Ensembles of Abstaining Classifiers based on Rule Sets

Jerzy Błaszczyński, Jerzy Stefanowski, Magdalena Zając

Institute of Computing Science, Poznań University of Technology,
60-965 Poznań, Poland
{jerzy.blaszczynski, jerzy.stefanowski, magdalena.zajac}@cs.put.poznan.pl

Abstract. The role of abstaining from prediction by component classifiers in rule ensembles is discussed. We consider bagging and Ivotes approaches to construct such ensembles. In our proposal, component classifiers are based on unordered sets of rules. In these classifiers we use an appropriate classification strategy that solves ambiguous matching of the object's description to the rules. We propose to induce these rule sets by a sequential covering algorithm and to apply classification strategies using either rule support or discrimination measures. We adopt these strategies to abstaining by not using partial matching strategy. The other contribution of this paper is an experimental evaluation of this proposal. Results of comprehensive comparative experiments show that abstaining rule sets classifiers improve the accuracy, however this effect is more visible for bagging than for Ivotes.

1 Introduction

In recent years there has been much research on multiple classifiers also called ensembles of classifiers. Various approaches have been proposed to construct them with respect to either a phase of generating component classifiers or a phase of aggregating predictions of these classifiers (see [8] for a review).

We are particularly interested in ensembles containing classifiers based on *sets of rules* induced from diversified training samples. This interest results from previous research of two authors on various rule induction algorithms, which have been earlier used to build single classifiers and, then, also successfully applied inside some ensembles of classifiers [13, 14]. Furthermore, one can notice that *rule ensembles* such as SLIPPER or LRI proved to be competitive classifiers to more popular decision tree ensembles, see e.g., [4, 12, 15]. A rule classifier has an interesting property. Let us note that a rule assigns class only to these objects that it covers. Moreover, it covers only a bounded part of problem space as opposed to a decision tree. This property makes a rule classifier interesting to concern the research topic of this paper. Namely, studying changes in the aggregation phase of the ensemble when some component classifiers may *abstain* from predicting class labels.

Naturally, a set of rules does not have to maintain the ability to refrain that is typical for a single rule. Most rule sets classifiers are designed to always assign a class label for a new object, e.g., by using ordered priority lists of rules with a default class label [11] or specialized strategies for solving ambiguous conflicts with unordered rules [6]. However, there exist solutions where the classifier may not produce its class prediction in case of uncertainty as to the classified objects. Such classifiers called *abstaining classifiers* have already been studied in the framework of ensembles. Most of the research concerns refraining from the final decision in case of disagreement between votes of component classifiers, e.g., see a study [12] showing that it may improve the final accuracy. Some researches allow single classifiers to give no answer. For instance, rule ensembles like SLIPPER [4] are based on a weighted combination of *single rules* (being component classifiers) and a rule is excluded from voting if the new objects is not covered by it. However, according to our best knowledge there are no similar abstaining solutions for ensembles where component classifiers are based on *sets of unordered rules* induced by sequential covering algorithms (which are the most popular techniques for inducing rules [7]).

Therefore, the first aim of our paper is to present a framework for constructing such an ensemble of rule classifiers. To achieve it we propose to use sets of rules induced by the MODLEM algorithm [13]. MODLEM proved to be competitive to other rule and tree classifiers and was successfully applied inside pairwise coupling and extended bagging multiple classifiers [14]. Moreover, it naturally joins with classification strategies based on matching a description of a classified object to rules. Such strategies can be adopted to abstaining from a class prediction. To become independent of one specific solution we choose two different classification strategies: the first, introduced by Grzymala, based on rule support [6] and the other proposed by An [1], which employs a rule discrimination measure. Following similar motivations we decide to investigate two different approaches to constructing the ensemble: *bagging* [2] based on bootstrap sampling and *Ivotes* [3] using a sequential adaptive approach.

The second, not least important, aim of our paper is to experimentally evaluate the influence of abstaining component classifiers based on rule sets induced by MODLEM on the final accuracy of the ensemble. Although following some related results one could expect an improvement of classification accuracy, our contribution include a comprehensive comparative study on several benchmark data sets, where we want to examine more deeply the degree of changes with reference to different classification strategies and different ensembles.

2 Related Research

The idea of classifiers refraining from class predictions has been considered in machine learning, in particular in cases when classification is uncertain. The classified object may be located either in the boundary between classes or very far from any class. Some techniques as threshold classifiers producing distributions of membership to several class, e.g., neural networks, may naturally abstain from classification if none of predictions exceeds a preferred threshold. Such an unknown decision may be suitable in some domains, e.g., in medical diagnosis.

The concept of abstaining in ensembles of classifiers is considered at two levels. At the first level, abstaining occurs in the final decision of the ensemble. In this case, the ensemble may abstain from classification for uncertain objects, which are characterized by the smallest difference between the number of indications for the most often predicted class label and the number of indications for the second predicted label. Such a kind of abstaining classifier has been analysed in [12] showing theoretically that it may improve PAC bounds of error for rule ensembles. The same work also contains preliminary experiments with the authors' proposal of stochastic algorithm showing benefits of abstaining from making uncertain predictions and comparing it to bagged versions of popular rule induction algorithms RIPPER and PART. A method for optimization of similar abstaining classifiers using ROC analysis was presented in [10]. Moreover, this work contains an interesting review of other previous works on refraining from classification. As they are not directly related to our research, we skip them. At the second level, component classifiers may refrain from prediction. This has been postulated in [4, 5]. However, the result of abstaining at this level on the accuracy of the ensemble is, to our best knowledge, not deeply studied.

The most similar ensemble to presented in our paper is the one produced by SLIPPER [4]. More precisely, such a conclusion can be drawn with respect to presented further ensemble of rule sets classifiers produced by Ivotes. Nevertheless, SLIPPER is different to the approach we take to create an ensemble. It uses a single rule as a component classifier while we use a set of rules. Moreover, it aggregates component classifiers by a linear combination. We apply majority voting between component classifiers and we use mentioned before classification strategies inside each of component classifiers.

3 Our Framework for Abstaining Rule Ensembles

To construct ensembles we first chose *bagging*. It was introduced by Breiman [2] with a key concept of bootstrap sampling, i.e., uniform sampling with replacement objects, from the original learning set. Having several independent bootstrap samples, a set of classifiers is generated by the same learning algorithm and the final decision is formed by aggregating predictions of classifiers with the majority equal weight voting scheme.

The second considered ensemble comes from the *Pasting Small Votes* idea. Its original motivation was handling massive data which does not fit into computer's memory. Breiman proposed using the so-called pasting votes where many component classifiers are trained on relative small subsets of the original training data [3]. He introduced two strategies for implementing this idea: *Rvotes* and *Ivotes*. In Rvotes training sets are sampled randomly from large data sets (similarly to bagging). Ivotes sequentially generates small data sets using the *importance* sampling. According to this sampling each new training data sample should have approximately 50% objects that were misclassified by the ensemble including previously generated classifiers. The content of the particular training data sample for each subsequent classifier relies on sampling with replacement where the sampling probability results from the out-of-bag estimate [3]. We chose Ivotes as it is more similar to boosting idea and may be more accurate than standard bagging [8].

We decided to generate sets of rules by the MODLEM algorithm, which was originally introduced by Stefanowski in [13]. Due to the space limit we skip its more precise presentation (see [14] for details). Briefly speaking, it is based on the scheme of a *sequential covering* and it generates an *unordered minimal set of rules* for every decision concept. It is particularly well suited for analysing data containing a mixture of numerical and qualitative attributes, inconsistent descriptions of objects or missing attribute values. Searching for the best single rule and selecting the best condition is controlled by a criterion based on a modified entropy measure. As it will be further explained, MODLEM unordered sets of rules are better suited to introduced abstaining with partial matching strategies than ordered lists of rules.

Let us remind that both considered ensembles are based on manipulating presence of objects in bootstrap samples to produce *diversified training* samples. MODLEM is an *unstable algorithm* in the sense of Breiman's postulate [2], i.e., small perturbation of data may results in large changes in the induced rules, see also [14]. This is a desirable property for ensembles like bagging. Using unprunned structure should increase diversity of component classifiers, as it was also noted by Breiman and others [9].

The set of induced rules needs to be combined with a specific *classification strategy* to constitute a classifier. Most of these strategies are based on matching the new object's description to condition parts of rules. If rules are ordered into a priority list (as it is done in e.g., in popular C4.5rules [11]; another kind of exception list with default rule is used in RIPPER) the first matched rule from the list is "fired" to classify a new object. Unlike this option, in our case the set of rules is *unordered* and all rules are tested for matching. This may lead to three situations: a unique match (to one or more rules from the same class); *matching more rules* from different classes or *not matching* any rules at all. In both last situations a suggestion is ambiguous, thus, proper resolution strategy is necessary. We skip descriptions of some early proposals of solving it, e.g., by Michalski in AQfamily or Clark et al. in CN2. Review of the different strategies, which could be combined with MODLEM is given in [14].

For our experiment we choose the strategy introduced by Grzymala-Busse in [6] as it has been successfully applied in many experiments. Briefly speaking it is based on a voting of matching rules with their supports. The total *support* for a class K is defined as: $sup(K) = \sum_i^m sup(r_i)$, where r_i is a matched rule that indicates K, m is the number of these rules, and $sup(r)$ is the number of learning objects satisfying both condition and decision parts. A new object is classified to the class with the highest total support. In the case of not-matching, so called *partial matching* is considered where at least one of rule conditions is satisfied by the corresponding attributes in the new object's description x. In this case, a matching factor *match(r,x)* is introduced as a ratio of conditions matched by the object x to all conditions in the rule r. The total support is modified to

$sup(K) = \sum_i^p match(r, x) \times sup(r_i)$, where p is the number of partially-matched rules, and object x is assigned to the class with its highest value.

As an alternative strategy we apply proposal of Aijun Ann [1] because it also considers partial matching and its experimental verification with ELEM algorithm (simpler sequential covering than in MODLEM) showed that it is competitive to C4.5rules and CN2. It uses a rule quality measure different than the rule support, i.e., a *measure of discrimination*: $Q_{MD} = log\frac{P(r|K)\times(1-P(r|\neg K)}{P(r|\neg K)\times(1-P(r|K))}$, where P denotes probability. For more technical details of estimating probabilities and adjusting this formula to prevent zero division see [1]. Its interpretation says that it measures the extend to which rule r discriminates between positive and negative objects of class K. Inside the ELEM2 classification strategies it is used in similar formulas for decision scores as in the Grzymala's strategy - the only difference concerns putting Q_{MD} in place of $sup(r)$. Therefore, the difference between classification strategies is choosing another rule quality measure.

We propose to adopt both strategies to the abstaining from prediction by *switching off the partially matching* phase. It corresponds to the fact the induced rules establish an area of expertise for a classifier (i.e., a subspace of problem space that is covered by the rules). If an object completely matches a rule, it may be treated as being close to this area. Otherwise, in case when it is not matched by any rule, it is far from the area of expertise and it can be classified as unknown. Moreover, assuming that classifiers are generated from diversified samples it is more likely that their areas of expertise does not overlap. This should result in an ensemble of experts being able to classify new objects better than any of component classifiers.

4 Experiments

The main goal of experiments is to evaluate the influence of abstaining of component classifiers on the final accuracy of the ensemble. As we discussed in the previous section, we are conducting experiments for two different approaches to construct ensembles: bagging and Ivotes. Moreover, we study the use of two different classification strategies with matching object's descriptions to rule (either An's proposal with discrimination measure and Grzymala's proposal of using rule support). We additionally carried out the experiments for the MODLEM classifier, to show that these classification strategies are useful for working with the single classifier. In all versions the classifiers were based on unprunned sets of rules induced by MODLEM - it was always induced with standard options as described in [14].

A number of classifiers in bagging ensemble was 20 because we noticed that Ivotes ensemble usually consisted of less but close to 20 base classifiers. Size of the learning sample used by Ivotes algorithm was set to 50%. This value was chosen because of the data sets used in experiments. As they were not so big as idea of Pasting Small Votes assumes, we have to keep reasonable size of training sample. Moreover, one of the author's experience shows that default version of Ivotes classifier has good results on smaller data sets when size of the learning sample is higher than 40%. They also showed that Ivotes is competitive with the standard bagging when size of the learning sample is set close to 50%.

Table 1. Characteristics of data sets

Data set	Objects	Attributes	Classes
breast-w	699	9	2
bupa	345	6	2
credit-german	1000	20	2
crx	690	15	2
diabetes	768	8	2
ecoli	336	7	8
glass	214	9	7
heart-cleveland	303	13	5
hepatits	155	19	2
ionosphere	351	34	2
pima	768	8	2
sonar	208	60	2
vehicle	846	18	4
vowel	990	13	11

All experiments were carried out on 14 data sets from the UCI repository[1]. Their characteristics are given in Table 1. We chose them because they were often used by other researchers working with rule ensembles.

The classification accuracy was estimated by the stratified 10-fold cross-validation, which was repeated several times. Tables with results always contain an average classification accuracy with a standard deviation. Moreover in brackets we present a rank of the best performance among all variants of classifiers for the given data set (the smaller, the better). We show them because they are used in the statistical test further described. Last row of each table shows an average rank scored by a given classification strategy.

In the first experiment we evaluate the use of both classification strategies in the single classifier based on MODLEM induced rule. In Table 2 we show two variants: abstaining (i.e., classification without partial matching) and no abstain (use of complete strategies with partial matching). The second experiment concerns using abstaining inside ensembles of MODLEM based classifiers. Results of this experiment are presented in Tables 3 and 4.

We use a statistical approach to compare difference in performance between classifiers in variants which we mentioned above. First, we apply Friedman test to globally compare performance of four different classifiers on multiple data sets [7]. The null-hypothesis in this test is that all compared classifiers perform equally well. It uses ranks of each of classifiers on each of the data sets. The lower rank, the better classifier. We started from analyzing results of single MODLEM

[1] see http://www.ics.uci.edu/ mlearn/MLRepository.html

Table 2. Accuracy of a single classifier with different classification strategies.

Data set	Classification strategy			
	Discrimination measure		Rule support	
	abstain	no abstain	abstain	no abstain
breast-w	92.73±1.00 (3)	94.71±0.67 (1)	92.53±1.06 (4)	93.88±0.70 (2)
bupa	59.88±2.00 (4)	66.96±2.64 (2)	60.41±1.70 (3)	68.35±1.95 (1)
credit-german	62.72±1.41 (4)	68.26±1.51 (2)	63.56±1.51 (3)	71.26±0.92 (1)
crx	77.28±0.87 (4)	81.97±1.22 (2)	77.33±0.76 (3)	83.33±0.90 (1)
diabetes	64.17±1.29 (3)	70.73±0.90 (2)	63.85±1.12 (4)	71.20±0.87 (1)
ecoli	74.52±1.48 (3)	77.56±1.65 (1.5)	74.05±1.74 (4)	77.56±1.42 (1.5)
glass	62.71±2.18 (3)	70.28±2.10 (1)	61.96±2.60 (4)	70.09±1.72 (2)
heart-cleveland	71.42±2.03 (4)	76.83±1.71 (2)	71.75±2.09 (3)	77.76±1.79 (1)
hepatits	56.65±3.32 (3)	70.19±2.53 (2)	56.39±2.96 (4)	80.90±0.77 (1)
ionosphere	88.21±1.59 (3)	90.20±1.43 (2)	87.81±1.19 (4)	90.83±0.58 (1)
pima	64.87±1.32 (3)	71.95±1.41 (2)	64.61±0.88 (4)	72.01±0.98 (1)
sonar	67.21±2.16 (4)	75.77±3.38 (1)	67.79±1.72 (3)	75.00±1.01 (2)
vehicle	66.50±0.80 (3)	71.39±1.03 (1)	66.19±1.04 (4)	67.54±1.06 (2)
vowel	75.076±0.67 (3)	75.80±0.57 (2)	74.63±0.84 (4)	76.14±0.76 (1)
average rank	3.36	1.68	3.64	1.32

classifiers presented in Table 2. Friedman statistics for these results gives 59.59 which exceeds the critical value 2.84 (for confidence level 0.05). We follow the same procedure with results of bagging presented in Table 3 and results of Ivotes presented in Table 4. In case of bagging, Friedman statistics gives 6.03. In case of Ivotes, Friedman statistics gives 2.47. Thus, we can reject the null hypothesis, at given confidence level 0.05, for single classifier and bagging. On the other hand, the value of Friedman statistic for Ivotes is close to critical value (p-value for this test is 0.076). We have not presented complete post-hoc analysis of differences between classifiers. However, we show the average ranks of each of classifiers in tables. The results of Friedman test and observed differences in average ranks between classifiers allow us to state that there is a significant difference between them.

We continue our comparison with examination of importance of difference in classification performance between each pair of classifiers. We apply Wilcoxon test [7] with null-hypothesis that the medians of results on all data sets of the two compared classifiers are equal. Let us remark, that in the paired tests ranks are assigned to the value of difference in accuracy between compared pair of classifiers. When we apply this test to results of single MODLEM classifiers, it detects statistically important difference in pairs between classifiers that abstain and those that does not (p-values for both classification strategies are around 0.0001). In case of bagging, Wilcoxon test indicates an important difference between classifiers that abstain regardless of classification strategy and this that use discrimination measure while not abstaining (p-values in this case are

Table 3. Comparison of using different classification strategies in Bagging.

Data set	Classification strategy			
	Discrimination measure		Rule support	
	abstain	no abstain	abstain	no abstain
breast-w	96.28±0.52 (1)	96.17±0.28 (2)	96.08±0.45 (3)	95.34±0.29 (4)
bupa	73.28±1.15 (3)	73.51±1.74 (2)	73.10±1.57 (4)	74.67±0.75 (1)
credit-german	76.10±0.99 (2)	75.26±0.74 (4)	76.30±0.55 (1)	75.50±0.45 (3)
crx	86.35±0.25 (1)	85.54±0.42 (4)	86.26±0.17 (3)	86.32±0.30 (2)
diabetes	75.05±0.68 (3)	75.00±0.44 (4)	75.36±0.71 (1)	75.26±0.68 (2)
ecoli	84.29±0.35 (2)	82.56±0.30 (3)	84.70±0.24 (1)	81.01±0.29 (4)
glass	77.29±1.09 (2)	75.98±1.05 (3.5)	77.38±1.44 (1)	75.98±1.27 (3.5)
heart-cleveland	80.92±1.44 (3)	80.40±1.87 (4)	81.19±0.86 (2)	81.52±1.52 (1)
hepatits	81.42±1.89 (3)	77.81±1.90 (4)	81.68±2.26 (2)	82.19±1.33 (1)
ionosphere	93.33±0.39 (2)	92.54±0.71 (4)	93.50±0.38 (1)	93.22±0.33 (3)
pima	75.47±0.62 (3)	74.92±0.78 (4)	75.91±0.83 (1)	75.76±0.63 (2)
sonar	83.56±0.71 (2.5)	83.56±1.23 (2.5)	84.04±0.64 (1)	81.73±1.43 (4)
vehicle	75.67±0.70 (1)	75.08±0.66 (3)	75.53±0.80 (2)	72.70±0.57 (4)
vowel	94.34±0.26 (1.5)	88.18±0.55 (4)	94.34±0.18 (1.5)	91.86±0.19 (3)
average rank	2.14	3.43	1.75	2.68

around 0.005). A difference is also reported between abstaining classifier that use rule support and the one that is not abstaining and use rule support (p-value 0.058). Moreover, there is a statistically important difference between abstaining classifier that use discrimination measure and abstaining classifier that use rule support (p-value equal to 0.043). In case of examining Ivotes results, the situation is slightly different. In this case, statistically important differences are found only when abstaining classifiers that use discrimination measure are compared in pairs with not abstaining classifiers (p-value around 0.05).

5 Conclusions

Let us summarize results of experiments. First of all, we conclude that introducing abstaining of classifiers by excluding partial matching for rule sets has improved the total accuracy of the ensemble. However, the statistical analysis clearly shows that the range of this improvement depends on the type of ensemble and classification strategy.

First, we conclude that classification improvements are more significant for bagging than for Ivotes. This conclusion is further confirmed by values of average ranks. We can attribute this effect to the adaptive nature of Ivotes. In importance sampling consecutive classifiers should be more focused on learning objects misclassified by classifiers constructed in previous iterations. This may reduce the effect of abstaining. We suspect that similar behavior may be observed for other boosting approaches. On the other hand, in bagging, classifiers are constructed on independent samples. Moreover, each of the classifiers is constructed

Table 4. Comparison of using different classification strategies in Ivotes.

Data set	Classification strategy			
	Discrimination measure		Rule support	
	abstain	no abstain	abstain	no abstain
breast-w	96.76±0.13 (1)	96.33±0.24 (2)	95.80±0.24 (3)	95.18±0.36 (4)
bupa	72.08±1.43 (3)	71.30±0.63 (4)	72.17±0.47 (2)	73.82±1.98 (1)
credit-german	75.37±0.09 (3)	75.87±0.21 (1)	75.67±0.25 (2)	75.23±0.76 (4)
crx	86.33±0.14 (2)	86.28±0.48 (3)	85.99±1.20 (4)	86.71±0.60 (1)
diabetes	75.74±0.93 (2)	75.17±0.32 (3.5)	75.17±0.12 (3.5)	76.13±0.37 (1)
ecoli	84.42±1.25 (2)	83.43±0.70 (3)	85.32±1.38 (1)	81.55±0.24 (4)
glass	74.92±0.79 (2.5)	74.92±1.88 (2.5)	75.08±1.23 (1)	73.52±2.10 (4)
heart-cleveland	82.95±0.56 (1.5)	82.95±0.82 (1.5)	81.63±1.12 (1)	81.96±2.02 (3)
hepatits	84.09±0.80 (2)	75.27±1.69 (4)	84.95±0.80 (1)	82.15±0.80 (3)
ionosphere	93.16±0.23 (3)	92.78±0.59 (4)	93.73±0.23 (1)	93.45±0.23 (2)
pima	76.22±0.75 (1)	75.56±0.06 (4)	75.91±0.38 (2)	75.82±0.85 (3)
sonar	79.49±1.26 (2)	78.37±3.74 (3)	79.65±1.94 (1)	76.60±1.63 (4)
vehicle	74.23±0.60 (3)	74.55±0.20 (2)	75.14±0.22 (1)	73.68±0.15 (4)
vowel	91.78±0.50 (1)	86.13±0.62 (4)	91.58±0.98 (2)	91.18±0.59 (3)
average rank	2.07	2.96	2.04	2.93

separately. Errors made by each of the component classifiers in learning phase do not affect the other classifiers. This makes the effects of abstaining more visible as it is not compensated during learning.

Although the aim of our experiment was not to compete with other rule ensembles, we also refer our best results against literature results of SLIPPER and other variants of bagging with rules [4, 12], noticing comparable accuracy.

Analysing the results of using classification strategies, with discrimination measures or rule support, we claim that abstaining helped for both of them. The advantage depends on the particular ensemble. Generally speaking, Grzymala's strategy with rule support is a bit more effective, in particular for bagging. However, the other strategy also works surprisingly well in abstaining ensembles. Although its classification performance is worse than for using rule support, the difference of accuracies between variants with and without abstaining mechanism are larger than for the strategy with rule support. We can interpret it by specificity of evaluating discrimination measures. For unprunned sets of rules, which is a case in our experiment, the values of this measures are very similar among rules (most of them equal to 1). So the matching strategy is not powerful, as just counts nearly equally important rules. In the other strategy values of rule support are strongly diversified and may more contribute to class prediction. We also noticed that this difference much influences the performance of the single classifier (see Table 2) where the partial matching significantly improved the accuracy, however, using rule support is definitely more effective.

Finally, in our experiments we also recorded the average number of classifiers that refrain from predictions. We can conclude that for data sets, where abstaining has improved accuracy, the number of these abstaining classifiers is not very high – on average usually between 2 and 4 classifiers (e.g., with respect to 20 components in bagging). This observation can lead us to a research question whether it is worth to further increase the level of abstaining. In our framework, it is possible to modify multiple matching part of classification strategy and produce unknown answer in case of uncertainty between two competitive class assignments. This could be a topic of future research.

References

1. An Aijun: Learning classification rules from data. Computers and Mathematics with Applications, 45 (2003) 737–748.
2. Breiman, L.: Bagging predictors. Machine Learning, 24 (2) (1996) 123–140.
3. Breiman, L.: Pasting small votes for classification in large databases and on-line. Machine Learning, 36 (1999) 85–103.
4. Cohen W, Singer Y.: A simple, fast and effective rule learner. In Proc. of the 16th National Conference on Artificial Intelligence AAAI-99, (1999), 335–342.
5. Freund Y., Schapire R. E., Singer Y., Warmuth M. K.: Using and combining predictors that specialize. In Proceedings of the 29th ACM symposium on Theory of Computing (1997) 334–343.
6. Grzymala-Busse, J. W.: Managing uncertainty in machine learning from examples. Proc. 3rd Int. Symp. in Intelligent Systems (1994) 70–84.
7. Kononenko, I., Kukar, M.: Machine Learning and Data Mining. Horwood Pub. 2007
8. Kuncheva L.: Combining Pattern Classifiers. Methods and Algorithms. Wiley (2004).
9. Mease, D. , Wyner, A.: Evidence Contrary to the Statistical View of Boosting. Journal of Machine Learning Research 9 (Jun. 2008), 131–156.
10. Pietraszek T.: Optimizing abstaining classifiers using ROC analysis. In Proc. of the 22st Int. Conf. on Machine Learning, ICML2005, (2005), 665 - 672
11. Quinlan J.R., C4.5: Programs for Machine Learning, Morgan Kaufmann, 1992.
12. Rucket U., Kramer S.: Towards tight bounds for rule learning. In Proc. of the 21st Int. Conf. on Machine Learning, ICML2004, (2004), 711-718.
13. Stefanowski J.: The rough set based rule induction technique for classification problems. In Proc. of the 6th European Conf. on Intelligent Techniques and Soft Computing EUFIT-98, (1998), 109–113.
14. Stefanowski, J.: On combined classifiers, rule induction and rough sets. In: Transactions on Rough Sets VI, LNCS, vol. 4374, Springer, 2007, 329–350.
15. Weiss S. M., Indurkhya N.: Lightweight rule induction, In Proc. of the 17st Int. Conf. on Machine Learning, ICML2000, (2000), 1135–1142.