

Concept Drift Detection in Data Stream Mining: The Review of Contemporary Literature

B. Ramakrishna¹

¹ JNTU

Received: 10 December 2016 Accepted: 3 January 2017 Published: 15 January 2017

Abstract

Mining process such as classification, clustering of progressive or dynamic data is a critical objective of the information retrieval and knowledge discovery; in particular, it is more sensitive in data stream mining models due to the possibility of significant change in the type and dimensionality of the data over a period. The influence of these changes over the mining process termed as concept drift. The concept drift that depict often in streaming data causes unbalanced performance of the mining models adapted. Hence, it is obvious to boost the mining models to predict and analyse the concept drift to achieve the performance at par best. The contemporary literature evinced significant contributions to handle the concept drift, which fall in to supervised, unsupervised learning, and statistical assessment approaches. This manuscript contributes the detailed review of the contemporary concept-drift detection models depicted in recent literature. The contribution of the manuscript includes the nomenclature of the concept drift models and their impact of imbalanced data tuples.

Index terms— concept drift, change point, data stream mining, ensemble classifiers, class imbalance, misclassification, supervised and unsupervised learning

1 Introduction

apid expansion of Information and Communication Technologies (ICT) has led to exponential growth in the volume of data generated. According to a survey from IDC, in the near future the quantum of data that is generated will run to trillions of gigabytes [1]. It is imperative that there is need for robust tools and solutions for handling the huge volume of data generated from varied range of applications. Eventually the huge volumes of data that are generated demand more effective techniques of data management.

Data mining is one of the major process in the data management. Profoundly the data mining solutions are about initially gathering data and processing them in offline mode. Predictive models are usually trained on the basis of historical data provided as pair of data (input and output). The trained models are used for prediction of output for new unseen input data.

Streaming data can't be processed simultaneously due to the quantum of data that is generated regularly. It is highly complex to accommodate the data in to machine's main memory and the online processing of data is the only right method that could be adapted. Predictive models can be trained either in an incremental manner by continuous update or by ensuring retaining of the model using batches of data.

In the constantly changing environments, the data distribution might change over course of time and it could lead to conditions of concept drift [2], [3]. Concept drift is the changes in the conditional distribution of varied output (for example, the target variable for the input features) despite of the input remaining unchanged.

A classic example of real concept drift is about the change in user level interests whilst following an online news stream. For instance, though the distribution of a news documents that are often relayed might remain the same, still the conditional distribution of interesting news documents for the user might undergo changes. The

7 3) DRIFT BY POSTERIOR PROBABILITY (A CHANGE INFLUENCED BY THE CONFLICT OF OLD AND NEW DECISION BOUNDARY):

process of adaptive learning reflects upon the predictive models online while their operation might be responding to concept drifts.

Phenomenal developments have taken place in terms of concept drift and there were many drift-aware adaptive learning algorithms were developed. The scope of the problem being very wide and spans over varied topics, not much of comprehensive survey is envisaged. Though the concept is relatively new, still there was some kind of adaptive learning algorithms that were proposed earlier.

Considering the quantum of developments that has taken place in the subject of concept drift, this paper focus on comprehensive summary of research done for gaining insights in to concepts unification and terminology and also to survey of contemporary methodologies and techniques that are investigated in the past.

2 II.

3 The Problem Description a) Misclassification

Profoundly, in the case of misclassification, the minority class is highly complex than the majority class. For instance, the spam class in the spam filters and the fraud class in a credit card application. Hence, misclassifying-class example is highly a costlier factors.

the facets of misclassification, as they presume only balanced class distribution. Training procedure with target of maximizing overall accuracy usually leads to higher levels of probability of induced classifier that predicts as the instance of majority class along with low. It is imperative to envisage majority class has higher accuracy levels while the minority class has much lower accuracy amidst ranging between 0-10% [4]. Misclassification resulting because of imbalance due to classifiers like the decision trees [5], KNN (K-Nearest Neighbour) [6], [7], [8], Neural Networks [9] and SVM [10], [11] that were reviewed. Classifiers offer a balanced degree of predictive performance over all classes that are required.

Profoundly the percentage of minority class in a data set is used in the researches to detail the level of imbalance in the data [12], list of varied illustrations in every class [13], size ratio amidst of classes [14]. Coefficient of variance is used in [15] that are less straight forward. Detailing of imbalance status might not be a crucial issue in offline learning, but it becomes more significant for online learning as there is hardly any static data over the online scenarios.

It is very essential to have some automatic evaluation for detailing the updated imbalance degree and techniques that monitor the changes over the Misclassification status. The facets of changes in the misclassification are directly coherent to concept drift.

4 b) Concept drift

Concept drift could take place when the joint probability $P(x, y)$ changes [16], [17]. Concept drift will manifest three fundamental forms of changes pertaining to three key variables in the Bayes' theorem [18].

5 1) Drift by prior probability (a change in learnt decision boundary):

The prior probability of circle class is reduced and the change can lead to misclassification. Identification of drift using the prior probability is simple and it is distance between two concepts that are estimated depending on the distance assessment methods like the Total Variation Distance and Hellinger Distance.

6 2) Drift by condition (decision boundary change influenced by condition):

True decision boundary remains unaffected. In the earlier research, authors have claimed that such types of drift are result of incomplete representation of true distribution over the current data that profoundly needs supplement data information for the learning model [19].

The subset of covariate attributes will have conditional probability distribution over varied possible values of covariate attributes for every specific class.

Conditional drift is weighted sum of distances amidst every probability distributions from varied time period, wherein the weights are average probability of class amidst the time periods in sequence.

7 3) Drift by posterior probability (a change influenced by the conflict of old and new decision boundary):

True boundary amidst classes varies post drift and the earlier learnt discrimination function does not apply the changes more. In a different dimension, it can be stated that the old function becomes completely or partially unfit and the learning models are required to adapt to new dimensions.

For every subset of covariate attributes, there is a probability distribution amidst the class labels for every combination of covariate values at every time period. Hence the posterior drift can be estimated as weighted sum of distances amidst the probability distributions wherein the weighted sum of such distances amid the

probability distributions wherein the weights are average probability over two periods of specific value to an covariate attribute subset.

Posterior distribution changes signify the fundamental changes among the data generation function, which is classified as a true concept drift. The other two types relate to virtual concept drift [20] that do not change the decision boundaries. In real-term conditions, one type of concept drift might appear with combination with other kind of concept drifts.

8 III. Review of State of the art models

In this section, the contemporary models pertaining to concept drift detection of streaming data mining from the contemporary literature. Overall models that were reviewed in varied context and it is a joint detection of concept drift and misclassification. Concept drift detection using incremental learning and concept drift detection based on statistical methods.

9 a) Joint detection of Concept Drift and Misclassification

Few of the researches have made effort to address the joint problem of concept drift and misclassification, because of rising need from practical problems [21], [22].

Uncorrelated Bagging is one of the old algorithm that is built to ensemble classifiers that are trained over a more balanced set of data based on re sampling and overcome the concept drift passively by weighing the base classifier having discriminative power [23], [24], [25].

Selectively the recursive approaches like REA [26] and SERA [27] adapt same kind of ideas for Uncorrelated Bagging of developing an ensemble weighted classifiers but with a smarter level of oversampling technique. Learn ++NIE and the Learn++CDS are some of the contemporary algorithms that tackle misclassification based on oversampling technique SMOTE [13] or sub-ensemble technique and the concept drift based on a dynamic weighting strategy [28].

HUWRS.IP [29] develops HUWRS [30] to handle the imbalanced data streams by focusing on instance propagation scheme that relies on Naïve Bayes classifier and it uses Hellinger distance as a major weighting measure for concept drift detection.

All such approached relate to chunk-based learning algorithms, and the core techniques work over a batch of data that is received at every step. It is very complex to develop a true online algorithm for concept drift due to the issues like measuring minority class statistics based on one illustration at a time [31].

In order to handle the misclassification and concept drift in the form of an online fashion, some of the methods were proposed in the recent past. DDM -OCI [32] is among one of the contemporary algorithms that were proposed for detection of concept drift actively over the imbalanced data streams online. It tracks the reduction over the minority-class recall and upon observing any kind of significant drop, the drift shall be reported. The solution was very effective when minority-class recall is impacted by concept drift but when majority class might be adversely impacted.

LFR (Linear Four Rates) is the other approach proposed for improving the DDM-OCI that monitors four rates based on confusion matrix-wherein the minority class recall and the precision, majority-class recall and precision that is statistically-supported bounds towards any kind of drift detection [33]. If any of the four rates if found exceeding the bound, the drift shall be confirmed. In PAUC (Prudential AUC) [34], [35] the emphasis is on developing an overall performance measure for online scenarios, and used as concept drift indicator. But accessibility to the historical data is imperative for the system. DDM-OCI, PAUC and LFR are very active drift detectors that are designed for imbalanced data streams and they are independent over the classification algorithms. Such significant constraint of such models is reset of the learning process if concept drift is considered. It could be infeasible in terms of handling the mis classification.

In addition to the above set of approaches, perceptr on oriented algorithms like ESOS-ELM [36], RLSACP [37] and ONN [38] focus on the classification model for non-stationery environments in a passive manner and comprise mechanisms to address misclassification. RLSACP and ONN are some of the single model approaches comprising similar set of modelling and framework.

CID (Class Imbalance Detection) approach was proposed with a varied objective towards concept drift [39]. For defining the imbalanced degree that is suitable for online learning, a real-time indicator was proposed which is based on time decayed class size, the size pertaining to every class in the data stream. It is updated incrementally at every time using the time decay factor that emphasizes current status of data and it weakens the effect of old data. Any kind of current imbalance status is reported and it provides information pertaining to which classes belong to minority and majority classes.

10 b) Concept Drift Detection by Incremental learning

Incremental learning is a new dimension in which the concept drift is identified with. Many of the models that were proposed earlier focused on incremental learning wherein the historical models were considered for forming the ensembles. Following are some of the contemporary incremental learning models SEA (Streaming Ensemble Algorithm) [40] uses simple majority voting, the DIC (Dynamic Integration of Classifiers) approach [41] combines

11 C) CONCEPT DRIFT DETECTION BY STATISTICAL MEASURES

historical models with novel model of data training using the dynamic selection (DS), DVS (Dynamic Voting with Selection) and Dynamic Voting (DV).

The other benchmark called AUE2 (Accuracy Updated Ensemble) [42] adapts weighted voting as a combination scheme, where the weights that assigned to individual models are defined in terms of mean squared errors of the models. Learn++ algorithm [43] unlike DIC and AUE2 weighs on the current performance over the Non-Stationary Environments, which assign weights to varied individual models depending on the current and the earlier data.

The model discussed in [44] reflects upon Inductive Transfer (TIX) approach which works on varied methods to gain insights to historical models like given a new chunk of data and the outputs of historical models over training data that are used as features of training data, and a new model is developed with augmented training data. In the instance of building linear models that are built on learning process, TIX can be perceived as one of the special weighted voting scheme, a linear combination of original features of training data can be perceived as outcome of a linear model based on the training data that is original.

The other ensemble model DDD (Diversity for Dealing with Drifts) method is discussed in [16]. The method focus on using the ensembles as single model for a chosen time step.

Existence of concept drift leads to various models with positive and negative effects in terms of learning the current concept. It is very important that whilst getting the positive effects, preventing the negative impacts is also very important. Preserving historical models induce overheads for both storage and computation. Such issues are not usually addressed in DIC, TIX, Learn++NSE. Despite that DS/DVS scheme of DIC and time-adjust error schemes of Learn++NSE shall be resourceful for choosing historical models for preserving, and such adaptations need not be evaluated.

SEA and AUE2 usually control the number of preserved models in the conditions of a predefined threshold. Both SEA and AUE2 assess the quality of individual models based on accuracy perspective. Major difference in the way SEA and AUE2 assess is that, SEA takes to account overall accuracy of ensembles of current training data, but AUE2 takes in to account every individual model in consideration over the training data in direct manner.

Not many of the existing methods of ensemble that are used for incremental learning has focused on ensemble diversity in an explicit manner, though the diversity is considered to play a critical role in ensembles [45], [46].

11 c) Concept Drift Detection by Statistical Measures

In [33], the related statistical change detection model was proposed to handle the imbalanced data streaming, wherein the proposed model monitor multiple performance metrics. The technique monitors true positive rate and false positive rate, the true negative rates and the false positive rates attained from the confused matrix of the classification. Unlike the traditional matrix that reflects a biased majority class, the confusion matrix depicts more detailed view that is essential for addressing imbalance class problems. DOD (Degree of Drift) which is an windowbased model identifies the drifts by computing the distance map of all the samples over the current chunk and its nearest neighbours from the earlier chunk [48]. The DOD is computed based on distance map and if it increases by a parameter, the drift is signalled.

In [49], the Paired Learners approach is proposed which adapts a pair of reactive learner which is trained based on the chunk of data. The model is a stable learner and trained over all the earlier data. The variation of accuracy amidst the two approaches depicts the drift. The disagreement is captured over binary value circular list. Also, the increase in the quantum of ones that are beyond change threshold is signalled as a drift, which is managed by replacing the stable model with a reactive one.

In [50], a contemporary model was proposed that depends on the observation of randomly chosen training and testing the samples from a chunk of data which should lead to good accuracy of prediction, unless the window have any kind of non-stationary data. The usually adapted model of classifier's cross validation evaluation [51] is the fundamental for the aforesaid model.

The OLINDDA (Online Novelty and Drift Detection Algorithm) adapts the K-means data clustering for monitoring continuously and adapt to the emerging data [52]. The short term memory queue holds the unknown samples and they are clustered periodically and merged to existing similar cluster profiles or the modern profile to the pool of clusters.

In [53], the MINAS were proposed which relies on micro clusters to obtain incremental stream clustering algorithm. It is an extended model to OLINDDA's approach used for multi class problem. Some of the similar techniques that rely on clustering for defining the boundaries of the known data are Woo Ensemble [54], ECS Miner [55], and DETECTNOD algorithm [56]. Samples falling out from the clusters are treated as suspicious [54] [56] or the other way as outliers [55]. The difference or similarity amidst the defined clusters and suspicious samples are estimated on the account of density that is observed. If similarity attains the suspicious or outlier samples that are incorporated towards corresponding clusters, it concludes the concept drift.

In [57], the GC3 approach is an improvisation with a grid density oriented clustering algorithm, wherein the novelty is estimated by considering newly appearing grids in the data space. Such methods [52], [54], [53], [56], [55], [57] face challenge of curse of dimensionality and issues of distance methods detection of concept drift in the binary data spaces. Such techniques are effective for multi-class classification problems and many classes might emerge or wane during the process of stream.

In [58], the COC (Change of Concept) treats every feature as an independent stream of data and screens the correlation amidst current chunk and the training chunk that has to be referred. The change observed in the average correlation is used for signal of change. Pearson correlation is used which makes the normality assumption to a distribution.

In [59], the non-parametric unlabelled approach was proposed in the model of HDDDM (Hellinger Distance Drift Detection Methodology). Hellinger distance is used a measure to change in the distribution.

In [60] and [61], the PCA (Principal Component Analysis) is used for drift detection computationally efficient for high dimensional data streams. Such techniques reduce set of features essentially to be monitored. Both the methods are contrary in the selection of principal components. PCA models envisage issues because of considerable false alarming rate when compared to the other kind of multivariate distribution models.

In [62], [63], [64], [65] consider the classification process by taking in to account the posterior probability estimates of classifiers, for identifying the drift. It can be used with probabilistic classifiers that have output of class probabilities before thresholding them to generate any kind of final class label. By tracking the posterior probability estimates, the intensity of drift detection task is reduced to the levels of monitoring univariate stream of values, which enable the process more computationally efficient.

Such methods are very effective in reducing the false alarm rates, but their dependence on using the probabilistic models creates implications in terms of the applicability. The methods also impact any kind of change to the posterior distribution in terms of margin samples. Changes that deviate from the margin of classifier are considerably less critical than the classification process, however, none of the approaches offer robustness against them.

12 d) Observations

In many of the existing studies the focus is on development of drift detection methods and techniques for addressing the real drift. There is not much of research that has taken place in the domain which might impact the classification purpose and the performance. Despite that afore discussed drifts do not impact the true decision boundaries, it can lead to a better levels of decision boundary. The current techniques for handling the real drift might not be effective for any virtual drift, however, they offer different scenarios to learn and need varied solutions. In the instance of methods to address the real drift that is chosen to reset and retrain the classifier, the old concepts are ignored and the new concepts are learnt, which might not be an appropriate strategy to be used in virtual drift.

It could be more effective for calibrating the existing classifier rather than retraining them. Also, the techniques for handling the actual drift very much depends on feedback about the classifier's performance, whilst the techniques towards handling virtual drift shall operate even without the feedback [66]. It is imperative from the above review that all the three types discussed has significance and still there is scope for improvement in the models.

13 e) Future Research Objectives

The future challenges in concept drift could be attributed to the scope of scalability, sturdiness and efficacy of the models right from the levels of adaptation to more interpretable solutions, which can reduce dependency over the time and improve accurate feedback. Even moving from the adaptive algorithms to adaptive systems which could impact the complete knowledge discovery process apart from automating adaptation of the decision models. Few of the challenges envisaged in the process are discussed in [67].

The outline of the key issues have to be addressed by the research to ensure that a significant progress in the area of pre-processing techniques for the data stream problems.

Limited number of online and supervised discretises that were proposed in the literature reflect that the adjustments turnout to be more abrupt. But the problem is addressed to an extent by inclusion of class information in the discrete zation process. The tweaks that are abrupt and the ones that are labelling are two of the key concerns which must be addressed in the future researches.

There is integral need for wrapper-based solutions that were not explored much in the earlier researches. Pure wrapper based on the online learning solutions could limit the computational costs because of the discriminative ability and adaptability to drift. Also, there is need for further research in terms of feature and the instance selection methods which can directly impact the problem of concept drift.

IV.

14 Conclusion

In this paper, the categorization of the existing models of adaptive learning strategies based on the conceptual models, and the ones that are able to adapt the concept drift in addition to the contemporary techniques. Majority of the concept drift models presume that the changes taken place in the hidden contexts that are complex to be identified in the adaptive learning system. Because of the aforesaid reasons, concept drift is considered as an unpredictable and its detection is profoundly a reactive approach.

Numerous application settings wherein the concept drift is considered to be reappearing based on time line and based on varied objects over the modelled domain. The seasonal effects comprising vague periodicity towards a certain subgroup of object can be very common.

Availability of external levels of contextual information or the extraction from the hidden contexts based on the predictive features might assign in handling the recurrent concept drift in an effective manner.

Majority of the earlier works on the concept drift detection models reviewed in this survey do not address the issues of representation bias which is prevalent in many of the adaptive systems that can direct a specific kind of behaviour. However, when there is any kind of reinforcement feedback or any kind of closed loop control towards learning mechanism, it can't evaluate and compare the performance of the concept drift based on the historical data. Hence, it can be stated that there is need for more emphatic studies that support in embedded concept drift handling in real operational settings towards proper kind of validation. While majority of the works towards handling concept drift has considered the supervised settings having immediate availability, still the actual problems looms much wider.

In the process of a supervised learning that emerges over data, and the case of delayed set of ondemand labelling over the supervised learning,^{1 2}

¹© 2017 Global Journals Inc. (US)

²© 2017 Global Journals Inc. (US) Year 2017 Concept Drift Detection in Data Stream Mining: the Review of Contemporary Literature

adaptation mechanisms are to be investigated. The related research in the domain of concept drift is much beyond the application of machine learning, pattern recognition and the data mining solutions and there could be more explorative solutions in the domain.

[Hayat and Hashemi ()] ‘A dct based approach for detecting novelty and concept drift in data streams’. Morteza Hayat , Mahmoud Reza Hashemi . *Soft Computing and Pattern Recognition (SoCPaR), 2010 International Conference of*, 2010. IEEE.

[Gao ()] ‘A general framework for mining concept-drifting data streams with skewed distributions’. Jing Gao . *Proceedings of the 2007 SIAM International Conference on Data Mining*, Society for Industrial and Applied Mathematics (the 2007 SIAM International Conference on Data Mining) 2007.

[Sethi et al. ()] ‘A grid density based framework for classifying streaming data in the presence of concept drift’. Tegjyot Sethi , Mehmedkantardzic Singh , Hanquing Hu . *Journal of Intelligent Information Systems* 2016. 46 p. .

[Wang et al. ()] ‘A learning framework for online class imbalance learning’. Wang , Leandro L Shuo , Xin Minku , Yao . *Computational Intelligence and Ensemble Learning (CIEL)* 2013. 2013. IEEE. (IEEE Symposium on)

[Sousa et al. ()] ‘A new dynamic modeling framework for credit risk assessment’. Maria Sousa , João Rocha , Elísiobrandão Gama . *Expert Systems with Applications* 2016. 45 p. .

[Qahtan ()] ‘A pca-based change detection framework for multidimensional data streams: Change detection in multidimensional data streams’. Abdulhakim A Qahtan . *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining) 2015. ACM.

[Street et al. ()] ‘A streaming ensemble algorithm (SEA) for large-scale classification’. W Street , Yong Seog Nick , Kim . *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, (the seventh ACM SIGKDD international conference on Knowledge discovery and data mining) 2001. ACM.

[Kohavi ()] ‘A study of cross-validation and bootstrap for accuracy estimation and model selection’. Ron Kohavi . *Ijcai* 1995. 14 (2) .

[Batista et al. ()] ‘A study of the behavior of several methods for balancing machine learning training data’. Gustavo Eapa Batista , C Ronaldo , Maria Carolina Prati , Monard . *ACM Sigkdd Explorations Newsletter* 2004. 6 p. .

[Gama ()] ‘A survey on concept drift adaptation’. João Gama . *ACM Computing Surveys (CSUR)* 2014. 46 p. 44.

[Dries and Rückert ()] ‘Adaptive concept drift detection’. Anton Dries , Ulrich Rückert . *Statistical Analysis and Data Mining* 2009. 2 p. .

[Kubat and Matwin ()] ‘Addressing the curse of imbalanced training sets: one-sided selection’. Miroslav Kubat , Stan Matwin . *ICML* 1997. 97.

[Tang et al. ()] ‘An analysis of diversity measures’. E Tang , Ponnuthurai N Ke , Suganthan , Yao . *Machine learning* 2006. 65 (1) p. .

[Ryu and Woo ()] ‘An efficient method of building an ensemble of classifiers in streaming data’. Joung Ryu , Woo . *International Conference on Big Data Analytics*, (Berlin Heidelberg) 2012. Springer.

[López ()] ‘An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics’. Victoria López . *Information Sciences* 2013. 250 p. .

[?liobaite ()] ‘Change with delayed labeling: When is it detectable?’. Indre ?liobaite . *Data Mining Workshops (ICDMW), 2010 IEEE International Conference on*, 2010. IEEE.

[Wu and Chang ()] ‘Class-boundary alignment for imbalanced dataset learning’. Gang Wu , Edward Y Chang . *ICML 2003 workshop on learning from imbalanced data sets II*, (Washington, DC) 2003.

[Masud ()] ‘Classification and novel class detection in concept-drifting data streams under time constraints’. Mohammad Masud . *IEEE Transactions on Knowledge and Data Engineering* 2011. 23 p. .

[Gao ()] ‘Classifying data streams with skewed class distributions and concept drifts’. Jing Gao . *IEEE Internet Computing* 2008. 12 (6) .

[Wu ()] ‘Classifying imbalanced data streams via dynamic feature group weighting with importance sampling’. Ke Wu . *Proceedings of the 2014 SIAM International Conference on Data Mining*, (the 2014 SIAM International Conference on Data Mining) 2014.

[Ditzler ()] ‘Computational Intelligence in Dynamic and Uncertain Environments (CIDUE)’. Gregory Ditzler , Robipolikar . *IEEE Symposium on* 2011. 2011. IEEE. (Hellinger distance based drift detection for nonstationary environments)

14 CONCLUSION

- [Wang ()] ‘Concept drift detection for online class imbalance learning’. Shuo Wang . *The 2013 International Joint Conference on*, 2013. IEEE. (Neural Networks (IJCNN))
- [Wang and Abraham ()] ‘Concept drift detection for streaming data’. Heng Wang , Zubin Abraham . *2015 International Joint Conference on*, 2015. IEEE.
- [Harel] *Concept Drift Detection Through Re sampling*, Maayan Harel . ICML. 2014.
- [Minku and Yao ()] ‘DDD: A new ensemble approach for dealing with concept drift’. Leandro L Minku , Xin Yao . *IEEE transactions on knowledge and data engineering* 2012. 24 p. .
- [Lee and Magoules ()] ‘Detection of concept drift for learning from stream data’. Jeonghoon Lee , Frederic Magoules . *High Performance Computing and Communication & 2012 IEEE 9th International Conference on Embedded Software and Systems (HPCC-ICES)*, 2012. 2012. IEEE. (IEEE 14th International Conference on)
- [Brown ()] ‘Diversity creation methods: a survey and categorisation’. Gavin Brown . *Information Fusion* 2005. 6 p. .
- [Lindstrom et al. ()] ‘Drift detection using uncertainty distribution divergence’. Patrick Lindstrom , Brian Mac Namee , Sarah Jane Delany . *Evolving Systems* 2013. 4 p. .
- [Tsymbal ()] ‘Dynamic integration of classifiers for handling concept drift’. Alexey Tsymbal . *Information fusion* 2008. 9 p. .
- [Mirza et al. ()] ‘Ensemble of subset online sequential extreme learning machine for class imbalance and concept drift’. Bilal Mirza , Zhiping Lin , Nan Liu . *Neuro computing* 2015. 149 p. .
- [Van Hulse et al. ()] ‘Experimental perspectives on learning from imbalanced data’. Van Hulse , Taghi M Jason , Amri Khoshgoftaar , Napolitano . *Proceedings of the 24th international conference on Machine learning*, (the 24th international conference on Machine learning) 2007. ACM.
- [Ross ()] ‘Exponentially weighted moving average charts for detecting concept drift’. Gordon J Ross . *Pattern Recognition Letters* 2012. 33 p. .
- [Pan and Shirui ()] ‘Graph ensemble boosting for imbalanced noisy graph stream classification’. Pan , Shirui . *IEEE transactions on cybernetics* 2015. 45 p. .
- [Hoens et al. ()] ‘Heuristic updatable weighted random subspaces for non-stationary environments’. T Hoens , Nitesh V Ryan , Robi Polikar . *Data Mining (ICDM), 2011 IEEE 11th International Conference on*, 2011. IEEE.
- [Schlimmer and Granger ()] ‘Incremental learning from noisy data’. Jeffrey C Schlimmer , Richard H Granger . *Machine learning* 1986. 1 p. .
- [Ditzler ()] ‘Incremental learning of concept drift from streaming imbalanced data’. Gregory Ditzler , Robi Polikar . *IEEE transactions on knowledge and data engineering* 2013. 25 p. .
- [Elwell ()] ‘Incremental learning of concept drift in nonstationary environments’. Ryan Elwell , Robi Polikar . *IEEE Transactions on Neural Networks* 2011. 22 p. .
- [Visa ()] ‘Issues in mining imbalanced data sets-a review paper’. Sofia Visa , Ancaralescu . *Proceedings of the sixteen midwest artificial intelligence and cognitive science conference*, sn (the sixteen midwest artificial intelligence and cognitive science conference) 2005. 2005.
- [Mani and Zhang ()] ‘kNN approach to unbalanced data distributions: a case study involving information extraction’. Inderjeet Mani , I Zhang . *Proceedings of workshop on learning from imbalanced datasets*, (workshop on learning from imbalanced datasets) 2003.
- [Polikar ()] ‘Learn++: An incremental learning algorithm for supervised neural networks’. Robi Polikar . *IEEE transactions on systems, man, and cybernetics, part C (applications and reviews)*, 2001. 31 p. .
- [Hoens et al. ()] ‘Learning from streaming data with concept drift and imbalance: an overview’. T Hoens , Robi Polikar Ryan , Nitesh V Chawla . *Progress in Artificial Intelligence* 2012. 1 (1) p. .
- [Hoens et al. ()] ‘Learning in non-stationary environments with class imbalance’. Thomas Hoens , Nitesh V Ryan , Chawla . *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, (the 18th ACM SIGKDD international conference on Knowledge discovery and data mining) 2012. ACM.
- [Ditzler ()] ‘Learning in nonstationary environments: A survey’. Gregory Ditzler . *IEEE Computational Intelligence Magazine* 2015. 10 p. .
- [Widmer ()] ‘Learning in the presence of concept drift and hidden contexts’. Gerhard Widmer , Miroslav Kubat . *Machine learning* 1996. 23 p. .
- [Hoens and Ryan ()] *Living in an imbalanced world*, Thomas Hoens , Ryan . 2012. University of Notre Dame
- [Kubat et al. ()] ‘Machine learning for the detection of oil spills in satellite radar images’. Kubat , Robert C Miroslav , Stan Holte , Matwin . *Machine learning* 1998. 30 p. .

-
- [Sobhani and Beigy ()] *New drift detection method for data streams*, Parinaz Sobhani , Hamid Beigy . 2011. Berlin Heidelberg: Springer.
- [Zliobaite ()] ‘Next challenges for adaptive learning systems’. Indre Zliobaite . *ACM SIGKDD Explorations Newsletter* 2012. 14 p. .
- [Faria et al. ()] ‘Novelty detection algorithm for data streams multi-class problems’. Elaine R Faria , João Gama , André Cplf Carvalho . *Proceedings of the 28th annual ACM symposium on applied computing*, (the 28th annual ACM symposium on applied computing) 2013. ACM.
- [Spinosa et al. ()] ‘Olindda: A clusterbased approach for detecting novelty and concept drift in data streams’. Eduardo J Spinosa , André Ponce De Leon F De Carvalho , João Gama . *Proceedings of the 2007 ACM symposium on Applied computing*, (the 2007 ACM symposium on Applied computing) 2007. ACM.
- [Yan ()] ‘On predicting rare classes with SVM ensembles in scene classification’. Rong Yan . *Proceedings.(ICASSP’03). 2003 IEEE International Conference on*, (.ICASSP’03). 2003 IEEE International Conference on) 2003. 2003. IEEE. 3.
- [Ghazi Khani et al. ()] ‘Online neural network model for nonstationary and imbalanced data stream classification’. Adel Ghazi Khani , Reza Monsefi , Hadi Sadoghi Yazdi . *International Journal of Machine Learning and Cybernetics* 2014. 5 p. .
- [Bach and Malo ()] ‘Paired learners for concept drift’. Stephen H Bach , Marcus A Malo . *Data Mining, 2008. ICDM’08. Eighth IEEE International Conference on*, 2008. IEEE.
- [Kuncheva and Faithfull ()] ‘PCA feature extraction for change detection in multidimensional unlabeled data’. Ludmila I Kuncheva , William J Faithfull . *IEEE transactions on neural networks and learning systems*, 2014. 25 p. .
- [Brzezinski and Wski ()] *Prequential AUC for classifier evaluation and drift detection in evolving data streams*, Dariusz Brzezinski , Jerzy Stefano Wski . 2014. Springer International Publishing. (International Workshop on New Frontiers in Mining Complex Patterns)
- [Brzezinski and Wski] *Prequential AUC: Properties of the Area Under the ROC Curve for Data Streams with Concept Drift*, Dariusz Brzezinski , Jerzy Stefano Wski .
- [Brzezinski and Stefanowski ()] ‘Reacting to different types of concept drift: The accuracy updated ensemble algorithm’. Dariusz Brzezinski , Jerzy Stefanowski . *IEEE Transactions on Neural Networks and Learning Systems* 2014. 25 p. .
- [Ghazikhani and Monsefi ()] ‘Recursive least square perceptron model for non-stationary and imbalanced data stream classification’. Adel Ghazikhani , Reza Monsefi , Hadisadoghiyazdi . *Evolving Systems* 2013. 4 p. .
- [Chen and He ()] ‘Sera: selectively recursive approach towards non stationary imbalanced stream data mining’. Sheng Chen , Haibo He . *IJCNN 2009. International Joint Conference on*, 2009. 2009. IEEE.
- [Chawla ()] ‘SMOTE: synthetic minority over-sampling technique’. Nitesh V Chawla . *Journal of artificial intelligence research* 2002. 16 p. .
- [Forman ()] ‘Tackling concept drift by temporal inductive transfer’. George Forman . *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, (the 29th annual international ACM SIGIR conference on Research and development in information retrieval) 2006. ACM.
- [Japkowicz and Stephen ()] ‘The class imbalance problem: A systematic study’. Nathalie Japkowicz , Shaju Stephen . *Intelligent data analysis* 2002. 6 p. .
- [Gantz and Reinsel ()] ‘The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the far east’. John Gantz , David Reinsel . *IDC iView: IDC Analyze the future*, 2007. 2012. 2012. p. .
- [Kelly et al. ()] ‘The impact of changing populations on classifier performance’. Mark G Kelly , J David , Niall M Hand , Adams . *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining*, (the fifth ACM SIGKDD international conference on Knowledge discovery and data mining) 1999. ACM.
- [Minku et al. ()] ‘The impact of diversity on online ensemble learning in the presence of concept drift’. Leandro L Minku , P Allan , Xin White , Yao . *IEEE Transactions on Knowledge and Data Engineering* 2010. 22 p. .
- [Chen and He ()] ‘Towards incremental learning of non stationary imbalanced data stream: a multiple selectively recursive approach’. Sheng Chen , Haibo He . *Evolving Systems* 2011. 2 (1) p. .
- [Dredze et al. ()] ‘We’re not in Kansas anymore: detecting domain changes in streams’. Mark Dredze , Tim Oates , Christine Piatko . *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, (the 2010 Conference on Empirical Methods in Natural Language Processing) 2010. Association for Computational Linguistics.