Dataset Shift Detection in Non-Stationary Environments using EWMA Charts

Haider Raza, Girijesh Prasad, Yuhua Li
Intelligent Systems Research Center, University of Ulster, Londonderry, Northern Ireland, UK
raza-h@email.ulster.ac.uk, g.prasad@ulster.ac.uk, y.li@ulster.ac.uk

***Abstract*—Dataset shift is a major challenge in the non-stationary environments wherein the input data distribution may change over time. Detecting the dataset shift point in the time-series data, where the distribution of time-series changes its properties, is of utmost interest. Dataset shift exists in a broad range of real-world systems. In such systems, there is a need for continuous monitoring of the process behavior and tracking the state of the shift so as to decide about initiating adaptive corrections in a timely manner. This paper presents an algorithm to detect the shift-point in a non-stationary time-series data. The proposed method detects the shift-point based on an exponentially weighted moving average (EWMA) control chart for auto-correlated observations. This algorithm is suitable to be run in real-time and monitors the data to detect the dataset shift. Its performance is evaluated through experiments using synthetic and real-world datasets. Results show that all the dataset-shifts are detected without the delay.**

Index terms- Non-stationary, Dataset shift, EWMA, Online change-detection

I. INTRODUCTION

In the research community of statistics and machine learning, detecting abrupt and gradual changes in time-series data is called change-point detection [1]. Based on the delay of the detection, change-point detection problems can be classified into two categories: retrospective detection and real-time detection [2]. The retrospective change-point detection tends to give more accurate and robust detection; however, it requires longer reaction periods so it may not be suitable for the real-time applications where initiating adaptation closest to the change-point is of paramount importance. Also, in the real-time systems, each observation coming in the data stream maybe processed only once and then discarded; hence retrospective change point detection may not be possible to implement. The real-time detection is also called a single pass method and requires an immediate reaction to meet the deadline. One important application of such a method is in pattern classification based on streaming data, performed in several key areas such as electroencephalography (EEG) based brain-computer interface, spam-filtering, network intrusion detection and remote-sensing environments.

Classifying the data stream in non-stationary environments requires the developments of a method which should be computationally efficient and able to detect the shift-point in the underlying distribution of the data stream. The key difference between the streaming classification and conventional classification is that, in the streaming case, the input data distribution may change over time during the testing/operating phase due to the presence of non-stationarities, this phenomenon is known as covariate shift [3]. In a conventional classification problem, we assume that the input data distribution does not vary with time. Based on the shift detection, a classifier needs to account for the dataset shift. There are several methods presented in the literature to deal with dataset shift problem such as completely retraining the classifier on the new received data in an offline mode [4], retraining the classifier (sample-by-sample) in an online mode [4] and covariate shift adaptation [5]. In non-stationary environments there are several types of dataset shift, a brief review of the dataset shift and the types of dataset shift are presented in next section.

Some pioneering works [2][6][7] have demonstrated very good shift-point detection performance by monitoring the moving control charts and comparing the probability distributions of the time-series samples over past and present intervals. These methods follow different strategies such as CUSUM (Cumulative SUM) [8], Computational-Intelligence CUSUM (CI-SUSUM) [7], Intersection of Confidence Interval rule (ICI) for change-point detection [9] and the generalized logarithm-ratio method [10].

However, the aforementioned methods rely on pre-designed parametric models such as underlying probability distribution, auto-regressive models and state-space models, for tracking some specific statistics such as the mean, variance, and spectrum. Thus, they are not robust against different types of the shifts because of the delay in shift detection on account of the need for identifying models, which may significantly limit their range of applications in fast data streaming problems. However, to overcome these weaknesses, recently some researchers have proposed a different strategy, which estimates the ratio of two probability densities directly without density estimation [11].

In this paper we present an approach to detect the dataset shift based on an exponentially weighted moving average (EWMA) chart. It is demonstrated to outperform other approaches in terms of non-stationarity detection without any significant time delay. The approach is computationally efficient because of low computational cost and less memory requirements during online processing. So, this scheme can be deployed along with any classifier such as k-nearest neighbor (kNN), linear discriminant analysis (LDA), neural networks, or support vector machine (SVM) in an adaptive online learning framework.

This paper proceeds as follows: Section 2 presents a background of dataset shift and EWMA control chart.

Section 3 consists of methodology with the shift-detection algorithm. Section 4 presents the datasets used in the experiment. Finally, Section 5 presents the results and discussion.

II. Background

Dataset shift: The term dataset shift [12],[13] was first used in the workshop of neural information processing systems (NIPS, 2006). Assume a classification problem is described by a set of features or inputs $\boldsymbol{x}$, a target variable $\boldsymbol{y}$, the joint distribution $\boldsymbol{P(y,x)}$, the prior probability $\boldsymbol{p(x)}$ and conditional probability $\boldsymbol{p(y|x)}$, respectively. The dataset shift is then defined as "cases where the joint distribution of inputs and outputs differs between training and test stages, i.e., when $\boldsymbol{P_{train}(y,x)} \neq \boldsymbol{P_{test}(y,x)}$"[14]. Dataset shift was previously defined by various authors and they gave different names to the same concept such as, concept shift, changes of classification, changing environment, contrast mining, fracture point, and fracture between data.

There are three types of dataset shift that usually occur, (i) *covariate shift*, (ii) *prior probability shift* and (iii) *concept shift*; we will review these shifts in following subsections.

Covariate Shift: In a generic way, it is defined as, "*covariate shift appears only in X→Y problems, and the case where the conditional probability in training and testing remains the same ($p_{train}(y|x) = p_{test}(y|x)$), but the input distribution $p(x)$ changes between training and testing, i.e., $p_{train}(x) \neq p_{test}(x)$*" [13]. For example, the input distribution often changes in the session-to-session transfer of an EEG based brain-computer interface.

Prior Probability Shift: It appears in the problem $Y \rightarrow X$. The prior probability shift is defined alternatively in [13] as "*appears only in Y→X problems, the case where the conditional probability remains same ($p_{train}(x|y) = p_{test}(x|y)$), but prior probability changes i.e., ($p_{train}(y) \neq p_{test}(y)$)*". For example, the survivorship in a stock market, where the bias is there in obtaining the samples.

Concept Shift: In concept shift, the input data distribution of training and testing remains unchanged but the conditional relationship between them changes, e.g., $p(y|x)$ changes from $p_{train}(y|x)$ *to* $p_{test}(y|x)$, i.e., ($p_{train}(y|x) \neq p_{test}(y|x)$) [13].

The shifts discussed above are the most commonly present in the real-world problems, there are other shifts also that could happen in theory, but we are not discussing those because they appear rarely.

EWMA control chart: An exponentially weighted moving average (EWMA) control chart [15] is from the family of control charts in the statistical process control (SPC) theory. Control charts are the graphical representation of sample statistics for SPC. EWMA is an efficient statistical method in detecting the small shifts in the time-series data. The EWMA control chart outperforms the other control charts because it combines current and historical data in such a way that small changes in the time-series can be detected more easily and quickly. Other charts, such as the Shewhart chart [16], only consider the most current observations by forgetting the historical data. The EWMA uses a weighting constant, lambda (λ), which decides the importance of current and historical observations. The exponentially weighted moving average model is defined as

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

where λ is the smoothing constant ($0 < \lambda \leq 1$). Moreover, the EWMA charts are used for both uncorrelated and auto-correlated data. We are only considering the auto-correlated data in our study and simulation.

***EWMA for auto-correlated data*:** If data contain a sequence of auto-correlated observations, $x_{(i)}$, then the EWMA statistic in (1) can be used to provide 1-step-ahead prediction model of auto-correlated data. We have assumed that the process observations $x_{(i)}$, can be defined as the equation (2) below, which is a first-order integrated moving average (ARIMA) model. This equation describes non-stationary behavior, wherein the variable $x_{(i)}$ shifts as if there is no fixed value of the process mean.

$$x_{(i)} = x_{(i-1)} + \varepsilon_i - \theta \varepsilon_{i-1} \tag{2}$$

where ε_i is a sequence of independent and identically distributed (i.i.d) random signal with zero mean and constant variance. It can be easily shown that the EWMA with ($\lambda = 1 - \theta$) is the optimal 1-step-ahead prediction for this process [17]. That is, if $\hat{x}_{i+1}(i)$ is the forecast of the observation in period $(i+1)$ made at the end of period i, then, $\hat{x}_{i+1}(i) = z_i$, where equation (1) is the EWMA. The 1-step-ahead prediction errors $err_{(i)}$ are calculated as [18]:

$$err_{(i)} = x_{(i)} - \hat{x}_{(i)}(i-1) \tag{3}$$

Assume, that the 1-step-ahead prediction errors $err_{(i)}$ are normally distributed. It is given in [17], that it is possible to combine information about the statistical control and process dynamics on a single control chart. Then, the control limits of the chart on these errors satisfy the following probability statement by substituting the right hand side of equation (3) as given below.

$$P\left[-L.\ \sigma_{err} \leq err_{(i)} \leq L.\sigma_{err}\right] = 1-\alpha$$
$$P\left[-L.\ \sigma_{err} \leq x_{(i)} - \hat{x}_{(i)}(i-1)\ (i) \leq L.\ \sigma_{err}\right] = 1-\alpha$$
$$P\left[\hat{x}_{(i)}(i-1)\ - L.\sigma_{err} \leq x_{(i)} \leq \hat{x}_{(i)}(i-1)\ + L.\sigma_{err}\right] = 1-\alpha$$

where σ_{err} is the standard deviation of the errors. If the EWMA is a suitable 1-step-ahead predictor, then one could use $z_{(i)}$ as the center line for the period $i+1$ with UCL (Upper Control Limit) and LCL (Lower Control Limit) [17].

$$UCL_{(i+1)} = z_{(i)} + L.\sigma_{err} \tag{4.a}$$
$$LCL_{(i+1)} = z_{(i)} - L.\sigma_{err} \tag{4.b}$$

Whenever, the $x_{(i+1)}$ moves out of $UCL_{(i+1)}$ and $LCL_{(i+1)}$, the process is said to be out of control. This method is also known as *a moving center-line EWMA control chart*.

The EWMA control chart is robust to the non-normality assumption if properly designed for the t and gamma distributions [19]. So, the EWMA chart can be employed when there is concern about the non-normality assumption. Following the above assumption, we have designed an algorithm for the shift detection based on the EWMA in the process observation of auto-correlated data, which is discussed in the next section.

III. METHODOLOGY

A control chart is the graphical representation of the sample statistics. Commonly it is represented by three lines plotted along horizontal axis. The center line and two control lines (control limits) are plotted on a control chart, which correspond to target value (μ) and acceptable deviation ($L\sigma$) from either side of the target value respectively, where L is the control limit multiplier and σ is the standard deviation. The goal of the control chart is to monitor a process change, such as the shift in data generating process. When the plotting statistics falls outside the control limits, the process is said to be out of control. The proposed work employs an EWMA control chart to detect the dataset shift in the data stream. It works in two phases, the first phase is retrospective or training phase and the second phase is operation or testing phase. In the training phase, the parameters are calculated to decide the null hypothesis that there is no shift in the data. In the testing phase, the process observations are continuously monitored by the EWMA chart and when an observation falls outside the control limits of the control chart the point is said to be a point of dataset shift. In other words, the process observation falling outside the control limit means that the observed variable is not in the control limit and the null hypothesis is rejected in favor of alternative hypothesis and the shift is detected. An important point to note here is that we have assumed that non-stationarity occurs due to changes in the input distribution only. So, it is said to be covariate shift detection in non-stationary time-series. In the next paragraph, the designed algorithm is discussed in detail.

Shift Detection based on EWMA (SD-EWMA)

Our algorithm works in two different phases, the first phase is a retrospective/training phase in which the parameters $(\lambda, z_{(0)}, \sigma_{err}{}^2_{(0)})$ to be used are calculated followed by the operation/test phase for shift detection. The pseudo code of the algorithm is given in Table 1.

In the training phase, first obtain the sequence of observations and calculate the mean, use it as the initial value $z_{(0)}$ and obtain the EWMA statistic by equation (1). The sum of the squared one-step-ahead prediction error divided by the length of the training dataset is used as an initial value of $\sigma_{err}{}^2_{(0)}$, for the testing data. The choice of λ is discussed later in the result and discussion section.

In the testing phase, for each observation use equation (1) to obtain the EWMA statistics and follow the steps given in Table 1, Algorithm 1 and then compute the $UCL_{(i+1)}$ and $LCL_{(i+1)}$. Next, check if each observation $x_{(i+1)}$ falls within the control limits $\left[UCL_{(i+1)}, LCL_{(i+1)}\right]$, otherwise the shift is detected and alarm is raised. The standard deviation of 1-step-ahead error can be estimated in various way such as mean absolute deviation or directly calculate the smoothed variance [17] as given in the step 4 of the testing phase in SD-EWMA algorithm in Table 1.

As we have discussed earlier, this type of shift detection test could be applied in parallel with any classifiers in the non-stationary environments, for process monitoring and shift detection.

TABLE 1**:** Algorithm-SD-EWMA

Training Input: For each observation in the training dataset, go to training phase and compute the parameters for testing.
Testing input: Receive new data sample-by-sample and perform the check as follows.
IF (Shift detected)
THEN (Report the point of shift and initiate an appropriate corrective action)
ELSE
(Continue and integrate the upcoming information).
Output: Shift-detection points

<u>Training Phase</u>
1. Initialize training data: $x_{(i)}$ $for\ i = 1\ to\ n$, n is the size of training data.
2. Calculate the mean of input data ($\bar{x}$) and set it to the initial value of $z_{(0)}$.
3. Compute the z-statistics for each observation $x_{(i)}$ in training data
$$z_{(i)} = \lambda . x_{(i)} + (1-\lambda). z_{(i-1)}$$
4. Compute 1-step-ahead prediction errors
$$err_{(i)} = x_{(i)} - \hat{x}_i(i-1)$$
5. Compute the sum of 1-step-ahead prediction error divided by number of observations and use it as the initial value of variance $\sigma_{err}{}^2_{(0)}$ for the testing phase.
6. Set λ by minimizing the sum of squared prediction error on training data.

<u>Testing Phase</u>
1. For each data point $x_{(i+1)}$ in the operation/testing phase,
2. Compute $z_{(i)} = \lambda . x_{(i)} + (1-\lambda). z_{(i-1)}$
3. Compute $err_{(i)} = x_{(i)} - \hat{x}_{(i)}(i-1)$
4. Compute the estimated variance $\hat{\sigma}_{err}{}^2_{(i)} = \vartheta . err_{(i)}{}^2 + (1-\vartheta). \hat{\sigma}_{err}{}^2_{(i-1)}$
5. Compute $UCL_{(i+1)}$ and $LCL_{(i+1)}$
6. $UCL_{(i+1)} = z_{(i)} + L.\sigma_{err}$
7. $LCL_{(i+1)} = z_{(i)} - L.\sigma_{err}$
8. *IF* $(LCL_{(i+1)} < x_{(i+1)} < UCL_{(i+1)})$
9. *THEN*
10. Continue processing with new sample
11. *ELSE* (Covariate Shift detected)
12. Initiate an appropriate corrective action.

IV. DATASET AND FEATURES ANALYSIS

To validate the effectiveness of the suggested SD-EWMA algorithm, a series of experimental evaluations have been performed on four synthetic datasets and one real-world dataset. The datasets are described as follows.

4.1 Synthetic Data

Dataset 1-Jumping Mean (D1)*:* The dataset used here is same as the toy dataset given in [2] for detecting change point in a time-series data. The dataset is defined as $y(t)$ in which 5000 samples are generated $(i.e., t = 1, \ldots\ldots, 5000)$

$$y(t) = 0.6y(t-1) - 0.5y(t-2) + \varepsilon_t$$

where ε_t is a noise with mean μ and standard deviation 1.5. The initial values are set as $y(1) = y(2) = 0$. A change point is inserted at every 100 time steps by setting the noise mean μ at time t as

$$\mu_N = \begin{cases} 0 & N = 1 \\ \mu_{N-1} + \dfrac{N}{16} & N = 2 \ldots .49 \end{cases}$$

where N is a natural number such that $100(N-1) + 1 \leq t \leq 100N$.

Dataset 2-Scaling Variance (D2)*:* The dataset used here is same as the toy dataset given in [2], for detecting change-point in time-series data. The dataset is defined as the auto-regressive model, but the change point is inserted at every

100 time steps by setting the noise standard deviation σ at time t as

$$\sigma = \begin{cases} 1 & N = 1,3, \dots\dots ,49 \\ \ln\left(e + \frac{N}{4}\right) & N = 2,4, \dots 48 \end{cases}$$

where N is a natural number such that $100(N-1)+1 \leq t \leq 100N$.

***Dataset 3-Positive-Auto-correlated (D3)**:* The dataset is consisting of 2000 data-points, the non stationarity occurs in the middle of the data stream, shifting from $\mathcal{N}(x; 1,1)$ to $\mathcal{N}(x; 3,1)$, where $\mathcal{N}(x; \mu, \sigma)$ denotes the normal distribution with mean and standard deviation respectively.

Dataset 4-Auto-correlated (D4): The dataset is a time-series consisting of 2000 data-points using 1-D digital filter from matlab function ('filter'). The filter function creates a direct form II transposed implementation of a standard difference equation. In the filter, the denominator coefficient is changed from 2 to 0.5 after producing 1000 number of points.

4.2 Real-world Dataset

The real-world data used here are from BCI competition-III dataset (IV-b) [20]. This dataset, contains 2 classes, 118 EEG channels (0.05-200Hz), 1000Hz sampling rate which is down-sampled to 100Hz, 210 training trials, and 420 test trials. This data set was recorded from one healthy subject. He sat in a comfortable chair with arms resting on armrests. The training data set consists of the first 3 (non-feedback) sessions. Visual cues (letter presentation) indicated for 3.5 seconds required the subject to perform (L) left hand or (F) right foot motor imageries. The presentation of the target cues was intermitted by periods of random length, 1.75 to 2.25 seconds, in which the subject could relax. The test data was recorded more than 3 hours after the training data. The experimental setup was similar to the training sessions, but the motor imagery had to be performed for 1 second only, compared to 3.5 seconds in the training sessions. The intermitting periods ranged from 1.75 to 2.25 seconds as before.

We have selected a single channel (C3) and performed the band-pass filtering on the mu (μ) band (8-12 Hz) and extracted the features. Further, to show the session-to-session shift in the EEG signal, we have plotted the histogram in Figure 1.

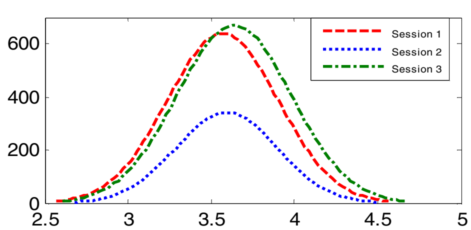


Figure 1**:** Histogram plot of 3 different sessions' data taken from the training dataset. It is clear from the plot that, in each session the distribution is changed by shifting the mean from session-to-session transfer.

It is clear from the Figure 1 that EEG data is non-stationary over multiple sessions of a BCI experiment. Further, to check the non-stationarities we have performed the two-sample hypothesis test, wherein the two distributions are matched to check whether they are coming from the same distribution. Moreover, from the Figure 1 we can easily mark the shift in the mean and therefore the change in the feature distribution in the EEG data obtained from three different sessions of a BCI experiment.

If these continuous shifts in the distribution are detected at the time of the occurrence, then an appropriate action could be taken to account for the dataset shifts. This real-world dataset is a good example to validate if the proposed SD-EWMA is able to detect these types of shifts in the data generating process. The results and discussion of the experiments are given in the next section.

V. Results and Discussion

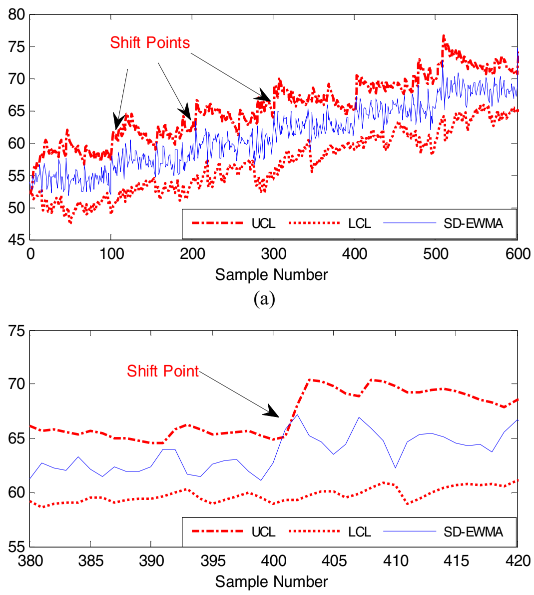
On each dataset, the SD-EWMA technique is evaluated by the number of false alarms, the number of hits and number of misses. A false alarm is the signal of false detection or type-I error. A hit is an alarm of shift detection and considered as true-positive. A false-negative is a miss and measured as type-II error. The following keys are suggested to measure the performance of the tests:

- *False positive (FP)*: it counts the times a test detects a shift in the sequence when it is not there i.e., (false alarm) or type-I error.
- *True positive (TP)*: it counts the times a test detects a shift in the sequence when it is there i.e., (hit alarm).
- *False negative (FN)*: it counts the times a test does not detects a shift in the sequence when it is there i.e., (miss) or type-II error.
- *Average delay (AD)*: it measures the average delay in shift detection, i.e., sum of delay in each shift detection divided by the number of shifts detection.

It is important to note that SD-EWMA is based on the current observation of the data to detect the shift, so it is coming without delay and this is the main advantage of this method. Figure 2 represents the SD-EWMA based shift detection test results for the Dataset 1.

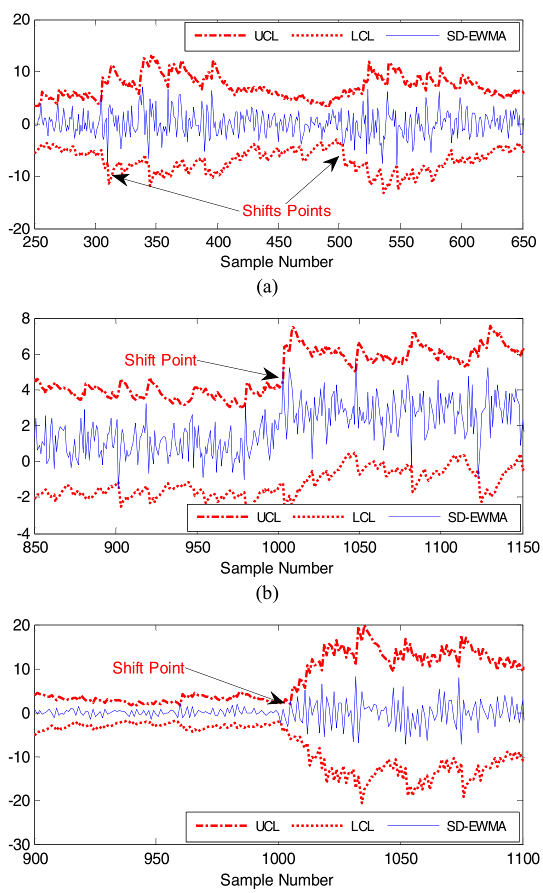
Table 2 shows the results in terms of true positive, false positive and false negative. Figure 3 shows the test results on each of synthetic datasets. The solid line is the observation plotted on the chart and the two dotted lines are the ULC and LCL, whenever the solid line crosses the dotted line (control limits), it is the shift-point detection. For each dataset the test has been performed on several values of lambda (λ), chosen as suggested in [17], [19], moreover, the lambda (λ), is chosen by minimizing the least mean square prediction error.

In the case of dataset 1 with $\lambda = 0.40$, it gives an optimal result with all the shift detection but with four false positive. For the dataset 2 with $\lambda = 0.70$, it gives an optimal result with all the shift detection along with six false positive. For $\lambda = 0.80$ it detects only 4 out of five shifts with eight false positive. Next, for the dataset 3 when $\lambda = 0.50$ it detects all the shifts but with nine false positives. On increasing and decreasing the value of λ, the number of false positives increases. Lastly, with the dataset 4, when $\lambda = 0.50$, it detects all shifts but with six false positives. Hence, the value of λ has to be carefully selected based on the type of experiment and by minimizing the least square error in training dataset.



(b)

Figure 2: Shift detection based on SD-EWMA: Dataset 1 (jumping mean): (a) the shift point is detected at every 100th point. (b) Zoomed view of figure a: shift is detected at 201st sample by crossing the upper control limit.



(c)

Figure 3: Shift detection based on SD-EWMA: (a) Dataset 2 (scaling variance): the shift is detected at 3 points. (b) Dataset 3 (positive auto-correlated): detects the shift after producing 1000 observations. (c) Dataset 4 (Auto-correlated): detects the shift after producing 1000 observations.

TABLE 2: SD-EWMA shift detection in time-series data

	Total shifts	Lambda (λ)	# TP	# FP	#FN	% Shift detected
D1	9	0.20	8	1	1	88.8
		0.30	8	3	1	88.8
		0.40	All	4	0	100
D2	5	0.60	All	11	1	100
		0.70	All	6	0	100
		0.80	4	8	1	80
D3	1	0.40	All	13	0	100
		0.50	All	9	0	100
		0.60	All	15	0	100
D4	1	0.40	All	7	0	100
		0.50	All	6	0	100
		0.60	All	7	0	100

TABLE 3: Simulation results on different tests

	ICI-CDT				SD-EWMA			
	# TP	# FP	#FN	AD	# TP	# FP	#FN	AD
D1	6	0	3	35	9	1	0	0
D2	1	0	4	60	5	6	0	0
D3	1	0	0	80	1	9	0	0
D4	1	0	0	60	1	6	0	0

To assess the performance of SD-EWMA against other shift detection methods, the ICI-CDT [6] is chosen because it is non-parametric sequential change-point detection test. Table 3 compares the results of ICI-based test and SD-EWMA in terms of the true positive, false positive, false negative and an average delay. The SD-EWMA provides better performance in terms of shift-detection without the delay compared to the ICI-CDT method. In the SD-EWMA, the numbers of true positives are much better for D1 and D2 datasets and for D3 and D4 there is only one shift which is detected by both the tests. The ICI-CDT test performs better by having no false positive for all datasets because of its hierarchal structure and validates its results of false positive from initial detection by another hypothesis test i.e., Hotelling's T-square test. In our SD-EWMA, we have applied the test only once and considered that detecting the true positive with no false negative is of paramount interest rather than running two different hypotheses tests for reducing the number of false positives. Another, advantage of the SD-EWMA is detecting the shifts without the delay.

On the real-world dataset, for better visualization purposes of SD-EWMA test, we have taken a window of 2000 samples and applied the test, it can be easily seen from the Figure 4 that the observations of an EEG signal are moving up and down as per the behavior of the user and correspondingly the UCL and LCL are computed, if the observation moves out of the control limits, the shift is detected in the non-stationary process. In the plot, the EEG signal is continuously moving in-between the UCL and LCL, the distribution of the time-series is not changing, so no shift is detected. Now, to perform the shift detection test, we have taken data from session one, it contains 70 trials and the parameters are calculated in training phase to be used in the testing/operational phase. Next, the test is applied to the sessions two and three and the results are given in Table 4.

Table 4 shows the result obtained from the test phase, as the number of shifts detected in each session. The initial statistics are calculated from the session-one of the training data and we have assumed that the process is in stationary state.

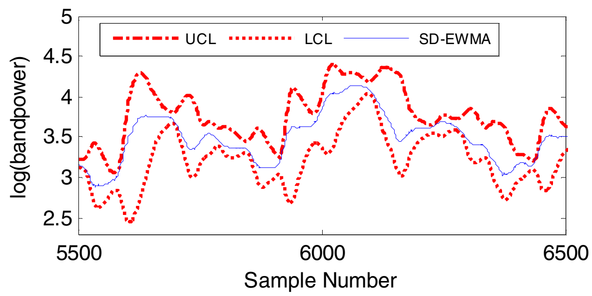


Figure 4: A window of 2000 samples obtained from real-world dataset.

TABLE 4: SD-EWMA shift detection in BCI data

Lambda (λ)	Number of Trials	# Shifts points SD-EWMA	
		Session 2	Session 3
0.01	1-20	*0*	1
	21-45	3	1
	46-70	4	2
0.05	1-20	*10*	4
	21-45	9	6
	46-70	7	5
0.10	1-20	16	9
	21-45	22	21
	46-70	25	17

Next, the test is applied to the sessions two and three, to detect the points of shift in the time-series. For various values of λ, the tests have been performed; the results show that the detection of shift in the time-series data depends upon the value of the smoothing constant. As, the value of the smoothing constant λ increases the number of shift detection increases because the test will become more sensitive to small shift. For evaluation purposes, we have performed the test on a fixed number of trials from each session and monitored the points where the shifts are detected. For the session-two, trials 1-20, there is no point where the shift is detected in the data with the value of λ=0.01. With increasing value of $\lambda =$ 0.05, the number of shifts increases to 10. In the session three with $\lambda = 0.01$, trials 1-20 and 21-45, the number of shifts is 0. But, for trials 46-70, the number of shifts is 2. Considering the EEG signals, no comparison can be done because the results are not provided in [20] for the shift points.

We have assumed that the data from session 1 is in stationary state. So, an optimal value of the smoothing constant λ is to be chosen from the training dataset. Next, we have tested several values of λ, in-between (0.01-0.1). For smaller value of λ, the test results in less false positive. As the value of lambda increases, the number of false positive increases.

Thus the smaller value of λ is a better choice for shift detection in EEG based BCI, as it avoids shift detections resulting from noise or spurious changes through much more intense smoothing of the EEG signal. Moreover, for correlated data, the smaller values of λ produce smaller prediction errors, thereby resulting in smaller estimated standard error. If λ is too small, the performance of the control chart results in less false alarm rate. In summary, the experimental results demonstrate that SD-EWMA control chart works well for shift detection in the non-stationary environments.

VI. CONCLUSION

This paper presented a method of SD-EWMA for detecting the shift in data stream based on the exponentially weighted moving average chart. Our method is focused on auto-correlated data, which contain non-stationarities. This algorithm can be used with classification method for detecting the shift in data generating process and based on the shift detection point necessary corrective action could be taken to account for the dataset shift. This method is computationally efficient because it does not require an additional memory to store the data. Experimental analysis shows that the performance of the approach is good in a range of non-stationary situations. This work is planned to be extended further by employing it into pattern recognition problems involving multivariate data and an appropriate classifier.

REFERENCES

[1] Y. Kawahara and M. Sugiyama, “Sequential Change-Point Detection Based on Direct Density-Ratio Estimation,” Statistical Analysis and Data Mining, vol. 5, no. 2, pp. 114–127, Apr. 2012.

[2] S. Liu, M. Yamada, N. Collier, and M. Sugiyama, “Change-Point Detection in Time-Series Data by Relative Density-Ratio Estimation,” Neural Networks, pp. 1–25, 2013.

[3] H. Shimodaira, “Improving Predictive Inference Under Covariate Shift by Weighting the Log-Likelihood Function,” Journal of Statistical Planning and Inference, vol. 90, no. 2, pp. 227–244, Oct. 2000.

[4] P. Angelov, Autonomous Learning Systems:From Data to Knowledge in Real Time. John Willey and Sons, Chichester., 2012, pp. 116–118.

[5] M. Sugiyama, “Learning under Non-stationarity : Covariate Shift Adaptation by Importance Weighting,” in Handbook of Computational Statistics: Concept and Methods, 2nd ed., Springer, 2012, pp. 1–27.

[6] C. Alippi, G. Boracchi, and M. Roveri, “Change detection tests using the ICI rule,” The 2010 International Joint Conference on Neural Networks (IJCNN), pp. 1–7, Jul. 2010.

[7] C. Alippi and M. Roveri, “Just-in-Time Adaptive Classifier--Part I: Detecting Nonstationary Changes,” IEEE Transactions on Neural Networks, vol. 19, no. 7, pp. 1145–1153, Jul. 2008.

[8] M. Basseville and I. Nikiforov, Detection of Abrupt Changes: Theory and Application. Prentice-Hall, 1993.

[9] C. Alippi, G. Boracchi, and M. Roveri, “A Just-In-Time Adaptive Classification System Based on the Intersection of Confidence Intervals Rule,” Neural networks : the official journal of the International Neural Network Society, vol. 24, no. 8, pp. 791–800, Oct. 2011.

[10] F. Gustafsson, “The Marginalize Likelihood Ratio st for Detecting Abrupt Changes,” IEE Transactions on Automatic Control, vol. 41, no. 1, 1996.

[11] M. Sugiyama, T. Suzuki, and T. Kanamori, Density Ratio Estimation in Machine Learning. Cambridge University Press, 2012, p. 344.

[12] A. J. Storkey, “When Training and Test Sets are Different : Characterising Learning Transfer,” in Dataset Shift in Machine Learning, no. Section 12, Press, MIT, 2010, pp. 3–28.

[13] J. G. Moreno-Torres, T. Raeder, R. Alaiz-Rodríguez, N. V. Chawla, and F. Herrera, “A Unifying View on Dataset Shift in Classification,” Pattern Recognition, vol. 45, no. 1, pp. 521–530, Jan. 2012.

[14] M. Sugiyama, A. Schwaighofer, and N. D. Lawrence, Dataset Shift in Machine Learning. The MIT Press, 2009.

[15] Roberts S W, “Control chart tests Based on Geometric Moving Averages,” Technometrics, 1959.

[16] W. A. Shewhart, Statistical Method from the Viewpoint of Quality Contro. Dover Publications, 1939.

[17] C. M. Douglas, Introdunction to Statistical Quality Control, 5th ed. John Wiley & Sons, 2007.

[18] N. Ye, S. Vilbert, and Q. Chen, “Computer Intrusion Detection Through EWMA for Autocorrelated and Uncorrelated Data,” IEEE Transaction On Reliability, vol. 52, no. 1, pp. 75–82, 2003.

[19] M. Connie, C. Douglas, and C. George, “Robustness of the EWMA control chart to non-normality,” Journal of Quality Technology, vol. 31,, no. 3;, p. pg. 309, 1999.

[20] B. B. Klaus-Robert Müller, “BCI Competition III: Data set IVb,” 2005. [Online]. Available: http://www.bbci.de/competition/iii/desc_IVb.html.