

Trouvez le traître grâce à l’argumentation abstrait(r)e

Projet de M1

Encadrants : Yann Munro*, Marie-Jeanne Lesot, Isabelle Bloch

1 Contexte

Imaginons une partie d’un jeu à rôle secret dans lequel un traître se cache parmi les joueurs et dont l’objectif pour eux est de le retrouver, comme par exemple *Undercover* ou bien *Among Us*. Avant de voter pour éliminer un joueur, une phase de discussion a lieu durant laquelle de nombreuses informations sont échangées de manière verbale et non verbale. Chacun va alors prendre une décision quant au joueur qu’il pense être le traître, c’est-à-dire qu’il évalue le degré de culpabilité des autres joueurs en fonction de ces différentes informations.

On peut modéliser de telles situations sous la forme d’un *graphe d’argumentation abstraite* [1] : ce formalisme logique permet de représenter et de raisonner sur des informations contradictoires. Pour cela, ces informations sont représentées sous la forme d’un ensemble d’éléments abstraits appelés *arguments*. Pour représenter un conflit entre arguments, on utilise une relation binaire appelée *relation d’attaque*. Il est ensuite possible, par l’intermédiaire de règles, de déterminer des ensembles d’arguments acceptables collectivement.

De nombreuses extensions ont été proposées depuis les travaux de Dung afin d’améliorer la représentativité de ce formalisme. Dans ce projet, on s’intéresse aux systèmes d’argumentation bipolaire pondérée (QBAF) [2] qui ajoutent une relation binaire supplémentaire, appelée *support*, ainsi qu’une force, représentée par un coefficient numérique, aux différents arguments. Dans ce formalisme, une méthode pour déterminer l’acceptabilité des arguments du modèle est de calculer un score, ou *degré d’acceptabilité*, à l’aide d’une fonction que l’on appelle une *sémantique graduelle*. Le choix de cette fonction est essentiel car c’est elle qui représente le “raisonnement argumentatif”. L’objectif du projet est d’étudier en particulier deux familles de sémantiques graduelles : les sémantiques *modulaires* [3] et les sémantiques *agrégatives* [4].

2 Objectifs du projet

L’objectif principal du projet est de comparer empiriquement les résultats fournis par ces deux familles de fonctions sur différents graphes d’argumentation bipolaire pondérée. Cela nécessite d’implémenter ces différentes sémantiques mais également de générer ces graphes d’argumentation.

Une implémentation des sémantiques modulaires existe déjà et est disponible dans un git¹. Outre la prise en main de ce code, une première étape du projet sera donc d’implémenter différentes sémantiques agrégatives. Cela passera principalement par l’implémentation de différentes fonctions d’agrégation [5], permettant de combiner les degrés d’attaque, de support et de force des arguments. Cela prendra la forme d’une librairie permettant d’ajouter et d’utiliser facilement de nouvelles fonctions de façon à ce que de nouvelles sémantiques agrégatives puissent être testées.

La deuxième étape consistera à permettre d’effectuer une comparaison quantitative en générant un grand nombre de QBAF de façon aléatoire, et en réfléchissant et implémentant des critères numériques d’évaluation des résultats obtenus, mais aussi une comparaison qualitative des résultats à l’aide de différents exemples préalablement choisis, inspirés par exemple de vos parties de jeux (réalisées en dehors du projet bien sûr...).

*yann.munro@lip6.fr

¹<https://github.com/TimKam/Quantitative-Bipolar-Argumentation>

References

- [1] Phan Minh Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial intelligence*, 77(2):321–357, 1995.
- [2] Pietro Baroni, Antonio Rago, and Francesca Toni. From fine-grained properties to broad principles for gradual argumentation: A principled spectrum. *International Journal of Approximate Reasoning*, 105:252–286, 2019.
- [3] Till Mossakowski and Fabian Neuhaus. Modular semantics and characteristics for bipolar weighted argumentation graphs. *arXiv preprint arXiv:1807.06685*, 2018.
- [4] Yann Munro, Isabelle Bloch, Mohamed Chetouani, Catherine Pelachaud, and Marie-Jeanne Lesot. Sémantique agrégative graduelle pour les systèmes d’argumentation bipolaires pondérés. In *Plate-Forme Intelligence Artificielle*, 2024.
- [5] Michel Grabisch, Jean-Luc Marichal, Radko Mesiar, and Endre Pap. *Aggregation functions*, volume 127. Cambridge University Press, 2009.