An Autonomous Drive Balancing Strategy for the Design of Purpose in Open-ended Learning Robots

Extended Abstract

Alejandro Romero
GII, CITIC Research Center
Universidade da Coruña
Spain
alejandro.romero.montero@udc.es

Francisco Bellas
GII, CITIC Research Center
Universidade da Coruña
Spain
francisco.bellas@udc.es

Richard J. Duro
GII, CITIC Research Center
Universidade da Coruña
Spain
richard@udc.es

ABSTRACT

This paper is concerned with designing purpose in autonomous robots for open-ended learning settings. Unconstrained human robot interaction situations and robotic systems that must operate in dynamic multi-robot scenarios are paradigmatic examples of open-endedness. An approach to the appropriate design and engineering of motivational structures to endow robots with a particular purpose is proposed and tested. This approach focuses on the drive structure and how it can be made to autonomously adapt to changing circumstances. Specifically, a simple evolutionary strategy for the autonomous regulation of multiple drives in order to optimize long-term operation is defined. The experimental results have been obtained on a Baxter robot facing changing situations in real setups.

KEYWORDS

Open-ended learning; Autonomous robotics; Motivation; Human-robot interaction

ACM Reference Format:

Alejandro Romero, Francisco Bellas and Richard J. Duro. 2021. An Autonomous Drive Balancing Strategy for the Design of Purpose in Open-ended Learning Robots: Extended Abstract. In *Proc. of the 20th International* Conference on *Autonomous Agents and Multiagent Systems (AAMAS 2021), Online, May 3–7, 2021.* IFAAMAS, 3 pages.

1 INTRODUCTION

The main issue in open-ended learning robotics is about designing autonomous robots able to operate in open-ended learning situations carrying out tasks that are useful for its own development and, more importantly, that serve a purpose to the human designer [1]. In other words, how can purpose be introduced in the robot cognitive structure without knowing beforehand the domains it will find itself in and, thus, the domain related goals it must achieve.

Addressing this question puts us squarely in the realm of motivation [2]-[4]. Most of the work in this field has concentrated on studying the effects of different motivational strategies [5][6], especially those related to intrinsic motivation [7]-[9] or how to construct mathematical representations that could be used to support them [10]-[12]. Very little work has been devoted to the design and engineering of motivational structures to endow autonomous robots with a particular purpose in open-ended settings.

In [13], we contemplated this problem by designing and testing, in a real setup, a motivational system based on drives and goals that are autonomously related in a dynamic graph. In that work, as a first approach, the balancing between drives was fixed by the designer. To improve the open-ended operation of the system, here we propose and test an autonomous strategy for such balancing.

2 DRIVES AND BALANCING STRATEGY

The motivational system proposed here is based on two different types of drives: *Operational drives* (${}_{op}D_j$) and *Cognitive drives* (${}_{cg}D_j$). Operational drives have to do with the purpose the designer wants to instill the robot with, and they can be classified into *survival* and *purpose-dependent* drives [13]. Cognitive drives, on the other hand, allow for more efficient cognitive operation and learning. Thus, they lead the robot towards efficiently collecting the information that is necessary to learn models of the domains it finds itself in. We can classify them into *exploration cognitive drives* ${}_{cg}D_{er}$ and *exploitation cognitive drives* ${}_{cg}D_{el}$ [13].

The set of drives that makes up the motivational system must be properly managed to allow the robot to fulfill its purpose in an open-ended setup. To do it, the strategy proposed here starts with a drive vector D that must be defined by the designer:

$$\boldsymbol{D} = \{c_1 \cdot {}_{op}D_1, c_2 \cdot {}_{op}D_2, c_3 \cdot {}_{op}D_3, c_4 \cdot {}_{op}D_4, \ldots, c_5 \cdot {}_{cg}D_{er}, c_6 \cdot {}_{cg}D_{el}\}$$

It is often unfeasible for the designer to establish the optimal coefficients of the drive vector, that is, which balance of drives leads to the optimal behavior in terms of global drive satisfaction. In addition, it is not always the best approach to maintain these coefficients fixed throughout the open-ended process, which is intrinsically dynamic. Consequently, a procedure to optimize the coefficients in real-time must be implemented. In order to explore these issues, we propose here, as a first approach, to apply a simple evolutionary strategy to address such optimization. Algorithm 1 shows the pseudocode of this approach.

Proc. of the 20th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2021), U. Endriss, A. Nowé, F. Dignum, A. Lomuscio (eds.), May 3–7, 2021, Online.

The values for the coefficients (c_i) proposed by the designer are used just as an initial balance of the relevance of the drives, but the evolutionary strategy will be responsible for adjusting them trying to minimize the global drive value (G_D).

Algorithm 1 Autonomous drive balancing strategy

e: current experiment
E: set of experiments to execute
p_e: probability of exploration versus exploitation
c: drive coefficients vector
c': candidate drive coefficients vector
d: current drive coefficient
G_c: global drive value associated to *c*
$G_{c'}$: global drive value associated to *c'*
p_e is initially set by the designer
c is initially set by the designer
G_c ← *EvaluatePerformance(c)*
for *e* ϵ *E* **do**
  exploration ← *ChooseBetweenExplorationExploitation*(p_e)
  if *exploration* is *True* **then**
    c' is initially set to *c*
    for *d* ϵ *c'* **do**
      c'[d] ← *ChooseNewCoefficientValue()*
    $G_{c'}$ ← *EvaluatePerformance(c')*
    if $G_{c'} > G_c$ **then**
      G_c is set to $G_{c'}$
      c is set to *c'*

3 ROBOT INTERACTION WITH HUMAN

To clarify how this autonomous drive balancing strategy can be applied in a real setup, we consider a Baxter robot that must operate in a human-robot interaction scenario (Figure 1). The main purpose the robot is endowed with, is to *assemble the maximum number of valid parts in environments where there are human supervisors that can modify or interfere with its behavior* [13].

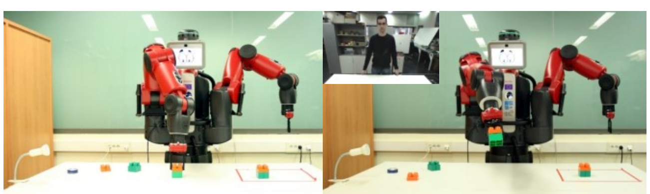

Figure 1. Experimental setup with the Baxter robot.

To achieve this purpose, we define 3 operational drives. First, an assembling task operational drive ($_{op}D_t$), associated to a sensor that detects whether the blocks are assembled and in the storage area (Figure 1 left). Secondly, an operational drive related to human satisfaction ($_{op}D_h$). The satisfaction of the drive will not be achieved until the supervisor makes an approval gesture (Figure 1 right). Finally, an energy drive ($_{op}D_e$) associated with the correct operation of the robot in the environment. The sensor associated with the drive is the ambient light sensor, since the robot perception system depends heavily on the use of cameras.

The experiment was carried out for 20000 time steps. The strategy we have proposed acts every 500 time steps, leading to 40 trials associated with them. To see how the autonomous drive balancing strategy responds to task changes, during the first 20 intervals the task consists only of assembling pieces. While in the next 20 intervals it consists in assembling parts and showing them to the human supervisor. The starting values that have been given to the coefficients are $c_{task} = c_{human} = c_{energy} = 0.5$.

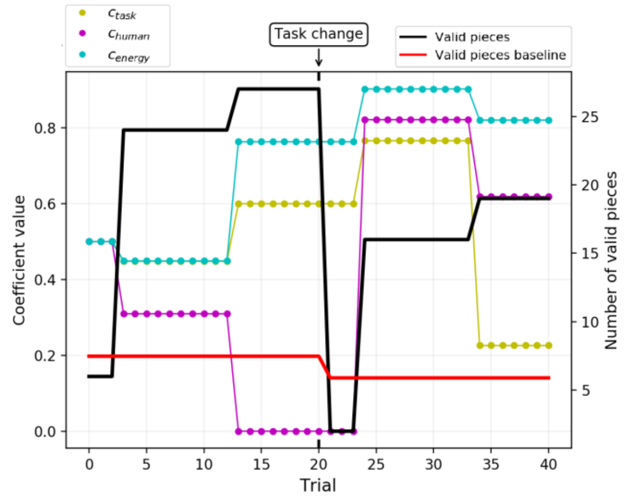

Figure 2. Evolution of the coefficient values and the number of valid pieces obtained after 40 trials of the experiment.

Figure 2 shows the evolution of the drive coefficient values and the number of valid pieces obtained (black line) for each coefficient combination in each execution. In addition, a solid red line shows the average number of valid pieces obtained if the coefficients remain static at their initial value (0.5), so that we can compare the autonomous balancing strategy with a non-adapting system.

Based on these results, it can be deduced that it is not the absolute value of the coefficients, but their relative order that will determine what the robot's correct operation will be in the environment. On the other hand, it has been possible to verify how the robot is able to autonomously adapt the relative strengths of its drives to task changes.

4 CONCLUSIONS

We have addressed one of the problems of engineering motivational structures for robots that must operate in open-ended learning settings. That is, how does one select and balance drives. We have focused on an on-line evolutionary mechanism for automatically balancing sets of drives so that the robots optimize their satisfaction over time and are able to adapt to changes in task definitions. The results of a series of experiments applying this mechanism on a robot, using drives that are either satisfied or not satisfied, have been very satisfactory.

ACKNOWLEDGMENTS

This by funded by, Ministerio de Ciencia, Innovación y Universidades of Spain/FEDER and Xunta de Galicia by grants RTI2018-101114-B-I00, ED431C 2017/12, CITIC research center grant ED431G 2019/01, and the FPU grant of Alejandro Romero.

REFERENCES

[1] S. Doncieux *et al.* 2018. Open-ended learning: a conceptual framework based on representational redescription. *Front. Neurorobot.*, vol. 12.

[2] G. Baldassarre and M. Mirolli. 2013. Intrinsically motivated learning in natural and artificial systems. 2013. *Intrinsically Motivated Learning Systems: an Overview*, Springer Berlin Heidelberg, 1–14.

[3] P. Oudeyer, A. Baranes, and F. Kaplan. 2013. Intrinsically Motivated Learning of Real-World Sensorimotor Skills with Developmental Constraints. *Intrinsically Motiv. Learn. Nat. Artif. Syst.*, 303--365.

[4] J. A. Starzyk. 2011. Motivated Learning for Computational Intelligence. *Computational Modeling and Simulation of Intellect: Current State and Future Perspectives*, IGI Global, 265–292.

[5] S. P. Singh, R. L. Lewis, A. G. Barto, and J. Sorg. 2010. Intrinsically Motivated Reinforcement Learning: An Evolutionary Perspective. *IEEE Trans. Auton. Ment. Dev.*, vol. 2, no. 2, 70–82.

[6] P. Oudeyer, V. V Hafner, and A. Whyte. 2005. The Playground Experiment: Task-Independent Development of a Curious Robot. *Proc. AAAI Spring Symposium on Developmental Robotics*, 42–47.

[7] U. Nehmzow, Y. Gatsoulis, E. Kerr, J. Condell, N. Siddique, and T. M. Mcginnity. 2013. *Intrinsically Motivated Learning in Natural and Artificial Systems.* Springer Berlin Heidelberg.

[8] E. Ugur and J. Piater. 2016. Emergent structuring of interdependent affordance learning tasks using intrinsic motivation and empirical feature selection. *IEEE Trans. Cogn. Dev. Syst.*, vol. 9, no. 4, 328–340.

[9] V. G. Santucci, G. Baldassarre, and M. Mirolli. 2013. Which is the best intrinsic motivation signal for learning multiple skills? *Front. Neurorobot.*, vol. 7.

[10] A. Wigfield and J. S. Eccles. 2000. Expectancy-value theory of achievement motivation. *Contemp. Educ. Psychol.*, vol. 25, no. 1, 68–81.

[11] K. Merrick. 2017. Value systems for developmental cognitive robotics: A survey. *Cogn. Syst. Res.*, vol. 41, 38–55.

[12] A. Prieto, A. Romero, F. Bellas, R. Salgado, and R. J. Duro. 2019. Introducing Separable Utility Regions in a Motivational Engine for Cognitive Developmental Robotics. *Integr. Comput. Aided. Eng.*, vol. 26, no. 1, pp. 3–20.

[13] A.Romero, F. Bellas, J.A. Becerra, R.J. Duro. 2020. Motivation as a Tool for Designing Lifelong Learning Robots. *Integr. Comput. Aided. Eng* vol. 27, no. 4, pp. 353-372, 2020.