig.1 Architecture of Unsupervised Feature Learning and Classification

Considering spatial distribution variance on scalp between the different EEG channels, as shown in Figure 1, we build a DBN for each EEG channel data to extract higher-level abstract features. Then, a discriminative RBM is built upon the combined h2 layer. The DBNs are first pre-trained with the training set in an unsupervised manner, followed by a supervised fine-tuning using the same data with labels conducted. To determine the DBNs' hyper-parameters, we tested different combinations of hyper-parameters and finally adopt the parameter combination with the minimal recognition error. Once trained, the activations of the DBNs' hidden units are extracted as a learned higher-level representation of raw EEG signals for the further validation.

3. EXPERIMENT AND EVALUATION
3.1 Experimental Data
The experiment was conducted on the publicly available multimodal dataset, namely DEAP [4], including EEG signals

collected from central nerves system and other types of physiological signals collected from peripheral nerves system. Different emotions were elicited by some carefully selected music and video stimuli, and the multimodal signals mentioned above were continually recorded during the trials. Table 1 gives an overview of the dataset.

Table 1. DEAP Dataset Content Summary

Number of Subjects	32
Number of Trials	40 per subject
Emotional Dimensions	Russell's valence-arousal scale [5] {Valence, Arousal, Dominance, Liking, Familiarity}
Rating Values	Familiarity: Discrete value of 1~5 Others: Continues value of 1~9
Channels & Signals	EEG: 32 channels Peripheral Signals: 8 channels Sampling rate: 128Hz Length: 60s

In this work, we focus on the EEG based emotional state recognition, thus the signals from EEG channels 1 to 32 were extracted in our analysis.

3.2 Data Preprocessing

Labeling the samples is critical for both traditional machine learning and deep learning methods. We divide the trials into two classes (respectively representing positive and negative emotion) according to each trial's rating value (positive: ≥ 5, negative: < 5, and only few trials' rating value equals to 5 in DEAP).

In each trial, we get 32 channels' EEG signals and divide each channel signal into 60 segments with 1s length per segment, for the purpose of increasing the number of training samples. Accordingly, we get a total of 2400 samples (40 trials $\times$ 60 segments) for each subject, where each sample is composed of 32 channels' 1s signal segments and the samples from the same trial are assigned with the same emotional label.

In order to compare the performance of the extracted features obtained by deep learning with the traditional manually generated features, the widely-applied power spectral density (PSD) [3, 4] feature was extracted for each 1s segment by using the Welch method. We compute the PSD feature for different frequency bands in theta band (4~8Hz), alpha band (8~12Hz), beta band (12~30Hz) and gamma band (30~45Hz). In summary, for each subject the input sample set consists of 2400 samples $\times$ 4096 dimensions (128Hz $\times$ 1s $\times$ 32 channels) for deep learning architecture, and 2400 samples $\times$128 dimensions (4 bands PSD $\times$32 channels) for the PSD feature.

3.3 Experimental Evaluation

The numbers of visible and hidden nodes of each DBN configured here are 128-10-10. as shown in Figure 1, after the unsupervised training and feature learning, the samples' feature dimensionality is reduced from 4096 to 320 (10 dimensions$\times$32 channels). After the feature extraction and reduction process, a SVM classifier with RBF kernel is applied on both PSD and DBN features. By using the same classifier we are able to carry out direct comparison between the two types of features. We evaluate the feature effectiveness and recognition accuracy on four emotional dimensions. The validation method adopted here is '10-folds cross-validation', i.e., 10% of samples are left for testing and the rest are used for training (training set: 2160 samples, test set: 240 samples). The results, in term of emotion recognition accuracy, for the two type of features are listed below for comparison.

Table 2. Recognition Accuracy Results (%)

Recognition Accuracy Results for Each Subject on 'Liking'											
Sub	**PSD**	**DBN**	**Sub**	**PSD**	**DBN**	**Sub**	**PSD**	**DBN**	**Sub**	**PSD**	**DBN**
1	95.0	95.0	***9***	74.9	75.0	***17***	62.4	62.5	***25***	62.3	62.5
2	67.0	70.0	***10***	72.4	72.5	***18***	56.7	57.5	***26***	77.6	77.5
3	82.4	85.0	***11***	51.7	52.5	***19***	67.5	67.5	***27***	85.0	85.0
4	50.5	52.6	***12***	59.8	60.0	***20***	72.5	72.5	***28***	62.4	62.5
5	59.6	60.0	***13***	65.0	67.5	***21***	62.1	62.5	***29***	57.3	57.5
6	70.0	70.0	***14***	70.0	70.0	***22***	59.5	65.0	***30***	53.0	52.5
7	92.5	92.5	***15***	66.5	67.5	***23***	52.5	52.4	***31***	72.5	72.5
8	85.0	85.0	***16***	50.9	52.6	***24***	52.2	55.0	***32***	52.8	48.5

Average Recognition Accuracy Results across Subjects on 'Valence', 'Arousal', 'Dominance' and 'Liking' Respectively				
	Valence	**Arousal**	**Dominance**	**Liking**
PSD	58.2	64.3	65.1	66.3
DBN	58.4	64.2	65.8	66.9

The preliminary experiment results indicate that the abstract features extracted by DBN are comparable to the manually extracted frequency features in emotion recognition.

4. CONCLUSIONS AND FUTURE WORK

This paper investigates on an unsupervised deep feature learning approach in an EEG-based emotion recognition task. The experimental results presented in Section 3.3 show that it is feasible to learn useful affective features from raw EEG data through a deep feature learning framework and its effectiveness is comparable to manually extracted features. In the future, we will further refine the model. Furthermore, the subjects' ambiguity in self rating after each trial and critical EEG channel selection for feature reduction should be taken into consideration. Finally, we aim to apply the real-time EEG analysis in the IR setting. For example, the users' affective responses to the retrieval results can be detected and used as affective features to train a personalized relevance model with the user tagged/clicked documents. Then the trained model can be used to predict the topical relevance of other documents. User's personalized profile get enriched and updated in this process.

5. ACKNOWLEDGEMENT

This work is funded in part by the Chinese National Program on Key Basic Research Project (973 Program, grant no. 2013CB329304 and 2014CB744604), the Chinese 863 Program (grant no. 2015AA015403), and the Natural Science Foundation of China (grant no. 61272265 and 61402324).

6. REFERENCES

[1] I. Arapakis, I. Konstas and J. M. Jose. Using facial expressions and peripheral physiological signals as implicit indicators of topical relevance. In ICMR, pages 461-470. ACM, 2009.

[2] G. E. Hinton and R. R. Salakhutdinov. Reducing the dimensionality of data with neural networks. Science 313(5786): 504-507, 2006.

[3] R. Jenke, A. Peer and M. Buss. Feature extraction and selection for emotion recognition from EEG. IEEE Transactions on Affective Computing, 5(3): 327-339, 2014.

[4] S. Koelstra, C. Muhl, M. Soleymani, et al. Deap: A database for emotion analysis; using physiological signals. IEEE Transactions on Affective Computing, 3(1): 18-31, 2012.

[5] J. A. Russell. A circumplex model of affect. Journal of Personality and Social Psychology, 39(6): 1161, 1980.