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

\[ S = \langle U, A \rangle \text{ – data table, where } U \text{ and } A \text{ are finite, non-empty sets} \]

\[ U \text{ – universe; } A \text{ – set of attributes} \]

\[ S = \langle U, C, D \rangle \text{ – decision table, where } C \text{ – set of condition attributes,} \]
\[ D \text{ – set of decision attributes, } C \cap D = \emptyset \]

**Decision rule** or **association rule** induced from \( S \)

is a consequence relation: \( \phi \rightarrow \psi \) read as \( \text{if } \phi \text{ then } \psi \)

where \( \phi \) and \( \psi \) 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

<table>
<thead>
<tr>
<th>U</th>
<th>Height</th>
<th>Hair</th>
<th>Eyes</th>
<th>Nationality</th>
<th>Support</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>tall</td>
<td>blond</td>
<td>blue</td>
<td>Swede</td>
<td>270</td>
</tr>
<tr>
<td>2</td>
<td>medium</td>
<td>dark</td>
<td>hazel</td>
<td>German</td>
<td>90</td>
</tr>
<tr>
<td>3</td>
<td>medium</td>
<td>blond</td>
<td>blue</td>
<td>Swede</td>
<td>90</td>
</tr>
<tr>
<td>4</td>
<td>tall</td>
<td>blond</td>
<td>blue</td>
<td>German</td>
<td>360</td>
</tr>
<tr>
<td>5</td>
<td>short</td>
<td>red</td>
<td>blue</td>
<td>German</td>
<td>45</td>
</tr>
<tr>
<td>6</td>
<td>medium</td>
<td>dark</td>
<td>hazel</td>
<td>Swede</td>
<td>45</td>
</tr>
</tbody>
</table>

\[ C \]  \[ D \]

E.g. decision rules induced from „characterization of nationalities“:

1) If \((Height=\text{tall})\), then \((\text{Nationality}=\text{Swede})\)

2) If \((Height=\text{medium}) \& (\text{Hair}=\text{dark})\), then \((\text{Nationality}=\text{German})\)
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:

\( \text{sup}(O) \) is the number of all objects from \( U \), having property \( O \)

E.g. \( \text{sup}(\phi) \), \( \text{sup}(\psi) \)
Basic quantitative characteristics of rules

- **Support** of rule $\phi \rightarrow \psi$ in $S$:

  $$sup(\phi \rightarrow \psi) = sup(\phi \land \psi)$$

- **Anti-support** of rule $\phi \rightarrow \psi$ in $S$:

  $$anti-sup(\phi \rightarrow \psi) = sup(\phi \land \neg \psi)$$
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 \(\Theta\) is a fraction constant between 0 and 1

- **Dependency Factor** (Pawlak 2002)

\[
\eta(\phi \rightarrow \psi) = \frac{sup(\phi \rightarrow \psi) - sup(\psi)}{\frac{sup(\phi)}{|U|} + \frac{sup(\psi)}{|U|}}
\]
Property M

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 $\phi$ and $\psi$ 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*

- $\phi$ is the property *to be a raven*, $\psi$ 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 \( \phi \) on \( \psi \) should be of the same strength, but of the opposite sign as the impact of \( \phi \) on \( \neg \psi \)

- Example: We draw cards from a standard deck.
  Let \( \phi \): the drawn card is *the seven of spades* and \( \psi \): *the card is black*. 
  \( \phi \) is conclusive for \( \psi \) and negatively conclusive for \( \neg \psi \)
Aim and scope of the conducted analysis

<table>
<thead>
<tr>
<th>Interestingness measure</th>
<th>Analyzed properties</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Property M</td>
</tr>
<tr>
<td>Rule Interest Function</td>
<td></td>
</tr>
<tr>
<td>Gain</td>
<td></td>
</tr>
<tr>
<td>Dependency Factor</td>
<td></td>
</tr>
</tbody>
</table>
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 \(\Delta > 0\), 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)
Dominated rules fall into this area

\[ \text{anti-support} = \sup (\phi \rightarrow \neg \psi) \]

No rules fall outside this 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
- dominoed rules
- Pareto-optimal border

$\rho_1$ represents rules optimal with respect to gain for $\Theta = 0.33$
and with respect to gain for $\Theta = 0.66$

$\rho_2$ represents rules optimal with respect to gain for $\Theta = 0.5$

$\rho_3$ represents rules optimal with respect to gain for $\Theta = 0.66$
and with respect to RI

$\rho_4$ represents rules optimal with respect to dependency factor
Conclusions
## Conclusions

<table>
<thead>
<tr>
<th>Interestingness measure</th>
<th>Properties</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Property M</td>
</tr>
<tr>
<td>Rule Interest Function</td>
<td>YES</td>
</tr>
<tr>
<td>Gain</td>
<td>YES</td>
</tr>
<tr>
<td>Dependency Factor</td>
<td>NO</td>
</tr>
</tbody>
</table>
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!
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)

<table>
<thead>
<tr>
<th>File</th>
<th>objects</th>
<th>atr+crit</th>
<th>classes</th>
<th>rules (alg)</th>
<th>consistency</th>
<th>length</th>
<th>coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buses</td>
<td>76</td>
<td>0+8</td>
<td>3</td>
<td>266 (all)</td>
<td>≥ 0.75</td>
<td>≤ 3</td>
<td>≥ 0.9</td>
</tr>
<tr>
<td>Nativity</td>
<td>342</td>
<td>0+33</td>
<td>2</td>
<td>64 (mc)</td>
<td>≥ 0.75</td>
<td>no limit</td>
<td>no limit</td>
</tr>
<tr>
<td>Urology</td>
<td>500</td>
<td>18+9</td>
<td>3</td>
<td>186 (mc)</td>
<td>≥ 0.96</td>
<td>no limit</td>
<td>no limit</td>
</tr>
</tbody>
</table>

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

„mc” = minimal-cover algorithm (DOMLEM)

*by Iza Brzezińska