35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

Feature Consistency-based Model Adaptation in Session-to-Session Classification: A Study Using Motor Imagery of Swallow EEG Signals

Huijuan Yang, Cuntai Guan, Kai Keng Ang, Chuanchu Wang, Kok Soon Phua,
Christina Tang Ka Yin and Longjiang Zhou

***Abstract*— The performance degradation for session to session classification in brain computer interface is a critical problem. This paper proposes a novel method for model adaptation based on motor imagery of swallow EEG signal for dysphagia rehabilitation. A small amount of calibration testing data is utilized to select the model catering for test data. The features of the training and calibration testing data are firstly clustered and each cluster is labeled by the dominant label of the training data. The cluster with the minimum impurity is selected and the number of features consistent with the cluster label are calculated for both training and calibration testing data. Finally, the training model with the maximum number of consistent features is selected. Experiments conducted on motor imagery of swallow EEG data achieved an average accuracy of 74.29% and 72.64% with model adaptation for Laplacian derivates of power features and wavelet features, respectively. Further, an average accuracy increase of 2.9% is achieved with model adaptation using wavelet features, in comparison with that achieved without model adaptation, which is significant at 5% significance level as demonstrated in the statistical test.**

***Index Terms*— model adaptation, clustering, cluster impurity, feature consistency**

I. INTRODUCTION

Motor imagery is expected to improve the functional recovery after stroke [1], hence motor imagery-based brain computer interface has been used for stroke rehabilitation[2][3]. The assumption is that motor imagery activates similar pathways as that of executed movements. Most machine learning methods work well with the assumption that the training and testing data are drawn from the same feature space or follow the same distribution [4]. However, this assumption is usually violated due to the non-stationarity of the EEG signals. This is due to the changes in the electrode locations and mental states of the subjects with long duration of experiments; movements of the eyes and muscles during experiments and the drying-up of the gels. Most importantly, the non-stationarity is also caused by the long breaks between two sessions, and the visual feedbacks during online feedback sessions. Existing methods for tackling the non-stationarity includes: covariate shift adaptation [5], [6]. The effects of non-class related non-stationarities in EEG during BCI sessions performed with motor imagery tasks are analyzed, subsequently, the parameters of a linear classifier without label information are adapted [6]. "ReTraining" and "ReBias" are proposed to recalculate the decision boundary or adjust the bias for onine feedback sessions [7] by utilizing the label information in the feedback. The results show that the simple "ReBias" is an effective method to correct the shift in session-to-session classification without recomputing the decision boundary. In this paper, we assume that a small amount of online calibration data is available to be used to adapt the model obtained from calibration for the classification of testing data. For this, the motor imagery of swallow EEG data collected are used to test the algorithm.

Dysphagia often occurs in acute stroke patients [8], which is the inability to swallow or difficulty in swallowing caused by stroke or other neuro-degenerative diseases [8]. Traditional methods used to treat swallowing disorders are ususally based on dietary changes, thermal stimulation, tongue strengthening exercises and pharyngeal maneuvers. Recently, neuro-muscular stimulation has been employed [9]. Strength training exercise may drive neural plasticity changes and muscular adaption [10]. In general, the adaption only occurs when the exercise efforts beyond the level of usual activity. An effortful swallow maneuver and tongue-holding maneuver can be used to elicit increased muscular effort and increase muscular work. These methods require the personal assistance from the therapists, hence, repetitive training requires high cost. We propose to use motor imagery of swallow to train the subjects in rehabilitation. For this, detection of motor imagery of swallow is important. This paper addresses the non-stationarity of the EEG signals between two offline sessions, or from offline calibration session to online feedback session. We propose a novel method for transfer learning of EEG signals in different sessions by considering the consitency of features between the training and calibration testing data. The assumption is that a suitable model can always be found to bring the distribution of features of training and testing data close to each other, so that a better performance can be achieved.

II. OUR PROPOSED METHOD

A. Feature Extraction Methods

Two methods are employed in feature extraction. The dual-tree complex wavelet transform (DT-CWT) [11] is employed in the first method due to its good localization in both time and frequency, shift-invariant, better direction selectivity and perfect reconstruction. Assume the EEG signal is decomposed into J levels. The features mainly consist of: power energy of the coefficients (F_w), the phase information at each level and direction (F_p), the coarse representation of the EEG

The authors are with Institute for Infocomm Research, Agency for Science, Technology and Research (A*STAR), Singapore 138632. Email: {hjyang, ctguan, kkang, ccwang, ksphua, kytang and zhoul}@i2r.a-star.edu.sg

978-1-4577-0216-7/13/$26.00 ©2013 IEEE

signals (F_a), and the statistical features (e.g., mean, variance, skewness and kurtosis) of power and phase. These features are effective in detecting the event-related synchronization and desynchronization (ERS/ERD) in brain EEG rhythms during preparation and performing motor imagery of swallowing. The power ($F^w_{s,d}$) and phase ($F^p_{s,d}$) at level (s) and direction (d) are given by

$$F^w_{s,d} = \sum_{i=1}^{n_s} A^f_{s,d}(i) \quad (1)$$

$$F^p_{s,d} = \sum_{i=1}^{n_s} P^f_{s,d}(i) \quad (2)$$

$$F^a_{s,d} = C_{J+1,d,1} \quad (3)$$

where n_s is the length of the coefficients at level s and direction d; $A^f_{s,d}$ and $P^f_{s,d}$ are calculated by

$$A^f_{s,d} = (C_{s,d,1})^2 + (C_{s,d,2})^2 \quad (4)$$

$$P^f_{s,d} = atan(\frac{C_{s,d,2}}{C_{s,d,1}+\epsilon}) \quad (5)$$

where ϵ is a small constant used to prevent the denominator becoming zero; $atan(x)$ is the arctangent of element x; $C_{s,d,r}$, where s ($s\in\{4,3,2,1\}$), d ($d\in\{1,2\}$) and r ($r\in\{1,2\}$) denote level of decomposition, directions, and real and imaginary parts of the complex coefficient $C_{s,d,r}$, respectively. The "asymmetry" and "peakedness" of the probability distribution of coefficients at level s and direction d are measured by skewness (S_{kw}) and kurtosis (K_{ur}) by

$$A^{sw}_{s,d} = S_{kw}(A^f_{s,d}) \quad (6)$$

$$P^{sw}_{s,d} = S_{kw}(P^f_{s,d}) \quad (7)$$

$$A^{cu}_{s,d} = K_{ur}(A^f_{s,d}) \quad (8)$$

$$P^{cu}_{s,d} = K_{ur}(P^f_{s,d}) \quad (9)$$

These features are further normalized and concatenated as the final features.

In the second feature extraction method, the band powers of laplacian derivatives (LAD) of the chosen electrode are used as the features. It is generated by subtracting the powers of neighboring four or two electrodes from that of the considered electrode, as illustrated in Fig. 1. Let's firstly denote the signal as: $S_e(k,n,m)$, where k=1,2,...,N_r, n=1,2,...,N_c and m=1,2,...,N_s represent indexes of the trials, channels and samples, respectively. The signal is firstly divided into N_f frequency bands, starting from 4Hz to 44Hz, with the bandwidth of each frequency and overlapping of two frequency bands as 4Hz and 2Hz, respectively. The signal is band-pass filtered by Chebyshev filter, with the resultant filtered signal denoted as S_f. The band power for the kth trial, nth channel in frequency band f_s is calculated by

$$P_w(k,n,f_s) = 10 * log_{10}(\sum_{m=1}^{N_s} S_f(k,n,m) * S_f(k,n,m)) \quad (10)$$

Let's denote the location indexes of the current processing electrode and its neighboring electrodes as: n_i and n_{ij},

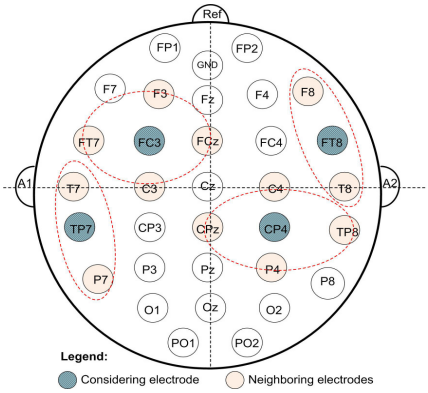

Fig. 1. Illustration of the laplacian derivatives of power features for channels "FC3" and "CP4"' (4 neighbors), and "TP7" and "FT8" (2 neighbors).

where j=1, 2,...,q denotes q neighboring electrodes. The LAD power of the current electrode is calculated by

$$P^d_w(k,n_i,f_s) = P_w(k,n_i,f_s) - P^n_w(k,\hat{n}_{ij},f_s) \quad (11)$$

$$P^n_w(k,\hat{n}_{ij},f_s) = \frac{1}{q}\sum_{\mathrm{j}=1}^{q} P_w(k,n_{ij},f_s) \quad (12)$$

$$\bar{P}^d_w(k,n_i) = \frac{1}{N_f}\sum_{f_s=1}^{N_f} P^d_w(k,n_i,f_s) \quad (13)$$

The final LAD power features ($\bar{P}^d_w(k,n_i)$) are the averaged power features across frequency bands.

B. Model Adaptation

The non-stationarity of the EEG signals leads to poor performance in session-to-session classification. Let's firstly assume that training data are partitioned into different partitions. Without loss of generality, the different partitions produced during cross-validation in calibration stage are utilized. The training model generated based on the r times and n folds cross validation (denoted as M_{dl}) is given by

$$M_{dl}(r,n) = M_g(F_a, I_{tr}(r,n), I_{te}(r,n)) \quad (14)$$

where $I_{tr}(r,n)$ and $I_{te}(r,n)$ denote the indexes of the training and testing data in cross-validation, respectively; F_a is feature vector and $M_g()$ is used to generate the training models based on the random partitions for subsequent optimal model selection. Two assumptions are made: 1) a suitable model can always be found to bring the distribution of testing data close to that of a selected subset of training data. 2) a small amount of labeled calibration testing data (e.g., 10-40 trials) collected on the same day as that of the testing data are available. The model is selected by measuring the feature consistency between the training and calibration testing data, in a cluster that is of minimum impurity. The detail steps for model selection are described as follows.

1) Cluster the whole set of features (F_t) into two clusters such that the within cluster errors are minimized. *K-means clustering* is chosen and the indexes of the ith cluster ($I_c(i)$)

are obtained by

$$I_c(i) = \arg \min_{C_s} \sum_{i=1}^{n_c} \sum_{F_t(j) \in C_s(i)} ||F_t(j) - u_s(i)||^2 \quad (15)$$

where C_s=$\{C_s(1),C_s(2),...,C_s(n_c)\}$ and $u_s(i)$ denote the total set of clusters and the feature mean of the ith cluster; j is the index of the feature vector F_t, where F_t is the feature set consisting of the training data from partition (r, n) and features of the calibration testing data, i.e., F_t=$F_{tr}(r, n)||F_{tc}$; F_{tc} denotes the features of online calibration testing data.

2) Calculate the number of features in class "k" for training (N_{tr}^k) and calibration testing data (N_{te}^k) in ith cluster by

$$I_{tr}(i,k) = I_{tr}(i)|(Y_{tr}(I_{tr}(i)) = k) \quad (16)$$
$$N_{tr}^k(i) = |I_{tr}(i,k)| \quad (17)$$
$$I_{te}(i,k) = I_{te}(i)|(Y_{tr}(I_{te}(i)) = k) \quad (18)$$
$$N_{te}^k(i) = |I_{te}(i,k)| \quad (19)$$

where k=$\{0, 1\}$ denotes class "0" and "1"; $I_{tr}(i, k)$ and $I_{te}(i, k)$ denote the feature indexes in kth class of ith cluster for training and calibration testing data, respectively; $Y_{tr}(i)$ and $Y_{te}(i)$ are class labels in cluster i for training and calibration testing data, respectively; $|x|$ gives cardinality.

3) Computer the dominant label ($L_c(i)$) for cluster i based on the majority of class labels of training data by

$$L_c(i) = \begin{cases} 0 \text{ if } N_{tr}^0(i) \geq N_{tr}^1(i) \\ 1 \text{ otherwise} \end{cases} \quad (20)$$

4) Calculate the number of consistent features for training (N_{tr}^c) and calibration testing (N_{te}^c) data, which is given by

$$N_{tr}^c(i) = N_{tr}^k(i)|Y_{tr}(I_{tr}(i,k)) = L_c(i) \quad (21)$$
$$N_{te}^c(i) = N_{te}^k(i)|Y_{te}(I_{te}(i,k)) = L_c(i) \quad (22)$$

5) Calculate *cluster impurity* of the ith cluster ($C_I(i)$) based on the numbers of features in each class of the training data, which is given by

$$C_I(i) = \frac{min(N_{tr}^0(i), N_{tr}^1(i))}{max(N_{tr}^0(i), N_{tr}^1(i))} \quad (23)$$

obviously, $C_I(i)$∈[0 1] always holds.

6) Repeat steps 2 to 5 till all the clusters are processed. The cluster with the minimum impurity in each iteration is chosen by

$$\hat{i} = \arg \min_i (C_I(i)) \quad (24)$$

Thereafter, the total number of consistent features in the chosen cluster ($\hat{i}$) between training ((n, r)th model) and calibration testing data $N_c(\hat{i}, n, r)$ is given by

$$N_c(\hat{i}, n, r) = N_{tr}^c(\hat{i}) + N_{te}^c(\hat{i}) \quad (25)$$

7) Finally, the model that has the maximum consistent features between training and calibration testing data is selected by

$$(\hat{n}, \hat{r}) = \arg \max_{(n,r)} (N_c(\hat{i}, n, r)) \quad (26)$$

It should be noted that all the models are generated during the training stage. The best model is selected to better classify the testing data.

III. Experimental Protocol and Results

The experiments consist of two tasks, i.e., motor imagery of swallow (MI-SW) and idle. In motor imagery of swallow, the subjects are instructed to imagine swallowing a cup of water, or a cup of juice, some bolus or pureed food, etc. The subjects are advised not perform any task and not to close their eyes in idle state. 10 healthy subjects without any history of respiratory, swallowing or neurological disorder participated in the experiments. Each session consists of two runs and each run consists of 40 trials for each action, yielding a total of 160 trials per session. The timing scheme is shown in Fig. 2. A resting state shown by a progress bar is firstly appeared on the black screen. A short acoustic tone was presented at 0s, followed by a preparation of 2s shown as "+". After that, the cue in the form of a virtual character for swallowing or a filled circle for idle will be shown for 3 seconds. With the disappearing of the visual cue, the subject will start to perform required action for 12 seconds. At the end of each trial, there will be 6 seconds rest time. Each trial lasts for 23 seconds. The EEG and EMG

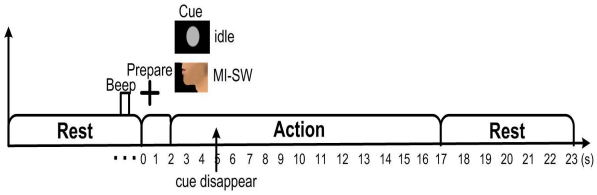

Fig. 2. Experimental protocol for motor imagery of swallow.

data are bandpass filtered between 0.5Hz and 100Hz, and the sampling rate was set to 250Hz. Four channels of HEOL, HEOR, VEOU and VEOL are used for EMG recording by taping the two pairs of electrodes beneath the skin of the submental and infrahyoid muscle groups. The EMG signal is used to monitor the muscle movements of the subjects. The other 32 channels are used for EEG recordings.

Experiments are conducted based on the motor imagery of swallow EEG data for 6 selected subjects whose cross-validation accuracies are above 62% in both sessions. To test the session-to-session classification accuracy, the model trained using features in one session is subsequently employed to classify the features in another session and vice versa. Support vector machines with linear kernel is used as the classifier. A small amount of calibration testing data ranging from 10-40 trials are employed to select the model, which is subsequently used to classify the rest of testing data. The results of using both wavelet features and LAD power features are shown in Table I. It is noted that the best time segment of 2 seconds are selected from 7 non-overlapping time segments in the interval of [2.5 10.5]s (from onset of action/idle). The results shown in the table demonstrate the effectiveness of our proposed feature consistency-based model selection technique. An average accuracy increase across subjects of 2.9% is achieved using wavelet features. Further, the use of model adaptation yields all 12 best performance for 6 subjects in 2 sessions as shown in boldface

TABLE I

SESSION-TO-SESSION CLASSIFICATION ACCURACIES FOR MOTOR IMAGEY OF SWALLOW WITH/WITHOUT MODEL ADAPTATION

Sub./Sess	**Power Features ($\mathcal{A}_{ss}$)**					Sub./Sess	**Wavelet Features ($\mathcal{A}_{ss}$)**				
		Adaptation (trials)						Adaptation (trials)			
	No adaptation	10	20	30	40		No adaptation	10	20	30	40
lj/1	73.13	74.00	74.29	74.62	**75.00**	lj/1	71.25	72.00	72.14	**76.15**	74.17
lj/2	**79.38**	79.33	75.00	78.46	78.33	lj/2	75.63	**76.67**	76.43	74.62	75.00
hj/1	**74.38**	74.00	73.57	70.77	72.50	hj/1	75.63	74.00	72.14	75.38	**77.50**
hj/2	76.88	78.00	**79.29**	77.69	78.33	hj/2	78.13	82.00	80.00	**83.85**	82.50
cr/1	77.50	80.67	81.43	83.08	**81.67**	cr/1	77.50	80.67	82.14	**83.08**	82.50
cr/2	85.63	86.00	85.71	85.38	**86.67**	cr/2	81.25	84.67	82.86	84.62	**85.83**
wy/1	68.75	70.67	70.71	**71.54**	70.00	wy/1	69.38	72.67	73.57	**74.62**	72.50
wy/2	79.38	82.00	**82.86**	80.77	80.83	wy/2	68.75	68.67	67.14	**70.00**	68.33
cc/1	63.75	**66.67**	65.71	65.38	**66.67**	cc/1	56.25	58.67	**60.71**	56.92	57.50
cc/2	**70.63**	68.67	70.00	66.92	68.33	cc/2	60.00	61.33	60.00	63.85	**65.00**
mt/1	61.25	64.00	63.57	**70.00**	65.00	zy/1	61.88	64.67	67.14	66.15	**67.50**
mt/2	65.00	65.33	63.57	**66.92**	66.67	zy/2	61.25	**63.33**	61.43	60.77	**63.33**
A_{as}	72.97	74.11	73.81	**74.29**	74.17	A_{as}	69.74	71.61	71.31	72.50	**72.64**
Significance		**	xx	xx	xx	Significance		**	xx	**	**
P value		0.0292	0.246	0.215	0.0677	P value		0.0023	0.0686	0.0032	0.0006

A_{ss}: Session-to-session accuracy (%), A_{as}: Average Accuracy across subjects. The best performance for each subject in each session is shown in bold. xx: not significant; **: significant

in the right portition of table. While using LAD power features, 9 out of 12 achieved best performance using model adaptation and an average accuracy increase of 1.29% across subjects is achieved compared with that of without model adaptation. Note that bias correction is carried out for LAD power features-based session-to-session classification. The bias is corrected so that the two-class proportion is closest to class prior of testing data. A paired sample t-test at significance level of 0.05 is conducted to test the hypothesis that the accuracies obtained using model adaptation and that obtained without model adaptation come from populations with equal means. The results show that the hypothesis is rejected for 3 out of 4 wavelet-based model adaptation and not rejected for 3 out of 4 LAD power feature-based model adaptaton, as shown in the last row of the table. This further demonstrates the significant difference in accuracies obtained by model adaptation using wavelet features compared those achieved without moodel adaptation.

IV. CONCLUSIONS

This paper proposes a novel model adaptation method for session-to-session classification of motor imagery of swallow EEG signals for dysphagia rehabilitation. The whole set of features including the training (from particular model) and calibration testing data are firstly clustered. The cluster that has the highest purity is selected and the number of features for both training and calibration testing data that are consistent with the dominant label in the cluster are calculated. Finally, the model with the largest consistent features is selected. Experiments conducted for EEG data of 6 healthy subjects in two sessions achieve the average accuracy of 74.29% and 72.64% using laplacian derivatives of power features and wavelet features, respectively. Further, significant average accuracy increase of 2.9% is achieved with model adaptation using wavelet features compared with that achieved without model adaptation.

REFERENCES

[1] N. Sharma, V. M. Pomeroy and J.-C. Baron, "Motor Imagery: A Backdoor to the Motor System After Stroke?" *Stroke,* vol. 37, pp. 1941-1952.

[2] J. J. Daly, J. R. Wolpaw, "Brain-Computer Interfaces in Neurological Rehabilitation," *Lancet Neurology*, vol. 7, no. 11, pp. 1032-1043, 2008.

[3] K.K. Ang, C. Guan, K.S.G. Chua, B.T. Ang, C.W.K. Kuah, C. Wang, K.S. Phua, Z.Y. Chin and H. Zhang, "A Large Clinical Study on the Ability of Stroke Patients in Using EEG-Based Motor Imagery Brain-Computer Interface," *Clinical EEG and Neuroscience*, vol. 42, no. 4, pp. 253-258, Oct. 2011.

[4] S. J. Pan and Q. Yang, "A Survey on Transfer Learning," *IEEE Trans. on KNowledge and Data Engineering,* vol. 22, no. 10, pp. 1345-1359, Oct. 2010.

[5] Y. Li, H. Kambara, Y. Koike and M. Sugiyama, "Applications of Covariate Shift Adaptation Techniques in Brain-Computer Interfaces," *IEEE Trans. On Biomedical Engineering,* vol. 57, no. 6, pp. 1318-1324, June 2010.

[6] Vidaurre C, Kawanabe M., von Bnau P., Blankertz B., Mller KR, "Toward unsupervised adaptation of LDA for brain-computer interfaces," *IEEE Trans Biomed Eng.,* vol. 58, no. 3, pp. 587-597, 2011 March.

[7] Shenoy P., Krauledat M., Blankertz B., Rao R.P., Mller K.R., "Towards adaptive classification for BCI," *Journal of Neural Eng.*, vol. 3, no. 1, pp. 13-23, March 1, 2006.

[8] Martin J. McKeowna, Dana C. Torpeyd, Wendy C. Gehmf, "Non-invasive monitoring of functionally distinct muscle activations during swallowing," *Clinical Neurophysiology,* vol. 113, pp. 354-366, 2002.

[9] M. Kiger, C. S. Brown and L. Watkins, "Dysphagia Management: An Analysis of Patient Outcomes Using VitalStim^{TM} Therapy Compared to Traditional Swallow Therapy," Dysphagia, vol. 21, no. 4, pp. 243-253, 2006.

[10] L. M. Burkhead, C. M. Sapienza, and J. C. Rosenbek, "Strength-Training Exercise in Dysphagia Rehabilitation: Principles, Procedures, and Directions for Future Research," *Dysphagia,* vol. 22, pp. 251-265, 2007.

[11] H. Yang, C. Guan, K.K. Ang, C. Wang, K.S. Phua and J. Yu, 2012, Dynamic initiation and dual-tree complex wavelet feature-based classification of motor imagery of swallow EEG signals, Proc. of IJCNN 2012, pp. 1-6, June 10-15, Brisbane, Australia.