Biologically Inspired Cognitive Architectures xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

Biologically Inspired Cognitive Architectures

journal homepage: www.elsevier.com/locate/bica

Research article

The parasitic manipulation of an animat's behavior

Valery Karpov *

Department of Neurocognitive Science and Intelligent Systems, Laboratory for Robotics, National Research Centre "Kurchatov Institute", Moscow, Russia
Moscow Institute of Physics and Technology, Russia

A R T I C L E I N F O

Article history:
Received 8 March 2017
Revised 8 May 2017
Accepted 8 May 2017
Available online xxxx

Keywords:
Robot
Temperament
Emotions
Parasitic manipulations

A B S T R A C T

This paper uses an analogy to parasitic zombiing in nature to show how a robot's goal-setting and behavior can be changed. The implementation of these parasite-like mechanisms is based on the emotional and temperamental architecture of the robot's control system. A number of possible mechanisms used by parasites to manipulate hosts' behavior have been considered, including changes to the host's behavioral characteristics, the direct re-ordering of behavioral responses, and the reorientation of hosts' reactions. It was determined that each of these mechanisms can be employed using a relatively primitive and restricted set of influences or parasite control signals. This approach led us to propose several variants of a high-level control system architecture. The main assumption was that the center of mass of a robot's control system lies on the level of basic behavioral procedures while high-level control can be realized using parasite-like methods.

Introduction

The creation of biologically inspired architectures, models, and methods is a rapidly developing area of robotics, which can be considered from two aspects: The first uses this inspiration in an attempt to create technical objects, and models, while the second creates commonality tests of developed models and methods and applies them to a wide range of tasks in a process of verification. This verification aspect of bio-inspiration is considered in this study.

This work continues research on the realization of temperament and emotions mechanisms in robots' control systems. We mean by it not the architecture of so-called artificial emotional intelligence described by Samsonovich (2012), for example. In this study, we are interesting the internal mechanisms defining behavior of the robot. Simonov's information theory of emotions forms the basis of the emotional component of robot architecture (Simonov, 1991). The realization of emotional component has evolved into the creation of temperament: a psychological level of robot behavior control. The study of the behavior of robots with an emotional–temperamental component has led to some conclusions, including the implementation of behavioral phenomena inspired by zombie parasites. The use of the word *zombie* underlines that the infected and manipulated individual represents a parasite's genome expressing parasite's behavior through the body of a host.

The main purpose of this work is to consider some possible mechanisms for control or manipulating the complex behavior of a technical device. At the same time it is assumed that this manipulation is carried out by some simple device, which is an analogue of a primitive parasite, controlling the behavior of the host's organism.

Research on the phenomena of parasitic zombiing was based on simulation modeling, while architectures of emotional and temperamental control systems were realized on real robotic devices. The architecture that describes the emotional and temperamental organization of a robot control system will be discussed further in this paper, as will the interpretation of the parasitic behavior in terms of its impact on the control system's parameters. The results of computational experiments will also be shown.

It is necessary to define some of the terminology that will be used in this study. The terms *robot* and *agent* will be used, in part, interchangeably. The former will be used where it is necessary to emphasize the hardware or technical aspects of described object, while the latter will be used more abstractly, such as in the discussion of model behavior. The term *animat*, referring to an artificial animal model, will also be used.

An agent is usually understood to be an abstract program essence. Maes defined autonomous agents as computational systems that inhabit complex dynamic environments, sense and act autonomously within these environments, and, by doing so, realize the goals or tasks for which they were designed (Maes, 1995). Con-

* Work address: National Research Centre "Kurchatov Institute", 1, Ac. Kurchatov Sq., Moscow 123182, Russia.
E-mail addresses: karpov.ve@gmail.com, karpov_ve@mail.ru

http://dx.doi.org/10.1016/j.bica.2017.05.002
2212-683X/© 2017 Elsevier B.V. All rights reserved.

versely, Russell and Norvig determined that an agent is anything that can be viewed as perceiving its environment through sensors and acting upon that environment through effectors (Russell & Norvig, 2010). At the same time, agents are divided into robotic agents and software agents. Nevertheless, a robot is mostly considered to be a hardware system.

It is therefore expedient to use the term *animat* to emphasize the biological inspiration of the studied principles and mechanisms. Wilson (1986) determined four basic characteristics of simple animals, but we will not stop on them. The main idea is that, according to Wilson (1987), "we also apply the term animat informally to the animals and autonomous robots themselves." It should be noted, that many years before Wilson, Bongard et al. made one of the first declarations on the creation of certain virtual essence modeling for the organization of behavior (Bongard, Losev, & Smirnov, 1975). They described an attempt to construct a model that reproduced, in rough lines, human behavior or, at least, that of "reasonable" animals. Importantly, the Animal Project also leaned on the problems of recognition and learning described by Bongard (1970).

The structure of the work is following. First, we will discuss the organization of the architecture of animat, which has signs of emotional and temperamental behavior. Next, we will consider some mechanisms of the host-parasite interaction which are observed in living nature, and possible interpretations of these mechanisms from the point of view of the technical system control.

The architecture of an emotional animat

The architecture of a robot control system that uses higher-level emotional mechanisms forms the basis for this study ((V.E. Karpov, 2014) and (V. Karpov, 2014)). This architecture is based on Simonov's information theory of emotions (Simonov, 1991), which makes the basic assumption that the strength and sign of an emotion are determined by an assessment of current needs and the possibility of their satisfaction. Qualitatively, this can be described by the following expression:

$$E = f(N, p(I_{need}, I_{has}))$$

where E represents an emotion (and its strength, quality, and sign), N represents the force and quality of an actual need, p represents the possibility of the need's satisfaction on the basis of the congenital mechanisms and the gained experience, I_{need} represents information on the available resources, necessary for need's satisfaction, and I_{has} represents information on the subject's existing means, resources, and time.

The input values for the control system are a set of current needs N and a set of sensor values S. At any moment, the robot can only perform a single action from some set P. It is assumed that the action is a kind of behavioral program: for example, sleep, random straying, running away from danger, or searching for food.

A set of behavior rules is presented in MYCIN-like form:

$$p_n: \ Cond_1 \ \& \ldots \& \ Cond_i \ \rightarrow A\ (a_n) \tag{1}$$

For instance, the rule "eat food" could be presented as follows:

$$IF\ "Need\ food"(N_{food})\ \&\ "Touch\ of\ food"(S_{food})\ THEN\ "Eat"(a_{eat})$$

where N_{food}, S_{food}, a_{eat} are certainty factors.

Further, we will give a qualitative mathematical model, the goal of which is to give a general presentation of the system components and the relations between them and to obtain general structural diagram of the control system.

Thus, the control system operates a variety of products using rules that define the robot's performance of a particular action P (1) depending on the existing set of current needs (N) and sensor values (S):

$$P = R(N, S) \tag{2}$$

Another component, which will determine the significance of this rule, will now be brought into the equation:

$$P = R(N, S, G) \tag{3}$$

The value G represents the influence of the agent's emotional state on the action performed. It is assumed that G depends on the agent's emotional state (E), as well as the values of current needs and sensors:

$$G = V(E, N, S) \tag{4}$$

The magnitude and sign of E is defined as the difference between the actions that a robot could perform given the current situation (based on the existing state of the current needs and sensor values) and the actual output vector Y:

$$E = k_{em}(Y - P) \tag{5}$$

Here, k_{em} is a certain coefficient. This expression (5) defines the character of emotions as a vector: Every action is related to its own private emotional value. At the conceptual level, this may be interpreted as follows: The system determines the weight of actions that could be accomplished at any given time based on current needs and existing sensor values. There are a lot of required (desirable) activities. However, as previously mentioned, the agent can only perform one of the many possible actions. The current action is associated with a positive emotion when it matches the desired action, while a negative emotion occurs where there is a discrepancy.

Vector Y identifies the only action a performed at the time and is calculated as follows:

$$Y_i = \begin{cases} 1, & i = a \\ 0, & else \end{cases} \tag{6}$$

Here, a is the number of actual actions performed by the robot:

$$a : Z_a = \max Z \tag{7}$$

$$Z(t) = P - k_{fb} \sum Z(t-1) \tag{8}$$

In (7) the intermediate value Z is used to calculate a. The mean of Z is to describe the need for a stabilization of the output variable Y.

The k_{ext} parameter can be added to expression (3). It determines the weight of the rules' significance index G, providing the final expression for P:

$$P = R(k_{ext}, N, S, G) \tag{9}$$

Fig. 1 shows an animat's control system in functional and formal forms. Models (3)–(9) are presented in block diagram form in Fig. 1b. It shows that emotion E is an element of the feedback loop.

The FAPs block defines realization of behavioral activities and will be discussed below. Here we'll note that the FAPs block can be considered as a set of behavioral subroutines, which are launched by elements of the Action block (in a way, we can consider these elements to be analogous to command neurons). The role of the SN-elements (Fig. 1a) is to determine a single output reaction. In some sense, we might consider them Kohonen's network elements realizing the winner-takes-all principle. Links of SN elements are fragmentally shown in Fig. 1a. Output signals of SN elements comes to the braking entrances of all SN elements. The role of the Gates elements is to take into account the influence of emotional feedback. And finally, the dashed lines ("Communication between action and need") reflect the direct impact of the needs on the behavioral procedures activation. For example, the