Honda Research Institute EU
Honda Research Institute Europe GmbH
https://www.honda-ri.de/

On the Performance of Oversampling Techniques for Class Imbalance Problems

Jiawen Kong, Thiago de Jesus de Araujo Rios, Wojtek Kowalczyk, Stefan Menzel, Thomas Bäck

2020

Preprint:

This is a post-peer-review, pre-copyedit version of an article published in PAKDD: Pacific-Asia Conference on Knowledge Discovery and Data Mining. The final authenticated version is available online at: https://doi.org/10.1007/978-3-030-47436-2_7

On the Performance of Oversampling Techniques for Class Imbalance Problems*

Jiawen Kong[1], Thiago Rios[2], Wojtek Kowalczyk[1], Stefan Menzel[2] and Thomas Bäck[1]

[1] Leiden University, Leiden, the Netherlands
{j.kong, w.j.kowalczyk, t.h.w.baeck}@liacs.leidenuniv.nl
[2] Honda Research Institute Europe GmbH, Offenbach, Germany
{thiago.rios,Stefan.Menzel}@honda-ri.de

* This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement number 766186.

The final publication is available at Springer via https://doi.org/10.1007/978-3-030-47436-2_7

On the Performance of Oversampling Techniques for Class Imbalance Problems*

Anonymous[1 [0000−1111−2222−3333]], Anonymous[1 [0000−1111−2222−3333]], Anonymous[1 [0000−1111−2222−3333]] and Anonymous[2,3 [1111−2222−3333−4444]]

[1] Anonymous
[2] Anonymous Anonymous
[3] Anonymous

Abstract. Although over 90 oversampling approaches have been developed in the imbalance learning domain, most of the empirical study and application work are still based on the "classical" resampling techniques. In this paper, several experiments on 19 benchmark datasets are set up to study the efficiency of six powerful oversampling approaches, including both "classical" and new ones. According to our experimental results, oversampling techniques that consider the minority class distribution (new ones) perform better in most cases and RACOG gives the best performance among the six reviewed approaches. We further validate our conclusion on our real-world inspired vehicle datasets and also find applying oversampling techniques can improve the performance by around 10%. In addition, seven data complexity measures are considered for the initial purpose of investigating the relationship between data complexity measures and the choice of resampling techniques. Although no obvious relationship can be abstracted in our experiments, we find F1v value, a measure for evaluating the overlap which most researchers ignore, has a strong negative correlation with the potential AUC value (after resampling).

Keywords: Class imbalance · Minority class distribution · Data complexity measures.

1 Introduction

The classification problem under class imbalance has caught growing attention from both the academic and the industrial field. Due to recent advances, the progress in technical assets for data storage and management as well as in data science enables practitioners from industry and engineering to collect a large amount of data with the purpose of extracting knowledge and acquire hidden insights. An example may be illustrated from the field of computational design optimization where product parameters are modified to generate digital prototypes which performances are evaluated by numerical simulations, or based

* Supported by organization x.

on equations expressing human heuristics and preferences. Here, many parameter variations usually result in valid and producible geometries but in the final steps of the optimization, i.e. in the area where the design parameters converge to a local/global optimum, some geometries are generated which violate given constraints. Under this circumstance, a database would contain a large number of designs which are according to specs (even if some may be of low performance) and a smaller number of designs which eventually violate pre-defined product requirements. By far, the resampling techniques have proven to be efficient in handling imbalanced benchmark datasets. However, the empirical study and application work in the imbalanced learning domain are mostly focusing on the "classical" resampling techniques (SMOTE, ADASYN, and MWMOTE etc.) [11, 15, 20], although there are many recently developed resampling techniques.

In this paper, we set up several experiments on 19 benchmark datasets to study the efficiency of six powerful oversampling techniques, including SMOTE, ADASYN, MWMOTE, RACOG, wRACOG and RWO-Sampling. For each dataset, we also calculate seven data complexity measures to investigate the relationship between data complexity measures and the choice of resampling techniques, since researchers have pointed out that studying the data complexity of the imbalanced datasets is of vital importance [15] and it may affect the choice of resampling techniques [20]. We also perform the experiment on our real-world inspired vehicle dataset. Results of our experiments demonstrate that oversampling techniques that consider the minority class distribution (RACOG, wRACOG, RWO-Sampling) perform better in most cases and RACOG gives the best performance among the six reviewed approaches. Results on our real-world inspired vehicle dataset further validate this conclusion. No obvious relationship between data complexity measures and the choice of resampling techniques is found in our experiment. However, we find F1v value, a measure for evaluating the overlap which most researchers ignore [15,20], has a strong negative correlation with the potential AUC value (after resampling).

The remainder of this paper is organized as follows. In Section 2, the research related to our work are presented, also including the relevant background knowledge on six resampling approaches and data complexity measures. In Section 3, the experimental setup is introduced in order to understand how the results are generated. Section 4 gives the results of our experiments. Further exploration through data from a real-world inspired digital vehicle model is presented in Section 5. Section 6 concludes the paper and outlines further research.

2 Related Works

Many effective oversampling approaches have been developed in the imbalanced learning domain and the synthetic minority oversampling technique (SMOTE) is the most famous one among all. Currently, more than 90 SMOTE extensions have been published in scientific journals and conferences [6]. Most of review paper and application work are based on the "classical" resampling techniques and

do not take new resampling techniques into account. In this paper, we briefly review six powerful oversampling approaches, including both "classical" ones (SMOTE, ADASYN, MWMOTE) and new ones (RACOG, wRACOG, RWO-Sampling) [2, 3, 5, 7, 24]. The six reviewed oversampling techniques can be divided into two groups according to whether they consider the overall minority class distribution. Among the six approaches, RACOG, wRACOG, and RWO-Sampling consider the overall minority class distribution while the other three not. Apart from developing new approaches to solve class-imbalance problem, various studies have pointed out that it is important to study the characteristics of the imbalanced dataset [13, 20]. In [13], authors emphasize the importance to study the overlap between the two-class samples. In [20], authors set up several experiments with the KEEL benchmark datasets [1] to study the relationship between various data complexity measures and the potential AUC value. It is also pointed out in [20] that the distinctive inner procedures of oversampling approaches are suitable for particular characteristics of data. Hence, apart from evaluate the efficiency for the six reviewed oversampling approaches, we also aim to investigate the relationship between data complexity measures and the choice of resampling techniques.

2.1 Resampling Technique

In the following, the six established resampling techniques SMOTE, ADASYN, MWMOTE, RACOG, wRACOG and RWO-Sampling are introduced.

SMOTE and ADASYN The synthetic minority oversampling technique (SMOTE) is the most famous resampling technique [3]. SMOTE produces synthetic minority samples based on the randomly chosen minority samples and their K-nearest neighbors. The new synthetic sample can be generated by using the randomized interpolation scheme above for minority samples. The main improvement in the adaptive synthetic (ADASYN) sampling technique is that the samples which are harder to learn are given higher importance and will be oversampled more often in ADASYN [7].

MWMOTE The majority weighted minority oversampling techniques (MWMOTE) improves the sample selection scheme and the synthetic sample generation scheme [2]. MWMOTE first finds the informative minority samples (S_{imin}) by removing the "noise" minority samples and finding the borderline majority samples. Then, every sample in S_{imin} is given a selection weight (S_w), according to the distance to the decision boundary, the sparsity of the located minority class cluster and the sparsity of the nearest majority class cluster. These weights are converted in to selection probability (S_p) in the synthetic sample generation stage. The cluster-based synthetic sample generation process proposed in MWMOTE can be described as, 1). cluster all samples in S_{imin} into M groups; 2). select a minority sample x from S_{imin} according to S_p and randomly select another sample y from the same cluster of x; 3). use the same equation employed

in k-NN-based approach to generate the synthetic sample; 4). repeat 1) – 3) until the required number of synthetic samples is generated.

RACOG and wRACOG The oversampling approaches can effectively increase the number of minority class samples and achieve a balanced training dataset for classifiers. However, the oversampling approaches introduced above heavily reply on *local* information of the minority class samples and do not take the overall distribution of the minority class into account. Hence, the *global* information of the minority samples cannot be guaranteed. In order to tackle this problem, Das et al. [5] proposed RACOG (RApidy COnverging Gibbs) and wRACOG (Wrapper-based RApidy COnverging Gibbs).

In these two algorithms, the n-dimensional probability distribution of minority class is optimally approximated by Chow-Liu's dependence tree algorithm and the synthetic samples are generated from the approximated distribution using Gibbs sampling. Instead of running an "exhausting" long Markov chain, the two algorithms produce multiple relatively short Markov chains, each starting with a different minority class sample. RACOG selects the new minority samples from the Gibbs sampler using a predefined *lag* and this selection procedure does not take the usefulness of the generated samples into account. On the other hand, wRACOG considers the usefulness of the generated samples and selects those samples which have the highest probability of being misclassified by the existing learning model [5].

RWO-Sampling Inspired by the central limit theorem, Zhang et al. [24] proposed the random walk oversampling (RWO-Sampling) approach to generate the synthetic minority class samples which follows the same distribution as the original training data.

In order to add m synthetic examples to the n original minority examples ($m < n$), we first select at random m examples from the minority class and then for each of the selected examples $\boldsymbol{x} = (x_1, ..., x_m)$ we generate its synthetic counterpart by replacing $a_i(j)$ (the ith attribute in x_j, $j \in 1, 2, ..., m$) with $\mu_i - r_i \cdot \sigma_i/\sqrt{n}$, where μ_i and σ_i denote the mean and the standard deviation of the ith feature restricted to the original minority class, and r_i is a random value drawn from the standard normal distribution. When $m > n$, we can repeat the above process until we reach the required amount of synthetic examples. Since the synthetic sample is achieved by randomly walking from one real sample, so this oversampling is called random walk oversampling.

2.2 Data Complexity Measures

In this section, we introduce the *feature overlapping measures* and *linearity measures* among various data complexity measures (Table 1).

Feature Overlapping Measures F1 measures the highest discriminant ratio among all the features in the dataset [14]. F1v is a complement of F1 and a

Table 1. Complexity measures information. "Positive" and "Negative" indicate the positive and negative relation between measure value and data complexity respectively.

Measure	Description	Complexity
F1	Maximum Fisher's Discriminant Ratio	Negative
F1v	The Directional-vector Maximum Fisher's Discriminant Ratio	Negative
F2	Volume of Overlapping Region	Positive
F3	Maximum Individual Feature Efficiency	Negative
L1	Sum of the Error Distance by Linear Programming	Positive
L2	Error Rate of Linear Classifier	Positive
L3	Non-Linearity of a Linear Classifier	Positive

higher value of F1v indicates there exists a vector that can separate different class samples after these samples are projected on it [19]. F2 calculates the overlap ratio of all features (the width of the overlap interval to the width of the entire interval) and returns the product of the ratios of all features [19]. F3 measures the individual feature efficiency and returns the maximum value among all features.

Linearity Measures L1 and L2 both measure to what extent the classes can be linearly separated using an SVM with a linear kernel [19], where L1 returns the sum of the distances of the misclassified samples to the linear boundary and L2 returns the error rate of the linear classifier. L3 returns the error rate of an SVM with linear kernel on a test set, where the SVM is trained on training samples and the test set is manually created by performing linear interpolation on the two randomly chosen samples from the same class.

3 Experimental Setup

The experiments reported in this paper are based on 19 two-class imbalanced datasets from the KEEL-collection [1] and six powerful oversampling approaches (using R package `imbalance` [4]), which have been reviewed in Section 2.1. The collected datasets are divided into 5 stratified folds (for cross-validation) and only the training set is oversampled, where the stratified fold is to ensure the imbalance ratio in the training set is consistent with the original dataset and only oversampling the training set is to avoid over-optimism problem [14].

The 19 collected datasets can be simply divided into 4 groups,*ecoli*, *glass*, *vehicle* and *yeast* (Table 2). IR indicates the imbalance ratio, which is the ratio of the number of majority class samples to the number of minority class samples. In this paper, we aim to study the efficiency of different oversampling approaches and investigate the relationship between data complexity measures and the choice of oversampling techniques. Therefore, we need to calculate the

Table 2. Information on datasets in 4 groups

Datasets	#Attributes	#Samples	Imbalance Ratio (IR)
ecoli{1,2,3,4}	7	336	{3.36, 5.46, 8.6, 15.8 }
glass{0,1,2,4,5,6}	9	214	{2.06, 1.82, 11.59, 15.47, 22.78, 6.38 }
vehicle{0,1,2,3}	18	846	{3.25, 2.9, 2.88, 2.99 }
yeast{1,2,3,4,5,6}	8	1484	{2.46, 8.1, 28.1, 32.73, 41.4 }

7 data complexity measures (shown in Table 1) for each dataset. In our 20 experiments for each dataset, we calculate the 7 data complexity measures for every training set (using R package `ECoL` [14]). Since we use 5 stratified cross-validations, we average each data complexity measures for these 5 training sets and make it the data complexity measure for the dataset.

In a binary classification problem, the confusion matrix (see Table 3) can provide intuitive classification results. In the class imbalance domain, it is widely admitted that *Accuracy* tends to give deceptive evaluation for the performance. Instead of *Accuracy*, the Area Under the ROC Curve (AUC) can be used to evaluate the performance [13] and can be computed as $AUC = \frac{1+TP_{rate}-FP_{rate}}{2}$, where $TP_{rate} = \frac{TP}{TP+FN}, FP_{rate} = \frac{FP}{FP+TN}$. Apart from the AUC value, there are some other measures to assess the performance for imbalanced datasets, such as geometric mean (GM) and F-measure (FM) [13].

Table 3. Confusion matrix for a binary classification problem

	Positive prediction	**Negative prediction**
Positive class	True Positives (TP)	False Negatives (FN)
Negative class	False Positives (FP)	True Negatives (TN)

4 Simulation Analysis and Discussions

Due to the limited space, only the AUC results for C5.0 decision tree in our experiments are presented in Table 4. We can observe that RACOG outperforms the other 5 oversampling techniques in 9 out of 19 datasets and MWMOTE is the 2nd best oversampling approaches. From our experimental results, we can conclude that, in most cases, oversampling approaches which consider the minority class distribution (RACOG, wRACOG and RWO-Sampling) perform better. It was expected that data complexity can provide some guidance for choosing the oversampling technique, however, from our experimental results, no obvious relationship between data complexity and the choice of oversampling approaches can be concluded. This is because the 6 introduced oversampling approaches are designed for common datasets and do not take a specific data characteristic into account.

According to our experimental results, although the data complexity measures cannot provide guidance for choosing the oversampling approaches, we find

Table 4. AUC results for C5.0 decision tree.

Dataset	Baseline	SMOTE	ADASYN	MWMOTE	RACOG	wRACOG	RWO
ecoli1	0.9418	0.9407	0.9364	0.9399	**0.9471**	0.9390	0.9423
ecoli2	0.8598	0.9019	0.9140	0.9071	0.9144	0.8959	**0.9168**
ecoli3	0.7795	0.9049	0.8943	0.8991	**0.9098**	0.8757	0.9001
ecoli4	0.8172	0.9247	0.9080	**0.9286**	0.9169	0.8875	0.9038
glass0	0.8286	0.8476	0.8377	0.8435	0.8442	0.8463	**0.8515**
glass1	0.7082	0.7496	0.7338	0.7489	**0.7500**	0.7421	0.6986
glass2	0.7264	0.8024	**0.8072**	0.7925	0.7990	0.7862	0.7173
glass4	0.8468	0.9291	0.9273	**0.9364**	0.9255	0.8648	0.9322
glass5	0.9905	0.9904	0.9903	0.9905	**0.9927**	0.9915	0.9916
glass6	0.9332	0.9337	0.9310	0.9357	**0.9384**	0.9378	0.9340
vehicle0	0.9734	0.9743	0.9725	0.9731	**0.9753**	0.9751	0.9675
vehicle1	0.7639	0.8031	0.7974	0.7987	0.8035	**0.8094**	0.7810
vehicle2	0.9742	0.9715	0.9741	0.9741	**0.9787**	0.9777	0.9772
vehicle3	0.7756	0.8045	0.8006	**0.8196**	0.8169	0.8166	0.7956
yeast1	0.7317	0.7437	0.7386	0.7449	**0.7585**	0.7109	0.7166
yeast3	0.9357	0.9584	0.9591	0.9600	**0.9647**	0.9564	0.9450
yeast4	0.7592	**0.9030**	0.9001	0.8940	0.8669	0.8245	0.8286
yeast5	0.9574	0.9774	0.9768	**0.9782**	0.9775	0.9727	0.9782
yeast6	0.7472	0.8760	0.8825	0.8825	0.8802	0.8085	**0.8851**

there is a strong correlation between the potential best AUC (after oversample) and some of the data complexity measures. From Figure 1 and Table 5, it can be concluded that the potential best AUC value that can be achieved through oversampling techniques has an extreme negative correlation with the F1v value and linearity measures. In the imbalanced learning domain, there are many researchers focus on studying data complexity measures. In [14], the authors propose that the potential best AUC value after resampling can be predicted through various data complexity measures. In [10], the authors demonstrate that F1 value has an influence on the potential improvement brought by oversampling approaches. However, they did not consider the F1v measure, which has the strongest correlation with AUC value. Hence, we recommend using F1v to evaluate the overlap in imbalanced dataset.

Fig. 1. Correlation matrix.

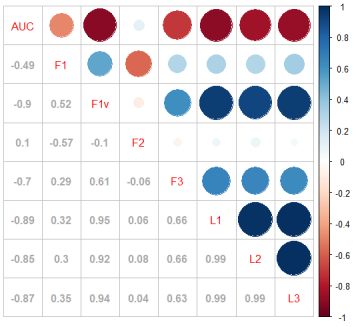

Table 5. Results of Hypothesis Test.

Measure	Correlation Coefficient	P-value	Correlation Level
F1	-0.4878	0.0341	medium
F1v	-0.9048	1.041×10^{-7}	extreme
F2	0.1018	0.6782	none
F3	-0.7019	0.0008	high
L1	-0.8913	3.054×10^{-7}	extreme
L2	-0.8471	4.735×10^{-6}	extreme
L3	-0.8693	1.354×10^{-6}	extreme

5 Efficient Oversampling Strategies For Improved Vehicle Mesh Quality Classification

In this section, we propose the application of the reviewed methods on the quality prediction of geometric computer aided engineering (CAE) models. In CAE applications, engineers often discretize the simulation domains using meshes (undirected graphs), i.e. a set of nodes (vertices), where the equations that describe the physical phenomena are solved, and edges connecting the nodes to form faces and volumes (elements), where the solution between nodes is approximated. The meshes are generated from an initial geometric representation, e.g. non-uniform rational B-Splines (NURBS) or stereolithography (STL) representations, using numerical algorithms, such as sweep-hull for Delaunay triangulation [23], polycube [12] etc.

In most cases the quality of the mesh plays an important role on the accuracy and fidelity of the results [9]. Engineers use different types of metrics to infer about the quality of the mesh, but it is common sense that increasing the number and uniformity of the elements in the mesh improves the accuracy of the simulation results. However, the computational effort associated with meshing is proportional to the target level of refinement. Therefore, a match between accuracy and available computational resources is often required, specially for cases that demand iterative geometric modifications, such as shape optimization.

Shape morphing techniques address this issue by operating on the mesh nodes through a polynomial-based lower-dimensional representation. Such techniques avoid re-meshing the simulation domain, speeding up the optimization process. Several cases of optimization using morphing techniques are published in the literature [16–18, 22]. For our experiments, we implemented the free form deformation (FFD) method presented in [21]. The FFD embeds the geometry of interest in a uniform parallelepiped lattice, where a trivariate Bernstein polynomial maps the position of the control points of the lattice to the nodes of the mesh, as an $\mathbb{R}^3 \rightarrow \mathbb{R}^3$ function. Therefore, by deforming the lattice, the nodes of the mesh are moved accordingly (Figure 2).

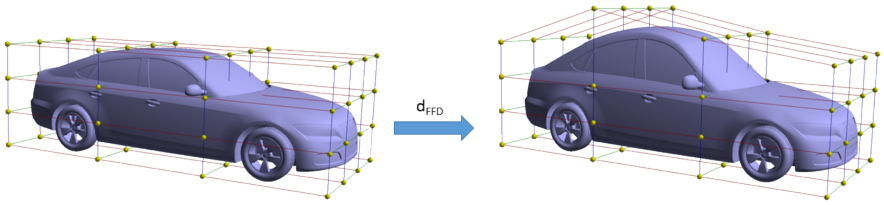

Fig. 2. Example of free form deformation applied to a configuration of the TUM DrivAer model [8] using a lattice with four planes in each direction.

The continuity of the surfaces is ensured by the mathematical formulation of the FFD up to the order of $k-1$, where k is the number of planes in the direc-

tion of interest, but the mesh quality is not necessarily maintained. The designer can either avoid models with ill-defined elements by applying constraints to the deformations, which might be unintuitive, or eliminate them by performing regular quality assessments. Addressing this issue, we propose the classification of the deformation parameters with respect to the quality of the output meshes, based on a data set of labeled meshes. Further than reducing the risk generating infeasible meshes for CAE applications, our approach avoids unnecessary computation to generate the deformed meshes, which is aligned with the objective of increasing the efficiency of shape optimization tasks.

5.1 Generation of a synthetic data set

For the experiments we adopted the computer fluid dynamics (CFD) simulation of a configuration of the TUM DrivAer model [8]. The simulation model is deformed using the discussed FFD algorithm, using a lattice with 7 planes in x- and z-directions, and 10 in y-direction (Fig. 3). The planes closer to the boundaries of the control volume are not displaced in order to enable a smooth transition from the region affected by the deformations to the original domain. Assuming symmetry of the shape with respect to the vertical plane (xz) and deformations caused by displacement of entire control planes only in the direction of their normal vectors, it yields a design space with 9 parameters. To generate the data set, the displacements x_i were sampled from a random uniform distribution and constrained to the volume of the lattice, allowing the overlap of planes.

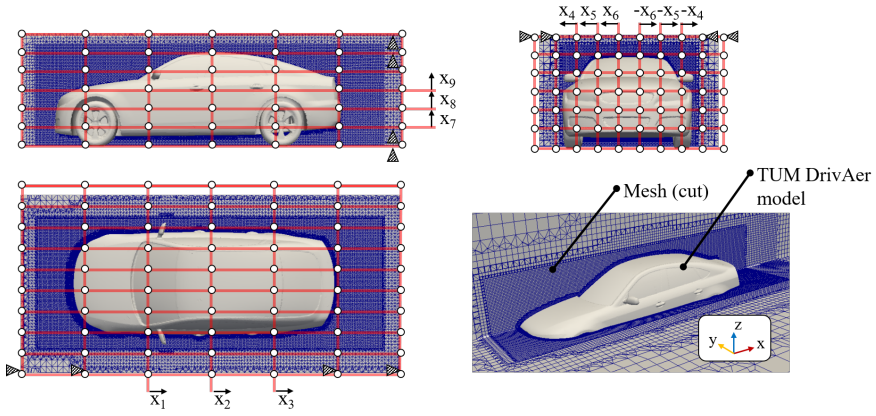

Fig. 3. Free form deformation lattice used to generate the data set for the experiments.

The initial mesh was generated using the algorithms *blockMesh* and *snappyHexMesh* of OpenFOAM®[4]. We automatically generated 300 meshes based

[4] https://www.openfoam.com

on the FFD algorithm implemented in python and evaluated them using the OpenFOAM checkMesh rounting. The quality of the meshes was verified using the *checkMesh* routine, also available in OpenFOAM®, and we generated 300 deformed meshes. In the process, 6 meshes were discarded due to errors in the meshing process. The metrics used to define the quality of the meshes were the number of warnings raised by the *meshCheck* algorithm, the maximum skewness and maximum aspect ratio. We manually labeled the feasible meshes according to the rules shown in Table 6. The imbalance ratios after manually labeling are also given in Table 6. Please note that the input attributes are exactly the same for all three sets of datasets, only the "class" labels are different. In this way, the values of data complexity measures for the three datasets vary from each other.

Table 6. Feasible meshes labeling rule.

Dataset	#Attribute	#Sample	#Warnings	Max skewness	Max aspect ratio	IR
set1	9	294	<4	<6	<10	6.35
set2	9	294	<4	<6.2	<10.5	2.54
set3	9	294	<2	<5.8	<10.3	12.36

5.2 Results and Discussion

The experimental results on the digital vehicle dataset are given in Table 7. It is consistent with the conclusion we draw in Section 4 that, RACOG outperforms the other 5 oversampling techniques in 2 out of 3 datasets. Therefore, combining our experimental results on both benchmark and real-world inspired datasets, we can conclude RACOG is the most powerful one of the considered 6 oversampling approaches. Moreover, we find that applying the oversampling techniques can improve the performance by around 10% for our digital vehicle datasets. We also calculate the data complexity measures for our digital vehicle datasets, our findings on the correlation between the potential AUC value and the data complexity measures remains consistent with the conclusion in Section 4.

Table 7. Experimental Results (AUC) on Digital Vehicle Dataset.

Dataset	Baseline	SMOTE	ADASYN	MWMOTE	RACOG	wRACOG	RWO
set1	0.7786	0.8412	0.8315	0.8354	**0.8543**	0.8406	0.8502
set2	0.6952	0.7575	0.7560	**0.7651**	0.7614	0.7421	0.7452
set3	0.6708	0.7780	0.7792	0.7660	**0.7823**	0.7534	0.7743

6 Conclusion and Future Work

In this work, we reviewed six powerful oversampling techniques, including "classical" ones (SMOTE, ADASYN and MWMOTE) and new ones (RACOG, wRACOG and RWO-Sampling), in which the new ones consider the minority class distribution while the "classical" ones not. The six reviewed oversampling approaches

were performed on 19 benchmark imbalanced datasets and an imbalanced real-world inspired vehicle dataset to investigate their efficiency. Seven data complexity measures were considered in order to find the relationship between data complexity measures and the choice of resampling techniques. According to our experimental results, two main conclusions can be derived:

1) In our experiment, in most cases, oversampling approaches which consider the minority class distribution (RACOG, wRACOG and RWO-Sampling) perform better. For both benchmark datasets and our real-world inspired dataset, RACOG performs best and MWMOTE comes to the second.
2) No obvious relationship between data complexity measures and the choice of resampling techniques can be abstracted from our experimental results. However, we find F1v value has a strong correlation with the potential best AUC value (after resampling) while rare researchers in the imbalance learning domain do not consider F1v value for evaluating the overlap between classes.

We only simply applying the oversampling techniques for our digital vehicle dataset and evaluate their efficiency in this paper. In future work, we will focus on adjusting the imbalance learning algorithms to solve the proposed engineering problem. Additionally, the effect of the interaction between various data complexity measures on the choice of resampling technique will be studied.

References

1. Alcalá-Fdez, J., Fernández, A., Luengo, J., Derrac, J., García, S., Sánchez, L., Herrera, F.: Keel data-mining software tool: data set repository, integration of algorithms and experimental analysis framework. Journal of Multiple-Valued Logic & Soft Computing **17** (2011)
2. Barua, S., Islam, M.M., Yao, X., Murase, K.: Mwmote–majority weighted minority oversampling technique for imbalanced data set learning. IEEE Transactions on Knowledge and Data Engineering **26**(2), 405–425 (2012)
3. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: Smote: synthetic minority over-sampling technique. Journal of artificial intelligence research **16**, 321–357 (2002)
4. Cordón, I., García, S., Fernández, A., Herrera, F.: Imbalance: oversampling algorithms for imbalanced classification in r. Knowledge-Based Systems **161**, 329–341 (2018)
5. Das, B., Krishnan, N.C., Cook, D.J.: Racog and wracog: Two probabilistic oversampling techniques. IEEE transactions on knowledge and data engineering **27**(1), 222–234 (2014)
6. Fernández, A., García, S., Galar, M., Prati, R.C., Krawczyk, B., Herrera, F.: Learning from imbalanced data sets. Springer (2018)
7. He, H., Bai, Y., Garcia, E.A., Li, S.: Adasyn: Adaptive synthetic sampling approach for imbalanced learning. In: 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence). pp. 1322–1328. IEEE (2008)
8. Heft, A.I., Indinger, T., Adams, N.A.: Experimental and numerical investigation of the drivaer model. In: ASME 2012 Fluids Engineering Division Summer Meeting. pp. 41–51. American Society of Mechanical Engineers Digital Collection (2012)

9. Knupp, P.: Measurement and impact of mesh quality. In: 46th AIAA Aerospace Sciences Meeting and Exhibit. p. 933 (2008)
10. Kong, J., Kowalczyk, W., Nguyen, D.A., Menzel, S., Bäck, T.: Hyperparameter optimisation for improving classification under class imbalance. In: 2019 IEEE Symposium Series on Computational Intelligence (SSCI). IEEE (2019)
11. Li, J., Liu, L.s., Fong, S., Wong, R.K., Mohammed, S., Fiaidhi, J., Sung, Y., Wong, K.K.: Adaptive swarm balancing algorithms for rare-event prediction in imbalanced healthcare data. PloS one **12**(7), e0180830 (2017)
12. Livesu, M., Vining, N., Sheffer, A., Gregson, J., Scateni, R.: Polycut: Monotone graph-cuts for polycube base-complex construction. Transactions on Graphics (Proc. SIGGRAPH ASIA 2013) **32**(6) (2013)
13. López, V., Fernández, A., García, S., Palade, V., Herrera, F.: An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics. Information sciences **250**, 113–141 (2013)
14. Lorena, A.C., Garcia, L.P., Lehmann, J., Souto, M.C., Ho, T.K.: How complex is your classification problem?: A survey on measuring classification complexity. ACM Computing Surveys (CSUR) **52**(5), 107 (2019)
15. Luengo, J., Fernández, A., García, S., Herrera, F.: Addressing data complexity for imbalanced data sets: analysis of smote-based oversampling and evolutionary undersampling. Soft Computing **15**(10), 1909–1936 (2011)
16. Menzel, S., Olhofer, M., Sendhoff, B.: Application of free form deformation techniques in evolutionary design optimisation. In: Herskovits, J., Mazorche, S., Canelas, A. (eds.) 6th World Congress on Structural and Multidisciplinary Optimization (WCSMO6). COPPE Publication, Rio de Janeiro (2005)
17. Menzel, S., Sendhoff, B.: Representing the Change - Free Form Deformation for Evolutionary Design Optimization, pp. 63–86. Springer Berlin Heidelberg (2008)
18. Olhofer, M., Bihrer, T., Menzel, S., Fischer, M., Sendhoff, B.: Evolutionary optimisation of an exhaust flow element with free form deformation. In: 4th European Automotive Simulation Conference, Munich (2009)
19. Orriols-Puig, A., Macia, N., Ho, T.K.: Documentation for the data complexity library in c++. Universitat Ramon Llull, La Salle **196**, 1–40 (2010)
20. Santos, M.S., Soares, J.P., Abreu, P.H., Araujo, H., Santos, J.: Cross-validation for imbalanced datasets: Avoiding overoptimistic and overfitting approaches [research frontier]. ieee ComputatioNal iNtelligeNCe magaziNe **13**(4), 59–76 (2018)
21. Sederberg, T.W., Parry, S.R.: Free-form deformation of solid geometric models. ACM SIGGRAPH computer graphics **20**(4), 151–160 (1986)
22. Sieger, D., Menzel, S., Botsch, M.: On shape deformation techniques for simulation-based design optimization. In: New Challenges in Grid Generation and Adaptivity for Scientific Computing, pp. 281–303. Springer (2015)
23. Sinclair, D.: S-hull: a fast radial sweep-hull routine for delaunay triangulation. arXiv preprint arXiv:1604.01428v1 [cs.CG] (2016)
24. Zhang, H., Li, M.: Rwo-sampling: A random walk over-sampling approach to imbalanced data classification. Information Fusion **20**, 99–116 (2014)