Analysis of monotonicity properties of new normalized rule interestingness measures

Roman Słowiński

Poznań University of Technology,

Institute for Systems Research, PAS, Poland

Salvatore Greco

University of Catania, Italy

Izabela Szczęch

Poznań University of Technology, Poland

HCP 2008, Delft

Presentation plan

- Introduction

- Basic quantitative characteristics of rules
- Normalization of interestingness measures
- Desired properties of interestingness measures

- Results of the conducted analysis
- Practical application of the results

- Conclusions

Introduction - motivations

The number of rules
induced from datasets is usually quite large

- **overwhelming for human comprehension,**
- **many rules are irrelevant or obvious**
(low practical value)

rule evaluation – interestingness (attractiveness) measures
(e.g. support, confidence, gain)

- **each measure was proposed to capture different characteristics of rules**
- **the number of proposed measures is very large**
- **easiness of interpretation of NORMALIZED measures**

Introduction - motivations

The choice of interestingness measure for a certain application is a difficult problem

- **the users expectations vary,**
- **the number of proposed measures is overwhelming**

properties of interestingness measures, which reflect users′ expectations towards the behavior of measures in particular situations

need to analyze measures with respect to their properties

Introduction – rule induction

- **Patterns in form of rules are induced from a data table**
- ***S*=⟨*U*, *A*⟩ – *data table*, where *U* and *A* are finite, non-empty sets**
  U* – *universe*; *A* – set of *attributes
- ***S*=⟨*U*, *C*, *D*⟩ – *decision table*, where C – set of *condition* *attributes*,**
  ***D* – set of *decision* *attributes*, *C*∩*D*=∅**

- ***Decision rule* or *association rule* induced from *S***
  is a *consequence relation*: f ® y *read as* *if* f *then* y
  where f is condition (evidence or premise)
  and y is conclusion (hypothesis or decision)
  formula built from attribute-value pairs (*q*,*v*)
- **If the division into independent and dependent attributes is fixed, then rules are regarded as decision rules, otherwise as association rules.**

Introduction – rule induction

Characterization of nationalities

U	*Height*	*Hair*	*Eyes*	*Nationality*	Support
1	*tall*	*blond*	*blue*	*Swede*	270
2	*medium*	*dark*	*hazel*	*German*	90
3	*medium*	*blond*	*blue*	*Swede*	90
4	*tall*	*blond*	*blue*	*German*	360
5	*short*	*red*	*blue*	*German*	45
6	*medium*	*dark*	*hazel*	*Swede*	45

C ***D***

- E.g. decision rules induced from „characterization of nationalities":
  1) **If (*Height=tall*), then (*Nationality=Swede*)**
  2) **If (*Height=medium*) & (*Hair=dark*), then (*Nationality=German*)**

Introduction – interestingness measures

- To measure the relevance and utility of rules, quantitative measures called attractiveness or interestingness measures, have been proposed (e.g. support, confidence, lift, gain, Piatetsky-Shapiro,...)
- Unfortunately, there is no evidence which measure(s) is (are) the best

- **Notation:**
  - $sup(\mathbf{0})$ is the number of all objects from *U*, having property °

    e.g. $sup(\phi)$, $sup(\psi)$

Basic quantitative characteristics of rules

- **Support of rule $\phi \rightarrow \psi$ in *S*:**

$$sup(\phi \rightarrow \psi) = sup(\phi \wedge \psi)$$

- **Anti-support of rule $\phi \rightarrow \psi$ in *S*:**

$$anti\text{-}sup(\phi \rightarrow \psi) = sup(\phi \wedge \neg\psi)$$

Basic quantitative characteristics of rules

- ***Rule Interest Function*** **(Piatetsky-Shapiro 1991)**

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

- ***Gain measure*** **(Fukuda et al. 1996)**

$$gain(\phi \to \psi) = sup(\phi \to \psi) - \Theta sup(\phi)$$

where Θ is a fraction constant between 0 and 1

- ***Dependency Factor*** **(Popper 1959, Pawlak 2002)**

$$\eta(\phi \to \psi) = \frac{\frac{sup(\phi \to \psi)}{sup(\phi)} - \frac{sup(\psi)}{|U|}}{\frac{sup(\phi \to \psi)}{sup(\phi)} + \frac{sup(\psi)}{|U|}}$$

Normalization of interestingness measures

Normalization of interestingness measures

- **We propose to normalize measures, so that they would distinguish between two completely different situations:**
  - **situation α in which confirmation occurs:** $Pr(\psi|\phi) \geq Pr(\psi)$ **which, under „the closed world assumption", can be estimated as:**

$$\frac{sup(\phi \to \psi)}{sup(\phi)} \geq \frac{sup(\psi)}{|U|}$$

  - **situation β in which disconfirmation occurs:** $Pr(\psi|\phi) < Pr(\psi)$ **which, under „the closed world assumption", can be estimated as:**

$$\frac{sup(\phi \to \psi)}{sup(\phi)} < \frac{sup(\psi)}{|U|}$$

Normalization of interestingness measures

- Inspired by Crupi et al. 2007 we propose to normalize *RI*, *gain* and *dependency factor* by dividing them
  - by the maximum value they obtain in case of confirmation, and
  - by the absolute minimum value they obtain in case of disconfirmation.
- In this way, we will obtain interestingness measures taking values from the interval [–1, 1].

Normalization of interestingness measures

- There are many approaches to determining the maximum and minimum values (in case of confirmation and disconfirmation, respectively), which eventually lead to different normalizations.
- Inspired by the approach of (Nicod 1923):
  - we consider only cases in which the evidence is true,
  - while we ignore cases where there is no evidence.

- For example, in case of rule: *„if x is a raven, then x is black"*, the evidence is "*raven*" and the hypothesis is "*black*".

  In this situation, a *raven* which is *black* supports the rule,
  a *raven* which is *not black* is against this rule,
  and everything which is *not a raven* can be ignored.

Normalization of interestingness measures – „Nicod approach"

- E.g. consider rule $\phi \rightarrow \psi$: ***if x is a raven then x is black***

  a=$sup(\phi \rightarrow \psi)$ – the number of objects in ***U*** which are black ravens

  b=$sup(\neg\phi \rightarrow \psi)$ – the no. of objects in ***U*** which are black non-ravens

  c=$sup\ (\phi \rightarrow \neg\psi)$ – the no. of objects in ***U*** which are non-black ravens

  d=$sup(\neg\phi \rightarrow \neg\psi)$–the no. of objects in ***U*** which are non-black non-ravens

- The maximal support of the rule will be obtained when:
  - ***c*=:0** (i.e. there are no non-black ravens) and
  - ***a*** takes over all observations from ***c***

    (i.e. each raven is black, ***a*=:*a*+*c***)

Normalization of interestingness measures – „Nicod approach"

- **E.g. consider rule $\phi \rightarrow \psi$: *if x is a raven then x is black***

  a=$sup(\phi \rightarrow \psi)$ – **the number of objects in *U* which are black ravens**

  b=$sup(\neg\phi \rightarrow \psi)$ – **the no. of objects in *U* which are black non-ravens**

  c=$sup(\phi \rightarrow \neg\psi)$ – **the no. of objects in *U* which are non-black ravens**

  d=$sup(\neg\phi \rightarrow \neg\psi)$–**the no. of objects in *U* which are non-black non-ravens**

- **The minimal support of the rule will be obtained when:**
  - ***a*=:0 (i.e. there are no black ravens) and**
  - ***c* takes over all observations from *a***

    (i.e. each raven is not black, *c*=:*a*+*c*)

Alternative approach to normalization – „likelihoodist”

- „Likelihoodist” approach to normalization:
  - the evidence confirms the hypothesis, if the evidence is more frequent with the hypothesis rather than with ¬hypothesis, and
  - the evidence disconfirms the hypothesis, if the evidence is more frequent with ¬hypothesis rather than with the hypothesis.

Alternative approach to normalization – „likelihoodist"

- The maximal support of the rule will be obtained when:
  - a takes over all observations from c
    (i.e. each raven is black, $a=:a+c$),
  - $b=:0$ (i.e. there are no black non-ravens),
  - $c=:0$ (i.e. there are no non-black ravens), and
  - d takes over all observations from b
    (i.e. each non-raven is not black, $d=:b+d$).

Alternative approach to normalization – „likelihoodist"

- The minimal support of the rule will be obtained when:
  - $a =: 0$ (i.e. there are no black ravens),
  - b takes over all observations from d

    (i.e. each non-raven is black, $b =: b + d$),
  - c takes over all observations from a

    (i.e. each raven is not black, $c =: a + c$), and
  - $d =: 0$ (i.e. there are no non-black non-ravens).

Alternative approach to normalization – „Bayesian"

- „Bayesian" approach to normalization:
  - the evidence confirms the hypothesis, if the hypothesis is more frequent with the evidence rather than with ¬evidence, and
  - the evidence disconfirms the hypothesis, if ¬hypothesis is more frequent with the evidence rather than with ¬evidence.

Alternative approach to normalization – „Bayesian"

- The maximal support of the rule will be obtained when:
  - *a* takes over all observations from *b*
    (i.e. each black object is a raven, $a=:a+b$),
  - $b=:0$ (i.e. there are no black non-ravens),
  - $c=:0$ (i.e. there are no non-black ravens), and
  - *d* takes over all observations from *c*
    (i.e. each non-black object is not a raven, $d=:c+d$).

Alternative approach to normalization – „Bayesian"

- The minimal support of the rule will be obtained when:
  - $a=0$ (i.e. there no black ravens),
  - b takes over all observations from a

    (i.e. each non-black object is not a raven, $b=a+b$),
  - c takes over all observations from d

    (i.e. each non-black object is a raven, $c=c+d$), and
  - $d=0$ (i.e. there no non-black non-ravens).

Normalized confirmation measures

- **Notation:** ***a*=**$sup(\phi\rightarrow\psi)$ **,** ***b*=**$sup(\neg\phi\rightarrow\psi)$**,** ***c*=**$sup\ (\phi\rightarrow\neg\psi)$**,** ***d*=**$sup(\neg\phi\rightarrow\neg\psi)$
- **Among confirmation measures that were normalized and analysed by Crupi et al. there are:**

$$D(\phi\rightarrow\psi)=\frac{a}{a+c}-\frac{(a+b)}{|U|}$$ **(Carnap 1950/1962)**

$$S(\phi\rightarrow\psi)=\frac{a}{a+c}-\frac{b}{b+d}$$ **(Christensen 1999)**

$$M(\phi\rightarrow\psi)=\frac{a}{a+b}-(a+c)$$ **(Mortimer 1988)**

$$N(\phi\rightarrow\psi)=\frac{a}{a+b}-\frac{c}{c+d}$$ **(Nozick 1981)**

$$C(\phi\rightarrow\psi)=\frac{a-(a+c)(a+b)}{a+b+c+d}$$ **(Carnap 1950/1962)**

$$R(\phi\rightarrow\psi)=\frac{a(a+b+c+d)}{(a+c)(a+b)}-1$$ **(Finch 1960)**

Z-measure

- **It can be observed that:**

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

- **Crupi et al. have therefore proposed to call them all by one name:**

_Z_-measure.

$$Z(\phi \rightarrow \psi) = \begin{cases} \dfrac{\dfrac{a}{a+c} - \dfrac{(a+b)}{a+b+c+d}}{\dfrac{c+d}{a+b+c+d}} = \dfrac{ad-bc}{(a+c)(c+d)} \\ \text{in case of confirmation} \\ \\ \dfrac{\dfrac{a}{a+c} - \dfrac{(a+b)}{a+b+c+d}}{\dfrac{a+b}{a+b+c+d}} = \dfrac{ad-bc}{(a+c)(a+b)} \\ \text{in case of disconfirmation} \end{cases}$$

Normalized Rule Interest Function

- **The normalized *Rule Interest Function*:**

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

Normalized gain measure

- **The normalized *gain* measure:**

$$gain_{norm} = \begin{cases} \dfrac{a - \Theta(a+c)}{(a+c) - (1-\Theta)} & \text{in case of confirmation} \\[2ex] \dfrac{a - \Theta(a+c)}{(a+c)\Theta} & \text{in case of disconfirmation} \end{cases}$$

Normalized dependency factor

- **The normalized *dependency factor*:**

$$\eta_{norm} = \begin{cases} \dfrac{\dfrac{a}{a+c} - \dfrac{a+b}{a+b+c+d}}{\dfrac{a}{a+c} + \dfrac{a+b}{a+b+c+d}} \times \dfrac{1 + \dfrac{a+b+c}{a+b+c+d}}{1 - \dfrac{a+b+c}{a+b+c+d}} \\ \textit{in case of confirmation} \\ \\ \dfrac{\dfrac{a}{a+c} - \dfrac{a+b}{a+b+c+d}}{\dfrac{a}{a+c} + \dfrac{a+b}{a+b+c+d}} \\ \textit{in case of disconfirmation} \end{cases}$$

Desired properties
of interestingness measures

Property M

- **Property M of monotonicity** **(Greco, Pawlak, Słowiński 2004)**
- **An interestingness measure $F(a, b, c, d)$ has the property M if it is a function non-decreasing with respect to a and d and non-increasing with respect to b and c**

where:

$a = sup(\phi \to \psi)$ **- the no. of objects in U for which ϕ and ψ hold together**

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

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

$d = sup(\neg\phi \to \neg\psi)$

Interpretation of the property M

- **E.g. consider rule $\phi \rightarrow \psi$:**

  if x is a raven then x is black

  - **non-decreasing with respect to *a* –**

    the more *black ravens* we observe,

    the more credible becomes the rule

  - **non-increasing with respect to *b* –**

    the more *black non-ravens* we observe,

    the less credible becomes the rule

  - **non-increasing with respect to *c***

  - **non-decreasing with respect to *d***

Results of the conducted analysis with respect to property M

- ***Theorem:***

  The normalized Rule Interest Function has the property M

- ***Theorem:***

  The normalized Gain measure has the property M

- ***Theorem:***

  The normalized Dependency factor does not have the property M

Results of the conducted analysis

- ***Theorem:*** **The normalized Rule Interest Function has the property M**
- ***Proof outline:***
  prove that the normalized *RI* both in case of confirmation and disconfirmation is:
  - ***non-decreasing*** **with respect to *a*,**
  - ***non-increasing*** **with respect to *b*,**
  - ***non-increasing*** **with respect to *c*,**
  - ***non-decreasing*** **with respect to *d***

Practical application of the results

Support - Anti-support Pareto border

- ***Theorem:***

  For a set of rules with the same conclusion,

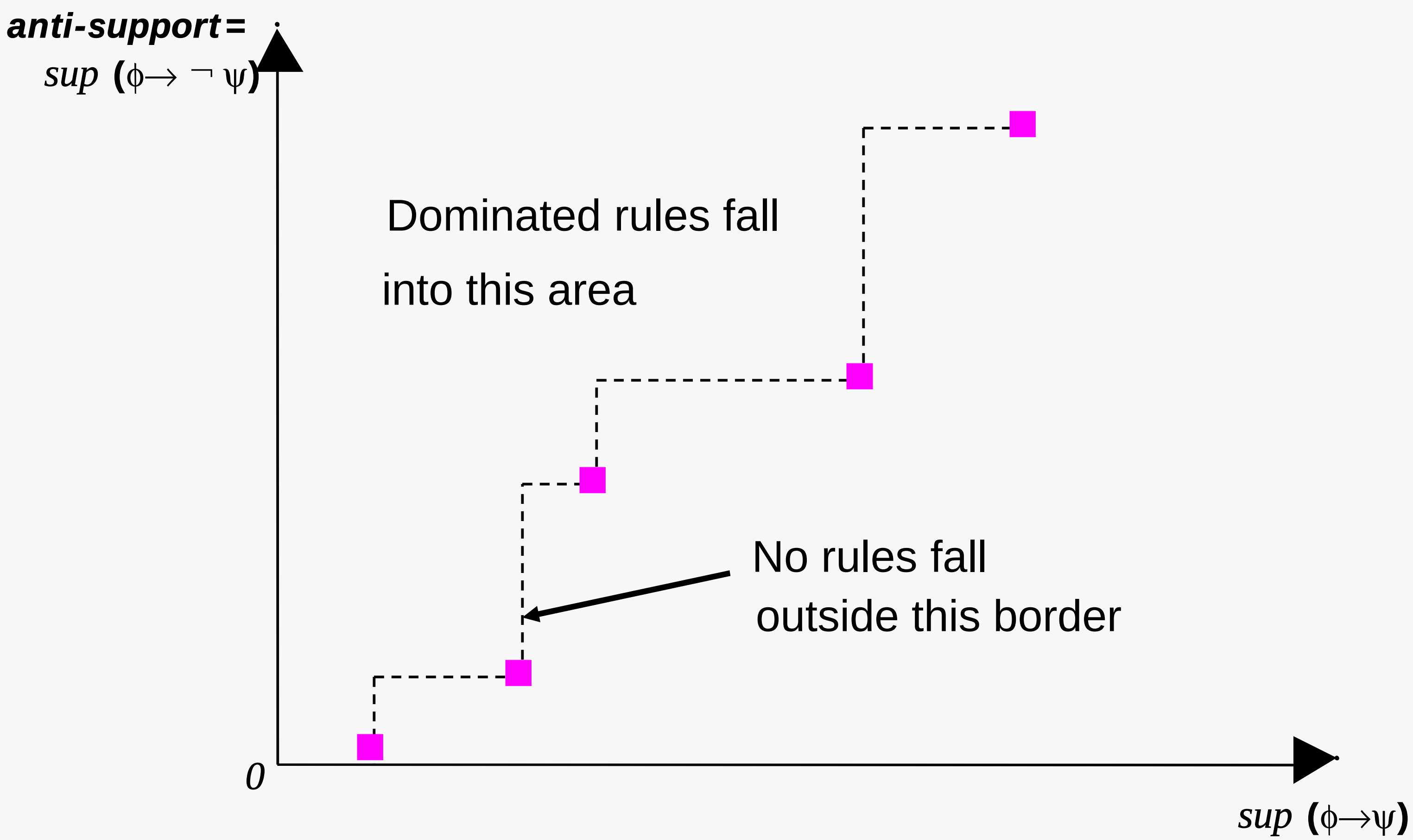
  due to (anti) monotonic dependencies between
  measures of support and anti-support on one hand
  and any interestingness measure with property M on the other hand

  the best rules according to any measure with the property M
  must reside on the support - anti-support Pareto optimal border

- **The support – anti-support Pareto border is a set of non-dominated rules with respect to those measures,**
  i.e. the set of rules for which there is no other rule
  with greater support and smaller anti-support

Brzezińska I., Greco S., Słowiński R.: "Mining Pareto-Optimal Rules with Respect to Support and Confirmation or Support and Anti-Support", Engineering Applications of Artificial Intelligence, vol. 20 no. 5 (2007) pp.587-600

Support - Anti-support Pareto border



The best rules according to any measure with the property M must reside on the support - anti-support Pareto border

Practical application of the results

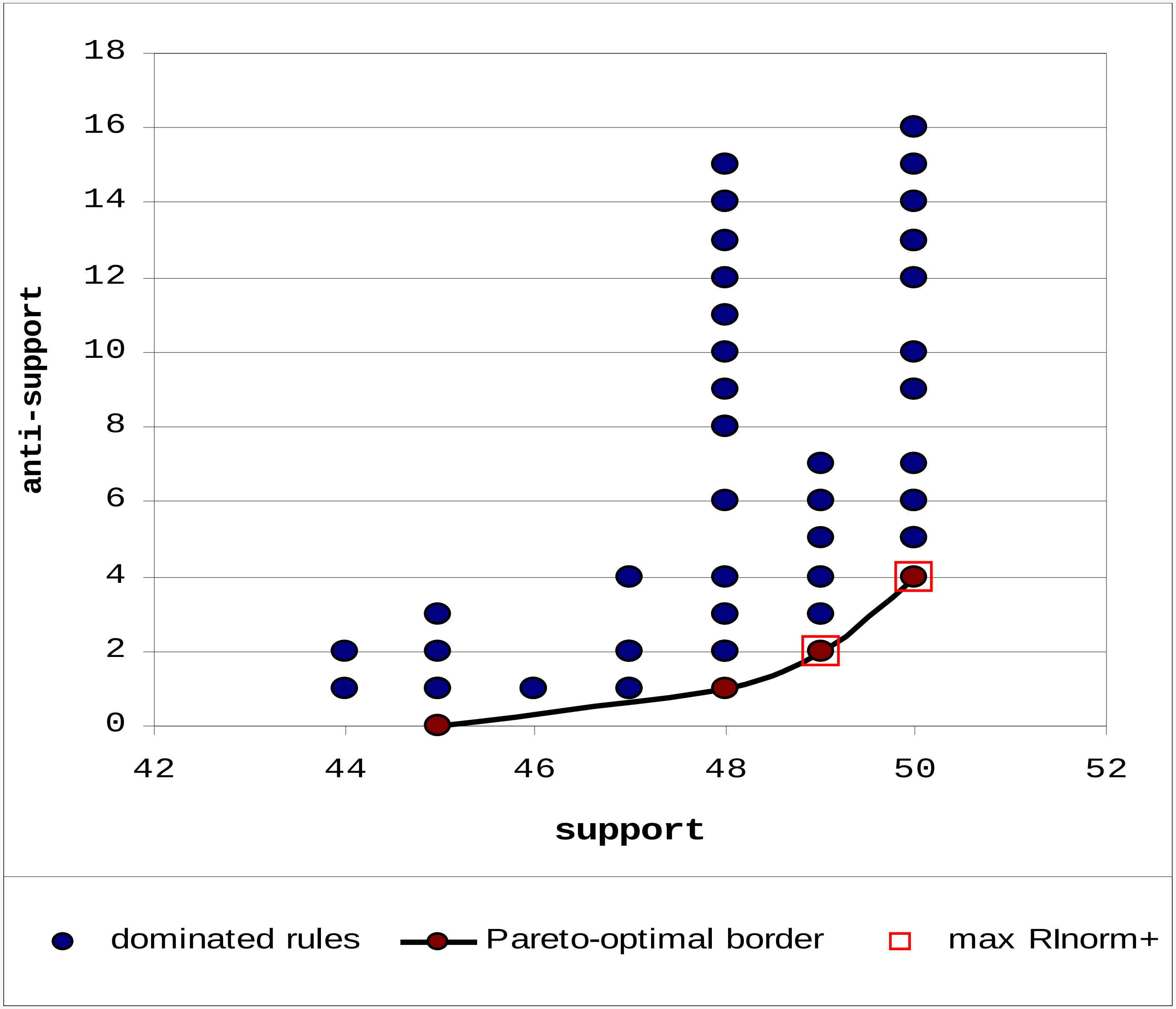
- Since the normalized *RI* and *gain* measures satisfy the property M we can conclude that
  rules optimal with respect to them will be found in the set of rules non-dominated according to support and anti-support.
  (considering rules with the same conclusion)

- Experiments illustrating the result:

  Dataset: *busses*, containing info. about technical state of buses

  Set of 85 rules with the same conclusion

Practical application of the results



Conclusions

Conclusions

Normalized interestingness measure	Property M
Rule Interest Function	YES
Gain	YES
Dependency Factor	NO

Conclusions

- **Properties explain how the measures behave in certain situations and thus, group them helping the user choose the measure relevant for his expectations**

  e.g. we know that the normalized *RI* is monotonically dependent on the number of objects supporting the rule
  or the number of objects supporting neither premise nor conclusion

Conclusions

- **Possession of property M implies potential efficiency improvement:**
  - **we can concentrate on mining only the support–anti-support Pareto set instead of conducting rule evaluation separately wrt to normalized *RI*, gain, or any other measure with property M**
  - **rules optimal wrt to normalized *RI*, gain or any other measure with property M can be mined from the support–anti-support Pareto set instead of searching the set of all rules**
  - **due to relationship between anti-support and any measure with property M, the rule order wrt anti-support (for fixed value of support) is the same for any other measure with M**

Thank you!