Poznań University of Technology

Experiments with induction of association rules with respect to support and anti–support

Roman Słowiński, Izabela Szczęch,
Mirosław Urbanowicz, Salvatore Greco

Research Report RA–018/06

Institute of Computing Science

Piotrowo 2, 60–965 Poznań, Poland

Experiments with Induction of Association Rules with respect to Support and Anti–support

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

[1] Institute of Computing Science, Poznań University of Technology, 60–965 Poznań, Poland
{Roman.Slowinski, Izabela.Szczech }@cs.put.poznan.pl

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

[3] Faculty of Economics, University of Catania, Corso Italia, 55, 95129 Catania, Italy
salgreco@mbox.unicit.it

Abstract. Evaluating the interestingness of decision rules or trees is a challenging problem of knowledge discovery and data mining. In recent studies, the use of two interestingness measures at the same time was prevailing. Mining of Pareto–optimal borders according to support and confidence, or support and anti–support are examples of that approach. In this paper, we consider induction of "*if…, then…*" association rules with a fixed conclusion. We investigate ways to limit the set of association rules non–dominated with respect to support and confidence on one hand and support and anti–support on the other hand, to a subset of truly interesting rules. Analytically, and through experiments on real life dataset, we show that both of the considered sets can be easily reduced by using the valuable semantic of confirmation measures.

Keywords: Association rules, Induction, Support, Anti–support, Confirmation, Confidence, Pareto–optimal border.

1. Introduction

In data mining and knowledge discovery, the discovered knowledge patterns are often expressed in a form of "*if…, then…*" rules. They are consequence relations representing correlation, association, causation etc. between independent and dependent attributes. When mining massive datasets, the number of the discovered patterns can easily exceed the capabilities of a human user to identify useful and interesting results. In order to increase the relevance and utility of selected rules and, thus, also limit the size of the resulting rule set, quantitative measures, also known as attractiveness or interestingness measures (metrics), have been proposed and studied (e.g. confidence and support, gain [10], conviction [3], lift [14]). Among widely studied interestingness measures, there is, moreover, a group of Bayesian confirmation measures, which quantify the degree to which a piece of evidence built of the independent attributes provides "evidence for or against" or "support for or against" the hypothesis built of the dependent attributes [9]. All the proposed

measures have been introduced to capture different characteristics of rules.
One of the simplest and most intuitive ways to deal with the problem of too large number of generated rules, is to set an interestingness measure threshold. This way a decision maker filters out all the rules for which the calculated interestingness measure threshold was not exceeded. Nevertheless, the resulting set can still remain difficult to handle and analyze due to its large size.

Another approach to evaluation of generated rules concentrates on the use of two different interestingness measures. In such a two dimensional perspective, the only objective information one can get about the quality of rules is the dominance relation in the set of rules. Non–dominated rules, also called Pareto–optimal, create a border in the two dimensional evaluation space. Valuable features of Pareto–optimal borders created with respect to many different interestingness measures have been widely studied in the literature. In particular, Bayardo and Agrawal [2] have proved that for a class of rules with fixed conclusion, the support–confidence Pareto–optimal border (i.e. the set of non–dominated rules with respect to both rule support and confidence) includes optimal rules according to several different interestingness measures, such as gain [10], conviction [3], an unnamed measure proposed by Piatetsky–Shapiro [16], etc. This practically useful result allows to identify the most interesting rules according to several interestingness measures by solving an optimized rule mining problem with respect to rule support and confidence only.
Among other important two–dimensional presentations of rules there is one relying on the rule support and rule anti–support measures. It has been shown in [5] that for a class of rules with fixed conclusion, all rules with optimal values of any confirmation measure with some desirable properties, can be found on the support–anti–support Pareto–optimal border. It is a very general and valuable result. Moreover, it has been proved in [5] that the support–anti–support Pareto–optimal border includes (i.e. is the superset) the whole support–confidence Pareto–optimal border. Thus, the support–anti–support Pareto optimal border expands the valuable features of support–confidence Pareto–optimal border. The support–confidence Pareto–optimal border presents a smaller number of rules (more precisely, a not greater number of rules) than the support–anti–support Pareto–optimal border. However, it does not present all the rules maximizing a confirmation measure satisfying some desirable properties.
Let us stress, nevertheless, that dominated rules are not without interest. It can be due to the fact that when inducing rules from data we are interested in a set of rules that characterize a given concept (conclusion), rather than in one rule being the best with respect to one or two interestingness measures. Thus, from the viewpoint of a good representation of a concept, the dominated rules may be found even more valuable than some non–dominated ones. Of course, the set of considered rules can de delimited by the thresholds set for the two interestingness measures, but still the resulting set of both dominated and non–dominated rules can exceed the capabilities to analyze it. Thus, a further reduction is desirable.

In this paper, we show a way to limit the set of rules generated with respect to pairs of measures: support–confidence and support–anti–support, by filtering out the rules for which the premise does not confirm the conclusion. This proposition is based on imposing the confirmation perspective on the analyzed two–dimensional evaluations.

The paper is organized as follows. In the next section, there are preliminaries on decision rules and their quantitative description. In section 3, we investigate the idea and the advantages of mining only rules with positive confirmation from Pareto–optimal border with respect to support and confidence. Section 4 concentrates on the proposal of limiting the set of rules generated with respect to support and anti–support. Theoretical considerations are supported by experimental results. The paper ends with conclusions.

2. Preliminaries

Since discovering rules from data is the domain of inductive reasoning, its starting point is a sample of larger reality often given in a form of a data table. Formally, a *data table* is a pair $S = (U, A)$, where U is a nonempty finite set of objects called *universe*, and A is a nonempty finite set of *attributes* such that $a: U \rightarrow V_a$ for every $a \in A$. The set V_a is a domain of a. Let us associate a formal language L of logical formulas with every subset of attributes. Formulas for a subset $B \subseteq A$ are built up from attribute–value pairs (a,v), where $a \in B$ and $v \in V_a$, using logical connectives $\neg$ (not), $\wedge$ (and), $\vee$ (or). A *rule* induced from S and expressed in L is denoted by $\phi \rightarrow \psi$ (read as "*if* ϕ, *then* ψ"). It consists of antecedent ϕ and consequent ψ, being formulas expressed in L, called *premise* and *conclusion*, respectively, and therefore it can be seen as a consequence relation (see critical discussion about interpretation of rules as logical implications in [12]) between premise and conclusion. The rules mined from data may be either *decision* rules or *association* rules, depending on whether the division of A into condition and decision attributes has been fixed or not.
In this paper, similarly to [2], we only consider rules with the same conclusion, which can be induced from a dataset.

2.1. Complete Preorder on a Set of Rules in terms of an Interestingness Measure

Let us denote by q any interestingness measure that quantifies the interestingness of a rule induced from an information table S. Application of q to a set of induced rules creates a complete preorder, denoted as $\underline{\mathbf{p}}_q$, 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 interestingness measure q 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 q if $r_1 \mathbf{f}_q r_2$ and, moreover, rule r_1 is *indifferent* to rule r_2 if $r_1 \sim_q r_2$.

2.2. Partial Preorder on Rules in terms of Two Interestingness Measures

Let us denote by $\underline{\mathbf{p}}_{qt}$ a partial preorder given by a dominance relation on a set X of rules in terms of any two different interestingness measures q and t, i.e. for all $r_1,r_2 \in X$ $r_1 \underline{\mathbf{p}}_{qt}\, r_2$ if $r_1 \underline{\mathbf{p}}_q\, r_2$ and $r_1 \underline{\mathbf{p}}_t\, r_2$. Recall that 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 complete 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). The partial preorder $\underline{\mathbf{p}}_{qt}$ can be decomposed into its asymmetric part $\mathbf{p}_{qt}$ and its symmetric part $\sim_{qt}$ in the following manner:
given a set of rules X and two rules $r_1,r_2 \in X$, $r_1\ \mathbf{p}_{qt}\ r_2$ if and only if

$$\begin{aligned} &q(r_1) \le q(r_2) \wedge t(r_1) < t(r_2), \text{ or} \\ &q(r_1) < q(r_2) \wedge t(r_1) \le t(r_2), \end{aligned} \tag{1}$$

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

$$q(r_1) = q(r_2) \wedge t(r_1) = t(r_2). \tag{2}$$

If for a rule $r \in X$ there does not exist any rule $r' \in X$, such that $r\ \mathbf{p}_{qt}\ r'$ then r is said to be non–dominated (i.e. Pareto–optimal) with respect to interestingness measures q and t. A set of all non–dominated rules with respect to q and t is also referred to as an *q–t Pareto–optimal border*.

2.3. Monotonicity of a Function in its Argument

Let x be an element of a set of rules X and let $g(x)$ be a real function associated with this set, such that $g{:}X \rightarrow \mathbf{R}$. Assuming an ordering relation $\mathbf{f}$ in X, function g is said to be monotone (resp. anti–monotone) in x, if for any $x,y \in X$, relation $x\ \mathbf{f}\ y$ implies that $g(x) \ge g(y)$ (resp. $g(x) \le g(y)$).

2.4. Support, Confidence and Anti–support Measures of Rules

Among measures very commonly associated with rules induced from information table S, there are *support* and *confidence*. The *support* of condition ϕ, denoted as $sup(\phi)$, is equal to the number of objects in U having property ϕ. The support of rule $\phi \rightarrow \psi$, denoted as $sup(\phi \rightarrow \psi)$, is equal to the number of objects in U having both property ϕ and ψ; for those objects, both premise ϕ and conclusion ψ evaluate to true.

The *confidence* of a rule (also called *certainty*), denoted as *conf*($\phi\rightarrow\psi$), is defined as follows:

$$conf(\phi \rightarrow \psi) = \frac{sup(\phi \rightarrow \psi)}{sup(\phi)}, \quad sup(\phi)>0 \, . \tag{3}$$

Note, that it can be regarded as a conditional probability $\Pr(\psi|\phi)$ with which conclusion ψ evaluates to true, given that premise ϕ evaluates to true, however, expressed in terms of frequencies.

Anti–support of a rule, denoted as *anti–sup*($\phi\rightarrow\psi$), is equal to the number of objects in U having the property ϕ but not having the property ψ. Thus, anti–support is the number of counter–examples i.e. objects for which the premise ϕ evaluates to true but which fall into a class different than ψ. Note, that anti–support can also be regarded as *sup*($\phi\rightarrow\neg\psi$).

2.5. Bayesian Confirmation Measures

Bayesian confirmation measures constitute a group of interestingness measures that quantify the degree to which a premise ϕ provides "support for or against" a conclusion ψ [8, 9]. In this context, a confirmation measure denoted by $c(\phi \rightarrow \psi)$ is required to satisfy the following definition:

$$c(\phi \rightarrow \psi) = \begin{cases} > 0 \;\; if \;\; \Pr(\psi \mid \phi) > \Pr(\psi), \\ = 0 \;\; if \;\; \Pr(\psi \mid \phi) = \Pr(\psi), \\ < 0 \;\; if \;\; \Pr(\psi \mid \phi) < \Pr(\psi). \end{cases} \tag{4}$$

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

$$c(\phi \rightarrow \psi) = \begin{cases} > 0 \;\; if \;\; conf(\phi \rightarrow \psi) > sup(\psi)/|U|, \\ = 0 \;\; if \;\; conf(\phi \rightarrow \psi) = sup(\psi)/|U|, \\ < 0 \;\; if \;\; conf(\phi \rightarrow \psi) < sup(\psi)/|U|. \end{cases} \tag{5}$$

For the confirmation measures a desired property of monotonicity (M) was proposed in [12]. This monotonicity property says that, given an information system S, a confirmation measure is a function non–decreasing with respect to *sup*($\phi\rightarrow\psi$) and

$sup(\neg\phi\rightarrow\neg\psi)$, and non–increasing with respect to $sup(\neg\phi\rightarrow\psi)$ and $sup(\phi\rightarrow\neg\psi)$. The property of monotonicity (M) of $c(\phi\rightarrow\psi)$ with respect to $sup(\phi\rightarrow\psi)$ (or, analogously, with respect to $sup(\neg\phi\rightarrow\neg\psi)$) means that any evidence in which ϕ and ψ (or, analogously, neither ϕ nor ψ) hold together increases (or at least does not decrease) the credibility of the rule $\phi\rightarrow\psi$. On the other hand, the property of monotonicity of $c(\phi\rightarrow\psi)$ with respect to $sup(\neg\phi\rightarrow\psi)$ (or, analogously, with respect to $sup(\phi\rightarrow\neg\psi)$) means that any evidence in which ϕ does not hold and ψ holds (or, analogously, ϕ holds and ψ does not hold) decreases (or at least does not increase) the credibility of the rule $\phi\rightarrow\psi$. The arguments for monotonicity property (M) given in [12] are the following. Given a probability *Pr*, an evidence ϕ confirms a hypothesis ψ, if $\Pr(\psi|\phi)>\Pr(\psi|\neg\phi)$. Expressing the probability in terms of confidence, one can say that an evidence ϕ confirms a hypothesis ψ, if $conf(\psi|\phi)>conf(\psi|\neg\phi)$. Greco, Pawlak and Slowinski [12] proved that it is possible to pass from one situation in which evidence ϕ does not confirm a hypothesis ψ, i.e. $conf(\psi|\phi)<conf(\psi|\neg\phi)$, to a situation in which evidence ϕ confirms a hypothesis ψ, i.e. $conf(\psi|\phi)>conf(\psi|\neg\phi)$, when $sup(\phi\rightarrow\psi)$ or $sup(\neg\phi\rightarrow\neg\psi)$ increases or $sup(\neg\phi\rightarrow\psi)$ or $sup(\phi\rightarrow\neg\psi)$ decreases. Thus, it is reasonable to expect that a confirmation measure $c(\phi\rightarrow\psi)$ is monotone with respect to $sup(\phi\rightarrow\psi)$ and $sup(\neg\phi\rightarrow\neg\psi)$and anti-monotone with respect to $sup(\phi\rightarrow\neg\psi)$ and $sup(\neg\phi\rightarrow\psi)$.

Among confirmation measures that have property (M) there are e.g. confirmation measure f [9] and confirmation measure s [6], defined as:

$$f(\phi\rightarrow\psi)=\frac{\Pr(\phi\mid\psi)-\Pr(\phi\mid\neg\psi)}{\Pr(\phi\mid\psi)+\Pr(\phi\mid\neg\psi)}\ , \tag{6}$$

$$s(\phi\rightarrow\psi)=\Pr(\psi\mid\phi)-\Pr(\psi\mid\neg\phi)\ . \tag{7}$$

2.6. Brief Description of a Dataset on which Experiments Were Conducted

For the purpose of running experiments we have used a dataset called *adult*, created by Becker and Kohavi [15] from a *census* dataset. The number of instances that were analyzed reached 32 561. They were described by 9 nominal attributes with different sizes of value sets. During the experiments, missing values occurring in the dataset were substituted by the most frequently appearing value. The experiments were conducted in order to illustrate the theoretical results surveyed in this paper and to show their application on a real–life dataset.

3. Support–Confidence Pareto–optimal Border

Bayardo and Agrawal in [2] have proposed evaluation of the set of rules induced from a dataset in terms of two popular interestingness measures being rule support and confidence. They have proved that for a class of rules with fixed conclusion, the support–confidence Pareto–optimal border includes optimal rules according to several different interestingness measures, such as gain [10], Laplace [7], lift [14], conviction [3], and unnamed measure proposed by Piatetsky–Shapiro [16]. Thus, by solving an optimized rule mining problem with respect to rule support and confidence one can identify a set of rules containing most interesting (optimal) rules according to several interestingness measures.

Despite those valuable features of the support–confidence Pareto–optimal border, one cannot, in general, claim that the set of dominated rules is without interest. An expert analyzing the set of induced rules can be interested in some dominated rules even more than in non–dominated rules, e.g. due to the fact that in order to cover the analyzed concept one has to use both dominated and non–dominated rules. Of course, a user can set some thresholds both to rule support and confidence, but still taking under the consideration both dominated and non–dominated rules can result in a large, difficult to analyze set of rules.

Hence, we pose a question whether there does not exist any way to limit the set of the analyzed rules. Are all the rules really worth analyzing? We will answer that question in the following paragraphs.

3.1. The Confirmation Perspective on the Support–Confidence Evaluations

The semantic utility of confidence in comparison with confirmation measures in general has been widely studied in [4, 5, 12]. It has been clearly shown that the utility of the scale of confirmation measures outranks the utility of confidence's scale. Confidence can obtain values between 0 and 1 (where 1 is regarded as the best) whereas confirmation measures take values between –1 and 1 (again 1 being the most desirable). The confidence measure, thus, has no means to show that the rule is useless when its premise disconfirms the conclusion. Such situation is expressed by a negative value of any confirmation measure. Thus, the rules for which the confirmation measures take negative values or, more generally, values below a non-negative significance threshold, can be filtered out. Due to those important semantic features of the family of confirmation measures, we find it valuable to impose the confirmation perspective on the analyzed support–confidence evaluations and limit in this way the set of rules to be analyzed.

It has been analytically proved in [5] that for a fixed value of rule support, confidence is monotone with respect to any confirmation measure $c(\phi \rightarrow \psi)$ having the desired property of monotonicity (M) proposed in [12].

Let us observe that according to definition (5) of a confirmation measure $c(\phi \rightarrow \psi)$, we have:

$$c(\phi \rightarrow \psi) > 0 \Leftrightarrow conf(\phi \rightarrow \psi) > \frac{sup(\psi)}{|U|} \quad . \tag{8}$$

Since, we limit our consideration to rules with the same conclusion, then $|U|$ and $sup(\psi)$ should be regarded as constant values. Thus, (8) shows that rules laying under a constant, expressing what percentage of the whole dataset is taken by the considered class ψ, are characterized by negative values of confirmation. For those rules ψ is satisfied less frequently when ϕ is satisfied rather than generically. Fig. 1 illustrates this point.

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

Theorem 1.

$$f(\phi \rightarrow \psi) \geq k \Leftrightarrow conf(\phi \rightarrow \psi) \geq \frac{sup(\psi)(k+1)}{|U| - k(|U| - 2sup(\psi))} \quad . \tag{9}$$

Proof. For the simplicity of presentation, let us use the following notation:

$a = sup(\phi \rightarrow \psi)$,

$b = sup(\neg\phi \rightarrow \psi)$,

$c = sup(\phi \rightarrow \neg\psi)$,

$d = sup(\neg\phi \rightarrow \neg\psi)$.

The analysis concerns only a set of rules with the same conclusion, thus the values of $|U| = a + b + c + d$ and $sup(\psi) = a + b$ are constant.

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

$a = sup(\phi \rightarrow \psi)$,

$b = sup(\psi) - sup(\phi \rightarrow \psi)$,

$c = \dfrac{1}{conf(\phi \rightarrow \psi)} sup(\phi \rightarrow \psi) - sup(\phi \rightarrow \psi)$,

$d = |U| - sup(\psi) - \dfrac{1}{conf(\phi \rightarrow \psi)} sup(\phi \rightarrow \psi) + sup(\phi \rightarrow \psi)$.

Thus, for given U and ψ, confirmation measure $f(\phi\rightarrow\psi)$ can be written in terms of confidence and support of rule $\phi\rightarrow\psi$ (effectively in terms of confidence only) as follows:

$$f(\phi \to \psi) = \frac{|U| conf(\phi \to \psi) - sup(\psi)}{(|U| - 2sup(\psi)) conf(\phi \to \psi) + sup(\psi)} \ . \tag{10}$$

Considering inequality $f(\phi \to \psi) \geq k$ from (10) we obtain the thesis of the theorem.

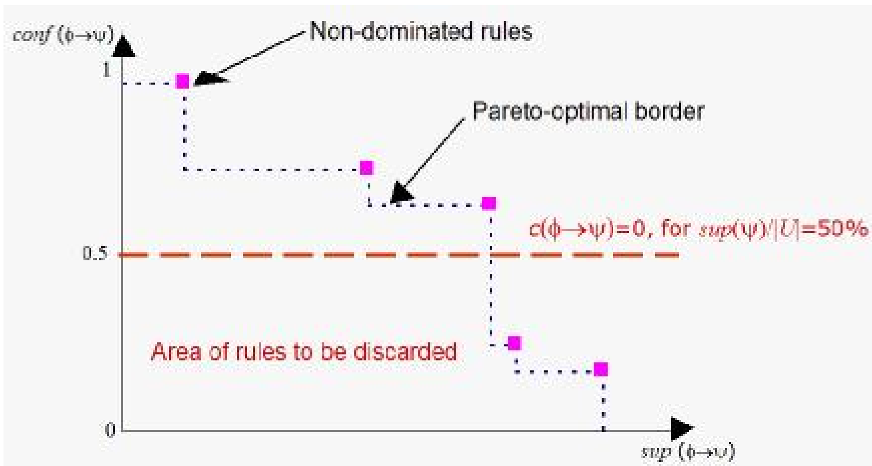

Fig. 1 An example of a constant line representing confirmation measure $c(\phi \to \psi)=0$ in a support–confidence space. Rules laying under this constant line should be discarded from further analysis.

3.2. Experiments with Rule Induction with respect to Support and Confidence

In this experiment, association rules were induced from an "*adult*" dataset. The induction proceeds in a two step Apriori-like framework:

- firstly, all conjunctions of elementary conditions (i.e. itemsets) that exceeded the minimum rule support threshold (i.e. frequent itemsets) were found;
- secondly, those frequent itemsets were used to generate association rules having confidence measure not smaller than the user's defined confidence threshold.

The detailed description as well as the efficiency comparison of the applied algorithms (based on [1, 13]) can be found in [17]. Throughout the experiment the values of support was expressed as a relative value between 0 and 1 (where 1 means that all objects from the dataset support the particular rule). Support thresholds was introduced. During the frequent itemset generation phase, only itemsets that occurred in more than 15% of objects (i.e. exceeded 0.15 support threshold) were approved. No confidence thresholds were applied.

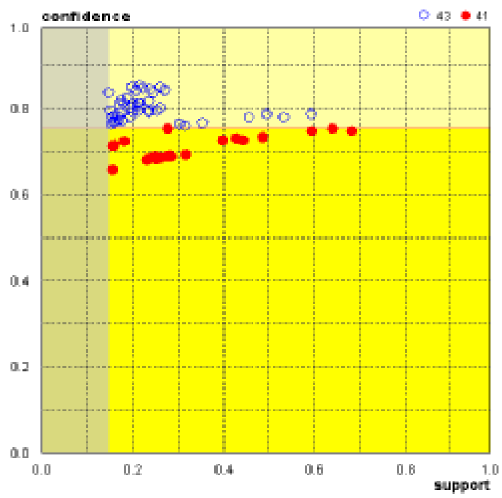

Fig. 2 Rules generated for a conclusion workclass='Private' with positive (empty circles) and non–positive confirmation measure value (solid circles) in a support–confidence space.

On Fig. 2 we have presented all the association rules generated, according to mentioned thresholds from the conclusion: workclass='Private'. This class contains information about people working in a private sector. Rules are presented in a support–confidence space. Rules for which the value of a confirmation measure is positive are marked by empty circles, whereas rules with non–positive confirmation measure are presented as solid circles. This experiment makes it evident that in practice even rules with high value of confidence (exceeding even 0.7) can be found useless as their premise disconfirms the conclusion (those rules are marked by solid circles). It is therefore clear, that the semantic scale of the confidence measure is not enough and that confirmation measures are very much needed. Only, the information brought by the sign of the confirmation measures can point out which rules are not valuable. As shown on Fig. 3, sometimes even rules from the Pareto–optimal border need to be discarded from further analysis as their value of confirmation is non–positive.

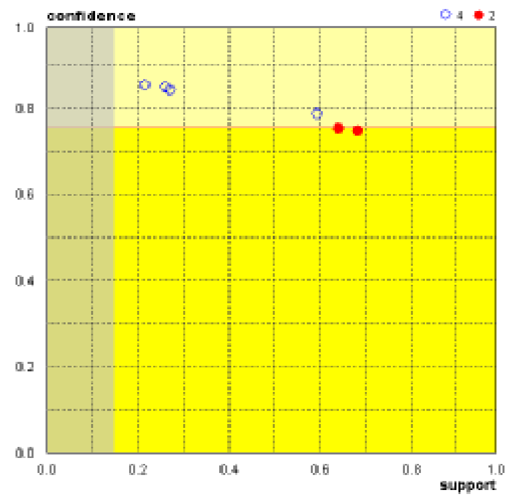

Fig. 3 Support–confidence Pareto–optimal border formed by rules induced for the conclusion: workclass='Private'. Rules with positive confirmation are marked by empty circles and rules with non-positive confirmation are marked by solid circles.

On Fig. 2 and Fig. 3 a constant line was placed separating the rules with positive confirmation (situated above the line) from those with non–positive confirmation (situated below the line). These figures visualize result (8) and say how big (in comparison to the whole dataset) is the considered class of rules for the analyzed conclusion workclass='Private'. This particular class is quite large, therefore the boundary line is quite high. For relatively smaller classes it was situated lower (see Fig. 4 for a class that constitutes less than 70% of the whole dataset).

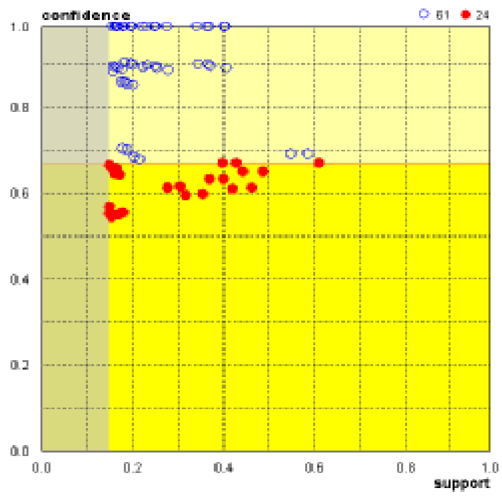

Fig. 4 Rules generated for a conclusion sex='Male' with positive (empty circles) and non–positive confirmation measure value (solid circles) in a support–confidence space.

By imposing the confirmation perspective, the number of rules to be analyzed by the domain expert can be significantly reduced. For the conclusion being workclass='Private', 41 out of 84 rules had to be discarded for disconfirming the conclusion. Thus, the set of potentially interesting and valuable rules was reduced by almost 50%! Table 1 shows results for other conclusions that we have considered.

Table 1. Information about the percentage of rules with non-positive confirmation in the set of all generated rules for different conclusions.

Considered conclusion y	**No. of all rules**	**No. of rules with non–positive confirmation**	**Reduction percentage**
workclass='Private'	84	41	49%
sex=Male	85	24	28%
race=White	108	29	27%
income<=50000USD	87	43	49%

Table 2 shows how many rules with non–positive confirmation laid on the support–confidence Pareto–optimal border for different considered conclusions. The result shows, that even Pareto–optimal borders, i.e. objectively the best sets of rules, contain rules that are misleading and should be discarded. In some cases, the support–confidence Pareto–optimal border could be reduced by even 33%, like for the first considered conclusion being workclass='Private'.

Table 2. Information about the percentage of rules with non-positive confirmation laying on the support–confidence Pareto–optimal border for different conclusions.

Considered conclusion y	No. of all rules on the Pareto border	No. of rules with non–positive confirmation	Reduction percentage
workclass='Private'	6	2	33%
sex=Male	6	1	17%
race=White	12	0	0%
income<=50000USD	5	1	20%

4. Support–Anti–support Pareto–optimal Border

Presentation of association rules in dimensions of rule support and rule anti–support was proposed in [5]. The idea of combining those two dimensions came from a critical remark towards support–confidence Pareto–optimal border. In [5], it was proved that a rule maximizing a confirmation measure satisfying the property (M) is on the support–confidence Pareto–optimal border only if a specific condition is satisfied. This means that, in general, not all rules maximizing a confirmation measure satisfying the property (M) are on the support–confidence Pareto–optimal border. However, on the basis of the observation that a confirmation measure is more meaningful than confidence, mining all rules that maximize confirmation measures satisfying the property (M), without taking into account the rule confidence, became an interesting problem. The solution to it is support–anti–support Pareto–optimal border. It was proved in [5] that any confirmation measure, denoted by *c*, satisfying the property of monotonicity (M) is monotone (non–decreasing) with respect to rule support *sup*($\phi\rightarrow\psi$) and anti–monotone (non–increasing) with respect to rule anti–support *anti–sup*($\phi\rightarrow\psi$). Therefore, the best rule according to any of these monotone confirmation measures 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.

Moreover, it was pointed out in [5] that the Pareto–optimal border of support–anti–support contains the support–confidence Pareto–optimal border. Thus, the support–confidence Pareto–optimal border presents a not greater number of rules than the support–anti–support Pareto–optimal border. However, it does not present all the rules maximizing a confirmation measure satisfying the property (M). In fact, all the rules constituting the difference between those two Pareto–optimal sets maximize some confirmation measure which is not monotone with respect to support because it does not satisfy the conditions mentioned in [5].

Despite all good characteristics of the support–anti–support Pareto–optimal border, one can still remain interested in the set of dominated rules. It can, for example, be due to the fact that the dominated rules are necessary in order to cover the analyzed concept, i.e. all instances having property ψ. Thus, analyzing whether one can also impose a confirmation perspective to the support–anti–support evaluations, and this way limit the area of the analyzed rules, by discarding those rules which do not confirm their conclusions, became an interesting task.

4.1. The Confirmation Perspective on the Support–Anti–support Evaluations

Without doubt, the semantic utility of the scale of confirmation measure outranks the utility of anti–support's scale. Anti–support, denoted as *anti–sup*(ϕ→ψ), is a measure showing cardinality of the set of rule counter–examples.

It has been analytically proved in [5] that for a fixed value of rule support, any confirmation measure $c(\phi \rightarrow \psi)$ having the desired property of monotonicity (M) is anti–monotone (i.e. non–decreasing) with respect to anti–support.

Let us observe that a simple transformation of definition (5) of a confirmation measure $c(\phi \rightarrow \psi)$ expressed in terms of confidence leads to the following result:

$$c(\phi \rightarrow \psi) \geq 0 \Leftrightarrow anti\text{-}sup(\phi \rightarrow \psi) \leq sup(\phi \rightarrow \psi)\left[\frac{|U|}{sup(\psi)} - 1\right] . \tag{11}$$

It is also interesting to investigate a more general condition $c(\phi \rightarrow \psi) \geq k$, $k \geq 0$, for some specific confirmation measures. In the following, we consider again the confirmation measure $f(\phi \rightarrow \psi)$.
Theorem 2.

$$f(\phi \rightarrow \psi) \geq k \Leftrightarrow anti\text{-}sup(\phi \rightarrow \psi) \leq sup(\phi \rightarrow \psi)(U - sup(\psi))\frac{1-k}{(1+k)sup(\psi)} . \tag{12}$$

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

$$f(\phi \rightarrow \psi) = \frac{\dfrac{sup(\phi \rightarrow \psi)}{sup(\psi)} - \dfrac{anti - sup(\phi \rightarrow \psi)}{|U| - sup(\psi)}}{\dfrac{sup(\phi \rightarrow \psi)}{sup(\psi)} + \dfrac{anti - sup(\phi \rightarrow \psi)}{|U| - sup(\psi)}} . \tag{13}$$

Considering the inequality $f(\phi \rightarrow \psi) \geq k$ from (13), we obtain the thesis.

Having limited our consideration to rules with the same conclusion, $|U|$ and $sup(\psi)$ should be regarded as constant values. Thus, the result (11) shows that a simple linear function bounds rules that are characterized by positive values of confirmation from those with non–positive confirmation values.

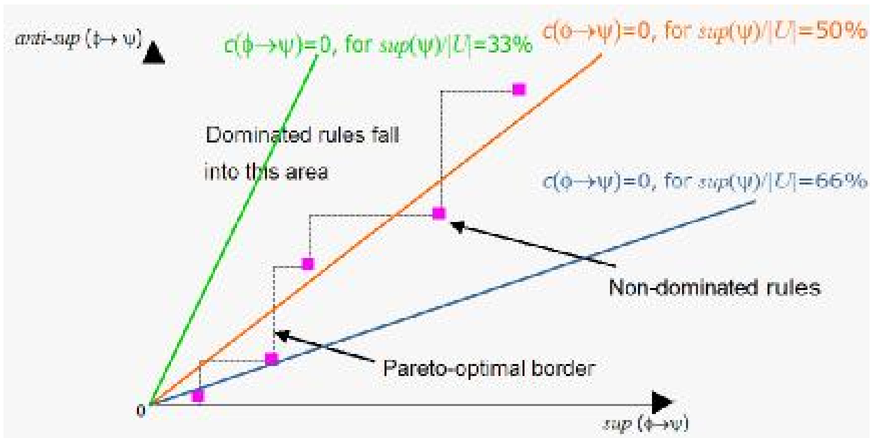

Fig. 5 Three examples of linear functions representing confirmation measure $c(\phi \rightarrow \psi)=0$ in a support–anti–support space. Each line was drawn according to a set of rules for conclusions different in cardinality. Rules laying above these functions should be discarded from further analysis.

4.2. Experiments with Rule Induction with respect to Support and Anti–support

In this experiment, association rules were again induced from an "***adult***" dataset. The induction proceed in a two step framework:

- firstly, all conjunctions of elementary conditions (i.e. itemsets) that exceeded the minimum rule support threshold (i.e. frequent itemsets) were found;
- secondly, those frequent itemsets were used to generate association rules having anti–support measure not greater than the user's defined confidence threshold.

The detailed description of the applied algorithms (inspired by [1, 13]) can be found in [17]. Throughout the experiment the value of support was expressed as a relative value between 0 and 1 (where 1 means that all objects from the dataset support the particular rule). During the frequent itemset generation phase, only itemsets that exceeded 0.15 support threshold were approved. No threshold for anti–support was introduced.

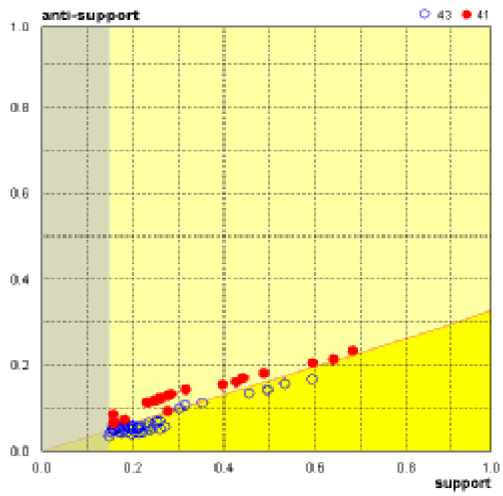

Fig. 6 Rules generated for a conclusion workclass='Private' with positive (empty circles) and non–positive (solid circles) confirmation measure value in a support–anti–support space.

On Fig. 6, we have presented all the association rules generated, according to mentioned threshold, for the conclusion: workclass='Private'. Rules are presented in a support–anti–support space. Rules for which the value of a confirmation measure is positive are marked by empty circles, whereas rules with non–positive confirmation measure are solid circles. This experiment makes it clear, that 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 confirmation measure with the property of monotonicity (M), it is necessary to take under consideration also the information brought by the sign of the confirmation measures. In the set of both dominated and non–dominated rules, there can be examples of rules with negative values of confirmation (see Fig. 6). Fig. 7 presents just the rules which form the support–anti–support Pareto–optimal border. It can be observed there that 4 out of 18 rules need to be discarded from further analysis as their value of confirmation is non–positive.

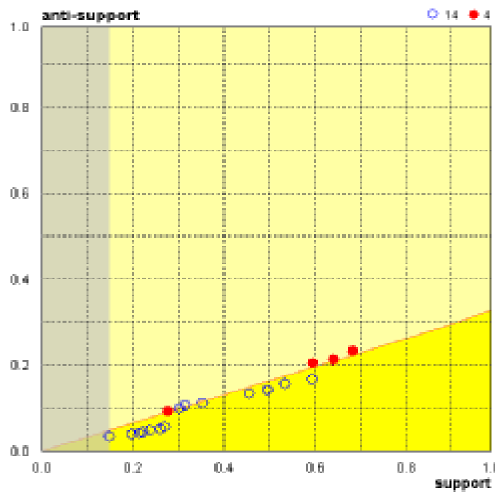

Fig. 7 Support–anti–support Pareto–optimal border formed by rules induced for the conclusion: workclass='Private'. Rules with positive confirmation are marked by empty circles and rules with non-positive confirmation are marked by solid circles.

On Fig. 6 and Fig. 7, 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). These figures visualize result (11). Table 3 presents the percentage of rules that should be discarded from the Pareto–optimal border with respect to support and anti–support, for different conclusions. 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 3, on the support–anti–support Pareto–optimal border there also came up other rules with non-positive confirmation values. In the conducted experiment the set of rules (from the support–anti–support Pareto–optimal border) to be analyzed could be reduced by e.g. about 22%, as it happened for the class workclass='Private'.

Table 3. Information about the percentage of rules with non-positive confirmation laying on the support–anti–support Pareto–optimal border for different conclusions.

Considered conclusion y	**No. of all rules on the Pareto border**	**No. of rules with non–positive confirmation**	**Reduction percentage**
workclass='Private'	18	4	22%
sex=Male	8	3	38%
race=White	19	1	5%
income<=50000USD	15	4	27%

5. Conclusions

Evaluation of induced rules with respect to two interestingness measures at the same time is an approach to measuring the relevance and utility of the induced rules. In this paper, we investigated rules induced for a fixed conclusion and evaluated in spaces of support–confidence and support–anti–support. The Pareto–optimal borders of those two–dimensional spaces are characterized by some valuable features. The support–confidence Pareto–optimal border contains rules optimal with respect to many other popular interestingness measures [2]. The support–anti–support Pareto–optimal border was introduced in [5] for its valuable property of containing all the rules that maximize any confirmation measure with the desired property of monotonicity (M).

However, these worthy features, do not assure that the number of induced rules would not exceed the human user capabilities to analyze them all. Inspired by the strength of the semantic scale of the family of confirmation measures, we show that it is reasonable to limit the set of rules to be presented to the domain expert, by eliminating those that are characterized by non–positive values of confirmation. We have shown analytically that a simple constant line imposed on the support–confidence space bounds the rules with positive values of confirmation measure from those with non–positive confirmation values. This is a very practical result as it allows to discard from further analysis all the rules laying below that constant line and this way limit the set of analyzed rules only to those with positive confirmation values, without actually calculating the value of a particular confirmation measure for each of the induced rules. Moreover, we have presented results from experiments on a large dataset called "adult". They show how greatly a set of induced rules can be reduced by throwing away the rules with non–positive values of confirmation.

Analogous analysis has been conducted for rules in support–anti–support space. We have shown that a simple linear function separates the rules with positive and non–positive values of confirmation. Again, this is an easy approach to limit the set of analyzed rules without calculating a value of a particular confirmation measure for each of the induced rules. Experimental results show how big the reduction of the rule set could be. The percentage of discarded rules just for the analyzed Pareto–optimal borders has also been presented.

References

1. Agrawal, R., Imielinski, T. and Swami, A.: 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 (1993).
2. Bayardo, R.J., Agrawal, R.: Mining the most interesting rules. [In]: *Proc. of the Fifth ACM–SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining*, 145–154 (1999).
3. Brin, S., Motwani, R., Ullman, J. and Tsur, S.: 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 (1997).

4. Brzezińska, I., Słowiński, R.: Monotonicity of a Bayesian confirmation measure in rule support and certainty. [In]: T. Burczyński, W. Cholewa, W. Moczulski (Eds.) *Recent Developments in Artificial Intelligence Methods*, AI–METH Series, Gliwice, 39–42 (2005).
5. Brzezińska, I., Greco, S., Słowiński, R.: Mining Pareto–optimal rules with respect to support and anti–support. [In]: *Engineering Applications of Artificial Intelligence* (to appear).
6. Christensen, D.: Measuring confirmation. *Journal of Philosophy XCVI*, 437–461 (1999).
7. Clark, P. and Boswell, P.: Rule induction with CN2: Some recent improvements. [In]: *Machine Learning: Proc. of the Fifth European Conference*, 151–163 (1991).
8. Eells, E., Fitelson, B.: Symmetries and asymmetries in evidential support. *Philosophical Studies*, 107 (2): 129–142 (2002).
9. Fitelson, B.: *Studies in Bayesian Confirmation Theory*. Ph.D. Thesis, University of Wisconsin, Madison (2001).
10. Fukuda, T., Morimoto, Y., Morishita, S. and Tokuyama, T.: 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 (1996).
11. Good, I.J.: The best explicatum for weight of evidence. *Journal of Statistical Computation and Simulation*, 19: 294–299 (1984).
12. Greco, S., Pawlak, Z. and Słowiński, R.: Can Bayesian confirmation measures be useful for rough set decision rules? *Engineering Applications of Artificial Intelligence*, 17: 345–361 (2004).
13. Han, J., Pei, J., Yin, Y.: Mining frequent patterns without candidate generation. [In]: *Proc. ACM SIGMOD Conference on Management of Data*, 1-12 (2000)
14. International Business Machines: *IBM Intelligent Miner User's Guide*, Version 1, Release 1 (1996).
15. Kohavi, R.: Scaling up the accuracy of Naive–Bayes Classifiers: a decision–tree hybrid. [In]: *Proc. of the Second International Conference on Knowledge Discovery and Data Mining*, (1996).
16. Piatetsky–Shapiro, G.: Discovery, Analysis, and Presentation of Strong Rules. Chapter 12 [In]: *Knowledge Discovery in Databases*, AAAI/MIT Press (1991).
17. Urbanowicz, M.: *Induction of association rules with given support, confidence and confirmation.* M.Sc. Thesis, Poznań University of Technology, Poznań (2007).