

Information Technology and Quantitative Management (ITQM2017)

Concept drift in Streaming Data Classification: Algorithms, Platforms and Issues

Janardan, Shikha Mehta*

Jaypee Institute of Information Technology, Noida Sector-62, India

Abstract

In this digital era we are surrounded by social media applications and the hardware devices (such as sensorsets) which are pouring data at an astonishing rate. This incoming data from heterogeneous sources is referred as data stream. Analysing data in motion (data streams) has become new challenge in order to meet the demands of real time analytics. Conventional mining techniques are proving inefficient since the behaviour of data itself has changed. Other challenges associated with data streams include resources constraints like memory and running time along with single scan of the data. Due to the time variant nature of data streams, applying any mining algorithm such as classification, clustering, indexing in a single scan of data is a tedious task. This paper focuses on concept drift problem in classification of streaming data. During classification a change in the concept or distribution of dataset over the time is termed as concept drift. The performance of a model/classifier degrades due to concept drift even in stationary data; dealing with this problem hence become more challenging in data streams. This paper presents categorization of existing streaming data classification algorithms along with their ability to solve concept drift problem. It also presents comparison of various tools available for simulating such problems along with their limitations. The paper also lists the various datasets and performance metrics that have been used in literature for performance analysis. Thus, this paper may serve as a complete roadmap for the researchers interested in designing new solutions for solving concept drift problem in streaming data classification. It also highlights the open research questions in this field.

© 2017 The Authors. Published by Elsevier B.V.

Peer-review under responsibility of the scientific committee of the 5th International Conference on Information Technology and Quantitative Management, ITQM 2017.

Keywords : Classification, data streams, concept drift, Apache Spark, Apache Storm, MOA, SAMOA

1. Introduction

The amount of data that has been created within the past 2 years is larger than the entire previous history of the human race. Data is growing quicker than ever before and according to IDC amount of data created

*Corresponding author.

Email address: mehtshikha@gmail.com

annually is expected to reach 180 zettabytes in 2025[1]. Data never sleeps; immense amount of data is being generated from diverse applications across the internet in every 60 seconds [2]. Every minute Google receives over 2,000,00 search queries and email users send 204166,667 messages. It is noticed that a massive growth in video data, where every minute up to 300 hours of videos are uploaded to YouTube alone [3]. These facts indicate the speed with which data is growing thereby forming the digital universe. Speed or velocity is one of the main characteristic of big data. Technological advancement in hardware is also responsible for rising popularity of data streams [4]. Many simple operations of everyday life, such as using a credit/debit card or smart phone, often lead to automated creation of data. Since these operations often scale over large numbers of participants, they lead to massive data streams. Similarly, telecommunications and social networks often contain large amounts of images, audio, video and text data streams. Mining data streams has now become major challenge for the research community as the data mining algorithms developed so far are more suitable for stationary data. Since the data itself is changing or evolving over time and is available only for very short duration, learning from data streams is another big challenge. The temporal properties (real time data generation) of data streams differentiate it from non-streaming data and hence derive the need for novel classification, clustering, pattern mining techniques and performance evaluation parameters. This paper focus on the problem of classification in data streams. Classical classification algorithms are not compatible with streaming data due to resource constraints (processing time, memory) and single scan of the data (one look, no random access) [47]. Therefore, to create new learning approaches like incremental or ensemble learning to classify the streaming data is another research issue. Classification of streaming data is not as simple as the classification on stationary data because the distribution of the dataset may change over the time during the training of model. This phenomenon is referred to as concept drift. All streaming models need to account for concept drift in the model construction process. Therefore, the construction of a training model in streaming and evolving scenarios is often very critical.

Research in this direction has started gaining more attention in the recent past due to the mushrooming size of data. Gaber et al [5] presented an overview of general stream mining algorithms. A survey by Gama and Rodrigues [6] focused on illustrating data stream mining with specific algorithms and applications. In recent survey papers, Hai-Long Nguyen et al. [7] and Bakshi Rohit Prasad et al. [8] discussed supervised and unsupervised learning algorithms for streaming data along with some available platforms to handle streaming data. However, these researchers mainly focus on classification and clustering algorithms in general and do not cover the specific problem with classification algorithms i.e. concept drift. This paper presents simplified categorization of the different types of algorithms, platforms and performance metrics for classification in streaming data, which are commonly used in the literature along with novel algorithms for classification which deal with concept drift problem. Therefore, this paper provides a road map for the researchers who intend to develop new algorithms for solving concept drift problem in classification. This work is also intended to aid the mining community who wish to utilize existing concept drift adapted stream mining algorithms for classifications in diverse domains.

Rest of the paper is organized as follows. Section 2 introduces the concept drift problem in classification and major issues of classification in drifted data. In section 3, categorization of classification algorithms which deals with concept drift is presented. Section 4 provides the details of available tools and dataset for the classification of streaming data. Section 5 discusses performance evaluators for classification under concept drift followed by future challenges and conclusion.

2. Concept drift in streaming data classification

In streaming data classification there is an infinite sequence of the form (x, y) where $y = \{y_1, y_2, \dots, y_k\}$, and the goal is to find a function $y = f(x)$ that can predict the y value for an unseen instance x , where the size of dataset is not known due to evolving nature of streaming data. There are some other constraints with streaming

data during mining tasks.

- Resource constraints (processing time, memory).
- Single scan of the data (one look, no random access).

With the above constraint, the main objective is to handle the problem of concept drift. Concept drift in data stream has a high level of importance especially in the case of classification problems. In this context, the word “concept” means the target class [9] and concept drift refers to changes in the underlying distribution of the target classes. In the real world applications, the underlying distribution of data are not stationary thus previously valid models are proving inefficient by the passage of time. There are two types of drift [48]:

- (1) Real Drift: the posterior probability $p_t(x|y)$ varies over time, independently from variations in the evidence $p_t(x)$.
- (2) Virtual Drift: the evidence or the marginal distribution of the data, $p_t(x)$, changes without affecting the posterior probability of classes $p_t(y|x)$.

Where $p_t(x|y)$ is the posterior probability of distribution of the classes, $p_t(x)$ is the class probability and $p_t(y|x)$ is posterior probability of distribution of the class membership. Concept drift can be handled using adaptive learning. Adaptive learning can be implemented using incremental and ensemble learning. Incremental learning follows a machine learning paradigm where the learning process takes place whenever new examples emerge, and then adjusts to what has been learned from the new examples. On the other hand, ensemble learning employs multiple base learners and combines their predictions [10]. The most popular evolving technique for handling concept drift in data streams is to use an ensemble classifier (a combination of classifiers). The outputs of several classifiers are combined to determine a final classification, which is often called fusion rules. Also the weights are assigned to the individual classifier's outputs at each point in time. Apart from incremental and evolving learning techniques, trigger based learning techniques [49] are also used in literature to handle the concept drift. Concept drift problem in classification could be handled through above discussed techniques in the literature. Ensemble learning approach is observed to better for dealing with concept drift problem efficiently.

3. Algorithms for classification of streaming data

As we know data stream is generated at very high speed and is infinite in size. It is impractical to store and process stream for training the systems. There are number of algorithm for classification of stationary datasets but these algorithms are not suitable for streaming data in terms of compatibility and performance. Therefore these algorithms need to be adapted to accommodate resource constraints and issue (like one pass, in memory, less time to access data, issue of concept drift) due to evolving nature of data streams. Table 1 categorises the streaming data classification algorithms [8] in different classes of classification algorithms such as tree based, ensemble based, neighbour, statistical and rule based. This table also highlights whether an algorithm handles concept drift problem or not.

Table 1. Categories of streaming data classification algorithms

SNO	Streaming Data Classification Algorithms	Year	Key Points	Classification Category
1	ITI [11]	1997	Require large storage hence, not suitable for large data stream.	Tree based
2	VFDT [12]	2000	Require lesser memory and does prediction in real time. It uses Hoeffding bound to assess the number of minimum instances require to grow the decision tree.	Tree based
3	CVFDT* [13]	2001	Advancement of VFDT with adaption for Concept drift.	Tree based

4	Streaming Algorithms* [14]	Ensemble	2001	It can handle the Concept drift but not good with high speed data streams.	Ensemble based
5	OLIN* [15]		2002	Require lesser memory and uses the info-fuzzy network for concept drift adaption.	Tree based
6	Weighted Ensemble* [16]	Classifier	2003	Deals well with concept drift by using ensemble of weighted classifier on chunk of data instances from data streams rather revising the model frequently.	Ensemble based
7	On demand Classifier* [17]		2004	Based on micro clustering, dynamically adapts and/or selects sliding window size for better performance and concept drift adaption.	Rule based
8	UFFT* [18]		2004	Use limited memory and generates a forest having binary tree for each pair of classes.	Tree based
9	Adaptive Nearest Neighbor classification Algorithm* [19]		2005	An incremental algorithm for adaptive learning with low cost for updating in model.	Nearest neighbour
10	Evolving Naïve Bayes* [20]		2006	An extended naïve bayes algorithm capable of learning from evolving data stream.	Statistical
11	Any Time Nearest Neighbor Algorithm* [21]		2006	A variation of Nearest neighbor algorithm and its capable of any time classification	Nearest neighbour
12	IOLIN* [22]		2008	Variation of OLIN that keeps on model updating until sufficient concept drift thereby saves computational effort significantly.	Tree based
13	ADWIN Bagging* [23]		2009	Employs ADWIN algorithm to detect changes as well as for estimating the weight for boosting method.	Tree based
14	ASHT Bagging* [23]		2009	Uses varying sized Hoeffding tree between small size tree is quickly adapts to changes.	Tree based
15	Random Forest based classifier algorithm* [24]		2011	Handles evolving data streams even with intermittent labeled data instances arrivals in one pass. Uses Entropy to detect concept drift.	Tree based
16	Vertical Hoeffding tree (VHT)* [25]		2013	A variation of VFDT that performs distributed parallel computation by vertically partitioning data sets(attribute wise)	Tree based
17	Fuzzy Passive-aggressive classification [26]		2013	It presents a novel strategy for online membership generation, which is particularly suitable for coping with unavailable outlier in online classification problems. This algorithm increase the accuracy of classification.	Rule based
18	Similarity based data stream classifier(SimC)* [27]		2014	Uses new insertion/removal approaches for quickly capturing and representing changes in data to improve performances.	Rule based
19	Prequential AUC based classifier* [28]		2014	It works better with highly imbalanced data streams.	Rule based
20	Online Stream classifier with incremental semi-supervised learning* [29]		2015	Utilizes the selective self-training based semi-supervised learning approach to achieve at the par classification accuracy even with availability of only 1% labeled data.	Rule based
21	Distance-based Ensemble online classifier with kernel clustering* [30]		2015	Use kernel-based clustering approach where a new instance is supplied for each of the iteration and prediction is made on it. An ensemble of classifier is constructed on the basis of portfolio of distance measure.	Ensemble based
22	One-class classifier with incremental learning and forgetting* [31]		2015	It used a modified weighted one-class SVM augmented with the principles of incremental learning and forgetting. These techniques allow adapting the decision boundary of the classifier to changes in the incoming data.	Rule based
23	Classifying re-curing concept using fuzzy similarity function* [32]		2015	It goal to show the feasibility of the solution in terms of reducing the number of records trained without losing precision of classification.	Rule based

Figure 1 depicts that the majority of the streaming data classification algorithms are either rule based or tree based. Very few algorithms are developed using ensemble, nearest neighbor and statistical approaches. These results depict that still there is lot of scope for research in this area as the performance of probability based and machine learning based classification approaches over streaming data is still an open research question.

*These algorithms are capable of handling concept drift.

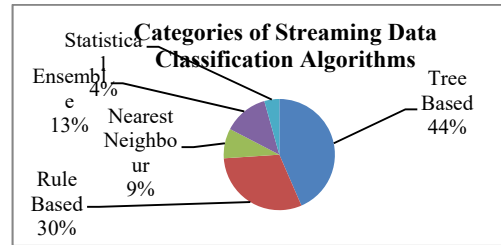


Fig. 1. Statistics of streaming data classification algorithms

Studies may be done on these categories of algorithms. For the researchers, it's very important to know about tools, datasets and performance evaluation measures in order to evaluate the efficacy of developed algorithms. Next section highlights the important features of the tools available for data streams mining research. Section 4 also presents the list of benchmark datasets suitable in this area as they have been widely used in literature.

4. Tools and datasets for data streams

Table 2 provides brief comparison of tools on the basis of different parameters like implemented in which language, supported programming API and processing model which are available online for data streams mining.

Table 2. Comparison of tools

	Apache Spark[33]	Apache Storm[34]	MOA[35]	SAMOA [36]
Definition	It is an alternative to Hadoop which is designed to overcome the disk I/O limitations and improve the performance of earlier systems	It is used for processing data in real-time and was initially conceived to overcome deficiencies of other processors in collecting and analyzing social media streams	A free open-source software specific for mining data streams with concept drift	It is a platform for machine learning from streaming data, was originally developed at Yahoo! Labs. Extension of MOA
Implemented In	Scala	Clojure	Java	Java
Programming API	Scala, Java, Python	Java API and usable with any programming language	Java API and works with WEKA	Java API and works with Storm, S4 and Samza
Development	A full top level Apache project	Undergoing Apache project	A free open source project	A free open source project but in development phase
Application	Good in data analytics pipeline for anomaly detection	Good for cleaning, normalization and resolve large amount of non-unique data points with low latency and high throughput	Good for creation and usage of benchmark setting for comparable and repeatable evaluation of mining algorithms	Users can develop distributed streaming ML algorithms once and execute them on multiple distributed stream processing engine like Spark, Storm etc.
Scaling	Yes (Horizontal and vertical)	Yes (Horizontal)	Yes	Yes (Horizontal)
Programming paradigm	Parallel computing	Parallel computing	Serial computing	Parallel computing

Processing Model	Batch processing and distributed	Stream processing framework that processes and dispatches messages as they arrive	Batch processing	Batch and distributed
Supported Algorithms	All machine learning algorithms	All machine learning algorithms	All machine learning algorithms	VHT, CluStream
Fault Tolerance	Yes	Yes	Yes	Yes

Data Sets: Table 3 shows some available datasets which are used for the different classification algorithms on streaming data in literature. These datasets are further categorized in real and artificial.

Table 3. Available dataset for streaming data classification

Sno.	Dataset Name	Description	Type
1	Airlines dataset [38]	This real dataset was first used for classification purposes in [37], and contains 539,384 records. It represents whether a flight was delayed or not from some information about it, i.e. the airline, the airports involved, or the day of week	Real Dataset
2	Electricity dataset [39]	The Electricity Market Dataset [40] is a real dataset that uses data collected from the Australian New South Wales Electricity Market, where the electricity prices are not stationary and are affected by the market supply and demand	Real Dataset
3	Poker [38]	Poker-Hand dataset is a real set of 829,201 instances composed by 11 attributes. Each record of the dataset is an example of a hand consisting of five playing cards drawn from a standard deck of 52.	Real Dataset
4	Sensor dataset [41]	Sensor stream is a real dataset that contains information (temperature, humidity, light, and sensor voltage) collected from 54 sensors deployed in Intel Berkeley Research Lab.	Real Dataset
5	SEA dataset [39]	This synthetic dataset is made up of 5M of records as a result of the merge process of five different SEA [42] datasets, each of them containing 1M of records.	Artificial Dataset
6	Hyperplane dataset [43]	A different synthetic dataset with gradual drifting concepts was created based on a moving hyperplane.	Artificial Dataset

5. Performance Metrics

One of the major challenges of data streams mining task is how to evaluate the performance of mining task since traditional performance metrics are not sufficient in streaming data mining scenario. The various performance metrics used for streaming data classification are as follows:

Table 4. Performance metrics for classification under concept drift

S.no	Evaluation Parameter	Major Purpose	Value Significance
1	Kappa measure [44]	Measure for benchmarking classification accuracy under class imbalance and is used in static as well as streaming data classification	Higher value means better performance
2	Temporal Statistics [44]	Measures the effectiveness of classification in temporal dependence in data streaming	Negative value means worse performance
3	Loss	Measuring how appropriate is the current model to the actual status of the nature	Low
4	Memory used	Learning algorithms run in fixed memory. We need to evaluate the memory usage over time, and the impact in accuracy when using the available memory	Minimum
5	Speed of Processing examples	Algorithms must process the examples as fast if not faster than they arrive	Higher value means better performance
6	Confusion Matrix	Performance of single and ensemble classifiers can be calculated by confusion matrix by using the testing data set. It measures misclassification rate.	Low
7	Prequential [45]	It measures accuracy with forgetting as a means of evaluating data stream classifiers and enhancing drift detection methods	Higher value means better performance

In data stream mining, predictive abilities of a classifier are evaluated by using a holdout test set, chunks of examples, or incrementally after each example [46]. More recently, Gama et al. [45] proposed prequential accuracy with forgetting as a means of evaluating data stream classifiers and enhancing drift detection methods. They have shown that computing accuracy only over the most recent examples, instead of the entire stream, is more appropriate for continuous assessment and drift detection in evolving data streams. Nevertheless, prequential accuracy inherits the weaknesses of traditional accuracy, i.e., variance with respect to class distribution and promoting majority class predictions.

6. Conclusion and Future scope

Classification algorithms that work well with stationary datasets are not compatible with the streaming data due to one pass, in memory and less time to access data. This paper elaborates various classification algorithms and platforms that are useful when the datasets are evolving and have problem of concept drift. In this paper we have also identified tools that are currently being used in field of big data analytics. Benchmark datasets used by streaming data researchers were also identified. This paper covers the comparative analysis of available data streaming frameworks. Many of these tools are very young, and more research is needed to properly benchmark and evaluate all of the different options. Only VHT, CluStream, Adaptive model rule are successfully implemented in SAMAO whereas distributed Naïve Bayes, Stochastic Gradient Descent, Adaptive VHT is still pending and need contribution from researchers.

References

- [1]. <http://techblog.comsoc.org/2016/03/09/idc-directions-2016-iot-internet-of-things-outlook-vs-current-market-assessment> [accessed july-2017].
- [2]. <https://www.domo.com/blog/how-much-data-is-created-every-minute/> [accessed july-2017].
- [3]. <https://www.forbes.com/sites/bernardmarr/2015/09/30/big-data-20-mind-boggling-facts-everyone-must-read/#4e98387f17b1> [Accessed april-2017].
- [4]. C. Aggarwal. *Data Streams: Models and Algorithms*, Springer, 2007.
- [5]. Gaber MM, Zaslavsky A, Krishnaswamy S (2005). Mining data streams: a review, *ACM Sigmod Rec* 34(2):18–26.
- [6]. Gama J, Rodrigues P (2009). An overview on mining data streams, vol 206 of *Studies in computational intelligence*, pp 29–45. Springer, Berlin.
- [7]. Nguyen, Hai-Long, Yew-Kwong Woon, and Wee-Keong Ng. A survey on data stream clustering and classification, *Knowledge and information systems* 45.3 (2015): 535–569.
- [8]. Prasad, Bakshi Rohit, and Sonali Agarwal. Stream Data Mining: Platforms, Algorithms, Performance Evaluators and Research Trends, *International Journal of Database Theory and Application* 9.9 (2016): 201–218.
- [9]. M.J. Hosseini, Z. Ahmadi and H. Beigy. Using a classifier pool in accuracy based tracking of recurring concepts in data stream classification, *Evolving Systems* 4(1) (2013), 43–60.
- [10]. Zang, Wenyu, et al. Comparative study between incremental and ensemble learning on data streams: Case study, *Journal Of Big Data* 1.1 (2014): 1–16.
- [11]. Utgoff, Paul E., Neil C. Berkman, and Jeffery A. Clouse. Decision tree induction based on efficient tree restructuring, *Machine Learning* 29.1 (1997): 5–44.
- [12]. Domingos, Pedro, and Geoff Hulten. Mining high-speed data streams, *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2000.
- [13]. Gama, João, Ricardo Fernandes, and Ricardo Rocha. Decision trees for mining data streams, *Intelligent Data Analysis* 10.1 (2006): 23–45.
- [14]. Street, W. Nick, and YongSeog Kim. A streaming ensemble algorithm (SEA) for large-scale classification, *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2001.
- [15]. Cohen, Lior, Gil Avrahami, and Mark Last. Incremental info-fuzzy algorithm for real time data mining of non-stationary data streams, *TDM Workshop*, Brighton UK. Vol. 43. 2004.
- [16]. Wang, Haixun, et al. Mining concept-drifting data streams using ensemble classifiers, *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2003.

- [17]. Aggarwal, Charu C., et al. On demand classification of data streams, *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2004.
- [18]. Gama, João, Pedro Medas, and Ricardo Rocha. Forest trees for on-line data, *Proceedings of the 2004 ACM symposium on Applied computing*. ACM, 2004.
- [19]. Law, Yan-Nei, and Carlo Zaniolo. An adaptive nearest neighbor classification algorithm for data streams, *European Conference on Principles of Data Mining and Knowledge Discovery*. Springer Berlin Heidelberg, 2005.
- [20]. Klawonn, Frank, and Plamen Angelov. Evolving extended naive Bayes classifiers, *Data Mining Workshops, 2006. ICDM Workshops 2006. Sixth IEEE International Conference on*. IEEE, 2006.
- [21]. Ueno, Ken, et al. Anytime classification using the nearest neighbor algorithm with applications to stream mining, *Data Mining, ICDM'06. Sixth International Conference on*. IEEE, 2006.
- [22]. Cohen, Lior, et al. Real-time data mining of non-stationary data streams from sensor networks, *Information Fusion* 9.3 (2008): 344-353.
- [23]. Bifet, Albert, et al. New ensemble methods for evolving data streams, *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2009.
- [24]. Abdulsalam, Hanady, David B. Skillicorn, and Patrick Martin. Classification using streaming random forests, *IEEE Transactions on Knowledge and Data Engineering* 23.1 (2011): 22-36.
- [25]. Prasad, Bakshi Rohit, and Sonali Agarwal. Critical parameter analysis of Vertical Hoeffding Tree for optimized performance using SAMOA, *International Journal of Machine Learning and Cybernetics* (2016): 1-14.
- [26]. Wang, Lei, Hong-Bing Ji, and Yu Jin. Fuzzy Passive-Aggressive classification: A robust and efficient algorithm for online classification problems, *Information Sciences* 220 (2013): 46-63.
- [27]. Mena-Torres, Dayrelis, and Jesús S. Aguilar-Ruiz. A similarity-based approach for data stream classification, *Expert Systems with Applications* 41.9 (2014): 4224-4234.
- [28]. Brzezinski, Dariusz, and Jerzy Stefanowski. Prequential AUC for classifier evaluation and drift detection in evolving data streams, *International Workshop on New Frontiers in Mining Complex Patterns*. Springer International Publishing, 2014.
- [29]. Loo, HuiRu, and Muhammad N. Marsono. Online data stream classification with incremental semi-supervised learning, *Proceedings of the Second ACM IKDD Conference on Data Sciences*. ACM, 2015.
- [30]. Jędrzejowicz, Joanna, and Piotr Jędrzejowicz. Distance-based ensemble online classifier with kernel clustering, *Intelligent Decision Technologies*. Springer International Publishing, 2015. 279-289.
- [31]. Krawczyk, Bartosz, and Michał Woźniak. One-class classifiers with incremental learning and forgetting for data streams with concept drift, *Soft Computing* 19.12 (2015): 3387-3400.
- [32]. Ángel, Abad Miguel, Gomes João Bartolo, and Menasalvas Ernestina. Predicting recurring concepts on data-streams by means of a meta-model and a fuzzy similarity function, *Expert Systems with Applications* 46 (2016): 87-105.
- [33]. <http://spark.apache.org/>
- [34]. Apache Storm. <https://storm.apache.org/>
- [35]. <http://moa.cms.waikato.ac.nz/>
- [36]. Morales GDF, Bifet A. SAMOA: Scalable Advanced Massive Online Analysis, *J Mach Learn Res*. 2015;16:149–53.
- [37]. Žliobaitė, Indrė, et al. Active learning with evolving streaming data, *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer Berlin Heidelberg, 2011.
- [38]. <https://sourceforge.net/projects/moa-datastream/files/Datasets/Classification/>
- [39]. http://www.inescporto.pt/~jgama/ales/ales_5.html
- [40]. Harries, Michael, and New South Wales. Splice-2 comparative evaluation: Electricity pricing, (1999).
- [41]. <http://www.cse.fau.edu/~xqzhu/stream>
- [42]. Street, W. Nick, and YongSeog Kim. A streaming ensemble algorithm (SEA) for large-scale classification, *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2001.
- [43]. Hulten, Geoff, Laurie Spencer, and Pedro Domingos. Mining time-changing data streams, *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2001.
- [44]. Bifet, Albert, et al. Pitfalls in benchmarking data stream classification and how to avoid them, *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, Berlin, Heidelberg, 2013.
- [45]. Gama, João, Raquel Sebastião, and Pedro Pereira Rodrigues. On evaluating stream learning algorithms, *Machine learning* 90.3 (2013): 317-346.
- [46]. Gama, J. *Knowledge Discovery from Data Streams*, Chapman and Hall (2010).
- [47]. Nguyen, Hai-Long, Yew-Kwong Woon, and Wee-Keong Ng. A survey on data stream clustering and classification, *Knowledge and information systems* 45.3 (2015): 535-569.
- [48]. Gama, João, et al. A survey on concept drift adaptation, *ACM Computing Surveys (CSUR)* 46.4 (2014): 44.
- [49]. Žliobaitė, Indrė. Learning under concept drift: an overview, *arXiv preprint arXiv:1010.4784* (2010).