Computational Models of Argument
P. Baroni et al. (Eds.)

doi:10.3233/978-1-61499-686-6-53

Preferences in Argumentation for Statistical Model Selection

Isabel SASSOON [a,1], Jeroen KEPPENS [a] and Peter MCBURNEY [a]
[a] *Department of Informatics, King's College London*

Abstract. The increase in routine clinical data collection coupled with an expectation to exploit this in support of evidence based decision making creates a need for an intelligent model selection system to support clinicians when analysing data because clinicians often lack the statistical expertise to do this independently. In a previous position paper, an argumentation based approach to devise a decision support system for such an application was introduced. This approach ignored the relative strength of arguments for and against alternative models. This paper demonstrates how an extended argumentation framework can be employed to capture and reason with statistical and research domain knowledge that affects the relative strength of arguments. The approach is validated by means of a real-world case study.

Keywords. argumentation schemes, preferences, automated analysis, decision support, model selection

1. Introduction

Answering a research question through statistical data analysis normally involves applying a particular statistical model or technique to the data. Software packages make the application of statistical methods easy but it is hard to determine which one to employ. The suitability of statistical approaches depends on the research question at hand, the assumptions underpinning the approaches and the extent to which they are satisfied by the data. Assessing the latter requires both statistical domain knowledge and an understanding of the data and how it has been collected.

This paper is part of a broader project that aims to address this problem by means an intelligent decision support system to aid with model selection. For the purposes of this paper, it is assumed that a clinician aims to analyse a research question by means of an existing data set. Sometimes, clinicians interact with statistical concepts at the design stage of a study, before data has been collected. Extending the approach to the latter scenario is left for future work. The research questions of interest extend beyond system identification by finding a ”best-fit” model for the data, and include hypothesis testing and other methods where the conclusions derived from statistical analysis are only valid in so far as an appropriate model has been applied.

Previously, we have proposed an approach to employ computational models of argumentation to identify the reasons to accept or reject the use of a statistical model [10].

[1]Corresponding Author: Isabel Sassoon, Department of Informatics,King's College London, Strand, WC2R 2LS London, UK; E-mail: isabel.sassoon@kcl.ac.uk.

Using Dung's argumentation framework [6], the resulting models enable identification of sets of accepted arguments and associated models. This approach ignores the relative strength of arguments. Statistical and application domain knowledge can inform appraisals of argument strength and can be modelled by means of preferences over arguments. This paper aims to address where such preferences emanate from, how they should be represented and how we might reason with them using existing argumentation approaches. The approach is validated by means of a case study in the medical domain, the initial results of which were published by Schilling et. al. [11].

2. Background

Within the clinical domain, clinicians are able to query and access many databases to explore and test research questions or perform hypothesis testing. A systematic review highlighted that while reporting of survival analysis results in journal publications had increased and the quality of the reporting of statistical analysis was improving slowly, only a low proportion of articles mention validation of model assumptions prior to use [1]. In our previous work we addressed the issue of the models to consider on the grounds of achieving the desired analytical objective and the underlying critical assumption testing. However as preferences are an important element in decision making, especially collective decision making, there is a need to leverage them as part of the model selection process. Within our model selection process preferences will be used to support the selection of the most appropriate model when more than one is possible, and given the clinician's research question and data.

In our position paper, we proposed an architecture for an argumentation based system to support the model selection process through the use of a knowledge base and an argumentation scheme [10]. A core component of this system is a statistical knowledge base (SKB) that defines the relations between research question type (R), research objectives (O), models (M) and assumptions (A). The SKB holds facts linking R, O, M, A in a way that will support the queries from the argumentation schemes. The SKB specifies multiple research question types. Each is linked to the objectives O that can fulfil that research question R. Models M are defined and linked to the respective objectives they are suitable for. For each model the critical assumptions that must be satisfied for the model to be applicable are identified. The relations and contents of the SKB are derived from statistical theory and best practice.
The elements of the SKB are denoted as follows:

- The set of *models*: $M = \{m_1, \dots, m_K\}$
- The set of *assumptions*: $A = \{a_1, \dots, a_P\}$
- The set of *objectives*: $O = \{o_1, \dots, o_Q\}$

The following relationships are defined in the SKB:

- $F : M \times O$ where $(m_k, o_q) \in F$ iff m_k fulfils objective o_q
- $C : M \times A$ where $(a_p, m_k) \in C$ iff a_p is a critical assumption for m_k
- $O : O \times O$ where $(o_r, o_q) \in O$ iff o_r is an alternative objective to o_q

A key benefit of the architecture proposed in [10] is that it differentiates knowledge into domain and problem specific information to be provided by the clinician, the problem independent domain specific statistical knowledge base and problem and domain independent argumentation schemes. This facilitates maintainability of the approach. However, our approach ignored subtle differences between the applicability of plausible models to a problem, such as the extent to which non-critical assumptions are not satisfied and contextual information that affects a model's suitability to meet the research objective. This work aims to capture such subtleties by modelling them by means of preferences over arguments.

A number of distinct approaches to represent and reason with preferences over arguments have been devised. Key approaches include Preference Argumentation Frameworks (PAF) [2, 4], Value-Based Argumentation Frameworks (VAF) [5] and Extended Argumentation Frameworks (EAFs) [9].

In VAFs, arguments are said to promote values and preferences over arguments derived from a preference ordering over values. Because the intelligent decision support system proposed herein aims to enable clinicians to answer research questions *objectively* supported by data, the choice of statistical model for performing an analysis rarely involves a conflict of values[2]. Thus, while VAFs enable a broad range of scenarios to be analysed, they are not a good fit for the problem at hand. Therefore, the remainder of this section focusses on PAFs, specifically in its incarnation of Argumentation Frameworks based on Contextual Preferences (CPAFs) [3], and EAFs [9].

3. Method

The objective of this paper is to define a preference ordering $Pref : M \times M$ over a set of models $M = \{m_1, \ldots, m_n\}$. However, such an ordering or orderings are not necessarily defined over the models directly. This section examines where the preferences for statistical model selection stem from, how they should be represented and which argumentation framework is suitable to infer decision support information based on those preference.

One source for preference orders is the statistical theory underpinning each model and dictating which models perform better when certain conditions are present in the data or the research question. For example, certain types of model are more resilient to particular features in the data, e.g. censoring or the proportion of case data lost to follow up, whereas others tend to become unreliable in such circumstances. Here, the presence of a particular feature causes a preference ordering over statistical models to arise. This relationship between a feature and an associated preference ordering is a matter of statistical knowledge. The presence of the feature may be determined by applying a test on the data or needs to be elicited from domain knowledge. In what follows, such preferences are called feature-based preferences.

A second source of preference orders is derived from model intent. There are different reasons for building a model when answering a research question. McBurney [8] explores the different purposes or reasons why a model can be used. In the context of statistical analysis the two most common intents for building a model on data are the need to predict or the need to explain (understand) the data. This is also covered in detail in [12]. In her article, Shmueli tackles the distinction between explanatory modelling and predictive modelling in detail and the implications these have on the choice of model

[2]It is understood this would be different in a scenario where statistical analysis aims to serve a *political* agenda.

CD1	Model	P_1
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 1. P_1 for model resilience to censoring

CD2	Model	P_2
predict	m_1 KM	avoid
	m_2 PH	suitable
	m_4 χ^2	avoid
explain	m_1 KM	suitable
	m_2 PH	suitable
	m_4 χ^2	neutral

Table 2. P_2 for model intent

to use. The definition of a good model will differ depending on whether we are looking for explanatory or predictive power, and this will reflect itself in an order of preference between models that can achieve a specific analytic objective. This preference order between models will change depending on the intent (purpose) of the analysis. In what follows, such preferences are called intent-based preferences.

Finally, there may be preference orders that are derived from the clinicians themselves. This could be due to the fact they are more familiar with a model, or that the literature they reference most makes use of a particular model. These preference orders can arise when more than one clinician is involved in an analysis and are an important factor within the decision making process. In what follows, such preferences are called domain-based preferences.

To incorporate preferences into the approach, the statistical knowledge base (SKB) introduced in [10] is extended with

- 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 $PF_i : CD_i \times M \mapsto P_i$.

For example, the feature-based preference ”resilience to censoring” can be modelled by a context domain $CD_1 = \{\text{absent}, \text{light}, \text{heavy}\}$ where the elements in the set correspond to features indicating distinct degrees to which censoring is present in the data. These can be defined more precisely in terms of proportion of records in the data affected but we avoid doing so to keep the example simple. The corresponding performance measure might be defined as $P_1 = \{\text{unaffected, affected}\}$. Table 1 presents an example of a performance function.

An example of an intent-based preference is $CD_2 = \{\text{predict}, \text{explain}\}$ where the performance measures would be defined as $P_2 = \{\text{suitable, neutral, avoid}\}$, as in Table 2. This can also be defined for domain-based preferences CD_3 where $P_3 = \{\text{preferred}, \text{neutral}\}$.

To construct an argumentation model based on the extended statistical knowledge base, first the set of context domains CD for the problem at hand must be established. CD contains contexts taken from the context domains in $\{CD_1, \ldots, CD_H\}$. Formally,

$CD \subseteq CD_1 \cup \ldots \cup CD_H$. Whether a context is relevant to a problem is derived by applying a test on the data, elicited from the domain expert/clinician or elicited from the research question. Where identification of the context is not straightforward, the contexts in CD provide hooks (conclusions) for further arguments about the appropriate statistical model.

Let $\langle Args, R \rangle$ be an argumentation framework produced using the method described previous in [10]. Such a model can now be extended to an EAF [9] $\langle Args, R', D \rangle$ by defining:

- $R' = R \cup \{(c_{ij}, c_{ik}) | c_{ij}, c_{ik} \in CD \cap CD_i, c_{ij} \neq c_{ik}\}$. Intuitively, R is extended with a symmetric attack relationship between each distinct pair of contexts in CD from the same context domain CD_i.
- $D = \{(c_{ij}, (m_1, m_2)) | c_{ij} \in CD, PF_i(c_{ij}, m_1) \prec PF_i(c_{ij}, m_2)\}$. Intuitively, an attack relationship $c_{ij} \twoheadrightarrow (m_1 \rightharpoonup m_2)$ is added for each attack of a model m_2 by a model m_1 where a context c_{ij} justifies a preference of m_2 over m_1.

The model can be enhanced further to take into account an importance order I of the context domains, if this is available. Let $\langle Args, R, D \rangle$ be the EAF, this can be extended to include I the importance of the context domains order by defining I as a complete or partial order on $CD \times CD$.

4. Case Study

The example used in this case study is derived from the ongoing collaboration with the Head and Neck Department at Guy's Hospital, King's College London (UK). The first published output of this work is in [11], and relies on a rich data set collected as part of the Sentinel European Node Trial (SENT). This data was collected as an observational study across 14 european centres and recruited a total of 415 patients who met the entrance criteria at diagnosis. The study commenced in 2005 and involves over 40 clinicians across the participating hospitals. The centres are periodically updating the current status of the patients in the trial. The main motivation for this trial was to assess whether sentinel node biopsy is a reliable and safe diagnostic technique in patients with early stage oral squamous cell carcinoma. The first output from this data answers the primary objective on patients with the potential for at least 3 years of follow up.

The data collected offers a cohort of data that can be exploited in support of answering many more clinician research questions or secondary objectives. There are a number of such analyses in progress initiated by different clinicians involved in SENT. An example of such a secondary analysis will be used as the case study in this paper. The research question is to identify whether there is a difference in survival between patients (within the SENT trial) who had so-called adjuvant therapy (such as Radiotherapy or Chemotherapy) to those that did not have any additional treatment.

By means of the approach presented in previous work [10], an argumentation framework $\langle Args, R \rangle$ is produced where $Args = \{m_1, m_2, m_4\}$, where each m_i is an argument supporting the use of a particular model and $R = \{(m_1, m_2), (m_1, m_4), (m_2, m_4), (m_2, m_1)$ $, (m_4, m_1), (m_4, m_2)\}$, which is the set of pairwise attacks between alternative models. Note that this is a substantial simplification of the argumentation model presented in [10]. The underlying assumptions have been omitted from the model as they are not necessary to understand how preferences are added.

To incorporate preferences over the models m_1, m_2 and m_4, four context domains need to be considered. The first (CD_1) corresponds to censoring. A query on the data has determined the presence of heavy censoring. Censoring can affect the reliability of the estimates obtained from some models, in this case both m_1 and m_4 are affected by heavy censoring. Using the context domain and performance function from Table 2, the following preference arguments c_{ij} arise: $c_{11} \twoheadrightarrow (m_1 \rightharpoonup m_2)$, $c_{12} \twoheadrightarrow (m_4 \rightharpoonup m_2)$.

The preference argument c_{11} is derived from the CD_1 and it attacks the attack of (m_1, m_2).

The second context domain (CD_2) corresponds to intent. In this case, the intent of the study is to explore or explain the data, therefore the context domain for model intent is relevant and preferences arising from the intent of explaining will be used. Using the context domain and performance function from Table 2, the following preference arguments arise: $c_{21} \twoheadrightarrow (m_4 \rightharpoonup m_1)$, $c_{22} \twoheadrightarrow (m_4 \rightharpoonup m_2)$.

The remaining context domains (CD_3 and CD_4) stem from clinician preferences. These are preferences expressed by different clinicians and result in a set of preference arguments that attack the attack of all arguments in support of all models except the one expressed by the clinician. The following preference arguments arise: $c_{31} \twoheadrightarrow (m_1 \rightharpoonup m_2)$, $c_{32} \twoheadrightarrow (m_4 \rightharpoonup m_2)$, $c_{41} \twoheadrightarrow (m_1 \rightharpoonup m_4)$, $c_{42} \twoheadrightarrow (m_2 \rightharpoonup m_4)$.

Finally, an importance ordering I, specifying that $CD_1 \succ CD_2 \succ CD_3 \succ CD_4$ is added to the argument framework. Depending on which context domains are pertinent to a specific analysis there may be an order on the context domain. The context domains that relate to statistical theory are more important in model selection than clinician preferences.

To recommend the most suitable model to apply for this analysis we would require a complete extension of this framework, which contains arguments in support of one model only. Without considering any preference arguments this argumentation framework contains only arguments that symmetrically attack each other. The introduction of preferences will enable the strengths of the arguments to the taken into consideration.

Applying CPAF to this argumentation framework using the above model yields a recommendation for the use of model m_2, irrespective of the approach used. The application of EAF to this situation does not yield any stable extensions, except the empty set. This is due to the relative importance of the preferences emanating from the different context domains not being exploited.

The preferences can be resolved in order to determine the recommended model by initially only considering the preference arguments from the most important context domain (CD_1). The preference arguments in the EAF attack the existing attacks between arguments in support of the models and their effect on the argumentation framework can be seen in Figure 1. In this case, m_2 is the only argument that is not strictly defeated and as such this would be the recommended model to be used, this would represent the stable extension to the argumentation framework. In this EAF, the justification to its choice over m_1 and m_4 is given by the context domain used in order to resolve this. In this case the recommendation of m_2 over the other models is explained by it being preferred under conditions of censoring.

If we assume that the order over the context domains is not known, then the extensions for the EAF can be computed for each CD_i in turn. The resulting extensions would be: $S_1 = \{m_2\}$, $S_2 = \{m_2\}$, $S_3 = \{m_2\}$ and $S_4 = \{m_4\}$ where S_i corresponds to the stable extension for CD_i. In other words model m_4 would only be selected in a situation where the preferences of clinician 2 are prioritised over all other contexts.

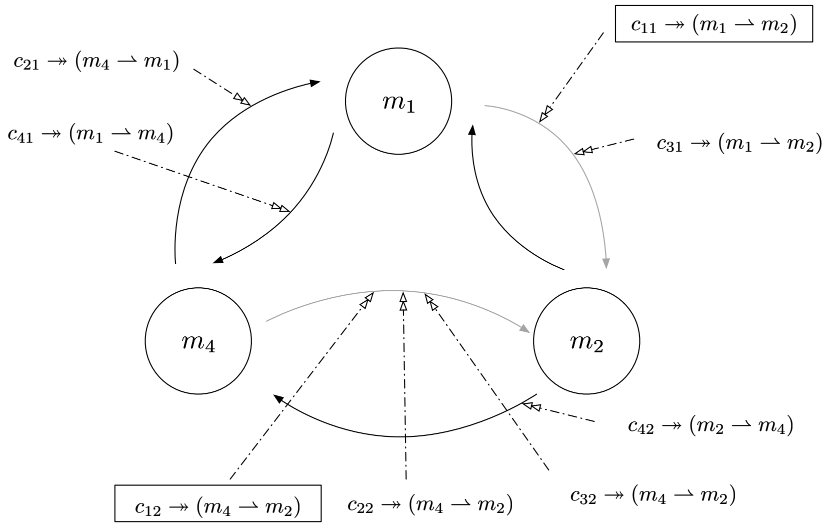

Figure 1. The preference arguments, considering only the preference arguments from context domain CD_1

5. Conclusion

This paper has presented an approach to develop a decision support tool to aid domain experts who collect data as part of their professional practice with choosing statistical techniques for analysis. This work has built on earlier work presented in [10]. Our proposed methods support the statistical model selection process by enabling contrasting preference orderings to be accounted for and reasoned with in order to recommend the most suitable model. This is achieved through EAFs and an extended statistical knowledge base. This approach can also take into account the relative importance of the different preference context domains, if this is applicable to the situation. Our proposed methodology for the inclusion of preferences enables the different types of preferences and their potential conflicts to be leveraged within the statistical model selection process, without statistical, informatics or administrative support.

The use of clinical preferences and argumentation to support decision making by clinicians has also been explored by Hunter et al [7]. In this paper, the aim is to offer the clinician the facility to aggregate evidence whilst taking into consideration the clinician's own assessment of the strength or weaknesses of each item of evidence. A clinician's preference may stem from the source of the evidence and is applied to the evidence used to evaluate the arguments, not on the arguments themselves. This method was evaluated by means of an actual trial with clinicians. The difference between our situation and the scenario considered in this paper is that in our case the preferences are not completely dependent on the clinician's view.

A prototype of the proposed system is being developed. This will offer the opportunity for the evaluation of the system using a range of case studies. Future work will focus on developing an ontology in support of a more flexible input method for the clinician's research question. This would enable clinicians to formulate their research questions using the terminology they may be more familiar with, as the ontology would relate it to the key concepts required by the proposed system to proceed with model selection. We also plan to address situations where the assumptions about the data available are removed. In such situations the data required to answer a research question may need to be extracted from multiple disparate sources, which may vary in provenance and quality. This would require methods able to handle multiple data sources, data matching, data quality and their impact on the proposed method for statistical model selection.

6. Acknowledgements

The authors would like to thank Mark McGurk for his generous support on this project.

References

[1] Abraira, V., Muriel, A., Emparanza, J. I., Pijoan, J. I., Royuela, A., Plana, M. N., Cano, A., Urreta, I., and Zamora, J. (2013). Reporting quality of survival analyses in medical journals still needs improvement. a minimal requirements proposal. *Journal of Clinical Epidemiology*, 66(12):1340–1346.

[2] Amgoud, L. and Cayrol, C. (2002). A reasoning model based on the production of acceptable arguments. *Annals of Mathematics and Artificial Intelligence*, 34(1-3):197–215.

[3] Amgoud, L., Parsons, S., and Perrussel, L. (2000). An argumentation framework based on contextual preferences. In *Proc. of FAPR'00, London*, pages 59–67.

[4] Amgoud, L. and Vesic, S. (2009). Repairing preference-based argumentation frameworks. In *IJCAI*, pages 665–670.

[5] Bench-Capon, T. J. (2003). Persuasion in practical argument using value-based argumentation frameworks. *Journal of Logic and Computation*, 13(3):429–448.

[6] P. M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77(2):321–358.

[7] Hunter, A. and Williams, M. (2010). Using clinical preferences in argumentation about evidence from clinical trials. In *Proceedings of the 1st ACM International Health Informatics Symposium*, IHI '10, pages 118–127, New York, NY, USA. ACM.

[8] McBurney, P. (2012). What are models for? In *Proceedings of the 9th European Conference on Multi-Agent Systems*, EUMAS'11, pages 175–188, Berlin, Heidelberg. Springer-Verlag.

[9] Modgil, S. (2009). Reasoning about preferences in argumentation frameworks. *Artificial Intelligence*, 173(9–10):901 – 934.

[10] Sassoon, I., Keppens, J., and McBurney, P. (2014). Towards argumentation for statistical model selection. In *5th International Conference on Computational Models of Argument*, pages 67–74. IOS Press.

[11] Schilling, C., Stoeckli, S. J., Sloan, P., McGurk, M., Sassoon, I., and others (2015). Sentinel european node trial (sent): 3-year results of sentinel node biopsy in oral cancer. *European Journal of Cancer*, 51(18):2777 – 2784.

[12] Shmueli, G. (2010). To explain or to predict? *Statistical Science*, pages 289–310.