Analysis of monotonicity properties of some rule interestingness measures

Izabela Szczęch

Poznań University of Technology

Roman Słowiński

Poznań University of Technology,

Institute for Systems Research, PAS, Poland

Salvatore Greco

University of Catania, Italy

KKNTPD 2007, Poznań

Presentation plan

- Introduction
- Basic quantitative characteristics of rules
- Properties of interestingness measures
- Results of the conducted analysis
- Application of the results
- Conclusions and lines of further investigation

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

need to analyze relationships between different 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 which measures have valuable and desired 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 **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.**

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, 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

- **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)**

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

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

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

where Θ is a fraction constant between 0 and 1

- **Dependency Factor** **(Pawlak 2002)**

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

Property M

- **Property M** **(Greco, Pawlak, Słowiński 2004)**
- **An interestingness 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=$sup(\neg\phi \rightarrow \neg\psi)$

Interpretation of the property M

- **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**

Property of hypothesis symmetry

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

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

- **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:** We draw cards from a standard deck.
  Let ϕ: the drawn card is *the seven of spades* and ψ: *the card is black.*
  ϕ is conclusive for ψ and negatively conclusive for $\neg\psi$

Aim and scope of the conducted analysis

Interestingness measure	Analyzed properties	
	Property M	Property HS
Rule Interest Function		
Gain		
Dependency Factor		

Results of the analysis with respect to property M

- ***Theorem:***

  Rule Interest Function has the property M

- ***Theorem:***

  Gain measure has the property M

- ***Theorem:***

  Dependency factor does not have the property M

Results of the analysis with respect to property M

- ***Theorem:*** **Rule Interest Function has the property M**

- ***Proof outline:***

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

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

To prove the monotonicity of *RI* wrt *a* we have to show that if *a* increases by Δ>0, then *RI* does not decrease i.e.

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

Analogous steps wrt *b*, *c* and *d*.

Results of the analysis with respect to property HS

- ***Theorem:***

  Rule Interest Function has the property HS

- ***Theorem:***

  Gain measure has the property HS iff $\Theta=1/2$

- ***Theorem:***

  Dependency factor does not have the property M

Application of the results

Support - Anti-support Pareto border

- ***Theorem:***

  For a set of rules with the same conclusion,

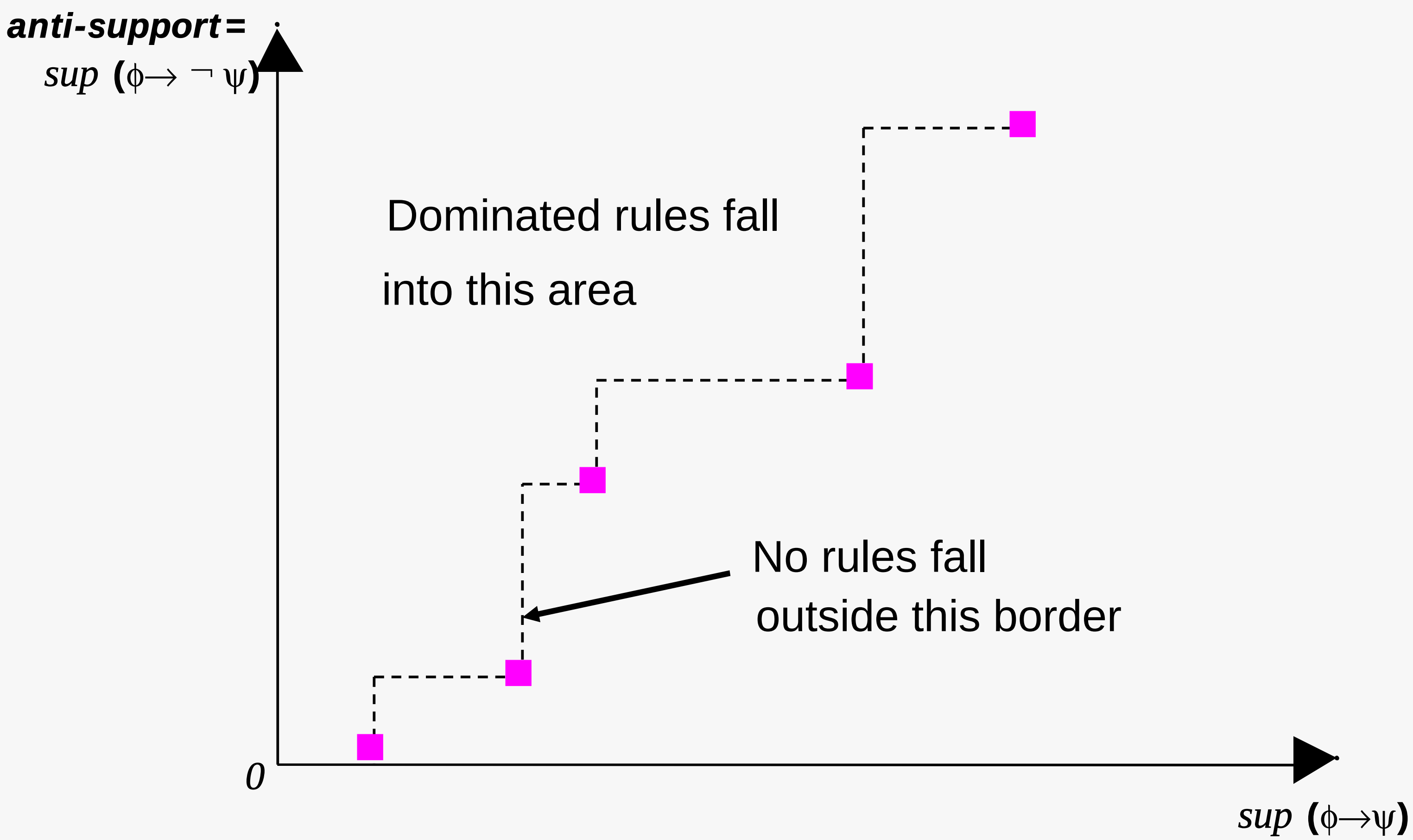
  due to (anti) monotonic dependencies between
  measures of support and anti-support on one hand
  and any interestingness measure with property M on the other hand

  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 (EAAI Journal, 2007)

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

Application of the result

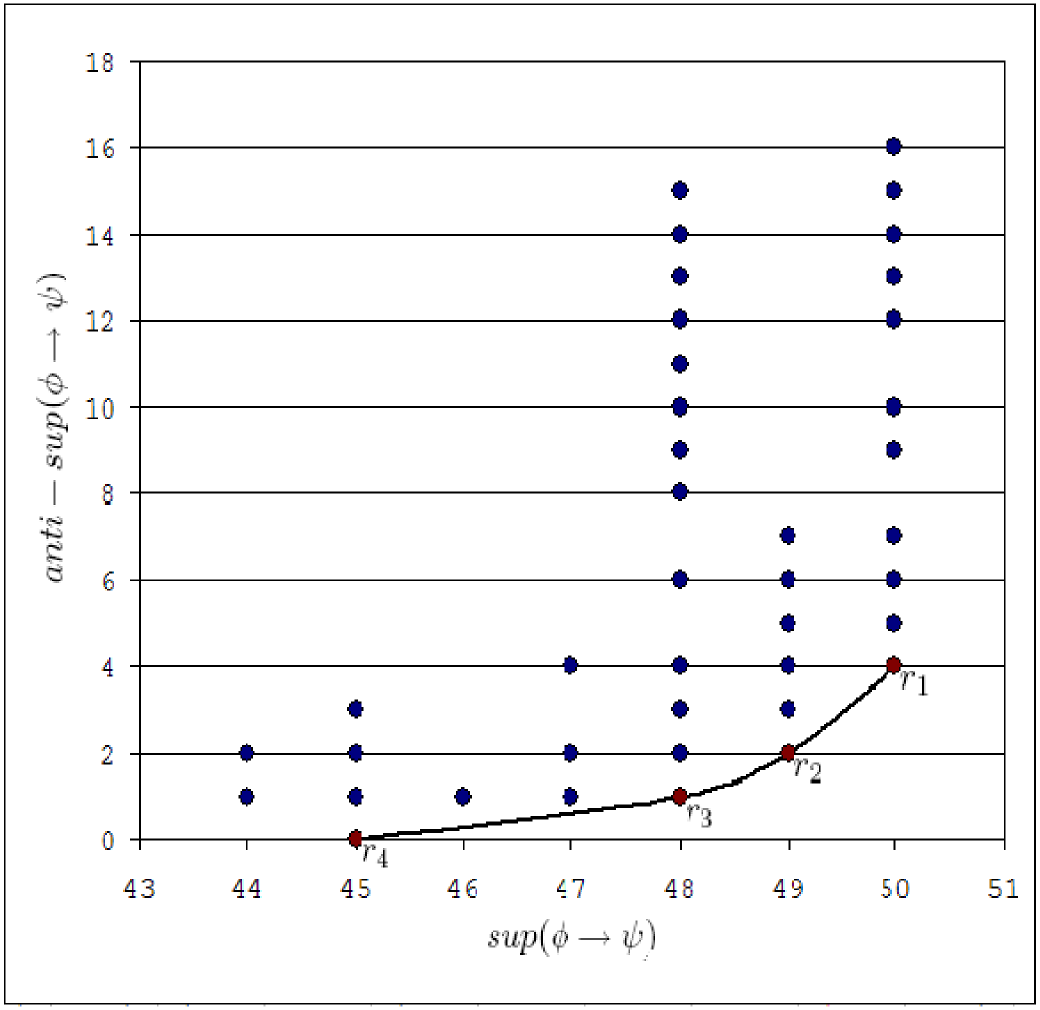
- **Since *RI* and *Gain* satisfy the property M we can conclude that rules optimal with respect to them will be found in the set of rules non-dominated according to support and anti-support.**
  (considering rules with the same conclusion)

- **Experiments illustrating the result:**

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

  Set of 85 rules with the same conclusion

Application of the result



Legend

dominated rules

Pareto-optimal border

r_1 represents rules optimal with respect to gain for $\Theta = 0.33$
and with respect to gain for $\Theta = 0.66$

r_2 represents rules optimal with respect to gain for $\Theta = 0.5$

r_3 represents rules optimal with respect to gain for $\Theta = 0.66$
and with respect to RI

r_4 represents rules optimal with respect to dependency factor

Conclusions

Conclusions

Interestingness measure	Properties	
	Property M	Property HS
Rule Interest Function	YES	YES
Gain	YES	iff $Q = 1/2$
Dependency Factor	NO	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 *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 *RI*, Gain, or any other measure with property M**
  - **rules optimal wrt to *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 is the same for any other measure with M**

Lines of further investigation

- Analysis of properties M and Hypothesis Symmetry with respect to other interestingness measures
- Analysis of other properties, eg. property of confirmation

Thank you!

Knowledge representation semantics – computational experiment*

- 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, Slowinski, Stefanowski 2001)

File	objects	atr+crit	classes	rules (alg)	consistency	length	coverage
Buses	76	0+8	3	266 (all)	$\geq$ 0.75	$\leq$ 3	$\geq$ 0.9
Nativity	342	0+33	2	64 (mc)	$\geq$ 0.75	no limit	no limit
Urology	500	18+9	3	186 (mc)	$\geq$ 0.96	no limit	no limit

Rule induction algorithms: „all" = all rules algorithm (DOMAPRIORI)

„mc" = minimal-cover algorithm (DOMLEM)

*by Iza Brzezińska