



Learning from Imbalanced Data Streams Using Rotation-Based Ensemble Classifiers

Ireneusz Czarnowski^(✉) 

Department of Information Systems, Gdynia Maritime University, Morska 83, 81-225 Gdynia, Poland

i.czarnowski@umg.edu.pl

Abstract. In this paper, the problem of learning from imbalanced data streams is considered. To solve this problem, an approach is presented based on the processing of data chunks, which are formed using over-sampling and under-sampling. The final classification output is determined using an ensemble approach, which is supported by the rotation technique to introduce more diversification into the pool of base classifiers and increase the final performance of the system. The proposed approach is called Weighted Ensemble with one-class Classification and Over-sampling and Instance selection (WECOI). It is validated experimentally using several selected benchmarks, and some results are presented and discussed. The paper concludes with a discussion of future research directions.

Keywords: Streaming Machine Learning · Data Stream · Imbalanced Data · Ensemble Learning · Rotation-Based Ensembles

1 Introduction

Nowadays, enormous amounts of data are generated by different devices, such as smartphones, wearables, computers, and Internet of Things (IoT) sensors, as well as by systems within businesses and corporations. These data have a massive character and are generated continuously, and can be seen as unbounded flows of data, which are known as data streams [1]. They can also be considered as a potential source of input to different systems, for example to support decision-making processes and in the use of machine learning algorithms for producing classifier models. However, such data streams pose challenges when used for machine learning.

One such challenge is to adapt the learning process to the presence of concept drift. Most of the data that are currently of interest have a non-stationary character and change over time, whereas traditional machine learning algorithms are not designed to monitor these changes. This means that machine learning tools are needed that can build models that evolve over time and are able to cope with concept drift in data. Thus, it is necessary to design machine learning algorithms that are able to adapt to these changes. In [5], it was highlighted that a data stream computational model can process only a small portion of the data, as it is not possible to keep and store all of the incoming data. In other words, the main challenge faced by machine learning algorithms is the ability to

process these streams, as traditional approaches are based on processing in batch mode. Updating the classification model when new data are input poses a further challenge. An appropriate solution to this problem is the use of incremental or online learning. It should be also underlined that when processing data streams, attention needs to be paid to the class imbalance problem, which can have a very detrimental influence on the learning process [2]. In summary, the answer to these challenges is found in the domain of streaming machine learning [3] or online machine learning [5], where the main aim is to design machine learning models that are able to work with streaming data sources. This is currently a very hot topic, and is still open to research.

One potential approach that is of interest in the area of streaming machine learning is based on ensemble learning. The motivation for using ensemble learning was discussed in [11], where the authors very clearly showed that no single classifier is appropriate for all tasks, and that an ensemble approach to online learning may be suitable for dealing with the restrictions arising from drifting data, the class imbalance problem or the availability and character of the data stream. An ensemble learning method is resilient and adaptable to changes in data [11].

In this paper, the problem of learning from an imbalanced data stream is considered, and the classification problem forms the main area of interest of this research work. To solve this problem, a dedicated framework is considered, which was originally presented in [4]. This framework has features that make it suitable for processing streaming data, as mentioned above. The assumptions on which this dedicated and proposed framework is founded are as follows:

- The process of classifier learning is based on data chunks, which are formed from incoming instances. These data chunks are also updated over time and when new instances are input. The size of the data chunk is limited by the settings of the input parameters of the system.
- The classification problem considered here is based on the decomposition of a multi-class classification problem into a set of subproblems involving one-class classification. The final decision output is produced using a weighted ensemble classification model.
- The problem of imbalanced class distribution is eliminated through the use of over-sampling (i.e. synthetic instance generation) and under-sampling (i.e. instance selection) techniques. These techniques are applied to create the data chunks that are used in the next step to create an ensemble of classifiers.
- The system implemented based on this framework can be adapted to work with concept drift.

The approach summarised above, called Weighted Ensemble with one-class Classification and Over-sampling and Instance selection (WECOI), is based on ensemble learning. However, this paper extends the original approach and focuses on providing more diversity to the ensemble and improving the quality of the WECOI system. The research question addressed in this paper concerns the possibility of increasing the performance of the WECOI system by introducing a rotation-based technique to the stage in which the ensemble classifier is formed.

The remainder of this paper is organised as follows. In Sect. 2, a review of selected schemes in the literature is presented, and the motivation for this work is discussed.

A general framework for the WECOI approach is presented in Sect. 3. In Sect. 4, the rotation-based technique is introduced as an extension of WECOI. The results of computational experiments are presented and discussed in Sect. 5. Some conclusions and future directions for research are given in the final section.

2 Related Work and Motivation

The problem of streaming machine learning is very current one in the literature. A broad discussion of the more important aspects of streaming data processing is given in [6]; the paper also discusses current research opportunities, different courses of action and phenomena that are important with respect to the processing of data streams, and ways of improving the performance of this learning method. The aspects described in this paper include feature transformation (summarisation sketches, feature scaling, feature discretisation), dimensionality reduction, feature selection, ensemble learning, imbalanced learning, detection of concept drift, feature evolution, concept evolution, and the evolution of data sources. A fundamental discussion of the problem of streaming machine learning is presented.

The authors of [1] discussed the different types of streaming data with respect to concept drift. Different forms of concept drift were reviewed, and the speed of drift was discussed. Although the main research thread followed in this work was the generation of synthetic data in order to eliminate the phenomenon of class imbalance, these authors' research results clearly also show that learning from streaming data can be effective when ensemble learning is applied. Several different approaches for streaming machine learning were compared, and the one conclusion was that learning based on ensemble methods can help in eliminating the negative impacts of imbalanced classes in a data stream.

In general, it has been experimentally shown that better results from the machine learning process, as well as better classification performance, can be achieved by using multiple classification methods. The combination of multiple different sampling algorithms and types of classifiers results in a competitive heterogeneous ensemble classifier. When the learning process is merged with the perturbation of instances and classifiers, this provides more diversity to the final ensemble-based classifier. It has also been shown that the effects of increasing the quality of learning based on an ensemble approach can be especially crucial when the data are imbalanced with respect to the class distribution. A review and an extensive study of the different approaches to imbalanced data classification were included in [7], and different heterogeneous and multi-balancing ensemble architectures were discussed. The paper also considered a weighted selection of classification methods in which various classifier distributions were taken into account.

A weighted ensemble learning approach was proposed in [8] with the aim of balancing diversity and accuracy in ensemble learning. This approach was based on the use of the particle swarm optimisation (PSO) algorithm, with the aim of manipulating the datasets and input features and obtaining a set of individual learners with appropriate diversity. In an earlier paper, the authors of [9] considered the comparable problem of balancing accuracy and diversity in ensemble learning using an artificial bee colony approach.

A heuristic dataset modification to obtain a diverse classifier ensemble was considered in [10]. Through computational experiments using different strategies, including those based on artificially generated training samples, it was shown that it was possible to significantly increase the generalisation performance of the classification models.

A summary of different types of ensemble approaches for data stream classification was also presented in [11]. The authors underlined that although many important research results have been reported in the field of streaming machine learning, and approaches based on ensemble learning are promising and can eliminate most of the problems of streaming machine learning, there are still a number of open research problems and challenges for learning ensembles from data streams. Finally, potential directions for future research in this domain were formulated.

The present research work was undertaken in answer to new questions related to learning from data streams and with the aim of increasing the performance of WECOI through the diversification of ensemble classifiers, which represented some of the research challenges identified in [11]. Thus, the main motivation of the work was to increase the performance of WECOI by diversifying the ensemble classifiers and introducing a rotation technique at the learning stage.

3 The WECOI Approach to Learning from Imbalanced Data

The implementation, background and some initial evaluation results were discussed earlier in [4]. As mentioned above, WECOI is based on several assumptions, and involves the three components of data summarisation, learning, and classification.

In WECOI, data processing is carried out in the form of data chunks. These data chunks are formed from incoming instances, and are created independently for each decision class. The data summarisation step is responsible for forming the data chunks extracted from the classification component. This component is also responsible for continuous updating of the data chunks by selecting suitable instances (called prototypes) from the incoming data, memorising them and forgetting/removing other (non-informative) instances from current data chunks. It also generates synthetic instances when necessary.

The updating of each data chunk means that new instances covered by this new data chunk are directed to the learning component, which is responsible for creating a new base classifier for this data chunk. The learning process is carried out independently for each decision class of the problem.

In general, it is assumed that the size of the data chunks is limited, and is determined by a predefined threshold. When the number of instances in a data chunk is equal to this predefined threshold, all new incoming instances trigger a process of updating the data chunk, which is carried out by the data reduction module. From an algorithmic point of view, the data reduction process is applied to the instances from both the current data chunk and the new incoming instances. The aim of the data reduction step is to eliminate from the set of instances those that do not meet the reduction criterion [12]. A further aim is to obtain an acceptable level of data compression, which results in a target data chunk with a size equal to the defined threshold. In general, this process may result in the data chunk being updated by removing some old instances and adding to them the new

incoming data. It is also possible that this process will not change the composition of the data chunk, if these new instances do not meet the quality criteria. In this step, various strategies for data reduction can be applied; for example, in [4], the condensed nearest neighbour (CNN) and edited nearest neighbour (ENN) algorithms were implemented¹.

When the size of the data chunk is smaller than the predefined threshold, all new instances corresponding to the decision class of the data chunk are allocated to it. However, one existing problem relates to completing a data chunk when the other data chunks are already complete. Without completing the data chunk, a situation arises where the learning process could proceed without finding a balance between the instances belonging to the considered classes, which is a feature of class imbalanced data. To obtain a more balanced distribution between the minority and majority instances, an over-sampling procedure is applied, with the aim of generating a synthetic instances within the set of the instances belonging to the minority class. Various approaches to the over-sampling procedure can be applied, including the well-known SMOTE algorithm and its different versions, including those dedicated to working with imbalanced data streams (see, for example, [1] and [7]). A new approach dedicated to the generation of synthetic instances for a chunk-based ensemble was presented in [4].

To sum up, the updating of each data chunk means that each newly formed data chunk is directed to the learning component, which carries out the process of induction of a new base classifier for such a new data chunk. The learning process is carried out independently for each decision class considered.

Data chunks represent collections of instances belonging to the different decision classes of the problem. Later data chunks are processed, merged, reduced and so on, to each decision class with the aim of inducing the base classifiers responsible for solving the given one-class classification problem. This process also aims to prepare suitable sets of instances, where one consists of positive instances and the other negative. The process is based on data reduction, and is applied to the negative instances with aim of compressing this subset to a level equal to that of the set of positive instances.

The learning component carries out continuous updating of a set of base classifiers. This also means that a base classifier is induced from each data chunk and that induction is carried out independently for each considered decision class. This approach was developed as a result of considering the single multi-class classification problem as a set of one-class classification problems. In other words, a multi-class classification problem is solved using an ensemble of single one-class classifiers, one for each target class. This also means that a pool of simple base classifiers is induced.

WECOI is also based on an ensemble of classifiers, consisting of base classifiers for the current step and base classifiers representing τ earlier steps, with respect to data chunks that no longer exist in the system and which have been forgotten. This also means that the ensemble consists of a fixed-size set of classifiers, depending on the value of τ , where τ is a WECOI parameter set by the user. The set of classifiers in WECOI (as originally described in [4]) can be denoted by the matrix Φ , consisting of $d \times \tau$ elements.

¹ In [4], it was shown that the use of ENN was superior to CNN, which held true for all of the datasets considered.

We have d one-class classifiers, one for each target class, which can be denoted as:

$$\Phi = \begin{bmatrix} \varphi_{t-\tau}^1 & \cdots & \varphi_{t-1}^1 & \varphi_t^1 \\ \vdots & \ddots & \vdots & \vdots \\ \varphi_{t-\tau}^d & \cdots & \varphi_{t-1}^d & \varphi_t^d \end{bmatrix}, \quad (1)$$

where t is the current step and the classifiers $\varphi_t^l (l = 1, \dots, d)$ are induced from the set of positive instances included in the arriving data chunks.

The ensemble given above is updated each time a new data chunk arrives. Equation 1 clearly shows that classifiers induced earlier are forgotten, and only τ base classifiers are considered when forming a prediction result for new, incoming instances, which is done by the classification component.

The prediction result produced by the ensemble classifier is determined through a weighted majority vote, which can be expressed as:

$$\Phi(x) = \operatorname{argmax}_l \sum_{i=1}^{\tau} \sum_l^d w(\varphi_i^l) (\varphi_i^l(x) = c_l). \quad (2)$$

Here, $w(\varphi_i^l)$ are the weights assigned to each base classifier, which are calculated [13] as follows:

$$w(\varphi_i^l) = \frac{\Lambda(\varphi_i^l)}{\sqrt{z}}, \quad (3)$$

where $\Lambda(\varphi_i^l)$ are the values for the frequency of correct class predictions by the classifiers φ_i^l , and z ($z < \tau$) defines the number of iterations for which each φ_i^l stayed in the ensemble. The value of a classifier's weight increases if the classifier has been taking the correct decisions.

4 Rotation-Based Technique for WECOI

The aims of this paper are to extend WECOI and to improve the quality of its performance by introducing a rotation-based technique. This technique belongs to a family of methods for increasing the diversification of the ensemble. A rotation-based technique provides the possibility of combining several machine learning techniques into one prediction model [14, 15].

The main idea of a rotation-based ensemble is to transform a original dataset into a new feature space through feature extraction. In the first step, the original data are split into subsets, the number of which is defined as a parameter of the approach (referred to here as a rotation level). Splitting is carried out in a feature dimension. In the second stage, feature extraction is carried out to generate the various individual classifiers. The dataset is divided in to two subsets using a bootstrap method, where the larger one includes 75% of the instances from the original dataset. Classifiers are induced based on this larger subset, and the smaller one is used as a test set.

In the original description of the method, a rotation-based ensemble applied principal component analysis (PCA) for feature extraction. However, there are other approaches

to feature extraction that can be applied to rotation-based ensembles, including maximum noise fraction (MNF), independent component analysis (ICA), and local Fisher discriminant analysis (LFDA) (see, for example, [16]).

This paper reports results obtained from using PCA and ICA. However, it should be noted that rotation is applied to data chunks that are adequately prepared for the induction of base classifiers and are well class-balanced.

Finally, an ensemble of classifiers, in which each base classifier is constructed using different training parameters (a set of instances diversified in a feature space), can be obtained. The procedure used for data processing within the data summarisation component for the current learning step is presented in Algorithm 1.

Algorithm 1. Rotation-based technique for WECOI

Input: D_1, \dots, D_d - data chunks with features described by a feature set A , where d , is a number of decision classes, and k is the value of the rotation level;

Output: φ_j^l - set of the base classifiers, where $l=1, \dots, d$ and $j=1, \dots, k$;

Begin

For $l=1$ **to** d **do**

Let $D'_l = D - D_l$

Apply data reduction to D'_l and obtain the reduced set of instances D'_l

Let $T_1 = D_1 \cup D'_l$

Randomly partition the feature set A into k subsets to obtain subsets T_l^1, \dots, T_l^k of instances, each with identical numbers of features. These are smaller than the number of features in the original dataset, and are described by the feature sets A_l^1, \dots, A_l^k .

For $j=1$ **to** k **do**

Generate the training set $T_l^{j'}$ through bootstrapping based on 75% of the original dataset, and create the testing set $T_l^{j''}$ from the remaining examples in the original set;

Apply **PCA** or **ICA** to the transformed $T_l^{j'}$ to produce new training datasets $rT_l^{j'}$, with features $A_l^{j'}$, using axis rotation;

Induce a base classifier φ_j^l based on $rT_l^{j'}$ and using a test set $T_l^{j''}$ described by a set $A_l^{j'}$.

End for

End for

Return φ_j^l , where $l=1, \dots, d$ and $j=1, \dots, k$;

End

Algorithm 1 shows how the base classifiers φ_j^l are produced. However, WECOI is based on a pool of base classifiers, which are induced in the current step and in earlier steps, remembered over τ iterations. Thus, the final prediction result is produced based on the following expression (see also Eq. 3):

$$\Phi(x) = \operatorname{argmax}_l \sum_{i=1}^{\tau} \sum_l^d \sum_j^k w(\varphi_{ij}^l) (\varphi_{ij}^l(x) = c_l). \tag{4}$$

5 Results of Computational Experiments

To validate the proposed approach, an computational experiment was carried out. The main goal was to answer the following question: can the rotation technique increase the performance of the WECOI approach, on average?

To assess the results reported in this section, the following versions of WECOI and other algorithms are compared:

- OLP (Online Learning based on Prototypes): This is the first version of the framework discussed in this paper. It is based on a simple ensemble model in which ensembles are updated by removing the oldest classifier; however, it is devoid of mechanisms for the decomposition of a multi-class classification problem into a set of subproblems involving one-class classification [17];
- WECU (Weighted Ensemble with one-class Classification and data chunk Updating): This is an earlier version of the framework discussed here, which is based on the use of an under-sampling procedure alone to form the data chunks [17];
- WECOI: This is earlier version of the proposed framework, based on a weighted ensemble with one-class classification, over-sampling and instance selection. An ENN-based instance selection approach was used in the under-sampling procedure [4];
- WECOI': This is an earlier version of the framework discussed here, which is based on an extended procedure for synthetic instance generation using a similarity-based approach for clustering and the ENN algorithm for under-sampling [18];
- WECOI⁽¹⁾: This an extended version of the WECOI'-based framework in which PCA is used for feature space rotation at the ensemble classifier formation stage (an approach proposed in this paper);
- WECOI⁽²⁾: This is a WECOI'-based framework that was extended by using the ICA for feature space rotation at the ensemble classifier formation stage (an approach proposed in this paper);
- OUOB (over-sampling and under-sampling online bagging) [21];
- OB (online bagging) [20],
- The Learn + +.NIE algorithm² [22],
- AWE (accuracy weighted ensemble) [23],
- HOT (Hoeffding option tree) [19];
- iOVFDT (incrementally optimised very fast decision tree [19]).

In all versions of WECOI, a POSC4.5 algorithm was applied to generate the base classifiers. In the case of OLP, the C4.5 algorithm was applied to induce all the base models for the ensemble classifiers. All versions of WECOI were run with the number of nearest reference vectors (a parameter for synthetic instance generation) set to two, and the number of neighbours for ENN was set to three. The number of base classifiers and the rotation level were set to five. These values were set arbitrarily, using a trial and error procedure.

The computational experiment was performed using synthetic and real datasets. The data benchmarks containing real data streams are referred to as Electricity (45,312

² For OUOB, OB and Learn + +.NIE, the Hoeffding tree was used for the base classifiers, and the base classifier pool was set to 10.

instances, eight features, two classes); Airlines (539,383 instances, seven features, two classes); Ozone (2,534 instances, 72 attributes, two classes); and Gas sensor array (13,910 instances, 128 features, six classes). All of these datasets were taken from the UCI Machine Learning Repository and the IDA repositories. For the synthetic data, the MOA framework was used [19], and the following generators were applied: SEA (10% noise, sudden concept drift and no class balancing); HYPERPLANE (standard MOA parameters, rotation of the decision boundary, and the incremental concept drift set to 0.01); and AGRAWAL (sudden concept drift, and no class balancing). In the synthetic dataset, the number of instances was set to 10,000,000, with two decision classes, and the number of attributes was set to 10, with all attributes subject to drift. The threshold that defined the size of the data chunks for the synthetic data, electricity, airlines and gas sensor array sets was set to 1,000 instances, whereas for the ozone dataset it was set to 250.

The performance of the algorithms was compared with respect to the classification accuracy, and the results are shown in Table 1 as average values. The performance was evaluated using the test-then-train procedure. Average values were calculated after 30 repetitions on each data benchmark.

The results in Table 1 show that no one algorithm was the best for all of the benchmark datasets; however, it can be concluded that introducing more diversification between the base classifiers in the ensemble can yield promising results. In general, introducing the rotation technique to achieve this diversification resulted in an increase in the performance of WECOI. This conclusion was drawn from an analysis of the results obtained from different versions of the WECOI, including the versions denoted as WECOI⁽¹⁾ or WECOI⁽²⁾. It should be also noted that PCA resulted in better accuracy.

It can be inferred that WECOI is competitive with the other algorithms in the literature that are suitable for use in learning from an imbalanced data stream.

Table 1. Average accuracy (in %)

ALGORITHM	ELECTRICITY	AIRLINES	OZONE	GAS	HYPERPLANE	AGRAWAL	SEA
OLP	70,52	60,58	78,92	80,08	75,48	82,5	80,67
WECU	75,31	65,42	80,56	82,61	83,46	88,67	83,22
WECOI	76,5	64,82	81,56	84,62	85,78	91,68	88,45
WECOI'	76,7	63,48	82,21	84,7	85,91	92,71	88,79
WECOI ⁽¹⁾	77,31	64,72	82,62	83,84	86,49	92,54	87,42
WECOI ⁽²⁾	76,54	63,43	82,46	83,15	85,78	92,05	86,9
OUB	75,42	65,42	82,31	83,42	72,27	92,48	86,69
OB	77,66	63,21	76,56	82	84,7	81,67	83,74

(continued)

Table 1. (continued)

ALGORITHM	ELECTRICITY	AIRLINES	OZONE	GAS	HYPERPLANE	AGRAWAL	SEA
LEARN++ NIE	70,7	66,8	82,1	83,5	84,51	92,42	89,3
AWE	71,06	63,4	66,54	82,42	70,15	80,2	77,81
HOT	74,02	62,34	81,03	82,6	80,07	94,99	88,03
iOVFDT	72,52	63,52	81,25	83,54	81,56	93,67	83,02

* The results presented on a gray background refer to the algorithms discussed in this paper

6 Conclusions

The main contribution of this paper is the presentation of a competitive approach to learning from imbalanced data streams. To address this problem, an approach was developed based on the processing of data chunks formed from incoming instances using over- and under-sampling. These procedures proved to be beneficial for the processing of imbalanced data. The proposed approach was also based on the assumption that the general problem could be solved by processing a set of independent one-class classification problems. The approach presented here can help in the recognition of concept drift, which is a challenge that arises when learning from data streams. The final output is calculated based on an ensemble approach, which in this study was extended by increasing the diversification between base classifiers by introducing a rotation in feature space in the data. This rotation offers benefits in terms of classification accuracy, as can be seen from a comparison of different alternative approaches to learning from data streams and earlier versions of the algorithm presented here.

The approach discussed here, called WECOI, needs to be further evaluated, particularly in terms of the influence of different parameters on its performance, including the rotation level. Important questions are also related to the impact of different rotation techniques on the performance of WECOI, and may provide directions for future research.

References

1. Bernardo, A., Valle, E.D.: An extensive study of C-SMOTE, a continuous synthetic minority oversampling technique for evolving data streams. *Expert Syst. Appl.* **196**, 116630 (2022). <https://doi.org/10.1016/j.eswa.2022.116630>
2. Khamassi, I., Sayed Mouchaweh, M., Hammami, M., Ghédira, K.: Discussion and review on evolving data streams and concept drift adapting. *Evol. Syst.* **9**(1), 1–23 (2018)
3. Shreya, S., Bernease, H., Aditya, G.P.: Rethinking streaming machine learning evaluation. *arXiv* (2022). <https://doi.org/10.48550/arxiv.2205.11473>
4. Czarnowski, I.: Weighted ensemble with one-class classification and over-sampling and instance selection (WECOI): an approach for learning from imbalanced data streams. *J. Comput. Sci.* **61**(1), 101614 (2022). <https://doi.org/10.1016/j.jocs.2022.101614>
5. Benczúr, A.A., Kocsis, L., Pálovics, R.: Online machine learning in big data streams. *arXiv* (2018). <https://doi.org/10.48550/ARXIV.1802.05872>

6. Gomes, H.M., Read, J., Bifet, A., Barddal, J.P., Gama, J.: Machine learning for streaming data: state of the art, challenges, and opportunities. *ACM SIGKDD Explor. Newsl.* **21**(2), 6–22 (2019). <https://doi.org/10.1145/3373464.3373470>
7. Ghaderi-Zefrehi, H., Altınçay, H.: Imbalance learning using heterogeneous ensembles. *Expert Syst. Appl.* **142**, 113005 (2020). <https://doi.org/10.1016/j.eswa.2019.113005>
8. You, G.-R., Shiue, Y.-R., Yeh, W.-Ch., Chen, X.-L., Chen, Ch.-M.: A weighted ensemble learning algorithm based on diversity using a novel particle swarm optimization approach. *Algorithms* **13**(10) 255 (2020). <https://doi.org/10.3390/a13100255>
9. Shiue, Y.-R., You, G.-R., Su, Ch.-T., Chen, H.: Balancing accuracy and diversity in ensemble learning using a two-phase artificial bee colony approach. *Appl. Soft Comput.* **105**, 107212 (2021). <https://doi.org/10.1016/j.asoc.2021.107212>
10. Jamalnia, H., Khalouei, S., Rezaie, V., Nejatian, S., Bagheri-Fard, K., Parvin, H.: Diverse classifier ensemble creation based on heuristic dataset modification. *J. Appl. Stat.* **45**(7), 1209–1226 (2018). <https://doi.org/10.1080/02664763.2017.1363163>
11. Krawczyk, B., Minku, L.L., Gama, J., Stefanowski, J., Woźniak, M.: Ensemble learning for data stream analysis: a survey. *Inf. Fusion* **37**, 132–156 (2017). <https://doi.org/10.1016/j.inf.fus.2017.02.004>
12. Czarnowski, I., Jędrzejowicz, P.: An approach to data reduction for learning from big datasets: integrating stacking, rotation, and agent population learning techniques. *Complexity* **7404627**, 1076–2787 (2018)
13. Wozniak, M., Cal, P., Cyganek, B.: The influence of a classifiers' diversity on the quality of weighted again ensemble. In: Nguyen, N.T., Attachoo, B., Trawinski, B., Somboonviwat, K. (eds.) *ACIIDS 2014*. LNAI, vol. 8398, pp. 90–99. Springer, Heidelberg (2014). https://doi.org/10.1007/978-3-319-05458-2_10
14. Adachi, K.: Rotation techniques. In: Adachi, K. (ed.) *Matrix-Based Introduction to Multivariate Data Analysis*, pp. 193–205. Springer, Singapore (2020). https://doi.org/10.1007/978-981-15-4103-2_13
15. Rodríguez, J.J., Alonso, C.J.: Rotation-based ensembles. In: Conejo, R., Urretavizcaya, M., Pérez-de-la-Cruz, J.-L. (eds.) *CAEPIA/TTIA -2003*. LNCS (LNAI), vol. 3040, pp. 498–506. Springer, Heidelberg (2004). https://doi.org/10.1007/978-3-540-25945-9_49
16. Xia, J.: Multiple classifier systems for the classification of hyperspectral data. Ph.D. thesis, University de Grenoble (2014)
17. Czarnowski, I., Jędrzejowicz, P.: Ensemble online classifier based on the one-class base classifiers for mining data streams. *Cybern. Syst.* **46**(1–2), 51–68 (2015). <https://doi.org/10.1080/01969722.2015.1007736>
18. Czarnowski, I., Martins, D.M.L.: Impact of clustering on a synthetic instance generation in imbalanced data streams classification. In: Groen, D., de Mulatier, C., Paszynski, M., Krzhizhanovskaya, V.V., Dongarra, J.J., Sloot, P.M.A. (eds.) *ICCS 2022*. LNCS, vol. 13351, pp. 586–597. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-08754-7_63
19. Bifet, A., Holmes, G., Kirkby, R., Pfahringer, B.: MOA: massive online analysis. *J. Mach. Learn. Res.* **11**, 1601–1604 (2010)
20. Oza, N.C.: Online bagging and boosting. In: *Proceedings of the 2005 IEEE International Conference on Systems, Man and Cybernetics*, Waikoloa, HI, USA, 10–12 October 2005, vol. 2343, pp. 2340–2345 (2005)
21. Wang, S., Minku, L.L., Yao, X.: Dealing with multiple classes in online class imbalance learning. In: *Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI 2016)* (2016)

22. Ditzler, G., Polikar, R.: Incremental learning of concept drift from streaming imbalanced data. *IEEE Trans. Knowl. Data Eng.* **25**(10), 2283–2301 (2013). <https://doi.org/10.1109/TKDE.2012.136>
23. Wang, H., Fan, W., Yu, P.S., Han, J.: Mining concept-drifting data streams using ensemble classifiers. In: *Proceedings of 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 226–235 (2003). <https://doi.org/10.1145/956750.956778>