ABA$^+$: Assumption-Based Argumentation with Preferences

Kristijonas Čyras
Imperial College London

Abstract

My work focuses on using argumentation theory to model common-sense reasoning with preferences. To this end, I have equipped a well-established structured argumentation formalism, Assumption-Based Argumentation, with a preference handling mechanism. I aim to advance the newly proposed formalism, called ABA$^+$, present its motivations and place among other argumentation formalisms, and discuss various properties of ABA$^+$.

1 Argumentation and Preferences

Dealing with preferences is an important topic in AI at large, and argumentation in particular (Kaci 2011). A principal issue regarding argumentation and preferences is the lack of consensus on how preferences should be accounted for. This is witnessed by a large number of argumentation formalisms handling preferences in different ways. Most of these can be broadly classified with respect to several main methods.

One method is to compile preferences into *the object level*, by encoding them within the existing components of a formalism: e.g. (Kowalski and Toni 1996; Thang and Luong 2014). Such an approach may, however, produce numerous additional objects from a compact preference relation (Wakaki 2014), and could be hard to generalize.

Another method, utilized by a majority of argumentation formalisms dealing with preferences, e.g. (Prakken and Sartor 1999; Amgoud and Cayrol 2002; Bench-Capon 2003; Kaci and van der Torre 2008; Brewka and Woltran 2010; Besnard and Hunter 2014; Modgil and Prakken 2014; García and Simari 2014), is to use preferences on *the argument level* to discard attacks from less preferred arguments: if an argument A attacks an argument B (written $\mathsf{A} \rightsquigarrow \mathsf{B}$) but B is preferred over A (written $\mathsf{A} < \mathsf{B}$), then $\mathsf{A} \rightsquigarrow \mathsf{B}$ fails. This may be problematic, as, for instance, $\{\mathsf{A}, \mathsf{B}\}$ can then be a subset of an acceptable extension, which would then not be conflict-free with respect to the original attack relation (see e.g. (Kaci 2010; Amgoud and Vesic 2014) for discussions). To avoid this issue, certain conditions can be imposed on argument frameworks: e.g. considering only symmetric attacks (Kaci 2010), contraposition on rules (Modgil and Prakken 2014). Such restrictions may, however, lead to other problems: e.g. rationality postulates unsatisfied, counter-intuitive limitations on expressiveness (see e.g. (Amgoud and Vesic 2014; Baroni, Giacomin, and Liao 2015)).

Yet another approach is to employ preferences on *the extension level* to select the most 'preferable' extensions, e.g. (Amgoud and Vesic 2011; Wakaki 2014). However, this may not always be adequate either. For example, if A, B are the only arguments and $\mathsf{A} \rightsquigarrow \mathsf{B}$ is the only attack, then $\{\mathsf{A}\}$ is the only (say, stable) extension to begin with, whence whatever the preferences over arguments, there is no choice to be made: even if $\mathsf{A} < \mathsf{B}$, $\{\mathsf{B}\}$ cannot be selected as the 'preferable' extension. Likewise, in the absence of extensions due to, for instance, odd cycles, preferences do not play a role.

Both methods of discarding attacks and selecting among extensions due to preference information, often involve preference aggregation mechanisms: object-level (or argument-level) preferences are lifted to argument- or extension-levels by means of element set-wise comparison, e.g. (Kaci and van der Torre 2008; Modgil and Prakken 2014; Amgoud and Vesic 2011; 2014; Wakaki 2014). This entails outcome dependency on the choice of comparison principle.

Finally, a couple of very recently proposed formalisms, namely *(Rich) Preference-based Argumentation Frameworks* (PAFs) (Amgoud and Vesic 2014) and *Assumption-Based Argumentation with Preferences* (ABA$^+$) (Čyras and Toni 2016a), represent the method of attack reversal. In PAFs, attacks are reversed in Abstract Argumentation (AA) (Dung 1995) setting: if $\mathsf{A} \rightsquigarrow \mathsf{B}$ and $\mathsf{A} < \mathsf{B}$, then $\mathsf{A} \rightsquigarrow \mathsf{B}$ fails and instead one obtains $\mathsf{B} \rightsquigarrow \mathsf{A}$. ABA$^+$ takes this idea further, to the realm of structured argumentation, by equipping Assumption-Based Argumentation (ABA) (Bondarenko et al. 1997; Toni 2014) with (object-level) preferences over assumptions and incorporating them directly into the attack relation so as to reverse attacks.

2 ABA$^+$

Background I sketch the necessary details on ABA$^+$, based on (Toni 2014; Čyras and Toni 2016a).

An *ABA$^+$ framework* is a tuple $(\mathcal{L}, \mathcal{R}, \mathcal{A}, ^-, \leqslant)$, where: $(\mathcal{L}, \mathcal{R})$ is a deductive system; $\mathcal{A} \subseteq \mathcal{L}$ is a non-empty set of *assumptions*; $^- : \mathcal{A} \to \mathcal{L}$ is a total *contrary* mapping; $\leqslant$ is a transitive binary relation on $\mathcal{A}$. A *deduction for $\varphi \in \mathcal{L}$ supported by $S \subseteq \mathcal{L}$ (and $R \subseteq \mathcal{R}$)*, denoted by $S \vdash^R \varphi$, is a finite tree with the root labelled by φ, leaves labelled by $\top$ or elements from S, the children of non-leaf nodes ψ labelled by the elements of the body of some rule from $\mathcal{R}$ with head ψ, and R being the set of all such rules.

For $A, B \subseteq \mathcal{A}$, A $<$-attacks B, written $A \rightsquigarrow_< B$, if:

- either there is a deduction $A' \vdash^R \overline{\beta}$, for some $\beta \in B$, supported by $A' \subseteq A$, such that $\forall \alpha' \in A'\ \alpha' \not< \beta$;
- or there is a deduction $B' \vdash^R \overline{\alpha}$, for some $\alpha \in A$, supported by $B' \subseteq B$, such that $\exists \beta' \in B'$ with $\beta' < \alpha$.

The first type of attack is called *normal*, and the second one *reverse*.

ABA$^+$ semantics (as well as conflict-freeness and defence with respect to $\rightsquigarrow_<$) are defined as for ABA, by replacing the notion of attack with that of $<$-attack.

The concept of $<$-attack reflects the interplay between deductions, contraries and preferences, by representing inherent conflicts among sets of assumptions while accounting for preferences. Normal attacks follow the standard notion of attack in ABA, preventing the attack to succeed when the attacker uses assumptions less preferred than the one attacked. Reverse attacks, meanwhile, manifest the conflict between sets of assumptions when one 'tries' to attack another but fails due to preferences. Extensions, representing coherent points of view, must recognize that the conflict is still present (see e.g. (Amgoud and Vesic 2014) for a discussion). ABA$^+$ ensures precisely that, while dealing with preferences solely on the object level, thus dispensing with preference aggregation mechanisms.

Progress (Consult (Čyras and Toni 2016a; 2016b).)

ABA$^+$ is a conservative extension of ABA: any ABA$^+$ framework $(\mathcal{L}, \mathcal{R}, \mathcal{A}, \overline{}, \emptyset)$ behaves exactly like its underlying ABA framework $(\mathcal{L}, \mathcal{R}, \mathcal{A}, \overline{})$. ABA$^+$ also preserves conflicts in the sense that $E \subseteq \mathcal{A}$ is $<$-conflict-free iff E is conflict-free, thus pre-empting criticisms regarding failure to adequately capture conflicts. Moreover, ABA$^+$ satisfies certain desirable properties, such as rationality postulates (Caminada and Amgoud 2007) and various preference handling principles. In terms of semantics, Fundamental Lemma and other familiar properties hold for ABA$^+$ assuming contraposition on rules; to relax this restriction, I propose a weaker version of contraposition that suffices to guarantee the same properties. ABA$^+$ also carries non-monotonic inference properties (Čyras and Toni 2015) over from ABA.

More generally, ABA$^+$ differs significantly from most approaches to argumentation with preferences by virtue of its attack reversal. ABA$^+$ generalizes the attack reversal of PAFs too, at the same time tackling the non-trivial task of handling object-level preferences in structured argumentation, without employing preference aggregation mechanisms. I also argue that ABA$^+$ yields more intuitive outcomes in various reasoning scenarios than some other argumentation formalisms.

Future Work Among exploring various properties and conducting a detailed comparison of ABA$^+$ to other formalisms of argumentation with preferences, I aim to investigate the following: preferences over rules; dynamic preferences (e.g. (Prakken and Sartor 1999)); complexity of reasoning problems; adaptation of computational mechanisms of ABA's dispute derivations to ABA$^+$.

References

Amgoud, L., and Cayrol, C. 2002. A Reasoning Model Based on the Production of Acceptable Arguments. *Ann. Math. Artif. Intell.* 34(1-3):197–215.

Amgoud, L., and Vesic, S. 2011. A New Approach for Preference-Based Argumentation Frameworks. *Ann. Math. Artif. Intell.* 63(2):149–183.

Amgoud, L., and Vesic, S. 2014. Rich Preference-Based Argumentation Frameworks. *Int. J. Approx. Reasoning* 55(2):585–606.

Baroni, P.; Giacomin, M.; and Liao, B. 2015. I don't care, I don't know ... I know too much! On Incompleteness and Undecidedness in Abstract Argumentation. In *Brewka Festschrift*, 265–280.

Bench-Capon, T. J. M. 2003. Persuasion in Practical Argument Using Value Based Argumentation Frameworks. *J. Log. Comput.* 13(3):429–448.

Besnard, P., and Hunter, A. 2014. Constructing Argument Graphs with Deductive Arguments: A Tutorial. *Arg.&Comp.* 5(1):5–30.

Bondarenko, A.; Dung, P. M.; Kowalski, R.; and Toni, F. 1997. An Abstract, Argumentation-Theoretic Approach to Default Reasoning. *Artif. Intell.* 93(97):63–101.

Brewka, G., and Woltran, S. 2010. Abstract Dialectical Frameworks. In *KR*.

Caminada, M., and Amgoud, L. 2007. On the Evaluation of Argumentation Formalisms. *Artif. Intell.* 171(5-6):286–310.

Čyras, K., and Toni, F. 2015. Non-Monotonic Inference Properties for Assumption-Based Argumentation. In *TAFA*, 92–111.

Čyras, K., and Toni, F. 2016a. ABA+: Assumption-Based Argumentation with Preferences. In *KR*.

Čyras, K., and Toni, F. 2016b. Properties of ABA+ for Non-Monotonic Reasoning. In *NMR*.

Dung, P. M. 1995. On the Acceptability of Arguments and its Fundamental Role in Nonmonotonic Reasoning, Logic Programming and n-person Games. *Artif. Intell.* 77:321–357.

García, A. J., and Simari, G. R. 2014. Defeasible Logic Programming: DeLP-servers, Contextual Queries, and Explanations for Answers. *Arg.&Comp.* 5(1):63–88.

Kaci, S., and van der Torre, L. 2008. Preference-Based Argumentation: Arguments Supporting Multiple Values. *Int. J. Approx. Reasoning* 48(3):730–751.

Kaci, S. 2010. Refined Preference-Based Argumentation Frameworks. In *COMMA*, 299–310.

Kaci, S. 2011. *Working with Preferences. Less is More*. Springer.

Kowalski, R., and Toni, F. 1996. Abstract Argumentation. *Artif. Intell. Law* 4(3-4):275–296.

Modgil, S., and Prakken, H. 2014. The ASPIC+ Framework for Structured Argumentation: A Tutorial. *Arg.&Comp.* 5(1):31–62.

Prakken, H., and Sartor, G. 1999. A System for Defeasible Argumentation, with Defeasible Priorities. In *Artif. Intell. Today*, 365–379.

Thang, P. M., and Luong, H. T. 2014. Translating Preferred Subtheories into Structured Argumentation. *J. Log. Comput.* 24(4):831–850.

Toni, F. 2014. A Tutorial on Assumption-Based Argumentation. *Arg.&Comp.* 5(1):89–117.

Wakaki, T. 2014. Assumption-Based Argumentation Equipped with Preferences. In *PRIMA*, 116–132.