Global Optimization of Exact Association Rules Relative to Length

Beata Zielosko

Institute of Computer Science
University of Silesia
39, Bedzinska St., 41-200 Sosnowiec, Poland
beata.zielosko@us.edu.pl

Abstract. In the paper, an application of dynamic programming approach to global optimization of exact association rules relative to length is presented. It is an extension of the dynamic programming approach to optimization of decision rules to inconsistent tables. An information system I is transformed into a set of decision tables $\{I_{f_1}, \ldots, I_{f_{n+1}}\}$. The algorithm constructs, for each decision table from the set $\{I_{f_1}, \ldots, I_{f_{n+1}}\}$, a directed acyclic graph $\Delta(I_{f_i})$, $i = 1, \ldots, n+1$. Based on the graph, the set of so-called irredundant (f_i)-association rules can be described. The union of sets of (f_i)-association rules, $i = 1, \ldots, n+1$, is considered as a set of association rules for information system I. Then, global optimization relative to length is made and sets of association rules with minimum length, for each row of information system I, are obtained. Preliminary experimental results with data sets from UCI Machine Learning Repository are included.

Key words: rough sets, association rules, length, dynamic programming, decision rules

1 Introduction

Association rule mining is one of the important fields of data mining and knowledge discovery. It aims to extract interesting correlations, associations, or frequent patterns among sets of items in data set.

There are many algorithms for construcion of association rules. The most popular algorithm for mining associaion rules is Apriori algorithm [1]. During years, based on the Apriori, many new algorithms were designed with some modifications or improvements, e.g., hash based technique [15], transaction reduction [7] and others [8, 13, 18]. Another approaches are frequent pattern growth [5] that adopts divide and conquer strategy, and algorithms that uses vertical data format [6].

In the paper, an application of dynamic programming approach to optimization of exact association rules relative to length is presented. Construcion of short rules is connected with the Minimum Description Length principle introduced by Rissanen [17]: the best hypothesis for a given set of data is the one that leads to the largest compression of the data. Short rules are more understandable and easier for interpreting by experts. Unfortunately, the problem of construction of rules with minimum length is

NP-hard [11, 14]. The most part of approaches, with the exception of brute-force and in some sense Apriori algorithm, cannot guarantee the construction of rules with the minimum length. The proposed approach allows one to construct optimal rules, i.e., rules with the minimum length.

Application of rough sets theory to the construction of rules for knowledge representation or classification tasks are usually connected with the usage of decision table [16] as a form of input data representation. In such a table one attribute is distinguished as a decision attribute and it relates to a rule's consequence. However, in the last years, associative mechanism of rule construction, where all attributes can occur as premises or consequences of particular rules, is popular. Association rules can be defined in many ways. In the paper, a special kind of association rules is studied, i.e., they relate to decision rules. Similar approach was considered in [11, 12], where greedy algorithm for minimization of length of association rules was studied.

This paper is an essential extension of the paper [3] in which consistent decision tables were considered, i.e., they do not contain equal rows with different decisions. When association rules for information systems are studied and each attribute is sequentially considered as the decision one, inconsistent tables are often obtained. So, the approach considered in [3] is extended to the case of inconsistent decision tables. It required changes in definitions, algorithms (new conditions of stop), proofs of algorithm correctness, and, especially, in the software.

Proposed approach by comparison with Apriori algorithm allows one to derive a reguired number of rules for a given row only. If we consider sequential optimization relative to coverage and length it is possible to find so-called totally-optimal rules, i.e., rules with maximum coverage and minimum length. Experimental resuls show that such rules have often good coverage and small length.

The aim of the paper is to create a research tool which is applicable to medium sized decision tables and allows one to construct exact association rules with minimum length. To this end, an information system I with attributes $\{f_1, \ldots, f_{n+1}\}$ is transformed into a set of decision tables $\{I_{f_1}, \ldots, I_{f_{n+1}}\}$. The algorithm constructs, for each decision table from the set $\{I_{f_1}, \ldots, I_{f_{n+1}}\}$, a directed acyclic graph $\Delta(I_{f_i})$, $i = 1, \ldots, n+1$. Based on the graph $\Delta(I_{f_i})$, the set of so-called irredundant (f_i)-association rules, $i = 1, \ldots, n+1$, can be described. The union of sets of irredundant (f_i)-association rules, $i = 1, \ldots, n+1$, is considered as the set of irredundant association rules for information system I. Using global optimization relative to length it is possible to obtain the set of irredundant association rules for information system I with minimum length. In [20], global optimization of association rules relative to coverage was presented.

The paper consists of six sections. Section 2 contains main notions. In Section 3, an algorithm for construction of a directed acyclic graph is presented. Section 4 contains a description of optimization procedure relative to length. Section 5 contains experimental results for decision tables from UCI Machine Learning Repository, and Section 6 - conclusions.

2 Main Notions

An *information system* I is a table with $n + 1$ columns labeled with attributes $f_1, \ldots, f_{n+1}$. Rows of this table are filled by nonnegative integers which are interpreted as values of attributes.

An association rule for I is a rule of the kind

$$f_{i_1} = a_1 \wedge \ldots \wedge f_{i_m} = a_m \rightarrow f_j = a,$$

where $f_j \in \{f_1, \ldots, f_{n+1}\}$, $f_{i_1}, \ldots, f_{i_m} \in \{f_1, \ldots, f_{n+1}\} \setminus \{f_j\}$, and a,a_1,…,a_m are nonnegative integers.

The notion of an association rule for I is based on the notion of a decision table and decision rule.

A *decision table* T is a table with n columns labeled with (conditional) attributes $f_1, \ldots, f_n$. Rows of this table are filled by nonnegative integers which are interpreted as values of conditional attributes. Each row is labeled with a nonnegative integer (decision) which is interpreted as a value of a decision attribute. It is possible that T contains equal rows with the same or different decisions.

For each attribute $f_i \in \{f_1, \ldots, f_{n+1}\}$, the information system I is transformed into a decision table I_{f_i}. The column f_i is removed from I and a table with n columns labeled with attributes $f_1, \ldots, f_{i-1}, f_{i+1}, \ldots, f_{n+1}$ is obtained. Values of the attribute f_i are attached to the rows of the obtained table which will be denoted by I_{f_i}.

$$I = \begin{array}{c|ccc} & f_1 & f_2 & f_3 \\ \hline r_1 & 1 & 1 & 2 \\ r_2 & 0 & 0 & 2 \\ r_3 & 1 & 1 & 1 \end{array} \Rightarrow I_{f_1} = \begin{array}{c|ccc} & f_2 & f_3 & \\ \hline r_1 & 1 & 2 & 1 \\ r_2 & 0 & 2 & 0 \\ r_3 & 1 & 1 & 1 \end{array}$$

Fig. 1. Decision table I_{f_1} obtained from information system I

The set $\{I_{f_1}, \ldots, I_{f_{n+1}}\}$ of decision tables obtained from the information system I is denoted by Φ. Let $T \in \Phi$. For simplicity, it is assumed that $T = I_{f_{n+1}}$.

The table T is called *degenerate* if it is empty or all rows of T are labeled with the same decision, or all rows of T are equal.

A minimum decision value that is attached to the maximum number of rows in T will be called the *most common decision for* T.

A table obtained from T by the removal of some rows is called a *subtable* of the table T. Let $f_{i_1}, \ldots, f_{i_m} \in \{f_1, \ldots, f_n\}$ and $a_1, \ldots, a_m$ be nonnegative integers. A subtable of the table T, which contains only rows of T that have numbers $a_1, \ldots, a_m$ at the intersection with columns $f_{i_1}, \ldots, f_{i_m}$, is denoted by $T(f_{i_1}, a_1) \ldots (f_{i_m}, a_m)$. Such subtables (including the table T) are called *separable subtables* of T.

The set of attributes from $\{f_1, \ldots, f_n\}$ which are nonconstant in T is denoted by $E(T)$. For any $f_i \in E(T)$, the set of values of the attribute f_i in T, is denoted by $E(T, f_i)$.

The expression

$$f_{i_1} = a_1 \wedge \ldots \wedge f_{i_m} = a_m \rightarrow f_{n+1} = d \qquad (1)$$

is called a *decision rule over* T if $f_{i_1}, \ldots, f_{i_m} \in \{f_1, \ldots, f_n\}$, and $a_1, \ldots, a_m, d$ are nonnegative integers. It is possible that $m = 0$. In this case (1) is equal to the rule

$$\rightarrow f_{n+1} = d. \qquad (2)$$

Let $r = (b_1, \ldots, b_n)$ be a row of T. Rule (1) will be called *realizable for* r, if $a_1 = b_{i_1}, \ldots, a_m = b_{i_m}$. If $m = 0$ then rule (2) is realizable for any row from T.

Rule (1) will be called *true for* T if the table $T' = T(f_{i_1}, a_1) \ldots (f_{i_m}, a_m)$ is degenerate and d is the most common decision for T'. If $m = 0$ then rule (2) is true for T if T is degenerate and d is the most common decision for T.

If rule (1) is true for T and realizable for r, then (1) will be called a *decision rule for* T *and* r.

Decision rules for T and r will be called (f_{n+1})-association rules for I and r. In general case, the notion of (f_i)-*association rule for* I *and* r coincides with the notion of decision rule for I_{f_i} and r, $i = 1, \ldots, n+1$. The union of sets of (f_i)-association rules, $i = 1, \ldots, n+1$, will be considered as the set of association rules for I and r.

Let $T = I_{f_{n+1}}$ and (1) be a decision rule over T. Rule (1) will be called an *irredundant* rule for T and r if (1) is a decision rule for T and r and the following conditions hold if $m > 0$:

(i) $f_{i_1} \in E(T)$, and if $m > 1$ then $f_{i_j} \in E(T(f_{i_1}, a_1) \ldots (f_{i_{j-1}}, a_{j-1}))$ for $j = 2, \ldots, m$;

(ii) if $m = 1$ then the table T is nondegenerate, and if $m > 1$ then the table $T(f_{i_1}, a_1) \ldots (f_{i_{m-1}}, a_{m-1})$ is nondegenerate.

If $m = 0$ then rule (2) is an *irredundant* decision rule for T and r if (2) is a decision rule for T and r, i.e., if T is degenerate and d is the most common decision for T.

Let $T = I_{f_{n+1}}$, τ be a decision rule over T, and τ be equal to (1). The *length* of τ is the number m of descriptors (pairs attribute=value) on the left-hand side of τ. It is denoted by $l(\tau)$. If $m = 0$ then the length of the rule τ is equal to 0.

3 Algorithm for Directed Acyclic Graph Construction

In this section, an algorithm for construction of a directed acyclic graph for a given decision table T is presented. Based on this graph it is possible to describe the set of irredundant rules for T and for each row r of T. This algorithm is repeated for each decision table I_{f_i}, $i = 1, \ldots, n+1$, obtained from the information system I.

Let $T = I_{f_{n+1}}$. The constructed graph is denoted by $\Delta(T)$. Nodes of the graph are some separable subtables of the table T. During each step, the algorithm processes one node and marks it with the symbol *. At the first step, the algorithm constructs a graph containing a single node T which is not marked with *. Let the algorithm have already performed p steps. Now the step $(p+1)$ will be described. If all nodes are marked with the symbol * as processed, the algorithm finishes its work and presents the resulting graph as $\Delta(T)$. Otherwise, choose a node (table) Θ, which has not been processed yet. If Θ is degenerate then mark the considered node with symbol * and proceed to the step

$(p+2)$. Otherwise, for each $f_i \in E(\Theta)$, draw a bundle of edges from the node Θ. Let $E(\Theta, f_i) = \{b_1, \ldots, b_t\}$. Then draw t edges from Θ and label these edges with pairs $(f_i, b_1), \ldots, (f_i, b_t)$ respectively. These edges enter to nodes $\Theta(f_i, b_1), \ldots, \Theta(f_i, b_t)$. If some of nodes $\Theta(f_i, b_1), \ldots, \Theta(f_i, b_t)$ are absent in the graph then add these nodes to the graph. Each row r of Θ is labeled with the set of attributes $E_{\Delta(T)}(\Theta, r) = E(\Theta)$. Mark the node Θ with the symbol * and proceed to the step $(p+2)$.

The graph $\Delta(T)$ is a directed acyclic graph. A node Θ of this graph will be called *terminal* if there are no edges leaving this node. A node Θ is terminal if and only if Θ is degenerate.

Later, a local optimization of the graph $\Delta(T)$ relative to the length will be described. As a result, a graph $G(T)$, with the same sets of nodes and edges as in $\Delta(T)$, is obtained. The only difference is that any row r of each nonterminal node Θ from $G(T)$ is labeled with a nonempty set of attributes $E_{G(T)}(\Theta, r) \subseteq E(\Theta)$, possibly different from $E(\Theta)$.

Let $G \in \{\Delta(T), G(T)\}$. Now, for each node Θ of G and for each row r of Θ, a set of rules $Rul_G(\Theta, r)$ will be described.

Let Θ be a terminal node of G. In this case Θ is a degenerate table and

$$Rul_G(\Theta, r) = \{\rightarrow f_{n+1} = d\},$$

where d is the most common decision for Θ.

Let now Θ be a nonterminal node of G such that, for each child Θ' of Θ and for each row r' of Θ', the set of rules $Rul_G(\Theta', r')$ is already defined. Let $r = (b_1, \ldots, b_n)$ be a row of Θ. For any $f_i \in E_G(\Theta, r)$, the set of rules $Rul_G(\Theta, r, f_i)$ is defined as follows:

$$Rul_G(\Theta, r, f_i) = \{f_i = b_i \wedge \gamma \rightarrow f_{n+1} = s :$$
$$\gamma \rightarrow f_{n+1} = s \in Rul_G(\Theta(f_i, b_i), r)\}.$$

Then $Rul_G(\Theta, r) = \bigcup_{f_i \in E_G(\Theta, r)} Rul_G(\Theta, r, f_i)$.

One can prove the following statement.

Theorem 1. *For any node Θ of $\Delta(T)$ and for any row r of Θ, $Rul_{\Delta(T)}(\Theta, r)$ is equal to the set of all irredundant rules for Θ and r.*

The algorithm for the directed acyclic graph construction is repeated for each decision table I_{f_i}, $i = 1, \ldots, n+1$, obtained from the information system I. In general, the obtained graph is denoted by $\Delta(I_{f_i})$, $i = 1, \ldots, n+1$. As a result, for $i = 1, \ldots, n+1$, the set $Rul_{\Delta(I_{f_i})}(I_{f_i}, r)$ of irredundant decision rules for I_{f_i} and r is obtained. This set will be called the set of irredundant (f_i)-association rules for I and r, $i = 1, \ldots, n+1$. The union of sets $Rul_{\Delta(I_{f_i})}(I_{f_i}, r)$ forms the set $Rul(I, r)$ of irredundant association rules for I and r:

$$Rul(I, r) = \bigcup_{i=1,\ldots,n+1} Rul_{\Delta(I_{f_i})}(I_{f_i}, r).$$

Example 1. To illustrate the presented algorithm the information system I depicted in Fig. 1 will be considered. Set $\Phi = \{I_{f_1}, I_{f_2}, I_{f_3}\}$ contains three decision tables obtained from I. Figure 2 presents a directed acyclic graph for decision table I_{f_1}. Based on the graph $\Delta(I_{f_1})$ the sets of rules attached to rows of I_{f_1} are described.

$Rul_{\Delta(I_{f_1})}(I_{f_1}, r_1) = \{f_2 = 1 \rightarrow f_1 = 1, f_3 = 2 \wedge f_2 = 1 \rightarrow f_1 = 1\};$

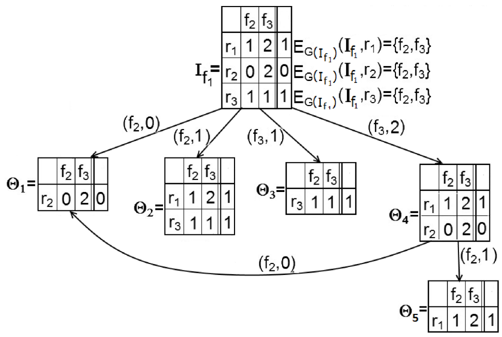

Fig. 2. Directed acyclic graph $\Delta(I_{f_1})$

$Rul_{\Delta(I_{f_1})}(I_{f_1}, r_2) = \{f_2 = 0 \to f_1 = 0, f_3 = 2 \wedge f_2 = 0 \to f_1 = 0\}$;
$Rul_{\Delta(I_{f_1})}(I_{f_1}, r_3) = \{f_2 = 1 \to f_1 = 1, f_3 = 1 \to f_1 = 1\}$.

Finally, based on the graphs $\Delta(I_{f_1})$, $\Delta(I_{f_2})$ and $\Delta(I_{f_3})$, it is possible to describe sets $Rul(I, r)$ of irredundant association rules for I and each row r of I.
$Rul(I, r_1) = \{f_2 = 1 \to f_1 = 1, f_3 = 2 \wedge f_2 = 1 \to f_1 = 1, f_1 = 1 \to f_2 = 1, f_3 = 2 \wedge f_1 = 1 \to f_2 = 1, f_1 = 1 \to f_3 = 1, f_2 = 1 \to f_3 = 1\}$;
$Rul(I, r_2) = \{f_2 = 0 \to f_1 = 0, f_3 = 2 \wedge f_2 = 0 \to f_1 = 0, f_1 = 0 \to f_2 = 0, f_3 = 2 \wedge f_1 = 0 \to f_2 = 0, f_1 = 0 \to f_3 = 2, f_2 = 0 \to f_3 = 2\}$;
$Rul(I, r_3) = \{f_2 = 1 \to f_1 = 1, f_3 = 1 \to f_1 = 1, f_1 = 1 \to f_2 = 1, f_3 = 1 \to f_2 = 1, f_1 = 1 \to f_3 = 1, f_2 = 1 \to f_3 = 1\}$.

4 Optimization Relative to Length

In this section, two kinds of optimization relative to the length l are presented: local optimization and global optimization.

Local optimization relative to the length is an optimization of the directed acyclic graph $\Delta(I_{f_i})$, $I_{f_i} \in \Phi$, constructed for a given decision table I_{f_i}.

Global optimization relative to the length is made for an information system I, i.e., among all graphs $G(I_{f_1}), \ldots, G(I_{f_{n+1}})$ constructed for decision tables from the set Φ and optimized locally relative to the length.

Let $T = I_{f_{n+1}}$. Now, a procedure of local optimization of the graph $\Delta(T)$ relative to the length will be described. For each node Θ in the graph $\Delta(T)$, this procedure assigns to each row r of Θ the set $Rul^l_{\Delta(T)}(\Theta, r)$ of decision rules with the minimum length from $Rul_{\Delta(T)}(\Theta, r)$ and the number $Opt^l_{\Delta(T)}(\Theta, r)$ – the minimum length of decision rule from $Rul_{\Delta(T)}(\Theta, r)$.

The algorithm moves from the terminal nodes of the graph $\Delta(T)$, which are degenerate tables, to the node T. It will attach the set $E_{G(T)}(\Theta, r)$ to each row r in Θ if Θ is a nonterminal node of $\Delta(T)$. The obtained graph is denoted by $G(T)$.

Let Θ be a terminal node of $\Delta(T)$. Then the number $Opt^l_{\Delta(T)}(\Theta, r) = 0$ is assigned to each row r of Θ.

Let Θ be a nonterminal node of $\Delta(T)$ and all children of Θ have already been treated. Let $r = (b_1, \ldots, b_n)$ be a row of Θ. The number

$$Opt^l_{\Delta(T)}(\Theta, r) = \min\{Opt^l_{\Delta(T)}(\Theta(f_i, b_i), r) + 1 : f_i \in E(\Theta, r)\}$$

is assigned to the row r in the table Θ and

$$E_{\Delta}(\Theta, r) = \{f_i : f_i \in E_{\Delta(T)}(\Theta, r), Opt^l_{\Delta(T)}(\Theta(f_i, b_i), r) + 1 = Opt^l_{\Delta(T)}(\Theta, r)\}.$$

Theorem 2. *For each node Θ of the graph $G(T)$ and for each row r of Θ, the set $Rul_{G(T)}(\Theta, r)$ is equal to the set $Rul^l_{\Delta(T)}(\Theta, r)$ of all rules with the minimum length from the set $Rul_{\Delta(T)}(\Theta, r)$.*

Now, a global optimization relative to the length is presented. It is made for the information system I.

The set of irredundant association rules for I and r with the minimum length from $Rul(I, r)$ is denoted by $Rul^l(I, r)$, and the minimum length of an association rule from $Rul(I, r)$ is denoted by $Opt^l(I, r)$.

To make global optimization relative to the length, the directed acyclic graph is constructed for each decision table $I_{f_i} \in \Phi$, and local optimization relative to the length of the graph $\Delta(I_{f_i}), i = 1, \ldots, n+1$, is made. As a result, the graph $G(I_{f_i})$ is obtained and each row r of I_{f_i}, $i = 1, \ldots, n+1$, has assigned the set $Rul_{G(I_{f_i})}(I_{f_i}, r)$ of (f_i)-association rules for I and r with the minimum length from $Rul_{\Delta(I_{f_i})}(I_{f_i}, r)$ and the number $Opt^l_{\Delta(I_{f_i})}(I_{f_i}, r)$, which is the minimum length of (f_i)-association rule from $Rul_{\Delta(I_{f_i})}(I_{f_i}, r)$.

Then, the value $Opt^l(I, r)$ is obtained, such that,

$$Opt^l(I, r) = \min\{Opt^l_{\Delta(I_{f_i})}(I_{f_i}, r) : i = 1, \ldots, n+1\},$$

and among all numbers, $i = 1, \ldots, n+1$, only these are selected, where

$$Opt^l_{\Delta(I_{f_i})}(I_{f_i}, r) = Opt^l(I, r).$$

These numbers forms the set $N(I)$. Then

$$Rul^l(I, r) = \bigcup_{i \in N(I)} Rul_{G(I_{f_i})}(I_{f_i}, r).$$

As a result of the global optimization relative to the length each row r of I has assigned the set $Rul^l(I, r)$ of association rules with the minimum length and the number $Opt^l(I, r)$.

Below one can find the set $Rul^l(I, r)$ and the value $Opt^l(I, r)$ for the information system I depicted in Fig. 1 and each row r of this system.
$Rul^l(I, r_1) = \{f_2 = 1 \to f_1 = 1, f_1 = 1 \to f_2 = 1, f_1 = 1 \to f_3 = 1, f_2 = 1 \to f_3 = 1\}$, $Opt^l(I, r_1) = 1$;
$Rul^l(I, r_2) = \{f_2 = 0 \to f_1 = 0, f_1 = 0 \to f_2 = 0, f_1 = 0 \to f_3 = 2, f_2 = 0 \to f_3 = 2\}$, $Opt^l(I, r_2) = 1$;

$Rul^l(I, r_3) = \{f_2 = 1 \rightarrow f_1 = 1, f_3 = 1 \rightarrow f_1 = 1, f_1 = 1 \rightarrow f_2 = 1, f_3 = 1 \rightarrow f_2 = 1, f_1 = 1 \rightarrow f_3 = 1, f_2 = 1 \rightarrow f_3 = 1\}$, $Opt^l(I, r_3) = 1$.

The problem of rule length minimization is NP-hard [11, 14]. The algorithms considered in this paper have polynomial time complexity depending on the size of decision table and the number of separable subtables in it. In general case, the number of separable subtables grows exponentially with the growth of table size. However, in [9, 10] classes of decision tables are described for each of which the number of separable subtables in tables from the class is bounded from above by a polynomial on the size of decision table.

5 Experimental Results

Experiments were made using data sets from UCI Machine Learning Repository [4] and modified software system Dagger [2].

Each data set was considered as information system I and, for each attribute $f_i \in \{f_1, \ldots, f_{n+1}\}$, the system I was transformed into a decision table I_{f_i}. The column f_i was removed from I and a table with n columns labeled with attributes $f_1, \ldots, f_{i-1}, f_{i+1}, \ldots, f_{n+1}$, was obtained. Values of the attribute f_i were attached to the rows of the obtained table I_{f_i}. The set $\{I_{f_1}, \ldots, I_{f_{n+1}}\}$ of decision tables obtained from the information system I is denoted by Φ.

Table 1 presents preliminary results of experiments connected with the minimum length of irredundant association rules (column "Association rules"). For each row r of I, the minimum length of an irredundant association rule for I and r was obtained. After that, for rows of I the minimum length of an association rule for I and r with the minimum length (column "Min"), the maximum length of such rule (column "Max") and the average length of association rules with the minium length - one for each row (column "Avg") were obtained. Column "Rows" contains the number of rows in I, column "Attr" contains the number of attributes in I. This table contains also, for the purpose of comparison, minimum, average and maximum length of exact irredundant decision rules (column "Decision rules") obtained by the dynamic programming algorithm.

Based on the results in Table 1, it is possible to see that the proposed approach allows one to obtain short association rules. The minimum value of minimum length (column "Min") is equal to 3 only for two data sets, for the rest of data sets, this value is equal to 1. In the case of comparison of length of association and decision rules, the minimum values (columns "Min") of minimum lenght of rules are the same. The average values (columns "Avg") and the maximum values (columns "Max"), often are smaller in the case of association rules. Only for data sets "Monks-1-test" and "Monks-3-test", the obtained results are the same for association and decision rules.

Table 2 presents the average number of nodes (column "Nodes") and the average number of edges (column "Edges") related to the data set I and the graph $\Delta(I_{f_i})$, $i = 1, \ldots, n+1$. For each data set I, the set Φ was obtained. For each decision table I_{f_i}, $i = 1, \ldots, n+1$, the graph $\Delta(I_{f_i})$ was constructed and the number of nodes and edges were calculated. Then, the average number of nodes and edges in the directed acyclic graphs $\Delta(I_{f_i})$, $i = 1, \ldots, n+1$, were computed.

Table 1. Minimum length of rules

			Association rules			Decision rules		
Data set	Rows	Attr	Min	Avg	Max	Min	Avg	Max
Adult-stretch	16	5	1	1.00	1	1	1.25	2
Balance-scale	625	5	3	3.14	4	3	3.20	4
Breast-cancer	266	10	1	1.83	3	1	2.67	6
Cars	1728	7	1	2.02	6	1	2.43	6
Hayes-roth-data	69	5	1	1.62	3	1	2.15	4
Lenses	24	5	1	1.25	3	1	1.40	3
Monks-1-test	432	7	1	2.25	3	1	2.25	3
Monks-3-test	432	7	1	1.75	2	1	1.75	2
Shuttle-landing	15	7	1	1.00	1	1	1.40	4
Teeth	23	9	1	1.00	1	1	2.26	4
Tic-tac-toe	958	10	3	3.00	4	3	3.02	4
Zoo-data	59	17	1	1.00	1	1	1.56	4

Table 2. Average number of nodes and edges

Data set	Rows	Attr	Nodes	Edges
Adult-stretch	16	5	48.0	104.0
Balance-scale	625	5	742.0	2386.0
Breast-cancer	266	10	6082.0	61063.6
Cars	1728	7	4335.3	17697.1
Hayes-roth-data	69	5	190.8	569.0
Lenses	24	5	70.8	174.8
Monks-1-test	432	7	1734.9	6760.1
Monks-3-test	432	7	1584.9	5770.4
Shuttle-landing	15	7	73.6	368.6
Teeth	23	9	112.3	952.7
Tic-tac-toe	958	10	31415.1	264362.9
Zoo-data	59	17	3595.4	57868.2

The proposed approach of rule induction is based on the analysis of the directed acyclic graph constructed for a given decision table. The structure of the graph depends on data set, i.e., number of attributes, distribution of values of attributes, number of rows. Such graph can be huge for larger data set. Therefore, possibilities of decreasing the size of the graph were studied by the author. In [19], the graph is constructed only for selected values of attributes contained in a decision table.

6 Conclusions

In the paper, an application of dynamic programming to global optimization of exact association rules relative to length was presented. It is based on the dynamic programming approach to optimization of decision rules. However, there are differences: (i) definitions are different, (ii) the information system is used, (ii) decision table can be inconsistent, and (iv) global optimization relative to length was studied. The presented approach can be considered as a research tool which allows one to construct association rules with minimum length.

Possible applications of association rules obtained using presented approach are construction of classifiers, inference process in knowledege base system, filling missing values of attributes.

Future works will be connected with future selection, construction of classifiers and possibilites of decreasing the size of the directed accyclic graph.

References

1. Agrawal, R., Imieliński, T., Swami, A.: Mining association rules between sets of items in large databases. In: SIGMOD '93, pp. 207–216. ACM (1993)
2. Alkhalid, A., Amin, T., Chikalov, I., Hussain, S., Moshkov, M., Zielosko, B.: Dagger: A tool for analysis and optimization of decision trees and rules. In: Computational Informatics, Social Factors and New Information Technologies: Hypermedia Perspectives and Avant-Garde Experiences in the Era of Communicability Expansion, pp. 29–39. Blue Herons (2011)
3. Amin, T., Chikalov, I., Moshkov, M., Zielosko, B.: Dynamic programming approach for exact decision rule optimization. In: Skowron, A., Suraj, Z. (eds.) Rough Sets and Intelligent Systems - Professor Zdzisław Pawlak in Memoriam - Volume 1, Intelligent Systems Reference Library, vol. 42, pp. 211–228. Springer (2013)
4. Asuncion, A., Newman, D.J.: UCI Machine Learning Repository (`http://wwwicsuciedu/~mlearn/`, 2007), `http://www.ics.uci.edu/~mlearn/`
5. Borgelt, C.: An implementation of the fp-growth algorithm. In: Proceedings of the 1st International Workshop on Open Source Data Mining: Frequent Pattern Mining Implementations, pp. 1–5. OSDM '05, ACM (2005)
6. Borgelt, C.: Simple algorithms for frequent item set mining. In: Koronacki, J., Raś, Z.W., Wierzchoń, S.T., Kacprzyk, J. (eds.) Advances in Machine Learning II, Studies in Computational Intelligence, vol. 263, pp. 351–369. Springer Berlin Heidelberg (2010)
7. Han, J., Fu, Y.: Discovery of multiple-level association rules from large databases. In: Dayal, U., Gray, P.M.D., Nishio, S. (eds.) VLDB, pp. 420–431. Morgan Kaufmann (1995)
8. Herawan, T., Deris, M.M.: A soft set approach for association rules mining. Knowledge-Based Systems 24(1), 186–195 (2011)
9. Moshkov, M., Chikalov, I.: On algorithm for constructing of decision trees with minimal depth. Fundam. Inform. 41(3), 295–299 (2000)
10. Moshkov, M.J.: On the class of restricted linear information systems. Discrete Mathematics 307(22), 2837–2844 (2007)
11. Moshkov, M.J., Piliszczuk, M., Zielosko, B.: Greedy algorithm for construction of partial association rules. Fundam. Inform. 92(3), 259–277 (2009)
12. Moshkov, M.J., Piliszczuk, M., Zielosko, B.: On construction of partial association rules. In: Wen, P., Li, Y., Polkowski, L., Yao, Y., Tsumoto, S., Wang, G. (eds.) RSKT, LNCS, vol. 5589, pp. 176–183. Springer (2009)
13. Nguyen, H.S., Nguyen, S.H.: Rough sets and association rule generation. Fundam. Inform. 40(4), 383–405 (1999)
14. Nguyen, H.S., Ślęzak, D.: Approximate reducts and association rules - correspondence and complexity results. In: Zhong, N., Skowron, A., Ohsuga, S. (eds.) RSFDGrC, LNCS, vol. 1711, pp. 137–145. Springer (1999)

15. Park, J.S., Chen, M.S., Yu, P.S.: An effective hash based algorithm for mining association rules. In: Carey, M.J., Schneider, D.A. (eds.) SIGMOD Conference, pp. 175–186. ACM Press (1995)
16. Pawlak, Z., Skowron, A.: Rudiments of rough sets. Inf. Sci. 177(1), 3–27 (2007)
17. Rissanen, J.: Modeling by shortest data description. Automatica 14(5), 465–471 (1978)
18. Wieczorek, A., Słowiński, R.: Generating a set of association and decision rules with statistically representative support and anti-support. Inf. Sci. 277, 56–70 (2014)
19. Zielosko, B.: Optimization of decision rules relative to coverage - comparative study. In: Kryszkiewicz, M., Cornelis, C., Ciucci, D., Medina-Moreno, J., Motoda, H., Raś, Z.W. (eds.) JRS 2014. LNCS, vol. 8537, pp. 237–247. Springer (2014)
20. Zielosko, B.: Global optimization of exact association rules relative to coverage. In: Kryszkiewicz, M., Bandyopadhyay, S., Rybiński, H., Pal, S.K. (eds.) PReMI 2015. LNCS, vol. 9124, pp. 428–437. Springer (2015)