MULTIMODEL EMOTION ANALYSIS IN RESPONSE TO MULTIMEDIA

*Wei-Long Zheng, Jia-Yi Zhu, and Bao-Liang Lu**

Department of Computer Science and Engineering
Key Lab. of Shanghai Education Commission for Intelligent Interaction and Cognitive Eng.
Shanghai Jiao Tong University, 800 Dong Chuan Rd., Shanghai, 200240, China

ABSTRACT

In this demo paper, we designed a novel framework combining EEG and eye tracking signals to analyze users' emotional activities in response to multimedia. To realize the proposed framework, we extracted efficient features of EEG and eye tracking signals and used support vector machine as classifier. We combined multimodel features using feature-level fusion and decision-level fusion to classify three emotional categories (positive, neutral and negative), which can achieve the average accuracies of 75.62% and 74.92%, respectively. We investigated the brain activities that are associated with emotions. Our experimental results indicated there exist stable common patterns and activated areas of the brain associated with positive and negative emotions. In the demo, we also showed the trajectory of emotion changes in response to multimedia.

***Index Terms*—** Emotion recognition, EEG, eye track, affective computing

1. INTRODUCTION

As more and more social media websites emerge, characterizing multimedia content with reliable tags become important for multimedia information retrieval. Implicit tagging of multimedia using emotion cues can help recommendation and retrieval system, because explicit tagging using words carries inaccurate information for different criteria and can not be recorded online. Moreover, in recent years, the field of affective computing (AC) has attracted more and more attention in human computer interface (HCI). The cognitive and emotional states of users become important as a key human factor to make machines work in a more human-like manner. These include basic researches on emotions during interactions with computers, emotion synthesis in computers, and systems that automatically detect and respond to users' emotional states [1]. Emotion detection and recognition can bridge the communicative gap between emotionally expressive human and the emotionally deficit computer. A human computer interface which is sensitive to users' emotional states is expected to be more natural and enjoyable.

This work was partially supported by the National Natural Science Foundation of China (Grant No. 61272248), the National Basic Research Program of China (Grant No. 2013CB329401) and the Science and Technology Commission of Shanghai Municipality (Grant No.13511500200). *Corresponding author: B. L. Lu (bllu@sjtu.edu.cn)

2. EXPERIMENT

In our experiment, emotional movie clips are chosen to help subjects elicit their emotions. We choose fifteen emotional movie clips in one experiment (five for each emotional state including positive, neutral and negative) and each one lasts for about 4 minutes. While subjects watched movie clips and their corresponding emotions were evoked, EEG data and eye tracking data were collected simultaneously. EEG was recorded using an ESI NeuroScan System at a sampling rate of 1000 Hz from 62-channel electrode cap according to the international 10-20 system. Eye tracking data was recorded using SMI eye track glasses with 30 Hz of temporal resolution. We also recorded the frontal face videos in the experiments. More detailed information about the experiment design can be found in [2].

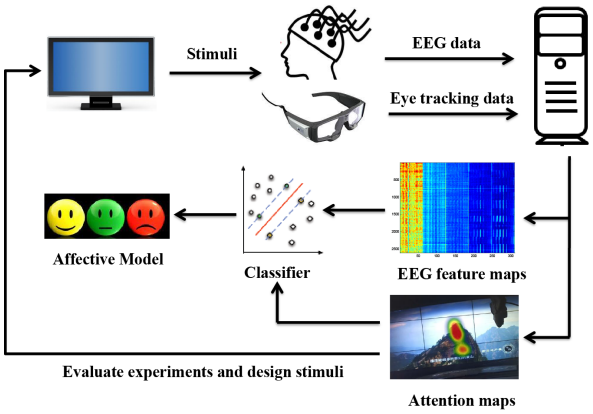

Fig. 1. The framework of our experiment processing.

The framework of our experiment processing is shown in Figure 1. For EEG data, we extract differential entropy features from five frequency bands (delta: 1-3Hz, theta: 4-7Hz, alpha: 8-13Hz, beta: 14-30Hz, gamma: 31-50Hz) as described in [2] and then train a classifier to build an affec-

tive model, which learns to identify users' current emotional states. For eye tracking data, we can get other physiological information such as blink frequency and pupil size. to build a multimodal fusion model and evaluate the models. Here, we extract mean value, standard deviation, spectral powers of frequency bands (0-0.2Hz, 0.2-0.4Hz, 0.4-0.6Hz and 0.6-1Hz) of pupil diameter [3].

Further, we can also get attention maps of subjects while participating in the experiments. This information will help to evaluate the experiments and design stimuli. The eye tracker records exactly where the users are looking and how long they spend looking in one place. While a standard usability study relies on audio-visual feedback alone, eye tracking provides qualitative evaluation for the experiments in a natural way and allows us to observe the thought processes of the users.

3. RESULTS

There were total 13 experiments evaluated here. We first projected DE features to the scalp to find the temporal dynamics of DE features and stable common patterns across subjects. Figure 2 depicts the average neural patterns of 13 experiments for positive and negative emotions. The result shows that neural signatures associated with positive and negative emotions do exist. The lateral temporal areas activate more for positive emotion than negative one in beta and gamma bands, while the energy of the prefrontal area enhance for negative emotion over positive one in beta and gamma bands. In the demo video, we demonstrated the dynamic changes of EEG patterns in response to emotional multimedia.

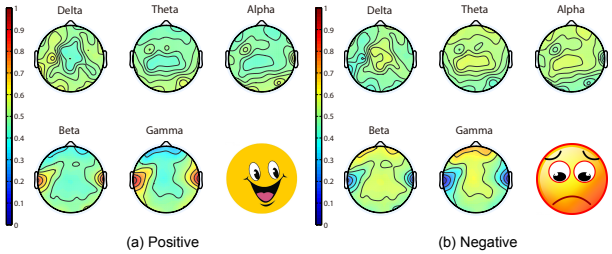

Fig. 2. The average neural patterns of 13 experiments for positive and negative emotions.

Table 1 shows the accuracies (%) of 13 experiments for different modalities using SVM as classifier. For SVM, we employed linear kernel and searched the parameter space to find the optimal value. As we can see from Table 1, we achieved 73.94% and 58.07% accuracies from EEG signals and eye gaze data, respectively. We can also see that the fusion model combining EEG signals and eye gaze data can improve the performance of emotion recognition. For feature level fusion, we obtained 75.62% accuracy. For decision level fusion, we used a max principle to combine two models trained with EEG signals and eye gaze data, respectively, which achieved 74.92% accuracy. The performance of multimodel emotion recognition outperformed single modality using EEG and eye tracking data. To track the trajectory of emotion changes in response to emotional multimedia, we used manifold learning methods to project high-dimension features into low-dimension space, as shown in Figure 3. In the demo video, we showed the trajectory of emotion changes as well as the regions of interest where subject focused while watching movie videos.

Table 1. The Accuracies (%) of 13 Experiments for Different Modalities

Modality	Mean Accuracy	Std.
EEG	73.94	13.93
Eye gaze	58.07	**11.88**
Feature level fusion	**75.62**	15.64
Decision level fusion	74.92	12.30
Random level	33.33	/

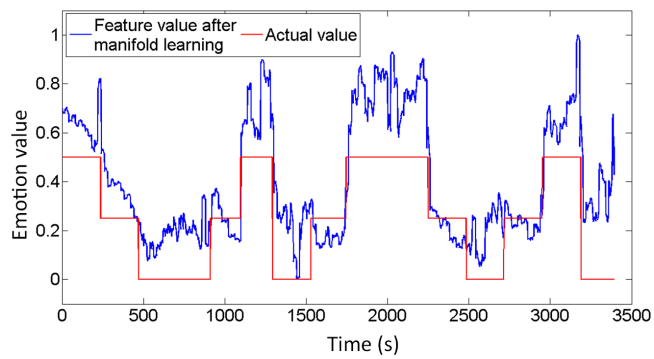

Fig. 3. The trajectory of emotion changes during one experiment.

4. REFERENCES

[1] Rosalind W Picard, "Affective computing: from laughter to IEEE," *IEEE Transactions on Affective Computing*, vol. 1, no. 1, pp. 11–17, 2010.

[2] Ruo-Nan Duan, Jia-Yi Zhu, and Bao-Liang Lu, "Differential entropy feature for EEG-based emotion classification," in *2013 6th International IEEE/EMBS Conference on Neural Engineering (NER)*. IEEE, 2013, pp. 81–84.

[3] Mohammad Soleymani, Maja Pantic, and Thierry Pun, "Multimodal emotion recognition in response to videos," *IEEE Transactions on Affective Computing*, vol. 3, no. 2, pp. 211–223, 2012.