Imbalanced Data Classification: A Novel Re-sampling Approach Combining Versatile Improved SMOTE and Rough Sets

Katarzyna Borowska, Jaroslaw Stepaniuk

Imbalanced data classification: a novel re-sampling approach combining Versatile Improved SMOTE and Rough Sets

Katarzyna Borowska, Jarosław Stepaniuk

Faculty of Computer Science
Bialystok University of Technology
Wiejska 45A, 15-351 Bialystok, Poland
{k.borowska,j.stepaniuk}@pb.edu.pl
http://www.wi.pb.edu.pl

Abstract. In recent years, the problem of learning from imbalanced data has emerged as important and challenging. The fact that one of the classes is underrepresented in the data set is not the only reason of difficulties. The complex distribution of data, especially small disjuncts, noise and class overlapping, contributes to the significant depletion of classifier's performance. Hence, the numerous solutions were proposed. They are categorized into three groups: data–level techniques, algorithm–level methods and cost–sensitive approaches. This paper presents a novel data–level method combining Versatile Improved SMOTE and rough sets. The algorithm was applied to the two–class problems, data sets were characterized by the nominal attributes. We evaluated the proposed technique in comparison with other preprocessing methods. The impact of the additional cleaning phase was specifically verified.

Keywords: Data preprocessing, Class imbalance, Rough Sets, SMOTE, Oversampling, Undersampling.

1 Introduction

Proper classification of imbalanced data is one of the most challenging problems in data mining. Since wide range of real–world domains suffers from this issue, it is crucial to find more and more effective techniques to deal with it. The fundamental reason of difficulties is the fact that one class (positive, minority) is underrepresented in the data set. Furthermore, the correct recognition of examples belonging to this particular class is a matter of major interest. Considering domains like medical diagnostic, anomaly detection, fault diagnosis, detection of oil spills, risk management and fraud detection [8],[21] the misclassification cost of rare cases is obviously very high. The small subset of data describing disease cases is more meaningful than remaining majority of objects representing healthy population. Therefore, the dedicated algorithms should be applied to recognizing minority class instances in these areas.

Over the last years the researchers' growing interest in imbalanced data contributed to considerable advancements in this field. Numerous methods were proposed to address this problem. They are grouped into three main categories [8],[21]:

- data–level techniques: adding the preliminary step of data processing - assumes mainly undersampling and oversampling,
- algorithm–level approaches: modifications of existing algorithms,
- cost–sensitive methods: combining data–level and algorithm–level techniques to set different misclassification costs.

In this paper we focus on data–level approaches: generating new minority class samples (oversampling) and introducing additional cleaning step (undersampling). Creating new examples of the minority class requires careful analysis of the data distribution. Random replication of the positive instances may lead to overfitting [8]. Furthermore, even applying methods like Synthetic Minority Oversampling Technique [5] (creation of new samples by interpolating several minority class examples that lie together) may not be sufficient for variety of real–life domains. Indeed, the main reason of difficulties in learning from imbalanced data is the complex distribution: existence of class overlapping, noise or small disjuncts [8], [13], [11], [15].

The VIS algorithm [4], incorporated into the proposed approach, addresses listed problems by applying dedicated mechanism for each specific group of minority class examples. Assigning objects into categories is based on their local characteristics. Although this solution considers additional difficulties, in case of eminently complex problems it may contribute to creation of noisy objects. Hence, the clearing mechanism is introduced as the second step of preprocessing. On the other hand, new preliminary step deals with uncertainty by relabeling ambiguous majority data. All negative (majority) instances belonging to the boundary region defined by the rough sets theory [20], [16] are relabeled to the positive class. Novel technique was developed to verify the impact of inconsistencies in data sets on the classifier performance. Only data sets described by nominal attributes were examined. However, discretization of attributes may allow applying proposed solutions to data including continuous values.

Although only the preprocessing techniques are discussed, we need to mention that there are numerous effective methods belonging to other categories, such as BRACID [14] (algorithm–level) or AdaC2 [21] (cost–sensitive).

2 Preprocessing algorithms overview

Since SMOTE algorithm [5] is based on the k–NN method, it is not deprived of some drawbacks related to the k–NN performance. Primarily, the k-NN technique is extremely sensitive to data complexity [9]. Especially class overlapping, noise or small disjuncts existing in imbalanced data negatively affects the performance of distance–based algorithms. Considering scenario of generating new minority examples by interpolating two minority instances that belong to different clusters

(but were recognised as nearest neighbors), it is likely that new object will overlap with an example of majority class [19]. Hence, applying SMOTE to some domains may cause creating incorrect synthetic samples that fall into majority regions [2]. Methods like MSMOTE [12], Bordeline–SMOTE [10], VIS [4] were developed to address this problem. They assume that there are inconsistencies in data set and identify specific groups of minority class instances to select the most appropriate strategy of preprocessing.

On the other hand, there are numerous proposals of hybrid re–sampling methods. They combine oversampling with undersampling to ensure that improper newly–generated examples will be excluded before applying classifier. SMOTE–Tomek links and SMOTE-ENN [3] introduce the additional cleaning step to original SMOTE processing. The SMOTE–RSB$_*$ algorithm [17] eliminates overfitting by application of the rough sets theory and lower approximation of a subset. Defining the lower approximation of the minority class enables to remove generated synthetic samples that are presumably noise.

The rough set theory was also the inspiration for developing techniques discussed below. They are dedicated to data sets described by nominal attributes.

2.1 Rough set based remove and relabel techniques

The method proposed in [18] considers applying the rough sets theory to obtain the inconsistences in imbalanced data. The fundamental assumption of the rough set approach is that objects from a set U described by the same information are indiscernible. This main concept is source of the notion referred as indiscernibility relation $IND \subseteq U \times U$, defined on the set U. Let $[x]_{IND} = \{y \in U : (x, y) \in IND\}$ be an indiscernibility class, where $x \in U$. For any subset X of the set U it is possible to prepare the following characteristics [16]:

- the lower approximation of a set X: all examples that can be certainly classified as members of X with respect to IND;

$$\{x \in U : [x]_{IND} \subseteq X\}, \tag{1}$$

- the boundary region of a set X: all instances that are possibly members of X set with respect to IND;

$$\{x \in U : [x]_{IND} \cap X \neq \varnothing \ \&[x]_{IND} \nsubseteq X\}. \tag{2}$$

In described method two filtering techniques based on the presented rough set concepts were developed. Both of them require calculation of boundary region of minority class. Next step depends on the chosen method. The first one removes majority class examples belonging to the minority class boundary region that contains inconsistent objects. The second technique relabels all majority objects that belong to the minority class boundary region.

The figure 1 illustrates results of applying two described methods on artificial data. It also demonstrates the boundary region (with 16 objects) of minority class in the original data set (dashed line).

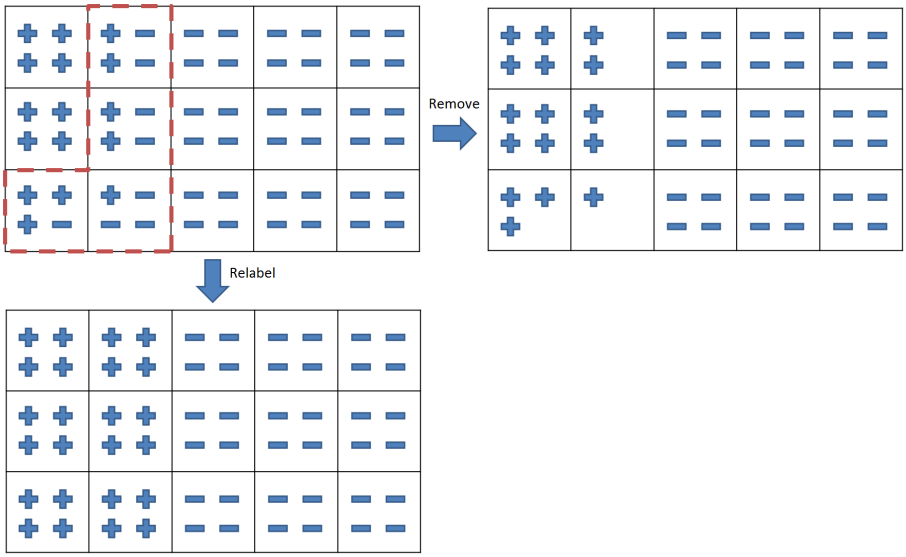

Fig. 1. Example of artificial data (60 objects, 15 indiscernibility classes, imbalance ratio $IR = 2.75$) described by two nominal attributes with three and five values. Data after filtering by the "Remove" technique ($IR = 2.25$). Data after applying "Relabel" technique ($IR = 1.5$).

2.2 Versatile Improved SMOTE and rough sets (VIS_RST)

The main idea of this new approach is to apply two preprocessing methods: oversampling and undersampling in order to generate minority class instances and ensure that no additional inconsistencies will be introduced to the original data set. This hybrid technique combines modified Versatile Improved SMOTE algorithm with the rough sets theory. Although the VIS method is considered as effective and flexible, introducing the step of removing noise from created minority examples may guarantee better results in classifying data with very complex distribution. The algorithm discussed in this paper is dedicated to data sets described by nominal attributes, however, it can be easily adjusted to the continuous data problems.

At the beginning of algorithm relabel technique is applied (described in subsection 2.1). It is based on rough set theory. Since numerous real–world data sets are imprecise (have nonempty boundary region), the relevancy of this process should be emphasized. Majority class samples belonging to the boundary region of minority class are transformed into minority class examples (their class attribute is modified). In other words, all examples that can be certainly classified neither as negative nor as positive samples are imposed to be considered as minority class members. Thus, the complexity of the problem becomes lower (by reducing inconsistencies) as well as the imbalance ratio is decreased.

Algorithm VIS_RST

INPUT: *DataSet*; Number of all instances S; Number of minority class samples M; Number of nearest neighbors k

OUTPUT: *resultDataSet*: minority and majority class instances after preprocessing

Calculate the boundary region. Modify the class attribute of the majority samples belonging to the boundary region - relabel to minority class. Add the number of relabeled instances to the overall number of minority class examples M.
for $i \leftarrow 1$ **to** M **do**
  Calculate the distance between minority class objects and all other examples using kNN method.
  Calculate the number of nearest neighbors that belongs to the majority class and save this value in $majorityClassNeighbors$ variable. Assign the positive instance i into one category (SAFE, DANGER or NOISE) considering the local characteristics (nearest neighbors):
  if $majorityClassNeighbors == k$ **then**
    $label[i] = NOISE$
  else if $majorityClassNeighbors < k/2$ **then**
    $label[i] = SAFE$
  else if $majorityClassNeighbors \geqslant k/2$ **then**
    $label[i] = DANGER$
  end if
end for
Calculate the total counts of objects belonging to each group and save these values in the following variables: $safe$, $danger$, $noise$. Based on these counts, choose the strategy ($mode$) of processing:
if $safe == 0$ **then**
  $mode := noSAFE$
else if $danger \geqslant 30\%M$ **then**
  $mode := HighComplexity$
else
  $mode := LowComplexity$
end if
Calculate the required number of minority class examples to create. The result save in N variable.
for $i \leftarrow 1$ **to** M **do**
  if $label \neq NOISE$ **then**
    Calculate the distance between minority class objects using kNN method, indexes of k nearest neighbors save in $nnarray$ array.
    Create the synthetic examples following rules specified for the appropriate $mode$.
  end if
end for
for $i \leftarrow 1$ **to** N **do**
  Using calculations made at the beginning of algorithm, verify whether newly created synthetic object i belongs to the lower approximation of the minority class - if yes: add the example i to the $resultDataSet$. In the other case, remove generated sample i.
end for

In the next step minority data is categorized into three groups. To obtain the proper group for each sample the k–NN technique is applied. In order to consider both numeric and symbolic attributes the HVDM metric [23] was chosen to calculate distance between objects. The Heterogeneous Value Distance Metric is defined as:

$$HVDM(x,y) = \sqrt{\sum_{a=1}^{m} d_a(v,v')^2} \tag{3}$$

where x and y are the input vectors, m is the number of attributes, v and v' are the values of attribute a for object x and y respectively. The distance function for the attribute a is defined as:

$$d_a(v,v') = \begin{cases} 1, & \text{if } v \text{ or } v' \text{ is unknown} \\ normalized\_vdm_a(v,v'), & \text{if } a \text{ is nominal} \\ normalized\_diff_a(v,v'), & \text{if } a \text{ is linear} \end{cases} \tag{4}$$

The distance function consists of two other functions conformed to different kinds of attributes. Hence, the following function is defined for nominal features:

$$normalized\_vdm_a(v,v') = \sqrt{\sum_{c=1}^{C} \left| \frac{N_{v,c}}{N_v} - \frac{N_{v',c}}{N_{v'}} \right|^2} \tag{5}$$

where N_v is the number of instances in the training set that have value v for attribute a, $N_{v,c}$ is the number of instances that have value x for attribute a and output class c, C is the number of classes.

On the other hand, the function appropriate for linear attributes is defined as:

$$normalized\_diff_a(v,v') = \frac{|v - v'|}{4\sigma_a} \tag{6}$$

where σ_a is the standard deviation of values of attribute a.

Definition 1. *Depending on the class membership of the sample's k nearest neighbors, the following labels for the minority class are assigned:*

- *NOISE, when all of the k nearest neighbors represent the majority class,*
- *DANGER, if half or more than half of the k nearest neighbors belong to the majority class,*
- *SAFE, when more than half of the k nearest neighbors represent the same class as the example under consideration (namely the minority class).*

The mechanism of detecting within–class subconcepts enables to customize the oversampling strategy for each specific type of objects. Moreover, depending on the number of samples in mentioned groups two main modes of preprocessing minority data are proposed in modified VIS algorithm.

The first one, "HighComplexity", represents the case when the area surrounding class boundaries can be described as complex (at least 30% of the minority class instances are the borderline ones – DANGER label) [15].

Definition 2. *Since generating most of the minority synthetic samples in this region may lead to the overlapping effect, the following rules of creating new objects are applied for particular kinds of nominal data:*

- *DANGER: only one new sample is generated by replicating features of the minority instance under consideration,*
- *SAFE: as the SAFE objects are assumed to be the main representatives of the minority class, a plenty of new data is created in these homogeneous regions using majority vote of k nearest neighbors' features,*
- *NOISE: no new instances created.*

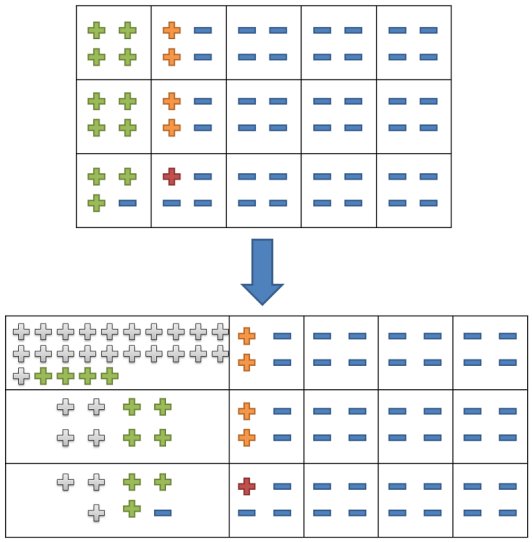

Fig. 2. Example of VIS_RSB preprocessing (relabel step is omitted): artificial data where minority objects are labeled as DANGER (orange), SAFE (green) and NOISE (red). The labels are assigned using $k = 3$ nearest neighbors and normalized_vdm metric. Grey objects are new minority class samples generated in respect of the assigned labels.

The second mode, "LowComplexity", is appropriate for less complex problems.

Definition 3. *When the number of minority samples labeled as DANGER does not exceed 30% of all minority class examples, the processing is performed according to the approach specified below:*

- *DANGER: many objects are created, because not sufficient number of minority class examples in this specific area may be dominated in the learning process by the majority class samples. Newly generated sample attributes' values are obtained by the majority vote of k nearest neighbors' features,*
- *SAFE: one new object for each existing instance is created. Therefore, number of SAFE examples is doubled. New sample has the same values of attributes as the object under consideration,*
- *NOISE: no new instances created.*

There is also one special strategy, namely "noSAFE". It was developed to ensure that the required number of synthetic samples will be created, even as any of the minority class instances belongs to SAFE category. Absence of the SAFE examples indicates that the problem is very complex and most of the objects are labeled as DANGER. In standard way of processing the "HighComplexity" mode is chosen, hence majority of the new objects are generated in safe regions. However, there are no SAFE instances, thus the safe regions are not specified. In order to consider this case, "noSAFE" mode assumes creation of all new examples in the area surrounding class boundaries.

The overall number of the minority class samples to be generated is obtained automatically. The algorithm is designed to even the number of objects from both classes.

The final synthetic minority data set is obtained by eliminating samples considered as noise. The algorithm inspired by rough set notions is applied to indicate which newly created examples are similar to the majority objects. Since only nominal attributes are considered in this analysis, the boundary region of the minority class is calculated. All synthetic samples that belong to the boundary region are removed. This additional cleaning step ensures that the generated data set is deprived of inconsistent objects. It is essential to select only these samples that are certainly members of the minority class.

3 Experiments

Six data sets were selected to perform experiments. All of them (except didactic) originally came from the UCI repository [22], but after conversions like adjusting them to the two–class problem they were published in Keel-dataset repository [1]. Only data sets described by the nominal attributes were chosen. They are presented in Table 1 (IR indicates the imbalance ratio).

The aim of this experiment was to prepare comparison of four preprocessing methods. The classification without any re–sampling step was performed to establish a reference point for evaluation of algorithms. The following assumptions were made considering SMOTE and VIS_RST techniques:

- the number of nearest neighbors (k) was set to 5,
- the HVDM distance metric was applied,
- the imbalance ratio after generating new samples was 1.0.

Table 1. Characteristics of evaluated data sets

dataset	objects	attributes	IR	boundary region
dermatology-6	358	34	16.90	empty
flare-F	1066	11	23.79	nonempty
lymphography-normal-fibrosis	148	18	23.67	empty
zoo-3	101	16	19.20	empty
car_good	1728	6	24.04	empty
didactic (see Fig. 1)	60	2	2.75	nonempty

The results of classification were evaluated by five measures:

- accuracy (Q) – the percentage of all correct predictions (both minority and majority class examples are considered),
- sensitivity (TP_{rate}) – the percentage of positive instances correctly classified,
- specificity (TN_{rate}) – the percentage of properly classified objects from the majority class.
- $F - measure$ – the average of sensitivity and precision. Precision is the number of correctly identified positive samples divided by the number of all instances classified as positive (both properly and erroneously),
- AUC – area under the ROC curve. The Receiver Operating Characteristics (ROC) graphic depicts dependency between TP_{rate} and FP_{rate}. The FP_{rate} means the percentage of negative examples misclassified.

The AdaBoost.M1 algorithm [7] with decision trees C4.5 as weak learners was applied as the classifier. This technique represents the group of ensemble methods. The main purpose of combining decisions of multiple classifiers to obtain the aggregated prediction is improvement of generalization [21]. A five–folds crossvalidation was performed. The final experiments' results (presented in Table 2) are the average values of results from five iterations of processing.

Results of these experiments show that the higher complexity of analysed data set is, the better outcomes from applying proposed technique are. VIS_RST algorithm indicates that three real–world data sets are the most complex: flare–F, zoo–3 and car–good. One of these data sets, namely flare–F, has nonempty boundary region. Method proposed in this paper outperformed other techniques for this complex example. In all experiments both SMOTE and VIS_RST achieve higher values of AUC measure than the classification without preprocessing step. Remove and Relabel filters perform better only in case of nonempty boundary region. Relabel technique may be considered as more effective. It is worth noting that all minority samples generated by the VIS_RST method were in the lower approximation. Therefore, undersampling cleaning step was not needed.

4 Conclusions and future research

Firstly, the experiments revealed that the new VIS_RST method is comparable to the SMOTE algorithm when applied to data sets described only by the nominal features. The AUC measure of VIS_RST was higher for the flare–F data set.

Table 2. Classification results for the selected UCI datasets: Q – accuracy, TP_{rate} – rate of true positives, TN_{rate} – rate of true negatives, F – F measure, AUC – area under the curve.

method	Q	TP_{rate}	TN_{rate}	F	AUC	Q	TP_{rate}	TN_{rate}	F	AUC
	`dermatology-6`					`flare-F`				
noPRE	99.44	95.00	99.70	0.95	97.35	94.65	11.63	98.14	0.15	54.89
SMOTE	99.85	100.00	99.70	1.00	99.85	97.26	96.38	98.14	0.97	97.26
VIS_RST	99.70	99.70	99.70	1.00	99.70	98.48	99.16	97.80	0.98	98.48
Remove	99.44	95.00	99.70	0.95	97.35	96.08	37.21	98.74	0.45	67.98
Relabel	99.44	95.00	99.70	0.95	97.35	97.09	87.61	98.22	0.86	92.91
	`lymphography-normal-fibrosis`					`zoo-3`				
noPRE	97.97	50.00	100.00	0.67	75.00	94.06	40.00	96.88	0.40	68.44
SMOTE	98.94	97.89	100.00	0.99	98.94	97.40	96.88	97.92	0.97	97.40
VIS_RST	98.24	97.89	98.59	0.98	98.24	97.40	97.92	96.88	0.97	97.40
Remove	97.97	50.00	100.00	0.67	75.00	94.06	40.00	96.88	0.40	68.44
Relabel	97.97	50.00	100.00	0.67	75.00	94.06	40.00	96.88	0.40	68.44
	`car_good`					`didactic`				
noPRE	98.38	73.91	99.40	0.78	86.66	83.33	68.75	88.64	0.69	78.69
SMOTE	99.43	99.64	99.22	0.99	99.43	89.77	88.64	90.91	0.90	89.77
VIS_RST	99.16	99.52	98.79	0.99	99.16	100.00	100.00	100.00	1.00	100.00
Remove	98.38	73.91	99.40	0.78	86.66	100.00	100.00	100.00	1.00	100.00
Relabel	98.38	73.91	99.40	0.78	86.66	100.00	100.00	100.00	1.00	100.00

Proposed algorithm outperformed other techniques when evaluated data sets had nonempty boundary regions (flare–F and didactic). Secondly, the Relabel filtering technique performed better than the Remove approach for data set which has the nonempty boundary region (flare–F). In future research the performance of the proposed algorithm adjusted for the Big Data may be investigated. The application of the MapReduce paradigm [6] seems to be promising solution for large imbalance data problem.

Acknowledgements

The research is supported by the Polish National Science Centre under the grant 2012/07/B/ST6/01504.

References

1. Alcala–Fdez J., Fernandez A., Luengo J., Derrac J., Garca S., Sanchez L., Herrera F., KEEL Data–Mining Software Tool: Data Set Repository, Integration of Algorithms and Experimental Analysis Framework. Journal of Multiple–Valued Logic and Soft Computing 17:2-3, 2011, 255–287.

2. Barua S., Islam M.M., Murase K., A novel synthetic minority oversampling technique for imbalanced data set learning, Proceedings of the 18th international conference on Neural Information Processing - Volume Part II. Springer-Verlag, Berlin, Heidelberg, 2011, 735–744.
3. Batista G.E.A.P.A., Prati R.C., Monard M.C., A study of the behavior of several methods for balancing machine learning training data. SIGKDD Explor. Newsl. 6(1), June 2004, 20–29.
4. Borowska K., Topczewska M., New Data Level Approach for Imbalanced Data Classification Improvement, Proceedings of the 9th International Conference on Computer Recognition Systems CORES 2015, Springer International Publishing, 2016, 283–294.
5. Chawla N.V., Bowyer K.W, Hall L.O., and Kegelmeyer W.P., SMOTE: synthetic minority over-sampling technique. J. Artif. Int. Res. 16(1), 2002, 321–357.
6. Dean J, Ghemawat S., MapReduce: Simplified Data Processing on Large Clusters, Commun. ACM 51, 1, 2008, 107–113.
7. Freund Y., Schapire R.E., Experiments with a new boosting algorithm. Machine Learning: Proceedings of the Thirteenth International Conference, 1996, 148–156.
8. 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), 42(4), 2012, 463–484.
9. Garca V., Mollineda R. A., Snchez J. S., On the k-NN performance in a challenging scenario of imbalance and overlapping. Pattern Anal. Appl. 11(3–4), 2008, 269–280.
10. Han H., Wang W–Y, Mao B–H., Borderline–SMOTE: a new over–sampling method in imbalanced data sets learning, Proceedings of the 2005 international conference on Advances in Intelligent Computing - Volume Part I (ICIC'05), Springer-Verlag, Berlin, Heidelberg, 878–887.
11. He H., Garcia E.A., Learning from Imbalanced Data. IEEE Trans. on Knowl. and Data Eng. 21(9), 2009, 1263–1284.
12. Hu S., Liang Y., Ma L., He Y., MSMOTE: Improving Classification Performance When Training Data is Imbalanced, Computer Science and Engineering. WCSE '09. Second International Workshop on, Qingdao, 2009, 13–17.
13. Jo T., Japkowicz N., Class imbalances versus small disjuncts. SIGKDD Explor. Newsl. 6(1), 2004, 40–49.
14. Napierała K., Stefanowski J., BRACID: a comprehensive approach to learning rules from imbalanced data, Journal of Intelligent Information Systems, 39, 2012, 335–373.
15. Napierała K., Stefanowski J., Wilk S., Learning from imbalanced data in presence of noisy and borderline examples, Proceedings of the 7th international conference on Rough sets and current trends in computing (RSCTC'10), Springer-Verlag, Berlin, Heidelberg, 158–167.
16. Pawlak Z., Skowron A., Rudiments of Rough Sets. Information Sciences, 177(1), 2007, 3–27.
17. Ramentol E., Caballero Y., Bello R., Herrera F., SMOTE-RSB$_*$: a hybrid preprocessing approach based on oversampling and undersampling for high imbalanced data-sets using SMOTE and rough sets theory. Knowledge and Information Systems, Springer-Verlag, 33(2), 2011, 245–265.
18. Stefanowski J., Wilk S., Rough Sets for Handling Imbalanced Data: Combining Filtering and Rule-based Classifiers. Fundam. Inf. 72(1–3), 2006, 379–391.

19. Stefanowski J., Wilk S., Selective Pre–processing of Imbalanced Data for Improving Classification Performance, Proceedings of the 10th international conference on Data Warehousing and Knowledge Discovery (DaWaK '08), Springer-Verlag, Berlin, Heidelberg, 283–292.
20. Stepaniuk J., Rough–Granular Computing in Knowledge Discovery and Data Mining, Springer–Verlag Berlin Heidelberg, Incorporated, 2008.
21. Sun Y., Kamel M. S., Wong A. K. C., Wang Y., Cost-sensitive Boosting for Classification of Imbalanced Data, Pattern Recogn., 40, 2007, 3358–3378.
22. UC Irvine Machine Learning Repository, http://archive.ics.uci.edu/ml/ (accessed 10.04.2016).
23. Wilson D.R., Martinez T.R., Improved heterogeneous distance functions, Journal of Artificial Intelligence Research, 6, 1997, 1–34.