International Journal of Approximate Reasoning 84 (2017) 75–109

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Decision bireducts and decision reducts – a comparison

Sebastian Stawicki [a,b], Dominik Ślęzak [a,b,*], Andrzej Janusz [a], Sebastian Widz [c]

[a] Faculty of Mathematics, Informatics and Mechanics, University of Warsaw, ul. Banacha 2, 02-097 Warsaw, Poland
[b] Infobright Inc., ul. Krzywickiego 34, 02-078 Warsaw, Poland
[c] Systems Research Institute, Polish Academy of Sciences, ul. Newelska 6, 01-447 Warsaw, Poland

ARTICLE INFO

Article history:
Received 18 July 2016
Received in revised form 13 January 2017
Accepted 21 February 2017
Available online 2 March 2017

Keywords:
Decision reducts
Decision bireducts
Classifier ensembles
Boolean reasoning
Computational complexity
Heuristic algorithms

ABSTRACT

In this paper we revise the notion of decision bireducts. We show new interpretations and we prove several important and practically useful facts regarding this notion. We also explain the way in which some of the well-known algorithms for computation of decision reducts can be modified for the purpose of computing decision bireducts. For the sake of completeness of our study we extend our investigations to relations between decision bireducts and so-called approximate decision reducts. We compare different formulations of those two approaches and draw analogies between them. We also report new results related to NP-hardness of searching for optimal decision bireducts and approximate decision reducts from data. Finally, we present new results of empirical tests which demonstrate usefulness of decision bireducts in a construction of efficient, yet simple ensembles of classification models.

1. Introduction

Decision reducts, developed as one of fundamental notions of the theory of rough sets [1], have a number of applications in attribute selection and knowledge discovery [2]. The notions analogous to decision reducts occur in many areas of science, such as Markov blankets in probabilistic modeling or irreducible multi-valued dependencies in relational databases [3]. There are a number of approaches utilizing both single decision reducts and their ensembles in the areas of data classification and representation [4]. There are also a number of thorough theoretical investigations [5] and software implementations letting better understand and apply decision reducts in practice [6].

Among a variety of extensions of decision reducts, there are approximate decision reducts introduced in [7] to deal with large, noisy data. An approximate decision reduct is an irreducible subset of attributes satisfying a kind of decision information preserving criterion, usually specified as an approximation threshold for a value of a function evaluating subsets of attributes against the data. Such function can reflect, e.g., a chance that a classifier constructed using a selected subset of attributes does not misclassify considered objects. It is further expected that values of such functions do not increase for smaller subsets of attributes, as classifiers based on less information have limited possibilities to distinguish between objects supporting different decision classes. Using such an approach, one can obtain subsets of attributes that are moderately less accurate than standard decision reducts but could be preferred in real-world applications as more robust and containing less attributes [8].

* Corresponding author.
E-mail addresses: sebastian.stawicki@gmail.com (S. Stawicki), slezak@mimuw.edu.pl (D. Ślęzak), janusza@mimuw.edu.pl (A. Janusz), sebastian.widz@gmail.com (S. Widz).

http://dx.doi.org/10.1016/j.ijar.2017.02.007
0888-613X/

In [9], it was discussed that ensembles of classifiers based on different approximate decision reducts can repeatedly misclassify the same data instances. This is because the above functions evaluate attribute subsets by means of overall summaries, without reporting what happens with particular objects. To address this issue, one might think about combining a process of searching for ensembles of approximate decision reducts with popular machine learning techniques, such as boosting or bagging [10]. Alternatively, decision bireducts were proposed as a new extension of decision reducts, wherein the emphasis is on both a subset of attributes which describes decisions and a subset of objects for which that description is valid. In other words, a decision bireduct is represented as a pair, where a subset of attributes can be evaluated by means of a subset of objects for which it assures good classification. Such a subset of objects can be assessed using many different criteria but, after all, it provides far more explicit information about the corresponding subset of attributes and its abilities to construct a good classifier than any evaluation function. In particular, information about objects whose decision values are validly described allows to verify whether classifiers designed using different selected subsets of attributes do not repeat mistakes on the same areas of the training data.

The idea of decision bireducts can be considered as a bridge between a few different views at the goals of knowledge discovery. In [9], ensembles of information bireducts were studied as a counterpart to the notion of concepts studied in the formal concept analysis and itemsets studied in the literature related to association rule mining. The concepts in the formal concept analysis are defined as non-extendable subsets of objects that behave in the same way with respect to non-extendable subsets of attributes [11]. Similarly, the itemsets (or patterns, templates) aim at describing a maximum number of objects with the same values on a maximum number of attributes [12]. In contrast, information bireducts correspond to non-extendable subsets of objects that can be said as different using irreducible subsets of attributes. Thus, they seem to refer to the most diversified (and thus most informative) fragments of data.

This paper aims at drawing analogies between decision bireducts, wherein decision classes are correctly determined by irreducible subsets of attributes over non-extendable subsets of objects, and other methods of representing dependencies in data. As a starting point, we refer to our previous research on decision bireducts with respect to their comparison to standard and approximate decision reducts [13]. We show correspondence between those notions and provide some new interpretations which constitute a basis for innovative algorithmic solutions to the problem of extracting the most interesting decision bireducts from data. We pay a special attention to interpretation of considered types of decision reducts and decision bireducts in terms of collections of decision rules which are able to neglect potentially noisy instances and, therefore, are more likely to remain robust when classifying new data.

For clarity, in this paper we only consider the classical setup from the theory of rough sets, i.e., we assume that all attributes are categorical. Consequently, we focus on decision bireducts formulated in terms of the standard discernibility. However, one could easily consider generalizations of this research by employing, e.g., fuzzy-rough models [14], similarity based models [15], or dominance based models [16]. It is also worth noting that most of the presented theoretical material concentrates on characteristics of single decision bireducts. We do it in order to set up firm mathematical and algorithmic foundations before proceeding with more advanced ways of utilizing decision bireducts in future. On the other hand, we discuss some ideas behind construction of ensembles of decision bireducts and we report previously unpublished experimental results showing performance of such ensembles in classification tasks.

The paper is divided into five sections. After a brief introduction and definition of preliminary notions related to bireducts in Section 2, we dwell on heuristic methods for finding them in decision tables in Section 3. In particular, we first show a data representation that allows to neatly compute all possible decision bireducts for a given table. Next, we recall an ordering algorithm [9] that is adapted to search for bireducts. In addition we propose to adapt for this purpose an algorithm that is based on random sampling of attributes and data objects. In Section 4 we focus on relations between the bireducts and different types of approximate decision reducts. In that section we also discuss possible criteria for selection of optimal bireducts for a given decision table and show some measures that can guide selection of an efficient ensemble of rule based classifiers. In the last subsection of Section 4 we present some new experimental results which compare the efficiency of classifier ensembles derived from bireducts with a few other popular ensemble construction techniques. Finally, we conclude the paper in Section 5.

2. Basics of decision reducts and decision bireducts

2.1. Decision reducts

In the following, we use standard representation of tabular data in a form of decision tables [1].

Definition 1. A decision table is a pair $\mathbb{A} = (U, A \cup \{d\})$ of non-empty, finite sets U and $A \cup \{d\}$, where U is a universe of objects, and $A \cup \{d\}$ is a set consisting of attributes such that every $a \in A \cup \{d\}$ is a function $a : U \to V_a$, where V_a denotes a's domain and is called the value set of a. The distinguished attribute d, such that $d \notin A$, is called a decision attribute and the elements of A are called conditional attributes.

For the decision attribute d, the values $v_d \in V_d$ correspond to decision classes that we want to describe using the values of attributes in A, in order to utilize such descriptions in the process of classification of objects outside U.