Analysis of monotonicity properties of new normalized rule interestingness measures

Słowiński Roman*[a] — Salvatore Greco — Szczęch Izabela***

** Institute of Computing Science, Poznan University of Technology,*
Piotrowo 2, 60–965 Poznań, Poland

Roman.Slowinski@cs.put.poznan.pl

Izabela.Szczech@cs.put.poznan.pl

[a] Systems Research Institute, Polish Academy of Sciences,
01–447 Warsaw, Poland

*** Faculty of Economics University of Catania,*
Corso Italia, 55, 95129 Catania, Italy

salgreco@mbox.unicit.it

ABSTRACT. *The paper considers interestingness measures for evaluation of relevance and usefulness of "if..., then..." rules induced from data. We propose a way to normalize three popular measures: rule interest function of Piatetsky-Shapiro, gain measure of Fukuda et al. and dependency factor used by Popper and Pawlak. The normalization transforms the measures to the interval [–1, 1], whose bounds correspond to maximal Bayesian confirmation and disconfirmation, respectively, and thus make them more meaningful. The new normalized measures are analyzed with respect to a 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 have a practical application as they lead to efficiency gains while searching for the best rules.*

KEY WORDS: *Data mining, Normalization, Rule Interest Function, Gain Measure, Dependency Factor, Monotonicity property M*

1. Introduction

Rules are popular patterns induced from data using various techniques of data mining. In this active research area, rule evaluation has been considered by many authors from different perspectives. 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, Bayesian confirmation measures) (Hilderman *et al*., 2000). 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 wide. Therefore, there naturally arises a need to analyze theoretical properties of measures. Properties of measures group them unveiling relationships between them, and are helpful in choosing an appropriate measure for a particular application (Lenca *et al*., 2008). Some of the properties have also very practical applications.

In this paper, we focus on three commonly known and used measures: rule interest function (Piatetsky-Shapiro 1991), gain measure (Fukuda *et al*., 1996) and dependency factor advocated by (Popper 1959) and (Pawlak2004). We propose a way to normalize these measures to the interval [-1, 1], whose bounds correspond to maximal Bayesian confirmation and disconfirmation, respectively, and thus make them more meaningful. We analyze the new normalized measures with respect to a valuable property M, introduced by (Greco *et al*. 2004), 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 these measures 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 the normalized rule interest function, gain measure and dependency factor 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*. For every attribute $a \in A$, let us denote by V_a the domain of a. By $a(x)$ we will denote the value of attribute $a \in A$ for an object $x \in U$. A *rule* induced from a data table S is denoted by $\phi \rightarrow \psi$ (read as "*if* ϕ, *then* ψ"), where ϕ and ψ are built up from elementary conditions using logical operator $\wedge$ (and). The *elementary conditions* of a rule are defined as ($a(x)$ *rel* v)) where *rel* is a relational operator from the set {=, <, ≤, ≥, >} and v is a constant belonging to V_a. The antecedent ϕ of a rule is also referred to as *premise* or *condition*. The consequent ψ of a rule is also called *conclusion*, *decision* or *hypothesis*. Therefore, a rule can be seen as a consequence relation (see critical discussion in (Greco *et al.*, 2004) about interpretation of rules as logical implications) between premise and conclusion. 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. The *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 equal to the number of objects in U having both property ϕ and ψ; for those objects, both premise ϕ and conclusion ψ evaluate to true.

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 miss the property ψ. Note that anti-support can also be regarded as $sup(\phi\to\neg\psi)$. Thus, it is considered as a cost-type criterion, which means that the smaller the value of anti-support, the more desirable the rule is.

2.2. *Piatetsky-Shapiro's Rule Interest Function, Gain and Dependency Factor*

The *rule interest* function RI introduced by (Piatetsky-Shapiro 1991) is used to quantify the correlation between premise and conclusion. It is defined by the following formula:

$$RI = sup(\phi \to \psi) - \frac{sup(\phi)\,sup(\psi)}{|U|}. \tag{1}$$

For rule $\phi\to\psi$, when RI=0, then ϕ and ψ are statistically independent and thus, such a rule should be considered as uninteresting. When $RI > 0$ ($RI < 0$), then there is a positive (negative) correlation between ϕ and ψ (Hilderman *et al.*, 2000).

The *gain* measure of (Fukuda *et al.*, 1996) is defined in the following manner:

$$gain(\phi \to \psi) = sup(\phi \to \psi) - \Theta\, sup(\phi). \tag{2}$$

where Θ is a fractional constant between 0 and 1. Note that, for a fixed value of $\Theta = sup(\psi)/|U|$, the *gain* measure becomes identical to the above *RI*.

The *dependency factor* considered in (Pawlak 2004) and also advocated by (Popper 1959), is defined in the following manner:

$$\eta(\phi \to \psi) = \frac{\dfrac{sup(\phi \to \psi)}{sup(\phi)} - \dfrac{sup(\psi)}{|U|}}{\dfrac{sup(\phi \to \psi)}{sup(\phi)} + \dfrac{sup(\psi)}{|U|}}. \qquad [3]$$

The *dependency factor* expresses a degree of dependency, and can be seen as a counterpart of correlation coefficient used in statistics. When ϕ and ψ are independent on each other, then $\eta(\phi\to\psi)=0$. If $-1< \eta(\phi\to\psi) <0$, then ϕ and ψ are negatively dependent, and if $0< \eta(\phi\to\psi) <1$, then ϕ and ψ are positively dependent.

2.3. *Normalization of interestingness measures*

Among widely studied and applied interestingness measures there is also a group of Bayesian confirmation measures which quantify the degree to which the premise provides "support for or against" the conclusion (Fitelson 2001). Thus, formally, a measure $c(\phi\to\psi)$ can be regarded as Bayesian measure of confirmation if it satisfies the following definition:

$$c(\phi \to \psi)\begin{cases} >0 & if \;\; Pr(\psi \mid \phi) > Pr(\psi), \\ =0 & if \;\; Pr(\psi \mid \phi) = Pr(\psi), \\ <0 & if \;\; Pr(\psi \mid \phi) < Pr(\psi). \end{cases} \qquad [4]$$

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 $\Pr(\psi|\phi) = \Pr(\phi\wedge\psi)/\Pr(\phi)$, and it can be regarded as $sup(\phi\to\psi)/sup(\phi)$. Thus, the above condition can be re–written as:

$$c(\phi \to \psi)\begin{cases} >0 & if \quad \dfrac{sup(\phi \to \psi)}{sup(\phi)} > \mathrm{sup}(\psi)/|U|, \\ =0 & if \quad \dfrac{sup(\phi \to \psi)}{sup(\phi)} = \mathrm{sup}(\psi)/|U|, \\ <0 & if \quad \dfrac{sup(\phi \to \psi)}{sup(\phi)} < \mathrm{sup}(\psi)/|U|. \end{cases} \qquad [5]$$

Since *RI*, *gain* measure (iff $\Theta = sup(\psi)/|U|$) and *dependency factor* are Bayesian confirmation measures (Szczęch 2007), we propose to normalize them, so that they would distinguish between two completely different situations: situation α in which confirmation occurs (i.e. when $sup(\phi\to\psi)/sup(\phi) \geq sup(\psi)/|U|$) and situation β in which disconfirmation occurs (i.e. when $sup(\phi\to\psi)/sup(\phi) < sup(\psi)/|U|$). Inspired by (Crupi *et al*. 2008), who have analyzed a group of normalized Bayesian confirmation measures, we propose to normalize *RI*, *gain* and *dependency factor* by dividing them by the maximum value they obtain in case of confirmation, and by the absolute minimum value they obtain in case of disconfirmation. In this way, we will obtain confirmation measures taking values from the interval [−1, 1]. The issue of normalizing measures keeps gaining significance in the literature and has also been taken up in (Diatta *et al.*, 2007).

There are many approaches to determining those maximum and minimum values, which eventually lead to different normalizations. In this paper, we present the normalization, inspired by the approach of (Nicod 1923), for which we consider only cases in which there is the evidence, while we ignore cases where there is no evidence. For example, in case of "*all ravens are black*", the evidence is "*raven*" and the hypothesis is "*black*". In this situation, a *black raven* supports the conclusion that *all ravens are black*, *non-black ravens* are against this conclusion, and everything which is *not a raven* can be ignored.

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

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

We assume that set U is not empty, so that at least one of a, b, c, d is strictly positive, and that any value in the denominator of any ratio is different from zero.

Based on the above notation, the *rule interest* function [1] can be expressed as:

$$RI = a - \frac{(a+b)(a+c)}{a+b+c+d} \qquad [6]$$

and thus the normalized *RI* should take the following form:

$$RI_{norm} = \begin{cases} \dfrac{a - \dfrac{(a+b)(a+c)}{a+b+c+d}}{a+c-\dfrac{(a+b+c)(a+c)}{a+b+c+d}} & \text{in case of confirmation} \\[2em] \dfrac{a - \dfrac{(a+b)(a+c)}{a+b+c+d}}{\dfrac{b(a+c)}{a+b+c+d}} & \text{in case of disconfirmation} \end{cases} \quad [7]$$

In case of the *gain* measure, its definition [2] can be expressed as:

$$gain = a - \Theta(a+c) \quad [8]$$

and the normalized *gain* should be defined as:

$$gain_{norm} = \begin{cases} \dfrac{a-\Theta(a+c)}{(a+c)(1-\Theta)} & \text{in case of confirmation} \\[1.5em] \dfrac{a-\Theta(a+c)}{(a+c)\Theta} & \text{in case of disconfirmation} \end{cases} \quad [9]$$

The *dependency factor* [3] takes the following form in the applied notation:

$$\eta = \frac{\dfrac{a}{a+c} - \dfrac{a+b}{a+b+c+d}}{\dfrac{a}{a+c} + \dfrac{a+b}{a+b+c+d}} \quad [10]$$

which determines the following definition of the normalized *dependency factor*:

$$\eta_{norm} = \begin{cases} \dfrac{\dfrac{a}{a+c} - \dfrac{a+b}{a+b+c+d}}{\dfrac{a}{a+c} + \dfrac{a+b}{a+b+c+d}} \times \dfrac{1+\dfrac{a+b+c}{a+b+c+d}}{1-\dfrac{a+b+c}{a+b+c+d}} & \textit{in case of confirmation} \\[2em] \dfrac{\dfrac{a}{a+c} - \dfrac{a+b}{a+b+c+d}}{\dfrac{a}{a+c} + \dfrac{a+b}{a+b+c+d}} & \textit{in case of disconfirmation.} \end{cases} \quad [11]$$

We have also considered, however not included in the further analysis in this paper, two other normalization approaches called "Bayesian" and "likelihoodist" (Fitelson 2007). The first one is related to the idea that the evidence confirms the hypothesis, if the hypothesis is more frequent with the evidence rather than with ¬evidence, and in this context, analogously, the evidence disconfirms the hypothesis, if ¬hypothesis is more frequent with the evidence rather than with ¬evidence. The second approach is related to the idea that the evidence confirms the hypothesis, if the evidence is more frequent with the hypothesis rather than with ¬hypothesis, and in this context, analogously, the evidence disconfirms the hypothesis, if the evidence is more frequent with ¬hypothesis rather than with the hypothesis.

2.4. *Property M of monotonicity*

Greco, Pawlak and Słowiński have proposed in (Greco *et al.*, 2004) 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:

$$F[\sup(\phi \to \psi), \sup(\neg\phi \to \psi), \sup(\phi \to \neg\psi), \sup(\neg\phi \to \neg\psi)] \qquad [12]$$

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* $\sup(\phi \to \neg\psi)$.

The property M with respect to $sup(\phi \to \psi)$ (or, analogously, with respect to $sup(\neg\phi \to \neg\psi)$) 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 $\phi\to\psi$. On the other hand, the property of monotonicity with respect to $\sup(\neg\phi \to \psi)$ (or, analogously, with respect to $\sup(\phi \to \neg\psi)$) means that any evidence in which ϕ does not hold and ψ holds (or, analogously, ϕ holds and ψ does not hold) decreases (or at least does not increase) the credibility of the rule $\phi\to\psi$.

Let us present the interpretation of property M on the following example used in (Hempel 1945). Let us consider a rule $\phi\to\psi$: *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 $F(\phi\to\psi)$ possesses the property M, then:

- the more black ravens there are in the dataset, the more credible is the rule, and thus $F(\phi\to\psi)$ obtains greater (or at least not smaller) values,
- with the increase of the number of non-black non-ravens $F(\phi\to\psi)$ also obtains greater (or at least not smaller) values,

- the more black non-ravens appear in the dataset, the less credible becomes the rule and thus, $F(\phi\rightarrow\psi)$ obtains smaller (or at least not greater) values,
- the more non-black ravens are the dataset, the less credible is the rule and thus, $F(\phi\rightarrow\psi)$ obtains smaller (or at least not greater) values.

2.5. *Support–Anti-support Pareto-optimal border*

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

$$sup(r_1) \leq sup(r_2) \wedge anti\text{-}sup(r_1) \geq anti\text{-}sup(r_2).$$

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 X and two rules r_1, $r_2 \in X$, $r_1 \prec_{s\sim a} r_2$ if and only if

[13]

$$\begin{aligned} &sup(r_1) \leq sup(r_2) \wedge anti\text{-}sup(r_1) > anti\text{-}sup(r_2), \text{ or} \\ &sup(r_1) < sup(r_2) \wedge anti\text{-}sup(r_1) \geq anti\text{-}sup(r_2), \end{aligned}$$

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

[14]

$$sup(r_1) = sup(r_2) \wedge anti\text{-}sup(r_1) = anti\text{-}sup(r_2).$$

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 with the same conclusion, having a greater support and a smaller anti-support.

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 (Brzezińska *et al.*, 2007), and also studied in (Słowiński *et al.*, 2006). 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 interestingness measure with the desirable property M.

Theorem 1. (Brzezińska *et al.*, 2007) 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.

The above theorem states that the best rules according to any interestingness measures with M are in the set of non-dominated rules (i.e. objectively, the best) with respect to support–anti-support. This valuable and practical result 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.

3. Analysis of normalized measures with respect to property M

In order to prove that a normalized measure has 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* both in case of confirmation and disconfirmation.

Theorem 2. The normalized *rule interest* function has the property M.

Proof: We will only present the proof that the normalized *RI* is non-decreasing with respect to *a* in case of confirmation, and omit the other proofs as they are analogous. Through simple mathematical transformation, we obtain the following form of the normalized *RI* in case of confirmation (for simplicity denoted by $RI_{\text{norm+}}$):

$$RI_{norm+} = \frac{ad - bc}{ad + cd} . \qquad [15]$$

$RI_{\text{norm+}}$ will be non-deceasing with *a* if and only if an increase of *a* by $\Delta > 0$ will not result in a decrease of $RI_{\text{norm+}}$. Simple algebraic transformations show that:

$$\frac{(a+\Delta)d - bc}{(a+\Delta)d + cd} - \frac{ad - bc}{ad + cd} = \frac{cd\Delta + bc\Delta}{d(a+\Delta+c)} > 0 . \quad \square \qquad [16]$$

Theorem 3. The normalized *gain* measure has the property M.

Proof: Analogous to the proof of Theorem 2.

Theorem 4. The normalized *dependency factor* does not have property M.

Proof: The normalized *dependency factor* in case of confirmation does possess property M (the proof is analogous to proof of Theorem 2), however, we can prove by a counterexample that the normalized *dependency factor* in case of disconfirmation (for simplicity denoted as η_{norm-}) does not have property M: Let us consider case α, in which a=7, b=2, c=3, d=3, and case α', in which *a* increases to 8 and b, c, d remain unchanged. The normalized *dependency factor* in case of disconfirmation does not have property M as such increase of a results in the decrease of the measure: $\eta_{norm-}(\phi \rightarrow \psi) = 0.0769 > 0.0756 = \eta'_{norm-}(\phi \rightarrow \psi)$. □

4. Consequences for the user - practical application of the results

In the previous section, we have proved that the normalized measures of *rule interest* function and *gain* possess the property M, while the normalized *dependency factor* does not have this property. These results are of practical value as they show that rules optimal with respect to normalized *RI* or *gain* reside on the Pareto-optimal border with respect to support and anti-support (when considering rules with the same conclusion). Moreover, they allow potential efficiency gains as:

- rules optimal with respect to normalized *RI* or *gain* can be found in the support–anti-support Pareto-optimal set instead of searching the set of all rules,
- rule evaluation can be narrowed down to mining only the support–anti-support Pareto-optimal set instead of conducting rule evaluation separately with respect to normalized *RI, gain*, or any other measure with property M, as we are sure that rules optimal according to normalized *RI*, *gain*, or any other measure with property M, are in that Pareto set.

To illustrate practical application of the above theoretical results, we have conducted several computational experiments analyzing rules optimal with respect to normalized *RI* in case of confirmation. In Figure 1, there is an exemplary diagram from those experiments presenting induced rules in the perspective of rule support and anti-support. For a real life dataset containing information about technical state of buses, a set of all possible rules for which the premise confirms the conclusion was generated. A set of rules with the same conclusion was then isolated and rules non-dominated with respect to support and anti-support were found (those rules form the support–anti-support Pareto-optimal border). The support–anti-support Pareto-optimal border is indicated in Figure 1 by circles connected by a line. In the generated set of rules, by empty red squares we have distinguished rules optimal according to the normalized *RI* in case of confirmation (these are rules with $sup(\phi \rightarrow \psi)=49$ and $anti\text{-}sup(\phi \rightarrow \psi)=2$, or with $sup(\phi \rightarrow \psi)=50$ and $anti\text{-}sup(\phi \rightarrow \psi)=4$). The diagram shows that, indeed, rules optimal with respect to the normalized *rule interest* function in case of confirmation lie on the support–anti-support Pareto-optimal border.

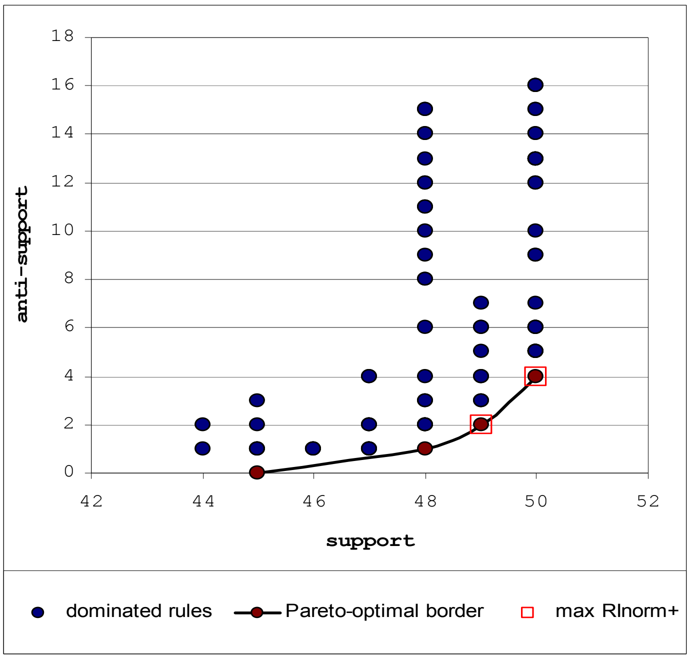

Figure 1. *Support–anti-support Pareto-optimal border*

5. Conclusions

In this paper, we have considered three popular measures: *rule interest* function, *gain* measure and *dependency factor*. The normalization consists in dividing the measures by the maximum value they obtain in case of confirmation, and by the absolute minimum value they obtain in case of disconfirmation. In this way, while keeping the interval of variation [-1, 1], the measures behave differently in case of confirmation and in case of disconfirmation.

A theoretical analysis of the new normalized measures with respect to valuable property M has been conducted. It has been proved that the normalized measure *RI* and *gain* satisfy property M, while the normalized *dependency factor* does not possess this property. The possession of property M implies that rules optimal with respect to the normalized *RI* and *gain* will be found on the support–anti-support Pareto-optimal border (when considering rules with the same conclusion). These results have also been illustrated on an exemplary dataset. It is, therefore, legitimate to conclude that rule evaluation can be narrowed down to mining only the support–anti-support Pareto-optimal set instead of conducting rule evaluation separately with respect to normalized *RI, gain*, or any other measure with property M, as we are sure that rules optimal with respect to normalized *RI*, *gain*, or any other measure with property M, are in that Pareto set.

6. References

Brzezińska I., Greco S., Słowiński R., “Mining Pareto–optimal rules with respect to support and anti-support”, *Engineering Applications of Artificial Intelligence*, vol. 20 no. 5. 2007, p. 587-600.

Crupi V., Tentori K., Gonzalez M., “On Bayesian measures of evidential support: Theoretical and empirical issues”, *Philosophy of Science* 2008.

Diatta J., Ralambondrainy H., Totohasina A., “Towards a Unifying Probabilistic Implicative Normalized Quality Measure for Association Rules”, *Quality Measures in Data Mining Series: Studies in Computational Intelligence,* vol. 43, 2007.

Fitelson B., *Studies in Bayesian Confirmation Theory*. Ph.D. Thesis, University of Wisconsin, Madison 2001.

Fitelson B., “Likelihoodism, Bayesianism, and Relational Confirmation”, *Synthese*, vol. 156 no. 3, 2007, p. 473-489. Springer Netherlands.

Fukuda T., Morimoto Y., Morishita S., Tokuyama T., „Data Mining using Two-Dimensional Optimized Association Rules: Schemes, Algorithms, and Visualization” *Proc. of the 1996 ACM SIGMOD Int'l Conference on Management of Data*, 1996, p. 13-23.

Greco S., Pawlak Z., Słowiński R., „Can Bayesian confirmation measures be useful for rough set decision rules?” *Engineering Applications of Artificial Intelligence*, vol. 17, 2004, p345–361.

Hempel C.G., “Studies in the logic of confirmation (I)” *Mind* vol. 54, 1945, p. 1-26.

Hilderman R.J., Hamilton H.J., “Applying Objective Interestingness Measures in Data Mining Systems”, *Fourth European Symposium on Principles of Data Mining and Knowledge Discovery (PKDD-2000),* Lyon, France, Springer Verlag, September 2000 p. 432-439.

Lenca Ph., Meyer P., Vaillant B., Lallich S., “On selecting interestingness measures for association rules: User oriented description and multiple criteria decision aid”, *European Journal of Operational Research*, Vol. 184, Issue 2, 2008, p. 610-626.

Nicod J., *Le problème de la logique de l’induction*, Alcan, Paris, 1923.

Pawlak Z., “Some issues on rough sets”, *Transactions on Rough Sets I,* LNCS, 2004, p. 1-58.

Piatetsky-Shapiro G., “Discovery, analysis and presentation of strong rules”, *Knowledge Discovery in Databases*, AAAI/MIT Press, 1991, pp. 229-248.

Popper K.R., *The Logic of Scientific Discovery*. Hutchinson, London 1959.

Słowiński R., Brzezińska I., Greco S., „Application of Bayesian confirmation measures for mining rules from support-confidence Pareto-optimal set”, *Invited paper in Artificial Intelligence and Soft Computing*. LNAI 4029, Springer, Berlin, 2006, pp. 1018-1026.

Szczęch I., *Multicriteria Attractiveness Evaluation of Decision and Association Rules*, Ph.D. Thesis, Poznań University of Technology, Poznań, Poland, 2007.