Cross-Subject Emotion Classification based on Dual-Attention Mechanism and Meta-Transfer Learning

Qian Zhang (qian_zhang@stu.xidian.edu.cn)
Liying Yang*(yangliying1208@163.com)
Gang Cao (gang_cao@stu.xidian.edu.cn)
Qiang Wang (qiang_wang@stu.xidian.edu.cn)
Department of Computer Science and
Technology, Xidian University, Xi'an, China

Abstract

Emotion recognition based on electroencephalogram (EEG) signals is a current focus in brain-computer interface research. However, due to the individual differences, how to build a simple and effective model and quickly adapt to the target subject are significant challenges in cross-subject emotion recognition. In this study, we proposed an approach by combining the Dual-Attention network and **M**eta-**T**ransfer **L**earning (MTL) strategy based on k-means clustering for meta-task sampling. The Dual-Attention network extracts EEG features through a channel attention block and a temporal attention block. The MTL strategy trains the model to learn both common and individual features among subjects. The meta-task sampling method based on k-means clustering adaptively groups the source domain samples, sampling support and query sets for meta-tasks from **D**ifferent **G**roups(DG sampler). The DG sampler allows the model to "grow in diversity", further enhancing its generalization capabilities. Binary classification experiments were conducted on the DEAP dataset, achieving accuracies of 72.35% and 71.77% in the arousal and valence dimensions, respectively. The results have reached the state-of-the-art level and demonstrated significant performance enhancement in cross-subject EEG-based emotion recognition.

Keywords: emotion recognition; EEG; attention mechanism; meta-transfer learning; k-means clustering

Introduction

Emotions serve as reflections of an individual's current physiological and psychological state, significantly influencing cognition, communication, and decision-making. A prominent focus in the field of human-computer interaction is affective computing, dedicated to exploring theories and methodologies for identifying and interpreting human emotions (Poria et al., 2017). Affective computing processes various input signals, including video, audio, and physiological signals,etc. Unlike facial expressions, physiological signals like electroencephalogram (EEG) are less susceptible to manipulation and offer a more genuine reflection of an individual's emotional state. Therefore, emotion recognition based on EEG signals plays an important role in areas such as clinical diagnosis and treatment (Liu et al., 2011).

In the early stages, researchers manually extracted EEG features and applied traditional machine learning methods for EEG-based emotion recognition (Wang et al., 2014; Petrantonakis & Hadjileontiadis, 2010). With the development of deep learning, how to achieve end-to-end emotion classification using raw EEG signals is a research hotspot. Neuroscience studies suggest the involvement of specific brain regions in the generation of emotions. Davidson (1993) and Hajcak et al. (2010) proposed the frontal lobe's role in regulating nervous system activity during emotion regulation. Additionally, emotions trigger distinct physiological signals over time, causing various ability of EEG signals to express emotions across different time intervals (Ma et al., 2019). Based on these neuroscience findings, Ning et al. (2021) introduced the convolutional block attention module (CBAM) (Woo et al., 2018) to effectively learn both channel and temporal representations from EEG signals.

Generalizing models to subjects unseen during training is a significant challenge in cross-subject EEG-based emotion recognition task. Transfer learning emerges as a solution, with many researchers employing domain adversarial networks to minimize the distribution gap between source and target domains (He et al., 2022; Pei et al., 2023). The above studies based on unsupervised learning require a substantial number of target domain samples for effective domain adaptation. Meta-learning has been used for few-shot learning, enhancing model generalization on target domain tasks by learning from multiple source domain tasks. Model-agnostic meta-learning (MAML) (Finn et al., 2017), a classic meta-learning algorithm, updates model parameters through dual gradient updates and has been employed in various cross-domain tasks (Qian & Yu, 2019; Guo et al., 2019).

Recent studies have applied meta-transfer learning(MTL) (Sun et al., 2019) to cross-subject EEG-based emotion recognition task (Duan et al., 2021; J. Li et al., 2022). This innovative approach trains models to learn to learn, enabling them to rapidly adapt to the data distribution of new subjects by using minimal labeled samples. J. Li et al. (2022) constructed a neural network with multiple convolution kernels to extract intricate emotional representations from connectivity features in EEG signals. Their use of MTL strategy enabled the model to learn shared emotion recognition patterns among individuals.

Due to significant variability in emotion recognition patterns based on EEG signals among different subjects, some researchers proposed sampling meta-tasks from various subjects during meta-transfer learning (J. Li et al., 2022; S. Li et al., 2022). Recent studies in neuroscience exploring Inter-Subject Correlation (ISC) suggest that EEG signals from individuals watching the same emotional video can reflect collective arousal, valence, and more (Dmochowski et al., 2012, 2014). These investigations provide crucial neurological evidence for constructing shared EEG-based emotional representations among individuals. Inspired by this research, Shen et al. (2022) generated positive pairs from EEG signals of different subjects viewing the same video and vice versa. While these studies leveraged differences among subjects or videos to design data samplers, their sampling methods relied on firm groupings, treating EEG samples from individual subject or video as identical data distribution. Therefore, these methods are challenging to effectively enhance the generalization capability of the model. Qu et al. (2020) used an unsupervised k-means algorithm to empirically stratify individuals into more homogeneous subgroups by clustering on EEG patterns. The results indicated that two subgroups identified in each of the three clustering models are highly consistent.

To enhance the model's generalization ability, we proposed a model that integrates the Dual-Attention network and MTL strategy, leveraging a grouping mechanism based k-means clustering for meta-task sampling. Inspired by the CBAM (Woo et al., 2018), we've designed a Dual-Attention network that extracts channels and temporal segments crucial for distinguishing emotions. Compared to traditional transfer learning, meta-transfer learning diminishes the reliance on target domain data. Therefore, our method adopts MTL to train model parameters, enabling the model to grasp shared EEG emotion patterns among subjects. Additionally, we proposed an innovative meta-task sampling method rooted in k-means clustering.The sampling method adaptively groups samples from source domain based on k-means clustering, heightening the distributional diversity between support and query sets within meta-tasks, thus enriching the model's learning capabilities.

Methodology

This section focuses on three key components of our approach: Dual-Attention model, MTL strategy, and the DG sampler. The Dual-Attention model effectively captures channel characteristics and temporal features from EEG signals using a channel attention block and a feature attention block. The MTL strategy updates parameters to enable the model to learn both common and individual traits among diverse subjects. Meanwhile, the DG sampler samples support and query sets across different groups, intentionally increasing the distribution discrepancy in meta-tasks and further enhancing the model's overall ability to generalize.

Dual-Attention Model

The researchers utilized multiple channels from various brain regions while collecting EEG datasets, yet studies indicate that only specific brain regions are associated with emotional responses. Additionally, emotions exhibit distinct and stable patterns during specific time intervals (T.-H. Li et al., 2019). To extract EEG features that better differentiate between emotions, we introduced the Dual-Attention model based on attention mechanism, as shown in Figure 1. The model employs a dual attention mechanism for channel and temporal features, automatically focusing on critical channels and time segments distinguishing emotions. It assigns higher weights to key channels and sampling points, effectively extracting EEG features while reducing model complexity.

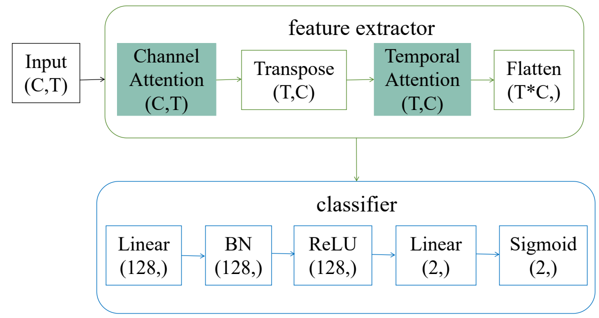

Figure 1: The framework of Dual-Attention model. The Dual-Attention model comprises a feature extractor and a classifier. The feature extractor consists of a channel attention block and a temporal attention block, responsible for extracting channel-specific and temporal features from EEG signals, respectively.

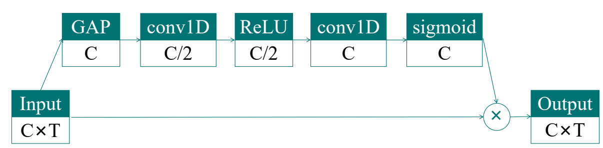

Figure 2: The framework of channel attention module, where C represents the number of channels and T represents the number of temporal features for each channel.

Channel Attention Module We calculated channel weights based on the inter-channel relationships of features. Figure 2 illustrates the architecture of the channel attention module. To effectively compute channel attention weights, we compressed the temporal feature dimension of the input signal utilizing global average pooling. Subsequently, two one-dimensional convolutional layers form the bottleneck layer, followed by a final sigmoid layer to train the channel weights. In essence, the computation within the channel attention block can be described using the following formula:

$$W_c(X) = \sigma(\mathrm{Conv1D}(\mathrm{ReLU}(\mathrm{Conv1D}(\mathrm{AvgPool}(X))))), \quad (1)$$

$$X^{'} = W_c(X) \otimes X, \quad (2)$$

where σ is sigmoid function, $\otimes$ represents element-wise multiplication. During the multiplication process, attention values are broadcasted accordingly: the channel attention values are broadcasted along the spatial dimension. $X \in \mathbb{R}^{C\times T}$, $X^{'} \in \mathbb{R}^{C\times T}$, $W_c(X) \in \mathbb{R}^{C}$, C represents the number of channels, T represents the number of temporal features for each channel, and $W_c(X)$ represents the computed channel weights.

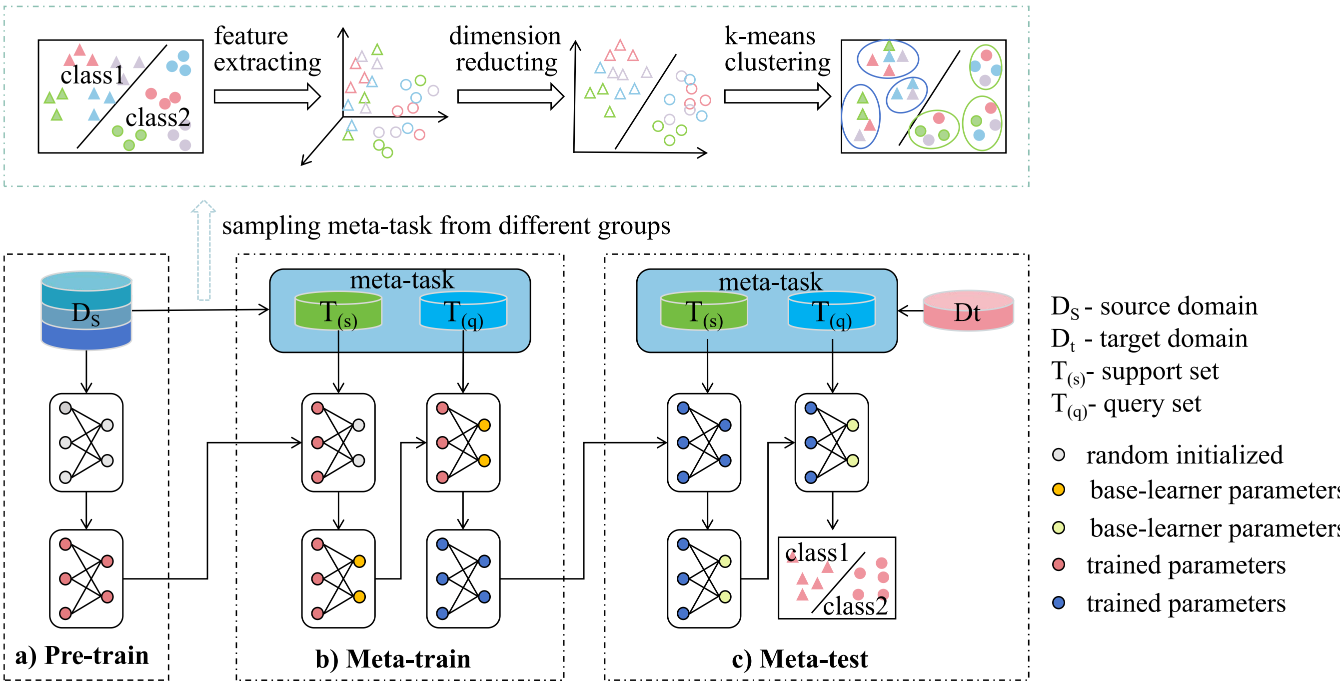

Figure 3: The overall process of MTL. The black dashed box sequentially presents three stages of MTL: pre-train phase, meta-train phase, and meta-test phase. The green dashed box illustrates the processing flow of the DG sampler: feature extraction, feature dimensionality reduction, and k-means clustering are sequentially applied to obtain $n\_way \times k$ groups. Subsequently, during the meta-train phase, support and query sets are sampled independently from distinct groups.

Temporal Attention Module Temporal attention block shares a structure similar to the channel attention block, except that the output from the channel attention block $X^{'}$ is transposed to $X^{''}$, where $X^{''} \in \mathbb{R}^{T \times C}$.

Meta-Transfer Learning (MTL)

Due to the substantial variability in EEG signals associated with the same emotion across different subjects, we've employed MTL strategy to train our model for rapid adaptation to the data distribution of target subject. MTL strategy combines the advantages of transfer learning and meta-learning. Transfer learning leverages knowledge learned by the model from a large number of labeled source domain samples to the target domain, while meta-learning trains the model's learning capabilities by sampling meta-tasks, particularly useful when the target domain has limited labeled samples. MTL strategy consists of three main stages: pre-training, meta-training, and meta-testing, as illustrated in Figure 3.

Pre-train Stage The pre-train stage aligns with the conventional pre-train stage in classic transfer learning paradigms, such as pre-training for object detection on large-scale image datasets like ImageNet (Russakovsky et al., 2015). During this phase, we trained the model using all data from the source domain, employing gradient descent to update the randomly initialized parameters of the model. This allows the model to learn the overall data distribution of all subjects.

We begined by initializing a feature extractor ϕ (e.g. the feature extractor in the Dual-Attention model) and a classifier θ (e.g. a fully connected layer) and performed gradient updates using the following formula:

$$[\phi;\theta] = [\phi;\theta] - \alpha \nabla L_{D_s}([\phi;\theta]), \tag{3}$$

$$L_{D_s}([\phi;\theta]) = \frac{1}{|D_s|} \sum_{(x,y)\in D_s} l(f_{[\phi;\theta]}(x), y), \tag{4}$$

where α denotes the learning rate, and $l(.)$ represents the loss function, typically the cross-entropy loss.

During the pre-train stage, the model efficiently fits the data distribution of source domain. The parameters of the feature extractor obtained during this stage are used to initialize the model in meta-train stage. However, since the model hasn't yet learned the differences between various subjects, the classifier, being the closest component to the output, is removed.

Meta-train Stage Due to significant differences in the data distribution between source and target domains, the model trained in the pre-train phase cannot generalize well to the target subject. Therefore, the meta-learning concept is employed to continue updating the model in the meta-train phase. Meta-tasks are sampled from the source domain data, consisting of support and query sets used to update both base-learner and meta-learner of the model. During this phase, the model parameters are updated through a dual gradient descent approach.

In each episode, the model computes the loss on the support set and updates the base-learner θ several times within an inner loop, as shown in Equation (5):

$$\theta^{'} = \theta - \beta \nabla L_{T_{(s)}}([\phi;\theta]), \tag{5}$$

where β denotes the base learning rate. $T_{(s)}$ represents support set, $T_{(s)} \in \mathbb{R}^{N_1 \times C \times T}$, $N_1 = n\_way \times n\_shot$, n_way is the number of classes, n_shot is the number of samples per class in the support set.

Following this, utilizing the feature extractor retained from previous episode and base-learner updated within the inner loop, we computed the loss on query set. Subsequently, we proceed to update meta-learner $[\phi;\theta]$ in the outer loop, as illustrated in Equation (6):

$$[\phi;\theta] = [\phi;\theta] - \gamma \nabla L_{T_{(q))}}([\phi;\theta^{'}]), \tag{6}$$

here γ denotes the meta learning rate, $T_{(q)}$ represents query set, $T_{(q)} \in \mathbb{R}^{N_2 \times C \times T}$, $N_2 = n\_way \times n\_query$, n_query is the number of samples per class in the query set.

In the meta-train phase, the dual gradient descent approach enables the model to grasp the distribution differences between support and query set samples. The model's base-learner adeptly extracts individual-specific features, while the meta-learner, particularly within the feature extractor component, learns shared characteristics among all subjects.

Meta-test Stage The meta-test stage involves sampling meta-tasks from the target domain. In this stage, the support set comprises labeled samples, while the query set lacks labels. Initially, the base-learner is updated using the support set, followed by the assessment of the model's accuracy in recognizing emotions within the query set.

The base-learner of the model, being the closest component to the output, can effectively capture intricate features within EEG signals, reflecting unique variations among different individuals. Therefore, during the meta-test stage, a simple fine-tuning of the model's base-learner enables adaptation to the specific data distribution of the target subject, leading to enhanced performance.

Sampling from Different Groups(DG sampler)

In the meta-train stage, the original MTL method randomly sampled meta-tasks from all samples in the source domain. However, due to substantial variations in EEG data among different subjects and even within various trials from the same subject, we introduced the DG sampler to enhance adaptation of the MTL strategy in cross-subject emotion recognition task. Figure 3 summarizes the workflow of DG sampler. Initially, we extracted features from the source domain data using a model trained in the pre-train phase. As the feature dimension was large, we performed PCA dimensionality reduction on the extracted features. Finally, k-means clustering was applied to the reduced features to obtain k clusters. For better meta-task sampling, we performed clustering on samples with different emotional labels. The method clustered the source domain samples into $n\_way \times k$ groups, from which the support and query sets for meta-tasks were sampled in different groups. The explicit augmentation of the data distribution disparity between support and query sets within meta-tasks aimed to enhance the model's generalization ability.

Expriments and Results

Dataset and Data Processing

The DEAP dataset comprises physiological signals elicited by music video stimuli, capturing recordings from 32 subjects while they watched 40 one-minute music videos. These recordings includes both the physiological responses and the subjects' subjective psychological assessments—Valence, Arousal, Dominance, and Liking—rated on a 1 to 9 scale. The physiological signals were initially sampled at 512Hz, followed by preprocessing and downsampling to 128Hz. Each subject's dataset matrix is 40 × 40 × 8064 (40 experimental music videos, 40 channels of physiological data, 8064 sampling points).

Utilizing 32 channels of EEG signals, the dataset was further segmented into 1-second intervals, removing the first 3s baseline. Consequently, each subject contributed 2400 samples, each represented as a matrix of dimensions 32 × 128. The samples were labeled based on a threshold of 5, with scores equal to or greater than 5 assigned a high valence/arousal label, and scores less than 5 assigned a low valence/arousal label. Binary classification experiments were conducted independently in both valence and arousal dimensions.

Implementation Details

Based on MTL strategy, the model undergoes training with a large number of labeled samples from the source domain, followed by fine-tuning using a limited amount samples from the target domain. Thus, we evaluated the proposed method's performance in cross-subject task using a leave-one-subject-out cross-validation technique.

In the pre-training phase, we used the Adam optimizer to update model parameters, iterating for 10 epochs. In the meta-training phase, each episode involved sampling meta-tasks from the source domain, with a batch size of 30 comprising 10-shot and 20-query, iterated over 100 episodes. During this phase, we utilized an Adam optimizer with a base-learning rate of 0.005 and a meta-learning rate of 0.0001, with 10 iterations for base-learner parameter updates. When performing k-means clustering on the source domain samples, the value of k was set to 25. In the meta-testing phase, meta-tasks were sampled from the target domain for testing. We set varying n-shot (10, 15, 20, 25) alongside a n-query of 10. Similar to the meta-training phase, we employed the Adam optimizer with consistent parameter settings.

Table 1: The classification accuracies of different methods.

Methods	Valence(%)	Arousal(%)
AD-TCN	64.33 ± 7.06	63.25 ± 4.62
ATDD-LSTM	69.06 ± 6.37	**72.90 ± 6.57**
MUPS-EEG	66.50 ± 4.70	/
MTL-MSRN	71.29 ± 7.67	71.92 ± 6.79
Ours	**72.35 ± 5.79**	71.77 ± 5.60

Results

We compared our proposed method with the current state-of-the-art works, as presented in Table 1. AD-TCN (He et al., 2022) utilizes a temporal model as a feature encoder, and integrates domain adversarial networks to mitigate distribution gaps between the source and target domains. ATDD-LSTM (Du et al., 2020), also employing domain adversarial networks for domain adaptation, constructs an LSTM-based model to capture spatial characteristics among distinct electrode EEG signals. In comparison to AD-TCN, ATDD-LSTM considers interrelationships among different EEG channels, resulting in a noticeable performance enhancement. However, these methods rely on unsupervised learning principles, requiring a considerable amount of unlabeled target domain data for domain adaptation. Furthermore, MUPS-EEG (Duan et al., 2021) and MTL-MSRN (J. Li et al., 2022) utilize MTL strategy to train models, fine-turning them with limited labeled target domain samples. Although MTL-MSRN obtains close accuracy with the proposed method, our model displays a smaller standard deviation. This suggests that our model, employing MTL strategy with DG sampler, offers superior generalization across different subjects, including those with more distinctive individual characteristics.

Ablation Experiments

Table 2: The accuracies(%) comparison of different models in the valence dimension.

Models	10-shot	15-shot	20-shot	25-shot
MLP	58.22	60.74	62.87	63.96
CNN	56.16	57.85	59.35	61.31
Transformer	59.81	61.80	64.37	65.40
Dual-Attention	**64.29**	**67.20**	**69.94**	**72.35**

Table 3: The accuracies(%) comparison of different models in the arousal dimension.

Models	10-shot	15-shot	20-shot	25-shot
MLP	59.75	61.97	64.14	66.20
CNN	58.15	60.60	61.42	62.72
Transformer	59.62	62.36	64.22	65.99
Dual-Attention	**64.23**	**67.49**	**69.70**	**71.77**

Comparison of Models To validate the superior performance of our proposed Dual-Attention model, we developed three additional neural network architectures: Multilayer Perceptron (MLP), Convolutional Neural Network (CNN), and a Transformer-based network, for comparative analysis. We adopted the MTL strategy for our proposed Dual-Attention model and these three models, and sampled meta-tasks using the DG sampler.

Tables 2 and 3 present the comparative results of different models in the valence and arousal dimensions within the DEAP dataset. Our proposed Dual-Attention model outperforms others in both dimensions. Transformer closely follows Dual-Attention, indicating the effectiveness of attention mechanisms in extracting features from EEG signals. Among these, CNN performs the poorest. Despite extracting temporal and channel features, its characteristics of local perception and parameter sharing may overlook inter-channel correlations and overall trends in temporal features. MLP performs better than CNN but remains suboptimal. It directly flattens the raw two-dimensional EEG signals into one-dimensional vectors, neglecting crucial channel information. In contrast, the Dual-Attention model compresses temporal and channel features through global average pooling, subsequently training channel and temporal weights to focus more intuitively on distinguishing key channels and temporal segments for emotions. Hence, the Dual-Attention model achieves the highest accuracy in both valence and arousal dimensions compared to other models.

Comparison of Meta-task Samplers To assess the effectiveness of our proposed DG sampler in improving model generalization, we conducted ablation experiments using three other meta-task sampling methods:

AS sampler: Randomly selecting support and query sets from all samples of source domain.

DS sampler: Each episode, randomly choosing a pair of subjects and sampling support set and query set from their respective samples.

DT sampler: Selecting a subject per episode and sampling support and query sets from various videos watched by this subject.

Table 4: The accuracies(%) comparison of different meta-task samplers in the valence dimension.

Samplers	0-shot	10-shot	15-shot	20-shot	25-shot
AS	53.36	61.13	64.65	66.70	68.61
DS	54.73	62.07	64.86	66.92	69.30
DT	**54.82**	61.98	65.37	67.33	69.26
DG	53.97	**64.29**	**67.20**	**69.94**	**72.35**

Table 5: The accuracies(%) comparison of different meta-task samplers in the arousal dimension.

Samplers	0-shot	10-shot	15-shot	20-shot	25-shot
AS	58.45	60.74	64.19	66.66	68.37
DS	**58.88**	61.47	64.09	67.18	68.93
DT	57.94	62.92	65.83	67.72	69.73
DG	58.63	**64.23**	**67.49**	**69.70**	**71.77**

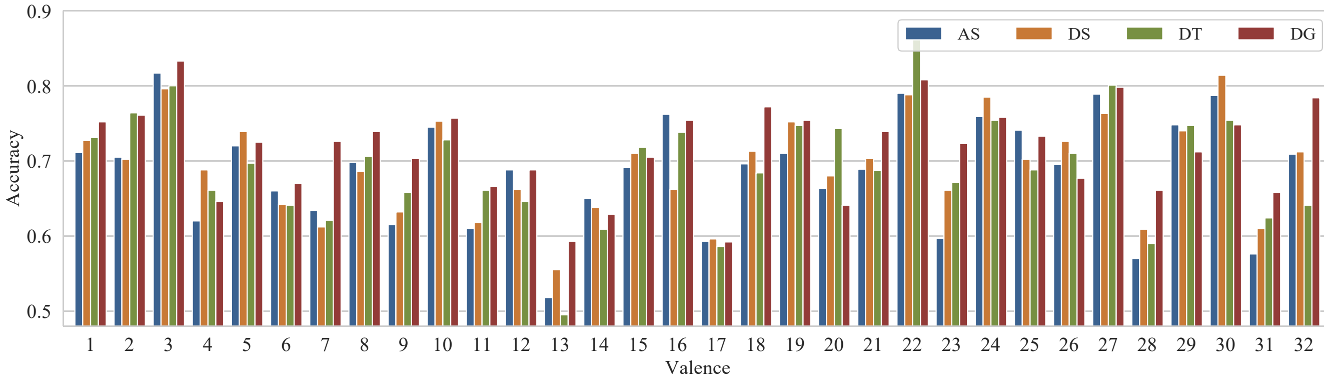

Figure 4: The results in the valence dimension based on different samplers when using 32 subjects individually as the target domain.

Tables 4 and 5 display the comparative results of four different meta-task samplers, with the DG sampler exhibiting the best performance. It can be observed that the accuracy rates at 0-shot among these samplers are close, differing by no more than 0.5%. However, the accuracy rate reaches only around 58%, indicating a relatively limited ability of the model to generalize to target subject without fine-tuning. After training the model with a small amount of labeled data, the accuracy rates of all four samplers improved. Among them, the DS and DT samplers show slightly higher results than the AS sampler, indicating that these two sampling methods allow the model to perceive the differences between different subjects or videos, thereby enhancing the model's learning ability amid sample variations. The DG sampler shows significantly superior results after fine-tuning compared to the other samplers. This is attributed to the DG sampler's more flexible grouping of samples based on the overall data distribution of the source domain using the k-means clustering. Figure 4 demonstrates the comparison of results obtained when 32 subjects are used as the target domain for different sampling samplers. It shows improved accuracy for the majority of subjects when utilizing the DG sampler, further confirming that the model trained with the DG sampler combined with MTL strategy exhibits superior generalization ability.

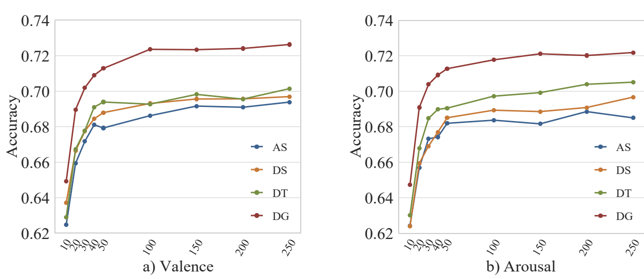

Figure 5: The test accuracies increase with update steps of the base-learner.

Comparison of Update Steps of the Base-learner in the Meta-test Stage Meta-transfer learning aims to rapidly adapt the model to the data distribution of the target domain through a few update steps. Hence, we examined the impact of the update steps of the base-learner during the meta-test stage on the accuracies of query sets, illustrated in figure 5. It reveals a consistent trend across the four different meta-task samplers as update steps increase. There's a noticeable accuracy improvement up to around 50 update steps, followed by marginal gains beyond 100 steps. This demonstrates the fast convergence of our proposed Dual-Attention model after a limited number of update steps. Additionally, the DG sampler consistently outperforms the other three samplers across various update steps, confirming its superior performance in cross-subject emotion recognition task.

Conclusion

In order to address the challenge of poor model generalization due to individual differences in EEG emotional patterns, we proposed an approach that combines the Dual-Attention network with MTL strategy using k-means clustering for meta-task sampling. The method effectively extracts channel-specific and temporal features distinguishing various emotions based on attention mechanism. Through MTL strategy, it learns shared emotional EEG patterns among individuals, enabling the model to rapidly adapt to target subjects via fine-tuning. Moreover, the DG sampler method based on k-means clustering amplifies differences between the support and query sets, further enhancing the model's generalization ability. Evaluating our method on the DEAP dataset yielded accuracy rates of 72.35% and 71.77% in the arousal and valence dimensions, respectively. Additionally, we conducted ablation experiments on the Dual-Attention network and DG sampler, validating their crucial contributions to cross-subject EEG emotion recognition.

Acknowledgments

This work was supported by the Natural Science Foundation of China (62374121,61974109).

References

Davidson, R. J. (1993). Cerebral asymmetry and emotion: Conceptual and methodological conundrums. *Cognition & Emotion*, *7*(1), 115–138.

Dmochowski, J. P., Bezdek, M. A., Abelson, B. P., Johnson, J. S., Schumacher, E. H., & Parra, L. C. (2014). Audience preferences are predicted by temporal reliability of neural processing. *Nature communications*, *5*(1), 4567.

Dmochowski, J. P., Sajda, P., Dias, J., & Parra, L. C. (2012). Correlated components of ongoing eeg point to emotionally laden attention–a possible marker of engagement? *Frontiers in human neuroscience*, *6*, 112.

Du, X., Ma, C., Zhang, G., Li, J., Lai, Y.-K., Zhao, G., ... Wang, H. (2020). An efficient lstm network for emotion recognition from multichannel eeg signals. *IEEE Transactions on Affective Computing*, *13*(3), 1528–1540.

Duan, T., Chauhan, M., Shaikh, M. A., Chu, J., & Srihari, S. N. (2021). Ultra efficient transfer learning with meta update for continuous eeg classification across subjects. In *Canadian conference on ai.*

Finn, C., Abbeel, P., & Levine, S. (2017). Model-agnostic meta-learning for fast adaptation of deep networks. In *International conference on machine learning* (pp. 1126–1135).

Guo, D., Tang, D., Duan, N., Zhou, M., & Yin, J. (2019). Coupling retrieval and meta-learning for context-dependent semantic parsing. *arXiv preprint arXiv:1906.07108*.

Hajcak, G., MacNamara, A., & Olvet, D. M. (2010). Event-related potentials, emotion, and emotion regulation: an integrative review. *Developmental neuropsychology*, *35*(2), 129–155.

He, Z., Zhong, Y., & Pan, J. (2022). An adversarial discriminative temporal convolutional network for eeg-based cross-domain emotion recognition. *Computers in biology and medicine*, *141*, 105048.

Li, J., Hua, H., Xu, Z., Shu, L., Xu, X., Kuang, F., & Wu, S. (2022). Cross-subject eeg emotion recognition combined with connectivity features and meta-transfer learning. *Computers in biology and medicine*, *145*, 105519.

Li, S., Wu, H., Ding, L., & Wu, D. (2022). Meta-learning for fast and privacy-preserving source knowledge transfer of eeg-based bcis. *IEEE Computational Intelligence Magazine*, *17*(4), 16–26.

Li, T.-H., Liu, W., Zheng, W.-L., & Lu, B.-L. (2019). 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 (ner)* (pp. 607–610).

Liu, Y., Sourina, O., & Nguyen, M. K. (2011). Real-time eeg-based emotion recognition and its applications. *Transactions on Computational Science XII: Special Issue on Cyberworlds*, 256–277.

Ma, J., Tang, H., Zheng, W.-L., & Lu, B.-L. (2019). Emotion recognition using multimodal residual lstm network. In *Proceedings of the 27th acm international conference on multimedia* (pp. 176–183).

Ning, R., Chen, C. P., & Zhang, T. (2021). Cross-subject eeg emotion recognition using domain adaptive few-shot learning networks. In *2021 ieee international conference on bioinformatics and biomedicine (bibm)* (pp. 1468–1472).

Pei, N., Yang, L., Chao, S., Sun, H., et al. (2023). User-independent emotion classification based on domain adversarial transfer learning. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 45).

Petrantonakis, P. C., & Hadjileontiadis, L. J. (2010). Emotion recognition from brain signals using hybrid adaptive filtering and higher order crossings analysis. *IEEE Transactions on affective computing*, *1*(2), 81–97.

Poria, S., Cambria, E., Bajpai, R., & Hussain, A. (2017). A review of affective computing: From unimodal analysis to multimodal fusion. *Information fusion*, *37*, 98–125.

Qian, K., & Yu, Z. (2019). Domain adaptive dialog generation via meta learning. *arXiv preprint arXiv:1906.03520*.

Qu, X., Liukasemsarn, S., Tu, J., Higgins, A., Hickey, T. J., & Hall, M.-H. (2020). Identifying clinically and functionally distinct groups among healthy controls and first episode psychosis patients by clustering on eeg patterns. *Frontiers in psychiatry*, *11*, 541659.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., ... others (2015). Imagenet large scale visual recognition challenge. *International journal of computer vision*, *115*, 211–252.

Shen, X., Liu, X., Hu, X., Zhang, D., & Song, S. (2022). Contrastive learning of subject-invariant eeg representations for cross-subject emotion recognition. *IEEE Transactions on Affective Computing*.

Sun, Q., Liu, Y., Chua, T.-S., & Schiele, B. (2019). Meta-transfer learning for few-shot learning. In *Proceedings of the ieee/cvf conference on computer vision and pattern recognition* (pp. 403–412).

Wang, X.-W., Nie, D., & Lu, B.-L. (2014). Emotional state classification from eeg data using machine learning approach. *Neurocomputing*, *129*, 94–106.

Woo, S., Park, J., Lee, J.-Y., & Kweon, I. S. (2018). Cbam: Convolutional block attention module. In *Proceedings of the european conference on computer vision (eccv)* (pp. 3–19).