Evaluation of Interestingness and Interaction of Conditions in Discovered Rules: Applications in Medical Data Analysis

Jerzy Stefanowski

Institute of Computing Sciences,
Poznań University of Technology
Poland

MLLS workshop, Sept. 23, 2016; Riva del Garda

General outline

- Medical applications of ML
- Rules induced from data

  → Interpretation
- Rule interestingness measures
  - Selection of complete rules
- Another perspective → focus on conditions inside a single rule and in sets of rules
- Studying most important conditions, their subsets and their interaction in rules
  - Set functions → Shapley and Banzhaf indices, Möbius representation
- Medical case studies

```
IF(BAO > 3)THEN disease A
IF(vol.ofgastric juice < 150)and(pain = high) THEN disease A
IF (Ot.gastric ≥ 100)and(duration=long)THEN disease B
```

Machine Learning for medical data

- Machine learning algorithms from beginning applied to analyse medical data
- Digitalization, new diagnostics tools → facilitate collecting and storing more data
- Many health units – collect, share large amounts of medical records
- Interest in automatic deriving medical diagnostic knowledge and interpreting results

See some surveys, such as:

I.Konenko: Machine Learning for Medical Diagnosis: History, State of the Art and Perspectives.

R.Bellazzi, F.Felazzi, L.Sachi: Predictive data mining in clinical medicine: a focus on selected methods and applications.

G.Magoulas, A.Prentza: Machine learning in medical applications.

…

However, difficulties with clinical acceptance - ECMLPKDD14 tutorial – P.Rodrigues: Knowledge discovery from clinical data

Requirements for systems supporting medical diagnosis

I.Kononeko’s postulates:

- Good performance
- Dealing with missing data
- Dealing with noisy data
- Transparency of diagnostic knowledge
- Explanation ability
- Reduction of the number of tests

This paper perspective:

→ Symbolic knowledge representation and its interpretation

Usually considered in medicine:

Decision trees, rules and partly Bayesian classifiers

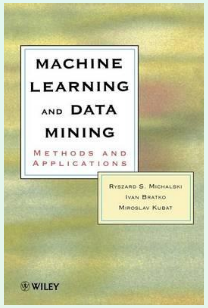
Rules – basics

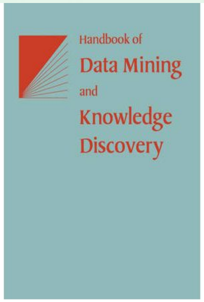
- Symbolic representation form

  IF Conditions THEN Class

IF Sex = male AND
Age > 46 AND
No_of_painful_joints > 3
AND
Skin_manif. = psoriasis
THEN Diagnosis =
Crystal_induced_synovitis

- Natural and easy for human → possible inspection and interpretation → descriptive perspective
  - Individual rules constitute "blocks" of knowledge
  - Rules directly related to facts in the training data
- Class predictions → easier to justify
- Rules could be integrated with domain knowledge
- Rules are more flexible than other representations
- Knowledge representations in AI / Intelligent Systems
  - Expert systems, Inference in IS
- Often used in medical applications

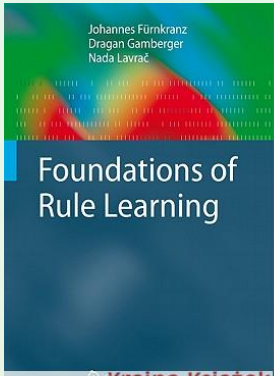




R. Michalski I.Bratko: Machine Learning and Data Mining.; W.Klosgen, J.Zytkow: Handbook of Data Mining and Knowledge Disc; J.Stefanowski: Rule discovery algorithms; C.Aggarwal: Data Classification 2015.

Different types of rules

- Various types of rules in data mining
  - Decision / classification rules
  - Association rules
  - Subgroup discovery → rule patterns
  - Logic formulas (ILP)
  - Rules in preference learning, rankings and ordinal classification
  - Multi-labeled classification
  - Sequential rule patterns
  - Other → action rules, …
- Other forms of rules in AI or MCDA,
- Comprehensive view:
  - Johannes Fürnkranz, Dragan Gamberger, Nada Lavrač: Foundations of Rule Learning, Springer 2012



How to learn rules?

- Typical algorithms based on the scheme of a sequential covering and heuristically generate a minimal set of rule to cover learning examples:
  - see, e.g., AQ, CN2, LEM, PRISM, MODLEM, Other ideas – PVM, R1, RIPPER, PART,..
- Other approaches to induce „richer" sets of rules:
  - Satisfying some requirements (Explore, BRUTE, or modification of association rules → „Apriori-like" CBA, OPUS,…)
  - Based on local „reducts" → Boolean reasoning or LDA
- Optimization problem (MP - Boolean rules)
- Meta-heuristics, e.g., genetic approaches
- Transformations of other representations:
  - Trees → rules
  - Construction of (fuzzy) rules from ANN

Case study - buses diagnostic rules

- A fleet of homogeneous 76 buses (AutoSan H9-21) operating in an inter-city and local transportation system [ack: J.Zak]
- 76 buses described by 8 technical symptoms and classified into 2 decision classes (good or bad technical condition)

s1 – maximum speed [km/h]	s5 – summer fuel consumption [l/100lm]
s2 – compression pressure [Mpa]	s6 – winter fuel consumption [l/100km]
s3 – blacking components in exhaust gas	s7 – oil consumption [l/1000km]
s4 – torque [Nm]	s8 – maximum horsepower

- Induction of a minimal set of rules (MODLEM)

  1. if ($s2 \geq 2.4$ MPa) & ($s7 < 2.1$ *l*/1000km) then (technical state=good) [46]

  2. if ($s2 < 2.4$ MPa) then (technical state=bad) [29]

  3. if ($s7 \geq 2.1$ *l*/1000km) then (technical state=bad) [24]

- The prediction accuracy → 98.7%.
- s2 → compression pressure, the most difficult measurement

Another set of rules – describe other conditions

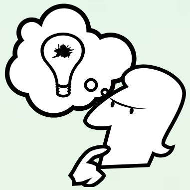
Rules by Explore with threshold (rule coverage > 50% in class):

1. if (*s1*>85 km/h) then (technical state=good) [34]
2. if (*s8*>134 kM) then (technical state=good) [26]
3. if (*s2*≥2.4 MPa) & (*s3*<61 %) then (technical state=good) [44]
4. if (*s2*≥2.4 MPa) & (*s4*>444 Nm) then (technical state=good) [44]
5. if (*s2*≥2.4 MPa) & (s7<2.1 *l*/1000km) then (technical state=good) [46]
6. if (*s3*<61 %) & (*s4*>444 Nm) then (technical state=good) [42]
7. if (*s1*≤77 km/h) then (technical state=bad) [25]
8. if (*s2*<2.4 MPa) then (technical state=bad) [29]
9. if (*s7*≥2.1 *l*/1000km) then (technical state=bad) [24]
10. if (*s3*≥61 %) & (*s4*≤444 Nm) then (technical state=bad) [28]
11. if (*s3*≥61 %) & (*s8*<120 kM) then (technical state=bad) [27]

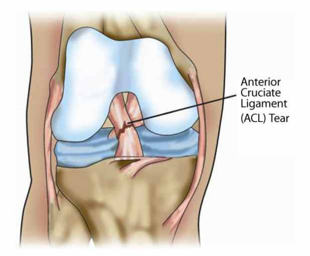
More appreciated by domain experts

Characteristic description / profile of buses

The prediction accuracy - still 97%



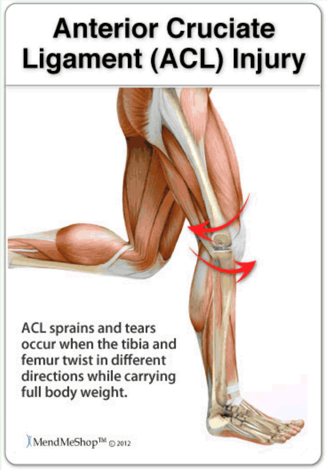
Evaluating conditions in ACL rules



- Diagnosing of an anterior cruciate ligament (ACL) rupture in a knee on the basis of magnetic resonance (MR) images (Slowinski K. et al. 2002)
- 140 patients described by 6 selected attributes
  - age, sex and body side and MR measurements (X, Y and PCL index)
- Patients categorized into two classes: „1” (with ACL lesion – 100 p.) and „2” (without ACL – 40 p.)
- Previous analysis of attribute importance
  - PCLINDEX; then age, sex, side
- Rule induction → (MODLEM) 15 rules (1- 4 conditions, with different supports)

  or richer set of rules (Explore) → over 30 rules

Predictive performance → accuracy 93.5%; G-mean 91.2%



ACL → minimal set of rules

rule 1. *if* (PCLINDEX < 3.225) *then* Class1 [26, 65%]
rule 2. *if* (AGE=[16.5,35))∧(PCLINDEX=[3.225,3.71)) *then* Class1 [6, 15%]
rule 3. *if* (SEX=MALE) ∧ (SIDE=RIGHT) ∧ (PCLINDEX=[3.225,3.71)) *then* Class1 [3, 7.5%]
rule 4. *if* (AGE=[16.5,35)) ∧ (PCLINDEX=[3.71,4.125)) ∧ (X≥14.5) *then* Class1 [2, 5%]
rule 5. *if* (X=[8.5,11.75)) ∧ (PCLINDEX=[4.125,4.535)) ∧ (SEX=MALE) *then* Class1 [1, 2.5%]
rule 6. *if* (X=[8.5,11.75)) ∧ (PCLINDEX=[3.225,3.71)) ∧ (AGE≥ 35) *then* Class1 [2, 5%]
rule 7. *if* (PCLINDEX=[3.71,4.125)) ∧ (X=[8.5,11.75)) ∧ (SEX=1) *then* Class1 [1, 2.5%]
rule 8. *if* (PCLINDEX≥4.535) *then* Class2 [75, 75%]
rule 9. *if* (SEX=FEMALE) ∧ (PCLINDEX=[4.125,4.535)) *then* Class2 [10,10%]
rule 10. *if* (PCLINDEX=[3.71,4.125)) ∧ (AGE≥ 35) *then* Class2 [6,6%]
rule 11. *if* (X=[11.75,14.5)) ∧ (Y=[2.75,3.75)) ∧ (SEX=FEMALE) *then* Class2 [8, 8%]
rule 12. *if* (SIDE=LEFT) ∧ (X=[11.75,14.5)) ∧ (Y=[2.75,3.75)) *then* Class2 [7, 7%]
rule 13. *if* (PCLINDEX=[3.225,3.71)) ∧ (AGE≥ 35) ∧ (SEX=MALE) *then* Class2 [2, 2%]
rule 14. *if* (AGE<16.5) *then* Class2 [14, 14%]
rule 15. *if* (PCLINDEX=[3.225,3.71))∧(Y=[3.75,4.75))∧(AGE≥ 35)∧(SIDE=LEFT) *then* Class2 [1,1%]

Clinical discussion → MR measurements are the most important.

- In particular, PCL< 3.23 (patients with ACL), PCL ≥ 4.53 (without ACL)
- Other PCL values → combinations with two other attributes age or sex indicate classes
  - Age below 16.5 years (so children or youth) characteristic class (without ACL lesion)
  - ACL injury more frequent for men and right side leg (sportsmen)!

Highly selective vagotomy (HSV)

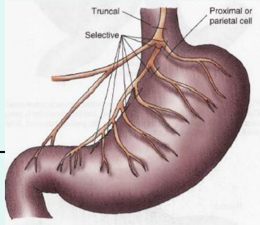


Figure 56-5 • The types of vagotomies.

- Highly selective vagotomy (HSV) - laparoscopic surgery for perforated Duodenal Ulcer Disease – stomach;
- An attempt to determine indications for HSV treatment (Slowinski K. et al. 1986);
- 122 patients × 11 pre-operating attributes and assigned to 4 target classes (long term result wrt. Visick grading)
  - Highly imbalanced and complex data
- LEM2 rule induction algorithm → 44 rules (1- 5 conditions)
- Predictive performance – accuracy 57% (not the main criterion)
- Focus on describing characteristic profiles of patients
- The previous results (e.g. very good prediction – class 1)
  - medium or longer duration of the disease,
  - without complications of ulcer or acute haemorrhage from ulcer,
  - medium or small volume of gastric juice per 1 hour (basic secretion),
  - medium volume of gastric juice per 1 hour under histamine,
  - high HCl concentration under histamine.

HSV – patient profiles for other classes

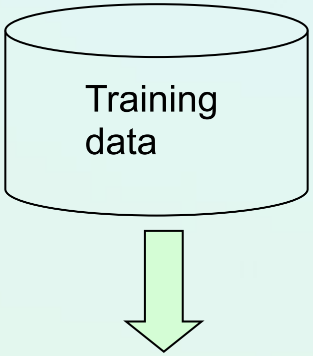
Other classes

- Satisfactory result of HSV treatment (class 2)
  - long or medium duration of disease,
  - multiple haemorrhages,
  - medium volume of gastric juice per 1 hour (basic secretion),
  - medium volume of gastric juice per 1 hour under histamine,
  - medium HCl concentration under histamine

- Unsatisfactory result of HSV treatment (class 3)
  - medium or short duration of the disease,
  - perforation of ulcer,
  - high or small volume of gastric juice per 1 hour (basic secretion),
  - high volume of gastric juice per 1 hour under histamine,
  - low HCl concentration under histamine.

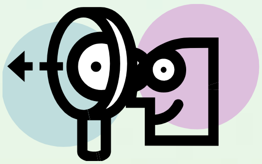
Motivations for interpreting rule patterns

- Description perspective → each rule evaluated individually - possibly an „interesting pattern”.
- Difficulties
  - Too many rules to be analyzed!
    - HSV (122 ob.×11 attr.) → 44 rules
    - Urology (500 ob. × 33 attr.) → 121 rules



r1. (A6 = 3) => (D1=1);
r2. (A1=2)&(A2=2)&(A4=2)&(A5 =3) => (D1=1);
r3. (A1 =2)&(A3=1)&(A4=2) & (A =3) => (D1=1)
r4. (A1 = 2) & (A5 = 1) => (D1=2);
r5. (A1 = 2) & (A6 = 1) => (D1=2);
r6. (A2 = 1) & (A4 = 3) => (D1=2);
r7. (A2 = 1) & (A5 = 1) => (D1=2);
....

- Related works → focus interest on some rules:
  - Subjective vs. objective perspective
  - Rule selection or ordering
    - Studies on rule evaluation measures
  - Interactive browsing
- Need for identification of characteristic attribute value pairs describing patients from particular classes



Rule interestingness measures

Rule *R*: IF ***P*** THEN ***K***

Objective measures → quantify R with the contingency table (learning data - *n*)

	K	¬*K*	
P	*a*	*c*	n_P
¬*P*	*b*	*d*	$n_{\neg P}$
	n_K	$n_{\neg K}$	*n*

$$sup(R) = a$$

$$conf(R) = \frac{a}{a+c}$$

$$G(P \wedge K) = \frac{a}{n}$$

$$IND(K,Q) = \frac{G(P \wedge K)}{G(P) \cdot G(K)}$$

$$K(K \mid P) = G(P)^{\alpha} \cdot (P(K \mid P) - G(K))$$

Many measures → see McGarry, Geng, L., Hamilton, H.et al. surveys;

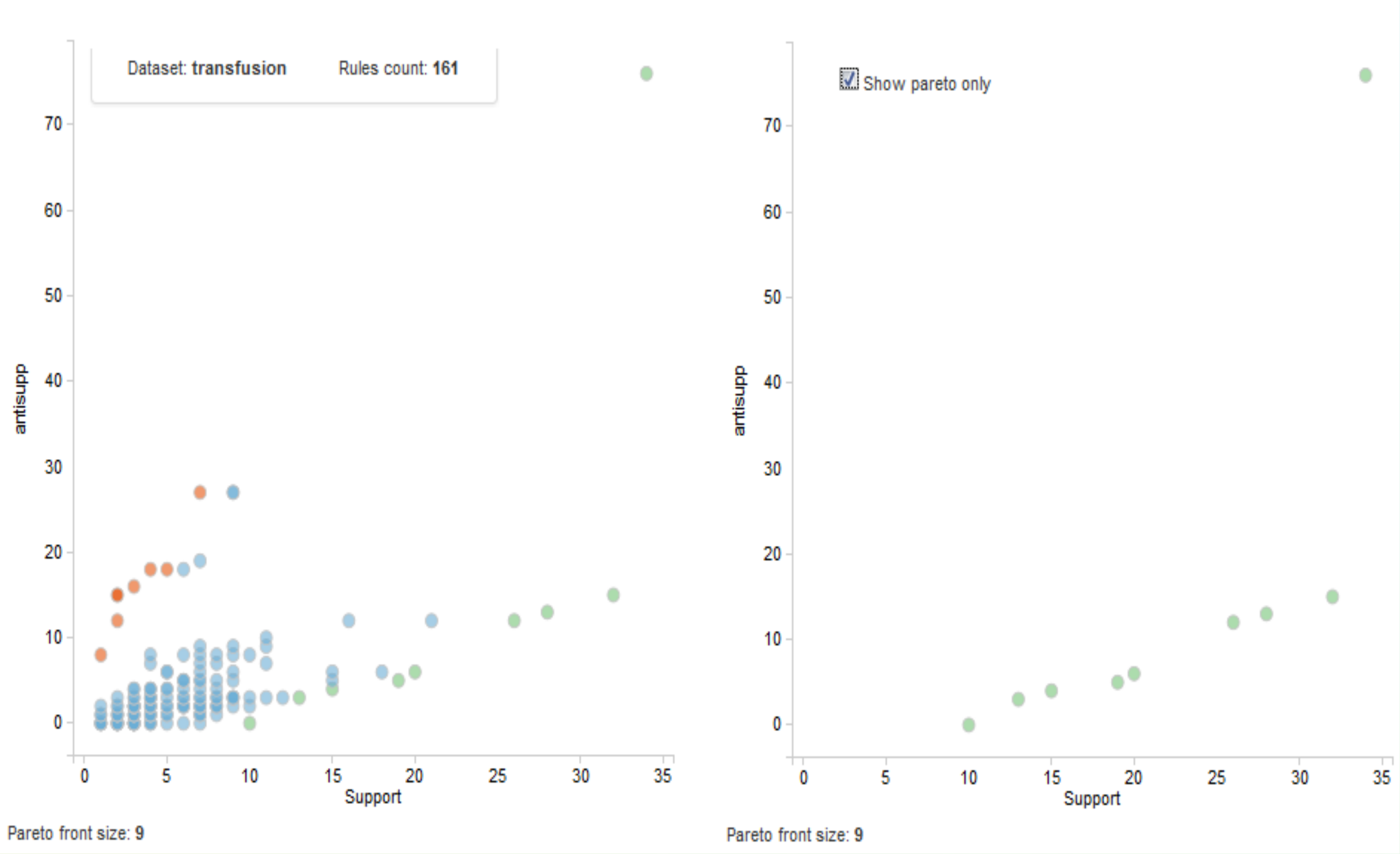
- Besides support → Bayesian confirmation measures (*K*,*P*)
- Study impact of the rule premise on its conclusion
  - Refer to class probabilities → imbalance

$$C(P,K) = conf(K \mid P) - P(K)$$

$$N(K,P) = P(P \mid K) - P(P \mid \neg K) = \frac{a}{a+b} - \frac{c}{c+d}$$

I.Szczech, S.Greco, R.Slowinski: Properties of rule interestingness measures. Inf. Scie. 2012

Minority class rules in the support–anti-support evaluation space → transfusion data and BRACID rules [Szczech,Stefanowski]



The best rules according to any monotonic measure are located on the support–anti-support Pareto border

Toward analysing conditions in rules

if $p_1 \wedge p_2 \wedge \ldots \wedge p_n$ *then class K*

if (blacking=medium) ∧ (oil_cons=low) ∧ (horsepower=high)
then (technical condition = good)

Current proposals:

→ Selecting a subset of rules from a larger set of many rules;
→ Focus on a „complete" condition part of a rule!

New view → evaluating an importance of elementary conditions and their interaction within the „if" part of the rule

Our aims:

- To propose a new approach based using set functions → Shapley, Banzhaf indices and Möbius representation
  - Start from a single rule → then generalize to the set of rules
- To verify the approach in rule discovery problems

Origins of the proposal

Shapley, Banzhaf indices / values and Möbius representation

- Previously considered in cooperative games, voting systems, party coalitions and multiple criteria decision aid:
- $X = \{1,2,...,n\}$ a set of elements / agents
  A set function $\mu : P(X) \rightarrow [0,1]$
  - A weighted average contribution of agent / element *i* in all coalitions
  - Conjoint importance of elements $A \subseteq X$
  - Measuring interaction of elements
- Main inspiration (Greco, Slowinski 2001) → a study of the relative value of information supplied by attributes to the quality of classification

Basics

- $X = \{1,2,...,n\}$ a set of elements (e.g. players in the game);
  $P(X)$ – the power set of X = the set of all possible subsets of X
  A set function $\mu : P(X) \rightarrow [0,1]$
- Function μ - a fuzzy measure satisfying:
  - $\mu(\emptyset) = 0$ and $\mu(X) = 1$
  - $A \subseteq B$ implies $\mu(A) \leq \mu(B)$
  - „1” could be treated as max value
- Interpretation of function μ in a particular problem
  - The profit obtained by players / agents
  - The importance of criteria in MCDA
- Transformations of function μ
  - Shapley and Banzhaf values refer to single elements $i \in X$, their interactions, subsets of elements $A \subseteq X$
  - Möbius representation $m: P(X) \rightarrow R$

Illustrative example – Möbius representation

Möbius representation $m: P(X) \to R$

For all $A \subseteq X$:

$$\sum_{B \subseteq A} m(B) = \mu(A)$$

$$m(A) = \sum_{B \subseteq A} \mu(B)(-1)^{|A|-|B|}$$

- m(A) – the contribution given by the conjoint presence of all elements from A to the function μ

Consider players 1,2,3, where the profits of their actions are $\mu(\{1\})=5$, $\mu(\{2\})=7$, $\mu(\{3\})=4$ and $\mu(\{1,2\})=15$ (by def. $\mu(\emptyset)=0$)

Calculate $m(\{1\})=5$, $m(\{2\})=7$ and $m(\{1,2\})=15-5-7=3$

Note - $\mu(\{1,2\})=15$ is greater than $\mu(\{1\}$ + $\mu(\{2\})$= 5+7

The contribution coming out from the conjoint presence of {1} and {2} in this coalition and it is equal to $m(\{1,2\})=3$

Illustrative example – Shapley value

- Shapley value – average contribution / importance of element
- Consider X={1,2,3} where the profits of the agent actions are $\mu(\{1\})=5$, $\mu(\{2\})=7$, $\mu(\{3\})=4$, $\mu(\{1,2\})=15$, $\mu(\{1,3\})=12$, $\mu(\{2,3\})=14$ and $\mu(\{1,2,3\})=30$
- How to fairly split the total profit of 30 units among the agents taking into account their contribution?
- Attribute to the conjoint presence of agents $A \subseteq X$, so split equally $m(A)$ among agents

$$m(A)/|A|$$

- Each agent should receive the value (Shapley)

$$\phi_i(\mu) = \sum_{A \subseteq X : i \in A} \frac{m(A)}{|A|}$$

Illustrative example – Shapley value

- X={1,2,3} and profits are μ({1})=5, μ({2})=7, μ({3})=4, μ({1,2})=15, μ({1,3})=12, μ({2,3})=14 and μ({1,2,3})=30

Möbius representations

- m({1})=5, m({2})=7, m({3})=4, m({1,2})=μ({1,2})- μ({1}) -μ({2}) =15-5-7=3, m({1,3})=3, μ({2,3})=3 and m({1,2,3})=μ({1,2,3})-μ({1,2})-μ({1,3})-μ({2,3})+μ({1})+ μ({2})+μ({3})=30-15-12-14+5+7+4=5

Shapley values for each agent

- ϕ_1(μ)=m({1})/1+m({1,2})/2+m({1,3})/2+m({1,2,3})/3=5+3/2+3/2+5/3=9.67
- ϕ_2(μ)=m({2})/1+m({1,2})/2+m({2,3})/2+m({1,2,3})/3=7+3/2+3/2+5/3=11.67
- ϕ_3(μ)=m({3})/1+m({1,3})/2+m({2,3})/2+m({1,2,3})/3=5+3/2+3/2+5/3=9.67

$$\phi_i(\mu) = \sum_{A \subseteq X : i \in A} \frac{m(A)}{|A|}$$

Other formulations

Shapley value:

$$\Phi_i(\mu) = \sum_{A \subseteq X - \{i\}} \frac{(|X - A| - 1)!\,|A|!}{|X|!} \cdot [\mu(A \cup \{i\}) - \mu(A)]$$

Banzhaf value:

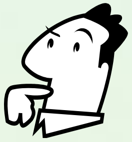
$$\Phi_{B_i}(\mu) = \frac{1}{2^{|X|-2}} \sum_{A \subseteq X - \{i\}} [\mu(A \cup \{i\}) - \mu(A)]$$

Both interpreted as an averaged contribution of element *i* to all coalitions *A*

Interaction indices (*i*,*j*) → Morofushi and Soneda; Roubens

$$I_{MS}(i,j) = \sum_{A \subseteq X - \{i,j\}} \frac{(|X - A| - 2)!\,|A|!}{(|X| - 1)!} \cdot [\mu(A \cup \{i,j\}) - \mu(A \cup \{i\}) - \mu(A \cup \{j\}) + \mu(A)]$$

$$I_R(i,j) = \frac{1}{2^{n-2}} \sum_{A \subseteq X - \{i,j\}} [\mu(A \cup \{i,j\}) - \mu(A \cup \{i\}) - \mu(A \cup \{j\}) + \mu(A)]$$



Adaptation to evaluate conditions in a single rule

- Consider a single rule *if* $p_1 \wedge p_2 \wedge \ldots \wedge p_n$ *then class K*
- Need to analyse its sub-rules *if* $p_{j1} \wedge p_{j2} \wedge \ldots \wedge p_{jl}$ *then class K* such that $\{p_{j1}, p_{j2}, \ldots, p_{jl}\} \subseteq \{p_1, p_2, \ldots, p_n\}$
  - sub-rules are more general than the first rule
- Choice of the characteristic function μ to evaluate a rule?
  - Confidence of the rule $\mu(W,K)=conf(r)$, where W is a set of conditions in r
  - Also – confirmation measures, …
- Then, for $Y \subset W$ we need to adapt set functions
  - $\mu(\emptyset,K)=?$ 0 or class prior

Indices for each condition in a rule

$p_i \in W$ - single condition in rule r, and $|W| = n$

- Shapley value:

$$\Phi_s(p_i, r) = \sum_{Y \subseteq W - \{p_i\}} \frac{(n - |Y| - 1)!|Y|!}{n!} \cdot [\mu(Y \cup \{p_i\}, K) - \mu(Y, K)]$$

- Banzhaf value:

$$\Phi_B(p_i, r) = \frac{1}{2^{n-1}} \sum_{Y \subseteq W - \{p_i\}} [\mu(Y \cup \{p_i\}, K) - \mu(Y, K)]$$

Both values Φ – a weighted contribution of p_i in rules generalized from r
For Shapley value - μ(W) is shared among all elements of W

Pairs – measures of an interaction resulted from putting p_i and p_j together in all subsets of conditions in rule r:

- Positive – complementary in increasing the confidence
- Negative – putting together provide some redundancy

$$I_{MS}(p_i, p_j) = \sum_{Y \subseteq W - \{p_i, p_j\}} \frac{(n - |Y| - 2)!|Y|!}{(n-1)!} \cdot [\mu(Y \cup \{p_i, p_j\}, K) - \mu(Y \cup \{p_i\}, K) - \mu(Y \cup \{p_j\}) + \mu(Y, K)]$$

Adapted indices for subsets - part 2

Generalized indices for a subset of conditions V⊂W [Grabisch]

Shapley generalized index

$$I_S(V,r) = \frac{1}{2^{n-|V|}} \sum_{Y \subseteq W-V} \frac{(n-|Y|-|V|)!|Y|!}{(n-|V|+1)!} \sum_{L \subseteq V} (-1)^{|V|-|L|} \mu(Y \cup L, K)$$

Banzhaf index of conditions V⊂W

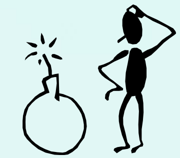
$$I_B(V,r) = \frac{1}{2^{n-|V|}} \sum_{Y \subseteq W-V} \sum_{L \subseteq V} (-1)^{|V|-|L|} \mu(Y \cup L, K_j)$$

Average conjoint contribution of the subset of conditions V⊂W to the confidence of all rules generalized from *r*

The Möbius representation of set functions μ :

$$m(V,r) = \sum_{B \subseteq V} (-1)^{|V-B|} \mu(B,K)$$

An intuitive example



HSV treatment – one of the rules

if (gastric_juice=*medium*)$\wedge$(HCL_conc.=*low*) then (*result*=*good*)

conf=1.0 , supp = 13 examples.

Möbius representation $m(1,2)$ = **-0.14493!!!**

- Rule generalizations and Möbius representation m:
  - Empty condition part → $m(0)=0$
  - *if* (gastric_juice=*medium*) then (*result* =*good*)
    $m(1)=0.16667$ and conf=0.16667
  - *if* (HCL_conc.=*low*) then (*result* =*good*)
    $m(2)=0.97826$ and conf= 0.97826
- An increase of rule confidence
  $1 = m(1) + m(2) + m(1,2)$
- Values of Möbius representation show the distribution of confidence among all coalitions of the considered conditions in the subset {(gastric_juice=*medium*),(HCL_conc.=*low*)}

Shapley value for single conditions

φ(gastric_juice=*medium*)=0.0942; φ((HCL_conc.=*low*) =0.908

Evaluating conditions in ACL rule

if (sex = female)$\wedge$(Y1 $<$ 2.75) $\wedge$(PCL$\in$[3.71,4.13)) *then* (no ACL) conf. =1.0

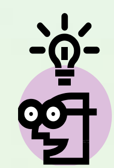
Sex	Y1	PCL	Banzhaf	Shapley	Mobius	conf.
∅	∅	√	0.43535	0.49575	0.28571	0.2857
∅	√	∅	0.24207	0.30246	0.04651	0.0465
∅	√	√	0.53015	0.53015	0.1766	0.5241
√	∅	∅	0.14139	0.1591	0.1452	0.1452
√	∅	√	0.1135	0.1135	-0.2316	0.2923
√	√	∅	0.1734	0.1734	-0.1034	0.1486
√	√	√	0.72476	0.72476	0.72476	1

Evaluating conditions in a set of rules

- The set of rules $R = \bigcup_{j=1}^{k} R(K_j)$, where *R*(*Kj*) a set of rules having as a consequence class *Kj*
- A given set of conditions Γ_f occur in many rules
- $FM_r(\Gamma_f)$ denote an evaluation of its contribution to the confidence of rule *r*
- The global contribution of Γ_f in a rule set *R* with respect to class *Kj* is calculated as:

$$G_{Kj}(\Gamma_j) = \sum_{r \in R(Kj)} FM_r(\Gamma_f) \cdot \sup(r) - \sum_{s \in \neg R(Kj)} FM_s(\Gamma_f) \cdot \sup(s)$$

- Conditions Γ_f are ranked according to $G_{Kj}(\Gamma_f)$ → identify the most characteristic combinations of conditions for rules from a given class
- Computational costs → start from the smallest sets of cond.



An illustrative example

An interest in condition (a7=0) in a set of several rules

It occurs in following rules with conf=1:

R1 if (a3=1)∧(a7=0) ∧(a3=1) then (D=1) sup 1

R2 if (a4=1)∧(a7=0) then (D=1) sup 45

R5 if (a4=0)∧(a7=0) then (D=2) sup 7

(Möbius representation of (a7=0)) in R1, R2 m=0.939
and in R5 m=0.184

A global contribution of (a7=0)

- (D=1) $0.939 \times 1 + 0.939 \times 45 = 43.194$
- (D=2) $0.184 \times 7 = 1.288$

Finally $G_{D=1}$(a7=0) = 43.194 – 1.288 = 41.906

Analysis of conditions in buses rules

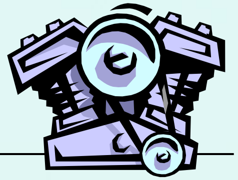


Table 1. Rankings of best conditions according to evaluation measures calculated for "buses" rules

busses in a good technical condition					
Möbius		Shapley		Banzhaf	
condition	value	condition	value	condition	value
comp-press=high	214.34	comp-press=high	116.91	comp-press=high	116.91
torque=high	163.36	torque=high	163.36	torque=high	163.36
blacking=low	161.33	blacking=low	87.86	blacking=low	87.86
oil cons.=low	132.36	oil cons.=low	70.88	oil cons.=low	70.80
MaxSpeed=high	122.66	MaxSpeed=high	63.71	MaxSpeed=high	63.71
busses in a bad technical condition					
Möbius		Shapley		Banzhaf	
condition	value	condition	value	condition	value
torque=low	48.33	torque=low	29.17	torque=low	29.17
blacking=high	46.70	comp-press=low	29.00	comp-press=low	29.00
comp-press=low	29.00	blacking=high	27.98	blacking=high	28.06
oil-cons.=high	27.00	oil-cons.=high	27.00	oil-cons.=high	27.00
summ-cons.=high	26.67	horsepower=low	26.00	horsepower=low	26.66
horsepower=low	26.66	MaxSpeed=low	25	MaxSpeed=low	25

- Pairs of conditions – much lower evaluations e.g. (horsepower=average) and (oil consumption=low) 0.166
- Previous analysis → „good" conditions: high compression pressure, torque, max-speed and low blacking components. Opposite values → characteristic for bad technical conditions. Blacking components in the exhaust gas and oil consumption more important than fuel consumption.

Evaluating conditions in ACL rules

- Diagnosing an anterior cruciate ligament (ACL) rupture in a knee on the basis of magnetic resonance (MR) images (Slowinski K. et al.)
- 140 patients described by 6 attributes
  - age, sex and body side and MR measurements (X, Y and PCL index).
- Patients classified into two classes „1” (with ACL lesion – 100) and „2” (without ACL – 40).
- LEM2 rule induction algorithm → 15 rules (1- 4 elementary conditions with different support, few possible rules).
- Clinical discussion → MR measurements are the most important.
  - In particular PCL< 3.23 (patients with ACL), PCL ≥ 4.53 (without ACL)
  - Other PCL values → combinations with two other attributes age or sex indicate classes.
  - Age below 16.5 years (so children or youth) characteristic for class (without ACL lesion).
  - ACL injury more frequent for men (sportsmen)!

ACL → minimal set of rules

rule 1. *if* (PCLINDEX $<$ 3.225) *then* Class1 [26, 65%]
rule 2. *if* (AGE=[16.5,35))$\wedge$(PCLINDEX=[3.225,3.71)) *then* Class1 [6, 15%]
rule 3. *if* (SEX=MALE) $\wedge$ (SIDE=RIGHT) $\wedge$ (PCLINDEX=[3.225,3.71)) *then* Class1 [3, 7.5%]
rule 4. *if* (AGE=[16.5,35)) $\wedge$ (PCLINDEX=[3.71,4.125)) $\wedge$ (X$\geq$14.5) *then* Class1 [2, 5%]
rule 5. *if* (X=[8.5,11.75)) $\wedge$ (PCLINDEX=[4.125,4.535)) $\wedge$ (SEX=MALE) *then* Class1 [1, 2.5%]
rule 6. *if* (X=[8.5,11.75)) $\wedge$ (PCLINDEX=[3.225,3.71)) $\wedge$ (AGE$\geq$ 35) *then* Class1 [2, 5%]
rule 7. *if* (PCLINDEX=[3.71,4.125)) $\wedge$ (X=[8.5,11.75)) $\wedge$ (SEX=1) *then* Class1 [1, 2.5%]
rule 8. *if* (PCLINDEX$\geq$4.535) *then* Class2 [75, 75%]
rule 9. *if* (SEX=FEMALE) $\wedge$ (PCLINDEX=[4.125,4.535)) *then* Class2 [10,10%]
rule 10. *if* (PCLINDEX=[3.71,4.125)) $\wedge$ (AGE$\geq$ 35) *then* Class2 [6,6%]
rule 11. *if* (X=[11.75,14.5)) $\wedge$ (Y=[2.75,3.75)) $\wedge$ (SEX=FEMALE) *then* Class2 [8, 8%]
rule 12. *if* (SIDE=LEFT) $\wedge$ (X=[11.75,14.5)) $\wedge$ (Y=[2.75,3.75)) *then* Class2 [7, 7%]
rule 13. *if* (PCLINDEX=[3.225,3.71)) $\wedge$ (AGE$\geq$ 35) $\wedge$ (SEX=MALE) *then* Class2 [2, 2%]
rule 14. *if* (AGE$<$16.5) *then* Class2 [14, 14%]
rule 15.*if* (PCLINDEX=[3.225,3.71))$\wedge$(Y=[3.75,4.75))$\wedge$(AGE$\geq$ 35)$\wedge$(SIDE=LEFT) *then* Class2 [1,1%]

Evaluating conditions in ACL rules

With ACL				Without ACL			
Möbius		Shapley		Möbius		Shapley	
PCL < 3.23	18.57	PCL < 3.23	18.57	PCL ≥ 4.53	21.42	PCL ≥ 4.53	21.42
PCL∈[3.23,3.7)	4.87	PCL∈[3.23,3.7)	5.06	Age < 16.5	4.0	Age < 16.5	4.0
(Age∈[16.5,35] & (PCL ∈[3.7,4.1)	1.58	Age∈[16.5,35)	1.63	Sex=female	3.23	Sex=female	2.85
(X1≥14.5) & (PCL ∈[3.7,4.1)	0.54	(X1≥14.5) & (PCL ∈[3.7,4.1)	0,92	PCL∈[4.13,4.5)	2.22	Y1∈[2.75,3.75)	1.84
X1 ∈[8.5,11.8)	0.52	Age∈[16.5,35 & (PCL ∈[3.7,4.1)	0.86	(Age≥35] & (PCL ∈[3.7,4.1)	1.78	(Age≥35] & (PCL ∈[3.7,4.1)	1.78
Sex=male	0.44	Sex=male	0.83	X1∈[11.8,14.5) PCL∈[3.23,3.7)	1.31	X1∈[11.8,14.5)	1.53
Age∈[16.5,35)	0.34	Y1<2,75 & (PCL ∈[3.7,4.1)	0.67	Y1∈[2.75,3.75)	1.28	PCL∈[4.13,4.53	1.48

Subsets of conditions → characteristic description of both diagnostic classes; PCL index with extreme intervals definitely the most important + its other values occur in some pairs, e.g (Age∈[16.5,35]) & (PCL ∈ [3.7,4.1)

Sex and age – young men (often sportsmen)

Evaluating conditions in ACL rules

- Rankings of conditions with respect to Shapley and Banzhaf values – top elements are the same.
- Top ranking with Möbius representation small re-ordering but PCL also dominates
- Pairs of conditions are higher evaluated than in the previous case
- Support for profiles of ACL patients
  - MR measurements are the most important

  Patients with ACL
  - PCL< 3.23 ; (Age∈[16.5,35]) & (PCL ∈ [3.7,4.1)
  - Sex=male and X1 ∈[8.5,11.8)

  Patients without ACL
  - PCL ≥ 4.53
  - Other MR measurements → combinations with two other attributes age or sex indicate classes.
  - Age below 16.5 years (so children or youth) or (age = much older) are characteristic for (without ACL)
- Profiles consistent with the earlier analyses and clinical knowledge

Highly selective vagotomy rules

Highly selective vagotomy (HSV) - laparoscopic surgery for perforated Duodenal Ulcer Disease.

- An attempt to determine indications for surgery treatment;
  - 122 patients described by 11 pre-operating attributes and assigned to 4 target class
  - 44 rules (1- 5 conditions)
- Focus on describing characteristic profiles of patients
- The previous results, e.g. very good prediction – class 1)
  - long or medium duration of the disease,
  - without complications of ulcer or acute haemorrhage from ulcer,
  - medium or small volume of gastric juice per 1 hour (basic secretion),
  - medium volume of gastric juice per 1 hour under histamine,
  - high HCl concentration under histamine.

Evaluating conditions in HSV rules – class 1 (good)

Möbius		Shapley		Banzhaf	
Cond	*Value*	*Cond*	*Value*	*Cond*	*Value*
A6=2	2,34	A6=2	3,85	A6=2	4,01
A9=3	2,31	A4=1	3,41	A4=1	3,57
A4=2	1,89	A4=2	3,16	A4=2	3,08
A4=1	1,58	A9=3	2,59	A9=3	2,72
A2=2	1,27	A2=2	1,65	A2=2	1,88

Möbius		Shapley		Banzhaf	
Cond	*Value*	*Cond*	*Value*	*Cond*	*Value*
A4=1 & A6=2	2,62	A4=1 & A6=2	2,82	A4=1 & A6=2	2,83
A4=1 & A8=1	1,95	A4=1 & A8=1	1,95	A4=1 & A8=1	1,95
A2=2 & A6=2	1,89	A5=2 & A6=1	1,49	A5=2 & A6=1	1,49
A2=2 & A9=3	1,53	A3=3 & A7=2	1,18	A3=3 & A7=2	1,18
A5=2 & A6=1	1,49	A2=2 & A6=2	1,01	A2=2 & A6=2	1,01

Attributes: A2 – age; A4 – complications of ulcer; A6 - volume of gastric juice per h; A9 - HCL concentration after histamine; A5 - HCL concentration; A3 - duration of disease

Subsets of conditions → closer to single conditions

HSV –patient class profiles

- Very good result of HSV (class 1)
  - without complications of ulcer or acute haemorrhage from ulcer,
  - medium or small volume of gastric juice per 1 hour (basic secretion),
  - medium volume of gastric juice per 1 hour under histamine,
  - high HCl concentration under histamine
  - / no medium duration of disease
- Satisfactory result of HSV (class 2)
  - long or medium duration of disease,
  - multiple haemorrhages,
  - medium or small volume of gastric juice per 1 hour (basic secretion),
  - medium volume of gastric juice per 1 hour under histamine,
  - medium or low HCl concentration under histamine
- Unsatisfactory result of HSV treatment (class 3)
  - medium or short duration of the disease,
  - perforation of ulcer,
  - high or small volume of gastric juice per 1 hour (basic secretion),
  - high volume of gastric juice per 1 hour under histamine,
  - No low HCl concentration under histamine condition in the rankings
- Bad result of HSV treatment (class 4)
  - Consistent profile
  - + new condition - low HCl concentration under histamine

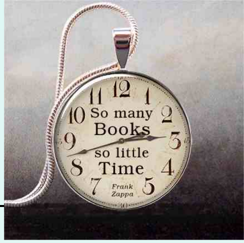
Working with larger set of rules

- „ESWL" – urological data
  - Urinary stones treatment by ESWL extracorporeal shock waves lithotripsy
- 500 patients × 33 attributes classified into two classes (imbalanced) – difficult to analyse (Antczak, Kwias et al. 2000)
- Explore rule induction algorithm → 484 rules (2-7 conditions with different support ≥ 5%, confidence ≥ 0.8).

ESWL rules

- Explore rule induction algorithm → 484 rules (2-7 conditions with different support ≥ 5%, confidence ≥ 0.8).
- Using the set functions we identify:
  - Class 1 → 8 single conditions, 12 pairs
    - (basic dysuric symptoms=1), (crystaluria=1), (location of the concrement=2),(stone size=2), …, (crystaluria=2)&(proteinurine=1), etc.
  - Class 2 → 10 single conditions, 13 pairs
    - (location of the concrement =3), (lumbar region pains=5), (operations in the past=3),…, (crystaluria=3)&(proteinurine=2),.., (cup-concrement=1)&(stone size=2), etc.
- More visible differences in Shapley and Banzhaf rankings ; triples less evaluated than single conditions and pairs.

Extensions to improve computability



- Limitations - computational for rules having more conditions
  - Both time and memory (to store temporary results)
- Possible heuristic approaches:
  - First filter and reduce the set of rules, then evaluate.
  - Iterative analysis, start from single conditions, pairs and work with smaller sets of conditions
- Modify calculations of measures (approximate them)

  M.Sikora: *Selected methods for decision rule evaluation and pruning* (2013)
  - Analyse only single conditions in rules
  - Do not consider all sub-rules (restrict to rules affected by dropping the single condition, or base sub-rules with the single condition)
  - Simpler forms of Baznhaf and Shapley indices

Possible re-using of best conditions in rule constructive induction

Final remarks

Interpretation of rule patterns

Our contribution:

- Evaluating the role of subsets of elementary conditions in rules discovered from data + their interaction and conjoint contribution
- An adaptation of measures based on set functions (not so frequent in ML)

Medical context:

- Identification of the most important conditions in single rules, sets of rules
- Support for characteristic descriptions of patients from different targets
- Using rules → order of applying diagnostic tests inside rules, complementary tests (use together), redundancy,..

Experimental observations:

- Identified conditions, pairs consistent with previous results (4 case studies)
- Rankings quite similar: Möbius has a wider range, Shapley and Banzhaf nearly the same – differences for larger sets of rules having more conditions

Approximate calculations + other applications

Co-operation with


Rules and set functions: Salvatore Greco (University of Catania) and Roman Słowiński (Poznan University of Technology)

S.Greco, R.Slowinski, J.Stefanowski: Evaluating importance of conditions in the set of discovered rules. In RFSDMGC Proc. (2007)

Med. applications: Krzysztof Słowiński, Dariusz Siwiński Andrzej Antczak, Zdzisław Kwias (Poznan Univ. of Medical Sciences)

Technical diagnostics: Jacek Żak et al. (PUT)

My master students (PUT)

- Bartosz Jędrzejczak (also soft. implementation)

Thank you for your attention

Questions and remarks?

Contact, remarks:
Jerzy.Stefanowski@cs.put.poznan.pl

or www.cs.put.poznan.pl/jstefanowski