Assessing the quality of rules with a new monotonic interestingness measure Z

Salvatore Greco[1], Roman Słowiński[2,3], and Izabela Szczęch[2]

[1] Faculty of Economics, University of Catania, Corso Italia, 55, 95129 Catania, Italy
salgreco@unict.it
[2] Institute of Computing Science Poznan University of Technology, 60-965 Poznan, Poland
{Izabela.Szczech, Roman.Slowinski}@cs.put.poznan.pl
[3] Systems Research Institute, Polish Academy of Sciences, 01-447 Warsaw, Poland

Abstract. The development of effective interestingness measures that help in interpretation and evaluation of the discovered knowledge is an active research area in data mining and machine learning. In this paper, we consider a new Bayesian confirmation measure for "*if..., then...*" rules proposed in [4]. We analyze this measure, called Z, with respect to valuable property M of monotonic dependency on the number of objects in the dataset satisfying or not the premise or the conclusion of the rule. The obtained results unveil interesting relationship between Z measure and two other simple and commonly used measures of rule support and anti-support, which leads to efficiency gains while searching for the best rules.

Keywords: Association rules, Data mining, Interestingness, Z-measure, Monotonicity property M.

1 Introduction

When mining large datasets, the number of knowledge patterns, often expressed in a form of "*if..., then...*" rules, can easily be overwhelming for the human capabilities to understand them and to find the useful results. To guide the data analyst identifying valuable rules, various quantitative measures of interestingness (attractiveness measures) have been proposed and studied (e.g. support, anti-support, measures of confirmation) [9]. They all reflect some different characteristics of rules. The problem of choosing an appropriate interestingness measure for a certain application is difficult because the number and variety of measures proposed in the literature is so big. Therefore, studies analyzing theoretical properties of these measures, as well as relationships among them, is worth consideration. Moreover, there are some theoretical properties of interestingness measures which are particularly valuable for practical applications. Properties of measures also naturally group them unveiling relationships between them, and are helpful in choosing an appropriate measure for a particular application.

In this paper, we focus on a new interestingness measure, from the category of Bayesian confirmation, proposed by Crupi et al. [4] and called the Z measure.

It is a measure that quantifies the degree to which the premise of a rule provides support for or against the rule's conclusion. We analyze it with respect to a valuable property M, introduced by Greco et al. [7], of monotonic dependency of the measure on the number of objects satisfying or not the premise or the conclusion of the rule. Moreover, on the basis of satisfying the property M, we draw some practical conclusions about very particular relationship between measure Z and two other simple but meaningful measures of rule support and anti-support.

The paper is organized as follows. In section 2, there are preliminaries on rules and their quantitative description. Next, in section 3, we analyze Z with respect to property M. Section 4 presents practical application of the obtained results. The paper ends with conclusions.

2 Preliminaries

Let us consider discovering rules from a sample of larger reality given in a form of a data table. Formally, a *data table* is a pair $S = (U, A)$, where U is a nonempty finite set of objects, called *universe*, and A is a nonempty finite *set of attributes*. The set V_a is the set of values of the attribute $a \in A$.

Let us associate a formal language L of logical formulas with every subset of attributes. Formulas for a subset $B \subseteq A$ are built up from attribute-value pairs (a, v), where $a \in B$ and $v \in V_a$, using logical operators $\neg$ (not), $\wedge$ (and), $\vee$ (or).

A *rule* induced from S and expressed in L is denoted by $\phi \to \psi$ (read as "if ϕ, then ψ"). It consists of antecedent ϕ and consequent ψ , being formulas expressed in L, called *premise* and *conclusion* (hypothesis or decision), respectively, and therefore it can be seen as a consequence relation between premise and conclusion (see critical discussion [7] about interpretation of rules as logical implications). The rules mined from data may be either *decision* or *association* rules, depending on whether the division of A into condition and decision attributes has been fixed or not.

2.1 Support and Anti-support Measures of Rules

One of the most popular measures used to identify frequently occurring association rules in sets of items from data table S is the *support*. Support of condition ϕ, denoted as $sup(\phi)$, is equal to the number of objects in U having property ϕ. The support of rule $\phi \to \psi$ (also simply referred to as support), denoted as $sup(\phi \to \psi)$, is the number of objects in U having property ϕ and ψ . Thus, it corresponds to statistical significance [9]. Naturally, support is a gain-type criterion, i.e. its higher values are more desirable.

Anti-support of a rule $\phi \to \psi$ (also simply referred to as anti-support), denoted as $anti - sup(\phi \to \psi)$, is equal to the number of objects in U having property ϕ but not having property ψ . Thus, anti-support is the number of counter-examples, i.e. objects for which the premise ϕ evaluates to true but which fall into a class different than ψ . Note that anti-support can also be regarded as $sup(\phi \rightarrow \neg\psi)$. Thus, it is considered as a cost-type criterion, which means that the smaller the value of anti-support, the more desirable it is.

In literature, there can also be found definitions of support and anti-support as relative values with respect to the number of objects in the whole dataset. In this paper, we will not take under consideration such interpretation of support and anti-support, however, doing so would not influence anyhow the generality of the conducted analysis and the obtained results.

2.2 Z measure

Among commonly used interestingness measures there is a large group of Bayesian confirmation measures which quantify the degree to which the premise provides "support for or against" the conclusion [6]. Thus, formally, a measure $c(\phi \rightarrow \psi)$ can be regarded as a measure of confirmation if it satisfies the following condition:

$$c(\phi \rightarrow \psi) \begin{cases} > 0 & if \quad Pr(\psi|\phi) > Pr(\psi), \\ = 0 & if \quad Pr(\psi|\phi) = Pr(\psi), \\ < 0 & if \quad Pr(\psi|\phi) < Pr(\psi). \end{cases} \tag{1}$$

Under the "closed world assumption" adopted in inductive reasoning, and because U is a finite set, it is legitimate to estimate probabilities $\Pr(\phi)$ and $\Pr(\psi)$ in terms of frequencies $sup(\phi)/|U|$ and $sup(\psi)/|U|$, respectively. In consequence, we can define the conditional probability as $Pr(\psi|\phi) = Pr(\psi \wedge \phi)/Pr(\phi)$, and it can be regarded as $sup(\phi \rightarrow \psi)/sup(\phi)$. Thus, the above condition can be re-written as:

$$c(\phi \rightarrow \psi) \begin{cases} > 0 & if \quad \dfrac{sup(\phi \rightarrow \psi)}{sup(\phi)} > sup(\psi)/U, \\ = 0 & if \quad \dfrac{sup(\phi \rightarrow \psi)}{sup(\phi)} = sup(\psi)/U, \\ < 0 & if \quad \dfrac{sup(\phi \rightarrow \psi)}{sup(\phi)} > sup(\psi)/U. \end{cases} \tag{2}$$

Over the years, many authors have proposed their own definitions of particular measures that satisfy condition 2 and now the catalogue of confirmation measures proposed in the literature is quite large. Among the most commonly used ones, there are those shown in Table (1).

Crupi et al. [4] have considered the above confirmation measures from the viewpoint of classical deductive logic [2] introducing function v such that for any argument (ϕ, ψ), v assigns it the same positive value (e.g., 1) iff ϕ entails ψ , i.e. $\phi| = \psi$, an equivalent value of opposite sign (e.g., -1) iff ϕ entails the negation of ψ, i.e. $\phi| = \neg\psi$, and value 0 otherwise. The relationship between the logical implication or refutation of ψ by ϕ, and the conditional probability of ψ by ϕ requires that $v(\phi, \psi)$ and $c(\phi \rightarrow \psi)$ should always be of the same sign. However, Crupi et al. [4] also argue that any confirmation measure $c(\phi \rightarrow \psi)$ should also

Table 1. Common confirmation measures

$D(\phi \to \psi) = \dfrac{sup(\phi \to \psi)}{sup(\phi)} - sup(\psi)$	Carnap [2]
$S(\phi \to \psi) = \dfrac{sup(\phi \to \psi)}{sup(\phi)} - \dfrac{sup(\neg\phi \to \psi)}{sup(\neg\phi)}$	Christensen [3]
$M(\phi \to \psi) = \dfrac{sup(\phi \to \psi)}{sup(\psi)} - sup(\phi)$	Mortimer [11]
$N(\phi \to \psi) = \dfrac{sup(\phi \to \psi)}{sup(\psi)} - \dfrac{sup(\phi \to \neg\psi)}{sup(\neg\psi)}$	Nozick [12]
$C(\phi \to \psi) = \dfrac{sup(\phi \to \psi)}{\lvert U \rvert} - \dfrac{sup(\phi)sup(\psi)}{\lvert U \rvert}$	Carnap [2]
$R(\phi \to \psi) = \dfrac{sup(\phi \to \psi)\lvert U \rvert}{sup(\phi)sup(\psi)} - 1$	Finch [5]
$G(\phi \to \psi) = 1 - \dfrac{sup(\phi \to \neg\psi)\lvert U \rvert}{sup(\phi)sup(\neg\psi)}$	Rips [13]

satisfy principle (3):

$$if \quad v(\phi_1, \psi_1) > v(\phi_2, \psi_2), \quad then \quad c(\phi_1 \rightarrow \psi_1) > c(\phi_2 \rightarrow \psi_2). \tag{3}$$

They have proved that neither of the above mentioned confirmation measures satisfies principle 3. However, their further analysis has unveiled a rather simple way to obtain a measure of confirmation that does fulfill this principle from either D, S, M, N, C, R, or G. They have normalized these measures by dividing them by the maximum they obtain in case of confirmation (i.e. when $sup(\phi \rightarrow \psi)/sup(\phi) \geq sup(\psi)/|U|$), and the absolute value of the minimum they obtain in case of disconfirmation (i.e. when $sup(\phi \rightarrow \psi)/sup(\phi) \leq sup(\psi)/|U|$). It has also been shown that those normalized confirmation measures are all equal:

$$D_{norm} = S_{norm} = M_{norm} = N_{norm} = C_{norm} = R_{norm} = G_{norm}. \tag{4}$$

Crupi et al. have therefore proposed to call them all by one name: **Z-measure**. They have proved that Z, and all confirmation measures equivalent to it, satisfy principle 3. Thus, Z is surely a valuable tool for measuring the confirmation of decision or association rules induced from datasets. Throughout this paper let us consider Z defined as follows:

$$Z(\phi \rightarrow \psi) \begin{cases} \dfrac{\dfrac{sup(\phi \rightarrow \psi)}{sup(\phi)} - \dfrac{sup(\psi)}{|U|}}{1 - \dfrac{sup(\psi)}{|U|}} & if \quad \dfrac{sup(\phi \rightarrow \psi)}{sup(\phi)} \geq \dfrac{sup(\psi)}{|U|}, \\ \\ \dfrac{\dfrac{sup(\phi \rightarrow \psi)}{sup(\phi)} - \dfrac{sup(\psi)}{|U|}}{\dfrac{sup(\psi)}{|U|}} & if \quad \dfrac{sup(\phi \rightarrow \psi)}{sup(\phi)} < \dfrac{sup(\psi)}{|U|}. \end{cases} \tag{5}$$

2.3 Property M of monotonicity

Greco, Pawlak and Słowiński have proposed in [7] property M of monotonic dependency of an interestingness measure on the number of objects satisfying or not the premise or the conclusion of a rule. Formally, an interestingness measure F satisfies the property M if:

$$F[sup(\phi \rightarrow \psi), sup(\neg\phi \rightarrow \psi), sup(\phi \rightarrow \neg\psi), sup(\neg\phi \rightarrow \neg\psi)] \tag{6}$$

is a function non-decreasing with respect to $sup(\phi \rightarrow \psi)$ and $sup(\neg\phi \rightarrow \neg\psi)$, and non-increasing with respect to $sup(\neg\phi \rightarrow \psi)$ and $sup(\phi \rightarrow \neg\psi)$.

The property M with respect to $sup(\phi \rightarrow \psi)$ (or, analogously, with respect to $sup(\neg\phi \rightarrow \neg\psi)$) means that any evidence in which ϕ and ψ (or, analogously, neither ϕ nor ψ) hold together increases (or at least does not decrease) the credibility of the rule $\phi \rightarrow \psi$. On the other hand, the property of monotonicity with respect to $sup(\neg\phi \rightarrow \psi)$ (or, analogously, with respect to $sup(\phi \rightarrow \neg\psi)$) means that any evidence in which ϕ does not hold and ψ holds (or, analogously, ψ

holds and ϕ does not hold) decreases (or at least does not increase) the credibility of the rule $\phi \rightarrow \psi$. In order to present the interpretation of property M let us use the following example used by Hempel [8]. Let us consider a rule $\phi \rightarrow \psi$:

if x is a raven then x is black.

In this case ϕ stands for the property of being a raven and ψ is the property of being black. If an attractiveness measure $I(\phi \rightarrow \psi)$ possesses the property M, then:

- the more black ravens or non-black non-ravens there will be in the dataset, the more credible will become the rule, and thus $I(\phi \rightarrow \psi)$ will obtain greater (or at least not smaller) values,
- the more black non-ravens or non-black ravens in the dataset, the less credible will become the rule and thus, $I(\phi \rightarrow \psi)$ will obtain smaller (or at least not greater) values.

2.4 Partial Preorder on Rules in terms of Rule Support and Anti-support

Let us denote by $\preceq_{s\sim a}$ a partial preorder given by the dominance relation on a set X of rules in terms of two interestingness measures *support* and *anti-support*, i.e. given a set of rules X and two rules $r_1, r_2 \in X$, $r_1 \preceq_{s\sim a} r_2$ if and only if

$$sup(r_1) \leq sup(r_2) \quad \wedge \quad anti - sup(r_1) \geq anti - sup(r_2).$$

Recall that a *partial preorder* on a set X is a binary relation R on X that is reflexive and transitive. The partial preorder $\preceq_{s\sim a}$ can be decomposed into its asymmetric part $\prec_{s\sim a}$ and its symmetric part $\sim_{s\sim a}$ in the following manner: given a set of rules X and two rules $r_1, r_2 \in X$, $r_1 \ {}_{s\sim a} r_2$ if and only if:

$$\begin{array}{lll} sup(r_1) \leq sup(r_2) & \wedge & anti - sup(r_1) > anti - sup(r_2), \quad or \\ sup(r_1) < sup(r_2) & \wedge & anti - sup(r_1) \geq anti - sup(r_2), \end{array} \tag{7}$$

moreover, $r_1 \sim_{s\sim a} r_2$ if and only if:

$$sup(r_1) = sup(r_2) \quad \wedge \quad anti - sup(r_1) = anti - sup(r_2). \tag{8}$$

If for a rule $r \in X$ there does not exist any rule $r' \in X$, such that $r \prec_{s\sim a} r'$, then r is said to be *non-dominated* (i.e. *Pareto-optimal*) with respect to support and anti-support. A set of all non-dominated rules with respect to these measures is also referred to as a *support–anti-support Pareto-optimal border*. In other words, it is the set of rules such that there is no other rule having greater support and smaller anti-support.

3 Analysis of Z-measure with respect to property M

For the clarity of presentation, the following notation shall be used from now on:

$$a = sup(\phi \to \psi), \quad b = sup(\neg\phi \to \psi), \quad c = sup(\phi \to \neg\psi), \quad d = sup(\neg\phi \to \neg\psi),$$
$$a + c = sup(\phi), \quad a + b = sup(\psi), \quad b + d = sup(\neg\phi), \quad c + d = sup(\neg\psi),$$
$$a + b + c + d = |U|.$$

We also assume that set U is not empty, so that at least one of a, b, c or d is strictly positive. Moreover, for the sake of simplicity, we assume that any value in the denominator of any ratio is different from zero. In order to prove that a measure has the property M we need to show that it is non-decreasing with respect to a and d, and non-increasing with respect to b and c.

Theorem 1. *Measure Z has the property M.*

Proof. First, let us consider Z in case of confirmation, i.e. when: $sup(\phi \to \psi)/sup(\phi) \geq sup(\psi)/|U|$:

$$Z = \frac{\frac{a}{a+c} - \frac{a+b}{a+b+c+d}}{\frac{c+d}{a+b+c+d}}. \tag{9}$$

Through simple mathematical transformations we obtain:

$$Z = \frac{ad - bc}{(a+c)(c+d)}. \tag{10}$$

Let us verify if Z is non-decreasing with respect to a, i.e. if an increase of a by $\Delta > 0$ will not result in decrease of Z. Simple algebraic transformations show that:

$$\frac{(a+\Delta)d - bc}{(a+\Delta+c)(c+d)} - \frac{ad - bc}{(a+c)(c+d)} = \frac{cd\Delta + bc\Delta}{(a+\Delta+c)(c+d)(a+c)} \geq 0. \tag{11}$$

Thus, Z (in case of confirmation) is non-decreasing with respect to a.

Clearly, Z is also non-increasing with respect to b, as increase of b by $\Delta > 0$ will result in decrease of the numerator of (10) and therefore in decrease of Z.

Now, let us verify if Z is non-increasing with respect to c, i.e. if an increase of c by $\Delta > 0$ will not result in increase of Z. Simple algebraic transformations show that:

$$\begin{aligned} &\frac{ad - b(c+\Delta)}{(a+c+\Delta)(c+\Delta+d)} - \frac{ad - bc}{(a+c)(c+d)} = \\ &= \frac{(c\Delta + \Delta^2)(bc - ad) - ad(b\Delta + c\Delta + a\Delta + d\Delta)}{(a+c+\Delta)(c+\Delta+d)(a+c)(c+d)}. \end{aligned} \tag{12}$$

Let us observe that: $c\Delta + \Delta^2 > 0$, $ad(b\Delta + c\Delta + a\Delta + d\Delta) > 0$, and $(bc - ad) < 0$ because we consider the case of confirmation. Thus, the numerator of (12) is negative. Since the denominator is positive, we can conclude that Z (in case of confirmation) is non-increasing with respect to c.

Finally, let us verify if Z is non-decreasing with respect to d, i.e. if an increase of d by $\Delta > 0$ will not result in decrease of Z. Simple algebraic transformations show that:

$$\frac{a(d+\Delta)-bc}{(a+c)(c+d+\Delta)} - \frac{ad-bc}{(a+c)(c+d)} = \frac{ac\Delta+bc\Delta}{(a+c)(c+d)(c+d+\Delta)} > 0. \qquad (13)$$

Thus, Z (in case of confirmation) is non-decreasing with respect to d.

Since all four conditions are satisfied, the hypothesis that Z measure in case of confirmation has the property M is true. The proof that in case of disconfirmation Z has the property M is analogous.

4 Practical application of the results

The approach to evaluation of a set of rules with the same conclusion in terms of two interestingness measures being rule support and anti-support was proposed and presented in detail in [1]. The idea of combining those two dimensions came as a result of looking for a set of rules that would include all rules optimal with respect to any confirmation measure with the desirable property M.

Theorem 2. *[1] When considering rules with the same conclusion, rules that are optimal with respect to any interestingness measure that has the property M must reside on the support–anti-support Pareto-optimal border.*

It means that the best rules according to any of confirmation measures with M are in the set of non-dominated rules with respect to support–anti-support. This valuable result unveils some relationships between different interestingness measures. Moreover, it allows to identify a set of rules containing most interesting (optimal) rules according to any interestingness measures with the property M simply by solving an optimized rule mining problem with respect to rule support and anti-support.

As we have proved, measure Z satisfies the property M. This result allows us to conclude that this interestingness measure is monotonically dependent on the number of objects satisfying both the premise and conclusion of the rule (or neither the premise nor the conclusion) and anti-monotonically dependent on the number of objects satisfying only the premise or only the rule's conclusion. Moreover, possession of property M means that rules optimal with respect to Z reside on the Pareto-optimal border with respect to support and anti-support (when considering rules with the same conclusion).

It is a very practical result as it allows potential efficiency gains:

- rules optimal with respect to Z can be mined from the support–anti-support Pareto-optimal set instead of searching the set of all rules,
- we can concentrate on mining only the support–anti-support Pareto-optimal set instead of conducting rule evaluation separately with respect to Z, or any other measure with property M, as we are sure that rules optimal according to Z, or any other measure with property M, are in that Pareto set.

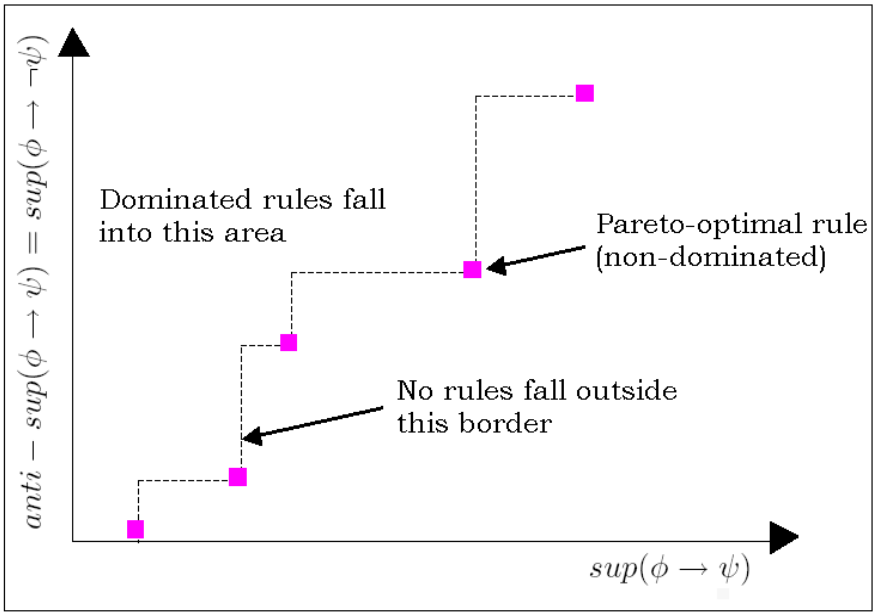

Fig. 1. Support–anti-support Pareto-optimal border

5 Conclusions

Measures of confirmation are an important and commonly used group of interestingness measures. The semantics of their scales is very useful for the purpose of elimination of rules for which the premise does not confirm the conclusion to the desired extent [14,15]. In this paper we considered a recently proposed confirmation measure Z [4]. A theoretical analysis of Z with respect to valuable property M has been conducted. It has been proved that measure Z does satisfy property M, which means that it is a function non-decreasing with respect to $sup(\phi \to \psi)$ and $sup(\neg\phi \to \neg\psi)$, and non-increasing with respect to $sup(\neg\phi \to \psi)$ and $sup(\phi\ to \neg\psi)$. Moreover, the possession of property M implies that rules optimal according to Z will be found on the support–anti-support Pareto-optimal border (when considering rules with the same conclusion). Thus, one can concentrate on mining the set of non-dominated rules with respect to support and anti-support and be sure to obtain in that set all rules that are optimal with respect to any measure with the property M, which includes measure Z.

References

1. Brzezińska, I., Greco, S., Słowiński, R., 2007. Mining Pareto-optimal rules with respect to support and anti-support. *Engineering Applications of Artificial Intelligence* Vol. 20(5) 587-600.

2. Carnap, R., 1962. *Logical Foundations of Probability* 2nd ed. University of Chicago Press, Chicago.
3. Christensen, D., 1999. Measuring confirmation. *Journal of Philosophy* 96, 437-461.
4. Crupi, V., Tentori, K., Gonzalez, M., (to appear). On Bayesian measures of evidential support: Theoretical and empirical issues. *Philosophy of Science*.
5. Finch, H.A., 1999. Confirming Power of Observations Metricized for Decisions among Hypotheses. *Philosophy of Science*, 27, 293-307 and 391-404.
6. Fitelson, B., 2001. *Studies in Bayesian Confirmation Theory*. Ph.D. Thesis, University of Wisconsin, Madison.
7. Greco, S., Pawlak, Z., Słowiński, R., 2004. Can Bayesian confirmation measures be useful for rough set decision rules? *Engineering Applications of Artificial Intelligence* 17: 345-361.
8. Hempel, C.G., 1945. Studies in the logic of confirmation (I). *Mind* 54, 1-26.
9. Hilderman, R. and Hamilton, H., 2001. *Knowledge Discovery and Measures of Interest*. Kluwer Academic Publishers.
10. Kemeny, J., Oppenheim, P., 1952. Degrees of factual support. *Philosophy of Science* 19, 307-324.
11. Mortimer, H., 1988. *The Logic of Induction*, Paramus, Prentice Hall.
12. Nozick, R., 1981. *Philosophical Explanations*, Clarendon Press Oxford (UK).
13. Rips, L.J., 2001. Two Kinds of Reasoning. *Psychological Science*, 12, 129-134.
14. Słowiński, R., Brzezińska, I., Greco, S., 2006. Application of Bayesian confirmation measures for mining rules from support-confidence Pareto-optimal set. [Invited paper in]: L.Rutkowski, R.Tadeusiewicz, L.A.Zadeh, J.Zurada (eds.), *Artificial Intelligence and Soft Computing (ICAISC 2006)*. LNAI 4029, Springer-Verlag, Berlin, pp. 1018-1026.
15. Słowiński, R., Szczęch, I., Urbanowicz, M., Greco, S., 2007. Mining association rules with respect to support and snti-support - experimental results. [In]: M.Kryszkiewicz, J.F.Peters, H.Rybiski, A.Skowron (eds.): *Rough Sets and Intelligent Systems Paradigms*. LNAI 4585, Springer-Verlag, Berlin, pp. 534–542.