Recent Advances in Decision Bireducts: Complexity, Heuristics and Streams*

Sebastian Stawicki[1] and Dominik Ślęzak[1,2]

[1] Institute of Mathematics, University of Warsaw
ul. Banacha 2, 02-097 Warsaw, Poland
[2] Infobright Inc.
ul. Krzywickiego 34, lok. 219, 02-078 Warsaw, Poland

Abstract. We continue our research on decision bireducts. For a decision system $\mathbb{A} = (U, A \cup \{d\})$, a decision bireduct is a pair (B, X), where $B \subseteq A$ is a subset of attributes discerning all pairs of objects in $X \subseteq U$ with different values on the decision attribute d, and where B and X cannot be, respectively, reduced and extended. We report some new results related to NP-hardness of extraction of optimal decision bireducts, heuristics aimed at searching for sub-optimal decision bireducts, and applications of decision bireducts to stream data mining.

Keywords: Bireducts, NP-hardness, Heuristic Search, Data Streams.

1 Introduction

Decision reducts have been found a number of applications in feature selection and knowledge representation [1]. Notions analogous to decision reducts occur in many areas of science, such as Markov boundaries in probabilistic modeling [2] or signatures in bioinformatics [3]. As one of extensions, approximate decision reducts are studied in order to search for irreducible subsets of attributes that *almost* determine decisions in real-world, noisy data sets [4].

Bireducts were proposed as a new extension of decision reducts in [5] and further developed in [6]. Their interpretation seems to be simpler than in the case of most of types of approximate decision reducts known from the literature. The emphasis here is on both a subset of attributes, which describes decisions, and a subset of objects, for which such a description is valid.

This paper continues our research on bireducts, both with respect to their comparison to classical and approximate decision reducts, and their applications in new areas. In Section 2, we recall basics of decision bireducts. In Section 3, we prove NP-hardness of one of possible optimization problems related to extraction of decision reducts from data. In Section 3, we show some new interpretations of decision bireducts, which are useful for their heuristic search. In Section 4, we outline how to apply decision bireducts in data stream analysis. In Section 5, we discuss some of future perspectives and conclude the paper.

* This research was partly supported by the Polish National Science Centre (NCN) grants 2011/01/B/ST6/03867 and 2012/05/B/ST6/03215.

P. Lingras et al. (Eds.): RSKT 2013, LNAI 8171, pp. 200–212, 2013.

Table 1. System $\mathbb{A} = (U, A \cup \{d\})$ with 14 objects in U, four attributes in A, and $d =$ Sport?

	Outlook	Temp.	Humid.	Wind	Sport?
1	sunny	hot	high	weak	no
2	sunny	hot	high	strong	no
3	overcast	hot	high	weak	yes
4	rain	mild	high	weak	yes
5	rain	cool	normal	weak	yes
6	rain	cool	normal	strong	no
7	overcast	cool	normal	strong	yes
8	sunny	mild	high	weak	no
9	sunny	cool	normal	weak	yes
10	rain	mild	normal	weak	yes
11	sunny	mild	normal	strong	yes
12	overcast	mild	high	strong	yes
13	overcast	hot	normal	weak	yes
14	rain	mild	high	strong	no

Table 2. Several examples of bireducts (B, X) for $\mathbb{A}$ in Table 1

(B, X)
({O},{1..5,7..8,10,12..13}) ({O},{1..3 6..8 12..14}) ({O},{3 6..7 9 11..14})
({O T},{1..4 6..10 12..13})
({O H},{1..3 6..9 11..14})
({O T W},{1..14})
({O H W},{1..14})
({O W},{2..7 9..10 12..14})
({T},{3..4 6 10..13})
({T H},{1..2 6 8 10..11 13..14})
({T W},{1..2 4..5 7 9..10 14}) ({T W},{2..6 9..13})
({H W},{1 5..6 8..10 12..13})
({W},{2..6 9..10 13..14})

2 Basics of Decision Bireducts

First formulation of decision bireducts occurred in [5], where their Boolean characteristics and simple permutation-based search algorithms were proposed in analogy to classical reducts [7]. It was also discussed in what sense ensembles of decision bireducts are better than ensembles of approximate reducts, which – although quite useful in practice [8] – do not allow for explicit analysis whether particular reducts repeat mistakes on the same cases.

We use a standard representation of tabular data in form of decision systems [9]. A decision system is a tuple $\mathbb{A} = (U, A \cup \{d\})$, where U is a set of objects, A is a set of attributes and $d \notin A$ is a decision attribute. For simplicity, we refer to the elements of U using their ordinal numbers $i = 1, ..., |U|$, where $|U|$ denotes the cardinality of U. We treat all attributes $a \in A \cup \{d\}$ as functions $a : U \to V_a$, V_a denoting a's domain. The values $v_d \in V_d$ correspond to decision classes that we want to describe using the values of attributes in A.

Definition 1. *[9] We say that $B \subseteq A$ is a decision reduct for decision system $\mathbb{A} = (U, A \cup \{d\})$, iff it is an irreducible subset of attributes such that each pair $i, j \in U$ satisfying inequality $d(i) \neq d(j)$ is discerned by B.*

As an example, for $\mathbb{A}$ in Table 1, there are two reducts: {Outlook, Temp., Wind} and {Outlook, Humid., Wind} (or {O, T, W} and {O, H, W} for short).

Definition 2. *[5] Let $\mathbb{A} = (U, A \cup \{d\})$ be a decision system. A pair (B, X), where $B \subseteq A$ and $X \subseteq U$, is called a decision bireduct, iff the following holds:*

- *B discerns all pairs $i, j \in X$ where $d(i) \neq d(j)$ (further denoted as $B \Rrightarrow_X d$);*
- *There is no $C \subsetneq B$ such that $C \Rrightarrow_X d$;*
- *There is no $Y \supsetneq X$ such that $B \Rrightarrow_Y d$.*

For the decision system $\mathbb{A}$ given in Table 1, some examples of decision bireducts are presented in Table 2.

A decision bireduct (B, X) can be regarded as the basis for an inexact functional dependency linking the subset of attributes B with the decision d in a degree X, denoted as $B \Rrightarrow_X d$ in Definition 2. Furthermore, the objects in $U \setminus X$ can be treated as outliers of $B \Rrightarrow_X d$.

Further in this paper, we focus on bireducts and their corresponding inexact dependencies formulated in terms of standard discernibility, where $B \subseteq A$ discerns objects $i, j \in U$ iff there is $a \in B$ such that $a(i) \neq a(j)$. However, as pointed out in Section 6, one can also consider some generalizations, such as e.g. bireducts based on fuzzy discernibility [10].

3 Decision Bireduct Optimization

There are a number of NP-hardness results related to extracting optimal decision reducts and approximate reducts from data [11]. In the case of decision bireducts, one may think about quite different optimization criteria with respect to a balance between the number of involved attributes and objects. The following form of a constraint for decision bireducts is somewhat analogous to those studied for frequent itemsets and patterns [12]. However, let us emphasize that this is just one of many ways of interpreting optimal decision bireducts.

Definition 3. *Let $\varepsilon \in [0, 1)$ be given. We say that a pair (B, X), $B \subseteq A$ and $X \subseteq U$, is a ε-bireduct, if it is a bireduct and the following holds: $|X| \geq (1-\varepsilon)|U|$.*

Definition 4. *Let $\varepsilon \in [0, 1)$ be given. By the Minimal ε-Decision Bireduct Problem (MεDBP) we mean a task of finding for each given decision system $\mathbb{A} = (U, A \cup \{d\})$ a ε-bireduct (B, X) with the lowest cardinality of B.*

In order to prepare the ground for the major result in this section, let us recall the following correspondence between decision bireducts and one of specific types of approximate decision reducts.

Definition 5. *[13] Let $\varepsilon \in (0, 1]$ and a decision system $\mathbb{A} = (U, A \cup \{d\})$ be given. For each $B \subseteq A$, consider the quantity $M_{\mathbb{A}}(B) =$*

$$= \frac{1}{|U|} \left| \left\{ u \in U : d(u) = \underset{v_d \in V_d}{\operatorname{argmax}} \left| \{ u' \in U : \forall_{a \in B}\, a(u') = a(u) \wedge d(u') = v_d \} \right| \right\} \right| \tag{1}$$

We say that $B \subseteq A$ is an (M, ε)-approximate reduct, iff

$$M_{\mathbb{A}}(B) \geq 1 - \varepsilon \tag{2}$$

and there is no proper subset of B, which would hold an analogous inequality.

Original formulation of the above definition in [13] was a bit different, with constraint $M_{\mathbb{A}}(B) \geq (1 - \varepsilon) M_{\mathbb{A}}(A)$ instead of $M_{\mathbb{A}}(B) \geq 1 - \varepsilon$. Thus, formally, we should refer to the above as to a *modified* (M, ε)-approximate reduct.

A way of defining $M_{\mathbb{A}}(B)$ is different as well, although mathematically equivalent to that in [13]. We rewrite it in the above form in order to emphasize that it is actually the ratio of objects in U that would be correctly classified by if-then decision rules learned from $\mathbb{A} = (U, A \cup \{d\})$ with the *attribute* = *value* conditions over B and *decision* = *value* consequences specified by identifying decision values assuring the highest confidence for each of rules.

For a consistent decision system, i.e. $\mathbb{A} = (U, A \cup \{d\})$, where A enables to fully discern all pairs of objects from different decision classes, there is $M_{\mathbb{A}}(A) = 1$. In such a case, original and modified conditions for an (M, ε)-approximate reduct are equivalent. Also, but only in consistent decision tables, (M, ε)-approximate reducts are equivalent to classical decision reducts for $\varepsilon = 0$.

In [6], a correspondence between decision bireducts and modified (M, ε)-approximate reducts was noticed. Consider a family of all subsets $X \subseteq U$ with which a given subset $B \subseteq A$ has a chance to form a bireduct:

$$X_B = \{X \subseteq U : \forall_{i,j \in X}\, d(i) \neq d(j) \Rightarrow \exists_{a \in B}\, a(i) \neq a(j)\} \tag{3}$$

Then the following equality holds:

$$M_{\mathbb{A}}(B) = \max_{X \in X_B} |X|/|U| \tag{4}$$

As a result, $B \subseteq A$ may be a modified (M, ε)-approximate reduct only if there is $X \subseteq U$ such that the pair (B, X) is an ε-bireduct. Given the computational complexity results reported in [13] for approximate decision reducts, we are now ready to formulate an analogous result for ε-bireducts:

Theorem 1. *Let $\varepsilon \in [0, 1)$ be given. $M\varepsilon DBP$ is NP-hard.*

Proof. In [13], it was shown that for each $\varepsilon \in [0, 1)$ treated as a constant, the problem of finding an (M, ε)-approximate reduct in an input decision system with minimum number of attributes is NP-hard. (Actually, in [13] it was presented for a far wider class of approximate decision reducts.)

The proof was based on polynomial reduction of the Minimal $\boldsymbol{\alpha}$-Dominating Set Problem ($M\alpha DSP$), aiming at finding minimal subsets of vertices that dominate at least $\alpha \times 100\%$ of all vertices in an input undirected graph. (NP-hardness of this problem was studied in [13] and later in [2].) For each $\varepsilon \in [0, 1)$, the formula for $\alpha(\varepsilon) \in (0, 1]$ can be constructed in such a way that for each graph $G = (V, E)$ being an input to $M\alpha(\varepsilon)DSP$ we can polynomially (with respect to the cardinality of V) construct a decision system with its minimal (M, ε)-approximate reducts equivalent to the $\alpha(\varepsilon)$-dominating sets in G.

Decision systems encoding graphs in the above reduction were consistent. Thus, following our earlier observation on equivalence of $M_{\mathbb{A}}(B) \geq (1 - \varepsilon)M_{\mathbb{A}}(A)$ and $M_{\mathbb{A}}(B) \geq 1 - \varepsilon$ in consistent decision systems, we can prove in the same way that finding of modified (M, ε)-approximate reducts is NP-hard too. As a result, by showing that the case of modified (M, ε)-approximate reducts can be polynomially reduced to $M\varepsilon DBP$ we will be able to finish the proof.

Such reduction is simple, as minimal modified (M, ε)-approximate reducts correspond to decision bireducts solving $M\varepsilon DBP$. Assume that a pair (B, X) is an ε-bireduct with the lowest cardinality of B for a given $\mathbb{A} = (U, A \cup \{d\})$. Then B needs to be a minimal (M, ε)-approximate reduct for $\mathbb{A}$. This is because, first of all, thanks to (4) we have that $M_{\mathbb{A}}(B) \geq |X|/|U| \geq 1 - \varepsilon$. Secondly, assume that there is a subset $B' \subseteq A$ such that $M_{\mathbb{A}}(B') \geq 1 - \varepsilon$ and $|B'| < |B|$. Then, however, there would exist at least one ε-bireduct (B', X') for some $X' \subseteq U$, so (B, X) would not be a solution of $M\varepsilon DBP$. □

4 Heuristic Search for Bireducts

There are a number of possible algorithmic approaches to searching for decision bireducts. One can, e.g., extend techniques introduced earlier for decision reducts, like it was done for permutation-based algorithms in [5], where instead of orderings on attributes the orderings on mixed codes of attributes and objects were considered. One can also translate some algorithms aiming at finding approximate decision reducts onto the case of decision bireducts, basing on connections between both those notions outlined in [6]. Finally, specifically for the problem of searching for minimal ε-bireducts, one can adapt some mechanisms known from other areas, such as association rules with a constraint for minimum support [14], basing on representations developed for decision reducts [15].

Let us recall the above-mentioned algorithm proposed in [5], which is an extension of one of standard approaches to searching for decision reducts [7].

Proposition 1. *[5] Let $\mathbb{A} = (U, A \cup \{d\})$ be given. Enumerate attributes and objects as $A = \{a_1, ..., a_n\}$, $n = |A|$, and $U = \{1, ..., m\}$, $m = |U|$, respectively. Put $B = A$ and $X = \emptyset$. Let permutation $\sigma : \{1, ..., n+m\} \to \{1, ..., n+m\}$ be given. Consider the following procedure for each consecutive $i = 1, ..., n+m$:*

1. *If $\sigma(i) \leq n$, then attempt to remove attribute $a_{\sigma(i)}$ from B subject to the constraint $B \setminus \{a_{\sigma(i)}\} \Rrightarrow_X d$;*
2. *Else, attempt to add $\sigma(i) - n$ to X subject to the constraint $B \Rrightarrow_{X \cup \{\sigma(i)-n\}} d$.*

For each σ, the final outcome (B, X) is a decision bireduct. Moreover, for each bireduct (B, X) there exists input σ for which the above steps lead to (B, X).

The above method follows an idea of mixing the processes of reducing attributes and adding objects during the construction of bireducts. If we consider a special case of permutations $\sigma : \{1, ..., n+m\} \to \{1, ..., n+m\}$ where all objects are added to X prior to starting removing attributes from B, we will obtain the permutation-based characteristics of standard decision reducts. In a general case, the approximation threshold $\varepsilon \in [0, 1)$ introduced in Definition 3 is not defined explicitly but it is somehow expressed in a way permutations are generated. We can define a parameter that controls probability of selecting an attribute in first place rather than an object during the random generation of σ. When σ contains relatively more attributes at its beginning, a bireduct having smaller number of attributes but also higher number of outliers is likely to be obtained.

In the remainder of this section, we present two examples of algorithmic constructions enabling to harness various attribute reduction heuristics directly to

Table 3. $\mathbb{A}^* = (U, A \cup A^* \cup \{d\})$ corresponding to $\mathbb{A} = (U, A \cup \{d\})$ in Table 1

	Outlook	Temp.	Humid.	Wind	a_1^*	a_2^*	a_3^*	a_4^*	a_5^*	a_6^*	a_7^*	a_8^*	a_9^*	a_{10}^*	a_{11}^*	a_{12}^*	a_{13}^*	a_{14}^*	Sport?
1	sunny	hot	high	weak	**1**	0	0	0	0	0	0	0	0	0	0	0	0	0	no
2	sunny	hot	high	strong	0	**1**	0	0	0	0	0	0	0	0	0	0	0	0	no
3	overcast	hot	high	weak	0	0	**1**	0	0	0	0	0	0	0	0	0	0	0	yes
4	rain	mild	high	weak	0	0	0	**1**	0	0	0	0	0	0	0	0	0	0	yes
5	rain	cool	normal	weak	0	0	0	0	**1**	0	0	0	0	0	0	0	0	0	yes
6	rain	cool	normal	strong	0	0	0	0	0	**1**	0	0	0	0	0	0	0	0	no
7	overcast	cool	normal	strong	0	0	0	0	0	0	**1**	0	0	0	0	0	0	0	yes
8	sunny	mild	high	weak	0	0	0	0	0	0	0	**1**	0	0	0	0	0	0	no
9	sunny	cool	normal	weak	0	0	0	0	0	0	0	0	**1**	0	0	0	0	0	yes
10	rain	mild	normal	weak	0	0	0	0	0	0	0	0	0	**1**	0	0	0	0	yes
11	sunny	mild	normal	strong	0	0	0	0	0	0	0	0	0	0	**1**	0	0	0	yes
12	overcast	mild	high	strong	0	0	0	0	0	0	0	0	0	0	0	**1**	0	0	yes
13	overcast	hot	normal	weak	0	0	0	0	0	0	0	0	0	0	0	0	**1**	0	yes
14	rain	mild	high	strong	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	no

the task of searching for decision bireducts, after reformulation of the input data. The first of considered methods refers to the following representation:

Proposition 2. *[5] Let $\mathbb{A} = (U, A \cup \{d\})$ be a decision system. Consider the following Boolean formula with variables $\overline{i}$, $i = 1, ..., |U|$, and $\overline{a}$, $a \in A$:*

$$\tau_{\mathbb{A}}^{bi} = \bigwedge_{i,j:\, d(i) \neq d(j)} \left(\overline{i} \vee \overline{j} \vee \bigvee_{a:\, a(i) \neq a(j)} \overline{a} \right). \tag{5}$$

An arbitrary pair (B, X), $B \subseteq A$, $X \subseteq U$, is a decision bireduct, if and only if the Boolean formula $\bigwedge_{a \in B} \overline{a} \wedge \bigwedge_{i \notin X} \overline{i}$ is the prime implicant for $\tau_{\mathbb{A}}^{bi}$.

The above result shows a way to utilize techniques known from Boolean reasoning to search for decision bireducts as prime implicants [16]. It also illustrates that attributes and objects are to some extent equally important while constructing bireducts, analogously to some other approaches to deriving knowledge from data [17]. This intuition has led us to the following observation:

Proposition 3. *Let $\mathbb{A} = (U, A \cup \{d\})$ be a decision system. Consider a new system $\mathbb{A}^* = (U, A \cup A^* \cup \{d\})$, where the number of objects in U as well as their values for attributes from the original $\mathbb{A}$ remain unchanged, and where new attributes in $A^* = \{a_1^*, ..., a_m^*\}$, $m = |U|$, are defined as $a_j^*(i) = 1$ if $i = j$, and 0 otherwise. Then, the pair (B, X), $B \subseteq A$, $X \subseteq U$, is a decision bireduct in $\mathbb{A}$, iff $B \cup X^*$, for $X^* = \{a_i^* \in A^* : i \notin X\}$, is the decision reduct in $\mathbb{A}^*$.*

Proof. The proof is straightforward and we omit it because of space limitations.

An illustrative example of the considered transformation can be seen in Table 3. Certainly, it should be treated just as a starting point for developing more efficient algorithms, because decision systems of the form $\mathbb{A}^* = (U, A \cup A^* \cup \{d\})$ cannot be constructed explicitly for large data. An appropriate translation of methods aiming at searching for decision reducts in systems with large amount of attributes can be especially useful in this case [18].

Another way to employ standard reduct computations in order to search for decision bireducts can be generally referred to sampling methods [19].

Table 4. Indiscernibility classes induced by randomly selected attributes $\{T, H\}$ for decision system in Table 1

	Temp.	Humid.	Sport?
1	hot	high	no
2	hot	high	no
3	hot	high	yes
13	hot	normal	yes
4	mild	high	yes
8	mild	high	no
12	mild	high	yes
14	mild	high	no
10	mild	normal	yes
11	mild	normal	yes
5	cool	normal	yes
6	cool	normal	no
7	cool	normal	yes
9	cool	normal	yes

Table 5. $\mathbb{A}' = (U', A' \cup \{d\})$ for randomly selected representatives $U' = \{1, 6, 8, 10, 13\}$. Decision reduct $\{T, H\}$ in $\mathbb{A}'$ corresponds to bireduct $(\{T, H\}, \{1, 2, 6, 8, 10, 11, 13, 14\})$ in $\mathbb{A}$.

	Temp.	Humid.	Sport?
1	hot	high	no
6	cool	normal	no
8	mild	high	no
10	mild	normal	yes
13	hot	normal	yes

Table 6. The case of $U' = \{3, 6, 11, 12, 13\}$. Decision reduct $\{T\}$ in $\mathbb{A}'$ corresponds to bireduct $(\{T\}, \{3, 4, 6, 10, 11, 12, 13\})$ in $\mathbb{A}$.

	Temp.	Humid.	Sport?
3	hot	high	yes
6	cool	normal	no
11	mild	normal	yes
12	mild	high	yes
13	hot	normal	yes

Proposition 4. *For a given $\mathbb{A} = (U, A \cup \{d\})$, consider the three-step procedure:*

1. *Randomly select a subset of attributes $A' \subseteq A$;*
2. *Choose a single object from each of partition blocks induced by A' – all chosen objects form a subset denoted by $U' \subseteq U$;*
3. *Find a standard decision reduct $B \subseteq A'$ for the system $\mathbb{A}' = (U', A' \cup \{d\})$.*

Then the pair (B, X), where $X = \big\{u \in U : \exists_{x \in U'} \forall_{a \in B \cup \{d\}}\, a(x) = a(u)\big\}$, is a decision bireduct for $\mathbb{A}$. Moreover, each decision bireduct for $\mathbb{A}$ can be obtained as a result of the above steps, no matter what method is used in the third stage.

Proof. Again, we omit the proof because of space limitations.

We illustrate the above procedure by Tables 4, 5, 6. Let us note that the reduced decision systems obtained in the third of above steps are compact representations of if-then rules generated by attributes in B, with their supports summing up to the overall support $X \subseteq U$ of decision bireduct (B, X). However, consequences of those rules are not necessarily chosen in a way aiming at maximizing $|X|$. Quite oppositely, when combined with appropriate mechanisms of sampling, this process can lead to ensembles of decision bireducts based on possibly diversified subsets of attributes and objects, with the underlying if-then rules paying attention to the cases not covered by rules corresponding to other bireducts rather than the cases that are easiest to describe.

The algorithm outlined in Proposition 4 could be also modeled within the framework sketched in Proposition 1, by considering more specific permutations

$\sigma : \{1, ..., n+m\} \to \{1, ..., n+m\}$ with some amount of attributes at their beginning, an ordering of all objects in their middle, and the remainder of attributes at their very end. Indeed, in such a case, all attributes at the very beginning of σ will be removed; then, within each partition class induced by the remaining attributes, objects corresponding to only one of possible decision values will be added (precisely, it will be the decision value of the first element of a given partition class occurring in σ); and finally the algorithm will try to remove each of the remaining attributes according to their ordering in σ, subject the discernibility criteria with respect to the previously-added objects.

5 Bireducts in Data Streams

The main motivation for introducing decision bireducts in [5] was to establish a simple framework for constructing rough-set-based classifier ensembles, as well as to extend capabilities of decision reducts to model data dependencies. Going further, in [10] it was noticed that algorithms for extracting meaningful bireducts from data could be utilized to integrate the tasks of attribute and instance selection. Such a potential is also illustrated by Proposition 4, where the objects in U' actually define a classifier based on the resulting $B \subseteq A$.

Some areas of applications were also pointed out for other types of bireducts. In [20], so called information bireducts were employed to model context-based object similarities in multi-dimensional data sets. Information bireducts may be also able to approximate data complexity analogously to some well-known mathematical tools [21]. Indeed, by investigating cardinalities of minimal subsets of attributes discerning maximal subsets of objects we can attempt to express a potential of a data source to define different concepts of interest.

In this section, we study one more opportunity in front of bireducts. Let us consider a stream of objects that is too large to be stored or represents data collected on-line [22]. For our purposes, let us focus on a stream interpreted as a decision system $\mathbb{A} = (U, A \cup \{d\})$, where there is no possibility to look at the entire U at any moment of processing time. Instead, given a natural order over U, we can access some buffered data intervals, i.e., the subsets of objects that occur consecutively in a stream. The question is how to design and efficiently conduct a process of attribute reduction in such a dynamic situation.

One of possibilities would be to fix the amount of objects in each data interval and compare decision reducts obtained for such narrowed down decision systems, in a kind of sliding window fashion. However, an arbitrary choice of the interval length may significantly influence the results. Thus, it may be more reasonable to adaptively adjust data intervals with respect to the currently observed attribute dependencies. Moreover, if our goal is to search for stable subsets of attributes that remain decision reducts for possibly wide areas of data, then we should tend to maximizing data intervals in parallel to minimizing the amounts of attributes necessary to determine decision classes within them.

Definition 6. *Let $\mathbb{A} = (U, A \cup \{d\})$ be given. Let U be naturally ordered with its elements indexed by integers. Consider a pair (B, X), where $B \subseteq A$ and $X = \langle first, last \rangle$. We say that (B, X) is a temporal decision bireduct, iff:*

- *An inexact dependency $B \Rrightarrow_X d$ holds;*
- *There is no $C \subsetneq B$ such that $C \Rrightarrow_X d$;*
- *$B \Rrightarrow_Y d$ is not true for neither $Y = \langle first{-}1, last \rangle$ nor $Y = \langle first, last{+}1 \rangle$.*

The above modification of Definition 2 can serve as a background for producing bireducts (B, X) with no *holes* in X with respect to a given data flow. Below we sketch an example of heuristic extraction of such bireducts from data. From a technical point of view, it resembles Proposition 4 with respect to a random choice of a subset of attributes to be analyzed. From a more strategic perspective, let us note that our goal is now to save the identified temporal bireducts analogously to micro-clusters [23] or data blocks [24] constructed within other applications for the purposes of further steps of on-line or off-line analysis. This way of data stream processing may open new opportunities for the task of scalable attribute subset selection. For instance, basing on frequent occurrence of a given subset of attributes in the previously-found temporal bireducts, one can reason about its ability to induce a robust decision model.

Proposition 5. *Let $\mathbb{A} = (U, A \cup \{d\})$ be given. Let U be naturally ordered with its elements indexed by integers. Select an arbitrary $A' \subseteq A$ and put $B = X = \emptyset$. Consider the following steps for each consecutive i-th object in U:*

1. *If $B \Rrightarrow_{X \cup \{i\}} d$, then add i to X;*
2. *Else, save (B, X), add i to X, and do the following:*
   (a) *Put $B = A'$ and remove the oldest objects from X until there is $B \Rrightarrow_X d$;*
   (b) *Heuristically reduce redundant attributes under the constraint $B \Rrightarrow_X d$.*

Then, all pairs (B, X) saved during the above procedure are temporal bireducts for $\mathbb{A}$. Moreover, each temporal bireduct can be obtained as one of saved pairs (B, X) for some $A' \subseteq A$, no matter what method is used in the last step.

Proof. Consider a pair (B, X), where $X = \langle first, last \rangle$, which was saved in the step 2. For such a case, we know that $B \Rrightarrow_{\langle first, last \rangle} d$ and $B \not\Rrightarrow_{\langle first, last+1 \rangle} d$. Also, there is $B \not\Rrightarrow_{\langle first-1, last \rangle} d$ because the oldest object in X is removed only when the newly joined object cannot be handled together with some elements of X even when using the whole A'. Therefore, X cannot be extended backwards beyond object $first$. Also, because of reduction of redundant attributes, B is irreducible for X. Hence, all saved pairs (B, X) are temporal bireducts.

Now, consider a temporal bireduct $(B, \langle first, last \rangle)$ and put $A' = B$. Consider the first buffer including object $first$, i.e., $\langle older, first \rangle$, $older \leq first$. Each next entry until object $last$ will be added with no need of removing $first$ (otherwise there would be no $B \Rrightarrow_{\langle first, last \rangle} d$). Moreover, when adding $last$, all objects older than $first$ (if any of them are still present) will be erased from the buffer (otherwise there would be $B \Rrightarrow_{\langle first-1, last \rangle} d$). Finally, when adding object $last + 1$ to $\langle first, last \rangle$, we will need to remove $first$ (otherwise there would be $B \Rrightarrow_{\langle first, last+1 \rangle} d$), which results in saving $(B, \langle first, last \rangle)$. □

#	O	T	H	d	State
1	sunny	hot	high	no	$S_1 = (\emptyset, \langle 1,1 \rangle)$
2	sunny	hot	high	no	$S_2 = (\emptyset, \langle 1,2 \rangle)$
					[save S_2]
3	overcast	hot	high	yes	$S_3 = (\{O\}, \langle 1,3 \rangle)$
4	rain	mild	high	yes	$S_4 = (\{O\}, \langle 1,4 \rangle)$
5	rain	cool	normal	yes	$S_5 = (\{O\}, \langle 1,5 \rangle)$
					[save S_5]
6	rain	cool	normal	no	$S_6 = (\emptyset, \langle 6,6 \rangle)$
					[save S_6]
7	overcast	cool	normal	yes	$S_7 = (\{O\}, \langle 6,7 \rangle)$
8	sunny	mild	high	no	$S_8 = (\{O\}, \langle 6,8 \rangle)$
					[save S_8]
9	sunny	cool	normal	yes	$S_9 = (\{O,T\}, \langle 6,9 \rangle)$
10	rain	mild	normal	yes	$S_{10} = (\{O,T\}, \langle 6,10 \rangle)$
					[save S_{10}]
11	sunny	mild	normal	yes	$S_{11} = (\{O,T,H\}, \langle 6,11 \rangle)$
12	overcast	mild	high	yes	$S_{12} = (\{O,T,H\}, \langle 6,12 \rangle)$
13	overcast	hot	normal	yes	$S_{13} = (\{O,T,H\}, \langle 6,13 \rangle)$
14	rain	mild	high	no	$S_{14} = (\{O,T,H\}, \langle 6,14 \rangle)$

#	T	H	W	d	State
1	hot	high	weak	no	$S_1 = (\emptyset, \langle 1,1 \rangle)$
2	hot	high	strong	no	$S_2 = (\emptyset, \langle 1,2 \rangle)$
					[save S_2]
3	hot	high	weak	yes	$S_3 = (\{W\}, \langle 2,3 \rangle)$
4	mild	high	weak	yes	$S_4 = (\{W\}, \langle 2,4 \rangle)$
5	cool	normal	weak	yes	$S_5 = (\{W\}, \langle 2,5 \rangle)$
6	cool	normal	strong	no	$S_6 = (\{W\}, \langle 2,6 \rangle)$
					[save S_6]
7	cool	normal	strong	yes	$S_7 = (\emptyset, \langle 7,7 \rangle)$
					[save S_7]
8	mild	high	weak	no	$S_8 = (\{W\}, \langle 7,8 \rangle)$
					[save S_8]
9	cool	normal	weak	yes	$S_9 = (\{H\}, \langle 7,9 \rangle)$
10	mild	normal	weak	yes	$S_{10} = (\{H\}, \langle 7,10 \rangle)$
11	mild	normal	strong	yes	$S_{11} = (\{H\}, \langle 7,11 \rangle)$
					[save S_{11}]
12	mild	high	strong	yes	$S_{12} = (\{H,W\}, \langle 7,12 \rangle)$
13	hot	normal	weak	yes	$S_{13} = (\{H,W\}, \langle 7,13 \rangle)$
					[save S_{13}]
14	mild	high	strong	no	$S_{14} = (\{W\}, \langle 13,14 \rangle)$

Fig. 1. Extraction of temporal bireducts from a data set in Table 1. The left- and right-side sequences correspond to subsets $A' = \{O, T, H\}$ and $A' = \{T, H, W\}$, respectively.

In Proposition 5, subsets $X \subseteq U$ are treated as the buffers of objects that appeared most recently in a data stream, within which a currently considered $B \subseteq A$ is sufficient to determine decision classes. As an illustration, consider the decision system in Table 1 and assume that we receive objects from $U = \{1, ..., 14\}$ one after the other. Let the i-th state of the process be denoted by $S_i = (B_i, X_i)$, where i is the number of objects already received from U and B_i is a decision reduct for the current buffer content X_i.

Figure 1 presents two examples of randomly chosen subsets of attributes. Let us concentrate on $A' = \{T, H, W\}$ and refer one more time to the permutation-based characteristics of decision reducts outlined e.g. in [7]. Namely, in the step 2(b) in Proposition 5, we are going to reduce attributes along $\sigma = \langle T, H, W \rangle$. In general, when following the same $\sigma : \{1, ..., n'\} \to \{1, ..., n'\}$, $n' = |A'|$, from the very beginning of a data stream, we can count on smoother evolution of subsets $B_i \subseteq A'$ for consecutive buffers. Furthermore, by working with a larger family of diversified subsets $A' \subseteq A$, we have a chance to witness the most representative changes of the observed temporal bireducts in time.

Let us now take a closer look at $A' = \{T, H, W\}$. The first two objects share the same decision. Thus, there is $S_2 = (\emptyset, \langle 1, 2 \rangle)$. Further, $\emptyset \Rightarrow_{\langle 1,3 \rangle} d$ is not valid, so we save the temporal bireduct $(\emptyset, \langle 1, 2 \rangle)$ and proceed with the step 2 in Proposition 5. As $\{T, H, W\}$ is insufficient to discern objects 1 and 3, we limit ourselves to $\langle 2, 3 \rangle$. Starting from $B = A'$ and given $\sigma = \langle T, H, W \rangle$, we reduce T and H, which results in the pair $S_3 = (\{W\}, \langle 2, 3 \rangle)$.

The next three objects do not break the dependency between $\{W\}$ and d. However, object 7 forces all earlier entries to be deleted. A different situation can be observed when adding the next two objects. In both cases, A' determines decision values, so we can keep buffers $\langle 7, 8\rangle$ and then $\langle 7, 9\rangle$. However, subsets of attributes generated by using the same σ will differ from each other. $\{W\}$ is not able to determine d within $\langle 7, 9\rangle$ although it was sufficient for $\langle 7, 8\rangle$. As a consequence, we need to restart from $B = A'$. We are allowed to remove T. Then, H turns out to be irreducible because of a need of keeping discernibility between objects 8 and 9. Finally, given the fact that H was not removed, W is not important any more, resulting in $S_9 = (\{H\}, \langle 7, 9\rangle)$.

6 Conclusions

In this paper, we attempted to establish better understanding of challenges and possibilities of searching for meaningful decision bireducts in data. We also outlined some examples of practical usage of decision bireducts in a new scenario of attribute subset selection in data streams. From this perspective, we need to remember that although decision bireducts were originally introduced in order to adopt some useful classifier ensemble principles, perhaps their major advantage lays in simple and flexible data-based knowledge representation.

In the nearest future, we intend to work on an enhanced interactive visualization of collections of decision bireducts, seeking for inspiration, e.g., in the areas of formal concept analysis [17] and visual bi-clustering [25]. We will also continue our studies on other types of bireducts, such as information bireducts which have been already successfully applied in [20]. Last but not least, following the research reported in [10], we are going to attempt to reconsider the discernibility-based bireduct construction criteria for the purposes of other rough set approaches, such as e.g. the dominance rough set model [26].

References

1. Świniarski, R.W., Skowron, A.: Rough Set Methods in Feature Selection and Recognition. Pattern Recognition Letters 24(6), 833–849 (2003)
2. Ślęzak, D.: Approximate Entropy Reducts. Fundamenta Informaticae 53(3-4), 365–390 (2002)
3. Abeel, T., Helleputte, T., de Peer, Y.V., Dupont, P., Saeys, Y.: Robust Biomarker Identification for Cancer Diagnosis with Ensemble Feature Selection Methods. Bioinformatics 26(3), 392–398 (2010)
4. Ślęzak, D.: Approximation Reducts in Decision Tables. In: Proc. of IPMU 1996, vol. 3, pp. 1159–1164 (1996)
5. Ślęzak, D., Janusz, A.: Ensembles of Bireducts: Towards Robust Classification and Simple Representation. In: Kim, T.-H., Adeli, H., Ślęzak, D., Sandnes, F.E., Song, X., Chung, K.-I., Arnett, K.P. (eds.) FGIT 2011. LNCS, vol. 7105, pp. 64–77. Springer, Heidelberg (2011)

6. Stawicki, S., Widz, S.: Decision Bireducts and Approximate Decision Reducts: Comparison of Two Approaches to Attribute Subset Ensemble Construction. In: Proc. of FedCSIS 2012, pp. 331–338. IEEE Computer Society (2012)
7. Bazan, J.G., Nguyen, H.S., Nguyen, S.H., Synak, P., Wróblewski, J.: Rough Set Algorithms in Classification Problem. In: Polkowski, L., Tsumoto, S., Lin, T.Y. (eds.) Rough Set Methods and Applications: New Developments in Knowledge Discovery in Information Systems. STUD FUZZ, vol. 56, pp. 49–88. Physica-Verlag, Heidelberg (2000)
8. Widz, S., Ślęzak, D.: Approximation Degrees in Decision Reduct-based MRI Segmentation. In: Proc. of FBIT 2007, pp. 431–436. IEEE Computer Society (2007)
9. Pawlak, Z., Skowron, A.: Rudiments of Rough Sets. Information Sciences 177(1), 3–27 (2007)
10. Mac Parthaláin, N., Jensen, R.: Simultaneous Feature And Instance Selection Using Fuzzy-Rough Bireducts. In: Proc. of FUZZ IEEE 2013 (2013)
11. Moshkov, M.J., Piliszczuk, M., Zielosko, B.: Partial Covers, Reducts and Decision Rules in Rough Sets – Theory and Applications. SCI, vol. 145. Springer, Heidelberg (2008)
12. Nguyen, S.H., Nguyen, H.S.: Pattern Extraction from Data. Fundamenta Informaticae 34(1-2), 129–144 (1998)
13. Ślęzak, D.: Normalized Decision Functions and Measures for Inconsistent Decision Tables Analysis. Fundamenta Informaticae 44(3), 291–319 (2000)
14. Sarawagi, S., Thomas, S., Agrawal, R.: Integrating Association Rule Mining with Relational Database Systems: Alternatives and Implications. Data Min. Knowl. Discov. 4(2/3), 89–125 (2000)
15. Kowalski, M., Stawicki, S.: SQL-based Heuristics for Selected KDD Tasks over Large Data Sets. In: Proc. of FedCSIS 2012, pp. 303–310. IEEE Computer Society (2012)
16. Nguyen, H.S.: Approximate Boolean Reasoning: Foundations and Applications in Data Mining. In: Peters, J.F., Skowron, A. (eds.) Transactions on Rough Sets V. LNCS, vol. 4100, pp. 334–506. Springer, Heidelberg (2006)
17. Ganter, B., Wille, R.: Formal Concept Analysis: Mathematical Foundations. Springer (1998)
18. Janusz, A., Ślęzak, D.: Utilization of Attribute Clustering Methods for Scalable Computation of Reducts from High-Dimensional Data. In: Proc. of FedCSIS 2012, pp. 295–302. IEEE Computer Society (2012)
19. Janusz, A., Stawicki, S.: Applications of Approximate Reducts to the Feature Selection Problem. In: Yao, J., Ramanna, S., Wang, G., Suraj, Z. (eds.) RSKT 2011. LNCS, vol. 6954, pp. 45–50. Springer, Heidelberg (2011)
20. Janusz, A., Ślęzak, D., Nguyen, H.S.: Unsupervised Similarity Learning from Textual Data. Fundam. Inform. 119(3-4), 319–336 (2012)
21. Blumer, A., Ehrenfeucht, A., Haussler, D., Warmuth, M.K.: Learnability and the Vapnik-Chervonenkis Dimension. J. ACM 36(4), 929–965 (1989)
22. Aggarwal, C.C. (ed.): Data Streams – Models and Algorithms. Advances in Database Systems, vol. 31. Springer (2007)
23. Zhang, T., Ramakrishnan, R., Livny, M.: BIRCH: An Efficient Data Clustering Method for Very Large Databases. In: Proc. of SIGMOD 1996, pp. 103–114. ACM Press (1996)
24. Ślęzak, D., Kowalski, M., Eastwood, V., Wróblewski, J.: Methods and Systems for Database Organization. US Patent 8,266,147 B2 (2012)

25. Havens, T.C., Bezdek, J.C.: A New Formulation of the coVAT Algorithm for Visual Assessment of Clustering Tendency in Rectangular Data. Int. J. Intell. Syst. 27(6), 590–612 (2012)
26. Słowiński, R., Greco, S., Matarazzo, B.: Dominance-Based Rough Set Approach to Reasoning About Ordinal Data. In: Kryszkiewicz, M., Peters, J.F., Rybiński, H., Skowron, A. (eds.) RSEISP 2007. LNCS (LNAI), vol. 4585, pp. 5–11. Springer, Heidelberg (2007)