The Impact of Local Data Characteristics on Learning from Imbalanced Data

Jerzy Stefanowski

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

Abstract. Problems of learning classifiers from imbalanced data are discussed. First, we look at different data difficulty factors corresponding to complex distributions of the minority class and show that they could be approximated by analysing the neighbourhood of the learning examples from the minority class. We claim that the results of this analysis could be a basis for developing new algorithms. In this paper we show such possibilities by discussing modifications of informed pre-processing method LN–SMOTE as well as by incorporating types of examples into rule induction algorithm BRACID.

1 Introduction

Many difficult learning problems, from a wide variety of domains, involve learning from imbalanced data, where at least one of the target classes contains a much smaller number of examples than the other classes. This class is usually referred to as the *minority class*, while the remaining classes are denoted as *majority ones*. For instance, in medical problems the number of patients requiring special attention is much smaller than the number of patients who do not need it. Similar situations occur in such domains as: fraud detection, risk management, technical diagnostics, image recognition, text categorization or information filtering. In all those problems, the correct recognition of the minority class is of key importance. However, class imbalance constitutes a great difficulty for most learning algorithms and often resulting classifiers are biased toward the majority classes and fail to recognize examples from the minority class.

While the difficulty with learning classifiers from imbalanced data has been known earlier from applications, this challenging problem has received a growing research interest in the last decade and a number of specialized methods have already been proposed, for their review see, e.g., [7,18,19,10].

In general, they may be categorized into *data level* and *algorithm level* ones. The first group includes classifier-independent methods that are used in the pre-processing step to modify the balance between classes, so that standard algorithms can be used to learn better classifiers. These methods are usually based on either adding examples to the minority class (called *over-sampling*) or removing examples from the majority class (*under-sampling*). The other main category of, so called, algorithmic methods involves modifications of either: learning phase

M. Kryszkiewicz et al. (Eds.): RSEISP 2014, LNAI 8537, pp. 1–13, 2014.

of the algorithm or classification strategy, construction of specialized ensembles, or adaptation of cost sensitive learning. For their reviews see [7,18,19,44].

Although several specialized methods already exist, the identification of conditions for their efficient use is still an open research problem. It is also related to more fundamental issues of better understanding the nature of the imbalance data and key properties of its underlying distribution.

Following related works [21,22,14,28] and earlier studies of J. Stefanowski and K.Napierala, J.Blaszczynski or Sz. Wilk [41,32,34,31] we claim that the high imbalance ratio between the minority and majority classes is not the only and not even the main reason of these difficulties. Other, as we call them, *data difficulty factors*, referring to characteristics of class distributions, are also influential. They include: decomposition of the minority class into many rare sub-concepts, the effect of too strong overlapping between the classes [36,14] or a presence of too many minority examples inside the majority class region. When these data difficulty factors occur *together* with class imbalance, they may seriously hinder the recognition of the minority class, see e.g. a study [28].

In our earlier paper [34] we propose to capture these data difficulty factors by considering the local characteristics of learning examples from the minority class. More precisely, it is achieved by analyzing the class distribution of examples from different classes inside a *local neighborhood* of the considered example. Finding how many examples from opposite classes are the neighbours of this example, the degree of its difficulty could be estimated.

We claim that the proper analyzing of this neighborhood of learning examples from the minority class could be the basis for developing new specialized algorithms for imbalanced data. In this paper we "implement" this postulate by considering representatives of two main categories of methods specialized for imbalanced data. Firstly, we will apply the analysis of neighbours into the new generalization of the most popular informed pre-processing method SMOTE. Secondly, we will show that data difficulty factors modeled by types of minority examples could be used inside the rule candidate generation phase of the rule induction algorithm BRACID. Finally, we will discuss other possible options of using the local information, in particular for ensembles and highlight other future research directions of studying imbalanced data.

2 Local Characteristics of Data Difficulty Factors and Identification of Example Types

Although many authors have experimentally shown that standard classifiers met difficulties while recognizing the minority class, it has also been observed that in some problems characterized by strong imbalance between classes standard classifiers are sufficiently accurate. Moreover, the discussion of data difficulty in imbalanced data still goes on, for its current review see, e.g., [10,19,29,34,40].

Some researchers have already noticed, that the *global class imbalance ratio* (i.e. the cardinality of the majority class referred to the total number of minority class examples) is not necessarily the only, or even the main, problem causing the decrease of classification performance and focusing only on this ratio may be insufficient for improving classification performance. Besides the imbalanced ratio other data difficulty factors may cause a severe deterioration of classifiers.

The experimental studies by Japkowicz *et al* with many artificial data sets have clearly demonstrated that the degradation of classification performance is also linked to the decomposition of the minority class into many sub-parts containing very few examples [21,22]. They have shown that the minority class does not form a homogeneous, compact distribution of the target concept but it is scattered into many smaller sub-clusters surrounded by majority examples. In other words, minority examples form, so called, *small disjuncts*, which are harder to learn and cause more classification errors than larger sub-concepts.

Other factors related to the class distribution are also linked to the effect of too strong *overlapping* between minority and majority class. Experiments with artificial data have shown that increasing overlapping has been more influential than changing the class imbalance ratio [36,14]. Yet another data factor, which influences degradation of classifiers performance on imbalanced data, is presence of noisy examples [2]. Experiments presented in [32] have also shown that single minority examples located inside the majority class regions cannot be treated as noise since their proper treatment by informed pre-processing may improve classifiers. Moreover, studies as [40] emphasize that several data factors usually occur together in real world imbalanced data sets.

These studies stress the role of the *local characteristics* of the class distribution. However it could be modeled in different ways. Here, we follow earlier works [24,25,32,31,29] and link data difficulty factors to *different types of examples* forming the minority class distribution. It leads us to a differentiation between safe and unsafe examples. *Safe examples* are ones located in the homogeneous regions populated by examples from one class only. Other examples are *unsafe* and more difficult for learning. Unsafe examples are categorized into *borderline* (placed close to the decision boundary between classes), *rare cases* (isolated groups of few examples located deeper inside the opposite class), or *outliers*. As the minority class can be highly under-represented in the data, we claim that the rare examples or outliers, could represent a very small but valid sub-concepts of which no other representatives could be collected for training. Therefore, they cannot be considered as noise examples which typically are then removed or re-labeled. Moreover, earlier works with graphical visualizations of real-world imbalanced data sets [34] have confirmed this categorization of example types.

The next question is how to automatically *identify these types of examples* in real world data sets (with unknown underlying class distributions). We keep the hypotheses [34] on role of the mutual positions of examples and the idea of assessing the type of example by analyzing class labels of the other examples in its *local neighbourhood*. This neighbourhood of the minority example could be modeled in different ways. In further considerations we will use an analysis of the class labels among *k-nearest neighbours* [34,31]. This approach requires choosing the value of k and the *distance function*. In our previous considerations we have followed results of analyzing different distance metrics [27] and chose the

HVDM metric (*Heterogeneous Value Difference Metric*) [45]. Its main advantage for mixed attributes is that it aggregates normalized distances for qualitative and quantitative attributes. In particular, comparing to other metrics it provides more appropriate handling of qualitative attributes as instead of simple value matching it calculates attribute value conditional probabilities by using a Stanfil and Valtz value difference metric [45]. Then, due to complexity of the distribution of the minority class, k should be rather a small value. Experiments from [34,31] over many UCI data sets have showed that $k = 5$ or 7 have led to good results.

Depending on the number of examples from the majority class in the neighbourhood of the minority example, we can evaluate whether this example could be safe or unsafe (difficult) to be learned. If all, or nearly all, its neighbours belong the minority class, this example is treated as the safe example, otherwise it is one of unsafe types. For instance, consider $k = 5$. In this case the type of example x is defined as: if 5 or 4 of its neighbours belong to the same class as x, it is treated as a safe example; if the numbers of neighbours from both classes are similar (proportions 3:2 or 2:3) – it is a borderline example; if it has only one neighbour with the same label (1:4) it is a rare case; finally if all neighbours come from the opposite class (0:5) – it is an outlier. Although this categorization is based on intuitive thresholding, its results are consistent with a probabilistic analysis of the neighbourhood, modeled with kernel functions, as shown in [31].

Our experiments with UCI imbalanced data sets [34,31] have also demonstrated that most of these real-world data do not include many safe minority examples. They rather contain all types of examples, but in different proportions. On the other hand, most of majority class examples have been identified as safe ones. Depending on the dominating type of identified minority examples, the considered datasets have been labeled as: safe, border, rare or outlier. As a large number of borderline examples often occurred in many data sets, some of these data sets could be assigned both to border and more difficult categories.

Moreover, the study [34] has shown that the classifier performance could be related to the category of data. First, for the safe data nearly compared single classifiers (SVM, RBF, k-NN, decision trees or rules) perform quite well with respect to sensitivity, F-measure or G-mean. The larger differentiation occurs for more unsafe data set. For instance, SVM and RBF work much better for safe category, while rare or outlier data strongly deteriorate their classification performance. On the other hand, unpruned decision trees and k-NN work quite well for these unsafe data sets. The similar analysis has been carried out for the most representative pre-processing approaches, showing that the competence area of each method depends on the data difficulty level, based on the types of minority class examples. Again in the case of safe data there are no significant differences between the compared methods - even random over-sampling works quite accurate. However, for borderline data sets Nearest Cleaning Rules (methods filtering difficult majority examples [25]) performs best. On the other hand, SMOTE [8] and SPIDER [41], which can add new examples to the data, have proved to be more suitable for rare and outlier data sets.

For more details on the competence of each studied single classifier and pre-processing methods see [31]. The similar analysis for different generalizations of bagging ensembles, included specialized solutions for class imbalances, have been carried out in the recent paper [5]. Finally, we will repeat our hypothesis that the appropriate treatment of these types of minority examples within new proposals of either pre-processing or classifiers should lead to improving classification performance. We will show it in the next sections.

3 Modifications of Informed Pre-processing Methods

The simplest data pre-processing techniques are random over-sampling, which replicates examples from the minority class, and random under-sampling, which randomly eliminates examples from the majority classes until a required degree of balance between classes is reached. However, random under-sampling may potentially remove some important examples and simple over-sampling may also lead to overfitting. Therefore, focused (also called *informed*) methods, which attempt to take into account internal characteristics of regions around minority class examples, were introduced, as e.g. SMOTE [8], one-side-sampling [24], NCR [25] or SPIDER[41] .

The most popular among the informed methods is SMOTE, which considers each example from the minority class and generates new synthetic examples along the lines between the selected example and some of its randomly selected k-nearest neighbors from the minority class. More precisely, let the training set S contain examples from the minority class P and other classes N. For each example $p_i \in P$ find its k nearest neighbours x from class P. Depending on the other parameter of this method – the amount of over-sampling – a given number of examples from these k nearest neighbours is randomly selected. Synthetic minority class examples are generated in the direction of each. For numerical attributes the new synthetic example is constructed as follows: compute the difference between attributes describing the example p_i and x – one of the selected k-nearest neighbours; multiply this feature vector difference by δ – a random number between 0 and 1; and add it to the attribute vector p_i creating a new vector $x_{new} = p_i + (x - p_i) \cdot \delta$. For qualitative attributes create a new example with the most common feature values among k nearest neighbours.

Although experiments have confirmed its usefulness (see e.g. [3,7]), some of the assumptions behind this technique could be still questioned. Two main shortcomings of SMOTE are: (1) treating all minority examples in the same way while they may not be equally important for learning classifiers (2) the possible over-generalization over the majority class regions as SMOTE blindly generalizes regions of the minority class without checking positions of the nearest examples from the majority classes. Some researchers solve these problem by integrating SMOTE with additional filtering steps (see e.g. [3,37]), while others modify SMOTE's internal strategies for selecting positions of synthetic examples.

In this paper we focus on the recent proposal called Local Neighbourhood extension of SMOTE (briefly LN-SMOTE) which is inspired by the analyzing local data characteristics and earlier modifications of SMOTE [4]. In this method the presence of the majority examples is taken into account before generating synthetic examples by calculating a special coefficient called a *safe level*. It is defined as the number of other minority class examples among its k nearest neighbours. The smaller its value, the more unsafe is this example. This level $sl(p)$ is calculated for the example p, which is a seed for oversampling and as $sl(x)$ for its randomly selected neighbour x. Unlike the standard SMOTE and its generalizations as [4,17], in LN-SMOTE the closest neighbours are calculated including also majority class examples. Having information about values of both safe levels $sl(p)$ and $sl(x)$, the range of positioning the synthetic example is modified. Only for equal both levels the examples will be generated along the whole line joining x and p in the same way as in the original SMOTE. If one safe level is greater than other the position new example will generated closer the safer example (more closer, the larger difference between these levels). Furthermore, in case of the neighbour from the majority class, the range of random overlapping is additionally limited not to come to close to the majority examples. Situations of outliers (safe level equal 0) are also distinguished by not putting the new examples for such a neighbour. Finally before starting over-sampling, all majority examples being outliers inside the minority class are identified by analysis content of k neighbourhood - they are removed from the learning set as they usually disturb the minority class distribution.

The LN-SMOTE was introduced and experimentally studied in [30]. Its comparative study against basic SMOTE and two other related generalizations called Borderline-SMOTE [17] and SL-SMOTE [4] applied with 3 different classifiers (J4.8, Naive Bayes and k-NN) showed that it was the best pre-processing method. For instance, Table 1 summarizes results of F-measure for these pre-processing methods applied to J4.8 decision tree, where Bord-SMOTE denotes Borderline SMOTE and SL-SM denotes Safe Level SMOTE.

Table 1. F-measure for the minority class for all compared methods used together with J48 classifier [%]

Data	None	SMOTE	Bord-SMOTE	SL-SM	LN-SMOTE
balance-scale	0.00	1.06	2.09	2.95	6.21
car	80.61	88.39	72.91	90.39	88.58
cleveland	19.29	21.35	22.89	21.33	26.70
cmc	40.81	41.46	42.05	41.66	44.75
ecoli	58.86	63.21	64.53	63.51	66.96
haberman	30.36	38.41	41.50	37.33	42.20
hepatitis	49.20	49.86	51.23	52.57	54.42
postoperative	5.84	11.90	15,06	12.08	16.18
solar-flare	28.79	27.13	28.32	29.62	31.6
transfusion	47.27	48.07	49.79	49.22	50.30
yeast	35.02	36.08	38.63	40.21	42.58

4 Incorporating Types of Examples in Rule Induction

Let us now consider using local characteristics of learning examples within approaches modifying the algorithms. Decision rules, being the most human readable knowledge representation, are particularly sensitive to class imbalance, see e.g. conclusions from [15,16,42]. Following some earlier methodological discussions [35,44] the standard algorithms for learning rule based classifiers share a number of their principles which are useful for classification with respect to the total accuracy but are limitations in case of class imbalances.

First, most algorithms induce rules using the top-down technique with *maximum generality bias*, which favors general rules however also hinders finding rules for smaller sets of learning examples, especially in the minority class. It is also connected with using *improper evaluation measures to guide the search* for best conditions. Typical measures, as presented, e.g., in [12,38], try to find a compromise between the accuracy and generality of the rule which achieve better values mainly for the majority class examples. Similar measures are also often used to *prune* induced rules, which are particularly inappropriate for small disjuncts or rare cases in the minority class where rules may be constructed as the conjunction of many elementary conditions. Thus, pruning of rules is guided mostly by measures referring more to the majority class examples, neglecting the minority class specific distributions [1].

Second, most algorithms use a *greedy sequential covering* approach [11], in which learning examples covered by the induced rule are removed from the current set of considered examples. This approach may increase the data fragmentation for the minority class and leads to the induction of *weaker* rules, i.e. supported by a smaller number of learning examples. The "weakness" of the minority rules could be also associated with a third factor: *classification strategies*, where minority rules have a smaller chance to contribute to the final classification decision, see discussions in [15,42,10].

The above limitations concerns difficulties inside the algorithm. Recall that one should also consider data difficulty factors referring to complex distributions of the minority class, as presented in section 2. Some researchers have already proposed the extensions of rule based approaches dedicated for class imbalance - for a comprehensive review see [35]. However, most of these proposals addresses only a single or at most a few of algorithmic or data factors.

Following these critical motivations K.Napierala and J.Stefanowski have introduced a new rule induction algorithm called BRACID (Bottom-up induction of Rules And Cases for Imbalanced Data) which is specialized for the classification of imbalanced data [35]. While constructing it we have addressed several limitations mentioned above.

First we have decided to induce rules by *bottom-up generalization* of the most specific rules representing single examples. Bearing in mind that local algorithms could better learn the difficult decision boundaries (which usually describe minority examples), this algorithm is a *hybrid* of rule-based and instance-based *knowledge representations*. Due to this representation it should also better deal with rare sub-concepts of the minority class. Overcoming the problem of data fragmentation is also connected with resigning from a greedy, sequential covering and top-down induction technique. Inside the crucial operation of the bottom-up generalization of the current rules the specific looking for the nearest example has been applied. The candidates for rules are temporarily added to the current classifier and evaluated with the F-measure in a leaving-one-out procedure. Such an approach allows us to evaluate and accept rules in a more appropriate way for recognizing the minority class. The final classifier uses a *nearest rule strategy* to classify new coming examples [39], which has proved to be more appropriate for recognizing minority classes than standard classification strategies too much oriented to majority examples.

An important component of BRACID is also using information about the nature of the neighbouring examples. Following the method presented in section 2, we identify types of learning examples in each class. Unsafe outlier examples from the majority classes are removed from the learning set as they may hinder fragmentation of the minority class and then the induction of more general minority rules. In case of an analogous situation for the minority class, this example is not removed but it is checked as a candidate for a rule generalization. Moreover for outliers, rare cases and borderline minority examples we allow to analyse k possible generalizations and to choose best ones according to the F-measure evaluation. It allows us to create more rules for the minority class in unsafe regions, as overlapping. It should diminish the possibility of overwhelming the minority class with the majority class rules in these difficult sub-regions.

Table 2. G-mean for BRACID and other rule induction algorithms [%]

Data	BRACID	CN2	RIPPER	MODLEM
abalone	65.0	39.6	42.1	48.4
balance-scale	56.7	2.9	1.9	0.0
cleveland	57.4	0.0	25.8	19.2
cmc	63.7	25.8	25.5	47.2
credit germ.	61.1	55.3	43.8	56.3
ecoli	83.1	28.4	58.7	56.8
haberman	57.6	34.5	35.6	40.1
ionosphere	91.2	87.0	87.4	89.2
vehicle	93.5	51.3	91.9	91.6
transfusion	63.9	34.2	26.6	52.9

The experimental studies with components of BRACID, in particular this way of incorporating types of examples have showed that it has improved the evaluation measures comparing to the plain option of treating all learning examples in the same way. For instance, for G-mean average improvements are 3.5 % [35]. Finally, BRACID has been compared to a number of state-of-the-art rule based classifiers (PART, RIPPER,C45rules,CN2, MODLEM, RISE), instance-based classifier (K-NN) and some approaches dedicated for class imbalances in the comprehensive experimental study over 22 imbalanced data sets. The results showed that BRACID can better recognize the minority classes than other compared algorithm - which is reflected by measures as F-measure and G-mean. The selected results of G-mean are presented in Table 2.

Furthermore, using the categorization of data sets obtained with analysis their local characteristics presented in section 2, we have found out that the best improvements of BRACID are observed for unsafe data sets containing many borderline examples from the minority class.

5 Other Perspectives of Applying Local Data Characteristics

The main message of this paper is to promote incorporating the information about the local neighbourhood of a chosen minority class example in the process of constructing and analyzing methods for learning classifiers from imbalanced data. Although we have shown two such possibilities, still other new directions are worth to be studied.

For instance, the proposed method for analysing k-neighbourhood of minority examples as well as practically all informed pre-processing methods have been considered for data with not so high number of attributes. However, some applications of data mining in bio-medicine, text or multimedia processing involve highly dimensional data sets. The use of typical dissimilarity measures and k-nearest neighbor classification on such data sets may suffer from the curse of dimensionality problem as it has been recently showed by Tomasev's research on, so called, *hubness*-aware shared neighbor distances for high-dimensional k-nearest neighbor classification [43]. Thus, studying generalizations of the presented methods for high dimensional data is still an open research challenge.

In case of pre-processing methods as SMOTE, there still remain some interesting questions on creation strategies for synthetic data, amount of new data to create, better identification of the most appropriate sub-regions of the minority class where to add new examples, avoidance of introducing noise instead of valuable instances of the under-represented class, distinguishing between noise and valuable outliers, and checking whether synthetic examples are equally important as the real ones by employing new evaluation measures while constructing and evaluating classifiers.

Another point of view on informed pre-processing methods and the role of generating large amounts of synthetic data could also come from specific applications. Such studies as [32,41] show that the degree of over-sampling is quite high for many data sets, even if it is tuned just to obtain a balance of cardinality of minority and majority classes. However, in medical problems physicians could be reluctant to analyse so a high number of artificial patients in their data set (i.e. their class of interest could include more non existing patient's descriptions than real clinical cases). Moreover, clinical experts often prefer to induce symbolic classifiers due to their potential interpretability. For instance, if rule induction algorithms are applied to data transformed by SMOTE, many rules in the final classifier could be supported just by synthetic learning examples. In spite of better classification performance such rules can be rejected by clinical experts due to their artificial supports and expert could still prefer to analyse rule referring to real facts in the original data.

These limitations open other perspectives on new informed preprocessing methods that do not introduce synthetic examples as well as on new specialized rule induction algorithms. Within the first perspective one can consider generalizing hybrid methods as SPIDER [41], which should better identify sub-regions where unsafe minority class examples should be amplified with different degrees as well as better filter the majority examples which too strongly influence these sub-regions. The other perspective is partly considered in such rule induction algorithms as EXPLORE [42], BRACID [35] or ABMODLEM [33] (which allows to directly incorporate expert explanations as to classifying difficult examples into the rule induction). However more extensive research and medical practical case studies are still needed.

Considering types of example could be also applied to new ensembles specialized for class imbalance. Most of current proposals are generalizations of known techniques as bagging, boosting or random forests; see their review in [13,26]. Their experimental results show that modifications of bagging often outperform boosting generalizations or more complex ensembles. While analyzing existing extensions of bagging one can also notice that most of them employ the simplest random re-sampling technique, as under-sampling or over-sampling, and, what is even more important, they just modify bootstraps to simply balance the cardinalities of minority and majority class. However, in all of these extensions (see, e.g., the Roughly Balanced Bagging [20]), all examples are treated as equally important while sampling them into bootstrap samples. We think that drawing of minority examples should not be done in a pure blind random way but it could be partly directed depending on the difficulty type of example.

In [5] we have already proposed to change probability of drawing different types of examples depending on the class distribution in the neighbourhood of the example. This has led us to the new type of bagging ensemble, called Nearest Neighborhood Bagging. The recent experiments [6] show that this new ensemble is significantly better than existing over-sampling bagging extensions and it is competitive to Roughly Balanced Bagging, which according to experiments [20,23] is the most accurate under-sampling extension of bagging. Nevertheless, several issues on: how much bootstrap samples should be modified, the influence of filtering majority class examples, diversity of bootstrap samples and the constructed classifiers, new techniques of their aggregation should be still studied.
The research on this paper is partially supported by NCN grant.

References

1. An, A.: Learning classification rules from data. Computers and Mathematics with Applications 45, 737–748 (2003)
2. Anyfantis, D., Karagiannopoulos, M., Kotsiantis, S., Pintelas, P.: Robustness of learning techniques in handling class noise in imbalanced datasets. In: Boukis, C., Pnevmatikakis, A., Polymenakos, L. (eds.) AIAI 2007. IFIP, vol. 247, pp. 21–28. Springer, Boston (2007)

3. Batista, G., Prati, R., Monard, M.: A study of the behavior of several methods for balancing machine learning training data. ACM SIGKDD Explorations Newsletter 6(1), 20–29 (2004)
4. Bunkhumpornpat, C., Sinapiromsaran, K., Lursinsap, C.: Safe-Level-SMOTE: Safe-Level-Synthetic Minority Over-Sampling TEchnique for Handling the Class Imbalanced Problem. In: Theeramunkong, T., Kijsirikul, B., Cercone, N., Ho, T.-B. (eds.) PAKDD 2009. LNCS, vol. 5476, pp. 475–482. Springer, Heidelberg (2009)
5. Błaszczyński, J., Stefanowski, J., Idkowiak, Ł.: Extending bagging for imbalanced data. In: Burduk, R., Jackowski, K., Kurzynski, M., Wozniak, M., Zolnierek, A. (eds.) CORES 2013. AISC, vol. 226, pp. 273–282. Springer, Heidelberg (2013)
6. Błaszczyński, J., Stefanowski, J., Szajek, M.: Local Neighbourhood in Generalizing Bagging for Imbalanced Data. In: Proc. of COPEM 2013 - Solving Complex Machine Learning Problems with Ensemble Methods Workshop at ECML PKDD 2013, Praque, pp. 10–24 (2013)
7. Chawla, N.: Data mining for imbalanced datasets: An overview. In: Maimon, O., Rokach, L. (eds.) The Data Mining and Knowledge Discovery Handbook, pp. 853–867. Springer, Heidelberg (2005)
8. Chawla, N., Bowyer, K., Hall, L., Kegelmeyer, W.: SMOTE: Synthetic Minority Over-sampling Technique. J. of Artificial Intelligence Research 16, 341–378 (2002)
9. Cost, S., Salzberg, S.: A Weighted Nearest Neighbor Algorithm for Learning with Symbolic Features. Machine Learning Journal 10(1), 1213–1228 (1993)
10. Fernández, A., García, S., Herrera, F.: Addressing the Classification with Imbalanced Data: Open Problems and New Challenges on Class Distribution. In: Corchado, E., Kurzyński, M., Woźniak, M. (eds.) HAIS 2011, Part I. LNCS, vol. 6678, pp. 1–10. Springer, Heidelberg (2011)
11. Furnkranz, J.: Separate-and-conquer rule learning. Artificial Intelligence Review 13(1), 3–54 (1999)
12. Furnkranz, J., Gamberger, D., Lavrac, N.: Foundations of Rule Learning. Springer (2012)
13. 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 99, 1–22 (2011)
14. García, V., Sánchez, J., Mollineda, R.A.: An empirical study of the behavior of classifiers on imbalanced and overlapped data sets. In: Rueda, L., Mery, D., Kittler, J. (eds.) CIARP 2007. LNCS, vol. 4756, pp. 397–406. Springer, Heidelberg (2007)
15. Grzymala-Busse, J.W., Goodwin, L.K., Grzymala-Busse, W., Zheng, X.: An approach to imbalanced data sets based on changing rule strength. In: Proceedings of Learning from Imbalanced Data Sets, AAAI Workshop at the 17th Conference on AI, pp. 69–74 (2000)
16. Grzymala-Busse, J.W., Stefanowski, J., Wilk, S.: A comparison of two approaches to data mining from imbalanced data. Journal of Intelligent Manufacturing 16(6), 565–574 (2005)
17. Han, H., Wang, W., Mao, B.: Borderline-SMOTE: A New Over-Sampling Method in Imbalanced Data Sets Learning. In: Huang, D.-S., Zhang, X.-P., Huang, G.-B. (eds.) ICIC 2005. LNCS, vol. 3644, pp. 878–887. Springer, Heidelberg (2005)
18. He, H., Garcia, E.: Learning from imbalanced data. IEEE Transactions on Data and Knowledge Engineering 21(9), 1263–1284 (2009)
19. He, H., Yungian, M. (eds.): Imbalanced Learning. Foundations, Algorithms and Applications. IEEE - Wiley (2013)

20. Hido, S., Kashima, H.: Roughly balanced bagging for imbalance data. Statistical Analysis and Data Mining 2(5-6), 412–426 (2009)
21. Japkowicz, N.: Class imbalance: Are we focusing on the right issue? In: Proc. II Workshop on Learning from Imbalanced Data Sets, ICML Conf., pp. 17–23 (2003)
22. Jo, T., Japkowicz, N.: Class Imbalances versus small disjuncts. ACM SIGKDD Explorations Newsletter 6(1), 40–49 (2004)
23. 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)
24. Kubat, M., Matwin, S.: Addresing the curse of imbalanced training sets: one-side selection. In: Proc. of the 14th Int. Conf. on Machine Learning, ICML 1997, pp. 179–186 (1997)
25. Laurikkala, J.: Improving identification of difficult small classes by balancing class distribution. Tech. Report A-2001-2, University of Tampere (2001)
26. Liu, A., Zhu, Z.: Ensemble methods for class imbalance learning. In: He, H., Yungian, M. (eds.) Imbalanced Learning. Foundations, Algorithms and Apllications, pp. 61–82. Wiley (2013)
27. Lumijarvi, J., Laurikkala, J., Juhola, M.: A comparison of different heterogeneous proximity functions and Euclidean distance. Stud Health Technol. Inform. 107 (pt. 2), 1362–1366 (2004)
28. Lopez, V., Fernandez, A., Garcia, S., Palade, V., Herrera, F.: An Insight into Classification with Imbalanced Data: Empirical Results and Current Trends on Using Data Intrinsic Characteristics. Information Sciences 257, 113–141 (2014)
29. Lopez, V., Triguero, I., Garcia, S., Carmona, C., Herrera, F.: Addressing imbalanced classification with instance generation techniques: IPADE-ID. Neurocomputing 126, 15–28 (2014)
30. Maciejewski, T., Stefanowski, J.: Local neighbourhood extension of SMOTE for mining imbalanced data. In: Proc. IEEE Symp. on Computational Intelligence and Data Mining, pp. 104–111 (2011)
31. Napierala, K.: Improving rule classifiers for imbalanced data. Ph.D. Thesis. Poznan University of Technology (2013)
32. Napierała, K., Stefanowski, J., Wilk, S.: Learning from Imbalanced Data in Presence of Noisy and Borderline Examples. In: Szczuka, M., Kryszkiewicz, M., Ramanna, S., Jensen, R., Hu, Q. (eds.) RSCTC 2010. LNCS, vol. 6086, pp. 158–167. Springer, Heidelberg (2010)
33. Napierała, K., Stefanowski, J.: Argument Based Generalization of MODLEM Rule Induction Algorithm. In: Szczuka, M., Kryszkiewicz, M., Ramanna, S., Jensen, R., Hu, Q. (eds.) RSCTC 2010. LNCS (LNAI), vol. 6086, pp. 138–147. Springer, Heidelberg (2010)
34. Napierala, K., Stefanowski, J.: Identification of different types of minority class examples in imbalanced data. In: Corchado, E., Snášel, V., Abraham, A., Woźniak, M., Graña, M., Cho, S.-B. (eds.) HAIS 2012, Part II. LNCS (LNAI), vol. 7209, pp. 139–150. Springer, Heidelberg (2012)
35. Napierala, K., Stefanowski, J.: BRACID: a comprehensive approach to learning rules from imbalanced data. Journal of Intelligent Information Systems 39(2), 335–373 (2012)
36. Prati, R., Batista, G., Monard, M.: Class imbalance versus class overlapping: An analysis of a learning system behavior. In: Proc. 3rd Mexican Int. Conf. on Artificial Intelligence, pp. 312–321 (2004)

37. 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 Inform. Systems 33(2), 245–265 (2012)
38. Sikora, M., Wrobel, L.: Data-driven adaptive selection of rule quality measures for improving rule induction and filtration algorithms. Int. J. General Systems 42(6), 594–613 (2013)
39. Stefanowski, J.: On combined classifiers, rule induction and rough sets. In: Peters, J.F., Skowron, A., Düntsch, I., Grzymała-Busse, J.W., Orłowska, E., Polkowski, L. (eds.) Transactions on Rough Sets VI. LNCS, vol. 4374, pp. 329–350. Springer, Heidelberg (2007)
40. Stefanowski, J.: Overlapping, rare examples and class decomposition in learning classifiers from imbalanced data. In: Ramanna, S., Jain, L.C., Howlett, R.J. (eds.) Emerging Paradigms in Machine Learning, pp. 277–306 (2013)
41. Stefanowski, J., Wilk, S.: Selective pre-processing of imbalanced data for improving classification performance. In: Song, I.-Y., Eder, J., Nguyen, T.M. (eds.) DaWaK 2008. LNCS, vol. 5182, pp. 283–292. Springer, Heidelberg (2008)
42. Stefanowski, J., Wilk, S.: Extending rule-based classifiers to improve recognition of imbalanced classes. In: Ras, Z.W., Dardzinska, A. (eds.) Advances in Data Management. SCI, vol. 223, pp. 131–154. Springer, Heidelberg (2009)
43. Tomasev, N., Mladenic, D.: Class imbalance and the curse of minority hubs. Knowledge-Based Systems 53, 157–172 (2013)
44. Weiss, G.M.: Mining with rarity: a unifying framework. ACM SIGKDD Explorations Newsletter 6(1), 7–19 (2004)
45. Wilson, D.R., Martinez, T.R.: Improved heterogeneous distance functions. Journal of Artifical Intelligence Research 6, 1–34 (1997)