ITC 3/50

Information Technology
and Control
Vol. 50 / No. 3 / 2021
pp. 485-494
DOI 10.5755/j01.itc.50.3.29242

Novel Algorithm for Agent Navigation Based on Intrinsic Motivation Due to Boredom

Received 2021/06/04 Accepted after revision 2021/07/15

http://dx.doi.org/10.5755/j01.itc.50.3.29242

HOW TO CITE: Loyola, O., Kern, J., Urrea, C. (2021). Novel Algorithm for Agent Navigation Based on Intrinsic Motivation Due to Boredom. *Information Technology and Control*, 50(3), 485-494. https://doi.org/10.5755/j01.itc.50.3.29242

Novel Algorithm for Agent Navigation Based on Intrinsic Motivation Due to Boredom

Oscar Loyola, John Kern, Claudio Urrea

Electrical Engineering Department, Faculty of Engineering,
University of Santiago of Chile (USACH), Av Ecuador 3519, Estación Central, Santiago 9170124, Chile;
e-mails: oscar.loyola@usach.cl; john.kern@usach.cl; claudio.urrea@usach.cl

Corresponding author: oscar.loyola@usach.cl

We propose a novel algorithm for the navigation of agents based on reinforcement learning, using boredom as an element of intrinsic motivation. Improvements obtained with the inclusion of this element over classic strategies are shown through simulations. Boredom is modeled through a chaotic element that generates conditions for the creation of routes when the environment does not offer any reward, allowing prompting the robot to navigate. Our proposal seeks to avoid what classical algorithms suffer in scenarios without rewards, generating losses of time in the resolution. We demonstrate experimentally that by adding the element of boredom it is possible to generate routes in scenarios in which rewards do not exist, allowing the use of these strategies in real circumstances and facilitating the robot's navigation towards its objective. The most important contribution sustained by this work corresponds to the fact that it is possible to improve navigation in completely adverse scenarios for a navigation algorithm based on rewards.

KEYWORDS: Reinforcement Learning, Intrinsic Motivation, Robotics, Boredom, Chaos.

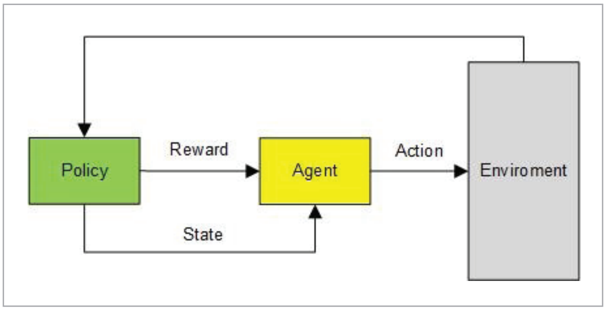
1. Introduction

Reinforcement learning (RL) is of the most common techniques in the field of machine learning [15, 29]. Its form of operation is based on how human beings learn, considering learning by conditioning one of its influences [7].

In general, reinforcement learning values the correct execution of actions and punishes the wrong decisions. However, the environment skews the options that the agent can take, therefore, it becomes an element that seeks to maximize an internal objective function, constantly learning from the problem through trial and error. A model of an agent based on reinforcement learning can be visualized in Figure 1.

Figure 1

Representation of a learning system based on reinforcement



There are multiple algorithms where extrinsic reward elements are considered to improve learning as indicated in [1, 25], however, a specific dependence of the entity is observed with respect to what the environment can offer it.

Currently the investigations developed by different authors integrate bio-inspired behaviors, such as the work [21] where a framework for the interaction of robots with humans is developed.

As robotic agent, what happens if the natural environment does not offer the rewards that the algorithm needs to function? In general, we can say that the entity could not operate creating havoc in the way of executing actions. On the contrary, human beings have the ability to determine their objectives considering their particular abilities [27], due to this the context in which the human finds himself does not determine how far he will be able to execute a certain task. This attribute can be linked to works developed on emotions as portrayed by [14, 26].

This capacity, which, based on the impulse to explore the environment spontaneously, is visualized in the works of [8] and widely discussed by [2], is known as intrinsic motivation (IM) and becomes an aspect to be considered to avoid failures in the algorithms that are only tested in simulation form.

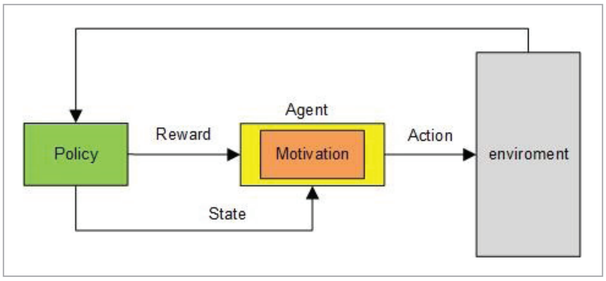
Learning by motivation is subject to learning by reinforcement. Considering the growing wave in which researchers have made important efforts to try to define emotions, which are basically rooted in the area of psychology, within the problems of computational learning, study models have been generated such as those exposed in [4, 11, 17] where indicators such as curiosity, novelty, pain, surprise, among others associated with motivation, are used. Other aspects where this type of bio-inspired algorithm is applied can be seen in [28], the proposal involves the development of an algorithm to emulate the cerebellum and ganglia interaction.

Psychologically it is accepted that learning is a process in which practical experience produces a change in behavior, therefore, there is an internal element that generates the expectations of this learning, we call this intrinsic motivation [19] and it is considered as a mechanism that encourages species to achieve objectives.

This mechanism can be seen associated with both internal and external factors Figure 2, this edge being an area of interest for research [22], since internal intentions or what really moves robots to fulfill a goal may not be the same for everyone, offering the possibility of establishing differences between a set of homogeneous robots since their constitution, but being absolutely heterogeneous regarding their condition towards the objective.

Figure 2

Representation of an agent system based on motivation



This concept is widely treated in different theories as portrayed in [10, 13] that associate motivation with intangible elements such as expectation, incentive or boredom [30].

Other authors have developed experiments with animals where they have sought to measure reinforcement learning considering intrinsic motivation [16] to find behavioral models.

The nature of boredom and the positives effects on motivation represent a starting point for this work [6], modeling this condition through a chaotic element that generates conditions for the creation of routes when the environment does not offer any reward.

Our objective is to compare navigation strategies for robots, through the application of a proprietary RL algorithm based on intrinsic motivation driven by boredom.

2. Methodology

Being then the problem of positioning a robot in space and how it will reach an objective point, the problem is defined in one related to dynamic programming using the Bellman equation, [3] where it is possible to define the algorithm as shown in Equation (1), where $V(s)$ it corresponds to the value of being in a state, $R(s, a)$ represents the reward function in a current state s and taking an action a, $V(s')$ represents the value of the new state s' if the action a is taken and γ is a discount factor that weighs the decisions that the entity will make allowing future decisions to be evaluated.

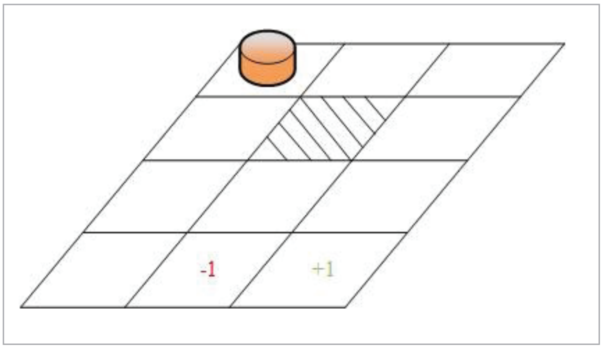
$$V(s) = \max_a \left(R(s,a) + \gamma V(s')\right) \tag{1}$$

Being then the case that the process of determining the direction of navigation will depend with total freedom on the entity, as specified in [18], it is possible to rewrite Equation (1) and express it as it is formalized in Equation (2), where the probabilities of all possible decisions are analyzed when the robot is at a certain point in space as represented in Figure 3.

$$V(s) = \max_a \left(R(s.a) + \gamma \sum_{s'} P(s,a,s') V(s)\right) \tag{2}$$

Considering the above, when the scenario does not offer alternatives that provide the algorithm with a reward for its execution of tasks, the entity begins to perform random actions, taking this as a basis, two classic RL algorithms are studied, this algorithm was previously compared without one intrinsic motivation [24] in mobile robot path planning.

Figure 3
Grid world visualized by the robot



2.1. Q-Learning

Learning algorithm that seeks to maximize the future reward through the exploration of all the possible solutions that could be had for a displacement, each iteration is stored in a table called Q table generating policies and displacement actions.

The model can be visualized as expressed in Equation (3), where $Q(s, a)$ represents a state-action set whatever where the robot is, ξ represents the learning coefficient, γ is a discount factor to weight the behavior of taking a new state called s'.

$$Q(s,a) = Q(s,a) + \xi\left[r + \gamma \max_a Q(s',a)\right] \\ -Q(s,a) \tag{3}$$

2.2. SARSA

Policy-based learning algorithm Markov decision process (MDP), bases its operation on updating a Q table that depends on the state and action selected by the agent $Q(s, a)$, the reward r will be selected according to that action and the new state is executed s' which involves a change to the new action a'. The system model can be visualized in Equation 4.

$$Q(s,a) = Q(s,a) + \xi\left[r + \gamma Q(s',a') - Q(s,a)\right] \tag{4}$$

3. Algorithm Based on Boredom Motivation

Through what is indicated in the literature, structures were developed using novelty as an element to give an stimulus [12] or the implementation of a dynamic controllers based on curiosity and boredom are demonstrated in [23] these related works are based in Q.learning algorithms with a method of intrinsic motivation, another case where is used boredom and curiosity is [31]. These research demonstrate that boredom is a enabled to curiosity.

It is possible to establish a model based on the Q-learning SARSA structure and apply new variables in decision making powered only by boredom considering the work [5], where the author exposed the boredom how a state that can motivate one to pursue a new goal when the actual state feeling is unsatisfactory.

Considering that the SARSA algorithm has a better response in growing scenarios [9], a condition called boredom is applied.

This condition occurs in the worst case scenario for the RL algorithms, which occurs when the medium does not offer any reward, therefore the matrix $R(s, a) = 0$, this condition implies that none of the available options attract you to something.

Taking what is stated by some authors in the theory of self-determination, it is possible to define boredom as an instance where creativity has its origin and therefore it is possible to use it as a catalyst towards intrinsic motivation.

Therefore, the state of boredom can be described as a random element that will lead us to two possible conditions, a) maintaining the current dissatisfied condition or b) propelling ourselves to a state of creativity This duality is portrayed in the completeness of the scenario of possible rewards, and this is represented in Equation (5), where are assigned either in the half or in the whole set $R(s, a) = 0$. The value of 0.5 is defined as the cut-off threshold for the Bored variable considering the criterion of maximum variance defined as M.

Assuming that the environment where the algorithm will be applied is unknown in size, this criterion provides guarantees by granting the same occurrence possibility to the situations in which the universe will be completed.

$$M = \begin{cases} \frac{1}{2} R_{chaos}(s,a), if, Bored \leq 0.5 \\ R_{chaos}(s,a), if, Bored > 0.5 \end{cases} \tag{5}$$

To avoid that the values used are distributed in a normal way, a chaotic function is used based on the Chua oscillator model in its discrete form as seen in Equation (6), where the term $f(x_{tk-1}))$ it is developed in the Equation (7) as exposed by [20]

$$\begin{aligned} & x(t_k) = \left(a\left(y(t_{k-1}) - x(t_{k-1}) - f\left(x(t_{k-1})\right)\right)\right) h^{q1} \\ & -\sum_{j=v}^{k} c_j^{(q1)} x(t_{k-j}) \\ & y(t_k) = \left(x(t_k) - y(t_{k-1}) + z(t_{k-1})\right) h^{q2} \\ & -\sum_{j=v}^{k} c_j^{(q2)} y(t_{k-j}) \\ & z(t_k) = \left(-\beta y(t_k) - \gamma z(t_{k-1})\right) h^{q3} \\ & -\sum_{j=v}^{k} c_j^{(q3)} z(t_{k-j}) \end{aligned} \tag{6}$$

$$\begin{aligned} & f\left(x(t_{k-1})\right) = m_1 x(t_{k-1}) + \frac{1}{2}(m_0 - m_1) \\ & \left(\left|x(t_{k-1}) + 1\right| - \left|x(t_{k-1}) - 1\right|\right) \end{aligned} \tag{7}$$

In Algorithm 1 the proposal for the integration of boredom in the SARSA flow is displayed.

Algorithm 1: Agent SARSA-Chaotic	
Step 0	**If** $R(s, a)$
Step 1	**Function** Sarsa_Agent(perception) **return** an action
Step 2	**Else if** R(s, a) == 0
Step 3	Boredom = Random
Step 4	**If** (Boredom>0.5)
Step 5	**Function** Chua **return** $M[x, y, z]$
Step 6	$R(s, a)= M[z]$
Step 7	**Else if** (Boredom <= 0.5)
Step 8	**Function** Chua **return** $M[x, y, z]$
Step 9	$R(\frac{s}{2}, a) = M[z]$

Step 10	**End if**
Step 11	**Function** SARSA-Agent(perception) **return** an action
Step 12	**End if**

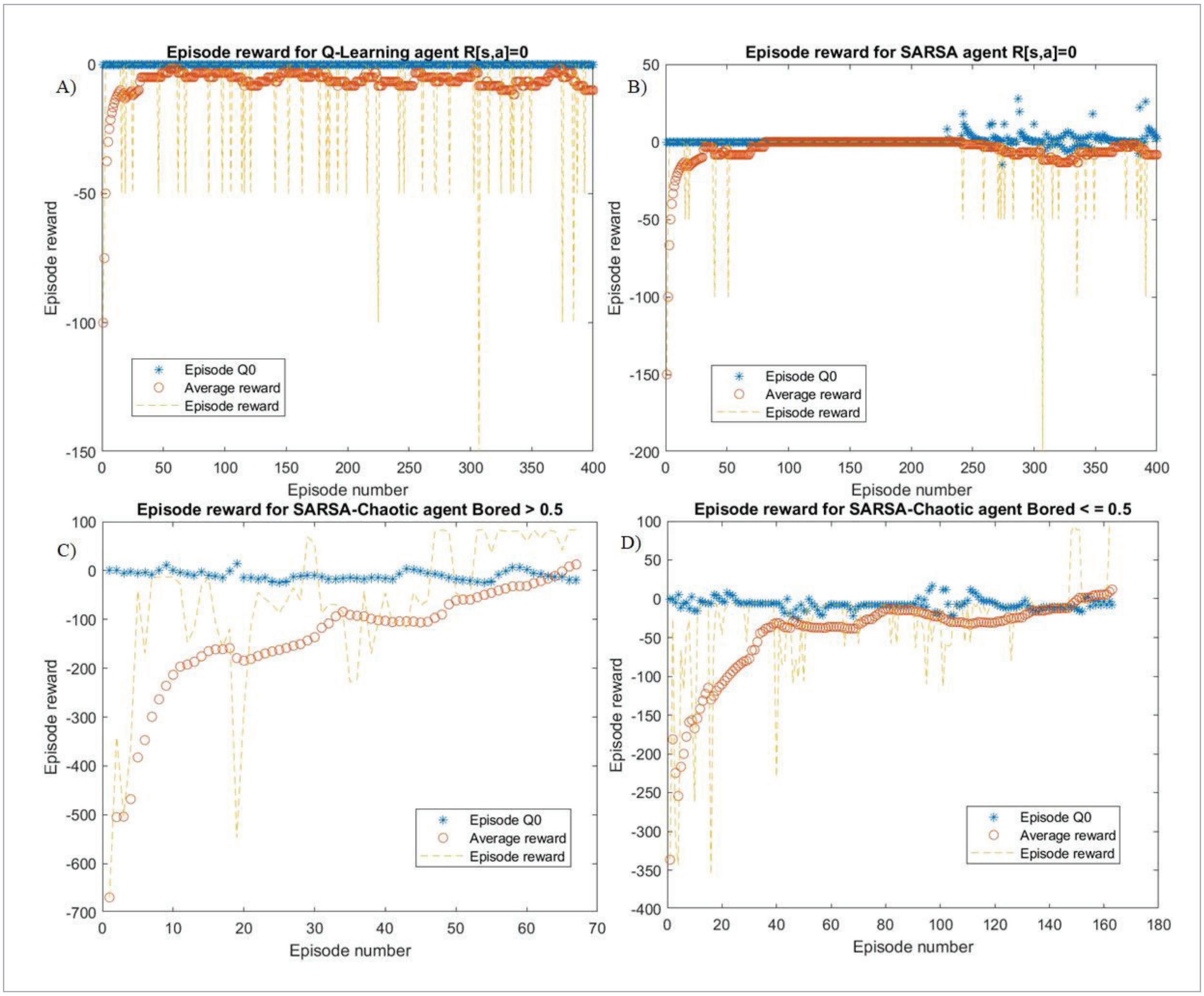
4. Results

Considering the training of 2 agents under normal operating conditions in a known world of size 7x7, it is possible to observe a slight superiority of the Q-Learning algorithm with respect to SARSA in the time of convergence towards a solution, however, this training process is performed under normal conditions with a specific reward.

When the universe does not deliver any reward, it is possible to observe how the algorithm tries to converge on some viable result, but they remain at 0 Figure 4 A) and B)), contrary to the proposed algorithm, since in any of the conditions that arise it generates training patterns.

Figure 4

Comparisons in the learning process when the reward is 0. A) Show how the Q-Learning algorithm falls to a minimum and is maintained until the time limit. B) Shows the behavior of the SARSA algorithm up to the time limit. C) It shows the behavior of SARSA with the chaotic component when the boredom indicator is greater than 0.5. D) It shows the behavior of SARSA with the chaotic component when the boredom indicator is less than 0.5



In C) the agent converges in a route at 70 iterations that are carried out with the entire reward matrix with values obtained from the Chua function, on the other hand in D) it is visualized that the system takes longer to generate a route, However, here only 50% of the matrix has rewards that are enough to take the learning system out of inertia and generate a route to the destination.

The routes traced in both cases are completely different, as can be seen in Figure 5 and Figure 6. This has effects on the way in which the agent faces the journey in the world, since the training and knowledge acquired in this process is vital. importance for behavior in the environment, two cases can be considered good, because the agent can arrive at the proposed destination and the reward is different than 0.

Figure 5

Agent behaviour in training world when boredom is greater 0.5

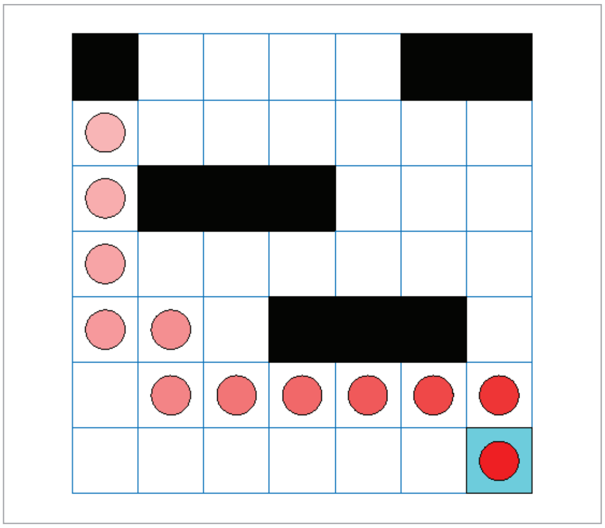
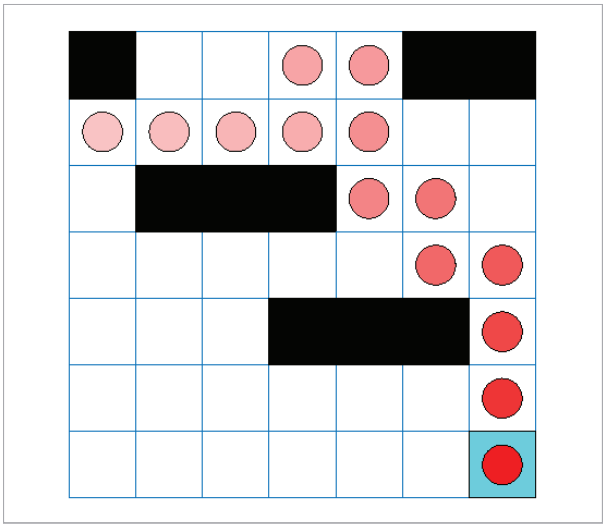


Figure 6

Agent behaviour in training world when boredom is less than 0.5



Taking this consideration, it is possible to generate a comparison between the different tests developed between the analyzed algorithms Table 1, where when focusing on points that the routes do not have within the training pattern, navigation failures are observed, as can be seen from what happened with Q-Learning and Boredom <0.5 when we refer to the point (6, 2).

Table 1

Behavior in navigation before target point (6, 2) outside the training routes

Algorithm	$R(s, a)$	Accumulate reward	Achievement
Q-Learning	0	0	No
Q-Learning	-5	0	No
SARSA	0	80	Yes
SARSA	-5	100	Yes
SARSA-Chaos Bored<0.5	0	-2680	No
SARSA-Chaos Bored<0.5	-5	-200	No
SARSA-Chaos Bored>0.5	0	93.98	Yes
SARSA-Chaos Bored>0.5	-5	89	Yes

The absolute failure of most of the algorithms is visualized in Table 2, where the Boredom <0.5 algorithm was the only one to navigate to that point.

The map used to carry out the tests had the same size, however, the non-displacement zones were modified, as can be seen in Figure 7, which shows the agent reaching the most complicated position for all the rest of the elements.

The most eloquent results on the effectiveness of the navigation method using boredom intrinsic element of motivation are observed when viewing the Q tables of each of the proposed methods.

Table 2

Behavior in navigation before target point (3, 5) outside the training routes

Algorithm	$R(s,a)$	Accumulate reward	Achievement
Q-Learning	0	0	No
Q-Learning	-5	-2500	No
SARSA	0	0	No
SARSA	-5	-2500	No
SARSA-Chaos Bored<0.5	0	-50.04	Yes
SARSA-Chaos Bored<0.5	-5	-65	Yes
SARSA-Chaos Bored>0.5	0	-2500	No
SARSA-Chaos Bored>0.5	-5	-8.65	No

Figure 7

Grid world used to perform algorithm tests

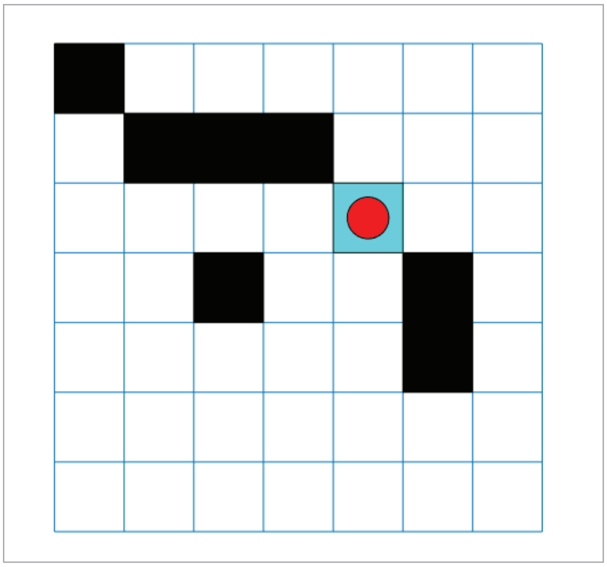


Figure 8

Behavior of possible alternatives according to the agent's decision to move when boredom is less than to 0.5

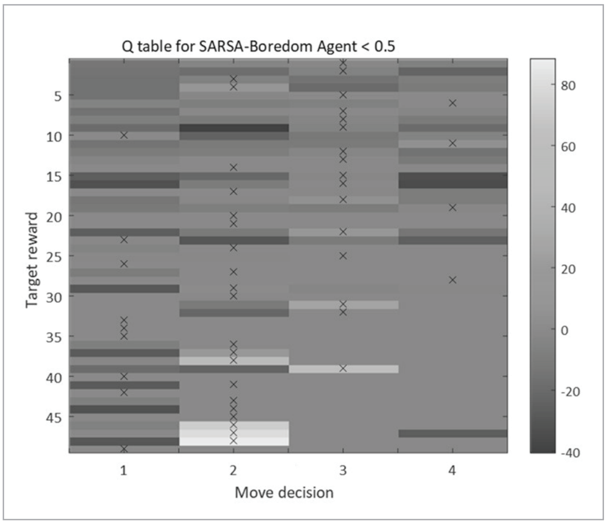
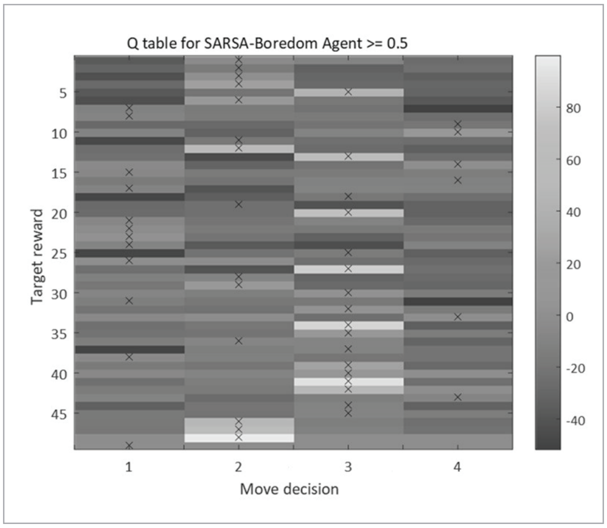


Figure 9

Behavior of possible alternatives according to the agent's decision to move when boredom is greather than or equal to 0.5



The Q tables show the relations between movements alternatives (forward, right, left, back) and target reward obtained for that decision. Figure 8 and Figure 9 portray the situation where even though the environment does not offer any rewards, the system generator allows navigation, giving the agent different alternatives to do a movement and show rewards different to cero.

Both responses differ in convergence times and the decisions that the robot executes. The latter can demonstrate how boredom influences the decisions, allowing the construction of two different solutions in the path planning.

In contrast to its peers such as Q-learning and SARSA that do not provide the robot with options to perform any movement portrayed in Figure 10 as the agent try to move selecting the movement 1 (forward) but in all cases the reward obtained is near to cero, that implies the agent in all cases is not going to nothing in the Figure 11, the case is different because the algorithm intends to give some possibilities of movement and the agent moves, but does not arrive at the destiny.

Figure 10

Behavior of the Q table in Q-learning algorithm when exposed to an enviroment of 0 values

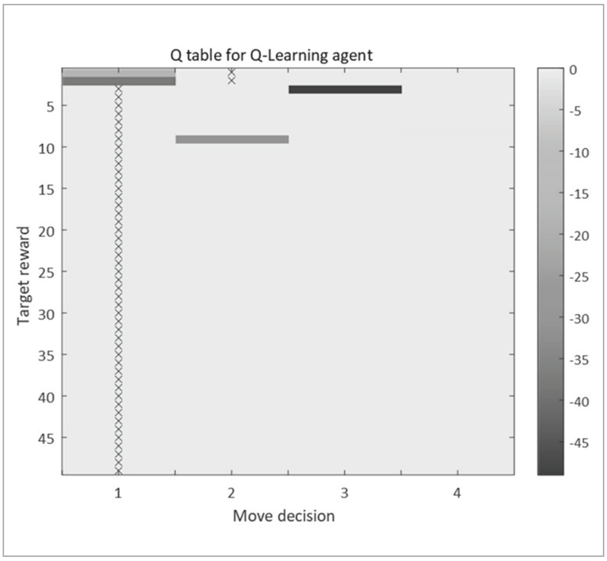
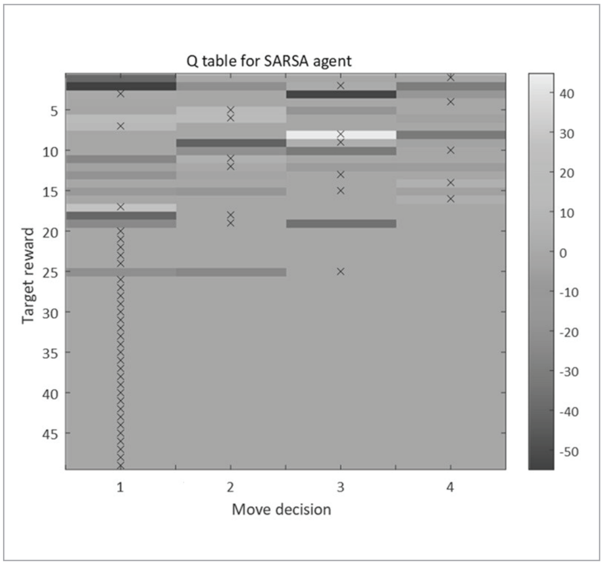


Figure 11

Behavior of the Q table in Q-learning algorithm when exposed to an enviroment of 0 values



5. Conclusion

From the different tests carried out, it is possible to deduce that the inclusion of the algorithm boredom motivation as a generator of intrinsic rewards proposes an improvement in the agent training process because under reward conditions 0, the system uses values that get around this problem providing the possibility of an algorithm where boredom powered by a chaos number generator is the main element to catalyze motivation.

The proposal generates a possible solution to navigation in aggressive environments for the algorithm, especially when environmental conditions offer no reward for travel, this can be used in path planning or in the training process.

Considering that the navigation system generates alternative routes in the training process, it is proposed to develop a mixture of both options as the algorithm have boredom greater than 0.5 and less 0.5 that allows, therefore, to know more travel options, this according to the data generated would allow that despite not finding the solution in one of the reward tables, you can jump to another that does contain it.

For future works, the application of these algorithms in a set of robots is proposed so that the navigation information is shared.

Acknowledgement

This research is sponsored by the Chilean government through its ANID program for advanced human capital formation.

References

1. Aubret, A., Matignon, L., Hassas, S. A Survey on Intrinsic Motivation in Reinforcement Learning. arXiv preprint arXiv:1908.06976,2019.
2. Barto, A. G. Intrinsic Motivation and Reinforcement Learning. Springer, 2013. https://doi.org/10.1007/978-3-642-32375-1_2
3. Bellman, R. The Theory of Dynamic Programming, 1954. https://doi.org/10.2307/1909830
4. Dörner, D., Güss, C. D. PSI: A Computational Architecture of Cognition, Motivation, and Emotion. Review of General Psychology, 2013, 17, 297-317. https://doi.org/10.1037/a0032947
5. Elpidorou, A. The Good of Boredom. Philosophical Psychology, 2018, 31(3), 323-351. https://doi.org/10.1080/09515089.2017.1346240
6. Elpidorou, A. The Significance of Boredom: A Sartrean Reading, 2015.
7. Gluck, M. A., Mercado, E., Myers, C. E. Learning and Memory: From Brain to Behavior, 2009.
8. Gopnik, A., Meltzoff, A. N., Kuhl, P. K. The Scientist in the Crib: Minds, Brains, and How Children Learn. William Morrow &amp; Co. 1999.
9. Harwin, L., Supriya, P. Comparison of SARSA Algorithm and Temporal Difference Learning Algorithm for Robotic Path Planning for Static Obstacles, Third International Conference on Inventive Systems and Control (ICISC), IEEE, 2019, 472-476. https://doi.org/10.1109/ICISC44355.2019.9036354
10. Heckhausen, J., Heckhausen, H. Motivation and Action. Springer, 2008. https://doi.org/10.1017/CBO9780511499821
11. Hester, T., Stone, P. Intrinsically Motivated Model Learning for Developing Curious Robots. Artificial Intelligence, 2017, 247, 170-186. https://doi.org/10.1016/j.artint.2015.05.002
12. Huang, X., Weng, J. Novelty and Reinforcement Learning in the Value System of Developmental Robots, 2002.
13. Hull, C. L. Principles of behavior. Appleton Century Crofts New York, 1943.
14. Klyubin, A. S., Polani, D., Nehaniv, C. L. Empowerment: A Universal Agent Centric Measure of Control. In 2005 IEEE Congress on Evolutionary Computation, 2005, 1, 128-135.
15. Laroche, R. Reliability in Reinforcement Learning, Retrieved from https://www.microsoft.com/en-us/research/blog/reliability-in-reinforcement-learning/, 2019.
16. Markou, A., Salamone, J. D., Bussey, T. J., Mar, A. C., Brunner, D., Gilmour, G., Balsam, P. Measuring Reinforcement Learning and Motivation Constructs in Experimental Animals: Relevance to the Negative Symptoms of Schizophrenia. Neuroscience and Biobehavioral Reviews, 2013, 37, 2149-2165. https://doi.org/10.1016/j.neubiorev.2013.08.007
17. Moerland, T. M., Broekens, J., Jonker, C. M. Emotion in Reinforcement Learning Agents and Robots: A Survey. Machine Learning, 2018, 107, 443-480. https://doi.org/10.1007/s10994-017-5666-0
18. Otterlo, M. V. Markov Decision Processes: Concepts and Algorithms. Course on Learning and Reasoning, 2009.
19. Oudeyer, P.-Y., Kaplan, F. What is Intrinsic Motivation? A Typology of Computational Approaches. Frontiers in Neurorobotics, 2009, 1, 6. https://doi.org/10.3389/neuro.12.006.2007
20. Petráš, I. Fractional Order Nonlinear Systems: Modeling, Analysis and Simulation. Springer Science and Business Media, 2011. https://doi.org/10.1007/978-3-642-18101-6
21. Samsonovich, A. V. Socially Emotional Brain Inspired Cognitive Architecture Framework for Artificial Intelligence. Cognitive Systems Research, 2020, 60, 57-76. https://doi.org/10.1016/j.cogsys.2019.12.002
22. Santucci, V., Baldassarre, G., Mirolli, M. Which is the Best Intrinsic Motivation Signal for Learning Multiple Skills? Frontiers in Neurorobotics, 2013, 7, 22. https://doi.org/10.3389/fnbot.2013.00022
23. Schmidhuber, J. Adaptive Confidence and Adaptive Curiosity. In Institut fur Informatik, Technische Universitat Munchen, Arcisstr, 21, 800 Munchen 2, 1991.
24. Sichkar, V. Reinforcement Learning Algorithms in Global Path Planning for Mobile Robot. International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM), IEEE, 2019, 1-5. https://doi.org/10.1109/ICIEAM.2019.8742915
25. Singh, S., Lewis, R. L., Barto, A. G., Sorg, J. Intrinsically Motivated Reinforcement Learning: An Evolutionary Perspective. IEEE Transactions on Autono-

mous Mental Development, 2010, 2, 70-82. https://doi.org/10.1109/TAMD.2010.2051031

26. Rosales, J. H., Rodríguez, L. F., Ramos, F. A General Theoretical Framework for the Design of Artificial Emotion Systems in Autonomous Agents. Cognitive Systems Research, 2019, 58, 324-341. https://doi.org/10.1016/j.cogsys.2019.08.003
27. Van Minkelen, P., Gruson, C., van Hees, P., Willems, M., de Wit, J., Aarts, R., Vogt, P. Using Self-Determination Theory in Social Robots to Increase Motivation in L2 Word Learning. In Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction, 2020, 369-377. https://doi.org/10.1145/3319502.3374828
28. Wang, D., Hu, Y., Ma, T. Mobile Robot Navigation with the Combination of Supervised Learning in Cerebellum and Reward Based Learning in Basal Ganglia. Cognitive Systems Research, 2020, 59, 1-14. https://doi.org/10.1016/j.cogsys.2019.09.006
29. Wason, R. Deep learning: Evolution and Expansion. Cognitive Systems Research, 2018, 59, 1-14.
30. Weissinger, E., Caldwell, L. L., Bandalos, D. L. Relation Between Intrinsic Motivation and Boredom in Leisure Time. Leisure Sciences, 1992, 14(4), 317-325. https://doi.org/10.1080/01490409209513177
31. Yu, Y., Chang, A., Kanai, R. Boredom-Driven Curious Learning by Homeo-Heterostatic Value Gradients. Frontiers in Neurorobotics, 2019, 12, 88. https://doi.org/10.3389/fnbot.2018.00088