Chapter 1

Covariate Shift Detection based Non-Stationary Adaptation in Motor-Imagery based Brain-Computer Interface

Haider Raza[1] *and*
Dheeraj Rathee[2]

Non-stationary learning (NSL) refers to the process that can learn patterns from data, adapt to shifts, and improve performance of the system with its experience while operating in the non-stationary environments (NSEs). Covariate shift (CS) presents a major challenge during data processing within NSEs wherein the input-data distribution shifts during transitioning from training to testing phase. CS is one of the fundamental issues in electroencephalogram (EEG) based brain-computer interface (BCI) systems and can be often observed during multiple trials of EEG data recorded over different sessions. Thus, conventional learning algorithms struggle to accommodate these CSs in streaming EEG data resulting in low performance (in terms of classification accuracy) of motor imagery (MI)-related BCI systems. This chapter aims to introduce a novel framework for non-stationary adaptation in MI-related BCI system based on covariate shift detection (CSD) applied to the temporal and spatial filtered features extracted from raw EEG signals. The chapter collectively provides an efficient method for accounting non-stationarity in EEG data during learning in NSEs.

1.1 Introduction

In EEG-based brain-computer interface (BCI) systems, the majority of the learning algorithms assume, either implicitly, or explicitly that the EEG data have statistically stationary/fixed distributions over different sessions and/or runs of the recording [1]. However, such an assumption is simply not true as the EEG data obtained over different sessions and/or runs ave possess non-stationary characteristics [2, 3]. In real-world BCI applications, non-stationarity is a ubiquitous phenomenon, especially when the system interacts with the dynamic and evolving environments,

[1]Institute for Analytics and Data Science, University of Essex, Colchester, England.
[2]Intelligent Systems Research Centre, Ulster University, Derry∼Londonderry, UK

e.g., driving robotic wheelchair or playing virtual games. Hence, developing machine learning models that are optimized for non-stationary environments (NSEs) is of great need. Machine learning methods are generally categorized into unsupervised, semi-supervised, and supervised learning methods, whereas adaption to non-stationarity is a common practice to all these three categories. Furthermore, the NSL methods are majorly grouped into passive and active approaches. In the passive approach to NSL, it is assumed that the input distribution should be continuously shifting over time. Thus, passive approach based NSL methods adapt to new data distributions in continuous fashion i.e., the model is updated for every new input observation or new batch of input observations from the streaming data. In contrast, an active approach based NSL method uses a shift detection test to detect the presence of shifts in the streaming data, and correspondingly, based upon the time of detected shift, an adaptive action is initiated.

A BCI system aims to provide an alternative means of communication or rehabilitation for the physically challenged population so as to allow them to express their intentions using brain signal modulations without explicit muscular exertion [4]. Non-invasive EEG-based BCI systems acquire neural signals at scalp level, analyse them to evaluate specific features of EEG activity that are related to voluntary imagery/execution tasks, and finally utilise the outcomes as control signals that are further relayed to an output device [5]. Such a system operates typically in two phases, namely the training phase and the evaluation (testing) phase [6]. The EEG signals are acquired through a multichannel EEG amplifier, and a pre-processing step is performed to reduce noise and enhances the signal-to-noise ratio. In the next step, the discriminable features are extracted from the artefact-cleaned signals using feature extraction techniques, such as spatial filtering (e.g., common spatial pattern (CSP), laplacian filtering, current source density) [7, 8, 9]. Further, the optimal features are employed for training the classifier model. With an EEG-based BCI system that operates online in real-time non-stationary/changing environments, it is required to consider the input features that are invariant to shifts of the data or the learning approaches that can track the changes repeating over time. However, due to the non-stationary nature of the brain response characteristics in the EEG signal, it may be difficult to classify the EEG patterns reliably using a conventional classification algorithms used in motor imagery (MI) related BCI systems [1, 6]. The non-stationarities of the EEG signals may be caused by various reasons, such as changes in the user attention level, fatigue, imagery preferences, and/or placement and aging of electrodes [10, 2, **?**]. These non-stationarities cause notable variations or shifts in the EEG signals both during trial-to-trial, and session-to-session transfers [11, 12]. As a result, these variations may often appear as covariate shifts, wherein the input data distributions differ significantly between training and testing phases, while the conditional distribution remains the same[**?**, 13, 14, **?**]. The traditional BCI systems are built upon a conservative approach to NSL i.e., passive approach, where the learning system is updated continuously, assuming that the environment is constantly shifting. Furthermore, both single and ensemble of classifier(s) based learning methods were developed for these passive approaches to improve the MI detection performance.

In contrast, an active approach provides a more intuitive option for NSL learning involving a covariate shift detection (CSD) test to capture the presence of covariate shifts (CSs) in the streaming EEG features followed by an adaptive action based on the correctly detected CSs.

The aim of this chapter is to present a CSD related non-stationary adaptation (CSD-NSA) algorithm for EEG-based BCI systems. Different from the existing methods, the CSD-NSA algorithm is an active approach to learning in NSEs, wherein a CSD test is applied to initiate adaptation by adding new information from the testing data to the existing training data and retrain the updated classifier. This active approach aims at better managing computational resources, by adapting the classifier only when it is necessary i.e., once the data from a novel distribution have to be processed. Specifically, for the detection of CSs in the EEG features, we considered an exponential weighted moving average (EWMA) model based CSD test that reacts to the CSs in non-stationary conditions [15]. At the point of shift detection, the classifier can be adapted either in supervised or unsupervised manner. To assess the performance of the proposed CSD-NSA algorithms (i.e., supervised and unsupervised), experimental evaluations have been performed on publicly available MI related EEG dataset.

1.2 Background

1.2.1 Covariate Shift in EEG Signals

CS is a case where the conditional probability distribution remains the same i.e. $(P_{train}(y|x) = P_{test}(y|x)$, whereas the input data distribution shifts i.e. $(P_{train}(x) \neq P_{test}(x))$, while transitioning from the training to testing stage. A typical example of the CS for ten overlapping frequency bands ([8-12], [10-14],...[26-20] Hz) in the feature set of EEG data is illustrated in Figure 1.1 for the subject $A07$ of BCI competition-IV dataset $2A$ (the description of the dataset is present in section IV). For each plot, the blue solid ellipse and line show the input data distribution $P_{train}(x)$ and the classification hyperplane for training dataset, respectively. Likewise, the red dashed ellipse and dash line show input data distribution $P_{test}(x)$ and the classification hyperplane for test dataset.

1.2.2 Adaptive Learning Methods in EEG-based BCI

To enhance the performance of MI related BCI systems, a large variety of signal processing methods have been proposed to extract features in the temporal and spatial domains to manage the non-stationarity in EEG signals. In the temporal domain, band-power and band-pass based filtering methods have been commonly used [12, 16], whereas in the spatial domain, common averaging, current source density, and CSP-based features were widely explored for the detection of MI related responses [17, 7, 18]. The issue of low classification accuracy with the existing BCI systems has been one of the main concerns in their rather low uptake among people

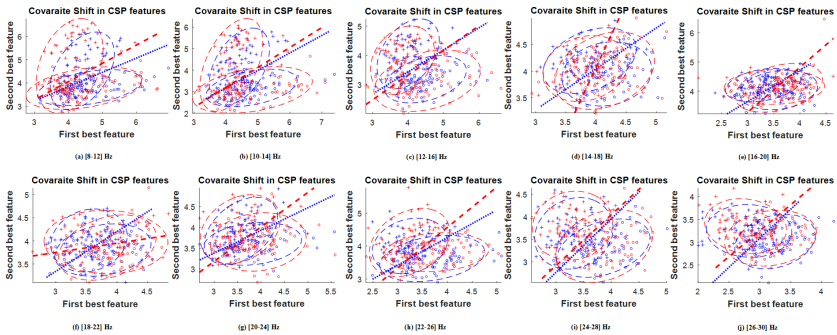

Figure 1.1 *Covariate shift between the training and test distributions of subject* A07 *of BCI competition-IV dataset* 2A. $(a)-(j)$ *illustrates the covariate shift in the [8-12], [10-14],...[26-20] Hz frequency bands. For each plot, the blue solid ellipse and sold line show the input data distribution* $P_{train}(x)$ *and the classification hyperplane for training dataset. Likewise, the red dashed ellipse and dash line show input data distribution* $P_{test}(x)$ *and the classification hyperplane for test dataset.*

with severe physical disability [17].

To tackle the issue of low performance of a BCI system due to the CSs, several adaptive learning algorithms have been proposed to devise adaptive BCI systems with encouraging results, such as Vidaurre et al. [19] have developed an adaptive classifier using an adaptive estimation of information matrix. Next, Shenoy et al. [1] have provided quantified systematic evidence of statistical differences in data recorded during multiple sessions and various adaptive schemes to enhance the BCI performance. Additionally, a covariate shift minimization (CSM) method was proposed for the non-stationary adaptation to reduce the feature set overlap and unbalance for different classes in the feature set domain [13]. More interestingly, Li et al.(2010) have contributed a covariate shift adaptation (CSA) method for the BCI system [10] based on a density ratio estimation technique. Density ratio based CSA is an unsupervised adaptation method that can adapt to the testing session without knowing the true labels. However, there exists a limitation that requires all the testing unlabeled data before starting the testing phase to estimate the importance for the non-stationarity adaptation, which makes this approach impractical in the real-time BCI systems for both communication and rehabilitation purposes.

Furthermore, since the last decade, ensemble based machine learning methods have become popular for NSL, where a set of classifiers is coupled to provide an overall decision. In the EEG-based BCI systems, ensemble learning based methods have been evaluated to improve the classification performance (e.g. bagging, boosting, and random subspace [20]). Impressively, a dynamically weighted ensemble classification (DWEC) method has been proposed, wherein an ensemble of multiple

classifiers is trained on cluster features to handle the issue of non-stationarity adaptation [21]. The DWEC method partitions the EEG data using clustering analysis and multiple classifiers were trained using different partitioned datasets. The final decision of the ensemble was then obtained by appropriately weighting the classification decisions of individual classifiers. In a recent study, the ensemble of common spatial pattern patches (CSPP) has shown a potential for improving the performance of online MI related BCI system [22]. Both single and ensemble of classifier(s) based approaches were developed based on the passive mechanism to improve the MI detection performance. In contrast, an active approach based NSL method in BCI system could offer a possible solution that involve an active shift detection mechanism, which leads to an efficient adaptation to non-stationarity in the streaming EEG features.

1.3 Covariate Shift Detection based Non-Stationary Adaptation (CSD-NSA) Algorithm

1.3.1 Problem Formulation

Given a set of training samples $X^{Tr} = \left\{x_i^{tr}, y_i^{tr}\right\}$, where $i \in \left\{1...n\right\}$ is the number of training data observations, $x_i^{tr} \in \mathbb{R}^d$ (d denotes the input data dimensionality) is a set of input training features drawn from a probability distribution with density $P_{tr}(x)$, and $y_i^{tr} \in \left\{C_1, C_2\right\}$ is a set of training labels, where $y_i^{tr} = C_1$, if x_i^{tr} belongs to class ω_1, and $y_i^{tr} = C_2$, if x_i^{tr} belongs to class ω_2. Here, we have assumed that the input data distribution remains stationary during the training phase. Furthermore, given an unlabelled test input data samples $X^{Ts} = \left\{x_i^{ts}\right\}$, where $i \in \left\{1...m\right\}$ is the number of test data observations, $x_i^{ts} \in \mathbb{R}^d$ is a set of test input features, drawn independently from a probability distribution with density $P_{ts}(x)$. Note that $P_{tr}(x) \neq P_{ts}(x)$ in general, and thus the input distributions may be different during the training and test phases, leading to covariate shifts.

1.3.2 Covariate shift detection (CSD) Test

The CSD test is an unsupervised method for detecting non-stationary changes in the unlabelled testing data (X^{Ts}) during the testing phase [23]. For this study we estimated the CSD parameters (i.e., smoothing constant (λ) and control limit multiplier (L)) during the training phase with an assumption that the training data distribution is different from testing data distribution $(P_{tr}(x) \neq P_{ts}(x))$. Next, during the testing phase, an exponentially weighted moving average $(EWMA)$ model was implemented for the detection of the covariate shifts in the incoming data-stream (trial-based estimation). The EWMA is a type of infinite impulse response filter that applies weighting factors, which decrease exponentially. The weight of each older observation decreases exponentially, however, never reaching zero value. The weighting factor is one of the strengths of the EWMA model. The EWMA control chart overtakes other control charts because it pools the present and the past data together in such a way that even small shifts in the time-series can be detected more

easily and quickly. Thus, the incoming observations are continuously examined to provide a 1-step-ahead prediction (1-SAP). Next, the 1-SAP error was plotted on the control chart and if the estimated error falls outside the control limits (L), the point was said to be a point of covariate shift. The EWMA model can be presented as,

$$z_{(i)} = \lambda x_{(i)} + (1-\lambda) z_{(i-1)} \tag{1.1}$$

where, $z_(i)$ is an EWMA statistics of the current trial, λ is a smoothing constant, which is selected based on minimizing 1-SAP error on the training dataset (X^{Tr}). The selection of the value of λ is a key issue in the CSD test. In particular, for the auto-correlated time series data, it is suggested to select a value of λ that minimizes the sum of the squares of the 1-SAP errors [24]. In the current approach, the value of λ was obtained by testing different values of λ in the range of $[0 : 0.01 : 1]$ on the training dataset. In particular, the CSD test further consists of two stages. In the first stage, it detects the covariate shift in the first principal component extracted from a principal component analysis (PCA) based EEG feature vector of each trial using the EWMA model and issue a covariate shift warning (CSW). In case of positive CSW outcome, a covariate shift validation (CSV) process is performed in the second stage in order to reduce the number of false alarms. During CSV stage, a multivariate two-sample Hotelling's T-Square statistical hypothesis test was performed to compare the two distinct samples with an equal number of observations generated before and at the CSW time point [25]. If the test rejects the null hypothesis, the existence of CS was confirmed during the CSV stage, otherwise, it was considered as a false alarm [23]. This CSD test has been successfully applied in our previous studies for the detection of covariate shifts in EEG features for different covariate shift adaptive methods [15].

1.3.3 Supervised CSD-NSA Algorithm

The Supervised Covariate Shift Detection-Non-Stationary Adaptation (S-CSD-NSA) algorithm combines the aforementioned CSD test and a supervised adaptation (SA) method as described in the Algorithm 1. In the training phase, a support vector machine (SVM) classifier f_s was trained using the dataset X^{Tr}. The total number of detected and validated shifts was denoted by s, whereas in the training phase s was set equal to 0, and the index of the current example is denoted by i. The next step involves the classification of the first trial of the testing data using classifier (f_s). Next, for each new unlabelled trial, the features were monitored to detect a covariate shift using the CSD test. If the test provides positive outcome, then a covariate shift was confirmed in the features of the current trial, otherwise, the trial was classified with the existing classifier f_s. In case of confirmed covariate shift, the value of s was incremented by 1. The next step is to select and store the correctly predicted trials in a buffer $X^{New} = \left\{ \left(x_v^{ts}, y_v^{ts} \right) \right\}_{v=1:l}$, where l is the number of correctly predicted trials before the ith trial in which a CS was most recently detected. Afterward, the training dataset X^{Tr} was merged with X^{New} to update the training dataset. Also, a new classifier f_s was trained on the updated dataset and discard the old classifier f_{s-1}. As the initial classifier was obtained from the training phase, wherein, the initial value of s

Algorithm 1 :Supervised CSD-NSA Algorithm

Input : $X^{Tr} = \{x_i^{tr}, y_i^{tr}\}$, where $i \in \{1...n\}$
: $X^{Ts} = \{x_i^{ts}\}$ where $i \in \{1...m\}$
Output : Y^{Ts}
TRAINING:
$Tr(f_0, X^{Tr})$
Set $s = 0$, where s counts the number of covariate shift detection (CSD)
TEST:
Start evaluation using testing dataset X^{Ts}
Set $i = 1$
$\hat{y}_i = f(x_i)$
for $i = 2$ to m **do**
  $p \leftarrow CSD(x_i^{ts})$
  if $(p < 0.05)$ **then**
    $s = s + 1$
    $X^{New} = \{(x_v^{ts}, y_v^{ts})\}_{v=1:l}$
    $X^{Tr} = (X^{Tr} \cup X^{New})$
    $Train(f_s, X^{Tr})$
  end if
  $\hat{y}_i = f(x_i)$
end for
return Y^{Ts}

was set to zero. This procedure of single-trial EEG classification was repeated for each new incoming trial until all the m trials were classified in the testing phase.

1.3.4 Unsupervised CSD-NSA Algorithm

The Unsupervised Covariate Shift Detection-Non-Stationary Adaptation (U-CSD-NSA) algorithm combines the aforementioned CSD test and an unsupervised adaptation method using a transductive-inductive classifier as described in the Algorithm 2. The idea of the proposed U-CSD-NSA algorithm is to adapt to the non-stationary changes by using both the training dataset and the new knowledge obtained in unsupervised mode from the testing phase. The transductive classifier was only used for adding new information to the existing training dataset and the inductive classifier was used for predicting the BCI outputs, after being retrained each time the CS was detected. It is thus a learning approach wherein the transductive and the inductive learning approaches were combined to update the training dataset and to adapt to the evolution of CS over the time period in the feature set of the testing phase. The transductive learning was implemented using a probabilistic weighted K-nearest neighbour (PWKNN) method (i.e. instance-based learning). The output from the transductive method was used to determine if a trial and its corresponding estimated label can be added to the training dataset and subsequently, the learning model was updated. Transductive learning combines induction and deduction in a single step and is related to the field of semi-supervised learning (SSL), which uses both labelled and unlabelled data during learning [26]. By eliminating the need to

Algorithm 2 :Unsupervised CSD-NSA Algorithm

Input : $X^{Tr} = \{x_i^{tr}, y_i^{tr}\}$, where $i \in \{1...n\}$
: $X^{Ts} = \{x_i^{ts}\}$ where $i \in \{1...m\}$
Output : Y^{Ts}
TRAINING:
$Tr(f_0, X^{Tr})$
Set $s = 0$, where s counts the number of covariate shift detection (CSD)
Set λ by minimizing 1-step-ahead-prediction error on the training dataset (X^{Tr})
TEST:
Start evaluation using testing dataset X^{Ts}
Set $i = 1, sl(0) = 1$
$\hat{y}_i = f(x_i)$
for $i = 2$ to m **do**
  $p \leftarrow CSD(x_i^{Ts}, \lambda)$
  if $(p < 0.05)$ **then**
    $s = s + 1$
    $sl(s) = i$
    $X^{New} = \{x_v^{ts}\}_{v=sl(s-1):sl(s)}$
    $(CR, \hat{y}) = PWkNN(X_{Tr}, X_{New}, CR_\alpha, k)$
    $X^{Tr} = (X^{Tr} \cup X^{New})$
    $Train(f_s, X^{Tr})$
  end if
  $\hat{y}_i = f(x_i)$
end for
return Y^{Ts}

construct a global model, transductive learning offers prospect to achieve higher accuracy. However, in order to make use of unlabelled data, it is necessary to assume some structure to the underlying distribution of the data. Additionally, it is essential that the SSL approach must satisfy at least one of the following assumptions such as smoothness, cluster, or manifold assumption [26], [27]. U-CSA-NDA algorithm makes use of the smoothness assumption (i.e. the points which are closest to each other are more likely to share the same label) to implement a transductive learning algorithm. The second classifier i.e. a linear support vector machine (SVM) classifier (f) was inductive and its outputs were used to determine the single-trial BCI outputs.

1.3.4.1 Probabilistic k-Nearest Neighbor

Probability theory plays a vital role in solving several of pattern recognition problems, as majority of these problems can be solved using a density estimation technique [27]. The task involved in this is to model a probability density function $P(x)$ of a random variable X, given a training input data X_{Tr}. There are two approaches for the density estimation, namely, parametric and non-parametric. One of the key limitations of the parametric approach is that it assumes a precise practical form for the distribution which may be incompatible for a specific application. An alternative approach is the non-parametric density estimation, which estimates density function

without applying any assumption about underlying data distribution. Here, we considered a non-parametric approach based on k-nearest-neighbors (kNN) being it a transductive learning methods, it uses the test data point to determine a decision. In the kNN algorithm, we considered a small sphere centered at the point x where the density $P(x)$ should be estimated. We allowed the radius of the sphere to grow until it contains k data points and the estimate of the density is then given by:

$$P(x) = k/(N \cdot V) \tag{1.2}$$

where, V is set to the volume of the sphere, and N is the total number of points. The parameter k governs the degree of smoothing. The technique of kNN density estimation may be extended to the classification task in which the kNN density estimation is obtained for each class and the Bayes' theorem is used to perform a classification task. Now, let's suppose that we have a data set comprising N_{ω_i} points in the class ω_i within the set of classes ω, where $i \in 1,2$, so that $\sum_{\omega_i} N_{\omega_i} = N$. If we wish to classify a new data point x, we draw a sphere centered on x containing precisely k points irrespective of their classes. Now, if this sphere has a volume V and contains $k_(\omega_i)$ from class ω_i as an estimate of the density associated with each class or likelihood can be obtained by:

$$P(x|\omega_i) = \frac{k_{\omega_i}}{N_{\omega_i} \cdot V} \tag{1.3}$$

Similarly, the unconditional density is given by $P(x) = k/(N \cdot V)$. The class prior probability is given by:

$$P(\omega_i) = N_{\omega_i}/N \tag{1.4}$$

Using the Bayes' theorem, we can obtain the posterior probability of the class membership:

$$P(\omega_i|x) = \frac{P(x|\omega_i)P(\omega_i)}{P(x)} = \frac{k_{\omega_i}}{k} \tag{1.5}$$

Further, we wish to minimize the probability of misclassification, this can be achieved by assigning the test point x to the class ω_i having the largest posterior probability, i.e. corresponding to the largest value of k_{ω_i}/k. Thus, to classify a new point, identify the k-nearest points from the training dataset and then assign the new point to the set having the largest number of representatives. This posterior probability is also known as the Bayesian belief or confidence ratio (CR). However, the overall estimate obtained by the kNN method may not be satisfactory because the resulting density is not a true probability density since its integral over all the samples space diverges [28]. Another drawback is that it considers only k points to build the density and each neighbor has an equal weight. An extension to the above kNN method is to assign the weight to each sample that depends on its distance to x. A radial basis function (RBF) kernel was used to obtain these weights. Using RBF Kernel, the nearest points have weights with the higher value than furthest points. A

PWkNN approach based on an RBF kernel is thus proposed to devise the transductive classifier with $RBF_{(p,q)}$, which is given as

$$RBF_{(p,q)} = exp(-\frac{d^2_{(p,q)}}{2\sigma^2}) \tag{1.6}$$

where $d_{(p,q)}$ is the Euclidean distance from the unlabelled data point x_p to the labeled data point x_q is computed as given below:

$$d_{(p,q)} = \sqrt{\sum_{i=1}^{m}(x_p(i) - x_q(i))^2} \tag{1.7}$$

and $x(i)$ is the ith feature of x and m is the number of features. For binary detection, the confidence ratio of CR_{ω_i} of the class ω_i, for a data point x_p, is defined by:

$$CR_{\omega_1} = \frac{\sum_{j=1}^{K} RBF_{(p,j)} \cdot (l_j == \omega_1)}{\sum_{j=1}^{K} RBF_{(p,j)}} \tag{1.8}$$

$$CR_{\omega_2} = 1 - CR_{\omega_1} \tag{1.9}$$

where $1 \leq j \leq k$, corresponds to the jth nearest neighbor of x_p. The outputs of PWKNN include the overall confidence of the decision, given by:

$$CR = max(CR_{\omega_1}, CR_{\omega_2}) \tag{1.10}$$

and the output class $\widehat{y}$ is equals to 1 if x_p is assigned to ω_1 otherwsie equals to 0.

1.4 Experimental Validation of the CSD-NSA Algorithms

1.4.1 EEG Dataset

The BCI Competition-IV dataset 2A [29] comprised of EEG signals acquired from nine healthy participants, namely $[A01 - A09]$, during two sessions on separate days using a cue-based MI paradigm. Each session consists of 6 runs where each run comprised of 48 trials (12 trials for each class). Thus, the complete study involved 576 trials from both sessions of the dataset. The total trial length is 7.5 s with variable inter-trial duration. The data were acquired from 25 channels (22 EEG channels along with three monopolar EOG channels) with a sampling frequency of 250 Hz and bandpass filtered between 0.5 Hz to 100 Hz (notch filter at 50 Hz). Reference and ground were placed at the left and right mastoid, respectively. Among the 22 EEG channels, 10 channels, which are responsible for capturing most of the MI related activations, were selected for this study (i.e. channels: $C3$, $FC3$, $CP3$, $C5$, $C1$, $C4$, $FC4$, $CP4$, $C2$, and $C6$). The dataset consist of four different MI tasks: left hand (class 1), right hand (class 2), both feet (class 3), and tongue (class 4). Only the classes corresponding to the left hand and right hand were considered in the present

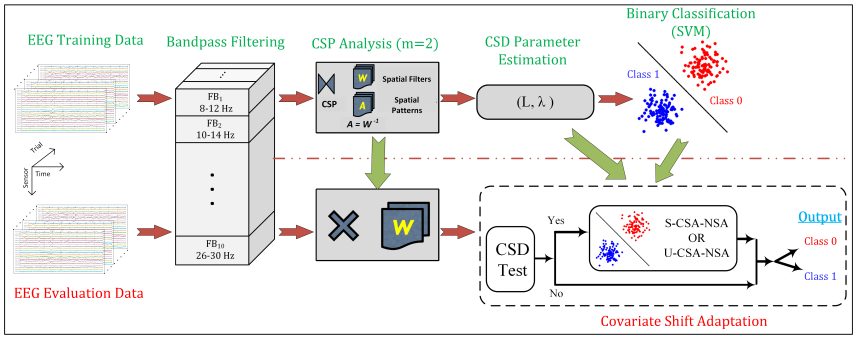

Figure 1.2 Schematic diagram of signal processing pipeline for implementation of CSD-NSA for MI-BCI system. The analysis consists of two phases. During the training phase, the features were extracted in the signal processing block from the filter-bank band-passed EEG signals using CSP algorithm and a classifier was trained on the labeled observation (i.e. XTrain). Also, two parameters (i.e., smoothing constant (λ) and control limit multiplier (L)) were estimated/set during this phase. In the testing phase, a similar signal processing approach was applied initially and CSP features were monitored by the CSA block. In the CSA block, the CSD test detects the presence of CSs and initiates adaptation using supervised or unsupervised adaptation. Finally, the classification was performed with the trained classifier model.

study. The MI data from the session-I was used for training phase and the MI data from the session-II was used for testing phase.

1.4.2 Signal Processing and Feature Extraction

Figure 1.2 illustrates the complete signal processing pipeline implemented in this study for CSD based NSA in MI related EEG patterns. The following steps have been executed: band-pass filtering, feature extraction (log variance of CSP), detection and validation of CSs, non-stationary adaptation in supervised or unsupervised manner, and finally the binary classification.

In the signal processing and feature extraction stage, a set of band-pass filters was used to decompose the EEG signals into different frequency bands (FBs) by employing an 8th order, zero-phase forward and reverse band-pass Butterworth filter. A total of 10 band-pass filters (i.e. filter bank) with overlapping bandwidths, including [8-12], [10-14], [12-16], [14-18], [16-20], [18-22], [20-24], [22-26], [24-28], and [26-30] Hz were used for temporal filtering of the data. Next, spatial filtering using CSP algorithm was performed to maximize the divergence of band-pass filtered signals under one class and minimize the divergence for the other class. In MI-related BCI systems, both physical and imaginary movements cause a growth of bounded neural rhythmic activity known as event related synchronization/desynchronization (ERD/ERS). The CSP algorithm has been widely implemented for estimating spa-

tial patterns for detecting ERD/ERS [21]. Each combination of the band-pass filter and extreme left and right components of the CSP filter provides the discriminative features that are specific to a particular frequency range. Next to CSP filtering, the discriminating features were extracted using a time window of 3 s after the cue onset so as to continue our further analysis on the MI-related features only. Finally, the obtained features from all FBs were merged to create the set of input features for a linear SVM classifier.

1.4.3 Feature Selection and Parameter Estimation

The parameters for the CSD test (i.e. λ and L) were computed during the training stage. The smoothing constant (λ) was selected by minimizing the sum of squared 1-step-ahead prediction errors. The control limit multiplier L was set equal to 2. The value of L is needed to be carefully selected because it has a major impact in the performance of the CSD test. A small value of L makes the system more sensitive to minor shifts in EEG features. Moreover, the shifts in the feature set generated due to the noise were smoothed by proper selection of λ. In the testing phase, the CSD test was applied on the multivariate inputs features of EEG data using the estimated values of parameters. Due to the high dimensionality of the EEG features, the PCA algorithm was used to reduce the dimensionality to a single component [30]. Next, the CSD test was applied to the PCA output features for detecting CSs at the first stage of the CSD test. Based on the positive outcome of the CSD test in stage I, stage II gets activated. In stage II, a window of 3 s of CSP features after the cue onset in the current trial was extracted to use as a first sample and a window of averaged CSP features from the previous data trials was used as the second sample to execute the multivariate two-sample Hotelling's T-Square statistical hypothesis test. If the p-value of the Hotelling's T-Square test is less than 0.05, a CSD was confirmed in the current trial and an adaptive action was initiated. In U-CSD-NSA algorithm, the value of CR_{Thres} was set equal to 0.70, which means if the probability of the classification is more than 0.70 then only the example was added to the training dataset.

1.4.4 Empirical Results

The performance of the CSD-NSA algorithm has been evaluated with active approach for covariate shift adaptation. With a single classifier, active approach was employed with both supervised adaptation and unsupervised adaptation (i.e. S-CSD-NSA and U-CSD-NSA, respectively). In the S-CSD-NSA algorithm, the adaptation was achieved after detection of each shift, wherein the data from the correctly predicted trials were merged with the existing training dataset to enrich the data distribution, and the classifier was re-trained on the updated dataset. Likewise, U-CSD-NSA algorithm has been implemented using the PWkNN method for the unsupervised adaptation, wherein the training dataset was updated at the instances of shift detection only.

We have compared the classification accuracies (in %) obtained by the two proposed algorithms with the baseline method (i.e., CSP features) for a binary classifica-

Table 1.1 Performance of the Covariate Shift Detection based Non-Stationary Adaptation (CSD-NSA) Algorithm

Subjects	**Baseline**	**S-CSD-NSA**	**U-CSD-NSA**
A01	87.50	91.75	93.06
A02	58.33	58.33	59.03
A03	84.72	91.75	95.14
A04	63.89	68.33	71.53
A05	67.55	68.33	71.53
A06	62.50	62.50	62.50
A07	70.83	71.53	71.53
A08	86.11	91.75	91.75
A09	86.11	89.58	89.58
Mean	74.17	77.09	78.41
Std	11.84	13.92	14.00
p-value		0.0156	0.0078

tion task of MI-related BCI system. Table 1.1 provides the classification accuracies (in %) for the nine healthy participants along with the mean and SD. The average binary classification accuracy (mean±SD) for the baseline method, S-CSD-NSA, and U-CSD-NSA are 74.17 ± 11.84, 77.09 ± 13.92, and 78.41 ± 14.00, respectively. Thus, both CSD-NSA algorithms enhances the discriminability of the MI features as compared to the non-adaptive system. Furthermore, the Wilcoxon signed rank statistical test provided significant p-values for both methods i.e., 0.0156 for S-CSD-NSA and 0.0078 for U-CSD-NSA in comparision with the baseline method.

1.5 Discussion and Future Prospects

Learning in NSEs provides a challenging and inspiring area of research in the fields of machine learning and computational intelligence, and secured escalating interest of the researchers globally because of its increasing prevalence in real-world applications involving streaming and big data. EEG-based BCI systems involve learning of features from brain generated electrical potentials which in turn provides a highly dynamic and complex environment. In such applications, using traditional approaches that either ignore the underlying shifts or their passive approach based handling are inevitably bound to low performance. Thus, an active approach based shift detection and subsequent adaptive measure is necessary to achieve high classification accuracies. In this chapter, we presented two algorithms for supervised and unsupervised adaptive learning in NSEs based on CSD and NSA framework. The results, with a benchmark MI-related BCI dataset, showed statistically significant improvement of the system performance.

The EWMA-CSD test is a good option for detecting covariate shifts because it circumvents false detections resulting from noise or spurious shifts through much more intense smoothing of the EEG signal. A central issue in the CSD test is the selection of the value for λ and L. For the auto-correlated data, it is suggested to select a value of λ that minimizes the sum of the squares of the 1-SAP errors. Moreover, for L, considering smaller value ($L = 2$) results in concentrating on trivial shifts, such as temporary disturbance in user concentration and muscular artefacts emerging during trial-to-trial transfer. Conversely, the long term covariate shift may be handled by fixing a large value ($L = 3$). The value of parameter CR_{Thres} resolves importance of the current information. If the value of CR is above this threshold, then it is beneficial to be into the existing knowledge base (KB) otherwise rejected. The rejected instance may belong to a shifted distribution but it is not providing higher confidence to be merged into the KB. Thus, the values of CR_{Thres} and L are needed to be cautiously selected in order to get greater classification accuracy and these hyper-parameters can be subject specific as well [31, 32, 33].

There are, however, a few limitations to be considered while using the active approach based non-stationary learning algorithm for BCI systems and can be taken into account in future studies. First, the CSD test has been applied on the combined CSP features of multiple frequency bands, which creates a high dimensional input vector and may affect the robustness of the covariate shift detection process. Furthermore, several other feature types have been utilised in MI-BCI systems and NSA for these feature sets may warrant further investigation of the proposed methods [8]. Second, we have used a supervised approach (i.e., S-CSD-NSA) for adaptation, which can be of limited usage in case of real-time BCI-based applications such as for rehabilitation or communication, where the labels are not available during the testing phase. However, the unsupervised adaptation method to learn with the upcoming shifts in unlabeled streaming EEG data. Third, the proposed system has been implemented on the feature set of two sessions only whereas practical BCI applications may have more number of sessions recorded on different days, and this condition can make the adaptive learning task further challenging. Moreover, single classifier based learning can be replaced by ensemble based learning by recruiting several classifiers. In that case, a recurrent concept handling method may be needed to dynamically replace the old classifier with the updated classifier in the ensemble and when the concept reappears, the old classifier will be re-activated. Fourth, datasets are assumed to be labeled when presented to a supervised algorithm or unlabeled for an unsupervised one. However, the data streams may contain a mixture of labeled and unlabeled data. Thus, one interesting prospect of further research is to examine the collective concept of supervised and unsupervised learning in NSEs. Fifth, one of the central issues with current BCI technology is the need to increase the number of classes for optimizing the practical application. Also hybrid BCI systems can provide further flexibility to the system [34]. However, these enhancements generate data with different characteristics, such as multi-dimensionality, multi-scale, and multi-label. Thus, future learning algorithms should include new modeling and adaptive strategies to be able to cope with such data and conditions.

References

[1] Shenoy P, Krauledat M, Blankertz B, et al. Towards adaptive classification for BCI. Journal of neural engineering. 2006 mar;3(1):13–23.

[2] Arvaneh M, Guan C, Ang KK, et al. Optimizing spatial filters by minimizing within-class dissimilarities in electroencephalogram-based brain-computer interface. IEEE Transactions on Neural Networks and Learning Systems. 2013;24(4):610–619.

[3] Arvaneh M, Guan C, Chai Quek. EEG Data Space Adaptation to Reduce Intersession Nonstationary in Brain-Computer Interface. Journal of Neural Computation. 2013;25:1–26.

[4] Wolpaw JR, McFarland DJ, Neat GW, et al. An EEG-based brain-computer interface for cursor control. Electroencephalography and Clinical Neurophysiology. 1991 mar;78(3):252–259.

[5] Wolpaw JR, Birbaumer N, McFarland DJ, et al. Brain-Computer Interfaces for Communication and Control. Clinical Neurophysiology : Official Journal of the International Federation of Clinical Neurophysiology. 2002 jun;113(6):767–91.

[6] Lotte F, Congedo M, Lécuyer A, et al. A review of classification algorithms for EEG-based brain-computer interfaces. Journal of neural engineering. 2007 jun;4(2):1–13.

[7] Ramoser H, Müller-Gerking J, Pfurtscheller G. Optimal spatial filtering of single trial EEG during imagined hand movement. IEEE Transactions on Rehabilitation Engineering. 2000 dec;8(4):441–446.

[8] Rathee D, Raza H, Prasad G, et al. Current Source Density Estimation Enhances the Performance of Motor-Imagery related Brain-Computer Interface. IEEE Transactions on Neural Systems and Rehabilitation Engineering. 2017;25(12):2461 – 2471.

[9] Rathee D, Cecotti H, Prasad G. Single-trial effective brain connectivity patterns enhance discriminability of mental imagery tasks. Journal of neural engineering. 2017;14(5):056005.

[10] Li Y, Kambara H, Koike Y, et al. Application of covariate shift adaptation techniques in brain-computer interfaces. IEEE Transactions on Biomedical Engineering. 2010 jun;57(6):1318–1324.

[11] Raza H, Cecotti H, Li Y, et al. Learning with Covariate Shift-Detection and Adaptation in Non-Stationary Environments : Application to Brain-Computer Interface. In: International Joint Conference on Neural Networks (IJCNN); 2015. p. 1–8.

[12] Blankertz B, Tomioka R, Lemm S, et al. Optimizing spatial filters for robust EEG single-trial analysis. IEEE Signal Processing Magazine. 2008;25(1):41–56.

[13] Satti A, Guan C, Coyle D, et al. A covariate shift minimization method to alleviate non-stationarity effects for an adaptive brain-computer interface. In: Proceedings - International Conference on Pattern Recognition. Ieee; 2010. p. 105–108.

[14] Sugiyama M, Krauledat M, Müller KR. Covariate shift adaptation by importance weighted cross validation. Journal of Machine Learning Research. 2007;8:985–1005.

[15] Raza H, Prasad G, Li Y. EWMA based two-stage dataset shift-detection in non-stationary environments. In: IFIP Advances in Information and Communication Technology. vol. 412. Springer Berlin Heidelberg; 2013. p. 625–635.

[16] Shahid S, Prasad G. Bispectrum-based feature extraction technique for devising a practical brain-computer interface. Journal of neural engineering. 2011 apr;8(2):025014.

[17] Suk HI, Lee SW. A novel bayesian framework for discriminative feature extraction in brain-computer interfaces. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2013;35(2):286–299.

[18] Raza H, Cecotti H, Prasad G. Optimising Frequency Band Selection with Forward-Addition and Backward-Elimination Algorithms in EEG-based Brain-Computer Interfaces. In: 2015 International Joint Conference on Neural Networks (IJCNN). IEEE; 2015. p. 1–7.

[19] Vidaurre C, Schlöogl A, Cabeza R, et al. A fully on-line adaptive BCI. IEEE Transactions on Biomedical Engineering. 2006;53(6):1214–1219.

[20] Sun S, Zhang C, Zhang D. An experimental evaluation of ensemble methods for EEG signal classification. Pattern Recognition Letters. 2007;28(15):2157–2163.

[21] Liyanage SR, Guan C, Zhang H, et al. Dynamically weighted ensemble classification for non-stationary EEG processing. Journal of neural engineering. 2013 apr;10(3):036007.

[22] Sannelli C, Vidaurre C, Müller KR, et al. Ensembles of adaptive spatial filters increase BCI performance: an online evaluation. Journal of neural engineering. 2016 may;13(4):046003.

[23] Raza H, Prasad G, Li Y. EWMA model based shift-detection methods for detecting covariate shifts in non-stationary environments. Pattern Recognition. 2015;48(3):659–669. Available from: http://dx.doi.org/10.1016/j.patcog.2014.07.028.

[24] Montgomery D, Mastrangelo C. Some Statistical Process Control Methods for Autocorrelated Data. Journal of Quality Technology. 1991;23(3):179–204.

[25] Kuncheva L, Faithfull W. PCA feature extraction for change detection in multidimensional unlabeled data. IEEE Transactions on Neural Networks and Learning Systems. 2014;25(1):69–80.

[26] Raza H, Prasad G, Li Y, et al. Toward Transductive Learning Classifiers for Non-Stationary EEG. In: Engineering in Medicine and Biology Society (EMBC), 2014 35th Annual International Conference of the IEEE; 2014. p. 1.

[27] Sugiyama M, Kanamori T, Suzuki T, et al. Density-difference estimation. Neural computation. 2013 oct;25(10):2734–75.

[28] Bishop CM. Pattern Recognition and Machine Learning. New York: Springer; 2006.

[29] Tangermann M, Müller KR, Aertsen A, et al. Review of the BCI competition IV. Frontiers in Neuroscience. 2012 jan;6:55.

[30] Kolter JZ, Maloof Ma. Dynamic Weighted Majority: A New Ensemble Method for Tracking Concept Drift. In: Third IEEE International Conference on Data Mining. IEEE Comput. Soc; 2003. p. 123–130.

[31] Raza H, Cecotti H, Prasad G. A combination of transductive and inductive learning for handling non-stationarities in motor imagery classification. In: International Joint Conference on Neural Networks (IJCNN); 2016. p. 763–770.

[32] Chowdhury A, Raza H, Dutta A, et al. EEG-EMG Based Hybrid Brain Computer Interface for Triggering Hand Exoskeleton for Neuro-Rehabilitation. In: Proceedings of the Advances in Robotics; 2017. p. 45:1—-45:6.

[33] Chowdhury A, Raza H, Meena YK, et al. Online Covariate Shift Detection based Adaptive Brain-Computer Interface to Trigger Hand Exoskeleton Feedback for Neuro-Rehabilitation. IEEE Transactions on Cognitive and Developmental Systems. 2017;PP(99):1.

[34] O'Doherty D, Meena YK, Raza H, et al. Exploring gaze-motor imagery hybrid brain-computer interface design. In: 2014 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). Ieee; 2014. p. 335–339.