Cross-phase Emotion Recognition using Multiple Source Domain Adaptation

Ke-Ming Ding[1], Tsukasa Kimura[2], Ken-ichi Fukui[2] and Masayuki Numao[2]
[1]*Graduate School of Information Science and Technology, Osaka University, Japan*
[2]*The Institute of Scientific and Industrial Research (ISIR), Osaka University, Japan*

Keywords: Electroencephalography (EEG), Emotion, Domain Adaptation, Generative Adversarial Network (GAN), Cross Phase.

Abstract: EEG signal, the brain wave, has been widely applied in detecting human emotion. Due to the human brain's complexity, the EEG pattern varies from different individuals, leading to low cross-subject classification performance. What is more, even within the same subject, EEG data also shows diversity for the same reason. Many researchers have conducted experiments to deal with the variance between subjects by transfer learning or domain adaptation. However, most of them are still low-performance, especially when the new subject does not share generality with training samples. In this study, we examined using cross-phase data instead of cross-subject data because the discrepancy of different phase data should be smaller than that of different subjects. Different phases represent data recorded multiple times from the same subject with the same stimuli. Two neural networks are adopted to verify the effectiveness of the cross-phase domain adaptation. As a result, experiments on the public EEG dataset showed approximation level accuracy compared to the state-of-the-art method but much lower standard derivation. Moreover, multiple source domains promote accuracy in contrast to one single domain. This study helps develop a more robust and high-performance real-time EEG system by transferring knowledge from previous data phases.

1 INTRODUCTION

Emotion plays an important role in cognitive science to decode the human brain. Traditional methods in detecting affective states include questionnaires or communication with the subject directly while these interactions intensely depend on individual perception. However, results are steadily influenced by external factors like environment, preference and tension. In recent years, using psychological signals, especially the electroencephalography (EEG), eye tracks, magnetoencephalography (MEG), is extensively employed to detect emotion state due to its objectivity (Bos et al., 2006). Among all these signals, EEG is mostly adopted because of its non-invasive and manageable solution (Coan and Allen, 2004). EEG electrodes placed at the scalp surface continuously record signals and trace affective states simultaneously. Machine Learning technique is thus adopted in detecting discriminative features and emotion classification depending on the electrical signals. Traditional machine learning algorithms include linear discriminant analysis (LDA), support vector machines (SVM) (Duan et al., 2013). With the development of deep learning, there are more and more deep neural networks employed in bio-signal processing as a feature extractor or classifier (Kahou et al., 2016). These deep learning architectures show more robust and better performance in many classification tasks. However, the diversity of EEG patterns within individual subjects limits the transferability of trained models between them, that is, model trained will not be necessarily profitable in a new subject (Jayaram et al., 2016). Although in a within-subject experiment, advanced classifiers obtain quite high accuracy that over 90%, in a subject-independent experiment, the same algorithm can only get accuracy about 70% (Luo et al., 2018), which further proves the high discrepancy across different subjects. Domain Adaptation (DA), is proposed to deal with the problem, which tries to align the distribution of different subjects to train a common classifier upon the new distribution (Margolis, 2011). For example, a DA algorithm named adaptive subspace feature matching (ASFM) (Chai et al., 2017) reduced dimension both on the source domain and target domain by PCA. Then the algorithm integrated

Ding, K., Kimura, T., Fukui, K. and Numao, M.
Cross-phase Emotion Recognition using Multiple Source Domain Adaptation.
DOI: 10.5220/0010200701500157
In *Proceedings of the 14th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2021) - Volume 4: BIOSIGNALS*, pages 150-157
ISBN: 978-989-758-490-9

both the marginal distributions with a unified transformation function. As a result, both the marginal and conditional discrepancies between the source and target domain were reduced. Finally, logistic regression was applied to the new source subspace and a classifier was adopted in the target domain. ASFM shows significant improvement in accuracy compared with a simple no-adaptation method, showing accuracy about 80%.

By now, many domain adaptation technologies have been applied in subject-independent emotion classification and achieve improvement in accuracy (Jayaram et al., 2016; Zheng and Lu, 2016). They want to take advantage of a large amount of data from different subjects and extract common knowledge that can apply in any new subject. However, we need to note that the variance within different subjects still exists and may lead to negative transfer learning——if the new subject is quite different from any other training samples (Li et al., 2018). Furthermore, collecting abundant training samples from different subjects can be time-consuming and costly, even impossible in build real-time emotion recognition system (Guger et al., 2000). To solve this problem, this study utilized data recorded from the same subject instead of data from different subjects so that the discrepancy will be smaller because of the correlation of emotion within the same subject. We used one of the most-cited EEG emotion dataset SEED and subjects in this dataset watched the same movie clip three times. Instead of data from different individuals as source domain, we choose data from the same person but in different periods as the source. The term phase is applied in reference to separate EEG signals while the subject was exposed to the same stimuli. Several unsupervised domain adaptation methods, domain-Adversarial Neural Network (DANN) (Ganin et al., 2016) and Multiple Source Domain Adaptation (MDAN) (Zhao et al., 2018) were operated and the results were compared with one state-of-the-art (SOTA) approach. The classification accuracy of MDAN was in approximation level with the SOTA model and MDAN showed much lower standard derivation. In addition, MDAN indicated more potential in real-world application. Increasing more source domains (collecting more data from the same subject) helps promote classification. What is more, we investigated what actually conclude negative transferring. Even exposure to the same movie, the familiarity and likeness will change while only similar affection enhances transfer learning. We explored the self-assessments when each subject exposing to the stimuli and compared the difference rating scores. The result further convinces that reaction of disparate subjects will be much more variant than reaction from one same subject, revealing the effectiveness of our research.

2 EXPERIMENT SETTING

2.1 SEED Dataset

This study was performed on the publicly available EEG dataset for emotion recognition. SEED dataset is proposed by Shanghai JiaoTong University (SJTU) (http://bcmi.sjtu.edu.cn/seed/). The SEED dataset consists of 15 participants, each of whom needs to watch 15 Chinese movie clips to induce three kinds of emotions: positive, neutral, and negative. All of them are native Chinese with 7 males and 8 females. The movie clips are well selected as the stimuli in the experiments. All movies need to be understandable and neither too long nor too short to induce enough emotional meanings. The EEG signals were recorded by an ESI NeuroScan System with a 62 electrode cap at a sampling rate of 1000 Hz. After each exploration of a movie clip, the subject needs to rate their emotion. Each film clip lasts about 4 minutes and there is a 5-s hint before each clip and 46-s for a questionnaire. Questions are as follows for self-assessment: 1) what they had felt in response to viewing the film clip; 2) have they watched this movie before; 3) have they understood the film clip. The procedure of each trail is shown in Figure 1.

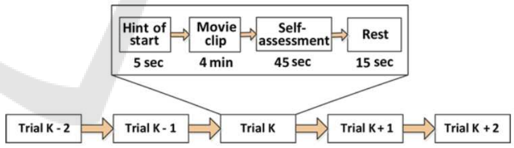

Figure 1: Protocal of each trail in the experiment.

Compared with many other EEG emotion datasets like DEAP (Koelstra et al., 2011), which only recorded data once when subjects were exposed to the stimuli, the SEED dataset has a special term of session. Each session includes data of all the subjects watching all the 15 film clips once. There were 3 sessions of data collected during the experiment. The difference between these sessions is the external environment like the time of the experiment. The mood of each subject and familiarity of each movie must be different since we can't strictly control the internal variables of each participant. However, the stimuli and subjects are the same, which guarantees the latent correlation of each session data. One of the main purposes of this study is to discover the latent correlation of EEG recording from the same person when exposing to the same stimuli.

2.2 Preprocess and Feature Extraction

The raw data was down-sampled with a frequency of 200 Hz and artifacts are removed manually. EEG signals that were seriously contaminated by EMG and EOG were checked visually and removed manually. Eye tracks (EOG) were also recorded to mark artifacts and removed later. A band-pass filter between 0.3 and 50 Hz was utilized to remove noise upon the EEG signals. Only EEG segments in duration of each film are selected. A 1s Hanning window without overlapping is operated on the original EEG data and finally, 3300 clean epochs are obtained in each channel for each experiment. Instead of raw data, frequency domain data, which calculated by a 512-point short-time Fourier transform in each time window is applied as input features. Various feature types are provided in SEED dataset like power spectral density (PSD), differential asymmetry (DASM), rational asymmetry (RASM), asymmetry (ASM), and differential entropy (DE). According to some studies (Zheng and Lu, 2015; Duan et al., 2013), a simple but discriminative feature named DE feature shows good performance in emotion recognition compared to other features. DE feature is defined as :

$$
\begin{aligned}
h(X) &= -\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \log\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right) dx \\
&= \frac{1}{2}\log\left(2\pi e\sigma^2\right)
\end{aligned}
\tag{1}
$$

where x obeys the Gaussian Distribution $N\left(\mu,\ \sigma^2\right)$. Obviously, original EEG signal doesn't follow the certain distribution. However, studies (Duan et al., 2013; Zheng and Lu, 2015) proved that the probability that certain sub-band signals meet Gaussian distribution is larger than 90 percent. It has been proved in that for a fixed EEG sequence, DE is equivalent to the logarithm average energy in a certain frequency band. After band-pass filtering in certain bands (delta: 1-3 Hz, theta: 4-7 Hz, alpha: 8-13 Hz, beta: 14-30 Hz, gamma: 31-50 Hz), Therefore, DE feature is calculated as the input feature (Duan et al., 2013) following the equation. Here, all 62 channels are used so the final input shape is 62 channels x 5 bands. One session of the total 15 subjects has a total of 27430 samples of data so the shape of all 3 sessions is (27430, 62, 5).

Table 1: Part of self-assessment from different subjects.

subject ID	session	score of each film clip					
		1	2	3	4	5	6
9	1	5	5	5	4	5	3
	2	4	5	5	4	4	4
	3	4	5	4	4	5	5
4	1	1	3	3	1	4	3
	2	2	3	2	2	4	3
	3	2	4	2	2	4	3
11	1	4	4	5	5	3	4
	2	4	4	5	5	4	4
	3	3	4	4	5	4	4
3	1	5	3	4	4	3	4
	2	5	3	5	5	4	4
	3	5	5	3	5	4	4
14	1	4	3	5	4	5	5
	2	5	5	5	4	5	5
	3	4	4	4	3	4	5

2.3 Emotion Labeling

Different from many other EEG emotion dataset, SEED notifies the label of each film clip as the ground truth, considering the selected film clip induced target emotion successfully. This labeling approach works better than traditional self-assessment from subjects. Participants do not have a good understanding of discrete indicators like 10-scale variance, who are not experts in psychology. The point they chose only shows the tendency of their mood contrary to a certain threshold rather than the correct tension to the stimuli. The SEED dataset provides another table that records the self-assessment after watching each film clip. We listed part of the rating score from different subjects and the reaction in each session in Table 1. Each score means variance of emotion in a range from 0 to 5. The table indicates that even when the subject watching the same movie, the rating score changes, but within a relatively stable range. This result suggests the discrepancy of session data exists even within the same subject and using movie labels instead of assessment labels helps build a more robust model.

3 METHODS

3.1 Domain Adaptation

Domain Adaptation is a subset of transfer learning, striving to align data distribution from different domains. It is widely applied in the field of image classification aiming to get high performance in the label-lack domain using label-rich domain data. In reverse to traditional transfer learning methods like fine-tuning, domain adaptation is more likely to be unsupervised machine learning (Pan and Yang, 2009). Assume that we have two datasets which are in different distribution. $\{(X_s^m, Y_s^m)\}$ represents the labeled source dataset and $\{X_t^n\}$ represents the non-label target dataset, where m and n represent the number of sample data in source and target domains respectively. We want to find a feature mapping that transforms the source X_s and target data X_t into a common subspace where X_s' and X_t' have the same data distribution, so that we can achieve good performance in both source domain and target domain.

3.2 Domain-Adversarial Neural Network (DANN)

Overview of GAN: Generative Adversarial Network (GAN) was proposed in 2014 (Goodfellow et al., 2014). Till now, GAN has been extensively applied in many fields like data segmentation, auto image generation, and domain adaptation. In a traditional GAN, two main components are optimized: the discriminator D and the generator G, both of which are composed by neural networks. G and D have different loss functions and the training process minimizes the loss separately, playing a minimax game. The discriminator tries to classify whether the generated data is true opposed to the original input data while the generator tries to fool D by producing generated samples that are similar to the real data. The final objective function is defined mathematically as follow:

$$\begin{aligned} \min_G \max_D \quad & \mathbf{E}_{x\sim p_{\text{data}}}(\log(D(x)) \\ & +\mathbf{E}_{z\sim p_{\text{noise}}} \log(1-D(G(z))) \end{aligned} \tag{2}$$

Since GAN shows remarkable capability in generating controllable data distribution, many researches implement domain adaptation with GAN in many fields. Domain Adversarial Neural Network (Ganin et al., 2016) is such a GAN-based network that trains in an unsupervised approach to get proper feature mapping space. Figure 2 shows the architecture of DANN. There are three main components in the DANN, the feature extractor, the label predictor, and the domain classifier. The feature extractor consists of several layers of CNN, which do feature extraction upon both source and target domain data. To normalize the feature mapping after each CNN layer, there's a Batch-Normalization layer which makes the feature mapping obey a certain distribution. After several CNN+BN layer combinations, there's an average pooling layer to reduce the feature shape. As for the label predictor, it's made by some full connection layers and final softmax classifier. Only source domain data will be sent into the label predictor since no-label target domain data can be optimized by the classification loss. The domain classifier is quite similar to the label predictor while the final softmax classifier does binary classification, taking human-made labels where data from the source has a label of 1 and data from the target has a label of 0. The domain classifier attempts to predict whether the data comes from the source or the target. Besides, a gradient reversal layer is set between the feature extractor and domain classifier to train the whole network in a backpropagation way.

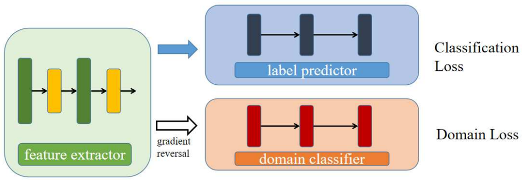

Figure 2: Architecture of DANN.

Assume that $\{X_s, Y_s\}$ represents data from the source and $\{X_t\}$ represents data from the target. The feature extractor processes both two domain data and yields feature mapping, the $\{X_s'\}$ and $\{X_t'\}$. The label predictor predicts emotion labels by optimizing a cross-entropy loss while the domain classifier makes source/target prediction using human-made labels. The gradient reversal layer represents a negative constant and the total loss function will be:

$$\begin{aligned} \tilde{E}(\theta_f, \theta_y, \theta_d) &= \tfrac{1}{n}\textstyle\sum_{i=1} \mathcal{L}_y\left(G_y\left(G_f\left(\mathbf{x}_i;\theta_f\right);\theta_y\right), y_i\right) \\ & -\lambda\left(\tfrac{1}{n}\textstyle\sum_{i=1}^{n} \mathcal{L}_d\left(G_d\left(\mathcal{R}\left(G_f\left(\mathbf{x}_i;\theta_f\right)\right);\theta_d\right), d_i\right)\right) \end{aligned} \tag{3}$$

Training in such an adversarial approach by minimizing the label loss and maximizing the domain loss, a high-performance label predictor and low-performance domain classifier were obtained, meaning the source data was well learned and the distributions were well aligned.

3.3 Multiple Domain Adaptation Network (MDAN)

MDAN is an extension of the DANN, which aligns numerous data distributions in an adversarial training process (Zhao et al., 2018). The difference is that MDAN uses two source domains with two domain classifiers. Data from two different source domains is notated by $\{X_{s1}, X_{s2}\}$ with labels $\{Y_{s1}, Y_{s2}\}$ and target domain data is X_t without label. There are two main purposes of this MDAN: (1) distinguishable enough for label predictor in all source domain data. (2) indistinguishable enough for the two domain classifiers to separate data between source domains and target domain. The architecture of MDAN is in Figure 3.

Similar to the DANN, the feature extractor consists of several layers of CNN and BN layers. It extracts low-level features from the original data space in both source and target domains. Optimization function in task label predictor is:

$$\frac{1}{2n}\sum_{i=0}^{2n}\mathcal{L}_y^i(\theta_f,\theta_d) \tag{4}$$

And the objective function in two domain classifiers is quite similar:

$$\frac{1}{2n}\sum_{i=0}^{n}\mathcal{L}_{d1}^i(\theta_f,\theta_d)+\frac{1}{2n}\sum_{i=n+1}^{2n}\mathcal{L}_{d1}^i(\theta_f,\theta_d) \tag{5}$$

The reason why we use this neural network is that: the discrepancy between EEG data in one single subject cross-phase exists, making it difficult to build a real-time emotion recognition system——the model trained by formerly collected data can't be adapted directly in present collected data. However, in contrast to cross-subject data, we note that cross-phase data has stronger correlation which can be strengthened by MDAN.

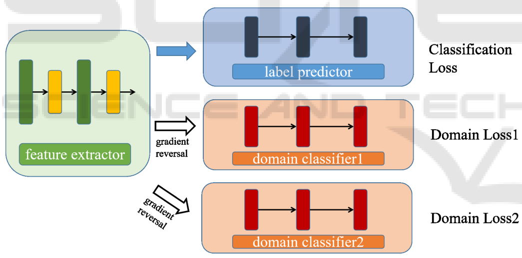

Figure 3: Architecture of MDAN.

4 EXPERIMENT EVALUATION

In this section, the effectiveness of DANN is described as well as the no-adaptation method and single domain adaptation method. All the experiments were performed with SEED dataset. There are in total 3 sessions data as independent domains for evaluation. This helped us investigate whether these DA methods overcome the non-stationarity of EEG signals. One session refers to all subjects exposed to stimuli once and 3 sessions were acquired by repetition at one week intervals. One significant contribution of the SEED dataset is that each subject repeated the same experiments three times. We have a way to further inspect the spatial-relationship of the affective state in each individual participant. Also, since

Table 2: Best accuracy (%) of MDAN and DANN.

Source /Target	no-DA	MDAN	DANN (s1)	DANN (s2)
1+2/3	72.40	85.84	81.32	74.80
1+3/2	70.72	84.89	82.09	74.32
2+3/1	73.01	83.64	79.44	81.76

SEED is a widely used dataset and some session-independent research has been done already, we made comparison with these methods.

4.1 One Source vs Two Source Domain Comparison

To build a real-time emotion recognition system, we hope to have good classification accuracy in the target domain which has no notation. However, since the discrepancy of different phase data exists, just concatenating different phase data together doesn't make sense, even leading to negative transfer learning. The DANN takes one session as source domains so we choose the best accuracy when trying either of two source domains. In contrast, MDAN utilizes the both two source domains. The classification accuracy is shown in Table 2. We can see a naive combination of different session data didn't have good performance and even one-source domain adaptation (DANN) outperformed a lot. MDAN performed much better than DANN using combined data. The result showed GAN-based domain adaptation has a good effect in aligning data distribution from different phase data and more data shows more improvement in classification accuracy.

Note that in DANN, choosing different source domain made influence to the accuracy to some degree. The reason may relay on change of affective states since session 1 refers to the first time subjects watching the movies and session 3 refers to the third time. Session 2 acts as a intermediary which has closer correlation to session 1 and 3. If we go further into the self-assessment from subjects and calculate the MSE error between every two sessions, we can have a better understanding of the cross-phase relationship. Results in Table 3 accounts for why DANN in 1 to 3 session adaptation performs poor. The assessment between these 2 sessions are much higher than others, showing quite massive transformation happened during the first experiment and the third.

The rating point of one typical subject in whole 3 sessions is plotted in Figure 4. In extreme case like movie 12, 5-scale rating can be thoroughly distinctive, as large as 3 points. Even the ground truth of the film is neural.

Table 3: MSE of self-assessment between every 2 sessions.

Combination	MSE
1 and 2	5.633
2 and 3	8.183
1 and 3	5.917

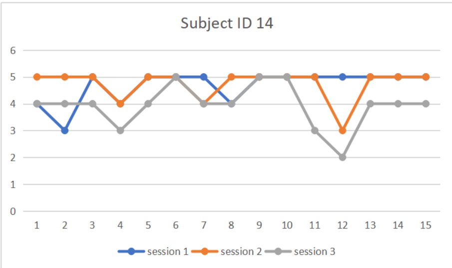

Figure 4: Rating score of subject 14.

We trained the model with a mini-batch of 64 using optimizer Adam. The total training epoch was 1000 until the accuracy curve was converged. The initial learning rate was 0.001 and decayed with running epoch. The accuracy curve of all three domains is plotted in Figure 5. As often the case when training with GAN, the domain accuracy fluctuates with high amplitude and will never be converged. The training accuracy in both source domains goes to 100% rapidly in contract with the target domain accuracy.

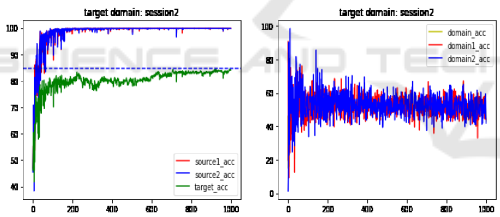

Figure 5: Training Accuracy curve of MDAN (prediction accuracy in the left and domain accuracy in the right).

Furthermore, we can have a look at the data distribution after dimension reduction algorithm of t-SNE (Maaten and Hinton, 2008). The original 310 dimension feature (62 channels x 5 bands) was decomposed into two dimension space. In Figure 6, original feature space are made by several clusters which are distinguishable. These clusters should be data from different subjects and different colors refer to the 3 sessions. The discrepancy across subjects is far more considerable than that across sessions, which makes it harder to transfer. In the right figure, distribution of entire 3 session data mix together instead of clustering separately, showing the MDAN successfully aligned the data distribution.

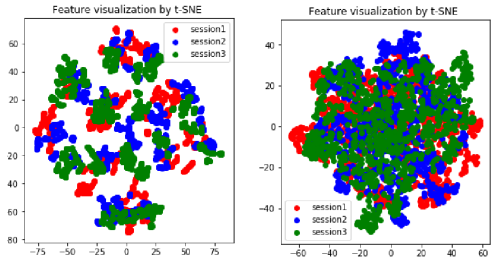

Figure 6: 2-Dimension Data Distribution of MDAN after t-SNE (red,blue,green colors refer to three session data).

5 DISCUSSION

Other session-independent methods are also applied with the SEED dataset, however, most of them can only get medium performance considering the variance of EEG signal. Instead of analysing the correlation of phase data within the same subject, they regard each session data as separate dataset. The purpose of all these methods is to get an aligned subspace between source and target domains. There are two main categories in related works, one is using adversarial training to get an aligned distribution, representative works like the DANN and MDAN. The other is a dimension-reduction based method, hoping to get a no-variance low dimension feature from original high dimension feature such as ASFM. Since we hope to evaluate the transferability between different phases and the enhancement of multiple source domain data, DANN and MDAN can be desirable methods. One state-of-the-art method and some well-known machine learning techniques are introduced here as baselines to the MDAN.

5.1 Adaptive Subspace Feature Matching (ASFM)

In this study (Chai et al., 2017), Principal Component Analysis (PCA) was chosen to do subspace alignment. Among the full space, d largest eigenvalues were selected by PCA and worked as the bases of both source and target subspace, Zs and Zt. Compared to large feature space domain adaptation which requires a large amount of computation, dimension reduction shows better performance in online implementations. The ASFM tries to find the best feature map from source space Zs to target space Zt in linear transformation. The main process of ASFM is in Figure 8.

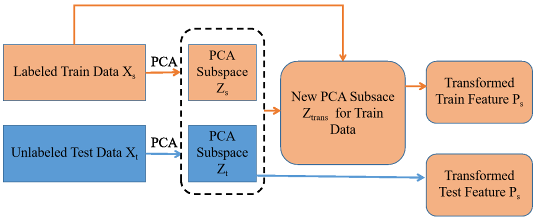

Figure 7: Architecture of ASFM.

5.2 Comparison with Other Methods

There are many other domain adaptation methods to deal with the EEG emotion recognition problem. Here, we compared with the state-of-the-art method, the Adaptive Subspace Feature Matching (ASFM). ASFM tends to be a traditional machine learning algorithm that Linear Regression works as the base classifier. Three famous machine learning methods, the Support Vector Machine (SVM), Linear Regression (LR) and Auto Encoder (AE) are adopted as baseline methods. Table 4 presents mean accuracy and standard deviation of all these methods (Chai et al., 2017). No adaptation methods like SVM shows average accuracy about 70%, higher than channel level (33%) but in a high standard deviation. ASFM achieves best mean accuracy and relatively small standard deviation in most of the cases. However, none of them has utilized the advantage of multiple source domains since all these methods don't have a way to deal with the multiple invariances between several source-target pairs. The MDAN we used here addressed this problem and gave a hint that more source domains help improve the performance. Moreover, both DANN and MDAN are deep neural networks, which are sensitive to hyper-parameter tuning and training epochs. To get a fair result, we repeated the same experiments five times to calculate the mean and standard deviation (SD). The mean accuracy and standard deviation of MDAN is 83.40% and 0.96%. The SD is much smaller than that in ASFM, demonstrating stable performance using MDAN and more potential in real-time recognition task.

Task Group No.	1 → 2	1 → 3	2 → 1	2 → 3	3 → 1	3 → 2	Average
SVM	67.63/12.73	64.21/13.00	67.62/14.73	70.37/19.48	72.69/14.21	66.64/14.38	68.19/11.41
LR	60.64/17.38	60.42/14.61	58.82/13.97	65.30/11.37	63.89/18.44	63.85/15.79	62.15/9.51
GELM * [38]	72.55/10.29	67.22/10.42	75.86/7.71	76.62/15.34	76.28/11.47	78.17/13.41	74.45/8.20
AE [28]	76.66/8.92	75.30/10.83	77.47/11.54	77.20/15.35	77.02/12.81	78.21/13.15	76.98/9.52
TSC [27]	79.85/12.12	80.71/10.70	82.07/8.08	80.24/10.32	79.92/7.71	77.99/11.36	80.13/8.52
TCA [24]	81.56/11.52	79.35/12.26	81.56/6.87	82.83/11.07	80.84/8.00	77.97/13.90	80.68/7.70
TJM [26]	82.43/10.90	80.89/12.06	82.36/5.69	84.04/10.38	80.57/9.97	79.09/11.69	81.56/7.47
SAAE * [34]	**84.47/10.06**	80.04/13.03	84.31/7.21	83.09/9.98	80.20/9.99	78.77/10.49	81.81/7.56
SFM	83.09/10.12	81.61/10.42	84.11/6.23	84.21/8.19	84.62/9.59	82.16/11.06	83.30/7.12
ASFM	84.34/10.18	**82.96/10.49**	**85.46/6.07**	**85.11/8.83**	**86.05/9.66**	**83.75/11.39**	**84.61/6.92**

Figure 8: Results of other DA methods(Chai et al., 2017).

6 CONCLUSION

The results presented above show the MDAN was a promising technique in transferring previous phase domain knowledge to predict in new phase domain data. There are two main contributions of this work. 1). We adopted the MDAN network which is an unsupervised machine learning method to align data distribution among different domains and get approximation level accuracy in comparison with the state-of-the-art method. 2). Instead of traditional subject-independent evaluation, experiments between sessions was performed. Experiment results reveal that transferring knowledge from different phases of data in one same subject helps build a more robust and high-performance real-time EEG system. For future work, we plan to explore what actually affect the emotion state in each subject when repeatedly exposed to a same stimuli. SEED dataset only consisted data from 15 subjects, which is far more than generalization. More subject data is required to validate the effectiveness of MDAN and further feature extraction method should take into consideration as general. As the case in some subject, the rating score switched a lot even when the film was well selected to induce specific emotion. To the end, since the main focus is on cross-phase domain adaptation, further data from same subject is also necessary to certify that more phase data brings about improvement in accuracy.

Table 4: Average accuracy and standard deviations (%) of other DA methods. (* mark means best accuracy).

Target Session	1	2	3	Average
SVM	72.69 /14.21	67.63 /12.73	70.37 /19.48	70.23 /15.47
LR	63.89 /18.44	63.85 /15.79	63.85 /15.79	63.86 /16.67
AE	77.47 /11.54	78.21 /13.41	78.21 /13.15	77.96 /12.70
ASFM*	86.05 /9.66	84.34 /10.18	85.11 /8.83	85.17 /9.56

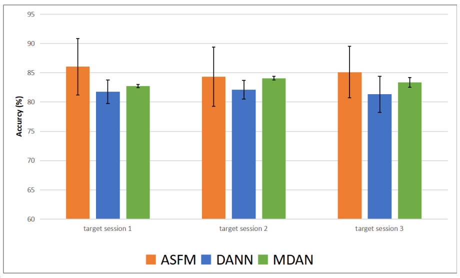

Figure 9: Comparison with SOTA method.

REFERENCES

Bos, D. O. et al. (2006). Eeg-based emotion recognition. *The Influence of Visual and Auditory Stimuli*, 56(3):1–17.

Chai, X., Wang, Q., Zhao, Y., Li, Y., Liu, D., Liu, X., and Bai, O. (2017). A fast, efficient domain adaptation technique for cross-domain electroencephalography (eeg)-based emotion recognition. *Sensors*, 17(5):1014.

Coan, J. A. and Allen, J. J. (2004). Frontal eeg asymmetry as a moderator and mediator of emotion. *Biological psychology*, 67(1-2):7–50.

Duan, R.-N., Zhu, J.-Y., and Lu, B.-L. (2013). Differential entropy feature for eeg-based emotion classification. In *2013 6th International IEEE/EMBS Conference on Neural Engineering (NER)*, pages 81–84. IEEE.

Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., Marchand, M., and Lempitsky, V. (2016). Domain-adversarial training of neural networks. *The Journal of Machine Learning Research*, 17(1):2096–2030.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680.

Guger, C., Ramoser, H., and Pfurtscheller, G. (2000). Real-time eeg analysis with subject-specific spatial patterns for a brain-computer interface (bci). *IEEE transactions on rehabilitation engineering*, 8(4):447–456.

Jayaram, V., Alamgir, M., Altun, Y., Scholkopf, B., and Grosse-Wentrup, M. (2016). Transfer learning in brain-computer interfaces. *IEEE Computational Intelligence Magazine*, 11(1):20–31.

Kahou, S. E., Bouthillier, X., Lamblin, P., Gulcehre, C., Michalski, V., Konda, K., Jean, S., Froumenty, P., Dauphin, Y., Boulanger-Lewandowski, N., et al. (2016). Emonets: Multimodal deep learning approaches for emotion recognition in video. *Journal on Multimodal User Interfaces*, 10(2):99–111.

Koelstra, S., Muhl, C., Soleymani, M., Lee, J.-S., Yazdani, A., Ebrahimi, T., Pun, T., Nijholt, A., and Patras, I. (2011). Deap: A database for emotion analysis; using physiological signals. *IEEE transactions on affective computing*, 3(1):18–31.

Li, X., Song, D., Zhang, P., Zhang, Y., Hou, Y., and Hu, B. (2018). Exploring eeg features in cross-subject emotion recognition. *Frontiers in neuroscience*, 12:162.

Luo, Y., Zhang, S.-Y., Zheng, W.-L., and Lu, B.-L. (2018). Wgan domain adaptation for eeg-based emotion recognition. In *International Conference on Neural Information Processing*, pages 275–286. Springer.

Maaten, L. v. d. and Hinton, G. (2008). Visualizing data using t-sne. *Journal of machine learning research*, 9(Nov):2579–2605.

Margolis, A. (2011). A literature review of domain adaptation with unlabeled data. *Tec. Report*, pages 1–42.

Pan, S. J. and Yang, Q. (2009). A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359.

Zhao, H., Zhang, S., Wu, G., Gordon, G. J., et al. (2018). Multiple source domain adaptation with adversarial learning.

Zheng, W.-L. and Lu, B.-L. (2015). Investigating critical frequency bands and channels for eeg-based emotion recognition with deep neural networks. *IEEE Transactions on Autonomous Mental Development*, 7(3):162–175.

Zheng, W.-L. and Lu, B.-L. (2016). Personalizing eeg-based affective models with transfer learning. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, pages 2732–2738.