UDC: 007.52
Original scientific paper

REASONING ABOUT KNOWLEDGE: SOME AGENT PROPERTIES

Mirko Maleković

University of Zagreb, Faculty of Organization and Informatics, Varaždin, Croatia
E-mail: mmalekov@foi.hr

In this paper we characterize some agent properties relating to reasoning about knowledge. The following facts are proved: if a reflexive agent has F, then F holds; if a transitive agent knows F, then he knows that he knows F (the positive introspection property); if an Euclidean agent does not know F, then he knows that he does not know F (the negative introspection property); a serial agent does not know F if F is a contradiction. Also, the concept of the more knowledgeable agent is introduced.

Keywords: agent properties, Euclidean agent, reflexive agent, serial agent, symmetric agent, transitive agent, reasoning about knowledge.

1. INTRODUCTION

The idea of a formal logical analysis of reasoning about knowledge (what do we know?, what can be known?, what does it mean that someone knows something?) is described in [1], [2], [3], [4], [5], and [6]. In this paper, we shall characterize in detail some important agent properties. We shall describe the following agent types: reflexive agent, transitive agent, Euclidean agent, symmetric agent, and serial agent. Also, we shall characterize what it means when one agent is more knowledgeable than another agent. Finally, we shall consider the situation where an agent is not reflexive. Such an agent can know F though F does not hold.

The paper consists of four sections and an appendix. In Section 2 , we introduce the basic notions of reasoning about knowledge. In Section 3, we characterize in detail agent types. We prove:

(ref) if a reflexive agent knows a formula F, then F is true;

(tra) if a transitive agent knows F, then he knows that he knows F;

(mka) an agent i is more knowledgeable than an agent j if $ki \subseteq kj$, where ki and kj are the possibility relations of the agent i and the agent j , respectively.

We also show that if an agent is not reflexive, then he can know something that does not hold.

The following propositions:

(Euc) if an Euclidean agent does not know F, then he knows that he does not know F;

(ser) a serial agent does not know F if F is a contradiction;
(symm) a symmetric agent knows that he does not know $\neg F$ if F is true;
(mp) an agent knows all the logical consequences of his knowledge, that is, if an agent knows F and $F \Rightarrow G$, then he knows G too;
are proved in the Appendix. Section 4 contains conclusions.

2. BASIC NOTIONS

In this section, we introduce some basic concepts and notations.

Suppose we have a group consisting of m agents, named 1, 2,.., m. An agent may be a man (a real agent), a software module or a communicating robot (an artificial agent). An agent may even be a component of a computer system (a wire or a message buffer). We assume these agents wish to reason about a world that can be described in terms of a nonempty set P of primitive propositions. A language is just a set of formulas, where the set of formulas LK, of interest to us, is defined as follows.

(1) The primitive propositions in P are formulas;

(2) If F and G are formulas, then so are $\neg F$, $(F \wedge G)$, $(F \vee G)$, $(F \Rightarrow G)$, $(F \Leftrightarrow G)$, and

Ki(F) for all $i \in \{1, 2, .., m\}$, where Ki is a modal operator.

A Kripke structure M for agent group {1, 2,.., m} over P is a $(m + 2)$-tuple

M = (S, I, k1, k2,.., km), where S is a set of possible worlds, I is an interpretation that associates with each world in S a truth assignment to the primitive propositions in P, and k1, k2,.., km are binary relations on S, called the possibility relations for agents 1, 2,.., m, respectively.

Given $p \in P$, the expression I[w](p) = true means that p is true in a world w in a structure M. The fact that p is false, in a world v of a structure M, is indicated by the expression I[v](p) = false

The expression $(u, v) \in ki$ means that an agent i considers a world v possible, given his information in a world u. Since ki defines what worlds an agent i considers possible in any given world, ki will be called the possibility relation of the agent i.

We will now define what it means for a formula to be true at a given world in a structure.

Let $(M, w) \models F$ mean that F holds or is true at (M, w). The definition of $\models$ is as follows:

(a) $(M, w) \models p$ iff I[w](p) = true, where $p \in P$;
(b) $(M, w) \models F \wedge G$ iff $(M, w) \models F$ and $(M, w) \models G$;
(c) $(M, w) \models F \vee G$ iff $(M, w) \models F$ or $(M, w) \models G$;
(d) $(M, w) \models F \Rightarrow G$ iff $(M, w) \models F$ implies $(M, w) \models G$;
(e) $(M, w) \models F \Leftrightarrow G$ iff $(M, w) \models F \Rightarrow G$ and $(M, w) \models G \Rightarrow F$;

(f) $(M, w) \models \neg F$ iff $(M,w) \not\models F$, that is, $(M, w) \models F$ does not hold;
(g) $M \models F$ iff $(M, w) \models F$ for all $w \in S$.

Finally, we shall define a modal operator Ki, where Ki(F) is read: Agent i knows F.

(h) $(M, w) \models Ki(F)$ iff $(M, t) \models F$ for all $t \in S$ such that $(w, t) \in ki$.

In (h) we have that an agent i knows F in a world w of a structure M exactly if F holds at all worlds t that agent i considers possible in w.

Example

We have a deck consisting of three cards labelled A, B, and C. Agents 1 and 2 each get one of these cards; the third card is left face down. The possible world is characterized by describing the cards held by each agent. A world w = (X, Y) indicates that agent 1 holds a card X and agent 2 holds a card Y.

There are six possible worlds:

w1 = (A, B) : 1 holds A, 2 holds B;
w2 = (B, A) : 1 holds B, 2 holds A;
w3 = (A, C) : 1 holds A, 2 holds C;
w4 = (C, A) : 1 holds C, 2 holds A;
w5 = (B, C) : 1 holds B, 2 holds C;
w6 = (C, B) : 1 holds C, 2 holds B.

The card which is left face down in a world wi is evident. Consequently, S = {w1, w2,.., w6} is the set of the possible worlds.

Let us now introduce the primitive propositions:

1A : 1 holds the card A, 1B : 1 holds the card B, 1C : 1 holds the card C

2A : 2 holds the card A, 2B : 2 holds the card B, 2C : 2 holds the card C.

Accordingly, P = {1A, 1B, 1C, 2A, 2B, 2C} is the set of the primitive propositions.

Interpretation I defines a truth assignment for the primitive propositions in p, for each world in S.

We have, for example, I[w1](1A) = true, that is, the proposition 1A holds in the world

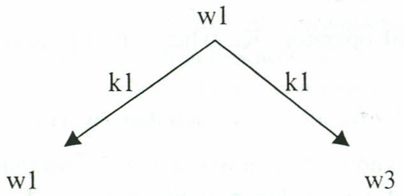
w1 = (A, B).

I[w5](2B) = false, that is, the proposition 2B is false in the world w5 = (B, C).

In the world w1 = (A, B), agent 1 thinks that two worlds are possible: w1 itself and

w3 = (A, C). Namely, in w1 = (A, B), agent 1 knows that he has the card A, but he considers it possible that agent 2 could hold either card B or card C.

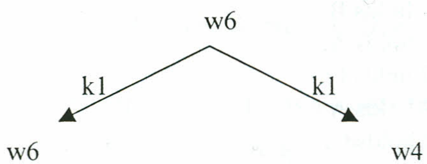
The above mentioned can be represented graphically as follows



k1 is the possibility relation for the agent 1:

k1 = {(w1, w1), (w1, w3), (w2, w2), (w2, w5), (w3, w3), (w3, w1), (w4, w4), (w4, w6),
(w5, w5), (w5, w2), (w6, w6), (w6, w4)}

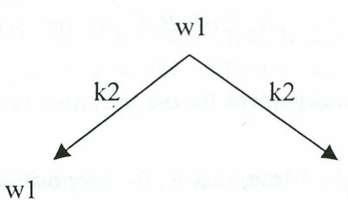
For example, in the world w6 = (C, B), agent 1 considers the two worlds w6 = (C, B) and w4 = (C, A) are possible. Graphically,



The possibility relation k2 for the agent 2 is

k2={(w1, w1), (w1, w6), (w2, w2), (w2, w4), (w3, w3), (w3, w5), (w4, w4), (w4, w2), (w5, w5), (w5, w3), (w6, w6), (w6, w1)}.

We can see from k2 that agent 2 in the world w1 = (A, B) thinks the two worlds w1 = (A, B) and w6 = (C, B) are possible. Graphically,



The Kripke structure M for our two agents 1 and 2 over P is the 4-tuple

M = (S, I, k1, k2). We can now describe the relation $\models$; we have:

(M, w1) $\models$ 1A because I[w1](1A) = true;

(M, w1) $\models$ 2B because I[w1](2B) = true;

(M, w1) $\not\models$ 1B because I[w1](1B) = false.

Additionally, (M, w1) $\models$ 1A $\vee$ 1B because (M, w1) $\models$ 1A ,

$(M, w5) \not\models 1B \wedge 2A$ because $(M, w5) \not\models 2A$, where $w5 = (B, C)$.

All the worlds that agent 1 considers possible worlds in the world w1 are w1 and w3.

Because $M, w1) \models 1A$ and $(M, w3) \models 1A$, we obtain $(M, w1) \models K1(1A)$.

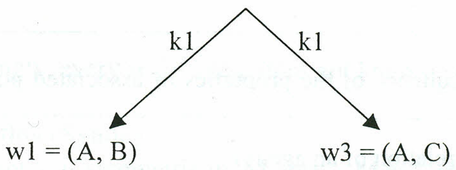
Consequently, we have that agent 1 knows 1A in the world w1.

Problem

Does $(M, w1) \models K1(2B)$ hold?

Solution

We have the graph $w1 = (A, B)$



Because $(M, w3) \not\models 2B$, we can conclude $(M, w1) \not\models K1(2B)$.

The result says: In the world w1 = (A, B), agent 1 does not know the proposition 2B (agent 1 does not know which card agent 2 holds).

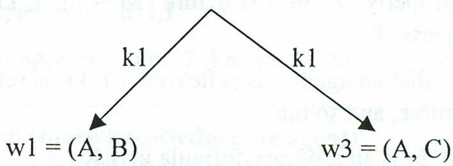
Problem

Does $(M, w1) \models K1(K1(2B))$ hold ?

Solution

We test whether agent 1 knows that agent 2 knows 2B .

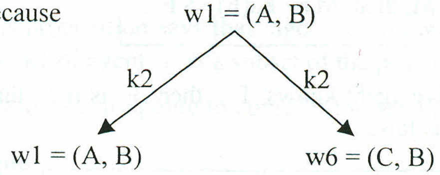
We have the graph $w1 = (A, B)$



Therefore, we need to solve $(M, w1) \models K2(2B)$ and $(M, w3) \models K2(2B)$.

First, we solve $(M, w1) \models K2(2B)$.

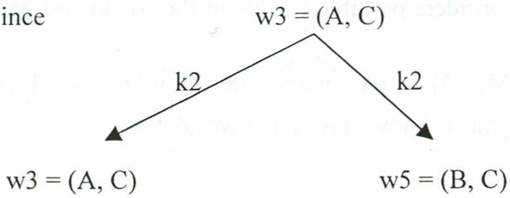
Because $w1 = (A, B)$



and because $(M, w1) \models 2B$ and $(M, w6) \models 2B$, we have that $(M, w1) \models K2(2B)$ holds.

Now, we test $(M, w3) \models K2(2B)$.

Since



and since $(M, w3) \not\models 2B$, we obtain $(M, w3) \not\models K2(2B)$.

Consequently $(M, w1) \not\models K1(K2(2B))$, that is, in the world w1 agent 1 does not know that agent 2 knows 2B.

3. AGENT PROPERTIES

What an agent knows is a consequence of the properties of associated possibility relation.

Let $ki \subseteq S \times S$ be a possibility relation of an agent i.

Definition1

(Ref) ki is reflexive iff (for all $t \in S$)[$(t, t) \in ki$]
(Symm) ki is symmetric iff (for all $u, v \in S$)[$(u, v) \in ki$ implies $(v, u) \in ki$]
(Tra) ki is transitive iff (for all $t, u, v \in S$) [$(t,u) \in ki$ and $(u,v) \in ki$ implies $(t,v) \in ki$]
(Euc) ki is Euclidean iff (for all $t,u, v \in S$)[$(t,u) \in ki$ and $(t,v) \in ki$ implies $(u,v) \in ki$]
(Ser) ki is serial iff (for all $t \in S$)(for some $u \in S$)[$(t, u) \in ki$]

Definition2

We say that an agent i has a property P in a structure $M = (S, I, ki)$ iff his possibility relation ki has the property P.

Definition2 means, for instance, that an agent i is reflexive iff ki is reflexive, an agent i is transitive iff ki is transitive, and so on.

From this point we suppose that F is an arbitrary formula in LK.

Proposition (Ref)
If an agent i is reflexive in a structure M, then $M \models Ki(F) \Rightarrow F$.

Proposition (Ref) states that if a reflexive agent knows F , then F is true, that is, a reflexive agent does not know F if F is false.

Proposition (Tra)
Let an agent i be transitive in a structure M. Then $M \models Ki(F) \Rightarrow Ki(Ki(F))$.

Proposition (Tra) declares that if a transitive agent knows F, then he knows that he knows F, that is, every transitive agent has the positive introspection property.

Proposition (Euc)
If an agent i is Euclidean in M, then $M \models \neg Ki(F) \Rightarrow Ki(\neg Ki(F))$.

Proposition (Euc) states that every Euclidean agent has the negative introspection property, that is, if he does not know F, then he knows that he does not know F.

Proposition (Ser)
If an agent i is serial in M, then $M \models \neg Ki(False)$.

Accordingly, every serial agent does not know a contradiction, named False.

Proposition (Symm)
If an agent i is symmetric in M, then $M \models F \Rightarrow Ki(\neg Ki(\neg F))$.

Proposition (Symm) states that every symmetric agent knows that he does not know $\neg F$ if F is true.

If we consider a two-agent group $G = \{1, 2\}$, then it is interesting to ask which agent is more knowledgeable.

Definition3

Let $M = (S, I, k1,k2)$ be a Kripke structure for a two-agent group $G = \{1, 2\}$.

Agent 1 is more knowledgeable in M than agent 2 iff $M \models K2(F) \Rightarrow K1(F)$, for each formula F in LK.

Accordingly, if agent 2 knows F, then agent 1 knows F too.

Proposition (more knowledgeable agent)
Let $M = (S, I, k1, k2)$ be a Kripke structure for agents 1 , and 2 .
If $k1 \subseteq k2$, then agent 1 is more knowledgeable than agent 2 .

This proposition says that agent 1 knows more than agent 2 if the possibility relation k1 of agent 1 is a subset of the possibility relation k2 of agent 2.

Finally, we are going to consider the situation where an agent is not reflexive.

Example

Let $M = (S, I, ki)$ be the structure of an agent i , where $S = \{s, t\}$, $ki = \{(s, s), (s, t), (t, s)\}$, and $I[s](F) = true$, $I[t](F) = false$, for some formula F.

Agent i is not reflexive because $(t, t) \notin ki$. We shall show that

$(M, t) \models Ki(F) \Rightarrow F$ does not hold.

It is easy to see that $(M, t) \models Ki(F)$, and $(M, t) \not\models F$. Consequently,

$(M, t) \models Ki(F) \Rightarrow F$ does not hold.

The result is very interesting as it states that an agent, who is not reflexive, can know something that is not true. The reason for this is the fact that agent i, in the world t, does not consider the world t is possible. 'He does not believe his eyes'.

We also have $M \not\models Ki(F)) \Rightarrow F$. In addition, we have $(M \models Ki(F)) \Rightarrow (M \models F\ )$ because

$M \not\models Ki(F)$.

Consequently, $M \models Ki(F) \Rightarrow F$ is different from $(M \models Ki(F)) \Rightarrow (M \models F)$.

4. CONCLUSIONS

We have described the reflexive, transitive, symmetric, Euclidean and serial agents. We have stated the proposition (Ref) [if a reflexive agent knows F, then F holds], the proposition (Tra) [if a transitive agent knows F, then he knows that he knows F (the positive introspection property)], the proposition (Euclidean) [if an Euclidean agent does not know F, then he knows that he does not know F (the negative introspection property)], the proposition (Ser) [a serial agent does not know F if F is a contradiction], and the proposition (Symm)[a symmetric agent knows that he does not know $\neg F$ if F holds]. Finally, we have characterized what it means when one agent is more knowledgeable than another agent. We have stated the proposition (more knowledgeable agent) [an agent i (with a possibility relation ki) is more knowledgeable than an agent j (with a possibility relation kj) if $ki \subseteq kj$] . In addition, we have considered the situation where an agent is not reflexive. Such an agent can know F though F does not hold. The proofs of the stated propositions are given in the Appendix.

As the problems of integrating knowledge bases are very severe, there is a need for a class of information sources (agent architecture) that stand between the user and the heterogeneous knowledge bases. It is clear that the properties of the agent types introduced previously are very important in synthesising an agent architecture.

In a forthcoming paper, we shall investigate the problem of knowledge integration for a group of agents that is also important in building an agent architecture.

APPENDIX

Proposition (Ref)

If an agent i is reflexive in a structure M, then $M \models Ki(F) \Rightarrow F$.

Proof

Let $w \in S$ be an arbitrary world. We have to show $(M, w) \models Ki(F) \Rightarrow F$.

Suppose $(M, w) \models Ki(F)$. We need to prove $(M, w) \models F$.

Since the agent i is reflexive, we have $(w, w) \in ki$. Now, from $(M,w) \models Ki(F)$ and $(w, w) \in ki$, we obtain $(M, w) \models F$.

Proposition (Tra)

Let an agent i be transitive in a structure M. Then $M \models Ki(F) \Rightarrow Ki(Ki(F))$.

Proof

We would like to prove $(M, w) \models Ki(F) \Rightarrow Ki(Ki(F))$, where $w \in S$ is an arbitrary world in S. Assume (M, w) Ł$\models Ki(F)$. We have to show $(M, w) \models Ki(Ki(F))$.

Since $(M, w) \models Ki(Ki(F))$ iff $(M, t) \models Ki(F)$ for all $t \in S$ such that $(w, t) \in ki$, and $(M, t) \models Ki(F)$ iff $(M, u) \models F$ for all $u \in S$ such that $(t, u) \in ki$, we have to prove $(M, u) \models F$ for all $u \in S$ such that $(t, u) \in ki$.

Because agent i is transitive and $(w, t) \in ki$, $(t, u) \in ki$, it follows $(w, u) \in ki$.

Since $(M, w) \models Ki(F)$, we obtain $(M, u) \models F$, as desired.

Proposition (Euc)

If an agent i is Euclidean in $M = (S, I, ki)$, then $M \models \neg Ki(F) \Rightarrow Ki(\neg Ki(F))$.

Proof

We have to prove $(M, w) \models \neg Ki(F) \Rightarrow Ki(\neg Ki(F))$, where $w \in S$ is an arbitrary world.

Assume $(M, w) \models \neg Ki(F)$. We would like to show $(M, w) \models Ki(\neg Ki(F))$, that is, $(M, t) \models \neg Ki(F)$, for all $t \in S$ such that $(w, t) \in ki$.

We have, by the assumption, $(M, u) \not\models F$ for some $u \in S$ such that $(w, u) \in ki$.

Since ki is serial, $(w, u) \in ki$ and $(w, t) \in ki$ implies $(t,u) \in ki$. Finally, since $(t, u) \in ki$ and $(M, u) \not\models F$, it follows that $(M, t) \models \neg Ki(F)$ holds.

Proposition (Ser)

If an agent i is serial in $M = (S, I, ki)$, then $M \models \neg Ki(False)$.

Proof

We have to prove $(M, w) \models \neg Ki(False)$, for all $w \in S$.

Suppose to the contrary, that is, $(M, t) \models \neg Ki(False)$ does not hold for some $t \in S$.

It follows $(M, t) \models Ki(False)$. Because ki is serial, then for $t \in S$ we have some $u \in S$ such that $(t, u) \in ki$. Hence, it follows $(M, u) \models False$, contradicting $I[u](False) = false$.

Proposition (Symm)

If an agent i is symmetric in $M = (S, I, ki)$, then $M \models F \Rightarrow Ki(\neg Ki(\neg F))$.

Proof

Assume $(M, w) \models F$, for an arbitrary $w \in S$. We have to prove $(M, w) \models Ki(\neg Ki(\neg F))$, that is, (A) $(M,u) \models \neg Ki(\neg F)$, for all $u \in S$ such that $(w, u) \in ki$.

Suppose that, contrary to (A), there exists some $v \in S$ such that $(w, v) \in ki$ and (B) $(M, v) \models \neg Ki(\neg F)$ does not hold. From (B) we obtain $(M, v) \models Ki(\neg F)$.

Because $(w, v) \in ki$ and ki is symmetric, it follows $(v, w) \in ki$. Hence we obtain $(M, w) \models \neg F$, contradicting our assumption $(M, w) \models F$.

Proposition (more knowledgeable agent)

Let $M = (S, I, k1, k2)$ be a Kripke structure for agents 1, and 2.

If $k1 \subseteq k2$, then agent 1 is more knowledgeable than agent 2.

Proof

We would like to prove $M \models K2(F) \Rightarrow K1(F)$, for each formula F in LK.

Let F be an arbitrary formula in LK and w an arbitrary world in S.

We shall prove $(M, w) \models K2(F) \Rightarrow K1(F)$. Assume $(M, w) \models K2(F)$.

It follows, by the definition of $\models$, that $(M, t) \models F$, for all $t \in S$ such that $(w, t) \in k2$.

Now, we prove $(M, w) \models K1(F)$, that is, $(M, u) \models F$, for all $u \in S$ such that $(w, u) \in k1$.

If $(w, u) \in k1$, then, since $k1 \subseteq k2$, it follows $(w, u) \in k2$. Therefore, $(M, u) \models F$.

ACKNOWLEDGMENT

I am grateful to the referees for their careful reading and useful comments.

REFERENCES

[1] R. Fagin et al. *Reasoning about Knowledge*. The MIT Press, London, 1995.

[2] R. Fagin and J. Y. Halpern. Belief, awareness, and limited reasoning. *Artificial Intelligence*, Vol. 34, No. 3, 1998, pp. 39-76.

[3] R. Fagin and J. Y. Halpern. Reasoning about knowledge and probability. *Journal of the ACM*, Vol. 41, No. 2, 1994, pp. 340-376.

[4] R. Fagin, J. Y. Halpern, and M. Y. Vardi. What can machines know ? On the properties of knowledge in distributed systems. *Journal of the ACM*, Vol. 39, No. 2, 1992, pp. 328-376.

[5] R. Fagin, J. Y. Halpern, and M. Y. Vardi. What is an inference rule?. *Journal of Symbolic Logic*, Vol. 57, No. 3, 1994, pp. 1018-1045.

[6] J. Y. Halpern, and M. R. Tuttle (1993). Knowledge, probability, and adversaries. *Journal of the ACM*, Vol. 40, No. 4, 1993, pp. 917-962.

Received: 31 March 1998
Accepted: 20 November 1998

Mirko Maleković

REZONIRANJE O ZNANJU: NEKA SVOJSTVA AGENTA

Sažetak

U ovom članku karakterizirana su svojstva agenta, koja se odnose na rezoniranje o znanju. Dokazane su sljedeće činjenice: ako refleksni agent zna F, onda F vrijedi; ako tranzitivni agent zna F, onda on zna da zna F (svojstvo pozitivne introspekcije); ako Euklidov agent ne zna F, onda on zna da ne zna F (svojstvo negativne introspekcije); serijski agent ne zna F ako je F kontradikcija. Također, uveden je koncept višeznajućeg agenta.

Ključne riječi: svojstva agenta, Euklidov agent, refleksivni agent, serijski agent, simetrični agent, tranzitivni agent, rezoniranje o znanju.