The robot consciousness based on empirical knowledge

Yuanxiu Liao¹,²,³, Mingrui Yan³*, Suqin Tang⁴

¹Guangxi Key Lab of Multi-Source Information Mining and Security, Guilin, Guangxi, 541004, China
²Guangxi Regional Multi-Source Information Integration and Intelligent Processing Collaborative Innovation Center, Guilin, Guangxi, 541004, China
³Faculty of Computer Science and Information Engineering, Guangxi Normal University, Guilin, Guangxi, 541004, China.
⁴Faculty of education, Guangxi Normal University, Guilin, Guangxi, 540114, China.

*corresponding author’s e-mail: yan-mingrui@outlook.com

Abstract. How to make robot conscious is an important goal of artificial intelligence. Despite the emergence of some very creative ideas, no convincing way to realize consciousness on a machine has been proposed so far. For example, the integrated information theory of consciousness proposes that consciousness can exist in any place that can reasonably process information, whether brain or machine. It points out that a physical system must satisfy two basic conditions to cause the emergence of consciousness: it must have rich information, and it must be highly integrated. However, the theory does not give an idea of how to realize consciousness on a machine. In this paper, we propose the robot consciousness based on empirical knowledge. We believe that the empirical knowledge of robots is an important basis for robot consciousness, any cognitive experience of robot can lead to the generation of consciousness.

We firstly propose a formal framework for describing robot empirical knowledge; then we discuss the robot consciousness based on its own empirical knowledge; finally, we propose a cost-oriented evolutionary method of robot consciousness.

1. Introduction

Researchers in the fields of artificial intelligence (AI) and robotics have struggled for decades to develop intelligent machine systems, and no machines similar to human intelligence have yet been created. Although AI systems are comparable to the functions of the human brain in some respects, none of the systems exhibit advanced cognitive functions similar to human consciousness and emotions. Let us look at a few typical cases. The first is Google’s Go program. Google’s deep thinking team developed an AI agent based on deep reinforcement learning (deep RL), called the Deep Q-network (DQN) [1], in 2015, and let it learn how to operate Atari video games. In 2016, the team combined Monte Carlo tree search and deep RL to develop the AlphaGo AI algorithm [2], which defeated two world champions, Li Shishi of South Korea and Ke Jie of China at the game Go. Second is image recognition. The deep convolutional neural network (CNN) has enabled breakthroughs in image recognition [3], which is used such as in facial recognition [4, 5]. The application effect of deep CNN in image recognition can be seen from the ImageNet Large-Scale Visual Recognition Challenge, which ran from 2010 through 2017 [6]. A third example is natural language processing (NLP) [7], where recurrent neural networks (RNNs) have dealt well with time series problems such as text
generation and machine translation [8, 9]. The final case is driverless technology. Wayve researchers [10] combined an end-to-end conditional imitation learning method with a vehicle's lateral and longitudinal control to enable it to drive on simple urban roads. Apple researchers [11] improved the effectiveness of deep RL by finding strategies to simultaneously explore, interact, and learn from the environment.

The above cases show that AI surpasses or approaches human capabilities in some respects. But these are not true AI systems because their functions are relatively simple, and they will not produce intelligent behaviors such as emotions or consciousness. For example, AlphaGo will be at a loss if the rules change, even just by adding two horizontal and vertical lines to the board or allowing a player to move two pieces at a time.

However, some new ideas of consciousness theory bring exciting news. For example, the information integration theory of consciousness (IIT) proposed by neuroscientist Tononi [12] states that consciousness can in principle exist in any place where information can be processed reasonably, whether brain or machine. According to the theory, a physical system must satisfy two basic physical conditions before consciousness can emerge. It must have rich information, and it must be highly integrated. Although IIT provides a theoretical framework to assess whether consciousness can appear in physical systems, no actual physical system exists to test the theory. In order to explore how to make robots have consciousness, we propose a formal framework for describing robot consciousness and ideas on how to realize robot consciousness on machine.

2. Sensor-based empirical knowledge acquisition

2.1. Robot sensors and sensing data acquisition

To enable a robot to complete complex tasks in an open environment requires certain innate conditions, such as sensing, action, knowledge acquisition, planning, and task understanding. We assume that a robot is equipped with enough sensors, such as object recognition (what is object a?), external environmental condition recognition (is the road surface flat?), obstacle recognition (are there obstacles in the road ahead?), and distance discrimination (how far is the target?). These are physical devices assembled on the robot. For ease of description, we regard each function of a sensor as an independent sensor. For example, a computer vision system can recognize a variety of objects such as tables, books, and pens. We regard the functions of recognizing the table, book, and pen as three independent sensors, and give each sensor a name. For example, $BOOK(x)$ is a sensor that recognizes whether object $x$ is a book, $BLUE(x)$ recognizes whether object $x$ is blue, $ON(x, y)$ recognizes whether object $x$ is on object $y$, and $dist(x, y)$ measures the distance between $x$ and $y$.

We divide sensors into two categories based on their functions.

1. A relational sensor identifies whether a certain relationship exists between two (or more) objects; it can also identify whether an object has a certain attribute (attributes are regarded as unary relationships). For example, $ON(x, y)$ is a binary relationship sensor that can determine whether a relationship “$x$ is on $y$” exists; $BOOK(x)$ is a unary relationship sensor to determine whether $x$ has the attribute of a book. This type of sensor outputs YES when a relationship is established, and otherwise NO.

Assume that a robot is located in the environment shown in Fig. 1. The robot is equipped with sensors that recognize objects and environmental conditions, whose partial results are shown in Fig. 2.
Next, the robot needs to convert the output of the sensor into perception data in its own system. For example, the output result of the sensors in Fig. 2 is converted to the perception data of the robot, as shown in Fig. 3.

(2) A functional sensor determines certain properties of an object. Its output value is like a mathematical function that varies within a certain range. For example, the sensor $dist(x, y)$ determines the distance between objects $x$ and $y$, and its value varies within $[0, +\infty)$; the sensor $temp(x)$ determines the temperature of object $x$, whose value is in the range $(-270, +\infty)$.

If the distance between $a$ and $e$ in Fig. 1 is 2 m, then $dist(a, e)$ measures the output value of objects $a$ and $e$ as 2 m, i.e., $dist(a, e) = 2$ m. The sensing result of $dist(a, e)$ is converted to the perception data of the robot, as shown in Fig. 4.

2.2. The cognitive experience of robot

Many consciousness theories about humans, such as KTC [13] and IIT [14], assume that experience is the basis for consciousness. Drawing from this, we regard experience as a necessary condition for robot consciousness. People experience things mainly through the sense organs to obtain perceptual information, and then through the neural network, level by level, to integrate and transmit it to the conscious area of the brain. This poses a challenge. Human sense organs and neural networks are innate, so to experience things is inherent. Robot sensing organs, such as image recognition, voice recognition, radar, infrared ranging, and thermometers, are man-made; hence, there is no inherent experience. To facilitate research, we must define the robot experience. In theory, every action of a robot can produce an experience. To simplify the discussion, here we only discuss the experience of using sensors to obtain external environmental information.

**Definition 1** The robot uses its own sensors to obtain relevant information about a thing and converts this information into its own perception data. This process is called a cognitive experience of the thing by the robot.
For example, a robot applies a physical sensor \( \overline{BOOK}(x) \), whose function is to recognize whether object \( x \) is a book, to an object \( b \) in the environment. If the sensor’s output is YES, then the robot system converts this sensing information into its own perception data that \( b \) is a book. This is a cognitive experience of a robot.

2.3. The empirical knowledge of robot

Empirical knowledge is acquired by participating in cognitive practice. Empirical knowledge is acquired independently by robots without external intervention as a kind of private knowledge. The empirical knowledge reflects experience through the cognitive process, and it is closely related to cognition. We give a more formal definition of robot empirical knowledge.

**Definition 2** Assume that the robot has an experience and obtains relevant perception data. If the robot converts these perception data into knowledge that its own system can express, then this knowledge is called the robot's empirical knowledge.

We outline how robots gain experience through sensors and integrate the information from these experiences into knowledge. First of all, we introduce some commonly used symbols and their meanings in this article.

- Symbols used to represent physical sensors: \( \overline{BOOK}(x) \), \( \overline{BLUE}(x) \), \( \overline{ON}(x,y) \), …
- Symbols used to represent sensing data: \( <BOOK, b, YES> \), \( <BLUE, d, NO> \), …
- Predicate formulas used to represent sensor function: \( \overline{BOOK}(x) \), \( \overline{BLUE}(x) \), \( \overline{ON}(x,y) \).
- Predicate formulas used to represent empirical knowledge: \( \overline{BOOK}(a) \), \( \overline{BOOK}(b) \), \( \overline{BOOK}(d) \), \( \overline{BLUE}(e) \), …

Taking Figs. 1, 2, and 3 as examples, we illustrate the process of the robot's experience gained by sensing the environment.

Assume that the robot activates sensors \( \overline{BOOK}(x) \) and \( \overline{BLUE}(x) \) to sense the environment. \( \overline{BOOK}(x) \) returns sensing data \( <BOOK(a),NO> \), \( <BOOK(b),YES> \); \( \overline{BLUE}(x) \) returns sensing data \( <BLUE(d), NO> \), \( <BLUE(e), YES> \). Then, the system integrates the sensing data \( <BOOK(a),NO> \), \( <BOOK(b),YES> \), \( <BLUE(d), NO> \), \( <BLUE(e), YES> \) into empirical knowledge: \( \overline{BOOK}(a) \), \( \overline{BOOK}(b) \), \( \overline{BOOK}(d) \), \( \overline{BLUE}(e) \). The process by which the robot acquires empirical knowledge is shown in Figure 5.

**Integrating information**

Fig. 5. The process for robot to acquire empirical knowledge.

3. The generation of robot consciousness

The study of human consciousness is different from the study of robot consciousness. Human perception and consciousness ability are innate. The study of human consciousness is to discover its
essential attributes through observation, experiment, analysis and other methods. The research of robot consciousness aims to construct robots with consciousness characteristics. To study human consciousness is to discover consciousness attributes, and to study robot consciousness is to construct consciousness attributes.

The structure and way of processing information between human brain and robot are very different, almost incomparable. Therefore, it is difficult to apply the research methods of human consciousness to robot consciousness. This is the main reason for the slow progress of machine intelligence research. A new research method and a theory of consciousness applicable to machines may help solve this problem. The focus of this paper is whether a robot consciousness framework or model can be implemented on existing robot systems. This research considers two questions: What is robot consciousness, and how to make the robot system aware? We believe that when the robot faces a new environment, if there are no preset decision-making options in the system and the robot can make appropriate responses based on its own empirical knowledge, then the robot will show a conscious response to new things.

3.1. Action-based robot empirical knowledge

Human consciousness can appear in a variety of different situations. Robot consciousness can also be designed to be produced in a variety of different situations. In this article, we discuss the robot consciousness generation method based on action results. To this end, we first introduce the formal representation of action-based robot empirical knowledge.

For the convenience of discussion, we assume that the robot has the following capabilities:
- Have sufficient perception of the environment and be able to discern the actual results of actions.
- Have sufficient ability to implement actions and be able to distinguish whether the actions implemented are successful. If the actual result after implementing the action is consistent with the result described by the result axiom, the action is successful. Otherwise, the action fails.
- Every action is assigned a cost C, indicating the price to be paid when the action is implemented. If the action succeeds, assign a smaller value to C; if the action fails, assign a larger value to C.

We use a specific example to illustrate the representation of action-based robot empirical knowledge.

**Example 1** Consider a situation: Suppose that when the robot is grabbing an item, it can choose 7 different grabbing strengths, in descending order: \( F_1, F_2, \ldots, F_7 \). It is assumed that the robot does not have any empirical knowledge about picking up eggs. The robot uses three different strengths to grab the egg, and the action results are as follows:

a) For the first time, the robot grabbed the egg with the strength \( F_5 \), and the egg was broken;

b) The robot grabbed the egg with the strength \( F_1 \) for the second time, but failed to grab the egg;

c) The robot grabs the egg with the strength \( F_3 \) for the third time, and the egg is intact.

Finally, the robot obtains the empirical knowledge of picking up eggs three times, which is expressed as follows with the predicate formula:

\[
S_{emp} = \{ \text{Grab (egg, } F_5, \text{ broken)), } \text{Grab (egg, } F_1, \text{ broken)), Grab (egg, } F_3, \text{ intact)} \}
\]

Among them, \( S_{emp} \) represents the robot’s empirical knowledge base; \( \text{Grab (egg, } F_5, \text{ broken)}) \), \( \text{Grab (egg, } F_1, \text{ broken)}) \), \( \text{Grab (egg, } F_3, \text{ intact)} \) represent the empirical knowledge gained by the robot after the action of picking up eggs with three different forces.
3.2 Methods of generating robot consciousness

Because the cognitive system of robots is too different from that of humans, it is impossible to transfer the way humans produce consciousness to robots. Therefore, we propose a cost-oriented algorithm that allows robots to evaluate events with unpredictable consequences and then make appropriate decisions based on existing empirical knowledge. The evaluation result of this algorithm can be regarded as the simplest and most primitive consciousness of the robot. With the accumulation of robot empirical knowledge and the evolution of algorithms, the level of consciousness of robots will continue to increase. The basic form of artificial neural network is also very simple. After continuous improvement and development, all kinds of neural networks with deep learning capabilities today have powerful functions. In this article, our purpose of studying robot consciousness is not to reproduce human consciousness on machines, but to make robots smarter.

Definition 3 In the absence of external information reference, if the robot can evaluate the consequences of a specific action (action completed) in the scene situation based on the integration of its own empirical knowledge, or evaluate the consequences of the planned action in the scene situation (action not Implementation) make predictions, and these evaluation results and prediction results are called robot awareness based on empirical knowledge.

The next time a similar situation occurs, if there are no preset decision-making options in the system, the robot will choose appropriate actions based on these consciousnesses.

Example 2 Consider the situation of Example 1. The actions Grab (egg, \(F_5\), broken), Grab (egg, \(F_1\), broken), Grab (egg, \(F_3\), intact) are assigned costs 10, 4, -2, respectively. Then the robot consciousness based on these empirical knowledge can be expressed as:

\[
S_{con} = \{\text{Grab (egg, } F_5, 10), \text{Grab (egg, } F_1, 4), \text{Grab (egg, } F_3, -2)\}
\]

among them, \(S_{con}\) represents the consciousness library of the robot; each element in \(S_{con}\) represents a kind of consciousness of the robot. That is, the robot evaluates the corresponding cost of implementing various actions. This is the most direct, simplest, and most basic consciousness of a robot based on empirical knowledge. With the accumulation of robot empirical knowledge, its consciousness should continue to evolve and its level of consciousness should continue to improve. To this end, we propose two simple algorithms for optimizing robot consciousness.

Definition 4 It is assumed that the robot performs the same action A many times and acquires multiple empirical knowledge. Integrating the sum of the cost of each action into a new cost C, then this new cost is called the robot's overall assessment of the action A, so it can be regarded as the robot's overall awareness of the implementation of the action A.

Here are two simplest cost integration algorithms to calculate the overall consciousness of the robot.

(1) Simple summation algorithm

We use an example to illustrate the robot's "simple summation" algorithm.

Example 3 assumes that the robot has "grabbed the egg" several times. The acquired empirical knowledge are shown in Table 1. Then the robot's comprehensive consciousness of implementing each action is as follows:

\[
S_{con} = \{\text{Grab (egg, } F_1, 0), \text{Grab (egg, } F_2, 10), \text{Grab (egg, } F_3, 10), \text{Grab (egg, } F_4, 8), \text{Grab (egg, } F_5, -24), \text{Grab (egg, } F_6, -2), \text{Grab (egg, } F_7, 4)\}
\]

Table 1 The empirical knowledge of the robot grabbing egg

| Action  | Strength | Cost of Success | Number of Successes | Cost of Failure | Number of Failures | Cost Sum |
|---------|----------|-----------------|---------------------|----------------|--------------------|----------|
| Grab egg | \(F_1\)  | -2              | 0                   | 10             | 0                  | 0        |
| Grab egg | \(F_2\)  | -2              | 0                   | 10             | 1                  | 10       |
| Grab egg | \(F_3\)  | -2              | 0                   | 10             | 1                  | 10       |
| Grab egg | \(F_4\)  | -2              | 6                   | 10             | 2                  | 8        |
| Grab egg | \(F_5\)  | -2              | 12                  | 10             | 0                  | -24      |
| Grab egg | \(F_6\)  | -2              | 7                   | 4              | 3                  | -2       |
| Grab egg | \(F_7\)  | -2              | 0                   | 4              | 1                  | 4        |
The comprehensive consciousness of a robot is a comprehensive evaluation of multiple empirical knowledge of a certain thing, and the cost reflects the strength of the robot's consciousness. In the next action, the robot will choose a low-cost action to complete the task. For example, when the robot has the comprehensive consciousness shown in Table 1, in the next action of picking up an egg, the robot will choose $F_5$ force to complete the task.

(2) Correlation derivation algorithm

The correlation derivation algorithm is based on the existing empirical knowledge to derive the result of actions that have not been implemented. For example, in the empirical knowledge in Table 1, the intensity has a correlation of "size": $F_1>F_2>\cdots>F_7$. On the other hand, we use $C_1$, $C_2$, $\cdots$, $C_7$ to represent respectively the corresponding costs of $F_1$, $F_2$, $\cdots$, $F_7$. It is easy to see that $C_5=-24$ is the minimum of the sum of various costs, and there is correlation: $C_1>C_2>C_3>C_4>C_5$. It can be predicted from the correlation: using $F_1$ force to grab the egg will fail. That is, the cost of $C_1$ should be 10.

(3) Other integration algorithms

Intelligent algorithms such as analogical reasoning and case-based reasoning can be used to integrate the cost of robot comprehensive consciousness. But the reasoning must be based on the robot's empirical knowledge, and can't use other external information. For example, preset the robot with "200g force for grabbing eggs" and "500g force for grabbing apples" in advance. When the robot meets these tasks, it will execute according to these options, so the robot's decision-making is not conscious.

4. Conclusion

In this paper, we clearly describe the robot consciousness based on empirical knowledge, and put forward the idea of constructing robot consciousness. Firstly, we propose a formal framework to describe the cognitive experience of robots based on their own sensors, and propose a method for robots to acquire empirical knowledge. These works lay a foundation for the construction of robot consciousness. Next, we propose a cost oriented method to generate robot awareness. We assume that robots pay for every action. The robot evaluates the role of action in completing the task through the results reflected by empirical knowledge, and measures the importance of action with cost. The robot's evaluation of the importance of action is a conjecture based on its own empirical knowledge, which is a kind of personal cognitive behavior, so it can be regarded as the robot's consciousness. Finally, we propose a method to generate the robot's comprehensive consciousness and an algorithm to integrate the cost of comprehensive consciousness. The main contribution of this paper is that when performing a task, if there are no preset decision options in the system, the robot can form a kind of consciousness based on the integration of its own empirical knowledge. And the robot can guide the next action according to its own consciousness, so that the task can be completed.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Nos. 61662007 and 61967002).

References

[1] V. Mnih, K. Kavukcuoglu, D. Silver, A.A. Rusu, J. Veness, M.G. Bellemare, A. Graves, M. Riedmiller, A.K. Fidjeland, G. Ostrovski, S. Petersen, Human-level control through deep reinforcement learning, Nature 518 (7540) (2015) 529–533. http://dx.doi.org/10.1038/nature14236.

[2] D. Silver, A. Huang, C.J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, Mastering the game of Go with deep neural networks and tree search, Nature 529 (7587) (2016) 484–489. http://dx.doi.org/10.1038/nature16961.

[3] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436–444. http://dx.doi.org/10.1038/nature14539.
[4] S. Balaban, Deep learning and face recognition: The state of the art, in: Proceedings Biometric and Surveillance Technology for Human and Activity Identification, Baltimore, Maryland, United States, 2015. http://dx.doi.org/10.1117/12.2181526.

[5] Y. Taigman, M. Yang, M.A. Ranzato, L. Wolf, Deepface: closing the gap to human-level performance in face verification, in: Proceedings Conference on Computer Vision and Pattern Recognition, 2014, pp. 1701–1708. http://dx.doi.org/10.1109/CVPR.2014.220.

[6] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A.C. Berg, Imagenet large scale visual recognition challenge, Int. J. Comput. Vis. 115 (3) (2015) 211–252. http://dx.doi.org/10.1007/s11263-015-0816-y.

[7] J. Zhang, C. Zong, Deep neural networks in machine translation: an overview, IEEE Intell. Syst. 30 (5) (2015) 16–25. http://dx.doi.org/10.1109/MIS.2015.69.

[8] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, Y. Bengio, Learning phrase representations using RNN encoder-decoder for statistical machine translation, in: Proceedings EMNLP Processing, 2014, pp. 1724–1734. arXiv:1406.1078

[9] Sutskever, J. Martens, G.E. Hinton, Generating text with recurrent neural networks, in: Proceedings 28th International Conference on International Conference on Machine Learning, Bellevue, Washington, USA, 2011, pp. 1017–1024.

[10] J. Hawke, R. Shen, C. Gurau, S. Sharma, D. Reda, N. Nikolov, P. Mazur, S. Micklethwaite, N. Griffiths, A. Shah, A. Kendall, Urban driving with conditional imitation learning, (2019) arXiv:1912.00177

[11] Y. Tang, Towards learning multi-agent negotiations via self-play, in: Proceedings IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), Seoul, Korea (South), 2019, pp. 2427–2435. http://dx.doi.org/10.1109/ICCVW.2019.00297.

[12] G. Tononi, An information integration theory of consciousness, BMC Neurosci. 5 (1) (2004) 42. http://dx.doi.org/10.1186/1471-2202-5-42.

[13] G. Tononi, Integrated information theory, Scholarpedia 10 (7) (2015) 4164. http://dx.doi.org/10.4249/scholarpedia.4164.

[14] V.A.F. Lamme, Separate neural definitions of visual consciousness and visual attention; a case for phenomenal awareness, Neural Netw. 17 (5–6) (2004) 861–872. http://dx.doi.org/10.1016/j.neunet.2004.02.005.

[15] R. Reiter, Knowledge in Action: Logical Foundations for Specifying and Implementing Dynamical Systems, MIT Press, Cambridge, MA, 2001.