A FORMAL DESCRIPTION OF AGENTS' EPISTEMIC STATES AND ENVIRONMENTS*

Yu Sun,[1,2] Cungen Cao[2] and Yuefei Sui[2]

[1]*School of Computer Science and Information Technology*
Yunnan Normal University, Kunming 650092
sunyu_km@hotmail.com

[2]*Key Laboratory of Intelligent Information Processing*
Institute of Computing Technology
Chinese Academy of Sciences, Beijing 100080
cgcao,yfsui@ict.ac.cn

Abstract To reason the interaction between agents' epistemic states and environments, the agents' epistemic states should include ontologies of agents (the agents' assumptions about environments), what the agents know, believe and desire, and also should include the agents' interpretations about these symbols in ontologies onto the environments. A formal description of such agents' epistemic states and several examples are given to show how such a description can be used to explain the puzzles in reasoning the modal sentences. For example, a sentence holding in a structure may have different consequences reasoned by different agents, because of the different epistemic states.

Keywords: Agent, Ontology, Epistemic state, Environment, Modality

1. Introduction

Multi-agent VSK logic ([1],[2]) is a multi-modal logic for reasoning about the epistemic states of agents in some environment. The logic is used to represent information visible, perceived and known to agents in an external environment and an agents' internal environment (epistemic state). Because the VSK logic is based on the propositional logic, it cannot express the interaction between agents' epistemic states and environments in which the agents are. What the agents perceive may change what the agents believe, know and

*This work is supported by the NSF (grant no. 60273019, 60496326, 60573063, and 60573064), the National 973 Programme (grants no. 2003CB317008 and G1999032701) and the Foundation of Yunnan Province (grants no. 04F00062, 2004YX42, 2004F0017Q, 03Y312D and 2005F0022Q).

desire; and what the agents' ontologies ([3],[4]) and interpretation of symbols in the ontologies may change the interpretation of statements holding in the environments.

An epistemic state of an agent in an environment includes statements about the real world, where the statements are interpreted in terms of the agent's own interpretation. It determines the interpretation of statements about the environment; and the environment changes what the agent perceives, hence, what the agent believes, knows and desires. Hence, an environment interacts with the epistemic states of the agent in the environment. Based on the above discussion, we shall give a logical description of agents' epistemic states and environments.

The paper is organized as follows: section 2 gives a formal definition of symbols representing the agents' epistemic states; in section 3, we give the interpretations of symbols and formulas in structures (environments), and the definition of inconsistency between what are true in structures and agents' epistemic states; section 4 uses examples to resolve the puzzles in modal logic. The last section concludes the paper.

2. The Epistemic States

An epistemic state E_a of an agent a should include the following ingredients:

• a logical language L which is sharable and common to every agent.

• an ontology O_a of agent a, where O_a consists of a set $C_a \subseteq L$ of names (concepts, denoted by α, β, etc.), a binary relation $\sqsubseteq$ on C_a (the subsumption relation), and a function F^a such that every name (concept) α is associated with a frame F^a_α, where F^a_α contains statements (formulas in L) about α which are known or assumed by agent a. Define $\alpha \equiv \beta$ if $\alpha \sqsubseteq \beta$ and $\beta \sqsubseteq \alpha$.

• a set $\mathbf{K}_a$ of statements known to agent a; $\mathbf{B}_a$ believed by agent a; $\mathbf{D}_a$ desired by agent a.

• an interpretation function $\mathbf{I}_a$ such that for any structure M (an environment), there is an interpretation I_a maps a name in L onto M.

• a set $\mathbf{W}$ of possible world (structure) names $\mathbf{M}$.

Assume that there is an ontology O and interpretation $\mathbf{I}$ independent of any agent.

Remark 2.1. The statements in F^a_α are different from these in $\mathbf{K}_a$ in that when agent a communicates with other agents, a assumes that F^a_α is known to other agents; but $\mathbf{K}_a$ may not be known to other agents.

Here, we assume that L and structures M are independent of agents. Any formula in L is called a statement.

The language $\mathcal{L}$ for representing epistemic states of agents consists of

• L as a sub-language;

- a set $\mathcal{A}$ of agents;
- epistemic modals $\mathbf{K}_a, \mathbf{B}_a$ and $\mathbf{K}_a$ for every $a \in \mathcal{A}$;
- a symbol $\mathbf{I}_a$ for every $a \in \mathcal{A}$;
- a set $\mathbf{W}$ of possible world names.

Definition 2.2. A string t of symbols in $\mathcal{L}$ is a term if either

(i) t is a term in L, or

(ii) $t = \mathbf{I}_a(s)$, where s is a term in L, or

(iii) $t = \mathbf{M}$ for some $\mathbf{M} \in \mathbf{W}$.

For any $a \in \mathcal{A},\ a$ should be a term. Because a occurs only as a subscript of $\mathbf{K}_a, \mathbf{B}_a, \mathbf{D}_a$ or $\mathbf{I}_a,\ a$ is not taken as a term any more.

Definition 2.3. A string φ of symbols is a formula in $\mathcal{L}$ if either

$\circ$ φ is a statement in L (given a structure M and an interpretation $I,\ \varphi$ is interpreted to be a property on M); or

$\circ$ $\varphi = \mathbf{K}_a\psi, \mathbf{B}_a\psi, \mathbf{D}_a\psi$, where ψ is a statement in L or formula in $\mathcal{L}$ (given a structure M and an interpretation $I_a,\ \mathbf{K}_a, \mathbf{B}_a$ and $\mathbf{D}_a$ are interpreted to be sets of properties about M); or

$\circ$ $\varphi = \mathbf{I}_a(t) = \mathbf{I}_a(s)$, or $\mathbf{I}_a(\psi)$, where t, s are terms in L and ψ is a formula in $\mathcal{L}$ (given a structure $M,\ \mathbf{I}_a$ is interpreted as an interpretation I_a, hence, $\mathbf{I}_a$ can be taken as a function from structures to interpretations); or

$\circ$ $\varphi = \neg\psi, \psi_1 \to \psi_2$, where ψ, ψ_1 and ψ_2 are statements in L or formulas in $\mathcal{L}$.

We can list some axioms about modals, such as: $\mathbf{K}_a\varphi \to \mathbf{B}_a\varphi$, etc.

Remark 2.4. $\mathbf{K}_a, \mathbf{B}_a, \mathbf{D}_a$ are not taken as modalities as in the BDI logic, but as sets of sentences. There are two reasons: one is to avoid using the possible world semantics, which is not compatible very well with our intuition that what a believes is just a set of statements. Another reason is that there is no appropriate modal predicate logic for the multi-agent systems, because of the propositional attitude reports ([5]).

3. Interpretations I and I_a

Given a structure M, assume that agent a is among M, that is, a is an object in M. There is an interpretation I independent of any agents, and an interpretation I_a for an agent a, where I_a satisfies the following conditions:

(3.1) $I_a(\mathbf{K}_a), I_a(\mathbf{B}_a)$ and $I_a(\mathbf{D}_a)$ are consistent sets of formulas such that $I_a(\mathbf{K}_a) \subseteq I_a(\mathbf{B}_a)$;

(3.2) $I_a(\mathbf{I}_a) = I_a$.

Then, there is an interpretation $\mathcal{I}$ of $\mathcal{L}$ such that $\mathcal{I}(\mathbf{I}) = I$; and $\mathcal{I}(\mathbf{I}_a) = I_a$.

Let $Th^I(M)$ be the statements in L which are true in M under interpretation I, that is, $Th^I(M) = \{\varphi \in L : M, I \models \varphi\}$. For the convenience, we assume that $Th^{I_a}(M) = Th^{I_a}(P_{a,M})$, where $P_{a,M}$ is a sub-structure of M that agent a can perceive. I_a maps every concept α in C_a to be a set of objects in M, denoted by $I_a(\alpha)$, such that for any $a \in M$, if $a \in I_a(\alpha)$ then we say that a is an instance of α in M under interpretation I_a.

Definition 3.1. Given a sentence φ and a structure M, we define the satisfaction of φ in M as follows:

If φ is a statement in L then $M \models \varphi$ if $\varphi \in Th^I(M)$;
$M, I_a \models \alpha \sqsubseteq \beta$ if $I_a(\alpha) \subseteq I_a(\beta)$;
$M, I_a \models \mathbf{K}_a\varphi$ if $\varphi \in I_a(\mathbf{K}_a)$;
$M, I_a \models \mathbf{I}_a(t) = \mathbf{I}_a(s)$ if $I_a(t) = I_a(s)$;
$M, I_a \models \mathbf{I}_a(\varphi)$ if $I_a(\varphi) \in Th^{I_a}(M)$;

What a perceives should be represented in the agent's epistemic states. Hence, for any object $x \in P_{a,M}$, there is a new name $\mathbf{x}_{a,M}$ added in O_a, and $Th^{I_a}(P_{a,M})$ is added in $\mathbf{K}_a$ and O_a. Two agents a and b can communicate information about $\mathbf{x}_{a,M}$ if (1) $a, b \in P_{a,M}, P_{b,M}$; and (2) $x \in P_{a,M}, P_{b,M}$. In this case, $\mathbf{x}_{b,M}$ is added in O_b too.

Remark 3.2. Even when the universes in $P_{a,M}$ and $P_{b,M}$ are equal, $P_{a,M}$ may not be equal to $P_{b,M}$. Because a relation symbol $\mathbf{r}$ may have different interpretations $I_a(\mathbf{r})$ and $I_b(\mathbf{r})$ on $P_{a,M}$.

Let $I_a(\mathbf{x}_{a,M}, M) = x$; and assume that $I_a(\mathbf{x}_{a,M}, M')$ may be undefined.

Notice that $\mathbf{K}_a\varphi \to \varphi$ is not an axiom, because φ is to be interpreted by a common interpretation I and $\mathbf{K}_a\varphi$ is interpreted by I_a, and $\mathbf{K}_a\varphi$ being satisfied under I_a does not imply the satisfaction of φ under I.

Given a structure M, I_a interprets L onto M. What an agent should notice is the difference and consistency between $\mathbf{I}_a(\mathbf{K}_a)$ and $Th^{I_a}(M)$, where $\mathbf{I}_a(\mathbf{K}_a) = \{\mathbf{I}_a(\varphi) : \varphi \in \mathbf{K}_a\}$. An autonomic agent could revise $\mathbf{B}_a, \mathbf{K}_a$ and $\mathbf{D}_a$ according to the inconsistency of $\mathbf{I}_a(\mathbf{K}_a) \cup Th^{I_a}(P_{a,M})$. If a is retrospective then if $\mathbf{I}_a(\mathbf{K}_a)$ and $Th^{I_a}(M)$ are inconsistent then a revises $\mathbf{K}_a$ such that $\mathbf{I}_a(\mathbf{K}_a) \subseteq Th^{I_a}(M)$. If it is necessary then a may revise O_a.

Remark 3.3. Notice that the revision done by a is not based on the inconsistence of $\mathbf{I}_a(\mathbf{K}_a)$ and $Th^I(M)$, but on the inconsistence of $\mathbf{I}_a(\mathbf{K}_a)$ and $Th^{I_a}(M)$. Because a knows that there is an incorrect statement in $\mathbf{K}_a, \mathbf{B}_a$ or $\mathbf{D}_a$, not because of the inconsistence between $\mathbf{I}_a(\mathbf{K}_a)$ and $Th^I(M)$ (a does not know $Th^I(M)$), but because of the inconsistence between $\mathbf{I}_a(\mathbf{K}_a)$ and $Th^{I_a}(M)$. What a perceives in M is $Th^{I_a}(P_{a,M})$, not $Th^I(P_{a,M})$.

4. Examples

In this section we give several examples to show how to use the description given above to reason what agents' know and what we know the agents' know.

Example 4.1. Let M_m be a snapshot of the real world in the morning, and M_e in the evening. Assume that $\alpha = morning\ star \in C_a$ and $\beta = evening\ star \in C_a$, and $\alpha \equiv \beta \notin O_a$. It is a basic fact that

$$M_m, I \models \alpha = \beta; \quad M_e, I \models \alpha = \beta.$$

For agent a, $I_a(\alpha)$ is defined in M_m and not defined in M_e; and $I_a(\beta)$ is defined in M_e and not defined in M_m. Hence, we have that

$$M_m, I_a \models \alpha \neq \beta; \quad M_e, I_a \models \alpha \neq \beta.$$

Because $I(\alpha) \in P_{a,M_m}$ and $I(\beta) \notin P_{a,M_m}$, agent a does not know $\alpha = \beta$ by perceiving α, so that agent a think that a need not revise $\mathbf{K}_a$ and O_a, because

$$\alpha \neq \beta \in \mathbf{K}_a, \quad \alpha \in \mathbf{M}_m \in \mathbf{K}_a, \quad \beta \notin \mathbf{M}_m \in \mathbf{K}_a.$$

In the common ontology O, we assume $\alpha, \beta \in C$, and $\alpha \sqsubseteq \beta, \beta \sqsubseteq \alpha \in O$. Notice that O may not be a perfect and complete ontology for the real world.

Even though the logical language L used to represent the real world is same to agents, conflicts and misunderstanding between agents may be resulted in by the difference in the following factors: (1) the statements about a concept; (ii) $\mathbf{K}_a, \mathbf{B}_a, \mathbf{D}_a$ and $\mathbf{I}_a$; and (iii) the environments perceived.

Because of the differences, a statement φ in L may be interpreted in different ways for different agents, i.e., it is possible that for some agents a and $b, M, I \models \varphi$; $M, I_a \models \varphi$ and $M, I_b \not\models \varphi$.

Example 4.2. Given a structure M and two agents a and b, it is possible that $P_{a,M} \neq P_{b,M}, O_a \neq O_b, I_a \neq I_b, F^a_\alpha \neq F^b_\alpha$ for some $\alpha \in L$. Hence, misunderstanding occurs between a and b.

Assume that α, β, M_m and M_e are the same as example 4.1; and $\alpha \equiv \beta \notin O_a, \alpha \equiv \beta \in O_b$. Let I_a and I_b be two interpretations such that $\mathcal{I}(\mathbf{I}_a) = I_a$ and $\mathcal{I}(\mathbf{I}_b) = I_b$. Then, I_a is a partial mapping and I_b is a total function, because $I_b(\alpha) \in M_e$ and $I_b(\beta) \in M_m$.

If agent b tells agent a in M_m that b perceives β in M_m, agent a does not believe in b, i.e.,

$$\begin{array}{ll} I_a(\alpha) \notin M_e; & I_a(\beta) \notin M_m; \\ M_m, I_b \models S_b(\beta, \mathbf{M}_m); & M_m, I_a \not\models \mathbf{B}_a S_b(\beta, \mathbf{M}_m), \end{array}$$

where $S_b(\beta, \mathbf{M}_m)$ means that b can see β in $\mathbf{M}_m$. Because agent a think it is impossible to perceive β in M_m.

Hence, the statement that b see the evening star in the morning is true in M_m for agent b, but false in M_m for agent a.

To explain that even when the universes in $P_{a,M}$ and $P_{b,M}$ are equal, as a structure, $P_{a,M}$ may not be equal to $P_{b,M}$, we use the following

Example 4.3. Let M be a structure consisting of a cubic object x which has one hole in three faces and no hole in other three faces. Assume that three faces with one hole, say x_1, x_2, x_3, are perceived only by agent a; and other three faces without hole, say x_4, x_5, x_6, are perceived only by agent b. Hence, defaultly, agent a think that every face of the cube has one hold on it, let $h_1, h_2, ..., h_6$ be the holes on $x_1, x_2, ..., x_6$ imaged by agent a; and agent b

think that there is no hole on any face of the cube. Hence,

$$\mathbf{x}_{a,M} = \mathbf{x}_{b,M}, \mathbf{x}_{a,M} \in O_a, \mathbf{x}_{b,M} \in O_b;$$
$$\mathbf{x}_{1,a,M}, \mathbf{x}_{2,a,M}, ..., \mathbf{x}_{6,a,M} \in O_a; \;\; \mathbf{x}_{1,b,M}, \mathbf{x}_{2,b,M}, ..., \mathbf{x}_{6,b,M} \in O_b;$$
$$\mathbf{x}_{1,a,M} \equiv \mathbf{x}_{1,b,M}, \mathbf{x}_{2,a,M} \equiv \mathbf{x}_{2,b,M}, ..., \mathbf{x}_{6,a,M} \equiv \mathbf{x}_{6,b,M} \in L;$$
$$\mathbf{h}_{1,a,M}, \mathbf{h}_{2,a,M}, ..., \mathbf{h}_{6,a,M} \in O_a, \notin O_b.$$

Let $on(x, y)$ denote a relation symbol that x is on y. Then,

$$on(\mathbf{h}_{1,b,M}, \mathbf{x}_{1,b,M}), ..., on(\mathbf{x}_{6,b,M}, \mathbf{x}_{6,b,M}) \in \mathbf{K}_a;$$
$$on(\mathbf{h}_{1,b,M}, \mathbf{x}_{1,b,M}), ..., on(\mathbf{x}_{6,b,M}, \mathbf{x}_{6,b,M}) \notin \mathbf{K}_b.$$

Notice that because $\mathbf{h}_{1,b,M} \notin O_b$, it is meaningless to b that $on(\mathbf{h}_{1,b,M}, \mathbf{x}_{1,b,M})$.

5. Conclusions

According to the difference of ontologies and interpretations of different agents, we can represent the agent' epistemic states in a more natural way which is similar to the way we perceive and revise our knowledge. The agents are in environments and perceive some properties holding in the environments, where the properties are interpreted by the agents under their own interpretations. Such a representation of epistemic states gives a natural way to describe the belief revision. An agent realizes to revise its belief set when the agent in some environment interprets its beliefs in an inconsistent way. Hence, the agent may not revise its belief set once when its belief set is inconsistent with the properties holding in a structure, as explained in the classical theory of belief revision, because the inconsistence may not be perceivable to the agent.

Further works should include the axiom systems for interpretations I and I_a; the theory of belief revision based on interpretations I and I_a; and the applications to emotional agents and the representation of speech acts.

References

[1] M. Wooldridge and A. Lomuscio, A computationally grounded logic of visibility, perception, and knowledge, *Logic Journal of the IGPL* 9(2001), 273-288.

[2] M. Wooldridge and A. Lomuscio, Multiagent VSK logic, in: *Proc. of the Seventh European Workshop on Logics in Artificial Intelligence (JELIAI-2000)*, Springer-Verlag.

[3] T. R. Gruber, Towards principles for the design of ontologies used for knowledge sharing, *International J. of Human-Computer Studies* 43(1995), 907-928.

[4] N. Guarino, Understanding, building, and using ontologies, *International J. of Human-Computer Studies* 46(1997), 293-310.

[5] G. E. Hughes and M. J. Cresswell, *A New Introduction to Modal Logic*, Routledge, London, New York, 1996.