Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14 Issue 05, May-2025

Reinforcement Learning Implementation on Bipedal Robot

Shivam Chavan
Industrial Engineering Department
Vishwakarma Institute of Technology
Pune, India

Abstract— **Reinforcement learning provides robotics with a framework and set of tools for creating complex and difficult-to-engineer behaviours. Robotic problems, on the other hand, give inspiration, impact, and validation for advancements in reinforcement learning. In this report contains all the major checkpoints of the flow of training a bi-pedal robot based on PPO, DDPG, A2C, and SAC algorithms. The entire simulation environment is in Gazebo and uses Ros-python as a backend to deploy an open-ai gym environment. We explain how reinforcement learning methodologies might be beneficially used by evaluating a basic situation in-depth, and we underline unresolved questions and the huge potential for future research throughout.**

Keywords— **Reinforcement learning, PPO, DDPG, Gazebo, open-ai gym.**

I. INTRODUCTION (HEADING 1)

When we consider the nature of learning, the first thought that typically comes to mind is that we learn by interacting with our surroundings. An infant has no explicit instructor when it plays, waves its arms, or looks around, but it does have a direct sensorimotor link to its surroundings. Exercising this link yields a wealth of information on cause and effect, the consequences of actions, and what to do to attain objectives. Reinforcement learning is just like this infant exploring the world around him and using the “hit and trial” method at first and then making wise predictions based on previous experiences. Sebastian Castro demonstrates an example of controlling humanoid robot locomotion using deep reinforcement learning, specifically the Deep Deterministic Policy Gradient (DDPG) algorithm. The robot is simulated using PyBullet while training the control policy using Reinforcement Learning Toolbox as mentioned in his book. Over the past three decades, the robotics research community around the world has shown considerable interest in the field of humanoid robotics [1], [2],[3]. One of the main reasons for this interest is that we humans tend to interact or relate more with human-like entities [3], [4]. Also, the domain of legged robots for traversing uneven, unstable terrains has intrigued several roboticists. Bipedal walking robots are one typical classification of humanoid robots that have garnered numerous research efforts in the past couple of decades. The legged-based locomotion of humanoid robots has a superior advantage over its conventional wheel-based counterparts, as it provides the possibility of either replacing or assisting humans in adverse environments [5]. Moreover, biologically inspired robots or modelled anthropomorphically render greater adaptability in diverse environments, especially ones requiring human intervention and needs [2]. The ease of overcoming random obstacles while travelling in complex dynamic environments has been advantageous for the bipedal robots compared to other legged robots like quadrupeds, etc [6]. From a biomechanics research point of view, understanding biped stability and walking mechanisms lays an important foundation for a better understanding of how humans traverse from one place to another [7]. Human locomotion, although simple as it appears to be, is a highly complex maneuver involving multiple degrees of freedom that are in turn coupled with complex non-linear dynamics generated due to various extensor and flexor muscle groups in the lower body. This served as one of the main motivations for proper understanding of physiology involved in human locomotion research and replicating the same on a BWR [7]. While the bipedal walking robots are known for their ease and flexibility in traversing over a wide range of terrains, stability is the main concern. BWRs pose exceptional challenges and concerns to control systems and designs mainly due to their nonlinearity and instability for which well-developed classical control architecture cannot be applied directly. The discrete phase change from a statistically stable double-stance position to the statistically unstable single-stance position in the dynamics of the BWR demand suitable control strategies [8]. Solving the stability issue of a bipedal walking system has aroused curiosity among many control scientists over the years [9], [10]. These conventional control theory approaches rely on complex deterministic and mathematical engineering models. Zero Moment Point (ZMP) is one of the conventional methods which is adopted as an indicator for dynamic stability in BWRs [11]. However, there are certain drawbacks associated with ZMP-based control methods that involve energy-inefficient walking, limited walking speed and poor resistance to external perturbations [12]. This method often relies on both a high level of mathematical computations and perfect knowledge of both the robot and environment parameters [13], [14]. Several machine learning practices have emerged over recent years that prove to have an edge over the conventional classical systems and control theory approaches to achieve stable bipedal walking. Reinforcement learning is a subdomain of machine learning, that could be applied as model-free learning of complex control systems [15]. Specifically, model-free learning of bipedal walking has mostly revolved around the implementation of several action policy learning algorithms based on the Markov Decision Process (MDP) [16], [17]. Several state-of-the-art reinforcement learning algorithms with MDP have produced significant results when used in visual simulations [18]. This has encouraged a growing number of computer scientists as well robotics researchers to use reinforcement learning (RL) ways to allow agents to perform powerful vehicle operations in complex and negative areas [19], [20]. Our contributions to this study:

- Raise the starting frame to tighten learning algorithms instead of the Gazebo simulator.
- Use the Deep Deterministic Policy Gradient-based on RL bipedal mobility and stability algorithm.
- Use the PPO algorithm on RL bipedal mobility.

PPO Background:
The Proximal Policy Optimization algorithm incorporates ideas from A2C (having more staff) and TRPO (using a trusted region for character development). The main idea is that after review, the new policy should not be too far away from the old policy. Thus, PPO uses cutting to avoid too large a review. Note that PPO contains a few changes from the original algorithm not written by OpenAI: the benefits are general, and the value function can also be cut.

II. LITERATURE REVIEW

There is a long history of using self-play in multiple agent settings. Preliminary work tested self-play using genetic algorithms. Sims (1994a) and Sims (1994b) have studied the complexities that emerge in morphology. and the behaviour of evolutionary creatures in a 3D simulation world. Open evolution was also observed in the Poly world (Yaeger, 1994) and Geb areas (Channon et al., 1998), where agents compete and marry in the world of 2D, and in Tierra (Ray, 1992) and Avida (Ofria & Wilke, 2004), when computer programs compete with computational resources. Tried the latest activity creates the necessary conditions for open evolution (Taylor, 2015; Soros & Stanley, 2014). Adaptation between agents and locations can also create an emerging complexity. In Foresterier et al. (2017); Xie et al. (2019) A real-world robot learns different solutions to jobs that require tools. In Bapst et al. (2019), the agent solves construction activities in a 2-D environment using both model-based and non-model methods. Allen et al. (2019) use a combination of man-made key elements and model-based policy development to solve a set of physics-based puzzles that require the use of the tool. However, in all of these activities, agents are clearly encouraged to work together and use the tools, and in our area, agents are clearly creating this motivation through. multi-agent competition.

III. MODELING AND SIMULATION

A. Gazebo

Gazebo is an open-source 3D robotics simulator. Includes ODE physics engine, OpenGL rendering, and support code for simulation and actuator control. A gazebo can use many very effective physics engines, such as ODE, Bullet, etc. (default is ODE). It offers a realistic rendering of environments that combine high-quality lighting, shadows, and performance. It can model "visual" sensors, such as laser range sensors, cameras (including wide-angle), Kinect-style sensors, etc. In 3D rendering, Gazebo uses the OGRE engine.

TABLE I. NOMENCLATURE OF THE LINKS OF BIPED

Link Name	**Terminology**
Waist	base
Right Thigh	right_thigh
Left Thigh	left_thigh
Right Shin	right_shin
Left Shin	left_shin

TABLE II. MASSES OF LINKS IN KG

Link Name	**Mass (kgs)**
Link name	0.36416
body	0.045155
thighs	0.069508
shins	0.05

TABLE III. MASSES OF LINKS IN KG

Parent Link	**Child Link**	**Joint type**
Base	right_thigh	Revolute
Base	left_thigh	Revolute
left_thigh	left_shin	Revolute
right_thigh	right_shin	Revolute

Bipedal Walking Robot is made in Gazebo, an open-source 3D robotics simulator capable of re-creating real environments for a variety of robotic applications based on robots [23]. 3D CAD biped walker model, designed with Blender was introduced into the Gazebo simulator file by file conversion from stl to urdf formats. Integrated Robot Description (URDF) is an extended Core language File format (XML) that is used to define links and integration to redesign and provide a robot of the Gazebo nature. The URDF file for the robot model contains the physical characteristics of each link such as property, size, length and the moment of inertia. And the place of origin (individually parent-child link compatible) and rotating axis of each link associated with the biped walker are described in the URDF file. Combined types and connecting positions are a few robotic links specified in this format. Different shared types of link links are drawn in Table. III. The link links are as follows: Ground is connected to a cylindrical stem with a constant member. The stump connected to a horizontal slide with a prismatic joint which is connected to the boom. The top of the waist is attached to a horizontal boom sliding back and forth and a bipedal person, limiting movement sagittal plane (i.e., near the Y-Z axis). The boom is empty weight compared to the biped walker's so it can be ignored. The boom detection was ignored, so the focus was on the bipedal interaction with nature. Apart from these links, two communication sensors, one below each shin is described in a URDF file. This was found ground contact times as you travel. Hips rotation, Hips speed, shin rotation, shin velocity, line speed in the sagittal plane and foot contact resulted in 12 state space and production space size 4. Robot Operating System (ROS) acts as a visual link between the control script and the Gazebo. Provinces were published like that in relevant articles and action instructions published back to control links. The communication level between the script and the Gazebo was 50 Hz.

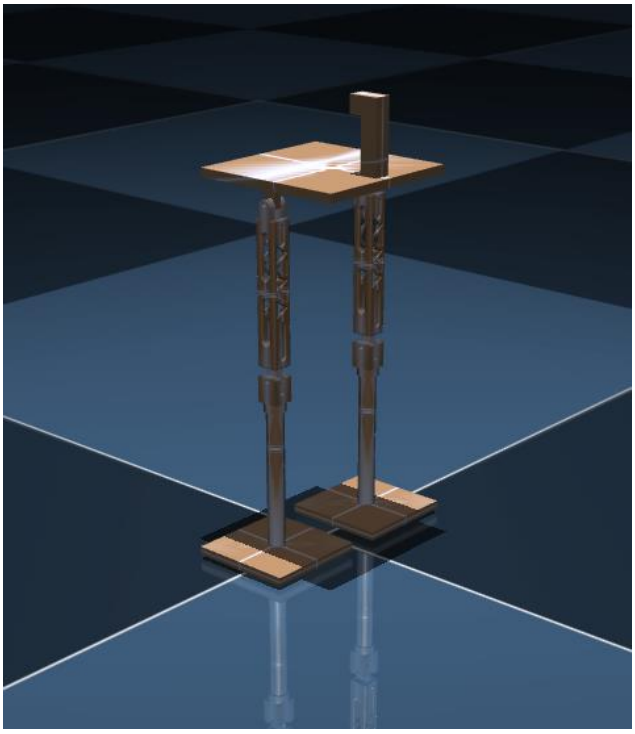

Fig. 1. Bipedal Walking Robot in Gazebo

IV. RESULTS

This section shows the effects of BWR simulation in order to achieve stable mobility. After bipedal training with the NVIDIA GeForce GTX 1660 Ti Graphical Processing Unit (GPU).

- After 20 million steps(n_frames), the robot started to struggle to keep its body position above 0.25 meters
- After 50 million steps(n_frames), The robot started to balance itself and started to stand.
- After 200 million steps, the robot started to take small steps and walk.

V. OBSERVATIONS

The implementation of reinforcement learning methods such as PPO, DDPG, A2C, and SAC on a biped walking robot with the Gazebo simulator and ROS backend revealed stepwise motor learning. Early training stages (up to 20 million steps) produced unstable behaviors where the robot could not keep its center of mass. At 50 million steps, balance began to emerge as the agent learned to stabilize itself. Approximately at 200 million steps, the robot showed the ability to take short, coordinated steps indicative of successful locomotion. The PPO algorithm showed promising results in policy stability and rate of convergence. State space (hip and shin angles, velocities, foot contact, and body height) captured sufficient dynamics relevant to learning control policies. GPU-based training facilitated feasible simulation of such high-dimensional control problems.

VI. CONCLUSION

Strengthening Reading can be used as an easy way to learn complex controls without prior knowledge. The power of the agent or the environment. It was so illustrated by the imitation of a planar bipedal walker in a Real-world Gazebo physics engine. In recent years, more effective reinforcement algorithms such as TD3 and Trust Regional Development TRPO are designed for complete control. They have turned out to be better than state of art due to the ease of better use and performance. Future work will include these algorithms to better learn to walk with bipedal. Efforts will be to reduce hardware usage and improve testing of the robot designed. Also, the current function has tested RL algorithms in planar bipedal walkers. In upcoming activities, two-dimensional obstacles will also be eliminated algorithms will be tested in a three-dimensional space.

VII. FUTURE SCOPE

The aim of the project is the implementation of a bipedal walker for last-mile delivery. On the Indian road, Terran a wheeled robot can do very small movements and will be restricted to certain areas itself for last-mile delivery.

REFERENCES

[1] Adams, B., Breazeal, C., Brooks, R.A. and Scassellati, B., 2000. Humanoid robots: A new kind of tool. IEEE Intelligent Systems and Their Applications, 15(4), pp.25-31.

[2] Swinson, M.L. and Bruemmer, D.J., 2000. Expanding frontiers of humanoid robotics [Guest Editor's Introduction]. IEEE Intelligent Systems and their Applications, 15(4), pp.12-17.

[3] Tanie, K., 2003, July. Humanoid robot and its application possibility. In Multisensor Fusion and Integration for Intelligent Systems, MFI2003. Proceedings of IEEE International Conference on (pp. 213-214). IEEE.

[4] Yamaoka, F., Kanda, T., Ishiguro, H. and Hagita, N., 2007, October. Interacting with a human or a humanoid robot?. In Intelligent Robots and Systems, 2007. IROS 2007. IEEE/RSJ International Conference on (pp. 2685-2691). IEEE.

[5] Huang, Q., Li, K. and Nakamura, Y., 2001. Humanoid walk control with feedforward dynamic pattern and feedback sensory reflection. In Computational Intelligence in Robotics and Automation, 2001. Proceedings 2001 IEEE International Symposium on (pp. 29-34). IEEE.

[6] Collins, S.H. and Ruina, A., 2005, April. A bipedal walking robot with an efficient and human-like gait. In Robotics and Automation, 2005. ICRA 2005. Proceedings of the 2005 IEEE International Conference on (pp. 1983-1988). IEEE.

[7] Raibert, M.H., 1986. Legged robots that balance. MIT press.

[8] Wang, S., Chaovalitwongse, W. and Babuska, R., 2012. Machine learning algorithms in bipedal robot control. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 42(5), pp.728-743.

[9] Ruiz-del-Solar, J., Palma-Amestoy, R., Marchant, R., Parra-Tsunekawa, I. and Zegers, P., 2009. Learning to fall: Designing low damage fall sequences for humanoid soccer robots. Robotics and Autonomous Systems, 57(8), pp.796-807.

[10] Fujiwara, K., Kanehiro, F., Kajita, S., Yokoi, K., Saito, H., Harada, K., Kaneko, K. and Hirukawa, H., 2003, October. The first human-size humanoid that can fall over safely and stand up again. In Intelligent Robots and Systems, 2003.(IROS 2003). Proceedings. 2003 IEEE/RSJ International Conference on (Vol. 2, pp. 1920-1926). IEEE.

[11] Vukobratovi, M. and Borovac, B., 2004. Zero-moment point thirty-five years of its life. International journal of humanoid robotics, 1(01), pp.157-173.

[12] Dallali, H., 2011. Modelling and dynamic stabilisation of a compliant humanoid robot, CoMan (Doctoral dissertation, The CICADA Project at the School of Mathematics, The University of Manchester).

[13] Kurazume, R., Tanaka, S., Yamashita, M., Hasegawa, T. and Yoneda, K., 2005, August. Straight-legged walking of a biped robot. In Intelligent Robots and Systems, 2005.(IROS 2005). 2005 IEEE/RSJ International Conference on (pp. 337-343). IEEE.

[14] Van der Noot, N. and Barrea, A., 2014, April. Zero-Moment Point on a bipedal robot under bio-inspired walking control. In Electrotechnical Conference (MELECON), 2014 17th IEEE Mediterranean (pp. 85-90). IEEE.

[15] Sutton, R.S. and Barto, A.G., 1998. Reinforcement learning: An introduction (Vol. 1, No. 1). Cambridge: MIT Press.

[16] Song, D.R., Yang, C., McGreavy, C. and Li, Z., 2017. Recurrent Network-based Deterministic Policy Gradient for Solving Bipedal Walking Challenge on Rugged Terrains. arXiv preprint arXiv:1710.02896.

[17] Peng, X.B., Berseth, G. and Van de Panne, M., 2015. Dynamic terrain traversal skills using reinforcement learning. ACM Transactions on Graphics (TOG), 34(4), p.80.

[18] Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D. and Riedmiller, M., 2013. Playing atari with deep reinforcement learning. arXiv preprint arXiv:1312.5602.

[19] Tedrake, R., Zhang, T.W. and Seung, H.S., 2004, October. Stochastic policy gradient reinforcement learning on a simple 3D biped. In Intelligent Robots and Systems, 2004.(IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on (Vol. 3, pp. 2849-2854). IEEE.

[20] Morimoto, J., Cheng, G., Atkeson, C.G. and Zeglin, G., 2004, April. A simple reinforcement learning algorithm for biped walking. In Robotics and Automation, 2004. Proceedings. ICRA’04. 2004 IEEE International Conference on (Vol. 3, pp. 3030-3035). IEEE.

[21] Lillicrap, T.P., Hunt, J.J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D. and Wierstra, D., 2015. Continuous control with deep reinforcement learning. arXiv preprint arXiv:1509.02971.

[22] Silver, D., Lever, G., Heess, N., Degris, T., Wierstra, D. and Riedmiller, M., 2014, June. Deterministic policy gradient algorithms. In ICML.

[23] Koenig, N. and Howard, A., 2004, September. Design and use paradigms for gazebo, an open-source multi-robot simulator. In Intelligent Robots and Systems, 2004. (IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on (Vol. 3, pp. 2149-2154). IEEE.