Updating Mental States from Informative Communication

Aldo Franco Dragoni
Istituto di Informatica
University of Ancona
Via Breccie Bianche
60131, Ancona - Italy
dragon@inform.unian.it

Paolo Giorgini
DISA
University of Trento
Via Inama, 5
38100, Trento - Italy
pgiorgini@cs.unitn.it

Luciano Serafini
ITC - IRST
Patè di Povo
38050, Trento - Italy
serafini@itc.it

1. INTRODUCTION

BDI agents (namely agents able to have Beliefs, Desires and Intentions) [9; 4] are supposed to be able to exchange information about: the environment, the agents' goals, intentions, and desires. Communication is also supposed to be intentional, i.e. activated by the speaker's reasoning about her own beliefs, desires and intention. In other words, it is generally possible to regard utterances as the "consequences" of the speaker's being in a particular mental state that provokes her desire to influence the hearer's mental state. This position is very general and is independent of the particular class of speech acts (Assertive, Commissive, Directive, Declarative or Expressive). If utterances are the effects of mental conditions it seems natural that a hearer tries to recognize the speaker's mental state through some form of backward reasoning (e.g. abduction) from the kind and the content of the received communication to the hypothetical mental state that originated it. If this were possible then we could develop logical theories that partially "predict" dialogue.

The main goal of this paper is to provide some abductive methods that a hearer can adopt to recognize the speaker's mental state. To do this we propose: i) a formal representation of mental states and ii) a causal theory which links mental states (the causes) to utterances (the effects).

2. AGENTS' MENTAL STATES

We use contexts to formalize agents' beliefs and intentions [1; 5; 6; 7]. A context α is a set of formulae closed under a set of inference rules (a theory). The beliefs of an agent i are represented by the context B_i and its intentions by a distinct context I_i. A formula ϕ in the context B_i ($B_i : \phi$) represents the fact that i believes ϕ and, analogously, a formula ϕ in the context I_i ($I_i : \phi$) represents the fact that i has the intention to bring about ϕ.

The language of a context α is denoted by L_α. It may be any kind of formal language (modal, first order, etc.), but here we will make use only of propositional ones.

To reason about beliefs and intentions of other agents, agent i recursively models the agents' mental states as pairs of contexts. In particular, i's beliefs about the beliefs and the intentions of an agent j are represented by the contexts B_iB_j and B_iI_j respectively. The same representation is used to formalize i's intentions regarding j's beliefs and intentions, that is the contexts I_iB_j and I_iI_j. Analogously i's beliefs about j's beliefs about another agent k are formalized by the pair of contexts $B_iB_jB_k$ and $B_iB_jI_k$. The formula ϕ in the context B_iI_j represents the fact that i believes that j has the intention to bring about ϕ

The interpretation of a formula depends on the context we consider. For instance, the formula ϕ in the context B_h, expresses the fact that agent h believes ϕ. The same formula in the context $B_hB_sI_h$ expresses the fact that h believes that s believes that h intends ϕ.

The interaction between contexts is formalized by *bridge rules*. Given a formula ϕ in a context α (precondition), a bridge rules produces another formula ψ in another context β (consequence). In particular, we use the following *reflection rules* [7]:

$$\frac{\alpha : B_i\phi}{\alpha B_i : \phi}\ \mathcal{R}_{dn.B} \qquad \frac{\alpha B_i : \phi}{\alpha : B_i\phi}\ \mathcal{R}_{up.B} \qquad \frac{\alpha : I_i\phi}{\alpha I_i : \phi}\ \mathcal{R}_{dn.I} \qquad \frac{\alpha I_i : \phi}{\alpha : I_i\phi}\ \mathcal{R}_{up.I}$$

and bridge rules between B_i and I_i to formalize the interaction between agent's beliefs and intentions.

Let MC be the multi-context system composed of the contexts and connected by the above bridge rules. Derivability in MC is defined in terms of Natural Deduction [8]. A wff $\alpha : \phi$ is derivable from a set of axioms AX in MC ($\mathsf{AX} \vdash_{\mathsf{MC}} \alpha : \phi$) if there is a deduction that ends with $\alpha : \phi$ and whose axioms are in AX. For a detailed description on the proof theory of MC we refer the reader to [7].

h's *mental state* is defined as the pair of sets containing her beliefs and her intentions derivable in the whole multi-context framework with the axioms AX. h's beliefs on s' beliefs and intentions, is called *h's image* of s' mental state and is composed of the pair formed by the set of h's beliefs on s' beliefs and the set of h's beliefs on s' intentions. Iterating the nesting, we define s' image of h's image of s' image of ... of h's image of s mental state.

3. MULTICONTEXT ABDUCTION AND REVISION

A *Simple Multi-Context Causal Theory* MT for a multi-context MC is a family $\{T_\alpha = \langle C_\alpha, E_\alpha, \Sigma_\alpha\rangle\}$ of Simple Causal Theories for each context α. We introduce the extra hypothesis that for all α, $E_\alpha \subseteq C_\alpha$. This is because we accept the fact that each effect can be regarded as the explanation of itself. A *causal literal* is defined to be a sentence of the form $\alpha : c$ or $\alpha : \neg c$, with $c \in C_\alpha$. We define $\mathbf{C} \triangleq \{\alpha : \sigma \mid \sigma \in C_\alpha\}$ $\Sigma \triangleq \{\alpha : \sigma \mid \sigma \in \Sigma_\alpha\}$ respectively, the *causes* and the *domain theory* of MT. An *Abductive Explanation* (ABE) for an observation $O \subseteq E_\alpha$ in a context α under the domain theory Σ, is a set $\mathbf{A} \subseteq \mathbf{C}$, such that:

1. $\forall\beta$, $\Sigma \cup \mathbf{A} \not\vdash_{\mathsf{MC}} \beta : \bot$. Or equivalently, $\mathbf{A}$ is consistent with Σ in any context.
2. $\Sigma \cup \mathbf{A} \vdash_{\mathsf{MC}} \alpha : O$. Or equivalently, the observation O can be derived in α from the set of axioms $\mathbf{A}$ and the domain theory Σ.
3. $\mathbf{A}$ is minimal w.r.t. the previous conditions, or equivalently, for any explanation $\mathbf{B}$ satisfying condition 1 and 2, $\mathbf{B} \not\subset \mathbf{A}$.

From the decidability of MC (see [7]) and the fact that the set of possible explanations is finite, we can conclude that the problem of finding an ABE of an observation O under the domain theory Σ is decidable.

In the following we provide specific criteria and methods for accomplishing such a task when the mental state represented by AX has been modified by a received communication. The change must consist on an observable effect O. These criteria are related to the nature of the speech act, i.e. to their preconditions (in the speaker's mental state) and their expected effects (on the hearer's mental state).

We suppose that $\mathsf{AX} = \Sigma \cup \mathbf{X}$, where $\mathbf{X} \subseteq \mathbf{C}$ is the only part that can be modified by the reception of the speech act (i.e., $\mathsf{AX}' = \Sigma \cup \mathbf{X}'$). To illustrate how $\mathbf{X}$ can be changed by the speech act let us introduce two operators: *abductive expansion* and *abductive contraction*.

Abductive Expansion Abductive Expansion, denoted by $+$, is applied when $\mathbf{X}$ must be extended to make it possible to derive the observation ϕ in a context α: $\mathbf{X} + \alpha : \phi \triangleq \mathbf{X} \cup \mathbf{A}$ where $\mathbf{A}$ is an ABE of $\alpha : \phi$ under the domain theory $\Sigma \cup \mathbf{X}$. If there are no ABEs for $\alpha : \phi$, then $\mathbf{X} + \alpha : \phi = \mathbf{X} \cup \{\alpha : \phi\}$ that is $\alpha : \phi$ is the explanation of itself. Expansion does not specify how $\mathbf{A}$ is chosen among the set of minimal explanations. This choice might be based on a partial order on ABEs. In this paper we do not consider the effect and the specific definition of such an order. This order strictly depends on the application domain and on the meaning and the degree of plausibility of the explanations. Similarly we do not consider methods to represent and compute such a partial order.

Abductive Contraction The operation of contraction, denoted by $-$, is applied if $\mathbf{X}$ is in contrast with a new observation. So, some formulas must be removed from $\mathbf{X}$. For any formula ϕ in a context α: $\mathbf{X} - \alpha : \phi \triangleq \mathbf{X} \setminus \mathbf{Y}$, where $\mathbf{Y}$ is a minimal hitting set[1] of $\{\mathbf{A}_1, \ldots, \mathbf{A}_n\}$, where $\{\mathbf{A}_1, \ldots, \mathbf{A}_n\}$ is the set of all the ABEs of ϕ in the context α under the domain theory Σ, which are contained in $\mathbf{X}$. Again we have not specified how the hitting set is chosen in all the $\mathbf{A}_i$. As before, this will be related to the ordering on the ABEs.

Abductive Revision The operation of revision, denoted by $*$, must be performed when $\mathbf{X}$ explains something which is inconsistent with the observation. For any formula ϕ in a context α, we define the operator $\mathbf{X} * \alpha : \phi = (\mathbf{X} - \alpha : \neg\phi) + \alpha : \phi$ This operative definition comes from Levi's identity (see [3; 2]).

Making the assumption that there is a causal relationship between a speaker's mental state and her uttering a sentence, we can use these opreators to update hearer's mental state and hearer's images of the speaker's mental state whenever she receives an utterance from a speaker.

4. REFERENCES

[1] A. Cimatti and L. Serafini. Multi-Agent Reasoning with Belief Contexts II: Elaboration Tolerance. In *Proc. 1st Int. Conference on Multi-Agent Systems (ICMAS-95)*, pages 57–64, 1996.

[2] A.F. Dragoni and P. Giorgini. Revising beliefs received from multiple source. In M A Williams and H Rott, editors, *Frontiers of Belief Revision*, Applied Logic. Kluwer, 1999.

[3] P. Gärdenfors. *Knowledge in Flux: Modeling the Dynamics of Epistemic States.* MIT Press., Cambridge Mass., 1988.

[4] M.P. Georgeff. Communication and interaction in multiagent planning. In *AAAI 83*, pages 125–129, 1983.

[5] F. Giunchiglia. Contextual reasoning. *Epistemologia, special issue on I Linguaggi e le Macchine*, XVI:345–364, 1993. Short version in Proceedings IJCAI'93 Workshop on Using Knowledge in its Context, Chambery, France, 1993, pp. 39–49.

[6] F. Giunchiglia and C. Ghidini. Local Models Semantics, or Contextual Reasoning = Locality + Compatibility. In *Proceedings of the Sixth International Conference on Principles of Knowledge Representation and Reasoning (KR'98)*, pages 282–289. Morgan Kaufmann, 1998.

[7] F. Giunchiglia and L. Serafini. Multilanguage hierarchical logics (or: how we can do without modal logics). *Artificial Intelligence*, 65:29–70, 1994.

[8] D. Prawitz. *Natural Deduction - A proof theoretical study.* Almquist and Wiksell, Stockholm, 1965.

[9] A.S. Rao and M. Georgeff. Bdi agents: from theory to prctice. In *Proc. of the First International Confernece on Multi-Agent Systems (ICMAS-95)*, pages 312–319, S. Francisco, CA, 1995.

[1]Given a collection of sets $S = \{\mathbf{A}_1, \ldots, \mathbf{A}_n\}$, a set $\mathbf{H}$ is a hitting set of S if for each $\mathbf{A}_i$, $\mathbf{H} \cap \mathbf{A}_i \neq \emptyset$. A hitting set is minimal if for any other hitting set $\mathbf{H}'$, $\mathbf{H}' \not\subseteq \mathbf{H}$.