The 33rd Annual Conference of the Japanese Society for Artificial Intelligence, 2019

Feasible Affect Recognition in Advertising based on Physiological Responses from Wearable Sensors

Taweesak Emsawas[*1] Ken-ichi Fukui[*2] Masayuki Numao[*2]

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

Recent studies in affective computing have facilitated and stimulated the development of systems and sensors that can recognize and interpret human affects. Affective computing has been applied in various domains, and one of the applied domains is in the marketing area to increase the consumers' appeal and attraction. In particular, advertisements (ads) can convey amounts of information in a short time. Therefore, using physiological responses can help to acquire a user's feedback and obtain an advantage. This study proposes non-invasive affect recognition in each scene of an advertising video using electroencephalogram (EEG), electrocardiogram (ECG) and eye-tracking. The preliminary analysis of EEG shows the relationship between scene feeling score and emotional affects regarding physiological responses. Hence, we also trained two types of recognition models: window recognition and sequence learning. The models learned from the physiological responses and questionnaires on a user's preference in each ad scene.

1. Introduction

In the marketing area, Marketer always tries to reach maximum customer for getting their attraction and sales. Consequently, the ads have been created in various ways such as articles, billboards, videos, and so on. Nowadays, TV commercial plays a role in human life, and consumer commonly watches many ad videos through various channels. The goal of creating ads video is to convey amounts of information in a short time. For this reason, all scene in the ads is essential information and should make an impression. To comprehend feedback and obtain an advantage, consumers' response is needed, and there can gather in many ways such as questionnaires, external appearances, and internal appearances.

Recently, the studies are interested in understanding the user's preferences regarding human affect or emotion. Hence, emotion recognition based on physiological signals has been a hot topic and applied in various domains such as health care, game and commercial [Shu 2018]. The emotion can be represented in various ways. In psychology, human affect is a concept to describe the experience of feeling or emotion, and affect transmits physiological responses with stimuli (e.g., pictures, audio and videos). To acquire the data from a subject, Sensors have been developed for different purposes such as for laboratory or wearable depend on the feasibility and accurate circumstances [Ragot 2017].

In recognition method based on physiological data, after extracting and selecting the features, various researches utilize these features to train a model and classify different emotional states. However, each model applies a different method to represent input data or physiological signals. General techniques train with the whole signal data and classify labels such as multilayer perceptron (MLP), support vector machine (SVM) and so on. While another technique, which studies continuous time interval of sequential behavior, or sequence learning in affective computing recognizes the continuous human affects appropriately. Long-short term memory (LSTM) networks are efficient sequence learner that suits to recognition tasks.

This study focuses on continuous ads-affect recognition that considers how the strength of changing states during watching ads. We propose subject-independent recognition [Chen 2017] using non-invasive wearable sensors, and a 15-second ad video was selected as stimuli. In this study, we study the subject affects as negative or positive (low or high) emotional affects each scene interval. First, in preliminary experiment, the result shows a relation between the scene-feeling score of questionnaires and the physiological responses. Thus, machine learning techniques including MLP, SVM and LSTM networks were performed to recognize these emotional human affects. The results of continuous affect recognition can support the interpretation of the scene interesting and how subjects interact with ad scenes.

2. Data Acquisition and Preprocessing

In this study, 130 subjects between 20 and 50 years of age including half of males and females participated in the experiments, and all subjects had never watched the ad video before, to assure a genuine reactions. The 15-second ad video of toothpaste was chosen as stimuli to elicit emotional affect from the subjects. This ad video contains 11 scenes which respectively present tooth problems, bacteria illustration, brushing teeth, cleaning up and toothpaste products. The subjects were asked to watch the video while wearing Neurosky's sensors. Then, after watching the video, all subjects were asked to fill out the questionnaires concerning scene feeling based on 7 scales of preference (-4 to 4). We represent the score of 0 to 4 and -1 to -4 as positive and

Contact: Taweesak Emsawas, Department of Information and Physical Sciences, Information Science and Technology, Osaka University, E-mail: taweesak@ai.sanken.osaka-u.ac.jp

negative labels respectively.

2.1 Electroencephalogram (EEG)

The frontal-brain activity plays a crucial role in emotion recognition and distinguishes emotional affects which were known to vary in affective valence such as positive and negative [Schmidt 2001]. EEG data were acquired from one electrode and consequently power spectral density (PSD) was employed to extract features from EEG signal into five frequency bands: delta (0-4 Hz) , theta (4-8 Hz), alpha (8-13 Hz), beta(13-30 Hz), and gamma (>30 Hz) from low to high frequencies [Jenke 2014]. Neurosky's sensor calculated and recorded five frequency bands in each second.

2.2 Electrocardiogram (ECG)

CardioChip/BMD10X-based devices of Neurosky's sensors recorded and converted to ECG signals. ECG signal consists of P-QRS-T waves in each one cardiac cycle. The features were extracted throughout the whole P-QRS-T segment as the standard deviation of the RR intervals (SNDD), the number of pairs of successive RR intervals that differ by more than 50 ms (NN50) and the proportion of NN50 that divides by the total number of NN intervals (PNN50)[Karpagachelvi 2010].

2.3 Eye-tracking

Video-based method of eye tracking is widely used in commercial eye trackers which measures horizontal and vertical components of the movements of both eyes [Chennamma 2013]. The distance of each X and Y pairs was calculated and used as an eye-tracking feature.

To synchronized all physiological features, the lower-frequency feature was duplicated multiple times unit it matches the timestamp with the higher-frequency feature. For example, in the case of all features, EEG and ECG features were duplicate until they match with eye-tracking features.

For the sake of simplicity, the targeted labels of affect recognition were extracted from questionnaires using mapping and rearrangement depend on the scene period. Thus, each second which involves more than 2 scenes would calculate the average of time-weighting by

$$f_t = \sum_{i}^{N_t} (sf_i * t_i) \tag{1}$$

where f_t is feeling score at time t, sf_i is feeling score at scene i, t_i is time weight at scene i, N_t is the number of scene which relates to time t and t is observed timestamp $(1, 2, 3, ..., 15)$

3. Research Methodology

In this study, we used trend value to comprehend the relation of physiological data over time. For affect recognition, the emotional affect was classified as positive and negative. The techniques can be divided into two types: window recognition and sequence learning. The input and output data were constructed from extracted features and questionnaires, which are shown in Figure 1.

3.1 Data Trends

To observe the relationship of physiological data and feeling score, representation of the same axis is required to indicates whether each particular band is increasing or decreasing over time, and how strong each data values. We present the data trend by using trend value or r which represents the trend of data over time. Each trend value calculated by using the 4-second sliding window technique. The positive value is a direct variation of data values over time, and the negative value is an inverse variation. The trend value can be calculated by

$$r = \frac{\sum_i^n (x_i - \bar{x})(t_i - \bar{t})}{\sqrt{\sum_i^n (x_i - \bar{x})^2}\sqrt{\sum_i^n (t_i - \bar{t})^2}} \tag{2}$$

where n is sliding window size, x is data value, t is time in second and i is running number of sliding window.

3.2 Window recognition

Window recognition recognizes emotional affects each second which this study applied the 4-second sliding window for segmentation technique to analyze temporal data and recognize emotional affects over time. The model input was constructed from the trend values in section 3.1. Then, the values were mapped to the time axis for data and label consistency. Meanwhile, the label or output was the feeling score value at each second. Therefore, MLP and SVM which are ubiquitous techniques were chosen in this experiment. However, these techniques are not able to transmit useful information between each window entirely but can learn along the values according to the sliding window size.

3.3 Sequence learning

We also applied sequence learning as LSTM networks which improve from a recurrent neural network (RNN). The main idea of the RNN is to memorize previous information in the network's internal states by using self-connections. The benefit of an RNN is the ability to use contextual information from the whole data sequence, and the dependencies of inputs can remain in the network. The LSTM architecture preserves information by using memory cells and gate units which allows the exhibiting of temporal dynamic behavior for a time series data. In the experiment of sequence learning, the LSTM networks were used to learn the dependencies of physiological responses. After constructing from the trend values in section 3.1, we perform a 1-second sliding window that contains the data values, and each feeling score in a second was used as the label. The sliding window is shifted along the data waves to construct an input vector for each window. After having been trained, the network can be used to predict the label sequence of feeling score.

4. Experiments and Results

4.1 Preliminary Experiment

In our preliminary experiment, we calculated and investigated the trends considering the five frequency bands of only EEG data over time. The subject's averages between EEG data over time are shown in Figure 2, and the average time-weighting of the feeling score is shown in Figure 3, respectively.

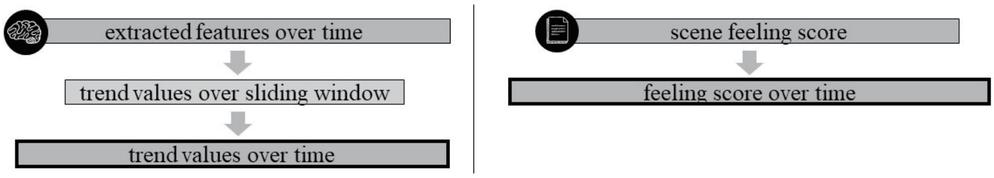

Figure 1: the input representation from extracted feature values (left) and output representation from questionnaires (right)

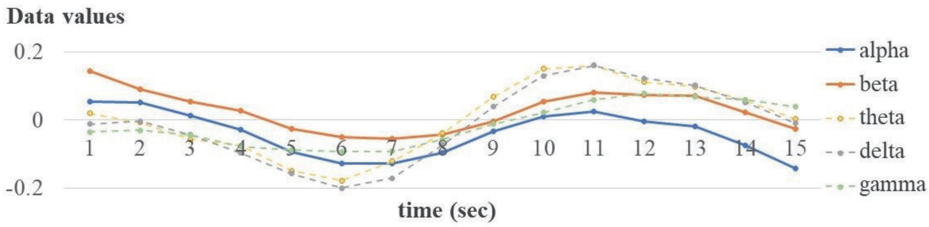

Figure 2: Average data trends of EEG data over time for all subjects

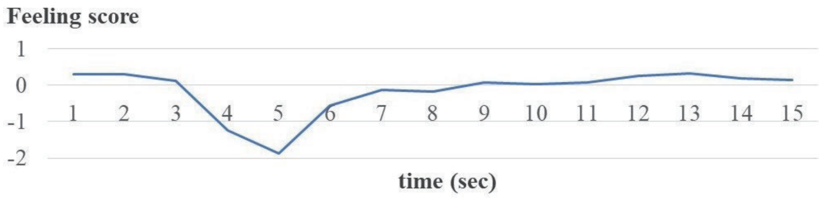

Figure 3: Average feeling score overtime for all subjects

Considering these two figures, we found the evidence that shows the relation between EEG responses and scene feeling score. Fortunately, we can indicate the unusual point around 3 to 7 seconds which is the disgusting scene of bacteria. Hence, we continued to examine the recognition experiment based on physiological responses using machine learning techniques.

4.2 Recognition Experiment

In this section, we divide the experiments, according to the features used, into two experiments: only EEG features and all features. These two experiments applied MLP, SVM and LSTM network to recognize the output label or feeling score each second. 5-fold cross-validation was performed for all subjects. The trend values and feeling score over time were used as the dataset and labels. This dataset was proportionally divided into 5 fold and used for train and test the models. In addition, MLP and LSTM need the validation set in the training process, so the training sets of each fold were divided into the training set and validation set with 80% and 20% respectively. The results are shown in Table 1. In the experiment of only EEG features, LSTM achieved the highest accuracy at 76.4%, and SVM and MLP achieved 70.3% and 68.8% respectively. Unfortunately, all features experiment achieved an accuracy of 72.8%, 69.2%, 61.9% respectively.

	MLP	**SVM**	**LSTM**
Only EEG features	68.8	70.3	76.4
All features	61.9	69.2	72.8

Table 1: Accuracy of affect recognition

5. Discussion and Conclusion

In this study, we present a study of emotional affect recognition based on physiological responses from wearable sensors. the models constructed through this study approach can be used for detecting the emotional affects change over time. Based on the recognition results, the best results are achieved in recognition when only EEG features are used. Considering ECG features and according to previous researches, we found that ECG features could not perform good results in a short duration. While eye tracking features in this experiment could not denote the emotional affects, the subjects just watched the interesting points in each scene. In recognition techniques, we present general recognition and sequence learning techniques. These two techniques could gain a continuous affect along the duration and apply to comprehend the feedback and obtain an advantage from the interesting points in ads. In addition, the sequence learning outperformed the accuracy through learning of data sequence.

References

[Shu 2018] L. Shu, J. Xie, M. Yang, Z. Li, Z. Li, D. Liao, X. Xu, X. Yang; A Review of Emotion Recognition Using Physiological Signals, Sensors, 18(7), 2018.

[Ragot 2017] M. Ragot, N. Martin, S. Em, N. Pallamin, J. Diverrez; Emotion Recognition Using Physiological Signals: Laboratory vs. Wearable Sensors, Advances in Human Factors in Wearable Technologies and Game Design, pp.15-22, 2017.

[Chen 2017] J. Chen, B. Hu, Y. Wang, P. Moore, Y. Dai, L. Feng, Z. Ding; Subject-independent emotion recognition based on physiological signals a three-stage decision method, BMC Medical Informatics and Decision Making, 17(3), 2017.

[Schmidt 2001] L.A. Schmidt and L.J. Trainor; Frontal brain electrical activity EEG distinguishes valence and intensity of musical emotions, Cognition and Emotion, 15(4), pp.487–500, 2001.

[Jenke 2014] R. Jenke, A. Peer, M. Buss; Feature Extraction and Selection for Emotion Recognition from EEG, IEEE Transactions on Affective Computing, 5(3), pp.327-339, 2014.

[Karpagachelvi 2010] S. Karpagachelvi, M. Arthanari, M. Sivakumar; ECG Feature Extraction Techniques - A Survey Approach, 2010.

[Chennamma 2013] H.R. Chennamma, X. Yuan; A Survey on Eye-Gaze Tracking Techniques, Indian Journal of Computer Science and Engineering, 4(5), pp. 388-393, 2013.