

Cadres d'argumentation bipolaire quantitatifs avec incertitude

Jordan Theyre¹, Caren Al Anaissy¹, Aurélie Beynier¹, Sébastien Destercke²,
Nicolas Maudet¹, Srdjan Vesic³

¹ LIP6 - CNRS, Sorbonne Université, 4 place Jussieu, F-75005 Paris, France

² UMR CNRS 7253, Heudiasyc, Sorbonne Université, Université de Technologie de Compiègne,
Compiègne, France

³ CRIL - CNRS - Univ. Artois

{jordan.theyre, caren.al-anaissy, aurelie.beynier, nicolas.maudet}@lip6.fr
sebastien.destercke@hds.utc.fr
vesic@cril.fr

Résumé

Les débats sur les plateformes de délibération en ligne permettant l'échange d'arguments et les votes peuvent être modélisés à l'aide de l'argumentation abstraite, mais la rareté des votes sur les contributions des utilisateurs reste un défi. Dans ce papier, nous proposons une généralisation des Quantitative Bipolar Argumentation Frameworks en y incorporant de l'incertitude. Concrètement, nous utilisons les votes sur les arguments pour initialiser les intervalles de poids à l'aide du modèle de Dirichlet imprécis, permettant une représentation plus fiable de la force des arguments en présence d'un nombre limité de données.

1 Introduction

La délibération vise à favoriser des décisions collectives équilibrées. Les plateformes de débat en ligne, comme Debatabase¹, Kialo² ou Debategraph³, facilitent ce processus en structurant les échanges autour d'arguments "POUR" et "CONTRE", et permettent parfois aux utilisateurs de voter sur la pertinence des arguments.

Avec l'essor de ces outils dans les pratiques démocratiques numériques, il devient crucial de comprendre comment évaluer les arguments à grande échelle, en prenant en compte leurs interactions et, lorsque disponibles, les votes des participants. Lire l'ensemble des contributions d'un débat étant en général peu réaliste, il est pertinent de recourir à des méthodes automatisées d'analyse.

L'argumentation computationnelle fournit un cadre formel pour traiter l'information conflictuelle, incomplète ou incertaine. Elle modélise les débats sous forme de graphes d'arguments reliés par des relations d'attaque et de support, et utilise des sémantiques pour évaluer la force de chaque argument en fonction de ces relations. Ces sémantiques abstraient le contenu des arguments, réduisant ainsi les biais potentiels dans l'évaluation.

Parmi les nombreux cadres proposés, les *cadres d'argumentation bipolaires quantitatifs* (QBAFs) sont particulièrement adaptés aux débats en ligne. Ils intègrent à la fois les relations bipolaires entre arguments et des poids initiaux, souvent issus des votes. Les *sémantiques graduelles bipolaires* [2, 5, 1, 4] permettent ensuite de calculer un degré d'acceptabilité finale pour chaque argument.

Problème de recherche. Un défi majeur dans les débats en ligne est la *parcimonie des votes* : la plupart des utilisateurs ne votent que sur une faible proportion des arguments [3]. Ce phénomène limite la fiabilité des évaluations fondées sur les seuls votes. Plutôt que de tenter de prédire les votes manquants ou de solliciter davantage les utilisateurs, une approche conservatrice consiste à reconnaître cette incertitude et à s'assurer qu'elle soit correctement prise en compte dans les résultats affichés.

Notre approche. Nous généralisons le cadre d'argumentation bipolaire quantitatif afin de prendre en compte l'incertitude sur les poids initiaux des arguments. Pour cela, chaque argument est associé à un *intervalle* de poids possibles, reflétant la variabilité et la rareté des votes exprimés. Ces intervalles sont initialisés à partir des votes observés grâce au *modèle de Dirichlet imprécis*, qui permet de quantifier l'incertitude tout en évitant les inférences prématurées.

Nos contributions sont les suivantes :

- Introduction du cadre d'argumentation bipolaire quantitatif imprécis (*IQBAF*) pour modéliser les poids initiaux sous forme d'intervalles ;
- Utilisation du modèle de Dirichlet imprécis pour estimer ces poids à partir de données de vote partiellement observées ;
- Proposition d'une méthode pour appliquer une famille de sémantiques graduelles bipolaires à ces *IQBAFs* ;
- Validation expérimentale sur un corpus réel issu de la plateforme Kialo.

1. <https://idebate.net/resources/debatabase>

2. <https://www.kialo.com/>

3. <http://debategraph.org>

2 Cadres d'argumentation bipolaire quantitatifs imprécis

Lorsqu'un argument dispose de peu d'informations — notamment en cas de votes clairsemés — il devient difficile de lui attribuer un poids initial précis reflétant l'opinion des utilisateurs. Pour remédier à cela, nous nous inspirons de la théorie des probabilités imprécises et proposons un cadre dans lequel le poids initial d'un argument est représenté non plus par une valeur unique, mais par un *intervalle de valeurs possibles*. La borne inférieure (resp. supérieure) indique la valeur minimale (resp. maximale) que peut prendre ce poids. Ce formalisme donne lieu à la définition des *cadres d'argumentation bipolaires quantitatifs imprécis* (IQBAFs) comme suit :

Définition 1. Soit $\mathcal{I}_{[0;1]}$ l'ensemble des intervalles inclus ou égaux dans $[0; 1]$. Un cadre d'argumentation bipolaire quantitatif imprécis (IQBAF) est un quadruplet $(\mathcal{A}, \mathcal{R}^-, \mathcal{R}^+, i^i)$ où : (i) \mathcal{A} est un ensemble fini d'arguments, (ii) $\mathcal{R}^- \subseteq \mathcal{A} \times \mathcal{A}$ représente une relation binaire acyclique d'attaque, (iii) $\mathcal{R}^+ \subseteq \mathcal{A} \times \mathcal{A}$ représente une relation binaire acyclique de support, (iv) $i^i : \mathcal{A} \rightarrow \mathcal{I}_{[0;1]}$ associe à chaque argument a un intervalle de poids initiaux $i^i(a) = [\underline{i}^i(a), \overline{i}^i(a)]$.

3 Calcul des intervalles d'acceptabilité finale

Plutôt que d'introduire une nouvelle sémantique graduelle bipolaire dédiée aux IQBAFs, nous proposons une méthode pour adapter les sémantiques existantes afin de calculer les intervalles d'acceptabilité finale i^f . Pour chaque argument a , $i^f(a) = [\underline{i}^f(a), \overline{i}^f(a)]$ est défini à partir des bornes de ses attaquants et supporters :

$$\underline{i}^f(a) = \min_{\substack{x \in i^i(a) \\ Y \in \{y \in i^f(b) \mid b \in \mathcal{R}^-(a)\} \\ Z \in \{z \in i^f(c) \mid c \in \mathcal{R}^+(a)\}}} s^f(x, Y, Z)$$

et

$$\overline{i}^f(a) = \max_{\substack{x \in i^i(a) \\ Y \in \{y \in i^f(b) \mid b \in \mathcal{R}^-(a)\} \\ Z \in \{z \in i^f(c) \mid c \in \mathcal{R}^+(a)\}}} s^f(x, Y, Z)$$

Nous montrons que si la sémantique satisfait un ensemble de propriétés peu contraignantes, alors les bornes de $i^f(a)$ peuvent être obtenues en combinant : poids initial minimal (resp. maximal), degré d'acceptabilité finale des attaquants maximaux (resp. minimaux), et ceux des supporters minimaux (resp. maximaux). L'intervalle d'acceptabilité finale d'un argument a peut alors être calculé directement à partir des bornes inférieure et supérieure de ses attaquants et de ses supporters comme suit :

$$\underline{i}^f(a) = \sigma(\underline{i}^i(a), \{\overline{i}^f(b) \mid b \in \mathcal{R}^-(a)\}, \{\underline{i}^f(c) \mid c \in \mathcal{R}^+(a)\})$$

et

$$\overline{i}^f(a) = \sigma(\overline{i}^i(a), \{\underline{i}^f(b) \mid b \in \mathcal{R}^-(a)\}, \{\overline{i}^f(c) \mid c \in \mathcal{R}^+(a)\})$$

Nous montrons également que l'ajout d'information (en particulier sous forme de nouveaux votes) ne réduit pas nécessairement les intervalles d'incertitudes. Nous avons donc mené une expérimentation sur un jeu de données de presque 3000 débats issus de la plateforme Kialo.

4 Résultats expérimentaux

Pour chaque argument, nous avons appliqué la méthode *IDM* pour estimer un intervalle de poids initial à partir des votes, puis utilisé la sémantique *QuEM* pour calculer l'intervalle final. Seuls quelques débats montrent une réduction nette de l'incertitude, souvent liés à un nombre suffisant de votes ou à une structure dominée par un type de relation (attaque ou support). Concernant l'acceptabilité, on observe un biais global vers une moindre acceptation finale des thèses. Certains débats présentent un basculement marqué, l'intervalle final ne recoupant pas l'initial. De façon générale, bien que la thèse centrale concentre de nombreux votes, la prise en compte des arguments "POUR" et "CONTRE" élargit l'éventail des conclusions possibles.

Remerciements

Ce travail a été soutenu par le projet AGGREEY (ANR-22-CE23-0005) de l'ANR (Agence Nationale de la Recherche), ainsi que par une subvention de l'État français gérée par l'Agence Nationale de la Recherche dans le cadre du programme France 2030, référence ANR-22-EXEN-0004 (PEPR eNSEMBLE / PC3 MATCHING).

Références

- [1] Leila Amgoud and Jonathan Ben-Naim. Evaluation of arguments in weighted bipolar graphs. *International Journal of Approximate Reasoning*, 99 :39–55, 2018.
- [2] Pietro Baroni, Marco Romano, Francesca Toni, Marco Aurisicchio, and Giorgio Bertanza. Automatic evaluation of design alternatives with quantitative argumentation. *Argument & Computation*, 6(1) :24–49, 2015.
- [3] Beth Goldberg, Diana Acosta-Navas, Michiel Bakker, Ian Beacock, Matt Botvinick, Prateek Buch, Renée DiResta, Nandika Donthi, Nathanael Fast, Ravi Iyer, Zaria Jalan, Andrew Konya, Grace Kwak Danciu, Hélène Landemore, Alice Marwick, Carl Miller, Aviv Ovadya, Emily Saltz, Lisa Schirch, Dalit Shalom, Divya Sid-darth, Felix Sieker, Christopher Small, Jonathan Stray, Audrey Tang, Michael Henry Tessler, and Amy Zhang. AI and the future of digital public squares, ArXiv, <https://arxiv.org/abs/2412.09988>, 2024.
- [4] Nico Potyka. Continuous Dynamical Systems for Weighted Bipolar Argumentation. In *Proceedings of the Sixteenth International Conference, KR 2018, Tempe, Arizona*, July 2018.
- [5] Antonio Rago, Francesca Toni, Marco Aurisicchio, Pietro Baroni, et al. Discontinuity-free decision support with quantitative argumentation debates. *Fifteenth International Conference on Principles of Knowledge Representation and Reasoning*, 16 :63–73, 2016.