



# The Problem of Finding the Simplest Classifier Ensemble is NP-Hard – A Rough-Set-Inspired Formulation Based on Decision Bireducts

Dominik Ślęzak<sup>1(✉)</sup> and Sebastian Stawicki<sup>2</sup>

<sup>1</sup> Institute of Informatics, University of Warsaw, Warsaw, Poland  
slezak@mimuw.edu.pl

<sup>2</sup> QED Software, Warsaw, Poland

**Abstract.** We investigate decision bireducts which extend the notion of a decision reduct developed in the theory of rough sets. For a decision table  $\mathbb{A} = (U, A \cup \{d\})$ , a decision bireduct is a pair  $(X, B)$ , where  $B \subseteq A$  is a subset of attributes which allows to distinguish between all pairs of objects in  $X \subseteq U$  labeled with different values of decision attribute  $d$ , and where  $B$  and  $X$  cannot be made, respectively, smaller and bigger without losing this property. We refer to our earlier studies on deriving bireducts  $(X, B)$  from decision tables and utilizing them to construct families of rule-based classifiers, where  $X \subseteq U$  is equal to total support of decision rules built using attributes in  $B \subseteq A$ . We introduce the notion of a correct ensemble of decision bireducts  $(X_1, B_1), \dots, (X_m, B_m)$ , where each  $u \in U$  must be validly classified by more than 50% of the corresponding models. We show that the problem of finding a correct ensemble of bireducts with the lowest cardinalities of subsets  $B_i \subseteq A$  is NP-hard.

**Keywords:** Decision bireducts · Classifier ensembles · Rule-based classifiers · Decision model simplification · NP-hardness

## 1 Introduction

There are a number of approaches based on ensembles of classifiers – or decision models, more generally – in the areas of knowledge discovery and data classification [1, 4]. One can name several reasons to use ensembles. First, we can count on stability and robustness of the collective. Moreover, it is expected that each of classifiers – a part of a bigger ensemble – can be simpler than a single, not ensemble-based decision model that would yield a similar level of accuracy. On the other hand, if we look at all ensemble parts as a whole, they often lose something out of their interpretability. It may be hard to set up their cooperation. Finally, a lot of computing power is needed to derive them from the data.

Several optimization tendencies repeat in case of numerous methods for learning classifier ensembles. For instance, it is widely assumed that components of

the ensemble should not misclassify too often the same training cases. Each single classifier is expected to make mistakes (in other words, each single part can be relatively weak with respect to its classification power) but for each training case, a (possibly weighted) majority of models in the ensemble should be correct. To address this aspect while building classifier ensembles, popular machine learning meta-algorithms such as boosting or bagging can be used.

One more aspect – besides tending to simplicity (and to some extent weakness) and complementarity (avoiding repeatable mistakes) in the ensemble – corresponds to diversification of attributes (features) that are used as inputs to learn particular models. For example, in case of rough-set-inspired approaches to knowledge discovery, it refers to computation of diverse decision reducts, i.e., irreducible (the smaller – the better) subsets of attributes that are sufficient to determine decision labels. If we want to put this idea together with diversification of cases which are classified correctly/wrongly by particular models, we can rely on ensembles of so-called decision bireducts [7, 8], whereby irreducibility of a subset of attributes is combined with non-extendability of a subset of objects (cases), for which those attributes let us form valid classification rules.

Rough-set-based approaches are also a good reference while considering formal optimization problems behind construction of decision models. Indeed, starting from fundamental works on NP-hardness of the problem of finding minimal (in terms of the number of attributes) decision reduct, a lot of attention is paid in the rough set literature to develop mathematical and algorithmic methods for operating with the simplest yet sufficiently accurate (and thus the most powerful) classifiers [2, 3]. However, even in the realm of rough sets, there are no studies on formulation of optimization problems related to classifier ensembles. In other words, optimization goals and their complexity characteristics are investigated only at the level of single models, rather than the whole ensembles.

Accordingly, in this paper we propose how to define the optimization problem related to searching for the simplest possible ensembles of decision models that meet specific accuracy constraints. We realize that there are plenty of ways of stating such constraints, with respect to various aspects of an ensemble as a whole or its single components. Similarly, there may be many ways of understanding the simplicity of an ensemble. Nevertheless, we believe that the introduced formulation – based on collections of decision bireducts that include the minimal amounts of attributes (the optimization goal) and in the same time, sufficiently cover all considered objects using the corresponding decision rules (the accuracy constraint) – can be a good starting point for further investigations.

The paper is organized as follows: Sect. 2 recalls decision bireducts  $(X, B)$ , whereby – for the training data set (referred as so-called decision table)  $\mathbb{A} = (U, A \cup \{d\})$  –  $B \subseteq A$  is a subset of attributes which allows to distinguish between all pairs of objects in  $X \subseteq U$  labeled with different values of decision attribute  $d$ , and where  $B$  and  $X$  cannot be reduced and extended, respectively. Section 3 introduces *correct* ensembles of bireducts  $(X_1, B_1), \dots, (X_m, B_m)$ ,  $m > 0$ , whereby there is inequality  $|\{i = 1, \dots, m : u \in X_i\}| > m/2$  for each  $u \in U$ . Section 4 presents our idea of expressing simplicity (which is the

optimization goal) of correct ensembles in terms of cardinalities of subsets  $B_1, \dots, B_m$ . Section 5 shows our main theoretical result, i.e., NP-hardness of finding the simplest correct ensembles of decision bireducts. Section 6 concludes the paper.

## 2 Decision Bireducts

We use tabular data representation by means of decision tables [3] – pairs  $\mathbb{A} = (U, A \cup \{d\})$  of finite sets  $U$  and  $A \cup \{d\}$ , where  $U$  is the universe of objects and  $A \cup \{d\}$  is the set consisting of attributes such that every  $a \in A \cup \{d\}$  is associated with function  $a : U \rightarrow V_a$ , where  $V_a$  is called the value set of  $a$ . The distinguished attribute  $d \notin A$  is called the decision. Elements of  $A$  are called conditional attributes. Values  $v_d \in V_d$  correspond to decision classes.

**Definition 1.** Let  $\mathbb{A} = (U, A \cup \{d\})$  and  $B \subseteq A$ ,  $X \subseteq U$  be given. We say that  $B$  determines  $d$  within  $X$ , further denoted as  $B \Rightarrow_X d$ , if and only if  $B$  discerns all pairs  $u_i, u_j \in X$  such that  $d(u_i) \neq d(u_j)$ . Further, we say that the pair  $(X, B)$  is a decision bireduct, if and only if the following holds:

1. There is  $B \Rightarrow_X d$ ,
2. There is no proper subset  $B' \subsetneq B$  such that  $B' \Rightarrow_X d$ ,
3. There is no proper superset  $X' \supsetneq X$  such that  $B \Rightarrow_{X'} d$ .

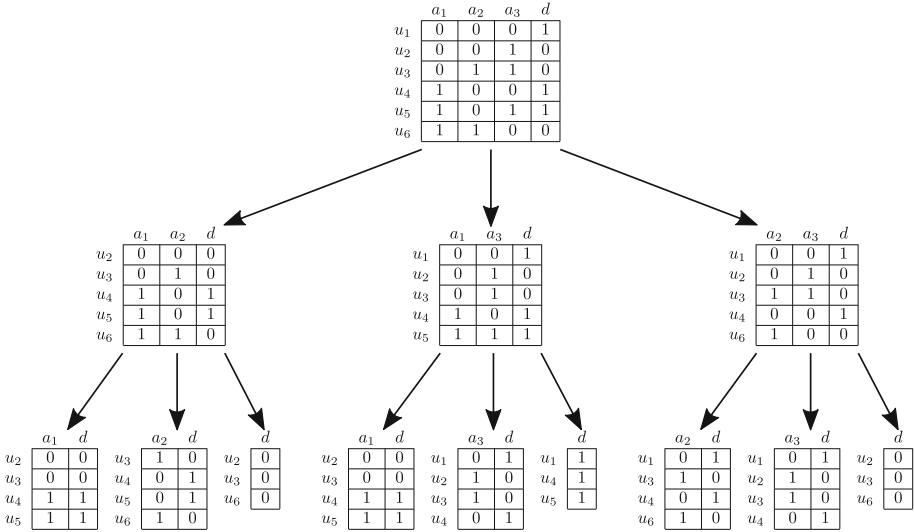
We will say that objects in  $X$  are covered by  $B \Rightarrow_X d$ .

Every decision bireduct  $(X, B)$  may be understood as a pair consisting of an irreducible subset of attributes that can be evaluated by means of an non-extendable subset of objects for which it provides good classification. It was shown in [8] that  $X$  is actually the set-theoretic sum of objects supporting deterministic rules using the values of attributes in  $B$  to describe the values of  $d$ .

Figure 1 displays some bireducts derived for decision table  $\mathbb{A}$  with objects  $U = \{u_1, \dots, u_6\}$  and conditional attributes  $A = \{a_1, a_2, a_3\}$ . For instance, consider the pair  $B = \{a_1, a_2\}$  and  $X = \{u_2, \dots, u_6\}$ . It corresponds to rules “if  $a_1 = 0$  and  $a_2 = 0$  then  $d = 0$ ” (supported by  $u_2$ ), “if  $a_1 = 0$  and  $a_2 = 1$  then  $d = 0$ ” (supported by  $u_3$ ), “if  $a_1 = 1$  and  $a_2 = 0$  then  $d = 1$ ” (supported by  $u_4$  and  $u_5$ ) and “if  $a_1 = 1$  and  $a_2 = 1$  then  $d = 0$ ” (supported by  $u_6$ ). Neither  $a_1$  nor  $a_2$  would be sufficient by itself to cover  $X$  with shorter rules. Moreover,  $u_1$  cannot be added to  $X$  because it is inconsistent with the first rule.

The first algorithms aimed at deriving decision bireducts from the data were proposed in [7]. They were based on random generation of mixed orderings of attributes and objects. Such orderings were utilized to encode sequences of attempts to remove attributes from  $B$  (starting with  $B = A$ ) and add objects to  $X$  (starting with  $X = \emptyset$ ) in order to obtain pairs  $(X, B)$  such that  $B \Rightarrow_X d$ , with possibly minimal  $B$  and maximal  $X$ . By using an appropriate process of generation of families of diverse orderings, one could derive collections of bireducts with quite different subsets of attributes and objects involved.

Probably the most efficient known technique for deriving decision bireducts consists of four phases: 1) Choose randomly a subset  $C \subseteq A$ ; 2) For each block of objects (so-called indiscernibility class) with the same values on  $C$ , add randomly one of its elements to  $Y \subseteq U$ ; 3) For table  $\mathbb{A} = (Y, C \cup \{d\})$ , run one of classical methods for finding standard decision reducts [2]; 4) For a decision reduct  $B \subseteq C$ , construct superset  $X \supseteq Y$  by adding to  $Y$  all objects that are consistent with decision rules generated from table  $\mathbb{A} = (Y, B \cup \{d\})$ .



**Fig. 1.** Examples of decision bireducts for table  $\mathbb{A} = (\{u_1, \dots, u_6\}, \{a_1, a_2, a_3\} \cup \{d\})$ . Bireducts in the middle layer form a correct ensemble (each object is validly classified by at least two bireducts out of three). For each “middle” bireduct treated as a new decision table, its corresponding correct ensemble is provided in the lowest layer.

### 3 Bireduct Ensembles

The notion of a decision bireduct allows us to operate with subsets of conditional attributes treated as classification descriptions, and with the associated subsets of objects for which those descriptions are valid. This gives us an elegant way to investigate complementarity of bireducts interpreted as classifiers in the ensemble. For instance, the following formulation expresses the idea of majority voting between ensemble components which – if properly tuned on the training data – gives us a chance of efficient performance over new cases too.

**Definition 2.** Let  $\mathbb{A} = (U, A \cup \{d\})$  and the ensemble of decision bireducts  $\mathcal{B} = (X_1, B_1), \dots, (X_m, B_m)$  be given. We say that  $\mathcal{B}$  is correct, if and only if there is inequality  $|\{i = 1, \dots, m : u \in X_i\}| > m/2$  for each  $u \in U$ .

The above inequality means that more than 50% of decision rules triggered for  $u \in U$  point at the valid decision  $d(u)$ . Figure 1 illustrates a kind of hierarchy of correct bireduct ensembles for  $m = 3$ . Alternatively, one can work with a “flat” collection of decision bireducts that are supposed to vote correctly on each of objects, even if some single bireducts are wrong for some single cases.

Figure 1 implicitly suggests a top-down way of constructing correct ensembles, whereby each of  $m$  bireducts is derived in the same time, with an option of further decompositions on even smaller pieces. Such algorithms have been already considered in [5] for another type of (bi)reducts, i.e., so-called generalized decision reducts. On the other hand, one can proceed with the aforementioned ordering-based methods [7], whereby – somewhat reflecting the mechanisms of bagging and boosting – each next ordering may take into account which objects were covered least frequently by decision bireducts derived up to now.

## 4 Ensemble Simplicity

The rough literature provides a great number of theoretical works on computational complexity of optimization problems focused on deriving the simplest possible decision models from the data [2]. Let us refer to a recent comparative study reflecting both decision bireducts and so-called approximate reducts with this respect [8]. By “the simplest” one can mean (bi)reducts involving the minimal amounts of attributes, generating minimal amounts of decision rules, having the minimal information entropy, etc. However, all those formulations refer to single (bi)reducts which correspond to single classifiers.

In other words, as it was emphasized in [3], simplicity is a crucial aspect of decision models, in relation to paradigms such as Occam’s Razor or the Minimum Description Length Principle. However, there is no clear guidance how to understand simplicity of ensembles. Thus, if we want to define optimization problems for ensembles, we need to know how to aggregate “complexities” of particular ensemble components (e.g.: the number of attributes in a single decision bireduct, the number of leaves in a single decision tree, etc.).

Intuitively, in case of ensembles of decision bireducts, the corresponding optimization problem should be stated by means of finding the smallest subsets  $B_1, \dots, B_m$  that satisfy – together with their counterparts  $X_1, \dots, X_m$  – the constraints of Definition 2. The question remains what we should mean by “the smallest” in case of a collection of subsets. In [6], for the analogous task related to the already-mentioned generalized decision reducts, it was proposed to look at it from the perspective of the maximum cardinality out of all involved subsets. For the purpose of bireducts it can be phrased as follows:

**Definition 3.** *Let decision table  $\mathbb{A} = (U, A \cup \{d\})$  and two correct ensembles of decision bireducts  $\mathcal{B} = \{(X_1, B_1), \dots, (X_m, B_m)\}$  and  $\mathcal{C} = \{(Y_1, C_1), \dots, (Y_n, C_n)\}$ ,  $m, n \geq 0$ , be given. We say that  $\mathcal{B}$  is simpler than  $\mathcal{C}$ , denoted as  $\mathcal{B} \prec \mathcal{C}$ , if and only if the following procedure yields it:*

1. Sort sequences of cardinalities of attribute subsets in a descending order.
2. Add a sentinel item with the value  $-1$  at the end of each of sequences.
3. Find the first position for which the sorted sequences differ from each other.
4. If the value in the above-found position is lower for  $\mathcal{B}$  than for  $\mathcal{C}$ , then  $\mathcal{B} \prec \mathcal{C}$ .

The above procedure – illustrated additionally by Fig. 2 – induces a linear order over ensembles of bireducts for a given  $\mathbb{A}$ . We therefore propose to search through a space of all correct ensembles  $\mathcal{B} = \{(X_1, B_1), \dots, (X_m, B_m)\}$ , paying special attention to cardinalities of their largest components along a kind of cardinality-based lexicographic order. This is because the largest subsets of attributes correspond to the largest collections of the longest rules, i.e., they affect complexity of the model more significantly than other subsets.

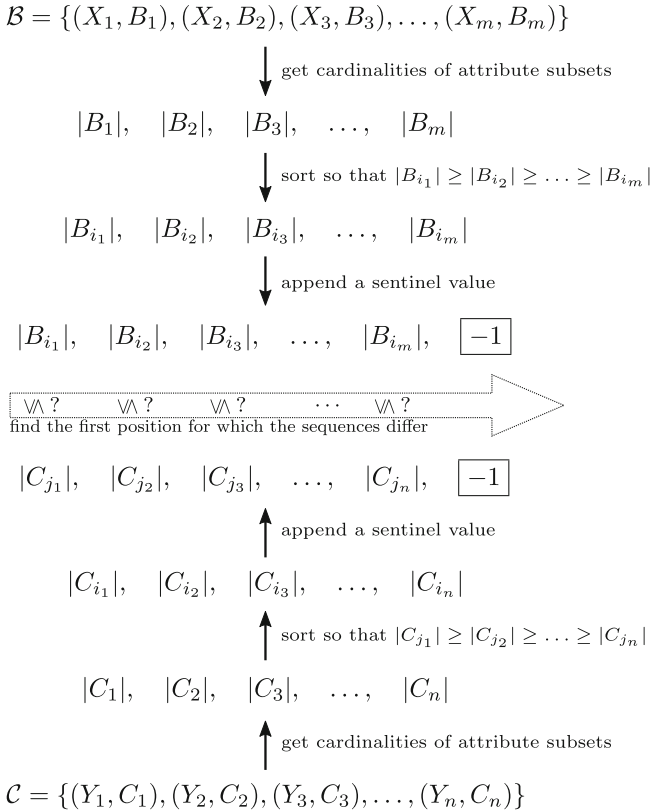


Fig. 2. Illustration of the procedure in Definition 3.

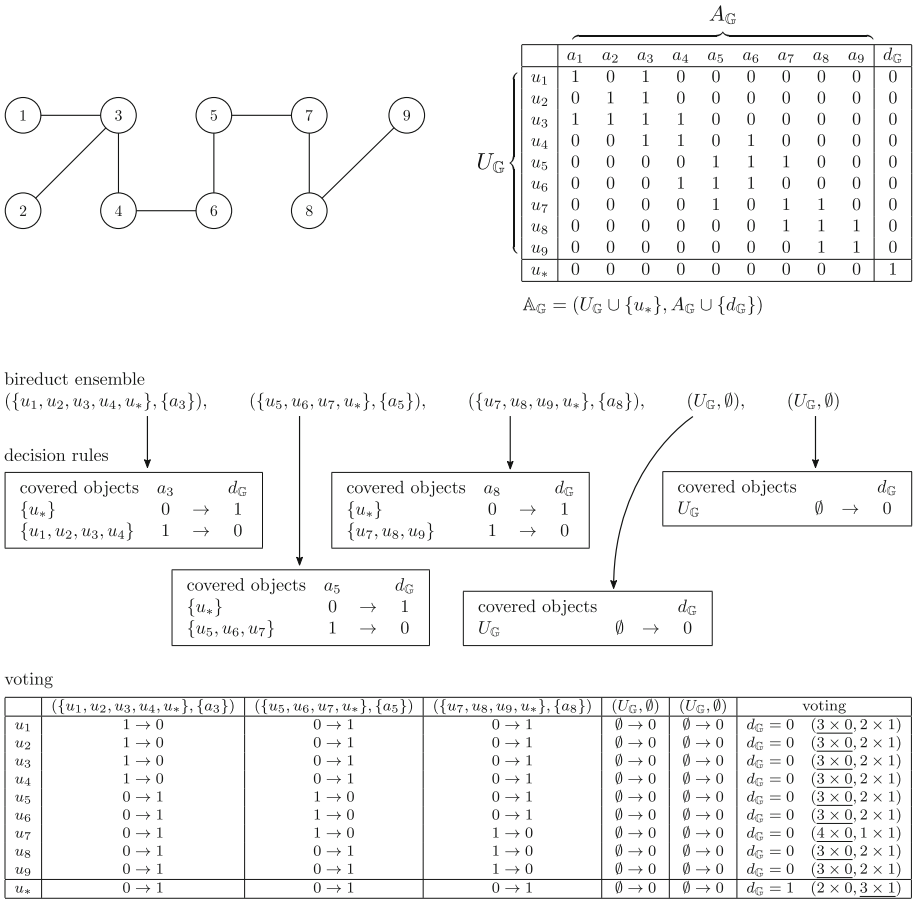
## 5 Main Result

Let us formalize the optimization goal that we drafted in the previous section:

**Definition 4.** *By the Simplest Correct Decision Bireduct Ensemble Problem (SCDBEP) we mean the task of finding – for each input decision table  $\mathbb{A}$  – the correct ensemble of decision bireducts  $\mathcal{B}$  such that there is no other correct ensemble for  $\mathbb{A}$  that would be simpler than  $\mathcal{B}$  according to Definition 3.*

**Theorem 1.** *SCDBEP is NP-hard.*

Before we present the proof, let us refer to Fig. 3. The proof is based on polynomial reduction of the problem of finding the smallest dominating sets in undirected graphs to SCDBEP. It requires encoding of each input graph  $\mathbb{G}$  to its corresponding decision table  $\mathbb{A}_{\mathbb{G}}$ . This encoding is analogous to those that were utilized for other (bi)reduct-related optimization problems [2, 8].



**Fig. 3.** Illustration for the proof of Theorem 1.

Figure 3 can also serve as one more illustration of creation of correct ensembles of decision bireducts. It displays how to interpret those bireducts as

rule-based classifiers. In particular, as it could be already noticed earlier in Fig. 1, some bireducts can correspond to empty sets of attributes. We can interpret them as “dummy” classifiers which point always at the same decision class. They may help to tune the majority voting mechanism in the ensemble.

*Proof.* As already stated, we intend to show NP-hardness of SCDBEP by polynomial reduction of the minimum dominating set problem. Let us consider an undirected graph  $G = (V, E)$  and create decision table  $A_G = (U_G \cup \{u_*\}, A_G \cup \{d_G\})$ , where  $a_v \in A_G$  corresponding to  $v \in V$  takes 1 on  $u_{v'} \in U_G$  corresponding to  $v' \in V$ , i.e.,  $a_v(u_{v'}) = 1$ , if and only if  $v = v'$  or  $(v, v') \in E$ , and where  $a_v(u_*) = 0$ . Let us also put  $d_G(u_{v'}) = 0$  and  $d_G(u_*) = 1$  (see Fig. 3).

Clearly, any  $B \subseteq V$  is a dominating set in  $G$ , if and only if it corresponds to a decision bireduct  $(U_G, B_G)$ . It is obvious that a single-element bireduct ensemble  $\{(U_G, B_G)\}$  is correct according to Definition 2. However, we can always construct a simpler (or equally simple if  $B_G$  is a singleton) correct ensemble.

Assuming that  $B_G = \{a_{v_1}, a_{v_2}, \dots, a_{v_n}\}$ , let us define new subsets of attributes as  $B_{G,1} = \{a_{v_1}\}, \dots, B_{G,n} = \{a_{v_n}\}, B_{G,n+1} = \emptyset, \dots, B_{G,2n-1} = \emptyset$  and new subsets of objects as  $X_{G,1} = \{u_*\} \cup \{u \in U_G | a_{v_1}(u) = 1\}, \dots, X_{G,n} = \{u_*\} \cup \{u \in U_G | a_{v_n}(u) = 1\}, X_{G,n+1} = U_G, \dots, X_{G,2n-1} = U_G$ . Then, the proposed simpler ensemble would be equal to  $\{(X_{G,1}, B_{G,1}), \dots, (X_{G,2n-1}, B_{G,2n-1})\}$  and it would be still correct according to Definition 2 (see Fig. 3 again).

The consequence of the above is that the simplest correct ensemble of decision bireducts corresponds to the smallest dominating set in the graph  $G$ .  $\square$

## 6 Conclusions

We investigated ensembles of so-called decision bireducts, which can be interpreted as rule-based classifiers. We introduced the notion of a correct ensemble, which means that every object (training case) must be validly recognized using the corresponding rules by more than majority of classifiers. We discussed how to specify a kind of simplicity criterion for such ensembles and we formulated an example of optimization problem related to extracting possibly simplest correct ensembles of decision bireducts from the input data. The main mathematical result of our paper is the NP-hardness of the considered problem.

In future, given such a sound theoretical framework, more attention should be paid to further extensions of our previous algorithmic approaches [7, 8] to deriving and applying decision bireducts for the real-life data. Moreover, some alternative formulations of optimization problems should be discussed as well, possibly referring to ensembles of other types of classifiers.

## References

1. Dietterich, T.G.: An experimental comparison of three methods for constructing ensembles of decision trees: bagging, boosting, and randomization. *Mach. Learn.* **40**(2), 139–157 (2000). <https://doi.org/10.1023/A:1007607513941>



2. Pawlak, Z., Skowron, A.: Rough sets and Boolean reasoning. *Inf. Sci.* **177**(1), 41–73 (2007)
3. Pawlak, Z., Skowron, A.: Rudiments of rough sets. *Inf. Sci.* **177**(1), 3–27 (2007)
4. Polikar, R., DePasquale, J., Mohammed, H.S., Brown, G., Kuncheva, L.I.: Learn++.MF: a random subspace approach for the missing feature problem. *Pattern Recogn.* **43**(11), 3817–3832 (2010)
5. Ślęzak, D.: Decomposition and synthesis of decision tables with respect to generalized decision functions. In: Pal, S.K., Skowron, A. (eds.) *Rough Fuzzy Hybridization - A New Trend in Decision Making*, pp. 110–135. Springer, Singapore (1999)
6. Ślęzak, D.: On generalized decision functions: reducts, networks and ensembles. In: Yao, Y., Hu, Q., Yu, H., Grzymala-Busse, J.W. (eds.) *RSFDGrC 2015. LNCS (LNAI)*, vol. 9437, pp. 13–23. Springer, Cham (2015). [https://doi.org/10.1007/978-3-319-25783-9\\_2](https://doi.org/10.1007/978-3-319-25783-9_2)
7. Ślęzak, D., Janusz, A.: Ensembles of bireducts: towards robust classification and simple representation. In: Kim, T., Adeli, H., Slezak, D., Sandnes, F.E., Song, X., Chung, K., Arnett, K.P. (eds.) *FGIT 2011. LNCS*, vol. 7105, pp. 64–77. Springer, Heidelberg (2011). [https://doi.org/10.1007/978-3-642-27142-7\\_9](https://doi.org/10.1007/978-3-642-27142-7_9)
8. Stawicki, S., Ślęzak, D., Janusz, A., Widz, S.: Decision bireducts and decision reducts - a comparison. *Int. J. Approximate Reasoning* **84**, 75–109 (2017)