Preferences in Argumentation for Statistical Model Selection

COMMA 2016 Short Paper presentation

Isabel Sassoon, Jeroen Keppens, Peter McBurney

Department of Informatics
King's College London
isabel.sassoon@kcl.ac.uk

14th September, 2016

KING'S College LONDON

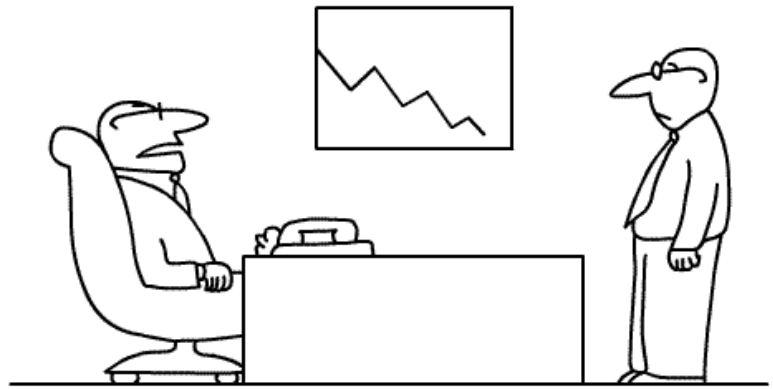
Agenda
Preferences in Argumentation for Statistical Model Selection

1. Background
2. Case study
3. Proposed method
4. Future work

KING'S College LONDON

Background

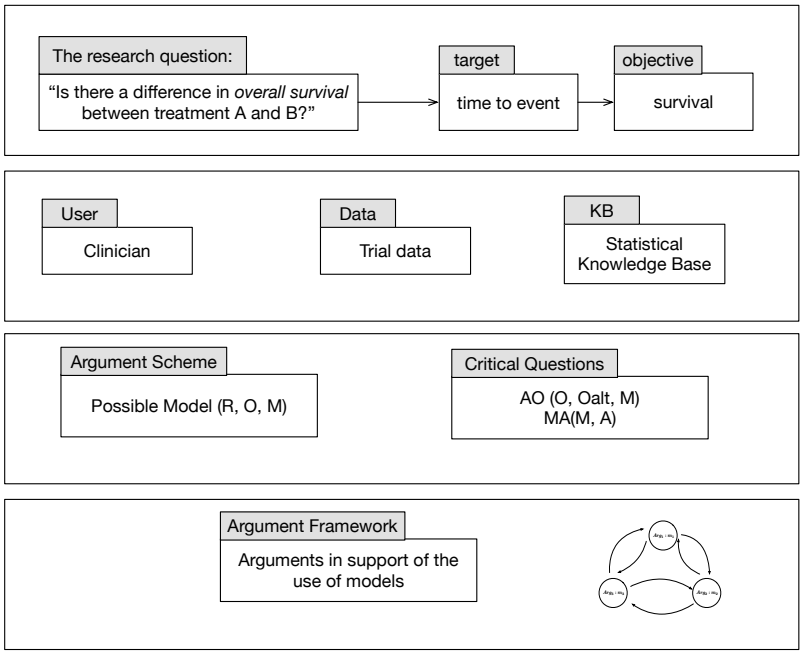
- Data collection is routine
- Easy to access but not easy to analyse robustly
- More than one possible analysis approach
- Justification for choice of approach
- Audit trail



"It would appear, Hopkins, that your gut feel was only indigestion"

KING'S College LONDON

Past research

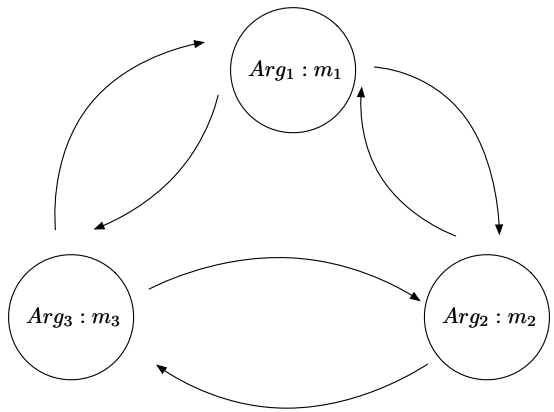


- Argument scheme for possible model given an objective
- Statistical Knowledge Base to hold all relevant theory
- Critical questions also instantiated as argument schemes

KING'S College LONDON

Case Study

- SENT multicentre Trial, 14 centres, 420 patients, 5 years of follow up.
- Research question: *is there a difference in survival between patients who had adjuvant therapy (Radiotherapy or Chemotherapy) to those that did not have any additional treatment?*
- Instantiating the argument schemes result in: $Args = \{m_1, m_2, m_4\}$, where each m_i is an argument supporting the use of a particular model
- Only one model can be applied - arguments symmetrically attack each other



KING'S College LONDON

Source of Preferences for Statistical Model Selection

A preference expressed in the context of statistical model selection refers to an order of priority between models.
If we have a set of models $\{m_1, \ldots, m_n\}$ then a preference order $Pref : M \times M$ where $pref_j$ on these models would be of the form: $Pref = \{m_a \succ m_b \succ m_c\}$ where $a, b, c \in \{1, \ldots, n\}$. Preferences in Statistical Model Selection can result from:

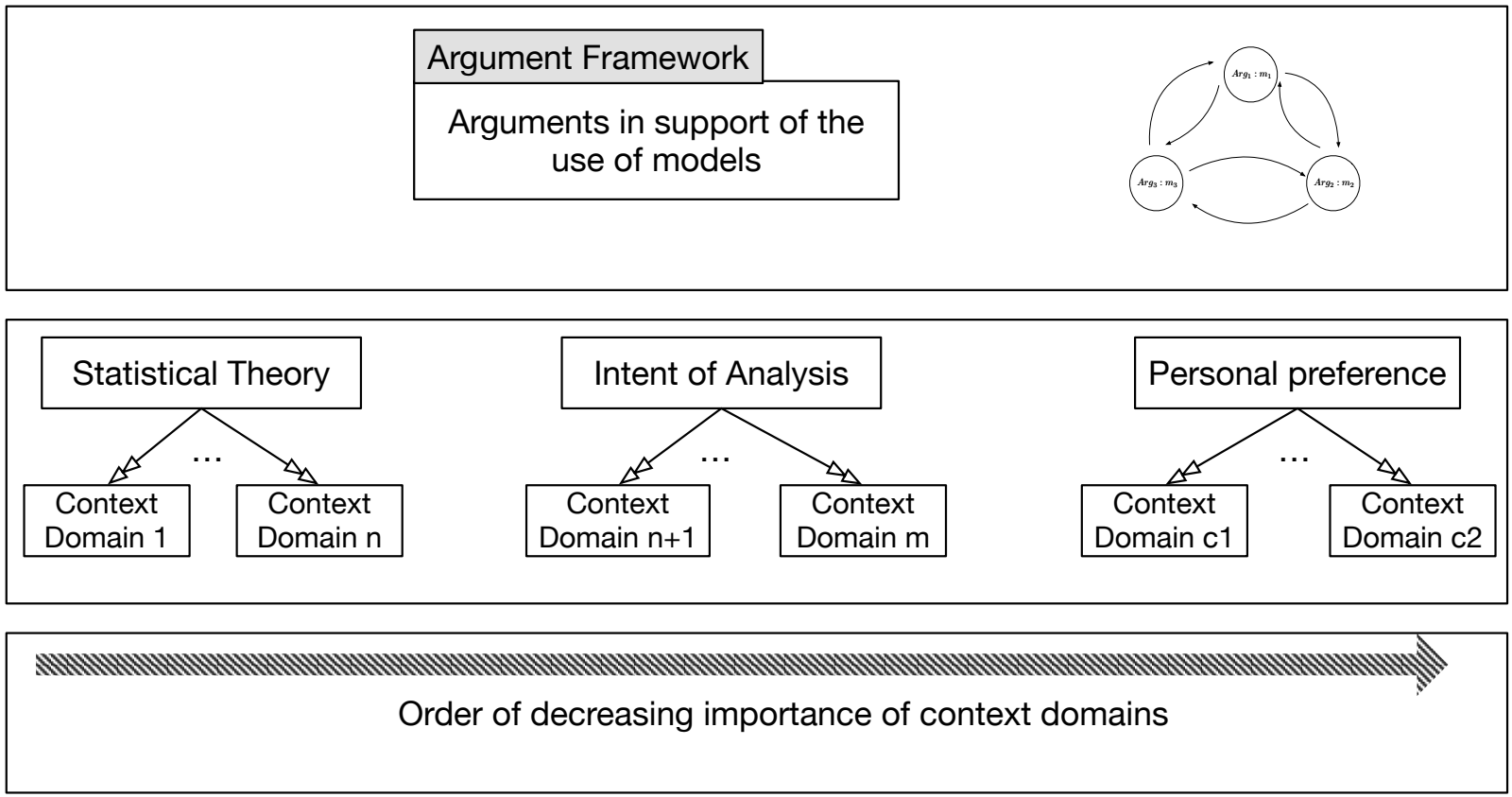
- Statistical Theory
- Intent of Analysis
- User preference

Each of these sources will result in one or more context domains (CD_i)

KING'S College LONDON

Importance order of Preferences for Statistical Model Selection

Some preferences are of higher importance than others:



Each preference order (more than one can be derived from each source results in a context domain (CD_i).

KING'S College LONDON

Formal Definitions - Extended Statistical Knowledge Base

- A set of context domains $CD = \{CD_1, \ldots, CD_H\}$. Each CD_h is a set of mutually exclusive contexts.
- A set of totally ordered sets of performance measures $P = \{P_1, \ldots, P_H\}$. Each P_h contains a set of measures $p_{h1} \prec \ldots \prec p_{hj_h}$ by means of which a model's performance is assessed in a specific context.
- A set of performance function $PF = \{PF_1, \ldots, PF_H\}$, such that each $PF_i : CD_i \times M \mapsto P_i$.

KING'S College LONDON

Sample Context Domain, performance measure

Context Domain	Model	Performance measure
absent	m_1 KM	unaffected
	m_2 PH	unaffected
	m_4 χ^2	unaffected
light	m_1 KM	unaffected
	m_2 PH	unaffected
	m_4 χ^2	affected
heavy	m_1 KM	affected
	m_2 PH	unaffected
	m_4 χ^2	affected

Table: Sample performance function for model resilience to censoring

KING'S College LONDON

Formal Definitions - Context domains preferences to an EAF

- To construct an argumentation model based on the extended statistical knowledge base: the set of contexts $CD \subseteq CD_1 \cup \cdots \cup CD_H$ for the problem at hand must be established.
- Let $\langle A, R \rangle$ be an argumentation framework produced by instantiating the argument scheme. It can now be extended to an EAF $\langle A, R', D \rangle$ by defining:
  - $R' = R \cup \{(c_{ij}, c_{ik}) | c_{ij}, c_{ik} \in CD \cap CD_i, c_{ij} \neq c_{ik}\}$.
  - $D = \{(c_{ij}, (m_1, m_2)) | c_{ij} \in CD, PF_i(c_{ij}, m_1) \prec PF_i(c_{ij}, m_2)\}$. If c_{ij} justifies a preference of m_2 over m_1 then an attack relationship $c_{ij} \twoheadrightarrow (m_1 \rightharpoonup m_2)$ is added.

KING'S College LONDON

Case Study contd...

- 4 Relevant context domains were identified
- The preference arguments from the censoring context domain CD1 are:

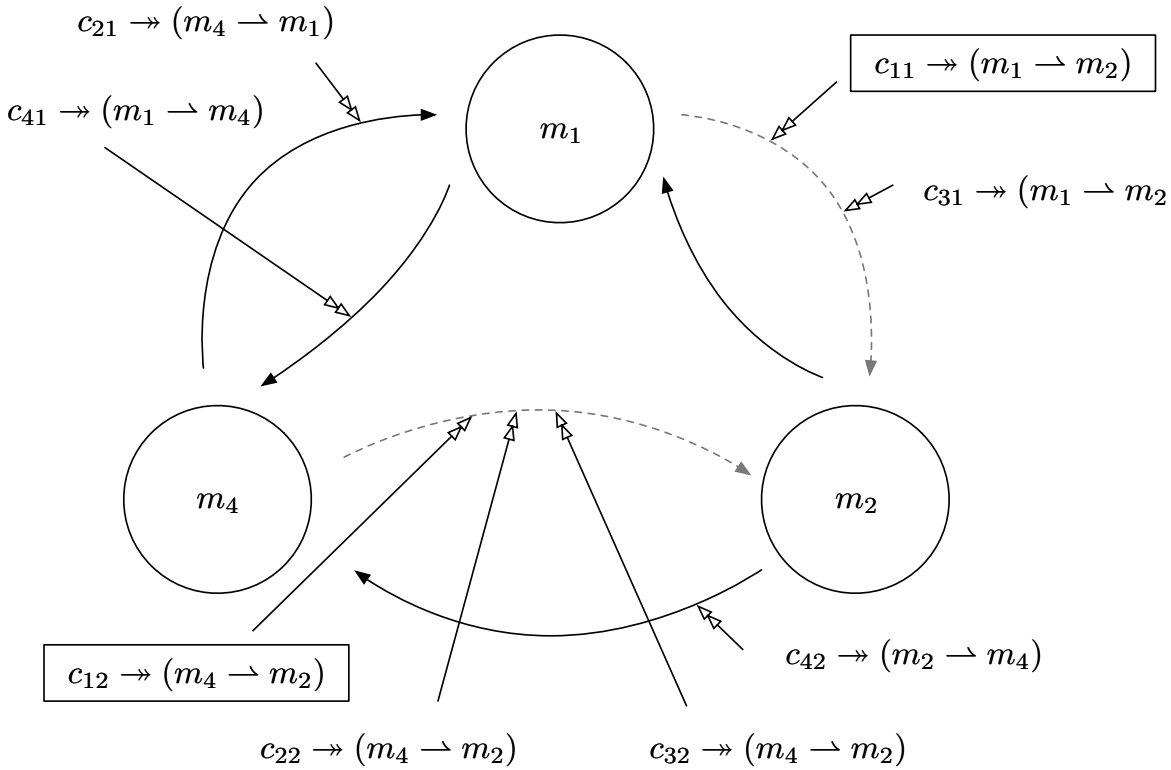
$$c_{11} \twoheadrightarrow (m_1 \rightharpoonup m_2)$$
$$c_{12} \twoheadrightarrow (m_4 \rightharpoonup m_2)$$

- Each CD results in a set of c_{ij}

KING'S College LONDON

Case Study contd...

- Assuming $CD_1 \succ CD_2 \succ CD_3 \succ CD_4$
- Applying only the preference arguments from CD1 results in m_2 being the only model argument that is acceptable with respect to $S'_{CD1} = \{c_{11}, c_{12}\}$



KING'S College LONDON

Future Work

Directions for further research:

- Prototype User survey
- Different domain
- Multiple Data sources

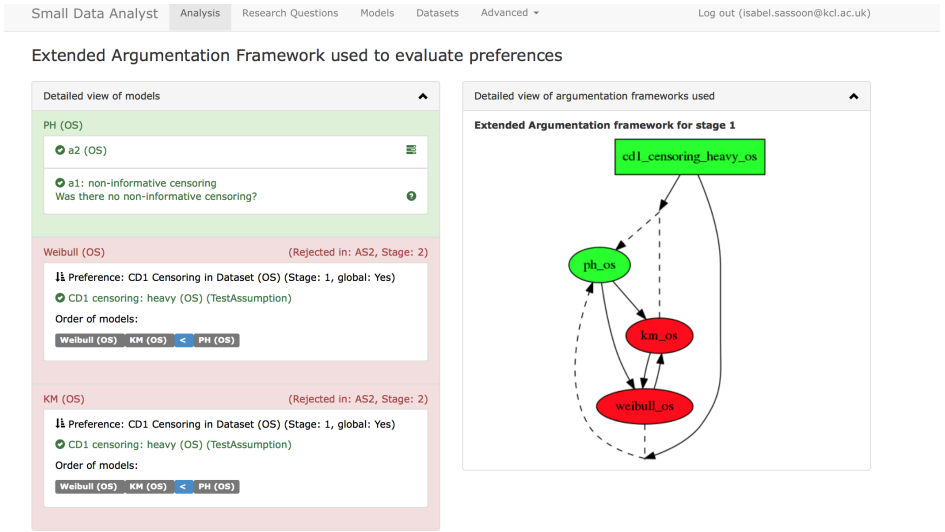
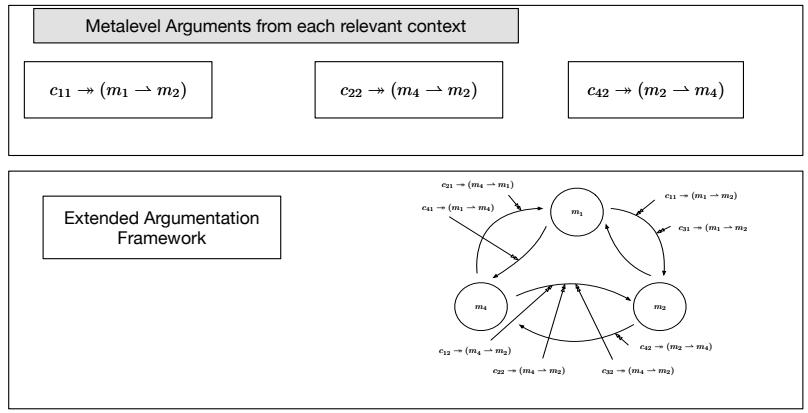


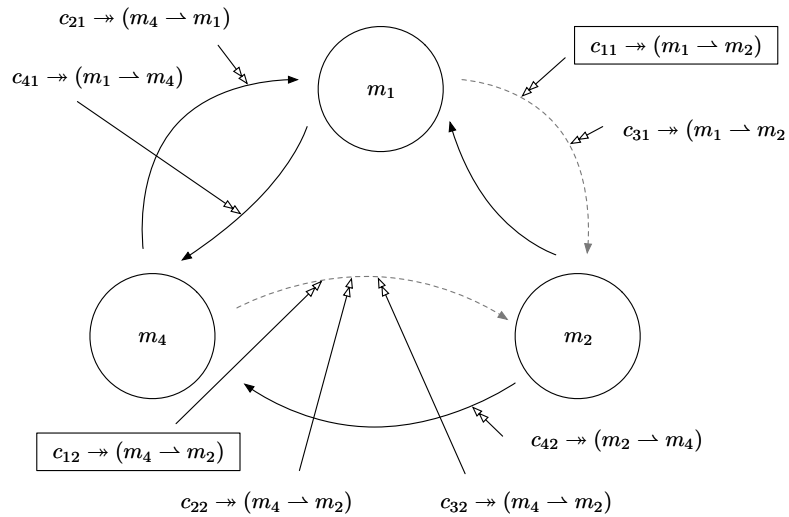
Figure: Screenshot of the prototype implementation - S Zillesen

KING'S College LONDON

Preferences in Argumentation for Statistical Model Selection

Questions?





KING'S College LONDON