Debate Games in Logic Programming

Chiaki Sakama

Department of Computer and Communication Sciences
Wakayama University, Sakaedani, Wakayama 640-8510, Japan
sakama@sys.wakayama-u.ac.jp

Abstract. A *debate game* provides an abstract model of debates between two players based on the formal argumentation framework. This paper presents a method of realizing debate games in logic programming. Two players have their knowledge bases represented by extended logic programs and build claims using arguments associated with those programs. A player revises its knowledge base with arguments posed by the opponent player, and tries to refute claims by the opponent. During a debate game, a player may claim false or incorrect arguments as a tactic to win the game. The result of this paper provides a new formulation of debate games in a non-abstract argumentation framework associated with logic programming. Moreover, it provides a novel application of logic programming to modelling social debates which involve argumentative reasoning, belief revision and dishonest reasoning.

1 Introduction

Logic programming and argumentation are two different frameworks for knowledge representation and reasoning in artificial intelligence (AI). In his seminal paper, Dung [4] points out a close connection between the two frameworks and shows that a logic program can be considered as a schema for generating arguments. Since then, several attempts have been made for integrating the two frameworks ([1, 12, 8, 20]; see [9] for an overview).

A line of research of formal argumentation is concerned with the dialectical process of two or more players who are involved in a discussion [3]. Along this line, Sakama [18] introduces a *debate game* between two players based on the formal argumentation framework. In a debate game, a player makes the initial claim, then the opponent player tries to refute it by building a counter-claim. A debate continues until one cannot refute the other, and the player who makes the last claim wins the game. A debate game has unique features such that (i) each player has its own argumentation framework as its background knowledge, (ii) during a debate each player revises its argumentation framework by new arguments provided by the opponent player, and (iii) a player may claim inaccurate or even false arguments as a tactic to win a debate. The study [18] formulates debate games using the abstract argumentation theory of [4].

The abstract argumentation theory has an advantage that it is not bound to any particular representation for arguments on the one hand, but on the other hand it does not specify how arguments are generated from the underlying knowledge base and what conclusions are yielded by those arguments. In [2] the authors argue that "Argumentation, as it happens in the world around us, is almost never completely abstract. ..."

Instead, the arguments one encounters in daily life consist of *reasons* that support particular *claims*. These reasons can formally be modelled in the form of *rules*, that are instances of underlying *argumentation schemes* [14]." In this respect, debate games based on the abstract argumentation theory need yet another formulation based on non-abstract argumentation frameworks.

With this motivation, this paper uses logic programming as an underlying representation language and formulates debate games in a non-abstract argumentation framework. In this framework, each player has a knowledge base represented by an extended logic program, and builds claims using arguments which can contain information brought by the opponent as well as information in the player's program. During a game, a player may use *dishonest* claims to refute the opponent, while a player must be self-consistent in its claims. The proposed framework provides an abstraction of real-life debates and realizes a formal dialogue system in logic programming. The rest of this paper is organized as follows. Section 2 reviews a framework of argument-based logic programming. Section 3 introduces debate games in logic programming and investigates formal properties. Section 4 discusses related issues and Section 5 concludes the paper.

2 Arguments in Logic Programming

In this paper we consider the class of extended logic programs [10]. An *objective literal* is a ground atom B or its explicit negation $\neg B$. We define $\neg\neg B = B$. A *default literal* is of the form $not\, L$ where L is an objective literal and not is *negation as failure* (NAF). An *extended logic program* (or simply a *program*) P is a finite set of *rules* of the form:

$$L_0 \leftarrow L_1, \ldots, L_m, not\, L_{m+1}, \ldots, not\, L_n$$

where each L_i $(0 \le i \le n)$ is an objective literal. The literal L_0 is the *head* of the rule and the conjunction $L_1, \ldots, L_m, not\, L_{m+1}, \ldots, not\, L_n$ is the *body* of the rule. A rule r is *believed-true* in P if $r \in P$. A rule containing default literals is called a *default rule*. A rule $L \leftarrow$ with the empty body is also called a *fact* and is identified with a literal L.

Let Lit be the set of all objective literals in the language of a program. A set S $(\subset Lit)$ is *consistent* if $L \in S$ implies $\neg L \notin S$ for any $L \in Lit$. The semantics of a program is given by its *answer sets* [10]. First, let P be a program containing no default literal and $S \subset Lit$. Then, S is an *answer set* of P if S is a consistent minimal set satisfying the condition that for each rule of the form $L_0 \leftarrow L_1, \ldots, L_m$ in P, $\{L_1, \ldots, L_m\} \subseteq S$ implies $L_0 \in S$. Second, given any program P (possibly containing default literals) and $S \subset Lit$, a *reduct* of P with respect to S (written P^S) is defined as follows: a rule $L_0 \leftarrow L_1, \ldots, L_m$ is in P^S iff there is a rule of the form $L_0 \leftarrow L_1, \ldots, L_m, not\, L_{m+1}, \ldots, not\, L_n$ in P such that $\{L_{m+1}, \ldots, L_n\} \cap S = \emptyset$. Then, S is an *answer set* of P if S is an answer set of P^S. A program may have none, one or multiple answer sets in general. A program is *consistent* if it has an answer set; otherwise, it is *inconsistent*.

Definition 2.1. ([12, 20]) An *argument* associated with a program P is a finite sequence $A = [r_1; \cdots; r_n]$ of rules $r_i \in P$ such that (i) for every $1 \le i \le n$, for every objective literal L_j in the body of r_i there is a rule r_k ($k > i$) such that the head of r_k is L_j. (ii) No two distinct rules in the sequence have the same head.

The head of a rule in an argument A is called a *conclusion* of A, and a default literal $not\,L$ in the body of a rule in A is called an *assumption* of A. We write $assum(A)$ for the set of assumptions and $concl(A)$ for the set of conclusions of an argument A. By the condition (i) of Definition 2.1, every objective literal in the body of a rule r_i is justified by the consequence of a rule that appears later in the sequence. The condition (ii) keeps an argument from containing circular sequences of rules. A *subargument* of A is a subsequence of A which is an argument. An argument A with a conclusion L is a *minimal argument for* L if there is no subargument of A with the conclusion L. An argument is *minimal* if it is minimal for some literal L. The minimality condition presents that an argument does not include rules which do not contribute to conclude some particular literal L.

Remark: In this paper, we slightly abuse the notation and use the same letter A to denote the *set* of rules included in an argument A. Thus, $P \cup A$ means the set of rules included either in a program P or in an argument A.

Example 2.1. Let P be the program:

$$\begin{aligned} & p \leftarrow q, \\ & \neg p \leftarrow not\ q, \\ & q \leftarrow, \\ & r \leftarrow s. \end{aligned}$$

Then, the following facts hold.

- The minimal argument for p is $A_1 = [\,p \leftarrow q\,;\, q \leftarrow\,]$, $concl(A_1) = \{p, q\}$, and $assum(A_1) = \emptyset$.
- The minimal argument for $\neg p$ is $A_2 = [\,\neg p \leftarrow not\,q\,]$, $concl(A_2) = \{\neg p\}$ and $assum(A_2) = \{not\,q\}$.
- The minimal argument for q is $A_3 = [\,q \leftarrow\,]$, $concl(A_3) = \{q\}$ and $assum(A_3) = \emptyset$.
- r and s have no minimal arguments.

Proposition 2.1. *Let P be a consistent program containing no default literal. Then, for any argument A associated with P, $concl(A) \subseteq S$ holds for the answer set S of P.*

Proof. Let P' be the program which is obtained by replacing every negative literal $\neg L$ in P with a new atom L' that is uniquely associated with $\neg L$. As P is consistent, P' has the least model S' iff P has the answer set S where $\neg L$ in S is replaced by the atom L' in S'. Let A' be an argument associated with P'. Then, $A' \subseteq P'$ implies $concl(A') \subseteq S'$ by the monotonicity of deduction. By replacing L' with $\neg L$, $A \subseteq P$ implies $concl(A) \subseteq S$. □

Definition 2.2. ([12, 20]) Let A_1 and A_2 be two arguments.

- A_1 *undercuts* A_2 if there is an objective literal L such that L is a conclusion of A_1 and $not\, L$ is an assumption of A_2.
- A_1 *rebuts* A_2 if there is an objective literal L such that L is a conclusion of A_1 and $\neg L$ is a conclusion of A_2.
- A_1 *attacks* A_2 if A_1 undercuts or rebuts A_2.
- A_1 *defeats* A_2 if A_1 undercuts A_2, or A_1 rebuts A_2 and A_2 does not undercut A_1.

An argument is *coherent* if it does not attack itself. A set S of arguments is *conflict-free* if no argument in S attacks an argument in S. Given a program P, we denote the set of minimal and coherent arguments associated with P by $Args(P)$.

If an argument A_1 undercuts another argument A_2, then A_1 denies an assumption of A_2. This means that the assumption conflicts with the evidence to the contrary, and A_1 defeats A_2 in this case. If A_1 rebuts A_2, on the other hand, two arguments support contradictory conclusions. In this case, the attack relation is symmetric and A_1 defeats A_2 under the condition that A_2 does not undercut A_1. The coherency condition presents self-consistency of an argument. By definition, if $A \in Args(P)$ then the set A of rules is consistent.

Example 2.2. In the program P of Example 2.1, the following facts hold.

- $Args(P) = \{A_1,\, A_2,\, A_3\}$.
- A_1 and A_3 undercut (and also defeat) A_2.
- A_1 rebuts A_2 and A_2 rebuts A_1.
- $\{A_1, A_3\}$ is conflict-free, but $\{A_1, A_2\}$ and $\{A_2, A_3\}$ are not.
- The argument $A_4 = [p \leftarrow q\,;\ \neg p \leftarrow not\, q;\ q \leftarrow]$ is incoherent.

Proposition 2.2. *Let P be a consistent program. For any argument $A \in Args(P)$, if A is not defeated by any argument associated with P, then $concl(A) \subseteq S$ for any answer set S of P.*

Proof. Let A^+ be the set of rules obtained from A by removing every default literal in A. When A is not defeated by any argument associated with P, $A^+ \subseteq A^S$ for any answer set S of P, where A^S is the reduct of A wrt S. By Proposition 2.1, for any argument A^S associated with P^S, $concl(A^S) \subseteq S$ for any answer set S of P. Since $A^+ \subseteq A^S$ implies $concl(A^+) \subseteq concl(A^S)$, the result holds. □

In Example 2.1, A_1 and A_3 are defeated by no argument, then $concl(A_1)$ and $concl(A_3)$ are subsets of the answer set $\{p, q\}$ of P.

3 Debate Games in Logic Programming

3.1 Debate Games

A debate game involves two players. Each player has its knowledge base defined as follows.

Definition 3.1 (player). A *player* has a knowledge base $K = (P, O)$ where P is a consistent program representing the player's belief and O is a set of rules brought by another player. In particular, the initial knowledge base of a player is $K = (P, \emptyset)$.

In this paper, we identify a player with its knowledge base. We represent two players by K_1 and K_2. For a player K_1 (resp. K_2), the player K_2 (resp. K_1) is called the *opponent*.

Definition 3.2 (revision). Let $K = (P, O)$ be a player and A an argument. Then, *revision* of K with A is defined as

$$rev(K, A) = (P \setminus R,\ O \cup A)$$

where $R = \{ r \mid$ there is a literal L in $concl(A)$ such that $not\, L$ is in the body of a rule r and A is not defeated by any argument associated with $P \cup O \cup A \}$.

The function rev is iteratively applied to a player. We represent the result of the i-th revision of K by $K^i = (P^i, O^i)$ $(i \geq 0)$, that is, $K^i = (P^i, O^i) = rev(K^{i-1}, A_i)$ $(i \geq 1)$ for arguments $A_1, \ldots, A_i$ and $K^0 = (P^0, O^0) = (P, \emptyset)$.

Note that we handle A as a set here. By definition, revision adds rules A to O while it removes default rules R from P. When a player K cannot defeat the new argument A, the player is obliged to accept it and removes default rules R that have assumptions conflicting with conclusions of A. The reason of separating P and O is to distinguish belief originated in a player's program from information brought by the opponent player. A player having a knowledge base after the i-th revision is represented by K^i, but we often omit the superscript i when it is unimportant in the context.

Definition 3.3 (claim). Let $K_1 = (P_1, O_1)$ and $K_2 = (P_2, O_2)$ be two players.

1. The *initial claim* is a pair of the form: $(\texttt{in}(A), \_)$ where $A \in Args(P_1)$. It is read that "the player K_1 claims the argument A."
2. A *counter-claim* is a pair of the form: $(\texttt{out}(B), \texttt{in}(A))$ where $A \in Args(P_k \cup O_k)$ and $B \in Args(P_l \cup O_l)$ $(k, l = 1, 2;\ k \neq l)$. It is read that "the argument B by the player K_l does not hold because the player K_k claims the argument A".

The initial claim or counter-claims are simply called *claims*. A claim $(\texttt{in}(A), \_)$ or $(\texttt{out}(B), \texttt{in}(A))$ by a player is *refuted* by the claim $(\texttt{out}(A), \texttt{in}(C))$ with some argument C by the opponent player.

Definition 3.4 (debate game). Let $K_1^0 = (P_1^0, O_1^0)$ and $K_2^0 = (P_2^0, O_2^0)$ be two players. Then, an *admissible debate* Δ is a sequence of claims: $[(\texttt{in}(X_0), \_), (\texttt{out}(X_0), \texttt{in}(Y_1)),$ $(\texttt{out}(Y_1), \texttt{in}(X_1)), \ldots, (\texttt{out}(X_i), \texttt{in}(Y_{i+1})), (\texttt{out}(Y_{i+1}), \texttt{in}(X_{i+1})), \ldots]$ such that

(a) $(\texttt{in}(X_0), \_)$ is the initial claim by K_1^0 where $X_0 \in Args(P_1^0)$.
(b) $(\texttt{out}(X_0), \texttt{in}(Y_1))$ is a claim by K_2^1 where $K_2^1 = rev(K_2^0, X_0) = (P_2^1, O_2^1)$ and $Y_1 \in Args(P_2^1 \cup O_2^1)$.
(c) $(\texttt{out}(Y_{i+1}), \texttt{in}(X_{i+1}))$ is a claim by K_1^{i+1} where $K_1^{i+1} = rev(K_1^i, Y_{i+1}) = (P_1^{i+1}, O_1^{i+1})$ and $X_{i+1} \in Args(P_1^{i+1} \cup O_1^{i+1})$ $(i \geq 0)$.

(d) $(\texttt{out}(X_i), \texttt{in}(Y_{i+1}))$ is a claim by K_2^{i+1} where $K_2^{i+1} = rev(K_2^i, X_i) = (P_2^{i+1}, O_2^{i+1})$ and $Y_{i+1} \in Args(P_2^{i+1} \cup O_2^{i+1})$ $(i \geq 0)$.
(e) for each $(\texttt{out}(U), \texttt{in}(V))$, V defeats U.
(f) for each $\texttt{out}(Z)$ in a claim by K_1^i (resp. K_2^i), there is $\texttt{in}(Z)$ in a claim by K_2^j such that $j \leq i$ (resp. K_1^j such that $j < i$).
(g) both $\bigcup_{i\geq 0}\{ X_i \mid X_i \subseteq P_1^0\}$ and $\bigcup_{j\geq 1}\{ Y_j \mid Y_j \subseteq P_2^0\}$ are conflict-free.

Let Γ_n $(n \geq 0)$ be any claim. A *debate game* Δ (*for an argument* X_0) is an admissible debate between two players $[\Gamma_0, \Gamma_1, \ldots]$ where the initial claim is $\Gamma_0 = (\texttt{in}(X_0), \_)$ and $\Gamma_m \neq \Gamma_{m+2k}$ $(m \geq 0; k > 0)$. A debate game Δ for an argument X_0 *terminates* with Γ_n if $\Delta = [\Gamma_0, \Gamma_1, \ldots, \Gamma_n]$ is an admissible debate and there is no claim Γ_{n+1} such that $[\Gamma_0, \Gamma_1, \ldots, \Gamma_n, \Gamma_{n+1}]$ is an admissible debate. In this case, the player who makes the last claim Γ_n *wins* the game.

By definition, (a) the player K_1^0 starts a debate with the claim $\Gamma_0 = (\texttt{in}(X_0), \_)$. (b) The player K_2^0 then revises its knowledge base with X_0, and responds to the player K_1^0 with a counter-claim $\Gamma_1 = (\texttt{out}(X_0), \texttt{in}(Y_1))$ based on the revised knowledge base K_2^1. In response to Γ_1, the player K_1^1 revises its knowledge base and builds a counter-claim $\Gamma_2 = (\texttt{out}(Y_1), \texttt{in}(X_1))$. A debate continues by iterating revisions and claims ((c),(d)), and (e) in each claim an argument V of $\texttt{in}(V)$ defeats an argument U of $\texttt{out}(U)$. (f) A player can refute not only the preceding claim of the opponent player, but any previous claim of the opponent. (g) During a debate game, arguments which come from a player's own program must be conflict-free, that is, each player must be self-consistent in its claims. Note that a player K_l^i $(l = 1, 2;\ i \geq 1)$ can construct arguments using rules included in arguments O_l^i posed by the opponent player as well as rules in its own program P_l^i. This means that conclusions of arguments claimed by a player may change nonmonotonically during a game. If a player K_l^i claims $(\texttt{out}(A), \texttt{in}(B))$ which is refuted by a counter-claim $(\texttt{out}(B), \texttt{in}(C))$ by the opponent, then the player K_l^j $(i < j)$ can use rules in the argument C for building a claim. Once the player K_l^j uses rules in C, it implies that K_l^j withdraws some conclusions of the argument B previously made by K_l^i (because B is defeated by C). Thus, two different claims by the same player may conflict during a game. The condition (g) states that such a conflict is not allowed among arguments which consist of rules from a player's original program P^0. In a debate game, a player cannot repeat the same claim ($\Gamma_m \neq \Gamma_{m+2k}$), otherwise arguments may go round in circles.

A debate game is represented as a directed tree in which the root node represents the initial claim, each node represents a claim, and there is a directed edge between two nodes Γ_i and Γ_j if the former refutes the latter. Figure 1 represents a debate game $\Delta = [\Gamma_0, \Gamma_1, \ldots, \Gamma_6]$ in which the player K_1^0 makes the initial claim Γ_0, the player K_2^1 makes a counter-claim Γ_1, the player K_1^1 refutes Γ_1 by Γ_2, and the player K_2^2 refutes Γ_2 by Γ_3. At this stage, K_1^2 cannot refute Γ_3 but refutes Γ_1 by Γ_4. The player K_2^3 cannot refute Γ_4 but refutes Γ_0 by Γ_5. Then, K_1^3 refutes Γ_5 by Γ_6. The player K_2^4 cannot refute Γ_6 and other claims by the opponent. As a result, the player K_1^3 wins the game. In what follows, we simply say "a debate game" instead of "a debate game for an argument X_0" when the argument X_0 in the initial claim is clear or unimportant in the context.

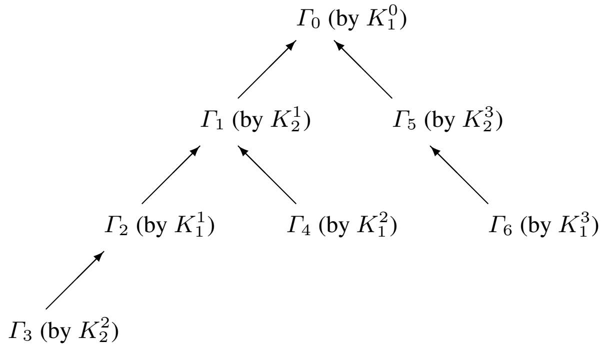

Fig. 1. Debate game

Proposition 3.1. *Let Γ be a claim of either $(\mathtt{in}(U), \_)$ or $(\mathtt{out}(V), \mathtt{in}(U))$ in a debate game. Then, U has a single answer set S such that $concl(U) = S$ and $concl(V) \not\subseteq S$.*

Proof. Since U is minimal and coherent, U has a single answer set S such that $concl(U) = S$. As U defeats V, there is a rule $r \in V$ such that the head of r is not included in S. □

Proposition 3.2. *Every debate game terminates.*

Proof. By definition, each player cannot repeat the same claim in a debate game. Since the number of minimal and coherent arguments associated with a propositional program is finite, the result holds. □

Example 3.1. Suppose a dispute between a prosecutor and a defense. First, the prosecutor and the defense have knowledge bases $K_1^0 = (P_1, \emptyset)$ and $K_2^0 = (P_2, \emptyset)$, respectively, where

$$\begin{aligned} P_1:\ & guilty \leftarrow suspect,\ motive, \\ & evidence \leftarrow witness,\ not\ \neg\, credible, \\ & suspect \leftarrow, \quad motive \leftarrow, \quad witness \leftarrow . \\ P_2:\ & \neg guilty \leftarrow suspect,\ not\ evidence, \\ & \neg\, credible \leftarrow witness,\ dark, \\ & suspect \leftarrow, \quad dark \leftarrow . \end{aligned}$$

A debate game proceeds as follows.

– First, the prosecutor K_1^0 makes the initial claim:

$(\mathtt{in}(X_0), \_)$ with $X_0 = [\,guilty \leftarrow suspect,\ motive;\ suspect \leftarrow;\ motive \leftarrow]$
("The suspect is guilty because he has a motive for the crime.")

where $X_0 \in Args(P_1)$.

- The defense revises K_2^0 into $K_2^1 = rev(K_2^0, X_0) = (P_2^1, O_2^1)$ where $P_2^1 = P_2$ and $O_2^1 = X_0$, and makes a counter-claim:

  $(\texttt{out}(X_0), \texttt{in}(Y_1))$ with $Y_1 = [\neg guilty \leftarrow suspect,\ not\ evidence;\ suspect \leftarrow]$
  ("The suspect is not guilty as there is no evidence.")

  where $Y_1 \in Args(P_2^1 \cup O_2^1)$ and Y_1 rebuts X_0.
- The prosecutor revises K_1^0 into $K_1^1 = rev(K_1^0, Y_1) = (P_1^1, O_1^1)$ where $P_1^1 = P_1$ and $O_1^1 = Y_1$, and makes a counter-claim:

  $(\texttt{out}(Y_1), \texttt{in}(X_1))$ with $X_1 = [evidence \leftarrow witness,\ not\ \neg\, credible;\ witness \leftarrow]$
  ("There is an eyewitness who saw the suspect on the night of the crime.")

  where $X_1 \in Args(P_1^1 \cup O_1^1)$ and X_1 undercuts Y_1.
- The defense revises K_2^1 into $K_2^2 = rev(K_2^1, X_1) = (P_2^2, O_2^2)$ where $P_2^2 = P_2$ and $O_2^2 = X_0 \cup X_1$. (Note that the first rule of P_2 is not removed by the revision because $P_2^1 \cup O_2^1 \cup X_1$ can defeat X_1). Then, the defense makes a counter-claim:

  $(\texttt{out}(X_1), \texttt{in}(Y_2))$ with $Y_2 = [\neg\, credible \leftarrow witness,\ dark;\ witness \leftarrow;\ dark \leftarrow]$
  ("The testimony is incredible because it was dark at night.")

  where $Y_2 \in Args(P_2^2 \cup O_2^2)$ and Y_2 undercuts X_1.
- The prosecutor revises K_1^1 into $K_1^2 = rev(K_1^1, Y_2) = (P_1^2, O_1^2)$ where $P_1^2 = P_1 \setminus \{evidence \leftarrow witness,\ not\ \neg\, credible\}$ and $O_1^2 = Y_1 \cup Y_2$. Since K_1^2 cannot refute the claim by K_2^2, the defense wins the game.

3.2 Dishonest Player

In debate games, each player constructs claims using rules included in its program or rules brought by the opponent. To defeat a claim by the opponent, a player may claim an argument which the player does not believe its conclusion.

Example 3.2. Suppose that the prosecutor in Example 3.1 has the program

$$P_1' = P_1 \cup \{\neg\, dark \leftarrow light,\ not\ broken,\quad light \leftarrow,\quad broken \leftarrow\}.$$

In response to the last claim $(\texttt{out}(X_1), \texttt{in}(Y_2))$ by the defense K_2^2, suppose that the prosecutor $K_1^2 = ({P_1'}^2, O_1^2)$ where ${P_1'}^2 = P_1'$ makes a counter-claim:

$$(\texttt{out}(Y_2), \texttt{in}(X_2)) \text{ with } X_2 = [\neg\, dark \leftarrow light,\ not\ broken;\ light \leftarrow].$$

("It was not dark because the witness saw the suspect under the light of the victim's apartment."). Then, X_2 defeats Y_2.

In Example 3.2, the prosecutor K_1^2 claims the argument X_2 but he/she does not believe its conclusion $concl(X_2)$. In fact, $\neg\, dark$ is included in no answer set of the program ${P_1'}^2 \cup Q$ for any $Q \subseteq O_1^2$. Generally, a player may behave dishonestly by concealing believed facts to justify another fact which the player wants to conclude. We classify different types of claims which may appear in a debate game.

Definition 3.5 (credible, misleading, incredible, incorrect, false claims). Let Γ be a claim of either $(\mathtt{in}(U), \_)$ or $(\mathtt{out}(V), \mathtt{in}(U))$ by a player $K_l^i = (P_l^i, O_l^i)$ $(l = 1, 2; i \geq 0)$. Also, let U^S be an argument which consists of rules in the reduct of U with respect to a set S.

- Γ is *credible* if $concl(U) \subseteq S$ for every answer set S of $P_l^i \cup Q$ for some $Q \subseteq O_l^i$ such that $P_l^i \cup Q$ is consistent and $concl(U) = concl(U^S)$.
- Γ is *misleading* if $concl(U) \subseteq S$ for every answer set S of $P_l^i \cup Q$ for some $Q \subseteq O_l^i$ such that $P_l^i \cup Q$ is consistent but $concl(U) \neq concl(U^S)$.
- Γ is *incredible* if $concl(U) \subseteq S$ for some (but not every) answer set S of $P_l^i \cup Q$ for any $Q \subseteq O_l^i$ such that $P_l^i \cup Q$ is consistent.
- Γ is *incorrect* if $concl(U) \not\subseteq S$ for any answer set S of $P_l^i \cup Q$ for any $Q \subseteq O_l^i$ such that $P_l^i \cup Q$ is consistent, and $concl(U) \cup S$ is consistent for some answer set S of $P_l^i \cup Q$ for some $Q \subseteq O_l^i$ such that $P_l^i \cup Q$ is consistent.
- Γ is *false* if $concl(U) \cup S$ is inconsistent for any answer set S of $P_l^i \cup Q$ for any $Q \subseteq O_l^i$ such that $P_l^i \cup Q$ is consistent.

A claim is called *dishonest* if it is not credible. A player K_l is *honest* in a debate game Δ if every claim made by K_l^i $(i \geq 0)$ in Δ is credible. Otherwise, K_l is *dishonest*.

During a game, a player K_l^i constructs an argument U using some rules $Q \subseteq O_l^i$. Then, U has the answer set which coincides with $concl(U)$ (Proposition 3.1), but this does not always imply that $concl(U)$ is a subset of an answer set of P_l^i.

Proposition 3.3. *Every claim in a debate game is classified as one of the five types of claims of Definition 3.5.*

Example 3.3.

- Given $K_1 = (\{\, p \leftarrow not\, q \,\}, \emptyset)$, the claim $\Gamma_1 = (\mathtt{in}([\, p \leftarrow not\, q \,]), \_)$ is credible.
- Given $K_2 = (\{\, p \leftarrow not\, q, \quad p \leftarrow q, \quad q \leftarrow \,\}, \emptyset)$, the claim $\Gamma_2 = (\mathtt{in}([\, p \leftarrow not\, q \,]), \_)$ is misleading.
- Given $K_3 = (\{\, p \leftarrow not\, q, \quad q \leftarrow not\, p \,\}, \emptyset)$, the claim $\Gamma_3 = (\mathtt{in}([\, p \leftarrow not\, q \,]), \_)$ is incredible.
- Given $K_4 = (\{\, p \leftarrow not\, q, \quad q \leftarrow \,\}, \emptyset)$, the claim $\Gamma_4 = (\mathtt{in}([\, p \leftarrow not\, q \,]), \_)$ is incorrect.
- Given $K_5 = (\{\, p \leftarrow not\, \neg p, \quad \neg p \leftarrow \,\}, \emptyset)$, the claim $\Gamma_5 = (\mathtt{in}([\, p \leftarrow not\, \neg p \,]), \_)$ is false.

In Example 3.3, Γ_1 is credible because $concl([p \leftarrow not\, q]) = \{p\}$ coincides with the answer set of the program $\{p \leftarrow not\, q\}$ in K_1. By contrast, Γ_2 is misleading because for $U = \{p \leftarrow not\, q\}$ it becomes $U^S = \emptyset$ by the answer set $S = \{p, q\}$ of the program in K_2, so that $concl(U) \neq concl(U^S)$. That is, a misleading claim does not use rules in a proper manner to reach conclusions. Γ_3 is incredible because p is included in some but not in every answer set of the program in K_3. Γ_4 is incorrect because p is included in no answer set of the program in K_4. Γ_5 is false because $\neg p$ is included in every answer set of the program in K_5.

The existence of dishonest claims is due to the nonmonotonic nature of a program. A player $K = (P, O)$ is *monotonic* if P contains no default literal. In this case, the following result holds.

Proposition 3.4. *Let Δ be a debate game between two monotonic players. Then, every claim in Δ is credible.*

Proof. Let Γ be a claim of either $(\mathtt{in}(U), \_)$ or $(\mathtt{out}(V), \mathtt{in}(U))$ by a player $K = (P, O)$. By $U \in Args(P \cup O)$, $U \subseteq P \cup Q$ for some $Q \subseteq O$ such that $P \cup Q$ is consistent. Since U is an argument associated with $P \cup Q$, $concl(U) \subseteq S$ holds for the answer set S of $P \cup Q$ by Proposition 2.1. By $U^S = U$, $concl(U) = concl(U^S)$. Hence, Γ is credible. □

Generally, it is unknown which player wins a debate game. In real life, a player who is more knowledgeable than another player is likely to win a debate. The situation is formulated as follows.

Proposition 3.5. *Let Δ be a debate game between two players $K_1^0 = (P_1, \emptyset)$ and $K_2^0 = (P_2, \emptyset)$. If K_1 (resp. K_2) is honest and $P_2 \subset P_1$ (resp. $P_1 \subset P_2$), then K_1^i (resp. K_2^i) $(i \geq 1)$ wins the game.*

Proof. Suppose that K_1 is honest and $P_2 \subset P_1$. Let Γ_m be a honest claim of either $(\mathtt{in}(X_0), \_)$ or $(\mathtt{out}(Y_i), \mathtt{in}(X_i))$ by $K_1^i = (P_1^i, O_1^i)$ $(i \geq 0)$ in Δ. By $O_1^i \subseteq P_1$, $P_1^i \subseteq P_1$ and every rule in $P_1 \setminus P_1^i$ is undercut by some argument $A \in Args(P_1)$ and A is not defeated by any argument in $Args(P_1)$. Then, every rule in $P_1 \setminus P_1^i$ is eliminated in the reduct P_1^S for any answer set S of P_1. Then, P_1^i and P_1 have the same answer sets. Thus, $concl(X_i) \subseteq S$ for every answer set S of P_1. Suppose that K_2^{i+1} makes a counter-claim $\Gamma_{m+1} = (\mathtt{out}(X_i), \mathtt{in}(Y_{i+1}))$, and K_1^{i+1} cannot refute Γ_{m+1} by any honest claim. In this case, P_1 has no rule to defeat Y_{i+1}. By $P_2 \subset P_1$, $Y_{i+1} \subset P_1$. Then, $concl(Y_{i+1}) \subseteq S$ for every answer set S of P_1. Since Y_{i+1} defeats X_i, either (i) Y_{i+1} undercuts X_i or (ii) Y_{i+1} rebuts X_i but X_i does not undercut Y_{i+1}. In either case, $concl(X_i) \not\subseteq S$ for any answer set S of P_1. This contradicts the fact that $concl(X_i) \subseteq S$. Hence, K_1^{i+1} can refute Γ_{m+1} by a honest claim in Δ. As such, every claim by K_2^i is honestly refuted by K_1^i. Hence, K_1^i wins the game. When K_2 is honest and $P_1 \subset P_2$, it is shown in a similar way that K_2^i wins the game. □

Proposition 3.5 presents that if a player K has information more than another player, K has no reason to behave dishonestly to win a debate. In fact, if a more informative player K behaves dishonestly, K may lose a game.

Example 3.4. Consider two players $K_1^0 = (P_1, \emptyset)$ and $K_2^0 = (P_2, \emptyset)$ where $P_1 = \{ p \leftarrow not\, q, \quad q \leftarrow \}$ and $P_2 = \{ q \leftarrow \}$. Then, $P_2 \subset P_1$. Suppose a debate game between K_1^0 and K_2^0 such that

$$K_1^0 : \ \Gamma_0 = (\mathtt{in}([\, p \leftarrow not\, q\,]), \_)$$
$$K_2^1 : \ \Gamma_1 = (\mathtt{out}([\, p \leftarrow not\, q]), \mathtt{in}([\, q \leftarrow])).$$

The claim Γ_0 by K_1^0 is incorrect because p is included in no answer set of P_1. Since the player K_1^1 cannot refute Γ_1, K_2^1 wins the game.

A player has an incentive to build a dishonest claim if the player cannot build a honest counter-claim in response to the claim by the opponent. Then, our next question is how a player effectively uses dishonest claims as a tactic to win a debate. We first show that among different types of dishonest claims, misleading claims are useless for the purpose of winning a debate.

Proposition 3.6. *Let Δ be a debate game between two players $K_1^0 = (P_1, \emptyset)$ and $K_2^0 = (P_2, \emptyset)$.*

1. *If the initial claim $\Gamma_0 = (\mathtt{in}(X_0), \_)$ by K_1^0 is misleading, there is a credible claim $\Gamma_0' = (\mathtt{in}(X), \_)$ by K_1^0 such that $concl(X) = concl(X_0)$.*
2. *If a claim $\Gamma_k = (\mathtt{out}(V), \mathtt{in}(U))$ by a player K_l^i $(l = 1, 2;\ i \geq 1)$ is misleading, there is a credible claim $\Gamma_k' = (\mathtt{out}(V), \mathtt{in}(W))$ by K_l^i such that $concl(W) = concl(U)$.*

Proof. (1) Since $concl(X_0) \subseteq S$ for every answer set S of P_1, there is a set $X \subseteq P_1$ of rules such that $concl(X_0) = concl(X) = concl(X^S)$. Selecting a minimal set X of rules satisfying the conditions of Definition 2.1, the result holds. (2) Since $concl(U) \subseteq S$ for every answer set S of $P_l^i \cup Q$ for some $Q \subseteq O_l^i$ such that $P_l^i \cup Q$ is consistent, there is a set $W \subseteq P_l^i \cup Q$ of rules such that $concl(U) = concl(W) = concl(W^S)$. Selecting a minimal set W of rules satisfying the conditions of Definition 2.1, the result holds. □

Thus, dishonest claims which are effectively used for the purpose of winning a debate game are either incredible, incorrect or false claims. Once a player makes a dishonest claim in a game, however, it will restrict what the player can claim later in the game. In Example 3.3, the player K_4 who makes the incorrect claim Γ_4 cannot subsequently use the believed-true fact $q \leftarrow$ which conflicts with Γ_4 (Definition 3.4(g)). To keep conflict-freeness of a player's claims in a game, dishonest claims would restrict the use of believed-true rules in later claims and may result in a net loss of freedom in playing the game. With this reason, it seems reasonable to select a dishonest claim only if there is no choice among honest claims. Comparing different types of dishonest claims, it is considered that incredible claims are preferred to incorrect claims, and incorrect claims are preferred to false claims. If a claim $\Gamma = (\mathtt{out}(V), \mathtt{in}(U))$ is incredible, the player does not *skeptically* believe the conclusion of U but *credulously* believes the conclusion of U. If Γ is incorrect, the player does not credulously believe the conclusion of U but the conclusion is consistent with the player's belief. If Γ is false, on the other hand, the conclusion of U is inconsistent with the player's belief. Thus, the degree of truthfulness (against the belief state of a player) decreases from incredible claims to incorrect claims, and from incorrect claims to false claims (Figure 2). Generally, a dishonest claim deviates from the reality as believed by a player, and a claim which increases such deviation is undesirable for a player because it increases a chance of making the player's claims conflict. A player wants to keep claims close to its own belief as much as possible, so the best-practice strategy for a debate game is to firstly use credible claims, secondly use incredible ones, thirdly use incorrect ones, and finally use false ones to refute the opponent.

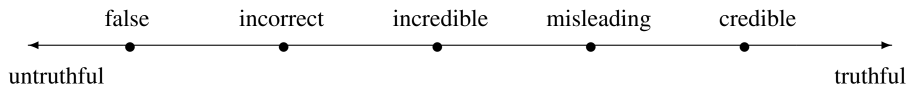

Fig. 2. Degree of truthfulness

4 Discussion

A formal argumentation framework has been used for modelling dialogue games or discussion games ([11, 13, 3]; and references therein). However, most of the studies use abstract argumentation and pay much attention on identifying acceptable arguments based on the topological nature of dialectical graphs associated with dialogues. On the other hand, the content of dialogue is important in human communication. Participants in debates are interested in why one's argument is defeated by the opponent, whether arguments made by the opponent are logically consistent, which arguments made by the opponent are unacceptable, and so on. In debate games proposed in this paper, each player can see the *inside* of the arguments in claims made by the opponent. As a result, a player can judge whether a counter-claim made by the opponent is grounded on evidences, and whether claims made by the opponent are consistent throughout a debate. Moreover, a player can obtain new information from arguments posed by the opponent.

In AI agents are usually assumed to be honest and little attention has been paid for representing and reasoning with dishonesty. In real-life debates, however, it is a common practice for one to misstate their beliefs or opinions [19]. In formal argumentation, [15] characterizes dishonest agents in a game-theoretic argumentation mechanism design and [18] introduces dishonest arguments in a debate game. These studies use the abstract argumentation framework and do not show how to construct dishonest arguments from the underlying knowledge base. In this paper, we show how to build dishonest arguments from a knowledge base represented by a logic program. Using arguments associated with logic programs, we argue that at least four different types of dishonest claims exist. In building dishonest claims, default literals play an important role—concealing known rules or facts could produce conclusions which are not believed by a player. Proposition 3.4 shows an interesting observation that players cannot behave dishonestly without default assumption. Dishonest reasoning in logic programs is introduced by [16] in which the notion of *logic programs with disinformation* is introduced and its computation by abductive logic programming is provided. An application of dishonest reasoning to multiagent negotiation is provided by [17] in which agents represented by abductive logic programs misstate their bargaining positions to gain one's advantage over the other. The current study shows yet another application of dishonest reasoning in argumentation-based logic programming.

Prakken and Sartor [12] introduce *dialogue trees* in order to provide a proof theory of argumentation-based extended logic programs. A dialogue tree consists of nodes representing arguments by the proponent and the opponent, and edges representing attack relations between arguments. Given the initial argument of the proponent at the root node of a dialogue tree, the opponent attacks the argument by a counterargument if any (called a *move*). Two players move in turn and one player wins a dialogue if the other player run out of moves in a tree. Comparing dialogue trees with debate games, a dialogue tree is constructed by arguments associated with a *single* extended logic program. In debate games, on the other hand, two players have different knowledge bases and build arguments associated with them. Dialogue trees are introduced to provide a proof theory of argumentation-based logic programs, and they do not intend to provide a formal theory of dialogues between two players. As a result, dialogue trees do no have mechanisms of revision and dishonest reasoning. Fan and Toni [6] propose a formal model for argumentation-based dialogues between agents. They use *assumption-based argumentation* (ABA) [5] for this purpose. In ABA arguments are built from rules and supported by assumptions, and attacks against arguments are directed at the assumptions supporting the arguments, and are provided by arguments for the contrary of assumptions. In their dialogue model, agents can utter claims to be debated, rules, assumptions, and contraries. A dialogue between the proponent and the opponent constructs a dialectical tree which represents moves by agents during a dialogue and outcomes. In their framework, two agents share a common ABA framework and assumed to have a common background knowledge. With this setting, an agent cannot behave dishonestly as one cannot keep some information from the other.

5 Conclusion

The contributions of this paper are mainly twofold. First, we developed debate games using a non-abstract argumentation framework associated with logic programming. We applied argumentation-based extended logic programs to formal modelling of dialogue games. Second, we showed an application of dishonest reasoning in argumentation-based logic programming. Debate games introduced in this paper realize dishonest reasoning by players using nonmonotonic nature of logic programs. To the best of our knowledge, there is no formal dialogical system which can deal with argumentative reasoning, belief revision and dishonest reasoning in a uniform and concrete manner. The current study contributes to a step toward integrating logic programming and formal argumentation.

The proposed framework will be extended in several ways. In real-life debates, players may use *assumptions* in their arguments. Assumptions are also used for constructing arguments in an assumption-based argumentation framework [5]. Arguments considered in this paper use assumptions in the form of default literals. To realize debate games in which players can also use objective literals as assumptions, we can consider a non-abstract assumption-based argumentation framework associated with *abductive logic programs*. In this framework, an argument associated with an abductive logic program can contain *abducibles* as well as rules in a program. A player can claim an argument containing abducibles whose truthfulness are unknown. This is an another type of dishonest claims called *bullshit* [7]. To realize debate games, we are now implementing a prototype system of debate games based on the abstract argumentation framework [16]. We plan to extend the system to handle non-abstract arguments associated with extended logic programs.

References

1. Bondarenko, A., Dung, P. M., Kowalski, R. and Toni, F. An abstract, argumentation-theoretic approach to default reasoning. *Artificial Intelligence* 93(1,2):63–101, 1997.
2. Caminada, M. and Wu, Y. On the limitation of abstract argumentation. In: *Proceedings of the 23rd Benelux Conference on Artificial Intelligence (BNAIC)*, Gent, Belgium, 2011.
3. Caminada, M. Grounded semantics as persuasion dialogue. In: *Proceedings of the 4th International Conference on Computational Models of Argument (COMMA), Frontiers in Artificial Intelligence and Applications* 245, IOS Press, pp. 478–485, 2012.
4. Dung, P. M. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence* 77(2):321–357, 1995.
5. Dung, P. M., Kowalski, R. A. and Toni, F. Assumption-based argumentation. In: I. Rahwan and G. R.. Simari (eds.), *Argumentation in Artificial Intelligence*, pp. 199–218, Springer, 2009.
6. Fan, X. and Toni, F. Assumption-based argumentation dialogues. In: *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI-11)*, pp. 198–203, 2011.
7. Frankfurt, H. G. *On Bullshit*. Princeton University Press, 2005.
8. García, A. J. and Simari, G. R. Defeasible logic programming: an argumentative approach. *Theory and Practice of Logic Programming* 4(1,2):95–138, 2004.
9. García, A. J., Dix, J. and Simari, G. R. Argumentation-based logic programming. In: I. Rahwan and G. R.. Simari (eds.), *Argumentation in Artificial Intelligence*, pp. 153–171, Springer, 2009.
10. Gelfond, M. and Lifschitz, V. Logic programs with classical negation. In: *Proceedings of the 7th International Conference on Logic Programming (ICLP)*, MIT Press, pp. 579–597, 1990.
11. Parsons, S., Wooldridge, M. and Amgoud, L. Properties and complexity of some formal inter-agent dialogues. *Journal of Logic and Computation* 13(3):347–376, 2003.
12. Prakken, H. and Sartor, G. Argument-based extended logic programming with defeasible priorities. *Journal of Applied Non-Classical Logics* 7(1):25–75, 1997.
13. Prakken, H. Coherence and flexibility in dialogue games for argumentation. *Journal of Logic and Computation* 15(6):1009–1040, 2005.
14. Prakken, H. On the nature of argument schemes. In: *Dialectics, Dialogue and Argumentation, An Examination of Douglas Walton's Theories of Reasoning and Argument*, pp. 167–185, College Pub, 2010.
15. Rahwan, I., Larson, K., and Tohmé, F. A characterisation of strategy-proofness for grounded argumentation semantics. In: *Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI-09)*, pp. 251–256, 2009.
16. Sakama, C. Dishonest reasoning by abduction. In: *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI-11)*, pp. 1063–1068, 2011.
17. Sakama, C., Son, T. C. and Pontelli, E. A logical formulation for negotiation among dishonest agents. In: *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 1069–1074, 2011.
18. Sakama, C. Dishonest arguments in debate games. In: *Proceedings of the 4th International Conference on Computational Models of Argument (COMMA), Frontiers in Artificial Intelligence and Applications* 245, IOS Press, pp. 177–184, 2012.
19. Schopenhauer, A. *The Art of Controversy*. Originally published in 1896 and is translated by T. Bailey Saunders, Cosimo Classics, New York, 2007.
20. Schweimeier, R. and Schroeder, M. A parameterized hierarchy of argumentation semantics for extended logic programming and its application to the well-founded semantics. *Theory and Practice of Logic Programming* 5(1,2), pp. 207–242, 2005.