Global Optimization of Exact Association Rules Relative to Coverage

Beata Zielosko^(✉)

Institute of Computer Science, University of Silesia,
39, Będzińska 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 coverage is presented.

Keywords: Rough sets · Dynamic programming · Coverage · Decision rules · Association rules

1 Introduction

Association rules are popular form of knowledge representation. They are used to discover patterns, associations, correlations among sets of items in data set. There are many approaches for mining association rules. The most popular, is Apriori algorithm based on frequent itemsets [1].

During years, many new algorithms were designed which are based on, e.g., closed itemset lattices [16], partitioning the data [19], and others [7, 15].

In the paper, an application of dynamic programming approach to optimization of association rules relative to coverage is presented. Coverage is one of popular rule's evaluation measure which denotes how many objects are covered by the rule. It allows one to discover major patterns in the data and also reduces the number of generated rules. Unfortunately, the problem of construction of rules with maximum coverage is *NP*-hard [6]. The most part of approaches, with the exception of brute-force and Apriori algorithm, cannot guarantee the construction of rules with the maximum coverage. The proposed approach allows one to construct optimal rules, i.e., rules with the maximum coverage.

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 [17] 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. Such rules can be interpreted as Horn rules [5, 8] where literals are coded using *attribute-value* pairs. This form of rule representation is widely used in data mining. Similar approach was considered in [13, 14], where greedy algorithm for minimization of length of association rules was investigated.

This paper is an essential extension of the paper [3] in which decision tables are considered that do not contain equal rows with different decisions (consistent tables). When association rules for information systems are studied and each attribute is sequentially considered as the decision one, inconsistent tables are often obtained, i.e., tables containing equal rows with different decisions. 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.

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 maximum coverage that cover given object (row). To this end, a directed acyclic graph is constructed. Based on this graph it is possible to describe all irredundant association rules with given attribute on the right-hand side. This graph is used also in an optimization procedure which changes the initial graph such that it describes, for each row, all association rules with the considered attribute on the right-hand side that cover given row, and have maximum cardinality.

Methods and results considered in this paper have common points with Formal Concept Analysis (FCA) [18]. In particular, for a given decision table T , a directed acyclic graph is constructed. It is used for the description and optimization of association rules. Nodes of this graph are so-called separable subtables of the table T given by systems of equations of the kind “attribute=value”. The considered construction is connected with FCA. Let the decision table T contain only conditional attributes with values from the set $\{0, 1\}$ and, for each attribute, contain also its negation. Then each subtable (node) from the directed acyclic graph corresponds to a concept for the context consisting of the set of rows of T and the set of conditional attributes of T . More common points are in results. In particular, exact association rules (called “implications”) are studied in FCA [9, 10, 15, 16].

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

2 Main Notions

An *information system* I is a rectangular table with $n + 1$ columns labeled with attributes $f_1, \dots, f_{n+1}$. 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

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

where $f_j \in \{f_1, \dots, f_{n+1}\}$, $f_{i_1}, \dots, f_{i_m} \in \{f_1, \dots, f_{n+1}\} \setminus \{f_j\}$, and $a, a_1, \dots, a_m$ 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 rectangular table with n columns labeled with (conditional) attributes $f_1, \dots, f_n$. 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 $f_i \in \{f_1, \dots, f_{n+1}\}$, 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 $f_1, \dots, f_{i-1}, f_{i+1}, \dots, f_{n+1}$ 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} .

The set $\{I_{f_1}, \dots, I_{f_{n+1}}\}$ 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 $f_{i_1}, \dots, f_{i_m} \in \{f_1, \dots, f_n\}$ and $a_1, \dots, a_m$ be nonnegative integers. A subtable of the table T , which contains only rows of T that have numbers $a_1, \dots, a_m$ at the intersection with columns $f_{i_1}, \dots, f_{i_m}$, is denoted by $T(f_{i_1}, a_1) \dots (f_{i_m}, a_m)$. Such subtables (including the table T) are called *separable subtables* of T .

The set of attributes from $\{f_1, \dots, f_n\}$ 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

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

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

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

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

Rule (1) will be called *true for T* if the table $T' = T(f_{i_1}, a_1) \dots (f_{i_m}, a_m)$ 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 , $i = 1, \dots, n + 1$. The union of sets of

(f_i) -association rules, $i = 1, \dots, n+1$, 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 $f_{i_j} \in E(T(f_{i_1}, a_1) \dots (f_{i_{j-1}}, a_{j-1}))$ for $j = 2, \dots, m$;
- (ii) if $m = 1$ then the table T is nondegenerate, and if $m > 1$ then the table $T(f_{i_1}, a_1) \dots (f_{i_{m-1}}, a_{m-1})$ 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 *coverage* of τ is the number of rows in T for which τ is realizable and which are labeled with the decision d . It is denoted by $c(\tau)$. If $m = 0$ then the coverage of the rule τ is equal to the number of rows in T which are labeled with the decision d .

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} , $i = 1, \dots, n+1$, 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 $E(\Theta, f_i) = \{b_1, \dots, b_t\}$. Then draw t edges from Θ and label these edges with pairs $(f_i, b_1), \dots, (f_i, b_t)$ respectively. These edges enter to nodes $\Theta(f_i, b_1), \dots, \Theta(f_i, b_t)$. If some of nodes $\Theta(f_i, b_1), \dots, \Theta(f_i, b_t)$ 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, i.e., Θ is degenerate.

Later, a local optimization of the graph $\Delta(T)$ relative to the coverage 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 $E(\Theta)$.

Now, for each node Θ of $G(T)$ and for each row r of Θ , a set of rules $Rul_{G(T)}(\Theta, r)$ will be described. It is possible that $G(T) = \Delta(T)$.

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

$$Rul_{G(T)}(\Theta, r) = \{\rightarrow f_{n+1} = d\}.$$

Let now Θ be a nonterminal node of $G(T)$ such that for each child Θ' of Θ and for each row r' of Θ' , the set of rules $Rul_{G(T)}(\Theta', r')$ is already defined. Let $r = (b_1, \dots, b_n)$ be a row of Θ . For any $f_i \in E_{G(T)}(\Theta, r)$, the set of rules $Rul_{G(T)}(\Theta, r, f_i)$ is defined as follows:

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

Then $Rul_{G(T)}(\Theta, r) = \bigcup_{f_i \in E_{G(T)}(\Theta, r)} Rul_{G(T)}(\Theta, r, f_i)$.

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} , $i = 1, \dots, n + 1$, obtained from the information system I . In general, the obtained graph is denoted by $\Delta(I_{f_i})$, $i = 1, \dots, n + 1$. As a result, for $i = 1, \dots, n + 1$, a 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 , $i = 1, \dots, n + 1$. The union of sets $Rul_{\Delta(I_{f_i})}(I_{f_i}, r)$ forms a set $Rul(I, r)$ of irredundant association rules for I and r :

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

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.

$$Rul_{\Delta(I_{f_1})}(I_{f_1}, r_1) = \{f_2 = 2 \wedge f_3 = 0 \rightarrow f_1 = 0, f_3 = 0 \rightarrow f_1 = 0\};$$

$$Rul_{\Delta(I_{f_1})}(I_{f_1}, r_2) = \{f_2 = 2 \wedge f_3 = 1 \rightarrow f_1 = 1, f_3 = 1 \rightarrow f_1 = 1\};$$

$$Rul_{\Delta(I_{f_1})}(I_{f_1}, r_3) = \{f_2 = 1 \rightarrow f_1 = 1, f_3 = 1 \rightarrow f_1 = 1\}.$$

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 . $Rul(I, r_1) = \{f_2 = 2 \wedge f_3 = 0 \rightarrow f_1 = 0, f_3 = 0 \rightarrow f_1 = 0, f_1 = 0 \rightarrow f_2 = 2, f_3 = 0 \rightarrow f_2 = 2, f_1 = 0 \rightarrow f_3 = 0, f_2 = 2 \wedge f_1 = 0 \rightarrow f_3 = 0\}$;

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

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

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

$Rul(I, r_2) = \{f_2 = 2 \wedge f_3 = 1 \rightarrow f_1 = 1, f_3 = 1 \rightarrow f_1 = 1, f_1 = 1 \rightarrow f_2 = 1, f_3 = 1 \rightarrow f_2 = 1, f_1 = 1 \rightarrow f_3 = 1, f_2 = 2 \wedge f_1 = 1 \rightarrow f_3 = 1\};$
 $Rul(I, r_3) = \{f_2 = 1 \rightarrow f_1 = 1, f_3 = 1 \rightarrow f_1 = 1, f_1 = 1 \rightarrow f_2 = 1, f_3 = 1 \rightarrow f_2 = 1, f_1 = 1 \rightarrow f_3 = 1, f_2 = 1 \rightarrow f_3 = 1\}.$

4 Optimization Relative to Coverage

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

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

Let $T = I_{f_{n+1}}$. Now, a procedure of local optimization of the graph $\Delta(T)$ relative to the coverage will be described. For each node Θ in the graph $\Delta(T)$, this procedure assigns to each row r of Θ the set $Rul_{\Delta(T)}^c(\Theta, r)$ of decision rules with the maximum coverage from $Rul_{\Delta(T)}(\Theta, r)$ and the number $Opt_{\Delta(T)}^c(\Theta, r)$ – the maximum coverage 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 each row r of Θ has assigned the number $Opt_{\Delta(T)}^c(\Theta, r)$ that is equal to the number of rows in Θ which are labeled with the most common decision for Θ .

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

$$Opt_{\Delta(T)}^c(\Theta, r) = \max\{Opt_{\Delta(T)}^c(\Theta(f_i, b_i), r) : f_i \in E(\Theta, r)\}$$

is assigned to the row r in the table Θ and

$$E_{G(T)}(\Theta, r) = \{f_i : f_i \in E_{\Delta(T)}(\Theta, r), Opt_{\Delta(T)}^c(\Theta(f_i, b_i), r) = Opt_{G(T)}^c(\Theta, r)\}.$$

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 $Rul_{\Delta(T)}^c(\Theta, r)$ of all rules with the maximum coverage from the set $Rul_{\Delta(T)}(\Theta, r)$.

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

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

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

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

$$Opt^c(I, r) = \max\{Opt_{\Delta(I_{f_i})}^c(I_{f_i}, r) : i = 1, \dots, n + 1\},$$

and among all numbers $Opt_{\Delta(I_{f_i})}^c(I_{f_i}, r), i = 1, \dots, n + 1$, only these are selected, where

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

The set $Rul^c(I, r)$ is the union of sets $Rul_{G(I_{f_i})}(I_{f_i}, r)$ for which the last equality holds:

$$Rul^c(I, r) = \bigcup_{Opt_{\Delta(I_{f_i})}^c(I_{f_i}, r) = Opt^c(I, r), Opt^c(I, r) = \max\{Opt_{\Delta(I_{f_i})}^c(I_{f_i}, r) : i = 1, \dots, n + 1\}} Rul_{G(I_{f_i})}(I_{f_i}, r)$$

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

Below one can find the sets of association rules, for information system I (depicted in Fig. 1) and row r , after global optimization relative to coverage, and the value $Opt^c(I, r)$.

$Rul^c(I, r_1) = \{f_2 = 2 \wedge f_3 = 0 \rightarrow f_1 = 0, f_3 = 0 \rightarrow f_1 = 0, f_1 = 0 \rightarrow f_2 = 2, f_3 = 0 \rightarrow f_2 = 2, f_1 = 0 \rightarrow f_3 = 0, f_2 = 2 \wedge f_1 = 0 \rightarrow f_3 = 0\}$, $Opt^c(I, r_1) = 1$;
 $Rul^c(I, r_2) = \{f_3 = 1 \rightarrow f_1 = 1, f_1 = 1 \rightarrow f_3 = 1\}$, $Opt^c(I, r_2) = 2$;
 $Rul^c(I, r_3) = \{f_3 = 1 \rightarrow f_1 = 1, f_1 = 1 \rightarrow f_3 = 1\}$, $Opt^c(I, r_3) = 2$.

The problem of rule coverage maximization is NP-hard [6]. 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 [11, 12] 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 $f_i \in \{f_1, \dots, f_{n+1}\}$, 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 $f_1, \dots, f_{i-1}, f_{i+1}, \dots, f_{n+1}$, was obtained. Values of the attribute f_i were attached to the rows of the obtained table I_{f_i} . The set $\{I_{f_1}, \dots, I_{f_{n+1}}\}$ of decision tables obtained from the information system I is denoted by Φ .

Table 1 presents preliminary results of experiments connected with the maximum coverage of irredundant association rules. For each row r of I , the maximum coverage of an irredundant association rule for I and r was obtained. After that, for rows of I the minimum coverage of an association rule for I and r with the maximum coverage (column “Min”), the maximum coverage of such rule (column “Max”) and the average coverage of association rules with the maximum coverage - 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 .

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 $\Delta(I_{f_i})$, $i = 1, \dots, n + 1$. For each data set I , the set Φ was obtained. For each decision table I_{f_i} , $i = 1, \dots, n + 1$, 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 related to the directed acyclic graphs $\Delta(I_{f_i})$, $i = 1, \dots, n + 1$, and data set I were computed.

The proposed approach of rule induction is based on the analysis of the directed acyclic graph constructed for a given decision table. A structure of the graph depends on data set, i.e., number of attributes, distribution of values of

Table 1. Maximum coverage of rules

Data set	Rows	Attr	Min	Avg	Max
adult-stretch	16	5	4	7.0	8
breast-cancer	266	10	5	41.143	65
cars	1728	7	1	336.319	576
hayes-roth	69	5	2	7.435	12
lenses	24	5	2	7.958	12
monks-1-test	432	7	12	45.0	108
monks-3-test	432	7	36	56.0	108
shuttle-landing	15	7	6	7.8	9
teeth	23	9	8	12.652	15
tic-tac-toe	958	10	12	67.282	90
zoo-data	59	17	31	34.814	35

Table 2. Average number of nodes and edges

Data set	Rows	Attr	Nodes	Edges
adult-stretch	16	5	48.0	104.0
breast-cancer	266	10	6082.0	61063.6
cars	1728	7	4335.3	17697.1
hayes-roth	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

attributes, number of rows. Such graph can be huge for larger data set. Therefore, possibilities of decreasing the size of the graph are studied by the author. In [20], 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 coverage was presented. It is based on the dynamic programming approach to optimization of decision rules. However, there are differences: (i) the information system is used, (ii) decision table can be inconsistent, (iii) definitions are different, and (iv) global optimization of rules was studied. The presented approach can be considered as research tool which allows one to construct association rules with maximum coverage.

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

Future works will be connected with the study of approximate association rules and sequential optimization of rules relative to the length and coverage.

Acknowledgements. The author wishes to thank the anonymous reviewers for helpful suggestions.

References

1. Agrawal, R., Imieliński, T., Swami, A.: Mining association rules between sets of items in large databases. In: SIGMOD 1993, 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: Ficarra, F.V.C. (ed.) *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, Bergamo (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. Volume 42 of Intelligent Systems Reference Library*, pp. 211–228. Springer, Heidelberg (2013)
4. Asuncion, A., Newman, D.J.: UCI Machine Learning Repository (2007). (<http://www.ics.uci.edu/~mllearn/>)
5. Balcázar, J.L., Garriga, G.C.: Horn axiomatizations for sequential data. *Theor. Comput. Sci.* **371**(3), 247–264 (2007)
6. Bonates, T., Hammer, P.L., Kogan, A.: Maximum patterns in datasets. *Discret. Appl. Math.* **156**(6), 846–861 (2008)
7. 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. SCI*, vol. 263, pp. 351–369. Springer, Heidelberg (2010)
8. Hammer, P.L., Kogan, A.: Essential and redundant rules in horn knowledge bases. *Decis. Support Syst.* **16**(2), 119–130 (1996)
9. Kuznetsov, S.O., Obiedkov, S.A.: Some decision and counting problems of the duquenne-guigues basis of implications. *Discret. Appl. Math.* **156**(11), 1994–2003 (2008)
10. Medina, R., Nourine, L.: Conditional functional dependencies: an FCA point of view. In: Kwuida, L., Sertkaya, B. (eds.) *ICFCA 2010. LNCS*, vol. 5986, pp. 161–176. Springer, Heidelberg (2010)
11. Moshkov, M., Chikalov, I.: On algorithm for constructing of decision trees with minimal depth. *Fundam. Inform.* **41**(3), 295–299 (2000)
12. Moshkov, M.J.: On the class of restricted linear information systems. *Discret. Math.* **307**(22), 2837–2844 (2007)
13. Moshkov, M.J., Piliszczyk, M., Zielosko, B.: Greedy algorithm for construction of partial association rules. *Fundam. Inform.* **92**(3), 259–277 (2009)
14. Moshkov, M.J., Piliszczyk, M., Zielosko, B.: On construction of partial association rules. In: Wen, P., Li, Y., Polkowski, L., Yao, Y., Tsumoto, S., Wang, G. (eds.) *RSKT 2009. LNCS*, vol. 5589, pp. 176–183. Springer, Heidelberg (2009)
15. Obiedkov, S.A., Duquenne, V.: Attribute-incremental construction of the canonical implication basis. *Ann. Math. Artif. Intell.* **49**(1–4), 77–99 (2007)
16. Pasquier, N., Bastide, Y., Taouil, R., Lakhal, L.: Efficient mining of association rules using closed itemset lattices. *Inf. Syst.* **24**(1), 25–46 (1999)
17. Pawlak, Z., Skowron, A.: Rudiments of rough sets. *Inf. Sci.* **177**(1), 3–27 (2007)
18. Poelmans, J., Ignatov, D.I., Kuznetsov, S.O., Dedene, G.: Formal concept analysis in knowledge processing: a survey on applications. *Expert Syst. Appl.* **40**(16), 6538–6560 (2013)
19. Savasere, A., Omiecinski, E., Navathe, S.B.: An efficient algorithm for mining association rules in large databases. In: Dayal, U., Gray, P.M.D., Nishio, S. (eds.) *VLDB*, pp. 432–444. Morgan Kaufmann (1995)
20. 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.) *RSEISP 2014. LNCS*, vol. 8537, pp. 237–247. Springer, Heidelberg (2014)