2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) | 978-1-7281-1179-7/21/$31.00 ©2021 IEEE | DOI: 10.1109/EMBC46164.2021.9630314

A Novel Experiment Setting for Cross-subject Emotion Recognition

Hao-Yi Hu, Li-Ming Zhao, Yu-Zhong Liu, Hua-Liang Li and Bao-Liang Lu*, *Fellow, IEEE*

Abstract— **Recently, cross-subject emotion recognition attracts widespread attention. The current emotional experiments mainly use video clips of different emotions as stimulus materials, but the videos watched by different subjects are the same, which may introduce the same noise pattern in the collected data. However, the traditional experiment settings for cross-subject emotion recognition models couldn't eliminate the impact of same video clips on recognition results, which may lead to a bias on classification. In this paper, we propose a novel experiment setting for cross-subject emotion recognition. We evaluate different experiment settings on four public emotion datasets, DEAP, SEED, SEED-IV and SEED-V. The experimental results demonstrate the deficiencies of the traditional experiment settings and the advantages of our proposed experiment setting.**

I. INTRODUCTION

Emotion plays an important role in human life and interpersonal interaction [1]. Although people have defined all kinds of emotions, it is difficult to directly quantify and measure the specific emotional state. There are two widely accepted models of emotion. The first emotion model divides emotions into six basic categories (anger, disgust, fear, happiness, sadness, and surprise) according to Ekman's theory [2]. The other emotion model called pleasure-arousal-dominance (PAD) model describes emotions using their underlying dimensions [3], which measures emotions from displeasure to pleasure, nonarousal to arousal, and submissiveness to dominance, respectively.

Emotion recognition refers to the process in which machines identify human emotions through various signals, such as facial expressions [4], voice [5], body postures [6], and physiological signals [7]. Recently, emotion recognition based on physiological signals, represented by electroencephalography (EEG), has attracted extensive attention for its information sufficiency and stable neural patterns [8]. In addition to EEG signals, eye movement signals are also widely used due to their easy acquisition.

Emotion recognition models can be broadly categorized into subject-dependent models [9] [10] and subject-independent (cross-subject) models [11]. Subject-dependent means that both the training and test data come from the same subject. In other words, an individual model needs to be trained for each subject. However, emotion recognition models trained on one subject tend to perform poorly on others because the EEG data varies significantly across different subjects [12] [13]. Therefore, it is necessary to develop subject-independent models by using transfer learning which are applicable to the real application scenario.

At present, the mainstream way to induce emotion is to use video clips [14] [15]. However, different subjects watching the same video may introduce the same noise pattern, which has no effect on the subject-dependent setting but would affect the cross-subject setting. The general experiment setting applied in the subject-independent models is leave-one-subject-out cross validation, i.e, take the data from one subject as the target domain and the rest data from other subjects as the source domain for each fold. So it is hard to say what the model categorizes are different emotions or different materials under the influence of the same stimulus materials.

In this paper, we reveal that the stimulus materials used to induce emotions have influence on the classification of emotion recognition under the general experiment setting applied in the subject-independent models and propose a new cross-subject experiment setting which can eliminate the impact of materials. We evaluate the experiment setting on four public emotion datasets and the experimental results demonstrate the advantages of the proposed experiment setting.

II. METHOD

A. Data Partition

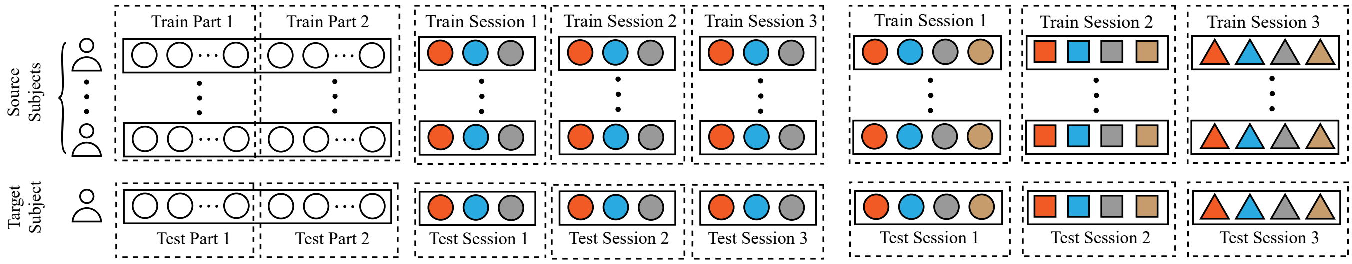
In subject-dependent emotion recognition, an individual model will be trained for each subject and each subject's data is divided into two parts: training set and test set, e.g., the model will be trained on first 9 trials and tested on the rest 6 trials in SEED dataset [16]. It is clear that the data in the training set and the test set are based on different materials. However, for cross-subject emotion recognition, the general experiment setting is leave-one-subject-out cross validation. In other word, the model will be trained on 14 subjects' data and evaluated on the remaining subject. In this experiment setting, the materials used to induce the data in the training set and the test set are exactly the same. Therefore, it is difficult to tell whether our trained model ultimately classifies the type of materials or the emotion that the materials evokes.

To avoid the problem mentioned above, we have proposed a novel experiment setting for cross-subject emotion recognition. In our method, we further divide the data on

[1]H. Hu, L. M. Zhao 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, the Brain Science and Technology Research Center, and Qing Yuan Research Institute, Shanghai Jiao Tong University, 800 Dong Chuan Rd., Shanghai 200240, China; the Center for Brain-Machine Interface and Neuromodulation and RuiJin Mihoyo Laboratory, RuiJin Hospital, Shanghai Jiao Tong University School of Medicine, 197 Ruijin 2nd Rd., Shanghai 200020, China.

[2]Y. Liu and H. Liang are with Key Laboratory of Occupational Health and Safety of Guangdong, Power Grid Co., Ltd, Electric Power Research Institute of Guangdong Power Grid Co., Ltd., Guangzhou, Guangdong, China.

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



(a) DEAP (b) SEED (c) SEED-IV

Fig. 1. The data partition of three datasets. Different shapes represent different stimulus material group, and different colors represent different emotions. DEAP is indicated in white because there is no clear emotion category. The partition of SEED-V is similar to SEED-IV. This figure shows only one fold in cross-validation.

the basis of leave-one-subject-out cross validation. We first divide the data into two or three parts based on the stimulus materials. Then we select the data of n-1 (n is the # subject) subjects from one part as the training set, and then select the remaining subject's data from another part as the test set. In the Experiment Settings section, we give detailed examples of data partitioning.

B. Classifier

Two different classifiers are used in our paper, Support Vector Machine (SVM) and Kernel principal component analysis (KPCA) [17].

SVM with a linear kernel acts as a benchmark model for cross-subject emotion recognition task. The reason why we choose SVM as the basic model in this paper is that we mainly focus on the impact of different experimental settings on classification results rather than the accuracy.

KPCA is the kernel version of principal component analysis and it can realize dimensionality reduction of linearly inseparable data. The general idea of KPCA is: for the data in the D-dimensional feature space, a nonlinear mapping $\phi(x)$ is used to map all samples to a M-dimensional feature space ($M \gg D$) to make it linearly separable. Because the computational complexity of using mapping $\phi(x)$ directly is very high, the kernel function is used instead. After that, the PCA dimension reduction is carried out in this high-dimensional space. For KPCA-based subject transfer, we concatenate the training data and test data and apply KPCA on it first. After that, we repartition the data into the training data and test data. Finally, the processed data is used as input to evaluate the classification model (SVM in this paper).

III. EXPERIMENTS AND RESULTS

A. Datasets

We evaluate our experiment setting on four public emoti datasets, DEAP [14], SEED [15] [16], SEED-IV [18] and SEED-V [19] [20]. Table I shows the comparison of these four datasets, where # subject refers to the number of subjects, # session refers to the number of times each subject took part in the experiment, # trial refers to the number of video clips in each session, category refers to whether the video clips used in different sessions are the same.

TABLE I
THE COMPARISON OF DEAP, SEED, SEED-IV AND SEED-V.

Dataset	# subject	# emotion	# session	# trial	category
DEAP	32	-	1	40	-
SEED	15	3	3	15	same
SEED-IV	15	4	3	24	different
SEED-V	20	5	3	15	different

The DEAP dataset is a multimodal dataset for analyzing the human affective states with data from 32 subjects. During the experiments, subjects watched 40 one-minute long excerpts of music videos and rated each video in terms of the levels of arousal, valence, like/dislike, dominance and familiarity. Note that each subject took part in the experiment only once, and different subjects watched the same music videos. In this paper, we label the data on the valence dimension (high valence: level > 5, low valence: level ≤ 5) and arousal dimension (high arousal: level > 5, low arousal: level ≤ 5).

The SEED dataset comprises EEG and eye movement data of 9 subjects (15 subjects for only EEG data) and contains three emotions: happy, sad and neutral. During the experiments, subjects watched fifteen rigorously screened Chinese movie clips. Each subject took part in the experiment three times (sessions) but watched the same movie clips in each session.

The SEED-IV dataset comprises EEG and eye movement data of 15 subjects and contains four emotions: happy, sad, neutral and fear. The SEED-V dataset comprises EEG and eye movement data of 20 subjects and contains five emotions: happy, sad, fear, disgust, and neutral. For each subject, three sessions are performed on different days, and each session contains 15 (24 for SEED-IV) movie clips. Each subject took part in the experiment three times and watched different movie clips in each session.

It should be noted that the subjects watched the same movie clips in three sessions in the SEED dataset, which are different in SEED-IV and SEED-V.

B. Feature Extraction

1) EEG feature: Datasets SEED, SEED-IV and SEED-V provide the differential entropy (DE) features of EEG by

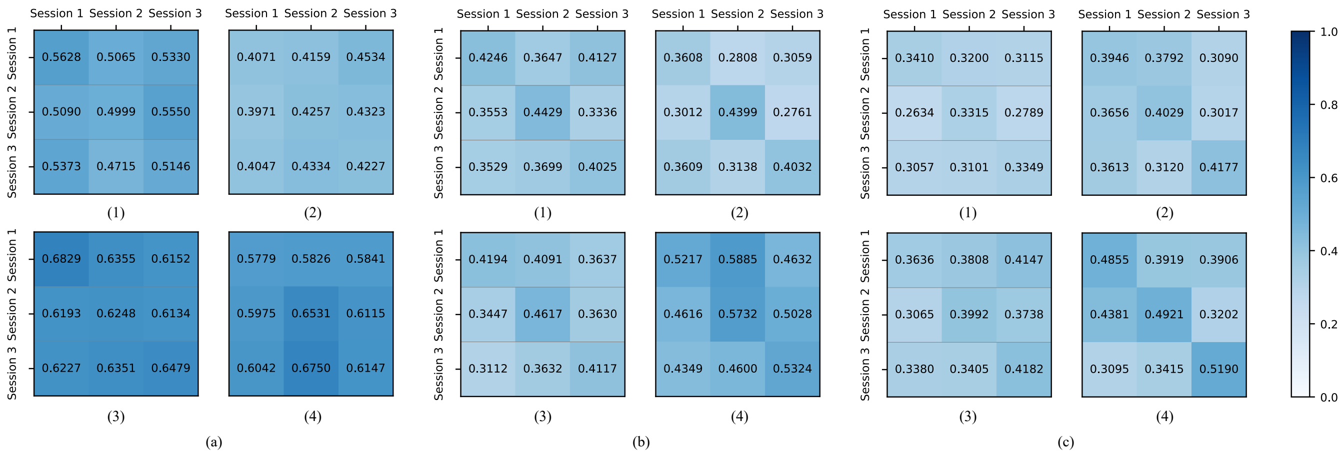


Fig. 3. The experimental results on (a) SEED, (b) SEED-IV, (c) SEED-V dataset by using the new experiment setting. For each subgraph, (1) refers to the results on EEG using SVM, (2) refers to the results on EYE using SVM, (3) refers to the results on EEG using KPCA and (4) refers to the results on EYE using KPCA.

using the short-term Fourier transforms (STFT) with a 4-second Hanning window without overlapping [21]. In order to maintain consistency, we also extract DE features from DEAP.

2) Eye movement features: Datasets SEED, SEED-IV and SEED-V also provide the eye movement features extracted from SMI ETG eye-tracking glasses [22]. The feature dimension is 33, including pupil diameter, dispersion, fixation duration, blink duration, saccade and so on.

C. Experiment Settings

Fig.1 shows the data partition of DEAP, SEED and SEED-IV. For DEAP dataset, we divide the data into two parts and each part contains the data when all subjects watched 20 (half of 40) music videos. We also follow the leave-one-subject-out cross validation idea here, i.e, during each fold, we train the model on 31 subjects' data in one part and test on the remaining subject's data in the other part. In order to make the comparison, we also test the model on the remaining subject's data in the same part. For SEED, SEED-IV and SEED-V dataset, we just divide the data by sessions for convenience and follow the approach we used in DEAP.

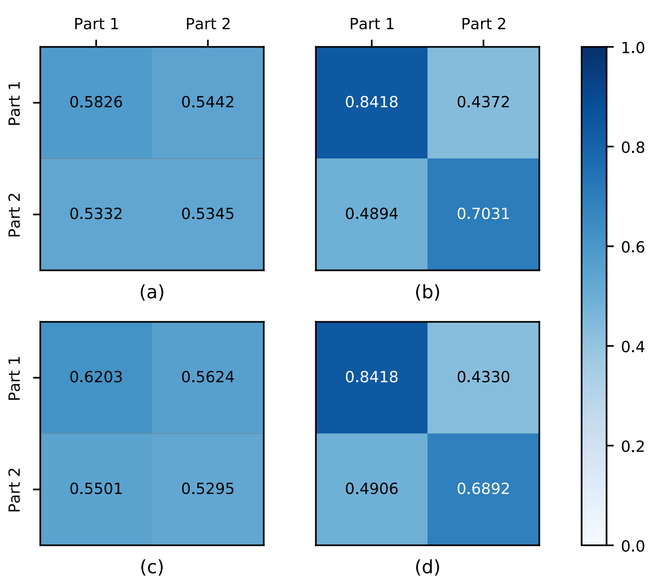


Fig. 2. The experimental results on DEAP dataset by using the new experiment setting, part 1 and part 2 are the results of bisection of DEAP in our experiment setting. (a) results on valence dimension using SVM. (b) results on arousal dimension using SVM. (c) results on valence dimension using KPCA. (d) results on arousal dimension using KPCA.

D. Experiment Results

In this section, we present the experimental results on DEAP, SEED, SEED-IV and SEED-V datasets by using SVM and KPCA-based subject transfer.

1) DEAP: Fig.2 shows the results on DEAP dataset. For each accuracy matrix, the label of each row refers to the training set, and the label of each column refers to the test set.

According to the idea we mentioned earlier, if emotional materials have an impact on classification accuracy, then the value on the diagonal of the accuracy matrix will be greater than the value on the other position because the training set and the test set used on the diagonal are generated from the same emotional material.

It can be seen from the experimental results that stimulus materials have a relatively large impact on arousal dimension and almost no impact on valence dimension. The results are well explained from the perspective of common sense that the same video had the same effect on EEG signals in different people. However, due to the differences in personal experience and physiology, there is a large difference in valence dimension, which eliminates the influence of whether the material is the same or not, resulting in a relatively average value in the accuracy matrix.

2) SEED, SEED-IV and SEED-V: Since the data composition and experimental settings of these three data sets are basically the same, we put their results in one section for comparison. Fig.3 shows the results of our experiment settings on SEED, SEED-IV and SEED-V datasets. For each accuracy matrix, the label of each row refers to the session used as the training set, and the label of each column refers to the session used as the test set.

From the results of SEED dataset, whether using SVM or KPCA-based subject transfer model, it is easy to find that the values in the accuracy matrix are close for EEG or eye movement features. This is because the data in SEED for each session is evoked by the same movie clips and the influence of movie clips on the signal is consistent from session to session.

For the results of SEED-IV and SEED-V dataset, we can find that the values on the diagonal of the accuracy matrix are greater than the values on the other position in general, which is similar to DEAP. This is because each session uses different movie clips to induce emotion, which is essentially the same as the partition we do on DEAP.

To summarize, the materials have an effect on the classification of emotion recognition and we can eliminate the effect by letting the training set and test set data be generated from different materials. For DEAP and SEED, the recommended method is to divide the data into 2-3 equal parts according to the material. For SEED-IV and SEED-V dataset, we need to use the division outside the diagonal. The final accuracy should be the average of the values in the accuracy matrix except the values on diagonal.

IV. CONCLUSIONS

In this paper, we have revealed that the materials used to induce emotions have an effect on the classification of emotion recognition under the traditional general experiment setting for the subject-independent models and have proposed a new cross-subject experiment setting which can eliminate the impact of stimulus materials. We evaluate our new experiment setting on four public emotion datasets (DEAP, SEED, SEED-IV and SEED-V) and find that when the data in the training set and test set is induced by the same video materials, the classification accuracy is often higher than that of the data induced by different video materials. Therefore, for cross-subject emotion recognition on any emotion dataset, the data should be partitioned according to our method and then the accuracy matrix should be calculated. The final accuracy should be the average of the values in the accuracy matrix except the values on diagonal.

V. ACKNOWLEDGMENTS

This work was supported in part by grants from the National Natural Science Foundation of China (Grant No. 61976135), SJTU Trans-Med Awards Research (WF540162605), the Fundamental Research Funds for the Central Universities, the 111 Project, and the China Southern Power Grid (Grant No. GDKJXM20185761).

REFERENCES

[1] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, and J. G. Taylor, "Emotion recognition in human-computer interaction," *IEEE Signal Processing Magazine*, vol. 18, no. 1, pp. 32–80, 2001.

[2] P. Ekman and D. Keltner, "Universal facial expressions of emotion," *Segerstrale U, P. Molnar P, eds. Nonverbal Communication: Where nature meets culture*, pp. 27–46, 1997.

[3] A. Mehrabian, "Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament," *Current Psychology*, vol. 14, no. 4, pp. 261–292, 1996.

[4] M. Pantic and L. J. M. Rothkrantz, "Automatic analysis of facial expressions: The state of the art," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1424–1445, 2000.

[5] J. Ang, R. Dhillon, A. Krupski, E. Shriberg, and A. Stolcke, "Prosody-based automatic detection of annoyance and frustration in human-computer dialog," in *Seventh International Conference on Spoken Language Processing*, 2002.

[6] A. Camurri, I. Lagerlöf, and G. Volpe, "Recognizing emotion from dance movement: comparison of spectator recognition and automated techniques," *International Journal of Human-Computer Studies*, vol. 59, no. 1-2, pp. 213–225, 2003.

[7] 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, pp. 1175–1191, 2001.

[8] 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.

[9] S. M. Alarcao and M. J. Fonseca, "Emotions recognition using EEG signals: A survey," *IEEE Transactions on Affective Computing*, vol. 10, no. 3, pp. 374–393, 2017.

[10] X.-W. Wang, D. Nie, and B.-L. Lu, "Emotional state classification from EEG data using machine learning approach," *Neurocomputing*, vol. 129, pp. 94–106, 2014.

[11] T. Song, W. Zheng, P. Song, and Z. Cui, "EEG emotion recognition using dynamical graph convolutional neural networks," *IEEE Transactions on Affective Computing*, vol. 11, no. 3, pp. 532–541, 2018.

[12] W. Samek, F. C. Meinecke, and K.-R. Müller, "Transferring subspaces between subjects in brain–computer interfacing," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 8, pp. 2289–2298, 2013.

[13] H. Morioka, A. Kanemura, J.-i. Hirayama, M. Shikauchi, T. Ogawa, S. Ikeda, M. Kawanabe, and S. Ishii, "Learning a common dictionary for subject-transfer decoding with resting calibration," *NeuroImage*, vol. 111, pp. 167–178, 2015.

[14] 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.

[15] R.-N. Duan, J.-Y. Zhu, and B.-L. Lu, "Differential entropy feature for EEG-based emotion classification," in *6th International IEEE/EMBS Conference on Neural Engineering*. IEEE, 2013, pp. 81–84.

[16] W.-L. Zheng and B.-L. Lu, "Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks," *IEEE Transactions on Autonomous Mental Development*, vol. 7, no. 3, pp. 162–175, 2015.

[17] B. Schölkopf, A. Smola, and K.-R. Müller, "Kernel principal component analysis," in *International Conference on Artificial Neural Networks*. Springer, 1997, pp. 583–588.

[18] 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, no. 3, pp. 1110–1122, 2018.

[19] T.-H. Li, W. Liu, W.-L. Zheng, and B.-L. Lu, "Classification of five emotions from EEG and eye movement signals: Discrimination ability and stability over time," in *2019 9th International IEEE/EMBS Conference on Neural Engineering*. IEEE, 2019, pp. 607–610.

[20] L.-M. Zhao, R. Li, W.-L. Zheng, and B.-L. Lu, "Classification of five emotions from EEG and eye movement signals: complementary representation properties," in *2019 9th International IEEE/EMBS Conference on Neural Engineering*. IEEE, 2019, pp. 611–614.

[21] 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.

[22] Y. Lu, W.-L. Zheng, B. Li, and B.-L. Lu, "Combining eye movements and EEG to enhance emotion recognition." in *International Joint Conference on Artificial Intelligence*, vol. 15. Citeseer, 2015, pp. 1170–1176.