# A General Model of Artificial General Intelligence

Zengkun Li

**Abstract:** This paper presents a general model of AGI, the model indicates how knowledge is represented and learned, how the knowledge is used to accomplish tasks such as inference, memory recalling and so on, and how the advanced intelligence phenomenons such as self conscious, language and emotions emerge.

**Keywords:** Artificial general intelligence, General model, Knowledge representation, Inference, Self conscious, Language, Emotion

## 1 Knowledge representation

Knowledge is composed of concepts and relationships between them. for concepts, they are represented by a multiple layered full connected forward neuron network, the network takes the input of the concept, sends it into the hidden layer, and generates an output, which is exactly equivalent to the input, and the state of the hidden layer is exactly the internal representation of the concept. The network is actually representing a **tautology** that 'A is A'.

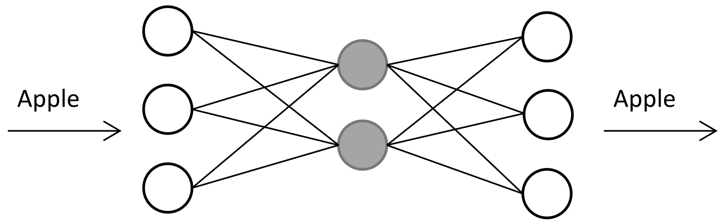


**Fig. 1 Tautology represented by neuron network**

Moreover, for concepts, there are specific concepts such as apples, cars, and abstract concepts such as fruits, transportation and so on. In fact no matter how specific or abstract one concept is, it is represented by a unified neuron network structure mentioned before but with different 'shape'. One aspect of the 'shape' is the number of neurons of the hidden layer, for specific concepts, the number is smaller, so the capacity of the network is also smaller, which means less inputs can be sent into the network and get the expected outputs; for the abstract concepts, the number is larger, so the capacity is larger, which means more inputs can be sent into the network and get the expected outputs. For instance, the network representing the specific concept 'apple', may only take the input of 'apple' and output 'apple', if 'pear' is sent into the network, it may give a random output with no meaning; for the network representing the abstract concept 'fruit', due to the larger capacity, it may takes both 'apple' and 'pear' and generate expected results, 'apple' for 'apple', and 'pear' for 'pear.

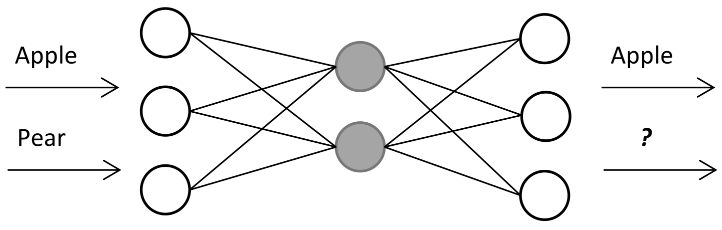


**Fig. 2 The specific network accepts 'Apple' while refuses 'Pear'**

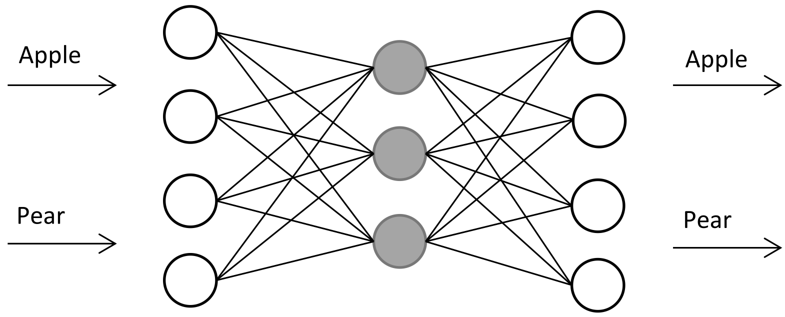


**Fig. 3 The abstract network accepts 'Apple' and 'Pear' simultaneously**

The other aspect of the 'shape' of the network is the numbers the input and output layers. Obviously, the larger the number is, the more complicated concepts can be taken by the network, and vice versa. And because the complexities of the input and output are generally the same, so the numbers of neurons of input and output layers are also equal. So far, we can make a quadrant map to summarize the relationship between concepts and networks representing them.

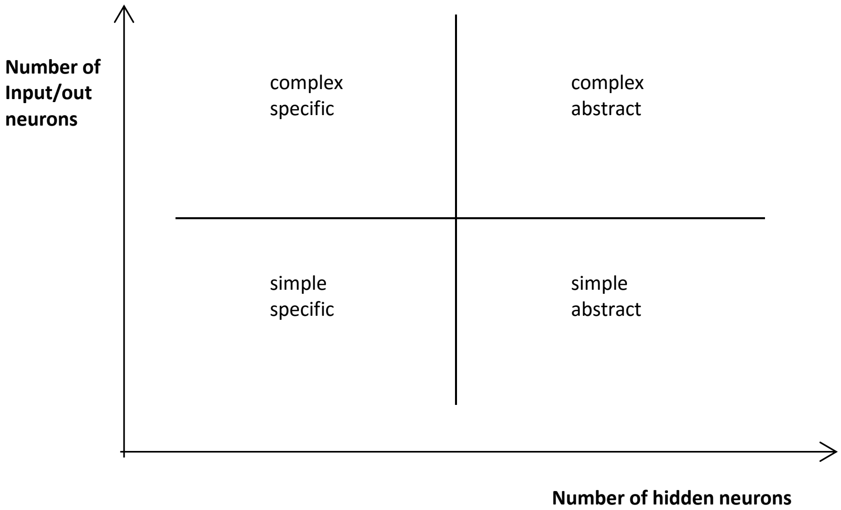


**Fig. 4 The quadrant map of the networks**

## 2 Knowledge processing

The entire network of an intelligence agent consists of multiple network structures mentioned in section 1, which are mentioned as network unit in the following discussion. Fig.4 indicates how the structure of an entire network is, in the horizontal direction, the network units under different abstract level are stacked, from left to right, taking more and more abstract inputs; and in the vertical direction, multiple network units under the same abstract level are aligned together to receive different inputs. More importantly, all of the layers have the same number of neurons, and the output layer at the right end is actually the same layer with the input layer at the left end, therefor the entire network is actually a huge recursive network.

So a concept is represented by the network in different abstract levels simultaneously, and when a complex concept which consists multiple simple concepts is represented as a concept tree, because the distribution of activated neurons of the hidden layers forms exactly a shape of a tree, and each path from left to right is a branch of the tree. In this way, the higher abstract layer are shaped by the output of the lower abstract layer, which is the learning process, and the exists higher abstract layers also constrain how the lower abstract layer act by feeding back to their inputs, which is the inference process.

Since the entire network is a recursive network, that means there are many 'loops' inside the network, but at the macro level, the network acts like a 'dynamic field' rather than a 'circle', because there is no start point or end point of the network in the timeline, the higher abstract and lower abstract layers can be activated simultaneously.

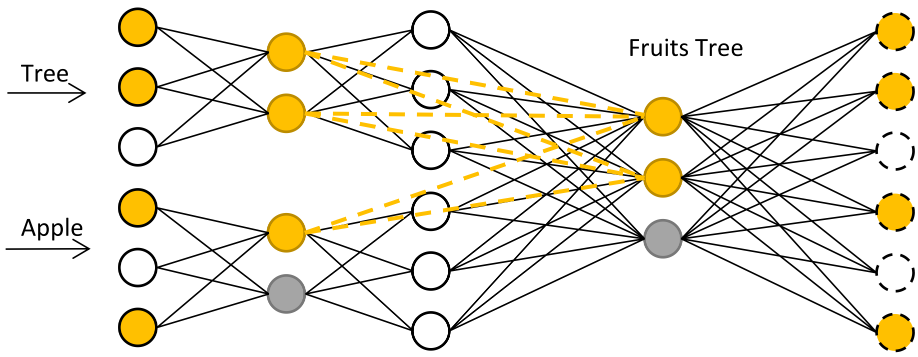


**Fig. 5 The entire network with the dotted lines representing a 'concept tree'**

As a dynamic field, there must be a dynamic system running on it, and the answer is Hebb's law[1]. Similar to Hebb's law, the learning algorithm strengthen or weaken the connections of neurons according to their activation modes. Look into one path of a concept tree, if it goes perfectly the right way from one layer to the next, then every connection of the path will be strengthened, and within a time frame, the loop formed by the path even accelerates this process; and if any step goes wrong, which meas the sibling neuron is not activated while the other is, then the connection between them will be weakened. However, most of inputs come from the outside world, so usually the lower abstract layers will be activated first, so there is actually a virtual 'phase' of the network, therefor the phase based part of Hebb's law is also playing an important role in the learning process.

Recursive network is well known by its power when dealing with time related issues, so is the network in this paper. But before that, another important feature of the network units is required, that is, the plasticity of the network units. Obviously, the lower abstract inputs the network unit is taken, the more plastic it is, and vice versa, because the more specific one concept is, the instances of the concept have more diversity, so the network units dealing with such concepts need more plasticity, while the higher abstract network units don't. But this feature is not only for static concept representing, it is even more important for processing time related information.

Due to the high plasticity of the lower abstract layers, they actually play a role as short term/working memory inside the entire network, they are easy to write and yet volatile, just like memory traces in the brain, or RAM in a computer, while the higher abstract layers are more like ROM in a computer, which storing long term/conception memories, or knowledge by short. When the short term/working memory comes out, time related information processing becomes easy, because the lower abstract layers are not just storing the outside world inputs, they are also storing the feed back information from the higher abstract layers as intermediate results, at the mean time, any information stored in it can be seen as contexts for the further step processing, in fact the entire network can be seen as a distributed, parallel solver, solving time-space related problems, of course, the static concepts are space related.

Currently we are just discussing about input, and the other question is where the output is, the answer is the output neurons are aligned with the input ones. It sounds wired but it is actually very reasonable, because the output is also very specific, imaging when an agent moves its body, the motions must be accurate and certain, just like inputs are, but still have very high diversity and are constrained/controlled by memory/knowledge, again like inputs are. Fig.5 shows how it works, the output neurons are exactly the same with the input ones, also cooperate with the entire network, but have another way out, at the same time, the actions themself also generate many inputs, just like a human is able to be aware of his/her own movements, and the inputs are another group of contexts/intermediate results to be solved/learned.

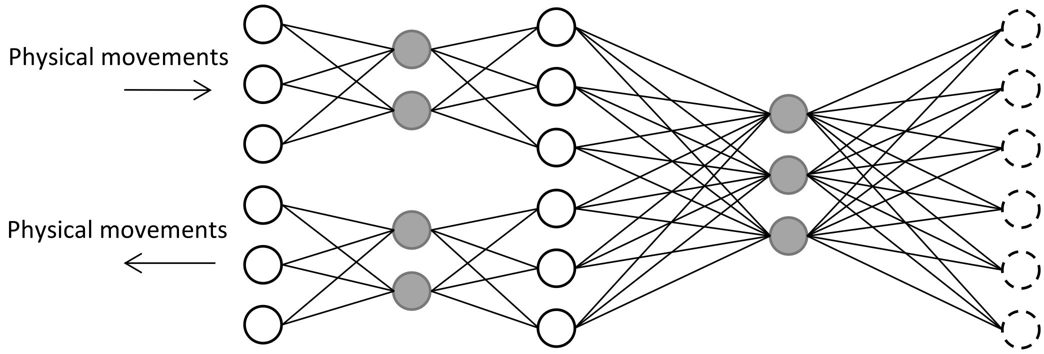


**Fig. 6 How the output such as physical movements works with the entire network**

# 3 Knowledge about intelligence

Language is one of the most significant phenomenons of intelligence, it is not just a tool for communicating, but also a tool for thinking. Obviously, the carriers of language are just visual or sound signals, no difference from the normal outside world input, when the signals are sent into the entire network, they become a part of concept trees and related to concepts which they refer to, that is how symbol systems, no matter natural language or formal language like mathematics works. But what makes natural language so extraordinary is, it can refer to concepts with no entity, even the processes of intelligence itself, such as 'reading' and 'thinking', and this is a strong evidence that an intelligence agent is aware of its own intelligent processes, which means the processes can be some kind of input to the entire network. This can be implemented by some global sensors which monitoring the state of the entire network, and sending results to the input layer of the entire network. The state can be the 'phase' of the entire network mentioned in section 2, which can indicate whether the network is mainly taking outside world inputs, in another word 'reading', or be dominated by higher abstract layers, generating intermediate results, in another word 'thinking'. Moreover, language not only has referential functions, but also has imperative functions, which means there must be executors to 'replay' the processes, in this case, intelligent processes are exactly the same with physical movements discussed in section 2, but happen inside the neuron network, not outside.

Another important state of the entire network must be monitored and replayed is the stability of the network. The definition of stability here is the measurement of how quick a group of inputs can be solved, in another word, can form a correct/acceptable concept trees, during the process, outputs are also be generated, and that is behaviour planning and actions. The stability of neuron network is very critical to intelligence agent, it indicates whether the agent is performing well or not, obviously, the more stable the network is, the better the agent is performing. In fact, the required sensor does not monitor the stability itself, instead it monitors the increment of it, so when the behaviour of the agent makes progress on solving problems, the 'good' signal is sent to the network and be recorded as concepts related to certain contexts, and the executor replays that behaviour again when proper contexts show up, and vice versa. So the conclusion is that this mechanism is actually a motivation system, driving the agent's behaviour, telling the agent what to do and what not to. On the other hand, it is also an emotion system, when 'good' signal is generated, the agent feels 'happy', when 'bad' signal is generated, it feels 'upset', and the global effects of the executor is very like the hormon secretion. Motivation systems of natural intelligence agents are usually very complex, containing a lot of evolution legacies, but in the case discussed in this paper, the agent is motivated only by the increment of stability, in another word, 'curiosity', which means the agent is purely rational, and the motivation of the agent, is pure rational motivation.

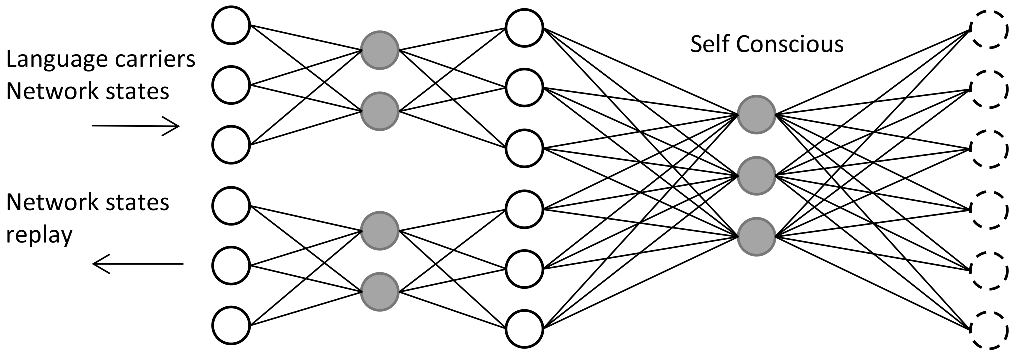


**Fig. 7 Advanced phenomenons of intelligence**

All the awareness mentioned before together forms the self conscious, so the entire network is a high ordered problem solver, it is self-referential, and according to the discussions before, all elements added to the input and output of the network have 3 common characteristics: **Appreciable**, **Recordable** and **Replayable**. Due to the general function of the network, any element with the 3 characteristics can be added to the network, and interacts with other elements, generates extremely complicated behaviour. So if all the elements are put into a tuple, then the state space of the network can be represented by a cartesian product of that tuple itself, because every element has an input and output space, and the two are homomorphic with each other. Although arbitrary element can be added to the network if needed, but for intelligence agents, there is a fix point of the state space, or in another word, there is a closure of the cartesian product, because the necessary basic elements of intelligence agents are limited, which basic here means non-self-referential elements, such as outside world inputs and physical movements, so the elements referring to the basic elements are limited, and the elements referring to the basic-elements-referential elements are also limited, so the increase of elements is convergent. For instance, the imperative sentence ‘read the book’ is corresponding a piece of knowledge which is in the pole set, because it is refers to the highly self-referential concept ‘reading’, which has already been discussed, so this paper is considered exhaustive.

## References

[1] Hebb, D.O. (1949). The Organization of Behavior. New York: Wiley & Sons.