Intelligent Information Systems 2008
ISBN 978-83-60434-44-4, pages 181–190

Diversity of Summaries for Interesting Action Rule Discovery

Angelina A. Tzacheva

Department of Informatics, University of South Carolina Upstate,
Spartanburg, SC 29303 USA. E-Mail: ATzacheva@USCUpstate.edu

Abstract

It has been recognized that a discovery system can generate plenty of patterns, which may be of no interest. Thus, one of the central problems in the field of knowledge discovery includes the reduction of volume with the discovered patterns, and the selection of good interesting measures. Other than general measures and domain specific measures, 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 interesting knowledge by themselves.

In this work, we present an approach which will decrease the space of action rules, by focusing on generalizing, or creating summaries. We introduce the notion of diversity of action rule summaries, and explore its application within the generalization process. Further, we evaluate the cost of the generalized rules, and additionally decrease the number of rules by removing the high cost (least useful) ones. We, therefore, discover summaries of interesting action rules.

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 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 association 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. We are unaware of any existing research on using diversity to measure the interestingness of classification rules, association rules [2], or action rules.

In this work, we present a theoretical approach which will decrease the space of action rules, by focusing on generalizing, or creating summaries. We introduce the notion of diversity of action rule summaries, and explore its application within the generalization process. Further, we evaluate the cost of the generalized rules, and additionally decrease the number of rules by removing the high cost, or least useful, ones. We, therefore, discover summaries of interesting action rules. In this way, we reduce the volume of the mined action rules, and present the user with the essence — the more general and more interesting actionable knowledge.

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 introduces the notion of diversity of action rule summaries; the cost of action rules is described in section 5; and, finally, in section 6 we conclude with discussion and 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. We 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, we 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.

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.

The notion of diversity of action rule summaries could possibly be useful during the generalization process. Finally, incorporating the cost of action rules into the generalized rules would further decrease the space of such rules, and leave only the most actionable or interesting rules.

3 Summaries of Action Rules

3.1 Mining Action Rules From Scratch

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

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

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 $r[d2 \rightarrow d1]$ associated with r and the reclassification task $(d, d_2 \rightarrow d_1)$ we mean the following expression:

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

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:

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

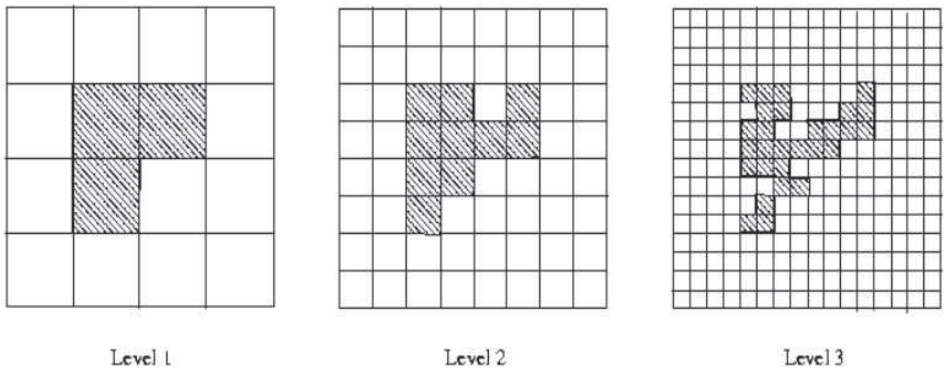

FIGURE 1: STING: A STatistical INformation Grid

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. The algorithm is illustrated on Fig. 1.

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 — Fig.2. 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 $i = 1, 2, ..., n$.

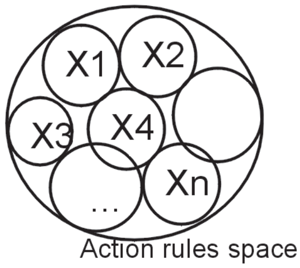

FIGURE 2: Clustered action rule space

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 — Fig. 3.a. 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 as shown on Fig. 3.b.

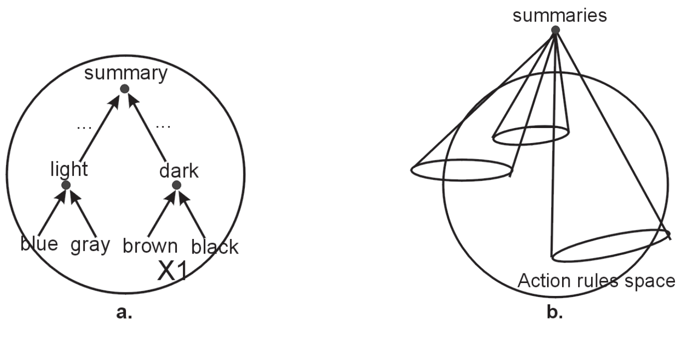

FIGURE 3: **a.** Generalizing the action rules for each cluster, example hierarchical attribute *eye color* in cluster X_1 **b.** Summaries — higher level rules

We will perform a generalization on every attribute. Thus, if we have two

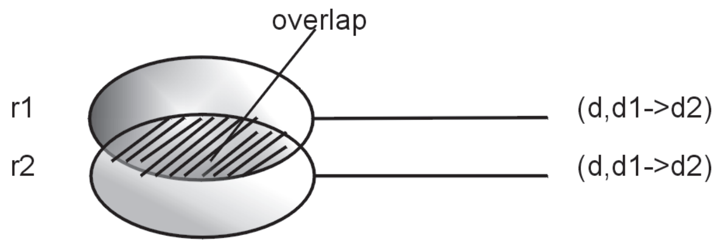

FIGURE 4: Diversity Δ of action rule r_1 with respect to r_2 is determined by the overlap between the two rules' domains.

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:

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

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

4 Diversity of Action Rule 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, association rules [2], or action rules.

In this work, we introduce the notion of diversity of action rule summaries, and suggest how it may be utilized within the generalization process. 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 as shown on Fig. 4.

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:

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

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 the 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):

$$r_3 = [(b_3 \wedge e_2)] \Rightarrow d_1$$
$$r_4 = [a_1 \wedge c_2] \wedge [(b_1 \wedge e_2)] \Rightarrow d_1$$
$$G(r_3, r_4) = [(b \wedge e_2)] \Rightarrow d_1$$

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

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

If we generalize b_3 and b_1 into b, as shown in $G(r_3, r_4)$, then that 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 r_3 and r_4 rules together.

5 Cost of Action Rules

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.

We adopt the following definitions of cost from [6]:

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.

6 Conclusions and Future Work

In this work, we present a theoretical approach which shows a way to decrease the space of action rules, by focusing on generalizing, or creating summaries through the use of hierarchical attributes. By incorporating the diversity and the cost, we discover summaries of interesting action rules. In this way, we reduce the volume of the mined action rules, and present the user with the essence — the more general and more interesting 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, classification [2], or action rules. In this work we introduced the notion of diversity of action rule summaries. The same notion may be used 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. In addition, the more diverse a set of low-level action rules is, the more interesting it is, since it will convey more knowledge to the user compared to a set with too many similar rules.

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. We disregard the ones with high cost as being of low interest to the user. In this way, we additionally decrease the space of action rules, and are left only with the more interesting summaries.

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.

Other future work could integrate the summarization of action rules into some of the approaches in the previous work — section 2, as means for decreasing 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.

References:

[1] He, Z., Xu, X., Deng, S., and Ma, R. (2005). *Mining action rules from scratch*, Expert Systems with Applications, 29(3): 691-699.

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

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

[4] Wang, K., Zhou, S., and Han, J. (2002). *Profit Mining: From Patterns to Actions*, In: Proceedings of the 8th International Conference on Extending Database Technology: Advances in Database Technology.

[5] 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 on Mining Complex Data (MCD 2007).

[6] Tzacheva, A. A. and Ras, Z. W. (2005). *Action rules mining*. International Journal Of intelligent Systems, 20(6):719-736.

[7] Tzacheva, A. A. and Ras, Z. W. (2007). *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.

[8] Ras, Z. and Wieczorkowska, A. (2000). *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, 587-592.

[9] 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.

[10] Dardzinska, A., and Ras, Z. W. (2006). *Cooperative discovery of interesting action rules*, In: Proceedings of the Seventh International Conference on Flexible Query Answering Systems (FQAS).

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

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

[13] Hilderman, R. J. and Hamilton, H. J. (2001). *Knowledge discovery and measures of interest*, Kluwer Academic.

[14] 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).