An Improved Ensemble Approach for Imbalanced Classification Problems

Bartosz Krawczyk
Department of Systems & Computer Networks
Wrocław University of Technology
Wrocław, Poland
bartosz.krawczyk@pwr.wroc.pl

Gerald Schaefer
Department of Computer Science
Loughborough University
Loughborough, U.K
gerald.schaefer@ieee.org

***Abstract*—Classification of imbalanced data is a challenging task in machine learning, as most classification approaches tend to bias towards the majority class, even though the minority class is often the one of greater importance. Consequently, methods that are capable of boosting the classification accuracy on the minority class are sought after. In this paper, we propose an improved ensemble approach for imbalanced classification. Our algorithm is based on undersampling of the majority class to create balanced object subspaces, on which individual classifiers are trained. As not all generated classifiers will be useful for the ensemble construction, we carry out a pruning procedure to discard irrelevant models. This classifier selection is based on a diversity measure to identify mutually complementary classifiers. The remaining predictors are combined using a trained fuser based on discriminants. Extensive experimental results on several benchmark datasets demonstrate our proposed method to adequately address class imbalance and to (statistically) outperform several state-of-the-art classifier ensembles dedicated to imbalanced classification.**

I. Introduction

The underlying class distribution can have a crucial effect on performance of pattern classification algorithms. While in many pattern recognition tasks the distribution is roughly equal among the classes, this does not hold for all applications. When samples of one of the classes (the majority class) significantly outnumber the remaining (minority) class(es), we deal with a problem known as imbalanced classification [1] which may occur in a variety of domains including anomaly detection [2], fault diagnosis [3], medical data analysis [4], and face recognition [5]. While the performance of classification algorithms is typically evaluated using predictive accuracy, clearly this is not appropriate when the data is imbalanced as this would favour the correct identification of majority class samples while neglecting recognition of the minority class.

Ensemble classification methods, or multiple classifier systems (MCSs) have been shown to be able to lead to both more robust as well as better performing classification approaches [6]. When designing a MCS, several important issues have to be considered, including

- which base classifiers should be used for the ensemble?
- should all classifiers be chosen as part of the committee or should should some classifiers be discarded?
- how should the classifier(output)s be combined to provide the best overall performance?

In this paper, we present an ensemble classification approach for imbalanced classification problems. In particular, we propose an extension of our previous algorithm – Under-Sampling Balanced Ensemble (USBE) [7] – which is based on the idea of creating base classifiers from balanced object subspaces, and add an ensemble pruning step to discard redundant classifiers from the ensemble. Our Pruned Under-Sampling Balanced Ensemble (PUSBE) is demonstrated to give excellent performance on several benchmark imbalanced datasets and is shown to outperform various other multiple classifier systems dedicated to imbalanced classification including the original USBE algorithm.

II. Multiple classifier systems

Multiple classifier systems (MCSs), or ensemble classifiers, combine a number of base classifiers to arrive at a classification system than can outperform the individual classifiers [6]. Assume we have N classifiers $\Psi^{(1)}, \ \Psi^{(2)}, \ \ldots \ , \ \Psi^{(N)}$. For a given sample $x \in \mathcal{X}$, each individual classifier makes a decision regarding class $i \in \mathcal{M} = \{1, \ ..., M\}$. Let $F^{(l)}\left(i, x\right)$ denote this decision of the l-th classifier $\Psi^{(l)}$, then a combined classifier Ψ can reach a decision based on [8]

$$\Psi\left(x\right) = i \quad \text{if} \quad \hat{F}\left(i, x\right) = \max_{k \in M} \hat{F}\left(k, x\right), \tag{1}$$

where

$$\hat{F}\left(i, x\right) = \sum_{l=1}^{N} w^{(l)}(i) F^{(l)}\left(i, x\right) \quad \text{and} \quad \sum_{i=1}^{N} w^{(l)}(i) = 1, \tag{2}$$

and the weights w are dependent on the classifier and the class, i.e. $w^{(l)}(i)$ is assigned to the l-th classifier and the i-th class.

III. Classification of imbalanced datasets

Imbalanced datasets occur when training samples are not approximately equally distributed among the classes. While the performance of pattern classification algorithms is conventionally evaluated using predictive accuracy, this is not appropriate when the data under consideration is imbalanced, since the decision boundary may be biased towards the majority class, leading to poor recognition of the minority class as illustrated in Fig. 1.

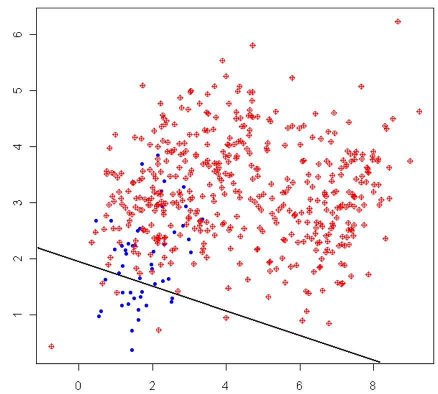

Fig. 1. Example of bias towards the majority class in linear classification of an imbalanced problem. The established decision boundary (line) would give poor prediction for minority class samples.

Techniques that address the problems associated with imbalanced datasets can in general be divided into three groups [9]:

- *Data level approaches* attempt to re-balance the class distribution. As they are independent of the actual classification stage, they can be used with basically any classification algorithm. The most popular approaches employ an oversampling strategy which generates artificial samples from existing ones [10].
- *Classifier level approaches* try to adapt existing algorithms to the problem of imbalanced datasets and bias them towards favouring the minority class. One possibility is to perform one-class classification, which can learn the concepts of the minority class by treating majority objects as outliers [11].
- *Cost-sensitive approaches* can use both data modifications (by adding a specified cost to the misclassification) and modifications of the learning algorithms (to adapt them to the possibility of misclassification) to address class imbalance by formulating the classification problem as a cost minimisation problem [12].

MCSs have also been adapted to handle imbalanced data, and typically combine an MCS algorithm with one of the above techniques. Examples of a combination of oversampling and classifier ensembles are SMOTEBagging [13] and SMOTEBoost [14] which introduce new objects into each of the bagging/boosting iterations separately. IIVotes [15] fuses a rule-based ensemble with a SPIDER pre-processing scheme so as to be more robust with respect to atypical data distributions and to automatically find an optimal number of bags. Cost-sensitive MCSs are mostly based on adjusting the object weights in a boosting schema [16], although schemes based on cost-sensitive decision trees have also been exploited [17]. EasyEnsemble [18] uses bagging as the main concept. Since for each of the bags AdaBoost is used as the base model, it can be viewed as an ensemble of ensembles.

IV. Pruned Under-Sampling Balanced Ensemble

In this paper, we present an ensemble based classification algorithm that works well on imbalanced datasets and is not based on either oversampling (which might lead to a class distribution shift) or cost-sensitive classification (since performance relies heavily on the correct specification of the cost matrix). For this, in [19], [7] we have presented Under-Sampling Balanced Ensemble (USBE) as an effective method for imbalanced classification. USBE is based on the idea of object space partitioning where each classifier is trained on a different subspace and constructed so as to counter the original imbalance in the dataset. In this paper, we add a classifier selection step based on a diversity measure to discard similar classifiers which do not contribute to the ensemble under consideration.

Pruned Under-Sampling Balanced Ensemble (PUSBE) proceeds in four main steps:

1) Creation of a number of balanced subspaces consisting of minority class and under-sampled majority class objects.
2) Construction of a pool of classifiers by training a single classifier on each of the subspaces. A feature selection algorithm [20] is employed and is applied independently for each of the subspaces/classifiers.
3) Diversity-based pruning of a pool of classifiers to select complementary models for the committee.
4) Fusion of outputs of the remaining classifiers.

In the following, we describe these stages in more detail.

A. Space partitioning and classifier generation

In USBE class imbalance is addressed based on object space division and proceeds in two steps:

1) Creation of a number of subspaces.
2) Construction of a pool of classifiers $\Pi^{\Psi} = \{\Psi^{(1)}, \Psi^{(2)}, \dots, \Psi^{(N)}\}$ by training single classifiers on each of the subspaces.

Space partitioning is employed to balance the unfavourable class distribution using a random undersampling method. Each of the newly created subspaces contains a smaller number of objects, randomly drawn from the dataset so that the number of objects in each of the subspaces is equal for both classes. The resulting subspaces contain all objects from the minority class and an equal number of objects from the majority class. Subspaces are created as long as there are unused objects from the latter group.

Feature selection is employed independently for each of the chosen subspaces. Therefore, in each of the subspaces the derived feature subsets may vary, leading to an increased overall diversity of the pool of classifiers, and consequently to a better ensemble. In our implementation, we employ the fast correlation-based feature filter (FCBF) [20]. In FCBF, the relations between features-classes and between pairs of features are considered. The algorithm proceeds at two levels. First, a ranking algorithm using the symmetric uncertainty coefficient (SUC) index is used to estimate class-feature relevance, and a threshold coefficient established to select predominant features. In the second part, features that are redundant to the predominant features are removed.

B. Ensemble pruning

While in USBE all base classifiers were using in the fusing stage, in this paper we employ an ensemble pruning stage to further improve the quality of the generated ensemble. Different base classifiers will have different areas of competence and hence may provide different contributions to the ensemble. Therefore, careful classifier selection should be conducted in order to choose the most valuable classifiers. There are several ways how such an ensemble pruning procedure can be implemented. One of the most popular criteria is to employ an ensemble diversity measure to select classifiers that are as different as possible from each other. Adding similar classifiers to the committee would not improve its quality but only increases its complexity. On the other hand, diverse models might be mutually supplementary and hence allow to exploit different areas of competence.

In PUSBE, we employ a pairwise double-fault diversity measure [21] to select classifiers and prune the ensemble. The diversity measure is based on the idea that it is more important to know when simultaneous misclassifications occur than when both classifiers are correct. This is also well aligned with the problem of imbalanced classification, since the main priority there is to minimise the number of misclassifications of the minority class.

Given two base classifiers h_i and h_j, let $n(a,b)$ denote the number of training objects on which the output of these classifiers is a and b respectively. The double-fault diversity measure can then be calculated as

$$DIV_{DF}(h_i,h_j) = \frac{n(-1,-1)}{n(1,1)+n(-1,1)+n(1,-1)+n(-1,1)}. \tag{3}$$

Diversity for an ensemble of L base classifiers is then calculated by averaging the measure over all classifier pairs in the ensemble

$$DIV_{DF}(\Psi) = \frac{2}{NL(L-1)} \sum_{j=1}^{L} \sum_{k=j+1}^{L} n_{j,k}(-1,-1), \tag{4}$$

where L is the number of training samples. The established diversity measure is in the interval $[0;1]$, where 1 corresponds to a set of identical classifiers and 0 to the highest possible diversity respectively.

Classifier selection is achieved by an exhaustive search over all possible combinations of committee members to minimise the diversity measure function.

C. Classifier fusion

Classifier fusion is an important aspect of classifier ensembles, and the choice of fusion method, which is responsible for the collective decision making, is hence crucial. In our approach, we use a trained fuser for the classifier ensemble to arrive at an optimal set of weights w for Eq. (2).

In PUSBE, we employ a neural network as a trained classifier fusion approach based on decisions obtained from discriminant functions for the classifiers [8], illustrated in Fig. 2. One perceptron fuser is constructed for each of the classes under consideration. Once trained (we employ the Quickprop algorithm in our implementation), the input weights established during the learning process give the weights assigned to each of the base classifiers.

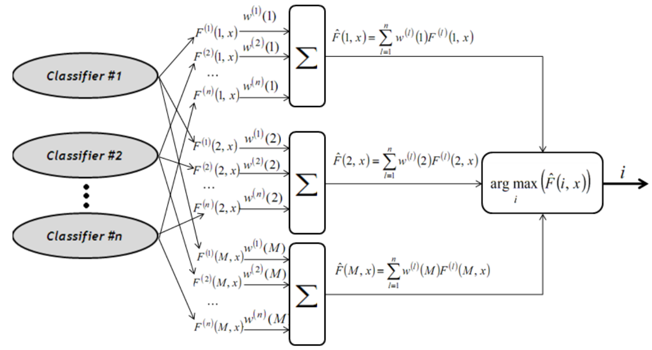

Fig. 2. Classifier fuser implemented as a one-layer neural network.

V. Experimental results

We evaluated our proposed ensemble approach on five datasets from the UCI repository[1], whose details are given in Table I.

As individual classifier, we used support vector machines (SVMs) [22] with a Gaussian RBF kernel, and performed classifier tuning [23] to obtain optimal parameters which were found to be $\sigma = 0.1$ and $C = 10$.

For comparison, we have also performed classification using several state-of-the-art ensembles dedicated to imbalanced classification, namely SMOTEBagging [13], SMOTEBoosting [24], EasyEnsemble [18], as well as the original USBE algorithm [7], all with identical base classifier models as used for PUSBE.

Results, based on a combined 5 x 2 CV F Test of statistical significance [25], are given in Table II, where we list the classification accuracies for the majority and minority class as well as the overall accuracy.

From Table II, we can see that our proposed method statistically outperformed all other ensembles dedicated to imbalanced classification in three out of five cases. At the same time, PUSBE never returned worse results than any of the examined methods. In most cases, we obtain a significant boost in terms of minority class recognition while maintaining overall classification accuracy. Using our proposed rebalancing scheme combined with feature selection is thus

[1] http://archive.ics.uci.edu/ml

TABLE I
Statistics of the datasets used in the experiments.

dataset	samples	features	class distribution [%]
Hepatitis	78	19	83–17
Heart Disease	296	13	54–46
Mushroom	300	22	62–38
Coil	2500	21	80–20
MRI	400	15	77–23

TABLE II

EXPERIMENTAL RESULTS ON THE FIVE UCI DATASETS. FOR EVERY ALGORITHM CLASSIFICATION ACCURACIES ON MINORITY AND MAJORITY CLASS ARE REPORTED TOGETHER WITH THE OVERALL CLASSIFICATION ACCURACY. EVERY SECOND LINE SHOWS COMPARED TO WHICH OTHER APPROACHES AN ALGORITHM WAS FOUND TO GIVE STATISTICALLY BETTER RESULTS (ON THE MINORITY CLASS).

	SMOTEBagging			SMOTEBoost			EasyEnsemble			USBE			PUSBE		
	minority	majority	overall	minority	majority	overall	minority	majority	overall	minority	majority	overall	minority	majority	overall
Hepatitis	62.56	75.56	73.35	63.46	76.12	73.97	65.23	76.48	74.57	65.45	75.11	73.46	66.75	75.73	74.20
		-			SMOTEBagging			SMOTEBagging,SMOTEBoost			SMOTEBagging,SMOTEBoost,EasyEnsemble			SMOTEBagging,SMOTEBoost,EasyEnsemble,USBE	
Heart Disease	73.43	82.11	78.11	76.75	82.76	80.00	76.21	82.64	79.68	76.68	81.89	79.49	77.03	82.04	79.73
		-			SMOTEBagging			SMOTEBagging			SMOTEBagging			SMOTEBagging	
Mushroom	84.05	93.20	89.72	84.78	93.86	90.40	86.74	93.53	90.95	86.97	93.03	90.72	87.90	93.12	91.13
		-			-			SMOTEBagging,SMOTEBoost			SMOTEBagging,SMOTEBoost,EasyEnsemble			SMOTEBagging,SMOTEBoost,EasyEnsemble,USBE	
Coil	64.95	94.11	88.28	66.94	95.26	89.60	69.43	94.76	89.69	67.88	93.79	88.61	69.18	94.44	89.38
		-			SMOTEBagging			SMOTEBagging,SMOTEBoost,USBE			SMOTEBagging,SMOTEBoost			SMOTEBagging,SMOTEBoost,USBE	
MRI	69.04	77.20	75.32	71.65	79.20	77.46	72.05	79.45	77.75	72.97	78.41	77.15	74.08	78.79	77.71
		-			SMOTEBagging			SMOTEBagging			SMOTEBagging,SMOTEBoost			SMOTEBagging,SMOTEBoost,EasyEnsemble,USBE	

shown to create a classifier pool of high quality which is further improved by employing the diversity-based classifier selection to prune the resulting ensemble which can then be effectively combined using the employed neural fuse. Finally, it should be noted that despite its impressive performance, our algorithm has no parameters that need to be set (apart from choosing and optimising the underlying base classifiers).

VI. CONCLUSIONS

In this paper, we have presented an effective multiple classifier system for imbalanced data classification. Our approach, Pruned Under-Sampling Balanced Ensemble (PUSBE), an extension to our earlier USBE algorithm, employs the concept of object space partitioning to train base classifiers on subspaces that contain an equal number of samples from both majority and minority classes. Redundant classifiers are then eliminated, based on a double-fault diversity measure, and the thus pruned ensemble fused using a perceptron neural network. Extensive experimental results on five benchmark datasets have shown our presented algorithm to outperform several state-of-the-art ensembles dedicated to imbalanced classification as well as to provide improved performance compared to the original USBE algorithm.

REFERENCES

[1] H. He and E. A. Garcia, “Learning from imbalanced data,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, pp. 1263–1284, 2009.

[2] W. Khreich, E. Granger, A. Miri, and R. Sabourin, “Iterative boolean combination of classifiers in the roc space: An application to anomaly detection with hmms,” *Pattern Recognition*, vol. 43, no. 8, pp. 2732–2752, 2010.

[3] Z. Yang, W. H. Tang, A. Shintemirov, and Q. H. Wu, “Association rule mining-based dissolved gas analysis for fault diagnosis of power transformers,” *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, vol. 39, no. 6, pp. 597–610, 2009.

[4] B. Krawczyk, L. Jelen, A. Krzyzak, and T. Fevens, “Oversampling methods for classification of imbalanced breast cancer malignancy data,” in *Computer Vision and Graphics*, ser. Lecture Notes in Computer Science, L. Bolc, R. Tadeusiewicz, L. Chmielewski, and K. Wojciechowski, Eds. Springer Berlin / Heidelberg, 2012, vol. 7594, pp. 483–490.

[5] Y. Liu and Y. Chen, “Face recognition using total margin-based adaptive fuzzy support vector machines,” *IEEE Transactions on Neural Networks*, vol. 18, no. 1, pp. 178–192, 2007.

[6] L. Kuncheva, *Combining pattern classifiers: Methods and algorithms*. Wiley-Interscience, New Jersey, 2004.

[7] B. Krawczyk and G. Schaefer, “Effective multiple classier systems for breast thermogram analysis,” in *21st Int. Conference on Pattern Recognition*, 2012.

[8] M. Woźniak and M. Zmyslony, “Designing combining classifier with trained fuser - analytical and experimental evaluation,” *Neural Network World*, vol. 20, no. 7, pp. 925–934, 2010.

[9] V. Lopez, A. Fernandez, J. G. Moreno-Torres, and F. Herrera, “Analysis of preprocessing vs. cost-sensitive learning for imbalanced classification. open problems on intrinsic data characteristics,” *Expert Systems with Applications*, vol. 39, no. 7, pp. 6585–6608, 2012.

[10] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “Smote: Synthetic minority over-sampling technique,” *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.

[11] B. Krawczyk, G. Schaefer, and M. Wozniak, “Combining one-class classifiers for imbalanced classification of breast thermogram features,” in *4th Int. Workshop on Computational Intelligence in Medical Imaging*, 2013, held as part of IEEE Symposium Series on Computational Intelligence.

[12] T. Nakashima, Y. Yokota, H. Ishibuchi, and G. Schaefer, “A cost-based fuzzy system for pattern classification with class importance,” *Artificial Life and Robotics*, vol. 12, no. 12, pp. 43–46, 2008.

[13] S. Wang and X. Yao, “Diversity analysis on imbalanced data sets by using ensemble models.”

[14] N. V. Chawla, A. Lazarevic, L. O. Hall, and K. W. Bowyer, “Smoteboost: Improving prediction of the minority class in boosting,” in *Lecture Notes in Artificial Intelligence*, vol. 2838, 2003, pp. 107–119.

[15] J. Blaszczynski, M. Deckert, J. Stefanowski, and S. Wilk, “Integrating selective pre-processing of imbalanced data with ivotes ensemble,” ser. Lecture Notes in Computer Science, 2010, vol. 6086 LNAI, pp. 148–157.

[16] Y. Sun, M. S. Kamel, A. K. C. Wong, and Y. Wang, “Cost-sensitive boosting for classification of imbalanced data,” *Pattern Recognition*, vol. 40, no. 12, pp. 3358–3378, 2007.

[17] B. Krawczyk, M. Wozniak, and G. Schaefer, “Improving minority class prediction using cost-sensitive ensembles,” in *16th Online World Conference on Soft Computing in Industrial Applications*, 2011.

[18] X. Liu, J. Wu, and Z. Zhou, “Exploratory undersampling for class-imbalance learning,” *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 39, no. 2, pp. 539–550, 2009.

[19] B. Krawczyk and G. Schaefer, “Evolutionary multiple classifier system based on space partitioning for breast thermogram analysis,” in *16th Online World Conference on Soft Computing in Industrial Applications*, 2011.

[20] L. Yu and H. Liu, “Efficient feature selection via analysis of relevance and redundancy,” *Journal of Machine Learning Research*, pp. 1205–1224, 2004.

[21] G. Giancinto and F. Roli, “Design of effective neural network ensembles for image classification purposes,” *Image Vision and Computing Journal*, vol. 19, pp. 699–707, 2001.

[22] V. Vapnik, *Statistical Learning Theory*. Willey, 1998.

[23] A. Karatzoglou, A. Smola, K. Hornik, and A. Zeileis, “Kernlab, an S4 package for kernel methods in R,” *Journal of Statistical Software*, vol. 11, no. 9, 2004.

[24] N. Chawla, A. Lazarevic, L. Hall, and K. Bowyer, “Smoteboost: Improving prediction of the minority class in boosting,” in *Lecture Notes in Artificial Intelligence*, vol. 2838, 2003, pp. 107–119.

[25] E. Alpaydin, “Combined 5 x 2 cv f test for comparing supervised classification learning algorithms,” *Neural Computation*, vol. 11, no. 8, pp. 1885–1892, 1999.