Learning to Self-Reconfigure for Freeform Modular Robots via Altruism Multi-Agent Reinforcement Learning

Extended Abstract

Lei Wu
Northwestern Polytechnical University
Xi'an, China
leiwu@mail.nwpu.edu.cn

Bin Guo
Northwestern Polytechnical University and Peng Cheng Laboratory
Xi'an, China
guob@nwpu.edu.cn

Qiuyun Zhang
Northwestern Polytechnical University
Xi'an, China
qiuyunzhang@mail.nwpu.edu.cn

Zhuo Sun
Northwestern Polytechnical University
Xi'an, China
zsun@nwpu.edu.cn

Jieyi Zhang
Northwestern Polytechnical University
Xi'an, China
jyzhang@mail.nwpu.edu.cn

Zhiwen Yu
Northwestern Polytechnical University
Xi'an, China
zhiwenyu@nwpu.edu.cn

ABSTRACT

Modular robots can change between different configurations to adapt to complex and dynamic environments. Therefore, performing accurate and efficient changes to modular robot system, known as the self-reconfiguration problem, is essential. Existing reconfiguration algorithms are based on discrete motion primitives. However, freeform modular robots are connected without alignment and their motion space is continuous, making existing reconfiguration methods infeasible. In this work, we design a parallel distributed self-reconfiguration algorithm based on multi-agent reinforcement learning for freeform modular robots. We introduce a collaboration mechanism into the reinforcement learning to avoid conflicts in continuous action spaces. Simulations show that our algorithm reduces conflicts and improves effectiveness compared to the baselines.

KEYWORDS

Modular Robots; Self-reconfiguration; Reinforcement Learning; Altruism Scale

ACM Reference Format:

Lei Wu, Bin Guo, Qiuyun Zhang, Zhuo Sun, Jieyi Zhang, and Zhiwen Yu. 2023. Learning to Self-Reconfigure for Freeform Modular Robots via Altruism Multi-Agent Reinforcement Learning: Extended Abstract. In *Proc. of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023), London, United Kingdom, May 29 – June 2, 2023,* IFAAMAS, 3 pages.

1 INTRODUCTION

Through reconfiguration between different configurations, a multi-robot system composed of a number of modular robots can adapt to different environments and tasks. Therefore, modular robots have drawn extensive attention from many fields [10, 18].

However, the strong reliance on the connectors with specific locations is a common problem for most types of modular robots, which may result in task failures [1, 20]. Freeform modular robots, inspired by the diverse connectivity mechanisms found in living organisms, have continuous freeform connectors that do not need to be aligned and can be reconfigured more freely in a continuous configuration space [9, 20, 22, 27]. Freeform designs increase the efficiency of self-reconfiguration and reduces connection errors [20].

The reconfiguration problem is one of the most important and challenging problems for freeform modular robots. The diverse ways in which modular robots are combined lead to a huge configuration space [14], and the search for a global optimal planning between any two configurations is NP-Complete [5, 26]. In the freeform modular robot system, the connection and movement methods are infinite. Consequently, the kinematic constraints are too complex, making existing reconfiguration methods infeasible.

In this paper, we propose a multi-agent reinforcement learning based distributed reconfiguration algorithm for freeform modular robots. The main challenge is that the motion trajectories conflict during the decentralised reconfiguration process due to the kinematic constraints. To address this problem, we let modular robots learn to avoid conflicts autonomously. Since it is difficult to synchronise global configuration information in real time in modular robot systems [23], we use local information for coordination. In this case, the method of maximising joint rewards is not suitable [12]. While several methods have been proposed for mixed-motive problem [2, 6, 11, 19, 24], most of them are only applicable to discrete action space problems. Inspired by the altruism scale, we design a personalized collaboration mechanism in proximal policy optimization (PPO) to avoid conflicts. We introduce personalised altruism factors for all modular robots to accommodate the dynamic dependencies among modular robots and find the optimal altruism of each modular robot through meta-reinforcement learning [4, 21]. Simulations show that our method has better reconfiguration effectiveness by avoiding conflict.

2 APPROACH

The goal of the reconfiguration algorithm is to achieve rapid changes between configuration pairs. Any given configuration is uniquely

Proc. of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023), A. Ricci, W. Yeoh, N. Agmon, B. An (eds.), May 29 – June 2, 2023, London, United Kingdom.

determined by the position information P_i of each modular robot i and the topological connection information $\delta_{i,j}$ of all modular robots.

We adopt the method of multi-agent reinforcement learning to solve the reconfiguration problem. To comply with the general design principles of modular robots, we take the reconfiguration problem of modular robots as a Decentralized Partially Observable Markov Decision Process (Dec-POMDP).

2.1 Reward Function

For the defined reconfiguration problem, we design a reward function as,

$$R_{i,t} = c_p \times R_{i,t}^{top} + c_q \times R_{i,t}^{geo} - c_t \tag{1}$$

which is used by all modular robots. This reward encourages the modular robots to approach the target configuration in both the geometric and topological connection directions. The topological reward encourages modular robots to take actions to approach their local topological relationships in the target configuration. The geometric reward measures the effective distance that the modular robot moves in the correct direction to the target position per time step, thus making the overall configuration consistent with the target configuration.

2.2 Altruism Proximal Policy Optimization

We introduce an altruism mechanism in PPO to achieve cooperation among modular robots and avoid conflicts.

We introduce the sociological mechanism of the altruism scale [16] to measure the tendency of each modular robot to benefit others and form the altruism reward

$$R_{i,t}^{AS} = R_{i,t} + \alpha_i R_{i,t}^{MF} \tag{2}$$

where α_i is an altruism factor in the interval $(-1, 1)$ that measures each modular robot's attitude that benefits others. The mean-field reward $R_{i,t}^{MF}$ represents the average reward of neighboring modular robots [13, 15]. Note that the altruism factor we introduce is more suitable for reconfiguration which is a non-zero-sum game, than the ring metric of social value orientation [12, 15] introduced by previous work in non-strict zero-sum game problems.

The vector $\boldsymbol{\alpha}$ is the set of α_i, representing the distribution of altruistic tendencies across the population. And the altruism factor α_i is an personalised attribute of each module M_i. Heterogeneity and diversity can improve performance [12].

We train PPO by way of CTDE(centralized training and decentralized execution) [11] to maximize R^{AS}:

$$L^{AS}(\theta_i, \alpha_i) = \hat{E}_{i,t}[\min(k\hat{A}_{i,t}^{AS}, clip(k, 1-\varepsilon, 1+\varepsilon)\hat{A}_{i,t}^{AS}] \tag{3}$$

where the altruism advantage function $A_{i,t}^{AS} = A_{i,t} + \alpha_i A_{i,t}^{MF}$.

But it is impractical to design each α_i manually. Therefore, referring to the meta reinforcement learning [4, 15, 21], we take Equation 4 as the optimization goal, and perform another layer of training to optimize an appropriate personalised α_i for each modular robot. A suitable set of personalised $\boldsymbol{\alpha}$ can lead to higher overall reconfiguration performance.

$$L_i^G(\theta_i|\theta_1, \theta_2 \cdots) = \mathrm{E}[\sum_t \frac{\sum_j R_{j,t}}{N}] \tag{4}$$

3 LEARNING TO RECONFIGURE

Referring to the setting of [23], we construct two-dimensional configuration pairs with 12 modular robots in unity-ml [7], as the basic environment for simulation experiments.

We implement the above method based on the PPO of Rllib [8] and share the same policy in all modular robots through the parameter sharing [3], which guarantees the scalability of the method.

We compare multiple baselines that can handle continuous motion control, including PPO [17], MFPPO [25] and CoPO [15].

Evaluation metrics include mean congruence and conflict rate. The mean congruence is the average of the similarity of the configuration obtained from the reconfiguration method to the target configuration, which represents the effectiveness of the method. The conflict rate is the proportion of modular robots that cause conflicts to the total number of modular robots. Conflict detection is simplified for easy detection in Unity into collision detection, loop formation detection and break detection.

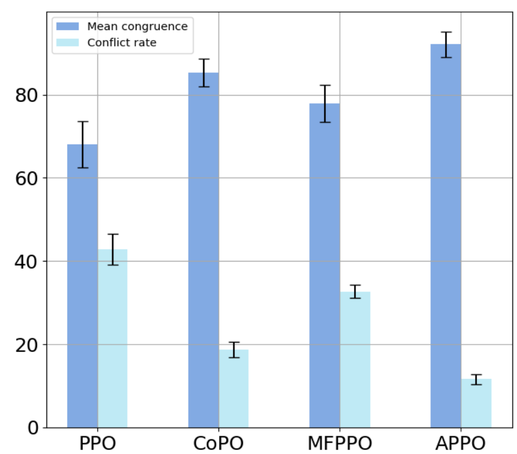

Figure 1: Comparison of different methods.

Figure 1 shows that our method has advantages in mean congruence and lower conflict rates compared with the baselines. The lower conflict rate guarantees the reconfiguration performance of our method. Due to the lack of a collaboration mechanism, both IPPO and MFPPO show poor performance, since conflicts between modular robots are more likely to occur during the reconfiguration of complex configurations. Compared with the CoPO algorithm with collaboration mechanism, the personalised altruism factors in our method enhance the diversity of modular robots and improve the effectiveness of reconfiguration. Collaboration mechanisms based on non-strict zero-sum games do not fit the nature of the reconfiguration problem.

4 CONCLUSION

In this paper, a multi-agent reinforcement learning algorithm based on altruism factors was designed to realize the continuous reconfiguration control of freeform modular robots. Our approach demonstrated the research prospects for the automatic design of reconfigurable motion controllers. Our study was limited to the two-dimensional reconfiguration problem.

ACKNOWLEDGMENTS

This work was partially supported by the National Science Fund for Distinguished Young Scholars (62025205), and the National Natural Science Foundation of China (No. 62032020, 62102322).

REFERENCES

[1] Hossein Ahmadzadeh and Ellips Masehian. 2015. Modular robotic systems: Methods and algorithms for abstraction, planning, control, and synchronization. *Artificial Intelligence* 223 (2015), 27–64.

[2] João Vitor Barbosa, Anna H Reali Costa, Francisco S Melo, Jaime S Sichman, and Francisco C Santos. 2020. Emergence of cooperation in n-person dilemmas through actor-critic reinforcement learning. In *Adaptive Learning Workshop (ALA), AAMAS*.

[3] Filippos Christianos, Georgios Papoudakis, Muhammad A Rahman, and Stefano V Albrecht. 2021. Scaling multi-agent reinforcement learning with selective parameter sharing. In *International Conference on Machine Learning*. PMLR, 1989–1998.

[4] Johannes Günther, Nadia M Ady, Alex Kearney, et al. 2020. Examining the use of temporal-difference incremental delta-bar-delta for real-world predictive knowledge architectures. *Frontiers in Robotics and AI* 7 (2020), 34.

[5] Feili Hou and Wei-Min Shen. 2014. Graph-based optimal reconfiguration planning for self-reconfigurable robots. *Robotics and Autonomous Systems* 62, 7 (2014), 1047–1059.

[6] Edward Hughes, Joel Z Leibo, Matthew Phillips, Karl Tuyls, Edgar Dueñez-Guzman, et al. 2018. Inequity aversion improves cooperation in intertemporal social dilemmas. *Advances in neural information processing systems* 31 (2018).

[7] Arthur Juliani, Vincent-Pierre Berges, Ervin Teng, Andrew Cohen, Jonathan Harper, Chris Elion, Chris Goy, Yuan Gao, Hunter Henry, Marwan Mattar, et al. 2018. Unity: A general platform for intelligent agents. *arXiv preprint arXiv:1809.02627* (2018).

[8] Eric Liang, Richard Liaw, Robert Nishihara, Philipp Moritz, Roy Fox, Ken Goldberg, et al. 2018. RLlib: Abstractions for distributed reinforcement learning. In *International Conference on Machine Learning*. PMLR, 3053–3062.

[9] Guanqi Liang, Haobo Luo, Ming Li, Huihuan Qian, and Tin Lun Lam. 2020. Freebot: A freeform modular self-reconfigurable robot with arbitrary connection point-design and implementation. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 6506–6513.

[10] Stefan B Liu and Matthias Althoff. 2020. Optimizing performance in automation through modular robots. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 4044–4050.

[11] Ryan Lowe, Yi I Wu, Aviv Tamar, Jean Harb, OpenAI Pieter Abbeel, and Igor Mordatch. 2017. Multi-agent actor-critic for mixed cooperative-competitive environments. *Advances in neural information processing systems* 30 (2017).

[12] Kevin R McKee, Ian Gemp, Brian McWilliams, Edgar A Duéñez-Guzmán, Edward Hughes, and Joel Z Leibo. 2020. Social diversity and social preferences in mixed-motive reinforcement learning. *arXiv preprint arXiv:2002.02325* (2020).

[13] Richard E Nisbett and Ziva Kunda. 1985. Perception of social distributions. *Journal of Personality and Social Psychology* 48, 2 (1985), 297.

[14] Michael Park, Sachin Chitta, Alex Teichman, and Mark Yim. 2008. Automatic configuration recognition methods in modular robots. *The International Journal of Robotics Research* 27, 3-4 (2008), 403–421.

[15] Zhenghao Peng, Quanyi Li, Ka Ming Hui, et al. 2021. Learning to simulate self-driven particles system with coordinated policy optimization. *Advances in Neural Information Processing Systems* 34 (2021), 10784–10797.

[16] Jack Sawyer. 1966. The altruism scale: A measure of co-operative, individualistic, and competitive interpersonal orientation. *Amer. J. Sociology* 71, 4 (1966), 407–416.

[17] John Schulman, Filip Wolski, Prafulla Dhariwal, et al. 2017. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347* (2017).

[18] Alexander Spröwitz, Rico Moeckel, Massimo Vespignani, Stéphane Bonardi, and Auke Jan Ijspeert. 2014. Roombots: A hardware perspective on 3D self-reconfiguration and locomotion with a homogeneous modular robot. *Robotics and Autonomous Systems* 62, 7 (2014), 1016–1033.

[19] Julian Stastny, Maxime Riché, Alexander Lyzhov, Johannes Treutlein, Allan Dafoe, and Jesse Clifton. 2021. Normative disagreement as a challenge for cooperative AI. *arXiv preprint arXiv:2111.13872* (2021).

[20] Petras Swissler and Michael Rubenstein. 2020. FireAnt3D: a 3D self-climbing robot towards non-latticed robotic self-assembly. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 3340–3347.

[21] Yunhao Tang, Tadashi Kozuno, Mark Rowland, et al. 2021. Unifying gradient estimators for meta-reinforcement learning via off-policy evaluation. *Advances in Neural Information Processing Systems* 34 (2021), 5303–5315.

[22] Yuxiao Tu, Guanqi Liang, and Tin Lun Lam. 2022. FreeSN: A Freeform Strut-node Structured Modular Self-reconfigurable Robot-Design and Implementation. In *2022 IEEE International Conference on Robotics and Automation (ICRA). IEEE.*

[23] Thadeu Knychala Tucci, Benoit Piranda, and Julien Bourgeois. 2018. A distributed self-assembly planning algorithm for modular robots. In *International Conference on Autonomous Agents and Multiagent Systems*.

[24] Jane X Wang, Edward Hughes, Chrisantha Fernando, Wojciech M Czarnecki, Edgar A Duéñez-Guzmán, and Joel Z Leibo. 2018. Evolving intrinsic motivations for altruistic behavior. *arXiv preprint arXiv:1811.05931* (2018).

[25] Yaodong Yang, Rui Luo, Minne Li, Ming Zhou, Weinan Zhang, and Jun Wang. 2018. Mean field multi-agent reinforcement learning. In *International conference on machine learning*. PMLR, 5571–5580.

[26] Zipeng Ye, Minjing Yu, and Yong-Jin Liu. 2019. NP-completeness of optimal planning problem for modular robots. *Autonomous Robots* 43, 8 (2019), 2261–2270.

[27] Da Zhao and Tin Lun Lam. 2022. SnailBot: a continuously dockable modular self-reconfigurable robot using rocker-bogie suspension. In *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 4261–4267.