Power Index-Based Semantics for Ranking Arguments in Abstract Argumentation Frameworks: an Overview

Carlo Taticchi[1] [0000−0003−1260−4672]

Gran Sasso Science Institute, L'Aquila, Italy - carlo.taticchi@gssi.it

Abstract. Ranking-based semantics for Abstract Argumentation Frameworks represent a well-established concept used for sorting arguments from the most to the least acceptable. This paper presents an overview of our ranking-based semantics that makes use of power indexes such as Shapley Value and Banzhaf Index. Such power index-based semantics is parametric to a chosen Dung semantics and inherits their properties.

Keywords: Argumentation · Ranking Semantics · Cooperative Game Theory · Power Indexes

1 Introduction

Argumentation Theory is a field of Artificial Intelligence that provides formalisms for reasoning with conflicting information. Arguments from a knowledge base are modelled by Dung [11] as nodes in a directed graph, that we call Abstract Argumentation Framework (AF in short), where edges represent attacks. Many *semantics* have been defined in order to establish different kinds of acceptability (see [2] for a survey). All these semantics return two disjoint sets of arguments: "accepted" and "not accepted". An additional level of acceptability is introduced in [9] with the reinstatement labelling, a semantics that marks as undecided the arguments that can be neither accepted nor rejected. Dividing the arguments into just three partitions could be not sufficient when dealing with very large AFs, so a different family of semantics has been defined for obtaining a broader range of acceptability levels for the arguments. Each of the defined *ranking-based* semantics [1,3,6,10,13,16,17] focus on a different criterion for identifying the best arguments in an AF.

In this paper, we give an overview of our work [4,5,6,7] towards the definition of a ranking-based semantics that relies on power indexes, like the Shapley Value and the Banzhaf index [14]. Our semantics is parametric to a chosen power index and allows for obtaining a ranking where the arguments are sorted according to their contribution to the acceptability of the other arguments in the various coalitions. To complete our study and support the research in this field, we also provide an online tool (ConArg[1]) capable of dealing with AFs and reasoning with our ranking-based semantics, besides classical ones.

[1] ConArg Website: http://www.dmi.unipg.it/conarg/

2 Preliminaries on Argumentation and Power Indexes

An *Abstract Argumentation Framework* [11] $\langle A, R\rangle$ consists of a set of arguments A and the relations among them $R \subseteq A \times A$. Such relations, which we call "attacks", are interpreted as conflict conditions that allow for determining the arguments in A that are acceptable together (i.e., collectively). An argumentation *semantics* is a criterion that establishes which are the acceptable arguments by considering the relations among them. The sets of accepted arguments with respect to a semantics are called *extensions*. Two leading characterisations can be found in the literature, namely *extension-based* [11] and *labelling-based* [9] semantics. While providing the same outcome in terms of accepted arguments, labelling-based semantics permits to differentiate between three levels of acceptability. In detail, a labelling of an AF is a total function $L : A \to \{in, out, undec\}$, with $in(L) = \{a \in A \mid L(a) = in\}$, $out(L) = \{a \in A \mid L(a) = out\}$ and $undec(L) = \{a \in A \mid L(a) = undec\}$. L is a reinstatement labelling if and only if it satisfies the following conditions:

- $\forall a, b \in A$, if $a \in in(L)$ and $(b, a) \in R$ then $b \in out(L)$;
- $\forall a \in A$, if $a \in out(L)$ then $\exists b \in A$ such that $b \in in(L)$ and $(b, a) \in R$.

A labelling-based semantics σ associates with an AF $F = \langle A, R\rangle$ a subset of all the possible labellings for F, denoted as $L_\sigma(F)$. For instance, we say that a labelling L of F is *admissible* if and only if the attackers of each *in* argument are labelled *out*, and each *out* argument has at least one attacker that is *in*[2]. The accepted arguments of F, with respect to a certain semantics σ, are those labelled *in* by σ. We refer to sets of arguments that are labelled *in*, *out* or *undec* in at least one labelling of $L_\sigma(F)$ with $in(L_\sigma)$, $out(L_\sigma)$ and $undec(L_\sigma)$, respectively.

In order to further discriminate among arguments, *ranking-based semantics* [8] can be used for sorting the arguments from the most to the least preferred. A ranking-based semantics associates with 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 $b \not\succcurlyeq_F b$). Such kind of semantics can be analysed in terms of properties defined on the obtained ranking of arguments [1]. For example, a ranking-based semantics satisfies *Cardinality Precedence* when arguments with more direct attackers are ranked lower than those with less direct attackers; and it satisfies *Totality* if all pairs of arguments can be compared.

In building our ranking-based semantics, we rely on power indexes for establishing a total order between the arguments of a framework. In game theory, cooperative games are a class of games where groups of players (or agents) are competing to maximise their goal, through one or more specific rules. In order to identify the "value" brought from a single player to a coalition, power indexes are used to define a preference relation between different agents, computed on

[2] There are other semantics that we consider in our work and which are omitted here due to space limitations.

all the possible coalitions. In our work [4,5,6,7], we studied and implemented Shapley Value, Banzhaf, Deegan-Packel and Johnston Index [14][3]. We provide here some intuition using the Banzhaf Power Index.

Every power index relies on a characteristic function $v : 2^N \rightarrow \mathbb{R}$ that, given the set N of players, associates each coalition $S \subseteq N$ with a real number in such a way that $v(S)$ describes the total gain that agents in S can obtain by cooperating with each other. The expected marginal contribution of a player $i \in N$, given by the difference of gain between S and $S \cup \{i\}$, is $v_{S_i} = v(S \cup \{i\}) - v(S)$. The Banzhaf Index $\beta_i(v)$ evaluates each player i by using the notion of *critical voter*: given a coalition $S \subseteq N \setminus \{i\}$, a critical voter for S is a player i such that $S \cup \{i\}$ is a winning coalition, while S alone is not. In other words, i is a critical voter if it can change the outcome of the coalition it joins.

$$\beta_i(v) = \frac{1}{2^{|N|-1}} \sum_{S \subseteq N \setminus \{i\}} v_{S_i} \tag{1}$$

The difference between the more famous Shapley Value and the Banzhaf index is that the latter does not take into account the order in which the players form the coalitions. Deegan and Packel assume that only minimal winning coalitions are formed, that they do so with equal probability, and that if such a coalition is formed it divides the (fixed) spoils of victory equally among its members. Finally, the Johnston index differs from Banzhaf's for the fact that critical voters in winning coalitions are rewarded with a fractional score instead of one whole unit.

3 Model Description

Our approach consists in assigning a value to each argument according to the labels *in* and *out* if it satisfies the considered classical semantics. An advantage of considering labelling-based semantics is that the characteristic functions only depend on the structure of a given AF, without adding to the picture other parameters, or external/computed values. Power indexes provides an a priori evaluation of the position of each player in a cooperative game, based on the contribution that each player brings to the different coalitions; in our ranking-based semantics, that we call *PI-based*, such coalitions are extension computed using classical Dung semantics.

Definition 1 (Characteristic function). *Consider an AF $F = \langle A, R \rangle$, a Dung semantics σ and the set L_σ of all possible labellings on F satisfying σ. For any $S \subseteq A$, the labelling-based characteristic functions $v^I_\sigma(S)$ and $v^O_\sigma(S)$ are defined as:*

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

[3] Other power indexes exist, such as the Public Good Index [15], that are relevant in cooperative game theory, and that we plan to study in the future.

The function $v^I_\sigma(S)$ takes into account the acceptability of a set of arguments S with respect to a certain semantics σ, assigning to such set a score equal to 1 if there exists a labelling L_σ in which all and only the arguments of S are labelled *in*. The higher the score of the power index, the better the rank of an argument. A second characteristic function, $v^O_\sigma(S)$, is also introduced to put attention on the negative effect of the attacks received by the arguments. The function $v^O_\sigma(S)$ considers the sets of arguments labelled *out* by σ, and the evaluation has the usual interpretation: the lower the score according to $v^O_\sigma(S)$, the worse the rank. Note that if $S \in in(L_\sigma)$ we can have $S' \subset S$ such that $S' \notin (L_\sigma)$. For instance, in Figure 1, $\{a, c\}$ is an extension of the admissible semantics, while $\{c\}$ is not.

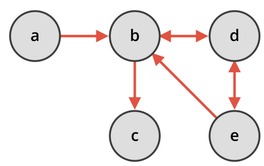

Fig. 1. Example of an AF in which the set of extensions for the admissible semantics is $ADM = \{\emptyset, \{a\}, \{d\}, \{e\}, \{a, c\}, \{a, d\}, \{a, e\}, \{c, d\}, \{c, e\}, \{a, c, d\}, \{a, c, e\}\}$.

In this paper, we use the Banzhaf Index β to provide an overview our ranking-based semantics, although any other power index π can be used for evaluating the arguments.

Definition 2 (PI-based semantics). *Let $F = \langle A, R \rangle$ be an AF, σ a Dung semantics, π a power index, and v_σ a characteristic function. The PI-based semantics associates to F a ranking $\succcurlyeq^\pi_\sigma$ on A, such that $\forall a, b \in A$,*

$$a \succcurlyeq^\pi_\sigma b \iff \pi_a(v_\sigma) \geq \pi_b(v_\sigma)$$

The strict relation is derived in the usual way.

A ranking-based semantics designed in this way has the further advantage of automatically inheriting the properties of the power indexes, like *efficiency*, *symmetry*, *linearity*, and *zero players* [19]. The lexicographic order on the pairs $(v^I_\sigma(S), v^O_\sigma(S))$ can be used to break possible ties in the final ranking, in the case two arguments of F have the same power index with respect to one of the two characteristic functions. An additional (partial) ordering can also be obtained as the Cartesian product of the two relations.

The PI-based semantics is capable of giving an overview of which are the most valuable arguments in a framework, from the point of view of their contribution to the existence of the various extensions belonging to different semantics. Indeed, it is reasonable to think that an argument which defend many other arguments should be given greater importance, when looking for sets satisfying the admissible semantics. In Table 1, we provide an example of ranking obtained through the PI-based semantics for the AF in Figure 1, with respect to the Banzhaf Index and the admissible semantics.

Table 1. Ranking for the AF in Figure 1 obtained through the Banzhaf Index.

	a	**b**	**c**	**d**	**e**	Semantics	Ranking
v^I_{ADM}	0.06250	−0.68750	−0.06250	−0.18750	−0.18750	$\beta - ADM$	$a \succ c \succ d \simeq e \succ b$
v^O_{ADM}	−0.25000	0.12500	−0.25000	−0.12500	−0.12500		

Besides conducting empirical experiments, in [7] we studied our semantics with respect to the properties introduced in [1], which describe and characterise the obtained rankings. We remark that those properties are not mandatory for obtaining a well-defined ranking and that different properties can be suitable for different applications [8]. For instance, the ranking produced by the power index β in combination with the characteristic function v^I_{ADM} satisfies the *Totality* property, but not the *Cardinality Precedence* (indeed, since we only take into account the acceptability of an argument, the number of direct attackers is not relevant to establish its value).

Finally, we implemented the PI-based semantics in the web interface of ConArg [7], a suite of tools developed with the purpose to facilitate research in the field of Argumentation in Artificial Intelligence. Four power indexes (Shapley Value, Banzhaf, Deegan-Packel and Johnston Index [14]) are available to compute the ranking, and can be chosen in combination with any Dung semantics for evaluating the arguments of a given AF. The output is provided for both the characteristic functions v^I_σ and v^O_σ. Besides ranking-based semantics, ConArg offers different functionalities to cope with various argumentation problems (like the computation of extensions).

4 Conclusion

In this paper we summarized the PI-based semantics presented in [4,5,6,7]. Differently from other ranking-based semantics defined in the literature, our approach allows for distributing preferences among arguments taking into account classical Dung/Caminada semantics. In this way, we obtain a more accurate ranking with respect to the desired acceptability criterion. We have also presented an online tool capable of dealing with ranking-based semantics, which implements the definition of the PI-based semantics [6].

The interest in solving argumentation problems has increased in the last few years, as also highlighted by the organisation of three editions of the *International Competition on Computational Models of Argumentation* (ICCMA 2015 [18], 2017 [12] and 2019). So far, ranking-based semantics have never been included in the competition and we believe that employing them in future editions can be useful to advance the research in this direction.

Acknowledgement

I want to thank with gratitude my supervisor, Professor Stefano Bistarelli, for his support in carrying out this work.

References

1. Amgoud, L., Ben-Naim, J.: Ranking-Based Semantics for Argumentation Frameworks. In: SUM 2013. LNCS, vol. 8078, pp. 134–147. Springer (2013)
2. Baroni, P., Caminada, M., Giacomin, M.: An introduction to argumentation semantics. Knowl. Eng Rev. **26**(4), 365–410 (2011)
3. Besnard, P., Hunter, A.: A logic-based theory of deductive arguments. Artif. Intell. **128**(1-2), 203–235 (2001)
4. Bistarelli, S., Faloci, F., Santini, F., Taticchi, C.: A Tool For Ranking Arguments Through Voting-Games Power Indexes. In: CILC 2019. CEUR Workshop Proceedings, vol. 2396, pp. 193–201. CEUR-WS.org (2019)
5. Bistarelli, S., Faloci, F., Taticchi, C.: Implementing Ranking-Based Semantics in ConArg. In: ICTAI 2019 (to appear). IEEE Computer Society (2019)
6. Bistarelli, S., Giuliodori, P., Santini, F., Taticchi, C.: A Cooperative-game Approach to Share Acceptability and Rank Arguments. In: AI^3@AI*IA 2018. CEUR Workshop Proceedings, vol. 2296, pp. 86–90. CEUR-WS.org (2018)
7. Bistarelli, S., Taticchi, C.: Power Index-Based Semantics for Ranking Arguments in Abstract Argumentation Frameworks. Intelligenza Artificiale (to appear) (2019)
8. Bonzon, E., Delobelle, J., Konieczny, S., Maudet, N.: A Comparative Study of Ranking-Based Semantics for Abstract Argumentation. In: AAAI-16. pp. 914–920. AAAI Press (2016)
9. Caminada, M.: On the Issue of Reinstatement in Argumentation. In: JELIA 2006. LNCS, vol. 4160, pp. 111–123. Springer (2006)
10. Dondio, P.: Ranking Semantics Based on Subgraphs Analysis. In: AAMAS 2018. pp. 1132–1140. IFAAMAS Richland, SC, USA / ACM (2018)
11. 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)
12. Gaggl, S.A., Linsbichler, T., Maratea, M., Woltran, S.: Introducing the second international competition on computational models of argumentation. In: SAFA@COMMA 2016. CEUR Workshop Proceedings, vol. 1672, pp. 4–9. CEUR-WS.org (2016)
13. Grossi, D., Modgil, S.: On the Graded Acceptability of Arguments. In: IJCAI 2015. pp. 868–874. AAAI Press (2015)
14. Keith, D.: Encyclopedia of Power. SAGE Publications, Inc. (2011)
15. Kurz, S.: Which criteria qualify power indices for applications? - A comment to "The story of the poor Public Good index". CoRR **abs/1812.05808** (2018)
16. Leite, J., Martins, J.: Social Abstract Argumentation. In: IJCAI 2011. pp. 2287–2292. IJCAI/AAAI (2011)
17. Matt, P.A., Toni, F.: A Game-Theoretic Measure of Argument Strength for Abstract Argumentation. In: JELIA. LNCS, vol. 5293, pp. 285–297. Springer (2008)
18. Thimm, M., Villata, S.: The first international competition on computational models of argumentation: Results and analysis. Artif. Intell. **252**, 267–294 (2017)
19. Winter, E.: The Shapley Value. In: Handbook of Game Theory with Economic Applications, vol. 3, pp. 2025–2054. Elsevier (2002)