EWMA Based Two-Stage Dataset Shift-Detection in Non-stationary Environments

Haider Raza, Girijesh Prasad, Yuhua Li

EWMA based Two-Stage Dataset Shift-Detection in Non-Stationary Environments

Haider Raza, Girijesh Prasad, Yuhua Li

Intelligent Systems Research Center, University of Ulster, UK
raza-h@email.ulster.ac.uk, {g.prasad, 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. In a time-series data, detecting the dataset shift point, where the distribution 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 a novel method to detect the shift-point based on a two-stage structure involving Exponentially Weighted Moving Average (EWMA) chart and Kolmogorov-Smirnov test, which substantially reduces type-I error rate. The algorithm is suitable to be run in real-time. Its performance is evaluated through experiments using synthetic and real-world datasets. Results show effectiveness of the proposed approach in terms of decreased type-I error and tolerable increase in detection time delay.

Keywords: Non-stationary, Dataset shift, EWMA, Online Shift-detection.

1 Introduction

In the research community of statistics and machine learning, detecting abrupt and gradual changes in time-series data is called shift-point detection [1]. Based on the delay of the detection, shift-point detection methods can be classified into two categories: retrospective detection and real-time detection. The retrospective shift-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 shift-point is of paramount importance. Also, in the real-time systems, each observation coming in the data stream may be processed only once and then discarded; hence retrospective shift-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, robot navigation, remote sensing and spam-filtering. 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 conventional classification and streaming classification is that, in a conventional classification problem, the input data distribution is assumed not to vary with time. In a streaming case, the input data distribution may change over time during the testing/operating phase [2]. Based on the detected shift, a classifier needs to account for the dataset shift. 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 the next section.

Some pioneering works [3][4] have demonstrated 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) [5], Computational-Intelligence CUSUM (CI-SUSUM) [4], Intersection of Confidence Interval rule (ICI) for change-point detection [6]. However, most of these methods rely on pre-designed parametric models such as underlying probability distribution and auto-regressive models for tracking some specific statistics such as the mean, variance, and spectrum. 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 [7]. Thus, the aforementioned methods are not robust against different types of the shift-detection tests because of the delay and the high rate of type-I error (i.e., false-positive). On account of the need for identifying models, this may significantly limit their range of applications in fast data streaming problems.

In this paper we present an approach consisting of a Two-Stage Dataset Shift-detection based on an Exponentially Weighted Moving Average chart (TSSD-EWMA). This TSSD-EWMA in an extension of our work on Dataset Shift-Detection based on EWMA (SD-EWMA) [8], which suffers from the high false-positive rate. In TSSD-EWMA, the stage-I consists of a shift-detection test that activates the stage-II when a shift is alarmed at the stage-I. At stage-II, the suspected shift from stage-I is validated and is either confirmed as a shift or declared as a false alarm. The structure of test is given in Figure 1. It is demonstrated to outperform other approaches in terms of non-stationarity detection with a tolerable time delay and decreased false-positive alarms. So, this scheme can be deployed along with any classifier such as k-nearest neighbor 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 details the shift-detection algorithm and the structure of two-stage test. Section 4 presents the datasets used in the experiment. Finally, Section 5 presents the results and discussion.

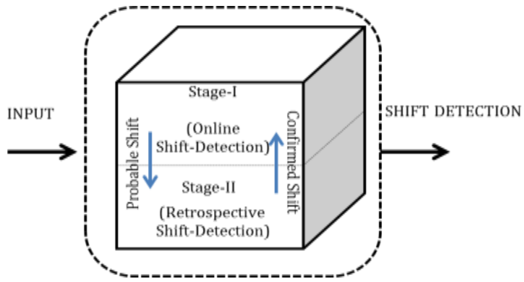

Fig. 1. The structure of the proposed two-stage shift-detection test

2 Background

2.1 Dataset shift

The term dataset shift [9] 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 x , a target variable y , the joint distribution $P(y, x)$, the prior probability $p(x)$ and conditional probability $p(y|x)$, 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 $P_{train}(y, x) \neq P_{test}(y, x)$"[10]. 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 briefly describe these shifts in following subsections.

Covariate Shift: It is defined as a case, where $(p_{train}(y|x) = p_{test}(y|x))$ *and* $(p_{train}(x) \neq p_{test}(x))$ [9] (e.g., the input distribution often changes in the session-to-session transfer of an EEG based brain-computer interface).
Prior Probability Shift: It is defined as a case, where $(p_{train}(x|y) = p_{test}(x|y))$ *and* $(p_{train}(y) \neq p_{test}(y))$ [9] (e.g., the survivorship in a stock market where there is bias in obtaining the samples).
Concept Shift: It is defined as a case, where $(p_{train}(y|x) \neq p_{test}(y|x))$ [9] (e.g., fraud detection).

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.

2.2 EWMA control chart

An exponentially weighted moving average (EWMA) control chart [11] is from the family of control charts in the statistical process control (SPC) theory. 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. The exponentially weighted moving average model is defined as

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

where λ is the smoothing constant $(0 < \lambda \leq 1)$ and z is a EWMA statistics. 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 statistics 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 [12]. That is, if $\hat{x}_{(i+1)}(i)$ is the forecast of the observation in the period $(i + 1)$ made at the end of the period i, then, $\hat{x}_{(i+1)}(i) = z_i$, where equation (1) is the EWMA. The 1-step-ahead error $err_{(i)}$ are calculated as [13]:

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

where $\hat{x}_{(i)}(i-1)$is the forecast made at the end of period $(i - 1)$. Assume, that the 1-step-ahead prediction errors $err_{(i)}$ are normally distributed. It is given in [12], 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 statements by substituting the right hand side of equation (3) as given below.

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

where σ_{err} is the standard deviation of the errors, L is the control limit multiplier and (1-α) is a confidence interval. The standard deviation of 1-step-ahead error can be estimated in various ways such as mean absolute deviation or directly calculate the smoothed variance [12]. 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).

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

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* [12]. The EWMA control chart is robust to the non-normality assumption if properly designed for the *t* and gamma distributions [14]. Following the above assumption, we have designed a two-stage algorithm for the shift detection in the process observation of auto-correlated data, which is discussed in the next section. So, the two-stage SD-EWMA (TSSD-EWMA) test can be employed when there is concern about the normality assumption.

3 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 proposed two-stage shift-detection based on EWMA (TSSD-EWMA) test works at two-stages. In the stage-I, the method employs a control chart to detect the dataset shift in the data stream. The stage-I works in an online mode, which continuously process the upcoming data from the data stream. The Stage-II uses a statistical hypothesis test to validate the shift detected by the stage-I. The stage-II operates in retrospective mode and starts validation once the shift is detected by the stage-I.

3.1 Stage-I

At the stage-I, the test works in two different phases. The first phase is training phase and the second phase is an operation or testing phase. In the first phase, the parameters $(\lambda, z_{(0)}, {\sigma_{err}}_{(0)}^{2})$ are calculated to decide the null hypothesis that there is no shift in the data. To calculate the parameters, first obtain the sequence of observations and calculate the mean. Use the mean as the initial value $z_{(0)}$ and obtain the EWMA statistics by equation (1). The sum of the squared 1-step-ahead prediction error divided by the length of the training dataset is used as an initial value of ${\sigma_{err}}_{(0)}^{2}$ for the testing data. The test has been performed on the several values of λ, chosen as suggested in [12], [14] and 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 algorithm TSSD-EWMA in Table 1. Next, check if each observation $x_{(i+1)}$ falls within the control its $\left[UCL_{(i+1)}, LCL_{(i+1)}\right]$, otherwise the shift is detected and alarm is raised at the stage-I. Furthermore, the shift detected by the stage-I is passed to the stage-II for validation.

3.2 Stage-II

In stage-II, the shift detected by the stage-I needs to be validated. This phase works in retrospective mode and it executes only when a shift is detected at stage-I. In particular, to validate the shift detected by the stage-I, the available information need to be partitioned into two-disjoint subsequences and then the statistical hypothesis test is applied. In literature, the statistical hypothesis tests are well established. The two-sample Kolmogorov-Smirnov test [15] is used to validate stationarity in the subsequences because of its non-parametric nature. The two-sample Kolmogorov-Smirnov test returns a test decision for the null hypothesis that the data in the subsequences are stationary with equal means and equal but unknown variances. The Kolmogorov-Smirnov statistics is briefly described as follows:

$$D_{n,n'} = sup_x \left|F_{1,n}(x) - F_{2,n'}(x)\right|$$

where sup_x is the supremum and $F_{1,n}(x)$ and $F_{2,n'}(x)$ are the empirical cumulative distribution function on the first and second samples respectively. The n and n' are the lengths of the two samples given as $n = \big((i-(m-1)):i\big)$ and $n' = ((i+1):(i+m))$ where m is the number of observations in the window of data used for the test. The null hypothesis is rejected at level α and (H=1) is returned if,

$$\sqrt{\frac{n.n'}{n+n'}}D_{n,n'} > K_\alpha$$

where K_α is the critical value and can be found in, e.g., [16].

Table 1. Algorithm-TSSD-EWMA

```
Input: A process x(i), generates independent and identically dis-
tributed observations over time (i).
Output: Shift-detection points
                              Stage-I
 Training Phase
   1. Initialize training data x(i) for i=1:n, where n is the num-
      ber of observations in training data
   2. Calculate the mean of x(i) and set as z(0)
   3. Compute the z-statistics for each observation x(i)
                 z(i)=lambda*x(i)+(1-lambda)*z(i-1)
   4. Compute the 1-step-ahead prediction error: err(i)=x(i)-z (i-
      1)
   5. Estimate the variance of error for the testing phase
   6. Set lambda by minimizing squared prediction error
 Testing Phase
   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)-z(i-1)
   4. Estimate the variance sigma_hat_err(i)^2=phi*err(i)^2 + (1-
      phi)* sigma_hat_err(i-1)^2
   5. Compute UCL(i+1) and LCL(i+1)
   6. IF(LCL(i+1)< x(i+1)< UCL(i+1))
         THEN (Continue processing)
      ELSE (Go to Stage-II)
                              Stage-II
   1. For each x(i+1)
   2. Wait for m observations after the time i, organise the se-
      quential observations around time i into two partitions, one
      containing x((i-(m-1)):i), another x((i+1):(i+m)).
   3. Execute the hypothesis test on the partitioned data
      IF(H==1)
         THEN (test rejects the null hypothesis): Alarm is raised
      ELSE(The detection received by stage-I is a false-positive)
```

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 by the TSSD-EWMA test. In

Table 1, the designed algorithm TSSD-EWMA is presented in the form of pseudo code.

4 Dataset and Feature Analysis

To validate the effectiveness of the suggested TSSD-EWMA test, a series of experimental evaluations have been performed on three 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 [1] 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 the same as the toy dataset given in [1], 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

$$\sigma = \begin{cases} 1 & N = 1,3, \ldots \ldots ,49 \\ \ln(\mathrm{e} + \dfrac{N}{4}) & N = 2,4, \ldots 48 \end{cases}$$

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 $N(x; 1,1)$ to $\mathcal{N}(x; 3,1)$, where $\mathcal{N}(x; \mu, \sigma)$ denotes the normal distribution with mean and standard deviation respectively.

4.2 Real-world Dataset

The real-world data used here are from BCI competition-III (IV-b) dataset [17]. 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. We have selected a single channel (C3) and performed the band-pass filtering on the mu (μ) band (8-12 Hz) to obtain bandpower features. This real-world dataset is a good example to validate if the proposed TSSD-EWMA is able to detect the types of shifts in the data generating process with decreased false-positive. The results and discussion of the experiments are given in the next section.

5 Results

On each dataset, the proposed TSSD-EWMA technique is evaluated by the following metrics 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); *True-positive (TP)*: it counts the times a test detects a shift in the sequence when it is there i.e., (hit alarm); *True*-negative (TN): it counts the times a test does not detect the shift when it is not there; *False-negative (FN)*: it counts the times a test does not detect a shift in the sequence when it is there i.e., (miss); *Not applicable (NA)*: it denotes not applicable situation, where the dataset is not suitable to be executed on the test; *Average delay (AD)*: it measures the average delay in shift-detection, i.e., sum of delay in each shift-detection is divided by the number of shifts detected; *Accuracy (ACC)*: it measures the accuracy of the results i.e.,

$$ACC = \frac{(\#True\ positive + \#Ture\ Negative)}{\left((\#True\ positive + \#False\ negative) + (\#False\ positive + \#True\ Negative)\right)}$$

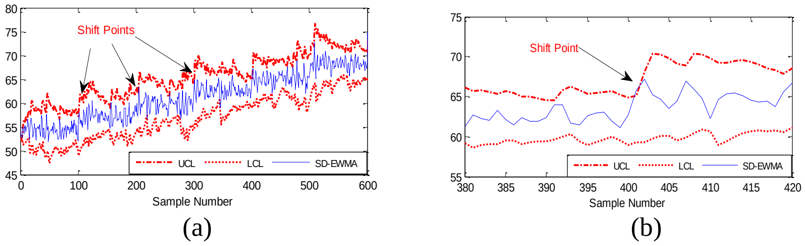

Fig. 2. Shift-detection based on TSSD-EWMA**:** Dataset 1 (jumping mean): (a) the shift point is detected at every 100th point. (b) Zoomed view of figure a: shift is detected at 401st sample.

It is important to note that the stage-I of the TSSD-EWMA is based on the current observation of the data to detect the shift. So, it is coming without the delay and this is the advantage of this approach that the stage-I of the test always operate in online mode. Once, the shift is detected by the stage-I, the test moves into the retrospective mode to validate the suspected shift in the process. Figure 2 represents the stage-I of the TSSD-EWMA based shift-detection test results from the Dataset 1. 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 detected. The tests have been performed on several values of λ as suggested in [12], [14].

According to Table 2, for the case of D1 with $\lambda = 0.40$, the TSSD-EWMA gives an optimal result with all the shift-detections and no false-positive, however with a delay of 10 samples, whereas, SD-EWMA detects four false-positive with no delay. The delay in TSSD-EWMA is resulting from the stage-II of the test. For the

D2, the TSSD-EWMA is not applicable because of the nature of the dataset. For D3 in the TSSD-EWMA, when $\lambda = 0.50$ it detects all the shifts with no false-positive and a low delay of 10 samples, whereas, the SD-EWMA test suffers twelve false-positives and no delay.

Table 2. TSSD-EWMA shift-detection

			SD-EWMA				TSSD-EWMA				
	Total shifts	**Lambda (λ)**	**# TP**	**# FP**	**# FN**	**ACC**	**# TP**	**# FP**	**# FN**	**AD**	**ACC**
D1	9	0.20	8	1	1	99.8	8	0	1	10	99.9
		0.30	8	3	1	99.6	8	0	1	10	99.9
		0.40	All	4	0	99.6	All	0	0	10	100
D2	5	0.60	All	11	1	99.8	NA	NA	NA	NA	NA
		0.70	All	6	0	99.4	NA	NA	NA	NA	NA
		0.80	4	8	1	99.1	NA	NA	NA	NA	NA
D3	1	0.40	All	13	0	99.1	All	0	0	10	100
		0.50	All	12	0	99.2	All	0	0	10	100
		0.60	All	15	0	99.0	All	0	0	10	100

To assess the performance and compare the results of the TSSD-EWMA, we have chosen other shift-detection methods such as the SD-EWMA [8] and the ICI-CDT [3] because these are state-of-the-art non-parametric sequential shift-point detection tests. Table 3 compares the results of ICI-CDT and SD-EWMA with TSSD-EWMA. The rate of false-positive is much reduced for D1 and D3 datasets. However, in the case of D2 the test is not applicable because of the nature of dataset. The delay in the shift-detection for the TSSD-EWMA is less than that for the ICI-CDT method.

TABLE 3**:** Comparison on different shift-detection tests

	ICI-CDT				SD-EWMA				TSSD-EWMA			
	# TP	**# FP**	**# FN**	**AD**	**# TP**	**# FP**	**# FN**	**AD**	**# TP**	**# FP**	**# FN**	**AD**
D1	6	0	3	35	9	4	0	0	9	0	0	10
D2	1	0	4	60	5	6	0	0	NA	NA	NA	NA
D3	1	0	0	80	1	12	0	0	1	0	0	10

TABLE 4**:** TSSD-EWMA shift-detection in BCI data

Lambda (λ)	**Number of Trials**	**# Shift-points SD-EWMA**		**# Shift-points TSSD-EWMA**	
		Session 2	**Session 3**	**Session 2**	**Session 3**
0.01	1-20	*0*	0	0	0
	21-45	3	1	2	0
	46-70	4	2	3	1
0.05	1-20	*10*	4	6	3
	21-45	9	6	6	4
	46-70	7	5	6	4
0.10	1-20	16	9	8	6
	21-45	22	21	15	15
	46-70	25	17	17	13

To perform the shift-detection test, on the real-world dataset we have made use of bandpower features from the EEG data of session one from the aforementioned BCI competition-III (IV-b) and assumed that it is in stationary state. 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, wherein TSSD-EWMA is compared with the SD-EWMA. 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 value of λ=0.01, and the trials 21-45 and 46-70, one false-positive is reduced by the TSSD-EWMA in each case. With increased value of $\lambda = 0.05$, the number of shifts increased, so the possibility of getting the number of false-positive also increased. Hence, by executing the TSSD-EWMA, the number of false alarms decreased. In case of the EEG signals, no comparison can be done because the results of actual shift points are not provided in [17].

6 Discussion and Conclusion

According to Table 3, the TSSD-EWMA provides better performance in terms of shift-detection with reduced number of false-positives compared to other state-of-the-art methods such as the SD-EWMA[8] and ICI-CDT [3], and it also has much smaller time delay compared to ICI-CDT. The TSSD-EWMA test outperforms over other methods in terms of low false-positive rate for all datasets because of its two-stage structure.

The choice of smoothing constant λ is an important issue in using EWMA control charts. The TSSD-EWMA results show that the detection of shift in the time-series data depends upon the value of the λ. The value of λ can be obtained by minimizing the least mean square prediction error on the datasets. However, selecting the value of λ by minimizing the least mean square prediction error does not always lead to a good choice. Moreover, if the value of the smoothing constant is large the test becomes more sensitive to spurious small shifts and it contains more number of false-positives. Hence, there is a trade-off between the smoothing constant and shift detection.

For the real-world dataset, we have assumed that the data from BCI session 1 is in stationary state and investigated for an optimal value of the smoothing constant λ considering the session 1 data as the training dataset. We have tested several values of λ in the range of (0.01-0.1). As the value of lambda increases, the number of detected shifts 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 test results in less false-positive rate but it tends toward getting more false-negatives, because chance of missing the shift is increased by much smoothing of the EEG signal. In summary, the experimental results demonstrate that TSSD-EWMA based test works well for shift-detection in the non-stationary environments.

This paper presented a method of TSSD-EWMA for detecting the shift in data stream based on the two-stage detection architecture. The advantage of using this method over other methods is primarily in terms of having a reduced number of false-positive detections. Our method is focused on auto-correlated data, which contain non-stationarities. 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. S. Liu, M. Yamada., N. Collier., M. Sugiyama.: Change-Point Detection in Time-Series Data by Relative Density-Ratio Estimation. In: Neural Networks, pp. 1–25, (2013).
2. H. Shimodaira.: Improving Predictive Inference Under Covariate Shift by Weighting the Log-Likelihood Function. In: Journal of Statistical Planning and Inference, vol. 90, no. 2, pp. 227–244, Oct. (2000).
3. C. Alippi,. G. Boracchi,. and M. Roveri.: Change Detection Tests Using the ICI Rule. In: The International Joint Conference on Neural Networks (IJCNN), pp. 1–7, Jul. (2010).
4. C. Alippi,. M. Roveri.: Just-in-Time Adaptive Classifier--Part I: Detecting Nonstationary Changes. In: IEEE Transactions on Neural Networks, vol. 19, no. 7, pp. 1145–1153, Jul. (2008).
5. M. Basseville,. I. Nikiforov.: Detection of Abrupt Changes: Theory and Application. Prentice-Hall, (1993).
6. C. Alippi,. G. Boracchi,. M. Roveri.: A Just-In-Time Adaptive Classification System Based on the Intersection of Confidence Intervals Rule. In: Neural networks: The Official Journal of the International Neural Network Society, vol. 24, no. 8, pp. 791–800, Oct. (2011).
7. M. Sugiyama,. T. Suzuki,. T. Kanamori.: Density Ratio Estimation in Machine Learning. Cambridge University Press, pp. 344, (2012).
8. H. Raza, G. Prasad, and Y. Li, “Dataset Shift Detection in Non-Stationary Environments using EWMA Charts,” in IEEE International Conference on Systems, Man, and Cybernetics, 2013 (Accepted).
9. J. G. Moreno-Torres,. T. Raeder,. R. Alaiz-Rodríguez,. N. V. Chawla,. F. Herrera.: A Unifying View on Dataset Shift in Classification. In: Pattern Recognition, vol. 45, no. 1, pp. 521–530, Jan. (2012).
10. M. Sugiyama., A. Schwaighofer., N. D. Lawrence.: Dataset Shift in Machine Learning. The MIT Press, (2009).
11. Roberts S W.: Control chart tests Based on Geometric Moving Averages. In: Technometrics, (1959).
12. C. M. Dougla.: Introduction to Statistical Quality Control, 5th ed. John Wiley & Sons, (2007).
13. N. Ye,. S. Vilbert,. Q. Chen.: Computer Intrusion Detection Through EWMA for Autocorrelated and Uncorrelated Data. In:IEEE Transaction On Reliability, vol. 52, no. 1, pp. 75–82, (2003).
14. M. Connie,. C. Douglas,. C. George.: Robustness of the EWMA Control Chart to Non-Normality. In: Journal of Quality Technology, vol. 31,, no. 3;, p. pg. 309, (1999).
15. G. W. Snedecor,. W. G. Cochran.: Statistical Methods, Eight. Iowa State University Press, (1989).
16. “Table of Critical Values for The Two-Sample Test.” [Online]. Available: http://www.soest.hawaii.edu/wessel/courses/gg313/Critical_KS.pdf
17. B. B. Klaus-Robert Müller.: BCI Competition III: Data set IVb. (2005). [Online]. Available: http://www.bbci.de/competition/iii/desc_IVb.html