ESANN 2019 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2019, i6doc.com publ., ISBN 978-287-587-065-0.
Available from http://www.i6doc.com/en/.

Recent trends in streaming data analysis, concept drift and analysis of dynamic data sets

Albert Bifet[1], Barbara Hammer[2] and Frank-M. Schleif[3]

1- LTCI, Telecom ParisTech,
Université Paris-Saclay
Paris, France

2- Bielefeld University,
CITEC centre of excellence
Bielefeld, Germany

3- University of Birmingham
School of Computer Science,
Edgbaston B15 2TT Birmingham,
United Kingdom.

Abstract. Today, many data are not any longer static but occur as dynamic data streams with high velocity, variability and volume. This leads to new challenges to be addressed by novel or adapted algorithms. In this tutorial we provide an introduction into the field of streaming data analysis summarizing its major characteristics and highlighting important research directions in the analysis of dynamic data.

1 Introduction

In many application domains data are given at large scale and with high velocity, requesting an in time analysis without the possibility to store large parts of the data or to process them multiple times. Examples include astronomical observations, earth sensing satellites and climate observation, genomics and post-genomics data, data gathered by smart sensors such as smart phones or wearable devices, IoT data, data gathered from assistive technologies such as Amazon's Alexa, data gathered in smart cities or smart factories, etc. Such data are widely referred to as streaming data since measurements arrive continuously as a data stream. In addition to the sheer data size, which typically prevents its processing in batches, streaming data can pose additional challenges to the models which renders standard techniques of machine learning unsuitable.

In recent years, quite a few approaches have been proposed in this context, most of which are different from current popular learning methods for batch learning, see e.g. [1, 2, 3, 4, 5, 6, 7, 8] for overviews. Besides the mere computational issues, online learning faces quite a few challenges which are fundamentally different from classical batch processing. In the following, we shortly define what we refer to as online learning first and we give an overview about challenges in this domain. We address two major tasks: (i) How to derive supervised models

ESANN 2019 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2019, i6doc.com publ., ISBN 978-287-587-065-0. Available from http://www.i6doc.com/en/.

for streaming data? (ii) How to best represent data in the streaming setting? We conclude with a glimpse on recent directions in this domain.

2 Problem statement and challenges

In *online learning*, a stream $s_1, s_2, s_3, \dots s_t, \dots$ of data $s_i \in \mathcal{S}$ is given, and the task is to incrementally infer a model h_t after having seen instances $s_1, \dots, s_t$. Unlike batch processing, the characteristics of the subsequent data $s_{t'}$ for $t' > t$ is unknown, and it is not guaranteed that future data obey the same characteristics as the already seen samples. Such settings pose a variety of challenges towards learning, which do not occur in this form in offline learning.

> **Challenge 1:** How to devise efficient online learning rules, which instantaneously adapt model parameters based on streaming data?

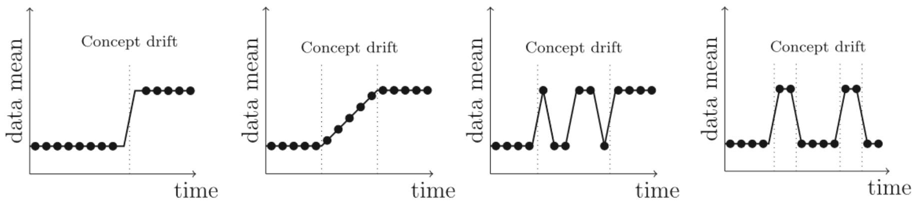
Typically, *efficient incremental learning* is required since models should be available in real time while observing the data stream. Hence, model h_t is typically inferred from the data point s_t (or a small window of previous samples only) and the previous model h_{t-1}, without explicitly storing all previous data points, i.e. *memory efficiency* is given. In addition, *computational efficiency* is aimed for, i.e. the learning rules should be efficient, yet *effective* in the sense that the update resemble the accuracy of batch learning as far as possible (provided data are i.i.d.) but without having access to the full data [9].

> **Challenge 2a:** How to find suitable model meta-parameters?

For offline learning of parametric models, typically, model meta-parameters, which characterize the complexity and degree of non-linearity (e.g. kernel width or neural architecture) are determined prior to training by Bayesian optimization of cross-validation. For online learning, this is not possible, since the required model complexity cannot be estimated prior to training without access to sufficient training data. Hence *model meta-parameters, which characterize the model complexity, become model parameters in online learning* — they need to be adapted while processing the data stream, unless non-parametric modeling (e.g. k-NN) or a priorly limited complexity (e.g. linear maps) are chosen. The problem of a priorly unclear model complexity requires suitable learning strategies for the meta-parameters, which characterize the model – usually a challenging task, since these meta-parameters are typically discrete-valued [4, 10].

> **Challenge 2b:** How to preprocess data?

Besides model complexity, many models typically come with a set of further meta-parameters such as step size, strength or regularization, etc., which need to be robustly chosen or adapted for online learning, since its suitability can hardly be determined prior to training. A related challenge concerns data preprocessing, such as data normalization, or dimensionality reduction for numerical stability and sampling in the presence of imbalanced classes. Similar to model meta-parameters, all data preprocessing steps become part of the actual processing pipeline and need to be adjusted while training [11].



(a) Rapid (b) Incremental (c) Gradual (d) Recurring

Fig. 1: Different drift types as they occur in streaming data analysis [5].

Challenge 3: How to deal with concept drift?

For online learning, it is usually not guaranteed that data come from a stationary data source; hence the typical assumption of batch processing, the fact that data are i.i.d., is violated. Whenever at least two points t, t' in time exist where the underlying probability distribution changes, i.e. $\mathcal{P}_t \neq \mathcal{P}_{t'}$ one speaks of *concept drift*. Thereby, drift characteristics can vary, as depicted in Fig. 1, yielding smooth or rapid drift, incremental, gradual or reoccurring drift. In addition to such structural changes, outliers can occur, i.e. data s_t deviate randomly from the underlying distribution $\mathcal{P}_t$. Such settings probably constitute the most fundamental difference of learning with streaming data to the batch setting, since models need to adequately react to changed underlying probability distributions during their whole lifetime. In particular if the type of drift is unknown, the classical *stability-plasticity dilemma* arises and learning faces an essentially ill-posed problem [12]: when is observed change caused by an underlying structure and should be taken into account, and when is it given by noise and should be neglected? Interestingly, if drift occurs, online learning can yield results which are superior to batch training, since online learning can react to changes and provide an optimal model at every given time step, while batch learning needs to restrict to an average model which fits an average time point only [6].

In practice, quite a number of different models have been proposed, which can roughly be decomposed into the following categories:

Supervised learning: Data have the form $s_t = (x_t, y_t)$ and the task is to learn a model h_t which predicts the subsequent output $y_{t+1} \approx f(x_{t+1})$ before actually seeing the true outcome. Such scenarios are relevant in practice whenever an early prediction is required such as predicting the behavior of participants in a traffic scene for early motion planning. For supervised learning, one distinguishes the notion of *real drift*, which refers to a change of the posterior distribution $\mathcal{P}(y|x)$ and *virtual drift* or *covariate shift*, which refers to a change of the input distribution $\mathcal{P}(x)$ only without affecting the posterior.

Unsupervised data representation: Data s_t are unlabeled, and the task is to solve a problem related to data representation such as online generative modeling, online compression, dimensionality reduction, clustering, or outlier detection. Interestingly, quite a few early methods of machine learning such as principal

ESANN 2019 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2019, i6doc.com publ., ISBN 978-287-587-065-0. Available from http://www.i6doc.com/en/.

component analysis via the Oja-learning rule or self-organizing maps have been inspired by biological counterparts and phrased as online learning algorithms for streaming data in its original form [13, 14].

Time characteristics of data: Some methods explicitly focus on the time characteristics of the data and tackle challenges, which can only be asked in the context of data streams. Partially, these tasks occur as sub-problems of supervised or unsupervised learning problems for streaming data. One central questions, which is often embedded in so-called *active* methods for streaming data, is *drift detection*, i.e. the task to detect points in time where a significant change of the underlying probability distribution can be observed [15]. Such detectors are often coupled by a strategy to adapt the model whenever a drift is detected [3]. Further challenges address the temporal characteristics, and focus on time-invariant motives or possible (Granger) causal relations [16, 17, 18].

3 Supervised learning in the streaming context

Interestingly, many classical learning rules such as perceptron learning, learning vector quantization, self-organizing maps, or neural back-propagation have been proposed as online learning rules, and these technologies, indeed, provide an astonishing robustness when learning with drift [19]. The SVM and, more generally, the focus on convex optimization methods and the possibility of their strong mathematical substantiation led to a dominance of batch learners in machine learning. With the advent of large data sets and the involved necessity to process data in patches, learning from streaming data became an issue again. One of the first methods which have been proposed in this context is the *very fast decision tree* [16], an incremental variant of a decision tree which resembles its batch counterpart. Another popular early method was *Learn++*, which constructs an ensemble very much in the spirit of adaptive boosting [20].

There exists a quite large variety of different approaches for supervised online learning, which rely on different principles and try to address different challenges. Quite a few methods focus on the *computational issue* how to efficiently transfer popular batch learners into online-learning counterparts, such that their result is equivalent to or (approximates) the result of the batch learner, a popular example being e.g. incremental variants of SVM [9], incremental decision trees or random forests [21], or incremental Bayesian learners [22]. Other methods investigate how to efficiently adjust the model complexity according to the observed data, whereby methods have to face a balance between [23, 24].

One of the most crucial challenges is given by possible *concept drift* and the question, how to deal with such drift. A variety of approaches centers on the detection and quantification of the observed drift in the first place, enabling a reaction to the drift in so-called *active approaches*. Examples are window-based approaches such as presented in the approaches [25, 15, 26], or the approaches [27, 28], which relies on a modeling via the β-distribution. Such detectors can be combined with random forests or other classifiers [29]. On the other side, *passive approaches* continuously adapt models according to the given data or

ESANN 2019 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2019, i6doc.com publ., ISBN 978-287-587-065-0. Available from http://www.i6doc.com/en/.

an active window, respectively, this way reacting also to smooth drift. Popular approaches often rely on non-parametric methods, such as extremely robust k-NN based methods [24] or prototype-based approaches [30]. Many modern technologies fall into the category of hybrid techniques, which combine active drift detection and continuous passive adaptation, this way combining the best of both worlds.

Many recent very popular different algorithmic approaches follow the idea of ensemble techniques, in particular tree ensembles [31] or ensembles of local models [32], this way displaying a high robustness and independence of model parametrization. From an application point of view, it is interesting to investigate which *types of drift* the models can deal with. By design, active drift detection methods are restricted to rapid drift, hence they are less suited for subtle incremental changes, but, on the other hand, typically react rapidly to the detected change. Passive or hybrid methods typically smoothly deal with continuously changing environment. Interestingly, there exist currently only very few approaches capable of dealing with reoccurring drift, a few of those having been proposed in the work [33, 34, 35]. With the advent of streaming data e.g. in personalized assistive systems such as investigated in the work [36], the relevance of such methodologies, which are capable of a flexible reaction to priorly unknown types of drift, will become even more prominent.

Quite a few further challenges have recently been addressed in the context of supervised learning for streaming data, a few keywords being the mathematical investigation of their convergence properties [37], learning in the context of more complex outputs such as structured predictions and multi-label learning [38], semi-supervised online learning [39], or learning for imbalanced data [40].

4 Unsupervised data representation in the streaming context

Similar like in classical data analysis the pre-processing can have a large impact on the effectiveness of a subsequent analysis. Still the majority of streaming data are (multi-dimensional) vectorial but the generating sources may have very different qualities. For example the data may be generated at different output rates, the noise level may be very different and the transmitted amount of information per time window can vary to a large degree. Also other type of non-numeric data by means of textual data, graphs or symbolic sequences are getting more and more common [41], asking for appropriate preprocessing to enable effective model building. Common to used pre-processing concepts is a (near) memory less approach. The incoming data are processed on the fly and the respective processing algorithms can only keep very limited information about the data seen so far. Hence in the most basic pre-processings like smoothing or noise reduction simple filter techniques like moving average are used. A survey can be found in [42].

Also the general problem of imbalanced data is a particular challenge in the streaming context [43], the distribution in general or with respect to class labels

ESANN 2019 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2019, i6doc.com publ., ISBN 978-287-587-065-0. Available from http://www.i6doc.com/en/.

may not be uniform and can evolve other time. One application of imbalanced learning is anomaly detection, where the problem consists in predicting when an anomaly appears. As anomalies appear with a very low frequency, it is a classical example of imbalanced learning [44].

Another important aspect is the existence of outliers. As discussed before streaming data are dynamic and hence show varying distributions and data characteristics. Outliers are particular challenging in the streaming context because the data distributions are naturally changing and it becomes very complicate to decide whether the observation is an outlier or due to a change in the data distributions. Early work addressing this point can be found in [45].

A classical data preprocessing approach is the principal component analysis (PCA) to characterize the variance in the data. Also in the streaming context initial work was provided to allow PCA like data processing [46]. The majority of those techniques go back to traditional power iteration methods to get an estimate of the underlying eigen functions with links to early neural network approaches, like Oja PCA [47].

In high-dimensional data, using all attributes is often not feasible, and we may need to preprocess the data to perform feature selection, or feature transformation. This can be done by a streaming PCA [46], but also other dimension reduction approaches are under research like multi dimensional scaling (MDS) [48] to obtain low dimensional data representations which keeps some distance preservation.

Another interesting approach is to systematically identify relevant input features by means of a weighting or relevance scheme and a metric adaptation concept [30].

The streaming domain shows also links to different types of the (contextual) *bandit* problem. In the contextual bandit problem a learner chooses an action among a set of available ones, based on the observation of action features or contexts and then receives a reward that quantifies the quality of the chosen action. This scenario can be used in vary different dedicated algorithms e.g. to adapt models, to switch between different input streams or other ways [49, 50].

The majority of the proposed streaming analysis methods make still use of linear or piecewise linear models due to the simplicity which is very desirable for large scale and high throughput streaming data. To overcome the limitation of linear methods also first kernelization strategies have been proposed, with some initial work in [51].

5 Future trends

Future trends in streaming data analysis, are based on how to develop data streaming methods that scale to Big Data like large deep neural networks, but work well in all domains. In the future, the quantity of data generated in real-time is going to continue growing, so there will be need to develop new methods using large distributed systems.

Deep Learning has become a very extreme successful use case for Machine Learning and Artificial Intelligence, due to the availability of massive quantities of data to build data models, and large computational resources. How to implement powerful methods such as deep learning, in a more green, low-emissions, sustainable way, is going to be an important scientific trend to fight against climate change. Standard deep learning techniques needs to do several passes over the data. How to build models only doing one pass over the data, without storing the data, will be an important future area of research [52].

Finally, when dealing with large quantities of data, an important trend will be how to do online learning using distributed streaming engines, as Apache Spark, Apache Flink, Apache Storm and others. Algorithms have to be distributed in an efficient way, so that the performance of the distributed algorithms does not suffer from the network cost of distributing the data [53].

6 Conclusions

In this tutorial we briefly reviewed challenges and approaches common in the field of streaming data analysis, concept drift and the analysis of dynamic data sets. The more recent proposals in these domains provide sophisticated algorithms and models to address the aforementioned challenges in streaming analysis and in particular the handling of concept drift. Also a variety of classical supervised and unsupervised analysis tasks like modeling non-linear decision planes or finding relevant dimensions in the data streams now came in the focus of recent research. Although the field has made much progress in preprocessing [42], concept drift detection [24, 26] and by means of generic frameworks for streaming analysis [54] there are still particular challenges as detailed before with a variety of open research perspectives.

Acknowledgment

BH gratefully acknowledges funding in the frame of the CITEC centre of excellence and the BMBF project ITS_ML, grant number 01IS18041A. FMS is thankful for support in the FuE program *Informations- und Kommunikationstechnik* of the StMWi, project *OBerA*, grant number IUK-1709-0011// IUK530/010.

References

[1] RR Ade and PR Deshmukh. Methods for incremental learning: a survey. *International Journal of Data Mining & Knowledge Management Process*, 3(4):119, 2013.

[2] Gregory Ditzler, Manuel Roveri, Cesare Alippi, and Robi Polikar. Learning in nonstationary environments: A survey. *Computational Intelligence Magazine, IEEE*, 10(4):12–25, 2015.

[3] R. Elwell and R. Polikar. Incremental learning of concept drift in nonstationary environments. *IEEE Transactions on Neural Networks*, 22(10):1517–1531, Oct 2011.

[4] Alexander Gepperth and Barbara Hammer. Incremental learning algorithms and applications. In *Proceedings of the European Sympoisum on Artificial Neural Networks (ESANN)*, 2016.

[5] João Gama, Indre Žliobaite, Albert Bifet, Mykola Pechenizkiy, and Abdelhamid Bouchachia. A survey on concept drift adaptation. *ACM Computing Surveys (CSUR)*, 46(4):44, 2014.

[6] Viktor Losing, Barbara Hammer, and Heiko Wersing. Incremental on-line learning: A review and comparison of state of the art algorithms. *Neurocomputing*, 275:1261–1274, 2018.

[7] Indre Zliobaite. Learning under concept drift: an overview, 2010.

[8] Adriana Sayuri Iwashita and Joao Paulo Papa. An Overview on Concept Drift Learning. *IEEE Access*, 7(Section III):1–1, 2018.

[9] C. P. Diehl and G. Cauwenberghs. Svm incremental learning, adaptation and optimization. In *Proceedings of the International Joint Conference on Neural Networks, 2003.*, volume 4, pages 2685–2690 vol.4, July 2003.

[10] Prachi Joshi and Parag Kulkarni. Incremental learning: areas and methods-a survey. *International Journal of Data Mining & Knowledge Management Process*, 2(5):43, 2012.

[11] Gregory Ditzler and Robi Polikar. Incremental learning of concept drift from streaming imbalanced data. *ieee transactions on knowledge and data engineering*, 25(10):2283–2301, 2013.

[12] Martial Mermillod, AurÃ©lia Bugaiska, and Patrick Bonin. The stability-plasticity dilemma: Investigating the continuum from catastrophic forgetting to age-limited learning effects. *Frontiers in Psychology*, 4(504), 2013.

[13] G. A. Carpenter, S. Grossberg, and J. Reynolds. Artmap: a self-organizing neural network architecture for fast supervised learning and pattern recognition. In *IJCNN-91-Seattle International Joint Conference on Neural Networks*, volume i, pages 863–868 vol.1, July 1991.

[14] Teuvo Kohonen. *Self-organizing maps.* Springer series in information sciences, 30. Springer, Berlin, 3rd edition, December 2001.

[15] Albert Bifet and Ricard Gavalda. Learning from time-changing data with adaptive windowing. In *SIAM International Conference on Data Mining (SDM)*, volume 7, page 2007. SIAM, 2007.

[16] Pedro Domingos and Geoff Hulten. Mining high-speed data streams. In *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 71–80. ACM, 2000.

[17] Pedro Domingos and Geoff Hulten. A general framework for mining massive data streams. *Journal of Computational and Graphical Statistics*, 12(4):945–949, 2003.

[18] Mohamed Medhat Gaber, Arkady Zaslavsky, and Shonali Krishnaswamy. Mining data streams: a review. *ACM Sigmod Record*, 34(2):18–26, 2005.

[19] Michiel Straat, Fthi Abadi, Christina Göpfert, Barbara Hammer, and Michael Biehl. Statistical mechanics of on-line learning under concept drift. *Entropy*, 20(10):775, 2018.

[20] R. Polikar, L. Upda, S. S. Upda, and V. Honavar. Learn++: an incremental learning algorithm for supervised neural networks. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 31(4):497–508, Nov 2001.

[21] Aiping Wang, Guowei Wan, Zhiquan Cheng, and Sikun Li. An incremental extremely random forest classifier for online learning and tracking. In *2009 16th IEEE International Conference on Image Processing (ICIP)*, Nov 2009.

[22] W. Jamil and A. Bouchachia. Online bayesian shrinkage regression. In Michel Verleysen, editor, *Proceedings of the 27. European Symposium on Artificial Neural Networks ESANN 2019*, page numbers to be obtained from ToC of this proceedings book, Evere, Belgium, 2019. D-Side Publications.

[23] Iain A D Gunn, Álvar Arnaiz-González, and Ludmila I Kuncheva. A taxonomic look at instance-based stream classifiers. *Neurocomputing*, 286:167–178, 2018.

[24] Viktor Losing, Barbara Hammer, and Heiko Wersing. Self-adjusting memory: How to deal with diverse drift types. *IJCAI International Joint Conference on Artificial Intelligence*, pages 4899–4903, 2017.

ESANN 2019 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2019, i6doc.com publ., ISBN 978-287-587-065-0. Available from http://www.i6doc.com/en/.

[25] Joao Gama, Pedro Medas, Gladys Castillo, and Pedro Rodrigues. Learning with drift detection. In *Advances in artificial intelligence–SBIA 2004*, pages 286–295. Springer, 2004.

[26] Roberto Souto Maior Barros and Silas Garrido T.Carvalho Santos. A large-scale comparison of concept drift detectors. *Information Sciences*, 451-452:348–370, 2018.

[27] L. Fleckenstein, S. Kauschke, and J. Fürnkranz. Beta distribution drift detection for adaptive classifiers. In Michel Verleysen, editor, *Proceedings of the 27. European Symposium on Artificial Neural Networks ESANN 2019*, page numbers to be obtained from ToC of this proceedings book, Evere, Belgium, 2019. D-Side Publications.

[28] Jeremias Knoblauch, Jack E Jewson, and Theodoros Damoulas. Doubly robust bayesian inference for non-stationary streaming data with \beta-divergences. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems 31*, pages 64–75. Curran Associates, Inc., 2018.

[29] Cesare Alippi and Manuel Roveri. Just-in-time adaptive classifiers-part i: Detecting nonstationary changes. *IEEE Transactions on Neural Networks*, 19(7):1145–1153, July 2008.

[30] Ilja Kuzborskij and Nicolò Cesa-Bianchi. Nonparametric Online Regression while Learning the Metric. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, 4-9 December 2017, Long Beach, CA, USA*, pages 667–676, 2017.

[31] Heitor Gomes, Albert Bifet, Jesse Read, Jean Paul Barddal, Fabrício Enembreck, Bernhard Pfharinger, Geoff Holmes, and Talel Abdessalem. Adaptive random forests for evolving data stream classification. *Machine Learning*, 106(9):1469–1495, Oct 2017.

[32] Junming Shao, Feng Huang, Qinli Yang, and Guangchun Luo. Robust Prototype-Based Learning on Data Streams. *IEEE Transactions on Knowledge and Data Engineering*, 30(5):978–991, 2018.

[33] Viktor Losing, Barbara Hammer, and Heiko Wersing. Tackling heterogeneous concept drift with the self-adjusting memory (SAM). *Knowl. Inf. Syst.*, 54(1):171–201, 2018.

[34] Viktor Losing, Barbara Hammer, and Heiko Wersing. KNN classifier with self adjusting memory for heterogeneous concept drift. In *IEEE 16th International Conference on Data Mining, ICDM 2016, December 12-15, 2016, Barcelona, Spain*, pages 291–300, 2016.

[35] C. Raab, M. Heusinger, and F.-M. Schleif. Reactive soft prototype computing for frequent reoccurring concept drift. In Michel Verleysen, editor, *Proceedings of the 27. European Symposium on Artificial Neural Networks ESANN 2019*, page numbers to be obtained from ToC of this proceedings book, Evere, Belgium, 2019. D-Side Publications.

[36] P. Siirtola, H. Koskimäki, and J. Röning. Importance of user inputs while using incremental learning to personalize human activity recognition models. In Michel Verleysen, editor, *Proceedings of the 27. European Symposium on Artificial Neural Networks ESANN 2019*, page numbers to be obtained from ToC of this proceedings book, Evere, Belgium, 2019. D-Side Publications.

[37] Lijun Zhang, Shiyin Lu, and Zhi-Hua Zhou. Adaptive online learning in dynamic environments. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems 31*, pages 1330–1340. Curran Associates, Inc., 2018.

[38] Rajasekar Venkatesan, Meng Joo Er, Shiqian Wu, and Mahardhika Pratama. A novel online real-time classifier for multi-label data streams. *CoRR*, abs/1608.08905, 2016.

[39] Yanchao Li, Yongli Wang, Qi Liu, Cheng Bi, Xiaohui Jiang, and Shurong Sun. Incremental semi-supervised learning on streaming data. *Pattern Recognition*, 88:383 – 396, 2019.

[40] Shuo Wang, Leandro L. Minku, and Xin Yao. A systematic study of online class imbalance learning with concept drift. *CoRR*, abs/1703.06683, 2017.

ESANN 2019 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2019, i6doc.com publ., ISBN 978-287-587-065-0. Available from http://www.i6doc.com/en/.

[41] Daniele Zambon, Cesare Alippi, and Lorenzo Livi. Concept Drift and Anomaly Detection in Graph Streams. *IEEE Transactions on Neural Networks and Learning Systems*, 29(11):5592–5605, 2018.

[42] Sergio Ramírez-Gallego, Bartosz Krawczyk, Salvador García, Michał Woźniak, and Francisco Herrera. A survey on data preprocessing for data stream mining: Current status and future directions. *Neurocomputing*, 239:39–57, 2017.

[43] Yan Yan, Tianbao Yang, Yi Yang, and Jianhui Chen. A framework of online learning with imbalanced streaming data, 2017.

[44] Swee Chuan Tan, Kai Ming Ting, and Fei Tony Liu. Fast anomaly detection for streaming data. In *IJCAI*, pages 1511–1516. IJCAI/AAAI, 2011.

[45] Alice Marascu and Florent Masseglia. Parameterless outlier detection in data streams. In *Proceedings of the 2009 ACM Symposium on Applied Computing (SAC), Honolulu, Hawaii, USA, March 9-12, 2009*, pages 1491–1495, 2009.

[46] Ioannis Mitliagkas, Constantine Caramanis, and Prateek Jain. Memory Limited, Streaming PCA. *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems*, pages 1–9, 2013.

[47] Erkki Oja. Simplified neuron model as a principal component analyzer. *Journal of Mathematical Biology*, 15(3):267–273, Nov 1982.

[48] Arvind Agarwal Jeff M. Phillips Hal Daume III Suresh Venkatasubramanian. Incremental Multi-Dimensional Scaling. `https://www.cs.utah.edu/~jeffp/papers/incrementalMDS.pdf`, 2010. Online; accessed 19 Feb. 2019.

[49] Thibault Gisselbrecht. Bandit algorithms for real-time data capture on large social medias. *CoRR*, abs/1808.10725, 2018.

[50] Thibault Gisselbrecht, Sylvain Lamprier, and Patrick Gallinari. Dynamic data capture from social media streams: A contextual bandit approach. In *Proceedings of the Tenth International Conference on Web and Social Media, Cologne, Germany, May 17-20, 2016.*, pages 131–140. AAAI Press, 2016.

[51] Joel A. Tropp, Alp Yurtsever, Madeleine Udell, and Volkan Cevher. Fixed-Rank Approximation of a Positive-Semidefinite Matrix from Streaming Data. *dvances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, 4-9 December*, (Nips):1–10, 2017.

[52] Ahmad M Mustafa, Gbadebo Ayoade, Khaled Al-naami, Latifur Khan, Kevin W Hamlen, Bhavani Thuraisingham, and Frederico Araujo. Unsupervised Deep Embedding for Novel Class Detection over Data Stream. *IEEE International Conference on Big Data (BIG-DATA) Unsupervised*, pages 1830–1839, 2017.

[53] Albert Bifet, Jesse Read, and Geoff Holmes. Efficient Online Evaluation of Big Data Stream Classifiers Categories and Subject Descriptors. *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 59–68, 2015.

[54] Albert Bifet, Ricard Gavald, Geoff Holmes, and Bernhard Pfahringer. *Machine Learning for Data Streams: With Practical Examples in MOA*. The MIT Press, 2018.