Annals of R.S.C.B., ISSN: 1583-6258, Vol. 25, Issue 6, 2021, Pages. 1244-1248
Received 25 April 2021; Accepted 08 May 2021.

Reward Shaping in Reinforcement Learning for Prosthetic Knees

Deepali Salwan[1],Shri Kant[2]

1 Associate Professor, Pt. Deen Dayal Upadhyaya National Institute for Physically Handicapped, New Delhi deepalisalwan@gmail.com
2 Professor, Research and Technology Development Centre Department of Computer Science, School of Engineering and technology Sharda University, Greater Noida
shrikant.ojha@gmail.com

Abstract. *:* In this paper we have looked into Reinforcement Learning Algorithm for Prosthetic Knee Joint. Model Based Aggregate Policy Optimization (MBAPO).Model Free Q Learning with Model Based Learning,It has been found out that with Reward Shaping is an important aspect of Reinforment Learning and can improve the learning in many ways.We have compared the different reward functions and compared the performance on each learning.This approach by far shows the better learning reward and learn faster.This is done in OpenSim and Prosthetics Environment was used.

Keywords: Reinforcement Learning, Artificial Intelligence, Deep Learning, challenges, Model-Based, Model-Free, Prosthesis,OpenSim

1 Introduction

Prosthetic Knee or the Artificial Knee helps the amputees to restore function and is a replacement of their lost leg. The amputation might occur due to various reasons like 1) Diabetes 2) Amputation due to war. There are multiple type of prosthetic knee which are available ranging from Active Knee to Semi Active Knee and Passive Knee/Mechanical Knee. Active Knee is the one in which the knee is automatic, it is a microprocessor-controlled knee. Semi Active Knee has few functions driven by microprocessor and others are carried out mechanically while in a passive knee all the functions are mechanical. The mechanical knee is affordable but is not energy efficient and while the active knee is more comfortable and provide gait symmetry and balance to the amputee. When the amputee moves, the prosthetic knee mimics the flexion and extension (bending and straightening) of an anatomical knee joint. The automatic design of a robotic prosthesis to meet the specifications and physical conditions poses a major technological challenge and is a barrier to the technology's adoption. The primary aim of designing prosthetic systems is to support people with their everyday activities. These devices have allowed amputees to go about these activities confidently and without any hinderance.When compared to passive prostheses, automatic prosthetic knee systems work better as mentioned above. However, tuning a larger number of control parameters to personalise the unit for individual amputee users is difficult. Traditional control methods fail to resolve this problem, and advanced robotic technologies such as Reinforcement learning (RL) are naturally appealing.

2 Experiment

In this study, an effective reinforcement learning algorithm for a sensor-based automatic prosthetic knee of a transfemoral amputee is proposed.

The suggested method uses a model-based and model-free approach to implement a reinforcement learning algorithm. The MF approach is used for *direct cost optimization*, while the MB approaches provide an *additional model for learning*.

In RL, the algorithms try to predict the result for the given problem based on the specific set of tuning parameters.The calculated output is then considered as an input parameter and the new output is evaluated till the optimal output is obtained. RL methods are mainly categorized as Model-Free (MF) and Model-Based (MB) approaches as stated above, while MB approaches often use a learning model. Model-free methods are more effective in studying complex environments, but they need more iterations to achieve convergence, resulting in local minima. Model-based methods, on the other hand, are capable of adapting to new tasks in dynamic environments. They also reduce the number of iterations needed to achieve convergence in real-time scenarios. However, since MB approaches cannot learn on their own, it is necessary to create an accurate model that can be studied or trained in order to generalise them. This is a tough problem to address since any changes to the model do not support policy changes.

As a result, MB approaches are limited to low-dimensional spaces and necessitate complex design to be successful. As a consequence, it's preferable to build models that take advantage of the benefits of current methods while still overcoming their weaknesses.

The proposed method's workflow included the use of a model-based approach as well as model-free learning. The model-based approach was developed using a novel Gaussian Mixture Based Gaussian Method. Model-free reinforcement learning model Trust Region Policy Optimization (TRPO) comparing TRPO, Proximal Policy Optimization (PPO), and Deep Deterministic Learning (DDL) were used.In order to find an efficient algorithm among the TRPO, PPO, and DDPG, a comparative analysis was performed. The simulation results showed that the TRPO algorithm archived the highest and maximum mean reward function of the three algorithms, so it was chosen to create a model-free approach.

The performance of the proposed RL-based MBMF method was calculated using different performance metrics, such as mean and standard deviation measures of cost with respect to iterations, in an OPENSIM prosthetics setting.OPenSim Simulation Environment was used to perform the learning of the Model.The output of model-free approaches like DDPG and PPO was evaluated using the entropy penalty, loss function, loss entropy, mean entropy reward, and reward for mean entropy reward. Other important factors that influence the efficiency of MBMF approaches, such as time steps and Velocity loss, were also assessed. The loss entropy of the MBMF method varies, as can be seen in the simulation.The value of loss entropy reaches a limit of 10.6. TRPO was used to construct the MF control model. To determine the value of model free approaches for updating the learning policy, the MBMF algorithm was evaluated and the model ensembled TRPO (ME-TRPO) was compared to the Vanilla Model-Based RL algorithm. The proposed solution improved the policy regularisation of the learning process, according to the results of the experimental study. It was also noticed that the resulting strategy outperformed all of the models significantly.

Annals of R.S.C.B., ISSN: 1583-6258, Vol. 25, Issue 6, 2021, Pages. 1244-1248
Received 25 April 2021; Accepted 08 May 2021.

The model's success with an automated prosthetic knee was evaluated in various simulation scenarios, including learning to stand, taking a step, and taking two steps forward, and the results are presented. Finally, the analysis addresses the main Skeletal Observations made when using the OPENSIM environment, as well as the limitations encountered during the execution.

8 *Results*

The use of reward shaping in reinforcement learning was illustrated in this study using a variety of learning scenarios, including learning to stand, learning to use muscles, learning to walk with a cross leg penalty, and reward shaping with increasing the velocity reward. The reward shaping mechanism confirmed the proposed theory of reward shaping by empirically demonstrating the analytical findings.

The reward function was performed iteratively using the following reward functions:

- Learning to stand: im_rew = 0.75 * p_re + 0.25 * v_re

Where im_rew is the mean reward, p_re = position reward and v_re = velocity reward.

- Learning to use muscles such as Hamstring for reward:

im_rew = 0.5 * p_re + 0.45 * v_re + 0.05 * m_re

Where p_re = position reward, v_re = velocity reward, and m_re = muscles_reward.

- Learning to walk with penalty of cross leg: im_rew = 0.6 * p_re + 0.20 * v_re + 0.20 * m_re - 0.01 * pe.

Where p_re = position reward, p_re = velocity reward, m_re = muscles_reward, and pe = penalty of cross leg.

The native learning process has a long reward horizon because it lacks the incentive benefit of a change. As a result, learning without incentive shaping can progress at a snail's pace. However, it ultimately hits the optimum value and achieves the best results.

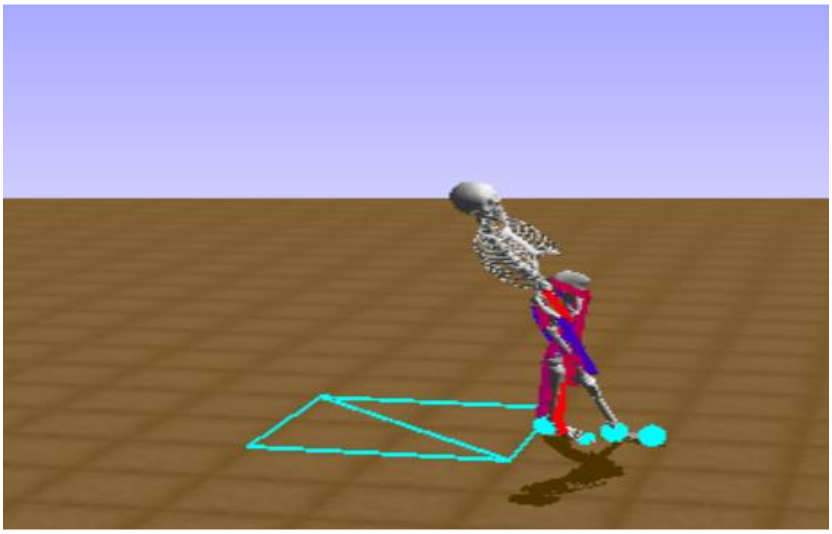

Figure 1: The figure above shows the penalty laid for cross leg in the reward function

Annals of R.S.C.B., ISSN: 1583-6258, Vol. 25, Issue 6, 2021, Pages. 1244-1248
Received 25 April 2021; Accepted 08 May 2021.

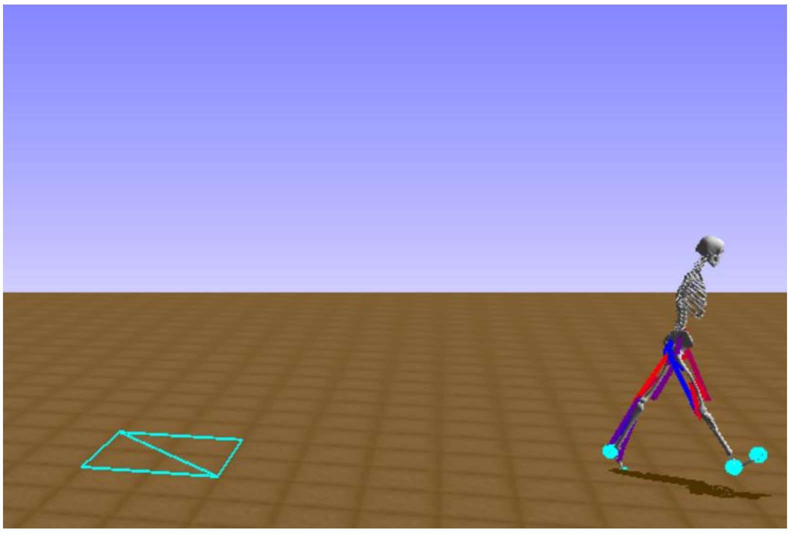

Fig 2: Figure above show the OpenSim Simulator learning to walk in ProstheticsEnv

9 *Conclusion and Future Work*

The following are some of the drawbacks faced when implementing the MBMF approach: The engines were put through their paces one by one. Each training session, which takes about 4 hours per engine, takes a long time to complete. This limits the approach's adaptability. The prosthetic model is unable to walk long distances. For all sessions, the same parameters were used. This applies to all the algorithms. The potential expansion of this research is constituted based on the found shortcomings of reinforcement learning models and model-based model-free approaches. The results indicate that the proposed solution had some design and operational problems, including increased complexity and the use of multiple datasets. This research effort will act as a base for future endeavours. This research's long-term goal is to improve the proposed probabilistic dynamics model, which was developed using both model-based and model-free approaches.

References

1. Deepali Salwan, Shri Kant Ojha, G. Pandian, Role of Artificial Intelligence and Machine Learning in Resolving the Issues and Challenges with Prosthetic Knees, ICICCT 2019 Springer CCIS Proceeding Vol.no 1025,2019.
2. Sutton, R. S. and Barto, A. G. Reinforcement learning: An introduction. MIT press, 2018.
3. Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L.Van Den Driessche, G., Schrittwieser, J., Antonoglou, I.,Panneershelvam, V., Lanctot, M., et al. Mastering the game of go with deep neural networks and tree search. Nature, 529(7587):484, 2016.
4. Kidznski, Lukasz and Mohanty, Sharada P and Ong, Carmichael and Hicks, Jennifer and Francis, Sean and Levine, Sergey and Salath\'e, Marcel and Delp, Scott, Learning to Run challenge: Synthesizing physiologically accurate motion using deep reinforcement learning ,Escalera, Sergio and Weimer, Markus, Springer,2018.

5. Matthew Botvinick, SamRitter, JaneX.Wang,ZebKurth-Nelson, CharlesBlundell,DemisHassabis, Reinforcement Learning, Fast and Slow , Trends in Cognitive Sciences 23(5), DOI: 10.1016/j.tics.2019.02.006,2019.
6. Ding Z., Dong H.,Challenges of Reinforcement Learning. In: Dong H., Ding Z., Zhang S. (eds) Deep Reinforcement Learning. Springer, Singapore,2020.
7. Alonso Marco, Felix Berkenkamp, Philipp Hennig, Angela P. Schoellig, Andreas Krause, Stefan Schaal and Sebastian Trimpe, Virtual vs. Real: Trading Off Simulations and Physical Experiments in Reinforcement Learning with Bayesian Optimization, IEEE International Conference on Robotics and Automation (ICRA), arXiv:1703.01250 [cs.RO],2017.
8. M. Deisenroth and C. Rasmussen, A model-based and data-efficient approach to policy search, in ICML, 2011.
9. Nagabandi, A., Kahn, G., Fearing, R. S., & Levine, S. Neural network dynamics for model-based deep reinforcement learning with model-free fine-tuning.2018 IEEE International Conference on Robotics and Automation (ICRA) (pp. 7559-7566). IEEE,2018.
10. Yue Wen, Jennie Si, Fellow, IEEE, Andrea Brandt, Xiang Gao, and He (Helen) Huang, Online Reinforcement Learning Control for the Personalization of a Robotic Knee Prosthesis, IEEE Transactions On Cybernetics, Vol. 50,2020.
11. Dalal, G., Dvijotham, K., Vecerik, M., Hester, T., Paduraru, C., and Tassa, Y. Safe exploration in continuous action spaces. CoRR, abs/1801.08757, 2018.
12. Ian Osband Charles Blundell, A.P.B.V.R.: Deep exploration via bootstrapped dqn ,2016
13. Chua, K., Calandra, R., McAllister, R., and Levine, S. Deep reinforcement learning in a handful of trials using probabilistic dynamics models. In Advances in Neural Information Processing Systems, pp. 4754–4765, 2018.
14. G. C. Nandi, A. Ijspeert and A. Nandi, Biologically inspired CPG based above knee active prosthesis, IEEE/RSJ International Conference on Intelligent Robots and Systems, Nice, 2008, pp. 2368-2373, doi: 10.1109/IROS.2008.
15. Bansal, Somil, et al. Mbmf: Model-based priors for model-free reinforcement learning. arXiv preprint arXiv:1709.03153,2017.
16. Yoshua, B., Jérome, L., Ronan, C., Jason, W.: Curriculum learning (2009).
17. Zhou, B., Zeng, H., Wang, F., Lian, R., & Tian, H., Efficient and Robust Learning on Elaborated Gaits with Curriculum Learning. In the NeurIPS'18 Competition (pp. 265-276). Springer, Cham,2020.
18. Dulac-Arnold, G., Evans, R., van Hasselt, H., Sunehag, P., Lillicrap, T., Hunt, J…. & Coppin, B. Deep reinforcement learning in large discrete action spaces. arXiv preprint arXiv:1512.07679,2015.
19. Dulac-Arnold, G., Mankowitz, D., & Hester, Challenges of real-world reinforcement learning. arXiv preprint arXiv:1904.12901.,2019.
20. Tiago Maia, Reinforcement learning, conditioning, and the brain: Successes and challenges, Cognitive, Affective, & Behavioural Neuroscience, 9 (4), 343-364, doi:10.3758/CABN.9.4.343,2009.
21. Petar Kormushev, Sylvain Calinon and Darwin G. Caldwel, Reinforcement Learning in Robotics: Applications and Real-World Challenges, Robotics, 2, 122-148; doi:10.3390/robotics2030122,2013.