WARES, S., ISAACS, J. and ELYAN, E. 2021. Burst detection-based selective classifier resetting. *Journal of information and knowledge management* [online], 20(2), article 2150027. Available from: https://doi.org/10.1142/S0219649221500271

# Burst detection-based selective classifier resetting.

WARES, S., ISAACS, J. and ELYAN, E.

2021

*Electronic version of an article published as WARES, S., ISAACS, J. and ELYAN, E. 2021. Burst detection-based selective classifier resetting. Journal of information and knowledge management [online], 20(2), articles 2150027. Available from: https://doi.org/10.1142/S0219649221500271 © copyright World Scientific Publishing Company. Journal homepage: https://www.worldscientific.com/worldscinet/jikm*

OpenAIR @RGU

This document was downloaded from https://openair.rgu.ac.uk

SEE TERMS OF USE IN BOX ABOVE

DISTRIBUTED UNDER LICENCE

# Burst Detection-based Selective Classifier Resetting

Scott Wares[1], John Isaacs[2], and Eyad Elyan[3]

[1] s.wares1@rgu.ac.uk
[2] j.p.isaacs@rgu.ac.uk
[3] e.elyan@rgu.ac.uk

**Abstract.** Concept drift detection algorithms have historically been faithful to the aged architecture of forcefully resetting the base classifiers for each detected drift. This approach prevents underlying classifiers becoming outdated as the distribution of a data stream shifts from one concept to another. In situations where both concept drift and temporal dependence are present within a data stream, forced resetting can cause complications in classifier evaluation. Resetting the base classifier too frequently when temporal dependence is present can cause classifier performance to appear successful, when in fact this is misleading. In this research a novel architectural method for determining base classifier resets, BD-SCR, is presented. **B**urst **D**etection-based **S**elective **C**lassifier **R**esetting (BD-SCR) statistically monitors changes in the temporal dependence of a data stream to determine if a base classifier should be reset for detected drifts. The experimental process compares the predictive performance of state-of-the-art drift detectors in comparison to the "No-Change" detector using BD-SCR to inform and control the resetting decision. Results show that BD-SCR effectively reduces the negative impact of of temporal dependence during concept drift detection through a clear negation in the performance of the "No-Change" detector, but is capable of maintaining the predictive performance of state-of-the-art drift detection methods.

**Keywords:** Data streaming · concept drift · temporal dependence.

## 1 Introduction

Concept drift detection is an area of data stream mining that has received considerable attention over the years. Various methods including statistical (Page, 1954, Gama et al., 2004, Baena-Garcıa et al., 2006, Yu & Abraham, 2017), ensemble (Street & Kim, 2001, Wang et al., 2003, Kolter & Maloof, 2003) and window based (Bifet & Gavalda, 2007, Domingos & Hulten, 2000, G. Liu et al., 2013) techniques have been published, adapted and evolved as result of research in the field. The notion and reasoning behind resetting a base classifier whenever a drift occurs is well known; as a drift occurs and the stream distribution shifts, the base classifier can become outdated and underperform. While there are numerous approaches for detecting and handling concept drift, the underlying architecture has remained unchanged.

More recent studies (Wares et al., 2019) have highlighted an issue that stems from the archaic formula of traditional concept drift, that is the problem of over-resetting a classifier when temporal dependence is present alongside concept drift. Temporal dependence is said to exist when arriving instances from a stream are not independent of the time of arrival, such that arriving instances are likely to have the same class label as those that arrived before them. This is given in Equation 1 (Zliobaite et al., 2015).

$$P(y_t, y_{t-1}) \neq P(y_t)P(y_{t-1}) \tag{1}$$

If a drift occurs when temporal dependence is present within the stream and the traditional process of classifier resetting and retraining is followed, then the classifier can produce misleading performance results. Since arriving class labels are likely to be identical during periods of temporal dependence, when the classifier is reset and subsequently evaluated it will appear to be performing with high classification accuracy.

This research proposes a novel approach which suggests an adaptation to the existing process for resetting base classifiers during detected drifts when temporal dependence is present. Burst Detection-based Selective Classifier Resetting (BD-SCR) challenges the assumption that classifiers must be reset for every detected drift and instead statistically monitors changes in the levels of temporal dependence over the entire stream to make informed decisions about classifier resetting. BD-SCR proposes a significant contribution to the research domain of concept drift detection. Recent studies (Zliobaite et al., 2015, Bifet, 2017, Wares et al., 2019) have identified the negative impact of temporal dependence upon the evaluation of concept drift detectors caused by the over-resetting of base classifiers. Currently there exists very little published research which proposes new methods for handling or processing temporal dependence data in a concept drift scenario.

This research provides a critical summary of existing metrics and methods for handling temporal dependence in Section 2. A concise statistical explanation of BD-SCR alongside an algorithmic description is given in Section 3. The experimental process, results and discussion are provided in Section 4. Finally, conclusions and suggestions for future research are outlined in Section 5.

## 2 Background

In the context of stream mining, arriving stream elements are innately assumed to be independent of their time of arrival, and that arriving class labels $y_t$ are dependent on their feature vectors $x_t$. This assumption is contradicted when temporal dependence is present in a data stream; arriving elements and their class labels become dependent on their time of arrival. According to Zliobaite et al. (2015), a class label is considered to be temporally dependent if

$$P(y_t, y_{t-1}) < P(y_t)P(y_{t-1}) \tag{2}$$

where $y_t$ is a class label and $t$ is the time of arrival. The authors also highlight that this is only first order temporal dependence; only the immediately previous label is used for observation. Temporal dependence of the $n$th order observes the previous $n$ labels. Temporal dependence itself is not a recent issue but has been the subject of research in other domains, such as time series analysis (Cheng et al., 2014, Feng et al., 2018). In the context of stream mining, however, the impact of temporal is relatively unexplored and is an emerging problem.

While concept drift detection techniques have become well established over the years, recent literature has identified that temporal dependence can cause misleading performance. The problem stems from resetting base classifiers when drifts are detected and temporal dependence is also present. Bifet (2017) effectively highlights this issue by comparing the performance of a "No-Change" drift detector to that of several state-of-the-art drift detectors. The "No-Change" detector does not implement any unique statistical concept drift detection, but instead emits a detected change every 60 instances. The justification for this is to demonstrate the how severely misleading the classification performance evaluation can be if temporal dependence is present in the data.

**Table 1.** State-of-the-art drift detectors with Naive-bayes

| Drift Detector | Elec2 | Forest Covertype |
|---|---|---|
| CUSUM | 79.21 | 81.55 |
| Page-Hinckley | 78.04 | 80.06 |
| DDM | 81.18 | 88.03 |
| EDDM | 84.83 | 86.08 |
| **No-Change** | **86.16** | **88.79** |

**Table 2.** State-of-the-art drift detectors with Hoeffding Tree

| Drift Detector | Elec2 | Forest Covertype |
|---|---|---|
| CUSUM | 81.71 | 83.01 |
| Page-Hinckley | 81.95 | 81.65 |
| DDM | 85.41 | 87.35 |
| EDDM | 84.91 | 86.00 |
| **No-Change** | **85.54** | **88.04** |

Table 1 (Bifet, 2017) shows the results of a Naive-bayes classifier using state-of-the-art drift detectors on two popular concept drift datasets; Electricity (Harries & Wales, 1999) and Forest Covertype (Blackard et al., 1998). As shown

above, the “No-Change” detector clearly appears to achieve higher performance accuracy than that of the state-of-the-art drift detectors, even though the “No-Change” detector performs no unique statistical drift detection. To ensure this anomaly is not caused by the base classifier itself, Bifet (2017) also shows the same results are found when using Hoeffding Tree in place of Naive-bayes (2). The author states the reason for this anomaly is due to the existence of temporal dependence within the datasets.

### 2.1 Advancements in Concept Drift Detection

While concept drift detection has several established state-of-the-art methods such as Page Hinckley (PH) (Page, 1954), Cumulative Sum (CUSUM) (Page, 1954), ADWIN (Bifet & Gavalda, 2007), Drift Detection Method (DDM) (Gama et al., 2004) and Early Drift Detection Method (EDDM) (Baena-Garcıa et al., 2006), there have been additional new proposed methods published in recent literature, such as ECHO (Haque, Khan, Baron, Thuraisingham, & Aggarwal, 2016), SAND (Haque, Khan, & Baron, 2016), RDDM (Barros et al., 2017), FPDD (de Lima Cabral & de Barros, 2018), FSDD (de Lima Cabral & de Barros, 2018) and FTDD (de Lima Cabral & de Barros, 2018). There have also been recent advancements in unsupervised detection, such as NN-DVI (A. Liu et al., 2018).

An interesting recent proposition which has inspired the architecture for the original work contained in this research is Hierarchical Hypothesis Testing (HHT) (Yu et al., 2019) which offers a unique framework for drift detection. HHT proposes a two layer architecture for drift detection where the first layer is responsible for detecting a drift whilst the second layer performs validation. This framework is shown in Figure x. The Hierarchical Linear Four Rates method (HLFR) (Yu et al., 2019) is a modern drift detection method developed under the HHT framework. The experimental results for HLFR show that this method outperforms both DDM and EDDM in terms of drift detection rate.

The DetectA method proposed by Escovedo et al. (2018) is another recent development in concept drift detection. Where historically drift detectors have always been reactive, that is a drift is detected and then some decision is made, DetectA instead proposed a reactive drift detection method. This is achieved by predicting class labels through distance based clustering of feature vectors in a process called Pattern Mean Shift. Experimental results for DetectA indicate the Pattern Mean Shift proactive approach to drift detection is capable of detecting drifts, however, when compared to traditional reactive drift detection methods the results do not yield significantly higher performance accuracy.

### 2.2 Temporal Dependence

A recent review by Wares et al. (2019) critically discusses and evaluates the issues surrounding temporal dependence and concept drift, echoing the ever emerging need for existing methods of drift detection to be able to cope with temporal

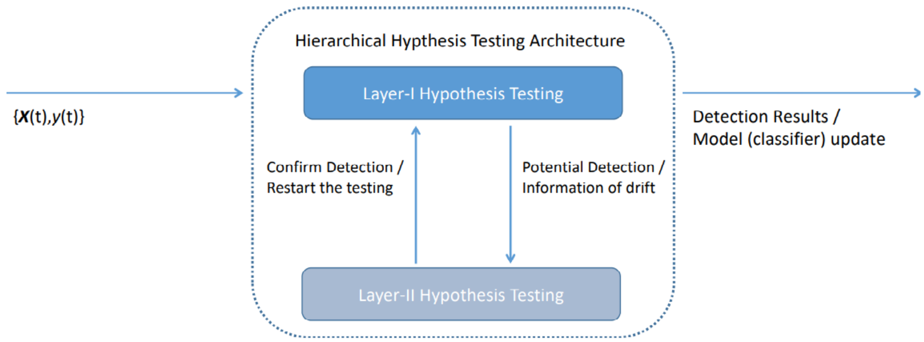


**Fig. 1.** HLFR Framework (Yu et al., 2019)

dependence. At the time of writing there exists little published work which addresses coping with temporal dependence in a data streaming scenario. A recent study by Duong et al. (2018) harnesses temporal dependence in streaming data to aid in change detection by way of a Candidate Change Point model. While this method makes use of the existence of temporal dependence in data streams, it does not aid in solving the issues presented by resetting base classifiers while temporal dependence is present. While existing methods for handling temporal dependence in such contexts are few, some published work does exist which provides important metrics and algorithmic solutions for handling temporal dependence.

One such metric is the Kappa Temporal Statistic (Zliobaite et al., 2015), which is given as

$$k_{per} = \frac{P - P_{per}}{1 - P_{per}}, \tag{3}$$

where $k_{per}$ is the Kappa Temporal value, $P$ is the accuracy of a base classifier and $P_{per}$ is the probability of the Persistent classifier which predicts that the next arriving class label will be identical to the immediately previous label. Classifiers achieving perfect classification performance will yield a result of 1, whilst those performing on equal terms with the Persistent classifier will result in 0. There is also potential for the Kappa Temporal value to fall below 0 in instances where the base classifier is performing worse than the Persistent classifier.

The Kappa Temporal statistic provides a valuable metric for indicating the severity of temporal dependence by how well a classifier performs in contrast to the Persistent classifier. However in the case of imbalanced datasets this metric can be ineffective. Instead the authors recommend the use of the Combined Measure which incorporates both the Kappa Temporal and Cohen's Kappa statistic. The Combined Measure (Zliobaite et al., 2015) is given as

$$K^{+} = \sqrt{max(0, k)max(o, k_{per})} \tag{4}$$

Two methods for handling temporal dependence are also proposed by Zliobaite et al. (2015); the Temporal Correction classifier and Temporally Augmented classifier. The Temporal Correction method involves estimating the posterior probabilities and is given as

$$\frac{P(y_t = i \mid y_{t-1})}{P(y_t = i)} P(y_t = i \mid X_t), \tag{5}$$

where $y_t$ is a class label at time $t$, $X_t$ is the corresponding feature vector and $i$ is some class from the set $i \in \{1, ..., k\}$. This provides a single score which indicates a classifier's ability to handle not only temporal dependence but imbalanced data. The key drawback to this Temporal Correction method is the assumption surrounding previous labels being known. If there is any delay in the arrival or retention of a previous label then the performance and effectiveness of this approach is negatively impacted.

The second, and more elegant, suggestion by Zliobaite et al. (2015) is the Temporally Augmented classifier. This method involves augmenting the features vector of an arriving instance with the last $x$ previously observed class labels. A classification model is trained using the augmented feature vectors, with the prediction $\hat{y}_t$ given as

$$\hat{y}_t = h_t(X_t, y_{t-1}, ..., y_{t-l}) \tag{6}$$

where $h_t$ is the classification model and $l$ is the length of the temporal dependence order. While this is a more elegant solution, it still suffers from the reliance of class labels arriving on time. The authors also note that the Persistent classifier already performs a similar job in predicting that the next arriving label will be the same as the last, and sometimes outperforms the Temporally Augmented classifier approach.

## 3 BD-SCR

This research suggests a novel approach for handling concept drift in streams which also suffer from temporal dependence. This proposed method, Burst Detection-based Selective Classifier Resetting (BD-SCR), challenges the assumption that base classifiers should be reset for each detected drift. Note that BD-SCR is not a drift detector itself but instead monitors the levels of temporal dependence within a stream to inform the decision of resetting the base classifier.

This section examines the construction of BD-SCR through definition and description of both the burst detection method and the process for selective classifier resetting. A full algorithmic overview of the complete method is also provided.

### 3.1 Burst Detection

Zhu & Shasha (2003) describes burst detection as the process of detecting abnormalities or outliers within data streams. In essence, burst detection involves the monitoring of events over specific time period and signals an alarm when an anomaly is detected.

In this research, BD-SCR utilises burst detection in order to determine when there has been a significant increase in the amount of temporal dependence within a data stream. BD-SCR uses a short-term average ($STA$) and long-term average ($LTA$) analysis, which is used to determine the burst value $B$. This is given in equation 7. This method of burst detection has been used extensively in stream based event detection research, particularly in the field of earthquake detection (Earle et al., 2012) (Ross & Ben-Zion, 2014) (Kong et al., 2015). This is an effective yet lightweight method for identifying sudden changes in normality for any given data stream.

$$B = STA/LTA \tag{7}$$

In BD-SCR, $STA$ is the average of a sliding window containing the Kappa Temporal values for all predictions since the last detected drift. This window is reset each time a drift is detected. Similarly, $LTA$ is the average of a window which stores the Kappa Temporal values for predictions made over the entire stream. In typical scenarios, values of $B$ exceeding 1.0 indicate a burst. $B$ is recalculated each time a drift is detected by the underlying drift detector. However, if $STA > 0$ at the point of detection then the classifier is allowed to reset as per traditional drift detection since a positive $STA$ value indicates that the temporal dependence presence level in the current window is low and that the base classifier is performing well.

### 3.2 Selective Resetting

The decision of when to reset the base classifier is of paramount importance. Resetting too frequently results in the same core issue presented by the results of Bifet (2017) and the “No-Change” detector in Section 1. However resetting too infrequently will risk the base classifier becoming outdated as more drifts are detected.

To overcome this problem BD-SCR does not reset the base classifier in every instance where $B > 1.0$. Instead $B$ is compared against a user specified parameter $T$ which defines a threshold indicating the maximum amount of increase, or maximum burst, of temporal dependence that is permitted between detected drifts. For any detected drift, the following conditions are evaluated to determine whether or not the base classifier is reset:

$$\begin{aligned} &\text{If } B < T \text{ reset base classifier} \\ &\text{If } B \geq T \text{ do not reset base classifier,} \end{aligned} \tag{8}$$

Algorithm 1 below provides an overview of the entire BD-SCR method. Note that the algorithm covers the burst detection and selective resetting process, base classifiers are trained using arriving instances as normal.

**Algorithm 1** BD-SCR Algorithm

$wSTA$: Sliding window of Kappa Temporal values since last detected drift
$wLTA$: Kappa Temporal values for the whole stream
$STA$: The average of $wSTA$
$LTA$: The average of $wLTA$
$B$: Burst value
$T$: Burst threshold
$KT_i$: Kappa Temporal value for arriving instance $i$
$C$: The base classifier

Add $KT_i$ to $wSTA$ and $wLTA$
**if** drift detected **then**
    **if** $STA > 0$ **then**
        Reset $C$
    **else**
        calculate $STA$ from $wSTA$
        calculate $LTA$ from $wLTA$
        calculate $B$
        **if** $B < T$ **then**
            reset $C$
        **end if**
        reset $wSTA$
    **end if**
**end if**

## 4 Experiment

BD-SCR was implemented using MOA (Bifet et al., 2010) with experimentation undertaken in the same environment. In order to evaluate BD-SCR and determine its effectiveness in overcoming the temporal issue presented by Bifet (2017), this experiment follows a similar in terms of drift detectors, classifiers and datasets. BD-SCR is tested using a Naive-bayes classifier in conjunction with the Drift Detection Method (DDM), Early Drift Detection Method (EDDM), Page Hinckley (PH) and Cumulative Sum (CUSUM) drift detectors. Since it is already proven that the anomaly is not caused by the base classifier, this experiment tests only with a Naive-bayes classifier. Datasets used are Electricity (Harries & Wales, 1999) and Forest Covertype (Blackard et al., 1998). $T$ values for experimentation range from 1 to 3 with increments of 0.1. Table 3 and Figure 2 contain the results for the Electricity dataset whilst Table 4 and Figure 3 portray the results for the Forest Covertype dataset.

In this experimentation classifier accuracy is the key performance metric used for evaluation. Whilst other metrics have become popular for the evaluation of statistical drift detectors, for example False Alarm Rate (FAR), Mean Time between False Alarms (MTFA) and Mean Time to Detection (MTD), these are not used used in this literature. The aforementioned metrics are used for evaluating the statistical performance of an individual drift detector, whereas BD-SCR is a novel framework for handling and coping with temporal dependence during concept drift detection. BD-SCR is not a statistical drift detector nor does it alter or amend the original statistical methods of any existing drift detectors.

**Table 3.** Numerical Results for Electricity Dataset

| $T$ | DDM | EDDM | CUSUM | PH | NO-CHANGE |
|---|---|---|---|---|---|
| 1 | 80.88 | 80.93 | 77.51 | 75.36 | 76.12 |
| 1.1 | 76.54 | 80.61 | 77.86 | 76.47 | 76.11 |
| 1.2 | 76.56 | 81.27 | 77.98 | 76.47 | 76.17 |
| 1.3 | 81 | 83.7 | 78.01 | 76.47 | 76.23 |
| 1.4 | 81.01 | 84.74 | 77.7 | 76.75 | 76.15 |
| 1.5 | 81.28 | 84.74 | 77.7 | 76.58 | 76.17 |
| 1.6 | 81.48 | 84.74 | 77.89 | 76.5 | 76.14 |
| 1.7 | 82.09 | 84.74 | 77.89 | 77.36 | 76.29 |
| 1.8 | 82.13 | 84.76 | 78.77 | 77.39 | 75.91 |
| 1.9 | 82.9 | 84.7 | 79.08 | 77.39 | 76.07 |
| 2 | 83.88 | 84.76 | 79.3 | 77.39 | 76.1 |
| 2.1 | 83.88 | 84.76 | 79.3 | 77.39 | 76.08 |
| 2.2 | 83.88 | 84.76 | 79.3 | 77.39 | 76.13 |
| 2.3 | 83.88 | 84.76 | 79.3 | 77.39 | 76.46 |
| 2.4 | 83.88 | 84.76 | 79.3 | 77.39 | 76.21 |
| 2.5 | 83.88 | 84.76 | 79.3 | 77.39 | 76.22 |
| 2.6 | 83.88 | 84.76 | 79.3 | 77.39 | 76.32 |
| 2.7 | 83.88 | 84.76 | 79.3 | 77.39 | 76.37 |
| 2.8 | 83.88 | 84.76 | 79.3 | 77.39 | 76.39 |
| 2.9 | 83.88 | 84.76 | 79.3 | 77.39 | 76.71 |
| 3 | 83.88 | 84.76 | 79.3 | 77.39 | 76.84 |

Two key evaluations can be drawn from the experimentation results using BD-SCR; the classification accuracy of the “No-Change” detector and how its effectiveness has been reduced, and the how the sensitivity of $T$ has an impact on performance in conjunction with statistical drift detectors.

Firstly the classification accuracy of the “No-Change” detector can be compared. As observed in Tables 1 and 2, the “No-Change” detector has been previously found to outperform state-of-the-art statistical classifiers due to the effect of over-resetting with temporal dependence. BD-SCR, however, severely impacts performance of “No-Change” with as much as a 10% reduction in classification accuracy with the Electricity dataset, and as much as 22% with Forest Covertype. This degradation in performance is positive for BD-SCR; since the “No-Change” detector performs no statistical evaluation of its own, it shouldn’t outperform other statistical detectors. BD-SCR effectively reduces the performance of the “No-Change” detector and by association provides a novel method for handling concept drift in the presence of temporal dependence.

Since BD-SCR restricts the resetting of the base classifier in certain situations, it follows logically that there may be an inherent risk of reduced classifier

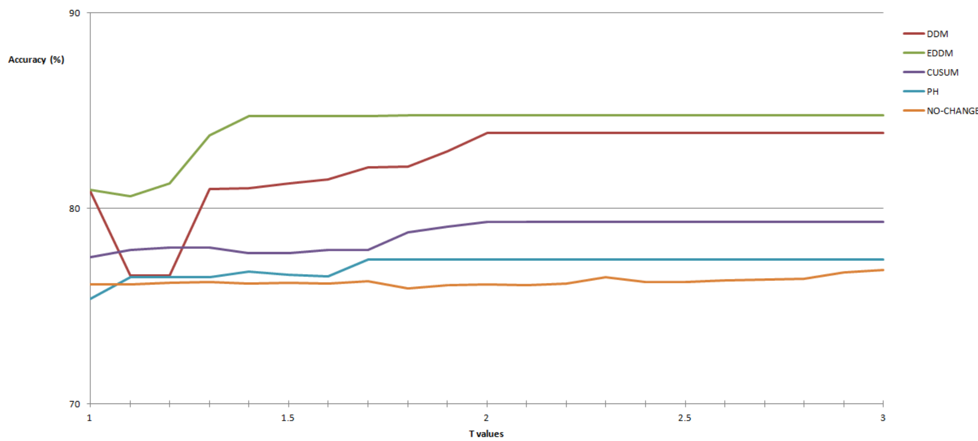


**Fig. 2.** Graphical Results for Electricity Dataset

accuracy. Base classifiers are traditionally reset upon detected drifts to prevent classifiers becoming outdated and irrelevant as stream distributions shift. Without resetting in these circumstances, classifier performance can deteriorate and under perform. This trade off in performance is directly related to the sensitivity of $T$ and underpins the importance of the applying the correct parameter value for $T$. As $T$ increases so does the number of permitted classifier resets, and as such it is expected that classifier performance will improve with increased values of $T$. It may also be the case that classifier performance plateaus at certain values of $T$. This will occur when the value of $T$ allows $B < T$ to always be true and therefore the the base classifier will be always be reset for a detected drift.

For the Electricity dataset results in Table 3 it can be observed that for the smaller values of $T$ there are corresponding small deteriorations in performance when compared to those reported by Bifet (2017) in Table 1. For CUSUM there is a small accuracy decrease of no more than 2% until it plateaus at $T = 2$ with an overall performance increase of 0.1%. Page Hinckley has a similar decrease in performance for values of $T < 1.7$, from which it plateaus with a slight decrease of 0.65%. DDM suffers from varying decreases in performance for all values of $T$ below 2.0, with the most severe decrease being 4.64% at $T = 1.1$. However, DDM does plateau at $T = 2$ with an improvement of 2.7%. EDDM follows a similar pattern to that of DDM with fairly considerable decreases in performance for early values of $T$, but plateaus much earlier at $T = 1.4$ with an overall performance decrease of 0.07%. EDDM does have a very small gain in performance of 0.02% between $T = 1.4$ and $T = 1.8$, but the improvement is so menial that this evaluation considers the plateau to occur at $T = 1.4$.

The Forest CoverType dataset, however, illustrates a more complex situation in regards to the sensitivity of $T$. As can be observed from Table 4, there is a much less obvious plateau of performance for increasing values of $T$, but a

**Table 4.** Results for Forest CoverType Dataset

| $T$ | DDM | EDDM | CUSUM | PH | NO-CHANGE |
|---|---|---|---|---|---|
| 1 | 65.17 | 65.21 | 64.77 | 64.07 | 65.2 |
| 1.1 | 65.21 | 65.21 | 64.77 | 64.07 | 65.31 |
| 1.2 | 65.22 | 65.22 | 64.77 | 64.07 | 65.32 |
| 1.3 | 65.22 | 65.24 | 64.77 | 64.07 | 65.34 |
| 1.4 | 65.2 | 65.26 | 64.77 | 64.07 | 65.39 |
| 1.5 | 65.2 | 65.29 | 64.77 | 64.07 | 65.43 |
| 1.6 | 65.2 | 65.3 | 64.77 | 64.07 | 65.51 |
| 1.7 | 65.22 | 65.36 | 64.77 | 64.07 | 65.6 |
| 1.8 | 65.24 | 65.44 | 64.83 | 68.08 | 65.63 |
| 1.9 | 68.89 | 67.31 | 69.58 | 71.15 | 66.06 |
| 2 | 71.67 | 71.23 | 70.98 | 77.09 | 71.69 |
| 2.1 | 75.58 | 72.9 | 73.2 | 77.09 | 73.44 |
| 2.2 | 78.54 | 75.51 | 77.26 | 77.17 | 76.36 |
| 2.3 | 81.02 | 79.54 | 79.33 | 77.35 | 80.44 |
| 2.4 | 81.46 | 80.58 | 79.35 | 77.35 | 81.88 |
| 2.5 | 81.01 | 80.74 | 79.35 | 77.35 | 81.99 |
| 2.6 | 82.05 | 82.51 | 80.76 | 79.36 | 83.55 |
| 2.7 | 83.68 | 82.74 | 81.15 | 79.43 | 84.16 |
| 2.8 | 83.67 | 82.85 | 81.15 | 79.43 | 84.24 |
| 2.9 | 84 | 82.87 | 81.17 | 79.43 | 84.33 |
| 3 | 84.23 | 84.47 | 81.17 | 79.43 | 84.42 |

clear plateau for lower values. This is caused by a much higher rate of temporal dependence in this dataset in comparison to that of Electricity. In this case, $T$ values below 1.9 provide only a clear degradation in performance. This is caused by much higher values of $B$ during detected drifts such that for lower values of $T$, $B < T$ is almost never true. It is not until much higher values of $T$, typically 2.6 or 2.7 in this case, that performance can be seen to begin to achieve close to the benchmark results in Table 1. It is also worth noting that as $T$ increases, so does the performance of "No-Change" since the base classifier is not being prevented from resetting. As such, it is important to identify the correct value $T$ which allows for high statistical performance but restricts the performance of "No-Change"". Specifically, values of $T$ higher than 3.0 could be tested to determine where results begin to plateau.

Table 5 compares the Kappa Temporal values for both datasets using a Naive-bayes classifier in conjunction with each of the previously mentioned state-of-the-art drift detectors. This clearly identifies the stark contrast in the levels of temporal dependence in both datasets.

The differences in severity of temporal drift between the two datasets affects the sensitivity of the $T$ parameter in BD-SCR. As previously discussed, the Electricity dataset results in Table 3 indicate clear preferred values for $T$ where the "No-Change" detector performance is reduced but the drift detector performance achieves classification accuracy in line with previously published results.

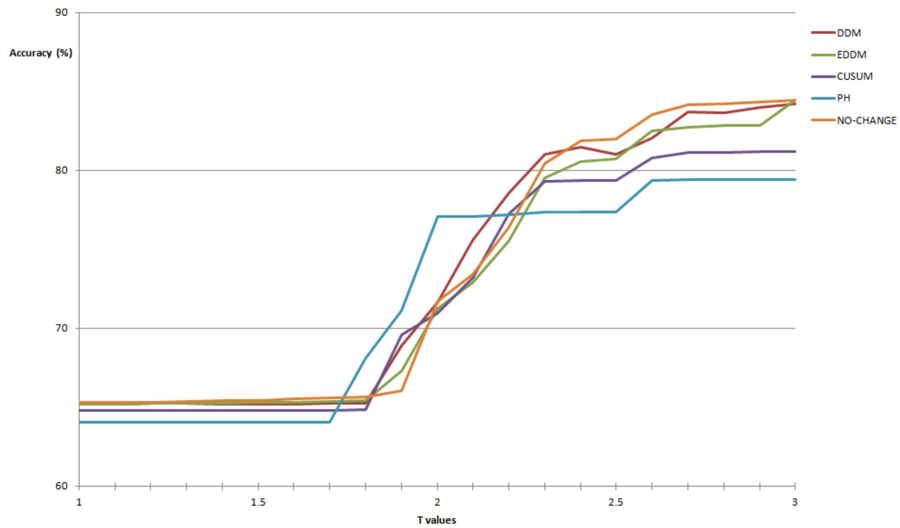


**Fig. 3.** Graphical Results for Forest CoverType Dataset

**Table 5.** Comparison of Kappa Temporal Values

| $T$ | DDM | EDDM | CUSUM | PH |
|---|---|---|---|---|
| Electricity | -0.2831 | -0.0342 | -0.4172 | -0.4968 |
| Forest Covertype | -1.4245 | -1.8181 | -2.7364 | -3.0375 |

In contrast, optimal $T$ values are obscured for the Forest Covertype dataset as highlighted in Table 4. By comparing this observation with the Kappa Temporal values given in Table 5, the affect of the severity of temporal dependence upon the sensitivty of $T$ becomes obvious. The Kappa Temporal values for the Electricity dataset are negative, indicating that the base classifier is performing worse than the persistent classifier,. However in comparison to the Kappa Temporal values of the Forest Covertype dataset it is trivial to notice the difference in severity of temporal dependence between the two datasets.

Results across both datasets show that BD-SCR does provide a novel approach for coping with temporal dependence when using concept drift detectors. Results show a clear deterioration in classifier accuracy for the “No-Change” detector when using BD-SCR, which clearly portrays that the selective resetting approach does aid in the problem of over-resetting. Any decrease in performance using statistical drift detectors is minimal, and in some cases performance is actually improved. However performance is directly related to the user defined parameter $T$, and calculating its optimal value to maximise statistical detector performance but minimising performance of the "No-Change" detector may be challenging.

## 5 Conclusion

This research presents BD-SCR as a novel approach to classifier resetting during concept drift detection in the presence of temporal dependence. Using a short term and long term average statistical event detection method, BD-SCR analyses the severity of temporal dependence to make an informed decision to reset the base classifier. Experimental results show that BD-SCR is effective at overcoming the problem of over-resetting during periods of temporal dependence by selectively resetting the classifier, and that there is minimal performance impact when choosing not to reset.

However, the sensitivity of the threshold parameter $T$ means finding the optimal balance between statistical detectors and "No-Change" is difficult and unclear. In simulated scenarios where static datasets are streamed, such as the Electricity and Forest Covertype datasets used in this research, it is fairly trivial to analyse the results and determine the optimal values for $T$. In a real-time situation, however, this is much more complex. Furthermore, the severity of the temporal dependence in the data has a direct affect on the sensitivity of the $T$ parameter. As shown through the results of the Electricity dataset in comparison to that of the Forest Covertype dataset, it is easier to identify optimal $T$ values in datasets with less temporal severity.

BD-SCR challenges the current architectural approach to drift detection to accommodate for other stream related anomalies. Future work in this field should look to improve BD-SCR by expanding the algorithm to automatically determine the optimal value for $T$. Additionally the proposed structure of BD-SCR could be used to develop more sophisticated ways of handling temporal dependence in temporally dependent concept drift streaming scenarios.

# References


Baena-Garcıa, M., del Campo-Ávila, J., Fidalgo, R., Bifet, A., Gavalda, R., & Morales-Bueno, R. (2006). Early drift detection method. In *Fourth international workshop on knowledge discovery from data streams* (Vol. 6, pp. 77–86).

Barros, R. S., Cabral, D. R., Gonçalves Jr, P. M., & Santos, S. G. (2017). Rddm: Reactive drift detection method. *Expert Systems with Applications*, *90*, 344–355.

Bifet, A. (2017). Classifier concept drift detection and the illusion of progress. In *International conference on artificial intelligence and soft computing* (pp. 715–725).

Bifet, A., & Gavalda, R. (2007). Learning from time-changing data with adaptive windowing. In *Proceedings of the 2007 siam international conference on data mining* (pp. 443–448).

Bifet, A., Holmes, G., Kirkby, R., & Pfahringer, B. (2010). MOA: massive online analysis. *Journal of Machine Learning Research*, *11*, 1601–1604. Retrieved from http://portal.acm.org/citation.cfm?id=1859903

Blackard, J. A., Dean, D. J., & Anderson, C. (1998). *The forest covertype dataset.*

Cheng, D., Bahadori, M. T., & Liu, Y. (2014). Fblg: a simple and effective approach for temporal dependence discovery from time series data. In *Proceedings of the 20th acm sigkdd international conference on knowledge discovery and data mining* (pp. 382–391).

de Lima Cabral, D. R., & de Barros, R. S. M. (2018). Concept drift detection based on fisher's exact test. *Information Sciences*, *442*, 220–234.

Domingos, P., & Hulten, G. (2000). Mining high-speed data streams. In *Proceedings of the sixth acm sigkdd international conference on knowledge discovery and data mining* (pp. 71–80).

Duong, Q.-H., Ramampiaro, H., & Nørvåg, K. (2018). Applying temporal dependence to detect changes in streaming data. *Applied Intelligence*, *48*(12), 4805–4823.

Earle, P. S., Bowden, D. C., & Guy, M. (2012). Twitter earthquake detection: earthquake monitoring in a social world. *Annals of Geophysics*, *54*(6).

Escovedo, T., Koshiyama, A., da Cruz, A. A., & Vellasco, M. (2018). Detecta: abrupt concept drift detection in non-stationary environments. *Applied Soft Computing*, *62*, 119–133.

Feng, S., Wang, X., Sun, H., Zhang, Y., & Li, L. (2018). A better understanding of long-range temporal dependence of traffic flow time series. *Physica A: Statistical Mechanics and its Applications*, *492*, 639–650.

Gama, J., Medas, P., Castillo, G., & Rodrigues, P. (2004). Learning with drift detection. In *Brazilian symposium on artificial intelligence* (pp. 286–295).

Haque, A., Khan, L., & Baron, M. (2016). Sand: Semi-supervised adaptive novel class detection and classification over data stream. In *Aaai* (pp. 1652–1658).

Haque, A., Khan, L., Baron, M., Thuraisingham, B., & Aggarwal, C. (2016). Efficient handling of concept drift and concept evolution over stream data. In *2016 ieee 32nd international conference on data engineering (icde)* (pp. 481–492).

Harries, M., & Wales, N. S. (1999). *Splice-2 comparative evaluation: Electricity pricing.*

Kolter, J. Z., & Maloof, M. A. (2003). Dynamic weighted majority: A new ensemble method for tracking concept drift. In *Data mining, 2003. icdm 2003. third ieee international conference on* (pp. 123–130).

Kong, Q., Kwony, Y.-W., Schreierz, L., Allen, S., Allen, R., & Strauss, J. (2015). Smartphone-based networks for earthquake detection. In *2015 15th international conference on innovations for community services (i4cs)* (pp. 1–8).

Liu, A., Lu, J., Liu, F., & Zhang, G. (2018). Accumulating regional density dissimilarity for concept drift detection in data streams. *Pattern Recognition*, *76*, 256–272.

Liu, G., Cheng, H.-r., Qin, Z.-g., Liu, Q., & Liu, C.-x. (2013). E-cvfdt: An improving cvfdt method for concept drift data stream. In *Communications, circuits and systems (icccas), 2013 international conference on* (Vol. 1, pp. 315–318).

Page, E. S. (1954). Continuous inspection schemes. *Biometrika*, *41*(1/2), 100–115.

Ross, Z. E., & Ben-Zion, Y. (2014). An earthquake detection algorithm with pseudo-probabilities of multiple indicators. *Geophysical Journal International*, *197*(1), 458–463.

Street, W. N., & Kim, Y. (2001). A streaming ensemble algorithm (sea) for large-scale classification. In *Proceedings of the seventh acm sigkdd international conference on knowledge discovery and data mining* (pp. 377–382).

Wang, H., Fan, W., Yu, P. S., & Han, J. (2003). Mining concept-drifting data streams using ensemble classifiers. In *Proceedings of the ninth acm sigkdd international conference on knowledge discovery and data mining* (pp. 226–235).

Wares, S., Isaacs, J., & Elyan, E. (2019). Data stream mining: methods and challenges for handling concept drift. *SN Applied Sciences*, *1*(11), 1412.

Yu, S., & Abraham, Z. (2017). Concept drift detection with hierarchical hypothesis testing. In *Proceedings of the 2017 siam international conference on data mining* (pp. 768–776).

Yu, S., Abraham, Z., Wang, H., Shah, M., Wei, Y., & Príncipe, J. C. (2019). Concept drift detection and adaptation with hierarchical hypothesis testing. *Journal of the Franklin Institute*, *356*(5), 3187–3215.

Zhu, Y., & Shasha, D. (2003). Efficient elastic burst detection in data streams. In *Proceedings of the ninth acm sigkdd international conference on knowledge discovery and data mining* (pp. 336–345).

Zliobaite, I., Bifet, A., Read, J., Pfahringer, B., & Holmes, G. (2015). Evaluation methods and decision theory for classification of streaming data with temporal dependence. *Machine Learning*, *98*(3), 455–482.