EEG-Based Sleep Quality Evaluation with Deep Transfer Learning

Xing-Zan Zhang[1], Wei-Long Zheng[1], and Bao-Liang Lu[1,2,3]*

[1] Center for Brain-Like Computing and Machine Intelligence,
Department of Computer Science and Engineering
[2] Key Laboratory of Shanghai Education Commission for
Intelligent Interaction and Cognition Engineering
[3] Brain Science and Technology Research Center,
Shanghai Jiao Tong University,
800 Dong Chuan Road, Shanghai 200240, China
{zxzsprinkle}@hotmail.com,{weilong,bllu}@sjtu.edu.cn

Abstract. In this paper, we propose a subject-independent approach with deep transfer learning to evaluate the last-night sleep quality using EEG data. To reduce the intrinsic cross-subject differences of EEG data and background noise variations during signal acquisition, we adopt two classes of transfer learning methods to build subject-independent classifiers. One is to find a subspace by matrix decomposition and regularization theory, and the other is to learn the common shared structure with the deep autoencoder. The experimental results demonstrate that deep transfer learning model achieves the mean classification accuracy of 82.16% in comparison with the baseline SVM (65.74%) and outperforms other transfer learning methods. Our experimental results also indicate that the neural patterns of different sleep quality are discriminative and stable: the delta responses increase, the alpha responses decrease when sleep is partially deprived, and the neural patterns of 4-hour sleep and 6-hour sleep are more similar compared with 8-hour sleep.

Keywords: Sleep quality, EEG, Neural pattern, Deep transfer learning

1 Introduction

Sleep quality evaluation has remarkable value in both scientific research and practical applications. Sufficient sleep is of great importance to human daily life, and the study of sleep mechanisms is an important part of brain science. An objective and effective measurement of sleep quality is quite valuable in transportation, medicine, health care, and neuroscience. For example, the tiredness of the drivers due to insufficient sleep imposes a severe threat to the public safety in the transportation industry.

Methods like polysomnography, actigraphy and smart bands have been shown as efficient approaches for evaluating sleep quality. However, these approaches

* Corresponding author

require the subjects to wear equipments such as EEG cap, eye sensors, nose sensors, elastic belt sensors during the whole sleep procedure. The whole-process signal acquisition requirement limits their feasibility in real-world applications.

In this paper, we propose an objective EEG-based approach to classify last-night sleep quality into three categories: poor, normal and good, which excludes whole-process physiological signal acquisition and expert knowledge. Rather than building subject-specific models, which require the collection of labeled data for each subject and thus is time-consuming and unfeasible in practice, we build a subject-independent model and then make inference on the new subjects. The performances of conventional algorithms, when applied to the cross-subject classification tasks, are unsatisfactory because of the intrinsic cross-subject differences and background noise variations. The cornerstone assumption of traditional machine learning methods is that the training data and test data are identically distributed, which is seldom satisfied if not never in sleep quality evaluation due to the variability of cross-subject and cross-session. Our previous work [10] only considers the total sleep time in the experiment setup and the subject-dependent evaluations. In this paper, we refine the experiments by taking deep sleep into consideration and apply transfer learning methods to deal with the cross-subject variations.

Transfer learning approaches have been proved to have the capability to reduce the differences of EEG data across subjects and sessions recently [11] [12]. In this paper, we explore two categories of subject-to-subject transfer learning. One is to find a subspace in Reproducing Kernel Hilbert Space (RKHS) in which the EEG data distributions of different subjects are drawn closer when mapped into this subspace. TCA [7] and ARRLS [5] belongs to this class. The other is to learn the common shared, higher-level structure underlying different categories of sleep quality among different subjects while eliminate the influences of background noise with deep learning, which refers to TLDA [13].

2 Transfer Learning Methods

2.1 TCA-based Subject Transfer

Transfer Component Analysis (TCA) [7] aims to find a set of transfer components across different subjects in a RKHS. In the new space spanned by the transfer components, the data distributions of different subjects are drawn closer, while the data variance properties within each subject are preserved. TCA assumes that there is a kernel function that can simultaneously adapt marginal distribution and conditional distribution. Under this assumption, the transfer components are found by minimizing the Maximum Mean Discrepancy (MMD) between training subject and test subject.

2.2 ARRLS-based Subject Transfer

Adaptation Regularization based Transfer Learning using Regularized Least Squares (ARRLS) [5] simultaneously optimizes the structural risk, the joint distribution and the manifold consistency of two subjects based on the structural risk minimization principle and the regularization theory.

Suppose that $f = \mathbf{w}^T \phi(x)$ is the prediction function where $\mathbf{w}$ is the classifier parameters and $\phi : \mathcal{X} \mapsto \mathcal{H}$ is the kernel function that projects the original feature space into a RKHS space $\mathcal{H}_K$. The prediction function f is learnt by

$$f = \min_{f \in \mathcal{H}_K} \sum_{i=1}^{n} \ell(f(x_i), y_i) + \sigma \|f\|_K^2 + \lambda D_{f,K}(J_s, J_t) + \delta M_{f,K}(P_s, P_t). \qquad (1)$$

where K is the kernel function induced by $\phi(\cdot)$, $\sum_{i=1}^{n} \ell(f(x_i), y_i) + \sigma \|f\|_K^2$ denotes the structural risk minimization of training subject, and $D_{f,K}(J_s, J_t)$ represents the minimization term of marginal distribution and conditional distribution. ARRLS measures the marginal distribution difference with MMD as same as TCA. Since there are no labels in the test subject, the conditional distribution adaption is achieved by the trick of pseudo target labels. The manifold regularization term $M_{f,K}(P_s, P_t)$ is computed by normalized graph Laplacian matrix. And σ, λ and δ are corresponding regularization parameters.

2.3 Transfer Learning with Deep Autoencoders

Deep autoencoders are effective and efficient in learning robust and higher-level representing features. Besides the optimization of individual reconstruction error as the ordinary autoencoders, Transfer Learning with Deep Autoencoders (TLDA) [13] learns a common feature representation shared by the training subject and test subject by explicitly minimizing the symmetrized Kullback-Leibler (KL) divergence, which is a non-symmetric measure of the divergence between two probability distributions, of the two subjects.

The framework of TLDA is shown in Fig. 1, which contains two encoding layers and two decoding layers. The first encoding layer is used to learn robust and high-level features of original EEG features and the second encoding layer encodes the hidden features into labels, which is in fact an classifier based on the softmax regression model. Given nonlinear activation function f (we adopt sigmoid function in our study), for $r \in \{s, t\}$, $\xi_i^r = f(W_1 x_i^r + b_1)$ is the hidden representations, $z_i^r = f(W_2 \xi_i^r + b_2)$ is the encoded labels, and $\hat{\xi}_i^r = f(W_2' z_i^r + b_2')$, $\hat{x}_i^r = f(W_1' \hat{\xi}_i^r + b_1')$ are the corresponding reconstructions of ξ_i^r and x_i^r.

The objective function to be minimized in TLDA can be formalized as

$$J = J_r(x, \hat{x}) + \alpha \Gamma(\xi_s, \xi_t) + \beta L(\Theta, \xi^s) + \gamma \Omega(W, b, W', b'), \qquad (2)$$

where α, β and γ are the trade-off parameters to balance between the four different optimization terms.

The first term in Eq. (2) is the reconstruction error for both training and test subjects, which can be calculated as $J_r(x, \hat{x}) = \sum_{r \in \{s,t\}} \sum_{i=1}^{n_r} ||x_i^r - \hat{x}_i^r||^2$.

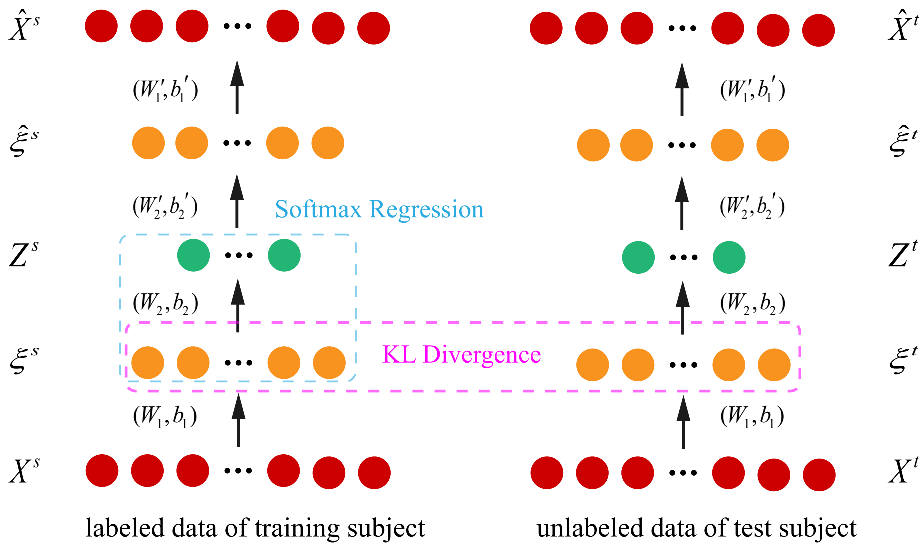

Fig. 1. The framework of TLDA. The encoding and decoding weights are shared by both the training and test subjects. We draw individual networks for each subject to better illustrate the idea that the distributions of two subjects are enforced to be similar in the hidden feature space by minimizing KL divergence.

The second term measures the symmetrized KL divergence among data distributions of two subjects in the embedded space.

$$\begin{aligned} \Gamma(\xi_s, \xi_t) = D_{KL}(E_s||E_t) + D_{KL}(E_t||E_s) = E_s ln(\frac{E_s}{E_t}) + E_t ln(\frac{E_t}{E_s}) \\ E_s = \frac{E_s^{'}}{\Sigma E_s^{'}}, E_s^{'} = \frac{1}{n_s}\sum_{i=1}^{n_s} \xi_i^s, E_t = \frac{E_t^{'}}{\Sigma E_t^{'}}, E_t^{'} = \frac{1}{n_t}\sum_{i=1}^{n_t} \xi_i^t. \end{aligned} \tag{3}$$

The third term denotes the loss function of the final softmax regression classifier. $L(\Theta, \xi^s) = -\frac{1}{n_s}\sum_{i=1}^{n_s}\sum_{j=1}^{c} sgn\{y_i^s = j\} log \frac{e\Theta_j^T \xi_i^s}{\Sigma_{l=1}^c e\Theta_j^T \xi_i^s}$, where Θ_j^T is the j-th row of W_2, sgn is the indicator function.

The last term is the regularization on the complexity of the model parameters which can be formulated as $\Omega(W, B, W^{'}, b^{'}) = ||W_1||^2 + ||b_1||^2 + ||W_2||^2 + ||b_2||^2 + ||W_1^{'}||^2 + ||b_1^{'}||^2 + ||W_2^{'}||^2 + ||b_2^{'}||^2$.

After taking the partial derivatives of the objective Eq. 2 with respect to W_1, b_1, W_2, b_2, $W_1^{'}$, $b_1^{'}$, $W_2^{'}$, $b_2^{'}$, we apply the the gradient descent methods to calculate the final weight matrices.

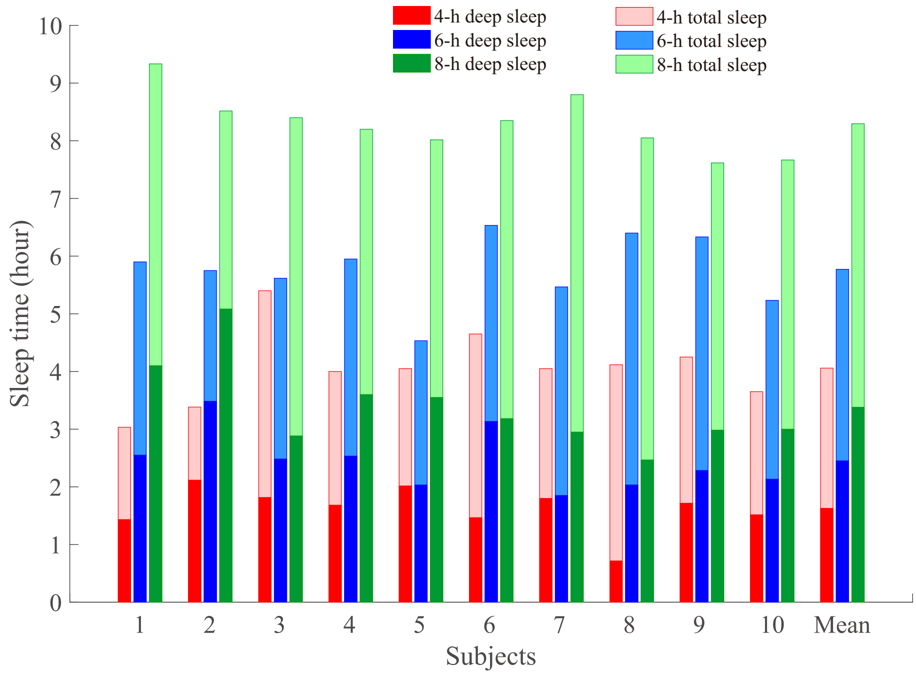

Fig. 2. The total sleep time and deep sleep time of all 10 subjects recorded by the smart bands.

3 Experiments

3.1 Deep Sleep Time

According to the findings of National Sleep Foundation (NSF), 8-hour sleep presents a high sleep quality for adults people, and a sleep time of less than 4 hours means an awful sleep quality [4]. Among whole night sleep procedure, the deep sleep part is the period that human brain gets a full rest, thus is the main factor that counts for sleep quality [2].

Based on the results of sleep medicine, we take 4-hour sleep, 6-hour sleep and 8-hour sleep with increasing deep sleep time as poor, normal and good in terms of the sleep quality in our study. And the sleep time and wake up time for three experiments are 3:00-7:00, 1:00-7:00 and 23:00-7:00, respectively.

3.2 Subjects and EEG data acquisition

Ten graduate and undergraduate students (six males and four females, age range: 21-26, mean: 23.57, std: 1.62) with self-reported healthy conditions and regular daily routines participate in the experiments. Each subject performs 4-hour, 6-

hour and 8-hour experiments wearing smart bands[4] with an interval of two days and declines coffee, drugs, alcohols and other medicines that may disturb human sleep during the experiments. Fig. 2 shows the total sleep time and deep sleep time of all subjects recorded by the smart bands. The EEG signals are recorded for 30 minutes in each experiment with a 62-channel electrode cap according to the international 10-20 system using the ESI NeuroScan system at a sampling rate of 1000 Hz. During the data acquisition procedures, the subjects are required to stare at a green dot on the screen and count numbers to keep a peaceful state. We used the mean signals of all 62 electrodes as the reference.

3.3 Feature Extraction

The raw EEG data are first down-sampled to 200 Hz and fed to a bandpass filter (0-50 Hz). Then the Infomax [1] denoising algorithms is applied to eliminate the noise and artifacts. We adopt DE feature [8] which is calculated in five frequency bands (delta: 1-3 Hz, theta: 4-7 Hz, alpha: 8-13 Hz, beta: 14-30 Hz, gamma: 31-50 Hz) using a Short-Time Fourier Transform with 200-point windows. Therefore, a 310 dimensional DE vector is extracted in each second. The DE features are further smoothed by the LDS algorithm [9] and normalized between 0 and 1 before traininging classifiers. Finally, there are 1800 data samples for each experiment, 5400 data samples for each subject, and 54000 data samples for all subjects.

3.4 Classifier traininging Details

A leave-one-subject-out cross validation scheme is applied for the evaluation. Each time, the 5400 samples from one subject without labels are treated as test data and the 48600 samples from the rest 9 subjects with labels are treated as training data. For TCA and ARRLS, it is impracticable to include all the available data due to limits of memory and time cost. Therefore, we randomly select 1/5 samples from 9 subjects (9720) as the traininging data each time.

We use SVM with linear kernel and $C = 0.01$ as the baseline. In TCA, $\mu = 1$ and the optimal dimension is 30 by searching from 5 to 100 with step 5. We adopt line search rather than grid search to avoid tremendous computation in ARRLS and TLDA. The best configuration for ARRLS is $p = 10$, $\sigma = 0.1$, $\lambda = 10$, $\gamma = 1$, and linear kernel. In TLDA, since the objective function is not convex, we first run Sparse Auto-Encoder (SAE) on training and test data, and initialize the weight matrices with the output of SAE to achieve a better local optimal solution. The optimal parameters are $k = 10$, $\alpha = 5$, $\beta = 1$ and $\gamma = 10^{-7}$.

4 Results and discussion

The classification accuracies of baseline SVM, TCA, ARRLS and TLDA for all 10 subjects are shown in Fig. 3. Accuracy means and standard deviations

[4] We use Mi band 2, the website is https://www.mi.com/shouhuan2

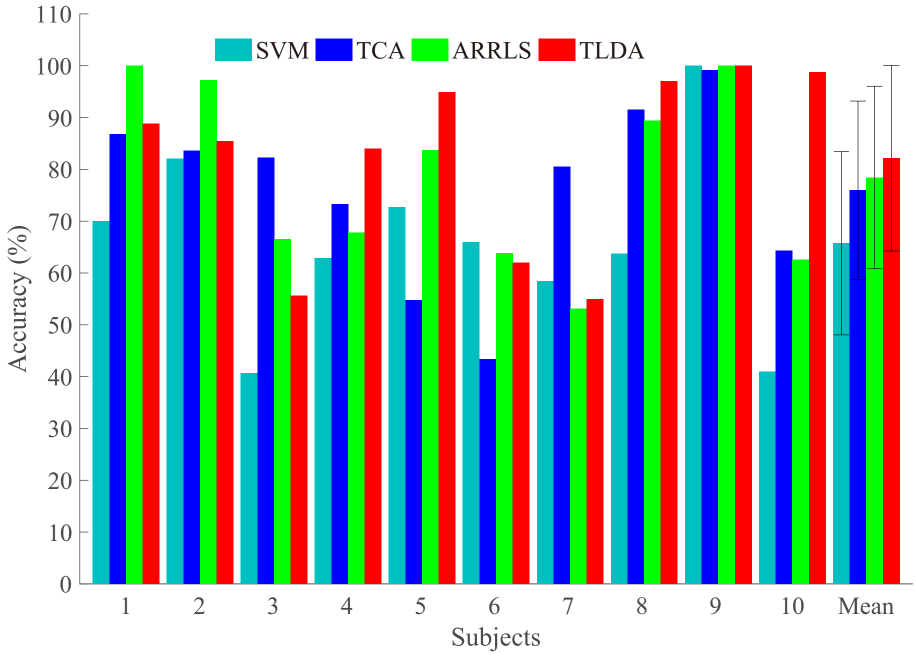

Fig. 3. Accuracy comparation of SVM, TCA, ARRLS and TLDA for each subject.

are 65.74%, 75.96%, 78.44%, 82.16% and 17.69%, 17.23%, 17.61%, 17.91%, respectively. Firstly, all three transfer methods have promotions compared with the baseline SVM. The promotions verify that transfer learning methods are effective at promoting generic classifier with the capability of capturing the underlying common structure shared by different subjects while eliminating sleep quality unrelated noise.

Among the transfer methods, TLDA outperforms other approaches with the highest accuracy of 82.16%. The reasons that TLDA achieves a better accuracy are two folds. The first one is that TLDA learns the mapping functions and hidden features from the trainingging and test data sets with impressive fitting capability, while TCA and ARRLS find the projection function with pre-defined kernels, which is seldom optimal for the data sets we are dealing with. The second reason lies in that TLDA learns the new features and classifiers altogether, while TCA does them in two steps separately.

The DE pattern comparison for three kinds of sleep quality on five bands of all subjects are shown in Fig. 4. The experimental results present that neural structures of different categories of sleep quality are discriminative and stable: the delta responses increase and the alpha responses decrease when sleep is partially deprived. Our findings are in accord with the studies from Borbely *et al.* [3] and Lorenzo *et al.* [6].

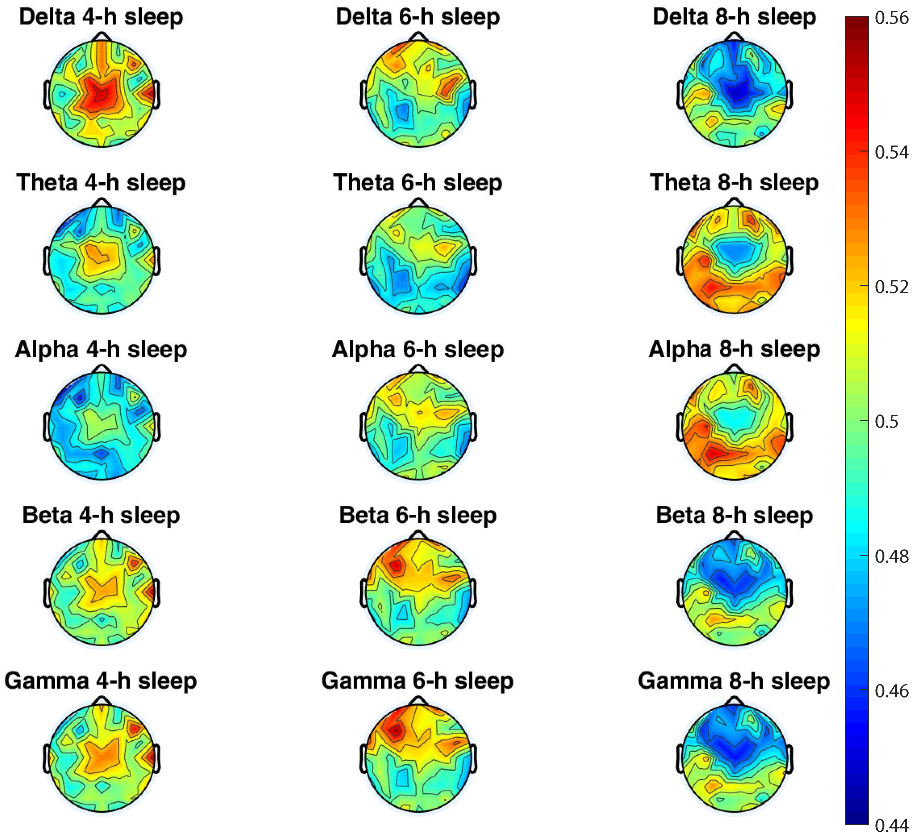

Fig. 4. Neural patterns of three kinds of sleep quality: 4-hour (poor), 6-hour (normal) and 8-hour (good).

Fig. 5 presents the confusion matrices. By examining the accuracies of different categories of sleep quality, we find that 8-hour sleep can be more effectively recognized with the comparatively high accuracy of 89%, while 4-hour and 6-hour sleep are more inclined to intertwine with each other with misclassification rates of 19% and 17%, respectively, in TLDA. This observation is consistent in SVM, TCA and ARRLS, which indicates that the neural patterns of 4-hour and 6-hour sleep are more similar. These findings can again be verified in Fig. 4 which shows that the DE patterns of 4-hour sleep and 6-hour sleep are much closer than the patterns of 8-hour sleep.

5 Conclusion

In this paper, we have adopted deep transfer learning approaches to build a subject-independent model to classify three kinds of sleep quality: poor, normal

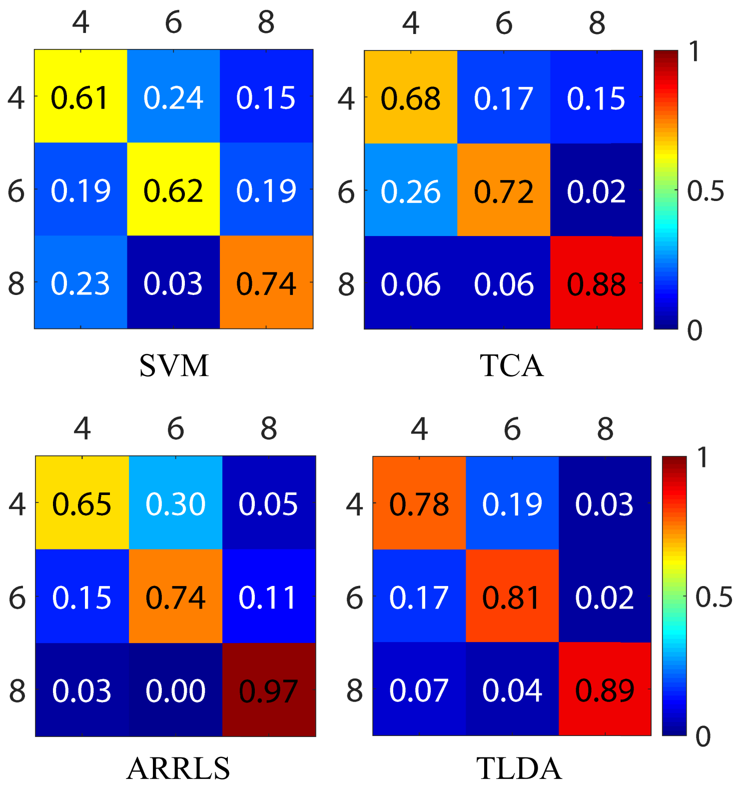

Fig. 5. The confusion matrices of SVM, TCA, ARRLS and TLDA for all subjects.

and good. Among the four evaluated methods (SVM, TCA, ARRLS and TLDA), the TLDA algorithm achieves the best performance with mean accuracy of 82.16%. Our experimental results demonstrate that the neural structures of different sleep quality among people are discriminative and stable: the energy of delta bands has an increasing trend while the alpha responses are depressed when sleep is partially deprived. The results also indicate that the neural patterns of 4-hour sleep and 6-hour sleep are more similar compared with 8-hour sleep.

Acknowledgments. This work was supported in part by grants from the National Natural Science Foundation of China (Grant No. 61673266 and No. 61272248), the National Basic Research Program of China (Grant No. 2013CB329401), the Major Basic Research Program of Shanghai Science and Technology Committee (Grant No. 15JC1400103), ZBYY-MOE Joint Funding (Grant No. 6141A02022604), and the Technology Research and Development Program of China Railway Corporation (Grant No. 2016Z003-B).

References

1. Bell, A.J., Sejnowski, T.J.: An information-maximization approach to blind separation and blind deconvolution. Neural Computation 7(6), 1129–1159 (1995)
2. Berry, R.B.: Fundamentals of sleep medicine. Elsevier Health Sciences (2011)
3. Borbely, A.A., Baumann, F., Brandeis, D., Strauch, I., Lehmann, D.: Sleep deprivation: effect on sleep stages and EEG power density in man. Electroencephalography and Clinical Neurophysiology 51(5), 483 (1981)
4. Hirshkowitz, M., Whiton, K., Albert, S.M., Alessi, C., Bruni, O., DonCarlos, L., Hazen, N., Herman, J., Katz, E.S., Kheirandish-Gozal, L., et al.: National Sleep Foundation's sleep time duration recommendations: methodology and results summary. Sleep Health 1(1), 40–43 (2015)
5. Long, M., Wang, J., Ding, G., Pan, S.J., Philip, S.Y.: Adaptation regularization: A general framework for transfer learning. IEEE Transactions on Knowledge and Data Engineering 26(5), 1076–1089 (2014)
6. Lorenzo, I., Ramos, J., Arce, C., Guevara, M.A., Corsi-Cabrera, M.: Effect of total sleep deprivation on reaction time and waking EEG activity in man. Sleep 18(5), 346–54 (1995)
7. Pan, S.J., Tsang, I.W., Kwok, J.T., Yang, Q.: Domain adaptation via transfer component analysis. IEEE Transactions on Neural Networks 22(2), 199–210 (2011)
8. Shi, L.C., Jiao, Y.Y., Lu, B.L.: Differential entropy feature for EEG-based vigilance estimation. In: 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. pp. 6627–6630. IEEE (2013)
9. Shi, L.C., Lu, B.L.: Off-line and on-line vigilance estimation based on linear dynamical system and manifold learning. In: 32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society. pp. 6587–6590. IEEE (2010)
10. Wang, L.L., Zheng, W.L., Ma, H.W., Lu, B.L.: Measuring sleep quality from EEG with machine learning approaches. In: International Joint Conference on Neural Networks. pp. 905–912. IEEE (2016)
11. Zhang, Y.Q., Zheng, W.L., Lu, B.L.: Transfer components between subjects for EEG-based driving fatigue detection. In: International Conference on Neural Information Processing. pp. 61–68. Springer (2015)
12. Zheng, W.L., Lu, B.L.: Personalizing EEG-based affective models with transfer learning. In: Proceedings of the 25th International Joint Conference on Artificial Intelligence. pp. 2732–3738 (2016)
13. Zhuang, F., Cheng, X., Luo, P., Pan, S.J., He, Q.: Supervised representation learning: Transfer learning with deep autoencoders. In: Proceedings of the 24th International Joint Conference on Artificial Intelligence. pp. 4119–4125 (2015)