Étude expérimentale concernant l'impact de l'ordre des agents dans les protocoles d'argumentation

Nassim Lattab
Université Paris Cité
M1 IAD
F-75006 Paris, France
nassim.lattab@etu.u-paris.fr

Mohamed Azzaoui
Université Paris Cité
M1 IAD
F-75006 Paris, France
mohamed.azzaoui@etu.u-paris.fr

Résumé

Dans le domaine de l'argumentation, les protocoles d'argumentation sont des outils qui contiennent un ensemble de "règles" à respecter permettant le bon déroulement d'un débat. Ces protocoles permettent aux agents d'échanger des arguments afin de fournir une évaluation collective de la valeur d'un argument spécifique (c'est-à-dire une question ou un problème). L'idée globale est que les agents participant au débat ne rendent pas de compte à une autorité centrale de l'ensemble de leur système d'argumentation, mais contribuent plutôt au débat étape par étape, guidés par leur évaluation individuelle de l'état actuel de la discussion, et sans coordination avec d'autres agents (exactement comme dans les débats en ligne). Bien que certains protocoles aient été proposés dans la littérature, l'effet de l'ordre dans lequel les agents interviennent sur le résultat du débat est largement inconnu. Dans cette étude expérimentale, nous nous intéressons spécifiquement à cette dynamique multi-agents dans les protocoles d'argumentation. Notre recherche vise donc à mieux comprendre comment l'ordre des interventions des agents influence la conduite des débats et l'évolution de la valeur des problèmes discutés.

1 Introduction

La théorie de l'argumentation a été largement étudiée dans le domaine de l'intelligence artificielle et des systèmes multi-agents. Malgré sa capacité de représentation plutôt simpliste, l'argumentation abstraite est influente en raison de sa généralité. Dung a développé plusieurs sémantiques pour évaluer les groupes d'arguments. Des sémantiques graduelles ont été proposées comme alternative quantitative. Elles sont considérées comme plus appropriées pour évaluer de manière plus nuancée les arguments dans des contextes tels que les débats en ligne. Ces sémantiques ont été récemment développées et ont été étudiées du point de vue formel et informatique. Cependant, peu d'attention a été accordée aux aspects dynamiques de ces contextes. Notre question dans cet article est donc la suivante : Comment l'ordre d'intervention des agents dans un débat impacte-t-il son évolution?

Il s'agit d'une question très générale et il existe un certain nombre d'hypothèses que nous souhaitons expliciter d'emblée par souci de clarté :

- Cohérence agent-système : nous supposons que les opinions des agents et l'évaluation du débat par le système sont basées sur la même sémantique d'argumentation;
- L'accord sur la structure argumentative : alors que les agents peuvent avoir des opinions différentes parce qu'ils détiennent des ensembles d'arguments, ils sont d'accord sur les relations d'attaque entre ces arguments;
- L'indépendance des agents : chaque agent se comporte indépendamment des autres, nous ne considérerons pas les questions de coalitions, de communication ou d'influence directe entre les agents.

Ces hypothèses constituent un point de départ naturel pour l'étude d'une telle dynamique et une sorte de test de pertinence minimale pour une telle sémantique dans des contextes multi-agents.

2 Contexte

2.1 Théorie de l'argumentation

Dans cette partie, nous allons définir la manière dont nous allons représenter les arguments ainsi que leurs différentes relations d'attaques.

Définition 1. Un cadre d'argumentation (CA) est une paire $AF = \langle A, R\rangle$ où A est un ensemble fini et non vide d'arguments (abstraits) et R est une relation binaire sur A, c'est-à-dire $R \subseteq A \times A$, appelée relation d'attaque (dans notre cas les attaques n'ont pas de poids). Pour tout $x, y \in A$, $(x, y) \in R$ signifie que x attaque y. Un CA peut être représenté par un graphe orienté, appelé graphe d'argumentation (GA), dont les nœuds sont des arguments et les arêtes représentent la relation d'attaque.

Définition 2. Soit $AG = \langle A, R\rangle$ un graphe d'argumentation et $x, y \in A$ deux arguments. Un chemin P de y à x, noté $P(y, x)$, est une séquence $\langle x_0, \ldots, x_n\rangle$ d'arguments telle que $x_0 = x$, $x_n = y$ et $\forall i \in \{0, 1, \ldots, n-1\}, (x_{i+1}, x_i) \in R$. Nous notons par $l_P = n$ la longueur de P. Un argument est un défenseur (resp. attaquant) de x s'il se trouve au début d'un chemin de longueur paire (resp. impaire) menant à x.

2.2 Sémantiques graduées

Les sémantiques de Dung évaluent les arguments au niveau d'un ensemble : soit un ensemble d'arguments est acceptable (et donc une extension, selon une sémantique donnée), soit il ne l'est pas. Cependant, cette classification peut être trop binaire pour certaines applications, en particulier pour les plateformes de débat en ligne où l'acceptabilité d'un ensemble d'arguments peut être plus ou moins nuancée. D'où les efforts pour fournir des moyens alternatifs d'évaluer l'acceptabilité d'un argument donné. Les sémantiques basées sur le classement classent chaque argument d'un système d'argumentation, afin de comparer leurs forces individuelles. Les sémantiques basées sur la gradation attribuent un score ou une note à chaque argument.

Dans le cadre de ce projet, nous allons nous concentrer exclusivement sur ces sémantiques dites "graduelles".

Définition 3. Une sémantique graduelle est une fonction qui associe à un cadre d'argumentation $AF = \langle A, R\rangle$ une notation $S : A \to R$.

Définition 4. Le h-categorizer pondéré est défini comme suit :

$$Hbs(a) = \frac{w(a)}{1 + \sum_{b \in Att(a)} Hbs(b)}$$

La formule du h-categorizer pondéré, notée $Hbs(a)$, fonctionne de la manière suivante :

- $w(a)$ représente le poids de l'argument a. Ce poids peut être déterminé selon différents critères, mais généralement il reflète l'importance ou la crédibilité de l'argument dans le contexte donné.
- $Att(a)$ représente l'ensemble des arguments qui attaquent a, c'est-à-dire les arguments qui sont dirigés vers a dans le graphe.
- $\sum_{b \in Att(a)} Hbs(b)$ calcule la somme des scores de force de soutien pour tous les arguments attaqués par a. En d'autres termes, cela prend en compte la force de soutien de tous les arguments qui s'opposent à a.

La formule divise le poids de l'argument a par 1 plus la somme des scores de force de soutien de tous les arguments attaqués par a. Cela peut être interprété comme une normalisation qui prend en compte à la fois la force de soutien de l'argument lui-même et la force des arguments contre lesquels il est dirigé.

En outre, cette formule évalue la force de soutien d'un argument en tenant compte de son poids propre (dans notre cas, le poids vaut toujours 1) et de la force des arguments qui s'y opposent. Cela permet de quantifier le degré de confiance que l'on peut accorder à un argument dans le contexte d'un débat ou d'une argumentation.

Notez que par construction, cette fonction renverra une valeur dans l'intervalle $[0, w_{\max}]$. Lorsqu'il s'agit de graphes plats, nous supposons simplement que les poids sont égaux à 1 pour tous les arguments. Dans ce cas, nous retrouvons la définition classique du h-categorizer. C'est ce que nous utiliserons dans le cadre de ce projet. Ainsi, la nouvelle formule sera :

$$Hbs(a) = \frac{1}{1 + \sum_{b \in Att(a)} Hbs(b)}$$

Cette formule sera largement utilisée dans le cadre du calcul des valeurs du public graph et des IOAG que nous définirons par la suite.

3 Notre modèle

3.1 Cadre des débats multi-agents

Dans le cadre de cette étude, les graphes d'argumentation utilisés pour modéliser les débats multi agents doivent satisfaire un certain nombre de conditions : tout d'abord, un argument spécifique (le problème) joue le rôle de la question principale du débat, et tous les arguments doivent être liés à ce problème.

Définition 5. Soit $AG = \langle A, R \rangle$ un graphe d'argumentation et $i \in A$ un argument. $DG = \langle A, R, i \rangle$ est un issue-oriented argumentation graph (IOAG) de problème i si i est la racine du graphe formé par les nœuds de A et les arêtes de R et si toutes les attaques sont dirigées vers la racine i, c'est-à-dire que pour tous $x, y \in A$, s'il existe un chemin entre x et y, alors ce chemin est un sous-ensemble d'un chemin de x à i, la racine de DG.

Ici, nous nous situons dans un cadre multi agent : nous allons donc travailler avec un ensemble $N = \{1, 2, \ldots, n\}$ d'agents. Nous supposons que chaque débat sera caractérisé par un graphe d'argumentation orienté par les problèmes UG, le graphe univers, contenant chaque argument pertinent pour la question du débat. Chaque agent est équipé d'un IOAG privé, composé d'un sous-ensemble de nœuds de UG, appelé son graphe d'opinion, et représentant sa propre vision du monde. Par conséquent, tous les agents sont d'accord sur les relations d'attaque entre les arguments qu'ils connaissent, et tous les graphes partagent la même question. Nous appellerons la collection de IOAG $\langle DG_1, DG_2, \ldots, DG_n \rangle$ le profil du jeu.

Définition 6. Soit $N = \{1, 2, \ldots, n\}$ un ensemble de n agents, $UG = \langle A, R, i \rangle$ le graphe univers, et $i \in A$ la question du débat. Le profil du jeu $\langle DG_1, DG_2, \ldots, DG_n \rangle$ est une collection de IOAG, un pour chaque agent, tel que $\forall k \in N, DG_k = \langle A_k, R_k, i \rangle$, où $A_k \subseteq A$ et $\forall x, y \in A_k, (x, y) \in R_k$ si et seulement si $(x, y) \in R$.

3.2 Stratégie de protocole

L'état du jeu à l'étape t est décrit par un tuple (U_G, P_{G_t}, N) où U_G est le graphe de l'univers (qui n'est jamais modifié) et P_{G_t} est le graphe du débat public à l'étape t, visible par tous les agents. Au début du jeu, $P_{G_0} = \langle i, \emptyset, i \rangle$ est composé uniquement de la question.

Étant donné que notre sujet d'étude porte sur l'impact de l'ordre des agents dans un débat, ils joueront l'un après l'autre dans un ordre fixe pour chaque tour. À chaque tour, chacun des agents cherche à jouer parmi ses arguments celui qui maximise sa zone de confort (la notion de confortabilité sera explicitée plus tard). Tous les arguments joués au tour t sont donc ajoutés au public graph permettant alors de mettre à jour sa valeur V_{Pt}.

Définition 7. Un agent k est à l'aise (dans sa zone de confort) à l'étape t si la valeur du graphe de débat public à cette étape se situe à l'intérieur de sa zone de confort, un intervalle autour de sa valeur idéale V_k ; c'est-à-dire, $V_{Pt} \in [V_k - cl, V_k + cl]$. avec $V_k = hbs(i)$, i la racine du public graph de k.

Si un agent k n'est pas à l'aise, il peut jouer n'importe quel argument présent dans son graphe d'opinion, qui attaque directement au moins un argument de PG_{t-1} et dont la valeur hypothétique est plus proche de son opinion que la valeur actuelle du graphe public (il choisit l'argument qui maximise cette valeur hypothétique).

Si un agent k est à l'aise, il ne joue pas et conserve ses arguments pour un futur tour. Tant qu'il reste à l'aise, il ne participe pas au débat et garde ses arguments en réserve, prêt à les utiliser si nécessaire ultérieurement (donc lorsqu'il ne sera plus à l'aise).

Si l'agent n'est pas à l'aise et a le choix entre deux arguments dont la valeur hypothétique est plus proche de son opinion que la valeur actuelle du graphe public et que ces deux valeurs sont identiques. Il choisit aléatoirement entre les deux un argument à jouer.

Le jeu s'arrête à l'étape T si la stratégie de chaque agent à cette étape est de ne rien faire. Comme le nombre d'arguments connus par les agents est fini, le jeu se termine toujours de manière triviale. Si A est le nombre total d'arguments distincts connus par les agents N, alors A est également trivialement une borne supérieure pour le nombre d'étapes avant la terminaison.

4 Conlusion et perspectives d'ouvertures

/// à rédiger