Computational Model of Mind for a Robot and Its Application to Spatial Language Understanding in Virtual Environment*

K. S. Jayakumar
Department of Mechanical Engineering
Sona College of Technology, Salem
Anna University, India-636005
ksjayakumar@sonatech.ac.in

Ming Xie
School of Mechanical and Aerospace Engineering
Nanyang Technological University
Singapore-639798
mmxie@ntu.edu.sg

Abstract Current researches in theory of mind focus on developing the model of mind by studying the human mental processes such as model of how human learns the language, how one understands the language and objects, how one represent the knowledge in the brain, etc. As a result, the model can be applied to a robot and it can think like humans. *The main basic thing that underlies in the theory of mind is how an agent in the physical world transforms the perceptions into mental actions (i.e. decision making) into physical actions.* Still, there are no theories of mind that answer the above issue. This paper proposes a new method to transform the perceptions into mental actions. Based on this, computational model of mind is proposed. The spatial language sentences are learned and understood using the proposed model of mind in virtual environment. The learning and understanding experiments demonstrate the effectiveness of the model.

1 Introduction

The main basic thing that underlies in the theory of mind is how an agent in the physical world transforms the perceptions into mental actions (i.e. decision making) into physical actions. Researchers in artificial intelligence have been trying to find solution to the above problem by studying the human mental processes ([1], [2], [3]). Let us consider the following experiment to understand the above problem. For the experiment, the situation is that there are three cubic blocks on the table and each of them is made of iron, wood and soft glass materials. They are of same size. When the iron or wood block is placed on top of the glass block, the glass block breaks up. The iron block can be placed on the wood block and vice versa.

The main task in the experiment is to stack the cube blocks in such a way that the blocks should not break. Let's assume that the task is performed by fifteen year old boy/girl. What he/she will do is first to look at the cubic blocks on the table, then to determine the weight and other properties of each block by holding in hand, then to decide which block be placed in bottom, which one in middle, and which one in top, and finally to carry out the task(i.e. stacking operation) by hand. In this task, looking at the objects correspond to perception(i.e. perceiving the objects through one's sensory systems and here it involves vision and touch). Then the perception goes to one's brain and decision making takes place there. The decision making is also referred to 'mental action'. Finally, the mental action is transformed to physical action(s) through one's limbs to carry out the decided or desired task. *Here, the big difficult task is how to transform the perception into mental action(s).* It is still an unknown process on how human brain maps the perception into mental action (i.e. decision making).

1.1 Entities Encoding and Decoding Processes - Basis for Intelligence

This research work proposes a new method to transform the perceptions into mental actions. The entities encoding and decoding processes play the important role in the transformation process. Fig. 1 shows the encoding and decoding processes of the entities that exist in the physical world. This is the fundamental philosophy behind the development of computational model of mind for a robot. The entities are transformed into properties through the sensors or programmed interface and this is called as encoding of entity. In other words, the entity can be modeled by properties, which can be classified as geometrical

*This work was done in School of Mechanical and Aerospace Engineering, Nanyang Technological University, Singapore

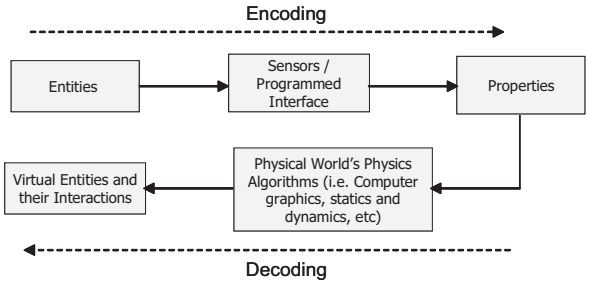

Figure 1: Entities encoding and decoding processes. Entities are encoded or mapped to properties through sensors or programmed interface. The properties can be decoded or mapped to virtual entities and their possible interactions with other entities through the physics model such as computer graphics, rigid body statics and dynamics, etc.

properties, mechanical properties, electrical properties, chemical properties, etc. The purpose of the sensors is to acquire the properties from the entity. If the sensors are not available, the properties are acquired through the programmed interface, which is nothing but graphical user interface through which user inputs the properties. Then the properties can be mapped to virtual entities and their possible interactions with other entities through physics model such as computer graphics, rigid body statics and dynamics, etc. This process is called as decoding of entity. The physics algorithms are not only used to virtually reconstruct the entity but also it determine what kind of interaction the entities has. *The perception process is equivalent to encoding process (i.e. mapping of entity into properties). The transformation process of perception into mental actions is equivalent to decoding process of the entity.*

The paper is organized as follows. Section 2 discusses the computational model of mind. Section 3 discusses the learning and understanding frame work for spatial language. Section 4 discusses the learning and understanding experiments. Section 5 discusses the related works in theory of mind and its implementation in robots. Finally, the paper ends with conclusion and future works.

2 Computational Model of Robot Mind

Figure. 2 shows the robot mental architecture. The robot mind consists of mental processes and organized memory and the both interact with each other in order to derive the mental actions. The mind is embodied in the body of the robot or vice versa. Real world consists of physical world and conceptual world. Physical world and conceptual world refers to our environment and texts of natural language respectively([4], [5], [6]).

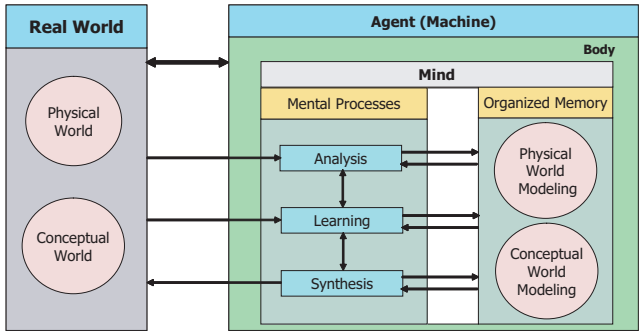

Figure 2: A robot's mental architecture. The interaction between mental processes and organized memory creates the mental actions

2.1 Organized Memory

Physical world or environment consists of man-made and natural-made entities, which exist in space and time. The entities are the basic building blocks of the physical world. The entities are never in isolation. Instead, they act and interact with one another through forces such as gravity, magnetic force, mechanical force, etc. The interaction between the entities creates the cause-effect phenomena in the physical world and forms the basis for creating the action, behavior, event and episode in the physical world.

The interaction between the entities is controlled by physical world physics or constraint physics. In physics, modeling the interaction between the entities helps us to study their cause-effect phenomenon, for example, rigid body statics and dynamics. The entity in the physical world can be modeled by properties such as mechanical, geometrical, chemical, etc. Because of the interaction among the entities in the physical world, some entities may produce certain behaviors, some may receive the actions and some may transform one form of energy into other. The capability of the entity to produce the above characteristics is called 'constraint'. So, each entity in the physical world has two important components such as properties and constraints. The properties and constraints can be seen as representation of the entity in terms of certain measurable attributes. Constraint physics models can reconstruct the entity from its properties as well as it can determine and control the interaction among the entities.

Similar to the physical world, word is the basic element in the conceptual world and words interact with one another through syntax or grammar rules in order to form a phrase or sentence. In the conceptual world, properties refer to noun, verb, adjective, etc and constraints are the grammar rules. The properties and constraints are represented or described in word or set of words in the conceptual world.

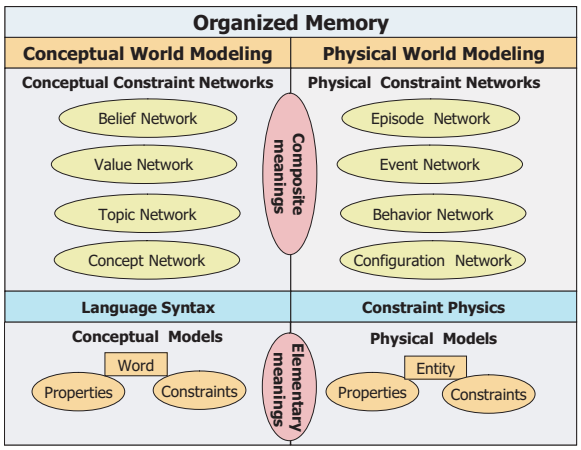

Figure 3: Architecture of an organized memory, in which modeling of the physical and conceptual world are represented. The properties are acquired through sensors and the constraints are learned by modeling the interaction among the entities.

Fig. 3 shows the architecture of the organized memory. The organized memory consists of elementary and composite meanings both for conceptual and physical world. Elementary meanings contain physical models and conceptual models. Physical models refer to entity's properties and constraints. Conceptual models refer to word's grammatical properties and constraints. Composite meanings in physical world (i.e. physical constraint networks) are constructed by combining the physical models through the mediation of constraint physics algorithm. Similarly, conceptual constraint networks, which is nothing but representation of configuration, action, event and episode in the form of symbols, are constructed by combining the conceptual models with help of grammar rules and physical constraint networks.

Physical constraint networks consist of configuration network, behavior network, event network and episode network. Configuration network refers to spatial arrangement of entities in space. Behavior network refers to chain of cause-effect phenomena of the configuration network or system of rigid bodies. It helps the agent to decide what kind of actions the entity produce. Interaction of set of events and episodes creates the event network and episode network respectively.

Conceptual constraint network consists of concept network, topic network, belief network and value network. Concept network is a sentence or set of sentences, which is the description of configuration network or behavior network. Topic network is a paragraph or set of paragraphs, which are the description of event or episode network. Value network is a concept or topic network, which tells the machine good and bad things about the activities in the physical world. Belief network makes the machine to accept or believe the concept or topic of other agents without any contradiction. The values and beliefs decide the morality of the robot.

3 Learning and Understanding Frame Work for Spatial Language

The configuration network refers to spatial arrangement of rigid bodies in certain pose with respect others. It helps the robot to learn and understand the spatial language sentences [7].

3.1 Learning Framework

The learning process consists of modeling, training or optimization and representation. The modeling refers to mathematical formulation of the problem and the formulation contains the set of modeling parameters. Training refers to assigning certain values to the modeling parameters with help of some examples. Representation refers to the finalized parameter-value pair, which can be used for recognition, classification and understanding tasks.

Fig. 4 shows the learning framework in which the words that correspond to the entity name and spatial positions are learned through modeling the entities and their interactions. First entity and its interaction are modeled by properties and constraints respectively. In the training stage, certain values are assigned to properties and constraints parameters with help of examples. During the representation stage, the finalized parameter-value pair is determined from the training stage and more details are given in the

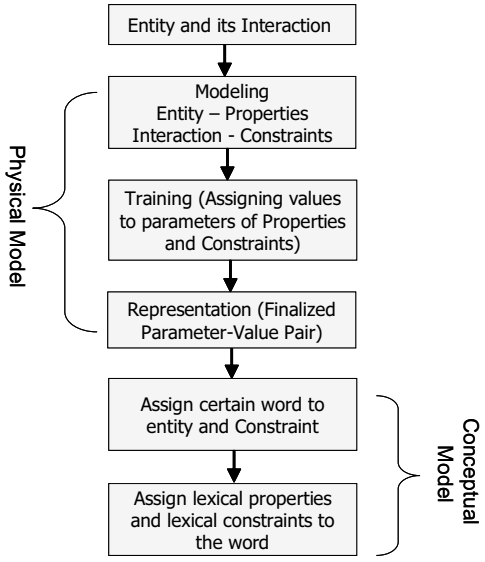

Figure 4: Learning framework in which the words that correspond to the entity name and spatial positions are learned through the modeling the entities and their interactions

section 'learning experiments'. Then the properties are assigned with certain entity name and the constraints are associated with certain spatial position word. Then for each word, the corresponding lexical properties and constraints are assigned. The lexical properties refer to part of speech of the word and the lexical constraints are nothing but grammar rules of the phrase and sentence formation. The finalized parameter-value pairs of properties and constraints of the entity are called physical model. The word's lexical properties and constraints are called conceptual model.

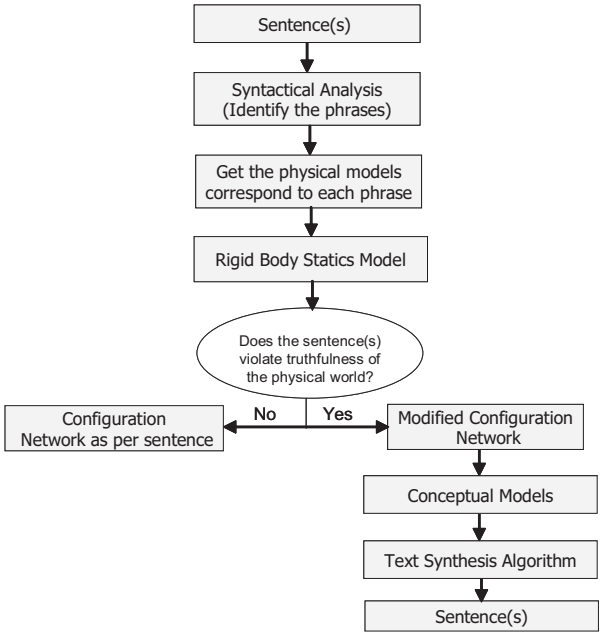

Figure 5: Understanding framework in which the spatial language sentences are understood with help of rigid body statics model

3.2 Understanding Framework

Fig. 5 shows the understanding framework in which the spatial language sentences are understood with help of rigid body statics model. First, using syntactical analysis, phrases such as noun phrase, verb phrase, prepositional phrase are identified in the sentence.

For each phrase, the corresponding physical models are retrieved and then are passed to the rigid body statics model in order to construct the configuration network. The configuration network refers to spatial arrangement of set of entities and it is the result of static interaction among the entities. While constructing the configuration network, two possible scenarios arise: a) the configuration network as per the sentence information, which does not violate the truthfulness of the physical world and b) the modified configuration network, which arises when the sentence information violates the truthfulness of the physical world. The violating information in the sentence are mainly due to incorrect subject and object, and wrong spatial information between the subject and object. The modified sentence will be generated by the text synthesis algorithm given the input the modified configuration network and conceptual models.

The rigid body statics model is a statics algorithm, which contains sub-algorithms such as 3D shape analysis, test for equilibrium, test for intersections, test for stability and different kinds of distance calculations. The rigid body statics model takes the input of entities properties such as geometrical and mechanical and static interaction parameter values(optional), then calculates the position vector for the target entity with respect to the reference entity.

4 Learning and Understanding Experiments

Sixty different entities are taken for the learning and understanding experiments and the interactions among the entities are used to learn the spatial position words such as *on, in, over, under, above, below, inside, outside, behind, in front of, back, front, left, right, beside, between, among, near and far.* The rigid body statics model is used to interpret or understand the meanings of spatial language sentences. The meaning of the spatial language sentences are represented in a 3D scene.

4.1 Learning of Spatial Position Words Through Modeling the Static Interaction Among the Entities

4.1.1 Computational Model-I for 'On'

The finalized parameters-values pairs of the particular configuration of the entities is called computational model of a spatial position word. The following discusses the learning process of 'on' such as modeling, training and representation.

Modeling Using the spatial constraints parameters, the interaction or the spatial constraint of 'one entity is in the top of the other entity' is modeled by

$\{S_{top}; E_r \cap E_t; VD_{O-ce_r} \geq 1; ST; E_t \cup E_r\}$

Where S_{top} - Whether the patch of the reference entity is a top surface and it takes the values such as 'true' or 'false'. $E_r \cap E_t$ - Intersection between the reference entity and the target entity and it takes the values such as 'true' or 'false'. VD_{O-ce_r} - Vertical distance between the observer and the centroid of the patch of the reference entity along the positive y axis and its value should be greater than or equal to one. ST - Whether the stability exists between the target entity and the reference entity and it takes the values such as 'true' or 'false'. $E_t \cup E_r$ - Whether the target entity is rested on the reference entity based on

the gravity and it takes the values such as 'true' or 'false'. Sometimes during the modeling stage, minimum value such as $\geq$ is assigned for the vertical distance VD_{O-ce_r}. The remaining parameters values are learned through training examples i.e. interaction between the entities.

Training Fig. 6 shows the graphical user interface for learning the spatial position words. First, the user is prompted to select the entities. The spatial configuration between the entities are constructed in two modes: manual mode and autonomous mode. In manual mode, using the translation and rotation values the desired configuration is constructed. In autonomous mode, the different configurations will be generated and the user can select the desired configuration.

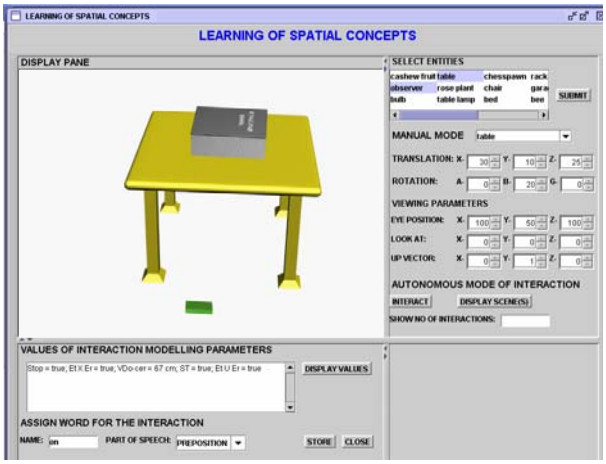

Figure 6: The graphical user interface for learning the spatial concept

In this example, the reference entity is taken as 'table' and the target entity as 'book'. The observer is assumed to be small 'box'. The ground is assumed to be infinite surface area and infinite strength, and its color is white. In the list box, the entities 'table','book' and 'observer' are selected. The desired configuration is constructed by the manual mode. Then pressing the 'Display Values' button shows the values for the parameters. The displayed values are as follows:

$\{S_{top} = true; E_r \cap E_t = true; VD_{O-ce_r} = 67cm; ST = true; E_t \cup E_r = true\}$

The vertical distance between the observer and the reference entity is 67cm. The units are arbitrary and the user can choose any units such as millimeter(mm), feet(ft), meter(m), etc.

Similarly, using the different interaction examples (i.e ten examples), the values are assigned to the parameters. In that, only the parameter values of VD_{O-ce_r} varies between 20 to 100. The range satisfies $\geq$ condition in the modeling and also all other parameters in each example have the same values. The training continues till at least parameters of ten interaction examples attain the similar values.

Representation: Computational Model After the training, the parameters-values pairs are finalized. Then the representation is assigned with spatial position word 'On'. The conceptual word 'On' is called 'spatial concept'. Then, the part of speech tag such as 'Preposition' is assigned to that word . Finally, pressing 'Store' button stores the spatial concept.

Based on the above, the computational model for the spatial position word 'On' is given by

$\{S_{top} = true; E_r \cap E_t = true; VD_{O-ce_r} \geq 1; ST = true; E_t \cup E_r = true\} \implies On$

This computational model is used to identify the spatial relationship between the entities and to understand the spatial language sentences. Similar to this, the computational model for other spatial position words are developed.

4.2 Understanding of Spatial Language Sentences

Understanding deals with interpreting and verification of communicated information, which is in the form of text. The grammatical analysis is used to partially interpret the meaning of the sentences. It gives the information about agent, action, patient, spatial and timing relationships. But it will not tell whether given spatial relationship is true to the agent and patient and will not tell the cause-effect relationship between the entities and properties of the entities. The main purpose of the grammatical analysis is to get *'the reference entity-constraint-the target entity'* sequences from the sentences. They are identified from the sentences using the physical and conceptual models information of the words.

Understanding Experiment-I The compound sentences such as *'The table is in the ground. The desktop is left of the table. The monitor is right of the table. The keyboard is in front of the monitor'* are considered for the understanding. Fig. 7 shows the understanding of the above sentences.

Understanding Experiment-II The compound sentences such as *'The bed is in the ground. The chair is beside the bed. The table lamp is on the chair. The table fan is behind the bed. The chair is in the ground'.* are considered for the understanding. Fig. 8 shows the configuration network and the scene of the network.

5 Related Works

Leslie [1] proposed the theory of mind, which discusses the mental processes in human child. The mind consists of three modules such as Theory of Body module (ToBY), which deals with understanding of entities, Theory of Mind module-I (ToMM-I), which describes the behavior of the agent, and Theory of

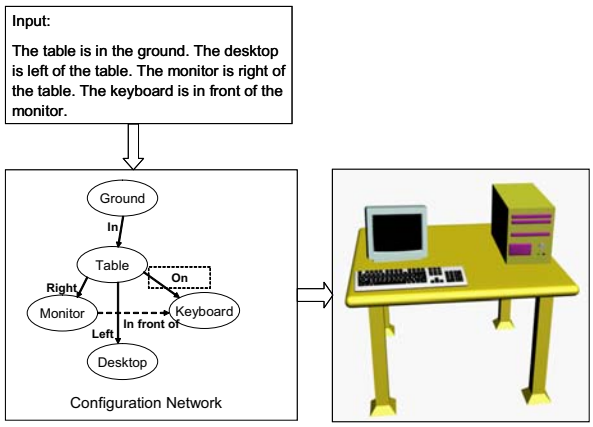

Figure 7: Understanding experiment-I

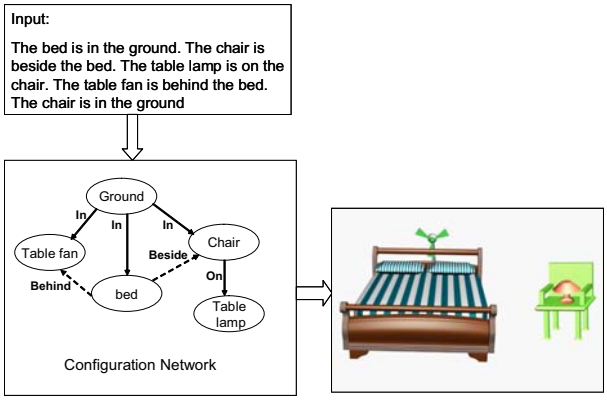

Figure 8: Understanding experiment-II

Mind module-II (ToMM-II) which discusses the beliefs and values of the agent. The proposed model almost similar to the Leslie's work, but it lacks the modeling of the entities and their interactions. Baron-Cohen's [2] model of mind consists of intentionality detector, eye direction detector, shared attention mechanism, and theory of mind mechanism. Scassellati [3] implemented the above mind theories in Cog humanoid robot. Similar to this, Deb Roy [8] implemented grounded language learning model called CELL (Cross-channel Early Lexical Learning). CELL learns the meaning of words from untranscribed acoustic and video input signal. As a result of learning, the word is represented in terms of sensory feature values of the entity corresponds to the word.

6 Conclusion and Future Works

The main basic thing that underlies in the theory of mind is how an agent in the physical world transforms the perceptions into mental actions (i.e. decision making) into physical actions. The computational model of mind for the robot is proposed based on entities encoding and decoding processes. It is shown this work that robot mind does not necessarily follows the principles of human mental processes. The spatial language learning and understanding experiments shows the effectiveness of the computational model. The spatial language sentences meaning are interpreted using rigid body statics model, which acts as a semantics model.

The future works will be to perform the understanding experiments with other natural languages such as Tamil, Hindi, French, German, etc, language generation or commonsense knowledge acquisition from set of entities and their interactions, learning and understanding of words related to dynamic interactions of the entities and real time implementation of this concept in robot.

References

[1] A. M. Leslie, "Spatiotemporal continuity and the perception of causality in infants", Perception, 13:287305, 1984

[2] S. Baron-Cohen, *Mindblindness*, MIT Press: Cambridge, MA, 1995.

[3] B. Scassellati, "Theory of Mind for a Humanoid Robot", Autonomous Robots 12, 1324, 2002.

[4] M. Xie, J. S. Kandhasamy, and H. F. Chia, "Meaning-centric framework for natural text/scene understanding by robots," International Journal of Humanoid Robotics, vol. 1, June 2004.

[5] M. Xie, J. S. Kandhasamy and K. H. Leong, Response to dialogue "Can robots learn languages the way children do?", IEEE Autonomous Mental Development Newsletter, Vol. 2(2), ISSN: 1550-1914, October 2005

[6] K. S. Jayakumar, "Organized memory for natural text understanding and its meaning visualization by machine", PhD Thesis, Nayang Technological University, Singapore, 2006 (Under examination).

[7] K. S. Jayakumar and M. Xie, "Spatial language learning and understanding in virtual environment", Artificial Intelligence, 2006 (Under Review)

[8] D. Roy and A. Pentland, "Learning words from sights and sounds: A computational model," Cognitive Science, vol. 26, no. 1, pp. 113–146, 2002.