Multicriteria Attractiveness Evaluation of Decision and Association Rules

Izabela Szczęch

Institute of Computing Science, Poznan University of Technology,
Piotrowo 2, 60-965 Poznan, Poland
Izabela.Szczech@cs.put.poznan.pl

Abstract. The work is devoted to multicriteria approaches to rule evaluation. It analyses desirable properties (in particular the property M, property of confirmation and hypothesis symmetry) of popular interestingness measures of decision and association rules. Moreover, it analyses relationships between the considered interestingness measures and enclosure relationships between the sets of non-dominated rules in different evaluation spaces. It's main result is a proposition of a multicriteria evaluation space in which the set of non-dominated rules will contain all optimal rules with respect to any attractiveness measure with the property M. By determining the area of rules with desirable value of a confirmation measure in the proposed multicriteria evaluation space one can narrow down the set of induced rules only to the valuable ones. Furthermore, the work presents an extension of an apriori-like algorithm for generation of rules with respect to attractiveness measures possessing valuable properties and shows some applications of the results to analysis of rules induced from exemplary datasets.

1 Introduction

1.1 Knowledge discovery

Computer systems are commonly used nowadays in vast number of application areas, including banking, telecommunication, management, healthcare, trade, marketing, control engineering, environment monitoring, research and science, among others. We are witnessing a trend to use them anytime and anywhere. As a result a huge amount of data of different types (text, graphics, voice, video) concerning in fact all human activity domains (business, education, health, culture, science) is gathered, stored and available. These data may contain hidden from a user interesting and useful *knowledge* represented (defined) by some non-trivial and not-explicitly visible patterns (relationships, anomalies, regularities, trends) [21], [48], [72], [87], [96], [97].

With the growth of the amount and complexity of the data stored in contemporary, large databases and data warehouses, the problem of extracting knowledge from datasets emerges as a real challenge, increasingly difficult and important. This problem is a central research and development issue of *knowledge* *discovery* that generally is a non-trivial process of looking for new, potentially useful and understandable patterns in data [21], [61], [4].

The discovered knowledge is represented by patterns which can take the form of decision or association rules, clusters, sequential patterns, time series, contingency tables, and others [14], [24], [33], [61], [62], [70], [90], [89], [95]. In this article, we shall consider patterns expressed in the form of "*if. . . , then. . .*" rules. The representation of knowledge in form of rules is considered as easier to comprehend than other forms (for discussion see [14], [55], [54], [75]). Rules are usually induced from a dataset being a set of objects characterized by a set of attributes. They can be described as consequence relations between the condition (the "*if* part") and decision (the "*then* part") formulas built from attribute-value pairs. The condition formulas are called the *premise* of the rule and the decision formulas are referred to as the *conclusion* or *hypothesis* of the rule. Objects from a dataset support the rule if the attribute-value pairs from the rule's premise and conclusion match respectively the values of the object on each of the attributes mentioned in the rule, i.e. if the premise and conclusion of the rule is satisfied by the object. The more objects support the rule, the stronger the rule is. Rules can be induced from different datasets, e.g. from a dataset containing information about patients of a hospital. Such data can be gathered in the process of diagnostic treatment. Rules that could potentially be induced from such dataset could describe co-occurrence of certain symptoms and a disease: *if* symptom s_1 is present *and* symptom s_2 is absent *then* disease d_1.

1.2 Attractiveness measures and their properties

Typically, the number of rules generated from massive datasets is quite large, but only a few of them are likely to be useful for the domain expert. It is due to the fact that many rules are either irrelevant or obvious, and do not provide new knowledge [10]. Therefore, in order to measure the relevance and utility of the discovered rules, quantitative measures, also known as attractiveness or interestingness measures (metrics), have been proposed and studied (for review see, e.g. [26], [37], [53], [81], [94]). They allow to reduce the number of rules that need to be considered by ranking them and filtering out the useless ones. Since there is no single attractiveness measure that captures all characteristics of the induced rules and fulfils the expectations of any user, the number of interestingness measures proposed in literature is large. Each of them reflects certain characteristics of rules and leads to an in-depth understanding of their different aspects. Among widely known and commonly applied attractiveness measures there are such as support and confidence [2], gain [25], conviction [6], rule interest function [72], dependency factor [66], entropy gain [58], [59], laplace [12], [93], lift [40], [6], [16].

While choosing an attractiveness measure(s) of rules for a certain application, the users also often take into consideration properties (features) of measures which reflect the user's expectations towards the behaviour of the measures in particular situations. For example, one may demand that the used measure will increase its value for a given rule (or at least will not decrease) when the number of objects in the dataset that support this rule increases. In this article, we shall focus on the following properties of attractiveness measures, well motivated in the recent literature [28], [9], [22], [15], [17], [10]:

– *the property M* of monotonic dependency of the measure on the number of objects supporting or not the premise or the conclusion of the rule [28], [9],
– *the property of confirmation* quantifying the degree to which the premise of the rule provides evidence for or against the conclusion [22], [15],
– *the property of hypothesis symmetry* arguing that the significance of the premise with respect to the conclusion part of a rule should be of the same strength, but of the opposite sign, as the significance of the premise with respect to a negated conclusion [17], [10].

Analyses verifying if popular interestingness measures possess the above listed properties widen our understanding of those measures and of their applicability, and help us learn about relationships between different measures. The obtained results are also useful for practical applications because they show which attractiveness measures are relevant for meaningful rule evaluation.

1.3 Aim and scope of the article

The problem of choosing an appropriate attractiveness measure for a certain application is difficult not only because of the number of measures but also due to the fact that a single measure of interestingness is often an insufficient indicator of the quality of the rules. Therefore, a multicriteria evaluation, i.e. using at the same time more than one attractiveness measure (criterion), has become a common approach solving this issue [3], [23], [30], [31], [50]. In case of a multicriteria rule evaluation, objectively, the best rules are the *non-dominated* ones (also known as *Pareto-optimal* rules), i.e. those for which there does not exist any other rule that is better on at least one evaluation criterion and not worst on any other. The set of all non-dominated rules, with respect to particular evaluation criteria, is referred to as the *Pareto-optimal set* or the *Pareto-optimal border*.

The popular measures of *rule support* and *confidence*, have been considered by Bayardo and Agrawal [3] as sufficient for multicriteria evaluation of rules. They express the number of objects in the dataset that support the rule and the probability with which the conclusion evaluates to *true* given that the premise is *true*, respectively. Those measures are used in the well-known apriori algorithms [2] and permit to benefit from the main advantage of these algorithms, which consists in reduction of the frequent itemset search space.

In the literature, a group of attractiveness measures called *Bayesian confirmation measures* has also been thoroughly investigated ([11], [17], [22], [28]). The reason for that was different than the motivation for using support–confidence measures in connection with apriori algorithms. It followed from semantic consideration of attractiveness measures. In general, Bayesian confirmation measures quantify the degree to which a premise of a rule "provides arguments for or against" the rule's conclusion. Therefore, their semantic meaning allows to distinguish the meaningful rules for which the premise confirms the conclusion. The measure of confidence does not have the means to do that and there may occur situations in which rules that are characterised by high values of confidence are in fact misleading because the premise disconfirms the rule's conclusion. In this context, there arises a natural need to search for a substituting evaluation space that would include a confirmation measure. Moreover, since the property M of monotonic dependency of an attractiveness measure on a number of objects supporting or not the premise or conclusion of a rule, proposed by Greco, Pawlak and Słowiński [28], has been recognised as crucial especially for confirmation measures, it is desirable for a new evaluation space to include measures that are not only confirmation measures but also have the property M.

The most general goal of the article is to find an evaluation space such that its set of non-dominated rules would include rules that are optimal with respect to any measure with the property M. Of course, the confirmation semantics would also need to be included in such space to avoid analysing uninteresting and misleading rules.

The problem of choosing an adequate multicriteria evaluation space is non trivial. It naturally leads to an important issue, not yet thoroughly discussed in the literature, of comparing different evaluation spaces, as well as determining the relationships of enclosure between their sets of non-dominated rules. If such relationships were discovered, it would mean that inducing non-dominated rules with respect to one evaluation space, one can guarantee that it contains optimal or Pareto-optimal rules with respect to combination of other measures. Such results would have a significant practical value as they would allow to determine a limited set of interesting rules more effectively, because instead of numerous rule evaluations in different spaces, one could conduct such an evaluation once only finding the most general set of non-dominated rules which contains other optimal or Pareto-optimal rules.

In the above context the general aim of this work has be formulated as:

Analysis of properties and relationships between popular rule attractiveness measures and proposition of multicriteria rule evaluation space in which the set of non-dominated rules will contain all optimal rules with respect to any attractiveness measure with the property M.

To attain this aim the following detailed tasks should be completed:

1. Analysis of rule support, rule anti-support, confidence, rule interest function, gain, dependency factor, f and s attractiveness measures with respect to the property M, the property of confirmation and the property of hypothesis symmetry.
2. Analysis of relationships between the considered interestingness measures and analysis of enclosure relationships between the sets of non-dominated rules in different evaluation spaces.

3. Proposition of a multicriteria evaluation space in which the set of non-dominated rules will contain all optimal rules with respect to any attractiveness measure with the property M.
4. Determining the area of rules with desirable value of a confirmation measure in the proposed multicriteria evaluation space.
5. Extension of an apriori-like algorithm for generation of rules with respect to attractiveness measures possessing valuable properties and presentation of application of the results to analysis of rules induced from exemplary datasets.

The plan of this article follows the above tasks. In particular, in Section 2 preliminaries on rules and their basis quantitative description as well as the definitions of the considered properties of attractiveness measures are presented. Section 3 is devoted to analysing whether considered interestingness measures possess the property M, the property of confirmation and the property of hypothesis symmetry. Section 4 describes different multicriteria evaluation spaces and discusses their advantages and disadvantages. Section 5 presents our proposition of the support–anti-support evaluation space, for which the set of non-dominated rules contains rules that are optimal with respect to any attractiveness measure that has the property M. Next, in Section 6 there is a presentation of an association mining system developed for showing applications of the results on exemplary datasets. Finally, Section 7 summarises the article with a discussion on the completed work and possible lines of further investigations.

2 Basic quantitative rule description

The discovery of knowledge from data is done by induction. It is a process of creating patterns which are true in the world of the analyzed data. However, it is worth mentioning, as Karl Popper [74] did, that *one cannot prove the correctness of generalizations of specific observations or analogies to known facts, but can refute them.*

In this article we consider discovering knowledge represented in form of rules. The starting point for such rule induction (mining) is a sample of larger reality often represented in a form of a data table. Formally, a *data table* is a pair

$$S = (U, A) \tag{1}$$

where U is a nonempty finite set of objects (items) called *universe*, and A is a nonempty finite set of *attributes* [66], [68], [70].

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 $x \in U$.

The are many scales of attributes describing objects. The classical hierarchy of attribute scales is the following [88], [5], [57], [13], [87]:

– nominal,
– ordinal,

– interval,
– ratio.

Nominal scale can be regarded as names assigned to objects as labels. The domain of attributes with the nominal scale is an unordered set of attribute values and therefore, the only comparison that can be performed between two such values is equality and inequality. Relations such as “less than” or “greater than” and operations such as addition or subtraction are inapplicable for such attributes. For practical data processing values of attributes with nominal scale can take the form of numerals, but in that case their numerical value is irrelevant. Examples of attributes with nominal scale can include: the marital status of a person, the make of a car, religious or political-party affiliation, birthplace.

The domain of attributes with the *ordinal scale* is an ordered set of attribute values. The values assigned to objects represent the rank order (1^{st}, 2^{nd}, 3^{rd} etc.) of the objects. In addition to equality/inequality, one can also perform “less than” or “greater than” comparisons on ordinal attribute values. Nevertheless, conventional addition and subtraction remain meaningless. Examples of ordinal scales include the results of a horse race, which only express which horse arrived first, second, etc., or school grades.

The domain of attributes with the *interval scale* is defined over numerical scale, in such way that differences between arbitrary pairs of values can be meaningfully compared. Therefore, equal differences between interval attribute values represent equal intervals and operations of addition or subtraction on interval attribute values are meaningful. Moreover, obviously, all the relation comparisons valid for nominal or ordinal attributes can also be performed. Operations such as multiplication or division, however, cannot be carried out as there exists an arbitrary zero point on the interval value scales, such as the 0 degrees of Celsius or Fahrenheit on the temperature scale. Therefore, one cannot say that one interval scale attribute value is e.g. double of another one. This limitation of interval scales is well illustrated by a popular joke: “If today we measure 0 degrees Celsius and tomorrow is twice as cold, what temperature do we measure tomorrow?” Among examples of attributes with the interval scale one can also mention year dates in many calendars.

The domain of attributes with the *ratio scale* is defined over numerical scale, such that ratios between arbitrary pairs of values are meaningful. Thus, on ratio scales operations of multiplication or division can be performed, as well as all the operations and comparisons valid for interval scales. The zero value on a ratio scale is non-arbitrary, like in e.g. the Kelvin temperature scale which is proportional to heat content, and zero on that scale really means there is zero heat (zero is absolute). Therefore, one can multiply and divide meaningfully on ratio scales, e.g. removing half the heat content of the air would cause the Kelvin thermometer to register twice as small temperature value. Examples of attributes with ratio scale also contain many physical quantities such as mass or length. Social ratio scales include age, number of class attendances in a particular time, etc.

Apart from the above mentioned attributes, there also exists a type of attributes called criteria, for which the domains are preference-ordered (e.g. from the least wanted to the most wanted). Among scales of criteria one can distinguish ordinal, interval or ratio scales [82]. One can also have partial order on the value set of attribute, e.g. representing a group of attributes.

Some authors also distinguish other attribute types e.g. structural, for which the domain values are characterized by a taxonomy [33], [54].

Association and decision rules. A *rule* induced from a data table S is denoted by $\phi \to \psi$ (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 $\{=, <, \leq, \geq, >\}$ 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 [28], [94] about interpretation of rules as material implications) between premise and conclusion. The attributes that appear in elementary conditions of the premise (conclusion, resp.) are called *condition attributes* (*decision attributes*, resp.). Obviously, within one rule, the sets of condition and decision attributes must be disjoint. The rules induced (mined) from data may be either *decision* or *association* rules, depending on whether the division of A into condition and decision categories of attributes has been fixed or not.

One of the classical examples of data table used in the literature to illustrate algorithms of rule induction concerns playing golf (see Table 1) and was originally introduced by Quinlan [75], [76]. The dataset uses weather information to decide whether or not to play golf. It contains 14 objects (items) described by four attributes concerning the weather state: *outlook* (with nominal values *sunny*, *overcast* or *rain*), *temperature* (with ordinal values *hot*, *mild* or *cold*), *humidity* (with ordinal values *high* or *normal*) and *windy* (with nominal values *true* or *false*). Moreover, there is also a decision attribute *play*? with nominal values *yes* or *no*.

Exemplary decision rules induced from this dataset could be the following:

- ***if*** outlook=overcast ***then*** play=yes,
- ***if*** outlook=sunny ***and*** humidity=normal ***then*** play=yes,
- ***if*** outlook=sunny ***and*** humidity=high ***then*** play=yes,
- ***if*** outlook=rain ***and*** windy=true ***then*** play=no,
- ***if*** outlook=rain ***and*** windy=false ***then*** play=yes.

Such rules could have a descriptive function helping to describe under what weather conditions people are willing to play golf, or a predictive function helping to forecast whether people will tend to play golf if certain weather conditions occur.

Table 1. A data table describing the influence of the weather conditions on the decision whether or not to play golf

outlook	temperature	humidity	windy	play?
sunny	hot	high	false	no
sunny	hot	high	true	no
overcast	hot	high	false	yes
rain	mild	high	false	yes
rain	cold	normal	false	yes
rain	cold	normal	true	no
overcast	cold	normal	true	yes
sunny	mild	high	false	no
sunny	cold	normal	false	yes
rain	mild	normal	false	yes
sunny	mild	normal	true	yes
overcast	mild	high	true	yes
overcast	hot	normal	false	yes
rain	mild	high	true	no

2.1 Attractiveness measures for decision and association rules

The number of rules induced from massive datasets usually is so large that it overwhelms the human comprehension capabilities, and, moreover, vast majority of them have very little value in practice. Thus, in order to increase the relevance and utility of selected rules and limit the size of the resulting rule set, *quantitative* measures, also known as *attractiveness* or *interestingness* measures, have been proposed and widely studied in literature [1], [45], [50], [91]. The variety of proposed measures comes as a result of looking for means of reflecting particular characteristics of rules or sets of rules. Most of the attractiveness measures are *gain-type* criteria, which means that the higher values they obtain, the greater is the utility, interestingness of an evaluated rule. However, in literature there are also measures which are considered as *cost-type* criteria, i.e. the smaller the value of the measure for a given rule, the more attractive the rule is.

Below, there are definitions of the attractiveness measures that are analyzed in this article.

Rule support. One of the most popular measures used to identify frequently occurring rules in sets of items from data table S is the support considered by Jan Łukasiewicz in [51] and later used by data miners e.g. in [2]. *Support* of condition ϕ (analogously, ψ), denoted as $sup_S(\phi)$ (analogously, $sup_S(\psi)$), is equal to the number of objects in U satisfying ϕ (analogously, ψ). The *support of rule* $\phi \rightarrow \psi$ (also simply referred to as support), denoted as $sup_S(\phi \rightarrow \psi)$, is the number of objects in U satisfying both ϕ and ψ. Thus, it corresponds to statistical significance [37]. The domain of the measure of support can cover any natural number. The greater the value of support for a given rule, the more desirable the rule is, thus, support is a gain-type criterion.

Example 2.1. Let Table 1 represent a data table S. Let us consider two rules induced from Table 1:

– r_1: ***if*** outlook=overcast ***then*** play=yes,
– r_2: ***if*** outlook=sunny ***and*** humidity=normal ***then*** play=yes.

On the basis of Table 1 we can calculate that $sup_S(r_1) = 4$ as there are four objects supporting r_1 (i.e. objects with "overcast" value of the attribute outlook and at the same time with "yes" value for the decision attribute). In case of the second rule: $sup_S(r_2) = 2$. Thus, rule r_1 is more interesting (attractive) than r_2 in the sense of support.

Some authors define support as a relative value with respect to the number of all objects in the dataset U. Then, the rule support can be interpreted as the percentage of objects satisfying both the premise and conclusion of the rule, in the dataset. Throughout this article we will only consider the former definition of support.

Rule anti-support. *Anti-support* of a rule $\phi \to \psi$ (also simply referred to as *anti-support*), denoted as $\textit{anti-sup}_S(\phi \to \psi)$, 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 in the data table S, i.e. objects for which the premise ϕ evaluates to true but whose conclusion is different than ψ. Note that anti-support can also be regarded as $sup_S(\phi \to \neg\psi)$.

Similarly to support, the anti-support measure can obtain any natural value. However, its optimal value is 0. Any value greater than zero means that the considered rule is not certain i.e. there are some counter-examples for that rule. The less counter-examples we observe in the dataset, the better, and therefore anti-support is considered a cost-type criterion.

Example 2.2. Let Table 1 represent a data table S. Let us consider two rules induced from Table 1:

– r_1: ***if*** outlook=overcast ***then*** play=yes,
– r_2: ***if*** humidity=high ***then*** play=no.

On the basis of Table 1 we can observe that there are no counter-examples for r_1 (there are no objects in Table 1 for which outlook = overcast and play$\neq$yes), and thus $\textit{anti-sup}_S(r_1) = 0$. Rule r_2, however, is not pure as there are three counter-examples, which means that $\textit{anti-sup}_S(r_2) = 3$. Thus, from the view point of anti-support, r_1 is more attractive than r_2.

Similarly to support, one can also define anti-support as a relative value with respect to the number of all objects in the dataset U. Then, the rule anti-support can be considered as the percentage of counter-examples, in the dataset. Again, throughout this work we will only focus on the former definition of anti-support.

Confidence. Among measures very commonly associated with rules induced from data table S, there is also confidence [51], [2]. The *confidence* of a rule (also called *certainty*), denoted as $conf_S(\phi \to \psi)$, is defined as follows:

$$conf_S(\phi \to \psi) = \frac{sup_S(\phi \to \psi)}{sup_S(\phi)}. \tag{2}$$

Obviously, when considering rule $\phi \to \psi$, it is necessary to assume that the set of objects having property ϕ in U is not empty, i.e. $sup_S(\phi) \neq 0$.

Under the "closed world assumption" [77] (which is the presumption that what is not currently known to be true is false) adopted in rule induction, and because U is a finite set, it is legitimate to express probabilities $Pr(\phi)$ and $Pr(\psi)$ in terms of frequencies $sup_S(\phi)/|U|$ and $sup_S(\psi)/|U|$, respectively. In consequence, the confidence measure $conf_S(\phi \to \psi)$ can be regarded as conditional probability $Pr(\psi|\phi) = Pr(\phi \wedge \psi)/Pr(\phi)$ with which conclusion ψ evaluates to *true*, provided that premise ϕ evaluates to *true*.

Moreover, let us point out the relationship between confidence and another attractiveness measure called *coverage*, denoted as $cov_S(\phi \to \psi)$ and defined for a data table S in the following manner:

$$cov_S(\phi \to \psi) = \frac{sup_S(\psi \to \phi)}{sup_S(\psi)}. \tag{3}$$

Since $sup_S(\phi \to \psi) = sup_S(\psi \to \phi)$ as they both express the number of objects satisfying both ϕ and ψ, it is clear that confidence of a rule $\phi \to \psi$ can be regarded as coverage for rule $\psi \to \phi$.

Confidence takes any value between 0 and 1. It is a gain-type criterion and thus, the most desirable value is 1, which reflects the situation in which all objects that satisfy the premise also support the whole rule (i.e. both the premise and conclusion). Let us note that confidence is equal to 1 only when the anti-support is 0.

Example 2.3. Let Table 1 represent a data table S. Let us consider two rules induced from Table 1:

- r_1: ***if*** outlook=overcast ***then*** play=yes,
- r_2: ***if*** humidity=high ***then*** play=no.

Since there are no counter-examples for rule r_1 ($anti\text{-}sup_S(r_1) = 0$), $conf_S(r_1) = \frac{4}{4} = 1$. For rule r_2, however, there are 3 counter-examples, which implies that confidence for this rule will not be 1. There are 7 objects supporting r_2's premise but only four of them support the whole rule ($sup_S(r_2) = 4$). Thus, $conf_S(r_2) = \frac{4}{7}$. It is, therefore, clear that in the considered S, r_1 is better than r_2 with respect to confidence.

Rule interest function. The *rule interest function RI* introduced by Piatetsky-Shapiro in [72] is used to quantify the correlation between the premise and conclusion in the data table S. It is defined by the following formula:

$$RI_S(\phi \to \psi) = sup_S(\phi \to \psi) - \frac{sup_S(\phi)\ sup_S(\psi)}{|U|}. \tag{4}$$

For rule $\phi \to \psi$, when $RI = 0$, then ϕ and ψ are statistically independent (i.e. the occurrence of the premise makes the conclusion neither more nor less probable) 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 [37]. Obviously, it is a gain-type criterion as greater values of RI reflect stronger trend towards positive correlation.

After simple algebraic transformation, RI can also be expressed as:

$$RI_S(\phi \to \psi) = \frac{sup_S(\phi \to \psi)\, sup_S(\neg\phi \to \neg\psi) - sup_S(\neg\phi \to \psi)\, sup_S(\phi \to \neg\psi)}{|U|}. \tag{5}$$

Now, one can also analyse RI's sign (interpreted as positive or negative correlation) as verification of the sign of the above nominator, i.e. the sign of the difference: $sup_S(\phi \to \psi)sup_S(\neg\phi \to \neg\psi) - sup_S(\neg\phi \to \psi)sup_S(\phi \to \neg\psi)$.

Example 2.4. Let Table 1 represent a data table S. Let us consider two rules induced from Table 1:

- r_1: ***if*** outlook=overcast ***then*** play=yes,
- r_2: ***if*** outlook=sunny ***and*** humidity=normal ***then*** play=yes

From Table 1 we obtain:

$$RI_S(r_1) = 4 - (4 * 9/14) = 1.42 \text{ and } RI_S(r_2) = 2 - (2 * 9/14) = 0.71.$$

These results show that in both of the considered rules the premises are positively correlated with the conclusions, however the correlation in r_1 is stronger.

Gain function. For a data table S the *gain* function of Fukuda et al. [25] is defined in the following manner:

$$gain_S(\phi \to \psi) = sup_S(\phi \to \psi) - \Theta\ sup_S(\phi) \tag{6}$$

where Θ is a fractional constant between 0 and 1.

Note that, for a fixed value of $\Theta = sup_S(\psi)/|U|$, the gain measure becomes identical to the presented above rule interest function RI. Moreover, if Θ is zero then, gain boils down to calculation of the support of the rule, and when Θ is equal to 1, gain will take negative values unless all objects satisfying ϕ also satisfy ψ (in that case gain will be 0). Thus, gain can take any integer value depending on what value Θ is set at. For a fixed Θ, greater values of gain are more desirable, thus it is a gain-type criterion.

Example 2.5. Let Table 1 represent a data table S. We consider two rules induced from Table 1:

– r_1: ***if*** outlook=overcast ***then*** play=yes,
– r_2: ***if*** humidity=high ***then*** play=no.

Let us assume $\Theta = 0.5$ (such value means that the value of $sup_S(\phi \rightarrow \psi)$ is twice as important to us as the value of $sup_S(\phi)$). Then, from Table 1 we obtain:

$$gain_S(r_1) = 4 - 0.5 * 4 = 2 \text{ and } gain_S(r_2) = 4 - 0.5 * 7 = 0.5.$$

In this example both of the considered rules had the same value of $sup_S(\phi \rightarrow \psi)$, however, for r_2 there were also some counter-examples. The existence of counter-examples causes the difference between the value of $sup_S(\phi \rightarrow \psi)$ and $sup_S(\phi)$, which directly influences the value of the gain measure. In this example, for the same value of Θ, $gain_S(r_1) > gain_S(r_2)$, thus we can conclude that r_1 is a more interesting rule with respect to gain.

Dependency factor. For a data table S the *dependency factor* of Pawlak [69] (also considered by Popper [74]) is defined in the following manner:

$$\eta_S\left(\phi \rightarrow \psi\right) = \frac{\frac{sup_S(\phi \rightarrow \psi)}{sup_S(\phi)} - \frac{sup_S(\psi)}{|U|}}{\frac{sup_S(\phi \rightarrow \psi)}{sup_S(\phi)} + \frac{sup_S(\psi)}{|U|}}. \tag{7}$$

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 $\eta_S(\phi \rightarrow \psi) = 0$. If $-1 < \eta_S(\phi \rightarrow \psi) < 0$, then ϕ and ψ are negatively dependent (i.e. the occurrence of ϕ decreases the probability of ψ), and if $0 < \eta_S(\phi \rightarrow \psi) < 1$, then ϕ and ψ are positively dependent on each other (i.e. the occurrence of ϕ increases the probability of ψ). The dependency factor is a gain-type criterion.

Example 2.6. Let Table 1 represent a data table S. Let us consider two rules induced from Table 1:

– r_1: ***if*** outlook=overcast ***then*** play=yes,
– r_2: ***if*** outlook=sunny ***and*** humidity=normal ***then*** play=yes.

From Table 1 we obtain:

$$\eta_S\left(r_1\right) = \frac{\frac{4}{4} - \frac{9}{14}}{\frac{4}{4} + \frac{9}{14}} = \frac{5}{23} = 0.217 \text{ and } \eta_S\left(r_2\right) = \frac{\frac{2}{2} - \frac{9}{14}}{\frac{2}{2} + \frac{9}{14}} = \frac{5}{23} = 0.217.$$

The results show that, from the viewpoint of dependency factor, both of the considered rules are of equal attractiveness. It is due to the fact that they have the same conclusion and do not have any counter-examples. The positive value of η reflects positive correlation between the premises in r_1 and r_2, and the conclusion.

Measures f and s. Among the best-known and widely studied confirmation measures (see the definition in Section *Property of Bayesian Confirmation.*), there are measures denoted by f and s, defined as follows (for a data table S):

$$f_S(\phi \to \psi) = \frac{Pr(\phi|\psi) - Pr(\phi|\neg\psi)}{Pr(\phi|\psi) + Pr(\phi|\neg\psi)}, \tag{8}$$

$$s_S(\phi \to \psi) = Pr(\psi|\phi) - Pr(\psi|\neg\phi). \tag{9}$$

Taking into account that conditional probability $Pr(\circ|*) = conf_S(\circ \to *)$, measures f and s can be re-written as:

$$f_S(\phi \to \psi) = \frac{conf_S(\psi \to \phi) - conf_S(\neg\psi \to \phi)}{conf_S(\psi \to \phi) + conf_S(\neg\psi \to \phi)}, \tag{10}$$

$$s_S(\phi \to \psi) = conf_S(\phi \to \psi) - conf_S(\neg\phi \to \psi). \tag{11}$$

Among authors advocating for measure f, there are Kemeny and Oppenheim [46], Good [27], Heckerman [35], Horvitz and Heckerman [39], Pearl [71], Schum [80] and Fitelson [22]. Measure s has been proposed by Christensen [11] and Joyce [42]. It is worth noting that confirmation measure f is monotone (and therefore gives the same ranking) with respect to the Bayes factor originally proposed by Jeffrey [41] and reconsidered as an interestingness measure by Kamber and Shingal [43]. The *Bayes factor* is defined by the following formula:

$$k_S(\phi \to \psi) = \frac{conf_S\,(\psi \to \phi)}{conf_S\,(\neg\psi \to \phi)}.$$

Measures f and s are regarded as gain-type measures quantifying the degree to which the premise ϕ provides "support for or against" the conclusion ψ. Thus, they obtain positive values (precisely, [1, 0]) iff the premise ϕ confirms the conclusion ψ i.e. iff $Pr(\psi|\phi) < Pr(\psi)$. Measures f and s take negative values (precisely, [-1, 0[) when the premise ϕ disconfirms the conclusion ψ, i.e. iff $Pr(\psi|\phi) < Pr(\psi)$. In literature, these measures are found as a powerful tool for analyzing the confirmation of conclusion by a rule's premise.

Example 2.7. Let Table 1 represent a data table S. Let us consider two rules induced from Table 1:

- r_1: ***if*** outlook=overcast ***then*** play=yes,
- r_2: ***if*** humidity=high ***then*** play=yes.

On the basis of Table 1 we can calculate that:

$$f_S(r_1) = \frac{\frac{4}{9} - \frac{0}{5}}{\frac{4}{9} + \frac{0}{5}} = 1 \text{ and } f_S(r_2) = \frac{\frac{3}{9} - \frac{4}{5}}{\frac{3}{9} + \frac{4}{5}} = -\frac{21}{51} = -0.41$$

$$s_S(r_1) = \frac{4}{4} - \frac{5}{10} = 1/2 \text{ and } s_S(r_2) = \frac{3}{7} - \frac{6}{7} = -3/7.$$

For the first rule there are no counter-examples (the rule is certain) which means that the premise confirms the conclusion. This fact is reflected by positive values of measures f and s. In case of r_2, one can observe in Table 1 even more counter-examples than examples actually supporting the rule ($sup_S(\phi \to \neg\psi) > sup_S(\phi \to \psi)$). Thus, in r_2 the premise disconfirms the conclusion, which is expressed by negative values of measures f and s.

2.2 Desirable properties of attractiveness measures

While choosing attractiveness measures for a certain application one also considers their properties (features), which express the user's expectations towards the behavior of measures in particular situations. Those expectations can be of various types, e.g. one can desire to use only such measures that have the property of not going further away (or even of coming closer to) from their optimal value for a certain induced rule when the number of objects supporting the pattern increases.

Properties group the attractiveness measures according to similarities in their characteristics. Using the measures which satisfy the desirable properties one can avoid considering unimportant rules. Therefore, knowledge of which commonly used interestingness measures satisfy certain valuable properties, is of high practical and theoretical importance.

Property M. Greco, Pawlak and Słowiński in [28] analyzed measuring of attractiveness of rules. They proposed a valuable *property M* of monotonic dependency of an attractiveness measure on the number of objects satisfying or not the premise or the conclusion of a rule. Property M makes use of elementary parameters of the considered dataset (numbers of objects satisfying some properties) and therefore is an easy and intuitive criterion helping to choose an appropriate attractiveness measure for a certain application [28], [8].

Formally, an attractiveness measure

$$\begin{aligned} &I_S(\phi \to \psi) = \\ &= F\left[sup_S\left(\phi \to \psi\right), sup_S\left(\neg\phi \to \psi\right), sup_S\left(\phi \to \neg\psi\right), sup_S\left(\neg\phi \to \neg\psi\right)\right] \end{aligned} \tag{12}$$

being a gain-type criterion (i.e. the higher the value of the measure, the better) has the property M iff it is a function:

- non-decreasing with respect to $sup_S(\phi \to \psi)$, and
- non-increasing with respect to $sup_S(\neg\phi \to \psi)$, and
- non-increasing with respect to $sup_S(\phi \to \neg\psi)$, and
- non-decreasing with respect to $sup_S(\neg\phi \to \neg\psi)$.

Respectively, an attractiveness measure

$$\begin{aligned} &I_S(\phi \to \psi) = \\ &= F\left[sup_S\left(\phi \to \psi\right), sup_S\left(\neg\phi \to \psi\right), sup_S\left(\phi \to \neg\psi\right), sup_S\left(\neg\phi \to \neg\psi\right)\right] \end{aligned} \tag{13}$$

being a cost-type criterion (i.e. the lower the value of the measure, the better) has the property M iff it is a function:

- non-increasing with respect to $sup_S(\phi \to \psi)$, and
- non-decreasing with respect to $sup_S(\neg\phi \to \psi)$, and
- non-decreasing with respect to $sup_S(\phi \to \neg\psi)$, and
- non-increasing with respect to $sup_S(\neg\phi \to \neg\psi)$.

Most of the considered attractiveness measures are gain-type criteria and therefore, below we will present the interpretation of property M only for this type of measures (for cost-type criteria, the considerations are analogous).

The property M with respect to $sup_S(\phi \to \psi)$ (or, analogously, with respect to $sup_S(\neg\phi \to \neg\psi)$) means that any object in the dataset for which ϕ and ψ (or, analogously, neither ϕ nor ψ) hold together, increases (or at least does not decrease) the attractiveness of the rule $\phi \to \psi$. On the other hand, the property M with respect to $sup_S(\neg\phi \to \psi)$ (or, analogously, with respect to $sup_S(\phi \to \neg\psi)$) means that any object for 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 attractiveness of the rule $\phi \to \psi$.

Let us use the following example mentioned by Hempel [36] to show the interpretation of the property. Consider a rule $\phi \to \psi$: *if x is a raven then x is black*. In this case ϕ stands for being a raven and ψ stands for being black. If an attractiveness measure $I_S(\phi \to \psi)$ (being a gain-type criterion) possesses the property M then:

- the more black ravens or non-black non-ravens there will be in the data table S, the more attractive the rule will become, and thus $I_S(\phi \to \psi)$ will obtain greater value,
- the more black non-ravens or non-black ravens in the data table S, the less attractive the rule will become and thus, the value of $I_S(\phi \to \psi)$ will become smaller.

Greco, Pawlak and Słowiński [28] have considered attractiveness measures with respect to property M. The results they obtained show that measures f and l [46], [27], [35], [39], as well as s [11], [42] possess the property M, while measures d [17], [18], [41] , [78], r [38], [44], [52], [79], [73], b [10] do not.

Property of Bayesian confirmation. Formally, an attractiveness measure $c_S(\phi \to \psi)$ has the property of Bayesian confirmation (or simply confirmation) iff it satisfies the following conditions:

$$c_S(\phi \to \psi) \begin{cases} > 0 \;\; if \; Pr(\psi|\phi) > Pr(\psi), \\ = 0 \;\; if \; Pr(\psi|\phi) = Pr(\psi), \\ < 0 \;\; if \; Pr(\psi|\phi) < Pr(\psi). \end{cases} \tag{14}$$

Since the conditional probability $Pr(\psi|\phi) = Pr(\phi \wedge \psi)/Pr(\phi)$ can be regarded as the confidence measure $conf_S(\phi \to \psi)$, the above definition can be re–written as:

$$c_S(\phi \to \psi) \begin{cases} > 0 \ \ if\ conf_S(\phi \to \psi) > sup_S(\psi)/\left|U\right| \ , \\ = 0 \ \ if\ conf_S(\phi \to \psi) = sup_S(\psi)/\left|U\right| \ , \\ < 0 \ \ if\ conf_S(\phi \to \psi) < sup_S(\psi)/\left|U\right| \ . \end{cases} \tag{15}$$

Measures that possess the property of confirmation are referred to as *confirmation measures* or *measures of confirmation.* According to Fitelson [22], measures of confirmation quantify the degree to which a premise ϕ provides "support for or against" a conclusion ψ. When their values are greater than zero, it means that the conclusion is satisfied more frequently when the premise is satisfied, rather than generally in the whole dataset. Measures of confirmation equal to zero, reflect that fulfilment of the premise imposes no influence on fulfilment of the conclusion. Analogously, when the value of confirmation measure is smaller than zero, it means that the premise only disconfirms the conclusion as the conclusion is satisfied less frequently when the premise is satisfied, rather than generally in the whole dataset. Thus, for a given rule $\phi \to \psi$, attractiveness measures with the property of confirmation express the credibility of the following proposition: *ψ is satisfied more frequently when ϕ is satisfied, rather than when ϕ is not satisfied.* This interpretation stresses the very valuable semantics of the property of confirmation. By using the attractiveness measures that possess this property one can filter out rules which are misleading and disconfirm the user, and this way, limit the set of induced rules only to those that are meaningful.

Among commonly used and discussed Bayesian confirmation measures there are the following measures: f and l [46], [27], [35], [39], [22], s [11], [42] d [17], [18], [41] , [78], r [38], [44], [52], [79], [73], and b [10].

Property of hypothesis symmetry. Many authors have also considered properties of symmetry of attractiveness measures. Eells and Fitelson have analyzed in [17] a set of best-known confirmation measures from the viewpoint of the following four properties of symmetry introduced by Carnap in [10]:

1. evidence symmetry (ES): $I_S(\phi \to \psi) = -I_S(\neg\phi \to \psi)$
2. commutativity symmetry (CS): $I_S(\phi \to \psi) = I_S(\psi \to \phi)$
3. hypothesis symmetry (HS): $I_S(\phi \to \psi) = -I_S(\phi \to \neg\psi)$
4. total symmetry (TS): $I_S(\phi \to \psi) = -I_S(\neg\phi \to \neg\psi)$

Eells and Fitelson remark in [17] that given CS, ES and HS are equivalent i.e. provided that $I_S(\phi \to \psi) = I_S(\psi \to \phi)$, $-I_S(\neg\phi \to \psi) = -I_S(\phi \to \neg\psi)$. Moreover, they show that that TS follows from the conjunction of ES and HS.

They also conclude that, in fact, only HS is a desirable property, while ES, CS and TS are not. The meaning behind the hypothesis symmetry is that the influence of the premise on the conclusion part of a rule should be of the opposite sign, as the influence of the premise on a negated conclusion.

The arguments against ES, CS and TS can be presented by an exemplary situation of randomly drawing a card from a standard deck ([17], [28]). Let ϕ stand for that the drawn *card is the seven of spades*, and let ψ be the hypothesis that *the card is black*. Despite the strong confirmation of ψ by ϕ, the negated premise is useless to the conclusion as the evidence the *card is not the seven of spades* ($\neg\phi$) is practically of no value to the conclusion *the card is black* (ψ). Thus, the ES is not valid. Continuing this example one can observe that the evidence that the *card is black* (ψ) does not confirm the hypothesis that the *card is the seven of spades* (ϕ) to the same extent as the evidence that the *card is the seven of spades* (ϕ), confirms the hypothesis that *the card is black* (ψ). This means that CS is not valid. Analogously, arguments against TS can be shown.

The above mentioned example is also an argument <u>for</u> the hypothesis symmetry as, obviously, the evidence that the *card is the seven of spades* (ϕ) is negatively conclusive for the hypothesis that the *card is not black* ($\neg\psi$).

Having considered popular confirmation measures with respect to symmetry properties, Fitelson [22] concluded that measures f and l [46], [27], [35], [39], as well as s [11], [42] and d [17], [18], [41] , [78], satisfy the property of hypothesis symmetry.

3 Analyses of properties of particular attractiveness measures

Analyses verifying whether popular attractiveness measures possess valuable properties widen our understanding of those measures and of their applicability. Moreover, through such property analysis one can also learn about relationships between different measures. The obtained results are useful for practical applications because they show which interestingness measures are relevant for meaningful rule evaluation. Using the measures which satisfy the desirable properties one can avoid analysing unimportant rules.

Many authors have considered different attractiveness measures with respect to several properties ([17], [22], [28]). However, analysis of property M, property of confirmation and property of hypothesis symmetry for many popular attractiveness measures still remains an open problem. In the following section we shall provide answers for some of those open questions.

For the sake of the clarity of presentation, the following notation shall be used throughout the next sections:

$$\begin{array}{lll} a = sup_S(\phi \to \psi), & b = sup_S(\neg\phi \to \psi), & c = sup_S(\phi \to \neg\psi), \\ d = sup_S(\neg\phi \to \neg\psi), & a + c = sup_S(\phi), & a + b = sup_S(\psi), \\ b + d = sup_S(\neg\phi), & c + d = sup_S(\neg\psi), & a + b + c + d = |U|. \end{array} \tag{16}$$

We also assume that set U is not empty, so that at least one of a, b, c, d is strictly positive.

3.1 Analysis of measures with respect to property M

In order to prove that for a data table S a gain-type measure $I_S(\phi \to \psi)$ 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. It means that all of the following conditions must be satisfied:

$$\left.\begin{array}{l}\text{1. the increase of } a \text{ does not result in decrease of the measure,}\\ \text{2. the increase of } b \text{ does not result in increase of the measure,}\\ \text{3. the increase of } c \text{ does not result in increase of the measure,}\\ \text{4. the increase of } d \text{ does not result in decrease of the measure.}\end{array}\right\} \quad (17)$$

In case of a cost-type measures $J_S(\phi \rightarrow \psi)$, we will say that it possesses the property M iff the following conditions will be fulfilled:

$$\left.\begin{array}{l}\text{1. the increase of } a \text{ does not result in increase of the measure,}\\ \text{2. the increase of } b \text{ does not result in decrease of the measure,}\\ \text{3. the increase of } c \text{ does not result in decrease of the measure,}\\ \text{4. the increase of } d \text{ does not result in increase of the measure.}\end{array}\right\} \quad (18)$$

During the analysis of measures with respect to property M we consider an increase of only one parameter at a time, e.g. if a increases, b, c and d remain unchanged. An increase of a by $\Delta > 0$ (and analogously an increase of b, c or d) is a result of adding to U objects that satisfy both ϕ and ψ. It means that the data table $S = (U, A)$ changes to $S' = (U', A)$, where $|U| = a + b + c + d$ and $|U'| = (a + \Delta) + b + c + d$.

Rule support. According to the notation in (16) rule support $sup_S(\phi \rightarrow \psi)$ is a. Thus, obviously, rule support, being a gain-type criterion, increases with a and does not change (i.e. neither decreases nor increases) with b, c, or d. Therefore, it is legitimate to conclude that the measure of rule support has the property M.

Rule anti-support. Anti-support is a cost-type criterion and therefore the conditions (18) need to be verified. Since anti-support can be regarded as the number of counter examples, $anti\text{-}sup_S(\phi \rightarrow \psi) = c$. Thus, obviously, $anti\text{-}sup_S(\phi \rightarrow \psi)$ increases with c and does not does not change with a, b, or d. Therefore, it can be concluded that $anti\text{-}sup_S(\phi \rightarrow \psi)$ has the property M.

Confidence. Now, let us consider confidence with respect to the property M. Confidence is a gain-type criterion dependent only on $sup_S(\phi \rightarrow \psi)$ and $sup_S(\phi)$, therefore the analysis of $conf_S(\phi \rightarrow \psi)$ with respect to the property M can be practically narrowed down to analysis of its dependence on the number of objects supporting both the premise and conclusion, and on the number of objects satisfying the premise but not the conclusion.

Theorem 3.1. *Confidence measure has the property M.*

Proof. Let us consider confidence expressed in the notation (16):

$$conf_S\,(\phi \rightarrow \psi) = \frac{sup_S\,(\phi \rightarrow \psi)}{sup_S\,(\phi)} = \frac{a}{a+c}. \tag{19}$$

As $conf_S(\phi \rightarrow \psi)$ does not depend on b nor d, it is clear that conditions (17).2 and (17).4 are satisfied. However, conditions (17).1 and (17).3 require verification:

Condition (17).1:
Let us assume that the considered data table $S = (U, A)$ is extended to $S' = (U', A)$ by adding to S some objects satisfying both ϕ and ψ. This addition increases the value of a to $a + \Delta$ where $\Delta > 0$. Condition (17).1 will be satisfied if and only if

$$conf_S(\phi \rightarrow \psi) = \frac{a}{a+c} \leq conf_{S'}(\phi \rightarrow \psi) = \frac{(a+\Delta)}{(a+\Delta)+c}.$$

We can easily calculate that

$$\frac{a}{a+c} \leq \frac{(a+\Delta)}{(a+\Delta)+c} \Leftrightarrow a(a+c+\Delta) \leq (a+c)(a+\Delta) \Leftrightarrow$$

$$a^2 + ac + a\Delta \leq a^2 + ac + a\Delta + c\Delta \Leftrightarrow c\Delta \geq 0.$$

Since both c and Δ are numbers greater than 0, the last inequality is always fulfilled, and therefore condition (17).1 is satisfied.

Condition (17).3:
Let us consider confidence given as:

$$conf_S(\phi \rightarrow \psi) = \frac{a}{a+c}.$$

Since c is in the denominator of the confidence measure, the increase of c will result in the decrease of confidence, and therefore condition (17).3 is satisfied.

Since all four conditions are satisfied, the hypothesis that confidence has the property M is true. □

Rule interest function. Let us now focus on the rule interest function and analyze it with respect to the property M. Such analysis will require verification of RI's dependency on the change of a, b, c and d, i.e. all (17) conditions.

Theorem 3.2. *Rule interest function has the property M. [29]*

Proof. Let us observe that according to notation (16) measure RI can be rewritten as:

$$RI_S(\phi \rightarrow \psi) = a - \frac{(a+b)\,(a+c)}{a+b+c+d}. \tag{20}$$

After some simple algebraic transformation, we obtain

$$RI_S(\phi \to \psi) = \frac{ad - bc}{a + b + c + d} \tag{21}$$

Taking into account condition (17).1, to prove the monotonicity of RI with respect to a we have to show that if a increases by $\Delta > 0$, then RI does not decrease, i.e.

$$\frac{(a + \Delta)\, d - bc}{a + b + c + d + \Delta} - \frac{ad - bc}{a + b + c + d} \geq 0.$$

The increase of a is a result of adding to U objects that satisfy both ϕ and ψ, i.e. a result of extending the data table $S = (U, A)$ to $S' = (U', A)$. After few simple algebraic passages, and remembering that a, b, c, d are non-negative, we get

$$\frac{(a + \Delta)\, d - bc}{a + b + c + d + \Delta} - \frac{ad - bc}{a + b + c + d} = \frac{d\,(b + c + d)\,\Delta + bc\Delta}{(a + b + c + d)\,(a + b + c + d + \Delta)} > 0$$

such that we can conclude that RI is non-decreasing (more precisely, strictly increasing) with respect to a. Analogous proofs hold for the monotonicity of RI with respect to b, c and d. $\square$

Gain function. We shall now consider gain function with respect to property M. Similarly to confidence, the gain function is a gain-type criterion that only depends on $sup_S(\phi \to \psi)$ and $sup_S(\phi)$. Thus, analysis of gain function with respect to the property M boils down to analysis of its dependence on a and c.

Theorem 3.3. *Gain function has the property M.*

Proof. Let us consider gain function expressed in notation (16):

$$gain_S(\phi \to \psi) = a - \Theta(a + c) \tag{22}$$

where Θ is a fractional constant between 0 and 1.

As $gain_S(\phi \to \psi)$ does not depend on b nor d, it is clear that the change of b or d does not result in any change of $gain_S(\phi \to \psi)$. Thus, we only have to verify if conditions (17).1 and (17).3 hold.

Condition (17).1:
Let us assume that the considered data table $S = (U, A)$ is extended to $S' = (U', A)$ by adding to S some objects satisfying both ϕ and ψ. Those objects increase the value of a to $a + \Delta$ where $\Delta > 0$. The condition will be satisfied if and only if

$$gain_S(\phi \to \psi) = a - \Theta(a + c) \leq gain_{S'}(\phi \to \psi) = (a + \Delta) - \Theta(a + \Delta + c).$$

Let us observe that

$$a{-}\Theta(a{+}c) \le (a{+}\Delta){-}\Theta(a{+}\Delta{+}c) \Leftrightarrow a - a\,\Theta{-}c\,\Theta \le a{+}\Delta{-}a\,\Theta{-}c\Theta - \Theta\Delta \Leftrightarrow$$

$$\Delta - \Theta\Delta \ge 0 \Leftrightarrow \Delta(1-\Theta) \ge 0.$$

The last inequality is always satisfied as $\Delta > 0$ and $(1-\Theta) \ge 0$ because Θ is a fractional constant between 0 and 1. Thus, condition (17).1 is satisfied.

Condition (17).3:
Let us assume that the considered data table $S = (U, A)$ is extended to $S' = (U', A)$ by adding to S some objects satisfying ϕ but not satisfying ψ. That addition increases the value of c to $c + \Delta$ where $\Delta > 0$. Condition (17).3 will be satisfied if and only if

$$gain_S(\phi \to \psi) = a{-}\Theta(a+c) \ge gain_{S'}(\phi \to \psi) = a{-}\Theta(a+\Delta+c).$$

Let us observe that:

$$a{-}\Theta(a+c) \ge a{-}\Theta(a+\Delta+c) \Leftrightarrow$$

$$\Leftrightarrow a - a\,\Theta{-}c\,\Theta \ge a - a\,\Theta{-}c\,\Theta - \Theta\Delta \Leftrightarrow \Theta\Delta \ge 0.$$

The last inequality is always satisfied as $\Delta > 0$ and $\Theta \ge 0$. Thus, condition (17).3 is satisfied.

Since all four conditions are satisfied, the gain function has the property M.

$\square$

Dependency factor. Let us, now analyse dependency factor with respect to the property M. The measure will satisfy the property only when all conditions (17) are fulfilled.

Theorem 3.4. *Dependency factor $\eta_S(\phi \to \psi)$ does not have the property M [29].*

Proof. Let us consider the dependency factor rewritten in notation (16):

$$\eta_S\,(\phi \to \psi) = \frac{\frac{a}{a+c} - \frac{a+b}{a+b+c+d}}{\frac{a}{a+c} + \frac{a+b}{a+b+c+d}}. \tag{23}$$

It will be shown by the following counterexample that $\eta_S(\phi \to \psi)$ does not satisfy the condition that the increase of a results in non-decrease of the dependency factor, thus this measure does not have the 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 dependency factor does not have the property M as such increase of a results in the decrease of the measure:

$$\eta_S(\phi \to \psi) = 0.0769 > 0.0756 = \eta_{S'}(\phi \to \psi).$$

$\square$

Measures f and s. Greco et al. have considered in [28] different measures from the perspective of property M. They have proved that measures f and s satisfy this property.

3.2 Analysis of measures with respect to property of confirmation

To prove that a measure has the property of confirmation the following conditions need to be verified:

$$\begin{array}{l} \text{1. the measure takes positive values iff } conf_S(\phi \to \psi) > sup_S(\psi)/|U| \\ \text{2. the measure value} = 0 \text{ iff } conf_S(\phi \to \psi) = sup_S(\psi)/|U| \\ \text{3. the measure takes negative values iff } conf_S(\phi \to \psi) < sup_S(\psi)/|U| \end{array} \tag{24}$$

If all those conditions are satisfied, then a measure is said to have the property of confirmation.

Rule support, anti-support and confidence. The domains of the attractiveness measures of support, anti-support and confidence are restricted to non-negative values only. Therefore, none of these measures can satisfy the last condition of (24). Hence, these simple measures do not have the property of confirmation.

Rule interest function. Let us consider the rule interest function with respect to the property of confirmation.

Theorem 3.5. *Rule interest function has the property of confirmation.*

Proof. Let us consider rule interest function given by formula (4). Let us observe that according to condition (24).1:

$$\begin{array}{c} RI_S\,(\phi \to \psi) = sup_S\,(\phi \to \psi) - \frac{sup_S(\phi)\; sup_S(\psi)}{|U|} > 0 \\ \Leftrightarrow \\ conf_S\,(\phi \to \psi) > \frac{sup_S(\psi)}{|U|}. \end{array} \tag{25}$$

Since

$$conf_S\,(\phi \to \psi) = \frac{sup_S\,(\phi \to \psi)}{sup_S\,(\phi)},$$

we can observe that:

$$conf_S\,(\phi \to \psi) > \frac{sup_S\,(\psi)}{|U|} \Leftrightarrow sup_S\,(\phi \to \psi) - \frac{sup_S\,(\phi)\, sup_S\,(\psi)}{|U|} > 0$$

which means that equivalence (25) is always true. Analogous proofs hold for conditions (24).2 and (24).3. □

Gain function. We will now analyse the gain function with respect to the property of confirmation.

Theorem 3.6. *Gain function has the property of confirmation iff* $\Theta = \frac{sup_S(\psi)}{|U|}$.

Proof. Let us consider gain function given by formula (6). Let us first consider condition (24).2, according to which:

$$\begin{gathered} gain_S(\phi \to \psi) = sup_S(\phi \to \psi) - \Theta sup_S(\phi) = 0 \\ \Leftrightarrow \\ conf_S(\phi \to \psi) = \tfrac{sup_S(\psi)}{|U|}. \end{gathered} \tag{26}$$

Since

$$conf_S(\phi \to \psi) = \frac{sup_S(\phi \to \psi)}{sup_S(\phi)},$$

we can observe that:

$$conf_S(\phi \to \psi) = \frac{sup_S(\psi)}{|U|} \Leftrightarrow sup_S(\phi \to \psi) = \frac{sup_S(\phi)\, sup_S(\psi)}{|U|}$$

which means that equivalence (26) can be transformed in the following manner:

$$gain_S(\phi \to \psi) = 0 \Leftrightarrow \frac{sup_S(\phi)\, sup_S(\psi)}{|U|} - \Theta sup_S(\phi) = 0. \tag{27}$$

It is easy to observe that equivalence (27) holds only for $\Theta = sup_S(\psi)/|U|$. In that situation gain function actually boils down to rule interest function, which was proved to be a confirmation measure. Hence, gain function has the property of confirmation if and only if $\Theta = sup_S(\psi)/|U|$. □

Dependency factor. Let us now focus on the dependency factor and analyze it with respect to the property of confirmation.

Theorem 3.7. *Dependency factor has the property of confirmation.*

Proof. Let us consider dependency factor given by formula (7). Let us observe that according to condition (24).1:

$$\eta_S(\phi \to \psi) = \frac{\frac{sup_S(\phi \to \psi)}{sup_S(\phi)} - \frac{sup_S(\psi)}{|U|}}{\frac{sup_S(\phi \to \psi)}{sup_S(\phi)} + \frac{sup_S(\psi)}{|U|}} > 0 \Leftrightarrow conf_S(\phi \to \psi) > \frac{sup_S(\psi)}{|U|}. \tag{28}$$

Since

$$conf_S(\phi \to \psi) = \frac{sup_S(\phi \to \psi)}{sup_S(\phi)} > \frac{sup_S(\psi)}{|U|},$$

it is clear that $\frac{sup_S(\phi \to \psi)}{sup_S(\phi)} - \frac{sup_S(\psi)}{|U|} > 0$.

Thus, both the nominator and the denominator of the dependency factor are positive and we can conclude that equivalence (28) is always true. Analogous proofs hold for conditions (24).2 and (24).3. □

Measures f and s. Among well recognized and established confirmation measures and important role is played by measures f and s. They have been considered as measures with property of confirmation since their introduction in literature and are widely discussed and analyzed by many authors ([9], [22], [49]).

3.3 Analysis of measures with respect to property of hypothesis symmetry

In order to prove that a certain measure has the property of hypothesis symmetry it must be checked if its values for rules $\phi \to \psi$ and $\phi \to \neg\psi$ are the same but of opposite sign.

Rule support, anti-support and confidence. Similarly as in confirmation analysis, in the analysis of property of hypothesis symmetry we sustain the limits introduced by the non-negative domains of support, anti-support and confidence. None of these attractiveness measures has the property of hypothesis symmetry as their values are never negative. E.g. $sup_S(\phi \to \psi) \neq -sup_S(\phi \to \neg\psi)$.

Rule interest function. Let us now analyse if the rule interest function satisfies the property of hypothesis symmetry.

Theorem 3.8. *Rule interest function has the property of hypothesis symmetry [29].*

Proof. Let us consider RI expressed as in (20):

$$RI_S(\phi \to \psi) = a - \frac{(a+c)(a+b)}{a+b+c+d}.$$

For a negated conclusion RI is defined as:

$$RI_S(\phi \to \neg\psi) = c - \frac{(a+c)(c+d)}{a+b+c+d}.$$

The hypothesis symmetry will be satisfied by RI iff:

$$a - \frac{(a+c)(a+b)}{a+b+c+d} = -\left[c - \frac{(a+c)(c+d)}{a+b+c+d}\right].$$

Through simple algebraic transformation we obtain that:

$$a - \frac{(a+c)(a+b)}{a+b+c+d} = -c + \frac{(a+c)(c+d)}{a+b+c+d} = \frac{ad-bc}{a+b+c+d}$$

and, therefore, we can conclude that RI has the property of hypothesis symmetry. $\square$

Gain function. We shall now consider the gain function with respect to the property of hypothesis symmetry.

Theorem 3.9. *Gain function has the property of hypothesis symmetry iff $\Theta = \frac{1}{2}$ [29].*

Proof. Let us consider gain function expressed as in (22):

$$gain_S(\phi \rightarrow \psi) = a - \Theta(a + c).$$

For a negated conclusion gain function is defined as:

$$gain_S(\phi \rightarrow \neg\psi) = c - \Theta(a + c).$$

The hypothesis symmetry will be satisfied by this measure iff:

$$a - \Theta(a + c) = -[c - \Theta(a + c)].$$

Through simple algebraic transformation we obtain that the equality above is satisfied only when $\Theta = 1/2$. □

Dependency factor. Let us now perform the analysis of dependency factor with respect to the property of hypothesis symmetry.

Theorem 3.10. *The dependency factor η does not have the property of hypothesis symmetry [29].*

Proof. Let us consider dependency factor expressed as in (23):

$$\eta_S(\phi \rightarrow \psi) = \frac{\frac{a}{a+c} - \frac{a+b}{a+b+c+d}}{\frac{a}{a+c} + \frac{a+b}{a+b+c+d}}.$$

For a negated conclusion it is defined as:

$$\eta_S(\phi \rightarrow \neg\psi) = \frac{\frac{c}{a+c} - \frac{c+d}{a+b+c+d}}{\frac{c}{a+c} + \frac{c+d}{a+b+c+d}}.$$

To prove that the dependency factor does not satisfy the hypothesis symmetry let us set $a = b = c = 10$ and $d = 20$. We can easily verify that $\eta_S(\phi \rightarrow \psi) = 0.11 \neq 0.09 = \eta_S(\phi \rightarrow \neg\psi)$. □

Measures f and s. Eells et al. have considered in [17] several confirmation measures from the perspective of properties of symmetry. They have proved that measures f and s satisfy the property of hypothesis symmetry.

4 Multicriteria attractiveness evaluation of rules

Application of a measure that quantifies the interestingness of a rule induced from a data table S creates a complete preorder (see formal definition in Section 4.1) on the set of rules. This way the rules are ranked and it is possible to filter out the unwanted ones by setting a threshold on the value of the attractiveness measure.

However, a single attractiveness measure is often not sufficient to evaluate the utility and attractiveness of rules. Thus, multicriteria attractiveness evaluation of rules has become very popular [3], [9], [23], [30], [31], [34], [50]. In this approach, the induced rules are evaluated with respect to many attractiveness measures (criteria) at once and as a result a partial preorder (see formal definition in Section 4.1) on the set of rules is obtained. The implication of many complete preorders by partial preorders is a very interesting problem both form theoretical and practical point of view, however not many authors have tackled it. The next sections will be devoted to discussion about the relationships between different complete and partial preorders and the comparison of sets of rules resulting from different evaluation approaches.

4.1 Definitions of orders and Pareto-optimal set

Complete preorder on set of rules with respect to a single attractiveness measure. Let us denote by v any attractiveness measure that quantifies the interestingness of a rule induced from a data table S. Application of v to a set of induced rules creates a complete preorder, denoted as $\preceq_v$, on that set. Recall that a *complete preorder* on a set X is any binary relation R on X that is strongly complete, (i.e. for all x, $y \in X$, xRy or yRx) and transitive (i.e. for all x, y, $z \in X$, xRy and yRz imply xRz). In simple words, if the semantics of xRy is "x is at most as good as y", then a complete preorder permits to order the elements of X from the best to the worst, with possible ex-aequo but without any incomparability. In other words, considering an attractiveness measure v that induces a complete preorder on a set of rules X and two rules r_1, $r_2 \in X$, rule r_1 is *preferred to* rule r_2 with respect to measure v if $r_1 \succ_v r_2$ and, moreover, rule r_1 is *indifferent* to rule r_2 if $r_1 \sim_v r_2$.

Partial preorder on rules with respect to two attractiveness measures. A *partial preorder* on a set X is any binary relation R on X that is reflexive (i.e. for all $x \in X$, xRx) and transitive. In simple words, if the semantics of xRy is "x is at most as good as y", then a partial preorder permits to order the elements of X from the best to the worst, with possible ex–aequo (i.e. cases of x, $y \in X$ such that xRy and yRx) and with possible incomparability (i.e. cases of x, $y \in X$ such that *not* xRy and *not* yRx).

Let us denote by $\preceq_{qt}$ a partial preorder given by a dominance relation on a set X of rules in terms of any two different attractiveness measures q and t, i.e. for all $r_1,\ r_2 \in X \quad r_1 \preceq_{qt} r_2$ if $r_1 \preceq_q r_2$ and $r_1 \preceq_t r_2$. The partial preorder $\preceq_{qt}$

can be decomposed into its asymmetric part $\prec_{qt}$ and its symmetric part $\sim_{qt}$ in the following manner:

given a set of rules X and two rules $r_1, r_2 \in X$, $r_1 \prec_{qt} r_2$ if and only if

$$\begin{array}{l} q(r_1) \leq q(r_2) \wedge t(r_1) < t(r_2),\ \mathit{or} \\ q(r_1) < q(r_2) \wedge t(r_1) \leq t(r_2), \end{array} \tag{29}$$

moreover $r_1 \sim_{qt} r_2$ if and only if

$$q(r_1) = q(r_2) \wedge t(r_1) = t(r_2). \tag{30}$$

Pareto-optimal border. If for a rule $r \in X$ there does not exist any rule $r' \in X$, such that $r \prec_{qt} r'$ then r is said to be *non–dominated* (i.e. *Pareto–optimal*) with respect to attractiveness measures q and t. A set of all non–dominated rules with respect to q and t forms a *Pareto-optimal border* (*Pareto-optimal set*) of the set of rules in the *q-t* evaluation space and is referred to as a *q-t Pareto–optimal border.*

Monotonicity of a function in its argument. Let $(X, \succ)$ be a pair where X is a set of rules and $\succ$ is an ordering relation over X. A function $g : X \to \mathbb{R}$ is monotone (resp. anti–monotone) with respect to $\succ$ (monotone in $\succ$) if and only if $x \succ y$ implies that $g(x) \geq g(y)$ (resp. $g(x) \leq g(y)$) for any x, y.

Implication of a complete preorder by a partial preorder. A complete preorder $\preceq_v$ is implied by a partial preorder $\preceq_{qt}$ if and only if given a set of rules X and any two rules $r_1, r_2 \in X$, $r_1 \preceq_{qt} r_2$:

$$\begin{array}{l} r_1 \prec_{qt} r_2 \Rightarrow r_1 \prec_v r_2,\ \ \mathit{and} \\ r_1 \sim_{qt} r_2 \Rightarrow r_1 \sim_v r_2. \end{array} \tag{31}$$

Moreover, Bayardo and Agrawal have shown in [3] that the following conditions are sufficient for proving that a complete preorder $\preceq_v$ defined over a rule value function $g(r)$ is implied by a partial preorder $\preceq_{qt}$:

$$\begin{array}{l} g(r) \text{ is monotone in } q \text{ over rules with the same value of } t\text{, and} \\ g(r) \text{ is monotone in } t \text{ over rules with the same value of } q. \end{array} \tag{32}$$

4.2 Support–confidence evaluation space

Bayardo and Agrawal in [3] have investigated the concept of rule evaluation with respect to two popular attractiveness measures being rule support and confidence. They have considered rules with the same conclusion and evaluated them in such two dimensional space.

It has been proved in [3] that, for a set of rules with the same conclusion, if a complete preorder $\preceq_v$ is implied by a particular support–confidence partial preorder $\preceq_{sc}$, then rules optimal with respect to $\prec_v$ can be found in the set of non-dominated rules with respect to rule support and confidence.

Adjusting (32), the following conditions are sufficient for proving that a complete preorder $\preceq_v$ defined over a rule value function $g(r)$ is implied by the support–confidence partial preorder $\preceq_{sc}$:

– $g(r)$ is monotone in rule support over rules with the same value of confidence, and
– $g(r)$ is monotone in confidence over rules with the same value of rule support. (33)

Analysing the above conditions we only consider rules with the same conclusion. Moreover, we do not apply any changes to the data table $S = (U, A)$ and therefore $|U|$ and $sup_S(\psi)$ are constant. In this context, the monotonicity of $g(r)$ with respect to rule support over rules with the same value of confidence means that for any two rules r_1 and r_2, such that $conf_S(r_1) = conf_S(r_2)$, if $sup_S(r_1) \leq sup_S(r_2)$ then $g(r_1) \leq g(r_2)$. Analogously, the monotonicity of $g(r)$ with respect to confidence over rules with the same value of rule support means that for any two rules r_1 and r_2, such that $sup_S(r_1) = sup_S(r_2)$, if $conf_S(r_1) \leq conf_S(r_2)$ then $g(r_1) \leq g(r_2)$.

Bayardo and Agrawal have shown that the support–confidence Paretooptimal border (i.e. the set of non-dominated rules with respect to support and confidence) includes rules optimal according to several different attractiveness measures, such as gain, Laplace [12], lift [40], conviction [6], rule interest function, and others. This practically useful result allows to identify the most interesting rules according to those measures by solving an optimized rule mining problem with respect to rule support and confidence only.

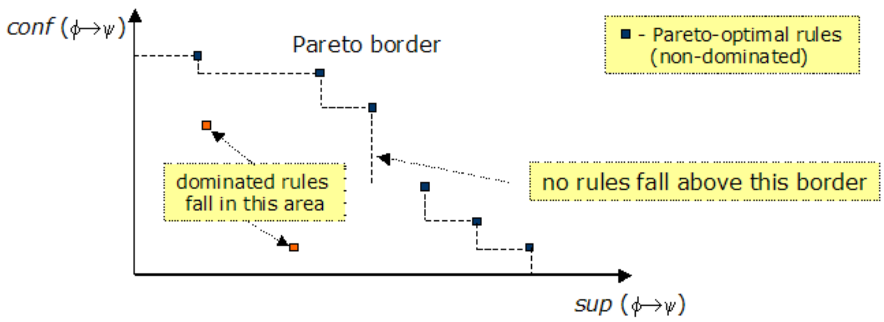


Fig. 1. Support–confidence Pareto-optimal border

Moreover, since the conditions (33) are general enough, the analysis of relationship of other attractiveness measures with support and confidence can be conducted. Due to the utility of the support–confidence Pareto-optimal border, the problem of proving which other complete preorders can be implied by $\preceq_{sc}$ remains an important issue both from theoretical and practical point of view.

Monotonic relationship of measure f with support and confidence. Due to valuable properties of measure f (property M, property of confirmation and hypothesis symmetry) our analysis aimed to verify whether (among rules with a fixed hypothesis) rules that are best according to measure f are included in the set of the non-dominated rules with respect to support and confidence. To fulfill the above objective, it has been checked whether conditions (33) hold when the confirmation measure f is the $g(r)$ rule value function.

Theorem 4.1. *Measure f is independent of rule support, and, therefore, monotone in rule support, when the value of confidence is held fixed [7].*

Proof. Let us consider measure f transformed such that, for given U and ψ, it only depends on confidence of rule $\phi \to \psi$ and support of ψ:

$$f_S(\phi \to \psi) = \frac{|U|\, conf_S(\phi \to \psi) - sup_S(\psi)}{(|U| - 2sup_S(\psi))conf_S(\phi \to \psi) + sup_S(\psi)}. \tag{34}$$

As we consider rules with a fixed conclusion ψ and we do not apply any changes to the data table S, the values of $|U|$ and $sup_S(\psi)$ are constant. Thus, for a fixed confidence, we have a constant value of measure f, no matter what the rule support is. Hence, confirmation measure f is monotone in rule support when the confidence is held constant. $\square$

Theorem 4.2. *Measure f is increasing in confidence, and, therefore, monotone with respect to confidence [7].*

Proof. Again, let us consider measure f given as in (34). For the clarity of presentation, let us express the above formula as a function of confidence, still regarding $|U|$ and $sup_S(\psi)$ as constant values greater than 0:

$$y = \frac{kx - m}{nx + m},$$

where $y = f_S(\phi \to \psi)$, $x = conf_S(\phi \to \psi)$, $k = |U|$, $m = sup_S(\psi)$, $n = |U| - 2sup_S(\psi)$.

It is easy to observe that $k = |U| > 0$, and $0 < m \leq |U|$. In order to verify the monotonicity of f in confidence, let us differentiate y with respect to x. We obtain:

$$\frac{\partial y}{\partial x} = \frac{m(k+n)}{(nx+m)^2}.$$

As $m > 0$, and $k + n = |U| + |U| - 2sup_S(\psi) = 2|U| - 2sup_S(\psi) > 0$ for $|U| \geq sup_S(\psi)$, the whole derivative is always not smaller than 0. Therefore, confirmation measure f is monotone in confidence. $\square$

Thus, both of Bayardo and Agrawal's sufficient conditions for proving that a total order $\preceq_v$ defined over a confirmation measure f is implied by partial order $\leq_{sc}$ are held. This means that, for a class of rules with a fixed conclusion, rules optimal according to measure f will be found in the set of rules that are best with respect to both rule support and confidence.

This result does not refer, however, to utility of scales in which $f_S(\phi \to \psi)$, having the property of confirmation, and $conf_S(\phi \to \psi)$, not having the property of confirmation, are expressed. While the confidence is the truth value of the knowledge pattern "*if* ϕ*, then* ψ", measure $f_S(\phi \to \psi)$ says to what extend ψ is satisfied more frequently when ϕ is satisfied rather than when ϕ is not satisfied. In other words, f says what is the "value of information" that ϕ adds to the credibility of ψ. For further discussion about weakness of confidence scale see [8], [72].

The difference of semantics and utility of $conf_S(\phi \to \psi)$ on one hand, and $f_S(\phi \to \psi)$ or $s_S(\phi \to \psi)$ as representatives of measures with the confirmation property on the other hand, can be shown on the following example. Consider the possible result of rolling a die: 1, 2, 3, 4, 5, 6, and let the conclusion ψ = "the result is divisible by 2". Given two different premises: ϕ_1 = "the result is a number from a set $\{1, 2, 3\}$", ϕ_2 = "the result is a number from a set $\{2, 3, 4\}$", we get, respectively: $conf_S(\phi_1 \to \psi) = 1/3$, $f_S(\phi_1 \to \psi) = -1/3$, $s_S(\phi_1 \to \psi) = -1/3$, $conf_S(\phi_2 \to \psi) = 2/3$, $f_S(\phi_2 \to \psi) = 1/3$, $s_S(\phi_2 \to \psi) = 1/3$. This example, of course, acknowledges the monotone link between confirmation measure f and confidence. However, it also clearly shows that the values of confirmation measures have a more useful interpretation than confidence. In particular, in the case of rule $\phi_1 \to \psi$, the premise actually disconfirms the conclusion as it reduces the probability of conclusion ψ from $1/2 = sup_S(\psi)$ to $1/3 = conf_S(\phi_1 \to \psi)$. This fact is expressed by a negative value of confirmation measure f and s , but cannot be concluded by observing only the value of confidence.

Finally, as semantics of $f_S(\phi \to \psi)$ is more useful than that of $conf_S(\phi \to \psi)$, and as both of these measures are monotonically linked, it is reasonable to propose a new rule evaluation space in which the search for the most interesting rules is carried out taking into account confirmation measure $f_S(\phi \to \psi)$ and rule support [84].

4.3 Support–*f* evaluation space

Combining of rule support and measure f in one rule evaluation space is valuable as f is independent of rule support in the sense presented in Theorem 4.1, and rules that have high values of measure f are often characterized by small values of the rule support.

Proposition of a new evaluation space, naturally, brings a question of comparison with the support–confidence evaluation space. To fulfil that objective a thorough analysis of monotonicity of confidence in support and measure f has been carried out (for details see [8]). It has been shown that rules optimal in confidence lie on the Pareto-optimal border with respect to support and measure f.

Moreover, it has been proved that all attractiveness measures for which the optimal rules are found on the support–confidence Pareto-optimal border also preserve that relationship with respect to support–f Pareto-optimal border. Thus, it is legitimate to conclude that the set of rules forming the support–confidence Pareto-optimal border is exactly the same as the set of rules constituting the support–f Pareto-optimal border [8], [9]. An illustration of this result on a real life dataset *census*, for conclusion workclass='Private', is presented on Figure 2 (for more details on the dataset refer to Section 6.2).

Those two Pareto sets can, in fact, be regarded as monotone transformations of each other. Hence, substitution of confidence by f does not diminish the set of interestingness measures for which optimal rules reside on the Pareto-optimal border. However, as semantics of measure f is more useful than of confidence, we are strongly in favor of mining the Pareto-optimal border with respect to rule support and confirmation f and not rule support and confidence as it was proposed in [3].

The advantage of the support–f evaluation space comes from the fact that, contrary to confidence, measure f has the property of confirmation and thus has the means to filter out the disconfirming rules (marked on Figure 2 by red circles; the blue triangles represent rules with positive confirmation value). Let us stress that even the non-dominated rules, which are objectively the best rules, might be characterized by a negative value of confirmation measure f, and therefore need to be discarded. Confidence measure cannot distinguish such useless rules.

The number of rules which are characterized by a negative value of any attractiveness measure with the property of confirmation (measure f is just a representative of the group of confirmation measures) depends on the dataset, but can potentially be quite large. Therefore, the reduction of the number of rules to be analyzed is another argument for the support–f evaluation space [86].

Table 2. Information about the percentage of rules with non-positive confirmation in the set of all generated rules for different conclusions, for minimal support=0.15 (*census* dataset)

Considered conclusion	No. of all rules	No. of rules with non-positive confirmation	Reduction percentage
workclass=Private	84	42	50%
sex=Male	84	23	27%
income<=50kUSD	85	43	51%
race=White	105	27	26%
native_country=USA	111	30	27%

Table 2 contains information about the number and percentage of rules with non-positive confirmation for few sets of rules with different conclusions from the

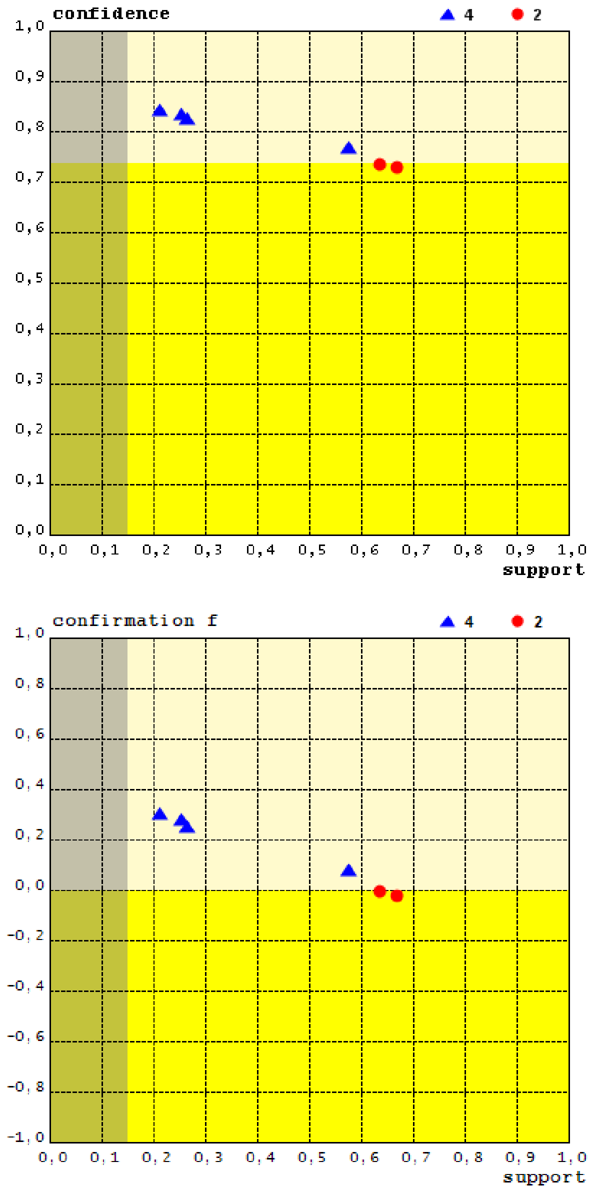


Fig. 2. Support–confidence Pareto-optimal border contains the same rules as support–f Pareto-optimal border (conclusion: workclass='Private', *census* dataset)

census dataset. For the conclusion being workclass='Private', 42 out of 84 rules had to be discarded for disconfirming the conclusion. Thus, the set of potentially interesting and valuable rules was reduced by 50%.

Table 3 shows how many rules with non–positive confirmation were situated on the support–f Pareto–optimal border (or support–confidence Pareto-optimal border) for different considered conclusions. Even the Pareto–optimal borders, i.e. the sets of objectively the best rules, contain rules that are misleading. In some cases, the Pareto–optimal border could be reduced by even 33%, like for the conclusion workclass='Private'. These reduction percentages also give weight to the need of taking into consideration the information brought by the confirmation property.

Table 3. Information about the percentage of rules with non-positive confirmation located on the support–f Pareto–optimal border for different conclusions, for minimal support=0.15 (*census* dataset)

Considered conclusion	**No. of all rules on Pareto border**	**No. of rules with non-positive confirmation**	**Reduction percentage**
workclass=Private	6	2	33%
sex=Male	6	1	17%
income<=50 kUSD	4	1	25%
race=White	10	0	0%
native_country=USA	10	0	0%

Confirmation perspective on support–confidence evaluation space. Inspired by the strength of the semantics of the family of confirmation measures, we find it valuable to try to impose a confirmation perspective on the support–confidence evaluation space and limit the set of rules by eliminating those that are characterized by non–positive or small values of a confirmation measure. Let us recall that non-positive values of a confirmation measure reflect that the premise of a rule disconfirms its conclusion, and small positive values of confirmation measures express that occurrence of the premise only slightly increases the probability of observing the conclusion.

Let us consider a confirmation measure $c_S(\phi \to \psi)$ with the property M (e.g. measure f) and let us observe that according to condition (24).1:

$$c_S(\phi \to \psi) > 0 \Leftrightarrow conf_S(\phi \to \psi) > \frac{sup_S(\psi)}{|U|}. \tag{35}$$

Moreover, it has been analytically proved in [9] (see Theorem 4.9) that for a fixed value of rule support, any measure $c_S(\phi \to \psi)$ having the property of confirmation and the property M is monotone with respect to confidence. Let us also stress that all confirmation measures (no matter whether having the property M or not) change their signs in the same situations. Thus, the possession or not of property M will not influence our further discussion.

Since, we limit our considerations to rules with the same conclusion and do not apply any changes to the data table S, $|U|$ and $sup_S(\psi)$ should be regarded as constant values. Thus, due to the monotonic link between $c_S(\phi \to \psi)$ and confidence, (35) shows that rules laying under a constant $sup_S(\phi \to \psi)/|U|$, expressing what percentage of the whole dataset is taken by the considered class ψ, are characterized by negative values of any measure with the property of confirmation. For those rules ψ is satisfied less frequently when ϕ is satisfied rather than generically. Figure 3 illustrates this analytical result. Of course, the more objects there are in the analyzed class with a particular conclusion, the more demanding is the position of the constant line separating rules with non-positive confirmation and the less rules are expected to lie above it.

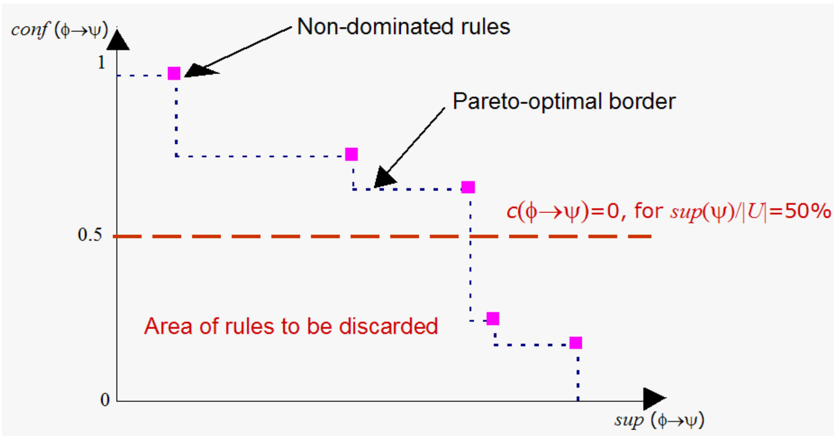


Fig. 3. An example of a constant line representing any confirmation measure $c_S(\phi \to \psi) = 0$ in a support–confidence space; rules laying under this constant line should be discarded from further analysis

It is also interesting to investigate a more general condition $c_S(\phi \to \psi) \geq k$, $k \geq 0$, for some specific measures with the property of confirmation. In the following, we consider confirmation measure $f_S(\phi \to \psi)$.

Theorem 4.3. *[86]*

$$f_S\left(\phi \to \psi\right) \geq k \Leftrightarrow conf_S\left(\phi \to \psi\right) \geq \frac{sup_S\left(\psi\right)\left(k+1\right)}{|U| - k\left(|U| - 2sup_S\left(\psi\right)\right)}. \tag{36}$$

Proof. The analysis concerns only a set of rules with the same conclusion, and since we do not apply any changes to the data table S, the values of $|U|$ and

$sup_S(\psi)$ are constant. For given U and ψ, let us consider confirmation measure $f_S(\phi \to \psi)$ written in terms of confidence and support of rule $\phi \to \psi$ (effectively in terms of confidence only) as in (34):

$$f_S(\phi \to \psi) = \frac{|U|\, conf_S(\phi \to \psi) - sup_S(\psi)}{(|U| - 2sup_S(\psi))conf_S(\phi \to \psi) + sup_S(\psi)}. \tag{37}$$

Transforming the above definition of f to outline how confidence depends on f we obtain:

$$conf_S(\phi \to \psi) = \frac{f_S(\phi \to \psi)sup_S(\psi) + sup_S(\psi)}{|U| - f_S(\phi \to \psi)(|U| - 2sup_S(\psi))}. \tag{38}$$

Considering inequality $f_S(\phi \to \psi) \geq k$ for (38) we obtain the thesis of the theorem. $\square$

Figure 4 in an exemplary application of the theoretical results on the *census* dataset (for more examples see also [85]). Rules with conclusion workclass='Private' are evaluated in support–confidence and support–f space. On the diagrams a constant line separates the rules with positive confirmation (blue circles situated above the line) from those with non–positive confirmation (red circles situated below the line). In the support–confidence evaluation space the position of the line had to be calculated according to result (35), whereas the same information is given straightforward in the support–f evaluation space as only the sign of f needs to be observed. This example points out the advantage of support–f space over support–confidence space, however, it also shows that result (35) provides means by which the support–confidence space can actually be made meaningful.

4.4 Support–s evaluation space

Having proved a monotonic relationship between confidence and measure f our analysis proceeded to verify the existence of such link between confidence and another attractiveness measure with the property M and the property of confirmation, hoping that these results would finally allow to generalize the result for the whole class of attractiveness measures possessing the property M. Below, we considered measure s, having the property M, property of confirmation and hypothesis symmetry, and verified whether (among rules with a fixed hypothesis) rules that are best according to measure s are included in the set of non-dominated rules with respect to support and confidence. To fulfil this objective, the monotonicity of s with respect to rule support over rules with the same value of confidence and the monotonicity of s with respect to confidence over rules with the same value of rule support was analyzed. Thus, in the first step it was verified whether for any two rules r_1 and r_2, such that $conf_S(r_1) = conf_S(r_2)$, if $sup_S(r_1) \leq sup_S(r_2)$ then $s_S(r_1) \leq s_S(r_2)$. Next, it was analyzed whether for any two rules r_1 and r_2, such that $sup_S(r_1) = sup_S(r_2)$, if $conf_S(r_1) \leq conf_S(r_2)$ then $s_S(r_1) \leq s_S(r_2)$.

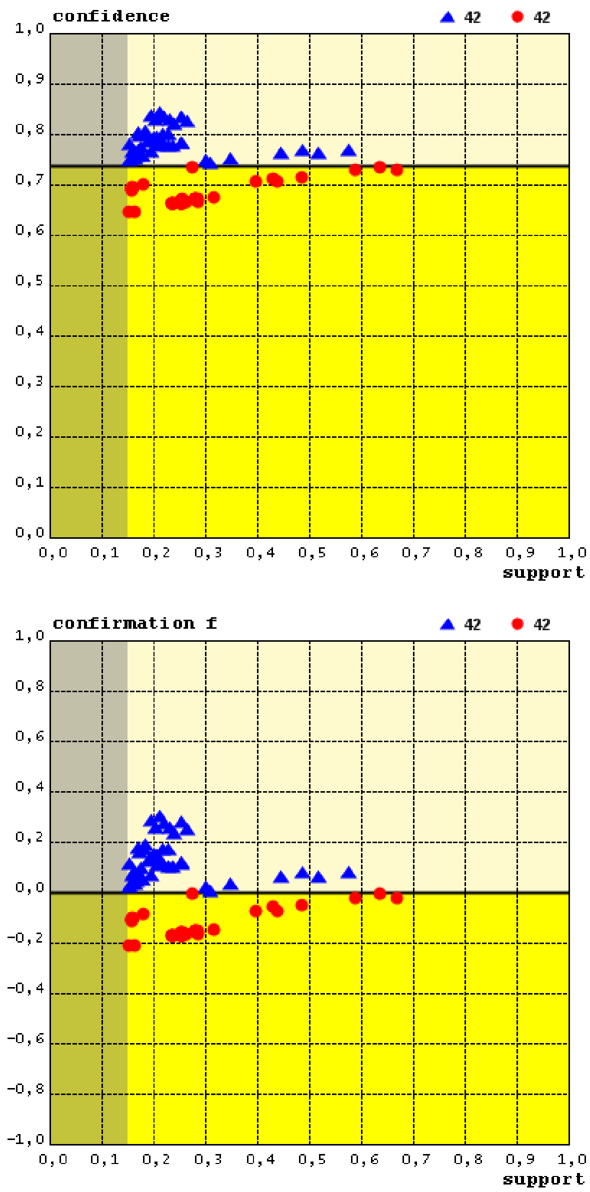


Fig. 4. Rules with positive (blue circles) and non–positive confirmation measure value (red circles) in a support–confidence and support–f space; for minimal support=0.15 (conclusion: workclass='Private', *census* dataset)

Theorem 4.4. *[8] When the confidence value is held fixed, then:*

1. *measure $s_S(\phi \rightarrow \psi)$ is increasing in rule support (i.e. strictly monotone) iff $s_S(\phi \rightarrow \psi) > 0$,*
2. *measure $s_S(\phi \rightarrow \psi)$ is constant in rule support (i.e. monotone) iff $s_S(\phi \rightarrow \psi) = 0$,*
3. *measure $s_S(\phi \rightarrow \psi)$ is decreasing in rule support (i.e. strictly anti-monotone) iff $s_S(\phi \rightarrow \psi) < 0$.*

Proof. Let us consider measure s expressed in the notation (16):

$$s_S(\phi \rightarrow \psi) = \frac{a}{a+c} - \frac{b}{b+d}. \tag{39}$$

Only the proof of part 1 shall be presented, as the other points are analogous.

Let us consider two rules: r_1: $\phi \rightarrow \psi$ and r_2: $\phi' \rightarrow \psi$ such that they have the same value of confidence and $sup_S(\phi \rightarrow \psi) < sup_S(\phi' \rightarrow \psi)$. For the first rule $sup_S(\phi \rightarrow \psi) = a$. Let us express the support of r_2 in the form of $sup_S(\phi' \rightarrow \psi) = a' = a + \Delta$, where $\Delta > 0$. Since confidence is to be constant, thus c should change into $c' = c + \epsilon$ in such a way that:

$$conf_S(\phi \rightarrow \psi) = \frac{a}{a+c} = conf_S(\phi' \rightarrow \psi) = \frac{a'}{a'+c'} = \frac{a+\Delta}{a+\Delta+c+\varepsilon}.$$

Simple algebraic transformation lead to the conclusion that:

$$\frac{a}{a+c} = \frac{a+\Delta}{a+\Delta+c+\varepsilon} \Leftrightarrow \frac{\Delta}{\Delta+\varepsilon} = \frac{a}{a+c}. \tag{40}$$

Let us observe that (40) implies that if $c = 0$ then $\epsilon = 0$ and moreover if $c > 0$ then $\epsilon > 0$. Since $|U|$ and $sup_S(\psi)$ must be kept constant, b and d need to decrease in such a way that $b' = b - \Delta$ and $d' = d - \epsilon$. In this situation, the confirmation measure s for r_2 will be:

$$s_S(\phi' \rightarrow \psi) = \frac{a'}{a'+c'} - \frac{b'}{b'+d'} = \frac{a+\Delta}{a+\Delta+c+\varepsilon} - \frac{b-\Delta}{b-\Delta+d-\varepsilon}.$$

Remembering that $conf_S(\phi \rightarrow \psi) = conf_S(\phi' \rightarrow \psi)$, let us observe that:

$$\begin{aligned} &s_S(\phi' \rightarrow \psi) > s_S(\phi \rightarrow \psi) \Leftrightarrow \tfrac{b}{b+d} > \tfrac{b-\Delta}{b-\Delta+d-\varepsilon} \Leftrightarrow \\ &d\Delta > b\varepsilon \Leftrightarrow d\Delta + b\Delta > b\varepsilon + b\Delta \Leftrightarrow \tfrac{\Delta}{\Delta+\varepsilon} > \tfrac{b}{b+d}. \end{aligned} \tag{41}$$

Considering (40) and (41) it can be concluded that:

$$s_S(\phi' \rightarrow \psi) > s_S(\phi \rightarrow \psi) \Leftrightarrow \frac{a}{a+c} > \frac{b}{b+d} \Leftrightarrow s_S(\phi \rightarrow \psi) > 0.$$

This proves that, for a fixed value of confidence, measure s is increasing with respect to rule support if and only if $s_S(\phi \rightarrow \psi) > 0$ and therefore in its positive range measure s is strictly monotone in rule support. □

Theorem 4.5. *When the rule support value is held fixed, measure s is increasing with respect to confidence (i.e. measure s is monotone in confidence) [8].*

Proof. Again, let us consider two rules $r_1 : (\phi \to \psi)$ and $r_2 : (\phi' \to \psi)$, and measure s given as in (39). For the hypothesis, the rule support is supposed to be constant, i.e. $sup_S(\phi \to \psi) = a = sup_S(\phi' \to \psi)$. Therefore, it is clear that $conf_S(\phi \to \psi) = \frac{a}{a+c}$ will be smaller than $conf_S(\phi' \to \psi) = \frac{a}{a+c'}$ only if we consider $c' = c - \Delta$, where $\Delta > 0$. Now, operating on c' the only way to guarantee that $|U|$ and $sup_S(\psi)$ remain constant (as we do not apply any changes to the data table S) is to increase d such that $d' = d + \Delta$. The values of a and b cannot change: $a' = a$ and $b' = b$. Now, the value of measure s for r_2 takes the following form:

$$s_S(\phi' \to \psi) = \frac{a'}{a' + c'} - \frac{b'}{b' + d'} = \frac{a}{a + c - \Delta} - \frac{b}{b + d + \Delta}.$$

Since $\Delta > 0$, it is clear that $s_S(\phi' \to \psi) > s_S(\phi \to \psi)$. This means that for a fixed value of rule support, increasing confidence results in an increase of the value of measure s and therefore measure s is monotone with respect to confidence. $\square$

As rules with negative values of measure s should always be discarded from consideration, the result from Theorem 4.4 states the monotone relationship just in the interesting subset of rules. It implies that rules for which $s_S(\phi \to \psi) \geq 0$ and which are optimal with respect to measure s will reside on the support–confidence Pareto-optimal border. They will also be found on the support–f Pareto-optimal border since those Pareto sets have the same contents.

Since confirmation measure s has the property of monotonicity M, we propose to generate interesting rules by searching for rules maximizing confirmation measure s and support, i.e. substituting the confidence in the support–confidence Pareto-optimal border with measure s and obtaining in this way a support–s Pareto-optimal border. This approach differs from the idea of finding the Pareto-optimal border according to rule support and confirmation measure f, because support–f Pareto-optimal border contains the same rules as the support–confidence Pareto-optimal border, while in general support–s Pareto-optimal border contains a subset of the support–confidence Pareto-optimal border as stated in the following theorem.

Theorem 4.6. *If a rule resides on the support–s Pareto-optimal border (in case of positive value of confirmation measure s), then it also resides on the support–confidence Pareto-optimal border, while one can have rules being on the support–confidence Pareto-optimal border which are not on the support–s Pareto-optimal border [9].*

Proof. Let us consider a rule $r : \phi \to \psi$ residing on the support–s Pareto-optimal border and let us suppose that measure s has a positive value. This means that for any other rule $r' : \phi' \to \psi$ we have that:

$$sup_S(\phi' \to \psi) > sup_S(\phi \to \psi) \Rightarrow s_S(\phi' \to \psi) < s_S(\phi \to \psi). \qquad (42)$$

On the basis of monotonicity of measure s with respect to support and confidence in case of positive value of s, we have that $sup_S(\phi' \to \psi) > sup_S(\phi \to \psi)$ and $s_S(\phi' \to \psi) < s_S(\phi \to \psi)$ implies that $conf_S(\phi' \to \psi) < conf_S(\phi \to \psi)$. This means that (42) implies that for any other rule r', $sup_S(\phi' \to \psi) > sup_S(\phi \to \psi) \Rightarrow conf_S(\phi' \to \psi) < conf_S(\phi \to \psi)$.

This means that rule r residing on the support–s Pareto-optimal border is also on the support–confidence Pareto-optimal border because one cannot have any other rule r' such that $sup_S(\phi' \to \psi) > sup_S(\phi \to \psi)$ and $conf_S(\phi' \to \psi) \geq conf_S(\phi \to \psi)$.

Now, we prove by a counter-example that there can be rules being on the support–confidence Pareto-optimal border which are not on the support–s Pareto-optimal border. Let us consider rules r and r' residing on the support–confidence Pareto-optimal border such that for rule r we have $sup_S(\phi \to \psi) = 200$ and $conf_S(\phi \to \psi) = 0.667$, while for rule r' we have $sup_S(\phi' \to \psi) = 150$ and $conf_S(\phi' \to \psi) = 0.68$. We have that $s_S(\phi \to \psi) = 0.167$ which is greater than $s_S(\phi' \to \psi) = 0.142$. Thus, rule *r'* is not on the support–s Pareto-optimal border because it is dominated with respect to support–s by rule r having a greater support and a greater value of measure s. $\square$

Theorem 4.6 states that some rules from the support–confidence Pareto-optimal border may be not present on the support–s Pareto-optimal border. Figure 5 is an exemplary illustration of this result on the *census* dataset for the conclusion being: income<=50kUSD. On the diagram, the support–s Pareto-optimal border contains three points, each representing one rule, whereas the set of non-dominated rules according to support and confidence has one more rule.

The result from Theorem 4.6 can be easily generalized by substituting measure s for any attractiveness measure monotone with respect to support and confidence. The following theorem states formally this point.

Theorem 4.7. *Given an attractiveness measure i, which is monotone with respect to support and confidence, if a rule resides on the support–i Pareto-optimal border, then it also resides on the support–confidence Pareto-optimal border, while the opposite assertion is not necessarily true. [9]*

Proof. Analogous to proof of Theorem 4.6.

4.5 Support and confirmation measures with the property M evaluation space

The investigation of a monotone link with confidence and rule support has also been extended to a general class of all measures that have the property M. Such measures monotonically linked with confidence and support are good candidates for substituting confidence in the support–confidence evaluation space if they also belong to the group of Bayesian confirmation measures.

For a set of rules with a fixed conclusion a general analysis has been conducted verifying when a measure possessing the property M holds conditions (33), i.e.:

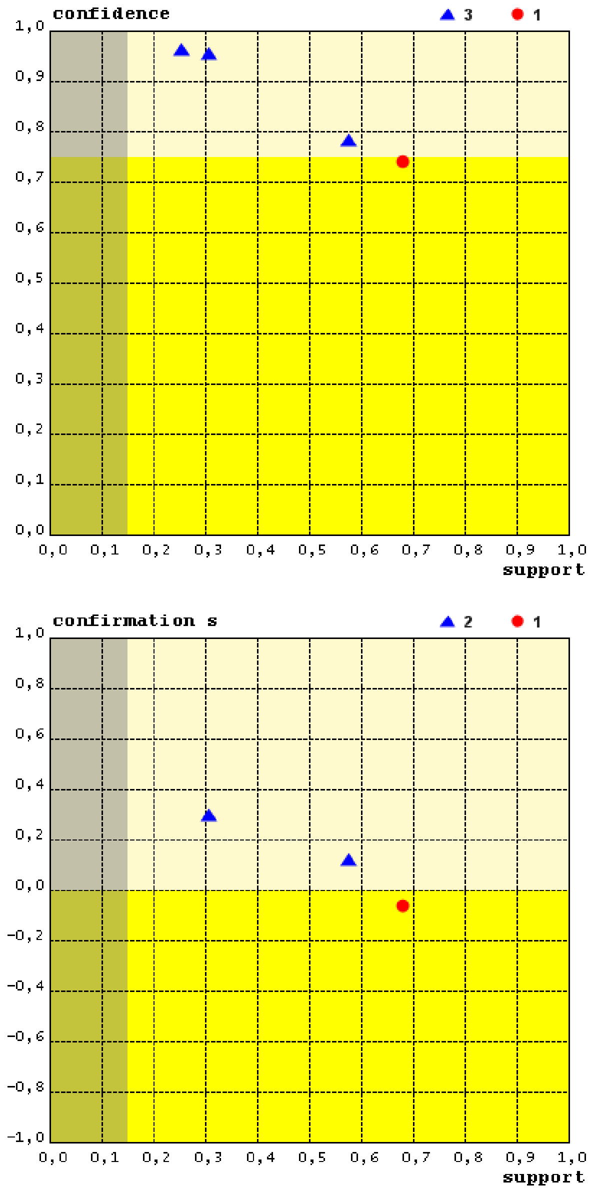


Fig. 5. Support–confidence Pareto-optimal border is the upper-set of support–s Pareto-optimal border (conclusion: income<=50kUSD, *census* dataset)

– is monotone in rule support when the value of confidence is held fixed,
– is monotone in confidence when the value of rule support is kept unchanged.

Let us consider an attractiveness measure $F(a, b, c, d)$ having the property M. The analysis concerns only a set of rules with the same conclusion, and as we do not apply any changes to the data table S, the values of $|U| = a + b + c + d$ and $sup_S(\psi) = a + b$ are constant.

One can observe that a, b, c, and d can be transformed in the following way:

$$\begin{aligned}
&a = sup_S(\phi \to \psi),\\
&b = sup_S(\psi) - sup_S(\phi \to \psi),\\
&c = \tfrac{1}{conf_S(\phi \to \psi)} sup_S(\phi \to \psi) - sup_S(\phi \to \psi),\\
&d = |U| - sup_S(\psi) - \tfrac{1}{conf_S(\phi \to \psi)} sup_S(\phi \to \psi) + sup_S(\phi \to \psi).
\end{aligned}$$

Then, measure F can be expressed as:

$$\begin{aligned}
&F(a, b, c, d) = F(sup_S(\phi \to \psi),\ \ sup_S(\psi) - sup_S(\phi \to \psi),\\
&\tfrac{1}{conf_S(\phi \to \psi)} sup_S(\phi \to \psi) - sup_S(\phi \to \psi),\\
&|U| - sup_S(\psi) - \tfrac{1}{conf_S(\phi \to \psi)} sup_S(\phi \to \psi) + sup_S(\phi \to \psi)).
\end{aligned} \tag{43}$$

Let us remark that we can say nothing in general about monotonicity with respect to support of $F(a, b, c, d)$ satisfying the property M. In fact $F(a, b, c, d)$ is clearly non-decreasing with respect to the value of support in variable a and b, however it is non-increasing with respect to the value of support in variable c and d. The latter point merits some explanations. We have that

$$c = \frac{1}{conf_S(\phi \to \psi)} sup_S(\phi \to \psi) - sup_S(\phi \to \psi) =$$
$$= sup_S(\phi \to \psi)\left(\frac{1}{conf_S(\phi \to \psi)} - 1\right)$$

and $\left(\frac{1}{conf_S(\phi \to \psi)} - 1\right)$ is non-negative. Since for the property M, F is non-increasing with respect to variable c, we get that F is non-increasing with respect to the value of support in variable c.

Analogously, we have that

$$d = |U| - sup_S(\psi) - \frac{1}{conf_S(\phi \to \psi)} sup_S(\phi \to \psi) + sup_S(\phi \to \psi) =$$
$$= |U| - sup_S(\psi) - sup_S(\phi \to \psi)\left(\frac{1}{conf_S(\phi \to \psi)} - 1\right)$$

and $\left(\frac{1}{conf_S(\phi \to \psi)} - 1\right)$ is non-negative. Since for the property of monotonicity M, F is non-decreasing with respect to variable d, we get that F is non-increasing with respect to the value of support in variable d.

Theorem 4.8. *[9] When the value of confidence is held fixed, then the measure $F(a,b,c,d)$ with property M, admitting derivative with respect to all its variables a,b,c and d, is monotone in rule support if:*

$$\frac{\partial F}{\partial c} = \frac{\partial F}{\partial d} = 0 \ \ or \ \ \frac{\frac{\partial F}{\partial a} - \frac{\partial F}{\partial b}}{\frac{\partial F}{\partial d} - \frac{\partial F}{\partial c}} \geq \frac{1}{conf_S(\phi \to \psi)} - 1. \tag{44}$$

Proof. Let us consider $F(a,b,c,d)$ expressed as in (43). Let us assume that $conf_S(\phi \to \psi)$ is constant. Let us differentiate $F(a,b,c,d)$ with respect to $sup_S(\phi \to \psi)$. Then, we obtain:

$$\frac{\partial F}{\partial sup_S(\phi \to \psi)} = \frac{\partial F}{\partial a} - \frac{\partial F}{\partial b} + \left(\frac{\partial F}{\partial c} - \frac{\partial F}{\partial d}\right)\ \left(\frac{1}{conf_S(\phi \to \psi)} - 1\right). \tag{45}$$

Since F is supposed to satisfy the property M, it must be non-increasing with respect to b, c and non-decreasing with respect to a, d, such that

$$\frac{\partial F}{\partial b} \leq 0, \frac{\partial F}{\partial c} \leq 0 \text{ and } \frac{\partial F}{\partial a} \geq 0, \frac{\partial F}{\partial d} \geq 0.$$

Hence, if

$$\frac{\partial F}{\partial c} = \frac{\partial F}{\partial d}, \text{ then } \frac{\partial F}{\partial sup_S(\phi \to \psi)} \geq 0.$$

It is clear, that due to the property M of F,

$$\frac{\partial F}{\partial c} = \frac{\partial F}{\partial d} \text{ if and only if } \frac{\partial F}{\partial c} = \frac{\partial F}{\partial d} = 0.$$

Let us observe, moreover, that if

$$\frac{\partial F}{\partial c} \neq \frac{\partial F}{\partial d}, \text{ then } \frac{\partial F}{\partial sup_S(\phi \to \psi)} \geq 0 \Leftrightarrow \frac{\frac{\partial F}{\partial a} - \frac{\partial F}{\partial b}}{\frac{\partial F}{\partial d} - \frac{\partial F}{\partial c}} \geq \frac{1}{conf_S(\phi \to \psi)} - 1.$$

□

Theorem 4.9. *When the value of rule support is held fixed, then measure $F(a,b,c,d)$ with property M is monotone in confidence. [9].*

Proof. Let us consider $F(a,b,c,d)$ expressed as in (43). Confidence determines the value of measure $F(a,b,c,d)$ through variables c and d. We have that variable c is non-increasing in confidence. In fact,

$$c = \frac{1}{conf_S(\phi \to \psi)} sup_S(\phi \to \psi) - sup_S(\phi \to \psi)$$

and $sup_S(\phi \to \psi)$ is non-negative.

Since for the property M, F is non-increasing with respect to variable c, we get that F is non-decreasing with respect to the value of confidence in variable c.

We also have that variable d is non-decreasing in confidence. In fact,

$$d = |U| - sup_S(\psi) - \frac{1}{conf_S(\phi \to \psi)} sup_S(\phi \to \psi) + sup_S(\phi \to \psi)$$

and $sup_S(\phi \to \psi)$ is non-negative.

Since for the property M, F is non-decreasing with respect to variable d, we get that F is non-decreasing with respect to the value of confidence in variable d. □

Theorem 4.9 states that for a set of rules with the same conclusion, any measure satisfying the property M is always non-decreasing with respect to confidence when the value of rule support is kept fixed. Moreover, due to Theorem 4.8, all those confirmation measures that are independent of $sup_S(\phi \to \neg\psi)$ and $sup_S(\neg\phi \to \neg\psi)$ are always found monotone in rule support when the value of confidence remains unchanged. However, for a constant value of confidence, measures which do depend on the value of $sup_S(\phi \to \neg\psi)$ and $sup_S(\neg\phi \to \neg\psi)$ are also non-decreasing with respect to rule support if and only if they satisfy the following condition:

$$\frac{\frac{\partial F}{\partial a} - \frac{\partial F}{\partial b}}{\frac{\partial F}{\partial d} - \frac{\partial F}{\partial c}} \geq \frac{1}{conf_S(\phi \to \psi)} - 1.$$

The general analysis in Theorem 4.8 and Theorem 4.9 outlines a method of verification whether there exists a monotone link between any measure with the property M, and rule support and confidence, respectively. Measures that positively undergo such ascertainment and, moreover, have the property of confirmation are, in our opinion, good candidates for substituting the confidence dimension in the Pareto-optimal border with respect to rule support and confidence proposed by Bayardo and Agrawal [3]. Theorem 4.8 and Theorem 4.9 consider general class of measures with property M, however taking into account the semantics of attractiveness measures, this result is especially interesting with respect to measures of confirmation.

5 Support–anti-support evaluation space

Let us observe that Theorem 4.8 can be regarded as a critical remark towards support–confidence Pareto optimal border as it says that a rule maximizing an attractiveness measure satisfying the property M is on the support–confidence Pareto-optimal border only if a specific condition is satisfied. Thus, in general, not all rules maximizing such a measure satisfying the property M are on the support–confidence Pareto–optimal border. However, due to valuable semantics of the property M, mining all rules that maximize any measure with M, is an interesting problem.

Let us consider an attractiveness measure $F(a,b,c,d)$ having the property M and induction of rules with a given conclusion ψ from a universe U such that $sup_S(\psi)$ and $|U|$ can be considered fixed. Again, for the simplicity of presentation, let us use the notation (16). One can observe that a, b, c, and d can be transformed in the following way:

$$\begin{aligned}
&a = sup_S(\phi \rightarrow \psi),\\
&b = sup_S(\psi) - sup_S(\phi \rightarrow \psi),\\
&c = sup_S(\phi \rightarrow \neg\psi),\\
&d = |U| - sup_S(\psi) - sup_S(\phi \rightarrow \neg\psi)).
\end{aligned}$$

Then, measure F can be expressed as:

$$\begin{aligned}
&F(a,b,c,d) = F(sup_S(\phi \rightarrow \psi),\ \ sup_S(\psi) - sup_S(\phi \rightarrow \psi),\\
&sup_S(\phi \rightarrow \neg\psi),\ \ |U| - sup_S(\psi) - sup_S(\phi \rightarrow \neg\psi)).
\end{aligned} \tag{46}$$

Let us stress that $sup_S(\phi \rightarrow \neg\psi)$ is the anti-support of rule $\phi \rightarrow \psi$. It represents the number of counter-examples to the rule $\phi \rightarrow \psi$. For example, if ϕ="x is a raven" and ψ="x is black", then $\phi \rightarrow \psi$ is the rule "if x is a raven, then x is black" and the anti-support=$sup_S(\phi \rightarrow \neg\psi)$ is the number of non-black ravens.

Theorem 5.1. *When the value of rule anti-support is held fixed, then measure $F(a,b,c,d)$ with the property M is monotone (non-decreasing) in rule support [9].*

Proof. Function $F(a,b,c,d)$ (46) depends on the rule support $sup_S(\phi \rightarrow \psi)$ through variables a and b. Observe that a is increasing with respect to $sup_S(\phi \rightarrow \psi)$ while b is decreasing. Remembering that for property M F is non-decreasing in a and non-increasing in b, we get the thesis. □

Theorem 5.2. *When the value of rule support is held fixed, then the measure $F(a,b,c,d)$ with property M is anti-monotone (non-increasing) in rule anti-support [9].*

Proof. Function $F(a,b,c,d)$ (46) depends on the anti-support $sup_S(\phi \rightarrow \neg\psi)$ through variables c and d. Observe that c is increasing while d is decreasing with respect to $sup_S(\phi \rightarrow \neg\psi)$. Remembering that for the property M $F(a,b,c,d)$ is non-increasing in c and non-decreasing in d, we get the thesis. □

Theorem 5.1 and Theorem 5.2 say that measure $F(a,b,c,d)$ with the property M is monotone (non-decreasing) with respect to rule support $sup_S(\phi \rightarrow \psi)$ and anti-monotone (non-increasing) with respect to rule anti-support $sup_S(\phi \rightarrow \neg\psi)$. Therefore, the best rules according to any measure having the property M must reside on the support–anti-support Pareto-optimal border being the set of rules such that there is no other rule having greater support and smaller anti-support.

Figure 6 presents the support–anti-support evaluation space. Since anti-support is a cost-type criterion (the smaller its value the better), the shape of the support–anti-support Pareto-optimal border is different than in the previously considered evaluation spaces. The set of non-dominated rules with respect to

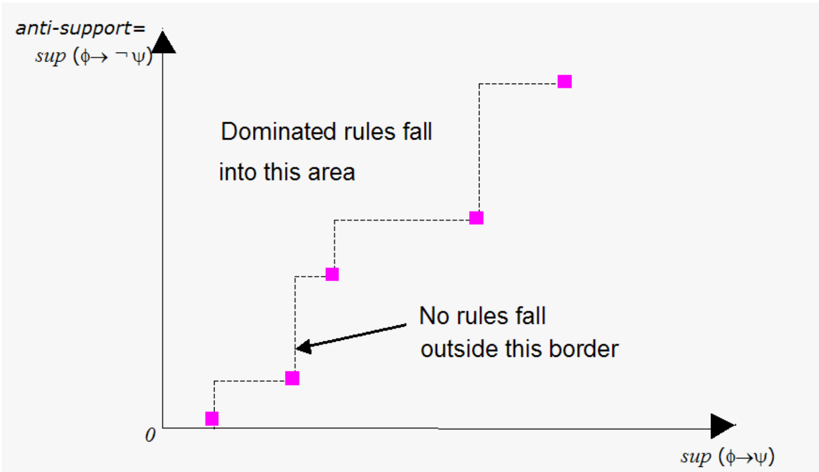


Fig. 6. Support–anti-support Pareto-optimal border

support and anti-support contains rules optimal with respect to any attractiveness measure with the valuable property M. Above the Pareto-optimal border there reside the dominated rules.

The following theorem formally describes the relationship between the Pareto optimal borders with respect to support and confidence on one hand, and support and anti-support on the other hand.

Theorem 5.3. *If a rule resides on the support–confidence Pareto-optimal border, then it also resides on the support–anti-support Pareto-optimal border, while one can have rules being on the support–anti-support Pareto-optimal border which are not on the support–confidence Pareto-optimal border [9].*

Proof. Let us consider a rule $r : \phi \rightarrow \psi$ residing on the support–confidence Pareto-optimal border. This means that for any other rule $r' : \phi' \rightarrow \psi$ we have that:

$$sup_S(\phi' \rightarrow \psi) > sup_S(\phi \rightarrow \psi) \Rightarrow conf_S(\phi' \rightarrow \psi) < conf_S(\phi \rightarrow \psi). \tag{47}$$

Observe that

$$conf_S(\phi' \rightarrow \psi) < conf_S(\phi \rightarrow \psi) \Leftrightarrow$$

$$\frac{sup_S\,(\phi' \rightarrow \psi)}{sup_S\,(\phi' \rightarrow \psi) + sup_S\,(\phi' \rightarrow \neg\psi)} < \frac{sup_S\,(\phi \rightarrow \psi)}{sup_S\,(\phi \rightarrow \psi) + sup_S\,(\phi \rightarrow \neg\psi)}$$

and since we are supposing $sup_S(\phi' \rightarrow \psi) > sup_S(\phi \rightarrow \psi)$, we get that $sup_S(\phi' \rightarrow \neg\psi) > sup_S(\phi \rightarrow \neg\psi)$.

This means that (47) implies that for any other rule r'

$$sup_S(\phi' \rightarrow \psi) > sup_S(\phi \rightarrow \psi) \Rightarrow sup_S(\phi' \rightarrow \neg\psi) > sup_S(\phi \rightarrow \neg\psi).$$

This means that rule r residing on the support–confidence Pareto-optimal border is also on the support–anti-support Pareto-optimal border because one cannot have any other rule r' such that $sup_S(\phi' \to \psi) > sup_S(\phi \to \psi)$ and $sup_S(\phi' \to \neg\psi) > sup_S(\phi \to \neg\psi)$.

Now, we prove with a counter-example that there can be a rule being on the support–anti-support Pareto-optimal border which is not on the support–confidence Pareto-optimal border. Let us consider two rules r and r' residing on the support–anti-support Pareto-optimal border such that for rule r we have support $sup_S(\phi \to \psi) = 200$ and anti-support $sup_S(\phi \to \neg\psi) = 100$, while for rule r' we have support $sup_S(\phi' \to \psi) = 150$ and anti-support $sup_S(\phi' \to \neg\psi) = 99$. We have that $conf_S(\phi \to \psi) = 0.667$ which is greater than $conf_S(\phi' \to \psi) = 0.602$. Thus, rule r' is not on the support–confidence Pareto-optimal border because it is dominated in the sense of support–confidence by rule r having a larger support and a larger confidence. □

Let us observe that the support–confidence Pareto-optimal border has the advantage of presenting a smaller number of rules (more precisely, a not greater number of rules) than the support–anti-support Pareto-optimal border. However, its disadvantage is that it does not present the rules optimizing any attractiveness measure satisfying the property M. In fact, all the rules which are present on the support–anti-support Pareto-optimal border and not present on the support–confidence Pareto-optimal border maximize an attractiveness measure which is not monotone with respect to support because it does not satisfy the condition of the above Theorem 4.8.

Summarizing illustration of comparison of the support–anti-support Pareto-optimal border with non-dominated sets from all previously mentioned evaluation spaces is presented on Figure 7. On the diagram there are the Pareto-optimal borders with respect to four evaluation spaces, for a fixed conclusion being income<=50 kUSD. We can observe on Figure 7 that, as it has been analytically shown, the support–anti-support Pareto-optimal border is the upper-set of all the other discussed Pareto-optimal sets. Moreover, the support–confidence Pareto-optimal border contains the same rules as the support–f Pareto-optimal set, whereas the set of non-dominated rules with respect to support and s is their subset.

5.1 Confirmation perspective on support–anti-support evaluation space

Since neither support nor anti-support has the property of confirmation, by using only these measures one cannot distinguish which rules are disconfirming and should be discarded from further analysis. Thus, it would be valuable to enrich the support–anti-support evaluation space by the strength of confirmation property. This would allow to limit the set of rules by eliminating those that are characterized by non–positive or small values of a confirmation measure.

Let us consider a confirmation measure $c_S(\phi \to \psi)$ with the property M (e.g. measure f). Let us observe that $anti\text{-}sup_S(\phi \to \psi) = sup_S(\phi \to \neg\psi) =$

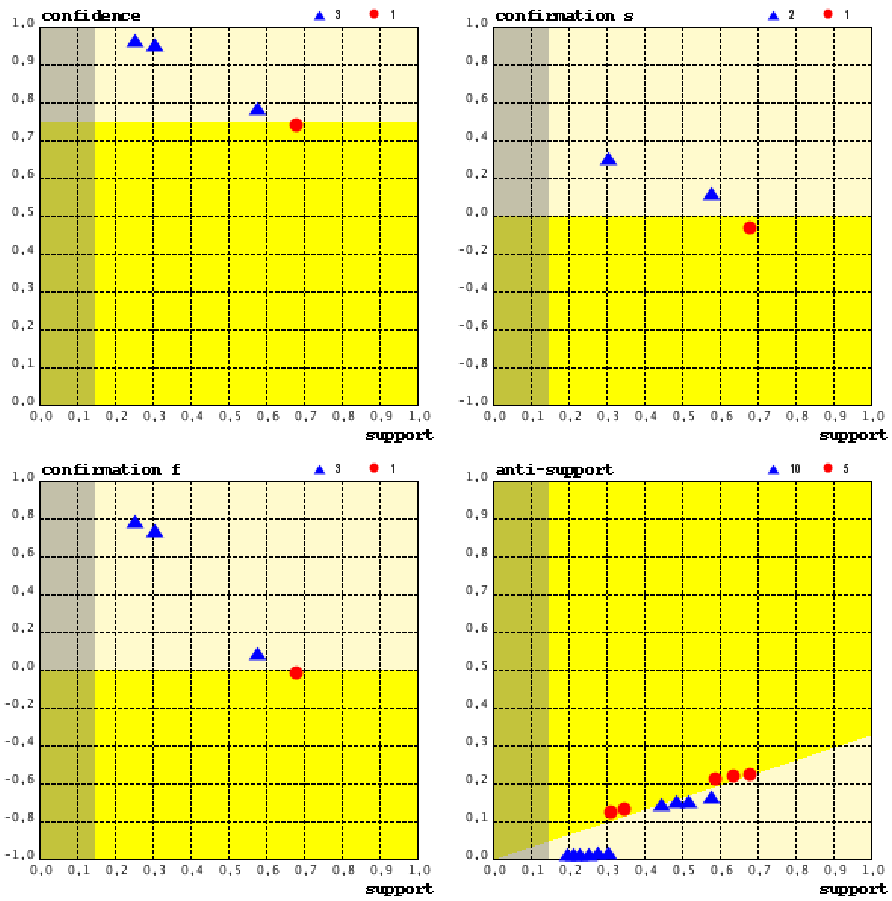


Fig. 7. Comparison of Pareto-optimal sets in different rule evaluation spaces, for minimal support=0.15 (conclusion: income<=50 k USD, *census* dataset)

$sup_S(\phi) - sup_S(\phi \to \psi)$. Thus, a simple transformation of condition (24).1 leads to the following result:

$$c_S(\phi \to \psi) \geq 0 \Leftrightarrow anti-sup_S(\phi \to \psi) \leq sup_S(\phi \to \psi)\left[\frac{|U|}{sup_S(\psi)} - 1\right]. \quad (48)$$

Moreover, due to Theorem 5.2, for a fixed value of rule support, any confirmation measure $c_S(\phi \to \psi)$ having the desired property M is anti–monotone (i.e. non–decreasing) with respect to anti-support. Let us also stress that all confirmation measures (no matter whether having the property M or not) change their signs in the same situations. Thus, the possession or not of property M will not influence our further discussion.

Since, we limit our consideration to rules with the same conclusion and we assume that the data table S does not change, $|U|$ and $sup_S(\psi)$ should be regarded as constant values. Thus, due to the anti–monotone link between $c_S(\phi \to \psi)$ and anti-support, (48) shows that a simple linear function bounds rules that are characterized by positive values of confirmation from those with non–positive confirmation values. For those rules ψ is satisfied less frequently when ϕ is satisfied rather than generically. Figure 8 illustrates this analytical result. The diagram shows three linear functions drawn for three conclusions varying in the number of objects supporting them. The blue line signifies the conclusion which is the largest in cardinality, the green one stands for the smallest class. The functions separate the rules with positive confirmation value (marked by blue triangles) from those with non-positive (marked by red circles). The angle of inclination of each of those functions is determined by the cardinality of the set representing the conclusion, and thus, the larger the class cardinality, the more demanding the function is (i.e. the smaller the angle).

It is also interesting to investigate a more general condition $c_S(\phi \to \psi) \geq k$, $k \geq 0$, for some specific confirmation measures. In the following, we consider again the confirmation measure $f_S(\phi \to \psi)$.

Theorem 5.4. *[86]*

$$\begin{gathered} f_S(\phi \to \psi) \geq k \\ \Leftrightarrow \\ anti\text{-}sup_S(\phi \to \psi) \leq sup_S(\phi \to \psi)\left(U - sup_S(\psi)\right)\tfrac{1-k}{(1+k)sup_S(\psi)}. \end{gathered} \quad (49)$$

Proof. For given U and ψ, measure $f_S(\phi \to \psi)$ can be written in terms of support and anti-support of rule $\phi \to \psi$ as follows:

$$f_S(\phi \to \psi) = \frac{\frac{sup_S(\phi\to\psi)}{sup_S(\psi)} - \frac{anti-sup_S(\phi\to\psi)}{|U|-sup_S(\psi)}}{\frac{sup_S(\phi\to\psi)}{sup_S(\psi)} + \frac{anti-sup_S(\phi\to\psi)}{|U|-sup_S(\psi)}}. \quad (50)$$

Transforming (50) as the dependency of $anti\text{-}sup_S(\phi \to \psi)$ on $f_S(\phi \to \psi)$ and considering the inequality $f_S(\phi \to \psi) \geq k$, we obtain the thesis. □

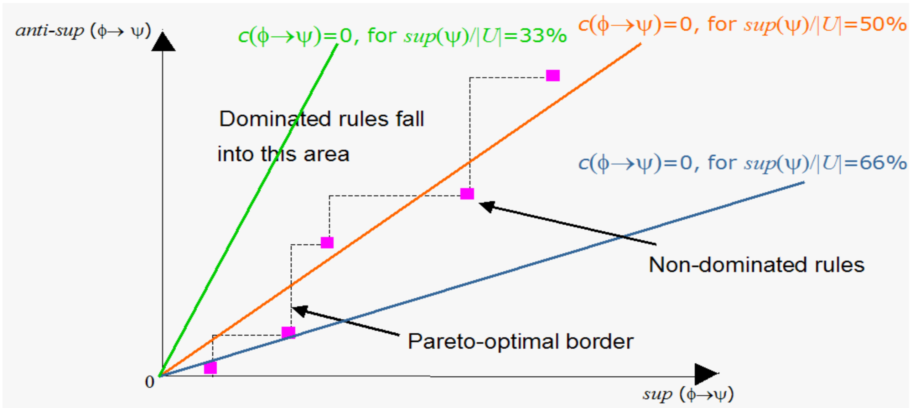
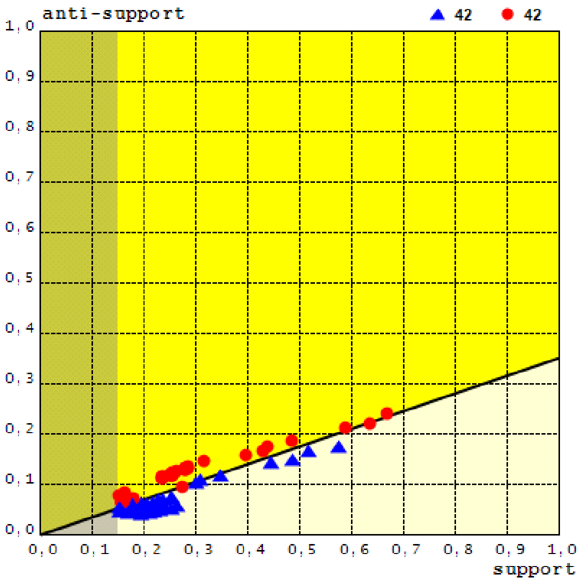


Fig. 8. Three examples of linear functions representing $c_S(\phi \rightarrow \psi)$=0 in a support–anti-support space. Each line was drawn according to a set of rules for conclusions different in cardinality. Rules laying above these functions should be discarded from further analysis

Figure 9 presents rules generated from the *census* dataset for the conclusion: workclass='Private' in a support–anti-support space. The semantic scale of anti-support is weaker than that of confirmation measures as it cannot show rules for which the premise disconfirms the conclusion. Therefore, despite the fact that the support–anti-support Pareto–optimal border contains all rules that are optimal according to any measure with the property M, it is necessary to take under consideration also the information brought by the sign of a confirmation measures. In the set of both dominated and non–dominated rules, there can be examples of rules with negative values of confirmation. The second chart on Figure 9. presents just the rules which form the support–anti-support Pareto–optimal border. Within the Pareto-optimal set presented on Figure 9, over 23% of rules need to be discarded from further analysis as their value of confirmation is non–positive. On the diagrams, a linear function was placed separating the rules with positive confirmation (situated under the line) from those with non–positive confirmation (situated above the line) to visualize result (48).

Table 4 presents the percentage of rules that should be discarded from the Pareto–optimal border with respect to support and anti-support, for different conclusions in the *census* dataset. The support–anti-support Pareto–optimal border is, in general, larger (or precisely, not smaller) than the support–confidence Pareto–optimal set. The first set fully contains the latter, and therefore it is obvious that if there appeared some confirmation–negative rules on the support–confidence Pareto–optimal border then they would also be present on the support–anti-support Pareto–optimal border. But as it can be observed in Table 4, on the support–anti-support Pareto–optimal border there also came up other rules with non-positive confirmation values.



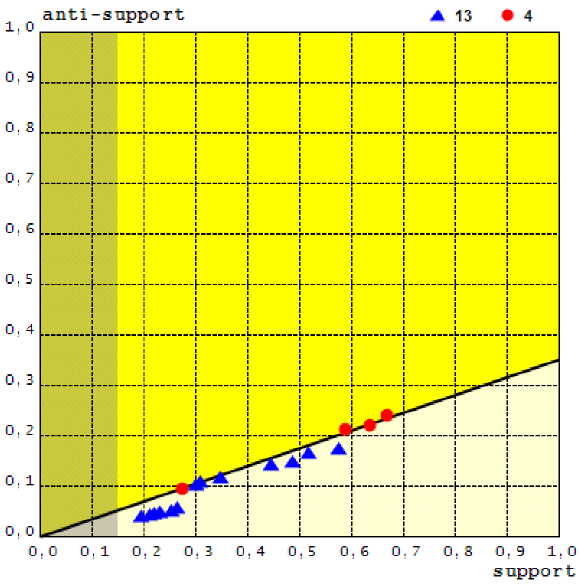


Fig. 9. Rules with positive (triangles) and non–positive (circles) confirmation measure value in a support–anti-support space. First diagram shows all generated rules, second one shows the Pareto-optimal border only (conclusion: workclass='Private', *census* dataset)

Table 4. Information about the percentage of rules with non-positive confirmation laying on the support-anti-support Pareto-optimal border for different conclusions (*census* dataset)

Considered conclusion	No. of all rules on Pareto border	No. of rules with non-positive confirmation	Reduction percentage
workclass=Private	17	4	24%
sex=Male	8	3	38%
income<=50 kUSD	15	5	33%
race=White	17	1	0.6%
native_country=USA	15	0	0%

6 Examples of application of attractiveness measures to multicriteria rule evaluation

The number of different attractiveness measures, which can be regarded as rule evaluation criteria, described in the literature is overwhelming. As we have shown in the previous sections, there exist many relationships between those criteria, despite which, however, the semantics of the measures still vary. Therefore, there arises a need to combine few interestingness measures to form a multicriteria evaluation space.

The monotonic or anti-monotonic relationships between different measures that have been pointed out in the previous sections (in particular Theorem 5.1 and Theorem 5.2) can be applied to allow an increase of efficiency while determining of the rules from the Pareto-optimal border or from the area close to it. It is a direct result of the anti-monotonic dependency of all of the considered measures on the anti-support. Practically, it means that, for a set of rules with the same conclusion, the order of rules according to confidence, f, s, or any other measure with the property M, for a fixed value of rule support, is in concordance (consistent) with the order determined by the measure of rule anti-support. Hence, having ordered a set of rules once (e.g. with respect to rule support and anti-support), we can "re-use" the same order many times in other evaluation spaces (e.g. in support–confidence, support–f, etc.).

Moreover, assuming induction of rules for any conclusion and according to a user-defined maximal acceptable rule anti-support and minimal rule support thresholds, the rule generation efficiency can be increased due to the following relationship between rule support and anti-support.

Remark 6.1. When generating decision or association rules from a frequent set it is advisable to first generate rules with few conclusion elements (for optimization reasons).

Explanation: Let us assume that a frequent (item)set is a set occurring in the dataset at least as many times as the value of the minimal support threshold requires. Let us consider three different rules constructed from the same frequent itemset $\{x, y, z, v\}$:

1. $r_1 : x \rightarrow yzv$ $\quad$ $anti\text{-}sup_S(r_1) = sup_S(x) - sup_S(xyzv)$,
2. $r_2 : xy \rightarrow zv$ $\quad$ $anti\text{-}sup_S(r_2) = sup_S(xy) - sup_S(xyzv)$,
3. $r_3 : xyz \rightarrow v$ $\quad$ $anti\text{-}sup_S(r_3) = sup_S(xyz) - sup_S(xyzv)$.

Clearly, $anti\text{-}sup_S(r_1) \geq anti\text{-}sup_S(r_2) \geq anti\text{-}sup_S(r_3)$. Therefore, we can conclude that if
$anti\text{-}sup_S(r_3) > max\_acceptable\ anti\text{-}support$ then
$anti\text{-}sup_S(r_2) > max\_acceptable\ anti\text{-}support$.
This observation means that if we generate r_3 and verify that its rule anti-support is unacceptable, because it is higher than the anti-support threshold, then we should skip the phase of r_1 and r_2 generation. This result allows us to limit the space of the rules to be generated from the frequent itemsets and this way gain on efficiency.

6.1 Multicriteria rule evaluation system

The system is available at http://www.cs.put.poznan.pl/iszczech/research.html

System concept. The system composes of the following modules:

1. File Processing Unit,
2. Frequent Itemset Generator,
3. Rule Generator,
4. Ordering and Optimization Unit,
5. Visualization Unit,
6. User Interaction Unit.

They cooperate according to the structure presented on Figure 10.

File Processing Unit is responsible for accessing the input datasets and adjusting them to the format the systems operates on. The acceptable input formats are .arff and .data, but the system can be easily extended to handle other input data formats as well.

Frequent Itemset Generator is a module searching for frequent itemsets (i.e. itemsets that occur in the input file at least as often as it is required by the minimal rule support threshold). The threshold is obtained from the user through the User Interaction Unit.

The user can choose the frequent itemset algorithm to be applied. There are two options: Apriori by Agrawal et al. [2] and FP-Growth by Han et al. [32]. For detailed presentation of those algorithms see also: [4], [60], [92].

The Apiori algorithm represents an iterative approach to association mining. In the first step, single-element frequent item sets are selected from the database

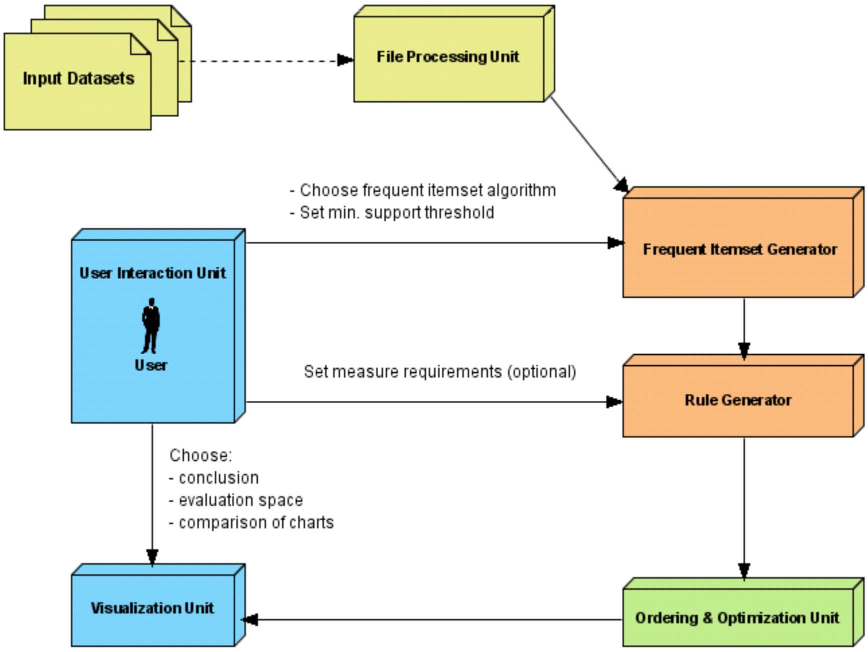


Fig. 10. The component diagram of the system

Apriori:

```
FSet1 = frequent 1-itemsets;
for(count = 2; FSetcount-1 ≠ ∅; count++)
{
    Candcount = GenerateCandidateSets(FSetcount-1);
    foreach(Record r in database)
    {
        foreach(CandidateSet cs in Candcount)
        {
            if(∀i i∈cs ⇒ i∈r)
                cs.support++;
        }
    }
    FSetcount = {cs ∈ Candcount : cs.support >= minsup};
}
Result = Ucount Fsetcount;
```

GenerateCandidateSets(*FSetcount*):

```
1:  foreach(pair of sets (s1, s2) in FSetcount)
2:  {
3:      union = s1 ∪ s2;
4:      if(union.count == count + 1)
5:          CandidateSets.Add(union);
6:      foreach(Set s in CandidateSets)
7:      {
8:          subsets = count-subsets of s;
9:          foreach (Set sub in subsets)
10:         if(sub ∉ FSetcount)
11:             CandidateSets.Remove(s);
12:     }
13: }
```

(Apriori line 1). Based on these sets, larger frequent sets are found by generating candidate sets (Apriori line 4) and by prunning (Apriori lines 5-13) them.

The first step of generating candidate sets is merging. Sets with size count are summed to create candidate sets of count+1 size (GenerateCandidateSets lines 3-5). Each of these newly created sets must be then verified to ensure it is a frequent set. This is done by checking if all count subsets of the candidate set are frequent (GenerateCandidateSets lines 6-11).

The database is then scanned to verify each candidate set's support and eliminate those which do not satisfy the minimal support threshold. Later frequent sets with size 2 are used to generate frequent sets of size 3 and so on. In each iteration the algorithm generates frequent sets that are one element bigger and with each iteration the database needs to be scanned. The algorithm ends when no more candidate sets can be generated.

The process of rule mining using the FP-Growth algorithm consists of two major steps. In the first step the database is transformed into a special structure called the FP-Tree. In the second step the FP-Tree is explored recursively in search of frequent sets.

To transform a database into an FP-Tree, single-element frequent sets must be selected and sorted by support in descending order for each row of the database (CreateTree lines 1-2). Then an empty FP-Tree is created with a null labeled root (CreateTree line 3). Reading the database for the second time we only read the frequent elements and create an FP-Tree as shown in CreateTree lines 6-17. If there is no child labeled as the item, a new node is created with the desired label and support = 1. Otherwise, we increment the child's support by 1 and consider it as the new root.

Upon reaching a new record with elements to add to the tree, we return to the null-labeled root (CreateTree line 6). During the generation of the tree a header list of item nodes is created in order to localize nodes containing the same item more rapidly and to identify each item's support. An FP-Tree constructed in such a way is then recursively explored by calling FP-Growth (Tree, null).

CreateTree():

```
FSet1 = frequent 1-itemsets;
SortDesc(FSet1);
Tree = new FPTree(null);
foreach(Row r in database)
{
    CurrentNode = Tree.Root();
    foreach(Item i in FSet1)
    {
        if(i ∈ r and i ∉ CurrentNode.Children)
        {
            CurrentNode.AddChild(i);
            CurrentNode = CurrentNode.Children(i);
        }
        if(i ∈ r and i ∈ CurrentNode.Children)
        {
            CurrentNode = CurrentNode.Children(i);
            CurrentNode.Support++;
        }
    }
}
```

FP-Growth(*Tree*, α):

```
if(Tree has a single path P)
{
    foreach(subset β of nodes in P)
    {
        fs = α ∪ β;
        fs.support = mini.support{i: i∈β}
        Result.Add(fs);
    }
}
else
{
    foreach(Item i in Header)
    {
        β = i ∪ α;
        β.support = i.support;
        Result.Add(β);
        create β's conditional pattern base;
        create β's conditional FPTree Treeβ;
        if (Treeβ ≠ ∅)
            FP-Growth(Treeβ, β);
    }
}
```

The recursive process of exploring the FP-Tree is based on distinguishing two situations. If the tree has only a single path then all non-empty subsets of that path are combined with the suffix pattern α (with which the function was called) and added to the result frequent pattern list (FP-Growth lines 1-9).

Otherwise, if the tree consists of more than one path then the header list is read in support ascending order (FP-Growth lines 12-21). The examined item is merged with suffix pattern and added to the result list with support equal to the item's support (FP-Growth lines 14-16). Based on paths containing nodes with the currently explored item a conditional pattern base is created and later a conditional FP-Tree (FP-Growth lines 17-18). The idea is to filter items that occur with the new suffix (the item currently read from header list) often enough to satisfy the minimal support value. Once these items are filtered a new FP-Tree is created in a similar way to the one created from the database (the support values of the nodes are incremented by the node's support value and not 1 like in CreateTree()). If the tree is not empty we recursively call FP-Growth with the new tree and suffix (FP-Growth lines 19-20).

Rule Generator is a unit that generates rules from the frequent itemsets found by the Frequent Itemset Generator. The basic rule generation algorithm is based on creating rules from the subsets of frequent itemsets:

GenerateRules(*FrequentSets*):

```
foreach (FrequentSet fs in FrequentSets)
{
    subsets = list of all non-empty subsets of fs;
    foreach (Subset sub in subsets)
        create rule: sub ⇒ (fs – sub);
}
```

The rules can be formed with respect to different attractiveness measures i.e. the user can optionally set the following thresholds: maximal acceptable rule anti-support, minimal acceptable confidence, f or s measure value. On the Rule Generator's output there will only be rules that satisfy the introduced thresholds. The results from Remark 6.1 are used to generate rules with a given anti-support threshold more effectively. For each generated rule the values of the following counters are obtained: $a = sup_S(\phi \to \psi)$, $b = sup_S(\neg\phi \to \psi)$, $c = sup_S(\phi \to \neg\psi)$ and $d = sup_S(\neg\phi \to \neg\psi)$. On their bases, for each rule, measures of rule support and anti-support, confidence, f and s are calculated. The user can see the output rules with their values of the considered attractiveness measures, in form of a table (see example on Figure 11).

The user can also choose a particular attribute to be a decision attribute and in that case only rules with that attribute in the rule's conclusion will be generated. If the user does not assign any decision attribute, association rules are generated.

premise	conclusion	supp	conf	s	f	a-supp
native-country is United-States	education is Bachelors	0,15	0,17	0,01	0,00	0,76
education is Bachelors	native-country is United-States	0,15	0,92	0,00	0,02	0,01
workclass is Private	education is Some-college	0,17	0,23	0,01	0,01	0,57
education is Some-college	workclass is Private	0,17	0,75	0,02	0,03	0,06
native-country is United-States	workclass is Private and education is Some-college	0,16	0,17	0,04	0,01	0,76
native-country is United-States and workclass is Private	education is Some-college	0,16	0,23	0,03	0,03	0,51

Fig. 11. Example of rule presentation format (*census* dataset)

Ordering and Optimization Unit is a module that divides the set of all rules into subsets according to their conclusions. Rules from each of such group are ordered with respect to their value of support and with respect to anti-support when the value of support is the same. Such ordering allows to optimize the phase of finding the Pareto-optimal border (or the area close to it) in the support–anti-support evaluation space because there is no need to perform n^2 of comparisons on rules (where n is the number of rules with the same conclusion). For each group of rules with the same conclusion, this Pareto-optimal border is found in the following manner (for reference see also JumpAlg):

1. the first element from the ordered list (i.e. the rules that have the highest rule support and the smallest anti-support) are placed on the Pareto-optimal border, for they are surely non-dominated,
2. we jump to the element on the list with the next highest value of support and search for a rule(s) whose anti-support is smaller than the anti-support of the element chosen in the previous step. If such rule(s) is found, it is added to the Pareto-optimal border. The procedure is then continued for the next highest values of support until the element with the smallest support is reached.

JumpAlg:

```
for (idx = 0; idx < Rules.count; idx += Rules[idx].EqualCount)
{
    if(idx ==0 ||Rules[idx].Measure > Rules[lastPareto].Measure)
    {
    ParetoOptimal.Add(Rules[idx]);
    lastPareto = idx;
    }
}
```

where: *Measure* has property *M*.

Due to relationships between anti-support and other considered measures (in particular Theorem 5.1 and Theorem 5.2) the fact that the rules are ordered with respect to anti-support implies that they are also ordered according to confidence, measure f or s, etc. This means that the above described way of searching for Pareto-optimal borders can also be applied to looking for Pareto-optimal rules with respect to support and confidence, support and f, etc. Moreover, since the support–anti-support Pareto-optimal border is the upper-set of all the considered Pareto-optimal sets, we can mine them simply from the support–anti-support Pareto-optimal border instead of searching the set of all rules.

Visualization Unit is responsible for presentation of the induced rules on diagrams. The user determines (through the User Interaction Unit) the conclusion with which the rules are to be displayed and the evaluation space. The user can view many charts at once in order to compare them. The thresholds for the evaluation criteria can be adjusted by the user, changing the set of rules that is presented. On the diagrams, rules with non-positive confirmation value are distinguished by color and shape from those with positive value. The user can limit the set of displayed rules only to the Pareto-optimal ones or view the whole set of rules with the chosen conclusion. The system presents the values of rule support and anti-support as relative values between 0 and 1.

User Interaction Unit is a module that provides communication between the user and other system components. All the user-set parameters (e.g. thresholds) are delivered to the system through this unit.

System functionality.

Association miner - general information. Association miner is an association rule mining program that utilizes the Apriori or FP-Growth algorithm. It enables the user to create and view charts presenting rules in support–confidence, support–s, support–f and support–anti-support planes. Rules can be saved in special structures for later use as well as be exported to Excel format. *Association miner* benefits are:

1. efficient Pareto-optimal rule generation in several measure evaluation spaces together in one step,
2. chart exemplification of important theoretical thesis,
3. rule export capabilities.

System Requirements. Microsoft .Net Framework 2.0 or higher, Windows XP

Installation. The program does not require any special installation steps and is ready to use as long as all of the system requirements are met.

Starting the program. To start *Association miner*, click the *Association miner* icon found in the installation directory.

Brief User Guide.

Opening data files. To open a data file simply choose **Open Data File** from the **File** menu option list or press Ctrl+O while using the main form of the program. You can choose from *.arff and MSWeb *.data file formats. When asked type in the desired minimal support and optionally other measure requirements to generate rules. If you decide to cancel at this point, no rules will be generated but the database will be loaded.

Managing rule files. Once rules are generated you can save them by accessing the **Save Rule File** option from the **File** menu. Rules can be saved to the default *.rff format or exported as an Excel sheet.

Rules can also be loaded through the **Open Rule File** option from the **File** menu. Only *.rff files are supported by this option.

Chart and multichart creation. To create a chart or multichart simply choose the desired option from the **Options** menu. Charts are only shown for rules with the same conclusion, so it is necessary to select one of the conclusions from the provided list.

When a chart has been created the user has the possibility to filter rules through various options:

1. Show only Pareto,
2. Show only selection,
3. Show invisible.

Figure 12 presents the main form of the program and gives access to all the options provided by the *Association Miner*:

– **premise**: if clicked, sorts the rules according to the alphabetical order of the premise column
– **conclusion**: if clicked, sorts the rules according to the alphabetical order of the conclusion column
– **supp**: if clicked, sorts the rules according to the support column
– **conf**: if clicked, sorts the rules according to the confidence column
– ***s***: if clicked sorts, the rules according to the confirmation-s value column
– ***f***: if clicked sorts, the rules according to the confirmation-f value column
– **a-supp**: if clicked sorts, the rules according to the anti-support column

An example of the form with generated rules is presented on Figure 13.

As it is shown on Figure 14, the algorithm for frequent itemset generation can be changed according to the user preferences.

During the rule generation phase the user can set measure thresholds and assign a decision attribute if there is one (see Figure 15). The following values can be set:

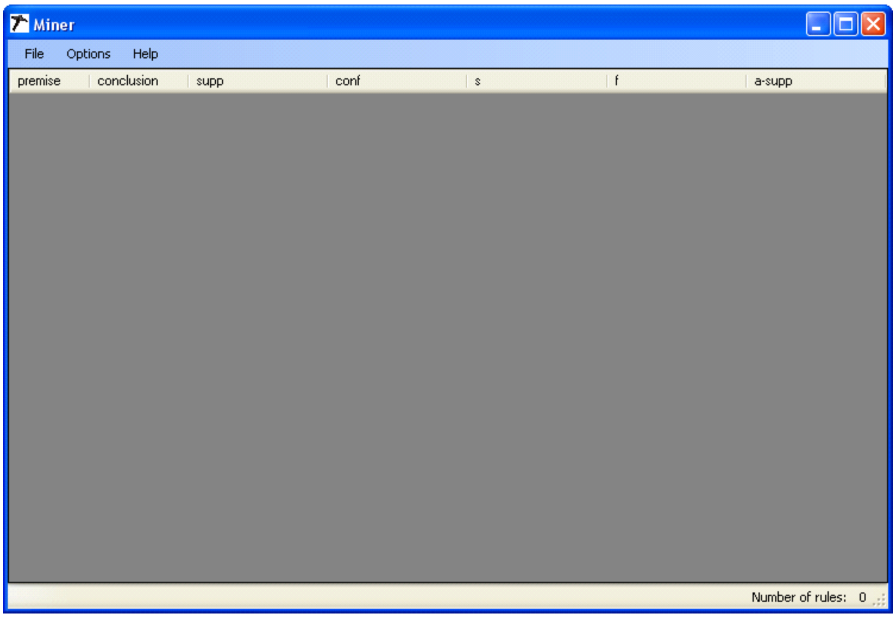
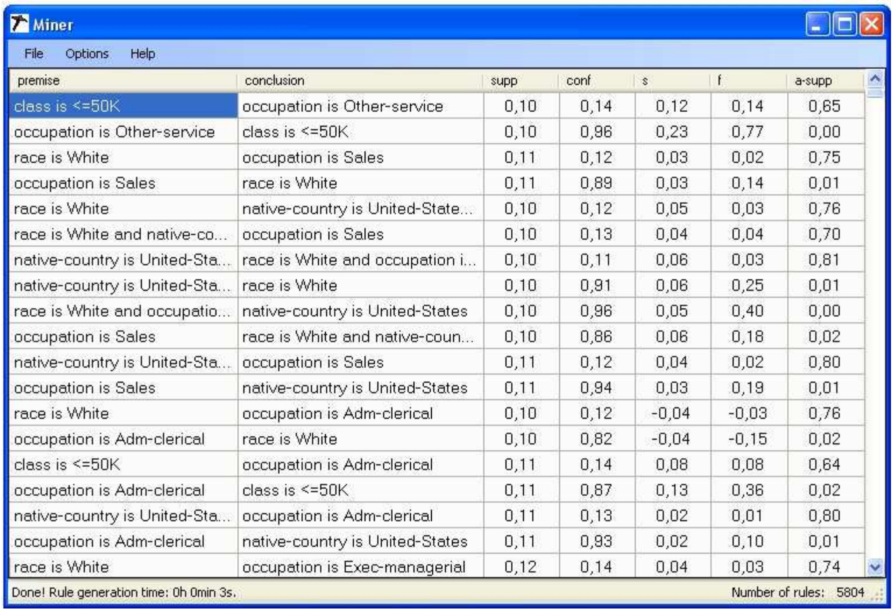


Fig. 12. Association Miner - main form



premise	conclusion	supp	conf	s	f	a-supp
class is <=50K	occupation is Other-service	0,10	0,14	0,12	0,14	0,65
occupation is Other-service	class is <=50K	0,10	0,96	0,23	0,77	0,00
race is White	occupation is Sales	0,11	0,12	0,03	0,02	0,75
occupation is Sales	race is White	0,11	0,89	0,03	0,14	0,01
race is White	native-country is United-State...	0,10	0,12	0,05	0,03	0,76
race is White and native-co...	occupation is Sales	0,10	0,13	0,04	0,04	0,70
native-country is United-Sta...	race is White and occupation i...	0,10	0,11	0,06	0,03	0,81
native-country is United-Sta...	race is White	0,10	0,91	0,06	0,25	0,01
race is White and occupatio...	native-country is United-States	0,10	0,96	0,05	0,40	0,00
occupation is Sales	race is White and native-coun...	0,10	0,86	0,06	0,18	0,02
native-country is United-Sta...	occupation is Sales	0,11	0,12	0,04	0,02	0,80
occupation is Sales	native-country is United-States	0,11	0,94	0,03	0,19	0,01
race is White	occupation is Adm-clerical	0,10	0,12	-0,04	-0,03	0,76
occupation is Adm-clerical	race is White	0,10	0,82	-0,04	-0,15	0,02
class is <=50K	occupation is Adm-clerical	0,11	0,14	0,08	0,08	0,64
occupation is Adm-clerical	class is <=50K	0,11	0,87	0,13	0,36	0,02
native-country is United-Sta...	occupation is Adm-clerical	0,11	0,13	0,02	0,01	0,80
occupation is Adm-clerical	native-country is United-States	0,11	0,93	0,02	0,10	0,01
race is White	occupation is Exec-managerial	0,12	0,14	0,04	0,03	0,74

Fig. 13. Association Miner - main form with generated rules

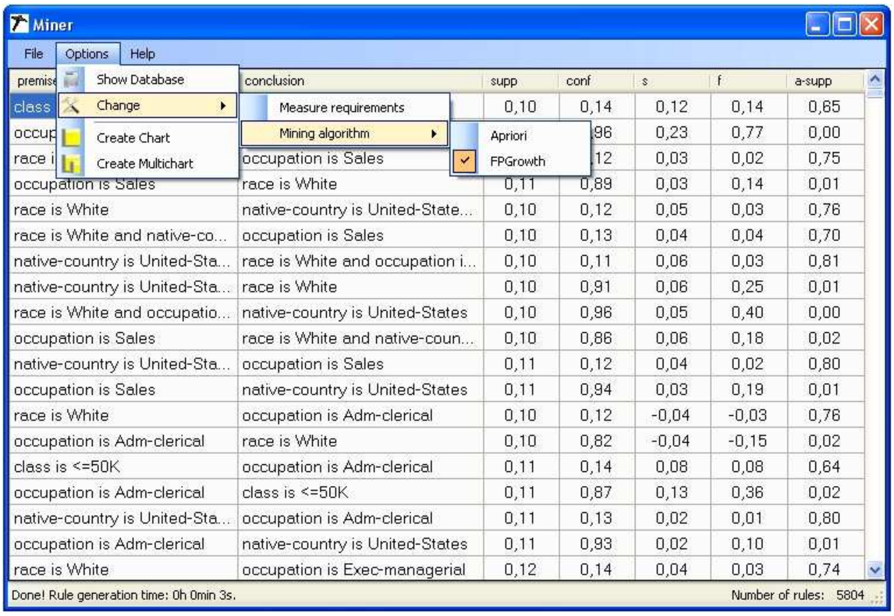


Fig. 14. Association Miner - settings

- **Minimal support value for rule generation**: specifies the minimal support value that will be used to generate rules. This field must be filled in order to generate rules. The provided values must be greater than 0 and less or equal to 1.
- **Minimal confidence**: specifies the minimal confidence value that will be used to generate rules. The provided values must be between 0 and 1.
- **Minimal confirmation-s**: specifies the minimal confirmation-s value that will be used to generate rules. The provided values must be between -1 and 1.
- **Maximal anti-support**: specifies the maximal anti-support value that will be used to generate rules. The provided values must be between 0 and 1.
- **Minimal confirmation-f**: specifies the minimal confirmation-f value that will be used to generate rules. The provided values must be between -1 and 1.
- **Decision attribute**: specifies an attribute that was chosen by the user to be a decision attribute. If no decision attribute is specified, then all possible association rules are generated.

For the generated set of rules the user can create charts presenting rules with a chosen conclusion in different evaluation spaces as shown on Figure 16. The following options can be set in a Chart creation dialog box (see Figure 17):

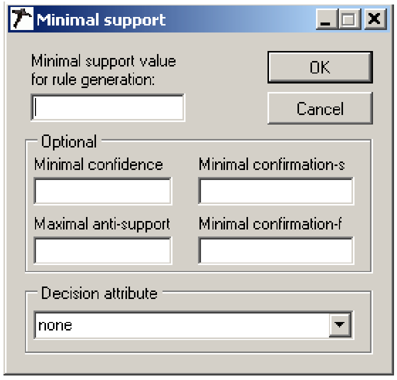


Fig. 15. Association Miner - thresholds

- **Choose conclusion**: allows the user to choose a conclusion for which the chart(s) will be created. Charts can only be created for rules with the same conclusion.
- **Choose chart type**
  - **confidence**: creates a chart showing rules in the support–confidence plane.
  - **confirmation** s: creates a chart showing rules in the support–s plane.
  - **confirmation** f: creates a chart showing rules in the support–f plane.
  - **anti-support**: creates a chart showing rules in the support–anti-support plane.
  - **all**: when selected creates all four of the above charts.

An example of created charts is presented on Figure 18.

The user can adjust the charts to his needs by operating on the following options/parameters:

- **Show border**: Shows the lines separating rules with negative confirmation value. Whether this line is shown or not, one can distinguish the rules with positive confirmation values by their shape of blue triangles (e.g. Figure 19).
- **Show only Pareto**: Shows only rules that are Pareto-optimal in the specified plane (e.g. Figure 20). If the **Show invisible** checkbox is selected then rules that are not Pareto-optimal will be shown as hollow shapes on the chart (e.g. Figure 21). Otherwise the rules that are not Pareto-optimal will not be shown at all.
- **Show only selection**: Shows on the chart only the rules that are within the selection lasso. The selection lasso is always drawn when the user presses the mouse on the chart picture and moves the mouse selecting rules (e.g. Figure 22).
- **Show invisible:** Forces rules that should not be drawn (they are not pareto optimal/not within selection/are in the shaded part of the chart) to appear on the chart as hollow shapes.

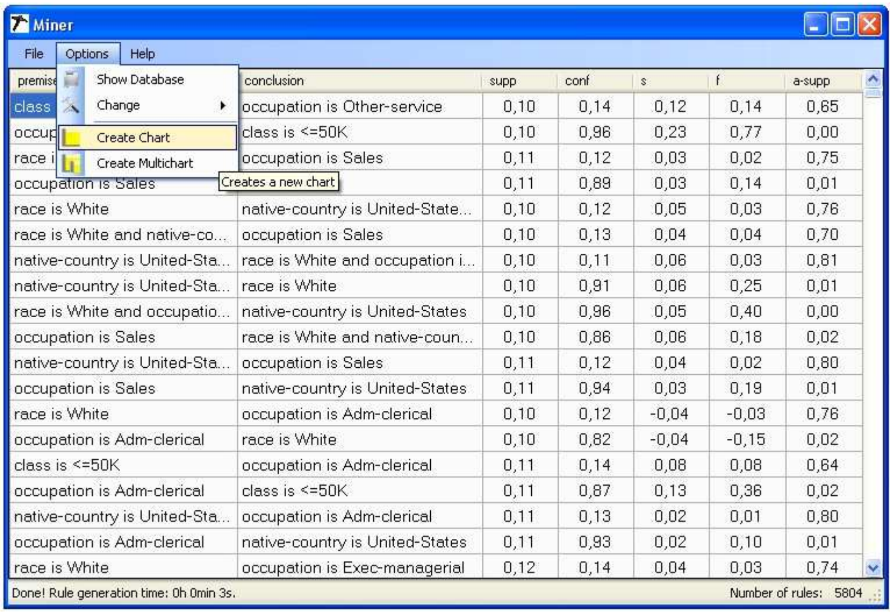


Fig. 16. Association Miner - charts

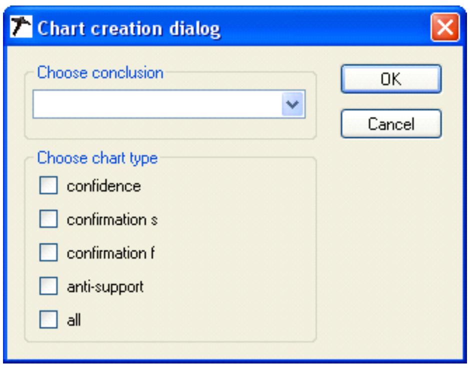


Fig. 17. Association Miner - chart creation parameters

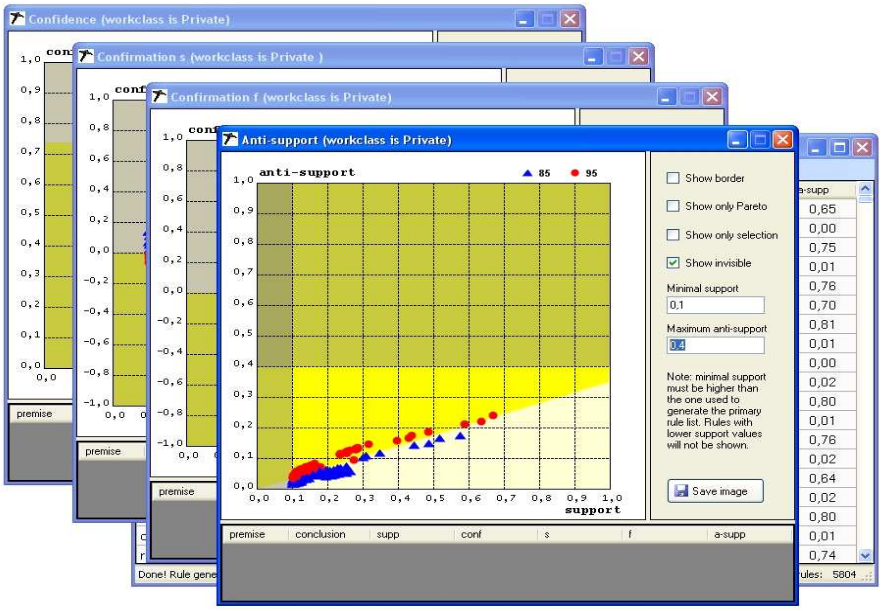


Fig. 18. Association Miner - example charts

– **Minimal support**: Defines the minimal support for the shown rules. If the minimal support provided by the user in the textbox is lower than the one used to generate the rules the field corresponding to that support will still stay shaded.
– **Minimal y-axis value**: Defines the minimal (maximal in case of anti-support) value of the measure in the y-axis of the chart. Rules that do not satisfy these conditions are not displayed on the diagram.
– **Save image**: Saves the chart image in *.bmp format.

The user can also create a multichart suitable for chart comparison (e.g. Figure 23).

6.2 Examples of the system's application

Census dataset. The *census dataset* is a selection from a dataset used by Kohavi et al. in [47]. It contains information about financial and social status of the questioned people. The number of analyzed instances reached 32 561. The chosen instances did not contain any missing values. They were described by 9 nominal attributes differing in domain sizes:

– workclass: Private, Local-gov, etc.;
– education: Bachelors, Some-college, etc.;
– marital-status: Married, Divorced, Never-married, etc.;
– occupation: Tech-support, Craft-repair, etc.;

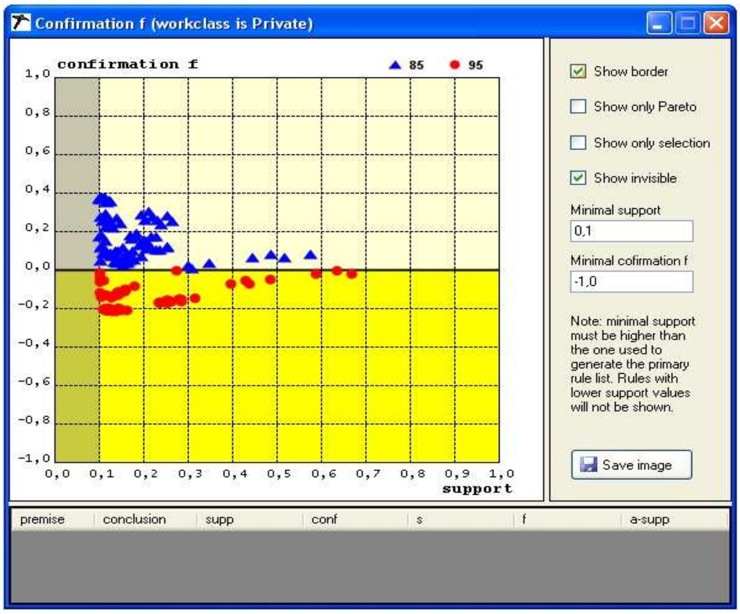


Fig. 19. Association Miner - border

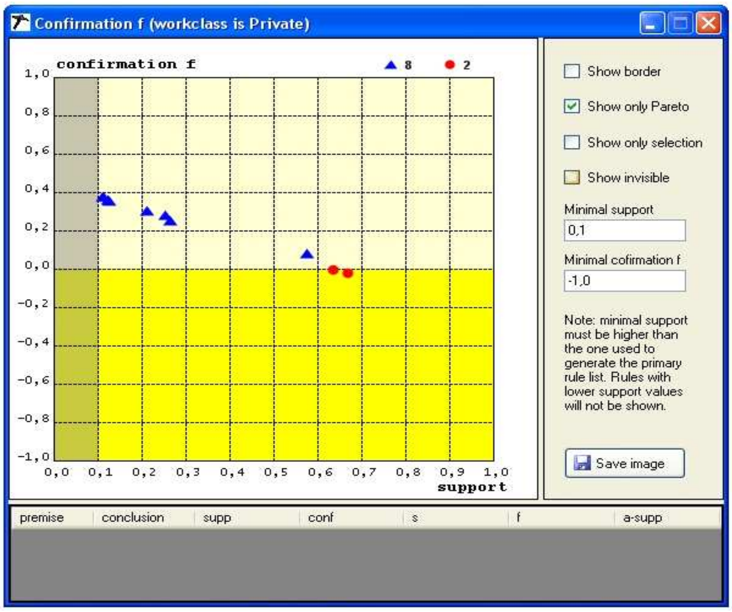


Fig. 20. Association Miner - Pareto-optimal rules

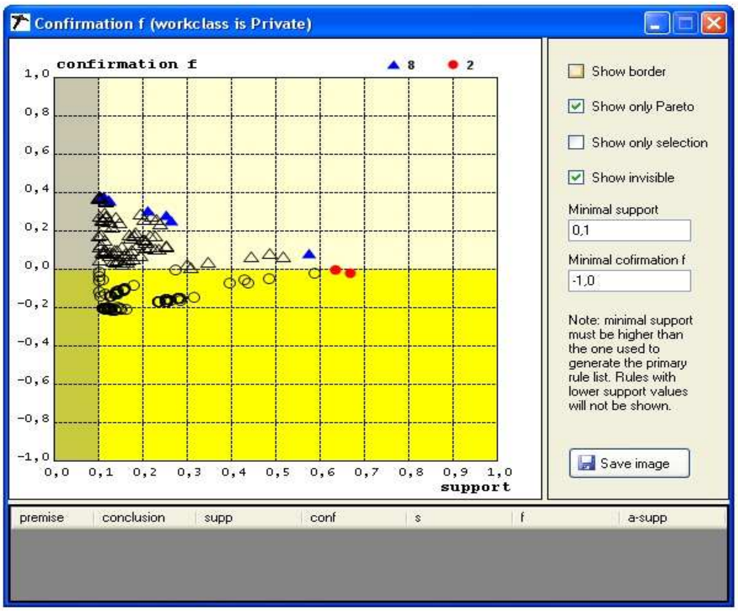


Fig. 21. Association Miner - hidden rules

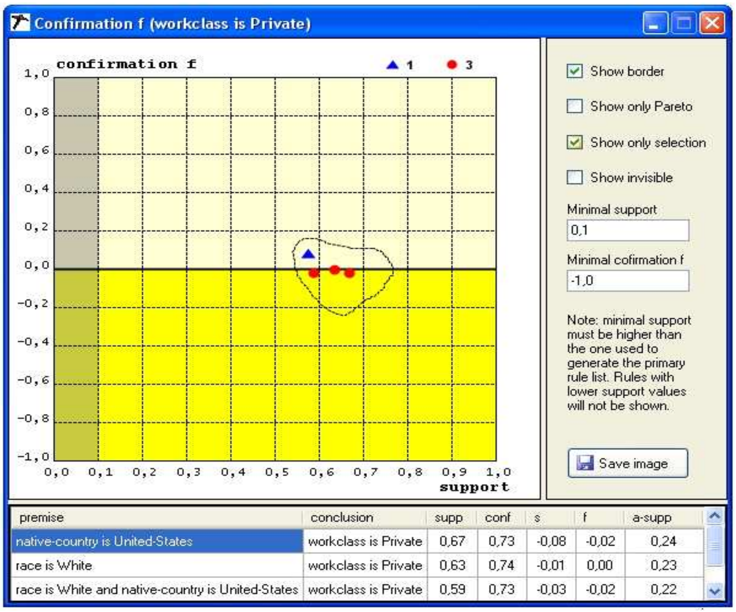


Fig. 22. Association Miner - selection

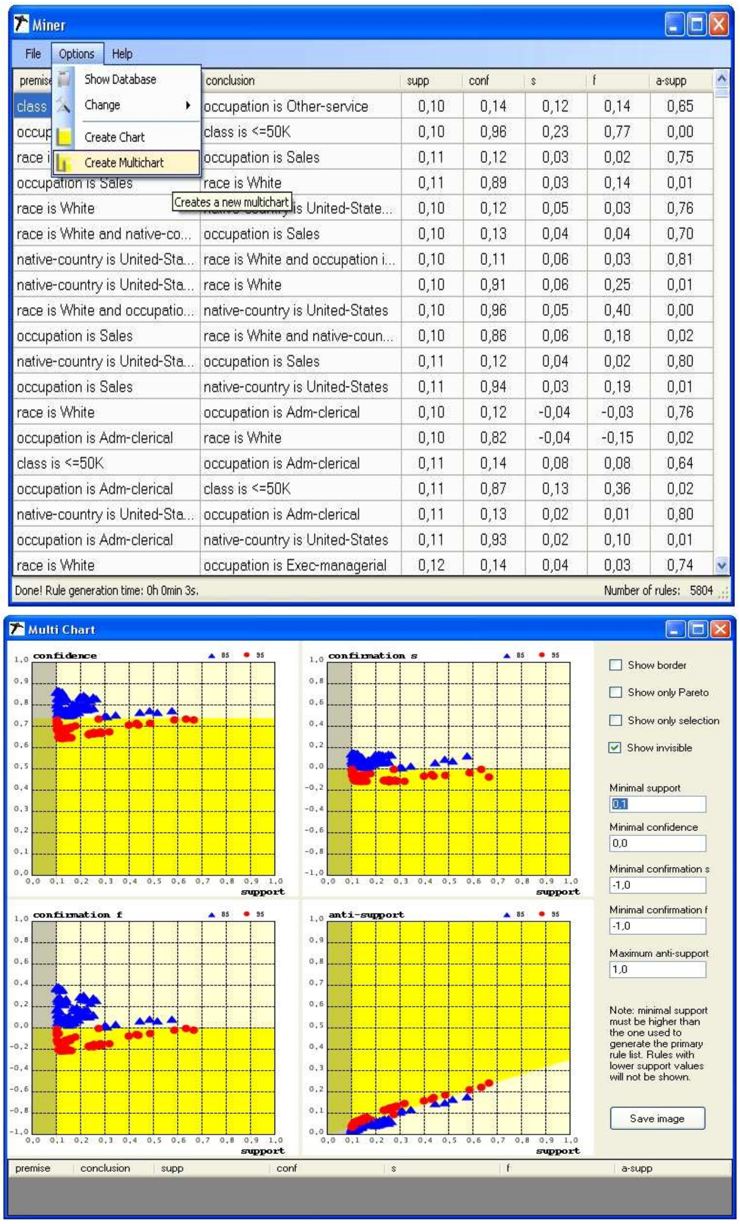


Fig. 23. Association Miner - multicharts

– relationship: Wife, Own-child, Husband, etc.;
– race: White, Asian-Pac-Islander, etc.;
– sex: Female, Male;
– native-country: United-States, Cambodia, England, etc.;
– salary: >50K, <=50K.

We have conducted several experiments with the *census* dataset and the most interesting results are presented below.

For the minimal support threshold set to 0.15, there were over 2200 rules generated. Figure 24 presents all of the rules with the conclusion workclass = 'Private' in different evaluation spaces. The confirmation semantics are added to the charts by the black lines that separate the rules with non-positive confirmation value (marked by red circles) from those with positive confirmation value (marked by blue triangles). It can be observed that 50% of the generated rules need to be discarded as their premises disconfirm the conclusion.

Figure 25 compares the Pareto-optimal borders (for conclusion: workclass = 'Private') obtained in different evaluation spaces. It shows that even the non-dominated rules can be disconfirming. Moreover, we can observe that the support–anti-support Pareto-optimal border is the upper-set of all the other considered non-dominated sets. The two circled points represent the rules with negative confirmation that are on the support–anti-support Pareto-optimal border but are not present on any other Pareto-optimal border. The system allows a detailed insight into each rule and helps to analyse thoroughly the differences between the evaluation spaces.

The system also allows to analyse how a change of one measure threshold is reflected in other evaluation spaces. On Figure 26 we have increased the minimal confidence threshold to 0.78, which resulted in decrease of the number of rules that satisfy this threshold (now only 21 rules are above the thresholds). We can observe how an increase of demands with respect to confidence influences other considered criteria as for any evaluation space the system displays only the rules that satisfy the most severe thresholds (in this case it is the confidence threshold).

MSweb dataset. The *msweb* dataset [64] is anonymous web data collected from www.microsoft.com during one-week timeframe in February 1998. The data was created by sampling and processing the www.microsoft.com logs. The data records the use of www.microsoft.com by 38000 anonymous, randomly-selected users. For each user, the data lists all the areas of the web site that the user visited. Users are identified only by a sequential number and the visited web sites by their title (e.g. "NetShow for PowerPoint") and URL (e.g. "/stream").

The data contains 32 711 transactions with 294 different web sites. The data is sparse and, on average in each transaction there are only 3 elements (web sites).

We have conducted several experiments with the *msweb* dataset and the most interesting results are presented below. Since the dataset is sparse, for most of

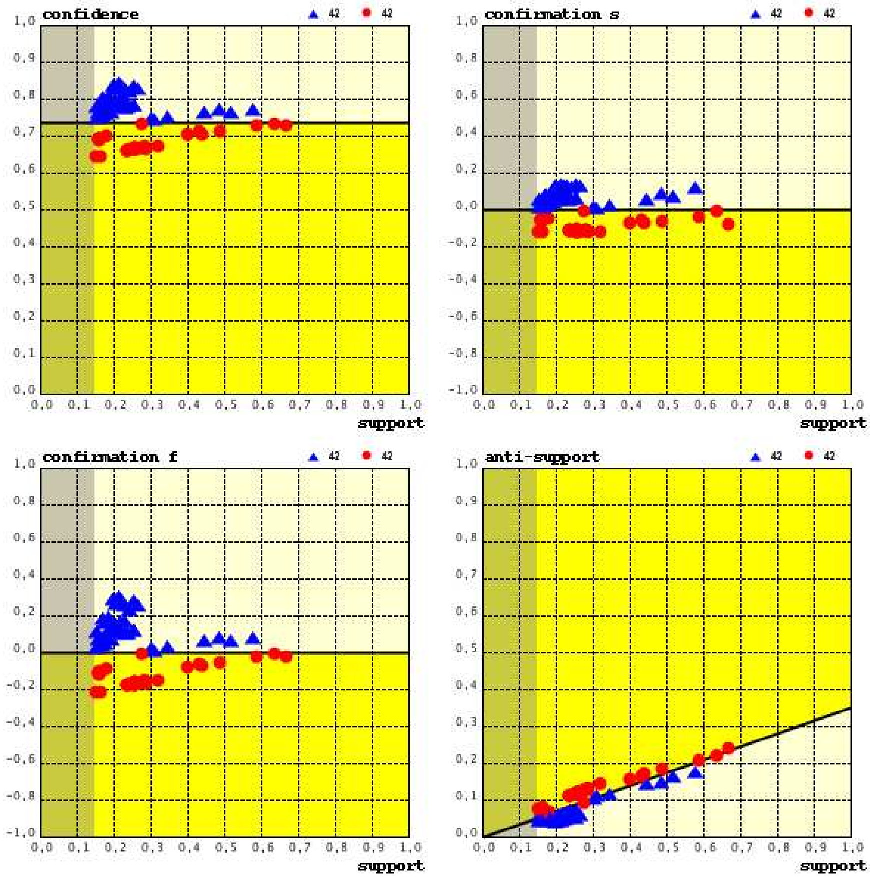


Fig. 24. Multichart presenting all rules with the conclusion workclass="Private" generated with the 0.15 minimal support threshold

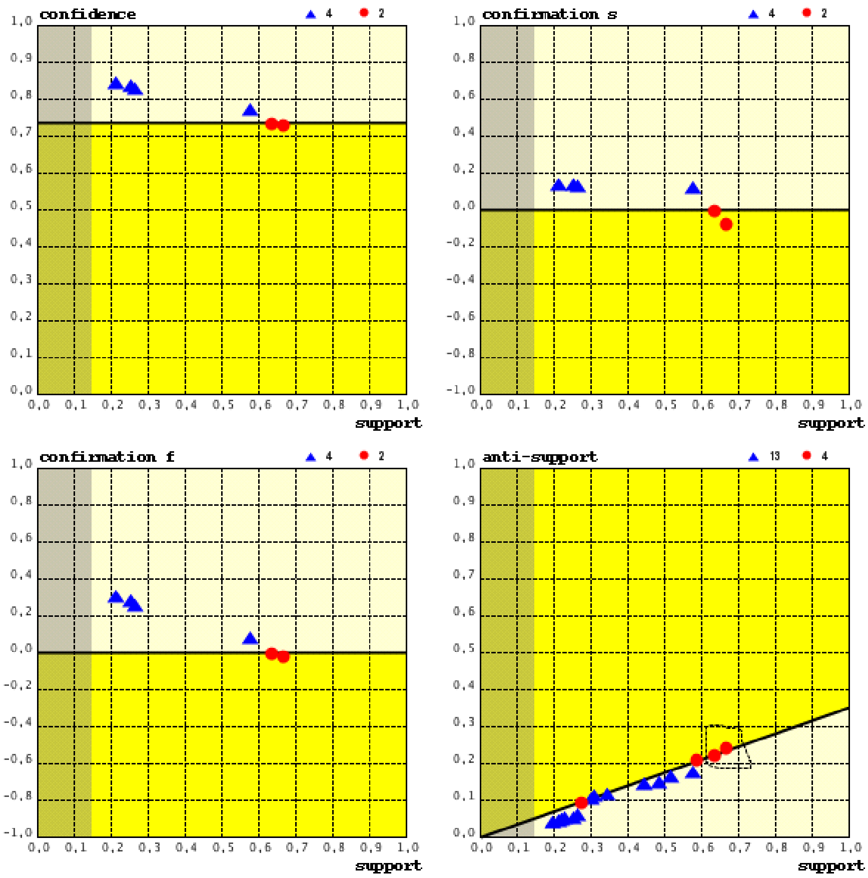


Fig. 25. Multichart presenting only the Pareto-optimal rules with the conclusion workclass="Private" generated with the 0.15 minimal support threshold

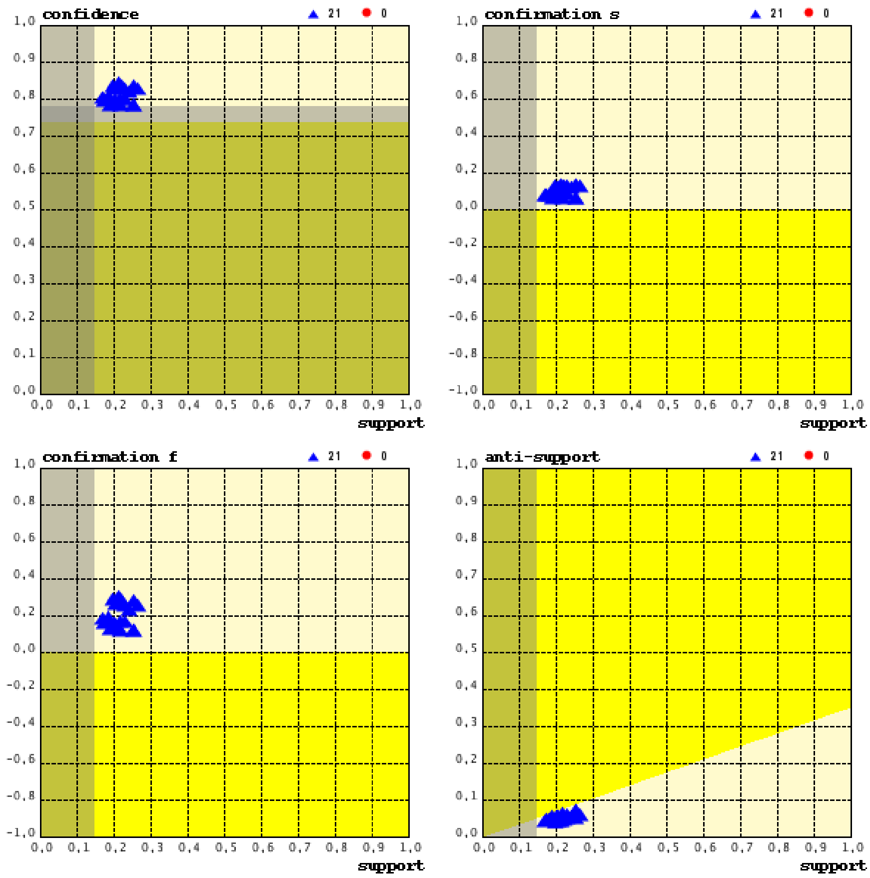


Fig. 26. Multichart presenting all rules with the conclusion workclass="Private" generated with the thresholds of minimal support = 0.15 and minimal confidence = 0.78

the conclusions there were only a few rules even if the minimal support threshold was lower than 0.1.

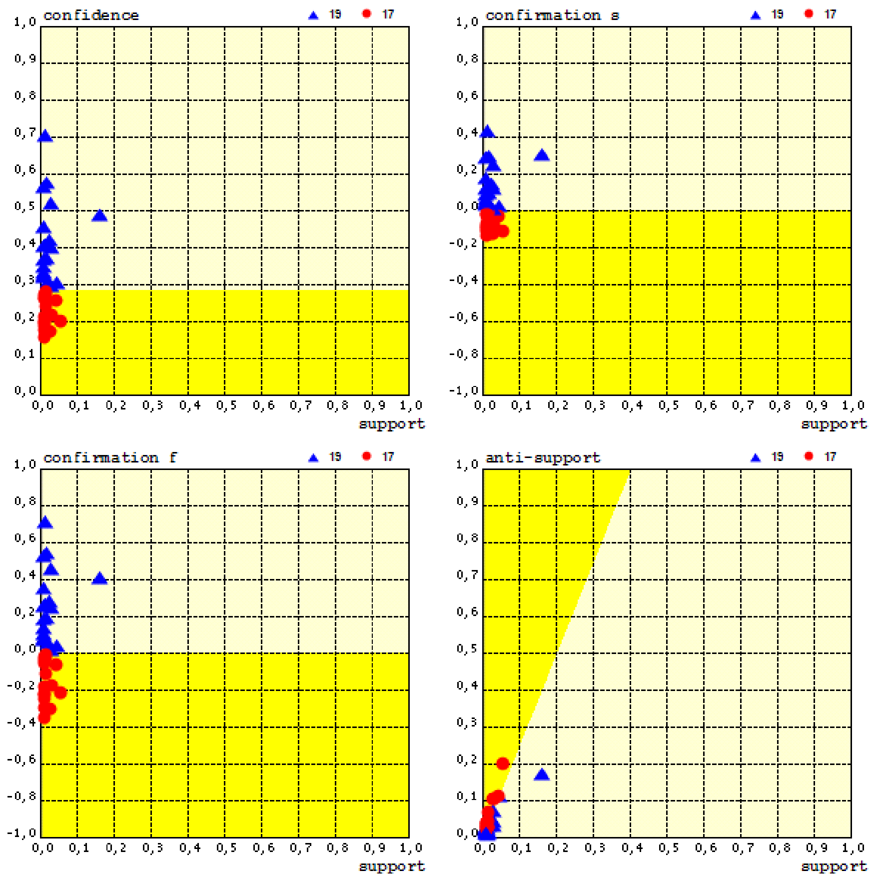


Fig. 27. Multichart presenting all rules with the conclusion web_site="Internet Explorer" generated with the 0.005 minimal support threshold

Figure 27 presents one of the largest classes, which was obtained for the conclusion being web_site='Internet Explorer' (this class represents transactions in which the anonymous users have visited the Internet Explorer web site). The diagram clearly shows that there are hardly any rules with rule support over 0.1 value. An insight into the points on the chart reveals that there is actually only one rule characterized by the rule support greater than 0.1. The investigated conclusion has quite a big percentage of rules with non-positive confirmation value. There are over 47% of rules that should be discarded. The Pareto-optimal borders, however, are completely free of such uninteresting rules, as it can be observed on Figure 28. The enclosure relationships of those non-dominated sets can also be analyzed using that multichart.

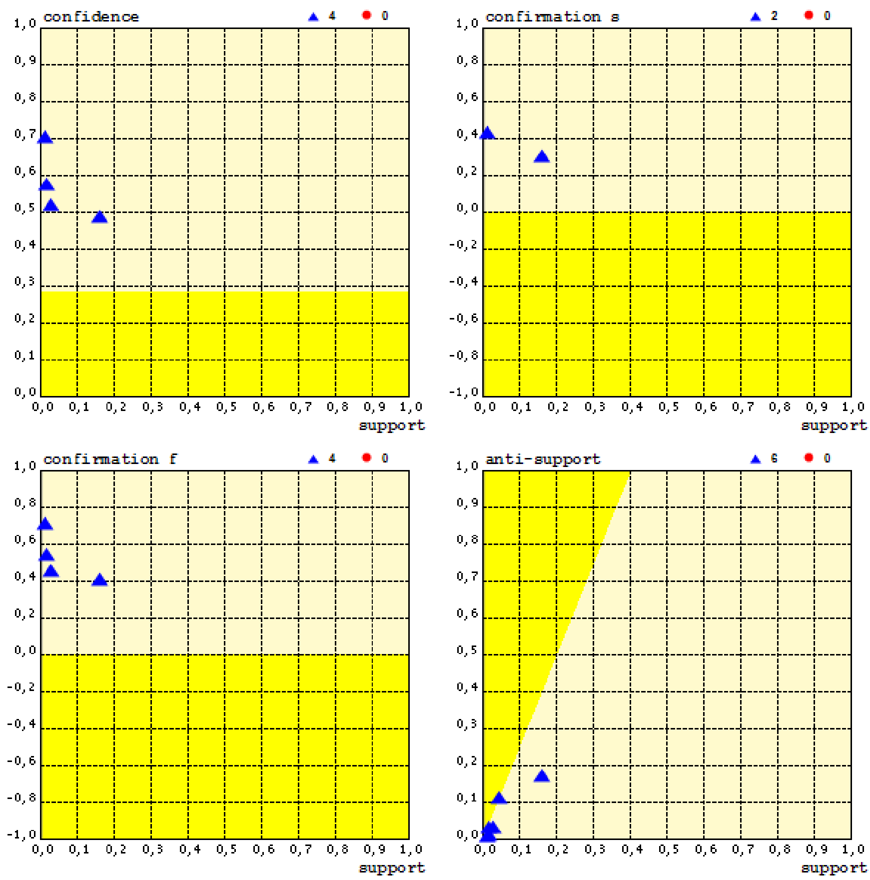


Fig. 28. Multichart presenting only the Pareto-optimal rules with the conclusion web_site="Internet Explorer" generated with the 0.005 minimal support threshold

HSV dataset. The *HSV* dataset [67] contains information about 122 patients with duodenal ulcer who were treated by highly selective vagotomy (HSV) in one of Poznań's hospitals in 1980's. Each object (patient) is described by 12 attributes: the first 11 attributes concern anamnesis and preoperative gastric secretion examined with the histaminic test of Kay, and the last attribute is a decision attribute defining classification of patients according to long term results of operation evaluated by a surgeon in the modified Visick grading. The eleven condition attributes contain such information about each patient:

– sex,
– age,
– duration of the disease,
– complications of ulcer,
– HCl concentration,
– basic volume of gastric juice per hour,
– volume of residual gastric juice,
– basic acid output (BOA),
– HCl concentration under histamine,
– volume of gastric juice per hour under histamine,
– maximal acid output (MAO).

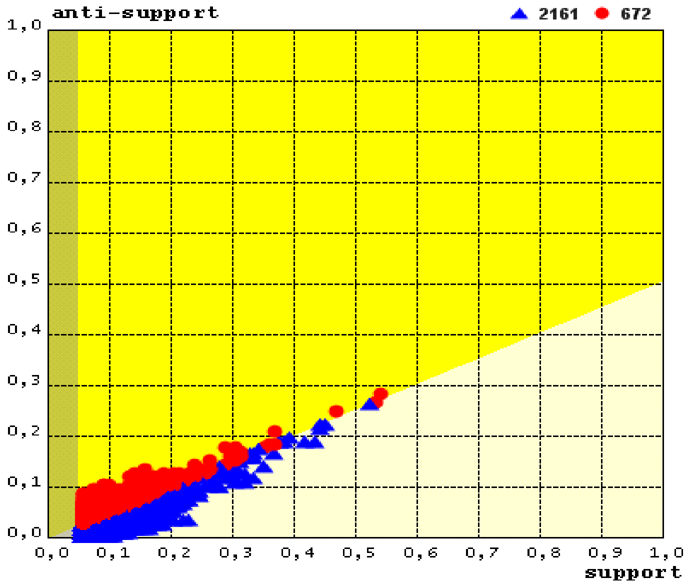
The twelfth attribute defines a long-term result of HSV, evaluated by a surgeon in the modified Visick grading. It obtains the following values:

1. excellent,
2. very good,
3. satisfactory,
4. unsatisfactory.

Several experiments concerning the HSV dataset have been carried out and the most interesting results for the conclusion: Visick_grading = "excellent" are presented below.

Originally, the minimal rule support threshold was set to 0.05, and the maximal acceptable rule anti-support was not set at all, which means that almost all possible (for that class) rules were generated. Our experiments aimed at showing how to determine an area of rules that are interesting from the point of view of support, anti-support, confirmation and property M.

Figure 29 presents decision rules generated for the class: Visick_grading = “excellent”. The first diagram contains all generated rules. It is clear that the number of rules is overwhelming and no surgeon could make any use of it. Thus, it is necessary to limit the set of rules by introducing more severe support and anti-support thresholds and by choosing only the rules that are characterized by positive values of confirmation measures. The determined area of interesting measures is marked on the second diagram of Figure 29. These are rules with support not lower than 0.3, anti-support not higher than 0.2, and a positive value of any confirmation measure. Let us stress, that this area is not limited only to the Pareto-optimal rules, because from the point of view of good representation of concept Visick_grading = “excellent” the dominated rules may also be useful and interesting for the user. The limited area contains 22 rules and includes rules that are optimal with respect to measures with the property M. A conducted analysis has shown that these rules cover almost 70% of all objects for which Visick_grading had the value “excellent”.



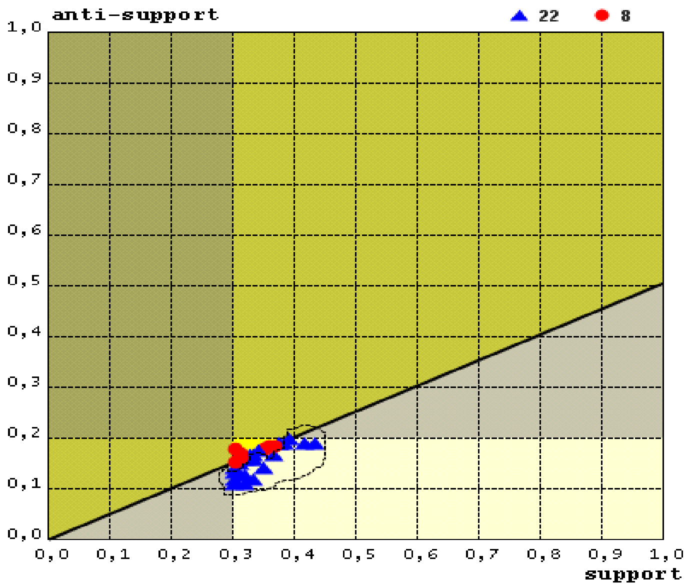


Fig. 29. Rules with positive (triangles) and non–positive (circles) confirmation measure value in a support–anti-support space. First diagram shows all generated rules, second one shows only a selection of interesting rules (conclusion: Visick_grading='excellent')

7 Conclusions

The work has been aimed at proposing a multicriteria rule evaluation space whose set of non-dominated rules contains all rules optimal with respect to any attractiveness measure having the property M, and at analyzing the relationships occurring between popular attractiveness measures (rule support, rule anti-support, confidence, rules interest function, gain, dependency factor and measures f and s).

To complete the main goal, the following specific tasks and problems have been solved:

1. An analysis of measures of rule support, rule anti-support, confidence, rule interest function, gain, dependency factor and measures f and s with respect to the property M, the property of confirmation and the property of hypothesis symmetry has been performed. It has been proved that in the set of the considered measures only the dependency factor does not possess the property M (Theorem 3.1 - Theorem 3.4). Moreover, we have analytically verified that rule support, anti-support and confidence cannot be considered as confirmation measures, while the gain measure is a confirmation measure only under certain conditions (Theorem 3.5 - Theorem 3.7). In case of the property of hypothesis symmetry, we have proved that, in addition to measures f and s which were found to be characterized by this property earlier by Eells et al. in [17], only rule interest function and gain possess it (Theorem 3.8 - Theorem 3.10). These results allow to group the attractiveness measures according to their properties and support the user in choosing a measure appropriate for his expectations.
2. An analysis of relationships between the considered attractiveness measures and analysis of the enclosure relationships between the sets of non-dominated rules in the evaluation spaces formed by different combinations of the concerned measures has been conducted. The analysis has been performed for a set of rules with the same conclusion. As a result, for a fixed value of rule support, we have proved existence of a monotonic link between confidence and measures f and s (Theorem 4.2, Theorem 4.5), and we have also generalized it to the whole class of measures with the property M (Theorem 4.9). Assuming a fixed value of confidence, we have also formulated the conditions under which there exists a monotonic relationship between rule support and measure f, s (Theorem 4.1, Theorem 4.4) and any measure with the property M (Theorem 4.8). Moreover, we have proved that any measure with the property M is monotone (anti-monotone) in rule support (rule anti-support) when the value of rule anti-support (rule support) is held fixed (Theorem 5.1, Theorem 5.2). The analysis of enclosure relationships between the sets of non-dominated rules in the evaluation spaces formed by different combinations of the concerned measures has revealed that the support–confidence and support–f Pareto-optimal sets of rules contain exactly the same rules. Moreover, we have proved that the set of the non-dominated rules with respect to support and confidence contains the support–s Pareto-optimal set,

while the support–anti-support Pareto-optimal set is the upper set of the support–confidence Pareto-optimal set (Theorem 4.6, Theorem 4.7, Theorem 5.3).
3. A proposition of a multicriteria evaluation space such that its set of the non-dominated rules contains all rules optimal with respect to any attractiveness measure that has the property M (Section 5). It is a two dimensional space formed by the measures of rule support and anti-support. The monotonic or anti-monotonic relationship, described above, between any attractiveness measure with the property M on one hand, and anti-support and support on the other hand, guarantees that rules optimal with respect to any measure possessing the property M will surely be found among the rules forming the support–anti-support Pareto-optimal border.
4. The support–confidence and support–anti-support evaluation spaces has also been enriched by the valuable semantics of confirmation measures. It has been proved that linear functions allow to distinguish the invaluable rules with non-positive or very small confirmation values in these evaluation spaces (Theorem 4.3, Theorem 5.4). This result helps to limit the set of induced rules only to those with an appropriate confirmation.
5. A multicriteria rule evaluation system has been designed and developed (Section 6). It is based on an apriori-like framework adjusted for generation of rules with respect to attractiveness measures possessing valuable properties. As the application of the system three datasets, *census*, *msweb* and *hsv*, have been analysed and discussed.

Completion of the tasks listed above pointed out also new interesting directions for future research. The most important and promising ones include:

1. verification of the above mentioned properties for other attractiveness measures,
2. development of algorithm for finding in the support–anti-support space a set of rules (both dominated and non-dominated) that covers the dataset in a certain percentage.

Acknowledgements The author would like to express her deep gratitude to her supervisor, Prof. Roman Słowiński for his motivating guidance, incalculable help and friendly attitude. She also wishes to thank Prof. Salvatore Greco for his valuable support, inspiring ideas and fruitful cooperation. The author gratefully acknowlegdes anonymous referees who provided her with really valuable and constructive remarks which were very helpful when preparing the final version of this work.

References

1. Azevedo, P.J., Jorge, A.M., 2007. Comparing rule Measures for Predictive Association Rules, in: *Proc of ECML Warsaw 2007*, LNAI 4701, Springer-Verlag Berlin Heidelberg, pp.510-517.

2. Agrawal, R., Imielinski, T., Swami, A.., 1993. Mining associations between sets of items in massive databases, in: *Proc. of the 1993 ACM-SIGMOD Int'l conf. on Management of Data*, 207-216.
3. Bayardo, R.J., Agrawal, R., 1999. Mining the most interesting rules, in: Proc. of the Fifth ACM-SIGKDD Int'l conf. on Knowledge Discovery and Data Mining, 145-154.Fayyad U.M., Piatetsky-Shapiro G., Smyth P., Uthurusamy R. (red.), *Advances in Knowledge Discovery and Data Mining*, AAAI/MIT Press, Cambridge Mass. 1996.
4. Bramer, M., 2007. *Principles of Data Mining*, Springer-Verlag, New York Inc.
5. Briand, L., El Emam, K., Morasca, S., 1995. On the Application of Measurement Theory in Software Engineering, in: *Empirical Software Engineering*, 1, 61-88.
6. Brin, S., Motwani, R., Ullman, J. and Tsur, S., 1997. Dynamic itemset counting and implication rules for market basket data, in: *Proc. of the 1997 ACM-SIGMOD Int'l conf. on the Management of Data*, 255-264.
7. Brzezinska, I., Slowinski, R., 2005. Monotonicity of a Bayesian confirmation measure in rule support and confidence, in: *Recent Developments in Artificial Intelligence Methods*, AI-METH Series, Gliwice, 39-42.
8. Brzezinska, I., Greco, S., Slowinski, R., 2005. Investigation of monotone link between confirmation measures and rule support and confidence. *Research Report* RA-025/05, Institute of Computing Science, Poznan University of Technology, Poznan.
9. 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) pp.587-600.
10. Carnap, R., 1962. *Logical Foundations of Probability*, second ed, University of Chicago Press, Chicago.
11. Christensen, D., 1999. Measuring confirmation. *Journal of Philosophy* XCVI, 437-461.
12. Clark, P., Boswell, P., 1991. Rule induction with CN2: some recent improvements, in: Machine Learning: *Proc. of the Fifth European Conference*, 151-163.
13. Cichosz, P., 2000. *Systemy uczące się*, Warszawa, WNT.
14. Cios, K., Pedrycz, W., Świniarski, R., 1999. *Data mining methods for knowledge discovery*, Kluwer Academic Publishers.
15. Crupi, V., Tentori, K., Gonzalez, M., (to appear). On Bayesian Theories of Evidential Support: Theoretical and empirical issues. *Philosophy of Science*.
16. Dhar, V., Tuzhilin, A., 1993. Abstract-driven pattern discovery in databases. *IEEE Transactions on Knowledge and Data Engineering*, 5(6).
17. Earman, J., 1992. Bayes or Bust: *A Critical Examination of Bayesian Confirmation Theory*. MIT Press, Cambridge, MA.
18. Eells, E., 1982. *Rational Decision and Causality*. Cambridge University Press, Cambridge.
19. Eells, E., Fitelson, B., 2002. Symmetries and asymmetries in evidential support. *Philosophical Studies*, 107 (2): 129-142.
20. Fayyad U.M., Piatetsky-Shapiro G., Smyth P., Uthurusamy R. (Eds.), 1996. *Advances in Knowledge Discovery and Data Mining*, AAAI/MIT Press, Cambridge Mass.
21. Fayyad U.M., Piatetsky-Shapiro G., Smyth P., Uthurusamy R., 1999. From data mining to knowledge discovery, in [20], pp.1-36.
22. Fitelson, B., 2001. *Studies in Bayesian Confirmation Theory*. Ph.D. Thesis, University of Wisconsin, Madison.

23. Francisci, D., Collard, M., 2003. Multi-criteria evaluation of interesting dependencies according to a data mining approach. in: *Congress on Evolutionary Computation, Canberra*, Australia, IEEE Computer Society Press, Los Alamitos, 12 pp.1568-1574.
24. Frawley, W.J., Piatetsky-Shapiro, G., Matheus, C.J., 1991. Knowledge discovery in databases: an overview, in: *Knowledge Discovery in Databases*, Eds. Piatetsky-Shapiro G., Frawley W.J., AAAI/MIT Press, Cambridge, MA, 1-27.
25. Fukuda, T., Morimoto, Y., Morishita, S., Tokuyama, T., 1996. Data mining using two-dimensional optimized association rules: scheme, algorithms, and visualization, in: *Proc. of the 1996 ACM-SIGMOD Int'l conf. on the Management of Data*, 13-23.
26. Geng, L., Hamilton, H.J, 2006. Interestingness measures for data mining: A survey, *ACM Computing Surveys*, 38(3).
27. Good, I.J., 1984. The best explicatum for weight of evidence. *Journal of Statistical Computation and Simulation*, 19: 294-299.
28. 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.
29. Greco, S., Słowiński, R., Szczęch, I., 2007. Analysis of monotonicity properties of some rule interestingness measures. in: Materiały II Krajowej Konferencji nt. *Technologie Przetwarzania Danych.* Poznań, 24-26.09.2007, Wydawnictwo Politechniki Poznańskiej, Poznań, 151-161.
30. Greco, S., Matarazzo, E., Słowiński, R., 2008. Parameterized rough set model using rough membership and Bayesian confirmation measures. *International Journal of Approximate Reasoning*, Vol 49(2) pp. 285-300.
31. Guillet, F., Hamilton, H.J., Eds., 2006. *Quality Measures in Data Mining.* Studies in Computational Intelligence, Volume 43, Springer.
32. Han, J., Pei, J., Yin, Y., 2000. Mining frequent patterns without candidate generation. in: *Proc. ACM SIGMOD Conference on Management of Data*, 1-12.
33. Han, J., Kamber, M., 2000. *Data mining: Concepts and techniques,* San Francisco, Morgan Kaufman Pub.
34. Hebert, C., Cremilleux, B., A Unified View of Objective Interestingness Measures, in: *Proc. of the 2007 MLDM Int'l conf. on Machine Learning and Data Mining*, LNAI 4571, Springer-Verlag pp. 533-547.
35. Heckerman, D, 1988. An axiomatic framework for belief updates, in: Kanal, L., Lemmer, J. (Eds), *Unconfidence in Artificial Intelligence 2.* Elsevier Science Publishers, New York, 11-22.
36. Hempel, C.G., 1945. *Studies in the logic of confirmation* (I). Mind 54, 1-26.
37. Hilderman, R. and Hamilton, H.: *Knowledge Discovery and Measures of Interest.* Kluwer Academic Publishers (2001).
38. Horwich, P., 1982. *Probability and Evidence.* Cambridge University Press, Cambridge.
39. Horvitz, E., Heckerman, D., 1986. The inconsistent use of certainty measures in artificial intelligence research, in: Kanal L., Lemmer J. (Eds), *Uncertainty in Artificial Intelligence* (1). Elsevier Science Publishers, New York, pp. 137-151.
40. International Business Machines, 1996. *IBM Intelligent Miner User's Guide*, Version 1, Release 1.
41. Jeffrey, H., 1935. Some tests of significance treated by theory of probability, in: *Proceedings of the Cambridge Philosophical Society*, 31: 203-222.
42. Joyce, J., 1999. *The Foundations of Causal Decision Theory.* Cambridge University Press, Cambridge.

43. Kamber, M., Shingal, R., 1996. Evaluating the interestingness of characteristic rules, in: *Proccedings of the 2nd International Conference on Knowledge Discovery and Data Mining* (KDD-96), Portland, Oregon, pp. 263–266.
44. Keynes, J., 1921. *A Treatise on Probability.* Macmillan, London.
45. Klemettinen, M., Mannila, H., Ronkainen, P., Toivonen, H., Verkamo, A.I., 1994. Finding interesting rules from large sets of discovered association rules, in: *Proc. of Third Int'l conf. on Information and Knowledge Management* (CIKM'94). ACM Press, pp. 401-407.
46. Kemeny, J., Oppenheim, P., 1952. Degrees of factual support. *Philosophy of Science* 19, 307-324.
47. Kohavi, R., 1996. Scaling up the accuracy of Naive–Bayes Classifiers: a decision–tree hybrid. [In]: *Proc. of the 2nd Int'l conf. on Knowledge Discovery and Data Mining.*
48. Kubat, M., Bratko, I., Michalski, R.S., Review of machine learning methods in [54] pp. 3-70.
49. Kyburg, H., 1983. Recent work in inductive logic, in: *Recent Work In Philosophy.* Lanham: Rowman &Allanheld, pp. 87–150.
50. Lenca, Ph., Meyer, P., Vaillant, B., Lallich, S., (to appear in Jan. 2008). On selecting interestingness measures for association rules: User oriented description and multiple criteria decision aid. *European Journal of Operational Research*, Vol. 184, Issue 2, 610-626.
51. Łukasiewicz, J., 1913. Die logischen Grundlagen der Wahrscheinlichkeitsrechnung, in: L. Borkowski (Ed.), *Jan Łukasiewicz –Selected Works.* North Holland Publishing Company, Polish Scientific Publishers,Amsterdam, London, Warsaw, 1970, pp. 16–63.
52. Mackie, J.L., 1969. The relevance citerion of confirmation. *The British Journal for the Philosophy of Science* 20, 27-40.
53. McGarry, K., 2005. A2005. A survey of interestingness measures for knowledge discovery, in: *The Knowledge Engineering Review*, vol. 20:1, Cambridge University Press, pp.39-61.
54. Michalski, R.S., 1983. A theory and methodology of inductive learning, in: *Machine Learning*: An Artificial Intelligence Approach, Eds: Michalski R.S., Carbonell J.G., Mitchell T.M., Morgan Kaufman, 83-134.
55. Michalski, R.S., 1997. Machine Learning, Data Mining and Knowledge Discovery. Principles and Applications, in: *Tutorials of Intelligent Information Systems*, IIS'97, Zakopane, IPI PAN Press.
56. Michalski, R.S., Bratko, I., Kubat, M (Eds), 1998. *Machine learning and data mining*, John Wiley & Sons.
57. Michell, J., 1986. Measurement scales and statistics: a clash of paradigms, in: *Psychological Bulletin*, 3, 398-407.
58. Morimoto, Y., Fukuda, T., Matsuzawa, H., Tokuyama, T. and Yoda, K., 1998. Algorithms for mining association rules for binary segmentation of huge categorical databases, in: *Proc. of the 24th Very Large Data Bases conf.*, 380-391.
59. Morishita, S., 1998. On classification and regression, in: *Proc. of the First Int'l conf. on Discovery Science* – Lecture Notes in Artificial Intelligence, 1532: 40-57.
60. Morzy, T., 2004. *Odkrywanie asocjacji: algorytmy i struktury danych.* Ośrodek Wydawnictw Naukowych, Poznań.
61. Morzy, T., Zakrzewicz, M., 2003. *Data mining, in: Handbook on data management in information systems*, Eds. Błażewicz J., Kubiak W., Morzy T., Rusinkiewicz M.E., Springer-Verlag, 487-565.

62. Morzy, M., 2006. Eksploracja danych - przegląd dostępnych metod i dziedzin zastosowań, in: *VI edycja Hurtownia danych i business intelligence*, Centrum Promocji Informatyki, Warszawa, Poland.
63. Morzy, M., 2005. Oracle Data Mining - odkrywanie wiedzy w dużych wolumenach danych, in: XI Krajowa Konferencja PLOUG'2005 "*Systemy informatyczne. Projektowanie, implementowanie, eksploatowanie*", Zakopane, Poland, 18-21.
64. Msweb dataset: http://kdd.ics.uci.edu/databases/msweb/msweb.html.
65. Nicod, J., 1923. *Le probleme de la logique de l'induction.* Alcan, Paris.
66. Pawlak, Z., 1982. Rough Sets. *Int. Journal of Computer and Information Science*, 11, 341-356.
67. Pawlak, Z., Słowiński, K., Słowiński, R., 1986. Rough classification of patients after highly selected vagotomy for duodenal ulcer. *International J. Man-Machine Studies*, 24, pp. 413-433.
68. Pawlak, Z., 1991. *Rough Sets: Theoretical Aspects of Reasoning about Data.* Kluwer, Dordrecht
69. Pawlak, Z., 2004. Some Issues on Rough Sets. *Transactions on Rough Sets I,* LNCS 3100, 1-58.
70. Pawlak, Z., Skowron, A., 2007. Rudiments of rough sets. *Information Sciences* 177(1), pp. 3-27; Rough sets: Some extensions. *Information Sciences* 177(1), pp. 28-40; Rough sets and boolean reasoning. *Information Sciences* 177(1),pp. 41-73.
71. Pearl, J., 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufman, San Francisco.
72. Piatetsky-Shapiro, G., 1991. Discovery, analysis, and presentation of strong rules. Chapter 12, in: *Knowledge Discovery in Databases*, AAAI/MIT Press.
73. Pollard, S., 1999. Milne's measure of confirmation. *Analysis* 59, 335-337.
74. Popper, K.R., 1959. *The Logic of Scientific Discovery.* Hutchinson, London.
75. Quinlan J.R., 1993. *C4.5: Programs for Machine Learning*, San Francisco, Morgan Kaufmann.
76. Quinlan J.R., 1986. Induction of decision trees. *Machine Learning* 1(1), pp81-106.
77. Reiter, R., 1978. On closed world data bases, in: *Logic and Data Bases*, H. Gallaire and J. Minker (eds), Plenum., New York pp. 119-140.
78. Rosenkrantz, R., 1994. Bayesian confirmation: paradise regained. *The British Journal for the Philosophy of Science* 45, 467-476.
79. Schlesinger, G., 1995. Measuring degrees of confirmation. *Analysis* 55, 208-212.
80. Schum, D., 1994. *The Evidential Foundations of Probabilistic Reasoning.* Wiley, New York.
81. Silberschatz, A., Tuzhilin, A., 1996. What Makes Patterns Interesting in Knowledge Discovery Systems, in: *IEEE Transactions on Knowledge and Data Engineering*, vol. 8 no. 6, pp. 970-974.
82. Słowiński, R., 2000. Rozszerzenie teorii zbiorów przybliżonych na atrybuty ze skalą preferencji, w: Materiały konferencji *Informatyka teoretyczna: metody analizy informacji niekompletnej i rozproszonej*, Białystok, 114-128.
83. Słowiński, R. and Greco, S., 2005. Measuring attractiveness of rules from the viewpoint of knowledge representation, prediction and efficiency of intervention, in: Szczepaniak, P., Kacprzyk, J., Niewiadomski, A. (Eds.), *Advances in Web Intelligence.* LNAI, vol 3528, Springer-Verlag, Berlin, 1-12.
84. Słowiński, R., Brzezińska, I., Greco, S., 2006. Application of Bayesian Confirmation Measures for Mining Rules from Support–confidence Pareto-optimal set, in: *Proc. of the 8th International Conference on Artificial Intelligence and Soft Computing.* LNCS, Springer-Verlag, Berlin 1018-1026.

85. Słowiński, R., Szczęch, I., Urbanowicz, M., Greco, S., 2006. Experiments with induction of association rules with respect to support and anti-support. *Research Report* RA-018/06, Institute of Computing Science, Poznań University of Technology, Poznań.
86. Słowiński, R., Szczęch, I., Urbanowicz, M., Greco, S., 2007. Mining Association Rules with respect to Support and Anti-support - Experimental Results, in: Proc. of RSEISP2007, *TheRough Sets and Emerging Systems Paradigms 2007 International Conference*. LNAI, vol. 4585, Springer-Verlag, Berlin, 534-543.
87. Stefanowski, J., 2001. *Algorytmy indukcji reguł decyzyjnych w odkrywaniu wiedzy,* Rozprawa habilitacyjna, Politechnika Poznańska, Instytut Informatyki, Poznań.
88. Stevens, S.S., 1946. On the theory of scales of measurement, in: *Science*, 103, 677-680.
89. Szczepaniak, P.S., Segovia, J., Kacprzyk, J., Zadeh, L.A. (Eds.): *Intelligent Exploration of the Web*. Physica Verlag, c/o Springer-Verlag, Heidelberg, New York, 2003.
90. Szczepaniak, P.S., 2004. *Obliczenia inteligentne, szybkie przekształcenia i klasyfikatory.* Akademicka Oficyna Wydawnicza EXIT, Warszawa.
91. Tan, P.-N., Kumar, V., Srivastava, J., 2004. Selecting the right objective measure for association analysis. *Inf.Syst.*29(4), pp.293-313.
92. Tan, P.-N., Steinbach, M., Kumar, V., 2006. *Introduction to Data Mining.* Pearson Education, Inc
93. Webb, G.I., 1995. OPUS: An efficient admissible algorithm for unordered search. *Journal of Artificial Intelligence Research*, 3: 431-465.
94. Yao, Y.Y., Zhong, N., 1999. An analysis of quantitative measures associated with rules, in: *Proceedings of the Third Pacific-Asia Conference on Knowledge Discovery and Data Mining*, LNAI vol. 1574, Springer, pp. 479-488.
95. Zakrzewicz M., 1997. Data Mining i odkrywanie wiedzy w bazach danych, in: *Materiały konferencyjne III Konferencji Polskiej Grupy Użytkowników Systemu Oracle*, Zakopane.
96. Zakrzewicz M., 2004. *Optymalizacja wykonania zapytań eksploracyjnych w systemach baz danych,* Rozprawa habilitacyjna, Politechnika Poznańska, Instytut Informatyki, Poznań.
97. Zembrowicz, R., Żytkow, J., From contingency tables to various forms of knowledge in databases, in [20], pp. 329-352.