Concept Drift in Streaming Data: A Systematic Literature Review

Tariq Mahmood[1] Tatheer Fatima

Abstract

World is generating immeasurable amount of data every minute, that needs to be analyzed for better decision making. In order to fulfil this demand of faster analytics, businesses are adopting efficient stream processing and machine learning techniques. However, data streams are particularly challenging to handle. One of the prominent problems faced while dealing with streaming data is concept drift. Concept drift is described as, an unexpected change in the underlying distribution of the streaming data that can be observed as time passes. In this work, we have conducted a systematic literature review to discover several methods that deal with the problem of concept drift. Most frequently used supervised and unsupervised techniques have been reviewed and we have also surveyed commonly used publicly available artificial and real-world datasets that are used to deal with concept drift issues.

Index Terms: Concept drift, machine learning, streaming data, unlabeled streaming data.

1 Introduction

The world is evolving and getting smarter day by day, new and intelligent technologies like internet of things (IoT), Big Data and cloud computing are playing a major role in this evolution. "IoT, refers to the millions of physical devices like chips, sensors, smart phones, cameras and many more, these devices are capable of collecting and sharing data over the internet anywhere in the world". It is one of the most popular technologies of this digital era. IoT has brought the revolution in how business operates, it is now one of the major sources of the data, it produces, directs and drives data continuously. Unlike traditional batch-based flows, streaming data is more time sensitive and larger in volume, due to this more robust and efficient tools are needed to effectively analyze the ever-growing streaming data in different application domains and fields.

Machine learning has been introduced in an era of data deluge, where data is continuously being generated and stored in large volumes. To deal with this enormous amount of data, industries have moved towards machine learning techniques for data driven solutions. Many companies have adopted classification to effectually perform the task of predictive analytics and have gotten rid of cumbersome traditionally used analytical methods. The capability of generalizing and foretelling using data have made classification a desirable technique to solve many data directed business problems.

However, this generalization ability of a classifier makes a powerful presumption about the stability of the underlying distribution of the arriving data. According to which the data that is being used for training and testing the model should be Identically and Independently Distributed

[1] *Institute of Business Administration, Karachi | tmahmood@iba.edu.pk*

(IID). Most of the machine learning algorithms assume that the relationship between the input (input features) and the output (i.e., target variable) remains static, but in real world data can change with time in unpredictable ways. This can degrade predictive performance of a model. The issue of varying the underlying relationships in the data is called concept drift. It was first proposed by (Jeffrey C. Schlimmer 1986), who aimed to point out that the same instances can be identified as noise data or non-noise information at separate times. The cause of these changes could be the variation in the hidden variables that could not be assessed directly. Areas where concept drift is involved includes recommendation systems, energy consumption, artificial intelligence systems with dynamic environment interaction, Fraud monitoring and anomaly detection systems, and biomedical signal analysis (e.g., neurogenerative diseases). For example, in cases of fraud detection, fraudsters are evolving and getting more creative with time and adopting new and advance techniques, machine learning model trained on the historical data will not be able to detect these changing patterns. The occurrence of concept drift has been perceived as the main reason of performance decaying in many data directed information systems that are being used for decision making and early warnings. In an evolving environment of big data, maintaining reliable and efficient data directed predictions and judgement facilities has become a critical problem. There is a need for detection of these changing concepts.

Researches to deal with concept drift have been increased a lot over the last decade, and many drift handling and adaptive learning techniques have been proposed. Adaptive learning means upgradation of the predictive model in an incremental manner to avoid encountering concept drift. Big data produces massive amount of data daily in a streamed fashion. But dealing with streaming data can be challenging. Machine learning can be used to perform tasks like querying, pattern recognition, real time analysis and predictive analytics. Industries have realized the potential of machine learning and are incorporating machine learning models to support business decisions. Many companies are using it to uncover hidden patterns, identify customer preferences, reveal market trends and analyze bigger and more complex data. Machine Learning models have made decision making easy and more robust. However, these learning models work in a dynamic setup, but within this era of technology things are changing at a fast pace and with this data is also changing, this can decompose the foretelling capabilities of a classifier with time and thus making it outdated.

Various researches related to drift-aware are available, however the focus is on supervised techniques. This survey aims to find out techniques for concept drift identification in not only labeled but also in unlabeled streams.

2 Background

Conventional machine learning has two main components: training/learning and prediction. Many of the algorithms assume that the data distribution stays the same over time. In other words, the examples that we see in the training set should resemble the observations we see in operation. However, streams are rarely stationary. The fast-changing surroundings of new products, new markets and new customer conduct can result in a population change or spontaneous change in the data which then changes the distribution of the data and can decrease model performance, this change in distribution is termed as Concept Drift. For an instance, in ecommerce the customer's shopping behavior may change over time. For e.g. If we are predicting monthly sales of each product or product category, looking at the historical data such as promotion budget, the model may give good accuracy currently but it can easily decay with time as the customer psychology changes, it can be due to external factors such as trends, celebrity influence, seasons etc.

Machine learning models should identify concept drift and adjust to it so that a satisfactory model performance is retained. Learning with the concept drift has been categorized in 3 parts: concept drift detection (identify if drift has occurred), drift understanding (when, how, where it occurs) and drift adaptation (how to react to the occurring drift), this survey is focuses on the techniques to detect concept drift.

A Concept drift definition

In the domain of predictive analytics, an unpredictable variation in the statistical properties of the target label, that the algorithm is trying to forecast is termed as concept drift. The target variable that is being predicted is referred as the term concept. The input variables can also be termed as concepts, mostly this term is used for the target variable. Formally, concept drift is defined as follows:

At a certain time duration [0,t], a group of examples is defined as $S_{0,t} = \{d_0,\dots,d_t\}$, where $d_i = (X_i, y_i)$ is a singular instance (or a data point), here X_i is denoting the input vector, y_i is the target variable, and $S_{0,t}$ follows a particular distribution $F_{0,t}(X,y)$. Concept drift occurs at timestamp t + 1, if there is a change in the distribution $F_{0,t}(X,y)$ that has $F_{0,t}(X,y) \neq F_{t+1,\infty}(X,y)$, denoted as $\exists t$: $P_t(X,y) \neq P_{t+1}(X,y)$.

Thus, the concept drift between time t and $t+1$ is $\exists t$: $P_t(X,y) \neq P_{t+1}(X,y)$ where P_t is the joint distribution at time t between the input vector X and the target label y. Variation in the relationship components of input and target variables can define the variation in data.

the change in the classes' $P(y)$ prior probability,

the change in the class $P(X|y)$ conditional probabilities,

Thus, the prediction will be affected due to the change in the classes $P(y|X)$ posterior probabilities.

This gives us the concept of real and virtual concept drift. Variation in $P(y|X)$ denotes the occurrence of **Real concept.** This change can either be affected with the change in $P(X)$ or not. **Virtual drift** happens when there is change in the distribution of the input vector $P(X)$ without affecting $P(y|X)$.

B Types of concept drift

Commonly drift is categorized in four types that are shown below:

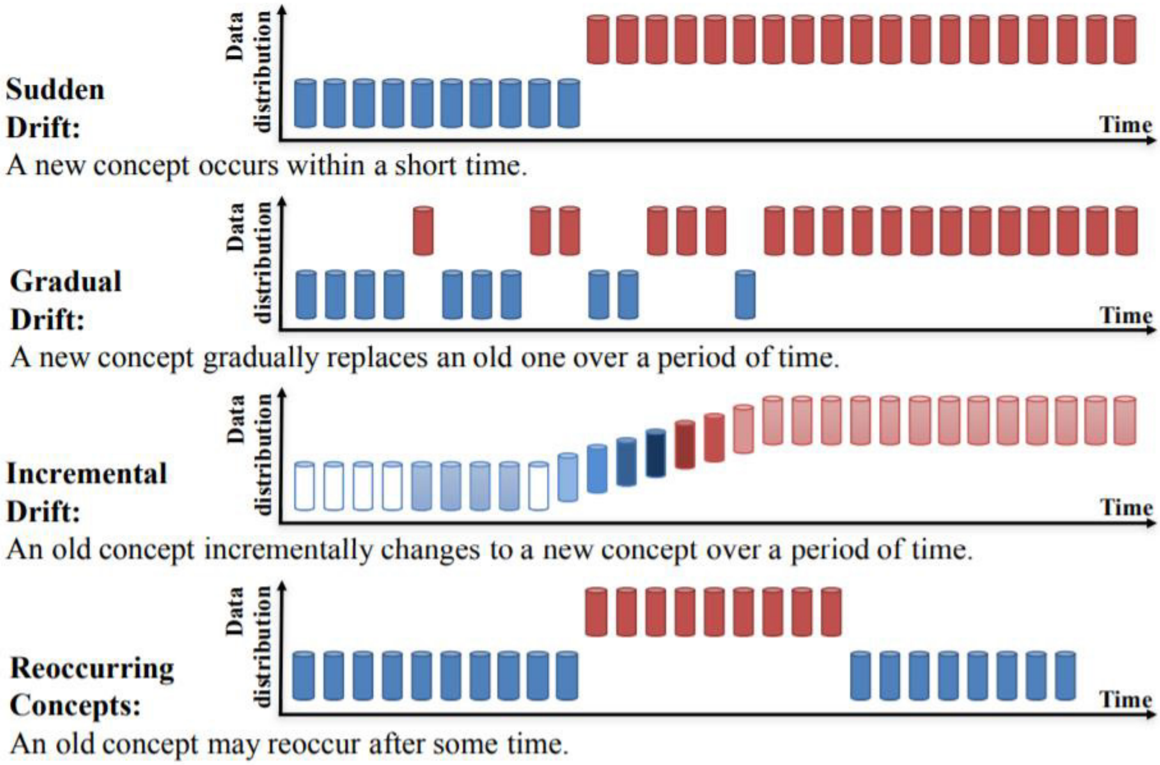


Figure 1: Categories of concept drift
Source: (Jie Lu 2018)

A sudden drift takes place when the distribution of the data is abruptly changed in a brief span of time. For example, a sudden drift may occur in a warehouse sensor data stream when an unseen abnormality is detected by it. An incremental drift occurs gradually in a longer span of time, an example of incremental drift is when the raiders change their fraudulent behaviors with time. As RFID chips are introduced, it changes patterns of raiders of committing fraud, as more customers get the new RFID card the new types of fraud will become more common until every customer has it, where the drift will stop. A gradual drift occurs when the starting distribution and the ending distribution happens simultaneously for some time, eventually transitioning to the ending distribution. An example would be customers shopping behavior, where sales figure fluctuates and shows this kind of drifts over different seasons. A recurring drift is any drift that repeats itself. An example would be energy consumption data where data fluctuates and reaches peaks during different times in a day. It is low at day and gets high at night. The focus of researches into concept drift adaptation in sudden/gradual/incremental drift is to increase the recovery rate and lower the drop-in accuracy during the concept transformation process. While, historical concepts are used in the case of recurring drift, the focus is on finding the best fitting historical concept, in the shortest span of time. The new concept may reemerge suddenly, incrementally, or gradually. The phenomenon of "intermediate concept" is used to better exhibit the different types of concept drifts. A concept drift doesn't have to occur at an exact timestamp, it can exist for a certain duration. Intermediate concept can be the mixture of initial and ending concepts. Thus, intermediate concepts can be used to represent the transition between initial concept and the final concept.

The above four categories are enough to define single class drift pattern, however class labels are not available in unsupervised drift detection, so there is a need to consider global data distribution change as well. To fulfill this purpose, two more concept drift types, stationary and non-stationary drifts are introduced as shown below:

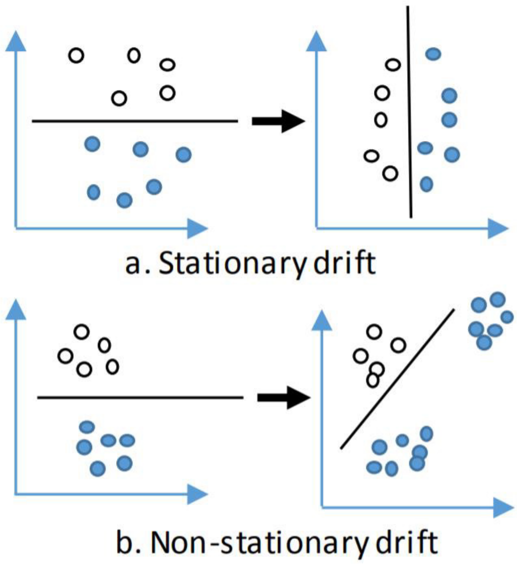


Figure 2: Stationary and Non-stationary drift
Source: (Hanqing Hu 2018)

In stationary drift, concept drift has changed the distributions of individual classes, thus resulted in the change of data model. Nonetheless, both classes still take the similar space globally. While in the scenario of non-stationary drift, global distribution of the data has also changed along with the individual classes. A stationary drift can be sudden, incremental or gradual. There can be an abrupt change in the decision boundary forming a new hyperplane within the data space or it can slowly change over time. Likewise, in a non-stationary drift a new data space can be formed suddenly if large amounts of data appear at a new region in a short time, or it can be formed gradually by shifting the shape of existing data space slowly.

C Concept Drift Detection Algorithms

Numerous performance-based drift identification methods and algorithms have been suggested in literature. These algorithms can detect all kinds of drift but require target labels.

I) Drift detection method (DDM)

This method is proposed by (João Gama 2004). It is based on the detection of probability distributions of samples using errors in the learning process. It models no of errors as a binomial random variable. This approach requires class labels to monitors the increase in error in the model predictions to identify drift. The stated method does not look at the change in distribution to detect drift that's why it can identify all kinds of drift.

The error rate is the probability of misclassifying (p_i), which is calculated for every point *i*. The standard deviation of the error rate is $s_i = \sqrt{p_i(1 - p_i)/i}$. While training the model DDM maintains 2 global variables, p_{min} and s_{min}, whenever a new sample is treated the values of these variables are updated if the condition $p_i + s_i > p_{min} + s_{min}$, is satisfied. This method is based on an assumption that as the count of the samples rises the error rate of the algorithm (p_i) will reduce in a stationary environment, if there is a noticeable rise in the error of the algorithm it means that there is a change in class distribution. Hence the values of p_i and s_i is stored and when $p_i + s_i \geq p_{min} + 2 * s_{min}$ a warning level is triggered and the examples are then stored in anticipation of a possible drift. And when $p_i + s_i \geq p_{min} + 3 * s_{min}$ drift level is triggered, the model is re-trained using the examples that were stored after the warning level was triggered.

Similarly, EDDM (M. Baena-Garcıa 2006) was proposed, it is also based on classification error monitoring, however to give better detection accuracy, EDDM uses the distance among the two errors, which makes it robust in identifying gradual drift. While DDM-OCI was proposed (S. Wang 2013) to address the limitation of DDM when data is imbalanced.

II) Adaptive windowing (ADWIN)

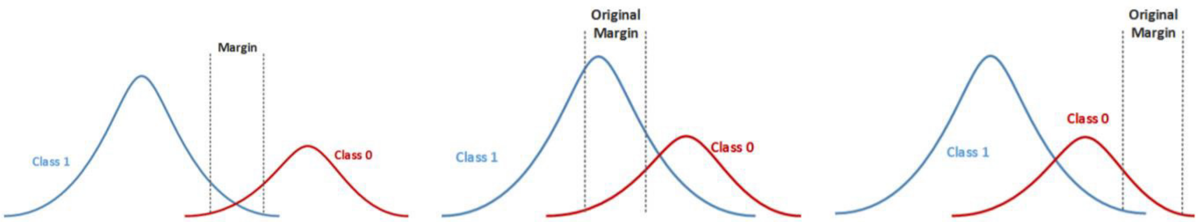
In contrast to the above techniques, which operate in an incremental fashion that takes one instance at a time, window-based approaches use a chunk based or sliding window approach over the recent samples, to detect changes. Windows of prediction errors for each partition it calculates mean error and compares the difference. (Albert Bifet 2007) presented the ADWIN2 algorithm, an improved version of ADWIN algorithm. ADWIN2 has a variable sized window: it grows or shrinks when no change or concept drift is detected, respectively.

III) Linear four rate (LFR)

LFR (Wang and Abraham 2015) is better for class imbalance then DDM and EDDM. The LFR strategy is straightforward that if data is stationary then the contingency table should stay constant. It uses four rates true positive rate (*tpr*), true negative rate (*tnr*), positive predictive value (*ppv*) and negative predictive value (*npv*) to identify the concept drift. During a stationary concept (i.e., $P(X_t, y_t)$ stays constant), $P_{tpr,}$ $P_{tnr,}$ $P_{ppv,}$ P_{npv} doesn't changed. Hence, a noticeable variation in any P_* ($* \in tpr, tnr, ppv, npv$) shows that the variation in underlying joint distribution $P(X_t, y_t)$, or concept has happened. The 4 rates can be computed as follows: $P_{tpr}=TP/(TP+FN)$, $P_{tnr}= TN/(TN+FP)$, $P_{ppv}= TP/(FP+TP)$ and $P_{npv}=TN/(TN+FN)$. All the mentioned characteristic rates in $P_*=\{P_{tpr,} P_{tnr,} P_{ppv,} P_{npv}\}$ are equal to 1, if there is no misclassification.

IV) Margin density drift detection (MD3)

(Kantardzic 2015) proposed MD3. In order to monitor the variation in the distribution of the arriving instances, MD3 traces the changes in classifier boundary. The thought behind this method is that, as the count of instances inside the margin will rise or reduce, a variation in *p(ylX)* will be observed. A gradual drift will occur if the number of instances grows inside the margin, causing the class distribution to move closer to the boundary, as shown in (Figure. 3b). However, if an entire class distribution moves towards a different area of a feature space, a reduction in the count of instances (Figure. 3c) will be seen, this drift is said to be sudden. Further inquiry is needed in both the cases. MD3 uses unlabeled samples and is good for identifying gradual and stationary drift.



(a) Initial distribution of samples as distance from the model boundary (b) Drift causing increase in margin density (c) Sudden drift causing drop in margin density

Figure 3: Drifting concepts and their effect on margin density
Source: (Kantardzic 2015)

A sliding window divides the stream in samples, the size of th e window is S, which moves at a rate of S_r samples. A Decision function is used to calculate each sliding window's margin density (ρ) e.g. for SVM, considering the kernel is linear margin density is defined as,

$$\rho = \frac{\#samples\ with\ abs(w \cdot x + b) \leq 1}{\#samples}$$

A block of unlabeled instances is given to the algorithm at time t. It calculates the margin density measure of the current block and triggers an alert if $\rho_{min} - \rho_{max} > \theta_{\rho}$. This means a drift has occurred. The range of margin density(ρ_{min} ρ_{max}) since the last drift is stored as well as the threshold to signal drift (θ_{ρ}). The current classifier (SVM with linear kernel the classification model is given as *w,b*) shows the concept at $t-1$

V) Clustering based drift detection

(Joung Woo Ryu 2012) presented method. The clusters are formed with the existing data. As new data arrives, model tries to classify the new instances into existing clusters. If the number of the new instances which didn't fit into any of the existing clusters have grown enough that they can create a new cluster on their own, then a drift is said to be detected. Density-based detection is based on an assumption that the samples within the same cluster belongs to the same class. Every cluster C is determined by a radius *radc* and a cluster density *dc*

radc = distance between cluster center and the furthest sample from the center
dc = number of samples in cluster divided by radc

radc is calculated using Euclidean distance. If the distance between the cluster center and the new sample is less than $radc$, then it is assigned to that cluster otherwise it is viewed as an un-allocated sample and is marked as $\sim s$. As time passes and the number of $\sim s$ have increased enough to form a new cluster then it can be said that concept drift has occurred.

VI) Grid based drift detection

This method is proposed by (Mohammad Masud 2011). Every feature/input of the data is divided into fixed intervals to create a grid of data space. On the basis of the feature values new samples are assigned to their specific cell in the grid. If the number of samples in a cell increases above a threshold level p a potential concept drift has occurred. Grid based detection monitors the shape of the data, if the new data arrives in the current dense cells of the grid it means the existing shape of the data has not changed and if the new data arrives in the less dense cells then it can be concluded that the shape of the data is changing and thus the concept drift is identified.

D Concept Drift Adaptation

Concept drift adaptation refers to updating the existing models in order to handle the detected drifts. It is also called drift handling and drift learning. The two main techniques for concept drift adaptation are single learning model adaptation and ensemble learning for concept drift adaptation.

I) Single learning model adaptation

In single learning model adaptation only one learning model is activated at a time, to perform the classification and prediction tasks. This approach further be classified in to two categories, instance selection and incremental learning.

- ***Instance selection and weighting***

The simplest way to handle concept drift is, to re-learn a fresh model with the most recent instances. A window-based approach is often adopted to store the most relevant instances with respect to the current concept and then retrain a fresh model with it to replace the outdated one. Paired Learners (Stephen H. Bach 2008) follows this method and uses two learners: the stable learner and the reactive. If the new instances are being falsely predicted by the stable learner and the same instances are being correctly predicted by the reactive learner, than the stable learner is substituted by the reactive one. Similarly, Self Adjusting Memory kNN (SAMkNN) (Viktor Losing 2016), gives weightage to the learners based on their performance on the most recent time frame.

Incremental Learning

A substitute to training an entire new model, is to build a model that incrementally updates itself and adapts to the recent changes in the data. This approach is more robust than retraining an entire new model when the drift is small or reoccurring. Majority of the algorithms in this group are based on decision trees, because trees are capable of examining each sub-region separately. Algorithms like Very Fast Decision Tree classifier (VFDT) and CVFDT (Geoff Hulten 2001) are proposed for this purpose.

II) Ensemble learning for concept drift adaptation

In contrast to single learning model adaptation, ensemble learning utilizes multiple models to make prediction at each point in time. The reason of using ensemble methods to deal with concept drift is that it can save the great deal of labor that is required to retrain a new model for recurring concepts. Ensemble is the most popular method being used recently. It consist of several base learners that may have distinct parameters. The final prediction is given by combining the output of each ensemble using votes of certain weights. A survey about this research topic is (Bartosz Krawczyk 2017).

3. Research Methodology

The purpose of the survey was to study the algorithms and methodologies proposed and are being used for concept drift detection. Research papers were searched on the most commonly used and popular platforms for Computer Science which include IEEE, Elsevier, Springer, Taylor and Francis, Wiley, ACM, Arxiv.org and Google Scholar using search queries:

- Concept drift
- Concept change
- Concept drift machine learning
- Concept change in machine learning
- Concept drift machine learning streaming data
- Concept change in machine learning streaming data
- Concept drift machine learning unlabeled streaming data
- Concept change machine learning unlabeled streaming data

After searching papers on mentioned platforms, the following filtration technique was applied to extract the most relevant papers.

Title Screening (56 papers)

→ **Abstract Screening (29 papers)**

→ **Content Screening (17 papers)**

A summary regarding all the work done is given below, Figure 4-7 represents the distribution of articles over the years and differentiated by digital sources.

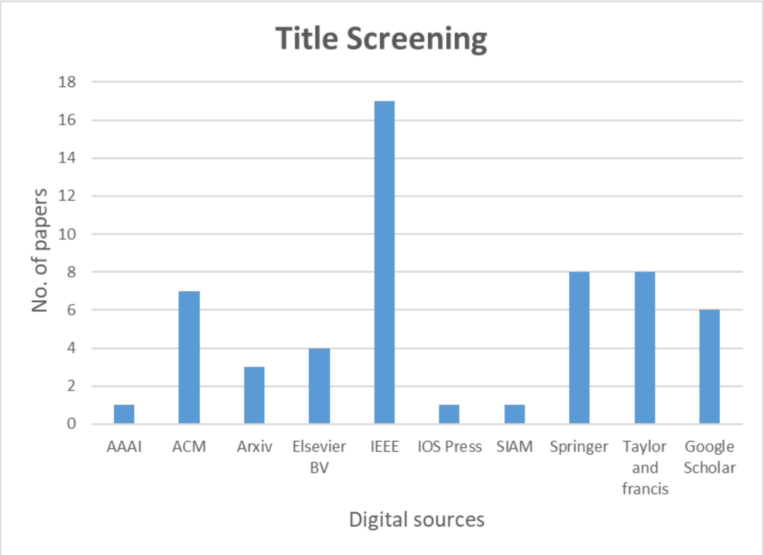


Figure 4: Distribution of 56 papers in the first stage of filtration differentiated by digital sources.

Figure. 5 shows the number of papers published by year. One can view that research in the field of concept drift has been focused even more in the recent years, hence showing that the scientific community is taking more interest in it now.

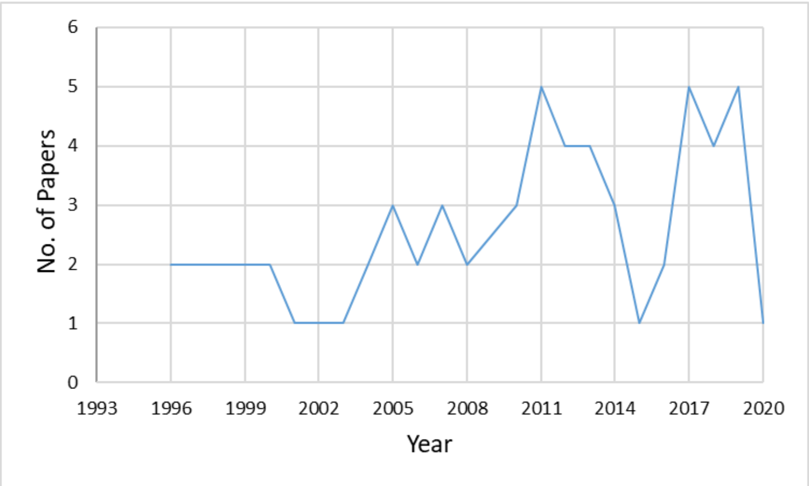


Figure 5: Papers separated by years

Figure. 6 shows the ratio of learning approaches used in the papers. It can be evidently seen, that in the area of concept drift supervised learning is by far the most used technique. The research in semi-supervised and unsupervised techniques for concept drift detection is very less as compared to supervised.

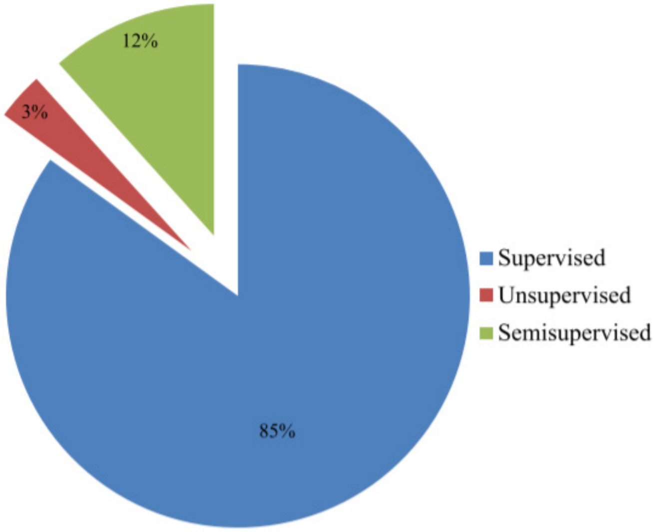


Figure 6: Ratios regarding distinct learning approaches applied for concept drift.

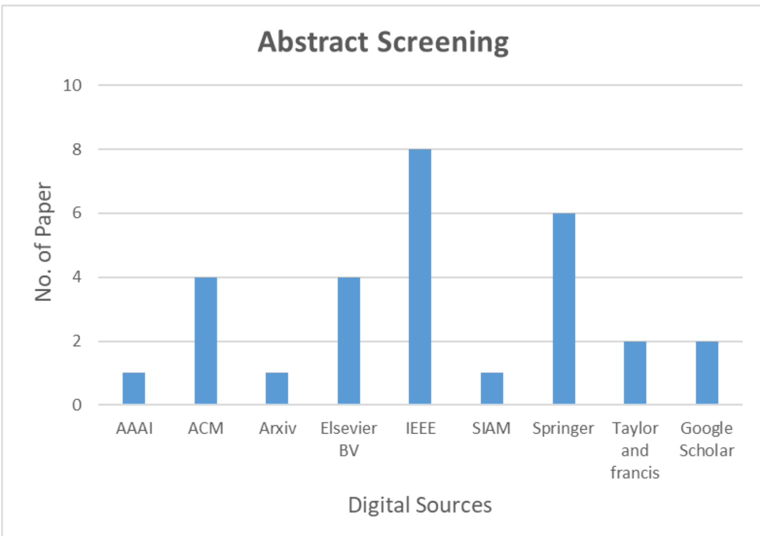


Figure 7: Distribution of 29 papers in the second stage of filtration with respect to digital sources.

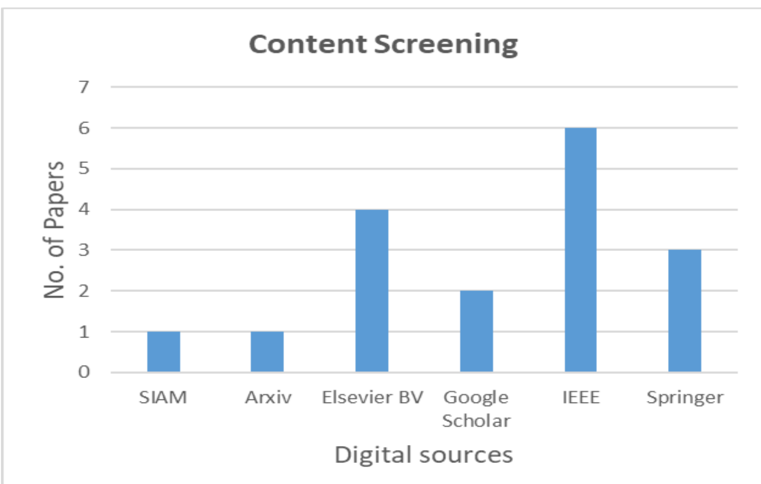


Figure 8: Distribution of 17 papers in the last stage of filtration with respect to digital sources.

4 Results

In this section, a summary of all the related articles is presented and further discussed. Table 3 contains all the basic information related to the final 17 papers that are reviewed and Table 4 contains the detailed information.

Table 1: Basic information about 17 papers

Paper	Publication Type	Venue	Publisher
(Shujian Yu 2017)	Book Chapter	Book Chapter published on 30th June 2017 in Proceedings of the 2017 SIAM International Conference on Data Mining	SIAM
(L. K. Geoffrey I. Webb 2017)	Journal article	Published in 2017	Arxiv
(João Gama 2004)	Book Chapter	Book chapter published 2004 in Advances in Artificial Intelligence – SBIA 2004 on pages 286 to 295	Springer
(M. Baena-Garcıa 2006)	Article	Published in 2006	Google Scholar
(Stanley 2003)	Journal article	Published in 2003	Google Scholar
(Hanqing Hu 2018)	Conference Paper	Conference Paper Published in 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)	IEEE
(Kantardzic 2015)	Journal article	Journal Article published 2015 in Procedia Computer Science volume 53 on pages 103 to 112	Elsevier BV
(Mohammad Masud 2011)	Journal article	Journal Article published Jun 2011 in IEEE Transactions on Knowledge and Data Engineering volume 23 issue 6 on pages 859 to 874	IEEE
(Joung Woo Ryu 2012)	Book Chapter	Book Chapter published 2012 in Convergence and Hybrid Information Technology on pages 533 to 540	Springer
(Maciej Jaworski 2017)	Proceedings article	Proceedings Article published Nov 2017 in 2017 IEEE Symposium Series on Computational Intelligence (SSCI)	IEEE
(Jie Lu 2018)	Journal article	Journal Article published 2018 in IEEE Transactions on Knowledge and Data Engineering	IEEE
(M. K. Tegjyot Singh Sethi 2017)	Journal article	Journal Article published Oct 2017 in Expert Systems with Applications volume 82	Elsevier BV

Paper	Publication Type	Venue	Publisher
(Xindong Wu 2012)	Journal article	Journal Article published Sep 2012 in Neurocomputing volume 92 on pages 145 to 155	Elsevier BV
(R. H. Geoffrey I. Webb 2016)	Journal article	Journal Article published Jul 2016 in Data Mining and Knowledge Discovery volume 30 issue 4 on pages 964 to 994	Springer
(M. K. Tegjyot Singh Sethi 2016)	Proceedings article	Proceedings Article published Jul 2016 in 2016 IEEE 17th International Conference on Information Reuse and Integration (IRI)	IEEE
(Adriana Sayuri Iwashita 2019)	Journal article	Journal Article published 2019 in IEEE Access volume 7 on pages 1532 to 1547	IEEE
(Niloofar Mozafari 2011)	Journal article	Journal Article published Aug 2011 in Computers & Mathematics with Applications volume 62 issue 4 on pages 1655 to 1669	Elsevier BV

Table 2: Detailed information of final 17 papers.

Paper Name	Contribution Level*	Data Type	Framework	Experiment	Algorithm
(Shujian Yu 2017)	Average	Streaming, labeled data	Yes	Yes	HLFR
(L. K. Geoffrey I. Webb 2017)	Good	-	No	No	-
(João Gama 2004)	Average	Streaming, labeled data	Yes	Yes	DDM
(M. Baena-Garcıa 2006)	Average	Streaming, labeled data	Yes	Yes	EDDM
(Stanley 2003)	Average	Streaming, Unlabeled data and Labeled Data	Yes	Yes	CDC
(Hanqing Hu 2018)	Good	Streaming, Unlabeled data	Yes	Yes	EFDD
(Kantardzic 2015)	Good	Streaming, Unlabeled data	Yes	Yes	MD3
(Mohammad Masud 2011)	Average	Streaming, Unlabeled data	Yes	Yes	ECSMiner
(Joung Woo Ryu 2012)	Good	Streaming, Unlabeled data	Yes	Yes	Active Ensemble Learning

Paper Name	Contribution Level*	Data Type	Framework	Experiment	Algorithm
(Maciej Jaworski 2017)	Good	Streaming, Unlabeled data and Labeled Data	Yes	Yes	RBM
(Jie Lu 2018)	Good	-	No	No	-
(M. K. Tegjyot Singh Sethi 2017)	Good	Streaming, Unlabeled data	No	Yes	MD3
(Xindong Wu 2012)	Average	Streaming, Unlabeled data	Yes	Yes	SUN
(R. H. Geoffrey I. Webb 2016)	Good	-	No	No	-
(M. K. Tegjyot Singh Sethi 2016)	Average	Streaming, Unlabeled data and Labeled Data	Yes	Yes	-
(Adriana Sayuri Iwashita 2019)	Good	-	No	No	-
(Niloofar Mozafari 2011)	Average	Streaming, Unlabeled data	No	Yes	PSCCD

***Contribution level is based on the personal opinion, regarding the relevancy with this survey**

Recently adaptive models and ensembles techniques are being researched more for concept drift adaptation while re-training of a model is being discouraged by most of the researchers. Very few unsupervised drift identification approaches have been discussed in literature. Techniques like Margin density, clustering based detection, grid-based detection and other ensemble frameworks provide satisfactory results. They give almost the same results in comparison to the benchmarking labeled techniques such as DDM. However, these techniques have been evaluated on synthetic datasets. More practical implementation of them is required to further comment on their change detection performance.

5 Conclusions and Future Work

This survey directly supports the researchers in understanding concept drift, it gives an overall view of the developments in concept drift learning. Most of the existing drift detection and adaptation techniques are proposed for labeled data streams. However, very few researches have been done to address the problem of concept drift in unsupervised or semi-supervised streams. In this survey we have tried to cover as many techniques as we can to address drift detection in unlabeled streaming data. We believe this paper provides researchers with basic understanding of concept drift and gives a know-how on applying drift identifying techniques on different domains persisting different challenges. We have tried to cover the maximum progress made in the research pertaining to this field.

Natural future research direction is obviously, experimentally comparing the above reviewed concept drift identification approaches for unlabeled data streams. Unsupervised drift detection methods need to be explored more. We have reviewed Restricted Boltzmann Machine for change detection, other areas of deep learning can also be explored for this purpose.

References

[1] Adriana Sayuri Iwashita, Joao Paulo Papa. 2019. "An Overview on Concept Drift Learning." IEEE Access (IEEE) 7: 1532 to 1547.

[2] Albert Bifet, Ricard Gavaldà. 2007. "Learning from Time-Changing Data with Adaptive Windowing." Proceedings of the 2007 SIAM International Conference on Data Mining. Society for Industrial and Applied Mathematics.

[3] AWS. n.d. What is Streaming Data? Accessed 2020. https://aws.amazon.com/streaming-data.

[4] Bartosz Krawczyk, Leandro L. Minku, João Gama, Jerzy Stefanowski, Michał Woźniak. 2017. "Ensemble learning for data stream analysis: A survey." Information Fusion (Elsevier BV) 37: 132 to 156.

[5] Geoff Hulten, Laurie Spencer, Pedro Domingos. 2001. "Mining time-changing data streams." Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '01. ACM Press.

[6] Geoffrey I. Webb, Loong Kuan Lee, François Petitjean, Bart Goethals. 2017. "Understanding Concept Drift." Arxiv abs/1704.00362 (Arxiv).

[7] Geoffrey I. Webb, Roy Hyde, Hong Cao, Hai Long Nguyen, Francois Petitjean. 2016. "Characterizing concept drift." Data Mining and Knowledge Discovery (Springer Science and Business Media) 30 (4): 964 to 994.

[9] Gerhard Widmer, Miroslav Kubat. 1996. "Learning in the presence of concept drift and hidden contexts." Machine Learning (Springer) 23 (1): 69 to 101.

[10] Hanqing Hu, Mehmed Kantardzic, Lingyu Lyu. 2018. "Detecting Different Types of Concept Drifts with Ensemble Framework." 17th IEEE International Conference on Machine Learning and Applications (ICMLA). IEEE.

[11] Jeffrey C. Schlimmer, Richard Granger. 1986. "Beyond Incremental Processing: Tracking Concept Drift." (AAAI).

[12] Jie Lu, Anjin Liu, Fan Dong, Feng Gu, Joao Gama, Guangquan Zhang. 2018. "Learning under Concept Drift: A Review." IEEE Transactions on Knowledge and Data Engineering (IEEE).

[13] João Gama, Pedro Medas, Gladys Castillo, Pedro Rodrigues. 2004. "Learning with Drift Detection." In Advances in Artificial Intelligence – SBIA 2004, 286 to 295. Springer.

[14] Joung Woo Ryu, Mehmed M. Kantardzic and Myung-Won Kim. 2012. "Efficiently Maintaining the Performance of an Ensemble Classifier in Streaming Data." In Convergence and Hybrid Information Technology, 533 to 540. Springer.

[15] Kantardzic, Tegjyot Singh Sethi and Mehmed. 2015. "Don't pay for validation: Detecting drifts from unlabeled data using margin density." Procedia Computer Science (Elsevier BV) 53: 103 to 112.

[16] M. Baena-Garcıa, J. del Campo-Avila, R. Fidalgo, A. Bifet, R. Gavalda, and R. Morales-Bueno. 2006. "Early Drift Detection Method." StreamKDD. 77 to 86.

[17] Maciej Jaworski, Piotr Duda, Leszek Rutkowski. 2017. "On Applying the Restricted Boltzmann Machine to Active Concept Drift Detection." Symposium Series on Computational Intelligence (SSCI). IEEE.

[18] Mohammad Masud, Jing Gao, Latifur Khan, Jiawei Han, Bhavani M. Thuraisingham. 2011. "Classification and Novel Class Detection in Concept-Drifting Data Streams under Time Constraints." IEEE Transactions on Knowledge and Data Engineering (IEEE) 23 (6): 859 to 874.

[19] Niloofar Mozafari, Sattar Hashemi, Ali Hamzeh. 2011. "A Precise Statistical approach for concept change detection in unlabeled data streams." Computers & Mathematics with Applications (Elsevier BV) 62 (4): 1655 to 1669.

[20] S. Wang, L. L. Minku, D. Ghezzi, D. Caltabiano, P. Tino and X. Yao. 2013. "Concept drift detection for online class imbalance learning." The 2013 International Joint Conference on Neural Networks (IJCNN). IEEE.

[21] Shujian Yu, Zubin Abraham. 2017. "Concept Drift Detection with Hierarchical Hypothesis Testing." In Proceedings of the 2017 SIAM International Conference on Data Mining, 768 to 776. SIAM.

[22] Stanley, Kenneth O. 2003. "Learning Concept Drift with a Committee of Decision Trees."

[24] Stephen H. Bach, Marcus A. Maloof. 2008. "Paired Learners for Concept Drift." 2008 Eighth IEEE International Conference on Data Mining. IEEE.

[25] Tegjyot Singh Sethi, Mehmed Kantardzic. 2017. "On the reliable detection of concept drift from streaming unlabeled data." Expert Systems with Applications (Elsevier BV) 82: 77 - 99.

[26] Tegjyot Singh Sethi, Mehmed Kantardzic, Elaheh Arabmakki. 2016. "Monitoring Classification Blindspots to Detect Drifts from Unlabeled Data." IEEE 17th International Conference on Information Reuse and Integration (IRI). IEEE.

[27] Viktor Losing, Barbara Hammer, Heiko Wersing. 2016. "KNN Classifier with Self Adjusting Memory for Heterogeneous Concept Drift." 2016 IEEE 16th International Conference on Data Mining (ICDM). IEEE.

[28] Wang, Heng, and Zubin Abraham. 2015. "Concept Drift Detection for Streaming Data." 2015 International Joint Conference on Neural Networks (IJCNN). IEEE.

[29] Xindong Wu, Peipei Li, Xuegang Hu. 2012. "Learning from concept drifting data streams with unlabeled data." Neurocomputing (Elsevier BV) 92: 145 to 155.