Katarzyna Budzyńska
Cardinal Stefan Wyszyński University in Warsaw

Magdalena Kacprzak
Bialystok University of Technology

MODEL CHECKING OF PERSUASION IN MULTI-AGENT SYSTEMS

Abstract: The paper presents the method of model checking applied to verification of persuasive inter-agent communication. The model checker Perseus is designed on the basis of a logic of actions and graded beliefs $\mathcal{AG}_n$ introduced by Budzyńska and Kacprzak. The software tool makes it possible to semantically verify satisfaction of $\mathcal{AG}_n$ formulas which describe different properties of a multi-agent system in a given model, and to perform parametric verification that enables searching for answers to questions about these properties.
Keywords: model-checking, modal logic, multi-agent systems, persuasive arguments, dialogue games

1. Introduction

The common method of verification of multi-agent systems is model checking technique (see e.g. [9, 11, 15]). The paper presents how this method can be applied to examine the properties of inter-agent persuasive communication. A software tool designed to verify persuasion in multi-agent systems is called **Perseus** [6]. It is built upon the Logic of Actions and Graded Beliefs $\mathcal{AG}_n$ [2]. The Perseus model checker offers two main options of investigation. First, it can semantically *verify satisfaction of formulas* of the $\mathcal{AG}_n$ language which describe properties of persuasion in a given model. In this case, the tool performs the standard model checking. Second, it can *search for answers to questions* of three kinds – questions about the degrees of uncertainty, questions about the sequence of arguments that should be executed and questions about the agents participating in the process of persuasion. In this case, the tool uses the new method of parametric verification introduced by Budzyńska, Kacprzak and Rembelski [6].

The most typical kind of inter-agent persuasive communication is **persuasion dialogue** [16]. It is a dialogue of which initial situation is a conflict of opinion and the aim is to resolve this conflict and thereby influence

ISBN 978–83–7431–305–6 ISSN 0860-150X

the change of agents' beliefs or commitments (i.e. beliefs declared by an agent). This rises a general question about what impact on beliefs and commitments has a given persuasion. In consequence, we may ask about the degree and the scenario of belief and commitment changes, the factors that influence them, the strength of different types of arguments and their arrangements, the credibility of persuader, the strategies that allow the victory in a dialogue game etc.

Formal systems for dialogues are often built in a game-theoretic style, i.e. speech acts performed in a dialogue are treated as moves in a **dialogue game** and rules for their appropriateness are formulated as rules of the game (see [13] for an overview). In this paper, we present the application of verification methods to a persuasion dialogue system introduced by Prakken [12]. A *dialogue system* for argumentation is defined as a pair $(\mathcal{L};\mathcal{D})$, where $\mathcal{L}$ is a logic for defeasible argumentation and $\mathcal{D}$ is a dialogue system proper. A *logic for defeasible argumentation* $\mathcal{L}$ is a tuple $(L_t, R, Args, \rightarrow)$, where L_t (the topic language) is a logical language, R is a set of inference rules over L_t, $Args$ (the arguments) is a set of AND-trees of which the nodes are in L_t and the AND-links are inferences instantiating rules in R, and $\rightarrow$ is a binary relation of defeat defined on $Args$. For any argument A, $prem(A)$ is the set of leaves of A (its premises) and $conc(A)$ is the root of A (its conclusion).

A *dialogue system proper* is a triple $\mathcal{D} = (L_c; P; C)$ where L_c (the communication language) is a set of locutions (utterances), P is a protocol for L_c, and C is a set of effect (commitment) rules of locutions in L_c, specifying the effects of the locutions on the participants' commitments. Agents may perform five types of **communication moves** (dialogue actions) in a dialogue game: $claim(\alpha)$ – the speaker asserts that α is the case, $why(\alpha)$ – the speaker challenges α and asks for reasons why it would be the case, $concede(\alpha)$ – the speaker admits that α is the case, $retract(\alpha)$ – the speaker declares that he is not committed (any more) to α, $argue(A)$ – the speaker provides an argument A. Every utterance from L_c can influence participants commitments. $C(d, i)$ denotes a player i's **commitments** at a stage of a dialogue d. In a dialogue game, an agent may adopt a strategy to achieve a desired goal which could be to make its adversary become committed to the agent's claim (see e.g. [10], [14], [1]).[1]

The remainder of this paper is organized as follows. Section 2 presents the model checking technique. Section 3 shows the $\mathcal{AG}_n$ logic. Section 4

[1] For present purposes a more detailed definitions are not needed. For the full details the reader is referred to [12].

presents the model checker Perseus. Finally, Section 5 discusses the types of properties of persuasion that may be verified by the Perseus system.

2. The model checking technique

The commonly applied method which allows for semantic verification of multi-agent systems and thus their communication is *model checking*. Model checking is considered as a one of the most spectacular applications of computer science. Testing the correctness of a given software system under their correctness conditions (i.e. specification) is crucial task which must be solved before this system will find the application in commercial exploitation. Potential non detected errors in programs like decision support for air traffic control or quality control can have financial effects or could be a threat to health or life of people. Computer simulations and computations allow for avoiding very expensive and time-consuming experiments. The main problem which appears in automatic verification is state explosion problem. There are diverse methods for dealing with this problem. However, in practice, the most effective results bring application of symbolic methods based on satisfiability of propositional formulas.

Model checking is the decision problem that takes either a program P or its more extensive representation $\mathcal{M}_{\mathcal{P}}$ (as a transition system), and a logical specification α which truth value should be determined. Then, the model checking problem for P asks whether $\mathcal{M}_P, s \models \alpha$ for a given state s of the model $\mathcal{M}_{\mathcal{P}}$. The model checking experiments executed by Walton have focused on the termination of Multi-Agent Protocol (in particular, auction protocols) [15]. The key difference between this approach and our research is that we want to verify the behavior of a whole multi-agent system, and not the properties of a language. We focus on the question what effects the dialogue actions can bring for a system, rather than the questions such as e.g. if an agent sends a sincere message. For instance, we want to ask what is the most effective (e.g. the shortest) sequence of speech acts that enables an agent i to achieve his goal (e.g. to persuade his opponent $\bar{i}$ to accept i's topic t).

For verification of dialogue games we use the model checker called Perseus [7]. Verification of epistemic and temporal formulas can be performed by other software tools like VerICS [9] or MCMAS [11], however, the important advantage of the Perseus system is that it performs not only pure model checking, but also parametric verification. It means that, given a model of a system, it can test automatically whether this model meets a given specification or given an input expressions with unknowns it can determine for which values of these unknowns the obtained logical formula is true in this model. Furthermore, Perseus is not limited to verification of formulas with epistemic and dynamic operators, but is already designed and adjusted to analyze phenomena related to agents persuasive actions with the use of graded doxastic modalities and with probabilistic modalities.

3. Logic of actions and graded beliefs $\mathcal{AG}_n$

In this section, we present the Logic of Actions and Graded Beliefs $\mathcal{AG}_n$ [2] extended with the components needed for representation of Prakken's dialogue system [3] and probabilistic beliefs [7].

3.1. Formal syntax of the language

Let $Agt = \{1, \ldots, n\}$ be a set of names of *agents*, V_0 be a set of *propositional variables*, Π_0^{ph} a set of *physical actions*, and Π_0^v a set of *verbal actions*. Further, let ; denote a programme connective which is a sequential composition operator. It enables to compose *schemes of programs* defined as finite sequences of atomic actions: $a_1; \ldots; a_k$. Intuitively, the program $a_1; a_2$ for $a_1, a_2 \in \Pi_0^{ph}$ means "Do a_1, then do a_2". The set of all schemes of physical programs we denote by Π^{ph}. In similar way, we define a set Π^v of schemes of programs constructed over Π_0^v. The set of F all *well-formed expressions* of the extended $\mathcal{AG}_n$ is given by the following Backus-Naur form:

$$\alpha ::= p|\neg\alpha|\alpha \vee \alpha|M_i^k\alpha|\mathbf{P}_i(\alpha) \geq q|\Diamond(i : P)\alpha|\mathbf{C}_i\alpha|$$

$$\langle i\rangle \mathrm{G}\alpha|\langle i\rangle^q \mathrm{G}\alpha|\langle\langle i\rangle\rangle \mathrm{G}\alpha|\langle i\rangle_f \mathrm{G}\alpha|\langle i\rangle_f^q \mathrm{G}\alpha|\langle\langle i\rangle\rangle_f \mathrm{G}\alpha|\langle\langle i\rangle\rangle \mathrm{X}\alpha|\langle\langle i\rangle\rangle \alpha \mathrm{U}\beta,$$

where $p \in V_0$, $k \in \mathbb{N}$, $q \in [0,1]$, $i \in Agt$, $P \in \Pi^{ph}$ or $P \in \Pi^v$ and f is a strategy.

We use also the following abbreviations:

- $B_i^k\alpha$ for $\neg M_i^k\neg\alpha$,
- $M!_i^k\alpha$ where $M!_i^0\alpha \Leftrightarrow \neg M_i^0\alpha$, $M!_i^k\alpha \Leftrightarrow M_i^{k-1}\alpha \wedge \neg M_i^k\alpha$, if $k > 0$,
- $M!_i^{k_1,k_2}\alpha$ for $M!_i^{k_1}\alpha \wedge M!_i^{k_2}(\alpha \vee \neg\alpha)$,
- $\Box(i : P)\alpha$ for $\neg\Diamond(i : P)\neg\alpha$.
- $\mathbf{P}_i(\omega) > q$, $\mathbf{P}_i(\omega) = q$, $\mathbf{P}_i(\omega) < q$, $\mathbf{P}_i(\omega) \leq q$ defined from $\mathbf{P}_i(\omega) \geq q$ in the classical way.

3.2. Some intuitions

Intuitions concerning the most frequently used $\mathcal{AG}_n$ formulas is described below. The **belief formula** $M!_i^{k_1,k_2}\alpha$ says that an agent i considers k_2 doxastic alternatives and in k_1 of them α holds. In other words, the agent i believes α with degree $\frac{k_1}{k_2}$. The **probability formula** $\mathbf{P}_i(\alpha) \geq q$ informally says that the agent i believes with probability higher or equal to q that α holds. The **commitment formula** $\mathbf{C}_i\alpha$ says that a claim α is a commitment of the agent i. The $\mathcal{AG}_n$ language contains also **dynamic formulas** which allow the representation of physical actions which modify states of a model, and verbal actions which change a whole model. It means that physical actions influence agents' environment while verbal actions influence their perception of the environment. The formula $\Diamond(i : P)\alpha$ says that after executing a sequence of (persuasion) actions P by the agent i, condition α may hold. Finally we explain the meaning of **strategic formulas**. $\langle i\rangle\mathrm{G}\alpha$ says that there *exists* a strategy of i and there *exists* a computation consistent with this strategy such that in all states of this computation α is true. The formula $\langle i\rangle^q\mathrm{G}\alpha$ expresses that agent i has such a strategy with degree of success which is higher than q. $\langle\langle i\rangle\rangle\mathrm{G}\alpha$ expresses that there *exists* such a strategy which *always* leads to success regardless of the other agents' actions, i.e. the agent i has a *winning strategy.* The operator $\langle i\rangle_f\mathrm{G}\alpha$ says that for the strategy f there *exists* a computation consistent with this strategy such that in all states of this computation α is true. The operator $\langle i\rangle_f^q\mathrm{G}\alpha$ expresses that the strategy f of agent i has degree higher or equal to q. The last operator we use is $\langle\langle i\rangle\rangle_f\mathrm{G}\alpha$ expresses that the strategy f is a winning strategy.

3.3. Kripke model

All $\mathcal{AG}_n$ formulas are interpreted over the semantic model which is an extended Kripke structure.

Definition 1

By a *semantic model* we mean a Kripke structure

$$\mathrm{M} = (Agt, S, RB, I^{ph}, P, C, v)$$

where

- Agt is a set of agents' names,
- S is a non-empty set of states (the universe of the structure),
- $RB : Agt \longrightarrow 2^{S\times S}$ is a doxastic function which assigns to every agent a binary relation,

- $I^{ph} : \Pi_0^{ph} \longrightarrow (Agt \longrightarrow 2^{S\times S})$ is an interpretation of physical actions,
- $P : Agt \to (S \times S \to [0,1])$ is a probability (partial) function defined for every $i \in Agt$ and $(s,s') \in RB(i)$ such that for every agent $i \in Agt$ and $s \in S$, $\sum_{\{s':(s,s')\in RB(i)\}} P(i)(s,s') = 1$,
- $C : S \times Agt \longrightarrow 2^F$ is a commitment function,
- v is a valuation function, $v : S \longrightarrow \{\mathbf{0},\mathbf{1}\}^{V_0}$.

Function I^{ph} can be extended in a simple way to define interpretation of any program scheme. Let $I^{ph}_{\Pi^{ph}} : \Pi^{ph} \longrightarrow (Agt \longrightarrow 2^{S\times S})$ be a function such that $I^{ph}_{\Pi^{ph}}(P_1;P_2)(i) = I^{ph}_{\Pi^{ph}}(P_1)(i) \circ I^{ph}_{\Pi^{ph}}(P_2)(i) = \{(s,s') \in S \times S : \exists_{s''\in S} \ ((s,s'') \in I^{ph}_{\Pi^{ph}}(P_1)(i) \text{ and } (s'',s') \in I^{ph}_{\Pi^{ph}}(P_2)(i))\}$ for $P_1, P_2 \in \Pi^{ph}$ and $i \in Agt$.

Further, we define a function I^v which is an interpretation for verbal actions.

Definition 2

Let $\mathcal{CM}$ be a class of models and $\mathcal{CMS}$ be a set of pairs (M, s) where $\mathrm{M} \in \mathcal{CM}$ and s is a state of the model M. An interpretation for verbal actions I^v is a function:

$$I^v : \Pi_0^v \longrightarrow (Agt \longrightarrow 2^{\mathcal{CMS}\times\mathcal{CMS}}).$$

We allow different verbal actions to be executed during persuasion process. Therefore, no restrictions on I^v are assumed in the general definition. An interpretation for verbal actions will obtain different specifications depending on the type of actions and the applications of the formal model. Moreover, verbal actions do not have to convey a true information. This is particularly important, if we want to use the formal framework to represent persuasion. Agents may try (successfully or not) to influence others using false messages (since they are insincere or have incomplete knowledge). Thus, we assume that I^v does not depend on the truth or falsity conditions of the announced formula. Interpretation $I^v_{\Pi^v}$ of all verbal programs is defined similarly to the function $I^{ph}_{\Pi^{ph}}$.

Before the semantics of several kinds of strategy modalities will be defined, we need to formalize the notion of a *strategy*. Let

$$\delta : \mathcal{CMS} \times Agt \to 2^{(\Pi^{Ph}\cup\Pi^v)}$$

be a function mapping a triple consisting of a model, a state of this model and an agent to a set of actions. These actions are assumed to be actions which the agent can perform next. In fact this function determines *transition* *function*, i.e. indicates models and states of these models reachable from a given state of a given model by a given agent.

Definition 3

A **computation** is a sequence

$$(\mathrm{M}_0, s_0), (\mathrm{M}_1, s_1), (\mathrm{M}_2, s_2), \ldots$$

such that for every $k \geq 0$, there exists an action a_k and an agent i_k such that $a_k \in \delta(\mathrm{M}_k, s_k, i_k)$ and $((\mathrm{M}_k, s_k), (\mathrm{M}_{k+1}, s_{k+1})) \in I(a_k, i_k)$ where I is the interpretation of action a_k, i.e., $I = I^{ph}$ if a_k is a physical action and $I = I^v$ if a_k is a verbal action.

Intuitively by a computation we mean a sequence of pairs (M_k, s_k), a model and a state of this model, such that for every position k, $(\mathrm{M}_{k+1}, s_{k+1})$ is a result of performing an action a_k by an agent i_k at the state s_k of the model M_k.

Definition 4

By a **strategy** for an agent i we call a mapping $f_i : M^{<\infty} \rightarrow 2^M$ which assigns to every finite dialogue $d = m_0, m_1, \ldots, m_k \in M^{<\infty}$ in which it is i's turn, i.e., $i \in T(d)$, a move $m \in M$ such that $m \in Pr(d)$.

In other words, a strategy function returns a move which is allowed by the protocol P after a dialogue d where i is to move. We say that a dialogue $d = m_0, m_1, \ldots$ is consistent with a strategy f_i iff for every $k \geq 1$ if $i = pl(m_k)$ then $m_k \in f_i(m_0, \ldots, m_{k-1})$ and for $k = 0$ if $i = pl(m_k)$ then $m_k \in f_i(\emptyset)$, i.e., every move of agent i is determined by the function f_i.

Next, we define the ***outcomes*** of f_i, i.e., a set of computations which are consistent with this strategy. Let $\lambda = (\mathrm{M}_0, s_0), (\mathrm{M}_1, s_1), (\mathrm{M}_2, s_2), \ldots$ be a computation, then

$$\lambda \in out((\mathrm{M}, s), f_i) \quad \text{iff} \quad (\mathrm{M}_0, s_0) = (\mathrm{M}, s) \quad \text{and}$$

there exists a dialogue $d = m_0, m_1, \ldots$ consistent with f_i such that

$$\text{for every } k \geq 0, \quad s(m_k) \in \delta(\mathrm{M}_k, s_k, pl(m_k))$$

and

$$((\mathrm{M}_k, s_k), (\mathrm{M}_{k+1}, s_{k+1})) \in I(pl(m_k), s(m_k)).$$

Intuitively, a computation is consistent with a strategy if it is determined by a dialogue consistent with the strategy.

3.4. Interpretation of formulas

The *semantics* of formulas of the $\mathcal{AG}_n$ logic is defined with respect to a model M, i.e., for a given structure $\mathrm{M} = (S, RB, I^{ph}, P, C, v)$ and a given state

$\mathrm{M}, s \models p$ iff $v(s)(p) = \mathbf{1}$, for $p \in V_0$,

$\mathrm{M}, s \models \neg\alpha$ iff $\mathrm{M}, s \not\models \alpha$,

$\mathrm{M}, s \models \alpha \vee \beta$ iff $\mathrm{M}, s \models \alpha$ or $\mathrm{M}, s \models \beta$,

$\mathrm{M}, s \models M_i^k \alpha$ iff $|\{s' \in S : (s, s') \in RB(i)$ and $\mathrm{M}, s' \models \alpha\}| > k$, $k \in \mathbb{N}$,

$\mathrm{M}, s \models \Diamond(i : P)\alpha$ iff $\exists_{s' \in S}$ $((s, s') \in I^{ph}_{\Pi^{ph}}(P)(i)$ and $\mathrm{M}, s' \models \alpha)$ for $P \in \Pi^{ph}$ or $\exists_{(\mathrm{M}', s') \in \mathcal{CMS}}$ $(((\mathrm{M}, s), (\mathrm{M}', s')) \in I^{v}_{\Pi^v}(P)(i)$ and $\mathrm{M}', s' \models \alpha)$ for $P \in \Pi^v$,

$\mathrm{M}, s \models \mathbf{P}_i(\alpha) \geq q$ iff $\sum_{\{s' \in S | (s,s') \in RB(i) \text{ and } \mathrm{M}, s' \models \alpha\}} P(i)(s, s') \geq q$,

$\mathrm{M}, s \models \mathbf{C}_i \alpha$ iff $\alpha \in C(s, i)$,

$\mathrm{M}, s \models \langle i \rangle \mathrm{G}\alpha$ iff there exists a strategy f_i such that for some computation $\lambda = (\mathrm{M}_0, s_0), (\mathrm{M}_0, s_0), \ldots \in out((\mathrm{M}, s), f_i)$, and for all positions $k \geq 0$, we have $(M_k, s_k) \models \alpha$,

$\mathrm{M}, s \models \langle i \rangle^q \mathrm{G}\alpha$ iff there exists a strategy f_i such that $\frac{k_1}{k_2} \geq q$ for k_2 being the number of all computations $\lambda \in out((\mathrm{M}, s), f_i)$ and k_1 being the number of computations $\lambda \in out((\mathrm{M}, s), f_i)$ in which every state satisfy α,

$\mathrm{M}, s \models \langle\langle i \rangle\rangle \mathrm{G}\alpha$ iff there exists a strategy f_i such that for all computations $\lambda = (\mathrm{M}_0, s_0), (\mathrm{M}_1, s_1), \ldots \in out((M, s), f_i)$, and for all positions $k \geq 0$, we have $\mathrm{M}_k, s_k \models \alpha$,

$\mathrm{M}, s \models \langle i \rangle_f \mathrm{G}\alpha$ iff for some computation $\lambda = (\mathrm{M}_0, s_0), (\mathrm{M}_1, s_1), \ldots \in out((M, s), f)$, and for all positions $k \geq 0$, we have $\mathrm{M}_k, s_k \models \alpha$,

$\mathrm{M}, s \models \langle i \rangle_f^q \mathrm{G}\alpha$ iff f is a strategy such that $\frac{k_1}{k_2} \geq q$ for k_2 being the number of all computations $\lambda \in out((\mathrm{M}, s), f)$ and k_1 being the number of computations $\lambda \in out((\mathrm{M}, s), f)$ in which every state satisfy α,

$\mathrm{M}, s \models \langle\langle i \rangle\rangle_f \mathrm{G}\alpha$ iff for all computations $\lambda = (\mathrm{M}_0, s_0), (\mathrm{M}_1, s_1), \ldots \in out((M, s), f)$, and for all positions $k \geq 0$, we have $\mathrm{M}_k, s_k \models \alpha$,

$\mathrm{M}, s \models \langle\langle i \rangle\rangle \mathrm{X}\alpha$ iff there exists a strategy f_i such that for all computations $\lambda = (\mathrm{M}_0, s_0), (\mathrm{M}_1, s_1), \ldots \in out((M, s), f_i)$, we have $\mathrm{M}_1, s_1 \models \alpha$,

$\mathrm{M}, s \models \langle\langle i \rangle\rangle \alpha \mathrm{U} \beta$ iff there exists a strategy f_i such that for all computations $\lambda = (\mathrm{M}_0, s_0), (\mathrm{M}_1, s_1), \ldots \in out(s, f_i)$, there exists a position $k \geq 0$ such that $\mathrm{M}_k, s_k \models \beta$ and for all positions $0 \leq j < k$, we have $\mathrm{M}_j, s_j \models \alpha$.

Based on commitment rules defined in [12] we introduce a specification of dialogue actions. Since speech acts are verbal actions they move a system from a model M to a new model M$'$. For instance, an action $claim(\alpha)$ performed at a state s of a model M moves a multi-agent system to a model M$'$ in which a new commitment function C' is defined in such a way that the new set of commitments of the performer of the action at s equals to the old one enriched with α. Notice that *claim* and *concede* are two different speech acts and are used with two different intentions. An agent "*claims* α" when he publicly announces that he is committed to α. Whereas, an agent "*concedes* α" when he agrees with his opponent that α holds. Therefore, it is a reply to an opponent's argumentation for α. Nevertheless, formal specification of these actions is exactly the same. Formally, they both add a formula α to the set of commitments of the performer. A formal definition of interpretation of dialogue actions I^v is as follows:

1. **claim**:
   $((\mathrm{M}, s), (\mathrm{M}', s)) \in I^v(claim(\alpha))(i)$ iff $\mathrm{M}' = (S, RB, I^{ph}, v, C')$
   where
   – $C'(s,i) = C(s,i) \cup \{\alpha\}$ and
   – $C'(s',i') = C(s',i')$ for $s' \neq s$ or $i' \neq i$,

2. **concede**:
   $((\mathrm{M}, s), (\mathrm{M}', s)) \in I^v(concede(\alpha))(i)$ iff $\mathrm{M}' = (S, RB, I^{ph}, v, C')$
   where
   – $C'(s,i) = C(s,i) \cup \{\alpha\}$ and
   – $C'(s',i') = C(s',i')$ for $s' \neq s$ or $i' \neq i$,

3. **retract**:
   $((\mathrm{M}, s), (\mathrm{M}', s)) \in I^v(retract(\alpha))(i)$ iff $\mathrm{M}' = (S, RB, I^{ph}, v, C')$
   where
   – $C'(s,i) = C(s,i) \backslash \{\alpha\}$ and
   – $C'(s',i') = C(s',i')$ for $s' \neq s$ or $i' \neq i$,

4. **argue**:
   $((\mathrm{M}, s), (\mathrm{M}', s)) \in I^v(argue(A))(i)$ iff $\mathrm{M}' = (S, RB, I^{ph}, v, C')$
   where
   – $C'(s,i) = C(s,i) \cup prem(A) \cup conc(A)$
   – $C'(s',i') = C(s',i')$ for $s' \neq s$ or $i' \neq i$,

5. **why**:
   $((\mathrm{M}, s), (\mathrm{M}', s)) \in I^v(why(\alpha))(i)$ iff $\mathrm{M}' = \mathrm{M}$.

4. Model checker Perseus

The Perseus system is a software tool designed for an automatic many-sided analysis of persuasive multi-agent systems. It was designed in 2008 by Budzyńska, Kacprzak and Rembelski and is still developed [6]. Its aim is to analyze persuasion ability of multi-agent systems given their formal model. Until now, Perseus can deal with features concerning graded beliefs of agents, probabilistic beliefs of agents and the impact of persuasive actions on agents' beliefs and activities.

Given a semantic model M of a system, the task of the Perseus system is to automatically analyze its properties. It could be done twofold: by model checking or by parametric verification. Application of **model checking** method allows for testing whether a $\mathcal{AG}_n$ formula is true in a given state of the model M. In other words, using model checking technique Perseus tests whether some specific property holds in a multi-agent system represented by the model M. **Parametric verification** was introduced by the authors of the Perseus system. This method allows Perseus to look for answers to questions about diverse properties of systems under consideration and, in consequence, allows to analyze these systems in an automatic way. In particular, questions can concern

- agents – is there an agent who can influence somebody's beliefs?, who can do it?, who can achieve a success?
- beliefs and degrees of beliefs – does an agent believe a claim?, what is a degree of his uncertainty about this claim?
- results of actions – whether a degree of agent's belief can change after execution of a given action or sequence of actions?, which actions should be executed in order to convince an agent that a claim is true?

The system input data of the Perseus tool, i.e. the **input question**, is a triple (M, s, ϕ), where M is a model described by an arbitrary specification of a model (see [6]), s is a state of the model M and ϕ is the **input expression**. The input expression is defined by the following BNF:

$$\phi ::= \omega|\neg\phi|\phi \vee \phi|M_i^d\phi|\Diamond(i:P)\phi|M_i^?\omega|\Diamond\,(i:?)\,\omega|\mathbf{P}_i\,(\omega) \geq ?|$$

$$M_?^d\omega|\Diamond\,(?:P)\,\omega|\mathbf{P}_?\,(\omega) \geq q,$$

where $\omega ::= p|\neg\omega|\omega \vee \omega|M_i^d\omega|\Diamond\,(i:P)\,\omega|\mathbf{P}_i(\omega) \geq q$ and $p \in V_0$, $d \in \mathbb{N}$, $P \in \Pi^{ph}$ or $P \in \Pi^v$, $i \in Agt$ as well as $q \in [0;1]$. Therefore the language of extended $\mathcal{AG}_n$ logic is a sublanguage of the Perseus system input expressions (what follows is that other modalities $B_i^d\omega$, $M!_i^d\omega$, $M!_i^{d_1,d_2}\omega$, $\Box\,(i:P)\,\omega$,

$\mathbf{P}_i(\omega) > q$, $\mathbf{P}_i(\omega) = q$, $\mathbf{P}_i(\omega) < q$, $\mathbf{P}_i(\omega) \le q$, can be derived in the standard way).

Perseus system accepts two types of the input expressions:

- **unknown free expressions**, where grammar productions

  $$M_i^? \omega | \Diamond (i :?) \omega | \mathbf{P}_i(\omega) \ge ? | M_?^d \omega | \Diamond (? : P) \omega | \mathbf{P}_? (\omega) \ge q$$

  are not allowed,

- **one-unknown expression**, where only one of the grammar productions

  $$M_i^? \omega | \Diamond (i :?) \omega | \mathbf{P}_i(\omega) \ge ? | M_?^d \omega | \Diamond (? : P) \omega | \mathbf{P}_? (\omega) \ge q$$

  is allowed.

Next the Perseus system executes a parametric verification of an input question, i.e. tests if (both unknown free and one-unknown expressions) and when (only one-unknown expressions) the expression ϕ becomes a formula of the extended $\mathcal{AG}_n$ logic ϕ^* such that $M, s \models \phi^*$.

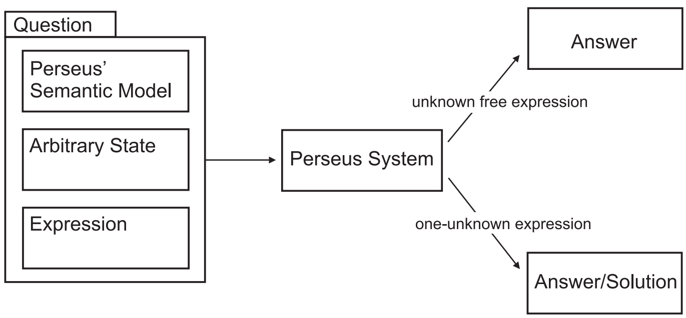

Figure 1. The idea of the Perseus system

In case of unknown free expressions we have $\phi^* = \phi$, i.e a standard model verification is done. In the other case a formula ϕ^* is obtained from ϕ by swapping all ? symbols for appropriate values either from set $\{0, 1, \ldots |S|\}$ or Agt or Π^{ph} or Π^v or $[0; 1]$. Finally the system output data, i.e the **output answer**, is given. The output answer is *true* if $M, s \models \phi^*$ and *false* otherwise (see Fig. 1). As soon as the output answer is determined, the **solution set** X for the one-unknown expression is presented, where:

- $X \subseteq \{0, 1, \ldots |S|\}$, for an expression ϕ with one unknown of type $M_i^? \omega$, $B_i^? \omega$, $M!_i^? \omega$, $M!_i^{?,d_2} \omega$, $M!_i^{d_1,?} \omega$,
- $X \subseteq \{0, 1, \ldots |S|\} \times \{0, 1, \ldots |S|\}$, for an expression ϕ with one unknown of type $M!_i^{?_1,?_2} \omega$,
- $X \subseteq Agt$, for an expression ϕ with one unknown of type $M_?^d \omega$, $B_?^d \omega$, $M!_?^d \omega$, $M!_?^{d_1,d_2} \omega$, $\Diamond (? : P) \omega$, $\Box (? : P) \omega$, $\mathbf{P}_? (\omega) \geq q$, $\mathbf{P}_? (\omega) > q$, $\mathbf{P}_? (\omega) = q$, $\mathbf{P}_? (\omega) < q$, $\mathbf{P}_? (\omega) \leq q$,
- $X \subseteq \Pi^{ph}$ or $X \subseteq \Pi^v$, for an expression ϕ with one unknown of type$\Diamond (i :?) \omega$, $\Box (i :?) \omega$,
- $X \subseteq [0; 1]$, for an expression ϕ with one unknown of type $\mathbf{P}_i (\omega) \geq ?$, $\mathbf{P}_i (\omega) > ?$, $\mathbf{P}_i (\omega) = ?$, $\mathbf{P}_i (\omega) < ?$, $\mathbf{P}_i (\omega) \leq ?$.

In order to find an answer to the input question (M, s, ϕ), the Perseus system executes the syntax analysis of the expression ϕ. The analysis is based on the standard descent recursive method. As a result a syntax tree of expression ϕ is created. All inner nodes of such a tree represent either Boolean operators or $\mathcal{AG}_n$ logic modalities while all outer nodes stand for either propositional variables or unknown. The solution for an arbitrary unknown is reached in the following way:

- if an unknown type is $M_i^? \omega$, $B_i^? \omega$, $M!_i^? \omega$, $M!_i^{?,d_2} \omega$, $M!_i^{d_1,?} \omega$, $M!_i^{?_1,?_2} \omega$, then the counting method is applied, i.e. all states, which are reachable via a doxastic relation of the agent i, and in which the claim ω is satisfied or refuted respectively, are counted,
- if an unknown type is $M_?^d \omega$, $B_?^d \omega$, $M!_?^d \omega$, $M!_?^{d_1,d_2} \omega$, $\Diamond (? : P) \omega$, $\Box (? : P) \omega$, $\mathbf{P}_? (\omega) \geq q$, $\mathbf{P}_? (\omega) > q$, $\mathbf{P}_? (\omega) = q$, $\mathbf{P}_? (\omega) < q$, $\mathbf{P}_? (\omega) \leq q$, say $\mathbf{P}_? (\omega) \geq q$, then for every agent $i \in Agt$ the property $\mathcal{M}, s \models \mathbf{P}_i (\omega) \geq q$ is tested,
- if an unknown type is $\Diamond (i :?) \omega$, $\Box (i :?) \omega$, then a nondeterministic finite automaton, which represents all possible argumentation $P \in \Pi$ such that respectively $\mathcal{M}, s \models \Diamond (i : P) \omega$ or $\mathcal{M}, s \models \Box (i : P) \omega$ holds, is created,
- if an unknown type is $\mathbf{P}_i (\omega) \geq ?$, $\mathbf{P}_i (\omega) > ?$, $\mathbf{P}_i (\omega) = ?$, $\mathbf{P}_i (\omega) < ?$, $\mathbf{P}_i (\omega) \leq ?$, then the **summing method** is applied, i.e. probabilistic coefficients of all states, which are reachable via doxastic relation of the agent i, and in which the claim ω is satisfied or refuted respectively, are add up.

If an unknown is a nested type, i.e. it is a part of claim of the extended $\mathcal{AG}_n$ logic operator, then its **solution set is bounded** by the outer modality/modalities. For example, if we consider an input question

$$\Big(M, s, \Box\,(i : P)\, M!^{1}_{j}\mathbf{P}_i\,(\omega) <?\Big),$$

then the solution of the unknown $\mathbf{P}_i\,(\omega) <?$ is reduced firstly by the operator $M!$ and secondly by the operator $\Box$.

5. The properties of persuasion in multi-agent systems

In this section, we present the important properties of persuasion in multi-agent systems which could be examined with the use of a model-checker.

5.1. Influence on degrees of beliefs

In order to formally verify the properties of persuasion in multi-agent systems, Perseus searches for answers to **questions** expressed in the $\mathcal{AG}_n$ language. The first group of questions ask about the properties of persuasion related to influencing agents uncertainty. In our model, uncertainty is represented by two types of operators: graded and probabilistic modalities. Each of them encodes slightly different information. The graded belief formula $M!^{3,5}_{John}p$ expresses that there are 5 John's doxastic alternatives and in 3 of them p holds, while the probabilistic formula $\mathbf{P}_{John}(p) = 0.6$ does not describe local properties of the model with such details, since equally John could allow 50 doxastic alternatives and in 30 of them p would hold. Thus, in the latter case we are dealing with a loss of the information. In other words, a probabilistic formula says what is the uncertainty of an agent about a claim, but does not give any reasons. On the other hand, the probabilistic operator allows the verification of questions in which such a detailed information is not needed, but instead we are interested in all cases when an agent is uncertain in a specific degree. For example, we may ask if it is possible that after a persuasion John will believe a claim with the degree 0.6 regardless of how many doxastic alternatives he allows (i.e. no matter if there are 5 John's doxastic alternatives and in 3 of them p holds or there are 50 John's doxastic alternatives and in 30 of them p holds, and so on).

Perseus can check the property of influencing agent uncertainty with respect to unknown-free expressions using the standard model-checking technique, e.g. the tool can check whether:

- $M!_i^3\omega$, exactly 3 doxastic alternatives of the agent i satisfy the claim ω,
- $M_i^4 B_j^2\omega$, in more than 4 doxastic alternatives of the agent i it is true that in at most 2 doxastic possibilities of the agent j the claim ω is refuted,
- $\Diamond(i : P)\,\omega$, the execution of the argumentation P by the agent i may cause that the claim ω is satisfied,
- $\Box(i : P)\, M!_j^{2,4}\omega$, the execution of the argumentation P by the agent i can not cause that it is not true that in exactly 2 doxastic alternatives of the agent j among exactly 4 his doxastic possibilities the claim ω is satisfied.

The Perseus tool can also check the property of influencing agent uncertainty with respect to one-unknown expressions using the parametric verification technique, e.g. it can check whether:

- $M_i^?\omega$, in more than how many doxastic alternatives of the agent i the claim ω is satisfied?
- $M_?^d\omega$, for which agent is it true that in more than d of his doxastic alternatives the claim ω is satisfied?
- $B_i^?\omega$, in at the most how many doxastic alternatives of the agent i the claim ω is refuted?
- $B_?^d\omega$, for which agent is it true that in at most d of his doxastic alternatives the claim ω is refuted?
- $M!_i^?\omega$, in exactly how many doxastic alternatives of the agent i the claim ω is satisfied?
- $M!_?^d\omega$, for which agent is it true that in exactly d of his doxastic alternatives the claim ω is satisfied?
- $M!_i^{?,d_2}\omega$, in exactly how many doxastic alternatives of the agent i, from exactly d_2 of his doxastic possibilities, the claim ω is satisfied?
- $M!_i^{d_1,?}\omega$, what is an exact number of all doxastic alternatives of the agent i, where in exactly d_1 of them the claim ω is satisfied?
- $M!_i^{?_1,?_2}\omega$, what is an exact number of all doxastic alternatives of the agent i and in exactly how many of them the claim ω is satisfied?
- $M!_?^{d_1,d_2}\omega$, for which agent is it true that in exactly d_1 doxastic alternatives among exactly d_2 of his doxastic possibilities the claim ω is satisfied?
- $\Diamond(i :?)\,\omega$, for what argumentation is it true that its execution by the agent i may cause that the claim ω is satisfied?

- $\Diamond\,(?:P)\,\omega$, for which agent is it true that his execution of the argumentation P may cause the claim ω is satisfied?
- $\Box\,(i:?)\,\omega$, for what argumentation is it true that its execution by the agent i can not cause that the claim ω is refuted?
- $\Box\,(?:P)\,\omega$, for which agent is it true that his execution of the argumentation P can not cause that the claim ω is refuted?

In particular, the expression $M_i^{?_1,?_2}(\Diamond(i:a)M!_j^{10,10}\alpha)$ asks "What is a degree of an agent i's belief that after using an argument a an agent j will believe α with a degree $\frac{10}{10}$?". The other question $\Diamond(i:a)M_j^{?_1,?_2}\alpha$ means "What will be a degree of an agent j's belief about α after an argumentation a performed by an agent i ?" and the question $\Diamond(i:?)M_j^{10,10}\alpha$ means "What argumentation should an agent i use to convince an agent j to believe about α with a degree $\frac{10}{10}$?".

5.2. Conflict of opinion and persuasiveness

The initial condition for persuasion is a **conflict of opinion** [16]. Consider two agents which task is to increase the environment's temperature as soon as it drops below 10^0C. Assume that one agent is not able to carry out the task on his own. It is possible only if agents cooperate. In our scenario a conflict will start when one of the agents believes that the temperature is lower than 10^0C while the other does not – possibly because agents use different sources of information and thereby derive different conclusions. Observe that a conflict may appear not only when a proponent is absolutely sure about the claim and an audience is absolutely against it. It can also arise from the fact that the degrees of agents' beliefs differ or belong to different intervals. Say that degrees from $(\frac{1}{2},1]$ mean accepting the claim and degrees from $[0,\frac{1}{2}]$ mean rejecting the claim. Then, Perseus can verify the property with respect to the conflict of opinion checking if e.g. such a $\mathcal{AG}_n$ formula holds in a given model:

$$M!_{prop}^{3,4}(p_{t<10}) \wedge M!_{aud}^{1,4}(p_{t<10})$$

where *prop* and *aud* mean the proponent and the audience, respectively, and $p_{t<10}$ is a propositional variable which expresses that the temperature is lower than 10^0C. This formula should be read as follows: "The proponent believes that the temperature is lower than 10^0C with the degree $\frac{3}{4}$ and the audience believes the temperature is lower than 10^0C with the degree $\frac{1}{4}$".

If the inter-agent conflict of opinion is resolved, we talk about the **success of persuasion**. The persuasion $P=(a_1;a_2;\ldots;a_k)$ can be successful when after performing actions a_1, a_2, …, a_k by the proponent, it is possible that the audience will believe the claim with some expected degree. We assumed that an agent accepts the claim if he believes it in a degree higher than $\frac{1}{2}$. Then, the Perseus software can check if the proponent's persuasion may be successful:

$$\Diamond(prop : P)(M!^{3,4}_{aud}(p_{t<10})).$$

This formula states "If the proponent performs arguments P then it is possible that the audience will believe the claim with the degree $\frac{3}{4}$".

In some circumstances, an agent may have a chance to achieve only the **subjective success**. That is, after execution of P the proponent may believe that he achieved a goal while he actually did not, which means that he wrongly evaluated the results of his persuasion. Perseus may provide an answer whether there is a risk of such a situation:

$$\Diamond(prop : P)[M!^{4,4}_{prop}(M!^{3,4}_{aud}(p_{t<10})) \wedge \neg M!^{3,4}_{aud}(p_{t<10})].$$

The formula states "If the proponent performs arguments P then it is possible he will believe that the audience is convinced with the degree $\frac{3}{4}$, but the audience will not believe the claim with this degree".

The other property that the model-checker can verify is whether the proponent **predicts** (believes) that he is able to succeed. Otherwise, he may not start persuasion even though he had all necessary means to win. Such a situation can be expressed by the formula:

$$M!^{0,4}_{prop}[\Diamond(prop : P)(M!^{4,4}_{aud}(p_{t<10}))] \wedge \Box(prop : P)(M!^{4,4}_{aud}(p_{t<10})).$$

If the formula holds in a given model, then the proponent is absolutely sure that his persuasion P will fail (the proponent believes with the degree $\frac{0}{4}$ that the audience may become convinced to the claim with the degree $\frac{4}{4}$), while P would actually lead him to success (after persuasion P the audience will believe the claim with the desired degree $\frac{4}{4}$).

The **persuasiveness** depends on the arguments (their quality as well as length and order of argument sequence) and on the credibility of proponent. The same proponent convincing the same audience with the use of **different quality arguments** may arrive at different results of the persuasion. Say that if the proponent gives verbal argument "One of your thermometers is placed wrongly since it is too close to a heater" (action a_1), then he will win with non-absolute strength:

$$M!^{1,4}_{aud}(p_{t<10}) \rightarrow \Box(prop : a_1)(M!^{3,4}_{aud}(p_{t<10})).$$

We read this formula as follows "If the audience believes the claim with the degree $\frac{1}{4}$ then always after the execution of the action a by the proponent, the audience will believe the claim with the degree $\frac{3}{4}$". On the other hand, if the proponent will perform a nonverbal persuasive action moving the thermometer to another place (action a_1) (proving this way that the temperature is lower than 10^0C), he may obtain audience's utter conviction:

$$M!^{1,4}_{aud}(p_{t<10}) \rightarrow \Diamond(prop : a_2)(M!^{4,4}_{aud}(p_{t<10})).$$

The next property of persuasion is the **length and order of argument sequence**. Say that the proponent is able to convince the audience to believe the claim in the degree $\frac{3}{4}$ with support of argument sequence $a_1; a_2; a_3$. The questions we may want to ask to the Perseus system are, firstly, whether it is possible to convince the audience using fewer than three arguments and obtain exactly the same result (or possibly even better), and, secondly, whether it is possible to obtain a better result performing these arguments in a different order, e.g. $a_3; a_1; a_2$? The first question is expressed by such a formula:

$$\Diamond(prop : a_1; a_2; a_3)(M!^{3,4}_{aud}(p_{t<10})) \wedge \Diamond(prop : a_4)(M!^{3,4}_{aud}(p_{t<10})).$$

This formula means that "If the proponent gives three arguments a_1, a_2, a_3 then the audience will believe the claim with the degree $\frac{3}{4}$ and if the proponent gives only one argument a_4 then the audience will believe the claim with the same degree of $\frac{3}{4}$". The second question is expressed by such a formula:

$$\Diamond(prop : a_1; a_2; a_3)(M!^{3,4}_{aud}(p_{t<10})) \wedge \Diamond(prop : a_3; a_1; a_2)(M!^{4,4}_{aud}(p_{t<10})).$$

This formula says that "If the proponent gives three arguments a_1, a_2, a_3 then the audience will believe the claim with the degree $\frac{3}{4}$ and if the proponent order them differently, i.e. $a_3; a_1; a_2$, then the audience will believe the claim with the higher degree of $\frac{4}{4}$".

Persuasiveness can be also affected by the **credibility of a proponent**. Assume that the audience finds a proponent $prop_1$ unreliable. As a result, it does not trust what the proponent says or acts and therefore none of $prop_1$'s arguments will convince aud. On the other hand, if another proponent $prop_2$ is a leader of a group of agents or a specialist, then his arguments may have great persuasive power. In other words, the same arguments can cause different results depending on an agent who performs them:

$$\neg\Diamond(prop_1 : P)(M!^{4,4}_{aud}(p_{t<10})) \wedge \Diamond(prop_2 : P)(M!^{4,4}_{aud}(p_{t<10})).$$

5.3. Verbal and non-verbal persuasive actions

In [5], the syntax and semantics of $\mathcal{AG}_n$ logic is enriched to allow the representation and verification of properties related to the type of actions performed in persuasion. Every persuasive action is described by 3-tuple (m, β, δ) which fixes a content of a message m sent in the action, a goal α of executing action and the way it is performed. Formally, the set of persuasive actions $\Pi_p \subseteq \Pi_0$ is defined as follows:

$$\Pi_p = \{(m, \beta, \delta) : m \in C, \beta \in F, \delta \in \Delta\}$$

where C is a set of contents, F is a set of formulas of $\mathcal{AG}_n$ and Δ is a set of symbols representing means of actions, i.e., ways they can be performed. This set can consist of the elements such as: *ver* – for verbal actions, *nver* – for nonverbal actions.

Consider the example given in [8]. John prepares shrimps for a dinner. Mary wants him to add some curry and says "Don't you think these shrimps need curry?". John is not convinced that it will make shrimps taste better and refuses to do it. Mary quits trying to persuade him verbally and tries nonverbal strategy. She goes to the kitchen cupboard, climbs onto the step-stool and begins searching through the upper shelves of the cupboard. Finally, she goes down with a smile and gives him a can of powder. John looks at her and says "Well, yeah, sure, O.K...". The effort that Mary put in finding curry finally made him add curry. In other words, the physical actions resulted in success while verbal action failed. Observe that both arguments – verbal and nonverbal – sends (more or less) the same message m of how important Mary thinks the curry is. That is, the **means of sending** the message can give different results in persuading a receiver of that message.

We say that a goal of an action is achieved if after execution of this action a state, in which it is satisfied, is reached (a formula expressing the goal is true in this state). For example, suppose an action $a = (m, \beta, \delta)$ executed by Mary. The goal of this action is achieved if there exists a state s such that s is a result of a and $s \models \beta$.

Perseus can verify if the same content sent with the same goal but by different means (verbal vs. non-verbal) brings about different results:

$$\Diamond((m, M!^{1,1}_{John}(p), \boldsymbol{nver}) : Mary)M!^{1,1}_{John}(p) \wedge$$
$$\neg\Diamond((m, M!^{1,1}_{John}(p), \boldsymbol{ver}) : Mary)M!^{1,1}_{John}(p).$$

where p means that "shrimps with curry are better". The property expresses that Mary's nonverbal action of sending m is successful, while verbal action of sending m is not. Moreover, it is possible to test which content can cause the success assuming that the goal and means are the same:

$$\Diamond((\boldsymbol{m}_1, \beta, \delta) : i)\beta \wedge \neg\Diamond((\boldsymbol{m}_2, \beta, \delta) : i)\beta.$$

5.4. Playing a dialogue game

In [3], we extend the $\mathcal{AG}_n$ logic to allow the formulation of questions referring to the properties of persuasion in agent dialogue games. Consider a dialogue given in [12]:

Paul: My car is safe. (making a claim)

Olga: Why is your car safe? (asking grounds for a claim)

Paul: Since it has an airbag. (offering grounds for a claim)

Olga: That is true, (conceding a claim) but this does not make your car safe. (stating a counterclaim)

Paul: Why does that not make my car safe? (asking grounds for a claim)

Olga: Since the newspapers recently reported on airbags expanding without cause. (stating a counterargument by providing grounds for the counterclaim)

Paul: Yes, that is what the newspapers say. (conceding a claim) OK, I was wrong that my car is safe. (retracting a claim)

The dialogue actions will move the system to a new model in which the sets of commitments will be enriched with new formulas. In our approach this dialogue is a sequence of the following actions:

$((Paul : claim(p)); (Olga : why(p)); (Paul : argue(q, q \to p; p));$
$(Olga : concede(q)); (Olga : claim(\neg p)); (Paul : why(\neg p));$
$(Olga : argue(r, r \to \neg p; \neg p)); (Paul : concede(r)); (Paul : retract(p))).$

Perseus will be able to verify that the dialogue satisfies a property. For example it can check whether the formula

$$\Diamond(Paul : claim(p))(\mathbf{C}_{Paul}p \wedge \Diamond(d)\neg\mathbf{C}_{Paul}p)$$

is true at the state s of the model M. $\Diamond(d)$ is the abbreviation for:

$\Diamond(Olga : why(p))\ \Diamond(Paul : argue(q, q \to p; p))\Diamond(Olga : concede(q))$
$\Diamond(Olga : claim(\neg p))\ \Diamond(Paul : why(\neg p))(Olga : argue(r, r \to \neg p; \neg p))$
$\Diamond(Paul : concede(r))\ \Diamond(Paul : retract(p)).$

The formula says that at the beginning Paul announces that his car is safe and after the dialogue d he withdraws from this statement.

The Perseus tool may also find the dialogue after which Paul is not committed to the proposition p. It is done by parametric verification of the expression

$$\Diamond(Paul : claim(p))(\mathbf{C}_{Paul}p \wedge \Diamond(?)\neg\mathbf{C}_{Paul}p).$$

In other words, Perseus will look for a sequence of actions such that if it is replaced with the symbol ? then this expression becomes a formula true at state s of M [6]. The parametric verification requires searching the whole dialogue game, i.e., all possible dialogues allowed by a given protocol.

5.5. Strategies and victory in a persuasion dialogue

In [4], the $\mathcal{AG}_n$ logic is extended to enable the formal verification of strategies allowing an agent to become victorious in a dialogue game. A victory can be specified in different ways, e.g. it may be assumed that the proponent i is the winner if the opponent has conceded i's main claim and the opponent is the winner if the proponent i has retracted i's main claim [12]. Let $win(i)$ mean that i is a winner of a given dialogue game, t be a topic (i.e. a conflict formula), *prop* – a proponent and *opp* – an opponent. Then: (1) $win(prop)$ is true in a state, in which $\mathbf{C}_{opp}(t)$ holds, and (2) $win(opp)$ is true in a state, in which $\neg\mathbf{C}_{prop}(t)$ holds. Using a given specification for a victory, we can ask: is it possible that after performing a dialogue d between i and $\bar{i}$ played in accordance with the protocol P, it will be the case that the proposition $win(i)$ will hold.

A more interesting question would be to ask **which dialogue** (sequence of moves) allows an agent to win a dialogue game. This question requires our model checker to perform parametric verification by searching for a legal dialogue (a dialogue played according to a given protocol) such that it is possible that after performing it the proposition $win(i)$ will hold. The answer to this question would allow an agent to plan how to play the dialogue game, however, it has a limitation. Unlike some other types of sequences of actions, a dialogue always consists of actions executed not only by one agent, but also by his adversary. It means that part of a sequence is not under control of a given agent. As a result, even though i knows that a particular sequence leads him to the victory, this sequence may not be performed in a dialogue, since $\bar{i}$ may execute the action allowed by a dialogue protocol, but other than considered by i. Say that a sequence *claim p*; *why p*; *p since q*; ...; *concede s* allows i to win a dialogue. Yet, in the second move $\bar{i}$ may execute *claim* $\neg p$ instead of *why p*.

The important property of a persuasion dialogue game is an existence of a strategy which allows an agent to win the game. A strategy for an agent i can be defined as a function from the set of all finite legal dialogues in which i is to move into L_c [13]. Intuitively, i has a strategy if he has a plan of how to react to any move of his adversary. Say that the first move *claim p* is performed by i. At this state, i considers how he will respond after all possible moves that $\bar{i}$ is allowed to make at the next stages of a dialogue. In particular, he may plan that at the subsequent stage if $\bar{i}$ executes *why* p, then his response will be: p *since* q (instead of, e.g., *retract* p), if $\bar{i}$ executes *claim* $\neg p$, then his response will be: *why* $\neg p$ (instead of, e.g., *concede* $\neg p$), and so on. An agent may want to know if a strategy that he adopted guarantees him victory in a given dialogue game regardless of what actions his opponent will perform, i.e. if his strategy is winning. A strategy is a **winning strategy** for i if in every dialogue played according to this strategy i accomplishes his dialogue goal [13].

The question about a winning strategy has some limitations, since in some systems of persuasion dialogue a player may avoid losing simply by never giving in [12, p. 1021], e.g. an opponent may repeat *why* α as a response to any assertion that the proponent performs, such as *claim* α or β *since* α. In such cases, we may want to ask if a strategy allows an agent to reason not about the guarantee but about the possibility of victory. Intuitively, a **strategy gives** i **chance for success** if there is a dialogue game played according to this strategy such that i accomplishes his dialogue goal. Knowing that the strategy has this feature allows an agent to make decision about which strategy he should adopt in order to have a chance to win. Even though the agent is not sure if he will win, the information that one strategy can bring him success and the other cannot is better than no information.

This type of question has also some limitations. Say that an agent knows that ten strategies allow him to be victorious. How can he decide which strategy to choose? Thus, an agent may wish to know in how many cases a strategy gives him a chance to win. Assume a class of dialogue games in which there is a finite number of possible game's scenarios. Let k_2 be the number of all dialogues played according to a given strategy, and k_1 be a number of dialogues played according to this strategy in which a given agent i accomplishes his dialogue goal. If k_1 and k_2 are finite, then we say that this **strategy gives** i **chance for success in a degree** $\frac{k_1}{k_2}$. If they are infinite, the degree of chance for success is not defined. Knowing that ten strategies allows him to be victorious and knowing their degrees of chance for success allows an agent to choose among them and, in consequence, to maximize his chance to win.

Formally, the properties of dialogue systems concerning the strategies in a dialogue game can be specified and verified with the use of the strategy operators, e.g.

- $\langle\langle i\rangle\rangle$ *true* U *win*(i) – there is a strategy for agent i which ensure that i will win the dialogue game, i.e. there exists a winning strategy in a game,

- $\langle i\rangle^{0.5}$ *true* U *win*(i) – there is a strategy which allows an agent i to win a dialogue game with degree higher than 0.5,
- $\langle\langle i\rangle\rangle$ *true* U $(M!^{1,1}_{\bar{i}})t$ – there is a strategy for agent i which ensure that i's adversary will believe the topic t with degree $\frac{1}{1}$,
- $\langle\langle i\rangle\rangle$ *true*U $C_{\bar{i}}(t)$ – there is a strategy for i which ensure that i's adversary will be committed to the topic t,
- $\langle\langle prop\rangle\rangle$G(*terminate* $\rightarrow$ *win*(*prop*)) – the proponent has such a strategy that he is a winner every time when the dialogue terminates (i.e. the proposition *terminate* is true),
- $\langle prop\rangle$G(*terminate* $\rightarrow$ *win*(*prop*)) – the proponent has such a strategy that if the dialogue terminates then he may be a winner.

6. Conclusions

The paper presents the Perseus model checker which allows the formal verification of persuasion in multi-agent systems. Perseus is built upon the $\mathcal{AG}_n$ logic and performs both the standard model checking method and the parametric verification. In the first case, the tool checks if a given $\mathcal{AG}_n$ formula is true in a given model, while in latter case it searches for answer to a question about a given property of persuasion in a multi-agent system. Formally it means that for an $\mathcal{AG}_n$ expression with unknowns Perseus searches for such values that if the unknowns are replaced with those values, the expression becomes true in a given state and a given model. Perseus is designed and adjusted to verify different properties of persuasion related to influence of persuasion on agent uncertainty, conflict of opinion which initiates persuasion, persuasiveness, the type of means of sending a persuasive message (verbal or non-verbal), playing a persuasion dialogue game and strategies allowing an agent to wine a dialogue game.

Acknowledgments

Katarzyna Budzyńska gratefully acknowledges the support from Polish Ministry of Science and Higher Education under grant N N101 009338.

References

[1] E. Black and A. Hunter. An inquiry dialogue system. *Autonomous Agents and Multi-Agent Systems*, 19(2): 173–209, 2009.

[2] K. Budzyńska and M. Kacprzak. A logic for reasoning about persuasion. *Fundamenta Informaticae*, 85: 51–65, 2008.

[3] K. Budzyńska and M. Kacprzak. Formal framework for analysis of agent persuasion dialogue games. In *Proc. of Concurrency, Specification and Programming*, pages 85–96, 2010.

[4] K. Budzyńska and M. Kacprzak. A logic for strategies in persuasion dialogue games. *to be published*, 2011.

[5] K. Budzyńska, M. Kacprzak, and P. Rembelski. Logic for reasoning about components of persuasive actions. *Proc. of ISMIS'09 (18th International Symposium on Methodologies for Intelligent Systems)*, LNAI (5722): 201–210, 2009.

[6] K. Budzyńska, M. Kacprzak, and P. Rembelski. Perseus. software for analyzing persuasion process. *Fundamenta Informaticae*, 2009.

[7] K. Budzyńska, M. Kacprzak, and P. Rembelski. Update of probabilistic beliefs: implementation and parametric verification. *Fundamenta Informaticae*, 102(1): 35–48, 2010.

[8] M. A. Gilbert. Modes, coalescence and pragma-dialectics. http://www.yorku.ca/gilbert/argthry/argthry/arg_readings.html, 1995.

[9] M. Kacprzak, W. Nabiałek, A. Niewiadomski, W. Penczek, A. Półrola, M. Szreter, B. Woźna, and A. Zbrzezny. Verics 2008 – a model checker for time petri nets and high-level languages. In *Proc. of the International Workshop on Petri Nets and Software Engineering (PNSE'09)*, pages 119–132. University of Hamburg, Department of Informatics. Paris, France, 2009.

[10] K. Larson and I. Rahwan. Welfare properties of argumentation-based semantics. In *Proceedings of the 2nd COMSOC*, 2008.

[11] A. Lomuscio, H. Qu, and F. Raimondi. MCMAS: A model checker for the verification of multi-agent systems. In *Proc. of the 21th International Conference on Computer Aided Verification (CAV 2009)*, volume 5643 of *Lecture Notes in Computer Science*, pages 682–688. Springer, 2009.

[12] H. Prakken. Coherence and flexibility in dialogue games for argumentation. *Journal of Logic and Computation*, (15): 1009–1040, 2005.

[13] H. Prakken. Formal systems for persuasion dialogue. *The Knowledge Engineering Review*, 21: 163–188, 2006.

[14] I. Rahwan, K. Larson, and F. Tohme. A characterisation of strategy-proofness for grounded argumentation semantics. In *Proceedings of the 21st IJCAI*, 2009.

[15] C. D. Walton. Verifiable agent dialogues. *Journal of Applied Logic*, 5: 197–213, 2007.

[16] D. N. Walton and E. C. W. Krabbe. *Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning*. State University of N.Y. Press, 1995.

Katarzyna Budzyńska
Institute of Philosophy
Cardinal Stefan Wyszyński University in Warsaw
ul. Dewajtis 5, 01-815 Warsaw, Poland
k.budzynska@uksw.edu.pl

Magdalena Kacprzak
Faculty of Computer Science
Biaystok University of Technology
ul. Wiejska 45A, 15-351 Białystok, Poland
mkacprzak@ii.pb.bialystok.pl

http://perseus.ovh.org/
http://argumentacja.pdg.pl/