Turkish Online Journal of Qualitative Inquiry (TOJQI)
Volume 12, Issue 7, July 2021: 6367- 6394

Research Article

Review of Incremental and Online Learning Methods for Network Anomaly Detection

Niharika Sharma
Ph.D. Scholar, Department of Computer Science & Information Technology, Central University of Jammu, Jammu, J&K
niharikasharma990@gmail.com

Bhavna Arora
Assistant Professor, Department of Computer Science & Information Technology, Central University of Jammu, Jammu, J&K
bhavna.aroramakin@gmail.com

Abstract

The internet has emerged as one of the rapidly growing and transformative technologies over the past two decades. Due to the widespread availability of advanced network technologies, there is a serious concern about rise in threat in information and communication technology. The intrusion detection system (IDS) automates the monitoring process in computer networks and analyse the network packets and logs. Due to the static behavior of traditional data packets; the performance of the detection model reduces significantly. The intrusion detection system (IDS) automates the monitoring process in computer networks and analyse the network packets and logs. IDS should be updated timely in order to avoid the system degradation gradually. In order to adapt the model continually, incremental learning mechanism comes into picture in order to analyse network streams in real time. Incremental learning is a continuous or adaptive learning which emphasize on how a model acquire and fine-tune its knowledge. In this paper, various incremental approaches employed for detection of network anomalies are critically analysed in order to provide gist of how these techniques have influence the detection model while mitigating the effects of concept drift, noisy data, stability-plasticity dilemma, and complexity of the adaptive model. Various incremental learning classifiers along with challenges faced by incremental and online learning have also been discussed. In addition, this paper also focusses on study of anomaly detection techniques in high dimensionality and big data; that presents the comparative analysis of existing approaches to address some of the challenges of high dimensionality in large scale networks.

Keywords: Incremental learning, online learning, catastrophic forgetting, concept drift, non-stationery environments, network anomaly detection, high dimensionality, big data

1. Introduction

In high-speed networks, anomaly detection (AD) classifiers must be updated periodically in order to keep their outputs reliable over the time. The activity of the network must be stored for further analysis while updating the classifiers regularly. Therefore, it is difficult task to achieve in real-time and high-speed networks. AD model builds a dynamic normal behavioral model for a user or network by employing machine learning technique for effective detection of unknown attacks. Intelligent IDS must be fast in monitoring and analysing network streams in real time.

Conventional IDS system often suffers from many problems such as comprehensiveness, scalability, continuous learning, high false alarm rate and lack of ability to work in online environment. Therefore, an intelligent and learnable anomaly detection model is recommended that incorporates adaptive intrusion detection model. Streaming data that is generated from computer networks is dynamic and continuously evolving over the time and infinite in incoming with a high speed. Consequently, these challenges can be overcome by building an effective model to classify unseen samples without any need of prior knowledge and should update itself on arrival of new data

Incremental learning (IL) techniques process the new instances and update the data instances automatically based on the changes that occurs with time without a need to retrain the model based on previous instances.

Classical batch learning rebuilds new models repeatedly all over again during the arrival of new instances instead of integrating all new information into the previously created models. This led to the outdated models and also very time consuming. To overcome this issue there is a shift from sequential data processing to a stream data processing. Stream processing models are all-time up to date models that utilize information as soon as it is available. Therefore, reduces the cost for storage of data and maintenance. The aim of incremental and online learning algorithms are minimal processing and space by incorporating information into their model continuously. There is a lot of ambiguity in defining incremental and online learning in the existing literature. Some authors use both of these terms reciprocally. Furthermore, some authors also termed as evolutionary or lifelong learning. Online learning algorithms can also be referred to as IL algorithms with an addition in model complexity and run-time. It qualifies for endless or life-long learning on a device with limited resources.

The emergence of new types of attacks requires continuous learning as network flows are stream-based. In the context of learning from real-data streams, incremental and online learning methods have gained more attention especially when the requirement of complete dataset is not possible. Therefore, two crucial requirements of IDS are new attack type learning and handling of concept drift and detection should be online. Streaming data often suffers from the issue of concept drift which exhibit unexpected changes in data distribution over the time. In addition, the concept of continuous learning emerges that involves adaptive algorithms which are capable of learning continually from such data streams; to find clusters in the data as reported in the literature [1][2][3]. Therefore, the structure of IL has the capability to handle concept drift in a dynamic network that changes with time as well and support adaptive learning.
Therefore, an enhanced version of Self-Organizing incremental neural network (ESOINN) is presented in the literature in order to learn arbitrary topology structure without experiencing the growth of endless growing of neurons [4].**The main contributions of this paper are listed as follows:**

- Comparative analysis of various incremental and online learning for anomaly detection is presented in order to gain insights of adaptive models that can be used in predictive and decision-making problems.
- Focus on various challenges that can be encountered while employing incremental and online learning models.

- Challenges pertaining to deployment of AD model in high dimensionality and big data along with analysis of AD approaches for handling high dimensionality and big data.
- The study of key works based on various factors such as concept drift, stability-plasticity dilemma, large scale datasets, dealing with noisy data, change detection, low-space requirement are conducted to provide gist of existing approaches applied for incremental and online learning.
- Discussion regarding open issues and future directions that focus on deployment of incremental and online learning models in non-stationery environments.

The rest of the paper is presented as follows: Section 2 presents the discussion on various classifiers for incremental and online learning. Section 3 focus on different types of IL. Various challenges in the incremental and online learning are presented in the section 4. Section 5 comprises discussion on challenges faced by learning models employed in high dimensionality and big data environments. Section 6 is the core section; which details literary works for incremental and online learning; contributed to deal with various factors such as catastrophic forgetting, concept drift, active and passive approaches for change detection, etc. In section 7, open challenges and forward-looking discussions are discussed followed by the conclusion in section 8.

2. Incremental and online learning model

The current research in IL is not just only about updating the new knowledge in the existing model, but also incorporate the mechanism that learns how to update the new data as knowledge for analysis of data and drawing predictions from the classifiers. The traditional machine learning approaches in practical application environment marked the training data and then training of the marked data is done to build a model for classification which is used to predict and classify data. When the new data instance arrives, it is updated with the previous training dataset and then retraining of new complete dataset is done. The shortcomings of the data processing and data analysis in the conventional machine learning approaches are described below.

- The acquisition of labeled data needs a lot of manpower and time.
- The training of the large dataset is very time consuming.
- When new data instances arrived; the traditional approaches employed for re-training the entire dataset are not suitable for online applications.

In order to overcome these shortcomings, IL is presented in the literature where there is no requirement to re-train the prior training data. The only need is to retrain the new samples in the incremental learning stage [5]. In the incremental learning stage, the classifier can learn from the change in ratio of circles and squares from the new samples as shown in the Figure1.

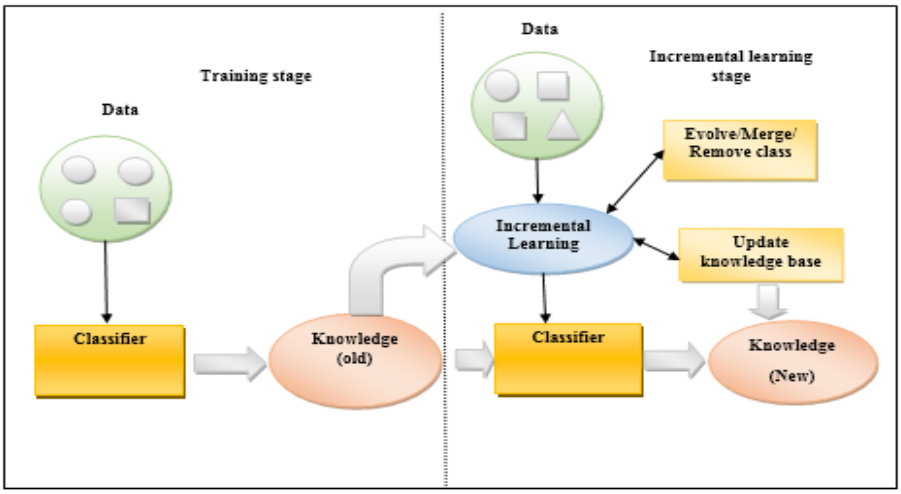

Fig1: Incremental learning model

For learning and training the new samples, the new knowledge can integrate with the old knowledge in order to form a more abundant knowledge system. The classifier can adapt, identify and update new classes by learning the triangle which is a new class that is never trained before as presented in the Figure 1 [5].

The various classifiers that support incremental learning are Naïve Bayesian, Support vector machine (SVM), Decision tree, Random forest, K-nearest neighbor (KNN), Artificial neural network (ANN), Learn ++ NC.

The characteristics of various incremental classifiers are presented in the table 1.

Table 1: Incremental classifiers and its characteristics.

Classifier	Incremental Learning	Characteristics	References
Naïve Bayesian	✓	Fusion of old and new knowledge.	[6]
Support Vector Machine (SVM)	✓	Learning of new support vectors based on previous support vectors that were reserved before.	[7]
Decision Tree	✓	The inheritable characteristics of a tree structure learn new knowledge by calculating the information gain in order to inherit the redundant instances or update the new information by reconstructing the sub-tree.	[8]
Random Forest	✓	It has characteristics of decision tree. The integrated idea of different decision trees make it possible to achieve incremental learning by learning new knowledge, train a decision tree, and	[9]

		then finally adding it into original random forest.	
Artificial Neural Network (ANN)	✓	It learns the incremental knowledge by itself i.e. the weights of large number of interconnected neurons is constantly changing based on the arrival of new data samples during incremental learning process.	[10]
K-Nearest Neighbor (K-NN)	✓	Finding the nearest neighbor in an incremental fashion. The nearest neighbors are computed one by one without the need to recomputed the query from scratch.	[11]
Learn ++ NC	✓	It is inspired by AdaBoost algorithm that inherits the characteristics of performance improvement of AdaBoost. It learns from the data that is continuously available even if the new data introduces the new classes that were not seen before.	[12], [13]

3. Types of incremental learning

Incremental learning (IL) can be segregated into three categories depending upon the difference in data change.

- **Data IL:** To improve the performance of prediction model; data IL utilizes newly available training samples on its arrival. It is also known as online learning.

- **Class IL:** Class IL learn from old models and exploit knowledge from previous classes to improve the learning of new classes. The basic idea of class incremental models is to reuse the old models i.e. storing a small number of useful previous tasks data. After learning new classes, a small amount of learned task from new classes could be accumulated while the rest can be discarded [14]. The key idea is to determine the instances of new class from that of old classes. Firstly, a small subset of classes is learned and then grow that set with inclusion of new classes incrementally [15].

- **Hybrid IL:** The union of data and class incremental learning is hybrid learning.

3.1 A review of types of IL

In this section, a brief overview of a few literary works is presented to differentiate the studies pertaining to different classes of incremental learning.

Data IL

Xie et al. proposed a data incremental learning approach that employ K-nearest neighbor algorithm. To increment the data; the concept of layer is introduced to the model cluster. When the new sample arrives; it is covered by many layers and the highest layer value of the model is selected to obtain the best classification results[16].

To address the issue of online learning in small sample data streams; Wang et al. [17] presented an incremental random forest approach in which the new data sample is stored on the current leaf node. Furthermore, experimental results on UCI datasets reveal that incremental learning algorithms outperform greedy reconstruction tree algorithm on small or moderate scale data streams.

Class IL

Marc et al. describe a set of classes as a sequence of task that are disjoint from classes in previous or future tasks. The tasks consist of a number of classes. Exemplars are used to store previous tasks that consider a small memory and the learners only has access to data from a single task during each training session. The current task can be processed multiple times by the learners to process the data during the training session in order to improve the learning of new ones. Furthermore, to learn knowledge from current task, IL approaches must learn how to balance between knowledge retaining from previous tasks and learning of new knowledge for the current task. This trade-off is termed as stability-plasticity dilemma.

One of the fundamental problems with class IL is that too much plasticity in the model leads to easily forgetting the old classes whereas high stability models are weak to learn new classes. In order to mitigate this issue; a novel network architecture is presented to balance the stability and plasticity problem dynamically [18].

IL often suffers from class imbalanced problem where the real-life applications often contain imbalanced datasets. Due to memory constraint, old classes are learned with fewer knowledge than new classes. This includes majority and minority classes where old classes represent majority classes and new classes represent majority classes. There is a prediction bias towards the majority class. Umang et al. [19] employ calibration methods to lessen the prediction bias between majority and minority classes. Furthermore, vanilla fine tuning is applied to only fine tune the classification layer. However, authors concluded that calibration results are useful in dealing with class imbalance problem; it still remains an open problem.

In computer vision, semantic segmentation is a hot topic in research field that focuses on predicting the class of each image in a given pixel. The inaccessibility of previous data is the main challenge for class-incremental semantic segmentation. In order to address this problem, self-training is applied to incremental semantic segmentation that uses auxiliary unlabelled data to alleviate forgetting by rehearsal of previous knowledge. Furthermore, conflict reduction strategy is presented to cater the conflicts of pseudo labels induce from both the old and new models [20].

Hybrid IL

A hybrid IL framework is proposed that incorporates data-incremental and class-incremental framework for recognition of hand-held objects. The proposed model is based on SVM that learn new concepts incrementally in order to improve the recognition ability. New classification planes are added and the existing classification planes are adjusted under the setting of SVM. In the proposed

model, prediction error is minimized that improves the recognition quality of known objects and the preceding model is transferred in order to recognize the unknown objects [21] .

4. Challenges in incremental and online learning approaches

- **Concept drift:** The shift in the data distribution over time is termed concept drift. The temporal structure of data samples is responsible for variations in the data distribution. Such changes can be reflected due to change in data distribution, changes in the underlying functionality itself.
- **Stability-plasticity dilemma:** When new information arrives, a quick update is made for swift adaptation rendering the new environment. However, the old information is also forgotten rapidly and easily. This trade-off of quickly adapting and forgetting is considered a stability-plasticity dilemma. Consequently, the adaption should be followed slowly in order to keep the useful old information.
- **Complexity of adaptive model:** The complexity of the model is difficult to assess when the samples of the data are unknown. Therefore, intelligent adaptive methods are required to overcome the limitations of the available resources. One such promising solution is batch learning in which the fundamental meta-parameters such as learning rate, degree of error, regularization constants, etc. are assessed prior to training[22]. However, concept drift affects the learning rate parameter and convert into modal parameters according to the change in data characteristics. Thus, an ensemble of robust parameters is required to address this problem.
- **Online learning:** In online learning, the adaption of the internal model depends on the small number of samples that are processed over mini-batch techniques. In fully online approaches the internal model promptly adapted upon processing an individual sample whereas batch learning processes definite number of samples internally. Therefore, it is critical to choose the required learning process based on the data and number of resources available.

5. High dimensionality and big data

The term “Dimensionality” assumes a critical part in real-world data analysis. It represents the number of attributes or features or variables present in the data that are available for analysis of data. The complexity of the data analysis is determined concerning the number of dimensions that require further advanced methodologies in order to process the data The increase in the number of dimensions or features creates a problem for anomaly detection due to high dimensionality that poses a considerable challenge for the detection of anomalies in large datasets [23]. The large and distributed nature of the datasets characterizes the term big data, which has become the key research problem for many applications employed in the real world. The high velocity and high volume comprise the big data that is generated by distinct datasets. The rate of the data which is generated in real-time may comprise many dimensions (high volume) that affect both the accuracy and performance of the existing techniques. In this section, various challenges faced by high dimensionality and big data in terms of various parameters are discussed in this section. In addition, a few key works is cited to address these challenges as mentioned in table 3 that are employed in high dimensionality and big data platforms as shown in table 4.

The various challenges pertaining to anomaly detection model while deploying in high dimensional and big data environments are discussed in table 2 below.

Table 2: Challenges contributed to anomaly detection in context of problems faced by high dimensionality and big data.

	High dimensional data	Big data
1.	**Relevant attribute identification:** Finding the relevant quantitative location of data points in high dimensional space is a complex task.	**Asynchronous instances:** The multiple data sources may come from many sources that contribute to generation of data points which arrive at different times [24] .
2.	**Distance concentration:** The data instance nearly become equidistant to each other based on the distance; due to the sparsity of data measure used [25][26].	**Dynamic relationship: The asynchronous behavior of data is due to the dynamic relationship of correlated data points that are continuously monitored from multiple data streams.**[24].
3.	**Subspace selection:** The increase in features of subspace results in exponential search space due to the increase in the dimensionality of data.	**Heterogeneous schema:** The data instances which are coming from multiple data sources exhibit various schemas. So compilation of multiple data instances over distinct schemas is a tedious task [24].
4.	**Hubness:** The hubs are described as behavior of data samples in high dimensional data which are constantly arriving in nearest neighbors.	**Concept drift:** The distribution of data changes over time. The prediction model may change over the time due to change in properties of the target variable [27].

Strategies for handling high dimensionality and big data in large scale networks using different anomaly detection approaches are investigated in the following table 3.

Table 3: Comparative analysis of anomaly detection approaches for handling high dimensionality and big data

Authors & Year	Discussion	Approaches	Distributed/ Parallel processing	Scalability	High dimensionality	Hubness	Multiple data streams	Performance
Lozano et al. [28] 2005	Presented parallel algorithms in order to classify distance-based anomalies by employing randomization and pruning strategies.	Local outlier factor	✓	✓	✓		✓	✓
Angiulli et al. [29] 2007	A stream outlier miner known as STORM is presented to handle multiple	Distance-based		✓	✓		✓	

	data streams and the anomalies are identified using the current sliding window							
Kontaki et al. [30] 2011	Effective anomaly monitoring is done to perform constant anomaly detection in data streams by applying sliding windows with lesser memory requirements.	Distance-based		✓	✓		✓	
Silva et al. [31] 2013	Address some of the issues related with multi data streams such as asynchronous instances, dynamic relationships, heterogeneous data using sliding window concept.	Distance-based		✓	✓		✓	
Tomasev et al. [32] 2014	The issue of clustering in high dimensional data is addressed by assessing the points that are frequently occurring known as hubs.	Clustering		✓	✓	✓		
Radovanovic et al. [33] 2015	To distinguish the connection between outliers, anti-hubs are examined to separate the normal instances from outliers.	Density-based		✓	✓	✓		
Angiulli et al. [34] 2016	Presented a set of distributed and parallel algorithms for distributed anomaly detection model in order to scale up the performance.	Distance-based	✓	✓	✓		✓	✓
Bai et al. [35]	A grid-based partition algorithm is	Distributed Local	✓	✓	✓		✓	✓

2016	employed to split the dataset into grids in order to identify density-based outliers.	outlier factor						
Colin et al. [36] 2016	Introduced a PCA based outlier detection method to identify anomalies in distributed environment	PCA	✓	✓	✓		✓	✓
Zhang et al. [37] 2017	The detection of anomalies using subspace selection from non-stationery high dimensional data streams.	Unsupervised		✓	✓		✓	

6. Related works

This section comprises the literary works that have been investigated to highlight the key works employed for incremental and online learning for anomaly detection mainly in networks. The literature survey is presented in various sub-sections comprising the various parameters that influence the working of incremental and online learning models. The sub-section part include how catastrophic forgetting affects the incremental learning and SOINN is proposed in the literature to evade the effects of catastrophic forgetting. Several works have been presented to update the learning model until the detection of concept drift. Furthermore, several studies related to different application areas are also discussed in the literature survey. Finally, summarization of key works that focus on adaptive and real-time learning in context of working with large datasets that require low-space requirements, high detection accuracy, dealing with noisy data, and stability-plasticity problem are presented at the end of this section.

6.1 Catastrophic forgetting

One of the simple ways of incremental learning is to tune the model to the data of new class. To make the model adaptive in nature; the two key points that need to be followed are updating the new instances and ignoring the outdated normal instances. To adhere the requirement of low memory space; simplification is performed to represent the data in bounded memory space. This could result in loss of some finer details while performing simplification which results in catastrophic forgetting. One of the main challenges with incremental learning is catastrophic forgetting, which attribute to abrupt loss in performance on prior learned task after learning a new task.

Many online learning methods with limited memory resources exhibits catastrophic behavior. It is also termed as stability-plasticity dilemma The key advantage of continuous and online learning system is to incorporate the new data without eliminating the need of previously accumulated knowledge i.e., without catastrophic forgetting [38]. However, online learning system often suffers

from the problem of stability-plasticity dilemma. “The well-known limitation for artificial and biological neural systems is the trade-off between stability and plasticity”[39]. The incorporation of new knowledge is referred to as plasticity whereas stability is required to avert the forgetting of past knowledge. Too much plasticity in the model results in constantly forgotten the previous knowledge and too much stability will deter the efficient coding of data. Stability-plastic dilemma is a major challenge for machine learning as well as deep learning techniques. In neural networks, the weights of new input data are changed in order to make the system adaptive or plastic. However, too many weight changes can infer to the loss of previously acquired knowledge. Generally, this type of problem is prevailing in various types of neural networks ranging from standard back-propagation neural networks to unsupervised neural networks like self-organizing maps.

Self-organizing maps: The concept of self-organizing neural networks is reported in the literature to evade the forgetting of past knowledge in incremental models. “Self-organizing maps (SOM) is a type of artificial neural network that employs unsupervised learning and groups the similar data together” [40] [41]. The training samples of the input space are represented as map in order to produce the low dimensional space and therefore this method is applied to dimensionality reduction. However, shortcomings of SOM is that it learn complex topologies without a proper predefined graph that affects its performance. This approach is not suitable in real-time datasets as the requirement of proper estimation of dimensions is not known a priori. Consequently, it adds the network complexity when the number of tasks increases.

Self-organizing incremental neural network: For continuous learning from non-stationery data, a class of neural networks are designed called self-organizing neural networks (SOINN). SOINN consists of one or more neural network that extract a topological structure in order to nearly emulate the data distribution of data streams [42]. In SOINN, the maximum number of nodes are not fixed in continual learning tasks unlike SOM. The number of nodes is usually smaller than the number of training examples as new training cases may be added as new nodes or they may be merged with the existing nodes. Several modifications have been proposed over the years for SOINN. Enhanced self-organizing incremental neural network (ESOINN) overcome the limitation of SOM that compromise its performance while learning complex topology structures [4]. ESOINN learns arbitrary topology structure without suffering from endless growing of neurons thus making it suitable to work in large scale networks. Several literary works have shown that ESOINN is more stable; that uses only one layer and fewer network parameters. Many more versions of SOINN have been proposed over the years such as mixture SOINN, load balancing SOINN, adjusted SOINN, kernel density estimation SOINN etc. For graceful forgetting, it should be considered as intrinsic part of the learning process.

6.1 A review of SOINN

In [43], authors proposed an unsupervised incremental learning neural network that depends on local distribution learning. It amalgamates the benefits of matrix learning and IL in order to discover the most appropriate nodes that fit the learning model. The presented approach does not require prior information like structure of the network and the new knowledge is updated automatically without letting the number of nodes to grow infinitely.

Xiang et al. [44] presented an incremental semi-supervised learning (SSL) framework for network intrusion detection in order to provide solution to concept drift problem and employ topological learning for low-space requirements. The authors combine incremental learning, modeling of non-linear data, and SSL. The modification of ESOINN is presented in the form of semi-supervised learning called mixture SOINN (MSOINN) in order to process the amalgamation of labeled and

unlabeled data incrementally. The knowledge re-use framework is applied to leverage the already stored knowledge in the trained neural networks. The kernel function is applied to train the SVM and then employ it to detect intrusions. The presented methodology reveals significant advantages over other supervised learning approaches such as transductive SVM (TSVM). Only a limited number of samples is needed and the proposed approach improves space complexity.

A mixture of kernel density estimation and self-organizing incremental neural network (KDESOINN) is proposed in order to work in real-time environments. The proposed approach first learns the observed samples using a modified SOINN by interpreting them as networks. The nodes in the network act as prototypes for the samples. The authors found that the information regarding underlying distribution of observed samples are collected by network structure of SOINN [45]. Chayut et al. [38] presents the modification of self-organizing incremental neural network (SOINN) called SOINN+ that represents how forgetting is modeled in the adaptive neural network system. The proposed method is different from the other models with respect to how edges are created and deleted and how nodes are deleted. It also determines clusters in noisy data streams that are subjected to persisting concept drift.

6.2 Concept drift

The changes in distribution of data with time is referred as concept drift. The changes in the input distribution over time are alluded as virtual concept drift or covariate drift whereas changes in the underlying functionality itself termed as real concept drift. The occurrence of real concept drift is problematic as the classification performance is affected until the model can be re-adapted accordingly.

In order to work in evolving environments and reacting to concept drift, just-in-time classifier (JIT) cope up with the process change. For an adaptive management of knowledge base during arrival of new information, JIT require temporal detection of process divergence that removes the obsolete information and insert the new one. The CI-CUSUM works well in detection of small abrupt changes and presence of drifts in the model. The change index is used to estimate the change in process over time i.e., identifying the obsolete knowledge. This index is employed in the k-nearest algorithm via weighting mechanism. The proposed adaptive weighted K-NN approach is applied along with the use of change index in order to react better to smooth variations. However, the proposed approach fails to react to abrupt variations in the model and K-NN classifier does not perform well with very large amount of data instances.

To work in the presence of concept drift; two approaches are presented in the literature in which the learning is based on active or passive approaches. The adaption algorithm learns the behavior of the model in the existence of concept drift.

6.2.1 Approaches for change detection in the presence of concept drift.

The two approaches are presented in the literature to work in the presence of concept drift in which the learning is based on active or passive approaches. The adaption algorithm learns the behavior of the model in the existence of concept drift. To detect the change in the learning model; the following approaches are described as below.

- Active approaches
- Passive approaches

Active approaches: In active approaches, algorithm aim at finding the concept drift and then update the learning model. Active approaches are based on change detection mechanism that learns under the

presence of concept drift. Whenever the change is detected, it triggers an adaptation mechanism that update or build a new classifier in order to react at the detected change. [46].

Active approaches do not operate directly on raw data. To detect the change independent and identically dependent features are extracted from the incoming data streams such as sample mean, sample variance and or the classification error [47]. Active approaches employed by various researchers as reported in the literature [48] [49] [50] [51] [52] to detect the change occurred due to concept drift and react to it accordingly. The methods employed for change detection can be further classified into four main families such as hypothesis tests, sequential hypothesis tests, change-point methods and change detection tests. These families inspect variations via theoretically grounded statistical techniques.

Passive approaches: In passive approaches, model is adapted continuously based on updating the model parameters every time whenever the new data arrives. This approach does not detect the drift directly but rather accepting the change in data distribution at any time with change in rate. The passive approaches work with updating a single classifier or ensemble-based classifiers.

The passive approaches work with updating a single classifier or ensemble-based classifiers. The selection of appropriate approach is specific to the application that further depends upon the parameters of the learning scenarios for e.g., to assess whether the data arrive online or in batches, drift rates, availability of computational resources (embedded systems or high-performance computers), presumption about distribution of data. Passive approaches are mostly used in environments with gradual drifts and recurring concept [53] and are quite effective in making predictions from the learning model. On the other hand, active approaches are suitable in settings where the drift is precipitous. Moreover, passive approaches are widely used for batch learning and active approaches are well suited to work in online environment.

6.2.2 Incremental learning approaches with respect to application areas

Comparative analysis of incremental learning approaches with respect to application areas are shown in the following table 4.

Table 4: Comparison of incremental learning approaches with respect to application areas

Author Name & Year	Methodology	Type of Algorithm	Approaches	Application areas
Jose et al. 2017 [54]	An incremental online approach is presented in order to adapt to accommodate new data and integrate new classes by the robots to operate in real environments.	Deep Learning	Convolutional Neural Networks (CNN), Self-organizing incremental neural networks	Robotics
Binhan Xu et al. 2017 [55]	Incremental K-NN SVM algorithm is proposed for cloud environments in order to deal with large scale	Supervised learning	Incremental K-NN, SVM	Intrusion detection

	and dynamic networks			
Mohammadreza et al. 2018 [8]	Outlier detection algorithm is presented to identify outliers in a sequential data stream.	Supervised learning	Incremental decision tree	Outlier detection
Akila et al. 2019 [56]	An ensemble based parallel bagging approach is presented for incremental learning that take care of concept drift and data imbalance	Transaction window bagging	Parallel and Incremental learning ensemble, Naïve Bayes, weighted voting-based combiner	Real-time credit card fraud detection
Dinithi et al. 2019 [57]	Traffic forecasting and real time concept drift is detected for intelligent traffic management in big data platforms such as IoT sensors and social media.	Unsupervised learning, Deep learning	Unsupervised incremental machine learning, Deep learning, deep reinforcement learning	Smart traffic management in big data.
Eden et al. 2019 [58]	Concept of dual memory is presented in order to store past class statistics that are gathered during initial learning. First exemplar images of past classes are stored in memory. The introduction of small memory is presented such that the classes are best modeled when all their data is available across different incremental states.	Deep learning	Deep neural network (DNN)	Computer vision
Bittencourt et al. 2020 [59]	The minimum description length principle is applied in multi-label text classification in order to classify multi-label documents which naturally supports online learning	Supervised learning	Minimum description length principle	Text classification
Ali Ayub et al.	A class incremental	Centroid-based	Few-shot	Image

2020 [60]	approach is proposed that is cognitively inspired approach. The novel strategy is proposed called centroid reduction mechanism that reduce the consumption of memory without notable loss in classification accuracy.	concept learning	incremental learning, centroid reduction method	processing
Marc et al. 2020 [61]	Emergency services receive valuable information from social media during disasters and emergencies. To filter out relevant information; active incremental learning is employed.	Active incremental learning	Active and online learning	Information processing and management

6.3 Key Works

In this section, some of the key works of authors are discussed to gain the insight of literary works mainly presented in table 5 along with some limitations of the existing works as proposed by various authors. Different incremental algorithms have been published so far along with its pros and cons. However, there is no depth study available that experimentally compare the most popular methods with the existing ones. This is due to the fact, that experimental studies are performed in specific settings which emphasize on the merits of their proposed method according to the suitable criteria and apply them in specific settings. Therefore, it provides a limited picture of various incremental and online learning algorithms till date that cater the needs of an incremental model which perform optimally in every scenario. There are a very few publications with practical focus in the field of IL in a general way. Instance incremental approaches are equally accurate as batch-incremental models; but it requires less resources and less computational overhead to process the accumulated instances in batch mode. The performance of lazy methods with a sliding window concept is exceptionally well.

Among the various incremental classifiers such as incremental support vector machine (ISVM), LASVM, online random forest (ORF), Incremental Learning Vector Quantization (ILVQ), Learn++ (LPP $_{CART)}$, Incremental Extreme Learning Machine (IELM), Naïve Bayes ($NB_{Gauss),}$ Stochastic Gradient Descent (SGD_{Lin}), SVM deliver the highest performance at the expense of model complexity due to training of large data samples. Due to the approximate nature of LASVM; it lessen the training time of classifier and is able to process large datasets faster than ISVM. However, LASVM cannot work well with noisy datasets [62]. For non-stationary environments, LAA and SVM work well by incorporating forgetting mechanism [63] [64] [65] [66]. Viktor et al. [67] performed experiments on various incremental classifiers such as ISVM, LASVM, ORF, ILVQ, Learn++, IELM, Naïve Bayes, and SGD; and reveal the results regarding the performance analysis of distinct set of classifiers. It is found that performance of ORF is slightly worse than ISVM and LSVM but possess very fast training and run time. In order to learn in endless streams; both SVM and ORF are not suitable as it grow

linearly with the rise in number of data instances than the rest of the remaining methods. The ILVQ serve an accurate and sparse alternative to SVM. Tree based models are suitable to work in high dimensional data due to its compressed representation and sub-linear run-time that is not dependent on the number of dimensions. Furthermore, SGD_{Lin}, NB_{Gauss} and IELM are suitable for online learning due to its applicability in life-long learning applications which shows constant performance in terms of complexity. For high dimensionality; the sparse models of $SGD_{Lin,}$ NB_{Gauss} are suitable for large scale learning. NB_{Gauss} and tree based methods require no or little hyper-parameters optimization and are easy to implement in practice. . In terms of complexity, SGD_{Lin} and NB_{Gauss} have linear complexity among the rest of the classifiers.

Gao et al. [68] proposed an adaptive IDS that employ incremental extreme learning machine (I-ELM) and an adaptive PCA. In the presented approach, the most important features of the network traffic are selected automatically and I-ELM is employed to achieve the best detection accuracy. ELM possess many advantages over ANN, SVM, CNN such as fast convergence and training, strong learning capability, approximating a linear function, faster training speed, and higher detection accuracy etc. This allows the detection model to work well in large datasets. I-ELM is an adjustable network structure that reduces the training error by joining nodes in order to mitigate the problem of overfitting and underfitting.

In the continuous stream of flow data, the biggest challenge is to acquire new knowledge from the raw data and transform it into the knowledge that is capable of representing information and accumulating experience overtime for further decision-making process. Haibo et al. [69] proposed a novel IL framework called ADAIN that learns from continuous raw data adaptively. The accumulation of knowledge from raw data over the time is useful to make prediction about performance and improve future learning. The proposed framework investigates how effectively the past learned knowledge that is integrated into the currently received data. To continuously learn from new data and to further improve the learning process; such experiences are accumulated over the time in order to support the future decision-making process.

The authors [70] proposed a novel approach in order to learn network traffic automatically using unsupervised learning. The proposed approach is online and real-time; that utilizes discrete-time sliding window and incremental grid clustering algorithm is applied for continuous detection of network anomalies. The usual clustering algorithm split the whole space when new data points are added or removed instead of partitioning only a few points. Each subspace is partitioned as units using incremental clustering algorithm and the authors named this methodology as Orunada. The grid clustering algorithm scales well with the number of points in the case of large-scale networks due to the partitioning of subspace into units rather than points. The presented approach employs a density-based clustering algorithm that works well with noise. Furthermore, DBSCAN considers isolated points as outliers and noise which do not belong to any cluster.

In order to address several challenges; that are encountered while employing anomaly detection techniques are: dealing with noisy data, processing speed, adapting to variations in a dynamic environment. For this; an efficient anomaly detection model is required which is noise-resilient, fast, and incremental. Bigdeli et al. [71] proposed an incremental approach, in which the first phase is to cluster the data. The second phase involves clustering the data in such a way that it can accommodate new data instances in order to classify them. In the third phase the representation of GMM is used to update the clusters and rejecting the repetitive data samples while detecting the anomalies. A novel strategy is employed called collective probabilistic labeling which is utilized to incrementally update the clusters. This makes the updated procedure fast. In this approach, all the instances of the cluster are not preserved; only the main components of the cluster are maintained without holding the direct

records of all data samples present within the cluster. GMM helps to identify whether a new instance belongs to a cluster set or not. The presented approach improves the false alarm rate 94% detection rate and a 4% false alarm rate.

Ensemble classifier models gain more popularity over single classifier-based system due to distinct advantages such as

- Accuracy
- Flexibility of incorporation of new data into a classification model on the arrival of new data instance by simply adding new ensembles
- To administer the balance between the stability-plasticity spectrum by adding or removing classifiers.

Ensemble based learning can also be applied to transfer learning, multi-task learning. It is also applied to approaches particularly suitable for handling concept drift.

In the work of [72], the authors employed ensemble learning that comprises transfer learning, deep learning and incremental learning. Deep learning is applied in source domain and make it possible to take the advantages of incremental learning in the target domain where knowledge learnt in the source domain is applied to target domain. In the proposed ensemble learning approach, the proposed model is trained by a group of source domain datasets in which the features are extracted using Convolutional Neural Network (CNN). The role of transfer learning in the proposed study is to extract important information from the different datasets using pre-trained CNN. The knowledge of CNN learned in previous object is utilized in order to reduce the modeling time. Some pre-trained CNN models are transferred to extract the features from other target domain. Features are extracted using Convolutional Neural Network (CNN) and then it is applied to ISVM to train these features without pre-processing.

Summary of key works for incremental and online learning is presented in table 5 to provide general comparison of various incremental approaches that are scalable, online, handle noisy data & concept drift, and which require less space requirements.

Table 5: Comparison of various incremental and online learning approaches for anomaly detection

Authors & Year	Techniques	Key points	Datasets	1	2	3	4	5	Accuracy	Limitations/Future Work
Haibo et al. [69] 2011	Multilayer perceptron (MLP), Learn++	To learn from continuous flow i.e. raw data adaptively in order to improve future learning and supports future	Spambase, Magic, Waveform				✓	✓		The problem of concept drift is not addressed in this study.

		decision making process.								
Dariusz et al. 2014 [73]	Ensemble classifier, Hoeffding trees	An ensemble classifier is proposed to respond to different types of concept drifts in data streams.	UCI repository	✓	✓				89	Not implemented in a truly incremental fashion especially in partially labeled streams.
Bosman et al. 2015 [74]	Recursive Least Square (RLS) method, Extreme learning machine (ELM)	A lightweight application independent framework to incorporate incremental learning in WSN.	Real and synthetic datasets	✓	✓		✓	✓	97	The memory requirement of the presented strategy is dependent upon the number of sensor nodes.
Fakhrodd in et al. 2015 [75]	Semi-supervise d	A stream classificatio n algorithm using semi-supervised learning is proposed in order to work with limited class labels and handle imbalanced data.	KDD99, real-time datasets				✓	✓	87	Active learning needs to be incorporated in order to deal with unavailability of labeled data.
Xiang et al. 2016 [44]	Semi-supervise d	To provide solutions to the problem of concept drift and employ topological learning for low-space	NSL-KDD		✓	✓	✓	✓	85	The proposed framework is not applicable to multi-class problems.

		requirements.								
Dromard et al. 2017 [70]	Incremental grid clustering	An anomaly detection algorithm is proposed for online and real-time learning using discrete time sliding window in order to learn behavior of network traffic automatically.	ONTS and MAWIlab	✓			✓	✓		Computational complexity is high.
Bigdeli et al. 2018 [71]	Spectral based and density based clustering	The network anomaly detection approach to update the clusters incrementally while ignoring the redundant instances.	KDD-99, DARPA98, NSL-KDD	✓		✓			94	The new information can be collected till detecting the concept drift in order to effectively using the updating strategies.
Viegas et al. 2019 [76]	Ensemble classifiers	In order to store the network activity for further analysis and updating the classifiers regularly by marinating the reliability in the outputs	MAWIflow				✓	✓		Need human assistance for labeling of rejected instances in order to collect more information about a new behavior.

		of the classifier.								
Gao et al. 2019 [68]	Incremental ELM	Relevant features of the network traffic are adaptively selected and incremental ELM is applied to achieve better detection accuracy.	NSL-KDD, UNSW-NB15				✓		82	To further improve the detection accuracy in real-time environment and extend this approach to industrial control system.
Chayut et al. 2020 [38]	Self-organizing incremental neural network	It represents how forgetting gracefully modeled in the adaptive neural network system by considering three parameters: idle time, trust worthiness, and un-utility of a node.	Synthetic datasets	✓	✓		✓	✓		The model is only designed for unsupervised learning

Note: 1, noisy data; 2, concept drift; 3, space complexity; 4, online; 5, scalability

In Figure 2 the accuracy rate of various incremental and online approaches are presented in which the work of [74] shows highest detection accuracy rate.

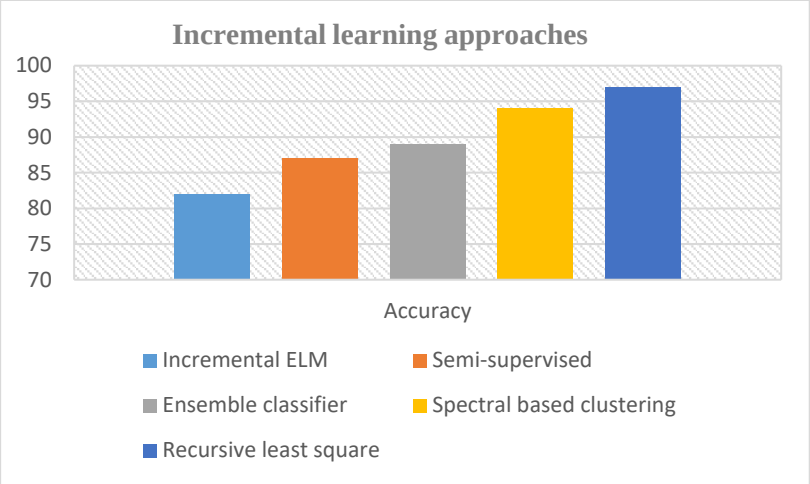

Figure 2: Accuracy rate of incremental and online learning approaches

7. Open issues and future directions

Incremental and online learning in non-stationery environment is a promising area of research in machine learning and computational intelligence due to its increasing pervasiveness in streaming and big data applications. Various studies have been presented in the literature to deal with challenges that are associated with incremental and online learning. However, existing work provides a limited picture of various incremental and online learning algorithms till date that cater the needs of an incremental model which perform optimally in every scenario. There are a very few publications with practical focus in the field of IL in a general way. Although, this list is not exhaustive, there are certainly several open issues and future directions that are found during the development of this paper. The following is the list of some open challenges and future directions that are discussed as below:

- **Ensemble of meta-parameters:** The essential parameters such as learning rate, degree of error, regularisation constants, etc. are estimated in batch learning prior to training. To handle concept drift; meta-parameters such as learning rate plays a critical role in order to adapt the internal model continually. Concept drift turns the learning rate into model parameters. Some incremental strategies adapt to changes till detecting concept drift; since their choice is to adapt the changes according to the change in data characteristics. For this purpose, ensemble of robust meta- parameters along with meta-heuristics are required to optimize the internal cost function of the learning model during training.
- **Unstructured and heterogeneous data streams:** To accommodate enormous amount of unstructured and heterogeneous data; big data possess challenges in order to mine data from heterogeneous data streams. The unstructured and heterogeneous data such as texts, images and graphs acquired from different data streams can possess different characteristics for e.g., multi-label, multi-dimensionality, spatial relationships and multi-scale. The focus of online research should include adaptive strategies that include new modeling under the presence of concept drift to handle such data.
- **Concept drift:** One of the challenges with IL models is dealing with concept drift at run time. Different techniques are employed for addressing concept drift based on its type for e.g., gradual drift is addressed by passive methods, abrupt concept changes is handled by active approaches. Furthermore, virtual concept drift concern with input distribution only whereas

real drift is directly associated with classifier performance. A lot of work needs to be investigated in this direction by incorporating such factors into the incremental model especially to address real concept drift.

- **Efficient memory models:** IL models work with limited data resources are bounded to store information in compact form provided by the observed data. "It is possible via various schemes such as classification error for explicit drift detection models" [46], model parameters in implied form (such as prototype or exemplar based methods) or explicit memory models [77] [78]. In order to store information in the form of examples; prototype or exemplar-based methods are exploited whereas explicit memory models represent memory in the form of a parametric model or depends on finite window of a particular training examples. Consequently, it is critical to carefully design the memory model for adaption in order to avoid the stability-plasticity dilemma.
- **Transient concept drift:** In evolving environments where the concept drift is transient and the samples related to the change are limited then it is more challenging to detect change detection than the permanent ones. This is due to the fact that very small sample size is available for estimating the features that are utilized by change detection strategies. Consequently, it adds an overhead to effectively learn the features that contribute the change detection.
- **Consensus maximization (CM):** The assumption for presenting labeled datasets for supervised algorithm or unlabeled for unsupervised one. Though, this assumption does not hold well in case of data streams that comprises a mixture of labeled and unlabeled data. The most widely used robust fitting parameters in computer vision is CM which is also an active research topic. In robust model fitting; outliers are removed in order to achieve robust and stable model. CM strives at designing a framework to construct and incorporate various supervised and unsupervised models for prediction. The use of Consensus maximization can be exploited for non-stationery data streams [79] [80] [81].

8. Conclusion

Incremental learning is a continuous or adaptive learning which emphasize on how model acquire and fine-tune its knowledge. In this paper, various incremental and online learning approaches employed for network anomaly detection model is critically analysed in order to provide gist of how these techniques have influence the detection model while mitigating the effects of concept drift, noisy data, stability-plasticity dilemma, and space complexity of the adaptive model. Several incremental classifiers along with the various challenges that come across while implementing incremental and online learning models are discussed. In large scale networks, the number of dimensions or features increases pose a great challenge for detection of anomalies in large datasets due to the amount of data require to generalize also increases. Comparative analysis of anomaly detection approaches pertaining to various challenges employed in high dimensionality and big data environments have been presented. This paper also summarizes key works of various researchers that focuses on adaptive and real-time learning in context of working with large datasets that require low-space requirements, high detection accuracy, dealing with noisy data, and stability-plasticity problem. Furthermore, open issues and future directions are underlined in order to provide an idea of challenges that need to be addressed associated with incremental and online learning models.

References

[1] M. Ghesmoune, M. Lebbah, and H. Azzag, "A new Growing Neural Gas for clustering data

streams," Neural Networks, vol. 78, pp. 36–50, Jun. 2016.

[2] G. I. Parisi, R. Kemker, J. L. Part, C. Kanan, and S. Wermter, "Continual lifelong learning with neural networks: A review," Neural Networks, vol. 113. Elsevier Ltd, pp. 54–71, 01-May-2019.

[3] G. I. Parisi, J. Tani, C. Weber, and S. Wermter, "Lifelong learning of human actions with deep neural network self-organization," 2017.

[4] S. Furao, T. Ogura, and O. Hasegawa, "An enhanced self-organizing incremental neural network for online unsupervised learning," Neural Networks, vol. 20, no. 8, pp. 893–903, Oct. 2007.

[5] J. Zhong, Z. Liu, Y. Zeng, L. Cui, and Z. Ji, "A Survey on Incremental Learning," no. Cape, pp. 166–174, 2017.

[6] S. Ren, Y. Lian, and X. Zou, "Incremental Naïve Bayesian Learning Algorithm based on Classification Contribution Degree," 2014.

[7] S. Rüping, "Incremental learning with support vector machines," in Proceedings - IEEE International Conference on Data Mining, ICDM, 2001, pp. 641–642.

[8] M. M. Neyshabouri and S. S. Kozat, "Sequential Outlier Detection based on Incremental Decision Trees."

[9] A. Wang, G. Wan, Z. Cheng, and S. Li, AN INCREMENTAL EXTREMELY RANDOM FOREST CLASSIFIER FOR ONLINE LEARNING AND TRACKING. .

[10] S. Ozawa, S. L. Toh, S. Abe, S. Pang, and N. Kasabov, "Incremental learning of feature space and classifier for face recognition," Neural Networks, vol. 18, no. 5–6, pp. 575–584, Jul. 2005.

[11] "(PDF) AN INCREMENTAL LEARNING SYSTEM FOR ON LINE KNN CLASSIFICATION: Application To Network Intrusion Detection." [Online]. Available: https://www.researchgate.net/publication/326609929_AN_INCREMENTAL_LEARNING_SYSTEM_FOR_ON_LINE_KNN_CLASSIFICATION_Application_To_Network_Intrusion_Detection.

[12] M. Muhlbaier, A. Topalis, and R. Polikar, "Learn++.MT: A new approach to incremental learning," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 3077, pp. 52–61, 2004.

[13] M. D. Muhlbaier, A. Topalis, and R. Polikar, "Learn++ .NC: Combining ensemble of classifiers with dynamically weighted consult-and-vote for efficient incremental learning of new classes," IEEE Trans. Neural Networks, vol. 20, no. 1, pp. 152–168, 200

[14] "Learning Deep Neural Networks Incrementally | by Arthur Douillard | Heuritech | Medium." [Online]. Available: https://medium.com/heuritech/learning-deep-neural-networks-incrementally-3e005e4fb4bc.

[15] M. Masana, X. Liu, B. Twardowski, M. Menta, A. D. Bagdanov, and J. Van De Weijer, "Class-incremental learning: survey and performance evaluation on image classification."

[16] "KNN Model Based Incremental Learning Algorithm." [Online]. Available: http://manu46.magtech.com.cn/Jweb_prai/EN/abstract/abstract9301.shtml.

[17] A. P. Wang, G. W. Wan, Z. Q. Cheng, and S. K. Li, "Incremental learning extremely random forest classifier for online learning," Ruan Jian Xue Bao/Journal Softw., vol. 22, no. 9, pp. 2059–2074, Sep. 2011.

[18] Y. Liu, B. Schiele, and Q. Sun, "Adaptive Aggregation Networks for Class-Incremental Learning."

[19] U. Aggarwal, A. Popescu, E. Belouadah, C. Hudelot, and M. Tools, "Multimedia Tools and Applications 1182: DEEP PROCESSING OF MULTIMEDIA DATA A comparative study of calibration methods for imbalanced class incremental learning."

[20] L. Yu, X. Liu, and J. Van De Weijer, "Self-Training for Class-Incremental Semantic Segmentation."

[21] C. Chen, W. Min, X. Li, and S. Jiang, "Hybrid incremental learning of new data and new classes for hand-held object recognition," J. Vis. Commun. Image Represent., vol. 58, pp. 138–148, Jan. 2019.

[22] A. Gepperth and B. Hammer, "Incremental learning algorithms and applications," 2016.

[23] S. Thudumu, P. Branch, J. Jin, and J. (Jack) Singh, "A comprehensive survey of anomaly detection techniques for high dimensional big data," J. Big Data, vol. 7, no. 1, p. 42, Dec. 2020.

[24] S. Sadik and L. Gruenwald, "Research issues in outlier detection for data streams," ACM SIGKDD Explor. Newsl., vol. 15, no. 1, pp. 33–40, Mar. 2014.

[25] J. Goldstein and U. Shaft, "When Is 'Nearest Neighbor' Meaningful? Trill: Streaming Query Processing View project Goldrush View project," 1997.

[26] J. Ahn, J. S. Marron, K. M. Muller, and Y.-Y. Chi, "The high-dimension, low-sample-size geometric representation holds under mild conditions," Biometrika, vol. 94, no. 3, pp. 760–766, 2007.

[27] N. Jiang and L. Gruenwald, "Research issues in data stream association rule mining," SIGMOD Rec., vol. 35, no. 1, pp. 14–19, Mar. 2006.

[28] E. Lozano and E. Acuña, "Parallel algorithms for distance-based and density-based outliers," in Proceedings - IEEE International Conference on Data Mining, ICDM, 2005, pp. 729–732.

[29] F. Angiulli and F. Fassetti, "Detecting distance-based outliers in streams of data," in International Conference on Information and Knowledge Management, Proceedings, 2007, pp. 811–820.

[30] M. Kontaki, A. Gounaris, A. N. Papadopoulos, K. Tsichlas, and Y. Manolopoulos, "Continuous Monitoring of Distance-Based Outliers over Data Streams," 2011.

[31] J. A. Silva, E. R. Faria, R. C. Barros, E. R. Hruschka, A. C. P L F De Carvalho, and A. P. Gama, "Data Stream Clustering: A Survey."

[32] N. Tomašev, R. Radovanovi´c, D. M. Mladeni´c, and M. I. Ivanovi´c, “The Role of Hubness in Clustering High-Dimensional Data,” IEEE Trans. Knowl. Data Eng., 2013.

[33] M. Radovanovi, A. Nanopoulos, and M. Ivanovi, “Reverse Nearest Neighbors in Unsupervised Distance-Based Outlier Detection,” 2014.

[34] F. Angiulli, S. Basta, S. Lodi, and C. Sartori, “GPU Strategies for Distance-Based Outlier Detection,” IEEE Trans. Parallel Distrib. Syst., vol. 27, no. 11, pp. 3256–3268, Nov. 2016.

[35] M. Bai, X. Wang, J. Xin, and G. Wang, “An efficient algorithm for distributed density-based outlier detection on big data,” Neurocomputing, vol. 181, pp. 19–28, Mar. 2016.

[36] C. Oreilly, A. Gluhak, and M. A. Imran, “Distributed Anomaly Detection Using Minimum Volume Elliptical Principal Component Analysis,” IEEE Trans. Knowl. Data Eng., vol. 28, no. 9, pp. 2320–2333, 2016.

[37] L. Zhang, J. Lin, and R. Karim, “Sliding Window-Based Fault Detection From High-Dimensional Data Streams,” IEEE Trans. Syst. Man, Cybern. Syst., vol. 47, no. 2, pp. 289–303, Feb. 2017.

[38] C. Wiwatcharakoses and D. Berrar, “SOINN+, a Self-Organizing Incremental Neural Network for Unsupervised Learning from Noisy Data Streams,” Expert Syst. Appl., vol. 143, p. 113069, 2020.

[39] M. Mermillod, A. Bugaiska, and P. Bonin, “The stability-plasticity dilemma: investigating the continuum from catastrophic forgetting to age-limited learning effects,” Front. Psychol., vol. 4, 2013.

[40] G. Cheng, Z. Song, J. Yang, and R. Gao, “On growing self - Organizing neural networks without fixed dimensionality,” in CIMCA 2006: International Conference on Computational Intelligence for Modelling, Control and Automation, Jointly with IAWTIC 2006: International Conference on Intelligent Agents Web Technologies ..., 2006.

[41] “Towards Growing Self-Organizing Neural Networks with Fixed Dimensionality.” [Online]. Available: https://publications.waset.org/7656/towards-growing-self-organizing-neural-networks-with-fixed-dimensionality.

[42] H. Yu, J. Lu, and G. Zhang, “Online Topology Learning by a Gaussian Membership-Based Self-Organizing Incremental Neural Network,” IEEE Trans. Neural Networks Learn. Syst., vol. 31, no. 10, pp. 3947–3961, Oct. 2020.

[43] Y. Xing, X. Shi, F. Shen, K. Zhou, and J. Zhao, “A Self-Organizing Incremental Neural Network Based on Local Distribution Learning,” 2016.

[44] Z. Xiang, Z. Xiao, D. Wang, and H. M. Georges, “Incremental semi-supervised kernel construction with self-organizing incremental neural network and application in intrusion detection,” in Journal of Intelligent and Fuzzy Systems, 2016, vol. 31, no. 2, pp. 815–823.

[45] Y. Nakamura and O. Hasegawa, “Nonparametric density estimation based on self-organizing incremental neural network for large noisy data,” IEEE Trans. Neural Networks Learn. Syst., vol. 28, no. 1, pp. 8–17, Jan. 2017.

[46] G. Ditzler, M. Roveri, C. Alippi, and R. Polikar, “Learning in Nonstationary Environments: A Survey,” IEEE Comput. Intell. Mag., vol. 10, no. 4, pp. 12–25, 2015.

[47] J. Gama, P. Medas, G. Castillo, and P. Rodrigues, “Learning with drift detection,” Lect. Notes

Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 3171, pp. 286–295, 2004.

[48] A. Bifet and R. Gavaldà, “Learning from time-changing data with adaptive windowing,” in Proceedings of the 7th SIAM International Conference on Data Mining, 2007, pp. 443–448.

[49] C. Alippi, G. Boracchi, and M. Roveri, “Just-in-time classifiers for recurrent concepts,” IEEE Trans. Neural Networks Learn. Syst., vol. 24, no. 4, pp. 620–634, 2013.

[50] A. Bifet and R. Gavaldà, “Kalman filters and adaptive windows for learning in data streams,” in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2006, vol. 4265 LNAI, pp. 29–40.

[51] C. Alippi and M. Roveri, “An adaptive CUSUM-based test for signal change detection,” in Proceedings - IEEE International Symposium on Circuits and Systems, 2006, pp. 5752–5755.

[52] C. Alippi, G. Boracchi, and M. Roveri, “Change detection tests using the ICI rule,” in Proceedings of the International Joint Conference on Neural Networks, 2010.

[53] R. Elwell and R. Polikar, “Incremental learning of concept drift in nonstationary environments,” IEEE Transactions on Neural Networks, vol. 22, no. 10. pp. 1517–1531, Oct-2011.

[54] J. L. Part and O. Lemon, “Incremental online learning of objects for robots operating in real environments,” in 7th Joint IEEE International Conference on Development and Learning and on Epigenetic Robotics, ICDL-EpiRob 2017, 2018, vol. 2018-January, pp. 304–310.

[55] B. Xu, S. Chen, H. Zhang, and T. Wu, “Incremental k-NN SVM method in intrusion detection,” in Proceedings of the IEEE International Conference on Software Engineering and Service Sciences, ICSESS, 2018, vol. 2017-November, pp. 712–717.

[56] A. Somasundaram and S. Reddy, “Parallel and incremental credit card fraud detection model to handle concept drift and data imbalance,” Neural Comput. Appl., vol. 31, no. 1, pp. 3–14, Jan. 2019.

[57] D. Nallaperuma et al., “Online Incremental Machine Learning Platform for Big Data-Driven Smart Traffic Management,” IEEE Trans. Intell. Transp. Syst., vol. 20, no. 12, pp. 4679–4690, Dec. 2019.

[58] E. Belouadah and A. Popescu, “IL2M: Class Incremental Learning With Dual Memory.”

[59] M. M. Bittencourt, R. M. Silva, and T. A. Almeida, “ML-MDLText: An efficient and lightweight multilabel text classifier with incremental learning,” in Applied Soft Computing Journal, 2020, vol. 96.

[60] A. Ayub and A. R. Wagner, “Cognitively-Inspired Model for Incremental Learning Using a Few Examples.”

[61] M. A. Kaufhold, M. Bayer, and C. Reuter, “Rapid relevance classification of social media posts in disasters and emergencies: A system and evaluation featuring active, incremental and online learning,” Inf. Process. Manag., vol. 57, no. 1, Jan. 2020.

[62] Ş. Ertekin, L. Bottou, and C. L. Giles, “Nonconvex online support vector machines,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 33, no. 2, pp. 368–381, 2011.

[63] R. Elwell, R. Polikar, and S. Member, “Incremental Learning of Concept Drift in Nonstationary Environments,” IEEE Trans. NEURAL NETWORKS, vol. 22, no. 10, 2011.

[64] G. Ditzler and R. Polikar, “Incremental learning of concept drift from streaming imbalanced data,” IEEE Trans. Knowl. Data Eng., vol. 25, no. 10, pp. 2283–2301, 2013.

[65] J. Zhao, Z. Wang, and D. S. Park, “Online sequential extreme learning machine with forgetting mechanism,” Neurocomputing, vol. 87, no. 87, pp. 79–89, Jun. 2012.

[66] Y. Ye, S. Squartini, and F. Piazza, “Online sequential extreme learning machine in nonstationary environments,” Neurocomputing, vol. 116, pp. 94–101, Sep. 2013.

[67] V. Losing, B. Hammer, and H. Wersing, “Incremental on-line learning: A review and comparison of state of the art algorithms,” Neurocomputing, vol. 275, pp. 1261–1274, Jan. 2018.

[68] J. Gao, S. Chai, B. Zhang, and Y. Xia, “Research on Network Intrusion Detection Based on Incremental Extreme Learning Machine and Adaptive Principal Component Analysis,” Energies, vol. 12, no. 7, p. 1223, Mar. 2019.

[69] H. He, S. Chen, K. Li, and X. Xu, “Incremental learning from stream data,” IEEE Trans. Neural Networks, vol. 22, no. 12 PART 1, pp. 1901–1914, Dec. 2011.

[70] J. Dromard et al., “Online and Scalable Unsupervised Network Anomaly Detection Method To cite this version : HAL Id : hal-01406273 Online and Scalable Unsupervised Network Anomaly Detection Method,” 2017.

[71] E. Bigdeli, M. Mohammadi, B. Raahemi, and S. Matwin, “Incremental anomaly detection using two-layer cluster-based structure,” Inf. Sci. (Ny)., vol. 429, pp. 315–331, Mar. 2018.

[72] J. Wang, Z. Mo, H. Zhang, and Q. Miao, “Ensemble diagnosis method based on transfer learning and incremental learning towards mechanical big data,” Meas. J. Int. Meas. Confed., vol. 155, Apr. 2020.

[73] D. Brzezinski and J. Stefanowski, “Reacting to different types of concept drift: The accuracy updated ensemble algorithm,” IEEE Trans. Neural Networks Learn. Syst., vol. 25, no. 1, pp. 81–94, Jan. 2014.

[74] H. H. W. J. Bosman, G. Iacca, A. Tejada, H. J. Wörtche, and A. Liotta, “Ensembles of incremental learners to detect anomalies in ad hoc sensor networks,” Ad Hoc Networks, vol. 35, pp. 14–36, Dec. 2015.

[75] F. Noorbehbahani, A. Fanian, R. Mousavi, and H. Hasannejad, “An incremental intrusion detection system using a new semi-supervised stream classification method,” Int. J. Commun. Syst., vol. 30, no. 4, p. e3002, Mar. 2017.

[76] E. Viegas, A. Santin, A. Bessani, and N. Neves, “BigFlow: Real-time and reliable anomaly-based intrusion detection for high-speed networks,” Futur. Gener. Comput. Syst., vol. 93, pp. 473–485, Apr. 2019.

[77] D. Liu, M. Cong, Y. Du, and X. Han, “Robotic cognitive behavior control based on biology-inspired Episodic memory,” in Proceedings - IEEE International Conference on Robotics and Automation, 2015, vol. 2015-June, no. June, pp. 5054–5060.

[78] V. Losing, B. Hammer, and H. Wersing, “Interactive online learning for obstacle classification on a mobile robot,” in Proceedings of the International Joint Conference on Neural Networks, 2015, vol. 2015-September.

[79] J. Gao, F. Liang, W. Fan, Y. Sun, and J. Han, “A graph-based consensus maximization

approach for combining multiple supervised and unsupervised models," IEEE Trans. Knowl. Data Eng., vol. 25, no. 1, pp. 15–28, 2013.

[80] "Graph-based consensus maximization among multiple supervised and unsupervised models — University of Illinois Urbana-Champaign." [Online]. Available: https://experts.illinois.edu/en/publications/graph-based-consensus-maximization-among-multiple-supervised-and-.

[81] S. Xie, J. Gao, W. Fan, D. Turaga, and P. S. Yu, "Class-Distribution Regularized Consensus Maximization for Alleviating Overfitting in Model Combination."