Modelling Uncertainty in Persuasion

Anthony Hunter[1]

Department of Computer Science, University College London,
Gower Street, London WC1E 6BT, UK

Abstract. Participants in argumentation often have some doubts in their arguments and/or the arguments of the other participants. In this paper, we model uncertainty in beliefs using a probability distribution over models of the language, and use this to identify which are good arguments (i.e. those with support with a probability on or above a threshold). We then investigate three strategies for participants in dialogical argumentation that use this uncertainty information. The first is an exhaustive strategy for presenting a participant's good arguments, the second is a refinement of the first that selects the good arguments that are also good arguments for the opponent, and the third selects any argument as long as it is a good argument for the opponent. We show that the advantage of the second strategy is that on average it results in shorter dialogues than the first strategy, and the advantage of the third strategy is that under some general circumstances the participant can always win the dialogue.

1 Introduction

Persuasion is a complex multifaceted concept. In this paper, we consider uncertainty in persuasion which is a topic that is underdeveloped in formal models of argument. We represent the uncertainty that an agent has over its own beliefs by a probability distribution over the models of the language. The agent uses this to judge which arguments are "good arguments" (arguments with premises with a probability on or above a "good argument" threshold), and which are "good targets" (arguments with premises with a probability on or below a "good target" threshold) and as such should be attacked if an attacker exists. The idea is that if an argument is a good argument but not a good target, then the agent considers the argument but ignores any attack on it. To illustrate, consider the following arguments[1]. Suppose each is a good argument.

- A_1 "The metro is the best way to the airport."
- A_2 "There is a strike today by metro workers."

Now consider the threshold for attack. It would be reasonable in this context to take a skeptical view (because we worry about missing the flight) and set the threshold for being a good target to be above the threshold for being a good argument. So even if the threshold for being a good argument might be set to a high level, we might want

[1] Note, we not proposing a formal model of argument-based decision making (c.f. [1]), but rather investigating criteria for selecting arguments and attacks to present in argumentation.

the threshold for being a good target to be even higher. Therefore, for this example, we would get an argument graph with both arguments where A_2 attacks A_1.

As an alternative example, consider A_3 and the potential counterargument A_2 where both are good arguments. Here we might take a credulous view (because we might not worry too much about the risk or consequences of delay on the metro when going home). So we set the threshold for being a good argument as above the threshold for being a good target. In this context, we may say that even though there exists the counterargument A_2, A_3 is not a good target because the threshold for being a good target is lower in this case. In other words, there is insufficient doubt in A_3 for A_2 to attack it. In this example, this is reasonable since often some trains still run when there are problems with the service.

- A_3 "The metro is the best way to go home"

As well as considering how an agent might judge its own arguments and counterarguments using its probability distribution, we also want to consider how it can be used in dialogue strategies. For this, we let an agent have an estimate of its opponent's probability distribution. This can be used to make the argumentation more efficient and/or more efficacious. There is no point in presenting arguments that are not going to persuade an opponent, particularly when there may be alternative arguments that could bring about the required outcome. Consider the following dialogue where participant 1 (husband) wants to persuade participant 2 (wife) to buy a particular car. Argument A_5 indicates that participant 2 does not believe argument A_4, and so participant 1 has not used a good argument to persuade participant 2.

- A_4 "The car is a nice red colour, and that is the only criterion to consider, therefore we should buy it."
- A_5 "It is a nice red colour, but I don't agree that that is the only criterion to consider."

Now consider argument A_6 which participant 2 sees as a good argument but not a good target. So if participant 1 has a good estimate of the probability distribution of participant 2, then it could see A_6 as better to posit than A_4, and that this could result in a more persuasive dialogue.

- A_6 "The car is the most economical and easy car to drive out of the options available to us, and those are the criteria we want to satisfy, so we should buy the car."

In this paper, we formalise good argument and good attack, and investigate their use in persuasion dialogues.

2 Preliminaries

We review abstract argumentation [2], probabilistic logic [3], and the use of probabilistic logic in argumentation [4].

2.1 Abstract argumentation

An **abstract argument graph** is a pair $(\mathcal{A}, \mathcal{R})$ where $\mathcal{A}$ is a set and $\mathcal{R} \subseteq \mathcal{A} \times \mathcal{A}$. Each element $A \in \mathcal{A}$ is called an **argument** and $(A, B) \in \mathcal{R}$ means that A **attacks** B (accordingly, A is said to be an **attacker** of B) and so A is a **counterargument** for B. A set of arguments $S \subseteq \mathcal{A}$ **attacks** $A_j \in \mathcal{A}$ iff there is an argument $A_i \in S$ such that A_i attacks A_j. Also, S **defends** $A_i \in S$ iff for each argument $A_j \in \mathcal{A}$, if A_j attacks A_i then S attacks A_j. A set $S \subseteq \mathcal{A}$ of arguments is **conflict-free** iff there are no arguments A_i and A_j in S such that A_i attacks A_j. Let Γ be a conflict-free set of arguments, and let $\mathsf{Defended} : \wp(\mathcal{A}) \mapsto \wp(\mathcal{A})$ be a function such that $\mathsf{Defended}(\Gamma) = \{A \mid \Gamma \text{ defends } A\}$. We consider the following extensions: (1) Γ is a **complete extension** iff $\Gamma = \mathsf{Defended}(\Gamma)$; and (2) Γ is a **grounded extension** iff it is the minimal (w.r.t. set inclusion) complete extension. For $G = (\mathcal{A}, \mathcal{R})$, let $\mathsf{Nodes}(G) = \mathcal{A}$ and let $\mathsf{Grounded}(G)$ be the grounded extension of G.

2.2 Probabilistic logic

We use an established proposal for capturing probabilistic belief in classical propositional formulae [3]. For this, we assume that the propositional language $\mathcal{L}$ is finite. The set of models (i.e. interpretations) of $\mathcal{L}$ is denoted $\mathcal{M}^{\mathcal{L}}$. Each **model** in $\mathcal{M}^{\mathcal{L}}$ is an assignment of $true$ or $false$ to the formulae of the language defined in the usual way for classical logic. So for each model m, and $\psi \in \mathcal{L}$, $m(\psi) = true$ or $m(\psi) = false$. For $\phi \in \mathcal{L}$, $\mathsf{Models}(\phi)$ denotes the set of models of ϕ (i.e. $\mathsf{Models}(\phi) = \{m \in \mathcal{M}^{\mathcal{L}} \mid m(\phi) = true\}$), and for $\Delta \subseteq \mathcal{L}$, $\mathsf{Models}(\Delta)$ denotes the set of models of Δ (i.e. $\mathsf{Models}(\Delta) = \cap_{\phi \in \Delta} \mathsf{Models}(\phi)$). Let $\Delta \models \psi$ denote $\mathsf{Models}(\Delta) \subseteq \mathsf{Models}(\psi)$.

Let $\mathcal{L}$ be a propositional language and let $\mathcal{M}^{\mathcal{L}}$ be the models of the language. A function $P : \mathcal{M}^{\mathcal{L}} \rightarrow [0, 1]$ is a **probability distribution** iff $\sum_{m \in \mathcal{M}^{\mathcal{L}}} P(m) = 1$. From a probability distribution, we get the **probability of a formula** $\phi \in \mathcal{L}$ as follows: $P(\phi) = \sum_{m \in \mathsf{Models}(\phi)} P(m)$.

Example 1. Let the atoms of $\mathcal{L}$ be $\{a, b\}$, and so $\mathcal{L}$ is the set of propositional formulae formed from them. Let m_1 and m_2 be models s.t. $m_1(a) = true$, $m_1(b) = true$, $m_2(a) = true$, and $m_2(b) = false$. Now suppose $P(m_1) = 0.8$ and $P(m_2) = 0.2$. Hence, $P(a) = 1$, $P(a \wedge b) = 0.8$, $P(b \vee \neg b) = 1$, $P(\neg a \vee \neg b) = 0.2$, etc.

For any probability distribution P, if $\models \alpha$, then $P(\alpha) = 1$, and if $\models \neg(\alpha \wedge \beta)$, then $P(\alpha \vee \beta) = P(\alpha) + P(\beta)$.

2.3 Logical arguments

We use deductive arguments based on classical logic to instantiate abstract argument graphs. Let $\Delta \subseteq \mathcal{L}$ be a set of propositional formulae and let $\vdash$ be the classical consequence relation. $\langle \Phi, \alpha \rangle$ is a **deductive argument** (or simply **argument**) iff $\Phi \subseteq \Delta$ and $\Phi \vdash \alpha$ and $\Phi \not\vdash \bot$ and there is no $\Psi \subset \Phi$ s.t. $\Psi \vdash \alpha$. For an argument $A = \langle \Phi, \alpha \rangle$, let $\mathsf{Support}(A) = \Phi$ and $\mathsf{Claim}(A) = \alpha$. Let $\mathsf{Arg}(\Delta)$ be the set of deductive arguments obtained from Δ. For counterarguments, we use direct undercuts [5, 6]. Argument A is a **direct undercut** of argument B when $\mathsf{Claim}(A)$ is $\neg\psi$ for some $\psi \in \mathsf{Support}(B)$. The set of direct undercuts is $\mathsf{Ucuts}(\Delta) = \{(A, B) \mid A, B \in \mathsf{Arg}(\Delta)$ and A is a direct undercut of $B\}$. The probability of an argument is the probability of its support.

Definition 1. *Let P be a probability distribution on $\mathcal{M}^{\mathcal{L}}$. The* **probability of an argument** $\langle\Phi, \alpha\rangle \in \mathsf{Arg}(\mathcal{L})$*, denoted $P(\langle\Phi, \alpha\rangle)$, is $P(\phi_1 \wedge \ldots \wedge \phi_n)$, where $\Phi = \{\phi, \ldots, \phi_n\}$.*

Example 2. Consider the following probability distribution over models (with atoms a and b) for each participant.

Model	a	b	Participant 1	Participant 2
m_1	true	true	0.5	0.0
m_2	true	false	0.5	0.0
m_3	false	true	0.0	0.6
m_4	false	false	0.0	0.4

Let $\Delta_1 = \{a, \neg b\}$ (resp. $\Delta_2 = \{b, \neg b, b \to \neg a\}$) be the knowlegebase for participant 1 (resp. 2). Below is the probability of each argument according to each participant.

Argument	Participant 1	Participant 2
$A_1 = \langle\{a\}, a\rangle$	1.0	0.0
$A_2 = \langle\{b, b \to \neg a\}, \neg a\rangle$	0.0	0.6
$A_3 = \langle\{\neg b\}, \neg b\rangle$	0.5	0.4

It is possible for the knowledgebase to be inconsistent and yet for the participant to have a probability distribution over the models, as illustrated by Example 2.

3 Good arguments and good attacks

Each agent has a knowledgebase Δ, and a probability distribution P, and these are used to identify good arguments.

Definition 2. *For a knowledgebase Δ, a probability distribution P, and a threshold $T \in [0, 1]$, the set of* **good arguments** *is* $\mathsf{GoodArg}(\Delta, P, T) = \{A \in \mathsf{Arg}(\Delta) \mid P(A) \geq T\}$.

Hence, if $T = 0$, then all arguments from the knowledgebase are good arguments (i.e. $\mathsf{GoodArg}(\Delta, P, T) = \mathsf{Arg}(\Delta)$). Whereas if $T = 1$, then only arguments that have premises that are certain are good arguments. Furthermore, if $T = 1$, then the premises of the good arguments are consistent together (i.e. $(\cup_{A \in \mathsf{GoodArg}(\Delta,P,T)} \mathsf{Support}(A)) \not\vdash \bot$), and so there are no $A, B \in \mathsf{GoodArg}(\Delta, P, T)$ such that A is a direct undercut of B when $T = 1$.

Good targets (defined next) are arguments for which there is sufficient doubt in their premises for us to want to attack them even if an attacker exists. If an argument is not a good target, then we will ignore attacks on it. This is a form of inconsistency/conflict tolerance allowing us to focus on the more significant inconsistencies/conflicts.

Definition 3. *For a probability distribution P, a threshold $S \in [0,1]$, and a knowledgebase Δ, the set of* **good targets** *is* $\mathsf{GoodTarget}(\Delta, P, S) = \{B \in \mathsf{Arg}(\Delta) \mid P(B) \leq S\}$.

If $S = 1$, then $\mathsf{GoodTarget}(\Delta, P, S) = \mathsf{Arg}(\Delta)$, whereas if $S = 0$, then only arguments with support with zero probability are good targets. Next, we use S to select the attacks.

Definition 4. *For a knowledgebase Δ, a probability function over arguments P, and a threshold $S \in [0,1]$, the set of* **good attacks** *is* $\mathsf{GoodAttack}(\Delta, P, S) = \{(A, B) \mid (A, B) \in \mathsf{Ucuts}(\Delta)$ *and* $P(B) \leq S\}$.

Given a knowledgebase, and a probability distribution, a good graph is the set of good arguments and good attacks that can be formed.

Definition 5. *For a knowledgebase Δ, thresholds $T, S \in [0,1]$, and a probability distribution P, the* **good graph** *is an argument graph,* $\mathsf{GoodGraph}(\Delta, P, T, S) = (\mathcal{A}, \mathcal{R})$, *where* $\mathcal{A} = \mathsf{GoodArg}(\Delta, P, T)$ *and* $\mathcal{R} = \mathsf{GoodAttack}(\Delta, P, S)$.

For considering whether or not a specific argument is in the grounded extension of a (good) graph, we only need to consider the component containing it, as illustrated next.

Example 3. Suppose $\Delta = \{a, \neg a\}$. Let $A_1 = \langle \{a\}, a \rangle$ and $A_2 = \langle \{\neg a\}, \neg a \rangle$. Suppose we want to determine whether A_1 is in the grounded extension of the good graph. Depending on the choice of P, S, and T, the component to consider is one of G_1 to G_6 where G_1 is $(\emptyset, \emptyset)$ when $P(A_1) < T$, and the constraints for G_2 to G_6 are tabulated below.

Graph	Structure	$P(A_1)?T$	$P(A_2)?T$	$P(A_1)?S$	$P(A_2)?S$
G_2	A_1	$P(A_1) \geq T$	$P(A_2) < T$	n/a	n/a
G_3	$A_1 \quad A_2$	$P(A_1) \geq T$	$P(A_2) \geq T$	$P(A_1) > S$	$P(A_2) > S$
G_4	$A_1 \leftarrow A_2$	$P(A_1) \geq T$	$P(A_2) \geq T$	$P(A_1) \leq S$	$P(A_2) > S$
G_5	$A_1 \rightarrow A_2$	$P(A_1) \geq T$	$P(A_2) \geq T$	$P(A_1) > S$	$P(A_2) \leq S$
G_6	$A_1 \leftrightarrow A_2$	$P(A_1) \geq T$	$P(A_2) \geq T$	$P(A_1) \leq S$	$P(A_2) \leq S$

Proposition 1. *If $T > S$, then $\forall$ Δ, P,* $\mathsf{GoodArg}(\Delta, P, T) \cap \mathsf{GoodTarget}(\Delta, P, S) = \emptyset$*. and if $T \leq S$, then $\exists$ Δ, P s.t.* $\mathsf{GoodArg}(\Delta, P, T) \cap \mathsf{GoodTarget}(\Delta, P, S) \neq \emptyset$*. Also if $T = 0$ and $S = 1$, then $\forall$ Δ, P,* $\mathsf{GoodArg}(\Delta, P, T) = \mathsf{GoodTarget}(\Delta, P, S)$.

Example 4. Consider the arguments $A_1 = \langle \{a\}, a \rangle$ and $A_2 = \langle \{\neg a\}, \neg a \rangle$ generated from Δ where $T = 0$ and $S = 1$. Whatever choice is made for P, either $P(A_1) < 1$ or $P(A_2) < 1$ or both $P(A_1) < 1$ and $P(A_2) < 1$. So if $T = 0$ then both arguments are good arguments, and if $S = 1$ then each argument attacks the other.

In the following, we consider how components in good graphs are constructed via dialogical argumentation. For this, we will assume $S < T$, and so T affects the choice of arguments to present, and S affects the choice of counterarguments to present.

4 Participants

We will assume two participants called 1 and 2 where 1 wants to persuade 2 about a claim ϕ which we refer to as the **persuasion claim**. Informally, for participant 1 to persuade participant 2 to accept the persuasion claim, it needs to give an argument with claim ϕ that is in the grounded extension of the argument graph produced during the dialogue. We formalize this in the next section.

For the good argument threshold T, and the good target threshold S, we assume $S < T$ so that each participant cannot attack its own good arguments. Each participant has a **position**: Participant 1 has position $\Pi_1 = (\Delta_1, P_1, P', T, S, \phi)$ containing its knowledgebase Δ_1, its probability distribution P_1, the probability distribution P' which is an estimate of the probability distribution P_2 of the other agent, the thresholds T and S, and the persuasion claim ϕ, and participant 2 has a position $\Pi_2 = (\Delta_2, P_2, T, S)$ containing its knowledgebase Δ_2, its probability distribution P_2, and the thresholds T and S. Note, position 1 has more parameters because participant 1 has the lead role in the dialogue. Also, note each participant does not know the position of the other participant (apart from S and T).

Participant 1 can build P' as an estimate of P_2 over time, such as by learning from previous dialogues. However, participant 1 does not know whether P' is a good estimate of P_2. But, we as external observers do know Π_1 and Π_2, and so we can measure how well P' models P_2. For this, we use a rank correlation coefficient which assigns a value in $[-1, 1]$ such that when P' and P_2 completely agree on the ranking of the arguments, the coefficient is 1, and when they completely disagree on the ranking of the arguments (i.e. one is the reverse order of the other), the coefficient is -1 (as defined next).

Consider the set of arguments $\mathsf{Arg}(\Delta)$ for some Δ and the threshold S. We compare P' and P_2 in terms of how they rank each argument in $A \in \mathsf{Arg}(\Delta)$ with respect to S. Let n_a be the number of arguments that P' and P_2 agree on (i.e. $n_a = |\{A \in \mathsf{Arg}(\Delta) \mid (P'(A) > S \text{ and } P_2(A) > S) \text{ or } (P'(A) \leq S \text{ and } P_2(A) \leq S)\}|$, and let n_d be the number of arguments that P' and P_2 disagree on (i.e. $n_d = |\{A \in \mathsf{Arg}(\Delta) \mid (P'(A) > S \text{ and } P_2(A) \leq S) \text{ or } (P'(A) \leq S \text{ and } P_2(A) > S)\}|$. From this, the **rank correlation coefficient** is

$$\mathsf{Correlation}(P', P_2) = \frac{n_a - n_d}{n_a + n_d}$$

Example 5. For $\mathsf{Arg}(\Delta) = \{A_1, A_2, A_3, A_4\}$, and $S = 0.5$, let $P'(A_1) = 1$, $P'(A_2) = 0.3$, $P'(A_3) = 0.7$, $P'(A_4) = 0.4$, $P_2(A_1) = 1$, $P_2(A_2) = 0.8$, $P_2(A_3) = 0.6$, and $P_2(A_4) = 0.2$. So, the coefficient is $(3 - 1)/4 = 1/2$.

Note, the coefficient is the same as the Kendall rank correlation coefficient [7], but the way we calculate n_a and n_d is quite different.

5 Dialogical argumentation

Participants take turns to contribute arguments and/or attacks, thereby constructing an argument graph. For this, we just record the additions to the graph as defined next.

Definition 6. *A* **dialogue state** *is a pair (X, Y) where X is a set of arguments, and Y is a set of attacks. Note, Y is not necessarily a subset of $X \times X$. A* **dialogue***, denoted D, is a sequence of dialogue states $[(X_1, Y_1), ..., (X_n, Y_n)]$.*

We use the function D to denote a dialogue, where for an index $i \in \{1, ..., n\}$, $D(i) = (X_i, Y_i)$ is a dialogue state. For a dialogue $D = [(X_1, Y_1), ..., (X_n, Y_n)]$, $\mathsf{Len}(D) = n$ is the index of the last step, and $\mathsf{Sub}(D, i) = [(X_1, Y_1), ..., (X_i, Y_i)]$ is the first i steps. For each step of the dialogue, there is an argument graph. We define this graph recursively with the base case being the empty graph.

Definition 7. *For dialogue D, s.t. $1 \leq i \leq \mathsf{Len}(D)$, and $D(i) = (X_i, Y_i)$, $\mathsf{Graph}(D, i)$ = $(\mathcal{A}_{i-1} \cup X_i, \mathcal{R}_{i-1} \cup Y_i)$ is the* **dialogue graph** *where if $i = 1$, then $(\mathcal{A}_{i-1}, \mathcal{R}_{i-1})$ is $(\emptyset, \emptyset)$, and if $i > 1$, then $(\mathcal{A}_{i-1}, \mathcal{R}_{i-1})$ is $\mathsf{Graph}(D, i-1)$.*

So for each step of the dialogue, we can construct the current state of the argument graph. So the sequence of states of the dialogue are all used to construct the current state of the graph. Clearly, this is monotonic: Arguments and attacks are added to the graph, and none are subtracted.

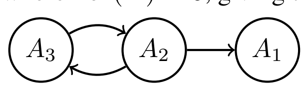
Example 6. Consider the following probability distribution over models for each participant, where $T = 0.5$ and $S = 0.3$.

Model	a	b	Participant 1	Participant 2
m_1	true	true	0.8	0.0
m_2	true	false	0.1	0.5
m_3	false	true	0.1	0.0
m_4	false	false	0.0	0.5

Hence, we get the following probabilities for A_1 to A_3.

Argument	Participant 1	Participant 2
$A_1 = \langle \{b, b \rightarrow a\}, a \rangle$	0.8	0.0
$A_2 = \langle \{\neg b\}, \neg b \rangle$	0.1	1.0
$A_3 = \langle \{b\}, b \rangle$	0.9	0.0

Now consider dialogue $D = [\ (\{A_1\}, \{\}), (\{A_2\}, \{(A_2, A_1)\}), (\{A_3\}, \{(A_3, A_2)\}), (\{\}, \{(A_2, A_3)\}), (\{\}, \{\})\ ]$, where $\mathsf{Len}(D) = 5$, giving the dialogue graph below.



A dialogue can be infinite since for example the contribution $(\emptyset, \emptyset)$ can be repeatedly added. So to draw the dialogue to a close, we restrict consideration to complete dialogues.

Definition 8. *A dialogue D is* **complete***, where $\mathsf{Len}(D) = n$, and the persuasion claim is ϕ, iff*

1. *$\forall i, j \in \{1, \ldots, n\}$, if $i \neq j$ & $D(i) = D(j)$, then $D(i) = (\emptyset, \emptyset)$ & $D(j) = (\emptyset, \emptyset)$*
2. *$\forall i \in \{1, \ldots, n-2\}$, if $D(i) = (\emptyset, \emptyset)$, then $D(i+1) \neq (\emptyset, \emptyset)$*
3. *if n is even, then $\exists A \in \mathsf{Grounded}(\mathsf{Graph}(D, n))$ s.t. $\mathsf{Claim}(A) = \phi$.*
4. *if n is odd, then $\nexists A \in \mathsf{Grounded}(\mathsf{Graph}(D, n))$ s.t. $\mathsf{Claim}(A) = \phi$.*
5. *$D(n) = (\emptyset, \emptyset)$*

We explain the above conditions as follows: Condition 1 ensures that the only state that can be repeated is the empty state; Condition 2 ensures that only the last two step of the dialogue can have the empty state followed immediately by the empty state; Conditions 3 and 4 ensure that if the last step is an even step then there is an argument with the claim in the grounded extension, whereas if the last step is an odd step then there is not an argument with the claim in the grounded extension; And condition 5 ensures that the last step is the empty state.

In the rest of the paper, for each step i, if i is odd (respectively even) participant 1 (respectively participant 2) will add $D(i)$. So intuitively, at the last step n, if n is odd (respectively even), participant 1 (respectively participant 2) has conceded the dialogue (perhaps because it has nothing more to add).

Proposition 2. *Let D be a complete dialogue and let $\mathcal{A}$ be a finite set of arguments. If for each i, $X_i \subseteq \mathcal{A}$, and $Y_i \subseteq \mathcal{A} \times \mathcal{A}$, and $D(i) = (X_i, Y_i)$, then D is finite (i.e. $\mathsf{Len}(D) \in \mathbb{N}$).*

Taking a simple view of persuasion, a participant is persuaded of a claim if the dialogue graph constructed is such that there is an argument for the claim in the grounded extension of the graph. We justify this in the next section.

Definition 9. *For a complete dialogue D, where $\mathsf{Len}(D) = n$, the* **outcome** *of the dialogue is specified as follows: If n is even, then participant 1* **wins***, whereas if n is odd, then participant 1* **looses***.*

So if n is even, then participant 1 is successful in persuading participant 2, otherwise participant 1 is unsuccessful. The dialogue D in Example 6 is a complete dialogue, and hence participant 1 looses. In the following sections, we present and justify three strategies for constructing complete dialogues.

6 Simple dialogues

In a simple dialogue, participant 1 can add arguments for the persuasion claim that are not in the current dialogue graph.

Definition 10. *For position Π_1, and dialogue D, a* **posit contribution** *by participant 1 is $(\{A\}, \{\})$ where $A \in \mathsf{GoodArg}(\Delta_1, P_1, T)$ and $A \notin \mathsf{Nodes}(\mathsf{Graph}(D, i))$ and $\mathsf{Claim}(A) = \phi$. The set of posit contributions by participant 1 for D at step i is $\mathsf{Posit}(\Pi_1, D, i)$.*

Both participants can add counterarguments. For this, the $\mathsf{NewAttackers}$ function identifies the good arguments that participant x has that are not in the current dialogue graph $G_i = \mathsf{Graph}(D, i)$ but that attack an argument in G_i.

$$\mathsf{NewAttackers}(\Pi_x, D, i) = \{A \in \mathsf{GoodArg}(\Delta_x, P_x, T) \mid A \notin \mathsf{Nodes}(G_i) \text{ and } \exists B \in \mathsf{Grounded}(G_i) \text{ s.t. } P_x(B) \leq S \text{ and } A \text{ is a direct undercut of } B\}$$

Definition 11. *For position $\varPi_x$, and a dialogue D, a* **counter contribution** *by participant x is (X_{i+1}, Y_{i+1}) s.t.*

$$\begin{array}{l} \textit{if there is an } A \in \mathsf{NewAttackers}(\varPi_x, D, i), \\ \textit{then } X_{i+1} = \{A\} \textit{ and } Y_{i+1} = \mathsf{NewArcs}(X_{i+1} \cup \mathsf{Nodes}(G_i), S) \\ \textit{else } X_{i+1} = \{\} \textit{ and } Y_{i+1} = \mathsf{NewArcs}(\mathsf{Nodes}(G_i), S) \end{array}$$

where $G_i = \mathsf{Graph}(D, i)$ and $\mathsf{NewArcs}(Z, S) = \{\ (A, B) \mid A, B \in Z$ and $P_x(A) \leq S$ and A is a direct undercut of $B\}$. The set of counter contributions by participant x for D at step i is $\mathsf{Counter}(\varPi_x, D, i)$.

So a counter contribution is zero or one argument and zero or more arcs, as illustrated in Example 6. Let $\mathsf{Part1}(i) = \mathsf{Simple}(\varPi_1, D, i) \cup \mathsf{Counter}(\varPi_1, D, i)$ (respectively $\mathsf{Part2}(i) = \mathsf{Counter}(\varPi_2, D, i)$) be the contributions for participant 1 (respectively participant 2) at step i. The next definition ensures that the participants take turns in the contributions.

Definition 12. *For positions $\varPi_1$ and $\varPi_2$, a dialogue D is* **turn taking** *iff for each $i \in \{1, \ldots, \mathsf{Len}(D)\}$, if i is odd, then $D(i) \in \mathsf{Part1}(i)$ and if i is even, then $D(i) \in \mathsf{Part2}(i)$.*

The next definition ensures that each agent gives a contribution other than $(\emptyset, \emptyset)$ if possible (i.e. there is a non-empty contribution) and needed (i.e. for participant 1, there is not an argument for the persuasion claim in the grounded extension of the current dialogue graph, and for participant 2, there is a argument for the persuasion claim in the grounded extension of the current dialogue graph). Note, $(\emptyset, \emptyset)$ is always available as a counter contribution.

Definition 13. *For positions $\varPi_1$ and $\varPi_2$, a complete dialogue D is* **exhaustive** *iff for each $i \in \{1, \ldots, \mathsf{Len}(D)\}$, where $G_i = \mathsf{Graph}(D, i)$, the following conditions hold.*

1. *If i is odd, and $\exists A \in \mathsf{Grounded}(G_i)$ s.t. $\mathsf{Claim}(A) = \phi$, then $D(i) = (\emptyset, \emptyset)$.*
2. *If i is odd, and $\nexists A \in \mathsf{Grounded}(G_i)$ s.t. $\mathsf{Claim}(A) = \phi$, and $|\mathsf{Part1}(i)| > 1$, then $D(i) \neq (\emptyset, \emptyset)$.*
3. *If i is even, and $\nexists A \in \mathsf{Grounded}(G_i)$ s.t. $\mathsf{Claim}(A) = \phi$, then $D(i) = (\emptyset, \emptyset)$.*
4. *If i is even, and $\exists A \in \mathsf{Grounded}(G_i)$ s.t. $\mathsf{Claim}(A) = \phi$, and $|\mathsf{Part2}(i)| > 1$, then $D(i) \neq (\emptyset, \emptyset)$.*

A **simple dialogue** is a dialogue that is turning taking and exhaustive. These definitions specify how the dialogue is constructed, and if the dialogue is complete it will terminate. The definitions ensure both agents only add good arguments and good attacks. Let $\mathsf{SD}(\varPi_1, \varPi_2)$ be the set of simple dialogues.

Example 7. For A_1, A_3 and A_5 from participant 1 and A_2 and A_4 from participant 2, D_1 is a simple dialogue for which Participant 1 wins.

$$\begin{array}{ll} A_1 = \langle \{b, b \rightarrow a\}, a \rangle & D_1(1) = (\{A_1\}, \{\}) \\ A_2 = \langle \{c, c \rightarrow \neg b\}, \neg b \rangle & D_1(2) = (\{A_2\}, \{(A_2, A_1)\}) \\ A_3 = \langle \{d, d \rightarrow \neg c\}, \neg c \rangle & D_1(3) = (\{A_3\}, \{(A_3, A_2)\}) \\ A_4 = \langle \{\neg d\}, \neg d \rangle & D_1(4) = (\{A_4\}, \{(A_4, A_3)\}) \\ A_5 = \langle \{e, e \rightarrow \neg c\}, \neg c \rangle & D_1(5) = (\{A_5\}, \{(A_5, A_2)\}) \\ & D_1(6) = (\{\}, \{\}) \end{array}$$

Example 8. Participant 1 has A_1 and A_3 and participant 2 has A_2. D = [$(\{A_1\}, \{\})$, $(\{A_2\}, \{(A_2, A_1)\})$, $(\{A_3\}, \{(A_3, A_2)\})$, $(\{\}, \{(A_1, A_3)\})$, $(\{\}, \{\})$] is a simple dialogue that participant 1 looses.

Example 9. Participant 1 has A_1, A_3, A_5 and A_6, and participant 2 has A_2 and A_4. D = [$(\{A_1\}, \{\})$, $(\{A_2\}, \{(A_2, A_1)\})$, $(\{A_3\}, \{(A_3, A_2)\})$, $(\{A_4\}, \{(A_4, A_3), (A_4, A_1)\})$, $(\{A_5\}, \{(A_5, A_2)\})$ $(\{\}, \{\})$ $(\{A_6\}, \{(A_6, A_4)\})$ $(\{\}, \{\})$] is a simple dialogue that participant 1 wins.

We use the joint graph (defined next) to show a type of correctness of the simple dialogues in the following result.

Definition 14. *For positions Π_1 and Π_2, the* **joint graph***, is an argument graph $(\mathcal{A}, \mathcal{R})$, denoted* $\mathsf{JointGraph}(\Pi_1, \Pi_2)$*, where* $\mathcal{A} = \mathsf{GoodArg}(\Delta_1, P_1, T) \cup \mathsf{GoodArg}(\Delta_2, P_2, T)$ *and* $\mathcal{R} = \{(A, B) \mid A, B \in \mathcal{A}$ *and* $(P_1(B) \leq S$ *or* $P_2(B) \leq S)$ *and* A *is a direct undercut of* B $\}$.

Proposition 3. *For positions Π_1 and Π_2, let G^* be* $\mathsf{JointGraph}(\Pi_1, \Pi_2)$*. For each* $D \in \mathsf{SD}(\Pi_1, \Pi_2)$*, participant 1 wins D iff there is an* $A \in \mathsf{Grounded}(G^*)$ *such that* $\mathsf{Claim}(A) = \phi$*.*

So a simple dialogue just involves each participant making contributions until one or other participant concedes. Both agents are selective in the sense that they only present good arguments and good attacks. But for participant 1, there is no consideration of what might be more likely to be persuasive (such as presenting arguments that are less likely to be attacked by participant 2). We address this next.

7 Bestfirst dialogues

The bestfirst dialogue involves participant 1 selecting its best arguments for positing first in the dialogue. Its best arguments, the bestfirst contributions, are its good arguments that it believes are not good targets for participant 2.

Definition 15. *For position Π_1, and dialogue D, the set of* **bestfirst contributions** *is* $\mathsf{Bestfirst}(\Pi_1, D, i) = \{(\{A\}, Y) \in \mathsf{Simple}(\Pi_1, D, i) \cup \mathsf{Counter}(\Pi_1, D, i) \mid P'(A) > S\}$.

Definition 16. *For Π_1, and Π_2, a simple dialogue D is* **bestfirst** *iff for each* $i \in \{1,\ldots,\mathsf{Len}(n)\}$*, if i is odd, and* $\mathsf{Bestfirst}(\Pi_1, D, i) \neq \emptyset$*, then* $D(i) \in \mathsf{Bestfirst}(\Pi_1, D, i)$*. Let* $\mathsf{BD}(\Pi_1, \Pi_2)$ *be the set of bestfirst dialogues.*

Example 10. Let D_1 = [$(\{A_1\}, \{\})$, $(\{\}, \{\})$] and D_2 = [$(\{A_2\}, \{\})$, $(\{A_3\}, \{(A_3, A_2)\})$, $(\{A_1\}, \{\})$, $(\{\}, \{\})$]. Also let $\mathsf{Correlation}(P', P_2) = 1$. If $P'(A_1) > S$ and $P'(A_2) \leq S$, then D_1 is bestfirst, and if $P'(A_1) \leq S$ and $P'(A_2) > S$, then D_2 is bestfirst. In both cases, participant 1 wins.

If the correlation is positive for P' and P_2, then the next result shows that on average the bestfirst dialogues are shorter than the simple dialogues.

Proposition 4. *For the majority of positions Π_1 and Π_2, s.t.* $\mathsf{Correlation}(P', P_2) > 0$*, then*

$$\left(\frac{\sum_{D\in\mathsf{BD}(\Pi_1,\Pi_2)}\mathsf{Len}(D)}{\mid \mathsf{BD}(\Pi_1,\Pi_2)\mid}\right) \leq \left(\frac{\sum_{D\in\mathsf{SD}(\Pi_1,\Pi_2)}\mathsf{Len}(D)}{\mid \mathsf{SD}(\Pi_1,\Pi_2)\mid}\right)$$

So the bestfirst dialogue captures a more efficient form of persuasion than the simple dialogue. Participant 1 presents its better arguments first, and if it does not succeed, then it will use its remaining arguments.

8 Insincere dialogues

The insincere dialogue is characterised by the proponent selecting its arguments based on what it believes the other participant believes (and therefore selecting the arguments that are less likely to be attacked by the other participant). Note, we do not assume that participant 1 actually believes these arguments. It is being manipulative by presenting arguments that it believes that the other participant will accept.

Definition 17. *For position Π_1, and a dialogue D, the set of insincere contributions by participant 1 for D is the following where $\Pi^{insincere} = (\Delta_1, P', P', S, T, \phi)$.*

$$\mathsf{Insincere}(\Pi_1, D, i) = \mathsf{Posit}(\Pi^{insincere}, D, i) \cup \mathsf{Counter}(\Pi^{insincere}, D, i)\}$$

Definition 18. *For positions Π_1 and Π_2, a simple dialogue D is* **insincere** *iff for each $i \in \{1, \ldots, \mathsf{Len}(n)\}$, if i is odd, then $D(i) \in \mathsf{Insincere}(\Pi_1, D, i)$, and if i is even, then $D(i) \in \mathsf{Counter}(\Pi_2, D, i)$. Let $\mathsf{ID}(\Pi_1, \Pi_2)$ be the set of insincere dialogues.*

So $D \in \mathsf{ID}(\Pi_1, \Pi_2)$ iff $D \in \mathsf{SD}(\Pi^{insincere}, \Pi_2)$. The advantage for participant 1 is that it is not restricted by its own probability distribution in making its contributions. Rather, the aim for participant 1 is to present any arguments it can with the sole aim of winning the dialogue. Though one would assume that participant 1 would have a high belief in the persuasion claim ϕ (i.e. $P_1(\phi)$ is high) for it to want to resort to an insincere dialogue.

Example 11. Let $m_1(a) = true$, $m_1(b) = true$, $m_1(c) = false$, $m_2(a) = true$, $m_2(b) = false$, and $m_2(c) = true$. For positions Π_1 and Π_2, where $\Delta_1 = \{b, b \rightarrow a, c, c \rightarrow a\}$, $P_1(m_1) = 1$, $\Delta_2 = \{\neg b\}$, and $P_2(m_2) = 1$, let ϕ be a. Also, suppose $P' = P_2$. So $A_1 = \langle\{b, b \rightarrow a\}, a\rangle$ is a good argument for participant 1, but a good target for participant 2. In a simple dialogue, participant 1 only has one argument for a, and it would loose the dialogue (because participant 2 would attack with $A_2 = \langle\{\neg b\}, \neg b\rangle$). In contrast, $A_3 = \langle\{c, c \rightarrow a\}, a\rangle$ is not a good argument for participant 1, but for participant 2, it is a good argument and not a good target. So, A_3 is an insincere contribution for participant 1, and it would win the insincere dialogue using it.

In the next example, we let $\Delta_1 = \mathcal{L}$. Since participant 1 is prepared to say anything that participant 2 believes, this just means that it is prepared to present any argument A available in the language $\mathcal{L}$ as long as the recipient believes it.

Example 12. Let $\Delta_1 = \mathcal{L}$ where $A_1, A_3 \in \mathsf{Arg}(\Delta_1)$ and $A_2 \in \mathsf{Arg}(\Delta_2)$, and assume the following regarding the probability distributions.

$$P_1(A_1) > T; P'(A_1) > T; P'(A_1) > S; P_2(A_1) < S$$
$$P_1(A_2) < T; P'(A_2) < T; P'(A_2) < S; P_2(A_2) < S$$
$$P_1(A_3) < T; P'(A_3) > T; P'(A_3) > S; P_2(A_3) > S$$

So P' only differs from P_2 on A_1. Hence, D = [$(\{A_1\}, \{\})$, $(\{A_2\}, \{(A_2, A_1)\})$, $(\{A_3\}, \{(A_3, A_2)\})$, $(\{\}, \{\})$] is an insincere dialogue that participant 1 wins.

The following definition of openness of a position just means that there is at least one atom in the language for which there are no strong arguments for or against it. In effect, it means that participant 2 has not got a position so constrained that it is impossible to persuade it.

Definition 19. *A position Π_2 is* **open** *iff there is an atom $\psi \in \mathcal{L}$, s.t. for all $A \in$ $\mathsf{GoodArgs}(\Delta_2, P_2, T)$, $\mathsf{Claim}(A) \neq \psi$ and $\mathsf{Claim}(A) \neq \neg\psi$.*

Proposition 5. *Let Π_1 and Π_2 be positions s.t. $\Delta_1 = \mathcal{L}$, the persuasion claim is ϕ, and $\mathsf{Correlation}(P', P_2) = 1$. For any $D \in \mathsf{ID}(\Pi_1, \Pi_2)$ if either ($P_2(\neg\phi) \leq S$ and $S < 0.5$) or Π_2 is open, then participant 1 wins D.*

The above result is a situation where the participant 1 has a very good model of participant 2. We can generalise the result to imperfect models of the opponent so that with high probably that participant 1 wins.

The idea of an insincere strategy is important; If a protocol for a argumentative dialogue allows for this strategy, then the above result shows that a participant can dominate in a quite negative way. It can manipulate the opponent, and the opponent may be oblivious to this manipulation. We are not proposing that we want to build agents who use the insincere strategy. But, we may want to build agents who are aware that there are other agents who do use the insincere strategy and protect against it. So we need to formalise and investigate the insincere strategy and developments of it.

9 Discussion

In this paper, we have introduced good arguments, good targets, and good attacks. We therefore provide a new approach to constructing argument graphs, drawing on probability theory, that allows us to drop arguments if there is too much doubt in them, and to drop attacks if there is insufficient doubt in them. There are other proposals that drop attacks (e.g. preference-based argumentation frameworks [8], value-based argumentation frameworks [9], and weighted argumentation frameworks [10]), but they do not drop arguments other than by attacking them, and they are not based on a quantitative theory of uncertainty. There are proposals for using probability theory in argumentation (e.g. [11–15, 4]) but they do not drop arguments or attacks, and there is a possibility theory approach [16]) but it is not based on argument graphs.

Our approach has been influenced by Amgoud et al [17] (a detailed protocol for exchanging logical arguments using preference-based argumentation). We go beyond that by providing a way to select arguments and counterarguments to be used, and for strategies that use selectivity. We can allow for instance for an agent to present the arguments it has greatest belief in and it thinks the other agent has high belief in. We also allow for tolerance of arguments by an opponent. For instance an opponent may choose to not attack an argument if it thinks the argument is not too bad.

There are a number of papers that formalize aspects of persuasion. Most approaches are aimed at providing protocols for dialogues (for a review see [18]). Forms of correctness (e.g. the dialogue system has the same extensions under particular dialectical semantics as using the agent's knowledgebases) have been shown for a variety of systems (e.g. [19–22]). However, strategies for persuasion, in particular taking into account beliefs of the opponent are under-developed. Using selection of arguments, based on probability distributions for the agents, and for modelling one agent by another, we can formalise interesting strategies. To illustrate the potential, we consider the bestfirst strategy with a clear proven advantage, and the more complex insincere strategy.

Strategies in argumentation have been analysed using game theory [23, 24]. This mechanism design approach assumes that all the agents reveal their arguments at the same time, and the resulting argument graph is evaluated using grounded semantics. This is a one step process that does not involve logical arguments, dialogues or opponent modelling. Mechanism design has also been used for comparing strategies for logic-based dialogical argumentation that may involve lying [25]. This complements our work since they do not consider the uncertainty of beliefs or modelling the opponent.

Finally, audience modelling has been considered in value-based argumentation frameworks [9, 26] and in deductive argumentation [27, 28]. However, they have not been harnessed in strategies in dialogical argumentation, and only [26] considers uncertainty in the form of a probability assignment that an argument will promote a particular "value" with an agent, which is a different kind of uncertainty to that considered here.

In conclusion, we provide a novel framework for modelling uncertainty in argumentation, and use this to give three examples of strategy for dialogical argumentation. In future work, we will develop further strategies, and investigate learning the probability distributions from previous interactions.

References

1. Amgoud, L., Prade, H.: Using arguments for making and explaining decisions. Artificial Intelligence **173**(3-4) (2009) 413–436
2. Dung, P.: On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming, and n-person games. Artificial Intelligence **77** (1995) 321–357
3. Paris, J.: The Uncertain Reasoner's Companion: A Methematical Perspective. Cambridge University Press (1994)
4. Hunter, A.: A probabilistic approach to modelling uncertain logical arguments. International Journal of Approximate Reasoning **54**(1) (2013) 47–81
5. Elvang-Gransson, M., Krause, P., Fox, J.: Acceptability of arguments as logical uncertainty. In: Symbolic and Quantitative Approaches to Reasoning and Uncertainty (ECSQARU'93). Volume 747 of LNCS., Springer (1993) 85–90
6. Cayrol, C.: On the relation between argumentation and non-monotonic coherence-based entailment. In: Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence (IJCAI'95). (1995) 1443–1448

7. Kendall, M.: A new measure of rank correlation. Biometrika **30**(1-2) (1938) 81–93
8. Amgoud, L., Cayrol, C.: A reasoning model based on the production of acceptable arguments. Annals of Mathematics and Artificial Intelligence **34** (2002) 197–216
9. Bench-Capon, T.: Persuasion in practical argument using value based argumentationframeworks. Journal of Logic and Computation **13**(3) (2003) 429–448
10. Dunne, P.E., Hunter, A., McBurney, P., Parsons, S., Wooldridge, M.: Weighted argument systems: Basic definitions, algorithms, and complexity results. Artificial Intelligence **175**(2) (2011) 457–486
11. Haenni, R.: Cost-bounded argumentation. International Journal of Approximate Reasoning **26**(2) (2001) 101–127
12. Dung, P., Thang, P.: Towards (probabilistic) argumentation for jury-based dispute resolution. In: Computational Models of Argument (COMMA'10), IOS Press (2010) 171–182
13. Li, H., Oren, N., Norman, T.: Probabilistic argumentation frameworks. In: Proceedings of the First International Workshop on the Theory and Applications of Formal Argumentation (TAFA'11). (2011) 1–16
14. Thimm, M.: A probabilistic semantics for abstract argumentation. In: Proceedings of the European Conference on Artificial Intelligence (ECAI'12). (2012) 750–755
15. Hunter, A.: Some foundations for probabilistic argumentation. In: Proceedings of the International Comference on Computational Models of Argument (COMMA'12). (2012) 117–128
16. Alsinet, T., Chesñevar, C., Godo, L., Simari, G.: A logic programming framework for possibilistic argumentation: Formalization and logical properties. Fuzzy Sets and Systems **159**(10) (2008) 1208–1228
17. Amgoud, L., Maudet, N., Parsons, S.: Arguments, dialogue and negotiation. In: Fourteenth European Conference on Artifcial Intelligence (ECAI 2000), IOS Press (2000) 338–342
18. Prakken, H.: Formal sytems for persuasion dialogue. Knowledge Engineering Review **21**(2) (2006) 163–188
19. Prakken, H.: Coherence and flexibility in dialogue games for argumentation. Journal of Logic and Computation **15**(6) (2005) 1009–1040
20. Black, E., Hunter, A.: An inquiry dialogue system. Autonomous Agents and Multi-Agent Systems **19**(2) (2009) 173–209
21. Fan, X., Toni, F.: Assumption-based argumentation dialogues. In: Proceedings of International Joint Conference on Artificial Intelligence (IJCAI'11). (2011) 198–203
22. Caminada, M., Podlaszewski, M.: Grounded semantics as persuasion dialogue. In: Computational Models of Argument (COMMA'12). (2012) 478–485
23. Rahwan, I., Larson, K.: Mechanism design for abstract argumentation. In: Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'08, IFAAMAS (2008) 1031–1038
24. Rahwan, I., Larson, K., Tohmé, F.: A characterisation of strategy-proofness for grounded argumentation semantics. In: Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI'09). (2009) 251–256
25. Fan, X., Toni, F.: Mechanism design for argumentation-based persuasion. In: Computational Models of Argument (COMMA'12). (2012) 322–333
26. Oren, N., Atkinson, K., Li, H.: Group persuasion through uncertain audience modelling. In: Proceedings of the International Comference on Computational Models of Argument (COMMA'12). (2012) 350–357
27. Hunter, A.: Towards higher impact argumentation. In: Proceedings of the 19th National Conference on Artificial Intelligence (AAAI 2004), MIT Press (2004) 275–280
28. Hunter, A.: Making argumentation more believable. In: Proceedings of the 19th National Conference on Artificial Intelligence (AAAI 2004), MIT Press (2004) 269–274