Encodage Logique et Explications Visuelles pour l'Argumentation

Théo Duchatelle[1]

[1] Université Toulouse III - Paul Sabatier, IRIT

theo.duchatelle@irit.fr

Résumé

En argumentation abstraite, le but est d'extraire des groupes d'arguments qui satisfont certaines contraintes, en se basant sur les conflits qui opposent les arguments. Ces groupes d'arguments peuvent être calculés en utilisant des outils logiques, dont certains prennent en compte des cas de généralisation du cadre de base. Une fois ces groupes d'arguments calculés, on peut vouloir chercher à les expliquer. On tâche d'apporter des éléments de réponse à cette question en se basant sur des critères visuels.

Mots-clés

Argumentation, Explicabilité, Logique, Graphes

Abstract

In abstract argumentation, the aim is to extract groups of arguments that satisfy some constraints, based on the conflicts that exist between the arguments. These groups of arguments can be computed using logical tools, some of them capturing cases of generalization of the basic framework. Once these groups of arguments are computed, we may seek to explain them. We try to bring some elements of answer, that are based on visual criteria, to this question.

Keywords

Argumentation, Explainability, Logic, Graphs

1 Introduction

L'argumentation abstraite est un mécanisme de prise de décision, introduit par Dung 1995, [4]. L'idée est de modéliser des entités abstraites appelées « arguments », ainsi que les conflits qui émergent entre eux (i.e. le fait qu'un argument en contredise un autre). Tout cela est appelé un cadre d'argumentation. On peut imaginer un cadre d'argumentation comme étant une sorte de débat, par exemple politique, pour décider de propositions à appliquer, ou juridique pour décider d'un jugement à rendre. Une fois un cadre établi, l'objectif est d'en extraire les arguments qui « gagnent » le débat, appelés collectivement extension, en utilisant des sémantiques qui définissent la manière de « gagner ».

Certains travaux se sont intéressés à généraliser cette notion. Ces travaux proposent ce qu'on appellera des « enrichissements » du cadre de base. On peut citer l'introduction d'une relation positive, de support, entre les arguments, le fait que plusieurs arguments contredisent en coalition un autre argument, ou le fait qu'un argument contredise le fait qu'un autre argument en contredise un troisième. On parle, dans le dernier cas, de relation d'ordre supérieur.

Il existe plusieurs méthodes pour calculer les extensions d'un cadre d'argumentation. L'une d'elles consiste à définir un encodage logique, pour obtenir les extensions via un mécanisme d'inférence. Dans cette lignée, Besnard et al. 2022a, [2] propose un encodage générique, qui permet de capturer plusieurs enrichissements et leurs combinaisons.

Une personne à qui une extension est présentée pourrait remettre en cause sa validité. Cette personne chercherait alors une explication au fait que les arguments présentés forment une extension. Čyras et al. 2021, [3] donne un état de l'art des méthodes d'explicabilité pour l'argumentation abstraite, ou l'utilisant pour expliquer. Les travaux de Besnard et al. 2022b, [1] exhibent des explications pour les processus d'argumentation qui reposent sur des critères visuels. Ces explications vont dans le sens de Vesic et al. 2022, [6], qui montre que des supports visuels aident à mieux comprendre le fonctionnement d'un processus argumentatif, et plus généralement de Miller 2019, [5] qui plaide pour des explications plus personnalisées et se reposant moins sur des mécanismes accessibles uniquement aux experts.

Ce papier a pour but de présenter brièvement Besnard et al. 2022a, [2] et Besnard et al. 2022b, [1].

2 Argumentation abstraite

Un *cadre d'argumentation* est un graphe orienté, où les noeuds représentent les *arguments* et la relation binaire représente les conflits. Un argument en *attaque* un autre s'il existe un arc du premier vers le deuxième (notion généralisable à l'attaque par un ensemble d'arguments).

Dans un contexte politique, les arguments pourraient être les propositions avancées et les conflits une incompatibilité entre les propositions. Dans un contexte juridique, les arguments pourraient être les jugements possibles et les éléments du procès, et les conflits une réfutation de véracité.

Un argument est *acceptable* selon un ensemble d'arguments si cet ensemble attaque tous les attaquants de l'argument. L'acceptabilité est au coeur des sémantiques classiques qui permettent de sélectionner les arguments qui « gagnent » le débat. Un ensemble d'arguments est dit *sans conflit* s'il n'existe pas d'arc qui relie deux de ses arguments (i.e. l'ensemble est cohérent). Il est dit *admissible* s'il est sans conflit et que tous ses arguments sont acceptables selon lui-même (i.e. il se défend contre tous ses attaquants).

Exemple. Dans le cadre d'argumentation de la figure 1,

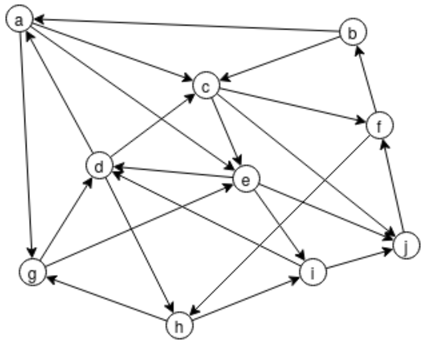

FIGURE 1 – Un exemple de cadre d'argumentation

l'ensemble $\{a, i, f\}$ est sans conflit et admissible.

3 Encodage logique

Pour calculer les extensions des sémantiques classiques d'un cadre d'argumentation, Besnard et al. 2022a, [2] a mis au point un encodage logique. Il utilise la logique du premier ordre, avec égalité, sans symbole de fonction. Il comprend deux parties : (i) une première qui encode uniquement le graphe ; (ii) une deuxième qui correspond à la sémantique dont il faut calculer les extensions.

On peut montrer que chaque modèle de la théorie encode une extension, et que pour toute extension, il existe un modèle de la théorie qui l'encode.

L'encodage logique est générique pour pouvoir capturer plusieurs enrichissements. Cela signifie que les formules de la partie qui encode la sémantique utilisent des prédicats laissés sans définition, dont il faut instancier la signification selon les enrichissements apportés. On illustre ce principe avec les formules correspondant à la sémantique sans conflit. Il s'agit des formules génériques (1).

$$\forall \alpha \in Att[Act(\alpha) \to \exists x \in E(T(\alpha, x) \land NA(x))] \quad (1a)$$

$$\forall x \in E\big(NA(x) \to \neg Acc(x)\big) \quad (1b)$$

$Att(x)$ signifie que x est une attaque, $Acc(x)$ signifie que x est dans l'extension, $NA(x)$ signifie que x ne peut pas être dans l'extension, $T(\alpha, x)$ signifie que x est la cible de α. $E(x)$ et $Act(x)$ sont ici des prédicats qui n'auront de signification que dans une instanciation particulière d'enrichissements. Ils représentent respectivement le fait que x peut faire partie de l'extension et que x est une interaction activable. Si le cadre possède des relations d'ordre supérieur, il faut instancier les prédicats E et Act avec les formules (2).

$$\forall x\big(E(x) \leftrightarrow Arg(x) \lor Att(x)\big) \quad (2a)$$

$$\forall \alpha[Act(\alpha) \leftrightarrow (\forall a \in Arg(S(\alpha, a) \to Acc(a)) \land Acc(\alpha))] \quad (2b)$$

Ici, on ajoute $Arg(x)$ qui signifie que x est un argument. Si le cadre d'argumentation a une relation de support, il faut instancier les prédicats E et Act avec les formules (3).

$$\forall x\big(E(x) \leftrightarrow Arg(x)\big) \quad (3a)$$

$$\forall \alpha[Act(\alpha) \leftrightarrow (\forall a \in Arg(S(\alpha, a) \to (Acc(a) \land Supp(a))))] \quad (3b)$$

Ici, on ajoute $Supp(x)$ qui signifie que x doit posséder un support. L'aspect modulaire de l'approche se ressent quand on combine les enrichissements. Par exemple, des formules (2) et (3), on peut facilement déduire les formules (4), qui correspondent au cas d'un cadre avec une relation de support et des relations d'ordre supérieur.

$$\forall x\big(E(x) \leftrightarrow Arg(x) \lor Att(x)\big) \quad (4a)$$

$$\forall \alpha[Act(\alpha) \leftrightarrow (\forall a \in Arg(S(\alpha, a) \to (Acc(a) \land Supp(a))) \land (Acc(\alpha) \land Supp(\alpha)))] \quad (4b)$$

4 Explications visuelles

Supposons qu'une personne utilise un système d'argumentation abstraite pour prendre une décision. Le système lui calcule une extension (sans conflit, admissible, ou autre selon des contraintes) dont les arguments servent de base à sa décision. Cependant, la personne est surprise par le résultat. Peut-être s'attendait-elle à autre chose, ou n'imaginait même pas que ce résultat était possible. Elle cherche donc à savoir ce qui en fait un résultat valable. Autrement dit, elle cherche à savoir pourquoi l'ensemble qui lui est présenté est une extension de la sémantique choisie.

Exemple. Une personne veut prendre une décision se basant sur le cadre d'argumentation de la figure 1. Les contraintes spécifiées sur la décision correspondent à une extension admissible. Le système lui renvoie l'ensemble $\{a, i, f\}$. Mais la personne est surprise car, par exemple, elle ne s'attendait pas à la présence de a dans l'extension. Elle cherche donc à savoir pourquoi $\{a, i, f\}$ est une extension admissible.

En ce sens, la réponse qui sera fournie à la personne, et qu'on appellera explication, doit pouvoir lui permettre de vérifier si un certain ensemble (a priori le résultat sur lequel elle s'interroge) est bien une extension de la sémantique choisie ou non. Si elle a accès au cadre d'argumentation initial, elle pourrait le vérifier elle-même en utilisant les définitions. Mais le graphe pouvant potentiellement être très grand et rempli, il semble pertinent de chercher à lui faciliter le travail. Afin de tirer parti de la nature *visuelle* des graphes, on cherchera donc à calculer une partie pertinente du cadre d'argumentation initial, telle que le processus de vérification puisse se reposer sur des conditions pouvant être *vues* dans cette partie (i.e. des conditions *structurelles*). Nos explications, telles que définies dans Besnard et al. 2022b, [1], sont donc des sous-graphes du cadre initial. On n'abordera ici que les cas du sans conflit et de l'admissibilité. Pour montrer qu'un ensemble est sans conflit, on calcule le sous-graphe induit par cet ensemble. On prouve que l'ensemble est sans conflit (propriété sémantique) si et seulement si le sous-graphe induit ne possède pas d'arcs (propriété structurelle / visuelle). Pour expliquer l'admissibilité

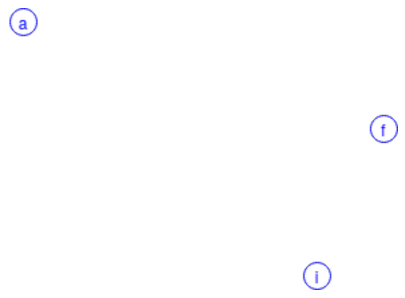

FIGURE 2 – Explication pour le fait que $\{a, i, f\}$ est sans conflit dans la figure 1 (aucun arc entre les arguments)

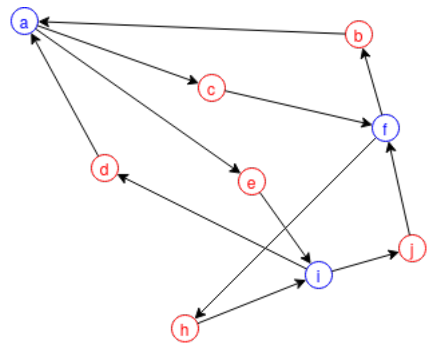

FIGURE 3 – Explication pour le fait que a, i, et f sont acceptables selon $\{a, i, f\}$ dans la figure 1 (ils ont des attaquants, en rouge, mais aucun n'est un noeud source)

d'un ensemble, on calcule le sous-graphe induit par l'ensemble et ses attaquants, puis on ne garde que les attaques qui vont de l'ensemble vers ses attaquants ou vice versa. On prouve que tous les arguments de l'ensemble sont acceptables (propriété sémantique) si et seulement s'il n'existe pas de noeud source parmi les attaquants dans l'explication (propriété structurelle / visuelle). Ainsi, ce sont les deux sous-graphes décrits ici qui, ensemble, forment une explication pour l'admissibilité d'un ensemble d'arguments.

Exemple. Les figures 2 et 3 *montrent* pourquoi (et donc forment une explication pour le fait que) $\{a, i, f\}$ est une extension admissible dans la figure 1. À l'inverse, la figure 4 *montre* pourquoi $\{i, f\}$ n'est pas une extension admissible dans la figure 1.

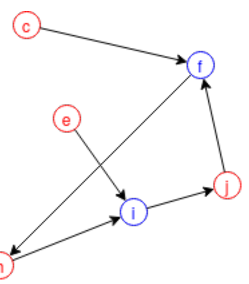

FIGURE 4 – Explication pour le fait que i et f ne sont pas acceptables selon $\{i, f\}$ dans la figure 1 (ils ont chacun un attaquant qui est un noeud source)

5 Conclusion et perspectives

Pour conclure, on a développé dans Besnard et al. 2022a, [2] une théorie logique de l'argumentation abstraite qui permet de calculer les extensions des sémantiques classiques. Cette théorie est générique car elle peut être paramétrée pour correspondre à des cadres plus ou moins enrichis. Cependant, certains enrichissements ne peuvent pas être retrouvés par cette théorie, par exemple les sémantiques graduelles, car ayant un fonctionnement fondamentalement trop différent. En plus de cela, on a défini dans Besnard et al. 2022b, [1] des explications qui permettent de certifier si un ensemble d'argument est une extension d'une sémantique ou non en se basant sur des propriétés visuelles. Cela permet de vérifier des résultats obtenus via des méthodes d'argumentation abstraite, même en étant non expert, car la vérification ne se base alors plus que sur des propriétés structurelles des graphes. En l'état ces explications ne permettent hélas pas encore de répondre à des questions plus raffinées, par exemple contrastives, comme « Pourquoi tel argument est présent dans l'extension et pas tel autre ? ».
Par la suite, on souhaiterait étendre notre outil logique pour lui permettre de calculer également les explications que l'on a définies. Cela aurait l'avantage supplémentaire de pouvoir étudier ces explications via le prisme des formalismes logiques et donc d'en extraire des qualités mieux exprimables mathématiquement que le fait de se baser sur des critères visuels, ou de pouvoir les comparer à des notions déjà existantes d'explications dans ces formalismes.

Remerciements

Je tiens à remercier Marie-Christine Lagasquie-Schiex et Sylvie Doutre pour leur bienveillance et leurs conseils avisés, ainsi que Philippe Besnard pour sa capacité à me guider durant mes débuts en tant que doctorant.

Références

[1] P. Besnard, S. Doutre, T. Duchatelle, and M.-C. Lagasquie-Schiex. Explaining semantics and extension membership in abstract argumentation. *Intelligent Systems with Applications*, 16 :200118, 2022.

[2] P. Besnard, S. Doutre, T. Duchatelle, and M.-C. Lagasquie-Schiex. Generic logical encoding for argumentation. *Journal of Logic and Computation*, 2022.

[3] K. Čyras, A. Rago, E. Albini, P. Baroni, and F. Toni. Argumentative XAI : A survey. In *Proc. of IJCAI*, pages 4392–4399, 2021.

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

[5] T. Miller. Explanation in artificial intelligence : Insights from the social sciences. *Artif. Intel.*, 267 :1–38, 2019.

[6] S. Vesic, B. Yun, and P. Teovanovic. Graphical representation enhances human compliance with principles for graded argumentation semantics. In *Proc. AAMAS*, 2022.