Application of Bayesian confirmation measures for mining rules from the support-confidence Pareto-optimal set

Roman Słowiński

Poznań University of Technology,

Systems Research Institute, PAS, Poland

Izabela Brzezińska

Poznań University of Technology

Salvatore Greco

University of Catania, Italy

ICAISC 2006, Zakopane

Plan

- Introduction
  - Basic quantitative characteristics of rules
  - Bayesian confirmation measures and their desirable properties
  - Confirmation measures *f* and *s*
  - Utility of confidence vs. utility of confirmation measures
- Support-confidence Pareto-optimal border
- New proposals:
  - support-confirmation *f* Pareto-optimal border
  - support-confirmation *s* Pareto-optimal border
  - support-anti-support Pareto-optimal border
- Conclusions

Introduction

- Discovering rules from data is the domain of inductive reasoning (IR)
- IR uses data about a sample of larger reality to start inference
- $S=\langle U, A\rangle$ – *data table*, where U and A are finite, non-empty sets
  U – *universe*; A – set of *attributes*
- $S=\langle U, C, D\rangle$ – *decision table*, where C – set of *condition attributes*,
  D – set of *decision attributes*, $C \cap D = \varnothing$

e.g.

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

Introduction

- With every subset of attributes $B \subseteq A$, one can associate a formal language of formulas ***L***, called *decision language*
- Formulas are built from attribute-value pairs (q,v), where $q \in B$ and $v \in V_a$ (domain of a), using logical connectives $\wedge$, $\vee$, $\neg$
- All formulas in ***L*** are partitioned into *condition* and *decision formulas* (called *premise* and *conclusion*, resp.)
- *Decision rule* or *association rule* induced from S is a *consequence relation*: $\phi \rightarrow \psi$ *read as* ***if*** ϕ**,** ***then*** ψ where ϕ and ψ are condition and decision formulas expressed in ***L***

Introduction

- 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*)
  3) If (*Height=medium*) & (*Hair=blond*), then (*Nationality=Swede*)
  4) If (*Height=tall*), then (*Nationality=German*)
  5) If (*Height=short*), then (*Nationality=German*)
  6) If (*Height=medium*) & (*Hair=dark*), then (*Nationality*=Swede)

- *Decision rule* or *association rule* induced from *S*

  is a *consequence relation*: $\phi \rightarrow \psi$ *read as* ***if*** ϕ, ***then*** ψ

  where ϕ and ψ are condition and decision formulas expressed in ***L***

Introduction

- The number of rules generated from massive datasets can be very large and only a few of them are likely to be useful
- In all practical applications, like medical practice, market basket, it is crucial to know how good the rules are
- 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,…)
- There is no evidence which measure(s) is (are) the best

Introduction – Basic quantitative characteristics of rules

- $\|\phi\|$ is the set of all objects from *U*, having property ϕ in *S*
- $\|\psi\|$ is the set of all objects from *U*, having property ψ in *S*

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

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

  - *Confidence* (called also *certainty factor*) of decision rule $\phi \rightarrow \psi$ in *S* (Łukasiewicz, 1913):

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

Introduction – Bayesian confirmation measures

- Among widely studied interestingness measures, there is a group of *Bayesian confirmation measures*
- 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”

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

where $Pr(\psi|\phi) = conf(\phi,\psi) = \dfrac{sup(\phi \to \psi)}{sup(\phi)}$

- Its meaning is different from a simple statistics of co-occurrence of properties ϕ and ψ in universe U

Introduction – Desirable properties of confirmation measures

- Desirable properties of $c(\phi,\psi)$:
  - *hypothesis symmetry* (Eells, Fitelson 2002):

    $c(\phi,\psi) = -c(\phi,\neg\psi)$

  - *monotonicity* (M) (Greco, Pawlak, Słowiński 2004):

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

    $c(\phi,\psi) = F(a,b,c,d)$, where F is a function non-decreasing with respect to a and d and non-increasing with respect to b and c

Introduction – Desirable properties of confirmation measures

- The property of monotonicity (M) takes into account four evidences in assesement of the impact of property ϕ on $\phi \rightarrow \psi$
- 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* and ψ is the property *to be black*
  - *a* – the number of objects in *S* which are black ravens
  - *b* – the number of objects in *S* which are black non-ravens
  - *c* – the number of objects in *S* which are non-black ravens
  - *d* – the number of objects in *S* which are non-black non-ravens

Introduction – Desirable properties of confirmation measures

- $c(\phi, \psi)>0$ means that property ψ is satisfied more frequently when ϕ is satisfied (then, this frequency is *conf*(ϕ, ψ)), rather than generically in *S* (where the frequency is *Pr*(ψ)),

- $c(\phi, \psi)=0$ means that property ψ is satisfied with the same frequency whether ϕ is satisfied or not

- $c(\phi, \psi)<0$ means that property ψ is satisfied less frequently when ϕ is satisfied, rather than generically

Introduction – Confirmation measure *f* and *s*

- As shown by (Greco, Pawlak, Słowiński 2004), confirmation measure *f* (Good 1984, Heckerman 1988, Pearl 1988, Fitelson 2001)

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

and confirmation measure *s* (Christensen 1999)

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

are the only ones that enjoy both hypothesis symmetry (HS) and monotonicity (M), among the most well known confirmation measures

Utility of confidence vs. utility of confirmation *measures* (1)

- Utility of scales:

  - *conf*($\phi \rightarrow \psi$) is the truth value of the knowledge pattern „*if* ϕ, *then* ψ”,

  - *f*($\phi \rightarrow \psi$), *s*($\phi \rightarrow \psi$) say to what extend ψ is satisfied more frequently when ϕ is satisfied rather than when ϕ is not satisfied

Utility of confidence vs. utility of confirmation measures Eg. 1

- Consider the possible result of rolling a die: 1,2,3,4,5,6, and let the conclusion be ψ=„the result is 6”
  - ϕ_1 ="the result is divisible by 3" $conf(\phi_1 \rightarrow \psi)=1/2$, $f(\phi_1 \rightarrow \psi)=2/3$
  - ϕ_2 ="the result is divisible by 2" $conf(\phi_2 \rightarrow \psi)=1/3$, $f(\phi_2 \rightarrow \psi)=3/7$
  - ϕ_3 ="the result is divisible by 1" $conf(\phi_3 \rightarrow \psi)=1/6$, $f(\phi_3 \rightarrow \psi)=0$
- In particular, rule $\phi_3 \rightarrow \psi$, can be read as „in any case, the result is 6”; indeed, the „any case” does not add any information which could confirm that the result is 6, and this fact is expressed by $f(\phi_3 \rightarrow \psi)=0$
- This example clearly shows that the value of *f* has a more useful interpretation than *conf*

Utility of confidence vs. utility of confirmation measures Eg. 2

- Consider the possible result of rolling a die: 1,2,3,4,5,6, and let the premise be ϕ=„the result is divisible by 2"
  - ψ_1="the result is 6" $conf(\phi\rightarrow\psi_1)=1/3$, $f(\phi\rightarrow\psi_1)=3/7$
  - ψ_2="the result is *not* 6" $conf(\phi\rightarrow\psi_2)=2/3$, $f(\phi\rightarrow\psi_2)=-3/7$
- In this example, rule $\phi\rightarrow\psi_2$ has greater confidence than rule $\phi\rightarrow\psi_1$
- However, rule $\phi\rightarrow\psi_2$ is less interesting than rule $\phi\rightarrow\psi_1$ because premise ϕ reduces the probability of conclusion ψ_2 from $5/6=sup(\psi_2)$ to $2/3= conf(\phi\rightarrow\psi_2)$, while it augments the probability of conclusion ψ_1 from $1/6=sup(\psi_1)$ to $1/3= conf(\phi\rightarrow\psi_1)$
- In consequence, premise ϕ disconfirms conclusion ψ_2, which is expressed by a negative value of $f(\phi\rightarrow\psi_2)=-3/7$, and it confirms conclusion ψ_1, which is expressed by a positive value of $f(\phi\rightarrow\psi_1)=3/7$

Support-confidence Pareto border

- In the set of rules induced from data, we look for rules that are optimal according to a chosen attractiveness measure

- This problem was addressed with respect to such measures as *lift*, *gain*, *conviction*, *Piatetsky-Shapiro*,…

- Bayardo and Agrawal (1999) proved, however, that

  given a fixed conclusion ψ, the *support-confidence Pareto border*

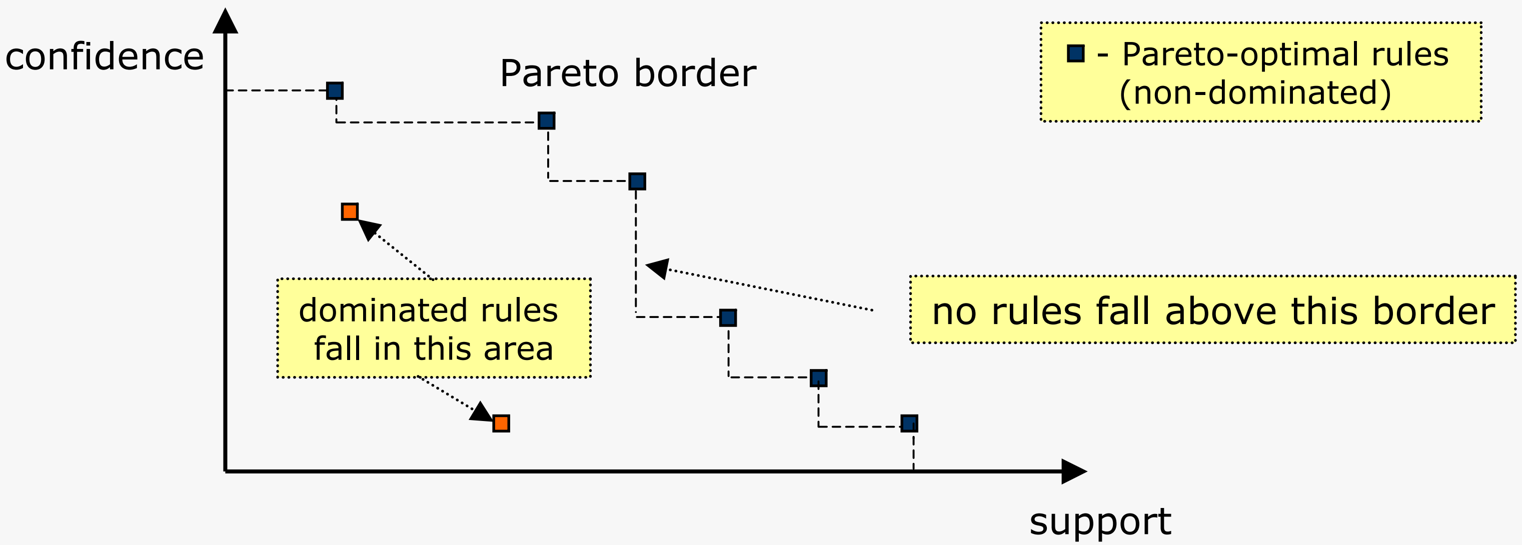
  (i.e. Pareto-optimal border w.r.t. rule support and confidence)

  includes optimal rules according to any of those attractiveness

  measures

Support-confidence Pareto border

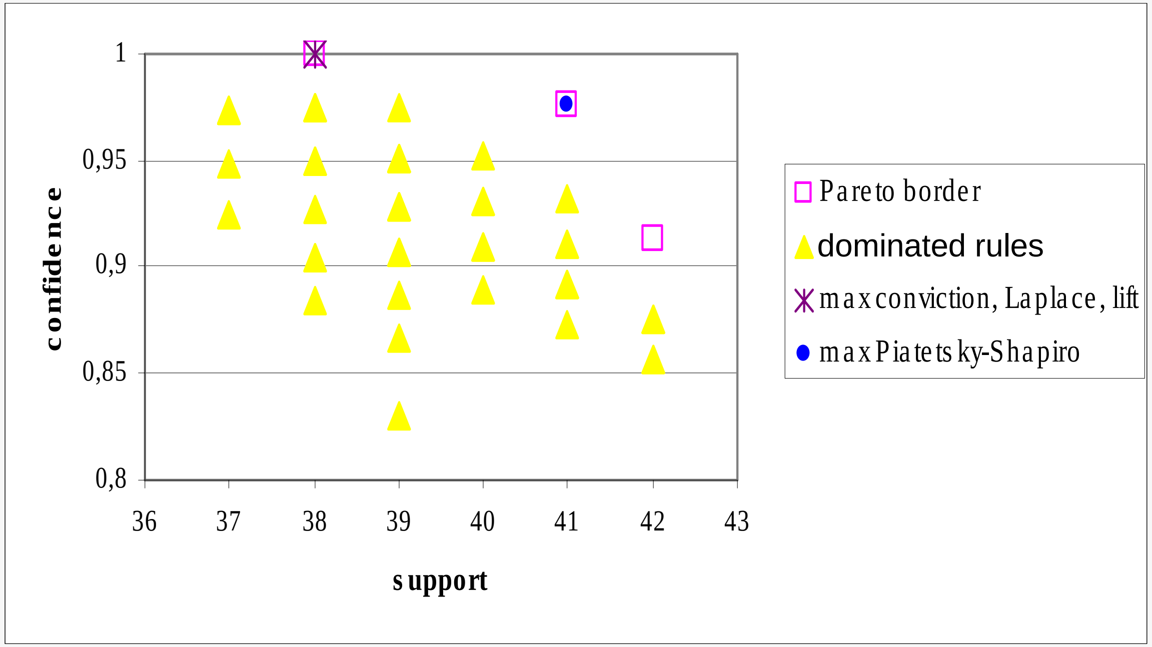
- Support-confidence Pareto border is the set of non-dominated, Pareto-optimal rules with respect to both rule support and confidence



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

Support-confidence Pareto border

E.g. „Buses" data set, class of „good state"



- Decision rules were generated from lower approximations of preference-ordered decision classes defined according to Variable-consistency Dominance-based Rough Set Approach (VC-DRSA) (Greco, Matarazzo, Słowiński, Stefanowski 2001)

  Rule induction algorithm: all rules algorithm (DOMAPRIORI)

Support-confidence Pareto border

- The following conditions are sufficient for verifying whether rules optimal according to a measure $g(x)$ are included on the support-confidence Pareto border:

  1. $g(x)$ is monotone in support over rules with the same confidence

  and

  2. $g(x)$ is monotone in confidence over rules with the same support

- A function $g(x)$ is understood to be *monotone* in x, if $x_1 \prec x_2$ implies that $g(x_1) \leq g(x_2)$

Monotonicty of *f* in support and confidence

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

- Theorem 1:
  Confirmation measure *f* is independent of support, and, therefore, monotone in support, when the value of confidence is held fixed

- Theorem 2:
  Confirmation measure *f* is increasing, and, therefore, monotone in confidence

- Conclusion:

  Rules maximizing *f* lie on the support-confidence Pareto border
  (rules with fixed conclusion)

Monotonicty of confidence in support and *f*

- The utility of confirmation measure *f* outranks utility of confidence

- **Claim 1:** Substitute the *conf*($\phi \rightarrow \psi$) dimension for *f*($\phi \rightarrow \psi$) in the support-confidence Pareto border

- Corollary 1:
  Confidence is independent of support, and, therefore,
  monotone in support, when the value of *f*($\phi \rightarrow \psi$) is held fixed

- Corollary 2:
  Confidence is increasing, and, therefore, monotone in *f*($\phi \rightarrow \psi$)

- Conclusion:

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

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

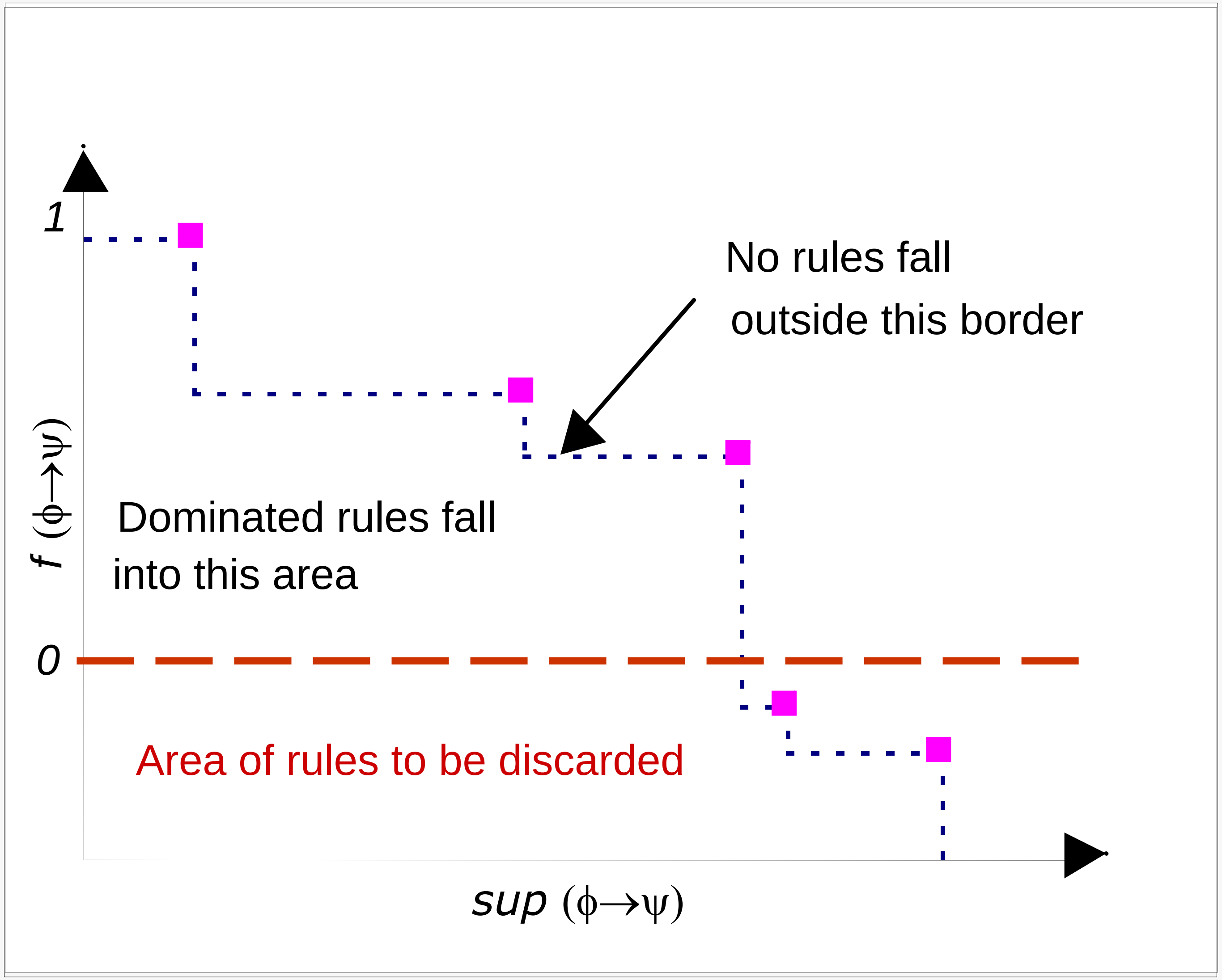
- All the other interestingness measures that were represented on the support-confidence Pareto border also reside on support-*f* Pareto border

- Any non-dominated rule with a negative value of $f(\phi\rightarrow\psi)$ must be discarded from further analysis as its premise only disconfirms the conclusion – such situation cannot be expressed by the scale of confidence

- Conclusion:

  The support-*f* Pareto border is more meaningful than the support-confidence Pareto border

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



Monotonicty of s in support and confidence

- Is confirmation measure s included in the rule support-confidence Pareto border?

- Theorem 3:
  Confirmation measure s is increasing, and, therefore, monotone in confidence when the value of support is held fixed

- Theorem 4:
  For a fixed value of confidence, confirmation measure s is:
  - increasing in $sup(\phi\to\psi) \Leftrightarrow s(\phi\to\psi)>0$
  - constant in $sup(\phi\to\psi) \Leftrightarrow s(\phi\to\psi)=0$
  - decreasing in $sup(\phi\to\psi) \Leftrightarrow s(\phi\to\psi)<0$

- Theorem 4 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

- Theorem 5:

  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

Confirmation measures with the property of monotonicity (M)

- What are the necessary and sufficient conditions for rules maximizing a confirmation measure $c(\phi,\psi)$ with the property of monotonicity (M) to be included in the rule support-confidence Pareto border?

- Reminder of the property of monotonicity (M):

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

  $c(\phi,\psi) = F(a,b,c,d)$, where F is a function non-decreasing with respect to a and d, and non-increasing with respect to b and c

Confirmation measures with the property of monotonicity (M)

- Let $F(a, b, c, d)$ be a confirmation measure with the property (M)

- Theorem 6:
  When the value of support is held fixed, then $F(a, b, c, d)$ is monotone in confidence.

- Theorem 7:
  When the value of confidence is held fixed, then $F(a, b, c, d)$ admitting derivative with respect to all its variables a, b, c and d, is monotone in support if:

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

Confirmation measures with the property of monotonicity (M)

- Conclusions:
  - *Theorem 6* states that for a set of rules with the same conclusion, any Bayesian confirmation measure satisfying the property of monotonicity (M) is always non-decreasing with respect to confidence when the value of support is kept fixed
  - Due to *Theorem 7*, all those confirmation measures that are independent of $c=sup(\phi\rightarrow\neg\psi)$ and $d=sup(\neg\phi\rightarrow\neg\psi)$ are always monotone in support when the value of confidence remains unchanged

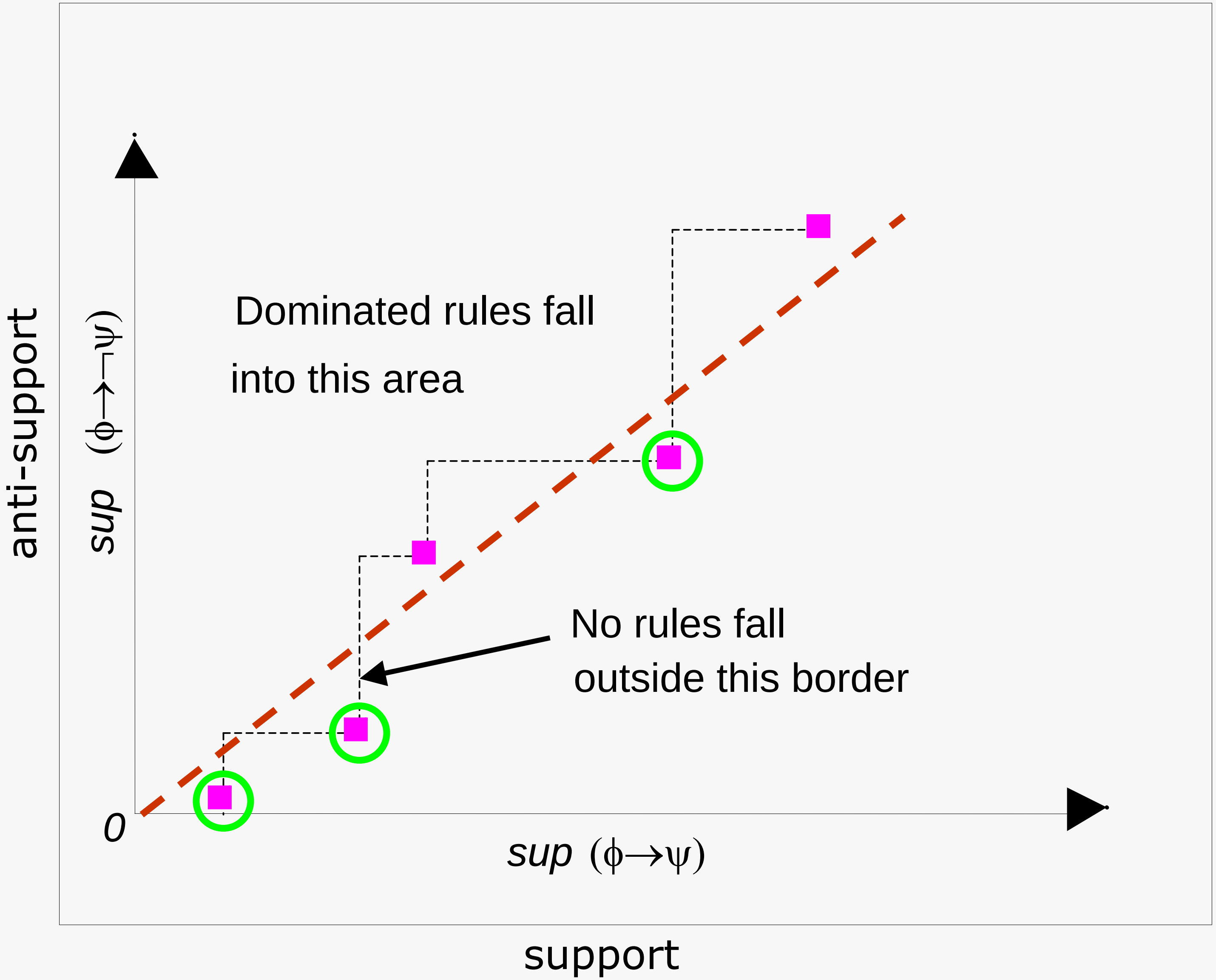
Support-anti-support Pareto border

- How to find rules optimal according to any confirmation measure with the property (M)?
- Theorem 8:
  When the value of support is held fixed, then $F(a, b, c, d)$ is anti-monotone (non-increasing) in anti-support
- Theorem 9:
  When the value of anti-support is held fixed, then $F(a, b, c, d)$ is monotone (non-decreasing) in support
- Anti-support is the number of examples which satisfy the premise of the rule but not its conclusion: $sup(\phi \rightarrow \neg\psi)$

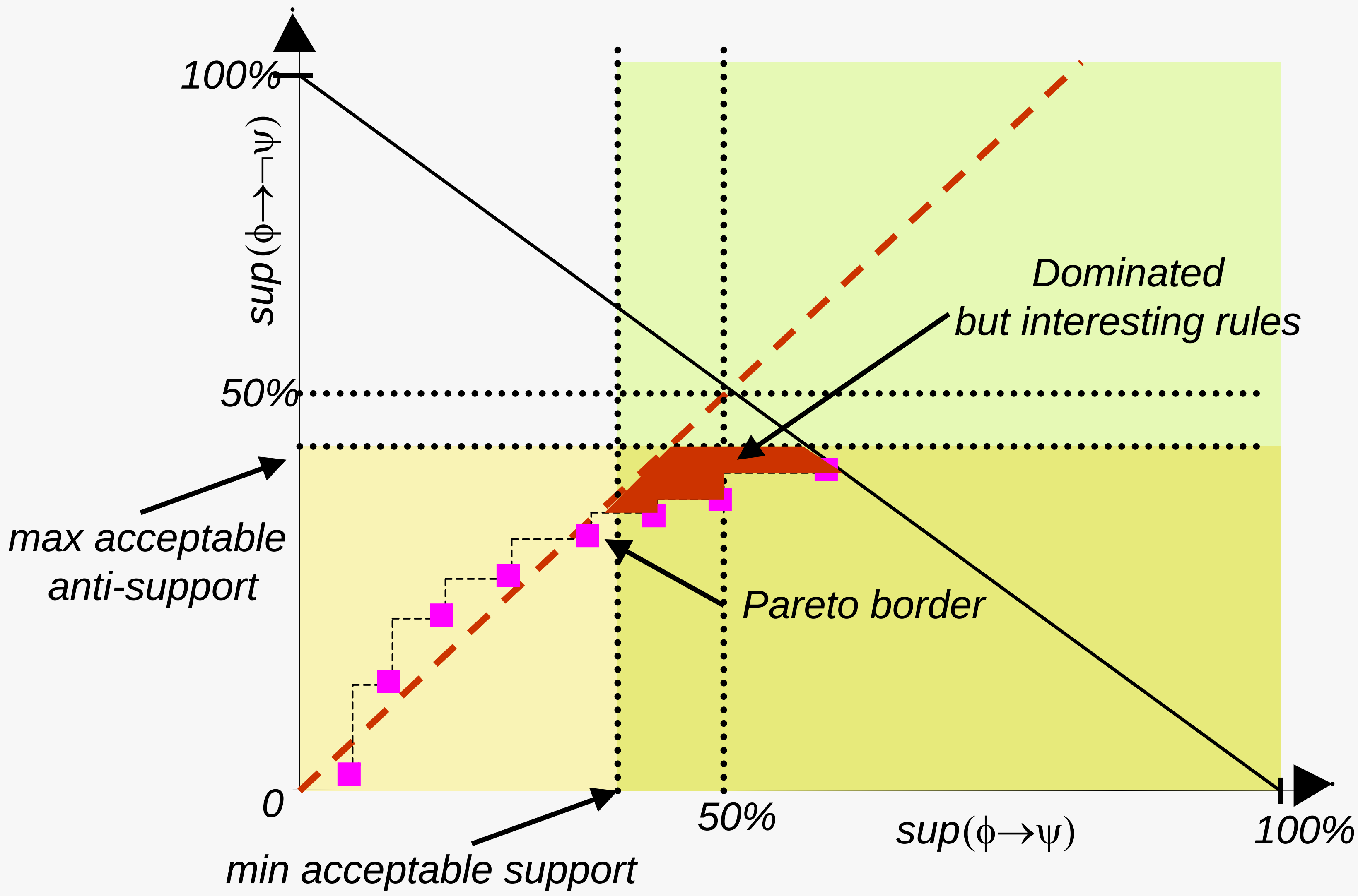
Support-anti-support Pareto border

- **Claim 2:**
  - The best rules according to any of the confirmation measures with the property of monotonicity (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

Support-anti-support Pareto border



Support-anti-support Pareto border



Conclusions

- Many attractiveness measures can be identified by mining the support-confidence Pareto border – very practical result
- The utility of confirmation measures outranks the utility of confidence
- Suggested new Pareto borders:
  - support-*f* Pareto border
  - support-*s* Pareto border
- Pareto border w.r.t. support and anti-support includes rules maximizing all confirmation measures with the property (M)

Conclusions

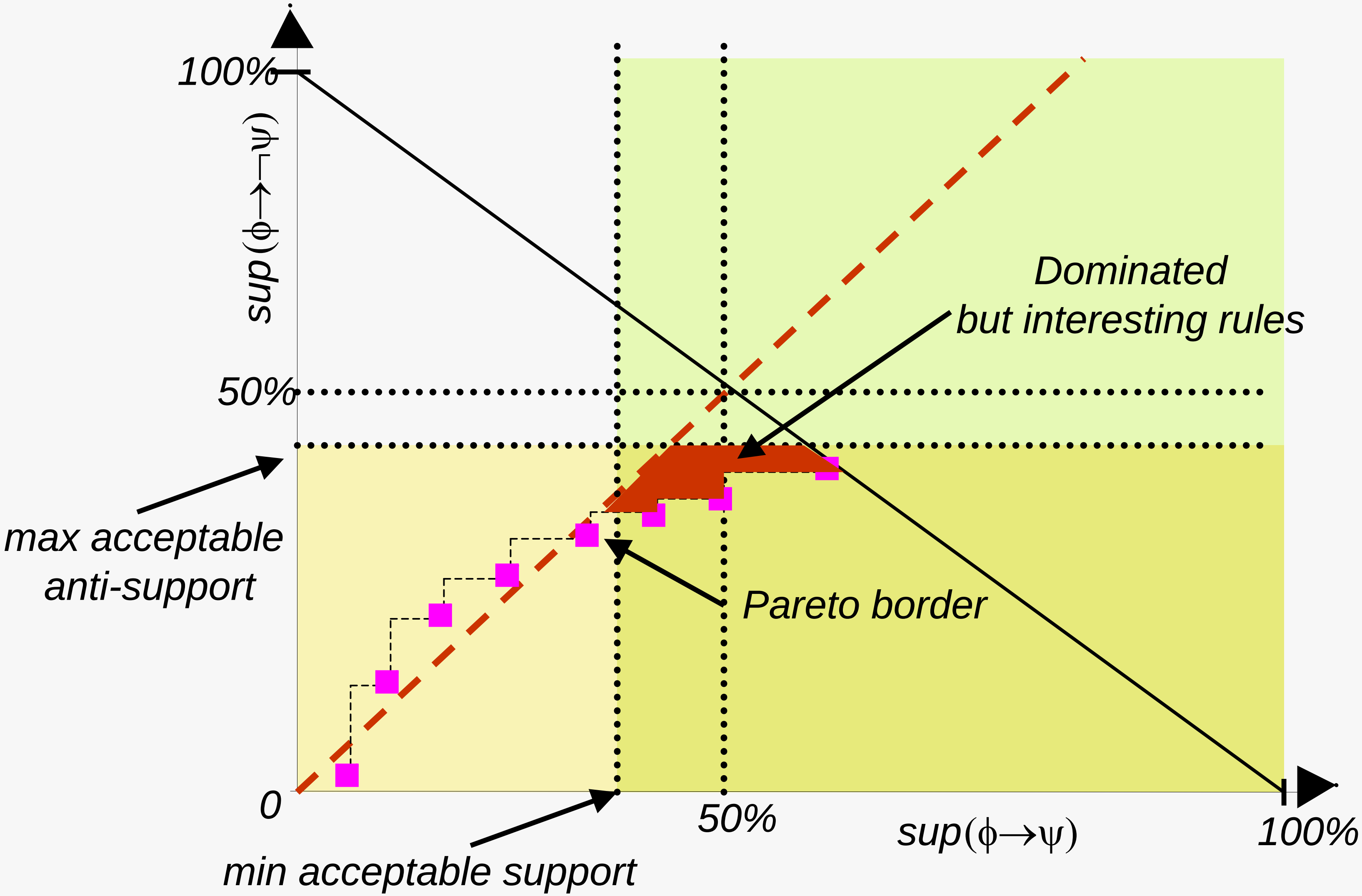
- Indeed, a rule $\phi \rightarrow \psi$ lying on the support-anti-support Pareto-optimal border is „maximally frequent" with respect to the pattern $\phi \wedge \psi$ and „minimally infrequent" with respect to the pattern $\phi \wedge \neg \psi$
- From an algorithmic viewpoint this is particularly useful because of the closure property of support and anti-support:

a) if an itemset is frequent, then all its subsets are also frequent,

b) if an itemset is infrequent, then all its supersets are also infrequent.

- Property a) means that support is downward closed, i.e. if an itemset has a required support, then all its subsets also have it.
- Property b) means that anti-support is upward closed, i.e. if an itemset has not a required support, then neither of its subsets has it.

Support-anti-support Pareto border



Thank you