Tracking Liking State in Brain Activity while Watching Multiple Movies

Naoto Terasawa
Graduate School of Information Science, NAIST
Ikoma-shi, Nara, Japan
terasawa.naoto.td3@is.naist.jp

Hiroki Tanaka
Graduate School of Information Science, NAIST
Ikoma-shi, Nara, Japan
hiroki-tan@is.naist.jp

Sakriani Sakti
Graduate School of Information Science, NAIST
Ikoma-shi, Nara, Japan
ssakti@is.naist.jp

Satoshi Nakamura
Graduate School of Information Science, NAIST
Ikoma-shi, Nara, Japan
s-nakamura@is.naist.jp

ABSTRACT

Emotion is a valuable information in various applications ranging from human-computer interaction to automated multimedia content delivery. Conventional methods to recognize emotion were based on speech prosody cues, facial expression, and body language. However, this information may not appear when people watch a movie. In recent years, some studies have started to use electroencephalogram (EEG) signals in recognizing emotion. But, the EEG data were entirely analyzed in each scene of movies for emotion classification. Thus, the detailed information of emotional state changes cannot be extracted. In this study, we utilize EEG to track affective state during watching multiple movies. Experiments were done by measuring continuous liking state during watching three types of movies, and then constructing subject dependent emotional state tracking model. We used support vector machine (SVM) as a classifier, and support vector regression (SVR) for regression. As a result, the best classification accuracy was 77.6%, and the best regression model achieved 0.645 of correlation coefficient between actual liking state and predicted liking state. These results demonstrate that continuous emotional state can be predicted by our EEG-based method.

CCS CONCEPTS

• **Computing methodologies** → *Cognitive science*;

KEYWORDS

Electroencephalogram (EEG), support vector regression (SVR), liking

ACM Reference Format:
Naoto Terasawa, Hiroki Tanaka, Sakriani Sakti, and Satoshi Nakamura. 2017. Tracking Liking State in Brain Activity while Watching Multiple Movies. In *Proceedings of 19th ACM International Conference on Multimodal Interaction (ICMI'17).* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3136755.3136772

ICMI'17, November 13–17, 2017, Glasgow, UK

ACM ISBN 978-1-4503-5543-8/17/11…$15.00
https://doi.org/10.1145/3136755.3136772

1 INTRODUCTION

Human emotion is a mental state that arises as pleasant or unpleasant reactions to the experiences. Automatic emotion recognition is one of important technologies that can be applied in various applications ranging from human-computer interaction to automated multimedia content delivery. In personalized content delivery, the system is able to create automatic video summary based on affective characteristics of the users. Many existing studies of emotion recognition have been done based on speech prosody cues and facial expression [1, 2]. However, this information may not appear when people watch movies. Thus, physiological indices such as heart rate, skin conductivity, and respiration were proposed to recognize emotion [3]. In particular, EEG receives attention to recognize emotion because EEG signal has significant relation to emotion [4].

As related works of emotion recognition with EEG data, Ahem and Schwartz found that spectral pattern is different between positive and negative emotions [5]. Furthermore, Khushaba et al. examined relationship between subjects' preference and EEG to understand physiological processes of decision making [4]. They asked subjects to choose a preferred object from some objects in order to predict preference. Nie et al. recognized positive and negative emotions during watching movies based on EEG information [6]. They measured an emotional content of each scene of movie clips using self-assessment manikin (SAM) [7]. They labeled each movie clip as positive or negative based on valence scores of SAM. Then, they classified positive or negative in each scene by using linear support vector machine (linear-SVM). In another study, Dmochowski et al. predicted viewership of TV commercial and movies from EEG data using a linear regression model [8]. They also divided the video into some scenes, and EEG data were analyzed for each scene. There are two publicly available datasets; DEAP [9] and MAHNOB-HCI [10] which were proposed for emotion recognition from EEG. These datasets are for about one minute movies, and changes of temporal emotional state may not be observed.

Despite various existing studies of EEG-based emotion recognition, the EEG data were entirely analyzed for each scene. Thus, the

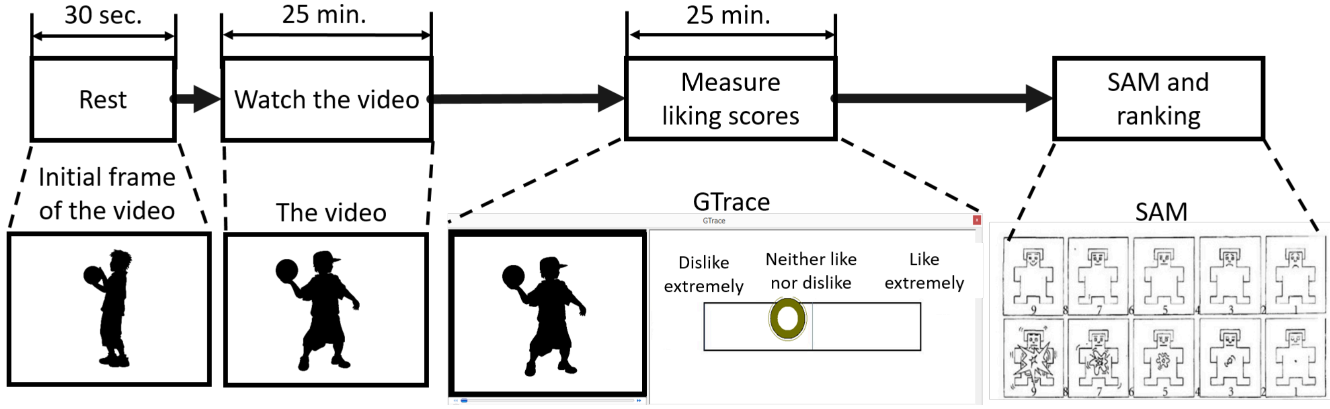

Figure 1: Experimental procedure

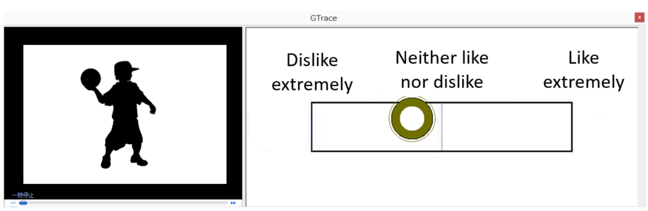

Figure 2: GTrace interface for a subjective evaluation of liking. Users move a cursor.

continuous information of emotional state changes cannot be extracted. For application such personalized movie summarization, continuous emotional state is necessary to measure because the movie contents are also continuous. Recently, there is an existing study that tracks continuous emotional states in theatrical improvisations of pairs of actors based on body language and speech prosody, and multimodal attributes predicted continuous emotion with 0.54 of correlation coefficient [11]. Nevertheless, such tracking emotional state has not been done from the analysis of brain activity.

In this study, our purpose is to track human continuous emotion during watching multiple movies based on EEG signal. To achieve it, first we selected liking as an emotional value, and we continuously measure the liking scores for continuous video sequence. We try to predict like and dislike emotions by using SVM, and we also applied regression models to estimate liking degree by using support vector regression (SVR).

2 METHODS

In this section, we describe our method to track continuous emotional state from EEG signal. The Research Ethic Committee of our institution has reviewed and approved this experiment. Written informed consent was obtained from all participants before the experiment.

2.1 Participants

The experiment was conducted with four participants who were healthy, right-handed, and graduate student of our institution (3 males and 1 female, ages: M=25.25, SD=2.28).

2.2 EEG

A dry EEG headset (HD-72, Cognionics) was used as an EEG recording device. 32 electrodes were located at Fp1, Fpz, Fp2, AF7, AF3, AF4, AF8, F5, F3, Fz, F4, F6, C5, C3, C1, Cz, C2, C4, C6, CP5, CPz, CP6, P7, P3, Pz, P4, P8, O1, Oz, O2 according to the International 10-20 electrode system. Two electrodes were located at both earlobes as reference and ground. A sampling rate of EEG signals was set as 300 Hz.

2.3 Movie Stimuli

To record variations of liking scores, three types of popular Japanese animations, One Piece (Movie 1), Pretty Cure (Movie 2), and Masked Rider Kabuto (Movie 3) which last about 25 minutes, were used in the experiment. The videos were displayed to a 20-inch monitor.

2.4 Data Recording

Figure 1 shows an experimental procedure. The experiment was conducted in a quiet shield room. We set a noiseless environment by directing subjects to turn their phones off. In the experiment, the participants were asked to sit a chair and watch a screen monitor. First, the participants rested for 30 seconds, and they watched a video for about 25 minutes. While the participants rested and watched the video, their EEG signals were recorded. After the participants watched the video, they evaluated the liking scores by using general trace program (GTrace) without EEG recording in order to prevent noise by movement of an arm (see Figure 2). We requested subjects to annotate by remembering their first liking state. And then, the participants evaluated emotional valence and arousal using SAM with nine scale for each scene [6]. Last, the participants sorted scenes of the video according to their liking as ranking. We did not use the ranking data in this paper. The procedure was repeated three times using different movies. We had a break between recordings.

2.5 Annotation

Here, we mention our annotation procedure. The participants continuously evaluated liking scores of the video using GTrace which is a feeling trace tool [12]. The participants can change liking scores by manipulating a cursor from side to side while they watched the video. A rightmost label is set as "like extremely," which has value one. On the other hand, a leftmost label is set as "dislike extremely," which has value zero. In addition, a central label is set as "neither like nor dislike," which has 0.5, and this position is initial state. This liking definition were proposed by Peryam and Pilgrim [13]. The liking scores were recorded at a sampling rate of 500 Hz. Obtained liking scores were down sampled from 500 Hz to 300 Hz by cubic spline interpolation. Finally, this liking value was normalized between 0 and 1. Annotation data possibly includes some mistakes due to user operation, but there were almost no errors as we checked.

2.6 Preprocessing of EEG Signal

All EEG signals were processed offline in the MATLAB software (Math Works, Natick, MA, USA). To decrease effects of noise, EEG signals were high-pass filtered with a cutoff 1 Hz, and notch filtered at 60 Hz. After filtering, noisy channels and samples of EEG signals were excluded. Channels whose average power exceeded four times of the mean channel power were excluded. This process repeated four times in an iterative scheme by referring to [8].

2.7 Feature Extraction and Selection

Power spectral densities (PSDs) of each channel were calculated by short-time Fourier transform (STFT) with three second and 50% overlapping of Hamming window[1]. PSDs were separated into theta (4-7 Hz), alpha (8-12 Hz), beta (12-30 Hz) and gamma (31-50 Hz) bands because PSDs of each frequency band relates to emotional brain activity [14]. In each frequency band and channel, we computed means of PSDs.

To find features having strong correlation with liking scores, Spearman's correlation coefficients between PSDs and liking scores were calculated for each participant and movie. We sorted combinations of channels and frequency bands in training data in descending order of correlation coefficients as candidate features, which were utilized for classification and regression.

2.8 Classification and Regression

For classification, three second movie clips were classified into two kinds of label, "like" or "not-like" label. When liking score was more than mean of entire liking scores in the participant, it was labelled as "like." In contrast, when liking scores was less than mean of liking scores, it was labelled as "not-like." Then, a classification model was trained by SVM with radial basis function (RBF) kernel for each subject and movie. Accuracies were computed by checking whether each estimated label belongs to original label using leave two sample out cross-validation considering overlap of window.

We also trained a regression model by using SVR with RBF kernel for three second movie clips. Training data and test data were prepared on a framework of leave two sample out cross-validation which was similar to the classification model. Pearson's correlation coefficients (r) between predicted liking score and actual liking score were calculated.

[1]We tested different seconds (one and five), and found that it didn't significantly affect later results.

Table 1: Classification accuracies (%). [-] denotes excluded data due to contaminated by noise.

	Movie 1	Movie 2	Movie 3
Subject 1	55.1	63.5	64.5
Subject 2	-	57.1	77.6
Subject 3	70.4	55.3	72.6
Subject 4	69.1	66.7	64.6

2.9 Relation to SAM

We calculated Pearson's correlation coefficients between liking score and emotional valence, and between liking score and emotional arousal, which come from SAM.

3 EXPERIMENTAL RESULTS

In this section, we report our experimental results in classification, regression, and SAM.

3.1 Classification

Classification accuracies of each subject and movie are shown in Table 1. We confirmed that 77.6% was achieved on the subject 2 - movie 3, and second-best classification accuracy of 72.6% was achieved on the subject 3 - movie 3. We found that classification accuracies of all subjects were over a chance rate (approximately 50%) and tested using binomial test ($p < 0.05$; excluding the case of subject 1 - movie 1).

3.2 Regression

Regression results of correlation coefficients is shown in Table 2. The best r of 0.645 was achieved on the subject 2 - movie 3, and second-best r of 0.549 was achieved on the subject 4 - movie 2. We found that correlation coefficients of all subjects were significant by using test for no correlation (all, $p < 0.05$). These results are almost consistent with classification results. Figure 3 shows that comparison between actual values and predicted values by SVR in the case of subject 2 - movie 3. It shows that temporal liking changes at times of around 600 and 1200 seconds were accurately tracked. Here, in the case of subject 2 - movie 3, the top ten features as selected in section 2.7 was as follows: 1) Oz-gamma, 2) Oz-beta, 3) P4-beta, 4) F5-gamma, 5) P3-gamma, 6) Pz-gamma, 7) P7-gamma, 8) CP6-gamma, 9) Pz-beta, 10) P3-beta.

3.3 Relation to SAM

Results of relationships between liking and valence and arousal are shown in Table 3 and 4. We confirmed that correlation to valence

Table 2: Correlation coefficients between true liking values and predicted liking values. [-] denotes excluded data due to contaminated by noise.

	Movie 1	Movie 2	Movie 3
Subject 1	0.221	0.283	0.229
Subject 2	-	0.350	0.645
Subject 3	0.377	0.176	0.541
Subject 4	0.347	0.549	0.382

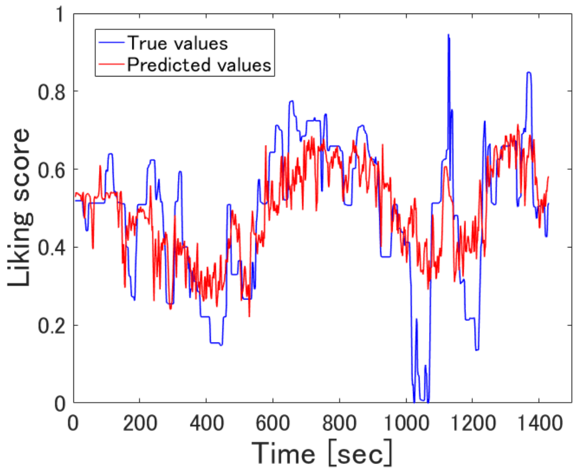

Figure 3: An example of comparison between true values and predicted values of best regression result (subject 2 - movie 3). The blue line shows actual liking score, and the red line shows predicted liking score. The predicted values were smoothed with 3-points moving average.

and arousal were relatively high on all participants (specifically valence). We can say that most of the correlations were exceeded to 0.5, which indicates high correlation between liking and SAM scores by using test for no correlation ($p < 0.05$; significant if absolute values were higher than 0.470).

4 DISCUSSION

For feature selection, as we showed in the case of subject 2 - movie 3, most of selected input features were from near parietal and central regions. In addition, beta and gamma bands were frequently selected. Cole and Ray [15] found that EEG beta activity in both temporal and parietal regions was associated with emotional valence. Furthermore, Lane et al. showed that the brain activity in the temporal region was related to the distinction between positive and negative emotions [16]. Our results also suggested subject and movie dependency. It is shown that classification and regression results are relatively good on the subject 2, the subject 4, and movie 3 because we assume that selected features in these setting were consistent with the previous works and associated with emotion.

Table 3: Correlation coefficients between liking and emotional valence.

	Movie 1	Movie 2	Movie 3
Subject 1	0.757	0.461	0.522
Subject 2	0.636	0.825	0.673
Subject 3	0.678	0.386	0.849
Subject 4	0.671	-0.244	0.727

Table 4: Correlation coefficients between liking and emotional arousal.

	Movie 1	Movie 2	Movie 3
Subject 1	0.549	-0.236	0.372
Subject 2	0.225	0.308	-0.791
Subject 3	0.569	0.359	0.510
Subject 4	0.513	-0.346	0.471

We examined relationships between liking and SAM scores, and confirmed that our liking score was correlated with traditional SAM measurement. North and Hargreaves found that liking was related to emotional states in musical stimuli [17]. We can say that even continuous emotional evaluation, liking score can be measured by GTrace.

5 CONCLUSION

In this study, we proposed a short time EEG-based liking prediction, which can be utilized many situations in order to track emotional state. To evaluate performance of the proposed method, we continuously classified liking and not-liking by a machine learning algorithm. Also, we continuously predicted liking degree by SVR. As a result, a best classification accuracy was 77.6%. For a regression model, we found that temporal changes of liking can be predicted for several subjects and movies.

For application of personalized movie summarization, this work is our first step and preliminary study. For our future directions, first we try to increase the number of participants and variations (genre) of video used in the experiment. Then, we attempt to normalize the EEG data across subject and movie, and use inter-subject-movie correlation features as referring to [8, 18] in order to obtain better subject independent classification accuracy.

ACKNOWLEDGMENTS

Part of this work was supported by a joint research project with Fujitsu Laboratories, as well as JSPS KAKENHI Grant Numbers JP17H06101, JP17K00237, and JP16K16172.

REFERENCES

[1] K. Anderson, P. W. McOwan, "A real-time automated system for the recognition of human facial expression," IEEE transactions on systems, man, and cybernetics, part B: Cybernetics, vol. 36, no. 1, Feb 2006, pp. 96-105.

[2] J. Ang, R. Dhillon, A. Krupski, E. Shriberg, and A. Stolcke, "Prosody-based automatic detection of annoyance and frustration in human-computer dialog," International conference on spoken language processing, 2002, pp. 2037-2039.

[3] R. W. Picard, E. Vyzas, and J. Healey, "Toward machine emotional intelligence: analysis of affective physiological state," IEEE transactions on pattern analysis

and machine intelligence, vol. 23, no. 10, Oct 2001, pp. 1175-1191.
[4] R. N. Khushaba, L. Greenacre, S. Kodagoda, J. Louviere, S. Burke, and G. Dissanayake, "Choice modeling and the brain: A study on the Electroencephalogram (EEG) of preference," Expert systems with applications, vol. 39, no. 16, Nov 2012, pp. 12378-12388.
[5] G. L. Ahem, G. E. Schwartz, "Differential lateralization for positive and negative emotion in the human brain: EEG spectral analysis," Neuropsychologia, vol. 23, no. 6, 1985, pp. 745-555.
[6] D. Nie, X. W. Wang, L. C. Shi, and B. L. Lu, "EEG-based emotion recognition during watching movies," Proceedings of the 5th international IEEE EMBS conference on neural engineering, May 2011, pp. 667-670.
[7] M. M. Bradley and P. J. Lang, "Measuring emotion: The self-assessment manikin and the semantic differenctial," Journal of behavior therapy and experimental psychiatry, vol. 25, no. 1, Mar 1994, pp. 49-59.
[8] J. P. Dmochowski, M. A. Bezdek, B. P. Abelson, J. S. Johnson, E. H. Schumacher, and L. C. Parra, "Audience preferences are predicted by temporal reliability of neural processing," Nature communications, vol. 5, no. 4567, July 2014.
[9] S. Koelstra, C. Muhl, M. Soleymani, J. S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, L. Patras, "DEAP: A database for emotion analysis using physiological signals," IEEE transactions on affective computing, vol. 3, no. 1, 2012, pp. 18-31.
[10] M. Soleymani, J. Lichtenauer, T. Pun, M. Pantic, "A multimodal database for affect recognition and lmplicit tagging," IEEE transactions on affective computing, vol. 3, no. 1, 2012, pp. 42-55.
[11] M. Angeliki, K. Athanassios, W. Yun and N. Shrikanth, "Tracking changes in continuous emotion states using body language and prosodic cues," IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2011, pp. 2288—2291.
[12] P. Petta, C. Pelachaud, and R. Cowie, "Emotion-oriented systems: The humaine handbook," 2011, pp. 215-244.
[13] D. R. Peryam and F. J. Pilgrim, "Hedonic scale method of measuring food preferences," Food technology, vol. 11, Sept 1957, pp. 9-14.
[14] W. Klimesch, "EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis," Brain research reviews, vol. 29, no. 2-3, Apr 1999, pp. 169-195.
[15] W. J. Ray and H. W. Cole, "EEG alpha activity reflects attentional demands, and beta activity reflects emotional and cognitive processes," Science, vol. 228, May 1985, pp. 750-752.
[16] R. D. Lane, E. M. Reiman, G. L. Ahern, G. E. Schwartz, and R. J. Davidson, "Neuroanatomical correlates of happiness, sadness, and disgust," The American journal of psychiatry, vol. 154, no. 7, Jul 1997, pp. 926-933.
[17] A.C. North and D. J. Hargreaves, "Liking arousal potential, and the emotions expressed by music," Scandinavian journal of psychology, vol. 38, no. 1, Mar 1997, pp. 45-53.
[18] H. Kang, Y. Nam, and S. Choi, "Composite common spatial pattern for subject-to-subject transfer," IEEE signal processing letters, vol. 16, no. 8, Aug 2009, pp. 683-686.