Computationally Grounded Model of BDI-Agents*

Kaile Su[1], **Abdul Sattar**[1], **Kewen Wang**[1] and **Guido Governatori**[2]
[1]Institute for Integrated and Intelligent Systems, Griffith University, Brisbane, Australia
[2]School of ITEE, The University of Queensland, Brisbane, Australia

1 Introduction

There are two main semantic approaches to formalizing agent systems via modal logics, the *possible worlds semantics* [Hintikka, 1962] and the *interpreted system* model [Fagin *et al.*, 1995]. The first approach includes the well-known theory of intension [Cohen and Levesque, 1990] and the formalism of the belief-desire-intension paradigm [Rao and Georgeff, 1998]. The second approach, offers a natural interpretation, in terms of the states of computer processes, to S5 epistemic logic. The significance of the second approach is that we are able to associate the system with a computer program, and formulas can be understood as properties of program computations. In this sense, the interpreted system model is *computationally grounded* [Wooldridge, 2000].

There are few computationally grounded models for formalizing general BDI-agents. A number of researchers have attempted to develop executable agent languages such as AgentSpeak(L) [Rao, 1996], but these agent languages have comparatively simple semantics [van der Hoek and Wooldridge, 2003]. The aim of this paper is to present a computationally grounded model of general BDI-agents.

The main idea of our BDI-model is that we characterize an agent's belief, desire and intention as a set of runs (computing paths), which is exactly a *system* in the interpreted system model. Let $\mathcal{B}_i$, $\mathcal{D}_i$ and $\mathcal{I}_i$ be sets of runs related to the beliefs, desires and intentions of agent i, respectively. Then, runs in $\mathcal{B}_i$ are possible computing paths from the viewpoint of the agent; runs in $\mathcal{D}_i$ are those computing paths that the agent desires; and runs in $\mathcal{I}_i$ are those computing paths with the agent's choices of the possible actions. Clearly, it is reasonable to assume that $D_i \subseteq B_i$, that is, every desired computing path is a possible one, but we need not to assume that $I_i \subseteq D_i$ or even $I_i \subseteq B_i$ because an agent's intention may fail to achieve its goal and the real computing path may be beyond the agent's belief even though the agent has chosen and completed an intentional series of actions.

The salient point of our work is that we present a general form of BDI agent programs, from which BDI-models can be generated and specifications in full BDI logics can be verified by symbolic model checking techniques.

*This work was supported by the Australian Research Council grant DP0452628, and partially by the NSFC grants 60496327, 10410638 and 60473004.

2 Computationally Grounded BDI Logic

We introduce a multimodal logic of belief, desire and intention, called OBDI logic, where the changes and computation of agents' beliefs, desires, and desires are based on agents' *observations* (i.e. local states). We assume the reader is familiar with the notion of interpreted system model, and we will follow the terminology of [Fagin *et al.*, 1995].

2.1 Syntax

Given a set Φ of propositional atoms, the language of OBDI logic is defined by the following BNF notations:

$$\begin{aligned}\langle wff\rangle ::= \quad & \text{any element of } \Phi \mid \neg\langle wff\rangle \mid \langle wff\rangle \wedge \langle wff\rangle \mid \\ & \bigcirc\langle wff\rangle \mid \langle wff\rangle \mathbf{U} \langle wff\rangle \mid \\ & B_i\langle wff\rangle \mid D_i\langle wff\rangle \mid I_i\langle wff\rangle\end{aligned}$$

Informally, $B_i\varphi$ and $D_i\varphi$ means that agent i believes and desires φ, respectively. While $I_i\varphi$ denotes that φ holds under the assumption that agent i acts based on his intention.

2.2 The BDI-system Model

Given a set G of global states and a system $\mathcal{K}$ over G, an agent's *mental state* over system $\mathcal{K}$ is a tuple $\langle \mathcal{B}, \mathcal{D}, \mathcal{I}\rangle$, where $\mathcal{B}$, $\mathcal{D}$ and $\mathcal{I}$ are systems (sets of runs over G) such that $\mathcal{I} \subseteq \mathcal{K}$ and $\mathcal{D} \subseteq \mathcal{B} \subseteq \mathcal{K}$. A *BDI-system* is a structure $\langle \mathcal{K}, \mathcal{M}_1, \cdots, \mathcal{M}_n\rangle$, where $\mathcal{K}$ is a system and for every i, $\mathcal{M}_i$ is agent i's mental state over $\mathcal{K}$.

Assume that we have a set Φ of primitive propositions. An *interpreted BDI-system* I consists of a pair $(\mathcal{S}, \pi)$, where $\mathcal{S}$ is a BDI-system and π is a valuation function, which gives the set of primitive propositions true at each point in G.

2.3 Semantics

In what follows, we inductively define the satisfaction relation $\models_{OBDI}$ between a formula φ and a pair of interpreted BDI-system and a point. Given an interpreted BDI-system $I = (\mathcal{S}, \pi)$, suppose that $\mathcal{S} = \langle \mathcal{K}, \mathcal{M}_1, \cdots, \mathcal{M}_n\rangle$ and for every i, $\mathcal{M}_i = \langle \mathcal{B}_i, \mathcal{D}_i, \mathcal{I}_i\rangle$. Let r be a run in $\mathcal{K}$ and u a natural number, then we have that:

- $(I, r, u) \models_{OBDI} B_i\varphi$ iff $(I, r', v) \models_{OBDI} \varphi$ for those $(r', v) \in \mathcal{B}_i$ such that $(r, u) \sim_i (r', v)$;
- $(I, r, u) \models_{OBDI} D_i\varphi$ iff $(I, r', v) \models_{OBDI} \varphi$ for those $(r', v) \in \mathcal{D}_i$ such that $(r, u) \sim_i (r', v)$;
- $(I, r, u) \models_{OBDI} I_i\varphi$ iff $(I, r', v) \models_{OBDI} \varphi$ for those $(r', v) \in \mathcal{I}_i$ such that $(r, u) \sim_i (r', v)$;

Proceedings of IJCAI 2005, Edinburgh, August, 2005.

Other cases are trivial and omitted here.

Proposition 1 *The following formulas are valid:*

- $X(\varphi \Rightarrow \psi) \Rightarrow (X\varphi \Rightarrow X\psi)$
  $X\varphi \Rightarrow YX\varphi$
  $\neg X\varphi \Rightarrow Y\neg X\varphi$
  where X and Y stand for B_i, D_i or I_i (for the same i).
- *Relationship between belief and desire*
  $B_i\varphi \Rightarrow D_i\varphi$
- *Temporal operators*
  $\bigcirc(\varphi \Rightarrow \psi) \Rightarrow (\bigcirc\varphi \Rightarrow \bigcirc\psi)$
  $\bigcirc(\neg\varphi) \Rightarrow \neg \bigcirc \varphi$
  $\phi \mathbf{U} \psi \Leftrightarrow \psi \vee (\varphi \wedge \bigcirc(\varphi \mathbf{U} \psi))$

2.4 Axiomatization

We give a proof system, called OBDI proof system, for those BDI-agents with perfect recall and a global clock, which contains the axioms of propositional calculus plus those formulas in Propositions 1. The proof system is closed under the propositional inference rules plus: $\frac{\vdash\varphi}{\vdash D_i\varphi}$ and $\frac{\vdash\varphi}{\vdash I_i\varphi}$ for every agent i.

Theorem 2 *The OBDI proof system is sound and complete with respect to interpreted BDI-systems with satisfaction relation $\models_{OBDI}$.*

3 Model Checking BDI-Agents

In order to make our model checking algorithm practically useful, we must consider where our model, an interpreted BDI-system comes from. To make the things simpler, we may consider some abstract programs such as *finite-state programs*, which are expressive enough from the standpoint of theoretical computer science. Moreover, to make our model checking system practically efficient, we use symbolic model checking techniques. Thus, a finite-state program in our approach is represented in a symbolic way.

3.1 Symbolic representation of interpreted BDI-agents

We formally define a (symbolic) *finite-state program with n agents* as a tuple $\mathcal{P} = (\mathbf{x}, \theta(\mathbf{x}), \tau(\mathbf{x}, \mathbf{x}'), O_1, \cdots, O_n)$, where $\mathbf{x}$ is a set of system variables; θ is a boolean formula over $\mathbf{x}$, called the *initial condition*; τ is a boolean formula over $\mathbf{x} \cup \mathbf{x}'$, called the *transition relation*; and for each i, $O_i \subseteq \mathbf{x}$, containing agent i's *local variables*, or *observable variables*. Given a state s, we define agent i's local state at state s to be $s \cap O_i$. We may associate with $\mathcal{P}$ an interpreted system $I_{\mathcal{P}} = (\mathcal{R}, \pi)$ called *the generated interpreted system of $\mathcal{P}$*.

For convenience, we may use $\mathcal{P}(\theta, \tau)$ to denote a finite-state program with n agents $(\mathbf{x}, \theta(\mathbf{x}), \tau(\mathbf{x}, \mathbf{x}'), O_1, \cdots, O_n)$, if $\mathbf{x}$ and $O_1, \cdots, O_n$ are clear from the context. Given a finite-state program $\mathcal{P}(\theta, \tau)$ with n agents, we define an *agent's internal program* (over $\mathcal{P}(\theta, \tau)$ as a tuple $\langle \mathcal{P}(\theta_1, \tau_1), \mathcal{P}(\theta_2, \tau_2), \mathcal{P}(\theta_3, \tau_3) \rangle$, where $(\theta_j \Rightarrow \theta) \wedge \tau_j \Rightarrow \tau$,for $j = 1, 2, 3$, and $(\theta_2 \Rightarrow \theta_1) \wedge \tau_2 \Rightarrow \tau_1$ are valid. Clearly, an agent's internal program is exactly related with an agent's mental state. Thus, we define a (symbolic) *BDI-program with n agents* as a tuple $P_A = (\mathcal{P}_K, P_1, \cdots, P_n)$, where $\mathcal{P}_K$ is a finite-state program with n agents and for each agent i, P_i is agent i's internal program over $\mathcal{P}$. We use I_{P_A} to denote the corresponding interpreted BDI-system.

3.2 Model checking OBDI logic

Theorem 3 *Given a BDI-program with n agents $P_A = (\mathcal{P}_K, P_1, \cdots, P_n)$, suppose that $\mathcal{P}_K = (\mathbf{x}, \theta(\mathbf{x}), \tau(\mathbf{x}, \mathbf{x}'), O_1, \cdots, O_n)$, and for every i, $P_i = \langle \mathcal{P}(\theta_1^i, \tau_1^i), \mathcal{P}(\theta_2^i, \tau_2^i), \mathcal{P}(\theta_3^i, \tau_3^i) \rangle$. Then, for every LTL formula φ and agent i, the following are valid in $I_{\mathcal{P}_A}$,*

1. $B_i\varphi \Leftrightarrow \forall(\mathbf{x} - O_i)(G(\mathcal{P}(\theta_1^i, \tau_1^i)) \Rightarrow \Gamma(\varphi, \theta_1^i, \tau_1^i))$
2. $D_i\varphi \Leftrightarrow \forall(\mathbf{x} - O_i)(G(\mathcal{P}(\theta_2^i, \tau_2^i)) \Rightarrow \Gamma(\varphi, \theta_2^i, \tau_2^i))$
3. $I_i\varphi \Leftrightarrow \forall(\mathbf{x} - O_i)(G(\mathcal{P}(\theta_3^i, \tau_3^i)) \Rightarrow \Gamma(\varphi, \theta_3^i, \tau_3^i))$

where $\Gamma(\varphi, \theta, \tau)$ is a boolean formula built from φ, θ, τ by using quantifications and fixed-point operations **lft** *and* **gft**.

Remark that Theorem 3 provides a reduction of OBDI to LTL, while Proposition 9 in [Su, 2004] gives an OBDD-based method of model checking LTL formulas. The complexity of our reduction of logic OBDI to LTL is $PSPACE$-complete. However, because quantifications of boolean functions and fixed-point operators can be dealt with in any OBDD package, the reduction can be based on OBDDs. In fact, we implemented a prototype of the OBDI model checker using CUDD, a very influential OBDD package developed by Fabio Somenzi, and achieved some preliminary experimental results.

4 Concluding Remarks

In this work, we have explored computationally grounded modal logics that characterize the internal attitudes of an agent–its beliefs, desires, etc, beyond S5 axioms and carried out a methodology on symbolic model checking for general BDI-agents.

References

[Cohen and Levesque, 1990] P.R. Cohen and H.J. Levesque. Intension is choice with commitment. *Artificial Intelligence*, 42:23–261, 1990.

[Fagin *et al.*, 1995] R. Fagin, J. Halpern, Y. Moses, and M. Vardi. *Reasoning about knowledge*. MIT Press, Cambridge, MA, 1995.

[Hintikka, 1962] J. Hintikka. *Knowledge and Belief.* Cornell University Press, Ithaca, NY, 1962.

[Rao and Georgeff, 1998] A.S. Rao and M.P. Georgeff. Decision procedures for BDI logics. *Journal of Logic and Computation*, 8(3):293–344, 1998.

[Rao, 1996] A.S. Rao. BDI agent speak out in a logical computable language. In *LNAI*, volume 1038, pages 42–55, 1996.

[Su, 2004] K. Su. Model checking temporal logics of knowledge in distributed systems. In *AAAI-04*, pages 98–103. AAAI, 2004.

[van der Hoek and Wooldridge, 2003] W. van der Hoek and M. Wooldridge. Towards a logic of rational agency. *L.J. of IGPL*, 11(2):135–159, 2003.

[Wooldridge, 2000] M. Wooldridge. Computationally grounded theories of agency. In E. Durfee, editor, *Proceedings of ICMAS-00*, pages 13–22. IEEE Press, 2000.