Proceedings of the 21st International Conference on Principles of Knowledge Representation and Reasoning
Main Track

Preference-Based Abstract Argumentation for Case-Based Reasoning

Adam Gould[1] , **Guilherme Paulino-Passos**[1] , **Seema Dadhania**[2] , **Matthew Williams**[2] , **Francesca Toni**[1]

[1]Department of Computing, Imperial College London, UK
[2]Department of Surgery and Cancer, Imperial College London, UK
{adam.gould19, gppassos, s.dadhania, matthew.williams, f.toni }@imperial.ac.uk,

Abstract

In the pursuit of enhancing the efficacy and flexibility of interpretable, data-driven classification models, this work introduces a novel incorporation of user-defined preferences with *Abstract Argumentation* and *Case-Based Reasoning* (CBR). Specifically, we introduce *Preference-Based Abstract Argumentation for Case-Based Reasoning* (which we call AA-CBR-$\mathcal{P}$), allowing users to define multiple approaches to compare cases with an ordering that specifies their preference over these comparison approaches. We prove that the model inherently follows these preferences when making predictions and show that previous abstract argumentation for case-based reasoning approaches are insufficient at expressing preferences over constituents of an argument. We then demonstrate how this can be applied to a real-world medical dataset sourced from a clinical trial evaluating differing assessment methods of patients with a primary brain tumour. We show empirically that our approach outperforms other interpretable machine learning models on this dataset.

1 Introduction

Abstract argumentation is a formalism for representing arguments and relationships between them, and for computing which arguments to accept (Dung 1995). It has been shown to be effective for recommendation systems (Rago et al. 2020), decision-making tasks (Amgoud and Prade 2009), reasoning with incomplete knowledge (Briguez et al. 2014) and approaches to explainable artificial intelligence (Čyras et al. 2021b).

Case-based reasoning (CBR) is a problem-solving methodology where new problems are solved by retrieving and adapting solutions from similar past cases. Approaches to combine CBR with abstract argumentation are successful for explaining the output of machine learning (Prakken and Ratsma 2022) or for making predictions (Cocarascu et al. 2020). *Abstract Argumentation for Case-Based Reasoning* (AA-CBR) (Čyras, Satoh, and Toni 2016) is a data-driven interpretable classification and explanation model which has been shown to have utility in many tasks, for example, as an interpretable binary classifier (Cocarascu, Čyras, and Toni 2018; Cocarascu et al. 2020; Paulino-Passos and Toni 2023), for cautiously monotonic reasoning (Paulino-Passos and Toni 2021a) or for explaining known legislative outcomes (Čyras et al. 2019).

As a result, AA-CBR represents a novel paradigm as an intrinsically explainable classification model. The need for such models in high-stakes decision-making is becoming increasingly apparent. In medical domains, for instance, decision-support tools need to be understood by a clinician for trust to be established and disagreements to be resolved (Amann et al. 2020). Moreover, legal requirements, such as those established by GDPR (Goodman and Flaxman 2017), require explainable AI to varying extents.

However, previous approaches to AA-CBR are missing the ability to compare cases by user-defined preferences. Preferences allow stakeholders to influence a decision-making tool and inject domain-specific knowledge, resulting in more desirable reasoning systems or better-performing models. Preferences have been integrated with argumentation systems through a variety of approaches, for example, preferences defined over abstract arguments in preference-based argumentation frameworks (Amgoud and Cayrol 1998); preferences defined over structured arguments in ASPIC+ (Modgil and Prakken 2013); preferences defined over values assigned to arguments as in value-based argumentation frameworks (Bench-Capon 2003); preferences defined by defeats of attack relations as with extended argumentation frameworks (Modgil 2009); preferences defined over constituents of an argument such as with Assumption-Based Argumentation with Preferences (ABA+) (Cyras and Toni 2016); and preferences over constituents of cases in which arguments are derived such as with case models[1] (Zheng, Grossi, and Verheij 2020).

Within healthcare, preferences allow for specialist expertise and patient goals to affect the decision-making process, leading to more agreeable courses of action (Politi et al. 2013). Moreover, preferences with argumentation have many applications within the medical domain (Čyras et al. 2021a; Kakas, Moraitis, and Spanoudakis 2018; Hunter and Williams 2012; Williams et al. 2015).

Despite this, there is yet to be an AA-CBR-based approach that can integrate preferences over the constituents of arguments. Considering the benefits of AA-CBR as an interpretable classification model, adding preferences will

[1]Literature on AA-CBR, including this work, uses the terms 'arguments' and 'cases' interchangeably, however, in work on case models, these concepts are separate.

improve the model's utility, flexibility, and performance. In this work, we address the limitations of AA-CBR in this space and contribute the following:

- Introduction of *Preference-Based Abstract Argumentation for Case-Based Reasoning* (AA-CBR-$\mathcal{P}$), a model that allows preferences to be defined over comparison methods of cases.
- Demonstration that AA-CBR-$\mathcal{P}$ cannot be captured by existing AA-CBR methods.
- Prove that AA-CBR-$\mathcal{P}$ inherently respects preferences when making classifications, where possible.
- Conduct an empirical study on a real-world medical data set sourced from a clinical trial on patients with a primary brain tumour.
- Showcase how AA-CBR-$\mathcal{P}$ models with preferences improve performance compared to existing AA-CBR approaches and outperform other interpretable machine learning models.

In Section 2 we provide the necessary preliminaries. We then demonstrate the motivation for integrating preferences with AA-CBR in Section 3. In Section 4, we introduce AA-CBR-$\mathcal{P}$. We explore properties in Section 5 and conduct an empirical evaluation of the models on a real-world medical dataset in Section 6. Section 7 discusses related work. We conclude in Section 8.

2 Preliminaries

2.1 Abstract Argumentation

An *abstract argumentation framework (AF)* (Dung 1995) is a pair $\langle Args, \rightsquigarrow \rangle$ where $Args$ is a set of arguments and $\rightsquigarrow \subseteq Args \times Args$ is a binary relation between arguments. For arguments, $\alpha, \beta \in Args$, α *attacks* β if $\alpha \rightsquigarrow \beta$. An AF can be represented as a directed graph where nodes are arguments and edges represent attacks. A set of arguments $E \subseteq Args$ *defends* an argument $\beta \in Args$ if for all $\alpha \rightsquigarrow \beta$ there exists $\gamma \in E$ such that $\gamma \rightsquigarrow \alpha$. For determining which arguments to accept, we focus on the *grounded extension* (Dung 1995), which can be iteratively computed as $\mathbb{G} = \bigcup_{i \geq 0} G_i$, where G_0 is the set of unattacked arguments and $\forall i \geq 0$, G_{i+1} is the set of all arguments that G_i defends.

2.2 Abstract Argumentation for Case-Based Reasoning (AA-CBR)

AA-CBR is a binary classification model utilising argumentation as the reasoner. It operates on a *casebase* D comprised of labelled examples of a generic *characterisation*. Given D and a new *unlabelled example*, N, AA-CBR assigns a label to the new case. Formally:

Definition 1 (Adapted from (Paulino-Passos and Toni 2021a; Cocarascu et al. 2020))**.** Let $D \subseteq X \times Y$ be a finite *casebase* of labelled examples where X is a set of *characterisations* and $Y = \{\delta, \bar{\delta}\}$ is the set of possible outcomes. Each example is of the form (x, y). Let (x_δ, δ) be the *default argument* with δ the *default outcome*. Let N be an *unlabelled example* of the form $(x_N, y_?)$ with $y_?$ an unknown outcome. The function AA-CBR(D, x_N) assigns the new case an outcome as follows:

$$\text{AA-CBR}(D, x_N) = \begin{cases} \delta & \text{if } (x_\delta, \delta) \in \mathbb{G}, \\ \bar{\delta} & \text{otherwise.} \end{cases}$$

where $\mathbb{G}$ is the grounded extension of the argumentation framework derived from D and x_N, known as $AF(D, x_N)$.

$AF(D, x_N)$ is constructed with cases of differing outcomes modelled as arguments. It is assumed that characterisations of the data points are equipped with a partial order $\succcurlyeq$, which determines the direction of attacks within the casebase and is used to ensure attacks occur between cases with minimal difference. The new case $(x_N, y_?)$ is added to the AF by attacking cases considered *irrelevant* to it, as defined by a provided irrelevance relation $\not\sim$.

Definition 2 (Adapted from (Paulino-Passos and Toni 2021a; Cocarascu et al. 2020))**.** Let $\succcurlyeq$ and $\not\sim$ be a partial order and binary relation defined over X, respectively. The argumentation framework $AF(D, x_N)$ mined from D and x_N is $(Args, \rightsquigarrow)$ in which:

- $Args = D \cup \{(x_\delta, \delta)\} \cup \{N\}$
- for $(x_\alpha, y_\alpha), (x_\beta, y_\beta) \in D \cup \{(x_\delta, \delta)\}$, it holds that $(x_\alpha, y_\alpha) \rightsquigarrow (x_\beta, y_\beta)$ iff
  1. $y_\alpha \neq y_\beta$, and
  2. either $x_\alpha \succ x_\beta$ and $\nexists (x_\gamma, y_\alpha) \in D \cup \{(x_\delta, \delta)\}$ with $x_\alpha \succ x_\gamma \succ x_\beta$
  3. or $x_\alpha = x_\beta$;
- for $(x_\alpha, y_\alpha) \in D \cup \{(x_\delta, \delta)\}$, it holds that $N \rightsquigarrow (x_\alpha, y_\alpha)$ iff $N \not\sim (x_\alpha, y_\alpha)$.

A casebase D is *coherent* iff there are no two cases $(x_\alpha, y_\alpha), (x_\beta, y_\beta) \in D$ such that $x_\alpha = x_\beta$ and $y_\alpha \neq y_\beta$, and it is *incoherent* otherwise.

The second bullet of Definition 2 defines attacks between the cases in the casebase. Condition **2** ensures that attacks occur from greater cases to smaller ones according to the provided partial order. When multiple possible attacks occur, we enforce that attacks originate from the case with minimal difference compared to the attacked case. We refer to this as the most concise possible attack. Condition **3** defines symmetric attacks for an incoherent casebase.

2.3 AA-CBR with Stages

An extension to AA-CBR (Čyras et al. 2019) employs *stages* in the characterisations to represent dynamic features. Stages represent the time at which a case was recorded. If two cases have the same set of features, then the time measure is used to distinguish the cases. If the only difference between two cases is the time measure and the outcome, then it is possible that the change in outcome by the latter case is a result of features that have not been recorded. This approach recognises that not all data may be available and so uses a time measure as a proxy to reason about unknown data. This version of AA-CBR is not defined generally for any characterisation but instead only for a set of features and stages.

Specifically, each case is characterised by $\mathbb{F} \times \mathbb{S}$, where $\mathbb{F}$ is a set of features and $\mathbb{S}$ represents the set of all subsequences of $\langle s_1, \ldots, s_n \rangle$. A subsequence refers to either the empty sequence $\langle\rangle$ or a contiguous sequence of elements, starting from s_1 and can contain any number of consecutive elements up to and including s_n. A case may be represented as $((F_\alpha, S_\alpha), y_\alpha)$, where F_α is the set of features, S_α is the stages and y_α is the outcome. For clarity, we will represent cases without the inner brackets, i.e. $(F_\alpha, S_\alpha, y_\alpha)$. The subsequence relation $\sqsubseteq$ can be defined as in (Čyras et al. 2019): for $S_\alpha, S_\beta \in \mathbb{S}$, $S_\beta \sqsubseteq S_\alpha$ iff either $S_\beta = \langle s_1, \ldots, s_k \rangle$, $S_\alpha = \langle s_1, \ldots, s_m \rangle$ and $k \leq m \leq n$ or $S_\beta = \langle\rangle$, the empty sequence. We can then define the $AF_\mathbb{S}(D, (F_N, S_N))$ mined from D and N for AA-CBR with Stages, as follows:

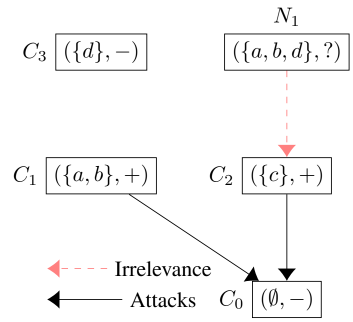
Definition 3 (Adapted from (Čyras et al. 2019))**.** The argumentation framework $AF_\mathbb{S}(D, (F_N, S_N))$ corresponding to a casebase D, a default outcome $\delta \in Y$ and a new case $N = (F_N, S_N, y_?)$, is $(Args, \rightsquigarrow)$ such that:

- $Args = D \cup \{(\emptyset, \langle\rangle, \delta)\} \cup \{(F_N, S_N, y_?)\}$
- $(\emptyset, \langle\rangle, \delta)$ is called the default argument
- For $(F_\alpha, S_\alpha, y_\alpha), (F_\beta, S_\beta, y_\beta) \in Args$, it holds that $(F_\alpha, S_\alpha, y_\alpha) \rightsquigarrow (F_\beta, S_\beta, y_\beta)$ iff
  1. $y_\alpha \neq y_\beta$, and (different outcomes)
  2. either
     (a) $F_\alpha \supset F_\beta$, and (specificity)
     (b) $\nexists (F_\gamma, S_\gamma, y_\alpha) \in D$ with (concision)
        i either $F_\alpha \supset F_\gamma \supset F_\beta$,
        ii or $F_\gamma = F_\alpha$ and $S_\alpha \sqsupset S_\gamma$,
        iii or $F_\beta = F_\gamma$ and $S_\alpha \sqsupseteq S_\gamma \sqsupset S_\beta$;
  3. or
     (a) $F_\beta = F_\alpha$ and $S_\alpha \sqsupset S_\beta$ and (advance)
     (b) $\nexists (F_\gamma, S_\gamma, y_\alpha) \in D$ with $F_\gamma = F_\alpha$ and $S_\alpha \sqsupset S_\gamma \sqsupset S_\beta$; (proximity)
- for $(F_\alpha, S_\alpha, y_\alpha) \in D \cup \{(\emptyset, \langle\rangle, \delta)\}$, it holds that $(F_N, S_N, y_?) \rightsquigarrow (F_\alpha, S_\alpha, y_\alpha)$ iff either $F_N \nsupseteq F_\alpha$ or $S_N \not\sqsupseteq S_\alpha$.

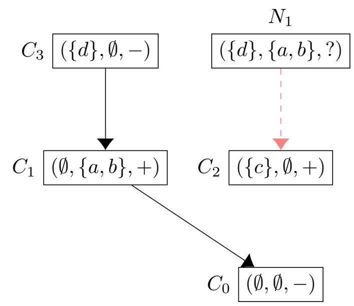
Intuitively, a case $(F_\alpha, S_\alpha, y_\alpha)$ attacks a case $(F_\beta, S_\beta, y_\beta)$ iff their outcomes differ (condition **1**) and one of the following conditions is met. Firstly, $(F_\alpha, S_\alpha, y_\alpha)$ is more specific than $(F_\beta, S_\beta, y_\beta)$ and is the most concise such case. If there are multiple such cases, $(F_\alpha, S_\alpha, y_\alpha)$ must be the closest by stages (condition **2**). Secondly, if $(F_\alpha, S_\alpha, y_\alpha)$ has advanced further and changed the outcome, it suggests that $(F_\beta, S_\beta, y_\beta)$ might be missing some features expected to be acquired as it progresses through the stages. Proximity identifies the earliest stage when the outcome change is expected (condition **3**).

3 Motivations

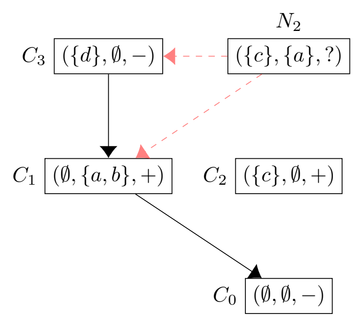
The partial order used to define attacks in the casebase is a key choice affecting the performance and explanations generated by an AA-CBR model. Selecting a different partial order at model construction may result in varying model performance even when using the same dataset. Additionally, different partial orders have different semantic meanings, so the explanations will differ when describing how an AA-CBR model made its final prediction. However, using a single partial order limits users' ability to express domain-specific preferences over how cases are compared.



(a) With $\succcurlyeq = \supseteq$ and $\not\sim = \nsupseteq$, the prediction for N_1 is $+$, ignoring case C_3, which a clinician might expect to be used if feature d is considered more important for this task.



(b) With $\succcurlyeq = \succcurlyeq_{lex}$ and $\not\sim = \not\sim_{lex}$, the prediction for N_1 is $-$, where C_3 is now instrumental in this classification.



(c) With $\succcurlyeq = \succcurlyeq_{lex}$ and $\not\sim = \not\sim_{lex}$, the prediction for N_2 is $-$, despite c being a high-priority feature associated with outcome $+$ used in case C_2, which is ignored in this classification.

Figure 1: A comparison of scenarios utilising AA-CBR with a naive attempt to introduce preferences

Example 1. Consider a simple medical assessment tool for assessing patients' well-being. The tool represents patient features using sets and classifies patients as either being in 'good' health $(+)$ or 'poor' health $(-)$.

A past patient (C_1) who gets regular exercise (feature a) and has a healthy diet (feature b) is in good health is represented by the case $(\{a, b\}, +)$. Similarly, a patient (C_2) who has adequate sleep (feature c) is represented by the case $(\{c\}, +)$. In contrast, a patient (C_3) who currently has an infection (feature d) and is considered in poor health is rep-

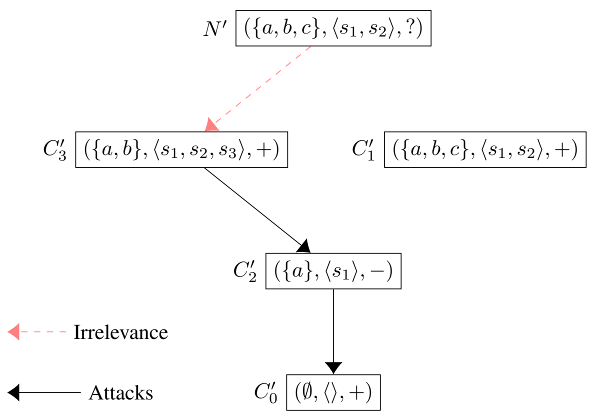


Figure 2: An AF derived from AA-CBR with Stages as defined in Definition 3. Despite C_1' having features and stages equal to N', the outcome predicted for N' is $-$, which differs from C_1''s outcome. The nearest cases property does not hold.

resented by the case $(\{d\}, -)$. If we have no information from a patient, we act cautiously and assume they have a poor health outcome, so the default case (C_0) is $(\emptyset, -)$.

Figure 1a shows the AF derived from AA-CBR to classify new case N_1 with features $\{a, b, d\}$. Here, we use $\succcurlyeq = \supseteq$ and $\not\sim = \not\supseteq$ as in (Čyras, Satoh, and Toni 2016) where casebase cases only attack those with an opposing outcome and a subset of features. N_1 attacks C_2 by irrelevance as C_2 presents with feature c, which N_1 does not present with. Case C_1 is unattacked, and attacks C_0, so the default case is not accepted. The grounded extension is $\{N_1, C_3, C_1\}$. As a result, we classify the new case with outcome $+$. Case C_3 is not used in this classification, and thus neither is feature d.

The outcome in Example 1 may be counter-intuitive to a clinician deeming that having an infection (d) and sleeping well (c) are more important for this health assessment. Thus, despite regular exercise (a) and a healthy diet (b), the clinician may wish for the new case to be classified as having a poor health outcome. Because C_3 contains an important feature (d), the clinician would expect the classification to utilise this case. Introducing preferences over constituent parts of the argument structure can thus change the ordering of the arguments to achieve the desired behaviour. The partial order used can be modified to create preferences between groups of features.

Example 2. Let us split the characterisation of each data point into two sets: a high-priority feature set and a low-priority feature set.[2] Between two cases, the one that has a superset of high-priority features attacks, but if the two cases have the same high-priority feature set, then the one that has a superset of low-priority features attacks. This lexicographic application can be achieved by a single partial order defined as follows:

Let $X \subseteq \mathbb{F}_H \times \mathbb{F}_L$ where $\mathbb{F}_H = \{c, d\}$ is the set of possible high-priority features and $\mathbb{F}_L = \{a, b\}$ is the possible set of low-priority features. A case is of the form $((H, L), y)$ where H contains the high-priority features and L the low-priority features. Throughout this paper, we will remove the inner brackets when presenting cases for clarity.

Let $(H_\alpha, L_\alpha) \supseteq_H (H_\beta, L_\beta)$ iff $H_\alpha \supseteq H_\beta$ and $(H_\alpha, L_\alpha) \supseteq_L (H_\beta, L_\beta)$ iff $L_\alpha \supseteq L_\beta$. The AA-CBR model can then be instantiated with a lexicographic application of these orders, $\succcurlyeq_{lex}$, where $(H_\alpha, L_\alpha) \succcurlyeq_{lex} (H_\beta, L_\beta)$ iff

1. $(H_\alpha, L_\alpha) \supset_H (H_\beta, L_\beta)$ or
2. $(H_\alpha, L_\alpha) =_H (H_\beta, L_\beta)$ and $(H_\alpha, L_\alpha) \supseteq_L (H_\beta, L_\beta)$.

We define the notion of irrelevance such that the attacks by the new case are the same as in Example 1:
$(H_\alpha, L_\alpha) \not\sim_{lex} (H_\beta, L_\beta)$ iff $H_\alpha \not\supseteq H_\beta$ or $L_\alpha \not\supseteq L_\beta$

The AA-CBR model is instantiated with $\succcurlyeq = \succcurlyeq_{lex}$ and $\not\sim = \not\sim_{lex}$ and the cases in Example 1 are modified moving features to their respective priority feature sets.

The resulting AF is shown in Figure 1b. Now C_3 defends the default case because d is a higher priority feature not present in C_1. The grounded extension is thus $\{N_1, C_3, C_0\}$. The classification for N_1 is therefore $-$, as desired.

However, C_1 is now considered a more concise case (in the sense of Definition 2, item 2) to the default case compared to C_2, preventing C_2 from attacking the default case. If we had another new case, N_2, with features $\{a, c\}$, the expected outcome should be $+$, as both these features are only ever associated with this outcome in the casebase. Furthermore, C_2 should contribute to the classification as c is a high-priority feature. This situation is modelled in Figure 1c. With $\succcurlyeq_{lex}$, C_2 is not used, and again, a high-priority feature is ignored. The grounded extension is $\{N_2, C_2, C_0\}$, resulting in the counter-intuitive outcome, $-$, predicted for N_2.

This illustrates how AA-CBR cannot effectively consider preferences over constituent parts of an argument. We require a new form of AA-CBR that respects these preferences by definition. Such a model should allow for an arbitrary number of preferences to be defined with no restrictions on how cases are characterised.

Definition 3 showcases a notion of preferences over comparisons of cases which introduced stages. This approach, however, is not generalised to work with characterisations beyond features and stages, nor allows for an arbitrary number of preferences to be defined. Furthermore, a useful property of AA-CBR relies on *nearest cases*, which are cases that are "closest" by $\succcurlyeq$[3]: if all cases that are nearest to the new case agree on an outcome, then this is the outcome predicted. Whilst no definition of nearest cases is provided for AA-CBR with Stages in (Čyras et al. 2019), we highlight an example where this property would clearly be violated.

Example 3. Consider the casebase $D' = \{C_0', C_1', C_2', C_3'\}$. The AF derived from AA-CBR with Stages for new case N' is shown in Figure 2. Note that the features and stages of C_1' are the same as in N'. We, therefore, expect the predicted outcome for N' to be the same $(+)$. However, this is not the case. The grounded extension is $\{N', C_1', C_2'\}$ and the

[2]This split is for presentation purposes only - each example contains the same features as in Example 1.

[3]We will provide a formal definition of nearest cases in our setting in Section 5.

predicted outcome for N' is $-$. Clearly, the nearest case condition does not hold.

Consequently, we cannot simply generalise the approach in AA-CBR with Stages. We need a new method that meets the requirements for allowing preferences to be defined and obeyed while ensuring that the nearest case property holds.

4 Preference-Based Abstract Argumentation for Case-Based Reasoning

We introduce AA-CBR-$\mathcal{P}$ and its regular variant AA-CBR-$\mathcal{P}_{\langle\rangle}$, where a user can employ various preorders[4] defined over constituent parts of an argument. The preorders are sorted by preferences and applied using a lexicographic strategy. We begin by defining the collection of preorders.

Definition 4 (Preference Ordering)**.** Let $\mathcal{P} = \langle \succcurlyeq_1, \ldots, \succcurlyeq_n \rangle$ be a sequence of preorders, each defined over X. Each preorder, $\succcurlyeq_i$, is a reflexive and transitive relation, with a corresponding strict preorder, $\succ_i$ (which is irreflexive and transitive), and an equivalence relation, $=_i$ (which is reflexive, symmetric, and transitive). For cases $(x_\alpha, y_\alpha), (x_\beta, y_\beta) \in D \cup \{(x_\delta, \delta)\}$, we define:

- $x_\alpha \succ_i x_\beta$ iff $x_\alpha \succcurlyeq_i x_\beta$ and $x_\beta \not\succcurlyeq_i x_\alpha$
- $x_\alpha =_i x_\beta$ iff $x_\alpha \succcurlyeq_i x_\beta$ and $x_\beta \succcurlyeq_i x_\alpha$

In simple terms, the ordering of $\mathcal{P}$ determines the preferences of the preorders. If two cases are equivalent by the first preorder, then a comparison under the second order can be used. If they are also equivalent by the second order, the third can be used. This process continues, using subsequent preorders when equivalence occurs. By incorporating multiple methods for comparing cases, we afford flexibility to select the approach that optimises performance and injects domain relevance. Consequently, we can create a new AA-CBR-based model that utilises $\mathcal{P}$. To do so, we make use of the following shorthand:

Notation 1. We use $x_\alpha \succcurlyeq_{[j:k]} x_\beta$, where $j \leq k$, to mean x_α is equivalent or larger than x_β on all preorders between orders $\succcurlyeq_j$ and $\succcurlyeq_k$ (inclusive). Formally:

$$x_\alpha \succcurlyeq_{[j:k]} x_\beta \text{ iff } \forall i \in [j,k], x_\alpha \succcurlyeq_i x_\beta.$$

Shorthands for $=_{[j:k]}$ and $\succ_{[j:k]}$ are defined analogously.

With this, we can create a definition similar to Definition 2 that enforces attacks in the direction defined by the sequence of partial orders and that attacks only originate from the most concise case possible, thus representing minimal change between cases. With the addition of multiple preorders, the concision condition must now consider the lexicographic application of the orders. We first define a notion of *potential attacks* for a particular order $\succcurlyeq_i$, which describes attacks that would transpire if we do not consider concision.

[4]Previous incarnations of AA-CBR use a partial order; as we use multiple orders to compare cases, each can be a preorder. For example, we have $(\{c,d\},\{a\}) \supseteq_H (\{c,d\},\{b\})$ and $(\{c,d\},\{b\}) \supseteq_H (\{c,d\},\{a\})$ but these two characterisations are not equivalent.

Definition 5 (Potential Attacks)**.** Let $\alpha = (x_\alpha, y_\alpha)$ and $\beta = (x_\beta, y_\beta)$. For $\alpha, \beta \in D \cup \{(x_\delta, \delta)\}$, we define a *potential attack* on order $\succcurlyeq_i$ as: $\alpha \overset{\text{p}}{\rightsquigarrow}_i \beta$ iff

i $y_\alpha \neq y_\beta$, and
ii $x_\alpha \succ_i x_\beta$, and
iii $x_\alpha =_{[1:i-1]} x_\beta$.

Intuitively, α potentially attacks β on order $\succcurlyeq_i$ when the outcomes of the cases are different (**i**), x_α is strictly greater than x_β by order $\succ_i$ (**ii**) and they are equivalent on all orders before $\succ_i$ (**iii**). Conditions (**ii**) and (**iii**) apply $\mathcal{P}$ lexicographically. We can subsequently restrict attacks to the most concise attacks.

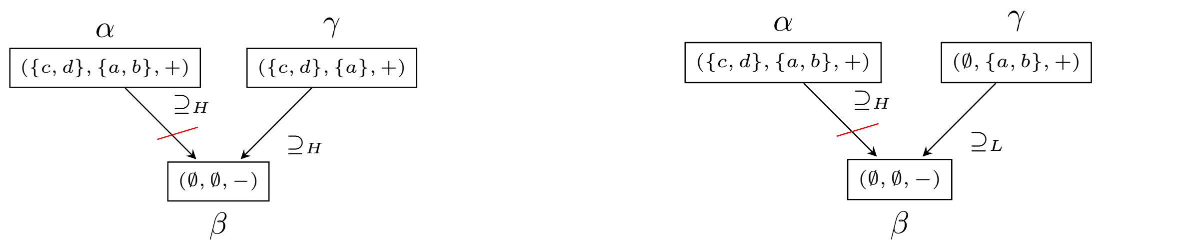
Definition 6 (Casebase Attacks)**.** Let $\alpha = (x_\alpha, y_\alpha)$ and $\beta = (x_\beta, y_\beta)$. For $\alpha, \beta \in D \cup \{(x_\delta, \delta)\}$, we define an attack on order $\succcurlyeq_i$: $\alpha \rightsquigarrow_i \beta$ iff

i $\alpha \overset{\text{p}}{\rightsquigarrow}_i \beta$ and
ii $\nexists \gamma = (x_\gamma, y_\alpha) \in D \cup \{(x_\delta, \delta)\}$ with $x_\alpha \succcurlyeq_{[1:n]} x_\gamma$ and
  (a) either $\gamma \overset{\text{p}}{\rightsquigarrow}_i \beta$ and $\exists l \geq i, x_\alpha \succ_l x_\gamma$,
  (b) or $\gamma \overset{\text{p}}{\not\rightsquigarrow}_i \beta$ and $\exists l > i, \gamma \overset{\text{p}}{\rightsquigarrow}_l \beta$.

This definition states that case α attacks case β on order $\succcurlyeq_i$ if two conditions are met. Firstly, α potentially attacks β on order $\succcurlyeq_i$ (**i**), and secondly, α is the most concise case capable of such an attack (**ii**). We illustrate the two conditions of concision in Figure 3, utilising the same characterisation approach as Example 2 and letting $\mathcal{P} = \langle \supseteq_H, \supseteq_L \rangle$.

(a) Condition **a** details the circumstance where both α and γ can potentially attack β on order $\succcurlyeq_i$ and x_γ is smaller than x_α on at least one of the orders in $\mathcal{P}$. As both α and γ are potential attackers of β, we know that they are both equivalent to β on all orders up to and including $\succcurlyeq_{i-1}$, that is $x_\alpha =_{[1:i-1]} x_\gamma =_{[1:i-1]} x_\beta$. Furthermore, as $x_\alpha \succcurlyeq_{[1:n]} x_\gamma$, we know there is an l' such that $l \geq l' \geq i$ in which $x_\alpha =_{[i-1:l'-1]} x_\gamma$ and $x_\alpha \succ_{l'} x_\gamma$. This is to say, both x_α and x_γ are equivalent on all orders up to some order $\succcurlyeq_{l'}$ in which x_α is greater than x_γ. As a result, we can take γ as the more concise case. In Figure 3a, α potentially attacks β by $\supseteq_H$ but is blocked by the more concise case γ because they are equivalent on the first preorder, $\supseteq_H$, i.e. both have features $\{c,d\}$, but γ is less specific on the second preorder as $\{a,b\} \supseteq \{a\}$.

(b) Condition **b** covers the instance where α is greater than γ on $\succcurlyeq_i$ but γ cannot potentially attack β on order $\succcurlyeq_i$ as γ and β are equivalent on all orders up to and including $\succcurlyeq_i$. γ is, therefore, only a more concise case if it can potentially attack by some order $\succcurlyeq_l$ where $l > i$. This is because, by $x_\alpha \succcurlyeq_{[1:n]} x_\gamma$, we have that α is greater than γ on all orders up to an including order $\succcurlyeq_l$ that is $x_\alpha \succcurlyeq_{[1:l]} x_\gamma$. Therefore, we know that $x_\alpha \succcurlyeq_l x_\gamma \succ_l x_\beta$ and thus can take γ as a more concise case. In Figure 3b α potentially attacks β by $\supseteq_H$ but is blocked by γ which has an equivalent high-priority feature set as β (both of which are $\emptyset$) and γ has a superset of β's low-priority features so γ attacks by order $\supseteq_L$.



(a) γ is equivalent to α by order $\supseteq_H$ and smaller by order $\supseteq_L$, blocking the attack from α.

(b) γ is smaller than α and equivalent to β by order $\supseteq_H$ but greater than β on order $\supseteq_L$, thus blocking the attack from α.

Figure 3: Two AFs showcasing potential attacks that are blocked by more concise cases according to the two concision conditions

Furthermore, the constraint $x_\alpha \not\succcurlyeq_{[1:n]} x_\gamma$ in (**ii**, Definition 6) ensures that the attacker α can never be blocked by a case that is greater by any preorder. If this did not hold, a nearest case may be irrelevant to a new case as in Example 3. This is discussed further in Section 5.

However, Definition 6 does not cover the circumstance where two cases may be equivalent to each other on all orders in $\mathcal{P}$, leading to an *incoherent* casebase.

Definition 7 (Coherent and Incoherent Casebases)**.** A casebase D is *coherent* iff there are no two cases $(x_\alpha, y_\alpha), (x_\beta, y_\beta) \in D$ such that $x_\alpha =_{[1:n]} x_\beta$ and $y_\alpha \neq y_\beta$, and it is *incoherent* otherwise.

Incoherence is handled with symmetric attacks:

Definition 8 (Incoherent Attacks)**.** For $(x_\alpha, y_\alpha), (x_\beta, y_\beta) \in D \cup \{(x_\delta, \delta)\}$, we define an incoherent attack: $(x_\alpha, y_\alpha) \overset{\text{I}}{\rightsquigarrow} (x_\beta, y_\beta)$ iff $x_\alpha =_{[1:n]} x_\beta$ and $y_\alpha \neq y_\beta$.

We have now defined all possible attacks between arguments in the casebase. We add the new case to the AF analogous to Definition 2, utilising a notion of irrelevance.

Definition 9 (New Case Attacks)**.** For $(x_\alpha, y_\alpha) \in D \cup \{(x_\delta, \delta)\}$ and new case $N = (x_N, y_?)$, we define a new case attack: $N \overset{\text{N}}{\rightsquigarrow} (x_\alpha, y_\alpha)$ iff $N \not\sim (x_\alpha, y_\alpha)$.

With these attacks defined, we can now define the argumentation framework for AA-CBR-$\mathcal{P}$.

Definition 10 (Argumentation Framework drawn from $\mathcal{P}$)**.** The argumentation framework $AF_\mathcal{P}(D, x_N)$ mined from dataset D and a new case N, with a sequence of preorders $\mathcal{P}$ of length n is $(Args, \rightsquigarrow)$ where

$Args = D \cup \{(x_\delta, \delta)\} \cup \{N\}$ and

$\rightsquigarrow = (\bigcup_{i=1}^{n} \rightsquigarrow_i) \cup \overset{\text{I}}{\rightsquigarrow} \cup \overset{\text{N}}{\rightsquigarrow}$.

As with Definition 1, the outcome predicted for N, written as AA-CBR-$\mathcal{P}(D, x_N)$, is δ if (x_δ, δ) is in the grounded extension $\mathbb{G}$ of the AF and $\bar{\delta}$ otherwise.

We also define a *regular* variant of AA-CBR-$\mathcal{P}$ that enforces that irrelevance is written in terms of the partial orders in $\mathcal{P}$ and that x_δ is the least element by all orders in $\mathcal{P}$. The regular variant naturally aligns with expectations that irrelevance is directly related to the partial order defining attacks. Moreover, regular AA-CBR-$\mathcal{P}$ has desired properties that we discuss in Section 5.

Definition 11 (Regular)**.** The AF mined from D and N, with default argument (x_δ, δ) is *regular* when:

1. The irrelevance relation $\not\sim$ is defined as $x_\alpha \not\sim x_\beta$ iff $\exists i, x_\alpha \not\succcurlyeq_i x_\beta$.
2. x_δ is the least element on all orders in $\mathcal{P}$.

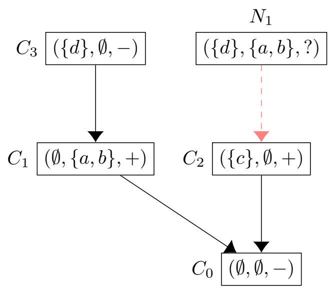
Definition 11 is a generalisation of the approach in (Paulino-Passos and Toni 2021a). We will use the notation AA-CBR-$\mathcal{P}_{\langle\rangle}$ to refer to the regular variant of AA-CBR-$\mathcal{P}$ with an arbitrary instantiation of $\mathcal{P}$. We will populate the subscript with a sequence of preorders to refer to an instantiation of the model. Similarly, we will use $AF_{\mathcal{P}\langle\rangle}(D, x_N)$ to refer to the regular variant of $AF_\mathcal{P}(D, x_N)$. To illustrate, consider the following:

Example 4. Based on Example 2, we can utilise AA-CBR-$\mathcal{P}_{\langle\rangle}$ to create a model that obeys preferences in all circumstances. We let $\mathcal{P} = \langle \supseteq_H, \supseteq_L \rangle$ and refer to this model as AA-CBR-$\mathcal{P}_{\langle \supseteq_H, \supseteq_L \rangle}$. $AF_{\mathcal{P}\langle \supseteq_H, \supseteq_L \rangle}(D, x_{N_1})$ is shown in Figure 4a and $AF_{\mathcal{P}\langle \supseteq_H, \supseteq_L \rangle}(D, x_{N_2})$ in Figure 4b. In both instances, the preferences are respected, and we see that the classification for N_1 and N_2 are determined by the most important features, d and c, respectively. C_1 is no longer considered a more concise case than C_2, thus, they can now both attack C_0. The grounded extension for $AF_{\mathcal{P}\langle \supseteq_H, \supseteq_L \rangle}(D, x_{N_1})$ is $\{N_1, C_3, C_0\}$ with the predicted outcome, $-$, as desired. For $AF_{\mathcal{P}\langle \supseteq_H, \supseteq_L \rangle}(D, x_{N_2})$ the grounded extension is $\{N_2, C_2\}$ with the predicted outcome, $+$, again as desired.

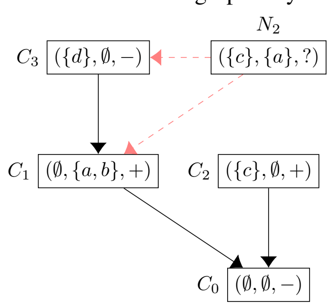
Example 4 illustrates how, in general, AA-CBR by Definition 2 cannot capture AA-CBR-$\mathcal{P}$. C_2 and C_3 are cases that should be treated with similar importance within the casebase; both have a single high-priority feature but differing outcomes. With a single partial order, as in AA-CBR, we can only support one of two possible scenarios:

1. either C_2 and C_3 are considered incomparable to C_1, allowing both C_2 and C_1 to attack C_0 but C_3 cannot attack C_1 as in Figure 1a,
2. or C_2 and C_3 are considered larger than C_1 by the partial order, allowing C_3 to attack C_1 but then C_1 is a more concise case than C_2, meaning only C_1 can attack C_0, as in Figure 1b and 1c.

AA-CBR-$\mathcal{P}$ allows for C_2 and C_3 to be treated with similar importance within the casebase, but have C_3 still attack



(a) With AA-CBR-$\mathcal{P}_{\langle \supseteq_H, \supseteq_L \rangle}$, the outcome predicted for N_1 is as expected, following the fact that d is a high-priority feature.



(b) With AA-CBR-$\mathcal{P}_{\langle \supseteq_H, \supseteq_L \rangle}$, the outcome predicted for N_2 is as expected, following the fact that c is a high-priority feature.

Figure 4: A comparison of scenarios utilising AA-CBR-$\mathcal{P}$ to introduce preferences as defined in Example 4.

C_1 and C_2 attack C_0. We can thus conclude the following theorem:

Theorem 1. There exists a sequence of partial orders $\mathcal{P}$ and casebase D, such that there is no partial order, $\succcurlyeq$, in which $AF(D, x_N)$ constructed with $\succcurlyeq$ is the same as $AF_{\mathcal{P}}(D, x_N)$ constructed with $\mathcal{P}$ and AA-CBR(D, x_N) is the same as AA-CBR-$\mathcal{P}(D, x_N)$.

Theorem 1 shows that AA-CBR-$\mathcal{P}$ is capable of constructing AFs that AA-CBR (as defined by Definition 2) is not able to and, as a result, predictions made by AA-CBR-$\mathcal{P}$ differ as well. Thus, AA-CBR-$\mathcal{P}$ is a novel contribution that improves on AA-CBR with the use of preferences.

5 Nearest and Preferred Cases

With a coherent casebase, D, and a regular AF, we can identify cases which, when in agreement on an outcome, precisely determine the outcome predicted by AA-CBR-$\mathcal{P}$. In this section, we will present this formally[5]. Note that when D is coherent, the argumentation framework is guaranteed to be *acyclic*.

Proposition 1. The argumentation framework corresponding to AA-CBR-$\mathcal{P}_{\langle\rangle}$ is guaranteed to be acyclic for a coherent D.

Previous approaches to AA-CBR have shown that for a coherent casebase if all cases that are *nearest* to the new case N have the same outcome, then that outcome will be predicted for N (Čyras, Satoh, and Toni 2016; Paulino-Passos and Toni 2021b). Here, we generalise this property for AA-CBR-$\mathcal{P}_{\langle\rangle}$.

Definition 12 (Nearest Case)**.** A case $(x_\alpha, y_\alpha) \in D \cup \{(x_\delta, \delta)\}$ is *nearest* to N iff $x_N \succcurlyeq_{[1:n]} x_\alpha$ and (x_α, y_α) is maximally so, that is, there is no $(x_\beta, y_\beta) \in D$ such that $x_N \succcurlyeq_{[1:n]} x_\beta$ and $x_\beta \succcurlyeq_{[1:n]} x_\alpha$ and $\exists i, x_\beta \succ_i x_\alpha$.

Intuitively, this states a case (x_α, y_α) that is smaller to N on all orders in $\mathcal{P}$ is nearest when there does not exist another case, (x_β, y_β), that is similarly smaller to N on all orders, no smaller than (x_α, y_α) on any orders and greater than (x_α, y_α) on at least one order. If (x_β, y_β) exists, it is a better candidate for being the nearest case than (x_α, y_α). We can define the relationship between nearest and new cases as follows:

Theorem 2. If D is coherent and every nearest case to N is of the form (x_α, y) for some outcome $y \in Y$, then AA-CBR-$\mathcal{P}_{\langle\rangle}(D, x_N) = y$.

The constraint $x_\alpha \succcurlyeq_{[1:n]} x_\gamma$ in **(ii)** in Definition 6 ensures that Theorem 2 holds. It enforces that a case is only more concise than another if it is smaller by all orders in $\mathcal{P}$. Without this constraint, an irrelevant case may prevent a nearest case from contributing to the classification by the concision. This explains the violation of the nearest cases theorem with AA-CBR with Stages in Example 3. AA-CBR-$\mathcal{P}_{\langle\rangle}$ resolves this.

Example 5. To illustrate, we instantiate AA-CBR-$\mathcal{P}_{\langle\rangle}$with $\mathcal{P} = \langle \supseteq, \sqsupseteq \rangle$ and case characterisations corresponding to that of Definition 3. For $x_\alpha = (f_\alpha, s_\alpha)$ and $x_\beta = (f_\beta, s_\beta)$:

- $x_\alpha \supseteq x_\beta$ iff $f_\alpha \supseteq f_\beta$,
- $x_\alpha \sqsupseteq x_\beta$ iff $s_\alpha \sqsupseteq s_\beta$.

We will refer to this model as AA-CBR-$\mathcal{P}_{\langle \supseteq, \sqsupseteq \rangle}$.

The argumentation framework, $AF_{\mathcal{P}\langle \supseteq, \sqsupseteq \rangle}(D, x_N)$ corresponding to the cases from Example 3 is shown in Figure 5. C'_3 is not smaller on all orders than C'_1 and therefore is no longer considered a more concise case. The grounded extension is $\{N', C'_1, C'_0\}$ and the predicted outcome is $+$ as

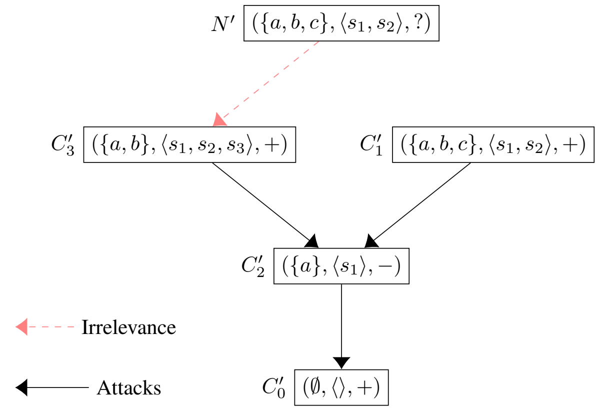


Figure 5: $AF_{\mathcal{P}\langle \supseteq, \sqsupseteq \rangle}(D, x_N)$ as defined in Example 5. The nearest cases property holds. C'_1 is not prevented from attacking as C'_3 is no longer considered a more concise case.

[5]Proofs of results are available in the Appendix at https://tinyurl.com/AA-CBR-P-with-appendix.

expected. As a result, C_1' is able to affect the determination of N', which results in the outcome predicted precisely agreeing. The nearest cases property clearly holds.

A modified variant of AA-CBR with Stages as described in Definition 3 can be derived from AA-CBR-$\mathcal{P}_{\langle \supseteq, \sqsupseteq \rangle}$ with condition **2. (b) i** altered to $F_\alpha \supset F_\gamma \supset F_\beta$ and $S_\alpha \sqsupseteq S_\gamma$. Moreover, by substituting $\mathcal{P} = \langle \succcurlyeq \rangle$, we can precisely capture AA-CBR as described in Definition 2 by AA-CBR-$\mathcal{P}$. As a result, our method effectively generalises both of these previous approaches whilst ensuring the nearest case property holds in all circumstances.

However, we can make a stronger claim that preferences themselves are obeyed and used in determining the outcome of a new case. In Figure 4a, both C_3 and C_1 are nearest cases but do not agree on an outcome. Yet, the outcome predicted is the same as that for C_3. This is not incidental; C_3 is a *preferred* case.

Definition 13 (Preferred Case)**.** A nearest case (x_α, y_α) is *preferred* for N if there is no other nearest case (x_β, y_β) such that $\exists j$, $x_\alpha =_{[1:j-1]} x_\beta$ and $x_\beta \succ_j x_\alpha$.

Preferred cases are nearest cases that are maximally close to N by the most preferred order possible. As a result, we can prove a similar but stronger condition than that provided by nearest cases: if preferred cases agree on an outcome, then that is the outcome predicted for a new case.

Theorem 3. If D is coherent and every preferred case to N is of the form (x_α, y) for some outcome $y \in Y$ then AA-CBR-$\mathcal{P}_{\langle\rangle}(D, x_N) = y$.

Preferred cases are a subset of nearest cases in which, by Theorem 3, the preferred cases alone are sufficient to decide the outcome of a new case when they agree on an outcome.

Example 6. Consider the argumentation framework, $AF_{\mathcal{P}\langle \supseteq_H, \supseteq_L \rangle}(D, x_{N_3})$ in Figure 6. C_1 and C_5 are each nearest cases to N_3 as they are maximally close to N_3 in the casebase as by Definition 12. However, these two nearest cases do not agree on an outcome. Only C_5 is a preferred case; it is maximally close to N_3 on the high-priority order $\supseteq_H$. By Theorem 3, the outcome of N_3 is the same as the outcome of C_5.

This occurs because C_1 is attacked by C_4, which is not considered nearest or preferred but is more concise than C_5. If C_4 were missing from this casebase, C_5 would fill the role and attack C_1. Therefore, the existence of C_5 as a preferred case precisely determines the outcome predicted for N_3.

Furthermore, by Definitions 5 and 6, we have that if there exists an attack from α to β, this attack must occur by the most preferred order of $\mathcal{P}$ possible.

Proposition 2. For cases $\alpha, \beta \in D$, if $\exists i, \alpha \rightsquigarrow_i \beta$ then $\nexists k < i, \alpha \rightsquigarrow_k \beta$.

This is easy to see since $\alpha \rightsquigarrow_i \beta$ implies $\alpha \overset{p}{\rightsquigarrow}_i \beta$, which in turn implies $\alpha =_{[1:i-1]} \beta$.

Consequently, the argumentation framework is constructed by obeying the preference orderings. Moreover, with Theorem 3, we know that preferred cases can determine the outcome of a new case. We can thus conclude that the outcome predicted for a new case follows the user-defined preferences by construction.

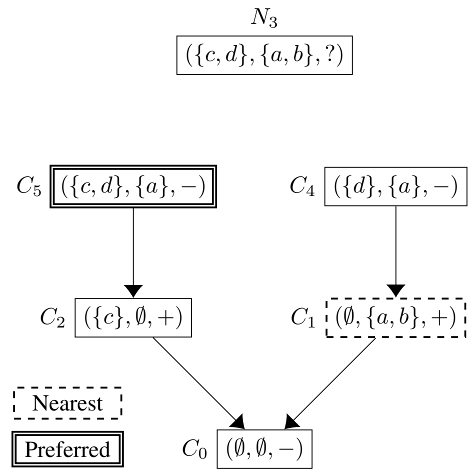


Figure 6: The argumentation framework as defined in Example 6. We see C_5 is the preferred case and despite disagreeing with nearest case C_1, the outcome predicted for N_3 is the same as C_5, as expected by Theorem 3.

6 Empirical Evaluation

In this section, we showcase how to utilise AA-CBR-$\mathcal{P}$ models in a medical classification task with preferences defined over features derived from varying data sources. BrainWear (Dadhania et al. 2021; Dadhania et al. 2023) is a study exploring the utility of *physical activity* (PA) data collected via wrist-worn accelerometers in patients with a primary brain tumour. PA data complements traditional questionnaires, that allow patients to report their health status, known as *patient reported outcomes* (PRO). We can predict disease progression in new patients using AA-CBR-$\mathcal{P}$, with preferences defined over features derived from PA and PRO data[6].

6.1 Methodology

Each completed questionnaire is matched with an 8-week average of PA data centred on the questionnaire date. The objective is to predict if the patient has *progressive disease* or not, as labelled by the outcome of the next MRI scan following the questionnaire or the patient's mortality status. Furthermore, we can incorporate the number of previous instances of progressive disease for each patient, serving as a proxy for cancer progression. We conduct two experiments to evaluate the models. The first experiment solely utilises PA and PRO data, while the second incorporates both PA and PRO data along with the proxy measurement as stages. Each experiment consists of multiple AA-CBR(-$\mathcal{P}$) models compared against similarly interpretable baseline models: a decision tree and a K-Nearest Neighbor (kNN).

In the first experiment, we compare AA-CBR with the $\supseteq$ relation as defined in Example 1 (we denote this model as AA-CBR$_\supseteq$) and AA-CBR with $\succcurlyeq_{lex}$ (denoted AA-CBR$_{\succcurlyeq_{lex}}$) as defined in Example 2 where PA and PRO features make up the full suite of features against,

[6]Details on data can be found in the appendix.

AA-CBR-$\mathcal{P}_{\langle \sqsupseteq_{pa}, \sqsupseteq_{pro} \rangle}$ defined analogously to Example 4, where we treat the PA feature set as the high-priority set and the PRO feature set as the low-priority set. We have chosen these models to showcase how a naive instantiation of preferences in AA-CBR compares against our more sophisticated approach using AA-CBR-$\mathcal{P}$.

For the dynamic models, we compare AA-CBR with Stages as in Definition 3 using the proxy measurement as the stages. We compare this against AA-CBR-$\mathcal{P}_{\langle \sqsupseteq, \sqsupseteq \rangle}$ as in Example 5. Finally, we look at AA-CBR-$\mathcal{P}_{\langle \sqsupseteq_{pa}, \sqsupseteq_{pro}, \sqsupseteq \rangle}$ where we combine both preferences over features with stages.

For all AA-CBR-based models, we use the empty set (and, where applicable, the empty sequence of stages) to characterise the default case. The default outcome is set to progressive disease.

6.2 Evaluation

We evaluate the performance of the models using accuracy, precision, recall and F1-score, with progressive disease as the 'positive' class (Hossin and Sulaiman 2015).

Table 1 shows results for the first experiment. We observe that AA-CBR-$\mathcal{P}_{\langle \sqsupseteq_{pa}, \sqsupseteq_{pro} \rangle}$ outperforms the other models in terms of accuracy, recall, and F1 score. The decision tree and AA-CBR$_{\sqsupseteq}$ perform similarly, but both are outperformed by the preference-based AA-CBR-$\mathcal{P}_{\langle \sqsupseteq_{pa}, \sqsupseteq_{pro} \rangle}$ model. The introduction of preferences, in this instance, leads to improved classification performance but this cannot be done naively as with AA-CBR$_{\succcurlyeq_{lex}}$, which has the worst performance by almost all metrics. Clearly, our novel implementation that introduces preferences with AA-CBR-$\mathcal{P}$ leads to an empirically better classifier for this task.

Table 2 shows the results for the second experiment. The decision-tree and k-nearest neighbor models are outperformed by the AA-CBR-$\mathcal{P}$ models. Notably, we see that AA-CBR-$\mathcal{P}_{\langle \sqsupseteq, \sqsupseteq \rangle}$ has an increase in performance compared to AA-CBR with Stages, highlighting how forcing the model to conform to Theorem 3 leads to more desirable empirical results. Furthermore, adding preferences over features with AA-CBR-$\mathcal{P}_{\langle \sqsupseteq_{pa}, \sqsupseteq_{pro}, \sqsupseteq \rangle}$ results in another increase in performance, again showcasing how preferences improve model performance.

Overall, we find that AA-CBR-$\mathcal{P}$ is not only successful at this classification task but can effectively leverage preferences to increase model performance. Moreover, our approach is flexible and generalised, so it can support varying representations, such as through the introduction of stages with preferences which improves performance further compared to AA-CBR. This comes at a cost of increasing time complexity by a factor of m, where m is the number of partial orders[7].

7 Related Work and Discussion

Preferences over arguments have been shown to resolve situations in which the direction of an attack is unclear. Approaches often introduce an ordering specifying preferences over arguments, as in (Amgoud and Cayrol 1998; Modgil 2009). Argumentation frameworks derived from

[7]Complexity results can be found in the appendix.

Model	Accuracy	Precision	Recall	F_1
Decision Tree	0.72	0.83	0.63	0.72
k-Nearest Neighbor	0.68	**0.88**	0.50	0.64
AA-CBR$_{\sqsupseteq}$	0.72	0.82	0.64	0.72
AA-CBR$_{\succcurlyeq_{lex}}$	0.60	0.67	0.57	0.62
AA-CBR-$\mathcal{P}_{\langle \sqsupseteq_{pa}, \sqsupseteq_{pro} \rangle}$	**0.78**	0.81	**0.79**	**0.80**

Table 1: Results of the first experiment where we utilise solely PA and PRO features to predict progressive disease in patients.

Model	Accuracy	Precision	Recall	F_1
Decision Tree	0.75	**0.94**	0.58	0.72
k-Nearest Neighbor	0.70	0.93	0.50	0.65
AA-CBR with Stages	0.74	0.86	0.64	0.73
AA-CBR-$\mathcal{P}_{\langle \sqsupseteq, \sqsupseteq \rangle}$	0.76	0.86	0.68	0.76
AA-CBR-$\mathcal{P}_{\langle \sqsupseteq_{pa}, \sqsupseteq_{pro}, \sqsupseteq \rangle}$	**0.80**	0.88	**0.75**	**0.81**

Table 2: Results of the second experiment where we utilise PA and PRO features in combination with a proxy measurement of previous disease progression to predict progressive disease in patients.

AA-CBR with a single order can be viewed similarly, where the order defines preferences over the casebase. However, as shown in Example 1, this alone is insufficient when introducing local preferences over constituents of an argument for classification. Existing approaches for accommodating preferences over constituents of arguments have also been proposed, e.g. as in (Modgil and Prakken 2013; Cyras and Toni 2016; Kakas, Moraitis, and Spanoudakis 2018; García and Simari 2014) for various structured argumentation frameworks. It is an open question as to whether these methods generalise or approximate our approach. As future work, it would be interesting to study whether any of these approaches correspond to our method.

Our empirical evaluation shows the usefulness of accommodating domain expertise in the form of preferences within the medical setting. It is widely acknowledged that healthcare providers have domain expertise that can influence the outcome of explainable AI systems (Pawar et al. 2020). Moreover, personalised healthcare and increased patient engagement can affect healthcare outcomes (Politi et al. 2013). Argumentation with preferences has already been shown to be effective for personalised healthcare. Specifically, in (Čyras et al. 2021a), the assumption-based argumentation model ABA$^+$G is used to resolve patient preferences with clinical guidelines. In that approach, preferences are defined over both assumptions, which dictate the direction of attacks and goals, in turn defining which extension(s) to accept. Also, the Gorgias system (Kakas, Moraitis, and Spanoudakis 2018) applies argumentation with preferences to multiple healthcare scenarios, by allowing experts to express preferences over outcomes of scenarios which lead to preference hierarchies then translated into arguments about outcomes. Like with AA-CBR-$\mathcal{P}$, both these approaches use preferences defined over constituents of arguments. However, these approaches derive rules from a knowledge base, e.g. clinical guidelines, rather than by case-based reasoning.

AA-CBR-$\mathcal{P}$ is also related to case-based reasoning with precedential constraints, which has been extensively researched within the legal domain (Horty 2011; Horty and Bench-Capon 2012). Dimension-based precedential constraints (Horty 2019; Rigoni 2017) represent an approach in which a set of values sampled from varying dimensions defines a case. Each dimension is equipped with a partial order in which values at either end of the partial order agree with opposing outcomes. In a classification task (Odekerken, Bex, and Prakken 2023), past cases constrain dimensions of a new case to an outcome consistent with the casebase. However, an outcome may not be predicted if the casebase cannot constrain a new case, and its outcome will remain undecided. Previous work has shown how to enforce factor-based precedential constraints on AA-CBR$_{\supseteq}$ (Paulino-Passos and Toni 2021a), deciding non-constrained cases, but it does not incorporate a notion of preferences between the factors, which is captured by the preferences in AA-CBR-$\mathcal{P}$. Indeed, each preorder in $\mathcal{P}$ can be considered a generalised form of dimension in the precedential constraint sense. By introducing a lexicographic ordering over these 'dimensions', we enforce that an outcome must be predicted, and preferred cases constrain new cases. Moreover, our approach does not enforce that either end of a partial order must agree with an outcome. This is suitable in a healthcare setting where such a restriction should not always apply; for example, sustained blood pressure that is either too high or too low could both have negative health consequences.

8 Conclusion

In this paper, we introduced AA-CBR-$\mathcal{P}$ as a novel method of including preferences with an AA-CBR-based approach. We have shown how preferences allow domain-specific knowledge or individual choices to guide the model and that existing approaches to AA-CBR cannot support this. Moreover, we prove that the desirable property of predictions abiding by the nearest cases, hold for AA-CBR-$\mathcal{P}_{\langle\rangle}$ and present a stronger condition for preferred cases. Given this and that individual attacks always occur by the most preferred order, we conclude that the model inherently abides by the preferences by construction. We then show that the addition of preferences and the enforcement of the nearest and preferred cases conditions can empirically lead to increased performance in a medical classification task, outperforming that of other interpretable baseline models.

We have not explored methods of automatically deriving a preference ordering that directly optimises the model's performance. Methods such as the DEAr Methodology (Cocarascu et al. 2020) could be applied to utilise AA-CBR-$\mathcal{P}_{\langle\rangle}$ to find characterisations and preference orderings that lead to the most optimal performance possible whilst still allowing for flexibility in user-defined preferences. Furthermore, we leave for future work the exploration of AA-CBR-$\mathcal{P}_{\langle\rangle}$ variants, such as a cumulative model akin to cAA-CBR of (Paulino-Passos and Toni 2021a).

Acknowledgements

Adam Gould is supported by UK Research and Innovation [UKRI Centre for Doctoral Training in AI for Healthcare grant number EP/S023283/1]. Francesca Toni and Guilherme Paulino-Passos were partially funded by the ERC under the EU's Horizon 2020 research and innovation programme (grant agreement No. 101020934). Toni was also partially funded by J.P. Morgan and the Royal Academy of Engineering, UK, under the Research Chairs and Senior Research Fellowships scheme. Seema Dadhania is supported by the Imperial College CRUK Centre Clinical Research Fellowship A27434. Matthew Williams is supported by the NIHR Imperial Biomedical Research Centre (BRC) and by ICHT NHS Trust.

We would like to thank the Computational Logic and Argumentation (CLArg) group and the Computational Oncology group, both at Imperial College London, for their feedback and support.

References

Amann, J.; Blasimme, A.; Vayena, E.; Frey, D.; Madai, V. I.; and Consortium, P. 2020. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC medical informatics and decision making* 20:1–9.

Amgoud, L., and Cayrol, C. 1998. On the acceptability of arguments in preference-based argumentation. In *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*, UAI'98, 1–7. Morgan Kaufmann Publishers Inc.

Amgoud, L., and Prade, H. 2009. Using arguments for making and explaining decisions. *Artificial Intelligence* 173(3):413–436.

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

Briguez, C. E.; Budán, M. C.; Deagustini, C. A.; Maguitman, A. G.; Capobianco, M.; and Simari, G. R. 2014. Argument-based mixed recommenders and their application to movie suggestion. *Expert Systems with Applications* 41(14):6467–6482.

Cocarascu, O.; Stylianou, A.; Čyras, K.; and Toni, F. 2020. Data-empowered argumentation for dialectically explainable predictions. In *ECAI 2020*. IOS Press. 2449–2456.

Cocarascu, O.; Čyras, K.; and Toni, F. 2018. Explanatory predictions with artificial neural networks and argumentation. In *Proceedings of the 2ndWorkshop onExplainable Artificial Intelligence (XAI 2018).*

Cyras, K., and Toni, F. 2016. ABA+: assumption-based argumentation with preferences. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Fifteenth International Conference, KR 2016, Cape Town, South Africa, April 25-29, 2016*, 553–556. AAAI Press.

Dadhania, S.; Pakzad-Shahabi, L.; Calvez, K. L.; Saleem, W.; Wang, J.; Mohammed, W.; Mistry, S.; and Williams, M. 2021. BrainWear: Longitudinal, objective assessment of physical activity in 42 high grade glioma (HGG) patients. *Neuro-Oncology* 23(Supplement_4):iv3–iv3.

Dadhania, S.; Pakzad-Shahabi, L.; Mistry, S.; and Williams, M. 2023. Triaxial accelerometer-measured physical activity and functional behaviours among people with high grade glioma: The BrainWear study. *PLOS ONE* 18(5):e0285399.

Dung, P. M. 1995. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence* 77(2):321–357.

García, A. J., and Simari, G. R. 2014. Defeasible logic programming: Delp-servers, contextual queries, and explanations for answers. *Argument & Computation* 5(1):63–88.

Goodman, B., and Flaxman, S. 2017. European union regulations on algorithmic decision making and a "right to explanation". *AI Magazine* 38(3):50–57.

Horty, J. F., and Bench-Capon, T. J. M. 2012. A factor-based definition of precedential constraint. *Artificial Intelligence and Law* 20(2):181–214.

Horty, J. F. 2011. Rules and reasons in the theory of precedent. *Legal Theory* 17(1):1–33.

Horty, J. 2019. Reasoning with dimensions and magnitudes. *Artificial Intelligence and Law* 27(3):309–345.

Hossin, M., and Sulaiman, M. N. 2015. A review on evaluation metrics for data classification evaluations. *International journal of data mining & knowledge management process* 5(2):1.

Hunter, A., and Williams, M. 2012. Aggregating evidence about the positive and negative effects of treatments. *Artificial Intelligence in Medicine* 56(3):173–190.

Kakas, A. C.; Moraitis, P.; and Spanoudakis, N. I. 2018. Gorgias: Applying argumentation. *Argument & Computation* 10(1):55–81.

Modgil, S., and Prakken, H. 2013. A general account of argumentation with preferences. *Artificial Intelligence* 195:361–397.

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

Odekerken, D.; Bex, F.; and Prakken, H. 2023. *Precedent-Based Reasoning with Incomplete Cases*. IOS Press. 33–42.

Paulino-Passos, G., and Toni, F. 2021a. Monotonicity and noise-tolerance in case-based reasoning with abstract argumentation. In *Proceedings of the Eighteenth International Conference on Principles of Knowledge Representation and Reasoning*, 508–518.

Paulino-Passos, G., and Toni, F. 2021b. Monotonicity and Noise-Tolerance in Case-Based Reasoning with Abstract Argumentation (with Appendix). arXiv:2107.06413 [cs].

Paulino-Passos, G., and Toni, F. 2023. *Learning Case Relevance in Case-Based Reasoning with Abstract Argumentation*. IOS Press. 95–100.

Pawar, U.; O'Shea, D.; Rea, S.; and O'Reilly, R. 2020. Explainable ai in healthcare. In *2020 International Conference on Cyber Situational Awareness, Data Analytics and Assessment (CyberSA)*. IEEE.

Politi, M. C.; Dizon, D. S.; Frosch, D. L.; Kuzemchak, M. D.; and Stiggelbout, A. M. 2013. Importance of clarifying patients' desired role in shared decision making to match their level of engagement with their preferences. *BMJ* 347(dec02 1):f7066–f7066.

Prakken, H., and Ratsma, R. 2022. A top-level model of case-based argumentation for explanation: Formalisation and experiments. *Argument & Computation* 13:159–194.

Rago, A.; Cocarascu, O.; Bechlivanidis, C.; and Toni, F. 2020. Argumentation as a Framework for Interactive Explanations for Recommendations. In *Proceedings of the Seventeenth International Conference on Principles of Knowledge Representation and Reasoning*, 805–815.

Rigoni, A. 2017. Representing dimensions within the reason model of precedent. *Artificial Intelligence and Law* 26(1):1–22.

Williams, M.; Liu, Z. W.; Hunter, A.; and Macbeth, F. 2015. An updated systematic review of lung chemo-radiotherapy using a new evidence aggregation method. *Lung Cancer* 87(3):290–295.

Zheng, H.; Grossi, D.; and Verheij, B. 2020. Case-based reasoning with precedent models: Preliminary report. In *Frontiers in Artificial Intelligence and Applications*. IOS Press. 443–450.

Čyras, K.; Birch, D.; Guo, Y.; Toni, F.; Dulay, R.; Turvey, S.; Greenberg, D.; and Hapuarachchi, T. 2019. Explanations by arbitrated argumentative dispute. *Expert Systems with Applications* 127:141–156.

Čyras, K.; Oliveira, T.; Karamlou, A.; and Toni, F. 2021a. Assumption-based argumentation with preferences and goals for patient-centric reasoning with interacting clinical guidelines. *Argument & Computation* 12(2):149–189.

Čyras, K.; Rago, A.; Albini, E.; Baroni, P.; and Toni, F. 2021b. Argumentative xai: A survey. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence*, IJCAI-2021. International Joint Conferences on Artificial Intelligence Organization.

Čyras, K.; Satoh, K.; and Toni, F. 2016. Abstract argumentation for case-based reasoning. In *Fifteenth International Conference on the Principles of Knowledge Representation and Reasoning*. AAAI Press.