Two-Component Scheme of Cognitive System Organization: the Hippocampus-Inspired Model

Ekaterina D. Kazimirova
AO Kaspersky Lab
Moscow, Russia
e-mail: Ekaterina.Kazimirova@kaspersky.com

***Abstract* – This paper presents a hypothesis on two-component principle of the cognitive system organization. We propose a biologically inspired architecture, which involves two subsystems, external and internal. Both subsystems are capable of compressing data by converting the images into symbols. They are connected at the symbol level, with necessary "relay" controlling their interaction. The External Subsystem reflects and processes the external image information. The Internal Subsystem reflects the internal states of the system and contains a "personal sense" of the external images; thus, it can be considered as a Library of Emotions. We propose a hypothesis that in living systems the role of a "relay" (a connector) between the external and internal libraries may be performed by the hippocampus. When applied to an artificial cognitive system, our hypothesis would imply the inclusion of certain modules (blocks), constructed in analogy with the hippocampus, into the system. This approach could be useful for designing self-regulatory systems that would account for both the external and internal factors. It may also be important for large industrial systems related to cyber-physical objects, which have hundreds of thousands of sensors; in this setting, decoupling the internal and external information may ensure efficient monitoring and protection. Technically, such two-component system could be represented as a block (modular) neural network.**

Keywords - emotion; symbol; cognitive architecture; hippocampus; industrial system

I. INTRODUCTION

Nowadays, robotic systems are growing more closely connected to humans. In the future, many of them may become an integral part of a human being for a certain time. A car equipped with an autopilot may be considered as an example of such integration. For such situations, simulation of "emotion" and "personal meaning" of events in artificial cognitive systems becomes very important.

Emotions are analyzed and interpreted within the framework of various scientific disciplines. Psychology is trying to define the basic mechanisms of emotions and their relation to personality. Neurophysiology explores the neural substrates of emotions (e.g., neural networks involved). Scientific efforts (both in robotics and in other domains) aimed at the development of General Artificial Intelligence seem to be especially active in the field of imitating human emotions. This includes the use of video recordings of human facial expressions for generating facial expressions of robots, mirroring human facial expressions, etc.

Obviously, the mere simulation of emotions is not enough, if we want an emotional unit to regulate the behavior of the robotic system. Thus, we need a certain understanding of the very essence and nature of emotions and we have to create an architectural solution in order to bring that understanding into reality.

There is a number of interesting attempts in this area. In some approaches, emotions are represented as noise [1]. In others, attention is drawn to the importance of the "mental states" of the intelligent agent [2], etc.

There are also efforts underway to draw parallels between the action of neurotransmitters in the living brain and the computational processes (in computers), such as computing power, memory distribution, learning and storage [3].

In this paper, we argue that a simple decoupling of any cognitive system into two subsystems could be useful for different disciplines. Such approach could advance our understanding of emotions as a reflection of internal states and regulation of behavior, both for robotic and living systems.

In this paper, we discuss the Symbol-Image model of cognitive system (Section II), present the two-component model (Section III), discuss the possible verification and application of the model (Section IV), and present conclusions and future perspectives (Section V).

II. SYMBOL-IMAGE MODEL OF COGNITIVE SYSTEM

First, we have to select a paradigm for solving the problem of emotion modeling. Let us examine the symbol-image model of the cognitive system [4]. The cognitive space consists of elements that can be described as symbols, images and attributes. The model describes our informational fields in terms of hierarchical structures of symbols and images. Previously, we have introduced the term "attribute" to describe the content of an image via the fields of attributes [5].

One of the important consequences of this simple model is that the attribute fields (different characteristics of the images) can overlap. In this model, symbols play a very important role, because they separate different images (the corresponding group of attributes). We propose that

 ISBN: 978-1-61208-531-9

symbols represent the memory of such a cognitive system, because they prevent mixing of images and provide a possibility for storing the images as entire units.

Concerning the living brain, this arrangement implies that the encoding neuron-symbols should be located in the structures responsible for memory. One of the key memory-related structures in the living brain is the hippocampus, as its lesions lead to inability to form memory of recent events.

III. THE TWO-COMPONENT MODEL

A. Internal and External Symbol Libraries

We suppose that the system, which is responsible for acquiring images from the external world, is only a part of the living cognitive system. Another part of the cognitive system has to encode its own inner states (see mental states, [2]). Here, we put forward a hypothesis that there are two cognitive subsystems – "external" and "internal". Our idea is that the basic principles of the organization of these systems may be similar. We assume that both of them contain images compressed into symbols.

Obviously, those subsystems have to be interconnected. Actually, it is this connection that provides appropriate reactions to external events, forms behavioral patterns and allows making predictions (forecasts). This interaction between the subsystems can be indirect – for example, through the decision-making unit – but it still has to be present.

In artificial systems, we can also try generating two subsystems – external and internal – and connecting them. This would constitute a simple architectural solution. Furthermore, we assume that, in the artificial cognitive systems, the internal subsystem may parallel (be analogous to) the emotional component of the live cognitive system. It is "emotional", because it automatically reflects some kind of "personality meaning" of the external images and because it forms a basis for generating prognoses and forming behavioral strategies.

We propose that the two cognitive subsystems are connected at the level of symbols.

B. Is the Internal Library an Emotional Library?

In a living organism, formation of such two-component system is a result of a life experience. It can serve as a basis for connection between the internal and external environments, for generating prognoses and forming behavioral patterns and strategies. Thus, a favorable context and images of external environment will correspond to "positive" internal images and symbols, while negative external events and images will correspond to "negative" internal images and symbols.

We may consider the "internal" library as an "emotional library", because it summarizes and reflects the internal states and the personal meanings of the external events and images.

C. Specific Role of the Hippocampus

Let us consider the arguments in favor of the idea that the hippocampus could play the role of a "relay", i.e., a connector between two subsystems, the external and internal symbols' libraries.

1) The hippocampus is connected with both, the cortex (that receives information from the outside world), and the limbic structures (that receive information from the internal organs and are responsible for emotions).

2) The hippocampus is involved in the memory consolidation [6]. We can assume that the hippocampus is the very place where the symbols are stored. The activation of the neuron-symbol stored in the hippocampus activates the image associated with that symbol in the cerebral cortex.

3) Neurogenesis (production of new neurons) was observed in the adult hippocampus [7] and was not detected in most of the other brain structures. We propose that new neurons may be required for marking (labeling) new images.

Taken together, these arguments make us suggest the hypothesis that the hippocampus is an integrator of the two (internal and external) subsystems in the brain at the symbol's level.

D. A Two-Component Hippocampus-Inspired Model

We propose a simple model of the cognitive system, which consists of two subsystems. One of them is responsible for processing and compressing the external information. The second one relates to internal information. They are connected by means of a "relay". In the living systems, the hippocampus could play the role of such a relay. This implies that everything that we see, hear, feel, and perceive via our sensory systems, results in formation of images in the brain cortex. This is similar to the appearance of images in a kaleidoscope. Later on, the images are to be converted into symbols.

At the same time, the internal system of receptors records the actual indices of the organism, its hormonal, physical, biochemical state, etc.

There are two information flows. One of them reflects the external stimuli, while another one reflects the internal changes.

The brain, in order to perform its functions effectively, has to process these two information flows simultaneously (subject to certain time intervals). The hippocampus appears to be a plausible candidate for coordinating these two data streams, since it is a key element connecting the cortex with the limbic system. Through the limbic system, the hippocampus is connected with the thalamic neural block, which is responsible for controlling the internal states.

Thus, in the artificial cognitive systems that are based on the emotional management model (e.g., [8]), it is possible to reproduce this type of data separation and integration. The internal state of the intelligent agent (IA) is described by data from the internal state sensors. This information has to

 ISBN: 978-1-61208-531-9

be combined with the data obtained from the outside world. Thus, the external data acquire personal meaning in terms of the internal states of IA.

IV. Possible Verification and Applications

Our hypothesis is that there are links between internal states of the system and corresponding external images. This assumption can be verified by explaining some psychological phenomena.

Let us consider a case of post-traumatic stress disorder, when a person throws himself into a ditch at a certain sound, e.g., a sound resembling a shell blast. According to our model, such behavior could be explained by activation of a single attribute and (selectively, despite the context) of the related symbol that means the "mortal danger" in the Internal Library.

In another example, an external image is ambiguous, i.e., it plays positive and negative roles simultaneously. So, it activates contrasting states (images) in the Internal Library of Images. In psychology, such situation is called a "double bind". E.g., if mother's attitude towards a child switches from overly affectionate to overly strict, it may lead to nervous and mental disorders up to schizophrenia. The Double Bind theory was described by Gregory Bateson and his colleagues in the 1950s [9]. The phenomenon of divarication (e.g., identical commands lead to different processes) in the artificial neuro-semantic graph is described in [10].

In the artificial cognitive system based on the Symbol-Images Cognitive Architecture (SICA), the absence or presence of instability (like divarication of images described above) or hyperstability may serve as a diagnostic factor. This implies that appearance of such double patterns could be treated as the indicator of anomaly.

Being applied to the goal of monitoring the state of industrial system, the model results in conclusion that certain set (sequence) of processes in the physical part of the system should correspond to a certain (identical) set of operator's commands. The observed divarication may be an indicator of a hacker's intrusion into the system.

V. Conclusions and Future Work

Since the robotic systems become more and more closely connected with humans, the importance of the intelligent systems that provide "personal sense" of the information, or the "emotional response" is constantly growing. However, modern neural networks, as a rule, do not provide any "personal interpretation" of the data received.

We propose a simple two-component hippocampus-inspired model of a cognitive system, which could fill that gap. In the artificial cognitive systems, this concept corresponds to embedding a certain module (block), which should be constructed to perform the main functions of the hippocampus.

The model has explanatory power for a range of psychological phenomena and is to be developed further. Furthermore, our hypothesis can be applied to industrial system for enhanced monitoring and protection.

Acknowledgments

The author is grateful to Sergey Sadovnikov, Andrey Lavrentyev, and Alexander Kharlamov for the fruitful discussions.

References

[1] O. D. Chernavskaya et al., "An architecture of the cognitive system with account for emotional component," Biologically Inspired Cognitive Architectures, vol. 12, pp. 144–154, 2015.

[2] A. V. Samsonovich, "Emotional biologically inspired cognitive architecture," Biologically Inspired Cognitive Architectures, vol. 6, pp. 109-125, 2013.

[3] M. Talanov, J. Vallverdú, S. Distefano, M. Mazzara, and R. Delhibabu, "Neuromodulating cognitive architecture: towards biomimetic emotional AI," IEEE 29th International Conference on Advanced Information Networking and Applications, pp. 587-592, 2015.

[4] O. D. Chernavskaya, D. S. Chernavskii, V. P. Karp, A. P. Nikitin, and D. S. Shchepetov, "An architecture of thinking system within the Dynamical Theory of Information," Biologically Inspired Cognitive Architectures, vol. 6, pp. 147--158, 2013.

[5] E. D. Kazimirova, "Elements of the symbol-image architecture of cognition and their parallelism to certain linguistic phenomena," "Нейрокомпьютеры" ("Neurocomputers"), vol. 4, pp 35-37, 2015. Available from: http://www.radiotec.ru/catalog.php?cat=jr7&art=16364 2017.02.06

[6] L. R. Squire, L. Genzel, J. T. Wixted, and R. G. Morris, "Memory consolidation," Digital Object Identifiers (DOIs): 10.1101/cshperspect.a021766 Available from: https://www.ncbi.nlm.nih.gov/pubmed/?term=26238360 2017.02.06

[7] J. T. Gonçalves, S. T. Schafer, and F. H. Gage, "Adult Neurogenesis in the Hippocampus: from Stem Cells to Behavior," Cell., vol. 167(4), pp. 897–914, 2016.

[8] S. M. Sadovnikov, S. V. Moiseev, and E. D. Kazimirova, "Utility function of intellectual agent and its self regulation," pp. 152—153, 2013 (in Russian). Available from: http://nd-cogsci.iapras.ru/2013/img/ND-2013.pdf 2017.02.06

[9] G. Bateson, D. D. Jackson, J. Haley, and J. Weakland, "Towards a Theory of Schizophrenia," Behavioral Science, vol. 1, pp. 251–264, 1956.

[10] A. B. Lavrentyev, "Neurosemantic approach and free energy minimization principle," The Sixth International Conference On Cognitive Science, pp. 68-70, 2014.

 ISBN: 978-1-61208-531-9