Monotonicity of a Bayesian confirmation measure in rule support and confidence

Izabela Brzezińska

Poznań University of Technology

Roman Słowiński

Poznań University of Technology,

Institute for Systems Research, Polish Academy of Science

AI-METH 2005, Gliwice

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***

- ***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
  (called premise and conclusion, resp.)

Introduction

- **The number of rules generated from massive datasets can be very large**
- **Only a few of them are likely to be *useful***
- **To measure the relevance and utility of selected rules, quantitative measures, also known as *attractiveness* or *interestingness measures* (metrics), have been proposed (e.g. support, confidence, lift, gain, *conviction*)**
- **Aim: find *the most interesting rules* with respect to some attractiveness measures**

Introduction

- $\|\phi\|$ **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 $\phi \to \psi$ in *S*:**

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

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

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

Introduction

- 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"

- An important role in literaure is played by a confirmation measure denoted by ***f***

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

Rule support-confidence Pareto optimal border

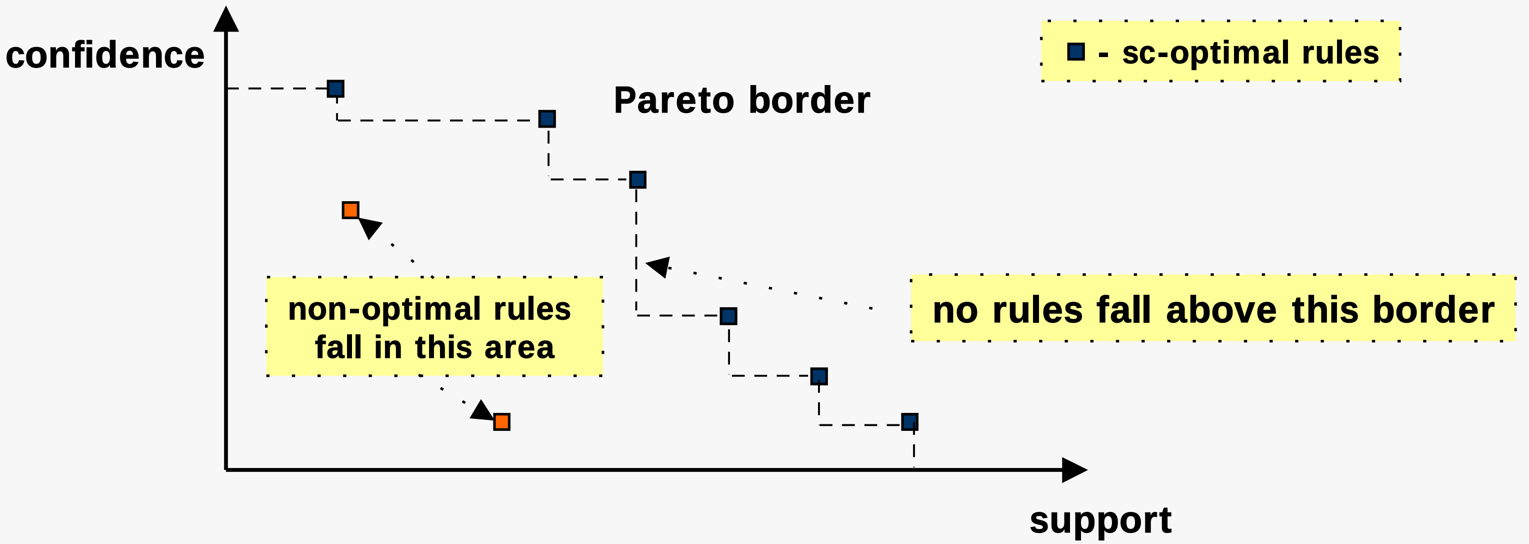
- In the set of rules induced from data sets, we look for rules that are ***optimal*** according to a chosen ***attractiveness measure***

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

- Bayardo and Agrawal (1999) proved, however, that, given a ***fixed conclusion*** ψ, the ***rule support-confidence Pareto border*** includes optimal rules according to any of those attractiveness measures

Rule support-confidence Pareto optimal border

- **Rule 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* etc.**

Rule support-confidence Pareto optimal border

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

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

- A function $g(x)$ is understood to be *monotone* in x,

  if $x_1 < x_2$ implies that $g(x_1) \leq g(x_2)$

Objective of work

- For a class of rules with fixed conclusion ψ, verify whether rules that are best according to the confirmation measure *f* are included in the rule support-confidence Pareto optimal border

- To fulfill this objective it must be checked wheter confirmation measure *f* is :

  1. monotone in rule support over rules with the same confidence, and
  2. monotone in confidence over rules with the same rule support

Monotonicity of *f* in rule support and confidence

- Let us consider a ***Bayesian confirmation measure f*** defined as follows:

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

- The measure *f* can be transformed into such a form:

$$f(\phi \to \psi) = \frac{|U| conf(\phi \to \psi) - sup(\psi)}{(|U| - 2sup(\psi))conf(\phi \to \psi) + sup(\psi)}$$

- It is assumed that **$|U|$** and ***sup*(ψ)** are ***constants*** as we consider only rules with a ***fixed conclusion*** (i.e. from one decision class)

Monotonicity of *f* in rule support for fixed confidence value

$$\boldsymbol{f}(\phi \to \psi) = \frac{|\boldsymbol{U}|\boldsymbol{conf}(\phi \to \psi) - \boldsymbol{sup}(\psi)}{(|\boldsymbol{U}| - 2\boldsymbol{sup}(\psi))\boldsymbol{conf}(\phi \to \psi) + \boldsymbol{sup}(\psi)}$$

- **Hypothesis:**

  ***f* is monotone in rule support for fixed confidence**

- **Verification:**

  f **is independent of rule support *sup*($\phi\to\psi$), so for *conf*($\phi\to\psi$)=const,**

  f **is constant and thus monotone in rule support**

- **Conclusion:**

  f **is monotone in rule support**

Monotonicity of *f* in confidence for fixed rule support

$$f(\phi \to \psi) = \frac{|U| conf(\phi \to \psi) - sup(\psi)}{(|U| - 2sup(\psi)) conf(\phi \to \psi) + sup(\psi)}$$

- **Hypothesis:**

  ***f* is monotone in confidence** **for fixed rule support**

- **Verification schema:**
  - **express *f* as a function of *conf*($\phi \to \psi$),**
  - **calcutate the derivative *f′* of *f* with respect to *conf*($\phi \to \psi$) and verify its sign**

- **Conclusions:**

  since *f′* is always $\geq$ 0 then *f* is monotone in confidence

Support-confidence monotonicity of *f* - conclusions

- The Bayesian confirmation measure *f* is:

1. independent of rule support and therefore monotone in rule support
2. and monotone in confidence

- Rules optimal with respect to

  ***f* lie on the rule support-confidence Pareto border**

  (sic: we consider rules with fixed conclusion)

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

- What's the use of looking for rules with optimal *f* since they lie on the Pareto border?
  - The result ***does not deny the interest of f*** in expressing the attractiveness of rules; it just states the monotonicity of *f* in confidence of rules for a fixed conclusion
  - ***Utility of scales***:

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

    the *f*($\phi \rightarrow \psi$) says 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 *f* (2)

- **Consider the possible result of rolling a die: 1,2,3,4,5,6, and let the conclusion 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

- **This example acknowledges the monotonicity of confirmation in confidence, it also clearly shows that the value of *f* has a more useful interpretation than *conf***
- **In particular, in case of rule $\phi_3 \rightarrow \psi$, which can also 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**

Utility of confidence vs. utility of confirmation *f* (3)

- **Consider the possible result of rolling a die: 1,2,3,4,5,6, and let the premise be kept fixed at 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*(ψ_2) to 2/3= *conf*($\phi\rightarrow\psi_2$), while it augments the probability of conclusion ψ_1 from 1/6=*sup*(ψ_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**

Conclusions

- **Confirmation measure *f is monotone* in rule support and confidence**

- **For a particular class of rules with a fixed conclusion ψ, rules *optimal with respect to f* are included on the rule support-confidence *Pareto optimal border***

- **As semantics of *f*($\phi \rightarrow \psi$) is more useful than that of *conf*($\phi \rightarrow \psi$), and as both these measures are monotonically linked while being independent of the support,**

  it would be reasonable to search for the most interesting rules taking into account *just confirmation f*($\phi \rightarrow \psi$) *and support sup*($\phi \rightarrow \psi$)

Thank you!