

UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

The Leverage of a Self Concept in Incremental Learning

Permalink

<https://escholarship.org/uc/item/02x2z1z6>

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 26(26)

ISSN

1069-7977

Authors

Samsonovich, Alexei V.
De Jong, Kenneth A.

Publication Date

2004

Peer reviewed

The Leverage of a Self Concept in Incremental Learning

Alexei V. Samsonovich (asamsono@gmu.edu)

Krasnow Institute for Advanced Study, George Mason University
4400 University Dr. MS 2A1, Fairfax, VA 22030 USA

Kenneth A. De Jong (kdejong@gmu.edu)

Department of Computer Science and Krasnow Institute for Advanced Study
George Mason University, 4400 University Dr., Fairfax, VA 22030 USA

Introduction

This work is an attempt to bring together three topics that belong to three different levels of science: (1) symbolic unsupervised learning, (2) the self of a cognitive system, and (3) a universal criterion for conscious experience. We study cases when a cognitive system develops new abilities by reinterpreting its own episodic memories, using the self-concept and the schema of a motivated voluntary action. Within our framework (Samsonovich & DeJong, 2003), the subject-self per se is not represented as a virtual entity in the cognitive system. Instead, the set of axioms that constitute a self-concept (Aleksander & Dunmall, 2003; Samsonovich & Nadel, in press) are implemented via dynamical rules and constraints. These principles are demonstrated in a model paradigm and have implications for the philosophy of mind.

Approach

The proposed approach is based on the general framework of schemas, charts and mental states described previously (Samsonovich & DeJong, 2003). The term "schema" was introduced by Kant (1781/1929). Here schemas are units of semantic knowledge, primitives of action, reasoning, sensation, etc. A schema has a header (specifying rules and conditions of binding and expected effects of execution) and a body (specifying how, if at all, the schema is executed).

Paradigm: Leveraging Self-Learning with the Self

In this paradigm, a set of specially designed virtual worlds is used as a "training facility" to help the virtual robot to develop useful and powerful schemas. Innate schemas may include elementary moves and senses, as well as relevant reasoning primitives. The robot "wakes up" in a first-level world and starts by repeating the following procedure:

1. Select an action schema and mutate its header to produce an idea of an action that is not straightforward.
2. Take the new header as a challenge and solve it in each of several encountered situations (execute the solutions).
3. Reinterpret own behavior: find an apparent common motivation in the performed intermediate steps in all cases.
4. Based on the above, write the body of the new schema and add the schema to semantic memory.

As the robot learns essentials at the first level, it is taken to the next level, and so on. At each new stage, previously developed schemas are used for solving new challenges.

Demonstration by Example

The scheme outlined above will be demonstrated in the poster by computer simulations based on a push-push puzzle setup. A minimal set of innate schemas includes a one-step move and some useful cognitive primitives, e.g., the notion of Euclidean distance. At the first stage the robot learns to move in an open space. Then it learns to navigate a maze, to push blocks, to avoid irreversible moves, etc. After that, when given a goal, it is capable of solving simple puzzle configurations and learns to deal with more complex ones.

Philosophical Implications

The above analysis has interesting implications for the philosophy of mind. Some philosophers believe that the phenomenon of conscious experience will always remain a mystery, while others maintain that this mystery is illusory. How can one decide, when and whether this phenomenon should occur? Chalmers (1994) answers with the Principle of Organizational Invariance. His answer sounds like this: *there is an abstract mathematical model M of a functional organization of a cognitive system, such that, whenever M can be mapped onto a given physical object, that object is conscious.* Therefore, a criterion for consciousness can be given in terms of M . We define it as follows: M must instantiate the self as a noumenon (a notion introduced by Kant, 1781/1929), which we understand as an imaginary thing that seems to determine system's dynamics, and yet it cannot be explicitly represented in the system due to its fundamental properties. The proposed framework in which the self is implemented via a set of self-axioms (constraints) about system's own dynamics conforms to this concept.

References

- Aleksander, I., & Dunmall, B. (2003). Necessary first-person axioms of neuroconsciousness. *Lecture Notes in Computer Science*, 2686: 630-637.
- Chalmers, D. J. (1994). On implementing a computation. *Minds and Machines*, 4, 391-402.
- Kant, I. (1781/1929) *Critique of pure reason*. Translated by N. K. Smith. New York: St. Martin's Press.
- Samsonovich, A. V., & De Jong, K. A. (2003). Meta-cognitive architecture for team agents. *Proceedings of the 25th Annual Meeting of the Cognitive Science Society* (pp. 1029-1034). Boston, MA: Cognitive Science Society.