Fundamenta Informaticae 172 (2020) 1–14
DOI 10.3233/FI-2020-1890
IOS Press

Action Rules of Lowest Cost and Action Set Correlations

Angelina A. Tzacheva*, **Ramya A. Shankar, Sridharan Ramachandran, Arunkumar Bagavathi**

Department of Computer Science
University of North Carolina at Charlotte, USA
aatzache@uncc.edu, rshanka7@uncc.edu, sramac11@uncc.edu, abagavat@uncc.edu

Abstract. A knowledge discovery system is prone to yielding plenty of patterns, presented in the form of rules. Sifting through to identify useful and interesting patterns is a tedious and time consuming process. An important measure of interestingness is: whether or not the pattern can be used in the decision making process of a business to increase profit. Hence, $actionable$ patterns, such as action rules, are desirable. Action rules may suggest actions to be taken based on the discovered knowledge. In this way contributing to business strategies and scientific research.

The large amounts of knowledge in the form of rules presents a challenge of identifying the essence, the most important part, of high usability. We focus on decreasing the space of action rules through generalization. In this work, we present a new method for computing the lowest cost of action rules and their generalizations. We discover action rules of lowest cost by taking into account the correlations between individual *atomic action sets*.

Keywords: action rules, interestingness, actionable knowledge discovery, generalization, cost of action rules

1. Introduction

Data mining, or knowledge discovery, is frequently referred to in the literature as the process of extracting interesting information or patterns from large databases. There are two major directions in data mining research: patterns and interest. The pattern discovery techniques include: classification,

*Address for correspondence: 9201 University City Blvd, Charlotte, North Carolina, 28223 - USA.

Received December 2018; revised October 2019.

association, and clustering. Interest refers to the pattern applications in business, or other organizations, being useful or meaningful [1].

Since the pattern discovery techniques often generate large amounts of knowledge, they require a great deal of expert manual to post-process the mined results. Therefore, one of the central research problems in the field relates to reducing the volume of the discovered patterns, and selecting appropriate interestingness measures.

These measures are intended for selecting and ranking patterns according to their potential interest to the user. Good measures also allow the time and space costs of the mining process to be reduced. Although much work has been conducted in this area, so far there is no widespread agreement on a formal definition of interestingness in this context. Based on the variety of definitions presented to-date, interestingness is perhaps best treated as a broad concept that emphasizes: *conciseness, coverage, reliability, peculiarity, diversity, novelty, surprisingness, utility,* and *actionability* [2].
In this work we focus on actionability and diversity.

Actionability - an important measure of interestingness is: how *actionable* the patterns are, i.e. to what extent the user can act on them to his/her advantage. For instance, whether or not the pattern can be used in the decision making process of a business to increase profit. Hence, recent research focuses on making it easier for the user to grasp the significance of the mined rules, in the context of a business action plan [1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12].

An action rule, provides hints to a business user to what changes within *flexible* attributes are needed in order to re-classify customers from low profitability to high profitability class, introduced by Ras an Wieczorkowska [8]. It is assumed that attributes in a database are divided into two groups: stable and flexible. By stable we mean attributes whose values cannot be changed (age, place of birth, number of children). On the other hand, attributes (like interest rate, or loan approval) whose values can be changed or influenced are called flexible. Each action rule was originally constructed from certain pairs of classification rules. The notion of action rule was extended by Tsay and Ras in [9], and a new simplified strategy for extraction was proposed by Ras and Wyrzykowska in [5].

Diversity - a pattern is diverse if its elements differ significantly from each other, while a set of patterns is diverse if the patterns in the set differ significantly from each other [2]. Diversity is a common factor for measuring the interestingness of summaries [13].

Summaries - summarization is one of the major tasks in knowledge discovery and the key issue in online analytical processing (OLAP) systems. The essence of summarization is the formation of interesting and compact descriptions of raw data at different concept levels, which are called *summaries*. For example, sales information in a company may be summarized to levels of area, such as City, Province, and Country. It can also be summarized to levels of time, such as Week, Month, and Year.

According to a simple point of view, a summary can be considered diverse if its probability distribution is far from the uniform distribution. The more diverse a summary, the more interesting it is, because in the absence of any relevant knowledge, a user commonly assumes that the uniform distribution will hold in a summary. Little research has been done on using diversity to measure the interestingness of classification rules, association rules [2], or action rules.

For this reason Tzacheva and Ras [6] introduce the notion of a cost and feasibility of an action rule. They suggest a heuristic strategy for creating new action rules, where objects supporting the new action rule also support the initial action rule but the cost of reclassifying them is lower or even much lower for the new rule. In this way the rules constructed are of more interest to the users.

The organization of this paper is as follows: first, we review related work that has appeared in section 2; the approach of creating action rules summaries is presented in section 3; section 4 discusses the algorithm for generalization of action rule summaries; the cost and the hierarchy is described in section 5; the new method for action set correlations and action rules of lowest cost is in section 6; and, finally, in section 7 we conclude and make future work remarks.

2. Related work

In the paper by Ras and Wieczorkowska [8], the notion of an action rule was introduced. The main idea was to generate, from a database, special type of rules which basically form a hint to users showing a way to re-classify objects with respect to some distinguished attribute (called a decision attribute). Values of some of attributes, used to describe objects stored in a database, can be changed and this change can be influenced and controlled by user. However, some of these changes (for instance "profit") can not be done directly to a decision attribute. In addition, the user may be unable or unwilling to proceed with the actions.

For this reason Tzacheva and Ras [6] introduce the notion of a cost and feasibility of an action rule. They suggest a heuristic strategy for creating new action rules, where objects supporting the new action rule also support the initial action rule but the cost of reclassifying them is lower or even much lower for the new rule. In this way the rules constructed are of more interest to the users.

Extended action rules, discussed in Tsay and Ras [9], form a special subclass of action rules. They construct them by extending headers of action rules in a way that their confidence is getting increased. The support of extended action rules is usually lower than the support of the corresponding action rules.

A new simplified strategy for action rule extraction was proposed by Ras and Wyrzykowska in [5]. In that work, they no longer use pairs of classification rules, but rather "grab" the objects. In this sense the action rules are mined directly from the database.

In [7] Tzacheva and Ras combine the approaches of [5], [6], and [9] leading to an improved constraint based action rule discovery with single classification rules. The minimum support, confidence, and feasibility parameters are specified by the user to produce an action rule of desirable low cost.

Yang and Cheng [12] aim for converting individuals from an undesirable class to a desirable class. The work proposes actions to switch customers to a more desirable class. It is rooted in case-base reasoning, where typical positive cases are identified to form a small and highly representative case base. This "role model" is then used to formulate marketing actions. The notion of cost of the action is also regarded. They use 1-NN classifier, 1-cluster-centroid classifier, or SVM. Such classifiers could become inadequate for disk-resident data due to their long computational time.

Ras et al.'s work on action rules is probably the pioneer in the action rule mining [5, 6, 7, 8, 9]. The notion of actionable attribute and the stable attribute is found from the beginning of their work. In most of their methods, they use a heuristic rule discovery method first to obtain as set of rules then they use a procedure which pairs a rule which predicts the positive class with a related rule which predicts the negative class. Unlike an exhaustive method, their method can miss important rules.

In [18] Hajja introduced object-driven action rules. They are action rules extracted from information systems with temporal and object-based nature, such as: systems that contain multiple observations for each object. A typical example of an object-based system is a system of patients recording multiple visits; each patient is considered a distinct object.

Mining action rules from scratch [1, 5, 12], i.e. directly from the database without using pairs of classification rules, or a similar approach which will present an exhaustive method, would supply us with all important rules. Clearly, the space of such rules is quite huge, so a generalization technique, such as creating summaries, would provide great means for reducing the space and furnish the user with the essence of the actionable knowledge.

Tzacheva [2] introduced a generalization technique, which creates summaries of action rules, by utilizing an exhaustive method. The author provided great means for reducing the space and furnished the user with the essence of the actionable knowledge. The author also introduced the notion of diversity of action rule summaries [2].

Tzacheva [2] suggested removing the high cost rules, when creating summaries, in order to additionally decrease the space of action rules. However, did not discuss how the cost will change, or how it will be computed when creating summaries. Tzacheva [17] suggested an average approach to computing the cost, which leads to cost decrease as we go up in the generalization hierarchy.

In this work, we present a new method for computing the lowest cost of action rules and their generalizations. We discover action rules of lowest cost by taking into account the correlations between individual *atomic action sets*.

3. Summaries of action rules

3.1. Exhaustive mining

In [5] Ras and Wyrzykowska propose a new simplified strategy for constructing action rules as follows:

Let us assume that $S = (U, A_{St} \cup A_{Fl} \cup \{d\})$ is a decision system, where $d \notin A_{St} \cup A_{Fl}$ is a distinguished attribute called the decision. Assume also that $d_1 \in V_d$, where V_d is the domain of d and $x \in U$. We say that x is a d_1-object if $d(x) = d_1$. Finally, we assume that $\{a_1, a_2, ..., a_p\} \subseteq A_{Fl}$, $\{b_1, b_2, ..., b_q\} \subseteq A_{St}$, $a_{[i,j]}$ denotes a value of attribute a_i, and $b_{[i,j]}$ denotes a value of attribute b_i for any i, j, and that

$$r = [[a_{[1,1]} \wedge a_{[2,1]} \wedge ... \wedge a_{[p,1]}] \wedge [b_{[1,1]} \wedge b_{[2,1]} \wedge ..... \wedge b_{[q,1]}] \to d_1]$$

is a classification rule extracted from S supporting some d_1-objects in S. By $sup(r)$ and $conf(r)$, we mean the support and the confidence of r, respectively. Class d_1 is a preferable class and our goal is to reclassify d_2-objects into d_1 class, where $d_2 \in V_d$.

By an action rule schema $r[d2 \to d1]$ associated with r and the reclassification task $(d, d_2 \to d_1)$ we mean the following expression:

$$r[d2 \to d1] = [[a_{[1,1]} \wedge a_{[2,1]} \wedge ... \wedge a_{[p,1]}] \wedge [(b_1, \to b_{[1,1]}) \wedge (b_2, \to b_{[2,1]}) \\ \wedge ..... \wedge (b_q, \to b_{[q,1]})] \Rightarrow (d, d_2 \to d_1)].$$

In a similar way, by an action rule schema $r[\rightarrow d_1]$ associated with r and the reclassification task $(d, \rightarrow d_1)$ we mean the following expression:

$$r[\rightarrow d1] = [[a_{[1,1]} \wedge a_{[2,1]} \wedge ... \wedge a_{[p,1]}] \wedge [(b_1, \rightarrow b_{[1,1]}) \wedge (b_2, \rightarrow b_{[2,1]}) \\ \wedge ..... \wedge (b_q, \rightarrow b_{[q,1]})] \Rightarrow (d, \rightarrow d_1)].$$

The term $[a_{[1,1]} \wedge a_{[2,1]} \wedge ... \wedge a_{[p,1]}]$ built from values of stable attributes, is called the header of the action rule $r[d2 \rightarrow d1]$ and its values can not be changed.

The next step is to partition the supporting set of the action rule schemas into classes, each one generating corresponding action rule.

We adopt this strategy as the first step in our proposed method, and as an approach which allows for mining action rules from scratch [1, 5, 12], i.e. directly from the database without using pairs of classification rules. We therefore use an exhaustive method that would supply us with all important rules as a start.

3.2. Clustering

We are constructing the actions rules, by ”grabbing” supporting objects into action rules schema, directly from the database. The next step is to cluster action rules into groups, i.e. groups of rules, which are similar. Such grouping would allow us to combine the similar rules together later in the process.

We use a grid-based method, STING: STatistical INformation Grid [14]. We choose this method because of its advantage of fast processing time and its typical independence of the number of objects (scales well).

The spatial area is divided into rectangular cells. There are usually several levels of such cells, which form a hierarchical structure. Each cell at high level is partitioned into a number of smaller cells in the next lower level. Statistical information of each cell is calculated and stored beforehand. When finish examining the current layer, proceed to the next lower level. Repeat this process until the bottom layer is reached.

Let $R = \{r_1, r_2, ..., r_k\}$ be the set of all action rules discovered by ARAS [5], and $X = \bigcup_{i=1}^{k} sup(r_i)$, where $sup(r_i)$ denotes the support of rule r_i.

Running STING clustering on X produces $\{X_1, X_2, X_3, ..., X_n\}$ as its partition representing the bottom layer R_i is defined as a set of action rules which are supported by objects in X_i. It means that:

$$R_i = \{r \in R : X_i \cap sup(r) \neq \emptyset\}$$

for $i = 1, 2, ..., n$.

By covering of R we mean $\{R_i : 1 \leq i \leq n\}$.

3.3. Generalization

Generalization of the data involves replacing low-level or ”primitive” (raw) data with higher-level concepts through the use of concept hierarchies. For example, categorical attributes, like *street*, can be generalized to higher-level concepts, like *city* or *country*. Similarly, values of numerical attributes, like *age*, may be mapped to higher-level concepts, like *young*, *middle-aged*, and *senior*.

In this way, we form compact descriptions of raw data at different concept levels, which are called *summaries*. For that purpose, in this work we assume that attributes are hierarchical.

Since we have clustered the action rule space, we have ended up with n clusters, where each cluster contains a set R_n of similar rules. Next, we will generalize the attributes of these rules to create a summary, or a higher-level action rule. Each such summary will cover a certain portion of the action rule space; and, it may go outside its cluster boundary or overlap with another summary.

We perform a generalization on every attribute. Thus, if we have two action rules r_1 and r_2 and the attribute value is not equal, then we go up in the hierarchy. If we have to go up to the highest/top level, then we drop the attribute. For example:

$r_1 = [[a_{112} \wedge b_{12} \wedge c_{134}] \wedge [(e, e_{121} \rightarrow e_{123}) \wedge (f, f_{12} \rightarrow f_{13})]] \Rightarrow (d, d_1 \rightarrow d_2)$ $r_2 = [[\quad b_{13} \wedge c_{135}] \wedge [(e, e_{132} \rightarrow e_{124})]] \Rightarrow (d, d_1 \rightarrow d_2)$ $G(r_1, r_2) = [[\quad b_1 \wedge c_{13}] \wedge [(e, e_1 \rightarrow e_{12})]] \Rightarrow (d, d_1 \rightarrow d_2)$

where $G(r_1, r_2)$ is the generalization, or the summary, of r_1 and r_2.

4. Algorithm for generalization of action rules to create summaries

We adopt the following algorithm from Tzacheva [17] which discovers summaries by maximizing the diversity [17] and minimizing the cost. The summaries fall within the user defined confidence threshold.

Algorithm:

$$\Delta(R_i) := \{\Delta(r_1, r_2) : r_1, r_2 \in R_i\}$$

$$for\ every\ i, j\ mark(r_i, r_j) := 0$$

$$\{$$

$$(r_i, r_j) := (r_1, r_2)\ where\ \Delta(r_1, r_2) = MAX\ (\Delta(R_i))$$

$$if\ conf(G(r_i, r_j)) \geq \lambda$$

$$R_i := R_i \cup \{G(r_i, r_j)\}$$

$$mark(r_i, r_j) := 1$$

$$\}\ while\ mark(r_i, r_j) = 0$$

$$for\ every\ r_i$$

$$if\ [r_i, G(r_i, r_j) \notin R_i]$$

$$R_i := R_i - r_i$$

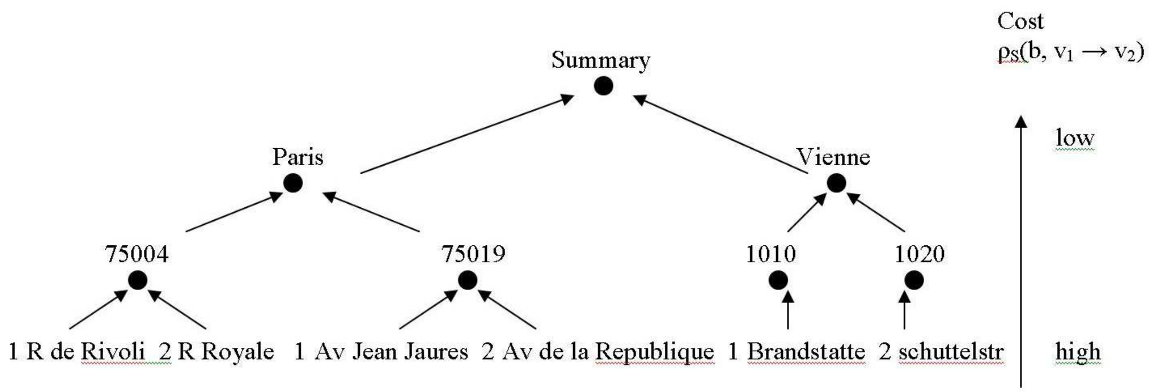

Figure 1. Generalized cost

5. Cost and the hierarchy

Typically, there is a cost associated with changing an attribute value from one class to another - more desirable one. The cost is a subjective measure, in a sense that domain knowledge from the user or experts in the field is necessary in order to determine the costs associated with taking the actions. Costs could be monetary, moral, or a combination of the two. For example, lowering the interest percent rate for a customer is a monetary cost for the bank; while, changing the marital status from 'married' to 'divorced' has a moral cost, in addition to any monetary costs which may be incurred in the process. Feasibility is an objective measure, i.e. domain independent.

According to the cost of actions associated with the classification part of action rules, a business user may be unable or unwilling to proceed with them.

The definition of cost was introduced by Tzacheva and Ras [6] as follows:

Assume that $S = (X, A, V)$ is an information system. Let $Y \subseteq X$, $b \in A$ is a *flexible* attribute in S and $v_1, v_2 \in V_b$ are its two values. By $\wp_S(b, v_1 \rightarrow v_2)$ we mean a number from $(0, \omega]$ which describes the average cost of changing the attribute value v_1 to v_2 for any of the qualifying objects in Y. These numbers are provided by experts. Object $x \in Y$ qualifies for the change from v_1 to v_2, if $b(x) = v_1$. If the above change is not feasible, then we write $\wp_S(b, v_1 \rightarrow v_2) = \omega$. Also, if $\wp_S(b, v_1 \rightarrow v_2) < \wp_S(b, v_3 \rightarrow v_4)$, then we say that the change of values from v_1 to v_2 is more feasible than the change from v_3 to v_4. Assume an action rule r of the form:

$$(b1, v_1 \rightarrow w_1) \wedge (b2, v_2 \rightarrow w_2) \wedge \ldots \wedge (bp, v_p \rightarrow w_p) \Rightarrow (d, k_1 \rightarrow k_2)$$

If the sum of the costs of the terms on the left hand side of the action rule is smaller than the cost on the right hand side, then we say that the rule r is *feasible*.

Once we have created the higher-level action rules, or the action rule summaries, we may examine the cost associated with each summary. Clearly, the summaries of low cost are more actionable, i.e. easier for the user to accomplish. Therefore, they are more interesting.

Hence, if the summary has high cost, we may disregard it as being of low interest to the user. In this way, we would further decrease the space of the mined action rules. In addition, it is possible that if the summary is not interesting, then we may make assumptions about the interestingness of the whole cluster, from which the summary was extracted. However, in order to make such determinations, the correlations of the attributes will need to be taken into consideration. Little, to no work has been done examining the correlations of attributes with action rule discovery.

The work in Tzacheva [15] suggested removing the high cost rules, when creating summaries, in order to additionally decrease the space of action rules. Tzacheva [15] computed the cost of the summaries by taking an average of the costs of the composing terms. In this way, cost decreases as we go up in the hierarchy, because some terms are removed with the generalization.

6. Action set correlations and lowest cost

In this work, we propose a new approach for discovering action rules of lowest cost by taking into account the correlations between individual *atomic action sets*.

An *atomic action set* is an expression that defines a change of attribute value for a single distinct attribute as described in Hajja [16]. For example, $(a, a_1 \rightarrow a_2)$ is an atomic action set which defines a change of value for the attribute a from a_1 to a_2, where $a_1, a_2 \in V_a$. The attribute a is a flexible attribute, since it changes its state from a_1 to a_2. In the case when there is no change, we omit the right arrow sign, so for example (b, b_1) means that the value of attribute b remains b_1, where $b_1 \in V_b$.

An *action set t* consists of all atomic action sets contained in t. For example, $t = (a, a_1 \rightarrow a_2) \wedge (b, b_1)$ is an action set that consists of two atomic action sets, namely $(a, a_1 \rightarrow a_2)$ and (b, b_1). We call that a $2 - pair$ action set. The action set can be $3 - pair$, $4 - pair$, ..., $n - pair$ set.

In this way, *action rules* are considered expressions of the form: $r = [t_1 \Rightarrow t_1]$, where t_1, t_2 are action sets. The action rule r means that: by applying the changes suggested in action set t_1, we would get, as a result, the changes in action set t_2. In other words, making t_1 happen, triggers t_2 to happen as well.

In this work, we propose a new approach for discovering action rules of lowest cost by taking into account the correlations between individual atomic action terms or sets. We start with a list of action rules, discovered from a dataset, and extract all atomic action sets from them. Next, we build a *Correlation Matrix* which shows the most frequent pairs of atomic action sets within the list of action rules as shown in Figure 2.

An atomic action set pair is said to be $frequent$ if it satisfies the minimum frequency treshold ϑ specified by the user. We scan the correlation matrix, and we *mark* all *1-pair sets* which are found to be *frequent*. Then a *2-pair* correlation matrix is built by combining the marked from the *1-pair* correlation matrix. The process is repeated with *3-pair, 4-pair, ... , n-pair* correlation matrix, until no more action set pairs are marked. See Figures 3. and 4.

If an action set pair is *frequent*, then we assume that there is a correlation between the changes which each individual atomic action set triggers. For example, consider the action rule r_1

$$r_1 \quad = [a_2 \wedge b_1] \quad \wedge [(c, c_1 \rightarrow c_2) \wedge (e, e_1 \rightarrow e_2)] \Rightarrow (d, d_1 \rightarrow d_2)$$

$e_? \rightarrow e_2$							
$e_1 \rightarrow e_2$	0						
$e_2 \rightarrow e_3$	0	0					
$f_? \rightarrow f_2$	1	0	0				
$f_3 \rightarrow f_2$	0	0	1	0			
$g_? \rightarrow g_1$	0	0	1	1	2		
$g_3 \rightarrow g_1$	1	1	0	2	0	0	
	$e_? \rightarrow e_2$	$e_1 \rightarrow e_2$	$e_2 \rightarrow e_3$	$f_? \rightarrow f_2$	$f_3 \rightarrow f_2$	$g_? \rightarrow g_1$	$g_3 \rightarrow g_1$

Figure 2. 1-Pair Correlation Matrix

$(f_? \rightarrow f_2)$^$(e_? \rightarrow e_2)$							
$(g_3 \rightarrow g_1)$^$(e_? \rightarrow e_2)$	1						
$(f_3 \rightarrow f_2)$^$(e_2 \rightarrow e_3)$	0	0					
$(g_? \rightarrow g_1)$^$(e_2 \rightarrow e_3)$	0	0	1				
$(g_? \rightarrow g_1)$^$(f_? \rightarrow f_2)$	0	0	0	0			
$(g_3 \rightarrow g_1)$^$(f_? \rightarrow f_2)$	0	0	0	0	0		
$(g_? \rightarrow g_1)$^$(f_3 \rightarrow f_2)$	0	0	0	0	0	0	
	$(f_? \rightarrow f_2)$ ^ $(e_? \rightarrow e_2)$	$(g_3 \rightarrow g_1)$ ^ $(e_? \rightarrow e_2)$	$(f_3 \rightarrow f_2)$ ^ $(e_2 \rightarrow e_3)$	$(g_? \rightarrow g_1)$ ^ $(e_2 \rightarrow e_3)$	$(g_? \rightarrow g_1)$ ^ $(f_? \rightarrow f_2)$	$(g_3 \rightarrow g_1)$ ^ $(f_? \rightarrow f_2)$	$(g_? \rightarrow g_1)$ ^ $(f_3 \rightarrow f_2)$

Figure 3. 2-Pair Correlation Matrix

$(g_3 \rightarrow g_1)$^$(e_? \rightarrow e_2)$^ $(f_? \rightarrow f_2)$		
$(g_? \rightarrow g_1)$^$(e_2 \rightarrow e_3)$^ $(f_3 \rightarrow f_2)$	0	
	$(g_3 \rightarrow g_1)$ ^ $(e_? \rightarrow e_2)$ ^ $(f_? \rightarrow f_2)$	$(g_? \rightarrow g_1)$ ^ $(e_2 \rightarrow e_3)$ ^ $(f_3 \rightarrow f_2)$

Figure 4. 3-Pair Correlation Matrix

If the 2-pair action set $[(c, c_1 \rightarrow c_2) \wedge (e, e_1 \rightarrow e_2)]$ is *frequent*, then the changes which occur when $(c, c_1 \rightarrow c_2)$ are considered correlated to the changes which occur when $(e, e_1 \rightarrow e_2)$. In other words, if $(c, c_1 \rightarrow c_2)$ happens, it is very likely to trigger $(e, e_1 \rightarrow e_2)$ to happen as well.

For this reason, if we would like to calculate the lowest cost of achieving our desired change in the class of the decision attribute d or $(d, d_1 \rightarrow d_2)$, and if we have a *frequent* action set pair in the action rule, then we can consider only the atomic action set of the lowest cost within the pair. Because paying the cost to make the changes of one atomic action set to occur, would very likely trigger the changes in the correlated atomic action terms to occur as well.

A cost $\wp_S(b, v_1 \rightarrow v_2)$ is assigned to each atomic action set, which is a number from $(0, \omega]$ describing the average cost of changing the attribute value for b from v_1 to v_2. The cost for each atomic action set is specified by an expert and recorded in an input file.

7. Experiment and results

We conducted experiments on the developed algorithms to find low cost action rules over the Mammographic mass dataset provided in [19]. Datasets provided in [19] are publicly available Machine Learning Repository generated by Department of Information and Computer Science of the University of California, Irvine. Mammography is the most effective method for breast cancer screening. The

```
BI-RADS(5->5)=>(1->0)
Shape(3->1)=>(1->0)
Shape(2->1)=>(1->0)
Density(3->2)=>(1->0)
Margin(3->1)=>(1->0)
Margin(2->3)=>(1->0)
Margin(2->1)=>(1->0)
BI-RADS(4->4),Shape(3->2)=>(1->0)
BI-RADS(4->4),Shape(3->1)=>(1->0)
Shape(3->2),Density(3->3)=>(1->0)
Shape(3->1),Density(3->3)=>(1->0)
BI-RADS(4->4),Density(3->2)=>(1->0)
Shape(3->2),Density(3->2)=>(1->0)
Shape(3->2),Margin(3->3)=>(1->0)
Shape(4->2),Margin(3->3)=>(1->0)
BI-RADS(4->4),Margin(3->1)=>(1->0)
Shape(3->3),Margin(3->1)=>(1->0)
Shape(3->2),Margin(3->1)=>(1->0)
Shape(3->1),Margin(3->1)=>(1->0)
Shape(4->3),Margin(3->1)=>(1->0)
Shape(4->2),Margin(3->1)=>(1->0)
Shape(4->1),Margin(3->1)=>(1->0)
Density(3->3),Margin(2->3)=>(1->0)
```

Figure 5. Action Rules for Mammographic mass Dataset

Mammographic mass dataset is to measure the severity of breast cancer based on BI-RADS (measures how severe the breast cancer), patients age, shape and density of the cancer. We consider that Severity of breast cancer (0-Less severe or 1-Highly severe) attribute as the decision attribute. We also consider that we need to transfer from 1(Highly Severe) to 0(Less severe). Figure 5 shows sample action rules taken into consideration for finding low cost action rules.

We randomly initialized the cost for each and every atomic action terms from the dataset. Say C be the set containing costs for all atomic action terms. After summarizing and finding correlation matrices for action rules, we obtain action sets that occurs most frequently together. We have set threshold as 2, to determine most frequently occurring action sets. From such correlation matrices and costs C, we obtained patterns as shown in the Figure 6 that we need to change to least cost combination when such patterns occur in the action rules.

```
Shape(3->2),BI-RADS(4->4)=====>BI-RADS(4->4)
Shape(3->1),BI-RADS(4->4)=====>BI-RADS(4->4)
Shape(4->3),BI-RADS(4->4)=====>Shape(4->3)
Shape(4->2),BI-RADS(4->4)=====>BI-RADS(4->4)
Shape(4->1),BI-RADS(4->4)=====>Shape(4->1)
Density(3->3),Shape(3->1)=====>Density(3->3)
Density(3->3),Shape(4->2)=====>Density(3->3)
Density(3->3),Shape(4->1)=====>Density(3->3)
Density(3->2),BI-RADS(4->4)=====>Density(3->2)
Density(3->2),Shape(3->2)=====>Density(3->2)
Density(3->2),Shape(4->2)=====>Density(3->2)
Density(3->?),BI-RADS(4->4)=====>Density(3->?)
```

Figure 6. Patterns from correlation matrices and Cost C

A pattern in Figure 6 specifies that *If this pattern occurs in the action rule =====¿ change it to this pattern* which is of low cost. We are converting all frequent action set pairs into single action term because cost to do one change would very like trigger changes in the correlated atomic action terms as well. When applying such idea on Action Rules generated from Mammographic mass dataset, we extract low cost action rules recommendations. Figure 7 shows some low cost action rules recommendations.

```
BI-RADS(5->5),Margin(5->4)=>(1->0) [Cost: 67]
BI-RADS(4->4),Margin(5->4)=>(1->0) [Cost: 130]

BI-RADS(4->4),Density(3->2)=>(1->0) [Cost: 91]
       =====> Density(3->2)=>(1->0) [Cost: 24]

Shape(3->2),Density(3->2)=>(1->0) [Cost: 202]
       =====> Density(3->2)=>(1->0) [Cost: 24]

Shape(4->2),Density(3->2)=>(1->0) [Cost: 103]
       =====> Density(3->2)=>(1->0) [Cost: 24]
```

Figure 7. Low Cost Action Rules Recommendations

8. Conclusions and future work

In this work, we present a new method for discovering action rules of lowest cost by taking into account the correlations between individual atomic action terms or sets. We suggest employing the method with short descriptions of action rules or summaries, and the use of hierarchical attributes.

The generalization algorithm used produces summaries by maximizing the diversity of rule pairs, and minimizing the cost of the suggested actions. We, therefore, provide means for reducing the volume of the mined results, and supply the user with short general descriptions of high interest actionable knowledge.

Diversity is a major criterion for measuring summaries, but no work has been done so far to study the diversity of either association, or classification [2] rules. Tzacheva [15] was the first to use the notion of diversity with action rules summaries. We use this notion within the generalization process in a sense that if two low-level action rules have a big overlap (low diversity), then they may be combined together to produce a high-level rule, or a summary. We merge the pair of two rules, which has the maximum diveristy.

With respect to the cost of action rules, the summaries of low cost are more actionable, i.e. easier for the user to accomplish, and therefore more interesting. Tzacheva [15] disregarded the ones with high cost as being of low interest to the user. Tzacheva [17] suggested an average approach to computing the cost, which leads to cost decrease as we go up in the generalization hierarchy.

In this work, we build a *correlation matrix*, which identifies pairs of individual atomic action sets, which are $frequent$ and marks them. If an action set pair is *frequent*, then there is a correlation between the changes which each individual atomic action set triggers. By using such correlated sets, we can consider only the atomic action set of the lowest cost within the pair, which significantly decreases the cost of achieving the desired change within the decision attribute.

Directions for the future include, further work with action rules in order to study the degree with which a suggested action succeeds in changing the class to a more desirable one; or, the prediction of unexpected effects/causes, which may occur after the action has been performed.

With respect to summaries, future work may employ a more generic approach for creating summaries, which would allow for using non-hierarchical attributes as well. For instance, taking intervals with numerical values, or a subset for non-numerical ones. Clearly, the effect on the precision and recall of summaries needs to be taken into consideration in such case.

The scalability issue with the number of action rules produced, may be dealt with by introducing user-defined thresholds on support and confidence. Semantic connections of the rules with the diversity measured require further studies; as well as, the computational complexity of the algorithm.

Applicable fields are: business, financial, medical, industrial.

References

[1] He Z, Xu X, Deng S, and Ma R. *Mining action rules from scratch*, Expert Systems with Applications, 2005. 29(3): 691-699. URL `https://doi.org/10.1016/j.eswa.2005.04.031`.

[2] Geng L, and Hamilton HJ. *Interestingness measures for data mining: a survey*, ACM Computing Surveys, 2006. 38(3): Article 9. doi:10.1145/1132960.1132963.

[3] Yang Q, Yin J, Lin C, and Chen T. *Postprocessing decision trees to extract actionable knowledge*, In: Proceedings of ICDM'03. 2003. doi:10.1109/ICDM.2003.1251008.

[4] Wang K, Zhou S, and Han J. *Profit Mining: From Patterns to Actions*, In: Proceedings of the 8th International Conference on Extending Database Technology: Advances in Database Technology. 2002 pp. 70–87. doi:10.1007/3-540-45876-X_7.

[5] Ras Z, Wyrzykowska E, and Wasyluk H. *ARAS: Action Rules discovery based on Agglomerative Strategy*, In: Post-Proceedings of 2007 ECML / PKDD Third International Workshop on Mining Complex Data (MCD 2007). 2007 pp. 196–208. doi:10.1007/978-3-540-68416-9_16.

[6] Tzacheva AA, and Ras ZW. *Action rules mining*. International Journal Of intelligent Systems, 2005. 20(6):719–736. URL `https://doi.org/10.1002/int.20092`.

[7] Tzacheva AA, and Ras ZW. *Constraint Based Action Rule Discovery with Single Classification Rules*, In: Proceedings of 2007 Joint Rough Set Symposium (JRS07) Conference on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing. 2007 pp. 322–329. doi:10.1007/978-3-540-72530-5_38.

[8] Ras Z, and Wieczorkowska A. *Action Rules: how to increase profit of a company*, In: Principles of Data Mining and Knowledge Discovery, (Eds. D.A. Zighed, J. Komorowski, J. Zytkow), Proceedings of PKDD'00, Lyon, France, LNAI, No. 1910, Springer, 2000 pp. 587–592. doi:10.1007/3-540-45372-5_70.

[9] Tsay L-S, and Ras ZW. *Action rules discovery system DEAR, method and experiments*, Journal of Experimental and Theoretical Artificial Intelligence, Taylor and Francis, 2005. 17(1-2):119–128. URL `https://doi.org/10.1080/09528130512331315855`.

[10] Dardzinska A, and Ras ZW. *Cooperative discovery of interesting action rules*, In: Proceedings of the Seventh International Conference on Flexible Query Answering Systems (FQAS). 2006 pp. 489–497. doi:10.1007/11766254_41.

[11] Jiang Y, Wang K, Tuzhilin A, and Fu AW-C. *Mining patterns that respond to actions*, In: Proceedings of the Fifth IEEE International Conference on Data Mining (ICDM'05). 2005. doi:10.1109/ICDM.2005.99.

[12] Yang Q, and Cheng H. *Mining case bases for action recommendation*, In: Proceedings of the IEEE International Conference on Data Mining (ICDM'02). 2002. doi:10.1109/ICDM.2002.1183997.

[13] Hilderman RJ, and Hamilton HJ. *Knowledge discovery and measures of interest*, Kluwer Academic. 2001. ISBN:0792375076.

[14] Wang K, Yang J, and Muntz R. *STING: A statistical information grid approach to spatial data mining*, In: Proceedings of 23rd International Conference on Very Large Data Bases (VLDB'97). 1997 pp. 186–195. ISBN:1-55860-470-7.

[15] Tzacheva AA. *Diversity of Summaries for Interesting Action Rule Discovery*, In: Proceedings of Intelligent Information Systems (IIS 2008). 2008.

[16] Hajja A, Wieczorkowska A, Ras ZW, and Gubrynowicz R. *Pair-based Object-driven Action Rules*, In: New Frontiers in Mining Complex Patterns, NFMCP 2012, ECML/PKDD Workshop, Bristol, United Kingdom, September, 2012, LNAI 7765, Springer, 2013 pp.79–93. doi:10.1007/978-3-642-37382-4_6.

[17] Tzacheva AA. *Algorithm for Generalization of Action Rules to Summaries*, in Special issue on Intelligent Information Processing and Web Mining, International Journal of Control and Cybernetics (invited paper), Klopotek, M. A., Przepirkowski A., Wierzchon S. T. (Eds), Systems Research Institute of Polish Academy of Sciences, 2010. 39(2):457–468.

[18] Hajja A, Ras ZW, Wieczorkowska A. *Hierarchical Object-driven Action Rules*, in Special issue on Mining Complex Patterns, A. Appice et al (Eds), Journal of Intelligent Information Systems, Springer, 2014. 42(2):207–232. doi:10.1007/s10844-013-0291-2.

[19] Lichman M. *UCI Machine Learning Repository*. Irvine, CA: University of California, School of Information and Computer Science, 2013. URL `http://archive.ics.uci.edu/ml`.