ONLINE PARAMETRIC AND NON PARAMETRIC DRIFT DETECTION TECHNIQUES FOR STREAMING DATA IN MACHINE LEARNING: A REVIEW

P.V.N. Rajeswari[1], Dr. M. Shashi[2]
[1]Assoc. Professor, Dept of Comp. Sc. & Engg.,
Visvodaya Engineering College, Kavali, AP, India
[2]Professor, Dept of Comp. Sc. & System Engg,
Andhra University, Visakhapatnam, AP, India

Abstract
Enormous amount of data streams are generated by large number of real time applications with the stupendous advancement in technology. Incremental and online - offline learning algorithms are more relevant to process the data streams in the data mining context. Detecting the changes and reacting against them is an interesting research area in current era in knowledge discovery process, in Machine Learning. The target function also is changed frequently, due to the occurrence of changes in data streams. This is an intrinsic problem of online learning, popularly known as Concept Drift. To handle the problem despite of the learning model, novel methods are needed to examine the performance metrics. This examination will be done during the learning process and prompt drift signals whenever a significant variation has been detected. The Concept Drift problem can be detected using various techniques. This paper focuses on reviewing on existing Parametric and non-Parametric techniques Online for drift detection.
Keywords: Stream mining; Concept drift; Parametric; Non-parametric; Machine Learning;

I. INTRODUCTION

Adaptation of science and technological outcomes including computerization in every walk of life leads to accumulation of voluminous amount of data that need to be processed by machines for pattern extraction for its effective usage. Machine Learning algorithms are widely applied in the areas like, image processing, data mining, medical diagnosis, search engines, etc. Machine learning categorized into two types namely Batch learning and Online learning based on representation and presentation of training samples. Batch learning systems accept a large number of examples and learn to build a model from all of them at once. In contrast, Online learning systems are given examples sequentially as streaming data [1] from which learning happens incrementally.

Learning from data streams aims to capture the target concept as it changes over time, and it is a current research area of growing interest. For instance, changes can emerge due to changing interests of people towards news topics, shopping preferences, energy consumption, etc. (e.g. influenced by seasonal changes). Spam filtering [2] is another example, wherein spammers try to elude filters by disguising their emails as genuine with newer tricks challenging spam filters to continuously enhance their capability to successfully identify spam over time. So it is possible that a learning model previously induced may be incompatible with the present data, making an update mandatory. This type of problem is commonly known as Concept Drift. Many online learning algorithms[18], which are variants of the base models such as Rule-based systems, Decision Trees, Naive Bayes, Support Vector Machines, Instance based learning, and ensemble of classifiers [3] [4], have been implemented for handling Concept Drift.

In supervised incremental learning, a well-extended approach continuously monitors a performance measure such as accuracy, or speed etc., of the learning model to handle Concept Drift. A Concept Drift is assumed if a significant fall in this measure is estimated and required actions are defined to update the model according to the latest data available. In this scenario, automatic/ inherent change detectors of the learning algorithm play a crucial role [5] [6] [7]. These detectors often operate over a stream of real values corresponding to a given performance measure. Most existing approaches monitor changes in terms of a suitable statistic (such as the mean or median) with standard deviation [8], to successfully capture the distributional changes associated with the stream of performance measures. Thus, the problem of concept drift detection is reduced to estimating significant changes in the statistic calculated from the sequence of values that measure a performance characteristic. Often, this stream of real values is also huge (probably infinite), as the learning model is monitored over longer times.

Therefore, it is very common to enforce restrictions on these online change detectors [8], [9]. The computational complexity required to process each performance value must be constant as well as methods should be single-pass, where each performance value is processed once and then discarded. These change detectors must also able to deal with common types of change widespread in much real-world data [4]. Under these conditions, many traditional statistical approaches that assume a fixed size of the input data for estimating distributional changes are not suitable.

Some of the most studied parametric schemes to detect changes online are:

- Shewhart's control charts [17]
- Exponentially Weighted Moving Average (EWMA) [11]
- Control charts and Page's cumulative sum (CUSUM) [10]
- Procedure and Non-parametric approaches are Drift Detection Method (DDM) [5]
- Early Drift Detection Method (EDDM) [6]
- Adaptive WINdow (ADWIN) [7]
- Adaptive WINdow2 (ADWIN2) [7]
- Detection With a Statistical Test of Equal Proportions STEPD [16]
- Exponentially Weighted Moving Average For Concept Drift Detection (ECDD) [15].

The comparative study of all algorithms is provided in Table 1.

II. ONLINE PARAMETRIC DRIFT DETECTION TECHNIQUES

A. *Shewhart's Control Charts*

To verify the validity of manufacturing or business control state, Shewhart's control charts (widely known as Process Behavior charts) are used as statistical tools [17]. In other words, certain statistical assumptions about the process data it produces are used to test the steadiness of some of the process properties over time. Mean, median, variance or standard deviation, distribution shape or fraction of non-conforming items are the commonly considered properties in the charts.

Shewhart's control charts are used to continuously monitor any deviations from the current state. These are aimed to identify events, which are very unlikely when the controlled process is in stable condition. An alarm is used to signal when any change is noticed in the stability of the process, so that possible steps should be taken for investigating the causes of the change as well as for correction.

For instance, in case of the control limits UCL (Upper Control Limit) or LCL (Lower Control Limit)) are exceeded. The control limits are constructed at $\pm 3\sigma$ from the mean of the performance measure, which can be exceeded with relative frequency of 0.27%.

In addition to the LCL and UCL, warning limits for LWL and UWL are also constructed at $\pm 2\sigma$ and as well as $\pm\sigma$. Having multiple limits defined at various levels around the mean statistic, provides fine-grain control through more sophisticated rules specifying low probability event possessing when the process is under control.

Some of the following rules can be selected in the Shewhart control chart dialog panel; by default all available rules are selected [17]

- One point exceeds LCL/UCL
- Two out of three points exceed LWL or UWL limits
- Four out of five points are above/below the central line and exceed ± σ limits
- Six consecutive points show increasing/decreasing trend
- Eight consecutive values are beyond ± σ limits
- Nine points above/below the middle line
- Difference of consecutive values alternates in sign for fourteen points
- Fifteen points are within ± σ limits

Application of Shewhart Control Charts for monitoring a control process involves two steps namely Chart construction and Chart application.

The purpose of construction step is to specify the middle/ central line (CL) and control limits to describe the interval of real process correctly. When these values are not provided in advance, assuming normally distributed data, they are set to mean and interval which contains 99.3% of available training data, with the statistical characteristics. These are based on observations of the process under control, excluding outliers or otherwise suspect data points. The chart is applied to control the process, by using a particular relevant rule set

A software tool named Quality Control Expert (QC. Expert) offers seven common types of Shewhart control charts namely X-bar and S, X-bar and R , X-individual for continuous variables, np, p, u, c for discrete quality attributes Figure 1 shows an example of process variant chart borrowed from Wikipedia.

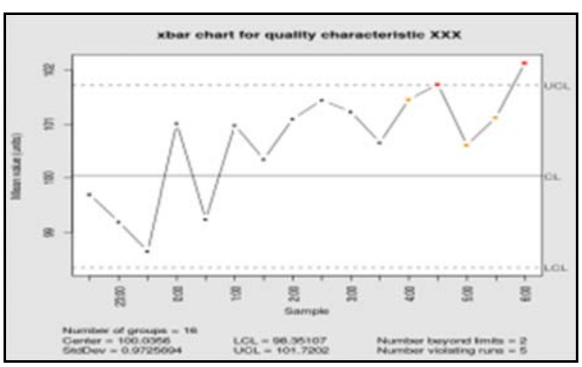

Figure 1: Process Control Chart Sample

B. CUSUM : Cumulative Sum Control Chart

CUSUM is commonly used as sequential analysis technique in statistical quality control. It is mainly used for monitoring abrupt or sudden change detection. In this, *'θ'* is used as the parameter of probability distribution referred to as "quality number". It is a technique to verify any changes in mean, median, etc., for deciding when to take corrective action. When the CUSUM [10] method is applied to changes in mean, it can be used for step detection of a time series. Alternative forms of CUSUM charts are given below:

- Direct form or Recursive forms
- One- or Two-sided forms

X[n] is defined as a discrete random signal with independent and identically distributed samples in CUSUM technique. Each sample follows a probability density function (PDF) represented as *$p(x[n], \theta)$* that depends on the 'θ' value known as 'deterministic parameter' i.e., it can be considered as either mean 'μ_x' or the variance 'σ^2_x' of X[n]. At the change time 'n', the discrete random signal may possess one abrupt change occurrence, indicated with 'c', and at the time n_c, the abrupt change is represented by an instant change in the value of 'θ'. Hence $\theta = \theta_0$ before n_c and $\theta = \theta_1$ from n_c in the current sample.

Therefore, according to the given hypothesis/ assumptions, the whole PDF of the signal p_x that is observed between the first sample x[0] and the current sample, i.e., x[k] can take two different types.

Let H_0 be the no change premise and the PDF of X[n] is given by:

$$Px|H_0 = \prod_{n=0}^{K} p(x[n], \theta_0)$$

And the change premise is represented as H_1, and the PDF is represented as:

$$Px|H_1 = \prod_{n=0}^{n_{c-1}} p(x[n], \theta_0) \prod_{n=n_c}^{k} p(x[n], \theta_1)$$

With the help of PDF of each incoming sample or example p(x[n], θ), it is possible to detect the abrupt change and also the values of the parameters before (θ_0) and after (θ_1), using this method. Abrupt changes for 1000 data values using CUSUM algorithm is shown in Figure 1. But in this method, some unknown values are also to be determined as mentioned below:

- There is a deficient in abrupt change rate between the values for $n = 0$ and $n = k$.
- The value of the possible change time n_c.

Following are the two different steps are performed in this algorithm:

1. Detection step: To decide the true hypothesis between H_0 and H_1.
2. Estimation step: To estimate the change time n_c, in case H_1 is true.

Here, H_0 and H_1 are the hypotheses for no change and change in the data respectively.

C. *EWMA : Exponentially Weighted Moving Averages Control Charts*

Like the CUSUM chart, even small deviation in the mean value of the process can be detected using EWMA control chart. These control charts observe the mean of a given process based on samples/data collected from the process at given time-intervals which span in terms of hours, days, weeks, or even months. Subgroups are constituted based on the measurements of the data samples taken.

The EWMA chart relies on the specification of a target value and on known or consistent approximate of the standard deviation of the input samples. For this rationale, the moving average chart is better used after process control has been established. This Exponential Moving Averages (EMA) can be obtained in two different behaviors named as Percent-based EMA and Period-based EMA.

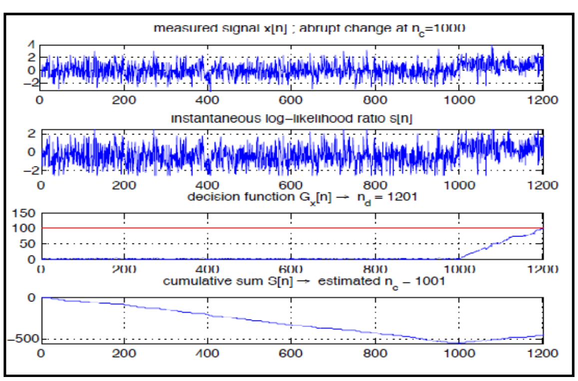

Figure 2 : Typical behavior of CUSUM algorithm using Guassian signal

Percentage is used as single parameter for percent-based EMA and duration is used as a parameter for period-based EMA, duration is the parameter. The formula for an EMA is:

EMA (present) = ((Price (present) – EMA (previous)) x (Multiplier) + EMA (previous).

The parameter “Multiplier” in the above equation has different notations in different EMAs. In percent-based EMA, specific percentage is used where as in period-based EMA, it is equal to $2 /(1 + N)$, where N is the specified number of periods.

The EWMA chart may also be used whenever there is only a single response is available at each time point. An alternative option for single responses is the Individuals and Moving Range (I-MR) control charts.

III. ONLINE NON-PARAMETRIC DRIFT DETECTION TECHNIQUES

A. *ADWIN : ADaptive WINdow*

An adaptive sliding window method, proposed by Bifet [12, 13], for detecting sudden drifts rapidly generating and voluminous data streams is termed as ADWIN. The working of the algorithm mainly depends on the parameter called as ‘W’value i.e., a sliding window for the most recently read data samples.

The main idea of ADWIN is as follows: whenever two ‘large enough’ sub-windows of size ‘W’ shows signs of different average values, one can conclude that the equivalent expected values are different from each other, and the older portion of the window is dropped. This gives the answer for the question whether “the average μ_t remained constant in W with confidence value δ” or not. The related pseudo-code for ADWIN [13] is given in Figure 3.

The important part of the algorithm lies in the definition of ‘ϵ_{cut}’ value and the test it is used for.

```
Input: S: a data stream of examples
       δ:confidence level
Output : W : a window of examples

initialize window W;
for all xi € S do
W ← W U { xi };
 repeat
drop the oldest element from W;
 until |µW0- µW1|< €cut holds for every
      split of  W into W=W0.W1
```

Figure 3 : ADWIN algorithm Psuedocode.

In this algorithm 'n' denote the size of W, where n0 and n1 are the sizes of W_0 and W_1 respectively.

So that n = n0+n1. Let, $\mu`w_0$ and $\mu`w1$ are the averages of W_0 and W_1 values, and also μw_0 and μw_1 as their predictable values.

The value of '€cut' is proposed as follows:

$$\epsilon_{cut} = \sqrt{\frac{1}{2m} \cdot \frac{1}{\delta`}}$$

where, $m = \frac{1}{\frac{1}{n_0}+\frac{1}{n_1}}$ and $\delta` = \frac{\delta}{n}$

But, ADWIN algorithm proved inefficient in terms of time and memory constraints.

B. ADWIN2 : ADaptive WINdow2

ADWIN2 is developed by Bifet and Gavalda [7] with the ideas from different data stream algorithms [20 - 23] to overcome the limitations of ADWIN algorithm in terms of time and memory. ADWIN algorithm computationally expensive as it verifies all "large enough" possible sub-windows of the current window exhaustively, for possible cuts which lead to increase in the memory size exponentially. The ADWIN2 algorithm retains a data structure with the following properties:

- For each available memory word which extends up to 'W' value, the space complexity results as O(M*log(W/M)).
- In O(1) amortized and with worst case value O(logW) it can process the arrival of a new element.
- It can provide the accurate counts of 1's for all the existed sub-windows whose lengths are of the form b(1+1/M), in O(1) time complexity per query.

ADWIN2 is an adaptive windowing scheme which is based on the use of the Hoeffding bound to detect concept change with less memory and time limits. ADWIN used a variation of exponential histograms and a memory parameter to limit the number of hypothesis tests done on a given window. ADWIN2 was shown to be superior to Gama's method and fixed size window with flushing [16] on almost all performance measures such as the false positive rate, false negative rate and sensitivity to slow gradual changes.

C. DDM : Drift Detection Method

The DDM method was proposed by Gama et al. [5] that includes Early Drift Detection Method (EDDM) [6], and Detection with a Statistical Test of Equal Proportions (STEPD) method [14], applies to single classifier learning methods. Only one classifier is used at any point of time in these methods, and old classifier is replaced with newly trained classifier after occurrence concept drift.

It is assured that error rate of the online classifies will decrease if the stability is maintained in the target concept even with the progression in time. Therefore, whenever there is a considerable increase in the error rate of the online classifier it's true that the concept is changing which results as concept drift. Warning level and drift level are the two types of thresholds mentioned to estimate the severity of the drift in inflow. Data examples or inflow are stored in short-term memory at the occurrence of warning level as well as all available samples are re-initialized at the occurrence of the drift. Also online classifier is remodeled automatically. DDM reacts fastly and positively when sudden and significant changes are detected, but reacts slowly and negatively for gradual and small changes. Sometimes DDM is ineffective because of short memory overflow and also sometimes incapable of detecting sudden change occurrences.

DDM employs binomial distribution, which is used to represent the form of underlying probability concept for concerned random variable. The random variable is meant to notate the total number of errors in 'n' samples. For each 'i', a sampled point in the input sequence, (pi), indicates the probability of misclassifying error rate and $si = \sqrt{pi(1-pi)/i}$ indicates the related standard deviation obtained. With (pi), value, it is assumed that the error rate of the learning algorithm will also decrease even with the increase in the number of samples, if distribution in examples stand stationary. It is stated in PAC (Probability Approximate Correct) learning model [10], [24]). So, with the considerable increase in the error of the algorithm, it suggests that the changing in the class distribution occurring and present decision learning model will become inappropriate for the current stream, and it checks when the following conditions set off:

i. $pi+si \geq p_{min}+2.s_{min}$ for the warning level
ii. $pi + si \geq p_{min} + 3. s_{min}$ for the drift level

Ahead of the drift level, occurrence of the changes in the concept is supposed to be true, then the model induced by the learning method need to be updated and a new learning model is learnt using the stored examples in the memory. And the values for p_{min} and s_{min} variables are reset as well. This approach suits well in detecting abrupt and gradual changes, when the gradual change is not very slow. But it may prove difficult in dealing when there is slow and gradual change. In that scenario, the stored examples are observed for long time and as a consequence triggering at drift level will happen after long time. In addition to that, the required memory for the examples can be exceeded which results need of more memory which proved as a disadvantage to this method.

D. EDDM : Early Drift Detection Method

EDDM was developed by Baena-Garcia et al. [6], to overcome the drawback in DDM, so that to improve the gradual concept drift detection. In parallel, it shows better performance in abrupt concept drift too. The basic scheme introduced in this technique is, instead of considering only the number of errors, the distance between two errors classification have to be taken into account for finding the drift. Predictions are measured by learning method and the distance between two errors will increase if there is drift. The adequate distance between two errors (represented as p^1_i) and its standard deviation (represented as s^1_i) are to be calculated. These values are to be stored if and only if $p^1_i+2.s^1_i$ reaches its maximum value by obtaining p^1_{max} and s^1_{max}. respectively. Thus, the value $p^1_{max}+2.s^1_{max}$ corresponds with the point where the distribution of distances between errors is maximum.

The method defines two thresholds too:

1. $p^1_i+2.s^1_i / p^1_{max}+2. s^1_{max} < \alpha$ for the warning level. There is a possibility of change in context further if the calculated value is more than this level. Ahead of this level, the samples are stored in advance of context change possibility.

2. $p^1_i+2.s^1_i / p^1_{max}+2.s^1_{max} < \beta$ indicates the drift level. It can be assured that the concept drift occurred and the learning model needs to be reset above this level. At the same time new model has to be built by using the examples stored since the warning level triggered.

The values for p^1_{max} and s^1_{max} also have to reset without any delay. The technique considers the thresholds and then searches occurrence of drift, if any, in given concept, only after the occurrence of a minimum of 30 errors.

E. ECDD : Exponentially Weighted Moving Average For Concept Drift Detection

Ross & Adams et al., proposed a drift detection method based on Exponentially Weighted Moving Average (EWMA) chart [11], and used for identifying an increase in the mean of a sequence of random variables to monitor the misclassification of the streaming data. ECDD method possess both single pass and computationally efficient with an overhead of O(1) properties. Here, means μ_0 and μ_1 are used as indications for, before and after change respectively. Change is occurred in the independent random variables $X_1 .....X_{n..,}$ where μ_t is the time of mean μ and σ_x is the standard deviation., The values of success and failure i.e., probability (1 and 0) are computed online using ECDD. The basic idea is on the classification accuracy of the base learner in the actual instance, together with an estimator of the expected time between false positive detections.

F. STEPD : Detection With a Statistical Test of Equal proportions

Nishida et al, proposed the STEPD technique [16], in which, monitoring of the predictive accuracy in single classifier is performed. Two different analytical accuracies are taken into account, one is which occurred recently and the other is overall one, in STEPD. The two accuracies are compared by using STEPD. In this, if the target concept is stationary, it is assumed that the accuracy of a classifier for recent W examples will be equal to the overall accuracy from the beginning of the learning; and whenever there is a significant decrease of recent accuracy indicates that there is a change in the concerned concept. To obtain the value for observed significant level, a chi-square test is performed by computing a statistic and the obtained value is compared to the percentile of the standard normal distribution. If this value is less than a significance level, then the null-hypothesis is rejected, assuming that a concept drift has occurred. In this, two significant levels viz. warning and drift are also used which are similar to the ones used in DDM, EDDM, PHT, and ECDD detection methods.

G. HDDM : Hoeffding Bound Drift Detection Method

HDDM method based on the design of statistical process control by using a Hoeffding inequality which assumes only independent and bounded random variables and no probability function is assumed [19]. Two important modules namely change detector and learning algorithm were proposed in HDDM for detection of the change in concepts. Loss function: $L(c^1{}_i, c_i)$ is used to detect changes in change detector module. Loss function results in two states:

- in-control ($L(c^1{}_i, c_i)=0$), if concept is stable
- out-control ($L(c^1{}_i, c_i)=1$), if changes are detected

In this technique, following two tests are performed:

- A-test or Bounded Moving Averages - implemented using window strategies approach
- W-test or Bounded Weighted Moving Averages -implemented using weighted instance approach.

The outcome of the status is presented in terms of three states termed as Stable, Warning and Drift state. Stable state occurs when there is no change in the generated input data stream. If any changes are observed in the streaming data with $Pr\{\mu-2\sigma \leq p \leq \mu+2\sigma\}\approx 0.95$, it results in a warning state as well as with $Pr\{\mu-2\sigma \leq p \leq \mu+2\sigma\}\approx 0.99$ results in drift level. Here 'p' is the statistic, computed from the last observed values and 'μ' *is* for expected value. Alternative classifier will replace the old classifier automatically with occurrence of the changes (drift) in the streaming data.

IV. CONCLUSION

In order to detect any occurrence of changes either in online or offline streaming data or samples i.e., whenever target functions are changing, a drift detection technique should be used to intimate the changes in the data which improves the accuracy of the classification process. As drift detection and classification algorithms are needed while changes occurs in data stream. There are numerous drift detection techniques like DDM, Drift detection through re-sampling, EDDM, ECDD, STEPD, DCDD, HDDM, etc.

Some of the drift detection techniques uses any of the windowing methods, statistical or ensemble methodologies. This Paper gives an overview about concept drift and different types of drift detection techniques which are classified in terms of Parametric and Non-Parametric methods for online process along with its requirements and details of changes.

Table 1: Comparative study of ONLINE & OFFLINE drift detection algorithms

S. No	Author	Method	Merit	Demerit
1	Walter A. Shewhart	Shewhart Control chart	Identifies when process is out of control by using mean, median or standard deviation methods.	It uses only the information from last sample provided & insensitive to small process shifts, violates likelihood principle and uses ARL(Average Run Lengths) for finding the deviation.
2	E. S. Page	CUSUM Control chart	Efficient for detecting small shifts in the process mean from shifts of 0.5 to 2 standard deviations from the target mean.	This CUSUM charts or schemes results as more complicated nature in design process. And it is slower to detect large shifts in the process mean.

3	S. S. Roberts	EWMA Control charts	Very efficient in detecting small shifts i.e., from .5 sigma to 2 sigma in process mean.	Slower in detecting large shifts in the process mean.
4	Bifet et al	ADWIN	Used to sudden drifts in voluminous data.	Gives high false positive rate and noise in-tolerance.
5	Bifet and Gavalda	ADWIN2	By using Hoeffding bounds results in less memory - time limits and guarantees on false positive and negative rates.	The drawback with this is, it is useful in dealing changes only for mean values.
6	Gama et al.	DDM	Well suits in detecting abrupt changes and gradual changes, but for only when the gradual change is not very slow	Need more memory and time when gradual change is very slow and triggers false alarms quickly and frequently
7	Baena-Garc et al.	EDDM	Proves efficient and improved for gradual drifts and it also keeps a better performance with sudden concept drift.	Less efficient for noisy data.
8	Ross & Adams et al	ECDD	Used for identifying an increase in the mean of misclassification of the streaming data. This method possess both single pass and computationally efficient with an overhead of O(1).	No guarantee of accurate performance.
9	Nishida et al	STEPD	Used to find the difference in accuracy on older and more recent training examples along with a statistical test.	The algorithm have to reset every time upon change detection.
10	Isvani Frıas Blanco and Jose del Campo Avila	HDDM	Very efficient for both abrupt and gradual changes detection using A-test & W-test in terms of WARNING, STABLE & DRIFT states.	In this, there is no distribution over the incoming samples.

V. REFERENCES

[1] Mohamed Medhat Gaber, Arkady Zaslavsky and Krishnaswamy, “Mining Data Streams: A Review”, SIGMOD Record, Vol. 34, No. 2, 2005.

[2] Christina V, Karpagavalli P, Suganya G: “A Study On Email Spam Filtering Techniques”, Intl. J. of Comp. Applications, 0975 – 8887 p, Vol. 12– Issue No.1, 2010.

[3] H. Wang, W. Fan, P. Yu, and J. Han, “Mining Concept-Drifting Data Streams using Ensemble Classifiers”, Proc. of 9th ACM SIGKDD Int. Conf. on Knowledge Discovery Data Mining, pp. 226– 235, 2003.

[4] J. Gama, I. Zliobaite, A. Bifet, M. Pechenizkiy, and Pouchachia, “A Survey on Concept Drift Adaptation, ACM Comp. Surv., Vol. 46, No. 4, 2014.

[5] J. Gama, P. Medas, G. Castillo, and P. Rodrigues, “Learning with drift detection”, in Proc. SBIA Adv. Artificial Intelligence, Vol. 3171, pp. 286–294, 2004.

[6] M. Baena, J. Del Campo, R. Fidalgo, A. Bifet, R. Gavalda, and R.Morales: “Early Drift Detection Method”, Proc. of 4th Intl. Conference on Knowledge Discovery Data Streams, pp. 77–86, 2006.

[7] A. Bifet and R. Gavalda, “Learning from Time-Changing Data with Adaptive Windowing”, Proc. of SIAM Intl. Conf. on Data Mining, pp. 443–448, 2010.

[8] G. Ross, D. Tasoulis, and N. Adams: “Non-Parametric Monitoring of Data Streams for Changes in Location and Scale, Technometrics”, Vol. 53, No. 4, pp. 379–389, 2011.

[9] M. Datar, A. Gionis, P. Indyk, and R. Motwani, “Maintaining Stream Statistics over Sliding Windows”, SIAM J. Comp., Vol. 31, No. 6, pp. 1794–1813, 2002.

[10] D. Montgomery, “Introduction to Statistical Quality Control”, 5th Edition, New York, USA: Wiley, 2001.

[11] E. Dalton, “EWMA charts”, NCSS Statistical Software chapter 249, pg 249-1 to 249-18, 2014.

[12] Albert Bifet and Ricard Gavald`a, Kalman: “Filters for an Adaptive Windows for Learning in Data Streams”, Discovery Science, Vol. 4265 of Lecture Notes in Comp. Sc., Springer, pages 29–40, 2006.

[13] Albert Bifet and Ricard Gavalda, “Learning from time changing data with adaptive window”,in SDM, SIAM, 2007 (Repeat of [7]).

[14] Nishida and K. Yamauchi, “Detecting Concept Drift Using Statistical Testing”, Proc. of 10th Intl. Conf. on Discovery Sc., pp. 264–269, 2007.

[15] G. J. Ross, N. M. Adams, D. Tasoulis, D. Hand, “Exponentially Weighted Moving Average Charts for Detecting Concept Drift”, Intl. J. of Pattern Recognition Letters, 191-198, 2014.

[16] K. Nishida, “Learning and Detecting Concept Drift, A Dissertation” Doctor of Philosophy in Info. Sc. and Tech., Graduate School of Info., 2008.

[17] M. B. (Thijs) Vermaat, Roxana A. Ion, Ronald J. M. M. Does1, and Chris A. J. Klaassen, “A Comparison of Shewhart Individuals Control Harts Based on Normal, Non-parametric, and Extreme- value Theory”, Quality & Reliability Engineering International, 2003.

[18] Joseph Carmona and Ricard Gavald,“Online Techniques for Dealing with Concept Drift in Process Mining”, Proc. of Int. Symp. on Data Analytics (IDA), pp 90-102, 2012.

[19] Isvani Frıas - Blanco Jose Del Campo, “Online Non – Parametric Drift Detection Methods on Hoeffding’s Bounds”, IEEE Transactions on Knowledge and Data Engg., Vol. 27, No.3, 2015.

[20] B. Babcock, S. Babu, M. Datar, R. Motwani, and J. Widom, “Models and Issues in Data Stream Systems”, Proc. of 21st ACM Symp. on Principles of Database Systems, 2002.

[21] B. Babcock, M. Datar, and R. Motwani, “Sampling from a Moving Window Over Streaming Data”, Proc. of 13th Annual ACM - SIAM Symp. on Discrete Algorithms, 2002.

[22] M. Datar, A. Gionis, P. Indyk, and R. Motwani, “Maintaining Stream Statistics over Sliding Windows”, SIAM J. on Computing, 14(1), pp 27–45, 2002.

[23] S. Muthukrishnan, “Data Streams, Alg’s and appl’s”, Proc. of 14th Annual ACM SIAM Symp. on Discrete Algorithms, 2003.

[24] Leslie Valiant, “Probability Approximate Correct (PAC), Universal Framework for Learning in SIAM News, Volume 46, No. 8 (2013).