Alternative normalization schemas for Bayesian confirmation measures

Salvatore Greco

University of Catania, Italy

Roman Słowiński

Poznań University of Technology,
Institute for Systems Research, PAS, Poland

Izabela Szczęch

Poznań University of Technology, Poland

Presentation plan

- Introduction
- Desired properties of interestingness measures
  - Property of Bayesian confirmation
  - Property Ex_1 of preservation of extremes
- Alternative normalization schemas for confirmation measures
- Results of applying normalization schemas to measures
- 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, measures of Bayesian confirmation)

Introduction - motivations

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

- there is no evidence which measure(s) is the best
- 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

In this work we focus on property of Bayesian confirmation and property Ex_1 of preservation of extremes

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*=∅**

- ***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*)

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 – used notation

- **For a rule $\phi \to \psi$, we shall use the following notation:**
  - $\boldsymbol{a}=sup(\phi\to\psi)$ **is the number of objects in the dataset $\boldsymbol{U}$ satisfying both the premise and the conclusion of the rule,**
  - $\boldsymbol{b}=sup(\neg\phi\to\psi)$ **,**
  - $\boldsymbol{c}=sup\ (\phi\to\neg\psi)$**,**
  - $\boldsymbol{d}=sup(\neg\phi\to\neg\psi)$.
- **Moreover, the following relations occur:**
  $\boldsymbol{a+c}=sup(\phi)$, $\boldsymbol{a+b}=sup(\psi)$, $\boldsymbol{b+d}=sup(\neg\phi)$, $\boldsymbol{c+d}=sup(\neg\psi)$,
  $\boldsymbol{a+b+c+d}=|\boldsymbol{U}|$
- **A 2x2 contingency table**

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

Desired properties
of interestingness measures

Property of Bayesian confirmation

- An attractiveness measure $c = (\phi \rightarrow \psi)$, has the property of Bayesian confirmation (i.e. is a confirmation measure) if is satisfies the following condition:

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

- 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 the following way:** $Pr(\psi) = \frac{a+b}{|U|}$

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

$$\Longrightarrow$$

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

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

- **The condition 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**

Rival Bayesian confirmation measures

- **Notation:** ***a***=$sup(\phi\to\psi)$, ***b***=$sup(\neg\phi\to\psi)$, ***c***=$sup(\phi\to\neg\psi)$, ***d***=$sup(\neg\phi\to\neg\psi)$
- **Among popular confirmation measures there are:**

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

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

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

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

$$C(\phi\to\psi)=\frac{a}{|U|}-\frac{(a+c)(a+b)}{|U|^2}$$ **(Carnap 1950/1962)**

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

$$G(\phi\to\psi)=1-\frac{c|U|}{(a+c)(c+d)}$$ **(Rips 2001)**

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. ϕ a ψ,**
  - **an equivalent value of opposite sign (e.g., -1) iff ϕ entails the negation of ψ, i.e. ϕ a $\neg\psi$, 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*(ϕ,ψ) 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): $if \;\; v(\phi_1, \psi_1) > v(\phi_2, \psi_2), \; then \;\; c(\phi_1 \rightarrow \psi_1) > c(\phi_2 \rightarrow \psi_2)$.

if x is seven of spades then x is black

Ex_1 guarantees that

- **any conclusively confirmatory argument (ϕ a ψ) 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 (ϕ a $\neg\psi$) 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

Crupi′s et al. normalization of measures

- Crupi et al. have proved that **neither of the above mentioned confirmation measures satisfies property (Ex$_1$)**
- However, their further analysis has unveiled a **normalization approach that makes those measures fulfill property (Ex$_1$)**

- The approach of Crupi et al. 2007 normalizes the chosen confirmation measures by dividing them by:
  - the **maximum** value they obtain in case of **confirmation**, and
  - the absolute **minimum** value they obtain in case of **disconfirmation**.

- This way, we obtain interestingness measures that distinguish between two completely different situations: of confirmation and disconfirmation

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.**

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

Alternative normalization schemas

Alternative approaches to normalization

- There are many approaches to understanding the meaning of the maximum and minimum of confirmation or disconfirmation, which eventually lead to different normalizations.

- We propose three alternative approaches to normalization:
  - an approach inspired by Nicod
  - Bayesian approach
  - likelihoodist approach

- Each of those approaches considers the concept of confirmation from different perspectives

Approach inspired by Nicod

- **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.**

Approach inspired by Nicod

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

  a$= sup(\phi \rightarrow \psi)$ **– the no. 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 confirmation 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, $a' = a + c$)

Approach inspired by Nicod

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

  a=$sup(\phi \to \psi)$ – the no. of objects in *U* which are black ravens

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

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

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

- **The minimal disconfirmation 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, $c'=a+c$)

Bayesian approach

- „Bayesian" approach assumes that:
  - 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.

Bayesian approach

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

- **The maximal confirmation of the rule will be obtained when:**
  - **all *black non-ravens* change into *black ravens* (i.e. $a'=a+b$ and $b'=0$),**
  - **and all *non-black ravens* change into *non-black non-ravens* (i.e. $d'=c+d$ and $c'=0$).**
- **when there are no *black non-ravens* (i.e. $b'=0$), the hypothesis of being *black* is more frequent with the premise of being a *raven* rather than with $\neg$premise of being a *non raven* (i.e. premise confirms conclusion)**
- **when there are no *non-black ravens* (i.e. $c'=0$), the $\neg$hypothesis of being *non-black* is disconfirmed as it is more frequent with the $\neg$premise of being a *non-raven* rather than with the premise of being a *raven*. Disconfirmation of $\neg$hypothesis is desirable as it results in confirmation of the hypothesis.**

Likelihoodist approach

- „Likelihoodist" approach assumes that:
  - 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.

Likelihoodist approach

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

- **The maximal confirmation of the rule will be obtained when:**
  - **all *non-black ravens* change into *black ravens* (i.e. $a'=a+c$ and $c'=0$),**
  - **and all *black non-ravens* change into *non-black non-ravens* (i.e. $d'=b+d$ and $b'=0$).**
- **when there are no *non-black ravens* (i.e. $c'=0$), the evidence of being *a raven* is more frequent with the hypothesis of being *black* rather than with $\neg$ hypothesis of being *non black* (i.e. premise confirms conclusion)**
- **when there are no *black non-ravens* (i.e. $b'=0$), the $\neg$evidence of being a *non-raven* is more frequent with the $\neg$hypothesis of being *non-black* rather than with the hypothesis of being *black*. Thus we can conclude that hypothesis is disconfirmed by the $\neg$premise and as a result of that the hypothesis is confirmed by the premise.**

Alternative normalization schemas - summary

Nicod's approach		Bayesian approach		Likelihoodist approach		Crupi's et al. approach	
Max	Min	Max	Min	Max	Min	Max	Min
$a'=a+c$	$a'=0$	$a'=a+b$	$a'=0$	$a'=a+c$	$a'=0$	$a'=a+c$	$a'=0$
$b'=b$	$b'=b$	$b'=0$	$b'=a+b$	$b'=0$	$b'=b+d$	$b'=b-c$	$b'=a+b$
$c'=0$	$c'=a+c$	$c'=0$	$c'=c+d$	$c'=0$	$c'=a+c$	$c'=0$	$c'=a+c$
$d'=d$	$d'=d$	$d'=c+d$	$d'=0$	$d'=b+d$	$d'=0$	$d'=c+d$	$d'=d-a$

Results of applying normalization schemas to measures

Results of applying alternative normalization schemas

Original confirmation measure	Nicod's norm.	Bayesian norm.	Likelihoodist norm.	Crupi et al. norm.
$D(\phi \to \psi) = \frac{ad-bc}{\lvert U \rvert (a+c)}$	$D_{N^+} = \frac{ad-bc}{d(a+c)}$	G	S	G
	$D_{N^-} = \frac{ad-bc}{b(a+c)}$	R	S	R
$S(\phi \to \psi) = \frac{ad-bc}{(a+c)(b+d)}$	D_{N+}	S	S	G
	D_{N^-}	S	S	R
$M(\phi \to \psi) = \frac{ad-bc}{\lvert U \rvert (a+b)}$	$M_{N^+} = \frac{(ad-bc)(a+b+c)}{d(a+b)(a+c)}$	N	$M_{L^+} = \frac{ad-bc}{(a+b)(b+d)}$	G
	R	N	R	R
$N(\phi \to \psi) = \frac{ad-bc}{(a+b)(c+d)}$	$N_{N^+} = \frac{(ad-bc)(a+b+c)}{(a+b)(c+d)(a+c)}$	N	N	G
	$N_{N^-} = \frac{(ad-bc)(a+c+d)}{(a+b)(c+d)(a+c)}$	N	N	R
$C(\phi \to \psi) = \frac{ad-bc}{\lvert U \rvert^2}$	D_{N+}	N	S	G
	D_{N^-}	N	S	R
$R(\phi \to \psi) = \frac{ad-bc}{(a+b)(a+c)}$	M_{N^+}	G	M_{L^+}	G
	R	R	R	R
$G(\phi \to \psi) = \frac{ad-bc}{(a+c)(c+d)}$	G	G	G	G
	$G_{N^-} = \frac{(ad-bc)(a+c+d)}{(a+c)(c+d)b}$	R	$G_{L^-} = \frac{ad-bc}{(c+d)(b+d)}$	R

Results of applying alternative normalization schemas

- Since the normalization of Crupi et al. was introduced as a tool for transforming the measures so they would satisfy the property Ex_1, we have analyzed the results of different normalizations of measures *D*, *S*, *M*, *N*, *C*, *R*, *G* from the view point of this property.

- Theorem 1:

  All measures *D*, *S*, *M*, *N*, *C*, *R*, *G* normalized using approach inspired by Nicod or approach of Crupi et al. satisfy property Ex_1. Moreover normalization using Bayesian approach gives measures satisfying Ex_1 only in case of measure *D*, *R* and *G*, whereas using likelihoodist approach, Ex_1 does not hold for any of the considered measures.

Results of applying alternative normalization schemas

- **The new measures obtained during normalization inspired by Nicod can be regarded as alternative ones with respect to measure *Z* advocated by Crupi et al. [4].**
- **D_N, S_N, M_N, N_N, C_N, R_N, and G_N are as valuable as *Z* in terms of possession of Ex_1 and, generally, produce different rankings on rules than *Z*.**
- **Theorem 2:**

  Measures D_N, S_N, M_N, N_N, C_N, R_N, and G_N

  (resulting from application of normalization inspired by Nicod)

  are ordinally non-equivalent to measure Z.
- **It is an important result widening the group of non-equivalent measures satisfying property Ex_1.**

Conclusions

Conclusions

- **Properties explain how the measures behave in certain situations and thus, group them helping the user choose the measure relevant for his expectations**
- **We have focused on possession of property Ex_1 in a group of popular confirmation measures**
- **We have proposed and analyzed three alternative approaches to understanding the meaning of confirmation, and eventually resulting in different normalization schemas:**
  - **an approach inspired by Nicod**
  - **Bayesian approach**
  - **likelihoodist approach**

Conclusions

- **We have proved that the approach inspired by Nicod (as well as approach of Crupi et al.) gives normalized measure satisfying property Ex_1 in case of all of the considered measures.**
- **Moreover, we have proved that measures obtained through those normalizations are ordinally non-equivalent.**

  Thus, we have extended the group of measures possessing valuable property Ex_1.

Thank you!

Bayesian approach

- **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).**

Likelihoodist approach

- 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).