Applicability of Roughly Balanced Bagging for Complex Imbalanced Data

Mateusz Lango and Jerzy Stefanowski

Institute of Computing Science, Poznań University of Technology,
60-965 Poznań, Poland

Abstract. Roughly Balanced Bagging is based on under-sampling and classifies imbalanced data much better than other ensembles. In this paper, we experimentally study its properties that may influence its good performance. Results of experiments show that it can be constructed with a small number of component classifiers, which are quite accurate, however, of low diversity. Moreover, its good performance comes from its ability to recognize unsafe type of minority examples better than other ensembles. We also present how to improve its performance by integrating bootstrap sampling with random selection of attributes.

Keywords: class imbalance, ensembles, Roughly Balanced Bagging, types of minority examples

1 Introduction

Learning classifiers from imbalanced data still reveals research challenges. However, difficulties are not caused by the unbalanced class cardinalities only. Deterioration of classification performance arises when other data difficulty factors occur together with the class imbalance ratio, such as decomposition of the minority class into rare sub-concepts, too extensive overlapping of decision classes or presence of minority examples inside the majority class regions [9, 13].

Several methods have been introduced to deal with imbalanced data; for their review see, e.g., [5]. New specialized *ensembles* are able to handle complex imbalanced distributions better than simpler approaches; see their review in [4, 12]. They are usually modifications of bagging or boosting and either employ pre-processing methods before learning component classifiers or embed the cost-sensitive framework in the learning process. However, the comparative studies of new ensembles are still too limited. Studies [1, 4, 10] have showed that extensions of bagging work better than generalizations of boosting and more complex solutions. The recent study [2] demonstrated that Roughly Balanced Bagging [6] has achieved the best results and is significantly better than other over-sampling extensions of bagging.

The key idea behind Roughly Balanced Bagging is a specific random under-sampling before generating component classifiers, which reduces the presence of the majority class examples inside each bootstrap sample. Although this ensemble has been successfully used in several papers, there are not enough attempts to check which of its characteristics are the most crucial for improving classification of complex imbalanced data. In our opinion, its properties should be examined more precisely.

The aim of this paper is to experimentally study the following issues: (1) the most influential aspects of constructing Roughly Balanced Bagging and its main properties (with respect to bootstrap construction, deciding on the number of component classifiers, their diversity, methods for aggregating predictions); (2) abilities of this ensemble to deal with different types of difficult distributions of the minority class; (3) directions for its further extension and improvements.

2 Related Works

For the reviews of ensembles dedicated for class imbalance consult [4, 5, 12]. Galar et al. in [4] distinguish mainly *cost-sensitive* approaches vs. integrations with *data pre-processing*. Below we briefly present under-bagging proposals only, which are most relevant to our study.

Breiman's algorithm for learning bagging samples a number of subsets from the training set, builds multiple base classifiers and aggregates their predictions to make a final decision. *Bootstraps* are generated by uniform random sampling with replacement of instances from the original training set (usually keeping the size of the original set). However, as this sampling is performed on all data elements, regardless their class labels (majority or minority), the imbalanced class distribution will be hold in each bootstrap and the ensemble will fail to sufficiently classify minority class. Most of current proposals overcome this drawback by applying pre-processing techniques to each bootstrap sample, which change the balance between classes – usually leading to the same, or similar, cardinality of the minority and majority classes. For instance, the over-sampling methods typically replicate the minority class data (either by random sampling or generating synthetic examples) to balance bootstraps.

In *under-bagging* the number of the majority class examples in each bootstrap is randomly reduced to the cardinality of the minority class (N_{min}). In the simplest proposals, as *Exactly Balanced Bagging*, the entire minority class is just copied to the bootstrap sample and then combined with the randomly chosen subset of the majority class to exactly balance cardinality between classes.

While such under-bagging strategies seem to be intuitive and work efficiently in some studies, Hido et al. [6] have claimed that they do not truly reflect the philosophy of bagging and could be still improved. In the original bagging the class distribution of each sampled subset varies according to the binomial distribution while in the above under-bagging each subset has the same class distribution as the desired balanced distribution. In *Roughly Balanced Bagging* (RBBag) the numbers of instances for both classes are determined in a different way by equalizing the sampling probability of each class. The number of minority examples (S_{min}) in each bootstrap is set to the size of the minority class N_{min} in the original data. In contrast, the number of majority examples is decided probabilistically according to the negative binomial distribution, whose parameters are the number of minority examples (N_{min}) and the probability of success equal to 0.5. In this approach only the size of the majority examples (S_{maj}) varies, and the number of examples in the minority class is kept constant since it is small. Finally, component classifiers are induced by the same learning algorithm from each i bootstrap sample ($S^i_{min} \cup S^i_{maj}$) and their predictions form the final decision with the equal weight majority voting.

Hido et al. compared Roughly Balanced Bagging with several algorithms showing that it was better on G-mean and AUC measures [6]. The study [10] demonstrated that under-bagging, including RBBag, significantly outperformed best extensions of boosting and the difference was more significant when data were more noisy. The results of [2] showed that Roughly Balanced Bagging was significantly better than best oversampling extensions of bagging and usually better than Exactly Balanced Bagging. These experiments also supported using sampling with replacement in RBBag – so, we will also use it in further experiments. However, there are not so many attempts to either to experimentally examine properties of this ensemble or to more theoretically explain why and when it should outperform other methods. Only the work [16] provides a probabilistic theory of imbalance and its reference to under-sampling classifiers.

3 Studying the Role of Components in Roughly Balanced Bagging

The first part of experiments aims at studying the following basic properties of constructing Roughly Balanced Bagging, which have not been studied in the literature yet: (1) Using different learning algorithms to built component classifiers; (2) The influence of the number of component classifiers on the final performance; (3) The role of diversity of component classifiers.

We extend the previous implementation of RBBag done by L. Idkowiak for the WEKA framework [2]. We choose 24 UCI datasets which have been used in the most related experimental studies [3, 10, 13, 14]. They represent different imbalance ratios and other data difficulty factors - so they constitute various difficulty levels for ensembles [3, 14]. Moreover, we consider binary class versions of these data as it is done in the related studies [4, 6, 10]. For their detailed characteristics the reader is referred to [3].

The performance of ensembles is measured using: *sensitivity* of the minority class (the minority class accuracy), its *specificity* (the majority classes accuracy), their aggregation to the *geometric mean* (G-mean) and *F-measure*. For their definitions see, e.g., [5]. We have chosen these point measures, instead of AUC as the most of considered learning algorithms produce deterministic outputs. These measures are estimated with the stratified 10-fold cross-validation repeated several times to reduce the variance.

3.1 Choosing Algorithms to Learn Component Classifiers

The related works show that Roughly Balanced Bagging as well as other under-sampling extensions of bagging are usually constructed with decision trees. In this study we check whether classification performance of this ensemble may depend on using other learning algorithms. Besides J48 unpruned tree we considered Naive Bayes tree, rule algorithms – Ripper and PART, Naive Bayes classifiers and SVM – all available in WEKA. The RBBag ensemble was constructed with different numbers (30, 50 and 70) of component classifiers.

Here, we summarize the Friedman test only. For all considered evaluation measures we were unable to reject the null hypothesis on equal performance of all versions of RBBag. For instance, average ranks in the Friedman test for G-mean (the smaller, the better) were the following: SVM 4.1; Ripper 4.12; NBTree 4.4; J48tree/PART 4.5; NB 4.8. Quite similar rankings were obtained for other measures. All these results did not show significant differences of using any of this algorithm inside RBBag.

Furthermore, for each single algorithm RBBag was significantly better than its standard bagging equivalent (according to the paired Wilcoxon test).

3.2 The Influence of the Number of Component Classifiers

Related works showed that Roughly Balanced Bagging was used with rather a high number of component classifiers. Hido et al. [6] tested it with 100 C4.5 trees. In the study [10] authors applied a dozen of components. Then, the study [2] showed that it also performed well with 30, 50 or 70 classifiers. Thus, we have decided to examine more systemically other (also smaller) sizes of this ensemble and its influence on the final performance. We stayed with learning components with J48 unpruned trees, and for each dataset we constructed a series of Roughly Balanced Bagging ensembles increasing its size one by one - so the number of component classifiers changed from 2 trees up to 100 ones.

For all considered datasets increasing the number of component classifiers improves the evaluation measures up to the certain size of the ensemble. Then, values of measures stay at a stable level or slightly vary around the certain level. Due to page limits, in Fig. 1 we present the most representative changes of G-mean values. Figures for few other datasets present similar trends.

What is even more surprised the RBBag ensemble achieves this good performance for a relatively small number of component classifiers. For most datasets, the stable highest value of G-mean is observed approximately between 10 and 15 trees. In case of the sensitivity or F-measure we noticed similar tendencies.

Moreover, we decided to examine confidence of the final decision of RBBag. We refer to a *margin* of the ensemble prediction. For standard ensembles, it is usually defined as a difference between the number of votes of component classifiers for the most often predicted class label and the number of votes for the second predicted label. Here, we modified it as: $marg = (n_{cor} - n_{incor})/n_{cptclas}$, where n_{cor} is the number of votes for the correct class, n_{incor} is the number of votes for the incorrect class and $n_{cptclas}$ is the number of component classifiers in the ensemble. The higher absolute value of $marg$ is interpreted as high confidence while values closer to 0 indicate uncertainty in making a final decision for a classified instance. In Fig 2 we present a representative trend of changes of the relative margin with the size of RBBag for `ecoli` and `cmc` data. For many other

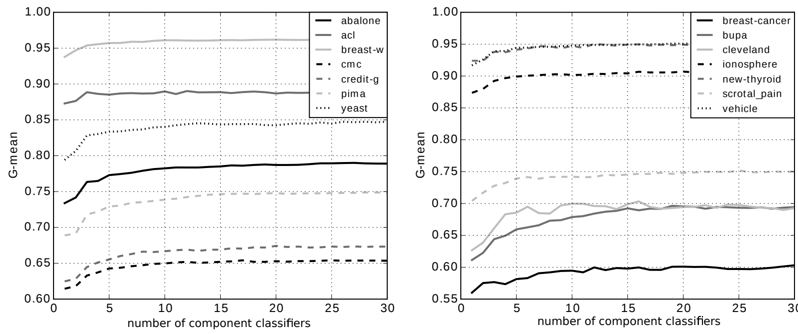

Fig. 1. G-mean vs. a number of component classifiers in RBBag for selected datasets.

datasets the trend line of the margin also stabilizes after a certain size (Note the resolution of the margin scale is more detailed than G-mean, so margin values achieve a satisfactory level also quite fast). We can conclude that the good performance of Roughly Balanced Bagging comes from rather a small number of component classifiers.

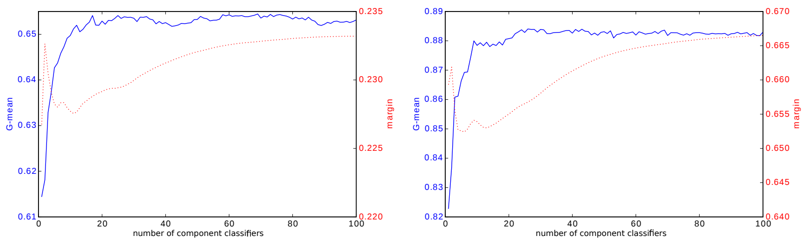

Fig. 2. G-mean and margin vs. a number of component classifiers in RBBag for `cmc` (left) and `ecoli` (right) datasets.

3.3 Diversity of Component Classifiers

The final accuracy of ensembles may be also related to their diversity - which is usually understood as the degree to which component classifiers make different decisions on one problem (in particular, if they do not make the same wrong decisions). Although, such an intuition behind constructing diverse component classifiers is present in many solutions, research concerns the total accuracy perspective [11]. It is still not clear how diversity affects classification performance especially on minority classes. The only work on ensembles dedicated for imbalance data [17] does not provide a clear conclusion. Its authors empirically studied diversity of specialized over-sampling ensembles and noticed that larger diversity improved recognition of the minority class, but at the cost of deteriorating the majority classes. However, nobody has analysed diversity of Roughly Balanced Bagging.

To evaluate diversity we calculated the disagreement measure [11]. For a pair of classifiers it is defined as a ratio of the number of examples on which both classifiers make different predictions to the number of all classified examples. This measure is calculated for each pair of component classifiers. Then the global, averaged disagreement D of an ensemble is averaged over all pairs of classifiers. The larger its value is, the more diverse classifiers are [11]. We calculated the global average disagreement D for predictions in both classes and also for the minority class only (denoted as D_{min}). These values are presented in Table 2 - two first columns for RBBag ensemble and next columns refer to its extension discussed in Sec. 5 – both ensembles were constructed with 30 component J48 trees. As this table concerns further extension of RBBag for a higher number of attributes, the list of datasets is reduced.

Notice that values of disagreement measures are relatively low. For nearly all datasets they are between 0.1 and 0.3. The small diversity concerns both class predictions (D) and minority class (D_{min}), although D_{min} is usually lower than D. Similar low values occurred for the remaining datasets, not included in Table 2. We also checked that changing the number of component classifiers in RBBag did not influence values of the disagreement measures.

To sum up, the high accuracy of Roughly Balanced Bagging is not directly related to its higher diversity. We have also analysed predictions of particular pairs of classifiers and noticed that they quite often make the same decisions (most often correct ones).

4 Influence of the Type of Examples

According to Napierala and Stefanowski [13, 14] the data difficulty factors concerning distributions of imbalanced classes can be modeled by the following types of examples: *safe examples* (located in the homogeneous regions populated by examples from one class only); *borderline* (placed close to the decision boundary between classes); *rare examples* (isolated groups of few examples located deeper inside the opposite class), or *outliers*. Following the method introduced in [13] the type of example can be identified by analysing class labels of the k-nearest neighbours of this example. For instance, if $k = 5$, the type of the example is assigned in the following way [13, 14]: 5:0 or 4:1 – an example is labelled as safe example; 3:2 or 2:3 – borderline example; 1:4 – labelled as rare example; 0:5 – example is labelled as an outlier. This rule can be generalized for higher k values, however, results of recent experiments [14] show that they lead to a similar categorization of considered datasets. Therefore, in the following study we stay with $k = 5$.

Repeating conclusions from experimental studies [13, 14] the most of datasets considered in this paper contain rather a small number of safe examples from the minority class. The exceptions are two datasets composed of many safe examples: `new-thyroid`, and `car`. Many datasets such as `cleveland`, `balance-scale` or `solar-flare` do not contain any safe examples but many outliers and rare cases.

In the current experiments we identified a type of the testing example and recorded whether it was correctly classified or not. Additionally, we refer types of examples in both (minority and majority) classes to the relative margins of the RBBag predictions (these are presented as histograms of numbers of testing examples with a given value of the margins). In Fig 3 and 4 we present a representative results of RBBag and the standard bagging for `cleveland` dataset. Histograms for other datasets present similar observations.

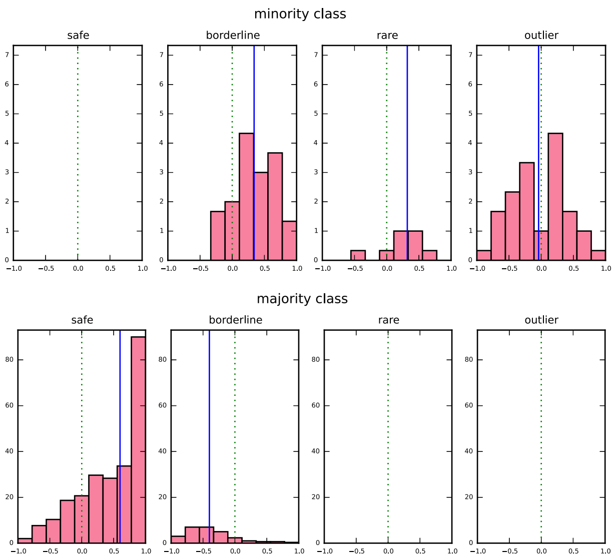

Fig. 3. Histogram of RBBag margins for `cleveland` dataset with respect to a class and a type of example. Blue vertical line shows the value of the margin's median.

Notice that RBBag quite well recognizes the borderline examples from the minority class. Rare minority examples are more difficult, however, on average RBBag can still recognize many of them. It classifies them much better than the standard bagging. Outliers are the most difficult, but RBBag classifies correctly some of them and again this is the main difference to standard bagging and other its over-sampling extensions evaluated in [3]. The similar tendency is observed for other unsafe datasets which are not visualized due to page limits. If the

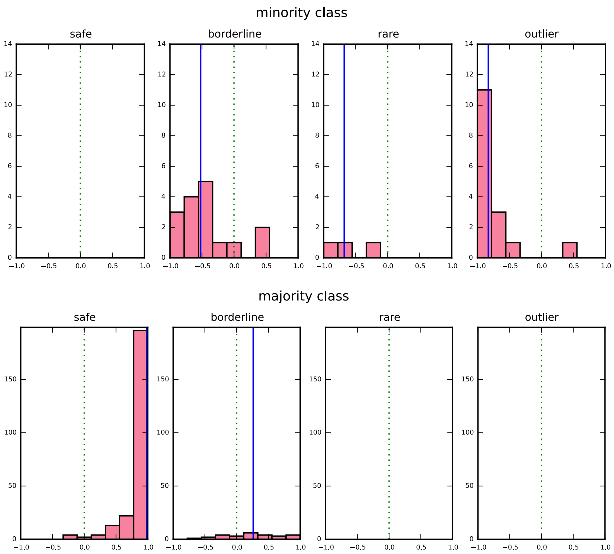

Fig. 4. Histogram of standard bagging margins for `cleveland` dataset with respect to a class and a type of example. Blue vertical line shows the value of the margin's median.

dataset contains some safe minority examples, nearly all of them are correctly classified with high margins.

On the other hand, for the majority class, one can notice that RBBag correctly classifies most of safe examples while facing difficulties with borderline ones. It also holds for other non-visualized datasests (where the margin's median for borderline majority examples is always worse than the median for borderline minority examples). The majority class does not contain any rare or outlying examples for nearly all considered datasets. For few exceptions, `pima`. `breast-cancer` or `cmc`, these rare majority examples are misclassified with the high negative margin.

In conclusion, we can hypothesize that Roughly Balanced Bagging improves recognition of unsafe minority examples, but at the cost of worse dealing with unsafe majority examples. However, as the number of unsafe examples is relatively small in the majority class, the final performance of RBBag (e.g., averaged by G-mean) is improved.

5 Applying Random Selection of Attributes

Although Roughly Balanced Bagging performs quite well, it can still be improved. Here, we have focused on modifications of constructing bootstrap samples. Observations of rather limited diversity of RBBag components have led us to considering inspirations from earlier research on applying random attribute selection while constructing component classifiers. Recall that Ho introduced in [7] *Random Subspace* method (RSM) for highly dimensional data, where in each iteration of constructing the ensemble a subset of all available attributes is randomly drawn and a component classifier is built using only this subset. Then, Breiman combined bootstrap sampling with random selection of attributes in nodes of trees inside the Random Forest ensemble. Recent experiments of [12] also demonstrated that combing re-sampling with Random Forests helps for class imbalance. However, we are more interested in adapting Random Subspace into the context of Roughly Balanced Bagging as it is a classifier independent strategy. To best of our knowledge it has not been considered for RBBag yet. In the only related work [8] authors successfully applied this method to SMOTE based oversampling ensemble.

In our extension of RBBag, after sampling each bootstrap we randomly select f attributes from the set of all attributes. Subsequently, we train base classifier on a sample from which we removed not selected ones. We denote this extension as RBBag+RSM.

Since RSM is a method designed for high-dimensional data, we have chosen to our experiments only these datasets from earlier phases of experiments, which contain more then 11 attributes. As this condition holds for 9 datasets only, we added 4 new, high-dimensional imbalanced datasets from UCI repository. Finally, in this experiment we examine 13 following datasets: `abdominal-pain` (13 attributes), `cleveland` (13), `credit-g` (20), `dermatology` (35), `hepatitis` (19), `ionosphere` (34), `satimage` (37), `scrotal-pain` (13), `segment` (20), `seismic-bumps` (19), `solar-flare` (12), `vehicle` (18) and `vowel` (14).

We tested with J48 decision tree (without pruning) and SVM as base classifiers. Following the literature review, we considered setting f parameter to $\lceil\sqrt{F}\rceil$, $\lceil\log_2 F+1\rceil$ and $\lceil{}^1/_2F\rceil$, where F is the total number of attributes in the dataset. Due to space limit we present results only for J48 decision trees and $f = \lceil\sqrt{F}\rceil$, since this parameter setting gives, on average, the highest increments.

We consider RBBag ensemble containing 30 component classifiers to be consistent with earlier experiments, in particular on diversity. However, following earlier observations, as e.g. [7], that randomization of attributes should increase the variance of bootstrap samples, we compare RBBag against the new proposed RBBag+RSM ensemble having also larger sizes (besides 30, also 50 and 70 components).

The values of G-mean and sensitivity are presented in Table 1. One can notice increases of both measures, in particular RBBag+RSM with more trees. For instance, the increase on sensitivity (`abdominal-pain`, `hepatitis` – above 6%) and G-mean (`abdominal-pain`, `hepatitis`, `scrotal-pain`, `seismic- bumps` – above 3%). We performed the paired Wilcoxon test to compare RBBag+RSM against RBBag. With the confidence $\alpha = 0.05$, RBBag+RSM is better on G-mean for 50 ($p = 0.007$) and 70 ($p = 0.003$) trees and nearly for 30 trees ($p = 0.054$). Similar results of this text hold for the sensitivity measure. Thus, it is better to construct RBBag+RSM with more trees than its RBBag equivalent.

	Sensitivity				G-mean			
	RBBag	RBBag+RSM			RBBag	RBBag+RSM		
Dataset	30	30	50	70	30	30	50	70
abdominal-pain	0.7955	0.8523	0.8623	0.8563	0.8077	0.8336	0.8411	0.8358
cleveland	0.7067	0.6800	0.7117	0.7567	0.7161	0.6938	0.7197	0.7410
credit-g	0.6610	0.6493	0.6407	0.6540	0.6735	0.6930	0.6923	0.7007
dermatology	0.9900	1.0000	1.0000	1.0000	0.9868	0.9986	1.0000	1.0000
hepatitis	0.7500	0.8200	0.8267	0.8267	0.7663	0.8131	0.8113	0.8029
ionosphere	0.8553	0.8660	0.8737	0.8796	0.9063	0.9068	0.9104	0.9152
satimage	0.8690	0.8738	0.8720	0.8777	0.8727	0.8677	0.8678	0.8698
scrotal-pain	0.7400	0.7467	0.7560	0.7453	0.7484	0.7869	0.7846	0.7884
segment	0.9863	0.9918	0.9933	0.9930	0.9892	0.9945	0.9955	0.9953
seismic-bumps	0.6312	0.6624	0.6629	0.6612	0.6824	0.7103	0.7153	0.7124
solar-flare	0.8690	0.8450	0.8670	0.8670	0.8499	0.8351	0.8437	0.8458
vehicle	0.9688	0.9990	0.9990	0.9990	0.9525	0.9590	0.9588	0.9599
vowel	0.9667	0.9911	0.9911	0.9900	0.9623	0.9751	0.9766	0.9789

Table 1. Sensitivity and G-mean for Roughly Balanced Bagging (RBBag) and its modification by random attribute selection (RBBag+RSM).

Additionally we calculated the disagreement measure for all examples (D) and also the minority class (D_{min}). The values presented in Table 2 are calculated for 30 trees. For reader convenience we present our results together with difference of disagreement between RBBag+RSM and original RBBag.

One can notice that Random Subspace method resulted in an increase of disagreement on almost all data sets (except `seismic-bumps`). Interestingly, despite a decline of the disagreement measure on this dataset we observed improvement on both G-mean and sensitivity.

	RBBag		RBBag+RSM		Difference	
Dataset	D	D_{min}	D	D_{min}	D	D_{min}
abdominal-pain	0.1564	0.1310	0.2995	0.2580	0.1431	0.1269
cleveland	0.2807	0.2470	0.3506	0.3050	0.0700	0.0581
dermatology	0.0211	0.0162	0.1815	0.1384	0.1604	0.1222
credit-g	0.2648	0.2279	0.4075	0.3951	0.1427	0.1672
hepatitis	0.2476	0.2127	0.3156	0.2915	0.0680	0.0788
ionosphere	0.0733	0.0909	0.1158	0.1650	0.0424	0.0741
satimage	0.1549	0.1160	0.1782	0.1448	0.0233	0.0288
scrotal-pain	0.1871	0.1670	0.3522	0.3139	0.1651	0.1469
semgent	0.0168	0.0106	0.0659	0.0293	0.0491	0.0187
seismic-bumps	0.2891	0.2373	0.2470	0.2383	−0.0421	0.0010
solar-flare	0.1062	0.0999	0.2362	0.2395	0.1300	0.1396
vehicle	0.0592	0.0509	0.1461	0.0972	0.0869	0.0463
vowel	0.0461	0.0251	0.2126	0.0825	0.1665	0.0574

Table 2. Disagreement measures, calculated for examples from both classes (D) and from the minority class only (D_{min}), for Roughly Balanced Bagging (RBBag) and its modification by random attribute selection (RBBag+RSM).

6 Discussion and Final Remarks

This study attempts to extend knowledge on properties of Roughly Balanced Bagging, which is one of the most accurate ensemble dedicated for class imbalances. Our experiments show that it can be constructed with a relatively small number of component classifiers (approx. 15 ones). It is an interesting observation, as this ensemble may require a heavy under-sampling. One could expect that due to such strong changes inside distributions in bootstrap samples, their variance will be high, and the ensemble should reduce it by applying many components. However, the experimental results have showed that it is not a case. Moreover, this can be a promising indication for mining complex, larger data and for constructing this ensemble in an iterative way (starting from a smallest size and stepwise adding a new component while testing it with the extra validation set). According to other experiments the choice of the considered algorithms for learning component classifiers does not influence the final performance of RBBag

Another discovery is quite low diversity of RBBag. We have also confirmed it by calculating Q statistics diversity measure [11, 17]. Comparing it to earlier results [2] we argue that RBBag is less diversified than over-bagging or SMOTE-based bagging. On the other hand, RBBag is more accurate than these more diversified ensembles. We have also checked that its components are quite accurate and pairs of classifiers often make the same correct decisions. It may open another research on studying the trade off between accuracy and diversity of ensembles for imbalanced data.

Studying the local recognition of types of classified examples shows that RBBag improves classification of unsafe minority examples. Its power for dealing with borderline, rare and outlying examples distinguishes it from other ensembles. Here, we recall experiments from [3], which were focused on analysing distributions of example types in bootstraps. Their results revealed that several unsafe minority examples from the original data were changed by RBBag bootstrap sampling into safer ones which was not a case for other bagging extensions.

In this study we advocate for further modifications of bootstrap sampling. Our experiments have demonstrated that an integration of random selection of attributes improves the classification performance. However, other modifications could be still considered. In [3] we have already introduced Nearest Balanced Bagging which exploits information on types of minority examples and directs sampling towards the more unsafe examples. Although its experimental results are encouraging (for some datasets even better than RBBag) it generates bootstrap samples containing more minority examples than majority ones. Thus, it may be still shifted too much to improving sensitivity at the cost of removing too many majority examples. Recall that experiments from Sec. 4 have shown that RBBag also improves recognition of unsafe minority examples while worsening classification of borderline majority examples. Considering different types of examples from both classes while modifying bootstrap sampling in Roughly Balanced Bagging is still an open challenge.

Furthermore, a decomposition of classes into sub-concepts [9] could be considered. In [15] authors applied k-means clustering to stratify sampling majority examples inside their modifications of standard bagging. Looking for another semi-supervised clustering to better handle complex boundaries of data distributions could be yet another direction for future research.

Ack. The research was supported by NCN grant DEC-2013/11/B/ST6/00963.

References

1. Anyfantis, D., Karagiannopoulos, M., Kotsiantis, S., Pintelas, P. : Creating ensembles of classifiers by distributing an imbalance dataset to reach balance in each resulting training set. In Proc. of the IEEE DHMS Conf., (2008).
2. Błaszczyński, J., Stefanowski, J., Idkowiak L.: Extending bagging for imbalanced data. Proc. of the 8th CORES 2013, Springer Series on Advances in Intelligent Systems and Computing 226, 269–278 (2013).
3. Błaszczyński, J., Stefanowski, J.: Neighbourhood Sampling in Bagging for Imbalanced Data. Neurocomputing, vol. 150 A, 184-203 (2015).
4. Galar, M., Fernandez, A., Barrenechea, E., Bustince, H. Herrera, F.: A Review on Ensembles for the Class Imbalance Problem: Bagging-, Boosting-, and Hybrid-Based Approaches. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, vol. 99, 1–22 (2011).
5. He, H., Yungian, Ma (eds): Imbalanced Learning. Foundations, Algorithms and Applications. IEEE - Wiley, (2013).
6. Hido S., Kashima H.: Roughly balanced bagging for imbalance data. In Proc. of the SIAM Int. Conference on Data Mining, 143-152 (2008) - an extended version in Statistical Analysis and Data Mining, vol. 2 (5-6), 412–426 (2009).
7. Ho, T.: The random subspace method for constructing decision forests. Pattern Analysis and Machine Intelligence 20(8), 832-844 (1998).
8. Hoens, T., Chawla, N.: Generating Diverse Ensembles to Counter the Problem of Class Imbalance. In Proc. PAKDD 2010, 488-499 (2010).
9. Jo, T., Japkowicz, N.: Class Imbalances versus small disjuncts. ACM SIGKDD Explorations Newsletter, vol. 6 (1), 40–49 (2004).
10. Khoshgoftaar T., Van Hulse J., Napolitano A.: Comparing boosting and bagging techniques with noisy and imbalanced data. IEEE Transactions on Systems, Man, and Cybernetics–Part A, 41 (3), 552–568 (2011).
11. Kuncheva, L.: Combining Pattern Classifiers. Methods and Algorithms. Wiley (2d Edition) (2014).
12. Liu A., Zhu Zh: Ensemble methods for class imbalance learning. In He, H., Yungian Ma. (eds): Imbalanced Learning. Foundations, Algorithms and Applications. Wiley, 61-82 (2013).
13. Napierala, K., Stefanowski, J.: The influence of minority class distribution on learning from imbalance data. In. Proc. 7th Conf. HAIS 2012, LNAI vol. 7209, Springer, 139-150 (2012).
14. Napierala, K., Stefanowski, J.: Types of Minority Class Examples and Their Influence on Learning Classifiers from Imbalanced Data. Journal of Intelligent Information Systems, on-line access DOI: 10.1007/s10844-015-0368-1 (2015).
15. Parinaz, S., Victor, H., Matwin, S.: Learning from Imbalanced Data Using Ensemble Methods and Cluster-based Undersampling. In Post- Proc. 3rd Workshop New Frontiers of Mining Complex Patterns at ECML-PKDD 2014, Nancy, LNAI vol. 8983, Springer, 69 – 86 (2015).
16. Wallace, B., Small, K., Brodley, C., Trikalinos, T.: Class Imbalance, Redux. Proc. 11th IEEE International Conference on Data Mining, 754 – 763 (2011).
17. Wang, S., Yao, T.: Diversity analysis on imbalanced data sets by using ensemble models. In Proc. IEEE Symp. Comput. Intell. Data Mining, 324-331 (2009).

ECML PKDD SEP 07 11 TEM BER 2015 PORTO, PORTUGAL

Proceedings of the 4th Workshop on

New Frontiers in Mining Complex Patterns (NFMCP 2015)

Editors

Michelangelo Ceci

Corrado Loglisci

Giuseppe Manco

Elio Masciari

Zbigniew W. Ras

Table of Contents

INVITED TALK