Multicriteria attractiveness evaluation of decision and association rules

Izabela Szczęch

Poznań University of Technology

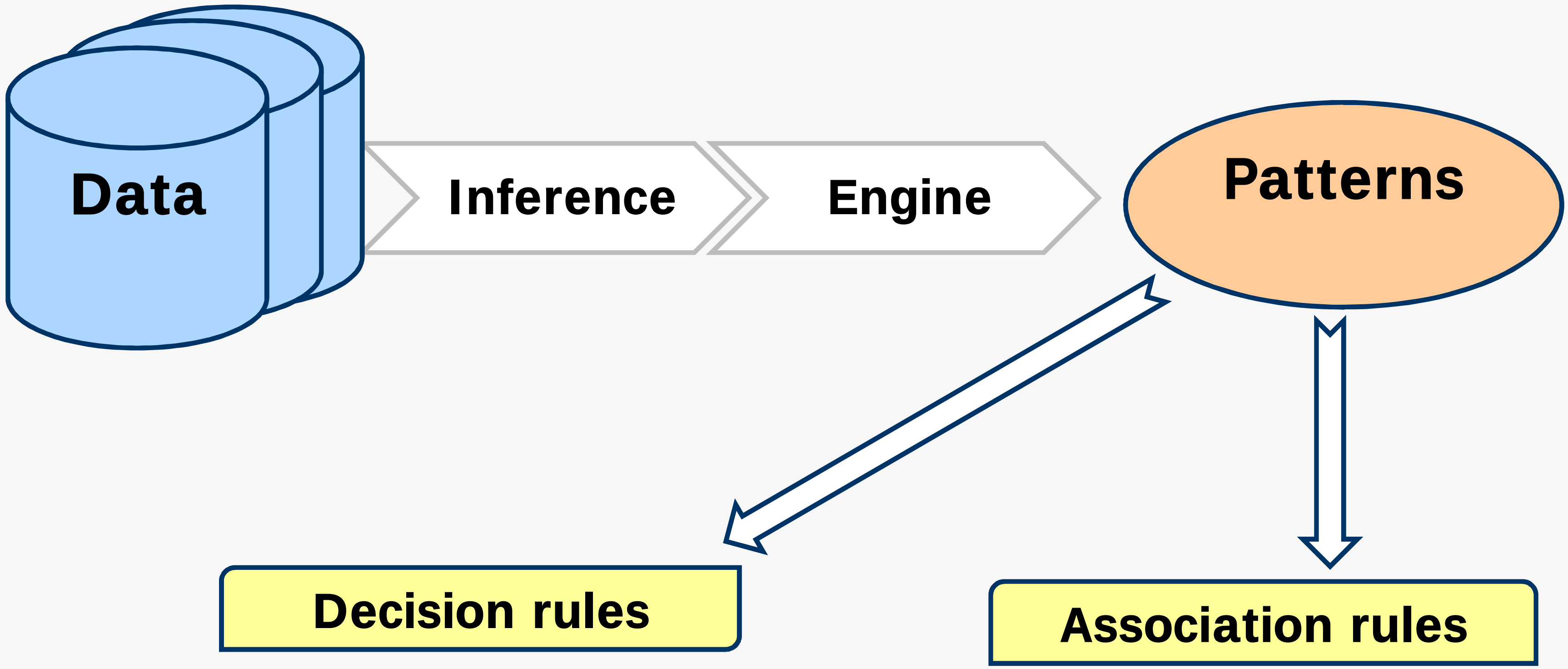
Wielokryterialna ocena atrakcyjności reguł decyzyjnych i asocjacyjnych

Poznań, 11.03.2008

Presentation plan

- Introduction
  - Rule induction
  - Rule evaluation and attractiveness measures
  - Desirable properties of attractiveness measures
  - Multicriteria rule evaluation
- Aim and scope of the thesis
- Property analysis of attractiveness measures
- Relationships between measures f, s, any measure with property M, and support-confidence Pareto border
- Support and anti-support rule evaluation space
- Summary

Rule induction



If symptom s1 is present
and symptom s1 is absent
then disease d1

If bread was bought
then butter *and* milk were bought

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 and y are condition and conclusion formulas
  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.**

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*)**

Attractiveness 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, conviction, 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

- **Basic quantitative characteristics of rules**
  - ***Support*** **of rule** $\phi \rightarrow \psi$ **in** ***S*****:**

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

  - ***Confidence*** **(called also *certainty factor*) of rule** $\phi \rightarrow \psi$ **in** ***S*****:**

$$conf(\phi \rightarrow \psi) = \frac{sup(\phi \rightarrow \psi)}{sup(\phi)}$$

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

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

Confirmation measure *f* and *s*

- **Confirmation measure** f **(Good 1984, Heckerman 1988, Pearl 1988, Fitelson 2001)**

$$f(\phi \to \psi) = \frac{conf(\psi \to \phi) - conf(\neg\psi \to \phi)}{conf(\psi \to \phi) + conf(\neg\psi \to \phi)}$$

- **Confirmation measure** s **(Christensen 1999)**

$$s(\phi \to \psi) = conf(\phi \to \psi) - conf(\neg\phi \to \psi)$$

- **Gain measure (Fukuda et al. 1996)**
- **Rule Interest Function (Piatetsky-Shapiro 1991)**
- **Dependency Factor (Pawlak 2002, Popper 1959)**
- **...**

Bayesian confirmation property

- **An attractiveness *c* measure has the property of confirmation if is satisfies the following condition:**

$$c(\phi,\psi)\begin{cases} >0 \;\; if \;\; Pr(\psi|\phi) > Pr(\psi) \\ =0 \;\; if \;\; Pr(\psi|\phi) = Pr(\psi) \\ <0 \;\; if \;\; Pr(\psi|\phi) < Pr(\psi) \end{cases}$$

- **Measures of confirmation quantify the strength of confirmation that premise ϕ gives to conclusion ψ**
- **„ψ is verified more often, when ϕ is verified, rather than when ϕ is not verified"**

Property M

- **Property M** **(Greco, Pawlak, Słowiński 2004)**
- **An attractiveness measure *I*(*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=$sup(\phi \rightarrow \psi)$

the number of objects in *U* for which ϕ and ψ hold together

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

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

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

Property M - interpretation

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

  if x is a raven then x is black

- ϕ **is the property *to be a raven*, ψ is the property *to be black***
  - ***a*** **– the number of objects in *U* which are black ravens**

    //the more black ravens we observe, the more credible becomes the rule
  - ***b*** **– the no. of objects in *U* which are black non-ravens**
  - ***c*** **– the no. of objects in *U* which are non-black ravens**
  - ***d*** **– the no. of objects in *U* which are non-black non-ravens**

Properties of symmetry

- **Properties of symmetry** **(Carnap 1962, Eells & Fitelson 2000):**
  - **Evidence symmetry:** $I(\phi \rightarrow \psi) = -I(\neg\phi \rightarrow \psi)$
  - **Commutativity symmetry:** $I(\phi \rightarrow \psi) = I(\psi \rightarrow \phi)$
  - **Hypothesis symmetry:** $I(\phi \rightarrow \psi) = -I(\phi \rightarrow \neg\psi)$
  - **Total symmetry:** $I(\phi \rightarrow \psi) = -I(\neg\phi \rightarrow \neg\psi)$

- **Only hypothesis symmetry is desirable**

Property of hypothesis symmetry

- **Property of hypothesis symmetry (HS)** (Carnap `62, Eells, Fitelson ′02)
- **An interestingness measure $I(\phi \rightarrow \psi)$ has the property HS if**

$$I(\phi \rightarrow \psi) = -I(\phi \rightarrow \neg\psi)$$

- **Interpretation:** the impact of ϕ on ψ should be of the same strength, but of the opposite sign as the impact of ϕ on $\neg\psi$
- **Example:** Let us consider a rule $\phi \rightarrow \psi$:

if x is $\boxed{}$ *then x is* $\boxed{}$

ϕ is conclusive for ψ and negatively conclusive for $\neg\psi$

Multicriteria rule evaluation

A single measure is often an insufficient indicator of the quality of rules, so there arises a natural **need for a multicriteria evaluation.**

Support–confidence evaluation space (Bayardo & Agrawal 1999)

semantic meaning of confidence does not allow
to distinguish rules for which
the premise disconfirms the conclusion

need to search for substituting evaluation spaces that would include

- confirmation measures
- measures with property M

General aim of the thesis

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.

Detailed tasks

- **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.**

- **Analysis of relationships between the considered interestingness measures and analysis of enclosure relationships between the sets of non-dominated rules in different evaluation spaces.**

Detailed tasks

- 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.

- Determining the area of rules with desirable value of a confirmation measure in the proposed multicriteria evaluation space.

- 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.

Performed analyses of properties of attractiveness measures

- **Analysis of measures wrt property of confirmation:**

  Theorems:

  RI, Dependency factor, and Gain (if $\Theta = sup(\psi)/|U|$) have the property of confirmation,

  while Rule support, Anti-support, Confidence do not have the property of confirmation

- **Analysis of measures wrt property M:**

  Theorems:

  Rule support, Anti-support, Confidence, RI, Gain have the property M,

  while Dependency factor does not have the property M

Performed analyses of properties of attractiveness measures

- **Analysis of measures wrt property hypothesis symmetry:**

  Theorems:

  RI and Gain (if $\Theta=1/2$) have the property of confirmation,

  while Rule support, Anti-support, Confidence and Dependency factor do not have the property HS

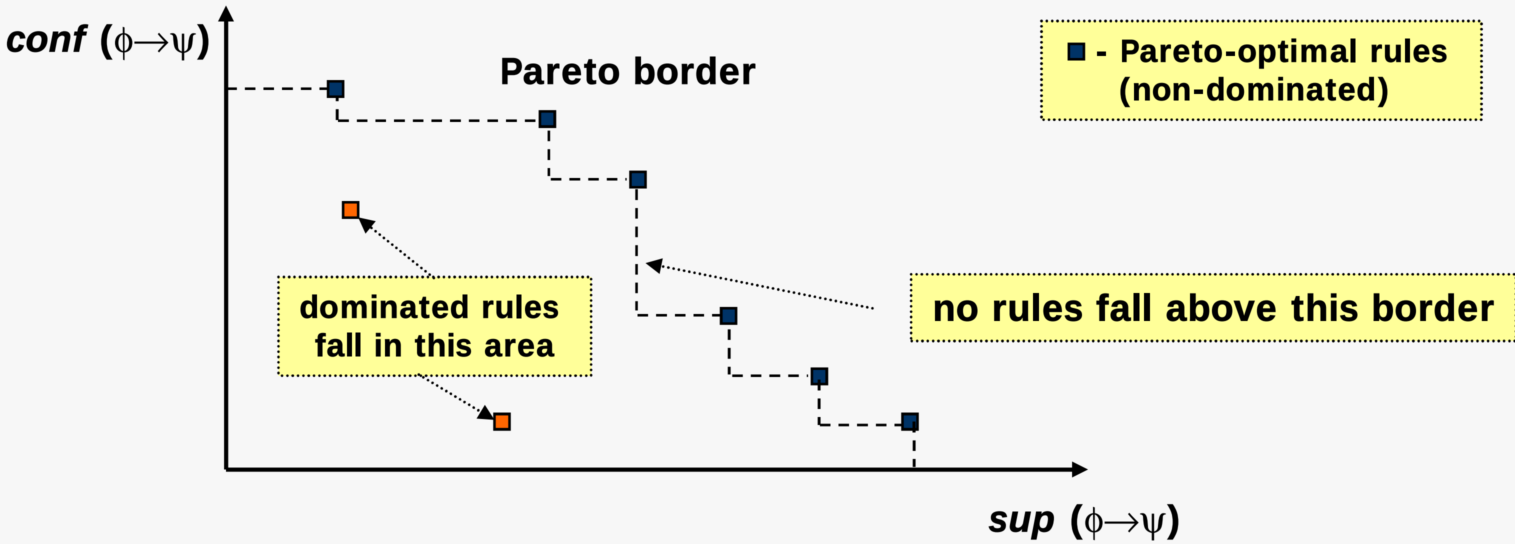
- ***Theorem*** **(Greco, Pawlak & Słowiński 2004):**

  Confirmation measures f, s have the property M and property of hypothesis symmetry

Support–confidence Pareto border

Support–confidence Pareto border

- **Support–confidence Pareto border is the set of non-dominated, Pareto-optimal rules with respect to both rule support and confidence**



- **Mining the border identifies rules optimal with respect to measures such as: *lift*, *gain*, *conviction*, *Piatetsky-Shapiro*,…**

Monotonicty of *f* in support and confidence

- **Is measure f included in the support-confidence Pareto border?**

- ***Theorem:***

  Confirmation measure f is independent of support, and, therefore, monotone in support, when the value of confidence is held fixed.

- ***Theorem:***

  Confirmation measure f is increasing, and, therefore, monotone in confidence

- **Conclusion:**

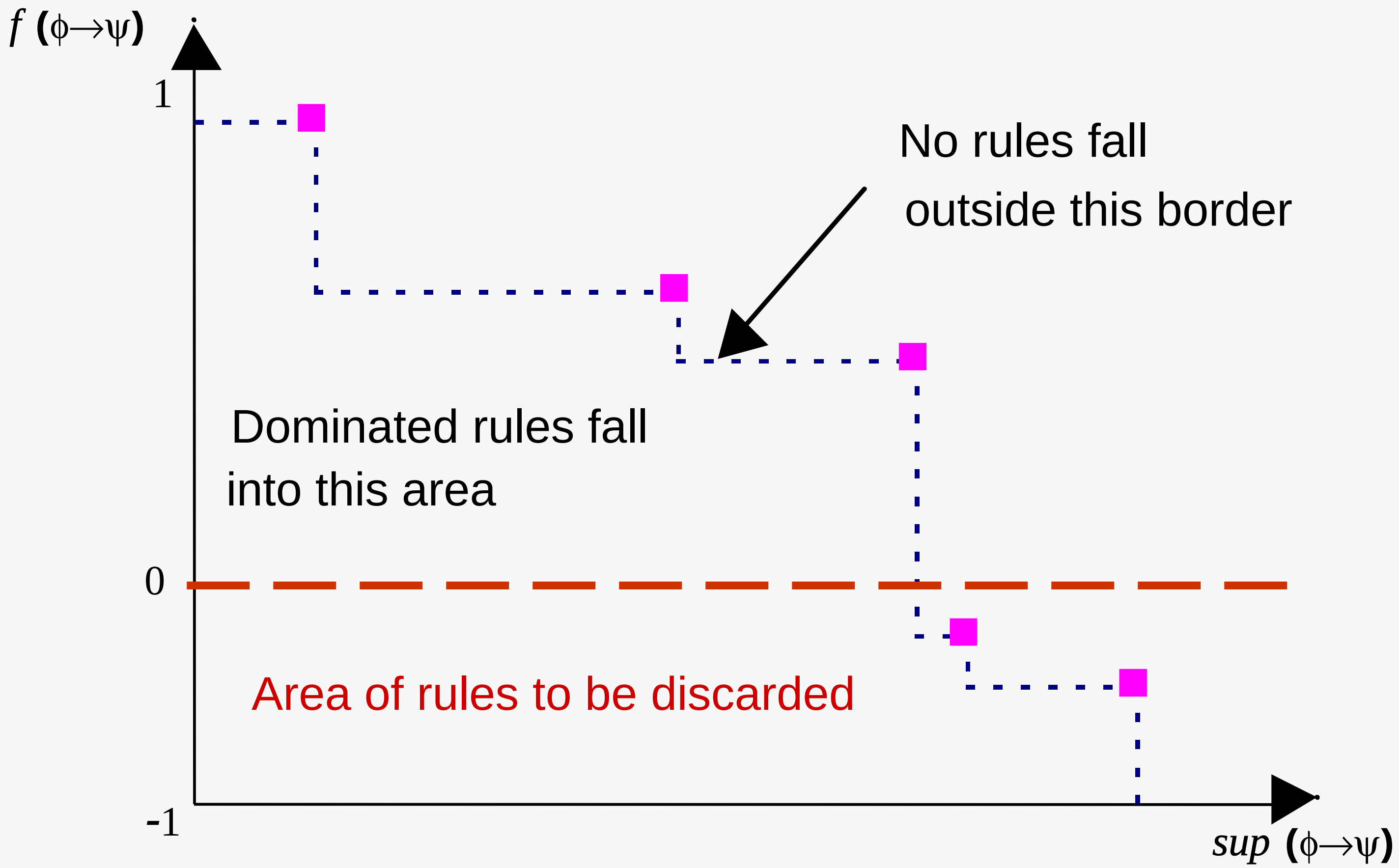
  Rules maximizing f lie on the support–confidence Pareto border

Support–confidence vs. support–*f* Pareto border

- **The utility of confirmation measure f outranks utility of confidence**
- **Claim: Substitute the $conf(\phi \rightarrow \psi)$ dimension for $f(\phi \rightarrow \psi)$**
- ***Theorem:***

  The set of rules located on the support–confidence Pareto border is exactly the same as on the support–f Pareto border

Support−*f* Pareto border is more meaningful



Confirmation perspective on support–confidence space

- **Is there a curve separating rules with negative value of any measure with the confirmation property in the support–confidence space?**

- ***Theorem:***

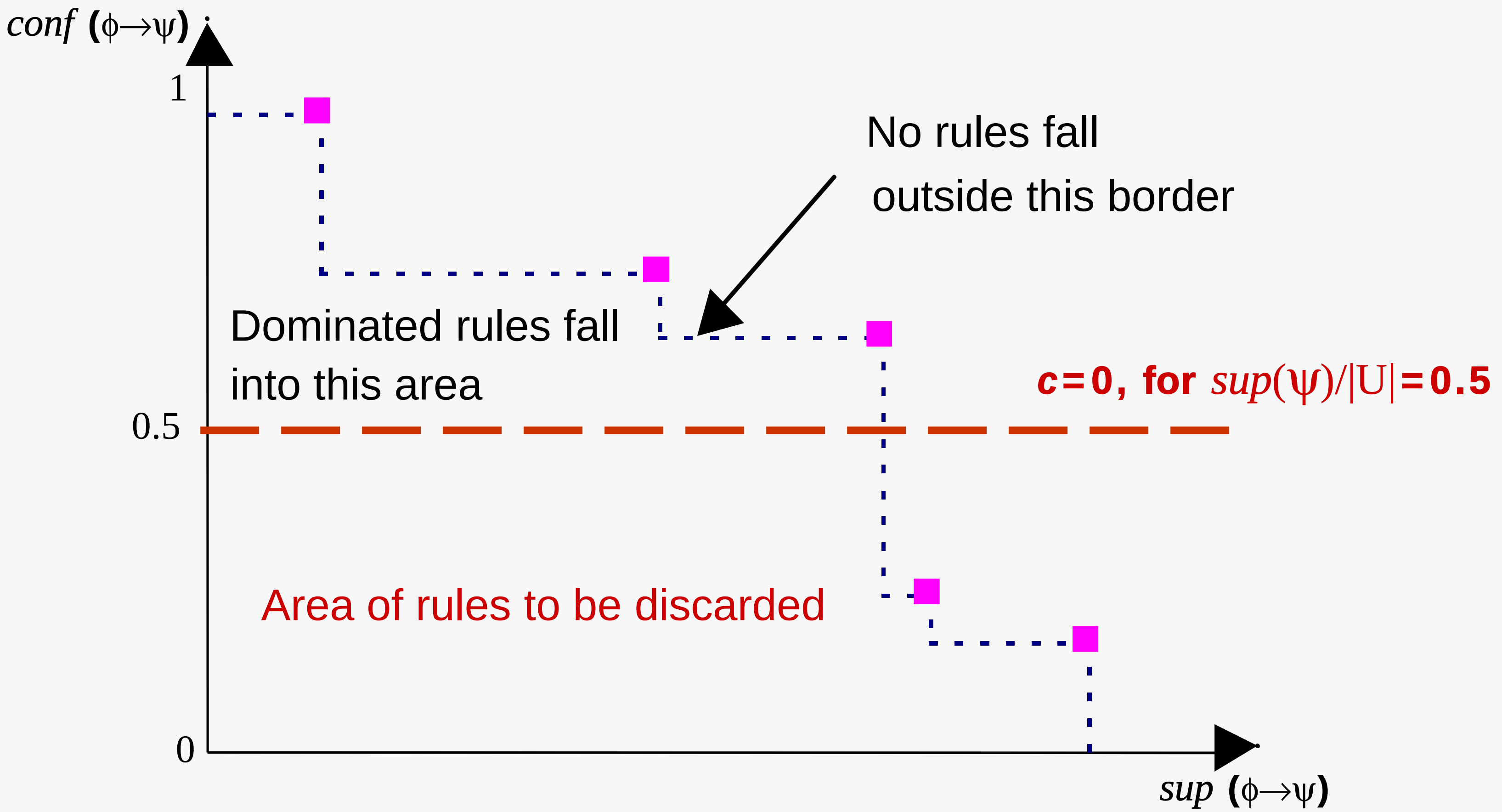
  Rules lying above a constant:

$$conf(\phi \rightarrow \psi) = sup(\psi)/|U|$$

  have a negative value of any confirmation measure.

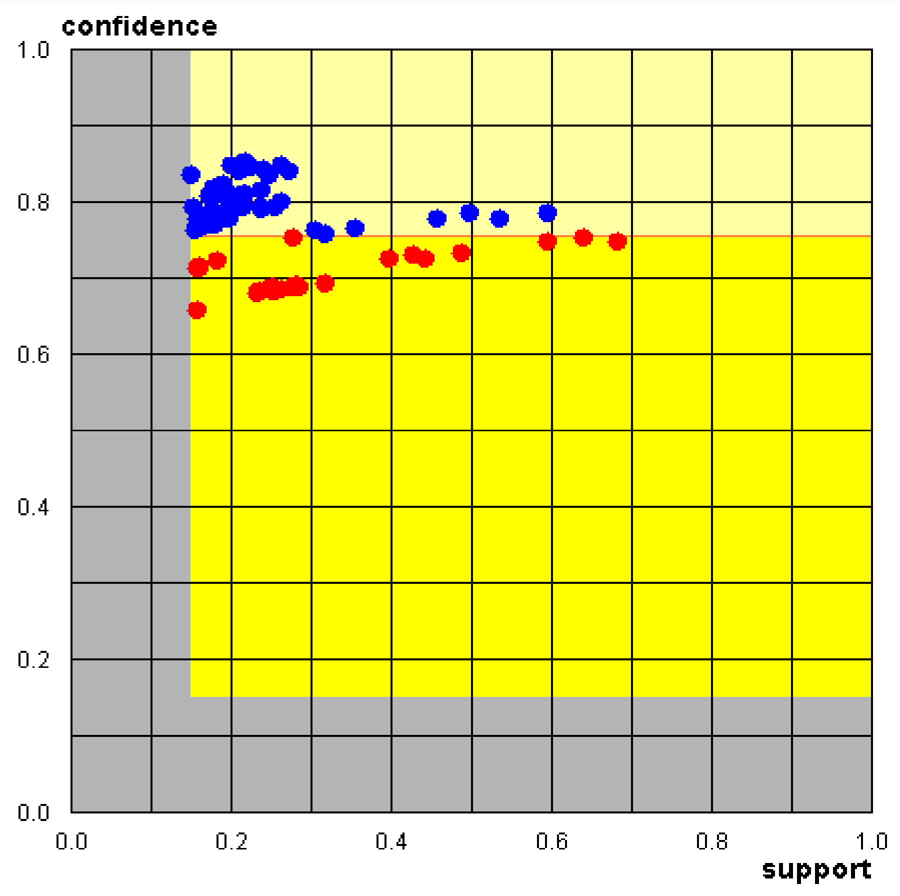
  For those rules, the premise only disconfirms the conclusion!

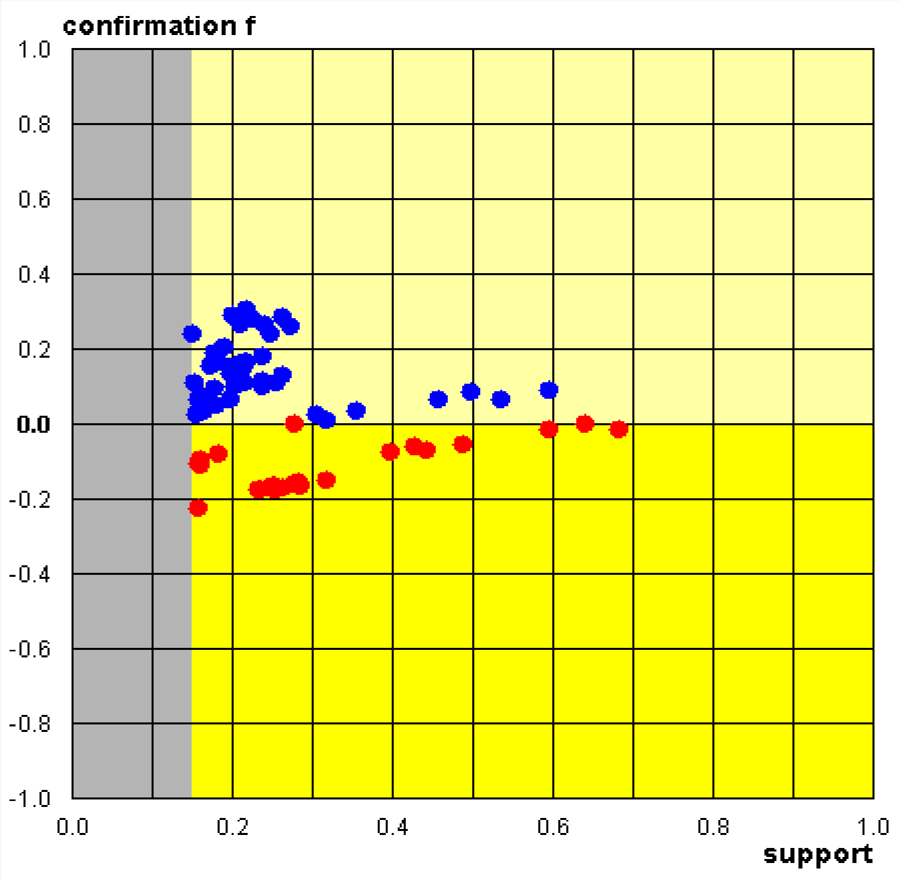
Confirmation perspective on support–confidence space



For rules lying below the curve for which *c*=0
the premise only disconfirms the conclusion

Support-confidence Pareto border vs. support-f





- ● indicates rules with negative confirmation
- the class constitutes over 70% of the whole dataset
- rules with high confidence can be disconfirming
- even some rules from the Pareto border need to be discarded

Support–s Pareto border

Monotonicty of *s* in support and confidence

- **Is measure s on rule support–confidence Pareto border?**

- **Theorem:**
  Confirmation measure s is increasing, and, therefore, monotone in confidence when the value of support is held fixed

- **Theorem:**
  For a fixed value of confidence, confirmation measure s is:
  - **increasing in** $sup(\phi \rightarrow \psi) \Leftrightarrow s(\phi \rightarrow \psi) > 0$
  - **constant in** $sup(\phi \rightarrow \psi) \Leftrightarrow s(\phi \rightarrow \psi) = 0$
  - **decreasing in** $sup(\phi \rightarrow \psi) \Leftrightarrow s(\phi \rightarrow \psi) < 0$

- **The above theorem states the monotone relationship just in the non-negative range of the value of s (i.e. the only interesting)**

Support–confidence vs. support–s Pareto border

- **<u>Theorem:</u>**

  If a rule resides on the support–s Pareto border

  (in case of positive value of s),

  then it also resides on the support–confidence Pareto border,

  while one can have rules being on the support–confidence Pareto border which are not on the support–s Pareto border.

- **Conclusion:**

  The support–confidence Pareto border is, in general, larger than the support–s Pareto border

Measures with the property M in support–confidence space

- What are the **conditions** for rules maximizing
  any measure with the property M
  to be included in the rule support-confidence Pareto border?

- Reminder of the **property M**:
  $a = sup(\phi \rightarrow \psi)$, $b = sup(\neg\phi \rightarrow \psi)$, $c = sup(\phi \rightarrow \neg\psi)$, $d = sup(\neg\phi \rightarrow \neg\psi)$
  $I(a,b,c,d)$ is a function **non-decreasing** with respect to a and d,
  and **non-increasing** with respect to b and c

Measures with the property M in support−confidence space

- **Theorem:**

  When the value of support is held fixed, then *I*(*a*, *b*, *c*, *d*) is monotone in confidence.

- **Theorem:**

  When the value of confidence is held fixed, then *I*(*a*, *b*, *c*, *d*) admitting derivative with respect to all its variables *a*, *b*, *c* and *d*, is monotone in support if:

$$\frac{\partial I}{\partial c} = \frac{\partial I}{\partial d} = 0 \quad or \quad \frac{\dfrac{\partial I}{\partial a} - \dfrac{\partial I}{\partial b}}{\dfrac{\partial I}{\partial d} - \dfrac{\partial I}{\partial c}} \geq \frac{1}{conf(\phi \to \psi)} - 1$$

- **There are some measure with property M whose optimal rules will not be on the support-confidence Pareto border.**

Support–anti-support Pareto border

Support–anti-support Pareto border

- **How to find rules optimal according to any measure with the property M?**

- **Theorem:**
  When the value of support is held fixed, then *I*(*a*, *b*, *c*, *d*) is anti-monotone (non-increasing) in anti-support

- **Theorem:**
  When the value of anti-support is held fixed, then *I*(*a*, *b*, *c*, *d*) is monotone (non-decreasing) in support

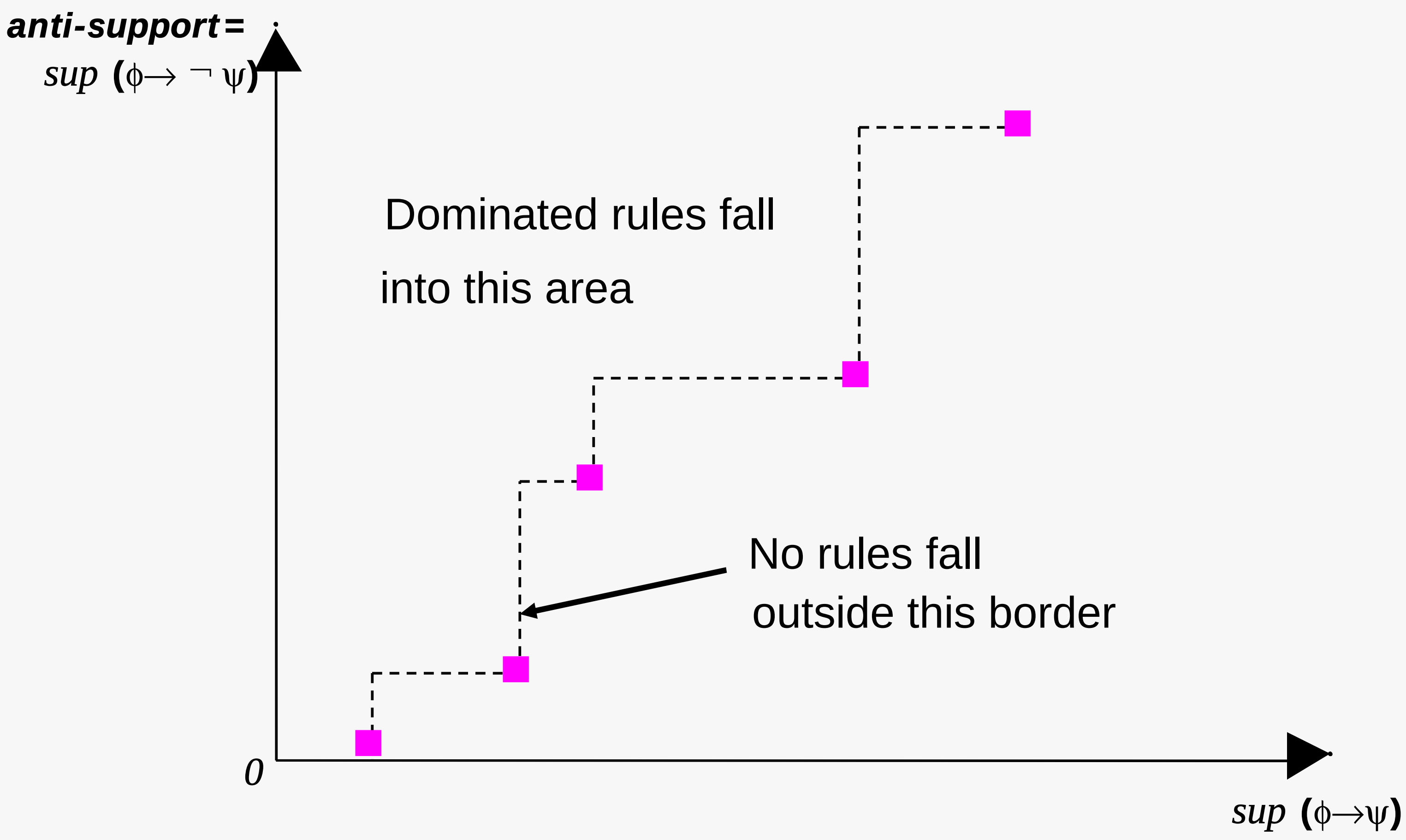
Support–anti-support Pareto border

- **Theorem:**
  For rules with the same conclusion,
  the best rules according to any measure with the property M
  must reside on the support–anti-support Pareto border

- The support–anti-support Pareto border is the set of rules such that there is no other rule having greater support and smaller anti-support

- **Theorem:**
  The support–anti-support Pareto border is, in general, not smaller than the support–confidence Pareto border

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

Confirmation perspective on the support–anti-support Pareto border

Confirmation perspective on support–anti-support border

- **Is there a curve separating rules with negative value of any confirmation measure in the support–anti-support space?**

- ***<u>Theorem:</u>***

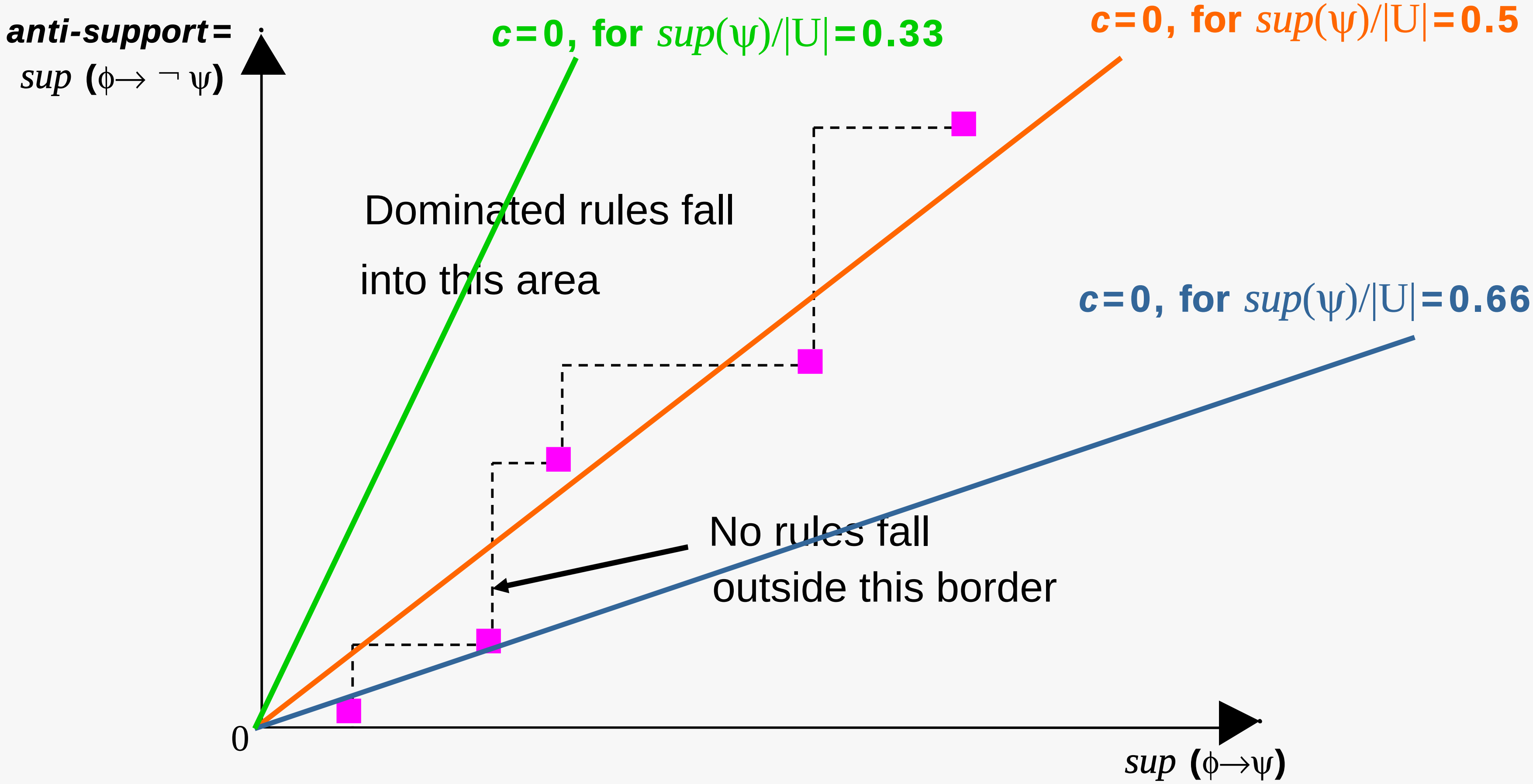
  Rules lying above a linear function:

$$anti\text{-}sup(\phi \rightarrow \psi) = sup(\phi \rightarrow \psi)[|U|/sup(\psi)-1]$$

  have a negative value of any confirmation measure.

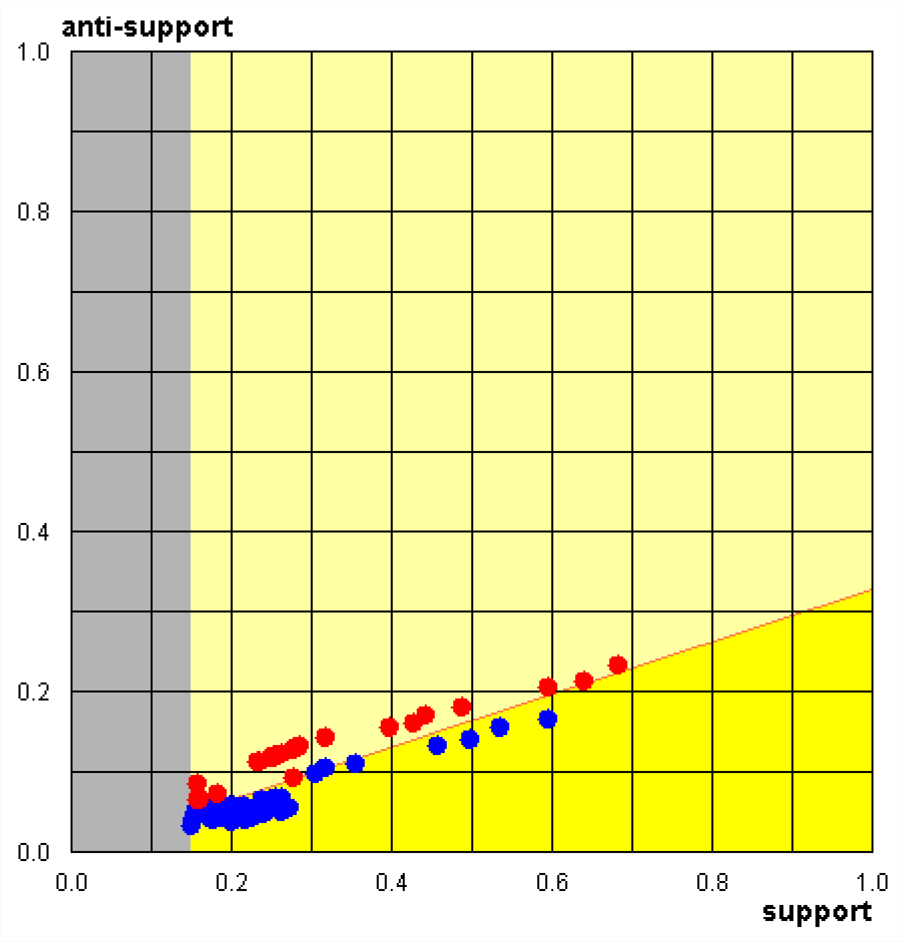
  For those rules, the premise only disconfirms the conclusion!

Confirmation perspective on support–anti-support border



For rules lying above the curve for which *c*=0
the premise only disconfirms the conclusion

Support - anti-support (workclass=Private)



- ● indicates rules with negative confirmation
- even some rules from the Pareto border need to be discarded

Inner monotonicity

in support - anti-support space

The gist of the algorithm for support–anti-support rules

- Traditional Apriori approach to generation of association rules (Agrawal et al) proceeds in a two step framework:
  - find frequent itemsets (i.e. sets of items which occur more frequently than the minimum support threshold),
  - generate rules from frequent itemsets and filter out those that do not exceed the minimum confidence threshold

- Generation of association rules regarding support and anti-support, in general, requires only the substitution of the parameter calculated in step 2. Confidence -> anti-support

The gist of the algorithm for support–anti-support rules

- **Claim: calculation of anti-support (instead of confidence) does not introduce any more computational overhead to the algorithm**

- **Let us observe that:** $\textit{anti-sup}(\phi \rightarrow \psi) = \textit{sup}(\phi \rightarrow \neg \psi) = \textit{sup}(\phi) - \textit{sup}(\phi \rightarrow \psi)$.
- **All the data required to calculate anti-support are also gathered in step 1 of Apriori**
- **The data needed to calculate anti-support is the same as to calculate confidence**

The gist of the algorithm for support–anti-support rules

- **Claim: When generating association rules from a frequent set it is advisable to first generate rules with few conclusion elements (for optimisation reasons)**
- **Let us observe three different rules constructed from the same frequent itemset {x, y, z, v}:**
  - r_1: $x \rightarrow yzv$ $\quad$ $anti\text{-}sup(r_1) = sup(x) - sup(xyzv)$
  - r_2: $xy \rightarrow zv$ $\quad$ $anti\text{-}sup(r_2) = sup(xy) - sup(xyzv)$
  - r_3: $xyz \rightarrow v$ $\quad$ $anti\text{-}sup(r_3) = sup(xyz) - sup(xyzv)$
- $anti\text{-}sup(r_1) \geq anti\text{-}sup(r_2) \geq anti\text{-}sup(r_3)$
- **Conclusion:** $anti\text{-}sup(r_3)$ **> *max_acceptable anti-support* =>**

  $anti\text{-}sup(r_2)$ **> *max_acceptable anti-support***

Generate and verify r_3 first!

Summary

Main results of the thesis

- **Analysis of 8 measures with respect to the property M, the property of confirmation and the property of hypothesis symmetry has been performed**

- **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**

Main results of the thesis

- A proposition of a support–anti-support 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

- The support–confidence and support–anti-support evaluation spaces has also been enriched by the valuable semantics of confirmation measures.

- A multicriteria rule evaluation system has been designed and developed. As the application of the system three datasets, *census*, *msweb* and *hsv*, have been analyzed and discussed

Lines of further investigation

- **Analysis of attractiveness measures with respect to other properties, in particular other forms of symmetry properties**

- **Development of algorithm for finding in support - anti-support space a set of rules (both dominated and non-dominated) that covers the objects in a certain percentage**

- **Analysis of properties of normalized measures (Crupi et al)**

Thank you!