Increasing the Interpretability of Rules Induced from Imbalanced Data by Using Bayesian Confirmation Measures

Krystyna Napierała[1,2], Jerzy Stefanowski[1], and Izabela Szczęch[1]

[1] Institute of Computing Science, Poznań University of Technology, 60-965 Poznań, Poland
[2] DATAX sp. z o.o. 53-609 Wroclaw, Poland

Abstract. Approaches to support an interpretation of rules induced from imbalanced data are discussed. In this paper, the rule learning algorithm BRACID dedicated to class imbalance is considered. As it may induce too many rules, which hinders their interpretation, their filtering should be applied. We introduce three different post-pruning strategies, which aim at selecting rules having good descriptive characteristics. The strategies are based on combining Bayesian confirmation measures with rule support, which have not yet been studied in the class imbalance context. Experimental results show that these strategies reduce the number of rules and improve values of rule interestingness measures at the same time, without considerable losses of classification accuracy, especially for the minority class.

Keywords: Bayesian confirmation measures, interpretability of rules, class imbalance, rule post-pruning

1 Introduction

Learning classification rules is one of mature and well studied tasks in machine learning. The popularity of rules comes from the fact that they directly provide a symbolic representation of knowledge discovered from data, which is more comprehensible and human-readable than other representations [4]. Many various algorithms for inducing rules have been already introduced (for their review see, e.g. [4]). Nevertheless, such aspects of data complexity as *class imbalance* still constitute difficulties [10]. The majority of standard rule algorithms are biased towards the majority classes and tend to neglect the minority class. Two kinds of reasons for poor performance of rule based classifiers for imbalanced data are usually pointed out – algorithmic and data level ones [10, 12].

Some extensions of rule classifiers for class imbalances have been already proposed, for their review see [12]. However, most of them address only a single or at most a few of algorithmic or data-related factors. In [12] we introduced a new rule induction algorithm, called BRACID (the acronym of Bottom-up induction of Rules And Cases for Imbalanced Data), which attempts to deal with more of the aforementioned factors. The previous experimental studies demonstrated that the rule classifier induced by BRACID significantly outperformed other popular rule classifiers as well as the extensions specialized to class imbalances, with respect to predictive measures [12]. On the other hand, BRACID may generate too many rules (see also experiments in Section 5). As it restricts human experts' abilities to analyze or interpret the rules, we are looking for a post-processing approach that could identify the most valuable rules. The first attempt, recently undertaken in [14], has shown that it is possible to select rules characterized by high supports and still leading to sufficient predictive performance.

Nevertheless, focusing attention on the most interesting rules should also take into account other rule characteristics than simply rule support. In particular, it is important not to neglect the descriptive abilities of rules, which are often overwhelmed by the need to increase the predictive performance. Note that the predictive and descriptive aspects often stand in opposition to each other. However, when human experts seek for a compact knowledge representation, improving the interpretability of each single rule (i.e. working within the descriptive perspective) can even justify some loses on the predictive performance.

Establishing when rules are interesting to users touches both subjective (user-based) and objective (data-driven) aspects [3]. In this paper we follow the latter aspect and consider *rule interestingness measures* which are often applied to filter the set of rules [6, 11]. They are calculated from learning examples and aim at quantifying the relationship between a rule's premise and its conclusion. A particular group of these interestingness measures, called *Bayesian confirmation measures*, is well suited for supporting rule interpretability, as it focuses on advancing rules for which the probability of the conclusion given the premise is greater than the genuine probability of the conclusion itself [2, 9].

Although the concept of confirmation has been firstly considered by philosophers of science in a very different context (see e.g. [1, 2, 15]), it has been adopted to rule interestingness measures, mainly for filtering association rules [7] and more recently for classification rules [8, 9]. Nevertheless, these measures have not been considered for imbalanced data yet. Their application should turn out to be particularly useful in the context of imbalance since considering the probability of each conclusion separately would be related to imbalance ratios.

For the purpose of this paper we focus on two particular confirmation measures called S [1] and N [15]. We have chosen them from a wider collection of confirmation measures discussed in the literature because of the desired properties that they possess [8, 9]. In our opinion, these measures satisfy properties that should influence the interpretability of rules [9].

The main aim of this paper is to introduce an approach that uses confirmation measures S and N to post-prune rules induced by BRACID. This approach should reduce the number of the rules while improving values of rule interestingness measures at the same time, especially within the minority class.

To achieve these aims, firstly we briefly review rule confirmation properties and formalize both S and N measures. Then, in section 3, the algorithm BRACID is summarized. The three new rule post-pruning strategies are introduced in section 4. Their usefulness to improve BRACID rules is evaluated in several experiments, which are described in section 5. In the final section we discuss these results and draw lines of future research.

2 Bayesian Confirmation Measures

Rules are consequence relations represented as *IF* (*condition part*) *THEN* (*target class*), where a condition part (premise) is a conjunction of elementary tests on values of attributes characterizing learning examples and a target class points to one of the predefined values of the decision attribute (represented in a rule conclusion). For simplicity, rules will be denoted as $E \to H$ or simpler as R.

If the number of induced rules exceeds human-expert's abilities to inspect them, the interestingness measures are applied to filter them (for their review see [6, 11]). For instance, in case of association rules or richer sets of classification rules, measures as *support* or $confidence$ are often exploited.

Interestingness measures quantify the relationship between E and H, and are defined as functions of four non-negative values that can be gathered in a 2×2 contingency table (see Table 1). For a particular data set, a is the number of objects that satisfy both the rule's premise and its conclusion, b is the number of learning examples for which only H is satisfied, etc. For instance, the support of $E \to H$ rule is defined as $sup(H, E) = a$ and its confidence as $conf(H, E) = a/(a{+}c)$. Note that a, b, c and d can also be regarded as frequencies for estimating probabilities: e.g. $P(E) = (a + c)/n$ or $P(H) = (a + b)/n$.

Table 1. An exemplary contingency table of the rule's premise and conclusion

	H	$\neg H$	Σ
E	a	c	$a+c$
$\neg E$	b	d	$b+d$
Σ	$a+b$	$c+d$	n

The need to achieve good interpretability and high descriptive abilities of rules drew our attention to a particular group of interestingness measures, called *Bayesian confirmation measures* (or simply *confirmation measures*). All those measures are characterized by a feature called *property of Bayesian confirmation*, which requires that an interestingness measure $c(H, E)$ obtains: positive values when $P(H|E) > P(H)$; 0 when $P(H|E) = P(H)$; and negative values when $P(H|E) < P(H)$.

Thus, confirmation measures are designed to depict simply through their scale the confirmatory, neutral or disconfirmatory impact of the rule's premise on its conclusion. Confirmation, interpreted as an increase in the probability of the conclusion H provided by the premise E, is a desirable situation. Let us stress that basic interestingness measures such as support or confidence do not possess the property of confirmation and thus, their utility is lower for the descriptive perspective of knowledge discovery.

Note that the property of confirmation leaves plenty of space for defining various, non-equivalent confirmation measures (for a review see [2, 8]). To guide the user towards the measures that reflect his expectations, researchers proposed special properties of confirmation measures. These properties express requirements for a measure behaviour in certain situations. We focus our interest on the property of monotonicity M [9], which favors these measures that are non-decreasing with respect to a and d, and non-increasing with respect to b and c. It is intuitively clear that we would like higher values of measures for rules that are supported by a greater number of positive examples (i.e. increase of a), and exactly the opposite when the number of counter-examples grows (i.e. increase of c). Nevertheless, some confirmation measures act contrarily to property M. In this paper we narrowed our study to two measures that have been verified as satisfying many desirable properties (including the property of monotonicity M) [8, 9]. The chosen measures $S(H, E)$ [1] and $N(H, E)$ [15] are defined as:

$$S(H, E) = P(H|E) - P(H|\neg E) = \frac{a}{a + c} - \frac{b}{b + d}, \quad (1)$$

$$N(H, E) = P(E|H) - P(E|\neg H) = \frac{a}{a + b} - \frac{c}{c + d}. \quad (2)$$

The values of $S(H, E)$ and $N(H, E)$ range from -1 (showing complete disconfirmation) to 1 (showing complete confirmation).

3 Rule Induction with BRACID

BRACID is a specialized algorithm to learn rules from imbalanced data. For its details see [12]. Here, we summarize its main characteristics:

– **Hybrid representation of rules and instances**: BRACID tries to create a general description in regions where the examples form large disjuncts (using rules) and exploit good properties of instances in the more difficult regions. BRACID allows some (difficult) examples to remain not generalized to rules. They can be treated as maximally specific rules.
– **Bottom-up rule induction**: Unlike a top-down strategy typical for rule induction, BRACID follows bottom-up (or specific-to-general) strategy as a more appropriate for imbalanced data. It starts from the set of most specific rules each covering a single learning example – which is called a seed of the rule. Then, in every iteration each rule is generalized in the direction of the nearest neighbour example from the same class, provided that it does not decrease the classification abilities of the whole rule set. The procedure is repeated until no rule in the set can be further generalized.
– **Resignation from greedy, sequential covering technique**: As this technique, popular in typical rule learning algorithms, increases the data fragmentation and is problematic for the minority examples, BRACID takes into account all the learning examples when evaluating new rule candidate.

- **Facing borderline minority examples**: Types of learning examples are evaluated and rules are generated differently depending on the type of the seed example of a rule [13]. The minority examples belonging to the borderline region are allowed to be generalized into more than one rule, to lessen the dominance of the majority class in this region.
- **Facing noisy examples from the majority class**: Noisy majority examples, present inside the minority class regions, may hinder the induction of general minority rules. BRACID has an embedded mechanism for detecting and removing such examples from the learning data set.
- **Less biased classification strategy**: BRACID employs a classification strategy based on nearest rules to diminish the domination of strong majority rules during solving conflict situations while a new instance matches condition parts of many rules.

Note that some mechanisms employed in this algorithm lead to the increase of the number of rules (mainly a bottom-up rule induction and generation of more rules in the borderline regions). However, the increased number of rules for the minority class, coupled with an increased rule support, are beneficial for final classification. The experimental evaluation of classification performance of BRACID showed indeed that it significantly outperformed other standard rule classifiers as well as approaches specialized for class imbalance such as PART algorithm combined with SMOTE preprocessing – see details in [12].

4 Selecting Rules with Respect to Confirmation

We want to select a subset of induced rules with respect to appropriate rule evaluation measures. In [14] we have already postulated that it would be profitable to find rules which cover diverse sets of examples referring to different sub-parts of the class distribution. Focusing the expert attention on a subset of rules having such characteristics should be particularly good for the minority class which is often decomposed into many rare sub-concepts.

Recall that several post-pruning techniques have already been proposed to order rules or to reduce their number. However, as we discussed in [14], it may not lead to diverse subsets of rules in BRACID, as e.g. high supports may characterize many rules having similar syntax and covering similar subsets of learning examples. Other post-pruning techniques considered in rule classifiers are focused on optimizing the predictive performance of the rules rather than on improving their descriptive properties [4].

Therefore, we follow a different inspiration, coming from using rules to represent patterns in *subgroup discovery*, where the task is to find subgroups of individuals that are statistically "most interesting" (e.g. covering as many examples as possible and having the most unusual statistical characteristics [4]). We have decided to generalize the algorithm originally proposed in [5] to find rules describing subgroups.

Our approach to select a given number of diverse rules with respect to a given rule evaluation measure is presented in Algorithm 4. It is run for each class separately and takes as an input the set of all rules induced for this class and their required number after selection – later on we discuss how to tune it.

Algorithm 1 Rule Filtering Algorithm

Input: Set of Rules S for class P, required NUMBER of rules; rule evaluation ev;
Output: Prunned set of rules FR

```
Delete rules with too low confirmation from S
FR ← ∅
for every example e ∈ P do
  c(e) ← 1
end for
repeat
  for each rule R ∈ S do
    calculate rule evaluation measure ev(R)
  end for
  Select R_max = arg max_R(ev(R))
  for each e covered by R_max do
    c(e) ← c(e) + 1
  end for
  Remove R_max from S
  FR = FR ∪ R_max
until size of FR = NUMBER
```

Firstly, we remove all rules with the non-positive value of a selected confirmation measure (except the option where rules are evaluated with the support only). The key idea of the algorithm is to assign a weight $c(e)$ to each learning example. It is initialized with $c(e) = 1$ for all examples from the given class. When rule R is selected, then weights for examples covered by this rule are increased by adding 1. Then, while evaluating the next rule being a candidate for selection, the example takes part in all calculation with the weight $1/c(e)$. For instance, the support of a rule is computed as a sum of $1/c(e)$ for all target class examples covered by this rule.

This weighted coverage causes that in the subsequent iterations of the algorithm, examples already covered by the selected rules contribute less to the evaluation of new rule. It promotes the rules referring to examples not yet covered and directs the search toward diverse regions of the class.

In this study we will consider three different versions of the rule evaluation $ev(R)$[3] for selecting rules:

1. a standard rule support $sup(R)$;
2. a product of support with a confirmation measure $N : sup(R) \times N(R)$;
3. a product $sup(R) \times S(R)$.

[3] For simplicity we will further use a notation of a rule as R instead of (H, E) in symbols of confirmation measures

The choice of rule support $sup(R)$ results from earlier experiments in [14] and we want to consider it as a baseline. The choice of both confirmation measures S and N has been justified in Section 2. We want to aggregate them with a rule support to represent a trade off in a bi-criteria evaluation where the user is interested in sufficiently strong patterns describing the classes; see also earlier experiences with multiple criteria evaluation of rules.

5 Experiments

In the experiments we will verify whether the proposed post-pruning strategies select a limited number of BRACID rules having better values of interestingness measures than in case of non-pruned rules.

As the evaluation criteria we choose the average values of confirmation measures S and N, rule support and rule confidence. We consider the last two measures due to their popularity in the previous rule filtering techniques and to their easy interpretation for the users. These criteria represent descriptive properties of single rules with respect to their possible interpretability and they are treated as primary criteria in our study. As a secondary criterion, we also evaluate the predictive ability of the rule set, which will be estimated by means of G-mean and sensitivity. We use this criterion to control whether pruning the set of rules does not dramatically deteriorate the performance compared to all rules produced by the BRACID algorithm. The predictive measures are evaluated in a repeated stratified 10-fold cross validation procedure while rule evaluation measures are calculated for a set of rules induced from the complete data set.

Table 2. Basic characteristics of datasets

Data set	#Examples	Minority class size	Imbalance ratio [%]	#Attributes (numeric)	Minority class name
balance-scale	625	49	7.84	4 (4)	B
breast-cancer	286	85	29.72	9 (0)	rec-events
cmc	1473	333	22.61	9 (2)	long-term
haberman	306	81	26.47	3 (3)	died
hepatitis	155	32	20.65	19 (6)	die
transfusion	748	178	23.80	4 (4)	yes

We analysed previous experiments from [12] and chose 6 data sets, where BRACID generated too many rules compared to other, standard rule induction algorithms. Although the imbalance ratios of some of these data sets are medium, all these data are also affected by different difficulty factors characterizing the distribution of examples from the minority class. According to experimental studies [13] these factors lead to difficulties while learning rules.

All these data sets come from the UCI repository. We analyzed them as binary problems – the minority class vs. majority one (which may aggregate others), as it a typical view of class imbalances with focusing attention on improving recognition of the class of special importance. The basic characteristics of these data sets are presented in Table 2.

We checked that for all data sets (except hepatitis), rule sets contained some rules with negative values of confirmation measures. For instance, balance-scale contained 8, cmc 19 and transfusion 14 such rules.

Table 3. Characteristics of pruned rules for the minority class

Data set	Pruning	#Rules	Avg.sup	Avg.conf	Avg.S	Avg.N
balance-scale	none	52	2.08	0.61	0.54	0.03
	sup	5	6.00	0.27	0.19	0.06
	sup*N	5	4.60	0.32	0.24	0.07
	sup*S	5	2.00	0.88	0.80	0.04
breast-cancer	none	77	3.36	0.71	0.42	0.03
	sup	8	9.63	0.71	0.43	0.09
	sup*N	8	10.13	0.74	0.46	0.10
	sup*S	8	9.13	0.82	0.54	0.09
cmc	none	354	6.59	0.72	0.50	0.02
	sup	35	14.91	0.67	0.45	0.04
	sup*N	35	18.57	0.65	0.43	0.05
	sup*S	35	12.69	0.78	0.56	0.03
haberman	none	122	6.05	0.72	0.46	0.06
	sup	12	9.92	0.65	0.41	0.10
	sup*N	12	12.25	0.78	0.55	0.14
	sup*S	12	9.42	0.90	0.66	0.11
hepatitis	none	66	7.42	0.99	0.82	0.23
	sup	7	12.00	0.97	0.83	0.37
	sup*N	7	12.57	1.00	0.86	0.39
	sup*S	7	12.57	1.00	0.86	0.39
transfusion	none	161	6.36	0.67	0.44	0.03
	sup	16	16.06	0.63	0.40	0.07
	sup*N	16	18.50	0.68	0.46	0.08
	sup*S	16	15.56	0.77	0.54	0.07

While using the algorithm for selecting rules we need to define a number of required rules as the stopping condition. In general, this parameter should represent the analyst's expectations and his abilities to inspect the rules. Here we recall our previous experiments [14], where we studied a wide range of values of this parameter (up to 30%). The results showed that the threshold 10% often led to rule sets having the good average rule support and comparable classification performance as the original set of BRACID rules.

Yet another option is to select all the rules which are necessary to cover all the learning examples in each class. We studied this coverage option in [14] and observed that it usually produced higher classification prediction (with respect to G-mean or sensitivity measure) than the percentage option. However, it also selected more rules than the percentage option. As in this study we aim at reducing the number of rules, we decided to consider the percentage option with the parameter tuned to 10% of the original set of rules for each class[4].

In our study, we will examine three proposed strategies to select rules with the rule evaluation $ev(R)$ (see Section 4), defined as: (1) a standard rule support $sup(R)$; (2) a product $sup(R) \times N(R)$; and (3) a product $sup(R) \times S(R)$.

The rule characteristics with respect to considered criteria are given in Table 3 and 4, for the minority and majority class, respectively. The column "pruning" corresponds to the selection strategy (note that results for using the standard version of BRACID without pruning is presented in the first row for each dataset with an abbreviation "none").

Table 4. Characteristics of pruned rules for the majority class

Data set	Pruning	#Rules	Avg.sup	Avg.conf	Avg.S	Avg.N
balance-scale	none	306	12.89	1.00	0.08	0.02
	sup	31	30.10	0.99	0.08	0.05
	sup*N	31	34.19	1.00	0.08	0.06
	sup*S	31	30.45	1.00	0.08	0.05
breast-cancer	none	75	4.97	0.96	0.26	0.02
	sup	8	11.75	0.93	0.23	0.05
	sup*N	8	13.38	0.99	0.31	0.07
	sup*S	8	12.50	0.99	0.30	0.06
cmc	none	401	7.30	0.97	0.20	0.01
	sup	40	21.73	0.98	0.20	0.02
	sup*N	40	22.98	0.99	0.22	0.02
	sup*S	40	21.50	0.99	0.22	0.02
haberman	none	60	6.38	0.98	0.25	0.03
	sup	6	15.83	0.99	0.27	0.07
	sup*N	6	15.83	0.99	0.27	0.07
	sup*S	6	15.83	0.99	0.27	0.07
hepatitis	none	52	18.62	1.00	0.24	0.15
	sup	5	59.60	1.00	0.34	0.49
	sup*N	5	65.20	1.00	0.36	0.53
	sup*S	5	59.60	1.00	0.34	0.49
transfusion	none	118	11.72	0.97	0.21	0.02
	sup	12	51.50	0.93	0.18	0.06
	sup*N	12	51.75	0.95	0.20	0.07
	sup*S	12	41.75	0.96	0.21	0.06

[4] More detailed experimental results, including also the coverage option are provided at the page http://www.cs.put.poznan.pl/iszczech/publications/nfmcp-2016.html; As one can check, rule post-pruning with the coverage option also improves the values of considered rule interestingness measures

Additionally, we constructed rule classifiers with the three selection strategies and evaluated their classification performance. The values of G-mean and sensitivity measures are presented in Tables 5 and 6, respectively.

Table 5. G-mean for BRACID with all rules vs. BRACID with pruned rules

Data set	BRACID	sup	sup*S	sup*N
balance-scale	0.57	0.63	0.59	0.60
breast-cancer	0.61	0.61	0.61	0.61
cmc	0.64	0.64	0.64	0.64
haberman	0.60	0.55	0.54	0.54
hepatitis	0.80	0.80	0.75	0.74
transfusion	0.65	0.64	0.63	0.65

Table 6. Sensitivity for BRACID with all rules vs. BRACID with pruned rules

Data set	BRACID	sup	sup*S	sup*N
balance-scale	0.49	0.59	0.52	0.63
breast-cancer	0.61	0.65	0.57	0.61
cmc	0.64	0.64	0.65	0.61
haberman	0.72	0.78	0.71	0.70
hepatitis	0.78	0.84	0.84	0.84
transfusion	0.75	0.78	0.71	0.76

6 Discussion and Final Remarks

First, we will discuss the results of the experiments. Each of post-pruning strategies improves the interestingness measure used in the given strategy. Note that all of them improve average rule supports for both minority and majority classes. For some data sets these improvements are quite high, for instance, for cmc data the average rule supports increase from 6.59 to 18.57 examples in the minority class, and from 7.3 to 21.0 examples in the majority class.

The second strategy (based on $sup(R)$ and $N(R)$) increases the average value of measure N for all data sets in both classes — see e.g. hepatitis data, where the improvements are from 0.23 to 0.39 for the minority class and from 0.17 to 0.49 for the majority class. Analogically, the third strategy improves the average values of the confirmation measure S – however, it is more visible for the minority class than for the majority one, for instance changes from 0.46 to 0.65 in the minority class and from 0.25 to 0.27 in the majority one for haberman data. Note that values of the confirmation measure S are always higher than N.

It is worth observing that the proposed strategies also improve rule evaluation measures other than the ones used in each strategy. In particular, the second strategy usually provides the highest values of the average support – in the majority of data sets it is better than the first strategy that uses the support only. Although it sometimes slightly improves the confirmation measure S, it usually decreases the average confidence of rules. On the other hand, the third strategy offers the highest increases of the rule confidence. It is more visible for the minority class as the confidence of majority rules is already quite high.

What is also interesting, classification performance of such pruned rules does not decrease too much compared to the original set of rules – see results in Tables 5 and 6. In particular, the sensitivity obtained by the first and the second strategy are close to the results of unpruned rules. Although it is not the main criterion of our experimental evaluation, we can say that these results show that the pruned rules can be treated as a representative subset of original rules.

The differences in results obtained by strategies using S and N measures could be explained by analyzing their formulae (see Equations 1 and 2). They exploit the contingency matrix in a different, although symmetric, way. S is more focused on considering a pair of numbers (a and c) decreased by (b and d), while N aggregates a different combination. As BRACID tries to induce rules with a very high confidence (which refers to a pair a and c), it is naturally oriented on obtaining higher values of the S measure. On the other hand, as N confirmation measure exploits complementary information to the one used in BRACID rule induction process, it may better co-operate with the rule support in the pruning strategy and may lead to better descriptive rule evaluation as well as classification results.

To sum up, our experiments have clearly demonstrated that all proposed post-pruning strategies lead to selecting a much smaller number of BRACID induced rules, which are characterized by better values of considered interestingness measures than in case of non-pruned rules.

As future research, we plan to extend the experimental evaluation with more imbalanced data characterized by additional data difficulty factors and potentially with other rule classifiers specialized for class imbalances. Furthermore, we plan to investigate a more local way of calculating the interestingness measures, which will be based on the analysis of neighbor examples to the given rule rather than on all data elements as it is currently done.

Ack. The research was supported by NCN grant DEC-2013/11/B/ST6/00963.

References

1. Christensen, D.: Measuring confirmation. Journal of Philosophy, 96, 437–461 (1999)
2. Fitelson, B.: The plurality of Bayesian measures of confirmation and the problem of measure sensitivity. Philosophy of Science, 66, 362–378 (1999)
3. Freitas, A.: On rule interestingness measures. Knowledge-Based Systems, 12, 309--315 (1999)
4. Furnkranz, J., Gamberger, D., Lavrac, N.: Foundations of Rule Learning, Springer Verlag (2012)

5. Gamberger, D., Lavrac, N. : Expert-guided subgroup discovery: methodology and application, J. Artif. Int. Res., 17(1), 501–527 (2002)
6. Geng, L., Hamilton, H.: Interestingness measures for data mining: A survey. ACM Computing Surveys, 38(3) (2006)
7. Glass, D.: Confirmation Measures of Association Rule Interestingness. Knowledge-Based Systems, 44, 65–77 (2013)
8. Greco, S., Slowinski, R., Szczech, I.: Properties of rule interestingness measures and alternative approaches to normalization of measures. Information Sciences, 216, 1–16 (2012)
9. Greco, S., Slowinski, R., Szczech, I.: Measures of rule interestingness in various perspectives of confirmation. Information Sciences 346-347C, 216–235 (2016)
10. He, H., Yungian, Ma (eds): Imbalanced Learning. Foundations, Algorithms and Applications, IEEE - Wiley (2013)
11. McGarry, K.: A survey of interestingness measures for knowledge discovery. The Knowledge Engineering Review, 20(1), 39–61 (2005)
12. Napierala, K., Stefanowski, J.: BRACID: a comprehensive approach to learning rules from imbalanced data, Journal of Intelligent Information Systems, 39(2), 335–373 (2012)
13. Napierala, K., Stefanowski, J.: Types of Minority Class Examples and Their Influence on Learning Classifiers from Imbalanced Data. Journal of Intelligent Information Systems, 46(3), 563–597 (2016).
14. Napierala, K., Stefanowski, J.: Post-processing of BRACID Rules Induced from Imbalanced Data. Fundamenta Informaticae, 145 - accepted for publication (2016)
15. Nozick, R.: Philosophical Explanations, Clarendon Press, Oxford, UK (1981)