What Should an Agent Know Not to Fail in Persuasion?

Shizuka Yokohama and Kazuko Takahashi(✉)

School of Science and Technology, Kwansei Gakuin University,
2-1, Gakuen, Sanda 669-1337, Japan
{yokohama-shizuka,ktaka}@kwansei.ac.jp

Abstract. This paper presents a strategy and conditions for non-failing persuasion using a dialogue model using argumentation. A concept of the predicted knowledge of the other agent participating in the dialogue is introduced. In the dialogue model, an agent's knowledge is updated as the dialogue proceeds; an argumentation framework is constructed from the current knowledge; and only the content of an acceptable argument can be offered as the next move. In this paper, a modified dialogue model is proposed in which the next move is determined using predicted knowledge and a strategy that navigates a non-failing persuasive argumentation is presented. Conditions under which persuasion never fails using this strategy when the prediction is equivalent to the actual knowledge of an opponent are described. Moreover, what the predicted knowledge should contain for non-failing persuasion are discussed. The introduction of predicted knowledge improves the formulation of real dialogue.

Keywords: Argumentation · Persuasion · Dialogue · Predicted knowledge base

1 Introduction

To achieve agreement during a dialogue between agents, it is important to resolve existing conflicts by exchanging protocols; persuasion is one dialogue type that has such characteristics. Each agent participating in a dialogue has their own knowledge, which changes as the dialogue proceeds. If dialogue is regarded as a game, then each agent is a player who determines their next move by considering the effect of the move based on a dialogue protocol. The agent's knowledge is updated with the utterance of an opponent, which may add knowledge that is inconsistent with their current belief. As an argumentation framework can handle inconsistency or nonmonotonicity of knowledge bases, it is useful for creating a dialogue model.

Amgoud et al. proposed a dialogue model using argumentation [2]. In their model, an agent's knowledge is updated as the dialogue proceeds; an argumentation framework is constructed from the current knowledge, and only the content of an acceptable argument can be asserted as the agent's beliefs. This approach

M. Rovatsos et al. (Eds.): EUMAS 2015/AT 2015, LNAI 9571, pp. 219–233, 2016.
DOI: 10.1007/978-3-319-33509-4_18

models argumentative agents who behave rationally; however, it lacks the viewpoint of predicting the opponent's inner states. On the other hand, in an actual dialogue, especially in the case of persuasion, we usually predict the opponent's knowledge or beliefs and create a strategy to succeed in persuasion.

Consider the following situation of students selecting their research laboratory. Alice and Bob want to apply to the same laboratory. Alice, who prefers a strict professor's laboratory, wants to apply to Charlie's laboratory. She knows that Charlie is generous as well as strict. On the other hand, Bob wants to apply to a generous professor's laboratory, but does not want to apply to a strict professor's laboratory. Bob does not know about the reputation of Charlie. In this example, if Alice has no idea about Bob's knowledge, then she may first say, "Let's apply to Charlie's laboratory because he is strict," which will fail to persuade Bob to accept Alice's proposal. However, if she knows that Bob does not like strict professors, then she could say, "Let's apply to Charlie's laboratory because he is generous," which will successfully persuade Bob to accept the proposal. This choice of utterance is based on the key knowledge that Bob does not want to apply to a strict professor's laboratory and on Alice having the correct key knowledge as her prediction.

In this paper, we revisit the dialogue model proposed by Amgoud et al. and enhance it to lead to non-failing persuasion by creating a strategy based on predicted knowledge. We propose a dialogue model in which each agent has predicted knowledge of their opponent as well as their own knowledge. In this strategy, an agent does not present an argument that s/he predicts will lead the opponent to refuse the proposal, and positively presents an argument that s/he predicts will lead the opponent to accept it. These decisions are made using an argumentation framework constructed from predicted knowledge.

We investigate the conditions under which persuasion succeeds, or at least does not fail using this strategy, when a prediction is equivalent to the actual knowledge of an opponent. Moreover, we discuss what the predicted knowledge should contain for persuasion not to fail.

This dialogue model using predicted knowledge, improves the formulation of real dialogue and can be extended to handle dialogues including a lie.

The rest of the paper is organized as follows. Section 2 describes the argumentation framework on which our model is based. Section 3 formalizes our dialogue model and proposes a persuasion strategy. Section 4 gives an example of a persuasive dialogue. Section 5, discusses the properties of this strategy. Section 6 compares our approach with other approaches. Finally, Sect. 7 presents our conclusions.

2 Argumentation Framework

Dung's abstract argumentation framework is defined as a pair of a set and a binary relation on the set [6]. We instantiate each argument by a set of formulas generated from a given knowledge base. In addition, *preference* is introduced to give relative strength to arguments.

Definition 1 (Argument). *Let Σ be a set of propositional formulas, called* knowledge base. *Σ may be inconsistent and not deductively closed.* An argument *on Σ is defined as a pair of support H and a conclusion h, (H, h), where either of the following conditions are satisfied: (i) $H = \emptyset$ and $h \in \Sigma$, or (ii) H is a consistent minimal subset of Σ in the sense of set inclusion, $H \vdash h$, and $\forall h' \in H$; $h' \not\equiv h$ where $\equiv$ represents logical equivalence.*

For an argument $A = (H, h)$, $supp(A)$ and $concl(A)$ denote H and h, respectively. $fml(A)$ denotes a set of formulas in A, that is, $fml(A) = H \cup \{h\}$. For a set of arguments Arg, $Fml(Arg)$ denotes $\bigcup_{A \in Arg} fml(A)$.

In an argumentation framework for a persuasive dialogue, it is often necessary to give relative strength to arguments to determine which formula is acceptable [1,4,8]. Similar to existing approaches, we define an argumentation framework with preferences.

The strength of each formula is assigned in advance, such that a higher level is more strong than a lower one. As a result, Σ is partially ordered with respect to strength. The preference of an argument is calculated depending on this strength, such that it depends on the least strong formula included in support of an argument. We do not discuss how to assign strength here, since it is out of the focus of this paper.

Definition 2 (Preference). *Let Σ be a set of formulas and str be a function that returns a natural number for an element of Σ. For each argument A, generated from Σ, $Pref(A)$ is defined as $min_{F \in supp(A)} str(F)$ if $supp(A) \neq \emptyset$, and $str(concl(A))$ if $supp(A) = \emptyset$.*

Let A_1 and A_2 be arguments. If $Pref(A_1) \leq Pref(A_2)$, it is said that A_2 *is preferable to* A_1.

Definition 3 (Attack). *For a pair of arguments $A_1 = (H_1, h_1)$ and $A_2 = (H_2, h_2)$, if $h_2 \equiv \neg h_1$, then it is said that A_2* rebuts *A_1 ; if there exists $h \in H_1$ such that $h_2 \equiv \neg h$, then it is said that A_2* undercuts *A_1; A_2 either rebuts or undercuts A_1 and A_2 is preferable to A_1, then it is said that A_2* attacks *A_1.*

Definition 4 (Argumentation Framework). An argumentation framework for a knowledge base Σ under strength str, *denoted by $AF(\Sigma, str)$, is defined as a pair $\langle AR, AT \rangle$ where AR is the set of arguments generated from Σ and AT is the set of attacks on AR based on str. If str is fixed throughout the discussion, then we denote $AF(\Sigma)$ in the form where str is omitted.*

Definition 5 (Acceptable). *Let $\langle AR, AT \rangle$ be an argumentation framework. For a set of arguments $S \subseteq AR$ and an argument A_1, for any argument $A_2 \in AR$ that attacks A_1, there exists an argument $A_3 \in S$ that attacks A_2; it is said that* A_1 is acceptable with respect to S.

Definition 6 (Grounded Extension). *Let $\mathcal{AF} = \langle AR, AT \rangle$ be an argumentation framework. For a set of arguments $S \subseteq AR$, let F be a function:*

$F(S) = \{\ A \in AR \mid A$ is acceptable with respect to $S\ \}$. Let S' be the least fixedpoint of F. Then S' is said to be the grounded extension of $\mathcal{AF}$, *and denoted by $Ext(\mathcal{AF})$.*

Note that there exists a unique grounded extension for any argumentation framework [6]. Hereafter, we use the term "extension" to mean a grounded extension, unless there is any confusion.

In addition to these well-known concepts, a few more new concepts are defined.

Definition 7 (Belief). *Let $\mathcal{AF}$ be an argumentation framework. A set of formulas appearing in arguments in the extension is said to be* a belief of $\mathcal{AF}$*, that is,* $Bel(\mathcal{AF}) = \bigcup_{A \in Ext(\mathcal{AF})} fml(A)$.

Definition 8 (NBA-Argument). *Let* $\mathcal{AF} = \langle AR, AT \rangle$ *be an argumentation framework. For an argument* $A_1 \in AR$*, if there does not exist an argument* $A_2 \in AR$ *that attacks* A_1*, then* A_1 *is said to be* not-being-attacked-argument *of* $\mathcal{AF}$*,* NBA-argument *in short.*

3 Dialogue Model

3.1 Dialogue Model Based on an Argumentation

Amgoud et al. proposed a dialogue model based on an argumentation [2]. An agent's knowledge and belief were distinguished by setting them as formulas in a knowledge base, and in an extension of an argumentation framework constructed from the knowledge base and an opponent's utterances, respectively. We modify this model by introducing a predicted knowledge base.

A dialogue is a sequence of utterances by agents along the protocol. Each agent constructs an argumentation framework from an initial knowledge base and the set of formulas provided so far. When an opponent makes an utterance, and new formulas are provided, then the argumentation framework is revised. First, s/he calculates the extension of the argumentation framework, that represents the consistent set of formulas that s/he currently believes. These are the formulas allowed for use as the next utterance. Next, s/he selects the best move from these allowed moves using a predicted knowledge base of an opponent.

Let X be a participant of a dialogue. Let Σ_X be X's initial knowledge base, Σ_Y be her opponent Y's initial knowledge base, and Π_Y be Y's initial knowledge base on X's prediction. That is, X has two knowledge bases Σ_X and Π_Y. It is usually assumed that common sense or widely prevalent facts on the subject are also known by the opponent. On the other hand, there is knowledge that only the opponent knows, or that the agent is not sure that the opponent knows. Therefore, we assume that the predicted knowledge base is a subset of the opponent's real knowledge base, that is, $\Pi_Y \subseteq \Sigma_Y$.

We consider acts of an agent.

Definition 9 (Act). An act *is either* $assert(p)$*,* $assert(S, p)$*,* $assertS(S, p)$*,* $challenge(p)$ *or* $pass$*, where* p *is a formula and* S *is a set of formulas.*

An act *assert* is asserting the statement with or without its ground, and an act *assertS* is asserting the ground itself. An act *challenge* is asking the reason for the assertion. An act *pass* is passing on the turn, without giving any information.

Let T be an act. We define the function $formula$ that returns a set of formulas for an act.

$$formula(T) = \begin{cases} \{p\} & \text{if} \quad T = assert(p) \\ \{p\} \cup S & \text{if} \quad T = assert(S,p) \\ S & \text{if} \quad T = assertS(S,p) \\ \emptyset & \text{otherwise.} \end{cases}$$

Definition 10 (Move). A move *is a pair of* (X, T)*, where* X *is an agent, and* T *is an act.*

Definition 11 (Dialogue). *When* $\Sigma_P, \Sigma_C, \Pi_P$ *and* Π_C *are given,* a dialogue d_k between a persuader P and their opponent C on a subject $\rho \in \Sigma_P$ *is a finite sequence of moves* $[m_0, \ldots, m_{k-1}]$ *where each* m_i *(*$0 \leq i \leq k-1$*) is in the form of* (X_i, T_i) *and the following conditions are satisfied:*

(i) $X_0 = P$ *and* T_0 *is either* $assert(\rho)$ *or* $assert(S, \rho)$*.*
(ii) For each i *(*$0 \leq i \leq k-1$*),* $X_i = P$ *if* i *is even,* $X_i = C$ *if* i *is odd.*
(iii) For each i *(*$0 \leq i \leq k-1$*),* m_i *is one of allowed moves.*

An allowed move is a move that obeys a dialogue protocol which is defined later.

Definition 12 (Complete Dialogue). *For a dialogue* $[m_0, \ldots, m_{k-1}]$ *between a persuader* P *and its opponent* C *on a subject* ρ*, if* $m_{k-2} = (X, pass)$ *and* $m_{k-1} = (Y, pass)$*, then it is said to be* a complete dialogue.

As a dialogue proceeds, formulas in each agent's knowledge base are disclosed. An agent's commitment store is a set of formulas which s/he has provided so far.

Definition 13 (Commitment Store). *For a dialogue* $d_k = [m_0, \ldots, m_{k-1}]$ *where each* m_i *(*$i = 0, \ldots, k-1$*) is in the form of* (X_i, T_i)*,* X*'s* commitment store for d_k*, which is denoted by* $CS_X^{d_k}$*, is defined as* $\emptyset$ *if* $k = 0$*, and* $\bigcup_{i=0,\ldots,k-1, X_i = X} formula(T_i)$ *if* $k \neq 0$*.*

Definition 14 (Argumentation Framework for a Dialogue). *For a dialogue* $d_k = [m_0, \ldots, m_{k-1}]$, an argumentation framework of agent X for d_k *is defined as* $AF(\Sigma_X \cup CS_Y^{d_k})$*, which is denoted by* $\mathcal{AF}_X^{d_k}$*.* A predicted argumentation framework of agent Y by X for d_k *is defined as* $AF(\Pi_Y \cup CS_X^{d_k} \cup CS_Y^{d_k})$*, which is denoted by* $\mathcal{PAF}_Y^{d_k}$*.*

A dialogue protocol is a set of rules for each act. For example, $assertS(S, p)$ is allowed if an agent has asserted p but not asserted S as its ground, $challenge(p)$ is allowed if p has been asserted by the opponent but its support has not. An agent is basically allowed to assert a proposition contained in the extension of the current argumentation framework, and not allowed to give a repetitive assertion. *An allowed move* is a move that obeys the rules.

Definition 15 (Allowed Move). *Let X, Y be agents, and $d_k = [m_0, \ldots, m_{k-1}]$ be a dialogue. The preconditions of each act of agent X for d_k are formalized as follows. If a move m_k satisfies the precondition, then m_k is said to be* an allowed move for d_k.

- *assert*(p):
  - *if* $k = 0$ *and* $\exists A \in Ext(\mathcal{AF}_X^{d_k})$; $p = concl(A)$.
  - *if* $k \neq 0$ *and* $\neg p \in CS_Y^{d_k}$ *and* $\exists A \in Ext(\mathcal{AF}_X^{d_k})$; $p = concl(A)$.
- *assert*(S, p):
  - *if* $k = 0$ *and* $\exists A \in Ext(\mathcal{AF}_X^{d_k})$; $p = concl(A), S = supp(A)$.
  - *if* $k \neq 0$ *and* $\neg p \in CS_Y^{d_k}$ *and* $(X, assert(p)) \neq m_i$ $(0 \leq i \leq k-1)$ *and* $\exists A \in Ext(\mathcal{AF}_X^{d_k})$; $p = concl(A), S = supp(A)$.
- *assertS*(S, p): *if* $p \in CS_X^{d_k}, (X, assert(S, p)) \neq m_i$ $(0 \leq i \leq k-1)$ *and* $\exists A \in Ext(\mathcal{AF}_X^{d_k})$; $S = supp(A), p = concl(A)$.
- *challenge*(p): *if* $p \in CS_Y^{d_k}$ *and* $(Y, assert(S, p)), (Y, assertS(S, p)) \neq m_i$ $(0 \leq i \leq k-1)$.
- *pass*: *if* $k \neq 0$.

There are two additional preconditions for m_k:

- *for every act: if not both of the acts of* m_{k-2} *and* m_{k-1} *are pass.*
- *for an act other than pass: if* $m_k \neq m_i$ $(0 \leq i \leq k-1)$.

After the move $m_k = (X, T)$, the following updates are undertaken: d_{k+1} is obtained from d_k by adding (X, T) to its end, $CS_X^{d_{k+1}} = CS_X^{d_k} \cup formula(T)$ and $CS_Y^{d_{k+1}} = CS_Y^{d_k}$.

Definition 16 (Win/Lose). *For a complete dialogue d_k between a persuader P and their opponent C on a subject ρ, the dialogue is said to be* win by P *if* $\rho \in Bel(\mathcal{AF}_C^{d_k})$, strongly win by P *if* $\rho \in Bel(\mathcal{AF}_P^{d_k}) \cap Bel(\mathcal{AF}_C^{d_k})$, *and* lost by P *if* $\neg\rho \in Bel(\mathcal{AF}_C^{d_k})$.

Definition 17 (Dialogue Tree). A dialogue tree *between P and C on ρ is a finite tree of which each node corresponds to a dialogue, and constructed in the following manner.*

1. *The root node corresponds to* ϵ *(an empty sequence).*
2. *For a node N corresponds to dialogue* $d_i = [m_0, \ldots, m_{i-1}]$,
   (a) *if the act of* m_{i-2} *and that of* m_{i-1} *are both pass, N has no child node;*
   (b) *otherwise, its child nodes* $N_1 \ldots, N_l$ *are the nodes corresponding to* $[m_0, \ldots, m_{i-1}, m_{i_j}]$ $(1 \leq j \leq l)$, *respectively, where* $\{m_{i_1} \ldots m_{i_l}\}$ *are the set of all allowed moves at N.*

A dialogue tree is a finite tree of which each leaf is a complete dialogue, and in which the depth of a node corresponding to dialogue d_k is k. It surveys all possible dialogues between P and C on ρ. Therefore, different branches may include the same move whereas a single branch never includes the same move with the exception of the *pass* act.

Definition 18 (Failure Tree). *Let Tr be a subtree of a dialogue tree. If all leaves of Tr are dialogues lost by P, then Tr is said to be* a failure tree.

Definition 19 (Fatal Move). *For a dialogue tree, let N be a node from which outgoing edges are P's moves and $N_1, \ldots, N_l$ be its child nodes. If there exists N_i $(1 \leq i \leq l)$ that is a root node of a failure tree, and there exists N_j $(1 \leq j \leq l)$ that is not a root node of a failure tree, then the move from N to N_i is said to be P's* fatal move at *N.*

Once a fatal move is taken, there is no possibility of P's winning a dialogue whatever move s/he makes afterwards. Therefore, strategy should be constructed in such a way that makes P avoid selecting a fatal move.

3.2 Strategy

Strategy is a function of $\mathcal{AF}_X^{d_k}$, $\mathcal{PAF}_Y^{d_k}$ and a set of allowed moves that returns a move $m_k = (X, T)$.

Definition 20 (Never Lose). *Let $\mathcal{S}$ be an arbitrary strategy. If P does not lose in all possible dialogues between P and C on ρ taken by $\mathcal{S}$, then it is said that P* never loses by *$\mathcal{S}$.*

We propose a strategy $\mathcal{S}_{\mathcal{NF}}$. This strategy is based on the principle that an agent will not make a risky move. An agent avoids making a move that causes their opponent to believe $\neg\rho$, whereas s/he positively makes a move that causes their opponent to believe ρ. S/he gives no more information if the goal is satisfied.

Strategy $\mathcal{S}_{\mathcal{NF}}$: Let $\mathcal{AF}_P^{d_k}$ and $\mathcal{PAF}_C^{d_k}$ be an argumentation framework of P for d_k and a predicted argumentation framework of C by P for d_k, respectively. Then the move $m_k = (P, T)$ is selected by the following rules.

The following rule 1 is prior to rule 2, and rule 2 is prior to rule 3.

1. If $\rho \in Bel(\mathcal{AF}_P^{d_k}) \cap Bel(\mathcal{PAF}_C^{d_k})$ where $d_k \neq d_0$, then $(P, pass)$ is selected.
2. For all possible actions where $d_k = d_0$, if $\neg\rho \in Bel(\mathcal{PAF}_C^{d_1})$, then $m_0 = (P, assert(\rho))$ is selected.
3. The descending order of priority on taking actions is $assert(p)$, $assert(S, p)$, $assertS(S, p)$, $challenge(p)$ and $pass$, that is, $assert(p)$ has the highest priority. If T is either $assert(p)$, $assert(S, p)$ or $assertS(S, p)$, then the following rules are applied.
   (a) If $\neg\rho \in Bel(\mathcal{PAF}_C^{d_k+1})$, then (P, T) is not selected.
   (b) If $\rho \in Bel(\mathcal{PAF}_C^{d_k+1})$, then (P, T) is selected.

If multiple moves that satisfy all of the above rules exist, then one of them is selected nondeterministically.

4 Example

We show the formalization of the example of selecting a laboratory discussed in Sect. 1. Let a, g and s represent propositions that applying to Charlie's laboratory, Charlie is generous, and Charlie is strict, respectively. In this dialogue, P (Alice) tries to persuade C (Bob) to believe a (to apply to Charlie's laboratory).

Assume that the strength of the formulas are given as follows: $str(g) = str(s) = str(s \rightarrow \neg a) = 3$, $str(g \rightarrow a) = str(s \rightarrow a) = 2$ and $str(a) = str(\neg a) = 1$. We show the case in which the predicted knowledge base of C by P is equivalent to C's actual knowledge base, that is, $\Pi_C = \Sigma_C$. Assume that knowledge bases are given as follows.

$$\Sigma_P = \{g, s, g \rightarrow a, s \rightarrow a, a\} \qquad \Pi_P = \{g \rightarrow a\}$$
$$\Sigma_C = \{g \rightarrow a, s \rightarrow \neg a, \neg a\} \qquad \Pi_C = \{g \rightarrow a, s \rightarrow \neg a, \neg a\}$$

Below we show relevant arguments from given knowledge bases. The number attached to each argument is its preference. More arguments can be constructed, but here we show only related ones to simplify an explanation.

$$A_1 = (\emptyset, g)[3] \qquad A_6 = (\emptyset, a)[1]$$
$$A_2 = (\emptyset, s)[3] \qquad A_7 = (\emptyset, \neg a)[1]$$
$$A_3 = (\{s, s \rightarrow \neg a\}, \neg a)[3] \qquad A_8 = (\{g \rightarrow a, \neg a\}, \neg g)[1]$$
$$A_4 = (\{g, g \rightarrow a\}, a)[2] \qquad A_9 = (\{s \rightarrow a, \neg a\}, \neg s)[1]$$
$$A_5 = (\{s, s \rightarrow a\}, a)[2] \qquad A_{10} = (\{s \rightarrow \neg a, a\}, \neg s)[1]$$

We show three possible dialogues in Table 1.

Let $\mathcal{PAF}_C^{d_{k+1}} = \langle PAR_C^{d_{k+1}}, PAT_C^{d_{k+1}} \rangle$ be a predicted argumentation framework of C by P for d_{k+1}, that is, obtained as a result of the move m_k in a dialogue $d_{k+1} = [m_0, \ldots, m_k]$. Here, $\mathcal{PAF}_C^{d_{k+1}} = AF(\Pi_C \cup CS_P^{d_{k+1}} \cup CS_C^{d_{k+1}})$. In these dialogues, $CS_C^{d_i}$ is $\emptyset$ for any i $(0 \leq i \leq k+1)$. Important transitions $PAR_C^{d_{k+1}}$, $Ext(\mathcal{PAF}_C^{d_{k+1}})$ and $CS_P^{d_{k+1}}$ are shown in the table, and the graph representation corresponding to $\mathcal{PAF}_C^{d_{k+1}}$ in each state is shown in Fig. 1(a)∼(e). In the figure, nodes represent arguments and edges represent attacks.

Initially, there is no attack, $PAR_C^{d_0} = \{A_7, A_8\}$, $Ext(\mathcal{PAF}_C^{d_0}) = \{A_7, A_8\}$, and $CS_P^{d_0} = \emptyset$ hold, represented in a graph AF1 (Fig. 1(a)). There are three allowed moves at the initial state. That is, P can give three acts: $assert(a)$, $assert(\{g, g \rightarrow a\}, a)$ or $assert(\{s, s \rightarrow a\}, a)$.

Dialogue1 shows the dialogue along the strategy $\mathcal{S}_{\mathcal{NF}}$. P first gives $assert(\{g, g \rightarrow a\}, a)$ from rules 3(a) and (b) (Fig. 1(c)). In this case, $a \in fml(A_4) \subseteq Bel(\mathcal{PAF}_C^{d_1})$. Next, C can provide only $challenge(g)$, $challenge(g \rightarrow a)$ or $pass$. The case in which $challenge(g)$ is given is shown in the table. P gives $pass$ along the strategy $\mathcal{S}_{\mathcal{NF}}$ against C's move. P continues to give $pass$ afterwards and finally wins. In case C gives $pass$ at any move, the result is the same.

If P does not have a strategy, she may make any one of three moves at the initial state. Dialogue2 and Dialogue3 are the ones P gives $assert(a)$ first (Fig. 1(b)). Next, C can provide only $challenge(a)$ except for $pass$. Next, P can

Table 1. Transitions of argumentation frameworks.

Dialogue1:

move m_k	$PAR_C^{d_{k+1}}$	$Ext(\mathcal{PAF}_C^{d_{k+1}})$	$CS_P^{d_{k+1}}$	graph
m_0: $(P, assert(\{g, g \rightarrow a\}, a)$	$\{A_7, A_8, A_6, A_{10},$	$\{A_1, A_4, A_6, A_{10}\},$	$\{a, g, g \rightarrow a\}$	AF3
m_1: $(C, challenge(g))$	$A_1, A_4\}$			
m_2: $(P, pass)$				
m_3: $(C, challenge(g \rightarrow a))$				
m_4: $(P, pass)$				
m_5: $(C, pass)$				

Dialogue2:

move m_k	$PAR_C^{d_{k+1}}$	$Ext(\mathcal{PAF}_C^{d_{k+1}})$	$CS_P^{d_{k+1}}$	graph
m_0: $(P, assert(a))$	$\{A_7, A_8, A_6, A_{10}\}$	$\emptyset$	$\{a\}$	AF2
m_1: $(C, challenge(a))$				
m_2: $(P, assertS(\{g, g \rightarrow a\}, a)$	$\{A_7, A_8, A_6, A_{10},$	$\{A_1, A_4, A_6, A_{10}\}$	$\{a, g, g \rightarrow a\}$	AF3
m_3: $(C, challenge(g))$	$A_1, A_4\}$			
m_4: $(P, pass)$				
m_5: $(C, challenge(g \rightarrow a))$				
m_6: $(P, pass)$				
m_7: $(C, pass)$				

Dialogue3:

move m_k	$PAR_C^{d_{k+1}}$	$Ext(\mathcal{PAF}_C^{d_{k+1}})$	$CS_P^{d_{k+1}}$	graph
m_0: $(P, assert(a))$	$\{A_7, A_8, A_6, A_{10}\}$	$\emptyset$	$\{a\}$	AF2
m_1: $(C, challenge(a))$				
m_2: $(P, assertS(\{s, s \rightarrow a\}, a)$	$\{A_7, A_8, A_6, A_{10},$	$\{A_2, A_3$	$\{a, s, s \rightarrow a\}$	AF4
m_3: $(C, challenge(s))$	$A_2, A_5, A_3, A_9\}$	$A_7, A_8\}$		
m_4: $(P, assertS(\{g, g \rightarrow a\}, a)$	$\{A_7, A_8, A_6, A_{10},$	$\{A_1, A_2$	$\{a, g, s,$	AF5
	A_2, A_5, A_3, A_9	$A_3, A_7\}$	$g \rightarrow a, s \rightarrow a\}$	
	$A_1, A_4\}$			

give either of $(assertS(\{g, g \rightarrow a\}, a)$ or $assertS(\{s, s \rightarrow a\}, a)$. If P gives the former one (Fig. 1(c)), $a \in fml(A_4) \subseteq Bel(\mathcal{PAF}_C^{d_3})$ holds. Dialogue2 shows this case. After that, if P gives *pass*, she finally wins. On the other hand, if P gives the latter one (Fig. 1(d)), $\neg a \in fml(A_3) \subseteq Bel(\mathcal{PAF}_C^{d_3})$ holds. Dialogue3 shows this case. Even if P gives $assertS(\{g, g \rightarrow a\}, a)$ afterwards (Fig. 1(e)), $\neg a \in fml(A_3) \subseteq Bel(\mathcal{PAF}_C^{d_5})$ holds, and P loses. In case C gives *pass* at any move, the result is the same.

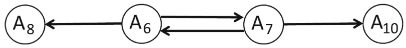
In this example, $assertS(a, \{s, s \rightarrow a\})$ is a fatal move.

5 Results

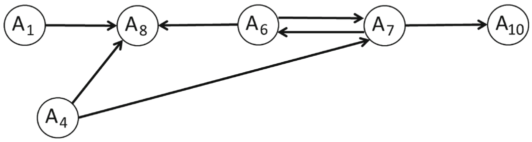
In this section, we discuss some properties of our model and what formulas should be included in a predicted knowledge base. All proofs are shown in the Appendix.

Note that hereafter N_i denotes a node in the depth i in a dialogue tree.

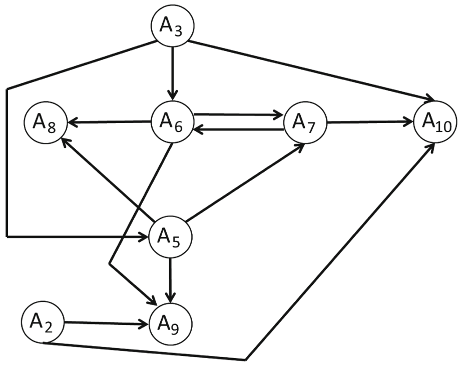
(a) AF1: initial state



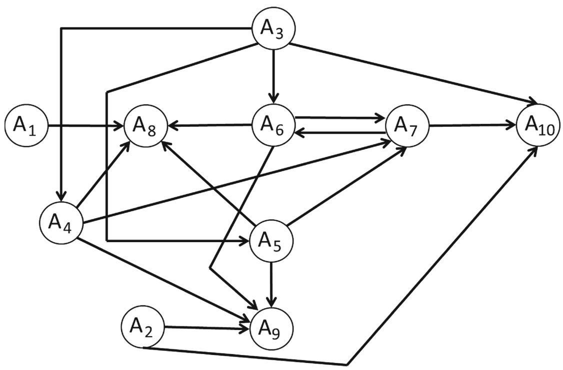
(b) AF2: after $(P, assert(a))$ in Dialogue2 and Dialogue3



(c) AF3: after $(P, assert(\{g, g \to a\}, a)$ in Dialogue1,
after $(P, assertS(\{g, g \to a\}, a)$ in Dialogue2



(d) AF4: after $(P, assertS(\{s, s \to a\}, a)$ in Dialogue3



(e) AF5: after $(P, assertS(\{g, g \to a\}, a)$ in Dialogue3

Fig. 1. Predicted argumentation frameworks of C by P.

Lemma 1. *For a failure tree of which the root is N_i corresponding to a dialogue d_i, $\neg\rho \in Bel(\mathcal{AF}_C^{d_i})$ holds.*

Here, we introduce the concept of *changing move (c-move)*. It represents the turning point of the move from the state in which C does not accept $\neg\rho$, to the state in which C accepts $\neg\rho$.

Definition 21 (c-move). *For a dialogue $d_{k+1} = [m_0, \ldots, m_k]$, if $\neg\rho \notin Bel(\mathcal{AF}_C^{d_k})$ and $\neg\rho \in Bel(\mathcal{AF}_C^{d_{k+1}})$, then m_k is said to be* changing move*, c-move in short.*

The following theorem and its corollary show a condition for a non-failing dialogue.

Theorem 1. *If $\Pi_C = \Sigma_C$, then P does not give a c-move at N_k for any k $(1 \leq k)$ by the strategy $\mathcal{S}_{\mathcal{NF}}$.*

Corollary 1. *If $\Pi_C = \Sigma_C$ and $\neg\rho \notin Bel(\mathcal{AF}_C^{d_k})$, then P can avoid a fatal move at N_k for any k $(1 \leq k)$ by the strategy $\mathcal{S}_{\mathcal{NF}}$.*

When the predicted knowledge base is equivalent to the real knowledge base, if there exists such an initial move that P predicts that C will not believe $\neg\rho$ next, then P never loses. It means that there is a case in which we can judge that P never loses under the strategy $\mathcal{S}_{\mathcal{NF}}$ simply from given knowledge bases.

Next, we consider the case in which the predicted knowledge base is a subset of the real knowledge base.

We show the condition in which P's strongly win can be judged only from an initially given C's real knowledge base. The following theorem shows that when the prediction is a subset of the real knowledge base, if there are no arguments which have $\neg\rho$ as its conclusion in C's initial argumentation framework, then X strongly wins by the strategy $\mathcal{S}_{\mathcal{NF}}$.

Theorem 2. *Let $AF(\Sigma_C)$ be $\mathcal{AF}_C^{d_0} = \langle AR_C^{d_0}, AT_C^{d_0}\rangle$. If $\Pi_C \subseteq \Sigma_C$ and $\{A \mid A \in AR_C^{d_0} \wedge concl(A) = \neg\rho\} = \emptyset$, then $\rho \in Bel(\mathcal{AF}_P^{d_k}) \cap Bel(\mathcal{AF}_C^{d_k})$ holds for a complete dialogue d_k by the strategy $\mathcal{S}_{\mathcal{NF}}$.*

Next, we discuss what formulas should be included in a predicted knowledge base Π_C.

The following theorem shows that it is insufficient to decide the condition for Π_C in order not to fail in P's persuasion simply from given knowledge bases, rather all dialogues must be surveyed.

Theorem 3. *Let S be the set of formulas in NBA-arguments of $AF(\Sigma_P \cup \Sigma_C)$, If $\Pi_C = S \cap \Sigma_C$, then P cannot always avoid the fatal move by the strategy $\mathcal{S}_{\mathcal{NF}}$.*

We show a condition for Π_C using the concept of *NBA-only move*.

For a dialogue d_k, let $\mathcal{AF}_X^{d_k} = \langle AR_X^{d_k}, AT_X^{d_k}\rangle$ and $\mathcal{PAF}_X^{d_k} = \langle PAR_X^{d_k}, PAT_X^{d_k}\rangle$. Then $PAR_X^{d_k} \subseteq AR_X^{d_k}$ holds.

Definition 22 (NBA-Only Move). *Assume that* $\Pi_Y \subseteq \Sigma_Y$. *Let* m_k *be* X*'s move,* $\mathcal{AF}_Y^{d_{k+1}} = \langle AR_Y^{d_{k+1}}, AT_Y^{d_{k+1}} \rangle$ *and* $\mathcal{PAF}_Y^{d_{k+1}} = \langle PAR_Y^{d_{k+1}}, PAT_Y^{d_{k+1}} \rangle$. *If there does not exist* $A \in AR_Y^{d_{k+1}} - PAR_Y^{d_{k+1}}$ *such that* $\exists C \in AR_Y^{d_{k+1}}; (C, A) \in AT_Y^{d_{k+1}}$ *holds, then the* m_k *is said to be* X*'s* NBA-only move.

An intuitive meaning of an NBA-only move is as follows: when we compare Y's argumentation framework and the predicted argumentation framework of Y by X, let S be a set of arguments that are included in the former but not in the latter; there is no argument in S that is attacked by some argument in the former.

For a complete dialogue $d_k = [m_0, \ldots, m_{k-1}]$ between P and C on ρ, let m_i $(0 \leq i \leq k-1)$ be a *c-move*, and SA_{d_k} be the set formulas in NBA-arguments in $\mathcal{AF}_C^{d_i+1}$. Let $SA = \bigcup_{d_k} SA_{d_k}$. It is clear that $SA \subseteq \Sigma_P \cup \Sigma_C$. Therefore, SA is divided into two disjoint subsets $SA_{P \backslash C}$ and SA_C, where $SA_{P \backslash C}$ is a set of formulas included in $\Sigma_P \backslash \Sigma_C$ and SA_C is a set of formulas included in Σ_C.

Theorem 4. *If* $\Pi_C = SA_C$ *and all c-moves in a dialogue tree are* P*'s NBA-only moves, then* P *does not give a c-move at* N_k *for any* k $(1 \leq k)$ *by the strategy* $\mathcal{S}_{\mathcal{NF}}$.

Corollary 2. *If* $\Pi_C = SA_C$, *all c-moves in a dialogue tree are NBA-only and* $\neg\rho \notin Ext(\mathcal{AF}_C^{d_k})$, *then* P *can avoid a fatal move at* N_k *for any* k $(1 \leq k)$ *by the strategy* $\mathcal{S}_{\mathcal{NF}}$.

6 Discussion

There have been many studies on Dung's abstract argumentation framework [12]. Among them, a dialogue model using argumentation based on this framework has been proposed.

Our model is based on the one studied by Amgoud et al. The model is set out and applied to several types of dialogues [2]. The strategy is defined and the dialogue according to the strategy is shown [3]. There, the strategy is based on the level of acceptance, strength of the argument and attitude of the agents. The various relationships between sets of knowledge, including that between the joint knowledge of agents and the outcomes of dialogues, are investigated [10]. The most significant difference between our work and theirs is the use of the predicted knowledge base. We construct a strategy using the predicted knowledge base, whereas their strategy is constructed without considering the opponent's inner state. Moreover, we have given an explicit definition to the argumentation framework for the current state of a dialogue, whereas formalization of the current argumentation framework is ambiguous in their works.

It is essential to consider an opponent's beliefs, especially in handling a strategic dialogue, which may include a lie. Several works have been undertaken regarding on this issue. Thimm et al. studied a strategy that reflects an opponent's belief [16] but they did not relate belief to an extension of an argumentation framework. Rienstra et al. showed a strategy of selecting the best move from multiple opponent models with probability [14], and Hadjinikolis et al. showed an approach of augmenting opponent models from accumulated dialogues with an agent's likelihood [7]. They evaluated their approaches experimentally, whereas we focus on giving a strategy and investigate its validity theoretically. Black et al. formally investigated usage and maintenance of opponent models illustrating a simple persuasion dialogue with different types of persuaders [5]. However, the order of utterances is out of their focus. Sakama presented the treatment of untrusted argumentation [15]. Rahwan et al. discussed hiding and lying in argumentation [13]. In these works, abstract argumentation frameworks are used, that is, arguments are not constructed from logical deduction from knowledge base, whereas a structured framework is used in our model.

ASPIC+ is a structured argumentation framework that generates arguments from a knowledge base using logical entailment [11]. However, only static argumentation can be handled in that framework and dynamically changing structures are not available. Okuno and Takahashi proposed a dynamic structured argumentation [9]. In their proposed method, each agent's argument is generated from their own knowledge base and commitment store, and the argumentation structure dynamically changes. Their model did not operate at the dialogue level, whereas we propose here a dialogue model based on an argumentation framework that changes at every move.

7 Conclusion

We have proposed a dialogue model that utilizes a predicted knowledge base and a strategy of withholding moves predicted to fail and only providing moves that avoid failure to persuade. We have investigated the conditions under which a persuasive dialogue never fails using this strategy, when the predicted knowledge base is equivalent to the actual knowledge base of an opponent. The introduction of prediction provides a model that better simulates real dialogue.

Moreover, we have discussed what a predicted knowledge base should include for a persuasive dialogue not to fail. Our main contribution is to set out the formalization of a dialogue using prediction and to propose a strategy for non-failing persuasion.

There are several issues that should be addressed in future work. The conditions presented herein for non-failing persuasion are relatively loose and inefficient and, therefore, more rigorous and efficient conditions should be explored. The next step is to determine conditions for successful persuasion rather than for non-failing persuasion. In addition, we will investigate a case in which a predicted knowledge base is not a subset of an actual one.

Because it is necessary to have an opponent's predicted knowledge base to construct a lie or to reveal it, our final goal is to develop a strategy to handle dialogue that includes a lie, and to investigate conditions of a predicted knowledge base that support the validity of the strategy.

Appendix

We show the sketch of the proofs because of the space limit.

Proof for Lemma 1. For any dialogue $d_i = [m_0, \ldots, m_{i-1}]$, if P can proceed with the dialogue just by giving *pass* as acts of $m_i, \ldots, m_k$, then P does not add any information to C. Therefore, a complete dialogue $[m_0, \ldots, m_{i-1}, m_i, \ldots, m_k]$ exists that satisfies $Bel(\mathcal{AF}_C^{d_{k+1}}) = Bel(\mathcal{AF}_C^{d_i})$. Thus, such a leaf node N_{k+1} exists that satisfies $Bel(\mathcal{AF}_C^{d_{k+1}}) = Bel(\mathcal{AF}_C^{d_i})$ in a subtree of which the root node is N_i. As N_i is the root node of a failure tree, $\neg\rho \in Bel(\mathcal{AF}_C^{d_{k+1}})$ holds. Therefore, $\neg\rho \in Bel(\mathcal{AF}_C^{d_i})$ holds. □

Proof for Theorem 1. For any dialogue d_k, an agent must not give a move at N_k if $\neg\rho \in Bel(\mathcal{PAF}_C^{d_{k+1}})$ holds by rule 3(a) of the strategy $\mathcal{S}_{\mathcal{NF}}$. It follows that $\neg\rho \notin Bel(\mathcal{AF}_C^{d_{k+1}})$ holds, since $\Pi_C = \Sigma_C$. It means that a move other than *c-move* should have been selected by the strategy $\mathcal{S}_{\mathcal{NF}}$. □

Proof for Corollary 1. If a fatal move is selected at N_k, there exists a failure tree of which the root is N_{k+1}. From Lemma 1, $\neg\rho \in Bel(\mathcal{AF}_Y^{d_{k+1}})$ holds. It means that this move is a *c-move*. It is a contradiction from Theorem 1. Therefore, an agent can avoid the fatal move by the strategy $\mathcal{S}_{\mathcal{NF}}$. □

Proof for Theorem 2. In this case, according to the strategy $\mathcal{S}_{\mathcal{NF}}$, agent P first gives $assert(\rho)$, and repeats *pass* against any move given by C afterwards. C cannot attack ρ since s/he cannot construct an argument of which a conclusion is $\neg\rho$. In this case, $\rho \in Bel(AF(\Sigma_C \cup \{\rho\})) = Bel(\mathcal{AF}_C^{d_k})$. □

Proof for Theorem 3. We show an example. Assume that the strength of each formula is given as follows: $str(a) = str(a \to \rho) = 5$, $str(b) = str(c) = str(b \to \neg\rho) = 4$, $str(b \to \rho) = str(c \to \rho) = 3$, $str(\neg\rho) = 2$ and $str(\rho) = 1$. Assume that knowledge bases are given as follows: $\Sigma_P = \{\rho, b, b \to \rho, c, c \to \rho, a\}$, $\Sigma_C = \{\neg\rho, b \to \neg\rho, a \to \rho\}$. Then, Π_C is defined as $\{a \to \rho\}$.

In this case, a dialogue in which P behaves according to the strategy $\mathcal{S}_{\mathcal{NF}}$ proceeds as follows. P gives $assert(\rho)$ as an initial move m_0. Then, C can give either $assert(\neg\rho)$, $challenge(\rho)$ or *pass*. Assume that C gives $assert(\neg\rho)$ as m_1. Then P can gives either $m_2 = assertS(\{b, b \to \rho\}, \rho)$ or $m_2' = assertS(\{c, c \to \rho\}, \rho)$. Let d_3 and d_3' dialogues $[m_0, m_1, m_2]$ and $[m_0, m_1, m_2']$, respectively. If P gives m_2, it causes C to make a new argument $(\{b, b \to \neg\rho\}, \neg\rho)$, which is an NBA-argument in $\mathcal{AF}_C^{d_3}$. Therefore, C believes $\neg\rho$ at the state. Since this argument is not attacked other than by $(\{a, a \to \rho\}, \rho)$ which never appears in any dialogue, $\neg\rho \in Bel(\mathcal{AF}_C^{d_k})$ holds for $d_k = [m_0, m_1, m_2, \ldots, m_{k-1}]$. On the other hand, if P gives m_2', it causes C to make a new argument $(\{c, c \to \rho\}, \rho)$, which attacks an argument $(\emptyset, \neg\rho)$ in $\mathcal{AF}_C^{d_3'}$. Therefore, C believes ρ at that state. Thus, m_2 is a fatal move. However, the strategy $\mathcal{S}_{\mathcal{NF}}$ cannot determine which is the best move between m_2 or m_2'. We should have $b \to \neg\rho$ in Π_C, instead of $a \to \rho$. □

Proof for Theorem 4. If $AR_C^{d_{k+1}} - PAR_C^{d_{k+1}} = \emptyset$, then *c-move* is never selected at N_k by the strategy $\mathcal{S}_{\mathcal{NF}}$, by the same reason with that of Theorem 1.

Therefore, there should exist an argument $A \in AR_C^{d_{k+1}} - PAR_C^{d_{k+1}}$. Assume that P gives a *c-move* at N_k.

Since A is an NBA-argument in $\mathcal{AF}_C^{d_{k+1}}$ from the assumption that all *c-moves* in a dialogue tree are P's NBA-only moves, $fml(A) \subseteq SA_C \cup SA_{P \setminus C}$. On the other hand, $fml(A) \cap SA_C \subseteq SA_C = \Pi_C$ and $fml(A) \cap SA_{P \setminus C} \subseteq CS_P^{d_{k+1}}$. Therefore, $fml(A) \subseteq \Pi_C \cup CS_P^{d_{k+1}}$. On the other hand, $\Pi_C \cup CS_P^{d_{k+1}} \subseteq \Pi_C \cup CS_P^{d_{k+1}} \cup CS_C^{d_{k+1}} = Fml(PAR_C^{d_{k+1}})$. It follows that $A \in PAR_C^{d_{k+1}}$, which is a contradiction.

Therefore, P never gives a *c-move* at N_k. □

Proof for Corollary 2. It is proved from Theorem 4 using similar logic to the proof of Corollary 1. □

References

1. Amgoud, L., Cayrol, C.: On the acceptability of arguments in preference-based argumentation. In: UAI 1998, pp. 1–7 (1998)
2. Amgoud, L., Maudet, N., Parsons, S.: Modeling dialogues using argumentation. In: ICMAS 2000, pp. 31–38 (2000)
3. Amgoud, L., Maudet, N.: Strategical considerations for argumentative agents (preliminary report). In: NMR 2002, pp. 399–407 (2002)
4. Bench-Capon, T.: Persuasion in practice argument using value-based argumentation frameworks. J. Logic Comput. **13**(3), 429–448 (2003)
5. Black, E., Hunter, A.: Reasons and options for updating an opponent model in persuasion dialogues. In: TAFA 2015 (2015)
6. Dung, P.M.: On the acceptability of arguments and its fundamental role in non-monotonic reasoning, logic programming and n-person games. Artif. Intell. **77**, 321–357 (1995)
7. Hadjinikolis, C., Siantos, C., Modgil, S., Black, E., McBurney, P.: Opponent modelling in persuasion dialogues. In: IJCAI 2013, pp. 164–170 (2013)
8. Modgil, S.: Reasoning about preferences in argumentation frameworks. Artif. Intell. **173**(9–10), 901–934 (2009)
9. Okuno, K., Takahashi, K.: Argumentation system with changes of an agent's knowledge base. In: IJCAI 2009, pp. 226–232 (2009)
10. Parsons, S., Wooldridge, M., Amgoud, L.: On the outcomes of formal inter-agent dialogues. In: AAMAS 2003, pp. 616–623 (2003)
11. Prakken, H.: An abstract framework for argumentation with structured arguments. Argum. Comput. **1**(2), 93–124 (2010)
12. Rahwan, I., Simari, G. (eds.): Argumentation in Artificial Intelligence. Springer, Heidelberg (2009)
13. Rahwan, I., Lason, K., Tohmé, F.: A characterization of strategy-proofness for grounded argumentation semantics. In: IJCAI 2009, pp. 251–256 (2009)
14. Rienstra, T., Thimm, M., Oren, N.: Opponent models with uncertainty for strategic argumentation. In: IJCAI 2013, pp. 332–338 (2013)
15. Sakama, C.: Dishonest arguments in debate games. In: COMMA 2012, pp. 177–184 (2012)
16. Thimm, M., García, A.J.: On strategic argument selection in structured argumentation systems. In: McBurney, P., Rahwan, I., Parsons, S. (eds.) ArgMAS 2010. LNCS, vol. 6614, pp. 286–305. Springer, Heidelberg (2011)