Tagging Continuous Labels for EEG-based Emotion Classification

Rong-Fei Gu, Li-Ming Zhao, Wei-Long Zheng and Bao-Liang Lu* *Fellow, IEEE*

***Abstract*— EEG-based emotion classification has long been a critical task in the field of affective brain-computer interface (aBCI). The majority of leading researches construct supervised learning models based on labeled datasets. Several datasets have been released, including different kinds of emotions while utilizing various forms of stimulus materials. However, they adopt discrete labeling methods, in which the EEG data collected during the same stimulus material are given a same label. These methods neglect the fact that emotion changes continuously, and mislabeled data possibly exist. The imprecision of discrete labels may hinder the progress of emotion classification in concerned works. Therefore, we develop an efficient system in this paper to support continuous labeling by giving each sample a unique label, and construct a continuously labeled EEG emotion dataset. Using our dataset with continuous labels, we demonstrate the superiority of continuous labeling in emotion classification through experiments on several classification models. We further utilize the continuous labels to identify the EEG features under induced and non-induced emotions in both our dataset and a public dataset. Our experimental results reveal the learnability and generality of the relation between the EEG features and their continuous labels.**

I. Introduction

Emotion acts as an indispensable part of human life, supporting our social interaction and personality formation. Therefore, the field of affective brain-computer interface (aBCI) [1] springs up with technology development, aiming at utilizing emotion for the sake of mankind. The first step towards getting emotion under control is to identify emotion, and a typical task is emotion classification.

Supervised learning is the most common solution to emotion classification. Researchers collect physiological measures of human brain, annotate these measures by the class of emotion, and learn the relationship between them with various models. Several methods for collecting the physiological measures have been adopted, and electroencephalogram (EEG) stands out among these methods for its objectivity and precision [2]. Some EEG datasets for emotion classification have been publicly released, employing various forms of stimulus materials. Among all the released EEG datasets, DREAMER [3], MAHNOB-HCI [4], DEAP [5], and SEED [6] are the most widely used and they all selected videos as the main form of stimulus material. In these datasets, the EEG data are labeled discretely. Data collected within the same video for about 5 minutes long are given the same label. The label is either the score of the material given by the annotators beforehand or the subjects' remarks afterward. Through both methods, the same labels are given to the great number of data pieces. This kind of discrete labeling violates the fact that emotion differs continuously. Thus the datasets are possible to be contaminated by mislabeled data and may bring down the classification performance.

A few works create and study on self-built continuously labeled EEG datasets where each sample is allocated a tailored label. Ding *et al.* created and revealed the significance of regression models on an EEG dataset with continuous valence and arousal labels obtained during the second and third times when the subjects watch the materials [7]. Soleymani *et al.* recorded the subjects' facial expressions during the trials and annotators afterward labeled these expressions which contribute to the ground truth labels [8]. Both works directly take advantage of the video clips used in forming the MAHNOB-HCI dataset which have been proven effective. Another popular stimulus material is music. Thammasan *et al.* investigated two feature extraction methods based on a dataset collected during the subjects listening to music clips and labeled when the subjects listen to the same music clips again [9], while Hasanzadeh *et al.* emphasized on different models, and the labels in their dataset are recorded along with the EEG signal [10]. These works have not state the superiority of continuous labeling over discrete labeling, and can hardly be applied into mainstream works.

So in this paper, we made the following contributions to prove the feasibility and necessity of continuous labeling, and to explore the general relationship between EEG data and the continuous labels: (1) design an efficient system for continuous labeling and formulate a continuously labeled EEG dataset (2) conduct contrast experiments using discrete and continuous labels (3) identify induced/non-induced emotion based on the continuously labeled EEG data and (4) apply the induced/non-induced classifier on the SEED dataset.

II. Experimental Setup

In this section, we first explain the process and the advantage of our paradigm for continuous labeling. Then we introduce the formulation of the continuously labeled EEG dataset based on the paradigm. Finally we design three tasks

This work was supported in part by grants from National Natural Science Foundation of China (Grant No. 61976135), STI 2030-Major Projects+2022ZD0208500, Shanghai Municipal Science and Technology Major Project (Grant No. 2021SHZDZX), Shanghai Pujiang Program (Grant No. 22PJ1408600), SJTU Global Strategic Partnership Fund (2021 SJTU-HKUST) and GuangCi Professorship Program of RuiJin Hospital Shanghai Jiao Tong University School of Medicine.

R. F. Gu, L. M. Zhao, W. L. Zheng and B. L. Lu are with the Center for Brain-Like Computing and Machine Intelligence, Department of Computer Science and Engineering, the Key Laboratory of Shanghai Education Commission for Intelligent Interaction and Cognitive Engineering, and Brain Science and Technology Research Center, Shanghai Jiao Tong University, 800 Dongchuan Rd., Shanghai 200240, People's Republic of China.

B. L. Lu is with the RuiJin-Mihoyo Laboratory, Clinical Neuroscience Center, RuiJin Hospital, Shanghai Jiao Tong University School of Medicine, 197 Ruijin 2nd Rd., Shanghai 200020, People's Republic of China.

*Corresponding author (bllu@sjtu.edu.cn)

to prove the advantage and necessity of continuous labeling, the learnability of the relationship between EEG data and the continuous labels, and the generality of the relationship.

A. Paradigm Design

We chose video clips as stimulus materials because of their superiority [6] [11]. We directly select the materials from those applied in forming SEED [6], which is composed of 5 video clips from each of positive, negative and neutral emotion. According to the subjects of SEED, their emotion is stable during neutral clips. So to better illustrate the changing emotion, we only utilize the positive and negative clips. Base on the selected 10 clips, we designed the following data collection paradigm, which has been approved by the local ethics committee. The paradigm is composed of the EEG collection phase and the continuous labeling phase. In the first phase, the video clips are presented to the subjects on the computer screen and the subjects wore an electrode cap that records their EEG signals throughout this phase. Each clip lasts for about 4 minutes and there is a 1-minute break between every two clips to help subjects calm down from the previous one. In the labeling phase, the subjects are asked to recall and continuously label their emotion during the first phase according to the degree of induction.

The advantages of our paradigm are listed as follows. First, the subjects label by themselves, retaining the subjectivity of EEG data. Second, the two-phase mode is preferred since labeling while collecting is sure to affect emotion induction and a third time of repetition may reduce the subjects' enthusiasm and lower the precision of labeling. Also, considering our goal of classification, we make the subjects focus on labeling the degree of induction. Later we classify the EEG signals into positive and negative according to the video clips since the class of emotion is stable during the video clips with a stable emotional keynote.

To facilitate the labeling phase, we developed the interactive interface exhibited in Figure 1. The video clips are shown in the upper left and listed on the right. To make the labeling direct and convenient, we enabled a mouse-wheel-controlled container icon in the bottom right. The history labels are visible below the video. Also, rotary knobs are provided on the bottom of the panel to control the volume and the playback speed. When all the clips are labeled, the system automatically processed the labeled curve and stored the labels aligned with the corresponding EEG signals collected in the first phase.

B. Dataset Formulation

We recruited 4 male and 4 female subjects whose ages range from 23 to 27. During the experiment, their EEG data were collected with ESI NeuroScan System at 1000 Hz from an electrode cap with 62 channels. After the raw EEG data and labels are collected, the EEG data go through a downsampling step from 1000 Hz to 200 Hz and a bandpass filtering of 0.5 Hz to 70 Hz to remove noise.

We extract the differential entropy (DE) [12] feature as it is better than the other EEG features in many tasks [6] [13]. The window size is 15 seconds and the stride is 1 second. After the features have been extracted, we smoothed the continuous labels and limited the value to [0, 5]. The smoothed labels of the 8 subjects on one positive clip and one negative clip can be seen in Figure 2. Individual differences are evident in labeling the emotion during the same clip, which verifies the inaccuracy of discrete labeling. We finally sampled from these labels in accordance with the DE features to form the continuously labeled EEG dataset.

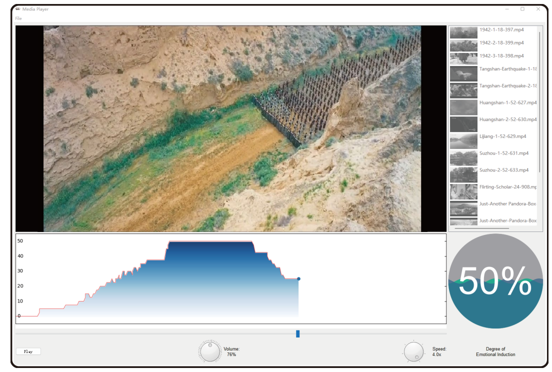

Fig. 1. The operation panel of the software contains a video exhibition window, a video list, a history label indicator, an annotation-assisting container, and two knobs for speed and volume control.

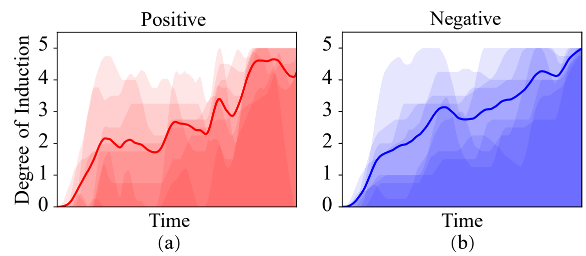

Fig. 2. The continuous labels annotated by the subjects during one positive (a) and one negative video clip (b). Each subject's labels are plotted separately in red and blue for positive and negative emotions with transparency, and the mean values are represented by solid lines.

C. Tasks

1) Comparison of Labeling Methods for Emotion Classification: From Figure 2 we have already observed intuitively the variability of emotion during the same video clips and the advantage of continuous labeling in precision. So in the first part of experiment, we further illustrate this idea by comparing the performance of continuous and discrete labeling in emotion classification. To make the result convincing, we train classifiers base on multiple models. The classical models we applied are SVM, MLP, CNN, LSTM, and CNN-LSTM. We also employ specific models that apply to EEG-based tasks including EEGNet [14] and GCN [15].

2) Induced/non-induced Classification: We put forward this task to investigate whether the differences between samples can be identified. We divide the EEG samples into induced and non-induced ones according to their continuous labels, and train classifiers base on EEGNet and GCN. By evaluating the predicted results, we explore the feasibility of mapping the EEG data to the continuous labels.

3) Cross-dataset Induced/non-induced Classification: The above task is conducted specifically on our dataset, and this task aims at classifying induced and non-induced samples based on EEG data from SEED [6], to show the universality of the relationship between EEG data and the continuous labels. SEED is a classic and popular EEG dataset for emotion classification. It is made up of 15 subjects who take the video-watching experiment. Since our stimulus materials are select from those videos used in forming SEED, we compare the continuous labels of our dataset and the classification result of the same video on SEED.

TABLE I

RESULT OF THE EMOTION CLASSIFICATION TASK, AVERAGED AMONG THE 8 SUBJECTS. THE BETTER ACCURACY AND STANDARD DEVIATION OF EACH METHOD ARE SHOWN IN BOLD FONT.

	Discrete		**Continuous**	
Method	Avg.	Std.	Avg.	Std.
SVM	78.39	16.28	**81.42**	**11.90**
MLP	72.82	13.66	**76.59**	**13.09**
CNN	81.75	14.58	**82.71**	**12.61**
LSTM	75.21	14.55	**76.07**	**13.17**
CNN-LSTM	87.57	12.33	**87.68**	**8.88**
EEGNet	85.39	13.04	**88.28**	**9.36**
GCN	91.83	8.83	**93.29**	**7.91**

III. EXPERIMENTAL RESULTS

A. Labeling Methods Comparison in Emotion Classification

We compare the emotion classification effects of positive and negative under continuous and discrete labeling of emotion. For continuous scenes, the EEG samples with emotion induction label no less than 2.5 were screened. For discrete scenes, all EEG samples are retained. For each subject, the first 3 positive and 3 negative clips form the train set, and the last 2 positive and 2 negative clips form the test set. To ensure the fairness of comparison, we uniformly use test sets in continuous scenarios. It is worth noting that the data volume of the discrete train sets are twice as large as that of the continuous ones.

The accuracy of each classifier is listed in Table I. According to the results, GCN outperforms the other models on both datasets, and more importantly, though the samples in the train sets are much fewer for the filtered datasets, the accuracy is improved apparently for all the classifiers and the stability over different subjects is better. These conclusions prove that omitting data low in induction degree contributes to the performance of the classifiers, which is a significant advantage of our continuously labeled dataset.

B. Induced/non-induced Classification

To explore the learnability of continuous labels, we train classifier to identify induced and non-induced samples first.

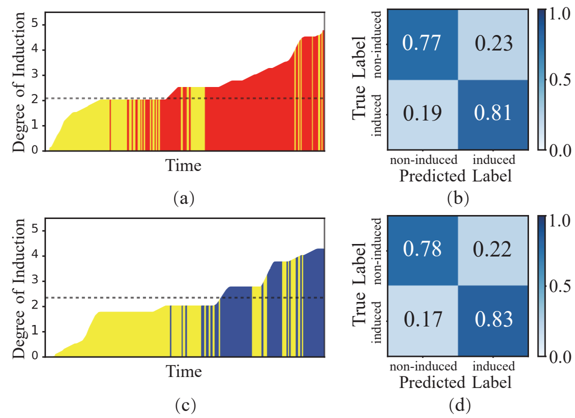

Fig. 3. (a)(c): The induced emotion identification result of one subject on a positive clip and a negative clip. The dashed lines are the thresholds. The upper bounds of the colored area are the continuous labels. The predicted non-induced features are colored in yellow and the induced ones are in red or blue separately for positive and negative. (b)(d): The confusion matrices averaged among all the subjects for positive emotion and negative emotion.

Features with induction value larger than the threshold are labeled as induced, and the rest are non-induced. For both positive and negative samples of each subject, we define the thresholds to be the medians of the continuous labels to ensure sample balance. The train sets and the test sets are formed by samples from the first four clips and the last clip.

Table II shows the accuracy of the classifiers, and we also plot the overall confusion matrices of positive and negative samples along with the comparison between the ground truth labels and the predicted labels of one subject in Figure 3. All the results reveal that certain differences do exist among the data collected during the same stimulus materials, and the learnable differences may contribute to better identification and understanding of emotion.

TABLE II

RESULT OF THE INDUCED EMOTION IDENTIFICATION TASK, AVERAGED AMONG THE 8 SUBJECTS. THE ACCURACY IS MORE THAN 78% FOR BOTH POSITIVE AND NEGATIVE EMOTION USING EEGNET AND GCN.

	Positive		**Negative**	
Method	Avg.	Std.	Avg.	Std.
EEGNet	78.61	9.69	79.81	9.40
GCN	79.05	8.84	81.22	9.19

C. Cross-dataset Induced/non-induced Classification

We train a cross-dataset GCN model for positive and negative emotion each on our dataset. The models classifying induced and non-induced samples are then applied on SEED. We define inductive periods to be the time slices of the videos when more than 50% (>4) of the subjects consider their emotion as induced in our dataset, as shown in Figure 4 (a)(c). We similarly define predicted inductive periods to be the time slices when more than 50% ($>$ 7.5) of the subjects have their emotion classified as induced in

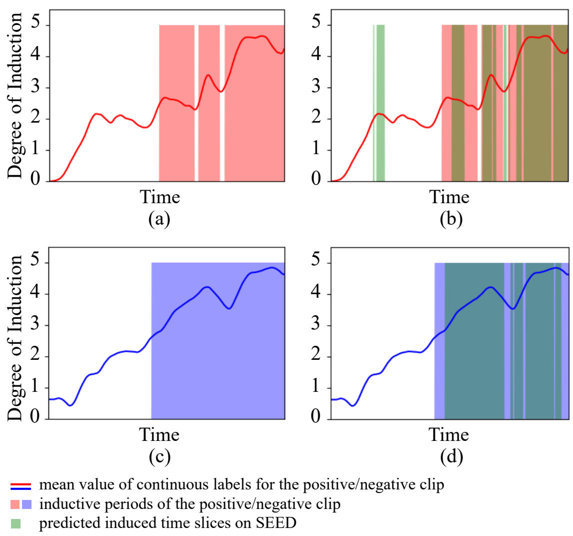

Fig. 4. (a)(c): Time slices of the positive/negative test video. The solid line is the mean value of the continuous labels from our dataset. Inductive periods are in light red/blue. (b)(d): The identification result of a specific subject from SEED on the positive/negative test video. Light green marks the time slices when the positive induced/non-induced classifier consider the subject's emotion as induced. The result agrees with not only the inductive periods but also the mean value.

SEED. The overlap ratio between the inductive periods and predicted inductive periods reflects the effectiveness of identifying induced and non-induced samples on SEED.

For either positive and negative emotion, the first four clips form the train set, and we label the samples by 'induced' and 'non-induced' with thresholds set as 2.5. The trained classifier is validated on the last clip, and then tested on the DE features extracted accordingly from the data collected during the last clip in SEED. The classification result of one subject in SEED is exhibited in Figure 4 (b)(d). Since the last clip is excluded during training, interference from the same material is avoided.

We denote the overlap ratio as O=(O_{in}, O_{non}), where O_{in}=$\frac{P\cap I}{I}$, and O_{non}=$\frac{(U-P)\cap(U-I)}{U-I}$ are separately the overlap ratio of inductive and non-inductive periods. P is the predicted inductive periods, I is the inductive periods, and U is the whole video clip. For the positive test video, O_{pos}=$(93.36\%, 92.87\%)$, while for the negative test video, O_{neg}=$(92.55\%, 54.50\%)$. The induced/non-induced classifiers are effective on SEED, while inter-individual differences are more evident for negative emotion. In general, the results verifies the universality of the relationship between EEG data and the continuous labels, and put forward the necessity of paying more attention to sample-wise differences in leading works.

IV. Conclusion

EEG-based emotion classification has been a stirring task in the field of aBCI. However, the mainstream work of emotion classification adopts discretely labeled datasets, which are imprecise and may decrease the performance of classifiers. In this work, we have proposed a continuous labeling paradigm and have developed a continuously labeled EEG dataset. Our experimental results indicate a general improvement in classification accuracies of several models using continuous labels. We have investigated the identification of EEG features under induced and non-induced emotions. The induced/non-induced classifier is demonstrated to be effective on the SEED dataset, which reveals the learnability and generality of the relation between EEG features and the continuous labels. Although we have only utilized limited characteristics of the continuously labeled dataset, its effectiveness and feasibility in enhancing emotion classification performance provide us a promising approach to aBCI.

References

[1] C. Mühl, B. Allison, A. Nijholt, and G. Chanel, "A Survey of Affective Brain Computer Interfaces: Principles, State-of-the-art, and Challenges," *Brain-Computer Interfaces*, vol. 1, no. 2, pp. 66–84, 2014.

[2] S. M. Alarcão and M. J. Fonseca, "Emotions Recognition Using EEG Signals: A Survey," *IEEE Transactions on Affective Computing*, vol. 10, no. 3, pp. 374–393, 2017.

[3] S. Katsigiannis and N. Ramzan, "DREAMER: A Database for Emotion Recognition Through EEG and ECG Signals from Wireless Low-cost Off-the-shelf Devices." *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 1, pp. 98–107, 2017.

[4] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A Multimodal Database for Affect Recognition and Implicit Tagging," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 42–55, 2011.

[5] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "DEAP: A Database for Emotion Analysis; Using Physiological Signals," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 18–31, 2011.

[6] W.-L. Zheng, J.-Y. Zhu, and B.-L. Lu, "Identifying Stable Patterns over Time for Emotion Recognition from EEG," *IEEE Transactions on Affective Computing*, vol. 10, no. 3, pp. 417–429, 2017.

[7] Y. Ding, X. Hu, Z. Xia, Y.-J. Liu, and D. Zhang, "Inter-Brain EEG Feature Extraction and Analysis for Continuous Implicit Emotion Tagging During Video Watching," *IEEE Transactions on Affective Computing*, vol. 12, no. 1, pp. 92–102, 2018.

[8] M. Soleymani, S. Asghari-Esfeden, Y. Fu, and M. Pantic, "Analysis of EEG Signals and Facial Expressions for Continuous Emotion Detection," *IEEE Transactions on Affective Computing*, vol. 7, no. 1, pp. 17–28, 2015.

[9] N. Thammasan, K. Moriyama, K. ichi Fukui, and M. Numao, "Continuous Music-emotion Recognition Based on Electroencephalogram," *IEICE Transactions on Information and Systems*, vol. 99, no. 4, pp. 1234–1241, 2016.

[10] F. Hasanzadeh, M. Annabestani, and S. Moghimi, "Continuous Emotion Recognition During Music Listening Using EEG Signals: A Fuzzy Parallel Cascades Model," *Applied Soft Computing*, vol. 101, pp. 107 028–, 2021.

[11] W.-L. Zheng, W. Liu, Y. Lu, B.-L. Lu, and A. Cichocki, "Emotionmeter: A Multimodal Framework for Recognizing Human Emotions." *IEEE Transactions on Cybernetics*, vol. 49, pp. 1110–1122, 2019.

[12] R.-N. Duan, J.-Y. Zhu, and B.-L. Lu, "Differential Entropy Feature for EEG-based Emotion Classification," in *2013 6th International IEEE/EMBS Conference on Neural Engineering*. IEEE, 2013, pp. 81–84.

[13] L.-C. Shi, Y.-Y. Jiao, and B.-L. Lu, "Differential Entropy Feature for EEG-based Vigilance Estimation," in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2013, pp. 6627–6630.

[14] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: a Compact Convolutional Neural Network for EEG-based Brain-Computer Interfaces," *Journal of Neural Engineering*, vol. 15, no. 5, p. 056013, 2018.

[15] R. Li, Y. Wang, and B.-L. Lu, "A Multi-Domain Adaptive Graph Convolutional Network for EEG-based Emotion Recognition," in *Proceedings of the 29th ACM International Conference on Multimedia*, 2021, pp. 5565–5573.