A Cooperative-game Approach to Share Acceptability and Rank Arguments*

Stefano Bistarelli[1], Paolo Giuliodori[2], Francesco Santini[1], and Carlo Taticchi[3]

[1] Department of Mathematics and Computer Science, University of Perugia, Italy
[2] School of Science and Technology, Computer Science Division, University of Camerino, Italy
[3] Gran Sasso Science Institute, L'Aquila, Italy

Abstract. We deploy a game-theoretic approach for analysing the acceptability of arguments in a generic Abstract Argumentation Framework. The result is a ranking-based semantics, which sorts arguments from the most to the least acceptable. In the computation of such a ranking, we adopt the Shapley Value formula, since it is usually used to fairly distribute costs to several entities in coalitions (labelled sets of arguments in our case). Finally, we show that some well-known properties are satisfied by the ranked-semantics we designed, and we provide an example of how our approach works.

Keywords: Argumentation, semantics, ranking, cooperative game theory.

1 Introduction

Argumentation represents a qualitative and logical method to deal with uncertain and defeasible reasoning. Defining the properties of an argumentation semantics [5] amounts to specifying the criteria for deriving subsets of arguments (called *extensions*) from an *Abstract Argumentation Framework* (*AAF*), which is defined by a set of arguments A and an attack relation R on A. On the basis of such extensions, three justification statuses can be assigned to each argument [7]: an argument is *justified*, w.r.t. a given semantics, if it belongs to all its extensions, *defensible* if it belongs to at least one (and it is not justified), or *overruled* if it does not belong to any extension. For some applications (e.g., Decision-making [1] or Strategic Games [6]), it is important to provide a ranking over the arguments. However, the three levels of justification previously introduced are not enough to obtain a detailed ranking. This is the main motivation behind *ranking-based semantics* [1], which define an order that can be interpreted as corresponding to a classification into finer acceptability levels.

The aim of our work is to design a ranking-based semantics based on the *Shapley Value* (*SV*) scheme taking advantage of *labelling semantics* [2]. The SV formula is a well-known concept in cooperative game theory, and it is usually deployed to share fairly cost or gains according to a valuation function. By exploiting the SV formula, we can build a ranking among arguments by taking into account how much an argument

* This work has been supported by: "ComPAArg" (Ricerca di base 2016–2018), "Argumentation 360" (Ricerca di Base 2017–2019) and "RACRA" (Ricerca di base 2018–2020).

participates to make an extension admissible (or complete, preferred, etc.). Designing a ranking-based semantics in this way has the advantage of automatically inheriting the properties of the SV, like *efficiency*, *symmetry*, *linearity*, and *zero players* [8].

2 Preliminaries

The most used fair division scheme used in cooperative game theory is the Shapley Value [8, 9], that takes a random ordering of the agents picked uniformly from the set of all $n!$ possible orderings, and charges each agent with her expected marginal contribution. Since for any agent $i \in G$ and any set $S_{-i} \subseteq G \setminus \{i\}$ with $|S_{-i}| = s$, the probability that the set of agents S_{-i} comes before i in a random ordering is $s!(n-1-s)!/n!$, the Shapley Value can be defined by the following formula, for each agent i:

$$\phi_i(v) = \sum_{s=0}^{n-1} \frac{s!(n-1-s)!}{n!} \sum_{\substack{S_{-i} \subseteq G \setminus \{i\} \\ |S_{-i}|=s}} (v(S_{-i} \cup \{i\}) - v(S_{-i})) \tag{1}$$

where $v : 2^n \rightarrow \{0,1\}$ for simple cooperation games. Computing the SV is $O(n^2)$.

An AAF [5] is a pair $\langle A, R \rangle$ consisting of a set A of arguments and of a binary relation R on A, called attack relation. Defining an argumentation semantics consists in providing criteria ruling which subsets of A can be accepted. Some well-known semantics are, for instance: conflict-free, admissible, complete, preferred, grounded and stable. Two main definition styles can be identified in the literature: *extension-based* and *labelling-based* ones [2]. In this section, we focus on reinstatement labelling [4].

Definition 1 (Labelling). *Let $F = \langle A, R \rangle$ be an argumentation framework and $\mathbb{L} = \{in, out, undec\}$. A labelling of F is a total function $L : A \rightarrow \mathbb{L}$ such that $in(L) \equiv \{a \in A \mid L(a) = in\}$, $out(L) \equiv \{a \in A \mid L(a) = out\}$ and $undec(L) \equiv \{a \in A \mid L(a) = undec\}$. We say that L is a reinstatement labelling if and only if it satisfies the following:*

- $\forall a \in A \mid a \in in(L)$, $\forall b \in A \mid (b,a) \in R, b \in out(L)$;
- $\forall a \in A \mid a \in out(L)$, $\exists b \in A \mid (b,a) \in R, b \in in(L)$.

The idea underlying the labelling-based approach is to give each argument a label, with the purpose to define a labelling-based semantics as follows.

Definition 2 (Labelling-based semantics). *A labelling-based semantics σ associates with an AAF* F *a subset of all the possible labellings for F, denoted as $L_\sigma(F)$. Let L be a labelling of $F = \langle A, R \rangle$, then L is*

- *conflict-free iff for each $a \in A$ it holds that:* if a *is labelled* in *then it does not have an attacker that is labelled* in*, and if a is labelled* out *then it has at least one attacker that is labelled* in*;*
- *admissible iff the attackers of each* in*-labelled argument are labelled* out*, and each* out*-labelled argument has at least one attacker that is* in*;*
- *complete iff for each $a \in A$, a is labelled* in *iff all its attackers are labelled* out*, and a is* out *iff it has at least one attacker that is labelled* in*;*

– *preferred/grounded if L is a complete labelling where the set of arguments labelled* in *is maximal/minimal (w.r.t. set-inclusion) among all complete labellings*
– *stable iff it is a complete labelling and* $undec(L) = \emptyset$.

In a framework F, the set of arguments labelled *in* for a labelling-based semantics σ corresponds to an extension of the semantics σ. We denote with L_σ a labelling L satisfying a semantics σ. Accordingly, $in(L_\sigma)$, $out(L_\sigma)$ and $undec(L_\sigma)$ refer to sets of arguments that are labelled in, out or undec, respectively, in at least one labelling of F. In Dung's framework [5], the acceptability of an argument depends on its membership to previously described sets. Another way to select a set of acceptable arguments is to rank arguments [1] from the most to the least acceptable ones.

Definition 3 (Ranking-based semantics). *A ranking-based semantics associates to any* $F = \langle A, R\rangle$ *a ranking* $\succcurlyeq_F$ *on A, where* $\succcurlyeq_F$ *is a pre-order (a reflexive and transitive relation) on A.* $a \succcurlyeq_F b$ *means that a is at least as acceptable as b* ($a \simeq b$ *is a shortcut for* $a \succcurlyeq_F b$ *and* $b \succcurlyeq_F a$, *and* $a \succ_F b$ *is a shortcut for* $a \succcurlyeq_F b$ *and* $b \not\succcurlyeq_F b$*).*

In the following we will use $\succcurlyeq$, $\simeq$, and $\succ$ omitting F when it is clear from the context. To describe some of the properties, we also need to define isomorphisms between AFs.

Definition 4 (AAF isomorphism). *An isomorphism* γ *between two AAFs* $F = \langle A, R\rangle$ *and* $F' = \langle A', R'\rangle$ *is a bijective function* $\gamma : A \longrightarrow A'$ *such that* $\forall a, b \in A, (a, b) \in R$ *iff* $(\gamma(a), \gamma(b)) \in R'$.

We now recall some of the logical properties for ranking-based semantics [1, 3] proposed in the literature.

Definition 5 (Properties). *Let* $F = \langle A, R\rangle$ *be an AAF and* $a, b \in A$ *and denote with* $P(b, a)$ *a path from b to a. The multi-set of* ***defenders*** *and* ***attackers*** *of a are* $R^+_n(a) = \{b \mid \exists P(b, a)$ *with length* $n \in 2\mathbb{N}\}$ *and* $R^-_n(a) = \{b \mid \exists P(b, a)$ *with length* $n \in 2\mathbb{N} + 1\}$, *respectively.* $R^-_1(a) = R^-(a)$ *is the set of direct attackers of a.*

Abstraction (Abs): *For any isomorphism* γ *s.t.* $F' = \gamma(F)$, $a \succcurlyeq_F b$ *iff* $\gamma(a) \succcurlyeq_{F'} \gamma(b)$.
Independence (Ind): $\forall F' \in cc(F)$, $\forall a, b \in Arg(F')$, *then* $a \succcurlyeq_{F'} b \Rightarrow a \succcurlyeq_F b$, *where* $cc(F)$ *denotes the set of connected components in* F.
Self-contradiction (SC): $(a, a) \notin R$ *and* $(b, b) \in R \Rightarrow a \succ b$.
Non-attacked Equivalence (NaE): $R^-(a) = \emptyset$ *and* $R^-(b) = \emptyset \Rightarrow a \simeq b$.
Argument Equivalence (AE): *for any* $F = \langle A, R\rangle$ *and* $\forall x, y \in A$, *for every isomorphism* λ *such that* $Anc_F(x) = \lambda(Anc_F(y))$, *then* $x \simeq^\sigma_F y$.
Total (ToT): $a \succcurlyeq b$ *or* $b \succcurlyeq a$.

3 Model Description

Our approach consists in assigning a boolean value to a subset of arguments according to the labels *in* and *out* if it satisfies the considered classical semantics. Compared to extension-based semantics, the use of labellings allows one to further distinguish among arguments by taking into account the ones that are not accepted (that is the *out* arguments). There is no convenience in taking into account also the label *undec* since it is derived directly from the other two.

Definition 6. *Let $F = \langle A, R \rangle$, σ be a semantics and L_σ the set of all possible labellings on F satisfying σ. Consider $S \subseteq A$. The "SV-based" ranking function is defined as:*

$$v^I_{\sigma,F}(S) = \begin{cases} 1, & \text{if } S \in in(L_\sigma) \\ 0, & \text{if otherwise} \end{cases} \qquad v^O_{\sigma,F}(S) = \begin{cases} 1, & \text{if } S \in out(L_\sigma) \\ 0, & \text{if otherwise} \end{cases}$$

We obtain a couple of values $\langle v^I_{\sigma,F}(S), v^O_{\sigma,F}(S) \rangle$ for each considered S. The ranking among arguments is then produced by considering a lexicographic ordering on the pairs, giving precedence to *in* and then to *out*. In particular, we can establish the rank of an argument $a \in A$ by computing its Shapley Values $\phi_a(v^{I/O}_{\sigma,F}(S))$.

Definition 7 (SV-based semantics). *The SV-based semantics associates to any framework $F = \langle A, R \rangle$ a ranking $\succcurlyeq^{SV}_F$ on A such that $\forall a, b \in A$, $a \succ^{SV}_F b$ iff*

- $\phi_a(v^I_{\sigma,F}) > \phi_b(v^I_{\sigma,F})$, *or*
- $\phi_a(v^I_{\sigma,F}) = \phi_b(v^I_{\sigma,F})$ *and* $\phi_a(v^O_{\sigma,F}) < \phi_b(v^O_{\sigma,F})$

and $a \simeq^{SV}_F b$ *iff* $\phi_a(v^I_{\sigma,F}) = \phi_b(v^I_{\sigma,F})$ *and* $\phi_a(v^O_{\sigma,F}) = \phi_b(v^O_{\sigma,F})$.

In the following, we show the properties that are satisfied by the proposed semantics.

Theorem 1 (Properties). *Considering two arguments $a, b \in A$ and the set $\Sigma = \{$conflict-free, admissible, complete, preferred, stable$\}$ with $\sigma \in \Sigma$, the SV ranked-semantics satisfies the following properties:* **Abs**, **Ind**, **NaE**, **AE**, **ToT** *for any $\sigma \in \Sigma$, and* **SC** *only if $\sigma =$ conflict-free.*

Theorem 2. *Given $a, b \in A$, if $\exists S \in in(L)$ s.t. $a \in S$ and $\nexists S' \in in(L)$ s.t. $b \in S'$ implies that $\phi_a(v^I_{\sigma,F}) > \phi_b(v^I_{\sigma,F}) \implies a \succ b$.*

Consider the example in Figure 1. The score of every argument in F, according to the considered semantics, is shown in Table 1, together with the final ranking for each SV-based semantics. We have that $v^I_{cf,F}(\bar{S}) = 1$ for $\bar{S} \in \{\{a\}, \{c\}, \{d\}, \{a,c\}, \{a,d\}\}$. The SV for the argument a is given by $\phi_i(v) = \frac{1!\cdot 3!}{5!} \cdot (v(\{a,b\}) - v(\{b\})) + \frac{2!\cdot 2!}{5!} \cdot (v(\{a,b,d\}) - v(\{b,d\})) = -0.084$. All other terms of the formula are equal to zero, since the gain given by a to any other sets of arguments is null. Due to the fact that the ranking function is non-monotone, the SV can be negative.

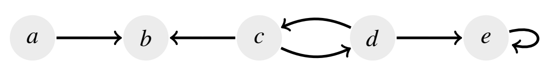

Fig. 1. Example of an AAF F. The computed sets of extensions for the conflict-fee, admissible, complete, preferred and stable semantics are: $CF = \{\{\}, \{a\}, \{b\}, \{c\}, \{d\}, \{a,c\}, \{a,d\}, \{b,d\}\}$, $ADM = \{\{\}, \{a\}, \{c\}, \{d\}, \{a,c\}, \{a,d\}\}$, $COM = \{\{a\}, \{a,c\}, \{a,d\}\}$, $PRE = \{\{a,c\}, \{a,d\}\}$ and $STA = \{\{a,d\}\}$, respectively.

	a	**b**	**c**	**d**	**e**	Semantics	Ranking
IN_{CF}	−0.084	−0.167	−0.167	−0.084	−0.5	SV-CF	$a \succ d \succ c \succ b \succ e$
OUT_{CF}	−0.167	0.25	0.0	−0.084	0.0		
IN_{ADM}	0.0	−0.417	−0.084	−0.084	−0.417	SV-ADM	$a \succ d \succ c \succ e \succ b$
OUT_{ADM}	−0.167	0.25	0.0	−0.084	0.0		
IN_{COM}	0.3	−0.117	−0.034	−0.034	−0.117	SV-COM	$a \succ c \simeq d \succ e \succ b$
OUT_{COM}	−0.133	0.283	−0.05	−0.05	−0.05		
IN_{PRE}	0.1	−0.0667	0.0167	0.0167	−0.0667	SV-PRE	$a \succ c \simeq d \succ e \succ b$
OUT_{PRE}	−0.0833	0.0833	0.0	0.0	0.0		
IN_{STA}	0.05	−0.0333	−0.0333	0.05	−0.0333	SV-STA	$a \simeq d \succ b \simeq c \simeq e$
OUT_{STA}	−0.05	0.0333	0.0333	−0.05	0.0333		

Table 1. SV for the arguments of the AAF in Figure 1, with final rankings for each semantics.

4 Conclusion

We have modelled a ranking-based semantics that takes advantage of two well-established concepts in the literature: Shapley Value [8] and semantics [2]. The semantics inherits all the good properties of SV and does not require external values to be computed. Moreover, our approach is able to distribute preferences among arguments by taking into account a particular semantics, allowing to obtain more precise rankings.

As future work, we would like to derive more SV functions than those presented in Section 3, with the purpose to further refine the ranking. Having more labels than just *in*, *out*, and *undec* would allow SV to distribute strength according to more levels of acceptance. Finally, we will check if all the properties in [3] are satisfied.

References

1. Amgoud, L., Ben-Naim, J.: Ranking-based semantics for argumentation frameworks. In: Proceedings of SUM 2013. LNCS, vol. 8078, pp. 134–147. Springer (2013)
2. Baroni, P., Caminada, M., Giacomin, M.: An introduction to argumentation semantics. Knowledge Eng. Review 26(4), 365–410 (2011)
3. Bonzon, E., Delobelle, J., Konieczny, S., Maudet, N.: A comparative study of ranking-based semantics for abstract argumentation. In: Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence. pp. 914–920. AAAI Press (2016)
4. Caminada, M.: On the issue of reinstatement in argumentation. In: Proceedings of JELIA 2006. Lecture Notes in Computer Science, vol. 4160, pp. 111–123. Springer (2006)
5. Dung, P.M.: On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. Artif. Intell. 77(2), 321–358 (1995)
6. Matt, P., Toni, F.: A game-theoretic measure of argument strength for abstract argumentation. In: Proceedings of JELIA 2008. LNCS, vol. 5293, pp. 285–297. Springer (2008)
7. Pollock, J.L.: How to reason defeasibly. Artif. Intell. 57(1), 1–42 (1992)
8. Shapley, L.S.: Contributions to the Theory of Games (AM-28), Volume II. Princeton University Press (1953)
9. Winter, E.: The shapley value. In: Aumann, R., Hart, S. (eds.) Handbook of Game Theory with Economic Applications, vol. 3, chap. 53, pp. 2025–2054. Elsevier, 1 edn. (2002)