2K3-OS-33a-3

Analysis of EEG response and Annotation Lag in EEG-based Emotion Recognition using Fusion Technique

Nattapong Thammasan[*1*2] Ken-ichi Fukui [*2*1] Masayuki Numao [*2*1]

[*1] Graduate School of Information Science and Technology, Osaka University
[*2] Institute of Scientific and Industrial Research (ISIR), Osaka University

Multimodality of electroencephalogram (EEG) and music information has been introduced to overcome challenges and limitation of EEG-based music-emotion recognition in recent years. In this study, we present empirical results suggesting that taking the EEG response and annotation delays into account by using multimodal fusion technique, the performance of music-emotion recognition can be improved.

1. Introduction

Emotion recognition based on EEG has been suffering from nonlinearity and nonstationarity of EEG signals. Recently, musical features extracted from stimuli have been employed in conjunction with EEG signals to enhance the performance using feature-level fusion [Lin et al. 2014] and decision level fusion [Thammasan et al. 2017] with the hypothesis that both modalities might play a complementary role in music-emotion recognition model. Interestingly, considering feature concatenation in feature-level fusion, the framework also allows shifting features in one modality before concatenating with features from another modality. Therefore, in this work, we present the results of shifting EEG features in a particular step to investigate any effect of EEG lag due to emotional response delay from the presented stimuli. Furthermore, previous affective computing research [Soleymani et al. 2016, Mariooryad&Busso 2013] reported the effect of annotation lag on continuous emotion recognition and that considering the delay of annotation improved emotion recognition. Hence, we also include the simultaneous analysis of annotation lags on continuous emotion recognition in this study.

2. Research Methodology

2.1 Experimental protocol

Twelve healthy male subjects were recruited to participate in our experiment. The task for each subject was to listen to the self-selected 16 songs. During music was presented to each subject, EEG signals were acquired from the 12 electrodes of Waveguard EEG cap placed in accordance with the 10-20 international system with Cz reference electrode (250 Hz sample frequency, 0.5-60 Hz bandpass filter). At the end of each song, each subject annotated the emotion by continuously clicking on a corresponding point in arousal-valence space shown on a monitor screen. For more details of experimental protocol, please refer to our previous work [Thammasan et al. 2017].

Contact: Nattapong Thammasan, Department of Architecture for Intelligence, ISIR, Osaka University, E-mail: nattapong@ai.sanken.osaka-u.ac.jp

2.2 EEG preprocessing and feature extraction

The acquired EEG signals were preprocessed by employing EEGLAB toolbox[*1]. Based on independent component analysis, independent components that were related to eye blinking, eye movement, muscle activity, and noise, were removed from EEG signals to eliminate unrelated artifacts. Then, power spectral density technique was employed to extract informative features from EEG signals. For each electrode, power spectral density was calculate and spectral features in delta (1–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), beta (13–30 Hz), and gamma (30–40 Hz) band within the boundary of non-overlapping 2-s window were obtained using power averaging in spectral band. In addition, asymmetry indexes calculated from the differential asymmetries of five left-right electrode pairs were also included in our EEG feature set. In total, we derived 85 features from EEG modality.

2.3 Musical features

To extract musical feature, we employed MIRtoolbox version 1.6.1[*2] which is a MATLAB toolbox containing integrated set of high-level musical features extraction function. In each non-overlapping 2-s sliding window, the converted MIDI-to-WAV (at a sampling rate of 44.1 kHz) provided totally 37 musical features, which could be categorized into dynamic, rhythm, timbre, tonal feature groups [Thammasan et al. 2017].

2.4 Feature fusion and Emotion Classification

To fuse features from different modality, we used feature-level fusion, which was firstly to extract features independently from each modality and then fuse them to form a composite feature vector to be inputted to classifiers. Finally, each feature vector was labeled with ground-truth emotion via timestamps. The majority approach was used to determine emotion of a particular window that contained variation in emotion annotation. To investigate the effects of EEG and annotation lags, we shifted EEG features and the annotation from musical features on the processes of the feature concatenation and labeling respectively. Using

*1 sccn.ucsd.edu/eeglab/
*2 www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials/mirtoolbox

dataset with 2-s sliding window, we varied EEG delays and annotation delays from 0 to 8 s at a step of 2 s.

For the sake of simplicity, emotion recognition was turned to be a binary classification of arousal (high vs. low) and valence (positive vs. negative). In this work, we used classifying trees built by MATLAB Statistics and Machine Learning toolbox*3 as emotion classifiers. We adopted leave-one-subject-out validation method to derive the subject-independent performance of classification. Prior to classification, each feature was independently normalized to the range of $[0, 1]$ using the min-max approach.

As self-reporting emotion annotation could lead to the imbalance in emotional classes, we employed Matthews correlation coefficient (MCC) [Matthews 1975] to reflect classification performance with consideration of class imbalance. Given a confusion matrix of binary classification, MCC can be calculated by

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}, \quad (1)$$

where TP is the number of true positives, TN is the number of true negatives, FP is the number of false positives and FN is the number of false negatives.

Shifting EEG or annotation lag could lead to inequality in the number of instances in training and testing sets, which could provide misleading results. Therefore, we used a technique of sub-sampling, which was to randomly remove a particular number of instances. The sub-sampled dataset had the same amounts of instances as in the set with maximal shifting range (8 s). As sub-sampling process relied on randomization, we repeated our experiment 10 times and derived the averages to reflect the overall performance.

3. Results

The emotion classification results considering EEG and annotation delays concurrently are shown in Figure 1. In general, considering EEG response delay of 4 s boosted performances of arousal classification. We also found the evidence suggesting that considering annotation delay of 8 s improved performance in arousal and valence classification and this is in line with a previous study [Bachorik et al. 2014] that reported the 8.31 s time requirement for participants to initiate music-emotional judgments. We further analyzed the trees beyond intensive tree pruning upon the conditions of best arousal and valence classification performance. We found EEG features appeared as 50.0% and 46.2% of the total number of features in the arousal and valence pruned classifying trees respectively, suggesting that both modalities played a complementary role in emotion classification. Overall, we might infer that there existed 4 s lag of arousal response to musical structures and emotion annotation lag of 8 s. Underlying physiological evidence of the existence of arousal EEG lag and the absence of valence EEG lag are encouraged to investigate in the future work. In addition, inter-subject variability that causes relatively low subject-independent performance is worthy to be addressed.

*3 www.mathworks.com/products/statistics

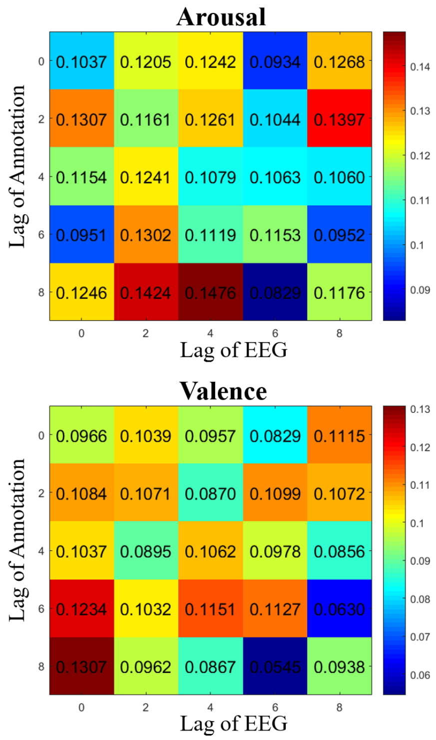

Figure 1: Averaged emotion classification MCCs across subjects using feature-level fused features and considering different EEG and annotation lags

References

[Bachorik et al. 2014] J. Bachorik, M. Bangert, P. Loui, K. Larke, J. Berger, R. Rowe, G. Schlaug; Emotion in Motion: Investigating the Time-Course of Emotional Judgments of Musical Stimuli, Music Perception, 26, pp.355-364, 2009,

[Mariooryad&Busso 2013] S. Mariooryad, C. Busso; Analysis and Compensation of the Reaction Lag of Evaluators in Continuous Emotional Annotations, Proc.ACII-2013, pp.85-90, 2013,

[Matthews 1975] B.W. Matthews; Comparison of the predicted and observed secondary structure of T4 phage lysozyme, Biochimica et Biophysica Acta (BBA) - Protein Structure, 405(2), pp.442-451, 1975,

[Soleymani et al. 2016] M. Soleymani, S. Asghari-Esfeden, Y. Fu, M. Pantic; Analysis of EEG Signals and Facial Expressions for Continuous Emotion Detection, IEEE Trans. Affective Computing, 7(1), pp.17-28, 2016,

[Thammasan et al. 2017] N. Thammasan, K. Fukui, M. Numao; Multimodal fusion of EEG and musical features in music-emotion recognition, Proc.AAAI-17, 2017,

[Lin et al. 2014] Y.P. Lin, Y.H. Yang, T.P. Jung; Fusion of electroencephalogram dynamics and musical contents for estimating emotional responses in music listening, Frontiers in Neuroscience, 8(94), 2014.