Human-robot communication through a mind model based on the Mental Image Directed Semantic Theory

Masao Yokota
Fukuoka Institute of Technology
yokota@fit.ac.jp

Masato Shiraishi
Fukuoka University of Education
siraisi@fukuoka-edu.ac.jp

Genci Capi
Fukuoka Institute of Technology
capi@fit.ac.jp

Abstract

The Mental Image Directed Semantic Theory (MIDST) has proposed a methodology for integrated multimedia information understanding, for example, cross-media translation. This paper describes a multi-agent model of human mind based on MIDST and its application to human-robot communication.

1. Introduction

On the way to grow up from infant to adult, people would sometimes encounter curious but instantly understandable sentences such as S1-S4. This curiosity perhaps comes from their apparently unscientific contents while such understandability perhaps comes from our everyday perceptive experiences in space and time.

(S1) Time passes swiftly (or slowly).
(S2) It is the longest day.
(S3) The Andes Mountains run south and north.
(S4) The road sinks to (or rises from) the basin.

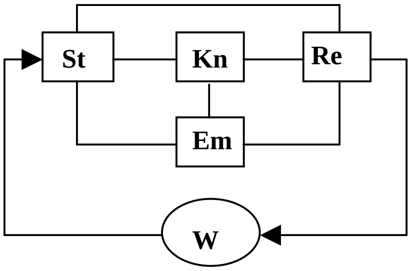
In near future, this kind of human mental phenomenon may lead to a certain barrier preventing humans and robots from comprehensible communication by natural language. This is because both entities can be equipped with sensors, actuators and brains of different performances and their vocabularies may well be grounded on quite different sensations, physical actions or mental actions. And in turn such a situation may bring inevitably different kinds of semantics to them, so called, “Natural Semantics (NS)” for humans and “Artificial Semantics (AS)” for robots.

The authors have been trying to develop such a methodology that can integrate NS and AS into a certain “Compatible Semantics (CS)” and that ultimately can lead to such a “Compatible Mind Model (CMM)” that is intended to organize CS autonomously [1]. The CMM is one kind of the multi-agent models [2]. Its most distinctively remarkable point is that it works by computing mental phenomena representations so called ‘Locus formulas’ based on the Mental Image Directed Semantic Theory (MIDST) [3], whose validity has been proven by the successful results of several versions of the intelligent system IMAGES [3], [6], [7], [9].

In this paper are presented a multi-agent model of human mind aimed at CMM, a brief description of MIDST as framework for CMM, several significant postulates for the basis of CS, formalization of communication, and their implementation on the intelligent system IMAGES-M.

2. Multi-agent model of human mind

The authors have proposed a prototype model of human mind consisting of Stimulus, Knowledge, Emotion and Response processing agents as shown in Figure1 [1]. This is a functional model of human central nervous system consisting of the brain and the spine.



St: Stimulus processing agent.
Kn: Knowledge processing agent.
Em: Emotion processing agent.
Re: Response processing agent.
W: World surrounding human mind,
including his/her body.

Figure 1. Multi-agent model of human mind.

The basic performances of the agents are as follows.

(1) Stimulus processing agent (**St**) receives stimuli from **W** and encodes them into mental images (i.e. encoded sensations) such as *“I sensed something oily.”* (if verbalized in English.)
(2) Knowledge processing agent (**Kn**) evaluates mental images received from the other agents based on its memory (i.e. knowledge), producing other mental images such as *“It is false that the earth is flat.”*
(3) Emotional processing agent (**Em**) evaluates mental images received from the other agents based on its memory (i.e. instincts), producing other mental images such as *“I like the food.”*

(4) Response processing agent (**Re**) converts mental images (i.e. encoded physical actions such as *"I'll walk slowly."*) received from the other agents into real physical actions against **W**.

A *performance* ***P*** against a *stimulus X* with a *result Y* at each agent can be formalized as a function by the expression (1).

$$Y=\boldsymbol{P}(X), \tag{1}$$

where

P : a combination of ***atomic performances*** described later,

X : a spatio-temporal distribution of stimuli from **W** to **St** or a mental image for another agent, and

Y : a series of signals to drive an actuator for **Re** or a mental image for another agent.

A performance ***P*** is assumed as a function formed either consciously or unconsciously. In a conscious case, a set of atomic performances are to be chosen and combined according to *X* by a meta-function, so called, '***Performance Selector*** (PS)' assumed as '***Conscience***'. On the contrary, in an unconscious case, such a performance as associated most strongly with *X* is to be applied automatically [8]

3. MIDST as framework for CMM

MIDST has modeled mental images as "Loci in Attribute spaces" [3], [7]. An attribute space corresponds with a certain measuring instrument just like a barometer, a map measurer or so and the loci represent the movements of its indicator. The performance of 'Attribute space' is the model of '***Atomic performance***' introduced in Section 2.

A general locus is to be articulated by "Atomic locus" formalized as the expression (2) in first-order logic, where "L" is a predicate constant.

$$L(x,y,p,q,a,g,k) \tag{2}$$

The expression (2) is called "Atomic locus formula" whose arguments are referred to as 'Event Causer', 'Attribute Carrier', 'Initial Attribute Value', 'Final Attribute Value', 'Attribute Kind', 'Event Kind' and 'Standard Attribute Value', respectively.

The interpretation of (2) is as follows, where "matter" means "object " or "event".

"Matter 'x' causes Attribute 'a' of Matter 'y' to keep (p=q) or change ($p \neq q$) its values temporally (g=Gt) or spatially (g =Gs), where the values 'p' and 'q' are relative to the standard 'k'."

When g=Gt and g=Gs, the locus indicates monotonous change or constancy of the attribute in time domain and in space domain, respectively. The former is called 'temporal event' and the latter, 'spatial event'.

For example, the motion of the 'bus' represented by S5 is a temporal event and the ranging or extension of the 'road' by S6 is a spatial event whose meanings or concepts are formalized as expressions (3) and (4), respectively, where the attribute is "physical location" denoted as *A12*. We think that the verb 'run' used in S6 must reflect the motion of the observer's attention [4].

(S5) The bus runs from Tokyo to Osaka.

$$(\exists x,y,k)L(x,y,\mathrm{Tokyo},\mathrm{Osaka},A12,Gt,k)\wedge \mathrm{bus}(y) \tag{3}$$

(S6) The road runs from Tokyo to Osaka.

$$(\exists x,y,k)L(x,y,\mathrm{Tokyo},\mathrm{Osaka},A12,Gs,k)\wedge \mathrm{road}(y) \tag{4}$$

The expression (5) is the conceptual description of the English word "fetch", implying such a temporal event that 'x1' goes for 'x2' and then comes back with it, where 'Π' and '•' are instances of the tempo-logical connectives, 'SAND' and 'CAND', standing for "Simultaneous AND" and "Consecutive AND", respectively.

In general, a series of atomic locus formulas with such connectives is called simply 'Locus formula'.

$$(\exists x1,x2,p1,p2,k)\ L(x1,x1,p1,p2,A12,Gt,k) \bullet (L(x1,x1,p2,p1,A12,Gt,k)\Pi L(x1,x2,p2,p1,A12,Gt,k)) \wedge x1\neq x2 \wedge p1\neq p2 \tag{5}$$

In order for complete representation of temporal relations, we have introduced a concept called 'Empty Event (EE)' and symbolized as 'ε' which stands exclusively for time collapsing. For example, (6) represents 'X_1 during X_2 '.

$$(\varepsilon_1 \bullet X_1 \bullet \varepsilon_2)\ \Pi\ X_2 \tag{6}$$

The image model presented here is also valid for formalizing word concepts (i.e. coding) of actions because any action must be measured with sensors for its formalization. That is, ***grounding words on actions is equivalent to grounding words on sensations of actions***

Sensors and actuators are assumed to collaborate very closely in feedback or feed-forward ways in cybernetics and there is a hypothesis that some kinds of sensations (or perceptions) and actions are encoded in the same way in organisms [5]. If not, real-time coordination of multiple sensors and actuators would be impossible. 'Mimicking' may be a good support for this hypothesis.

As easily imagined, if an attribute space corresponds to one of human senses, then its loci associated with certain words belong to NS, otherwise to AS.

4. Formalization of communication

At first, we formalize a piece of information (***I***) as a set of messages (*m*'s) in the expression (7).

$$\boldsymbol{I}=\{m_1, m_2, \ldots, m_n\} \tag{7}$$

In turn, a message (*m*) is defined in the expression (8), where *D, S, R* and *B* mean the duration, sender(s), receiver(s) and the body of the message, respectively.

$$m=(D, S, R, B) \tag{8}$$

The body (B) consists of the two elements shown in the expression (9), where *E* and *T* mean the event referred to and the task requested or intended by the sender, respectively.

$$B=(E, T) \qquad (9)$$

For example, each item of the message m_0: “Fetch me the book from the shelf, Tom” uttered by Jim during the time-interval [t_1, t_2] is as follows:

$m_o=(D_0, S_0, R_0, B_0)$, $B_0=(E_0, T_0)$,

$D_0=[t_1, t_2]$, $S_0=$ *“Jim”*, $R_0=$ *“Tom”*,

$E_0=$ *“Tom FETCH Jim BOOK FROM SHELF”*,

and $T_0=$ *“realization of* E_0*”*.

The authors have found that there are almost unique correspondences between the kinds of tasks (T’s) and the types of sentences as shown in Table 1, which are very useful for computation.

Table 1. Sentence types and Tasks.

Sentence type (Examples)	Task (T)
Declarative (It is ten o’clock now.)	To believe E.
Interrogative ([A] Is it ten o’clock now? [B] What time is it now?)	[A] To reply whether E is true or false. [B] To reply what makes E true.
Imperative (Show me your watch.)	To realize E.

5. Human-robot communication

The authors have planned to implement IMAGES-M [6] on real robots as a mind model. One of the most significant works of robots equipped with IMAGES-M is to help people by performing dialogs with them. For example, assume such a scenario as follows:

*…**A human ‘Masato’ and a humanoid robot ‘Robbie’ encounter at the terrace in front of the room where a Christmas party is going on merrymaking. Masato says “Robbie, please fetch me a colorful sweet soft scentless candy from the noisy room.” Robbie replies “OK, Masato.”….***

Robbie interprets Masato’s statement as the expression (10) that reads “If Robbie fetches Masato the candy (***E1***), then consecutively it makes Masato happier (***E2***),” or as its logical equivalent, the expression (11), reading “It is not the case that Robbie fetches Masato the candy and consecutively it does not make Masato happier.” Both the expressions are adopted in MIDST as the canonical conceptual structures of an imperative sentence.

$$\boldsymbol{E1} \rightarrow c\ \boldsymbol{E2} \qquad (10)$$

$$\sim(\boldsymbol{E1}\bullet\sim\boldsymbol{E2}) \qquad (11)$$

where

$$\boldsymbol{E1} \Leftrightarrow (\exists x1,x2,k1,\ldots,v,\mathbf{C})\ (L(R,R,M,x2,A12,Gt,k1)\bullet$$

$$(L(R,R,x2,M,A12,Gt,k1)\Pi L(R,x1,x2,M,A12,Gt,k1)))$$

$$\Pi(\underline{L(v,x1,c1,c2,A32,Gs,k2)\bullet}$$

$$\underline{L(v,x1,c2,c3,A32,Gs,k2)}$$

$$\underline{\bullet\cdot\ldots\bullet\cdot L(v,x1,c_{m-1},c_m,A32,Gs,k2)}\ )$$

$$\Pi\ L(v,x1,Sweet,Sweet,A29,Gt,k3)$$

$$\Pi\ L(v,x1,Soft,Soft,A24,Gt,k4)$$

$$\Pi\ L(v,x1,/,/,A30,Gt,k5)$$

$$\Pi L(v,x2,Noisy,Noisy,A31,Gt,k6)$$

$$\wedge\ candy(x1) \wedge room(x2) \wedge \mathbf{C}=\{c1,c2,\ldots,ci\} \wedge \#(\mathbf{C}) >1$$

$$\boldsymbol{E2} \Leftrightarrow (\exists e1,e2,k7)\ L(\boldsymbol{E1},M,e1,e2,B04,Gt,k7) \wedge e2>e1.$$

The special symbols and their meanings in the expressions above are:

‘X→c Y’ =‘If *X* then consecutively *Y*’, *‘R’*=‘Robbie’, *‘M’*=‘Masato’, ‘**C**’=‘the total set of colors’, *‘#(***X***)’*=‘cardinal number of set **X**’, *‘A29’*=‘taste’, *‘A24’*=‘touch’, *‘A30’*=‘smell’, *‘/’*=‘absence of value’, *‘A31’*=‘sound’, *‘A32’*= ‘color’, and *‘B04’*= ‘happiness (=degree of happiness)’

According to Table1, Robbie’s task (T) is only to make ***E1*** come true where each atomic locus formula is associated with his actuators/sensors. By the way, the underlined part of ***E1*** represents the spatial distribution of colors over the candy referred to by the word ‘colorful’. Of course, Robbie believes that he will become happier to help Masato, given by expression (12) where ‘B03’ is ‘trueness (=degree of truth)’and ‘K_B‘ is a certain standard of ‘believability’. That is to say *emotionally*, Robbie likes Masato. Therefore, this example is also very significant for intentional sensing and action of a robot driven by logical description of its belief.

$$(\exists p)L(R,\boldsymbol{E},p,p,B03,Gt,K_B) \wedge p>K_B$$

$$\wedge\ \boldsymbol{E = E1} \rightarrow c\ \boldsymbol{E2} \qquad (12)$$

For constructing a plausible CMM it is most essential to find out functional features of human mind and to formalize them as postulates that rule the performances of CMM and form the basis of CS [1]. APPENDIX shows some of such postulates and examples of dialog processing by IMAGES-M based on them.

6. Discussions and conclusions

The mind model proposed here is much simpler than Minsky’s [2] but the locus formula representation can work for representing and computing mental phenomena fairly well as shown in APPENDIX. One of the most important problems to be solved is how to realize the atomic performances corresponding to attribute spaces, including the meta-function ‘*conscience*’. In order to solve this problem, we will consider the application of soft computing theories such as neural network, genetic algorithm, fuzzy logic, etc. in the future.

References

[1] Yokota, M. & Shiraishi: “A multi-agent mind model for comprehensible communication between humans and robots”, Reports of CSL, FIT, 18, pp.1-9, 2004.

[2] Minsky, M: The society of mind, Simon and Schuster, New York, 1986.

[3] Yokota, M. *et al.*: "Mental-image directed semantic theory and its application to natural language understanding systems'', Proc. of NLPRS'91, pp.280-287, 1991.

[4] Rybak, I.A., *et al*: "A model of attention-guided visual perception and recognition", Vision Research, 38, pp.2387-2400, 1998.

[5] Prinz, W.: "A common coding approach to perception and action", In *Relationships between perception and action* (Neumann, O. and Prinz, W. eds.), pp.167-201, Springer-Verlag, 1990.

[6] Hironaka, D., Oda, S., Ryu, K., and Yokota, M.: "Mutual Conversion of Sensory Data and Texts by an Intelligent System IMAGES-M'', Proc. of the 8th International Symposium on Artificial Life and Robotics (AROB '03), pp.141-144, 2003.

[7] Yokota, M. and Hironaka, D.: "Cross-media translation based on Mental Image Directed Semantic Theory toward more comprehensible communication between humans and robots", Proc. of AINA '04 IEEE, Fukuoka, 2004.

[8] Brooks, R. A.: "A robust layered control system for a mobile robot", IEEE Journal of Robotics and Automation, RA-2, pp.14-23, 1986.

[9] Oda, S., Oda, M.and Yokota, M.: "Conceptual Analysis and Description of Words for Color and Lightness for Grounding them on Sensory Data'', Trans. of JSAI, 16-5-E, pp436-444, 2001.

APPENDIX

Examples of postulate-based dialog processing by IMAGES-M, where 'H' and 'S' mean a human's and IMAGES-M's utterance, respectively.

[Dialog 01]

Postulate 01: "Sensation precedes perception in humans."

H: Tom heard Mary sing "Hey, Jude".
H: Did he sense any sounds?
S: Yes, he did.

[Dialog 02]

Postulate 02: "There are sensor-specific pieces of knowledge in humans."

H: Tom knows Mary by sight.
H: Is he familiar with her?
S: No, he isn't.

[Dialog 03]

Postulate 03: "Perceived matters are memorized as facts belonging to knowledge in humans."

H: Tom saw Mary move to Tokyo.
H: What did Tom know about Mary?
S: He knew that she went to Tokyo.

[Dialog 04]

Postulate 04: "Intentional performances are necessarily accompanied by decisions in humans."

H: Tom sold his book to Mary.
H: Did Tom decide to sell his book?
S: Yes, he did.

[Dialog 05]

Postulate 05: " Desire precedes decision in humans."

H: Tom decided to sell his book.
H: What did he want?
S: He wanted someone to buy his book.

[Dialog 06]

Postulate 06: "For humans, a desire for something is a belief of becoming happier with it."

H: Tom wants to go to Tokyo.
H: Does Tom believes to become happier if he goes to Tokyo?
S: Yes, he does.

[Dialog 07]

Postulate 07: "Some emotion can coexist with another in humans."

H: Tom loves Jane.
H: Whom does he like?
S: He likes Jane.

[Dialog 08]

Postulate 08: "A physical object has never two values of an attribute."

H: Tom melted the ice (*into water*) and drank it.
H: Did he drink the ice?
S: No. He drank the water.

[Dialog 09]

Postulate 09: "Communication is not transfer but duplication of information."

H: Tom said to Mary that his mother was sick.
H: Who knew that Tom's mother was sick?
S: Tom and Mary did.

[Dialog 10]

Postulate 10: "A spatial event is reversible."

H: The path sinks to the brook.
H: Does the path rise from the brook?
S: Yes, it does.
H: The roads meet at the city.
H: Do the roads separate at the city?
S: Yes, they do.