# Combining rough sets and rule based classifiers for handling imbalanced data

Jerzy Stefanowski* and Szymon Wilk

Institute of Computing Science, Poznań University of Technology,
ul. Piotrowo 3A, 60–965 Poznań, Poland,
Jerzy.Stefanowski@cs.put.poznan.pl, Szymon.Wilk@cs.put.poznan.pl

**Abstract.** The paper presents two rough sets based filtering approaches combined with rule based classifiers suited for handling imbalanced data sets, i.e., data sets where the minority class of primary importance is under-represented in comparison to the majority classes. We introduced two techniques to detect and process inconsistent majority cases in the boundary between the minority and majority classes. The experiments showed that the best results were obtained for the relabel filtering, where inconsistent majority examples were reassigned to the minority class, combined with MODLEM rule induction algorithm.
**Keywords:** knowledge discovery, data mining, rough sets, classification, class imbalance, rule induction.

## 1 Introduction

The paper discusses problems of constructing rule based classifiers for the task of supervised learning from examples. There are several aspects that might cause difficulties for a learning algorithm and decrease performance of learned classifiers. One of these aspects is related to *class imbalance* in the input data, i.e. to a situation when one class (further called the *minority class*) includes much smaller number of examples comparing to other classes. A typical real life example is medicine, where databases with medical records regarding a rare (but important) disease usually contain a small group of patients requiring special attention while there is a much larger number patients from other classes. Similar situations occur in many other domains, e.g. in technical diagnostics, image analysis, fraud detection, text categorization, information retrieval and filtering. For more examples see, e.g., survey papers as [21, 9, 3].
Many learning system usually assume that the learning sets are balanced. However, this is not always the case and if the imbalance in the class distribution is high, i.e. some classes are *heavily under-represented*, these learning methods do not work properly. They are "somehow biased" to focus search on the more frequent classes while "missing" examples from the minority class. As a result final classifiers are also biased toward recognition of majority classes and they usually have difficulties (or even are unable) to classify correctly new unseen objects from the minority

* Corresponding author

class. In [13] authors described an information retrieval system, where the minority class (being of a primary importance) contains only 0.2% examples. Although the classifiers achieved the accuracy close to 100%, they were useless because they failed to deliver the requested documents from this class. The similar degradation of the classifier's performance for the minority classes is usually reported for other imbalanced problems, see e.g. [9, 11, 21].

The total classification accuracy is not the best characterization of the classifier's performance for imbalanced data sets as we usually do not have a *high enough recognition of the minority class.* Using again an example of medical diagnosis, the class of interest, being usually the minority class, is a critical one and costs of making wrong decisions, either false positive or false negative have different meaning. Therefore, diagnostic performance is characterized by *sensitivity* (the conditional probability of the set of correctly classified cases from the minority class, given the minority class - in other words, the ratio of correctly recognized patients from the critical class) and by *specificity* (the conditional probability of the set of correctly recognized cases from the majority class, given the majority class - in other words, the ratio of correctly excluded cases from not ill classes). In such applications more attention is given to sensitivity than to specificity [5, 7]. In general there is a kind of trade-off between these two measures and the *ROC (Receiver Operating Characteristics) curve* technique can be used to summarize classifier performance. The *Area Under Curve* (AUC) is used by many researchers to identify potentially good classifiers - for more details see, e.g., [3, 21].

The small number of examples in the minority class ("the lack of data") is not the only source of difficulties for inducing classifiers. Several researchers claim that besides the size of this class it is necessary to go deeper into its other characteristics. Quite often the minority class overlaps heavily the majority classes. In particular, boundaries between classes are ambiguous. Both boundaries and the inside of the minority class may be affected by noisy examples from other classes, which cause incorrect classification of many examples from the minority class. Their influence is more critical for this class than the majority ones, see e.g. experiments and discussions in [11, 12]. Japkowicz in her experimental study [9] also showed that the class imbalance becomes even a more difficult problem particularly when the minority class contains a number of very small subclusters, which are difficult to be learned (so called, a small disjunct problem). Other aspects, e.g. inappropriate evaluation measures or inductive biases of learning algorithms, are discussed in [21].

In recent years the problem of dealing with the class imbalance receives a growing research interest the machine learning and data mining communities. Although several methods have been proposed, see e.g. their review in [21], the research problem is still open.

In our previous work we attempted to modify the rule based classifier structure to increase its sensitivity for recognizing examples from the minority class [7]. We focused our interest on generating larger rule set of the minority class, while inducing minimal sets of rules for other majority classes. As a result of extending the number of minority class rules we increased the chance of predicting this class during the classification strategy for new objects.

However, this proposal is focused mainly on the uneven cardinalities of decision classes. As we discussed before it may be beneficial to consider more precisely boundary examples between classes. This leads us to a question about the possible use of *rough set theory* to capture this aspect. In general, rough set theory is claimed to be a well suited approach for handling vague information and inconsistent descriptions [10, 14]. Therefore, we could use it to detect inconsistent examples, which are located in class boundaries. Then, inconsistent examples from the majority classes could be removed while inducing rule sets for these classes or upper approximations could be used for inducing the minority class rules. Thus, a new contributions of this paper is an introduction of the rough set based approach and its experimental evaluation on several imbalanced data sets with different degree of inconsistency.

The paper is organized as follows. In Section 2 we shortly discuss previous related works. Then, in Section 3 we introduce the rough set based approach. In Section 4 we experimentally evaluate its usefulness in a comparative study with standard rough set based rule classifiers, induced by LEM2 and MODLEM algorithms, and finally we draw conclusions in Section 5.

## 2 Related works

We briefly discuss the most related proposals to our research, i.e. concerning either rule based approaches or identification of difficult examples. For more exhaustive reviews of other works, see, e.g., [21].

One of the most common technique for dealing with imbalance data is to transform the original class distribution into a more balanced by sampling. The basic approaches include random *over-sampling* or *under-sampling*. In the former approach the minority class examples are randomly replicated until a balance with cardinalities of majority classes is obtained. Random under-sampling goes in the opposite way - the majority class examples are randomly eliminated until obtaining the same cardinality as the minority class.

Drawbacks of the above simple random techniques are often reported [2, 3, 11, 21]. Random under-sampling can discard potentially useful majority class examples that could be valuable for learning a good classifier. On the other hand, simple over-sampling introduces copies of original examples only, which may lead to overfitting a classifier. Therefore, several more advanced heuristic techniques have been introduced.

In *one-side-sampling* [11] Kubat and Matwin selectively under-sampled the majority class while keeping the original content of the minority class. The examples were divided into four categories: *noisy* examples located inside the minority class region, *borderline* examples (i.e. these lying either on or very close to the decision border between classes), *redundant* examples (i.e. majority class examples which are quite distant from the decision border) and *safe* examples. The borderline and noisy examples from the majority class were detected using the Tomek links concept [20] (together with the condensed nearest neighbor rule) and then removed. Borderline examples were assumed to be unsafe since a small amount of noise could make them fall on the wrong side of the decision border between classes. Redundant majority class examples were also removed. Another approach to removing noisy and borderline examples is *Neighborhood Cleaning Rule* introduced by Laurikkala in [12]. First the Wilson's *Edited Nearest Neighbor Rule* is used to remove these majority class examples whose class labels differ from the class of at least two of its three nearest neighbors. Experimental studies [2, 12] showed that both above approaches provide better sensitivity and not worse total accuracy than a simple random over-sampling. Cleaning rule is usually better than one side sampling.

Yet another proposal is *SMOTE*, which over-samples the minority class by creating *new synthetic* examples. Its main idea is to create these new examples by interpolating several minority class examples that are close one to another. It widens decision boundaries for the minority class. The experimental results provided in [2, 3] indicate that SMOTE is often more efficient than other sampling methods considering AUC measure. Its mixture with elements of under-sampling may even improve the ability to predict the minority class - see [2]. Furthermore, there are interesting extensions of SMOTE for multiple classifiers. The aspects of modifying multiple classifiers for imbalanced data are also discussed in [21].

As for using rules, let us remind that typical rule induction algorithms exploit a *greedy search strategy* while looking for rule conjunctions which favors the majority class but may be ineffective in dealing with minority examples. Few researchers tried to develop *less greedy search* (an example is Brute algorithm [15] or a specific genetic algorithm [21]) or to change the *inductive bias of the algorithm*, e.g. Holte at al. modified the rule induction algorithm CN2 to improve its performance for small disjuncts referring to rare examples from the minority class [8]. Moreover, Weiss describes hybrid and two-phase rule induction [21], where one part focuses on optimizing sensitivity while the other corresponds to optimizing specificity. Other approaches may use knowledge about prior distribution of probabilities or transforming the task to cost sensitivity learning and to a deeper analysis of ROC convex hull, see , e.g, [21] or [3].

When discussing the role of rough set theory, we should notice that imbalance was not studied enough. Although a few authors in their application oriented papers calculated sensitivity or ROC measures, it seems that the most related research are Grzymala's works on increasing sensitivity of LEM2 rule classifiers by *changing rule strengths* [5, 6]. The strength of a rule is the number of learning examples, which satisfy both condition and decision part of this rule. The LEM2 algorithm [4] is used to induce the *minimal set of rules* covering examples from rough approximations of decision classes. Rules from the minority class have lower strength than rules from other classes. The main idea of the Grzymala's approach is to multiply the strength of all minority class rules by the same real number, called *strength multiplier*, while not changing the strength of rules from the majority classes. As a result, during classification of new cases, such minority class rules have an increased chance to classify correctly these examples. Another problem is selecting a proper value for the strength

multiplier. Grzymala proposed in [5] a procedure based on maximizing a measure called *gain* = *sensitivity* + *specificity* - 1. Experimental results confirmed that this approach outperformed the use standard LEM2 classifier for many imbalanced medical data sets [5, 6].
Moreover, in [7] we introduced another approach to improve the minority class prediction for rule base classifiers by *replacing rules*, which was inspired by previous studies with inducing more exhaustive set of rules [17–19]. It is based on a different principle than the Grzymala's proposal. A minimal set of rules for the minority class is replaced by a new set of rules with the strength greater than a certain threshold. Such rules are discovered by a special algorithm, called EXPLORE [19]. If the strength threshold is sufficiently low, EXPLORE may generate much more rules then LEM2. Thus, by using such rules for the minority class, while preserving the original set of rules for the majority classes, the chance that an example from the minority class is selected by a classifier is increased and sensitivity should improve.
In [7] we carried out a comparative study of the above both approaches and the standard LEM2 classifier on 9 imbalanced data sets coming mainly from UCI ML repository [1]. Both approaches performed similarly (differences between were globally insignificant) and better than standard LEM2 considering the sensitivity and gain measures without decreasing total accuracy, however, for some data sets it was accompanied by the decreased specificity.

## 3 Using rough sets to handle inconsistent examples before constructing rules

Both rule based approaches, presented in the previous section, are based either on modifying the classification strategy (strength multiplier) or changing structure of rule sets and they do not change the original distribution of data. However, coming back to approaches discussed in the previous section, as e.g. cleaning rule or SMOTE, one can notice that class overlapping, boundary region between them and noisy examples are also important for classifying examples from imbalanced classes. They can cause misclassification of k-NN and other classifiers [2, 3]. Furthermore, they may lead to overspecialized decision trees or rules, in particular for the minority class.
Taking into account the above inspiration we could pose a question: *is rough set theory able to identify such examples and to handle them during rule induction process in order to increase a prediction of minority class*?
Let us remind that rough sets start from establishing a binary relation $R$ between objects in a decision table $(U, A \cup \{d\})$, where $U$ is a set of objects, $A$ is a set of condition attributes describing objects and $d$ is a decision attribute expressing classification of objects. Originally the *indiscernibility relation* is considered, i.e. for any subset $B \subseteq A$ it is defined as $I(B) = \{(x, y) \in U \times U : a(x) = a(y), \forall a \in B\}$. Equivalence classes of the indiscernibility relation for object $x$ are denoted as $I_B(x)$. Some extensions of standard rough set theory use other relation as e.g. tolerance or similarity [10]. Objects described by indiscernible values of

attributes but belonging to different sets are *inconsistent*. As the result of inconsistency some sets of objects cannot be precisely defined using indiscernibility classes. But any set if objects $X \subseteq U$ is approximated by two sets *B-lower* and *B-upper approximations*, denoted as $\underline{B}X$ and $\overline{B}X$, respectively. The set $BN_B(X) = \overline{B}X - \underline{B}X$ is called *B-boundary* region of $X$ and contains inconsistent objects. One should notice that rough sets meaning of boundary refers to impossibility of classifying objects on the basis of their description, while boundary in the sense discussed in Section 2 corresponds rather to closest objects from different classes located on the other side of a decision border which could be affected by noise – however, it does not implicate that they have inconsistent descriptions. In spite of this difference we could check whether rough sets based technique of detecting inconsistent examples could be useful. So, we could handle these objects either by removing inconsistent examples from majority class or by considering the complete boundary region in a specific way. It leads us to two following *filtering approaches*:
The first approach:

1. Let $X$ be a minority class of interest. Calculate its approximations $\underline{B}X$ and $\overline{B}X$.
2. For the remaining classes of objects $Y_i \subseteq U \backslash X$ calculate their lower approximations $\underline{B}Y_i$.
3. Identify inconsistent objects from majority classes belonging to *B-boundary* region of $X$ and *remove them*.
4. For class $X$ induce rules using as *learning examples all objects* belonging to this class.
5. For remaining classes $Y_i$ induce rules using as learning examples *only* objects belonging to the lower approximations $\underline{B}Y_i$.

The second approach:

1. Perform the same operation as points 1 and 2 in the first approach.
2. Identify inconsistent objects from majority classes belonging to *B-boundary* region of $X$ and *change their labels* to the minority class $X$. In this way construct a new set $X' = \overline{B}X$.
3. For class $X$ induce rules using as learning examples objects from $X'$.
4. For remaining classes $Y_i$ induce rules using as learning examples *only* objects belonging to the lower approximations $\underline{B}Y_i$.

The first approach *removes* inconsistent examples from the majority class that make difficulties in the boundary region, therefore it will be denoted as *remove* hereafter, and we hope that this cleaning may give a chance for inducing less specific rules in this region. The second approach is more greedy as it *relabels* suspicious examples located on the wrong side of the border between classes. In the following text we will refer to it as *relabel*. The *relabel* approach is inspired by an edited nearest neighbor approach, e.g. Generalized Edited Algorithm proposed by Koplowitz and Brown. Moreover, another inspiration may be good experiences of Kubat and Matwin with the SHRINK system to detect oil spills in satellite images, see [11]. In general, the latter approach increases the number of examples in the minority class, so it may lead diminish the problem of small disjuncts while inducing stronger rules.

## 4 Experimental evaluation

We evaluated 8 data sets listed in Table 1. These data are coming either from the UCI repository [1] or from our applications in medicine (these are shortly characterized in [17]). Most of the considered data sets were originally composed of more than two decision classes, however, to simplify calculations we decided to group all majority classes into one. Two data sets (ecoli, cleveland) included real-valued attributes. As they might have caused a problem for rough sets based on the indiscernibility relation, we discretized them using a local approach based on minimizing entropy. Moreover, for two data sets (urology and breast-Poland) we used a reduced set of attributes to decrease the level of consistency.

**Table 1.** Characteristics of evaluated data sets

| Data set | Number of examples | Ratio of examples Minority | Ratio of examples Majority | Level of consistency |
|---|---|---|---|---|
| abdominal pain | 606 | 27.2% | 72.8% | 0.91 |
| acl | 140 | 28.6% | 71.4% | 0.88 |
| breast-Poland | 228 | 29.0% | 71.0% | 0.97 |
| cleveland | 303 | 11.6% | 88.4% | 0.94 |
| ecoli | 336 | 10.4% | 89.6% | 0.87 |
| hsv | 122 | 11.5% | 88.5% | 0.98 |
| scrotal pain | 171 | 28.6% | 71.4% | 0.89 |
| urology | 500 | 31.2% | 68.8% | 0.97 |

The experiment was based on the 10-fold cross validation with stratified selection, i.e., the distribution of classes in each fold was the same as in a whole data set. We used two proposed techniques to filter learning data sets: *remove*, where inconsistent majority examples were removed from a set, and *relabel*, where inconsistent majority examples were relabeled and assigned to the minority class. We employed two different algorithms for inducing decision rules: LEM2 [4] and MODLEM [17] and tested both with each filtering technique. Moreover, to establish a reference point for the results, we also induced and tested rules without prior filtering of learning data sets. In order to minimize the impact of splitting data into folds on the results, minimize their variances, and provide a reliable comparison, the split was conducted only once for each data set and the same pairs of learning and testing files were used for all calculations.

Tables 2 and 3 give the detailed results of classification performance of LEM2 and MODLEM combined approaches, respectively. It is measured for the minority class by its sensitivity, specificity, gain (expressed as ratio in [0,1]) and the total error for all classes (expressed as a percentage of tested examples). Table 4 presents characteristics of rules generated by those algorithms during calculations (all numbers were calculated as averages over 10 folds).

**Table 2.** Classification performance of LEM2 approaches

| Data set | Filtering | Sensitivity | Specificity | Gain | Error |
|---|---|---|---|---|---|
| abdominal-pain | none | 0.8118 | 0.8686 | 0.6804 | 14.69% |
| | remove | 0.7702 | 0.8777 | 0.6479 | 15.18% |
| | relabel | 0.8066 | 0.8641 | 0.6707 | 15.17% |
| acl | none | 0.7250 | 0.8900 | 0.6150 | 15.71% |
| | remove | 0.7500 | 0.8900 | 0.6400 | 15.00% |
| | relabel | 0.7500 | 0.8900 | 0.6400 | 15.00% |
| breast-cancer | none | 0.5738 | 0.7279 | 0.3017 | 31.52% |
| | remove | 0.5548 | 0.7088 | 0.2636 | 33.32% |
| | relabel | 0.5571 | 0.7217 | 0.2788 | 32.43% |
| cleveland | none | 0.3167 | 0.8732 | 0.1899 | 19.12% |
| | remove | 0.3500 | 0.8692 | 0.2192 | 19.11% |
| | relabel | 0.3167 | 0.8618 | 0.1785 | 20.11% |
| ecoli | none | 0.7167 | 0.9200 | 0.6367 | 10.13% |
| | remove | 0.7167 | 0.9200 | 0.6367 | 10.12% |
| | relabel | 0.7167 | 0.9233 | 0.6400 | 9.83% |
| hsv | none | 0.0500 | 0.8991 | -0.0509 | 19.49% |
| | remove | 0.0000 | 0.8991 | -0.1009 | 20.32% |
| | relabel | 0.0000 | 0.9082 | -0.0918 | 19.49% |
| scrotal pain | none | 0.6900 | 0.8276 | 0.5176 | 21.01% |
| | remove | 0.6900 | 0.7865 | 0.4765 | 23.92% |
| | relabel | 0.6700 | 0.7949 | 0.4649 | 23.92% |
| urology | none | 0.3717 | 0.6972 | 0.0689 | 40.40% |
| | remove | 0.3913 | 0.7123 | 0.1036 | 38.80% |
| | relabel | 0.4046 | 0.7061 | 0.1107 | 38.80% |

## 5 Conclusion

The aim of our experiments was to compare two rough set based approaches for filtering inconsistent examples in imbalanced data sets. The general observation is that their usefulness depends on a rule induction algorithm used on a filtered set.

The best results were obtained when filtering was coupled with the MODLEM algorithm. We recorded increased sensitivity and decreased the classification error for all tested data sets comparing to the results achieved for MODLEM and unfiltered data. The largest increase of sensitivity was observed for scrotal pain - 0.0850, abdominal pain - 0.0420, hsv - 0.1000, cleveland 0.0584 and ecoli - 0.0330. Moreover, the gain measure raised for all data sets, and the specificity measure was never decreased (for some data it was even improved, e.g. abdominal pain, acl, breast).

From the two filtering techniques used with MODLEM, the $r$elabel one turned out to be more effective in terms of increased sensitivity, as it could be observed for 6 of 8 data sets (for acl sensitivity was equal and for scrotal pain it was lower than for $r$emove).

LEM2 combined with any of the two filtering methods offered worse results than MODLEM. For sensitivity we observed increases for 3 data

**Table 3.** Classification performance of MODLEM approaches

| Data set | Filtering | Sensitivity | Specificity | Gain | Error |
|---|---|---|---|---|---|
| abdominal-pain | none | 0.7643 | 0.8709 | 0.6352 | 15.84% |
| | remove | 0.8063 | 0.8867 | 0.6930 | 13.54% |
| | relabel | 0.8121 | 0.8664 | 0.6785 | 14.86% |
| acl | none | 0.7750 | 0.8800 | 0.6550 | 15.00% |
| | remove | 0.8000 | 0.8800 | 0.6800 | 14.29% |
| | relabel | 0.8000 | 0.8800 | 0.6800 | 14.29% |
| breast-cancer | none | 0.5833 | 0.7346 | 0.3179 | 30.67% |
| | remove | 0.5405 | 0.7221 | 0.2626 | 32.81% |
| | relabel | 0.5905 | 0.7651 | 0.3556 | 28.46% |
| cleveland | none | 0.3083 | 0.8655 | 0.1738 | 19.76% |
| | remove | 0.3167 | 0.8843 | 0.2010 | 18.13% |
| | relabel | 0.3667 | 0.8734 | 0.2401 | 18.44% |
| ecoli | none | 0.7417 | 0.9167 | 0.6584 | 10.15% |
| | remove | 0.7500 | 0.9200 | 0.6700 | 9.84% |
| | relabel | 0.7750 | 0.9167 | 0.6917 | 9.85% |
| hsv | none | 0.0500 | 0.8900 | -0.0600 | 20.38% |
| | remove | 0.0500 | 0.8891 | -0.0609 | 20.45% |
| | relabel | 0.1500 | 0.9264 | 0.0764 | 16.35% |
| scrotal pain | none | 0.6500 | 0.8032 | 0.4532 | 23.92% |
| | remove | 0.7350 | 0.7705 | 0.5055 | 23.95% |
| | relabel | 0.6900 | 0.7942 | 0.4842 | 23.40% |
| urology | none | 0.3354 | 0.7179 | 0.0533 | 40.20% |
| | remove | 0.3542 | 0.7297 | 0.0839 | 38.80% |
| | relabel | 0.3604 | 0.6887 | 0.0491 | 41.40% |

sets only (acl, cleveland and urology); for other sets there was either the same performance as without filtering or even worse. The similar observation corresponds to gain and error.

MODLEM has an internal ability to construct more general syntax of elementary conditions (attribute in set of values), while LEM2 uses only a simplest form (attribute = value). Therefore, it should be able to induce more general (stronger) rules when supplied with a larger number of minority class examples. This presumption is supported by the characteristics of MODLEM rules presented in Tables 4, where the average strength increased for the $r$elabel filtering and 7 data sets (the most noticeable change was observed for ecoli, where the average strength raised from 1.8 to 10.4). On the contrary, Table 4 shows that it does not hold for LEM2. Moreover, it clearly shows that the $r$elabel filtering improved the minority class rules (in terms of their strength) without worsening the majority class rules – their numbers and strengths did not change.

Comparing the current work to our previous paper [7], we should admit that the approached described there led to greater increase of sensitivity and gain. However, the methods presented here are "light-weight" modifications of a learning data set that do not require extensive customization of parameters for rule induction algorithms and classification strategies.

**Table 4.** Characteristics of rules induced by LEM2 and MODLEM algorithms; 'Min' denotes the minority class and 'Oth' denotes majority classes; An average strength is calculated as a number of learning examples.

| | | Number of rules | | | | Strength of rules | | | |
|---|---|---|---|---|---|---|---|---|---|
| Data | Filtering | LEM2 | | MODLEM | | LEM2 | | MODLEM | |
| set | | Min | Oth | Min | Oth | Min | Oth | Min | Oth |
| abdominal | none | 24.0 | 35.9 | 26.3 | 33.1 | 10.2 | 31.4 | 8.3 | 32.6 |
| pain | remove | 24.8 | 36.6 | 24.7 | 33.5 | 12.7 | 30.9 | 12.2 | 32.6 |
| | relabel | 24.5 | 35.9 | 25.9 | 33.1 | 14.6 | 31.4 | 11.3 | 32.6 |
| acl | none | 6.5 | 8.7 | 6.5 | 7.7 | 4.6 | 19.5 | 4.6 | 24.3 |
| | remove | 7.6 | 8.8 | 7.2 | 7.7 | 5.9 | 19.2 | 5.8 | 23.8 |
| | relabel | 7.6 | 8.7 | 6.8 | 7.7 | 7.0 | 19.5 | 9.3 | 24.3 |
| breast-Poland | none | 30.5 | 31.6 | 21.2 | 26.2 | 1.9 | 6.3 | 3.1 | 7.8 |
| | remove | 31.2 | 31.6 | 23.5 | 27.0 | 2.0 | 6.3 | 2.8 | 7.9 |
| | relabel | 31.2 | 31.6 | 23.9 | 26.2 | 2.0 | 6.3 | 2.9 | 7.8 |
| cleveland | none | 17.1 | 23.9 | 16.8 | 22.5 | 1.0 | 24.3 | 1.0 | 28.1 |
| | remove | 19.2 | 23.6 | 17.3 | 22.1 | 1.0 | 25.5 | 1.5 | 29.2 |
| | relabel | 19.8 | 23.9 | 16.5 | 22.5 | 1.8 | 24.3 | 2.3 | 28.1 |
| ecoli | none | 8.9 | 13.4 | 6.9 | 8.9 | 1.2 | 36.4 | 1.8 | 59.6 |
| | remove | 8.6 | 13.3 | 5.9 | 9.0 | 4.2 | 36.0 | 7.1 | 59.9 |
| | relabel | 8.6 | 13.4 | 5.7 | 8.9 | 7.0 | 36.4 | 10.4 | 59.6 |
| hsv | none | 9.1 | 14.6 | 8.4 | 13.0 | 1.0 | 11.2 | 1.0 | 13.5 |
| | remove | 9.2 | 14.6 | 8.3 | 13.5 | 1.0 | 11.2 | 1.1 | 13.1 |
| | relabel | 9.1 | 14.6 | 8.3 | 13.0 | 1.0 | 11.2 | 1.3 | 13.5 |
| scrotal pain | none | 13.5 | 16.6 | 10.8 | 14.9 | 3.8 | 11.0 | 4.7 | 12.5 |
| | remove | 14.7 | 16.5 | 13.5 | 15.1 | 4.5 | 11.0 | 4.7 | 11.6 |
| | relabel | 14.9 | 16.6 | 13.8 | 14.9 | 4.9 | 11.0 | 5.4 | 12.5 |
| urology | none | 66.2 | 79.4 | 57.7 | 70.8 | 2.0 | 5.2 | 2.2 | 6.2 |
| | remove | 67.8 | 80.6 | 61.6 | 69.7 | 2.0 | 5.2 | 2.1 | 6.5 |
| | relabel | 67.0 | 79.4 | 60.7 | 70.8 | 2.0 | 5.2 | 2.6 | 6.2 |

Finally, we would like to note that the proposed approaches are based on the classical indiscernibility relation that may be insufficient for detecting all unsafe (borderline, but consistent) cases. More advanced methods of handling such cases are the subject of our ongoing research.

**Acknowledgment**: This research was partially supported by the State Committee for Research (KBN) of Poland, grant 3 T11C 050 26.

## References

1. Blake C., Koegh E., Mertz C.J.,: Repository of Machine Learning, University of California at Irvine 1999 [URL: http://www.ics.uci.edu/ mlearn/MLRepositoru.html].
2. Batista G., Prati R., Monard M., A study of the behavior of several methods for balancing machine learning training data. ACM SIGKDD Explorations Newsletter, vol. 6, no. 1, 2004, 20-29

3. Chawla N., Bowyer K., Hall L., Kegelmeyer W., SMOTE: Synthetic Minority Over-sampling Technique. Journal of Articifical Intelligence Research, vol. 16, 2002, 341–378.
4. Grzymala-Busse JW., LERS - a system for learning from examples based on rough sets. In Slowinski R. (eds.) Intelligent decision support. Handbook of application and advances of the rough sets theory. Kluwer Academic Publishers, 1992, 3-18.
5. Grzymala-Busse, J. W., Goodwin, L. K., and Zhang, X., Increasing sensitivity of preterm birth by changing rule strengths. Proceedings of the Eigth Workshop on Intelligent Information Systems (IIS'99), Ustron, Poland, June 14–18, 1999, 127–136.
6. Grzymala-Busse, J.W., Goodwin, L.K., Grzymala-Busse, W.J., Zheng, X., An approach to imbalanced data sets based on changing rule strength. Learning from Imbalanced Data Sets, AAAI Workshop at the 17th Conference on AI, AAAI-2000, Austin, TX, July 30–31, 2000, 69–74.
7. Grzymala-Busse J.W., Stefanowski J. Wilk Sz., A comparison of two approaches to data mining from imbalanced data. Proceedings of the KES 2004, 8-th International Conference on Knowledge-based Intelligent Information & Engineering Systems, Wellington, New Zealand, Springer LNCS vol. 3213, 2004,757-763.
8. Holte R.C., Acker L.E., Porter B., Concept learning and problem of small disjuncts. In Proceedings of 11th Joint Conf. on Artificial Intelligence, 1989, 813-819.
9. Japkowicz, N., Learning from imbalanced data sets: a comparison of various strategies. Learning from Imbalanced Data Sets, AAAI Workshop at the 17th Conference on AI, AAAI-2000, Austin, TX, July 30–31, 2000, 10–17.
10. Komorowski J., Pawlak Z., Polkowski L. Skowron A., Rough Sets: tutorial, In: Pal S.K., Skowron A. (eds.), Rough fuzzy hybridization. A new trend in decision-making. Springer-Verlag, 1999, 3–98.
11. Kubat M., Matwin S., Addresing the curse of imbalanced training sets: one-side selection. In Proceedings of 14th Int. Conf. on Machine Learning ICML 97, 1997, 179-186.
12. Laurikkala J., Improving identification of difficult small classes by balancing class distribution. Tech. Report A-2001-2, University of Tampere, 2001.
13. Lewis D., Catlett J., Heterogenous uncertainty sampling for supervised learning. In Proc. of 11th Int. Conf. on Machine Learning, 1994, 148-156.
14. Pawlak Z., Rough sets. Theoretical aspects of reasoning about Data. Kluwer Academic Publishers, 1991.
15. Riddle P., Segal R., Etzioni O., Representation design and Bruteforce induction in a Boening manufacturing fomain. Applied Artificial Intelligence Journal, vol. 8, 1994, 125-147.
16. Stefanowski J., The rough set based rule induction technique for classification problems. In Proceedings of 6th European Conference on Intelligent Techniques and Soft Computing EUFIT'98, Aaachen 7-10 Sept. 1998, 109-113.

17. Stefanowski J., Algorithims of rule induction for knowledge discovery. (In Polish), Habilitation Thesis published as Series Rozprawy no. 361, Poznan Univeristy of Technology Press, Poznan, 2001.
18. Stefanowski J., Wilk S., Evaluating business credit risk by means of approach integrating decision rules and case based learning. International Journal of Intelligent Systems in Accounting, Finance and Management, 10, 2001, 97–114.
19. Stefanowski J., Vanderpooten D., Induction of decision rules in classification and discovery-oriented perspectives. International Journal of Intelligent Systems, 16, 2001, 13–28.
20. Tomek I., Two modifications of CNN. IEEE Transactions of Systems, Man and Communications, SMC-6, 1976, 769-772.
21. Weiss G.M., Mining with rarity: a unifying framework. ACM SIGKDD Explorations Newsletter, vol. 6 (1), 2004, 7-19.