Properties of attractiveness measures for data mining – a survey

Izabela Szczęch

Poznań University of Technology

Przegląd własności miar oceny reguł

IDSS Poznań, 08.12.2009

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

Introduction - motivations

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

- each measure was proposed to capture different characteristics of rules
- 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

Introduction - motivations

Properties group the measures according to similarities in their characteristics

- objective vs. subjective properties,
- properties for „rule-measures” vs. properties for „itemset-measures”

need to analyze which properties are most desirable

Presentation plan

- Desirable properties of objective attractiveness measures
  - property of Bayesian confirmation
  - property M
  - symmetry properties
  - property Ex_1 of preserving extremes
- Critical survey on other properties in the literature
- Summary

Notation

- Patterns in form of rules are induced from a data table
- $S=\langle U, A\rangle$ – *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*, $C\cap D=\varnothing$

- *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 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.

Notation

- $a = sup(\phi \rightarrow \psi)$ is the number of objects in *U* satisfying both the premise ϕ and the conclusion ψ of a rule $\phi \rightarrow \psi$

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

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

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

- $a+c = sup(\phi)$, $a+b = sup(\psi)$, $b+d = sup(\neg\ \phi)$, $c+d = sup(\neg\psi)$, $|U| = a+b+c+d$

- A 2x2 contingency table

	ψ	$\neg\psi$	
ϕ	*a*	*c*	*a+c*
$\neg\ \phi$	*b*	*d*	*b+d*
	a+b	*c+d*	*U*

Property of Bayesian confirmation

- An attractiveness $c(\phi \rightarrow \psi)$ measure has the property of confirmation if it satisfies the following condition:

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

- 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 of Bayesian confirmation

- Under „the closed world assumption" adopted in inductive reasoning, and because *U* is a finite set, it is legitimate to estimate probabilities in terms of frequencies, e.g. $Pr(\psi) = \frac{a+b}{|U|}$

$$c(\phi \to \psi) \begin{cases} > 0 & if \ \frac{a}{a+c} > \frac{a+b}{|U|} \\ = 0 & if \ \frac{a}{a+c} = \frac{a+b}{|U|} \\ < 0 & if \ \frac{a}{a+c} < \frac{a+b}{|U|} \end{cases} \qquad \text{(BC)}$$

where: $a = sup(\phi \to \psi)$, $b = sup(\neg\phi \to \psi)$, $c = sup(\phi \to \neg\psi)$, $d = sup(\neg\phi \to \neg\psi)$, $|U| = a+b+c+d$

Rival Bayesian confirmation measures

- The condition

$$c(\phi \to \psi)\begin{cases} > 0 & if \quad \dfrac{a}{a+c} > \dfrac{a+b}{|U|} \\ = 0 & if \quad \dfrac{a}{a+c} = \dfrac{a+b}{|U|} \\ < 0 & if \quad \dfrac{a}{a+c} < \dfrac{a+b}{|U|} \end{cases} \qquad \text{(BC)}$$

does not put any constraint on the value to be assigned to confirmatory arguments (as long as they are positive) or disconfirmatory arguments (as long as they are negative)

- There are many alternative, non-equivalent measures of Bayesian confirmation with different scales

Rival Bayesian 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 popular confirmation measures there are:

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

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

1. non-decreasing with respect to *a* and
2. non-increasing with respect to *b* and
3. non-increasing with respect to *c* and
4. non-decreasing with respect to *d*.

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

* Greco, S., Pawlak, Z., Słowiński, R., 2004. Can Bayesian confirmation measures be useful for rough set decision rules? Engineering Applications of Artificial Intelligence, 17: 345-361.

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* (positive examples) we observe,

    the more credible becomes the rule

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

    the more *non-black ravens* (counter examples) we observe,

    the less credible becomes the rule

  - non-decreasing with respect to *d*

Interpretation of the property M

- A positive value of a confirmation measure means that the rule's conclusion ψ is satisfied more frequently when the premise ϕ is satisfied rather than when ϕ is not satisfied.
  From this viewpoint we can justify the relationship between I and *b*, *d* :
  - objects $\neg\phi\wedge\psi$ (i.e. objects represented by *b*) reflect the latter situation and therefore decrease the value of confirmation.
    Thus, measure I should be non-increasing with respect to *b.*
  - objects $\neg\phi\wedge\neg\psi$ (i.e. objects represented by *d*) decrease the frequency of ψ in the situations where ϕ is not satisfied, and therefore should increase the value of confirmation.
    Thus, measure I should be non-decreasing with respect to *d*

Properties of symmetry

- Properties of symmetry (Carnap 1962*, Eells & Fitelson 2002**):
  - Evidence symmetry (ES): $I(\phi \rightarrow \psi) = -I(\neg\phi \rightarrow \psi)$
  - Inversion symmetry (IS): $I(\phi \rightarrow \psi) = I(\psi \rightarrow \phi)$
  - Hypothesis symmetry (HS): $I(\phi \rightarrow \psi) = -I(\phi \rightarrow \neg\psi)$
  - Total symmetry (TS): $I(\phi \rightarrow \psi) = -I(\neg\phi \rightarrow \neg\psi)$

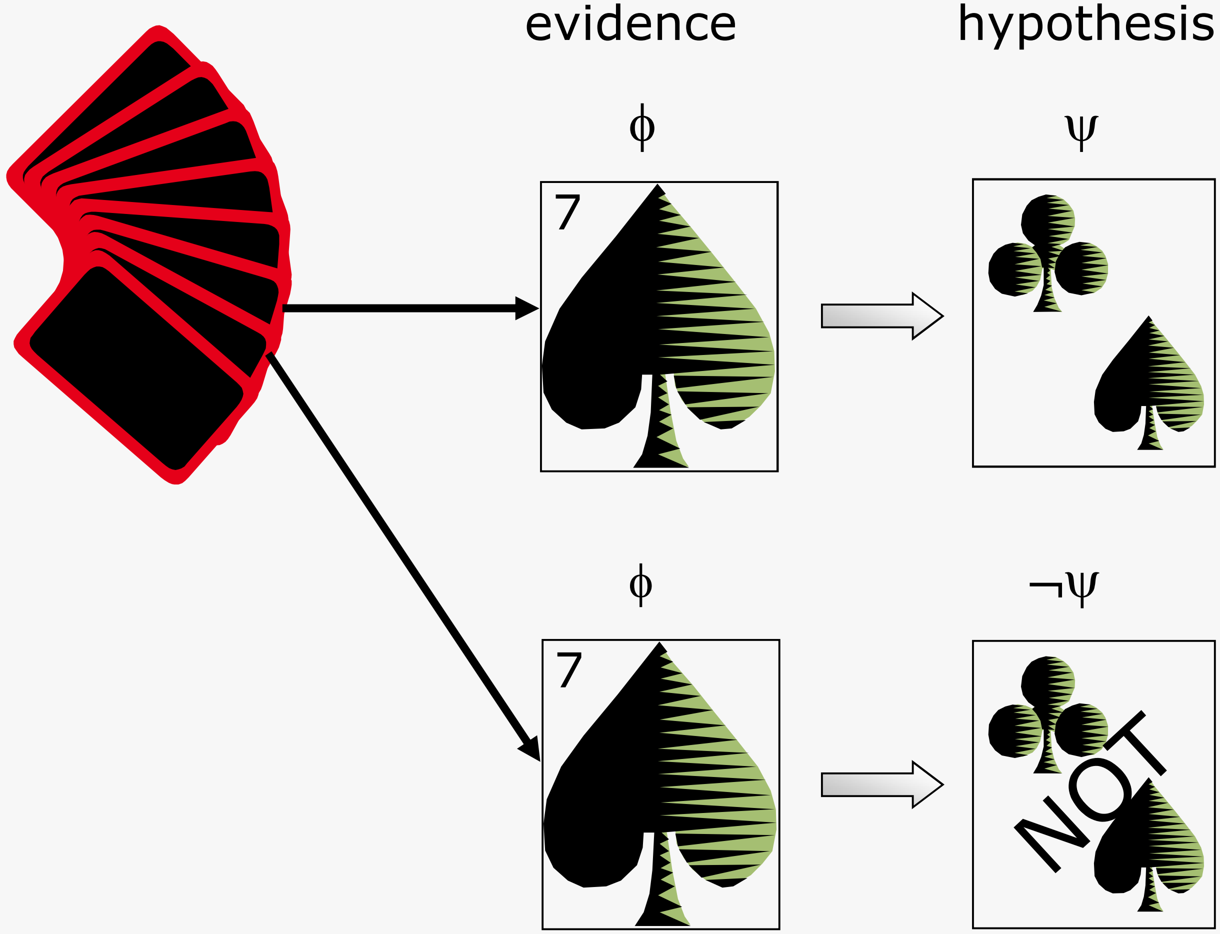
- Only hypothesis symmetry (HS) is desirable

  HS: the impact of ϕ on ψ should be of the same strength, but of the opposite sign, as the impact of ϕ on $\neg\psi$

* Carnap, R., 1962. Logical Foundations of Probability, Univ. of Chicago Press, Chicago.
** Eells, E., Fitelson, B., 2002. Symmetries and asymmetries in evidential support. Philosophical Studies, 107 (2): 129-142.

Hypothesis Symmetry (HS)

evidence — hypothesis („*the card is black*")



ϕ is conclusive for ψ

(„*the card is not black*")

ϕ is negatively conclusive for $\neg\psi$

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

Evidence Symmetry (ES)

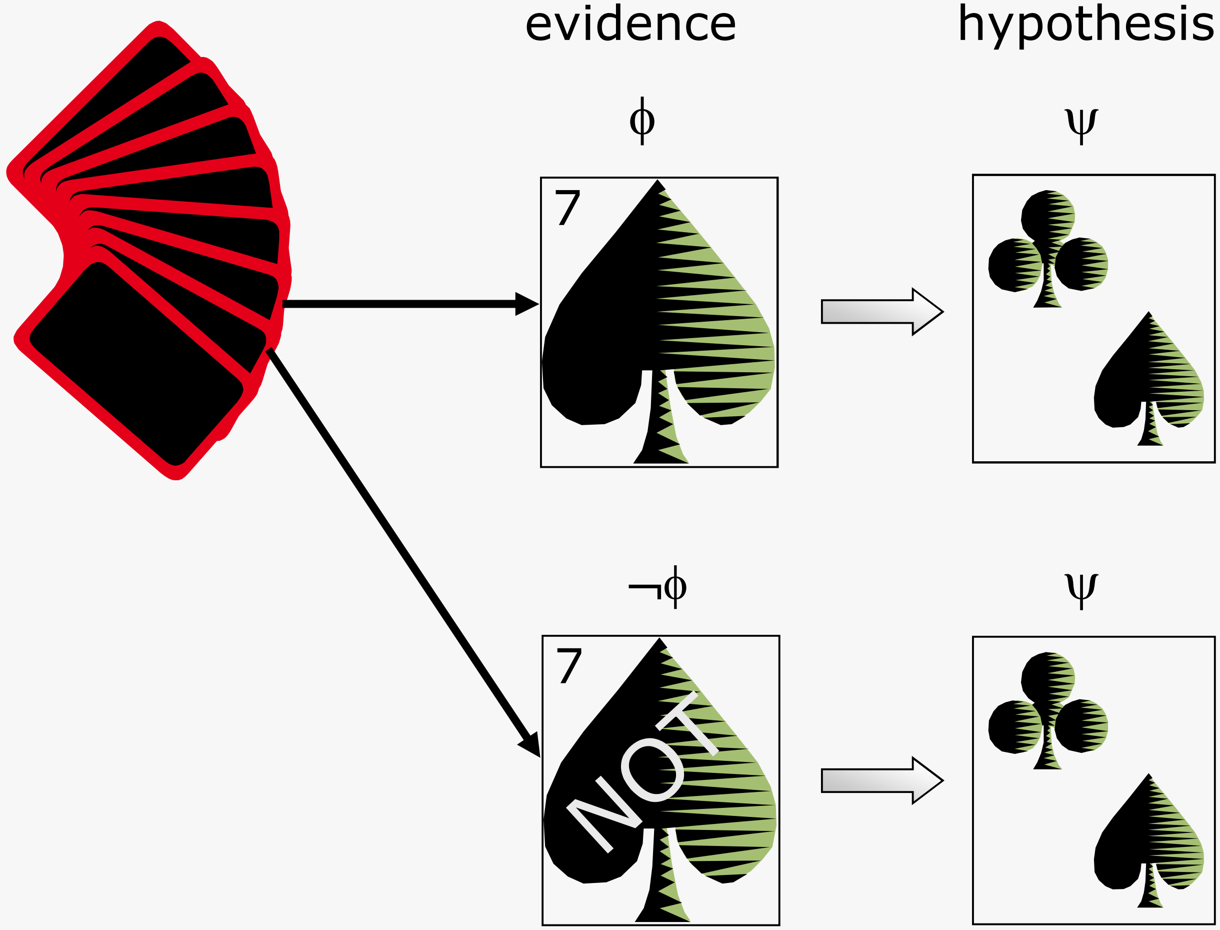
evidence — hypothesis („*the card is black*”)

ϕ — ψ

ϕ is conclusive for ψ

$\neg\phi$ — ψ

$\neg\phi$ is useless for ψ

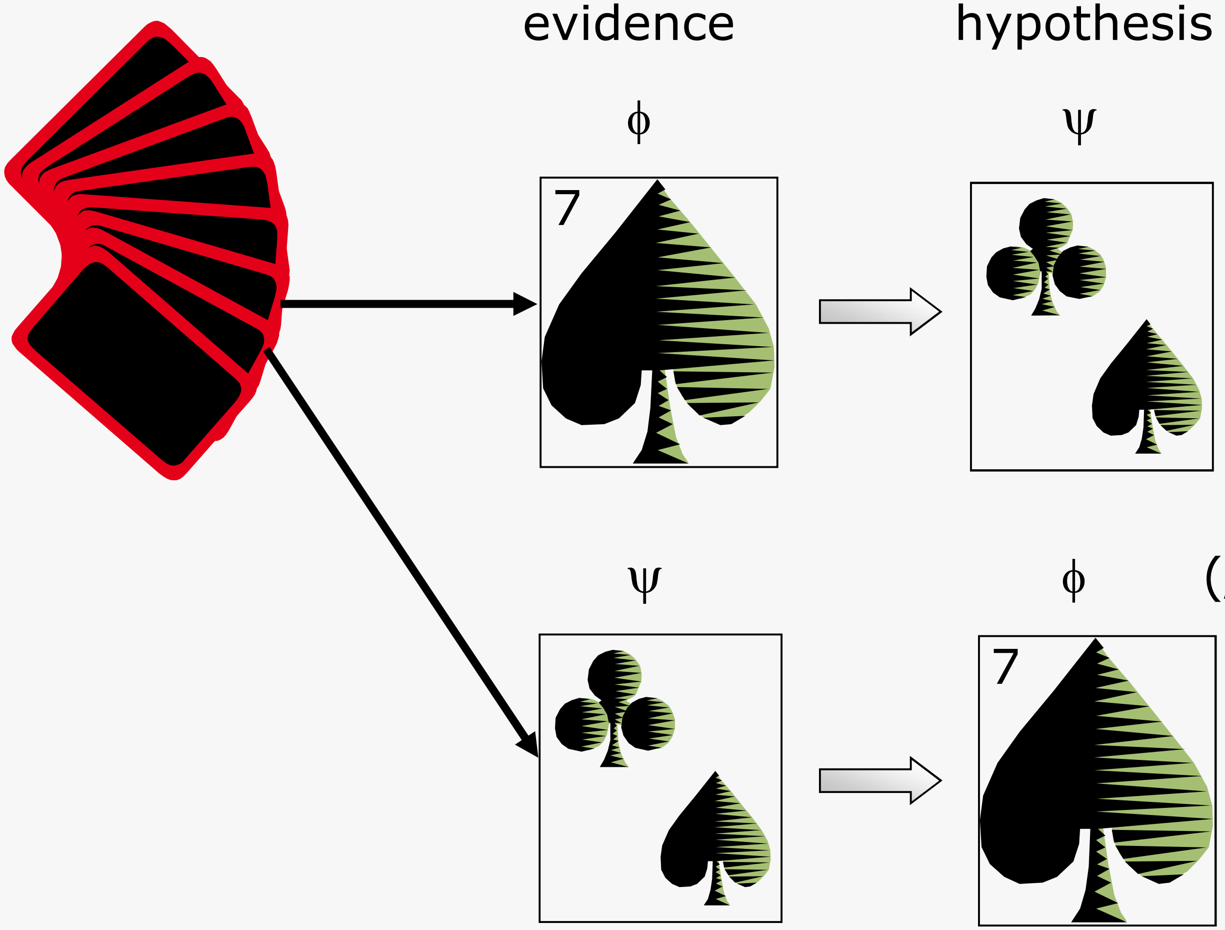


$$c(\phi \to \psi) = -c(\neg\phi \to \psi)$$

Inversion Symmetry (IS)

evidence — hypothesis („*the card is black*”)

ϕ — ψ



ϕ is conclusive for ψ

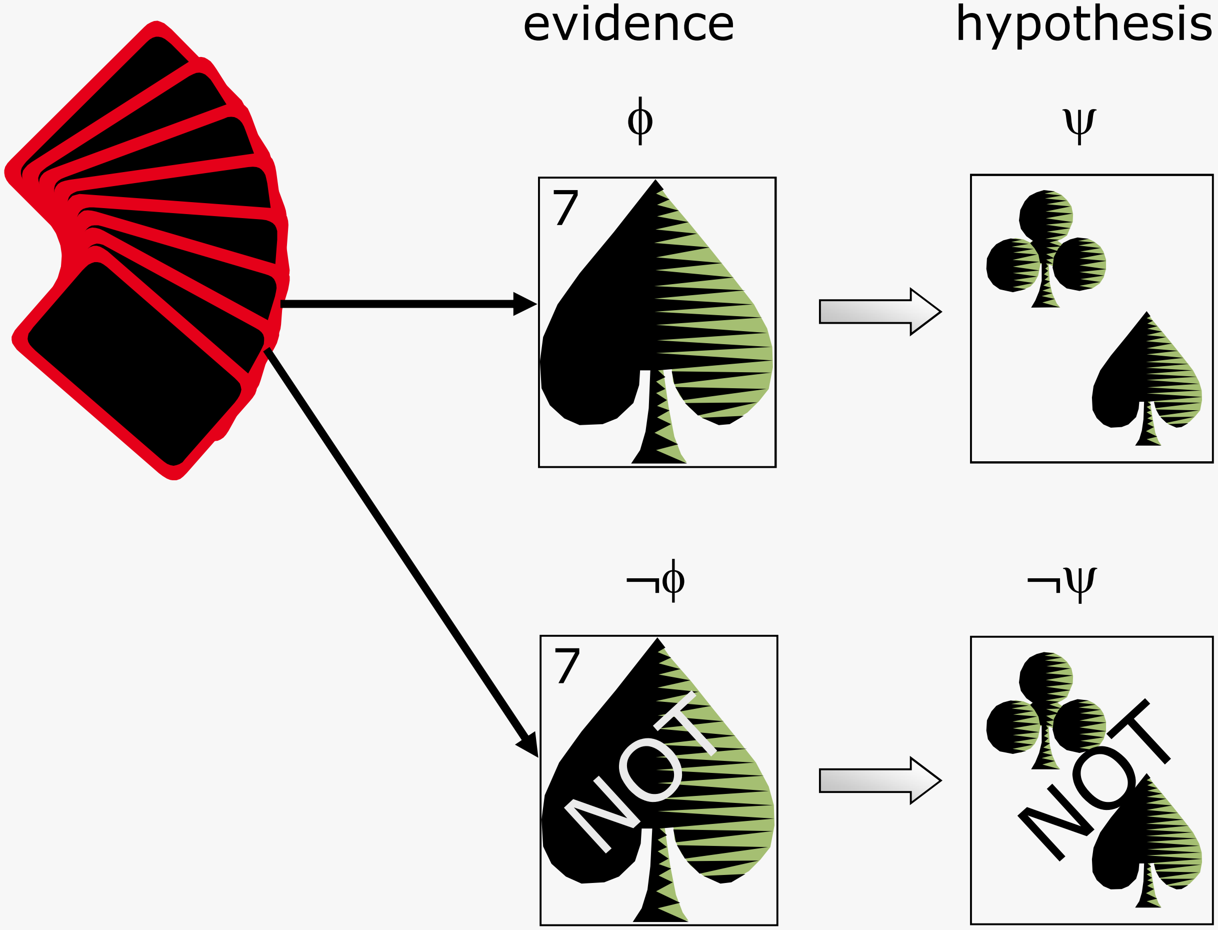
ψ — ϕ („*the card is 7 of spades*”)

ψ is less useful for ϕ than *vice versa*

$$c(\phi \to \psi) = c(\psi \to \phi)$$

Total Symmetry (TS)

evidence ϕ — hypothesis („*the card is black*") ψ



$\neg\phi$ — $\neg\psi$

ϕ is conclusive for ψ

$\neg\phi$ is not conclusive for $\neg\psi$

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

Property of preserving extremes (Ex_1)

- Crupi, Tentori and Gonzalez 2007* have considered the confirmation measures from the viewpoint of classical deductive logic introducing function *v* such that for any argument (ϕ,ψ):
  - *v* assigns it the same positive value (e.g., **1**) iff ϕ entails ψ, i.e. $\phi \mapsto \psi$,
  - an equivalent value of opposite sign (e.g., **-1**) iff ϕ entails the negation of ψ, i.e. $\phi \mapsto \neg\psi$, and
  - value **0**, otherwise.

* Crupi V., Tentori, K., Gonzalez, M., 2007. On Bayesian measures of evidential support: Theoretical and empirical issues. Philosophy of Science, 74, 229-252.

Property of preserving extremes (Ex_1)

- The relationship between the logical implication or refutation of ψ by ϕ, and the conditional probability of ψ subject to ϕ requires that any Bayesian confirmation measure $c(\phi \rightarrow \psi)$ agrees with $v(\phi,\psi)$ in the following sense:

(Ex_1): $$\textit{if}\ \ v(\phi_1 \rightarrow \psi_1) > v(\phi_2 \rightarrow \psi_2),\ \textit{then}\ \ c(\phi_1 \rightarrow \psi_1) > c(\phi_2 \rightarrow \psi_2).$$

1	0
1	-1
0	-1

Property of preserving extremes (Ex_1)

(Ex_1): $if \;\; v(\phi_1, \psi_1) > v(\phi_2, \psi_2),\; then \;\; c(\phi_1 \rightarrow \psi_1) > c(\phi_2 \rightarrow \psi_2)\,.$

Ex_1 guarantees that

if x is seven of spades then x is black

- any conclusively confirmatory argument ($\phi \mapsto \psi$) is assigned a higher value of $c(\phi \rightarrow \psi)$ than any argument which is *not conclusively confirmatory*,

if x is black then x is seven of spades

if x is seven of spades then x is red

- and any conclusively disconfirmatory argument ($\phi \mapsto \neg\psi$) is assigned a lower value of $c(\phi \rightarrow \psi)$ than any argument which is *not conclusively disconfirmatory*

if x is black then x is seven of spades

Four desirable properties

- Desirable properties of objective attractiveness measures
  - property of Bayesian confirmation
  - property M
  - property of hypothesis symmetry
  - property Ex_1 of preserving extremes

  - property of evidence symmetry, inversion symmetry and total symmetry are **undesirable**

Principles (properties) of Piatetsky-Shapiro

- Piatetsky-Shapiro* proposed three principles that shoud be obeyed by any objective measure, F:

  (P_1) F=0 if ϕ and ψ are statistically independent,
  i.e. $\Pr(\phi \wedge \psi) = \Pr(\phi)\Pr(\psi)$,

  (P_2) F monotonically increases with $\Pr(\phi \wedge \psi)$
  when $\Pr(\phi)$, and $\Pr(\psi)$ remain the same,

  (P_3) F monotonically decreases with $\Pr(\phi)$ (or $\Pr(\psi)$)
  when $\Pr(\phi \wedge \psi)$ and $\Pr(\psi)$ (or $\Pr(\phi)$) remain the same.

* Piatetsky-Shapiro, G., 1991. Discovery, analysis, and presentation of strong rules. Chapter 12, in: Knowledge Discovery in Databases, AAAI/MIT Press.

Principles (properties) of Piatetsky-Shapiro

- P_1: $F=0$ if ϕ and ψ are statistically independent,

  i.e. $Pr(\phi \wedge \psi) = Pr(\phi)\, Pr(\psi)$

- P_1 agrees with the „middle” condition of property of confirmation

$$c(\phi \to \psi) \begin{cases} > 0 & \text{if } \dfrac{a}{a+c} > \dfrac{a+b}{|U|} \\ = 0 & \text{if } \dfrac{a}{a+c} = \dfrac{a+b}{|U|} \\ < 0 & \text{if } \dfrac{a}{a+c} < \dfrac{a+b}{|U|} \end{cases} \qquad \text{(BC)}$$

Notation:
$a = sup(\phi \to \psi)$
$b = sup(\neg\phi \to \psi)$
$c = sup(\phi \to \neg\psi)$
$d = sup(\neg\phi \to \neg\psi)$

$$Pr(\phi \wedge \psi) = \frac{a}{|U|}, \; Pr(\phi) = \frac{a+c}{|U|}, \; Pr(\psi) = \frac{a+b}{|U|}$$

$$Pr(\phi \wedge \psi) = Pr(\phi)\, Pr(\psi) \to aU^2 = |U|(a+c)(a+b)$$

$$\frac{a}{a+c} = \frac{a+b}{|U|}$$

Principles (properties) of Piatetsky-Shapiro

- P_2: *F* monotonically increases with $Pr(\phi \wedge \psi)$
    - when $Pr(\phi)$ remains the same
    - and when $Pr(\psi)$ remains the same

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

$Pr(\phi \wedge \psi)$ increases while $Pr(\phi)$ remains unchanged when some observations $(\phi \wedge \neg\psi)$ change into $(\phi \wedge \psi)$, i.e. when *c* decreases and *a* increases.

Conclusion: P_2 agrees with III and I condition of property M

$Pr(\phi \wedge \psi)$ increases while $Pr(\psi)$ remains unchanged when some observations $(\neg\phi \wedge \psi)$ change into $(\phi \wedge \psi)$, i.e. when *b* decreases and *a* increases.

Conclusion: P_2 agrees with II and I condition of property M

Principles (properties) of Piatetsky-Shapiro

- P_3: *F* monotonically decreases with $Pr(\phi)$
  when $Pr(\phi \wedge \psi)$ and $Pr(\psi)$ remain the same,
- P_3: *F* monotonically decreases with $Pr(\psi)$
  when $Pr(\phi \wedge \psi)$ and $Pr(\phi)$ remain the same,

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

$Pr(\phi)$ increases while $Pr(\phi \wedge \psi)$ and $Pr(\psi)$ remain unchanged when the number of observations $(\phi \wedge \neg\psi)$ increases, i.e. when *c* increases.

Conclusion: P_3 agrees with III condition of property M

$Pr(\psi)$ increases while $Pr(\phi \wedge \psi)$ and $Pr(\phi)$ remain unchanged when the number of observations $(\neg \phi \wedge \psi)$ increases, i.e. when *b* increases.

Conclusion: P_2 agrees with II condition of property M

Properties of Tan et al.

	ψ	$\neg\psi$	
ϕ	*a*	*c*	*a+c*
$\neg\phi$	*b*	*d*	*b+d*
	a+b	*c+d*	*U*

- Tan et al.* proposed properties based on operations for 2x2 contingency tables:

  (O_1) *F* should be symmetric under variable permutation,

  (O_2) *F* should be the same when we scale any row or column by a positive factor,

  (O_3) *F* should become –*F* if either the rows or columns are permuted, i.e. swapping either the rows or columns in the contingency table makes interestingness values change their signs,

  (O_4) *F* should have no relationship with the count of the records that do not contain ϕ and ψ.

* Tan, P.-N., Kumar, V., Srivastava, J., 2002. Selecting the right interestingness measure for association patterns. In: Proc. of the 8th international Conf. on Knowledge Discovery and Data Mining (KDD 2002). Edmonton, Canada, pp.32-41.

Properties of Tan et al.

- O_1: *F* should be symmetric under variable permutation,
  i.e. rules $\phi \rightarrow \psi$ and $\psi \rightarrow \phi$ should have the same interestingness value.

  $$F(\phi \rightarrow \psi) = F(\psi \rightarrow \phi)$$

  O_1 is another formulation of inversion symmetry and should be considered as an **undesirable** property

  $$c(\text{Jack} \rightarrow \text{face}) \neq c(\text{face} \rightarrow \text{Jack})$$

- O_2: *F* should be the same when we scale any row or column by a positive factor

  Scaling of rows or columns effects the values of *a*, *b*, *c* or *d*. Any change of those values should be reflected by a measure.
  O_2 is **undesirable**.

Properties of Tan et al.

- O_3: *F should become –F if either the rows or columns are permuted,* i.e. swapping either the rows or columns in the contingency table makes interestingness values change their signs.

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

- $F(\phi\rightarrow\psi) = -F(\phi\rightarrow\neg\psi)$ is a hypothesis symmetry (**desirable**)

$$c(\text{Jack}\rightarrow\text{face}) =- c(\text{Jack}\rightarrow \neg\text{face})$$

$$100\% = -(-100\%)$$

- $F(\phi\rightarrow\psi) = -F(\neg\phi\rightarrow\psi)$ is an evidence symmetry (**undesirable**)

$$c(\text{Jack}\rightarrow\text{face}) \neq - c(\neg\ \text{Jack}\rightarrow \text{face})$$

$$100\% \neq -{}^{(52-4)}/_{12} *100\%$$

Properties of Tan et al.

- O_4: *F* should have no relationship with the count of the records that do not contain ϕ and ψ.

The number of observations that do not contain ϕ or ψ effects whether ψ is satisfied more frequently when the premise ϕ is satisfied rather than when ϕ is not satisfied. Thus, the count of records that do not contain ϕ and ψ should have a relationship with *F*.
O_4 is **undesirable**.

Properties of Lenca et al.

- Lenca et al.* proposed five properties to evaluate association rules:

  (Q_1) *F* is constant if there is no counterexamples to the rule,

  (Q_2) *F* decreases with $Pr(\phi \wedge \neg \psi)$ in a linear, concave or convex fashion around 0+,

  (Q_3) *F* increases as the total number of records increases assuming that $Pr(\phi)$, $Pr(\psi)$ and $Pr(\phi \wedge \psi)$ are held constant,

  (Q_4) The threshold is easy to fix,

  (Q_5) The semantics of the measure are easy to express.

* Lenca, Ph., Meyer, P., Vaillant, B., Lallich, S., 2008. On selecting interestingness measures for association rules: User oriented description and multiple criteria decision aid. European Journal of Operational Research, Vol. 184, Issue 2, 610-626

Properties of Lenca et al.

- Q_1: *F* is constant if there is no counterexamples to the rule
  i.e. rules with a confidence of 1 should have the same interestingness value, regardless of the support.

  It is desirable that the value of *F* is not only constant but maximal, which agrees with property Ex_1.

- Q_2: *F* decreases with $Pr(\phi \wedge \neg \psi)$ in a linear, concave or convex fashion around 0+

  Q_2 agrees with III condition of property M

* Lenca, Ph., Meyer, P., Vaillant, B., Lallich, S., 2008. On selecting interestingness measures for association rules: User oriented description and multiple criteria decision aid. European Journal of Operational Research, Vol. 184, Issue 2, 610-626

Properties of Lenca et al.

- Q_3: *F* increases as the total number of records increases assuming that Pr (ϕ), Pr(ψ) and Pr($\phi \wedge \psi$) are held constant,

$|U|$ increases while Pr(ϕ), Pr(ψ) and Pr($\phi \wedge \psi$) remain unchanged when new observations ($\neg\phi \wedge \neg\psi$) are added to the dataset, i.e. *d* increases.

Conclusion: Q_3 agrees with IV condition of property M

- Q_4: The threshold is easy to fix,

  Q_5: The semantics of the measure are easy to express.

  Q_4 and Q_5 are subjective properties.

* Lenca, Ph., Meyer, P., Vaillant, B., Lallich, S., 2008. On selecting interestingness measures for association rules: User oriented description and multiple criteria decision aid. European Journal of Operational Research, Vol. 184, Issue 2, 610-626

Summary

- Desirable properties of objective attractiveness measures
  - property of Bayesian confirmation
  - property M
  - property of hypothesis symmetry
  - property Ex_1 of preserving extremes

- Sets of properties proposed by Piatetsky-Shapiro, Tan et al., and Lenca et al. has been presented and commented showing which of them are desirable and which are concordant with the above properties.

Thank you!