

Algorithm for generalization of action rules to summaries*

by

Angelina A. Tzacheva

Department of Informatics, University of South Carolina Upstate
Spartanburg, SC 29303, USA
e-mail: ATzacheva@USCUpstate.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 paper, we propose an improved method for discovering short descriptions of action rules. The new algorithm produces summaries by maximizing the diversity of rule pairs, and minimizing the cost of the suggested actions.

Keywords: action rules, interestingness, actionable knowledge discovery, summaries, diversity, generalization, flexible attributes, 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, association, and clustering. Interest refers to pattern applications in business, or other organizations, being useful or meaningful (He et al., 2005).

Since pattern discovery techniques often generate large amounts of knowledge, they require a great deal of expert manual work to post-process the mined

*Submitted: December 2008; Accepted: September 2009.

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 done 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* (Geng and Hamilton, 2006).

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 (He et al., 2005; Yang et al., 2003; Wang, Zhou, and Han, 2002; Ras, Wyrzykowska and Wasyluk, 2007; Tzacheva and Ras, 2005, 2007; Ras and Wiczorkowska, 2000; Tsay and Ras, 2005; Dardzinska and Ras, 2006; Jiang et al., 2005; Yang and Cheng, 2002).

An action rule provides hints to a business user as to what changes within *flexible* attributes are needed in order to re-classify customers from low profitability to high profitability class, see Ras and Wiczorkowska (2000). 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 (2005), and a new simplified strategy for extraction was proposed by Ras, Wyrzykowska and Wasyluk (2007).

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 (Geng and Hamilton, 2006). Diversity is a common factor for measuring the interestingness of summaries (Hilderman and Hamilton, 2001).

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, called *summaries*. For example, sales information in a company may be summarized at different territorial levels, like as City, Province, and Country. It can also be summarized at the 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. We are unaware of any existing research on using diversity to measure the interestingness of classification rules, association rules (Geng and Hamilton, 2006), or action rules.

The organization of this paper is as follows: first, we review related work in Section 2; the approach to creating action rule summaries is presented in Section 3; Section 4 discusses the notion of diversity; the cost and the hierarchy are described in Section 5; the new algorithm for generalization of action rule summaries is given in Section 6; and, finally, in Section 7 we conclude and make future work remarks.

2. Related work

In Ras and Wieczorkowska (2000), the notion of action rule was introduced. The main idea was to generate, from a database, special type of rules which basically form a hint to the 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 the 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 (2005) introduced the notion of 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 higher interest to the users.

Extended action rules, discussed in Tsay and Ras (2005), form a special subclass of action rules. They are constructed by extending headers of action rules in a way that their confidence is 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 in Ras, Wyrzykowska and Wasyluk (2007). In that work, the authors no longer use pairs of classification rules, but rather “grab” the objects. In this sense the action rules are mined directly from the database.

Tzacheva and Ras (2007) combine the approaches of Ras, Wyrzykowska and Wasyluk (2007), Tzacheva and Ras (2005) and Tsay and Ras (2005), proposing 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 (2002) aim at 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 considered. 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.

The work of Ras and associates on action rules is probably pioneering in the action rule mining (Ras, Wyrzykowska and Wasyluk, 2007; Tzacheva and Ras, 2005, 2007; Ras and Wierzchowska, 2000; Tsay and Ras, 2005). The notions of actionable and stable attributes can be found from the beginning of their work. In most of their methods, they use a heuristic rule discovery method first to obtain a set of rules, then they use a procedure, which pairs a rule predicting the positive class with a related rule predicting the negative class. Unlike an exhaustive method, their method can miss important rules.

Mining action rules from scratch (He et al., 2005; Ras, Wyrzykowska and Wasyluk, 2007; Yang and Cheng, 2002), 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 significant help by reducing the space and furnish the user with the essence of the actionable knowledge.

Tzacheva (2008) introduced a generalization technique, which creates summaries of action rules, by utilizing an exhaustive method. The author also introduced in Tzacheva (2008) the notion of diversity of action rule summaries.

In this work, we present a new algorithm for generalization of action rules. It maximizes the diversity of rule pairs, and minimizes the cost of the suggested actions.

Tzacheva (2008) suggested removing the high cost rules, when creating summaries, in order to additionally shrink the space of action rules. However, the author did not discuss how the cost will change, or how it will be computed when creating summaries. In this work, we suggest an average approach to computing the cost, which leads to cost decrease as we go up in the generalization hierarchy.

3. Summaries of action rules

3.1. Exhaustive mining

Ras, Wyrzykowska and Wasyluk (2005) 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 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 $\{a_1, a_2, \dots, a_p\} \subseteq A_{Fl}$, $\{b_1, b_2, \dots, b_q\} \subseteq A_{St}$, $a_{[i,j]}$ denotes a value of attribute a_i , and $b_{[i,j]}$ denotes a value of attribute b_i for any i, j , and that

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

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 $r[d_2 \rightarrow d_1]$ associated with r and the reclassification task $(d, d_2 \rightarrow d_1)$ we mean the following expression:

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

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:

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

The term $[a_{[1,1]} \wedge a_{[2,1]} \wedge \dots \wedge a_{[p,1]}]$ built out of values of stable attributes, is called the header of the action rule $r[d_2 \rightarrow d_1]$ 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 (He et al., 2005; Ras, Wyrzykowska and Wasyluk, 2007; Yang and Cheng, 2002), 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 (Wang, Yang and Muntz, 1997). We choose this method because of its advantage of fast processing and its typical independence of the number of objects (good scalability).

The spatial area is divided into rectangular cells. There are usually several levels of such cells, forming a hierarchical structure. Each cell at a high level

is partitioned into a number of smaller cells at the next lower level. Statistical information of each cell is calculated and stored beforehand. Upon termination of examining the current layer, we proceed to the next lower level. We repeat this process until the bottom layer is reached.

Let $R = \{r_1, r_2, \dots, r_k\}$ be the set of all action rules discovered by ARAS (Ras, Wyrzykowska and Wasyluk, 2007), and $X = \bigcup_{i=1}^k \text{sup}(r_i)$, where $\text{sup}(r_i)$ denotes the support of rule r_i .

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

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

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

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 extend outside its cluster boundary or overlap with another summary.

We will 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:

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

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

4. Diversity of summaries

A pattern, or a set of patterns, is diverse if its elements (or patterns in the set) differ significantly from each other. Diversity has been used as factor for measuring the interestingness of summaries.

Statistically, 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 a user would commonly assume the uniform distribution. We are unaware of any existing research on using diversity to measure the interestingness of classification rules, or association rules (Geng and Hamilton, 2006).

Tzacheva (2008) was the first to use diversity for action rules. The notion of diversity of action rule summaries was introduced by Tzacheva (2008) as follows:

If we have two higher-level action rules r_1 and r_2 , which have a big overlap in terms of the conditional part, then we say that the diversity Δ of r_1 with respect to r_2 is low. Analogically, if the two action rules have a small overlap in their domains, then we say that the diversity Δ of r_1 with respect to r_2 is high.

The diversity $\Delta(r_1, r_2)$ is defined as: the number of different atomic terms including their positions, between the two rules. For example:

$$\begin{aligned} r_1 &= [a_2 \wedge b_1] \quad \wedge [(c, c_1 \rightarrow c_2) \wedge (e, e_1 \rightarrow e_2)] \Rightarrow (d, d_1 \rightarrow d_2) \\ r_2 &= [a_1 \quad \wedge e_1] \wedge [(c, c_1 \rightarrow c_2) \wedge (f, f_1 \rightarrow f_2)] \Rightarrow (d, d_1 \rightarrow d_2) \\ \Delta(r_1, r_2) &= \quad 3 \quad + \quad 0 \quad + \quad 2 \quad = \quad 5. \end{aligned}$$

The same concept of action rule diversity can be applied to low-level action rules (non-generalized). In that sense, if two action rules have a big overlap (or low diversity), then they may be combined together to produce a generalized, or high-level action rule.

When the overlap is large, we are more safe to perform the generalization (combining of two rules). However, we would prefer a risky small overlap, because of diversity. Since the diversity is higher with small overlap, then the generalization we create will be more interesting.

Some overlaps will be ruled out as “bad”, since with generalization it is possible that objects may shift into a class different from the original. Consider the following example (for clarity the action part is omitted):

$$\begin{aligned} r_3 &= \quad [(b_3 \wedge e_2)] \Rightarrow d_1 \\ r_4 &= [a_1 \wedge c_2] \wedge [(b_1 \wedge e_2)] \Rightarrow d_1 \\ G(r_3, r_4) &= \quad [(b \wedge e_2)] \Rightarrow d_1. \end{aligned}$$

At the same time we have another rule r_5 which implies a different class - d_2 :

$$r_5 = \quad [(b_2 \wedge e_2)] \Rightarrow d_2.$$

If we generalize b_3 and b_1 into b , as shown in $G(r_3, r_4)$, then b also covers b_2 from the rule r_5 , which, however, has a different decision part, i.e. implies d_2 .

It means that if we perform the generalization $G(r_3, r_4)$, some objects will shift into a class different from the intended one. Therefore, we cannot combine the rules r_3 and r_4 together.

5. Cost and hierarchy

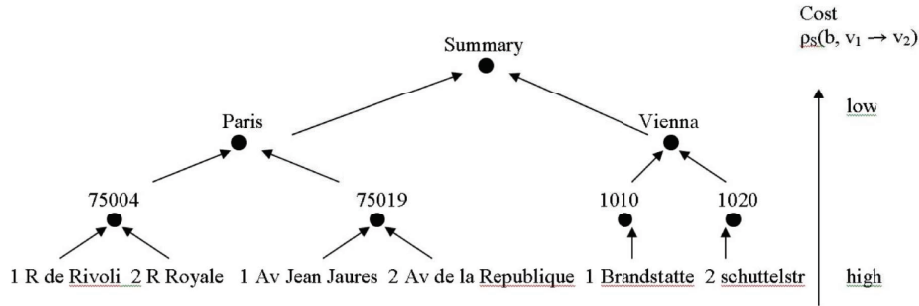


Figure 1. Generalized cost

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 the sense that domain knowledge from the user or expert 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 rate on a credit 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 (2005) as follows:

Assume that $S = (X, A, V)$ is an information system. Let $Y \subseteq X$, $b \in A$ be a *flexible* attribute in S and $v_1, v_2 \in V_b$ its two values. By $\varphi_S(b, v_1 \rightarrow v_2)$ we mean a number from $(0, \omega]$, corresponding to 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 $\varphi_S(b, v_1 \rightarrow v_2) = \omega$. Also, if $\varphi_S(b, v_1 \rightarrow v_2) < \varphi_S(b, v_3 \rightarrow v_4)$, 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:

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

If the sum of 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 infer about the interestingness of the whole cluster, from which the summary was extracted. However, in order to arrive at such implications, the correlations of the attributes will need to be considered. Little or no work has been done in examining the correlations of attributes with action rule discovery.

Tzacheva (2008) suggested removing high cost rules, when creating summaries, in order to additionally shrink the space of action rules. In this work, we propose computing the cost of the summaries by taking an average of costs of the composing terms. In this way, cost decreases as we go up in the hierarchy, because some terms are removed with generalization.

6. Algorithm for generalization of action rules to create summaries

The algorithm discovers summaries by maximizing diversity and minimizing cost. The summaries fall within the user defined confidence threshold.

Algorithm:

$$\begin{aligned} &\Delta(R_i) := \{\Delta(r_1, r_2) : r_1, r_2 \in R_i\} \\ &\text{for every } i, j \text{ mark}(r_i, r_j) := 0 \\ &\quad \{ \\ &\quad (r_i, r_j) := (r_1, r_2) \text{ where } \Delta(r_1, r_2) = \text{MAX}(\Delta(R_i)) \\ &\quad \text{if } \text{conf}(G(r_i, r_j)) \geq \lambda \\ &\quad \quad R_i := R_i \cup \{G(r_i, r_j)\} \\ &\quad \quad \text{mark}(r_i, r_j) := 1 \\ &\quad \} \text{ while mark}(r_i, r_j) = 0 \\ &\text{for every } r_i \\ &\quad \text{if } [r_i, G(r_i, r_j) \notin R_i] \\ &\quad \quad R_i := R_i - r_i \end{aligned}$$

7. Conclusions and future work

In this work, we present an improved method for shrinking the space of action rules, through creation of summaries, and the use of hierarchical attributes.

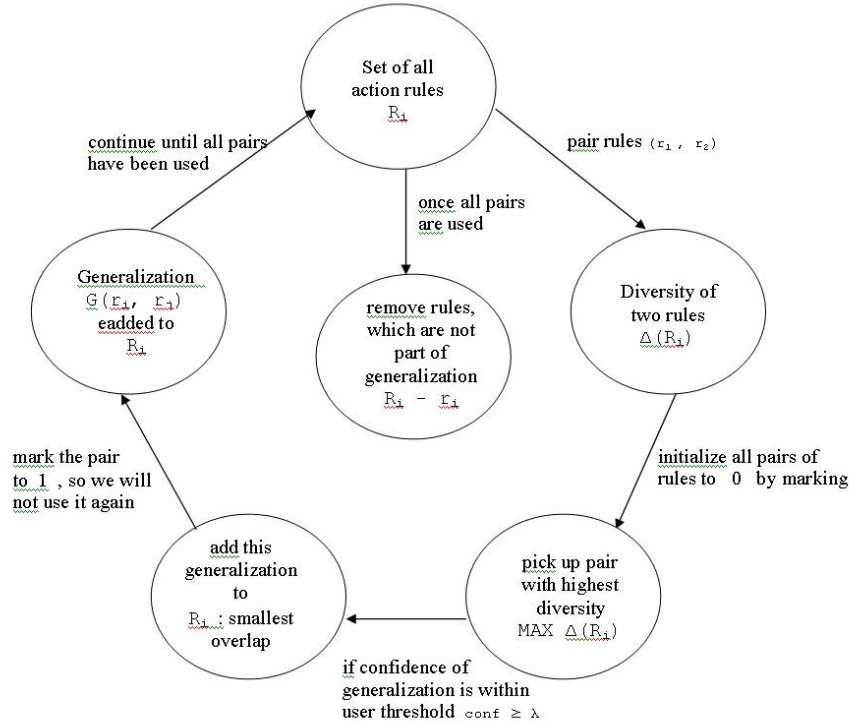


Figure 2. Algorithm for generalization of action rules

The new algorithm 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 rules (Geng and Hamilton, 2006). Tzacheva (2008) was the first to use the notion of diversity with action rule summaries. The notion is used within the generalization process in the 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. In addition, the more diverse a set of low-level action rules, the more interesting it is, since it will convey more knowledge to the user compared to a set with too many similar rules.

With the algorithm proposed in this work, we extend diversity of action rule summaries by choosing to merge the pair of rules, which has the maximum diversity.

We consider the cost of action rule summaries, as well. The summaries of low cost are more actionable, i.e. easier for the user to accomplish, and therefore more interesting. Tzacheva (2008) disregarded the ones with high cost as being of low interest to the user. In this work, we suggest an average approach to computing the cost, which leads to cost decrease as we go up in the generalization hierarchy.

Directions for the future include employing 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. Experiments with application domain data need to be performed to supplement this theoretical method.

Other future work could integrate the summarization of action rules into some of the approaches in the previous work - see Section 2, as a means for shrinking the action rule space of the mined results.

In addition, with action rules, further work is needed 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.

Application fields are: business, financial, medical and industrial.

References

- DARDZINSKA, A., and RAS, Z.W. (2006) Cooperative discovery of interesting action rules. In: H.L. Larsen, G. Pasi, D.O. Arroyo, T. Andreasen, H. Christiansen, eds., *Flexible Query Answering Systems, 7th International Conference, FQAS 2006, Milan, Italy, June 7-10, 2006*. LNCS 4027, Springer, 489–497.
- GENG, L. and HAMILTON, H.J. (2006) Interestingness measures for data mining: a survey. *ACM Computing Surveys* **38** (3).
- HE, Z., XU, X., DENG, S. and MA, R. (2005) Mining action rules from scratch. *Expert Systems with Applications* **29** (3), 691-699.
- HILDERMAN, R.J. and HAMILTON, H.J. (2001) *Knowledge Discovery and Measures of Interest*. Kluwer Academic.
- JIANG, Y., WANG, K., TUZHILIN, A. and FU, A.W.-C. (2005) Mining patterns that respond to actions. In: *Proceedings of the 5th IEEE International Conference on Data Mining (ICDM 2005), 27-30 November 2005, Houston, Texas, USA*. IEEE Computer Society, 669-672.

- RAS, Z. and WIECZORKOWSKA, A. (2000) Action Rules: how to increase profit of a company. In: D.A. Zighed, J. Komorowski, J. Zytkow, eds., *Principles of Data Mining and Knowledge Discovery*. **LNAI 1910**, Springer, 587-592.
- RAS, Z., WYRZYKOWSKA, E. and WASYLUK, H. (2007) ARAS: Action Rules discovery based on Agglomerative Strategy. In: *Post-Proceedings of 2007 ECML/PKDD Third International Workshop (MCD 2007)*, **LNAI 4944**, Springer, 196-208.
- TSAY, L.-S. and RAS, Z.W. (2005) Action rules discovery system DEAR, method and experiments. *Journal of Experimental and Theoretical Artificial Intelligence*, Taylor and Francis, **17** (1-2), 119-128.
- TZACHEVA, A.A. (2008) Diversity of Summaries for Interesting Action Rule Discovery. In: *Proceedings of 16th International Conference on Intelligent Information Systems*. Springer-Verlag, 181-290.
- TZACHEVA, A.A. and RAS, Z.W. (2005) Action rules mining. *International Journal Of Intelligent Systems* **20** (6), 719-736.
- TZACHEVA, A.A. and RAS, Z.W. (2007) Constraint Based Action Rule Discovery with Single Classification Rules. In: *Proceedings of the Joint Rough Sets Symposium (JRS07), May 14-16, 2007, in Toronto, Canada*. **LNAI 4482**, Springer, 322-329.
- WANG, K., ZHOU, S. and HAN, J. (2002) Profit Mining: From Patterns to Actions. In: C.S. Jensen et al., eds., *Proceedings of the 8th International Conference on Extending Database Technology: Advances in Database Technology*. Springer-Verlag, 70-87.
- WANG, K., YANG, J. and MUNTZ, R. (1997) STING: A statistical information grid approach to spatial data mining. In: *Proceedings of 23rd International Conference on Very Large Data Bases (VLDB'97)*. Morgan Kaufmann Publishers, 186-195.
- YANG, Q. and CHENG, H. (2002) Mining case bases for action recommendation. In: *ICDM '02: Proceedings of the 2002 IEEE International Conference on Data Mining*. IEEE Computer Society, 522.
- YANG, Q., YIN, J., LIN, C. and CHEN, T. (2003) Postprocessing decision trees to extract actionable knowledge. In: *ICDM '03: Proceedings of the Third IEEE International Conference on Data Mining*. IEEE Computer Society, 685.