THE PRINCIPLES OF ACTION OF INTELLIGENT SYSTEMS[1]

A. V. Gavrilov

Novosibirsk State Technical University
Novosibirsk, Russia
email: andr_gavrilov@yahoo.com

Introduction

In the field of artificial intelligence, the concept of hybrid intelligent systems [1-4], which combines various knowledge representation and processing methods as well as traditional AI paradigms and neural network models, is being actively developed. The final goal (if we can talk about a final goal here) is to develop a strong human-like AI reasoning and learning process. This goal is based on:

1) The need to understand and predict the behavior of artificial intelligent systems (e.g. robots) in the same way as human beings. This need to have the same understanding in AI systems as in human beings was addressed some time ago by N. Wiener [5]. Without human-like AI it would be impossible for robots and other artificial intelligent systems to become assimilated into human society.
2) The need for self-reproduction (and self-evolution) in smart machines without any prejudice to human society. This necessity comes from the divide between dramatically more complex information technologies (especially in terms of AI) and the relatively unchanged levels of education used to train qualified IT development specialists.

[1] This is a revised version of the paper “The principles of action of intelligent systems”, originally published in 2003 in the *Proc. of Int. Conf. IST*, Novosibirsk, Russia, v.3, pp. 91-94. I am grateful to Wendy Piemonte for her assistance in revising the English version.

The real problem therefore lies in creating an intelligent system theory that combines the knowledge and experience acquired in various fields of science related to acquiring intelligence and adaptive behavior. In this paper the author has tried to formulate the main working principles of intelligent systems which reflect the most fundamental aspects of simulated reasoning and learning when interacting with the environment.

1. The principle of associative recall

The ability to be trained (to adapt, expand and adjust a knowledge base, etc.) is an essential part of every truly intelligent system (artificial "intelligent" systems that do not possess this ability, certainly cannot be called "intelligent").

The training principle can be formulated as follows.

When interacting with the environment, the intelligent system stores the associations between different images (stimuli, signals, signs, actions, etc.) which it uses to plan and execute its behavior. The associations originate from a recall process (reading associations from memory) of images based on their fragments. The associations that are involved during the storing process are reinforced. When sufficiently consolidated, the associations can be used to denote the relations (attributive, cause and effect, case, etc.) between entities.

The recall process (reading associations from memory) increases the importance (strength) of associations and makes it easier to recall them in the future. The process of storing is supervised by previously stored information. In other words, we see what we are ready to see. At this level of abstraction, the storing (of associations) is similar to interpretation: that is, the context in which we see anything controls our perception and storing processes. This comes from a feedback between our high and low levels of perception. This process is linked to attention (see principle 2).

A more detailed description of this principle as a model of associative thinking suggested by the author can be found in [6, 7].

The principle of associative recall can be considered a variant of the holographic metaphor suggested by Pribram [8] in describing how the brain functions.

2. The principle of concentration and economy of resources

The principle of the concentration and economy of resources can be formulated as follows:

Intelligent systems (natural or artificial) have a mechanism that selects (recognition) and activates the information resources (neurons, neural ensembles, frames, rules, etc.) that are essential to the solution of an actual task by the intelligent system, and that deactivates the resources that are not essential to the solution of an actual task.

The principle of concentration and economy at a level of conscious thought is inextricably related to the concepts of goal and purposefulness [9] and is expressed as being the concentration of efforts (information and power resources) needed to achieve a goal. At a subconscious level, the principle of concentration expresses, for example, the concentration of attention needed to solve a task (Adaptive Resonance Theory of Grossberg [10] and Attentional Semantics [11]).

This principle is apparently connected with the mechanism of emotions [12] in intelligent systems. One can say that, on the one hand, the strength of emotions influences the degree of concentration of resources. On the other hand, the quality of emotions influences the choice of the goal on which the resources that are necessary to achieve it, are concentrated. It is also possible to assume that emotions are connected with the previous principle, and that the strength and "sign" of the emotions (positive or negative) stimulate the consolidation of those associations occurring along with positive emotions.

The level of attention depends on the level of emotion, e.g. if we feel a strong emotion (positive or negative) then we only focus our attention on the objects (subjects, situations, concepts and so on) that are connected with this emotion. Emotions influence both goal selection and achievement. The process of achieving a goal is source of positive emotions and the presence of any unexpected obstacles is source of negative emotions.

3. The principle of uncertainty [13]

In quantum physics the principle of uncertainty by Werner Heisenberg establishes a link between the accuracy with which the coordinates of an elementary particle are defined and its measured energy. In the same way, in the theory of intelligent systems it is possible to formulate the principle of uncertainty by establishing a relationship between the accuracy of definition (recognition) of the internal semantics of an object (syntax or structure of an image) and its interrelations with other objects (external or simply the semantics of an image).

This can be formulated as follows:

The more precise the recognition of the structure of an image, i.e. its internal semantics, the less accurate the recognition of its interaction with other images, i.e. its external semantics.

The principle of uncertainty means that any intelligent system cannot understand the syntax of images and their external semantics to the same extent. In other words, it is impossible to simultaneously study the detailed structure of an object and its interrelation with other objects. Or, the more the developer of knowledge representation (or the user or the interpreter) pays most of his/her attention to representing and recognizing the details of the structure of objects (concepts), the more he/she is forced to ignore the representation and recognition of the interrelations between objects. This principle holds true both when formalizing knowledge during the development of the intelligent system, and during the interpretation (application) of knowledge.

Just as errors can occur when defining the coordinates and energies of an elementary particle because of the characteristics of the "observer" (measurement resources), in the same way processing knowledge errors depend on the memory size of computing resources and on the speed of the inference used at interpretation. In a human brain, these characteristics apparently correspond to the brain's ability to store and recall images.

Heisenberg's principle, as well as the aforesaid principle, are based on the unavoidable unity of, and interconnection between, the observed object and the observer and the necessity to consider them a single system. Consequently, as far as intelligent systems are concerned, it is not important who the observer of knowledge is: that is, whether the observer is a part of them (the interpreter of knowledge) or an external observer (the user or the developer).

This analogy between images and elementary particles is probably not accidental. In fact, the analogy between hologram and image in biological memory was formulated in [8]. Likewise, the properties of the image recall process and those of the hologram restoring process are similar. The concepts of quantum physics in artificial intelligence [14-15] and frameworks of quantum computers [16] are now being widely developed.

This principle is very likely to be one of the reasons for the appearance (evolution) of the ability of abstraction in natural intelligent systems. This is because it limits the ability of the system to establish a relationship between cause and effect and to plan, forcing it to operate only with the concrete images of an external world.

This principle clearly applies when formalizing the knowledge base of an intelligent system (for example, the consulting model in Expert Systems): if the developer of a knowledge base wants to describe the interrelations of a given concept with other concepts to the greatest possible extent, the representation of this concept "is blurred". Its selection as an independent concept (data structure) becomes problematic. It becomes more and more a part of the other concepts and loses its value as a separate object with an internal structure.

The result of this principle in knowledge engineering is the appearance of various methods of structuring or systematizing knowledge, such as structured semantic networks, frames, ontologies, and multi-agent systems.

4. The principle of unity in fuzzy reasoning and certain other operations

In natural intelligent systems, the reasoning process is always based on using fuzzy or corrupted images, inexact concepts, incomplete descriptions, conditional outputs with incomplete information, etc. The reasons for this are the limited capacity of sense-organs, localized perceptions of the environment and the non-specific character of the environment (world), though the last factor is disputable since we judge an external world by means of the same limited sense-organs, even though this is amplified with every possible tool and instrument. On the other hand, the operations (actions) that can be performed by an intelligent system, have a specific character such as, "taking an object", "throwing an object", "raising a hand", "the birth of a child", "inclusion of the defined muscle", etc.

In the elementary case of the production of a conditioned reflex, an associative link is created between a situation and the operation to be performed when the situation occurs. Here, the reasoning process is reduced to recognizing a situation and the associative retrieval of an operation appropriate to it. It seems reasonable to assume that an attempt to reason in terms of such a "situational" thinking has led to abstract thinking, taxonomy, classification and classical logic. Fundamentally, the strict inference or solution of a task described in a mathematical, precise way is reduced to a decision-making chain using the "situation – operation" principle.

More creative processes of reasoning involve thought through images, which provide the lowest risk of losing information about the situation (task) in the decision-making process. This is because at all levels of task solution (before making a decision about the operation) a large number of tags describing an image of a situation are used. The greater the vector of tags used in an associative recall chain (reasoning), the more the brain is involved in the reasoning process, and the more likelihood there is of obtaining nontrivial solutions.

A principle describing the unity of fuzzy reasoning and certain other operations may be formulated as following:

The basis of reasoning lies in operating with fuzzy images by means of a process of associative recall of images (see Principle No. 1). At the end of the process, a choice of certain operations is carried out (recalling of it): it is therefore possible to associate the successful choice (the solved task) with the focusing of attention, the start of operation as programs of operation motor neurons, etc. Thus the selected operation is involved as a tag in the further process of reasoning.

Conclusion

The suggested principles, in opinion of the author, are fundamental in describing the operations (behavior) of intelligent systems. All the other principles and models that exist and can be further formulated are more specific and particular, and can be considered a consequence of the principles formulated above.

The principles formulated in this article are the result of twenty years' experience by the author and other researchers in the field of the development of hybrid intelligent systems, and can be considered part of the theory of intelligent systems.

References

1. Honavar V. Toward, “Learning Systems That Use Multiple Strategies and Representations”. In: *Artificial Intelligence and Neural Networks: Steps Toward Principled Integration*. Honavar, V. and Uhr, L. (Ed.) New York: Academic Press, 1994. – Pp. 615-644.
2. Wermter S. *Hybrid Connectionist Natural Language Processing*. Chapman and Hall, International Thomson Computer Press, London, UK, January, 1995.
3. Gavrilov A.V. *Hybrid Intelligent Systems*. NSTU, Novosibirsk, 2003. – 162p. (In Russian).
4. Gavrilov A.V., Chistyakov N.A. *An architecture of the toolkit for development of Hybrid Expert Systems.* – Int. Conf. IASTED ACIT-2005, Novosibirsk, June, 2005.
5. N. Winer. *God and Golem*. 1966.
6. Gavrilov A.V. “The model of associative memory of intelligent system”. - / *The 6-th Russian-Korean International Symposium on Science and Technology. Proceedings. - Novosibirsk, 2002. - Vol. 1*.- Pp. 174-177.
7. Gavrilov A.V. “The model of associative thinking”. / *Proceedings of Conf.. CAI*-2002. - Moscow, 2002, V. 1. - Pp. 464-472. (In Russian).
8. Pribram K. H. *Languages of the brain*. – Prentice-hall inc., Englewood Cliffs, N.J., 1975.
9. Ackof R.L., Emery F.E. *On Purposeful Systems*, Chicago, New York, 1972.
10. Carpenter G., A., Grossberg S. *Pattern Recognition by Self-Organizing Neural Networks*, Cambridge, MA, MIT Press, 1991.
11. Giorgio Marchetti. A presentation of Attentional Semantics*. Cogn. Proc. 2006, v. 7*, pp. 163-194.
12. Lewin D.I. “Why is that computer laughing?” – *IEEE Intelligent Systems, 2001, vol. 16, 5*. - Pp. 79-81.
13. Gavrilov A.V. “About Knowledge Representation and Processing in Intelligent Systems”. – *Proc. of 4-th Int. Symp. “KORUS-2000”, V. 2, Ulsan*. - Pp. 84-87.
14. Wilson R.A. *Quantum Psychology*. - New Falcon Publications, USA, 1996.
15. Stuart Hameroff M.D., Roger Penrose. *Quantum Theory and Human Consciousness*. - http://www.u.arizona.edu/~hameroff/
16. Blatter G.B., Geshkenbein V. B., Fauchere A. L., Feigelman M. V., Ioffe L. B. “Quantum computing with superconducting phase qubits”. - *Physica C 352, 105, (2001).*