Improving Online Bagging for Complex Imbalanced Data Streams

Bartosz Przybył[0009−0005−3912−9172] and Jerzy Stefanowski[0000−0002−4949−8271]

Institute of Computing Science, Poznań University of Technology, Poznań, Poland
bprzybyl98@gmail.com, jerzy.stefanowski@cs.put.poznan.pl

Abstract. Learning classifiers from imbalanced and concept drifting data streams is still a challenge. Most of the current proposals focus on taking into account changes in the global imbalance ratio only and ignore the local difficulty factors, such as the minority class decomposition into sub-concepts and the presence of unsafe types of examples (borderline or rare ones). As the above factors present in the stream may deteriorate the performance of popular online classifiers, we propose extensions of resampling online bagging, namely Neighbourhood Undersampling or Oversampling Online Bagging to take better account of the presence of unsafe minority examples. The performed computational experiments with synthetic complex imbalanced data streams have shown their advantage over earlier variants of online bagging resampling ensembles.

Keywords: data streams · class imbalance · data complexity · online bagging

1 Introduction

Although learning classifiers from concept drifting data streams has been intensively studied in the last decades it is still the subject of many new studies, including better consideration of various types of concept drifts or other data complexity [3,7]. This applies in particular to class imbalances in the stream, which also occur in some practical applications [1,6].

However, the current research on imbalanced and concept drifting streams is still not as developed as in the case of separately considered static data or streams. Moreover, the existing works mostly focus on re-balancing classes and reacting to changes caused by the *varying global imbalance ratio*. These works do not sufficiently consider more complex imbalanced stream scenarios, where these changes are additionally accompanied by the *local difficulty factors* already considered for the static imbalanced data such as the minority class split into sub-concepts, class overlapping, or occurrence of different types of unsafe minority examples (borderline, rare or outliers [11]). In evolving data streams these factors could also influence the changes in *local class distributions* and lead to distinguishing new types of *local drifts* in the stream [3,4].

Up to now, only two experimental studies have been conducted on the role of these local difficulty factors combined with other types of concept drifts [3,10]. The authors of [3] carried out a large series of comprehensive experiments with synthetic and real data streams showing the different influence of types of minority examples and class decomposition on predictions of representative online classifiers. They showed that especially re-sampling generalizations of online bagging coped sufficiently well with changes in the global class imbalance ratio in the stream and partially with the split of the minority class into smaller sub-concepts. However, drift associated with the greater presence of unsafe examples (especially rare examples) caused a deterioration in the performance of all tested online classifiers. Furthermore, combinations of multiple factors were demonstrated to be the most challenging for all classifiers, which were not able to recover from these drifts [3]. The experiments of [3] concerned binary imbalanced classes, however later their main observations were confirmed also for multiple classes in [10].

Therefore, the aim of this work is to propose new generalizations of under-sampling or over-sampling online bagging ensembles that would better handle the drifts of unsafe types of examples in such complex imbalanced data streams. The other aim is to evaluate them experimentally on multiple synthetic data and compare them with earlier versions of online re-sampling baggings.

2   Related Works

Due to the page limits the reader is referred to works such as [9] for a review of research on data streams or to [2,6,8] for imbalanced data. In the case of imbalanced and concept-drifting data streams, refer to [1,3,4].

In the case of static imbalanced data, it has been already shown that besides the global imbalance between classes other sources of classifier deterioration include: (1) the decomposition of the minority class into several sub-concepts, (2) the presence of small, isolated groups of minority examples located deeply inside the majority class region (it corresponds to *rare cases*), (3) the effect of strong overlapping between the classes.

The last two factors can be identified through, the so-called, *types of examples* [11], which distinguish between safe and unsafe examples. *Safe examples* are the ones located in homogeneous regions populated by examples from one class only. The *unsafe* examples are categorized into *borderline* (placed close to the decision boundary between classes), *rare cases* (isolated groups of few examples located deeper inside the opposite class), and *outliers* - singletons. See Figure 1 for their illustration. Following the method from [11] these types of examples can be identified based on the analysis of class labels of other examples in the local neighbourhood of the given instance. Most studies on imbalanced data streams focus on handling either static or drifting imbalance ratios only and they do not consider the aforementioned data factors.

The authors of [3] introduced an *extended concept drift categorization from imbalanced streams* that takes into account these local data factors and covers
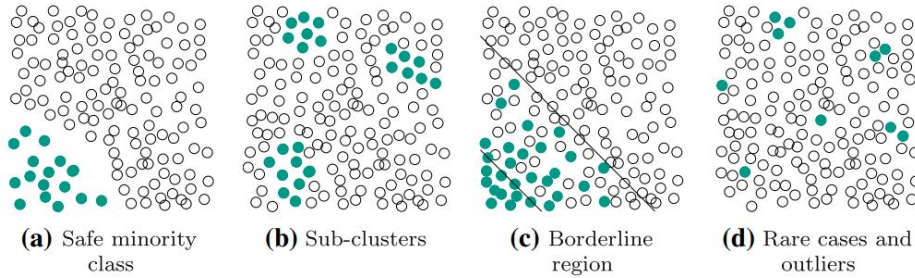
(a) Safe minority class   **(b)** Sub-clusters   **(c)** Borderline region   **(d)** Rare cases and outliers

Fig. 1. An illustration of different difficulty factors in imbalanced data - based on [4]

specific types of drifts inside them. These authors prepared a special generator of the synthetic data streams, where various scenarios of the occurrence of these elements are modeled - see its description in an electronic appendix of [3].

It was observed in [3] that specialized imbalanced stream classifiers (ie. two re-sampling versions of online bagging, i.e. over-sampling the minority class OOB or under-sampling the majority class UOB [14], and a specialized neural network ESOS-ELM) coped well with nearly all static and dynamically changing imbalance ratios. Non-specialized online classifiers (such as OB and VFDT) performed much worse, especially when the minority class ratio dropped to 1–5%. The rest of the experiments demonstrated that the imbalance ratio can play an important (amplifying) role only when combined with other factors. Then, the other aspects of drift locality and minority class composition affected all the analyzed classifiers. All classifiers suffered moderate performance drops when exposed to moving sub-clusters and even more substantial performance drops and limited recovery when exposed to minority class splits. However, the next experiments demonstrated that changes in the distribution of types of minority examples from safe to unsafe ones (i.e. borderline or rare) were very influential, in particular for introducing more rare examples as a result of the drift. Drifting proportions of rare examples were the only drift the best classifiers did not recover from in any way. Furthermore, the effects of all the analyzed difficulty factors are amplified when they occur together as combined *multiple drifts*.

These observations for binary imbalanced classes were confirmed in a similar study for multi-class data [10], which also showed a greater impact of the presence of borderline examples.

3   Neighbourhood Online Bagging

In our work, we decided to extend under or over-sampling ensemble approaches based on online bagging (originally proposed in [14]) because they performed best in experiments with difficult imbalanced data streams [3,10].

Online bagging [12], in contrast to the static Breiman's bagging ensemble, is capable of dealing with online environments because it processes each example only once upon arrival. The idea of online bagging is based on exploiting the

Poisson distribution and its parameter λ to estimate how many times the current incoming example should be sent to each component classifier (a default $\lambda = 1$, which works well for approximately balanced classes in the stream). Each classifier can incrementally update (trained) with such received examples depending on this estimate.

Oversampling-based Online Bagging (OOB) and *Undersampling-based Online Bagging* (UOB) are generalizations proposed in [13,14] to change the presence of examples from the particular class in this Poisson distribution operation concerning the current imbalance ratio in the stream.

The number of examples from each class in the given time moment in the stream is continuously updated and used to calculate λ as a function of the current imbalanced ratios. The value in the nominator of λ depends on whether a classifier should over- or under-sampled for a currently considered concept present in the stream. In OOB the value of λ is equal to the cardinality of the current biggest class size divided by the class size of the currently processed example. It increases the number of minority class examples (as $\lambda > 1$) and does not change the number of majority ones $\lambda = 1$. In the UOB the value of λ is computed as the ratio of the current size of the smallest class in the stream and the class size of the current example. Practically, it ($\lambda < 1$) reduces the number of majority examples sent to component classifiers.

To sum up, the both re-sampling generalizations modify sending an incoming example from a specific class for updating component classifiers in online bagging with respect to the imbalance ratio between classes in the stream.

In our current proposal, we modify the λ coefficient by incorporating information about the difficulty of the incoming example. Previously [5] in order to estimate the difficulty of the example in the static data, the local analysis of class labels of its k nearest neighbours was exploited. Following this inspiration we propose to define *unsafeness level* of a minority class example x as

$$L^2_{min} = \frac{(N'_{maj})^{\Psi}}{k},\qquad(1)$$

where N'_{maj} is the number of examples belonging to the majority class among k nearest neighbours of x which are calculated on a sliding window in the stream, k is the number of nearest neighbours taken for the analysis, Ψ is a parameter responsible for additional amplification of the impact of unsafe examples (i.e. value $\Psi > 1$ amplifies the role of rare cases and outliers). In the *Neighbourhood Oversampling Online Bagging* (NOOB) this coefficient is aggregated with the class sizes coefficient to increase the number of unsafe minority examples send to update the component classifiers, which is defined as follows:

$$\lambda = (N_{maj}/N_{min}) \cdot (L^2_{min} + 1),\qquad(2)$$

where N_{maj} denotes the number of examples from the majority class contained in the sliding window, N_{min} denotes the number of examples from the minority class.

In (NOOB) it increases the Poisson distribution estimate of the number of minority examples while for the incoming majority examples $\lambda = 1$ (as in the standard online bagging). The general pseudocode of the Neighbourhood Oversampling Online Bagging is presented in Algorithm 1.

Algorithm 1 Neighbourhood Oversampling Online Bagging (NOOB)

Input: S: stream of examples
n: number of classifiers in ensemble
W: window of examples
k: number of nearest neighbours
Ψ: additional coefficient for calculating safe level
Output: $\mathcal{E}$: an ensemble of classifiers

1: **for all** examples $x \in S$ **do**
2:　　calculate safe level of incoming example $L^2_{min} = \frac{(N'_{maj})^{\Psi}}{k}$
3:　　**if** $x \in minority\ class$ **then**
4:　　　　$\lambda \leftarrow (N_{maj}/N_{min}) \cdot (L^2_{min} + 1)$
5:　　**else**
6:　　　　$\lambda \leftarrow 1$
7:　　**end if**
8:　　**for all** classifiers $C_i \in \mathcal{E}$ **do**　　　　　　$\triangleright$ C_i - single classifier of ensemble
9:　　　　set $l \sim Poisson(\lambda)$
10:　　　　update C_i using x, l times
11:　　**end for**
12:　　$W \leftarrow W \cup \{x\}$
13:　　if necessary remove outdated examples from W
14: **end for**

In similar way, we propose *the Neighbourhood Undersampling Online Bagging* (NUOB), where for the incoming majority examples we define a *safeness level* as

$$L^2_{maj} = \frac{N_{maj}}{k}, \tag{3}$$

and as its consequence a new coefficient for the majority example

$$\lambda = (N_{min}/N_{maj}) \cdot (L^2_{maj})^{\Psi}, \tag{4}$$

which reduces the chance of using unsafe majority examples to update component classifiers and having $\lambda = 1$ for incoming minority examples it leads to such undersampling which removes rare cases, outliers and partly borderline majority examples. Due to the page limits of this paper, we moved its pseudocode to the electronic appendix.[1]

---

[1] Pseudo-codes of NUOB and Hybrid ensembles are available in an online appendix
http://www.cs.put.poznan.pl/jstefanowski/pub/imbstreams.pdf

Finally, as observed in experiments sometimes NUOB is better than NOOB and sometimes, on the contrary, NOOB is better than NUOB, so we decided to propose their hybrid version.

In *Hybrid Neighbourhood Online Bagging* (HNOB), both ensembles NUOB and NOOB are trained in parallel for each incoming example and their classification performance is continuously assessed using an evaluation metric suitable for class imbalanced data (in our experiment *G–mean* of both classes). The ensemble that had a better evaluation measure value in the last evaluation is selected to make class prediction for the current example in the stream.

4   Experiments

The aim of the experiment is to evaluate the usefulness of the proposed Neighbourhood Online Bagging ensembles and their reaction to considered data difficulty factors and drifts, in particular concerning borderline and rare minority examples. This is why we compare the 6 following classifiers: *basic online bagging, undersampling online bagging, oversampling online bagging, neighbourhood undersampling online bagging, neighbourhood oversampling online bagging, and hybrid neighbourhood online bagging.*

We apply the same set of hyperparameters as in [3] (as it was already tuning for considered synthetic data streams). So, all bagging ensembles contain 15 component Hoeffdings' trees (with their standard parameters). UOB and OOB use the forgetting degree ($\theta = 0.9$). For all neighbourhood variants, we compute the (un)safeness levels with $k = 5$ neighbours over the last sliding window W of size equal to 500 examples. These values were found by a grid search ($k = 5, 7, 9$ and $W = 500$ till 2000) in preliminary a experiment. Similarly, we tune the exponent degree Ψ to 2.0, which means that we pay more attention to rare examples than to safe ones.

We carried out the experiments in a controlled framework based on synthetic generated data streams, where each data factor can be modeled and parametrized according to different planned scenarios. We use a generator designed and used for the earlier related experiments from [3]. Our experiments cover the following difficulty factors and drifts:

- *Imbalanced ratios* – into two scenarios of either static data or with a single drift with changing the imbalances ratio from 1% to fully balanced classes.
- *Types of minority examples* being either *borderline* (class overlap) or *rare* ones (other examples in the minority classes are generated to be safe ones), i.e. with the given percentages of their occurrence in each class: 20%, 40%, 60%, 80%, and 100%.
- *Changes in class composition*, i.e. local drifts of the split of each minority class into 3, 5 and 7 sub-clusters; or merging them and moving these numbers of sub-clusters.

Besides modeling the single factor in the given stream, we considered their pairs and occurrences of multiple factors, which previously [3] were identified as the most demanding.

Due to the high number of experiments and the limited number of pages, we only include the most important, representative results and use summaries of many comparative experiments using the non-parametric Friedman tests.[2]

Firstly, we examined the impact of **single factors**. As expected the static global imbalanced ratio or the drift of this ratio was well coped by all compared versions of the online bagging ensemble, except the standard online bagging for the stronger imbalance ratios of 1% and 2%. Let's recall that both OOB and UOB worked well for such scenarios in [3,10]. Moreover, our neighbourhood proposals are also not exploiting their additional mechanisms if the data streams do not contain other difficulty factors, so may work similarly to OOB or UOB (what was observed in our results).

More interesting was studying the impact of the minority class decomposition. For the minority class split into smaller sub-concepts NUOB was the best classifier with respect to Recall, while NOOB worked better than other compared classifiers (also for moving sub-concept) but with respect to G-mean (the right-hand plot of Figure 2).

However, the most important experiments included data streams with different proportions of borderline and rare examples or their drifts. For both static and drift cases we noticed better performance of the new proposed neighbourhood baggings in comparison to earlier UOB and OOB ones. In general, NOOB was the best classifier for the higher proportions of rare minority examples, while NUOB worked better for borderline examples. Exemplary reactions to such drifts are shown in Figure 2.

	OB	UOB	OOB	NUOB	NOOB		OB	UOB	OOB	NUOB	NOOB
start	0.9959	0.9959	0.9959	0.9959	0.9940	start	1.0000	1.0000	1.0000	1.0000	0.9941
pre-dirft	0.9971	0.9966	0.9971	0.9967	0.9966	pre-dirft	0.9976	0.9975	0.9976	0.9987	0.9959
post-drift	0.5303	0.5305	0.5310	0.5327	0.6173	post-drift	0.9634	0.9565	0.9619	0.9707	0.9482
end	0.5058	0.4973	0.5205	0.5035	0.6213	end	0.9930	0.9919	0.9914	0.9946	0.9877

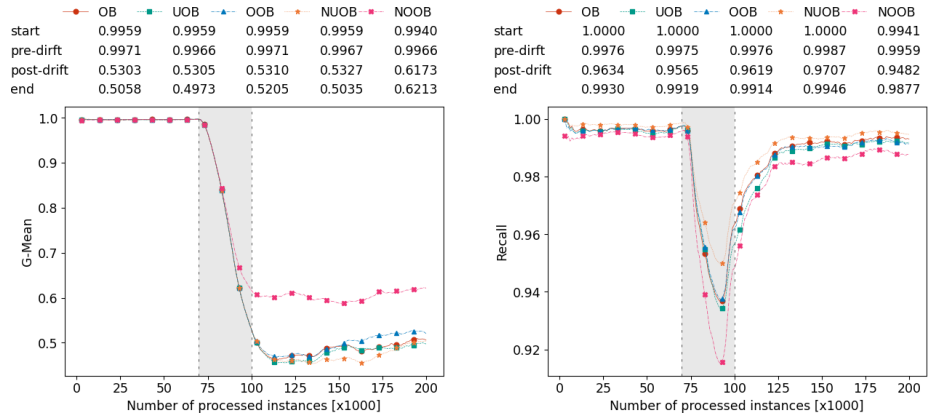
Fig. 2. Plots showing bagging variants reacting to two kinds of drift *80% rare minority examples and G-mean* measure (the left-hand figure) and *minority class split into 5 sub-concepts and Recall* measure (the right-hand figure)

---

[2] Additional results are provided in the online appendix to this paper, which is available at http://www.cs.put.poznan.pl/jstefanowski/pub/imbstreams.pdf

In Table 1 we summarize the results of Friedman tests carried out over aggregated results for single elements (static factors and drifts) for considered sub-categories concerning Recall measure. One can easily notice that nearly for all categories the proposed NUOB is either the best classifier or one of the best classifiers. In particular, for rare cases or borderline it is superior to the other classifiers. On the other hand, for G-mean measure performance NOOB is better (which indicates achieving a better balance of recognizing correctly both minority and majority classes) - in particular for different proportions of rare examples NOOB followed by OOB are the best classifiers (averaged ranks in the Friedman test are 1.9 and 2.1 respectively, while NUOB and UOB have 4.0 and 4.2 – for the critical difference $CD = 2.01$ – it is a significant dominance of NOOB.

Table 1. Average ranks in the Friedman test for different single factors in streams (the smaller, the better) – where classifiers evaluated by Recall metric

Data factor	OB	UOB	OOB	NUOB	NOOB	CD
Static imbalance	4.35	**2.41**	3.35	**1.00**	3.88	1.51
Class ratio changes	4.88	2.17	3.59	**1.00**	3.36	0.77
Sub-cluster merge	**2.67**	3.67	**2.67**	**1.00**	5.00	2.68
Sub-cluster move	**3.17**	3.17	**2.67**	**1.00**	5.00	2.68
Sub-cluster split	**2.50**	**2.50**	4.00	**1.00**	5.00	2.68
Borderline examples	**2.70**	3.80	**2.50**	**1.00**	5.00	2.01
Rare examples	**3.90**	**3.80**	**2.80**	**2.10**	**2.40**	2.01

Similar good performance of both proposed NUOB and NOOB classifiers was noticed for data streams with **pairs of factors**, in particular, if one of them was the presence of rare minority examples. An illustration of good reaction of NUOB to such drifts is presented in Figure 3 (the right-hand plot).

The usefulness of the new proposals is even more visible for the most difficult scenarios of generating **complex data streams** involving at least **three factors** occurring simultaneously. An example of such a reaction is shown in Figure 3 (the left-hand plot) for *StaticIm10+Im1+Rare80+Split5* combined factors. The global performance for different scenarios of such multiple factors is presented in Table 2.

Finally, the efficacy of Hybrid Neighbourhood Online Bagging was examined. The results showed its advantage over the best online bagging ensembles, in particular NUOB. Its good effectiveness was noticeable, especially for very complex scenarios involving **many factors** in the streams. An example of such a reaction is shown in Figure 4 for *StaticIm10+Im1+Rare80* and *Split5+Im1+Borderline40+Rare40* combined factors.
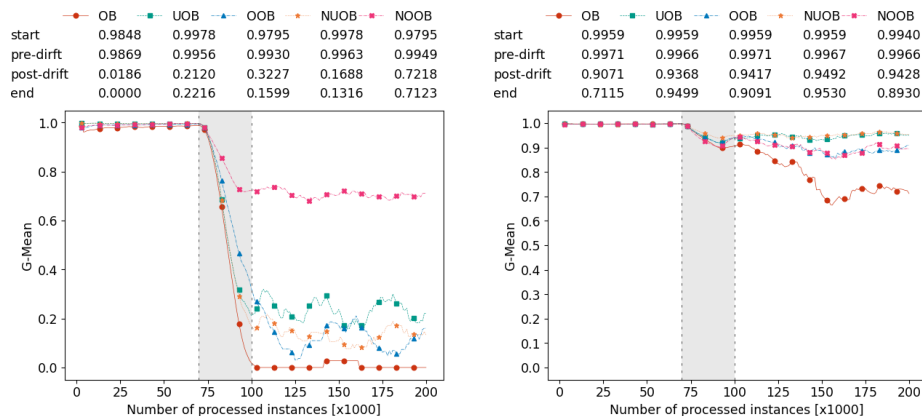
	OB	UOB	OOB	NUOB	NOOB			OB	UOB	OOB	NUOB	NOOB
start	0.9848	0.9978	0.9795	0.9978	0.9795		start	0.9959	0.9959	0.9959	0.9959	0.9940
pre-dirft	0.9869	0.9956	0.9930	0.9963	0.9949		pre-dirft	0.9971	0.9966	0.9971	0.9967	0.9966
post-drift	0.0186	0.2120	0.3227	0.1688	0.7218		post-drift	0.9071	0.9368	0.9417	0.9492	0.9428
end	0.0000	0.2216	0.1599	0.1316	0.7123		end	0.7115	0.9499	0.9091	0.9530	0.8930

Fig. 3. Plots showing *G-mean* measure of bagging variants reacting to two kinds of drift *80% rare minority examples and Split 5 and imbalanced ratio changing from 10% to 1%* (the left-hand figure) and *imbalance ratio 10% and 80% borderline minority examples* (the right-hand figure)

Table 2. Average ranks in Friedman test for multiple factors scenarios vs. all examined factors

Factors	Metric	OB	UOB	OOB	NUOB	NOOB	CD
Multiple	G-mean	4.60	2.69	3.27	**2.30**	**2.14**	0.43
All		4.39	**2.56**	2.99	**2.58**	**2.48**	0.31
Multiple	Recall	4.67	2.61	3.76	**1.79**	**2.17**	0.43
All		4.58	2.60	3.62	**1.53**	2.67	0.31

Furthermore, we summarize in Table 3 the results of the Friedman test for multiple vs. all studied factors in synthetic data streams. As one can notice Hybrid Neighbourhood Online Bagging is the best-performing classifier. For the G-mean measure and multiple factors, its performance clearly outperformed all compared variants of bagging (the post hoc analysis and critical difference support for it). For the Recall measure, it is also the winner however due to the post-hoc analysis (CD = 0.43) its advantage over both NUOB and NOOB (also new proposed ensembles) is not so significant.

This trend was also observed in the Recall measure when analyzing streams with pairs of difficulty factors (see Table 4). The proposed hybrid algorithm is the best in all of the analyzed scenarios but its superiority over other ensembles is not so significant, as confirmed by the post-hoc analysis.
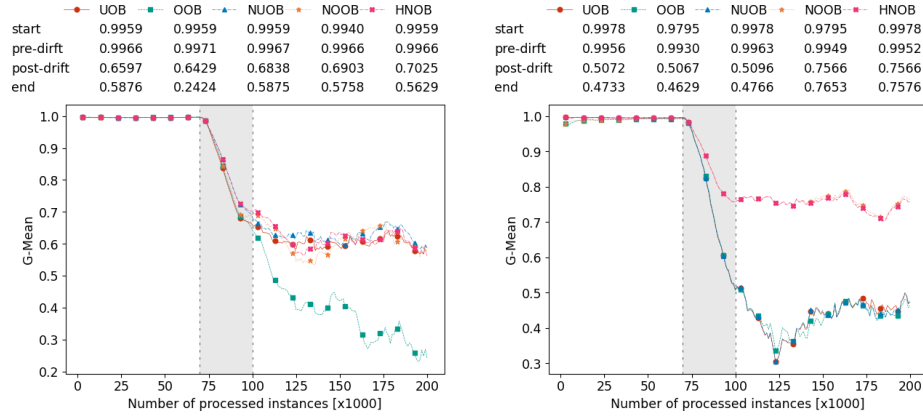
	UOB	OOB	NUOB	NOOB	HNOB
start	0.9959	0.9959	0.9959	0.9940	0.9959
pre-dirft	0.9966	0.9971	0.9967	0.9966	0.9966
post-drift	0.6597	0.6429	0.6838	0.6903	0.7025
end	0.5876	0.2424	0.5875	0.5758	0.5629

	UOB	OOB	NUOB	NOOB	HNOB
start	0.9978	0.9795	0.9978	0.9795	0.9978
pre-dirft	0.9956	0.9930	0.9963	0.9949	0.9952
post-drift	0.5072	0.5067	0.5096	0.7566	0.7566
end	0.4733	0.4629	0.4766	0.7653	0.7576

Fig. 4. Plots of *G-mean* measures for complex data streams *Split5+Im1+Borderline40+Rare40* (the left-hand figure) and *StaticIm10+Im1+Rare80* (the right-hand figure)

Table 3. Average ranks in the Friedman test for comparison of Hybrid Neighbourhood Online Bagging vs. other online bagging variants

Factors	Metric	HNOB	UOB	OOB	NUOB	NOOB	CD
Multiple	G-mean	**1.87**	3.31	4.16	2.86	2.80	0.43
All		**2.09**	3.02	3.76	3.03	3.10	0.31
Multiple	Recall	**2.25**	3.15	4.63	**2.32**	**2.66**	0.43
All		2.41	3.04	4.42	**1.89**	3.23	0.31

5   Discussion and Final Remarks

As the empirical studies have shown that previously known streaming online classifiers cannot cope well with imbalanced data streams affected by difficulty factors such as the presence or drift of unsafe minority types of examples (including borderline or rare examples), in this paper we have introduced specialized extensions of resampling online bagging, namely Neighbourhood Undersampling or Oversampling Online Bagging ensembles. The main element of their design is to use the safety assessment of the currently processed example based on the distribution of labels of the other examples in its local neighbourhood. This leads us to an appropriate modification of the degree of transfer of these processed examples to component classifiers in accordance with the Poisson distribution parameter. Depending on tuning an additional exponent Ψ parameter it is possible to more or less focus on the most difficult examples.

The results of experiments conducted on many synthetic data streams, modeling various difficulty factors or drifts, showed that both introduced online bag-

Table 4. Average ranks in the Friedman test for different pairs of factors in streams – where classifiers evaluated by Recall metric

Pairs of factors	HNOB	UOB	OOB	NUOB	NOOB	CD
Imbalance+Move	**2.50**	**2.50**	4.83	**1.58**	3.58	1.82
Imbalance+Merge	**2.42**	**2.32**	4.83	**1.67**	3.75	1.82
Imbalance+Split	**2.39**	**2.56**	4.83	**1.56**	3.67	1.47
Imbalance+Borderl.	**2.20**	3.10	4.88	**1.75**	3.08	0.97
Imbalance+Rare	**1.95**	3.58	4.83	2.90	**1.75**	0.97
Split+Borderline	**2.46**	3.00	4.48	**2.04**	3.02	0.87
Split+Rare	**2.14**	3.28	4.56	**2.76**	**2.26**	0.87

ging generalizations are quite effective. In particular, in line with the hypotheses made during their construction, they were most effective for scenarios with drifts in the types of examples that previous online re-sampling bagging could not cope with. Neighbourhood Undersampling Online Bagging is the best classifier for dealing with the presence of borderline examples - which could be explained by their internal mechanism of reducing the probability of learning from the most difficult examples and performing "cleaning" of the borderline region between classes. On the other hand, Neighbourhood Oversampling Online Bagging was sometimes better for dealing with rare examples - recall here that Ψ parameter was tuned to a higher value increasing the chance of copying many such examples. In general Neighbourhood Undersampling Online Bagging was the best classifier in most complex scenarios including multiple difficulty factors or their drift. However, the best choice is to use Hybrid Neighbourhood Online Bagging which dynamically uses the currently superior model of a pair of parallel trained online bagging ensembles.

Despite its definitely best experimental results (see e.g. Table 3), it opens the way for further research on more advanced ensembles of classification models and on stronger modifications of their structure as a result of detected drifts.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Aguiar, G., Krawczyk, B., Cano, A.: A survey on learning from imbalanced data streams: taxonomy, challenges, empirical study, and reproducible experi-

mental framework. Machine learning **113**, 1–79 (2023). https://doi.org/10.1007/s10994-023-06353-6

2. Branco, P., Torgo, L., Ribeiro, R.: A survey of predictive modeling under imbalanced distributions. ACM Comput Surv **49**(2), 31 (2016). https://doi.org/10.1145/2907070

3. Brzeziński, D., Minku, L.L., Pewiński, T., Stefanowski, J., Szumaczuk, A.: The impact of data difficulty factors on classification of imbalanced and concept drifting data streams. Knowledge and Information Systems **63**(6), 1429–1469 (2021)

4. Brzeziński, D., Stefanowski, J.: Ensemble Classifiers for Imbalanced and Evolving Data Streams, pp. 44–68. World Scientific (2018)

5. Błaszczyński, J., Stefanowski, J.: Neighbourhood sampling in bagging for imbalanced data. Neurocomputing **150**, 529–542 (10 2015). https://doi.org/10.1016/j.neucom.2014.07.064

6. Chen, W., Yang, K., Yu, Z., Shi, Y., Chen, C.L.P.: A survey on imbalanced learning: latest research, applications and future directions. Artificial Intelligence Review **57** (2024). https://doi.org/10.1007/s10462-024-10759-6

7. Ditzler, G., Roveri, M., Alippi, C., Polikar, R.: Learning in nonstationary environments: A survey. IEEE Comp. Int. Mag. **10**(4), 12–25 (2015)

8. Fernández, A., García, S., Galar, M., Prati, R.C., Krawczyk, B., Herrera, F.: Learning from Imbalanced Data Sets. Springer International Publishing (2018)

9. Krawczyk, B., Minku, L., Gama, J., Stefanowski, J., Woźniak, M.: Ensemble learning for data stream analysis: A survey. Information Fusion **37**, 132–156 (2017)

10. Lipska, A., Stefanowski, J.: The influence of multiple classes on learning from imbalanced data streams. In: Third International Workshop on Learning with Imbalanced Domains: Theory and Applications. ECMLPKDD 2022. pp. 187 – 198. PMLR (2022)

11. Napierała, K., Stefanowski, J.: Types of Minority Class Examples and Their Influence on Learning Classifiers from Imbalanced Data. Journal of Intelligent Information Systems **46**, 563–597 (2016)

12. Oza, N.C., Russell, S.J.: Experimental comparisons of online and batch versions of bagging and boosting. In: Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining, San Francisco, CA, USA, August 26-29, 2001. pp. 359–364 (2001)

13. Wang, S., Minku, L., Yao, X.: Dealing with multiple classes in online class imbalance learning. In: Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-16). pp. 2118—2124 (2016)

14. Wang, S., Minku, L.L., Yao, X.: Resampling-based ensemble methods for online class imbalance learning. IEEE Trans. Knowl. Data Eng. **27**(5), 1356–1368 (2015)