深度確定性策略梯度於田間履帶機器人之追跡及避障

研究生：李仁祐　　　　　　　　　　指導教授：陳顯禎 教授

國立陽明交通大學 影像與生醫光電碩士班

摘要

本論文以視覺化即時定位與地圖構建(visual simultaneous localization and mapping，vSLAM)以及 YOLO 物件辨識等技術為基礎，利用 MATLAB Simulink 之深度強化學習(deep reinforcement learning，deep RL)中的無模型(model-free approach)深度確定性策略梯度(deep deterministic policy gradient，DDPG)為學習控制架構，其擅長處理非線性時變控制問題以及具有連續動作空間決策優勢。在機器人作業系統(robot operating system，ROS)框架下，以智能體(intelligent agent)強化學習環境系統(environment)出最佳化控制器(actor)後，驅使機器人依目標函數來行走，在這裡主要是動態決策新的避障局部路徑來跟蹤，讓機器人能迅速避障後回到原有的全局路徑。

我們先於 Gazebo 模擬環境中去驗證所學習訓練出的 DDPG 智能體擁有「能夠在執行任務上具有高成功率」的前提條件後，將其引入實驗室所開發的履帶機器人於田間場域中進行實測實驗。此履帶機器人以 Intel RealSense D435i 深度相機用於 vSLAM 中的定位感測，以及結合 YOLO v3 進行障礙物感知定位的輸入，藉由訓練完成 DDPG 智能體來提供最佳化 actor。目前在 Gazebo 模擬中可完成 80%成功率之避障追跡，而在履帶機器

人於田間場域之避障追跡實作中，我們以 RTAB-Map (real-time appearance-based mapping)為 vSLAM 演算法，成功率已達到 60%以上。

關鍵字：vSLAM、YOLO、田間履帶機器人、Simulink、深度確定性策略梯度、避障。

Path Following and Obstacle Avoidance of Field Tracked Robots by Deep Deterministic Policy Gradient

Student: Jen-Yu Lee Advisor: Prof. Shean-Jen Chen
Institute of Imaging and Biomedical Photonics
National Yang Ming Chiao Tung University

Abstract

In this thesis, a deep deterministic policy gradient (DDPG) learning architecture in MATLAB Simulink is utilized to train intelligent agents. The DDPG is a kind model-free approach of deep reinforcement learning (deep RL). Based on the condition that the mature positioning technology and object detection we already have by using vSLAM (visual simultaneous localization and mapping) and YOLO (you only look once), the advantage of good at making decision and dealing with nonlinear time variant control problem of DDPG is used to train an agent as the system controller for controlling the robot to achieve the expected behaviors we want in the robot operating system (ROS) framework. The thing that the trained agent is mainly used to make decision dynamically for the new local path is to let the robot which is driven by agent can be returned rapidly to the reference path after avoiding the obstacle optimally.

Before doing the experiment of obstacle avoidance control on the field tracked robot we developed which will be driven by DDPG trained agent in field, we go first on doing the verification with the trained agent at the simulated environment called Gazebo, make sure of that the success rate of executing task by agent is certainly high. And the sensing of the tracked robot is using Intel RealSense depth camera D435i as both inputs of vSLAM for doing robot localization and YOLO v3 for doing obstacle sensing. Currently, we get 80% success rate in verification of obstacle avoidance control under Gazebo, and 60%

in verification of obstacle avoidance control on the field tracked robot in field by using RTAB-Map (real-time appearance-based mapping) as vSLAM.

Keywords: vSLAM, YOLO, field tracked robot, Simulink, DDPG, obstacle avoidance.

參考文獻

[1] M. Labbé and F. Michaud, "RTAB-Map as an open-source lidar and visual simultaneous localization and mapping library for large-scale and long-term online operation," *Journal of Field Robotics* 36, 416- 446 (2018).

[2] M. Leshno, V. Y. Lin, A. Pinkus, and S. Schocken, "Multilayer feedforward networks with a nonpolynomial activation function can approximate any function," *Neural Networks* 6, 861-867 (1993).

[3] L. Baird, "Residual algorithms: Reinforcement learning with function approximation," *Machine Learning: Proceedings of the Twelfth International Conference*, pages 30-37 (1995).

[4] E. Hernandaz and Y. Arkun, "Neural network modeling and an extended DMC algorithm to control nonlinear systems," *American Control Conference*, 2454-2459 (1990).

[5] I.-M. Chen and C.-Y. Chan, "Deep reinforcement learning based path tracking controller for autonomous vehicle," *Proc. IMechE, Part D: Journal of Automobile Engineering.* (2020). DOI: 10.1177/0954407020954591.

[6] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," *Proc. IEEE Conf. Comp. Vis. Patt. Recogn.*, 779-788 (2016).

[7] X. Gao, T. Zhang, and Y. Liu, *14 Lessons of Visual SLAM: From Theory to Practice[M].* Publishing House of Electronics Industry, Beijing (2017).

[8] D. Nistér, O. Naroditsky, and J. R. Bergen, "Visual odometry," *Intl. Conference on Computer Vision and Pattern Recognition*, 652-659 (2005).

[9] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," arXiv

preprint arXiv:1804.02767 (2018).

[10] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction (2nd Editio)*, MIT Press (2017).

[11] C. J. C. H. Watkins, *Learning from Delayed Rewards,* PhD thesis, University of Cambridge England (1989).

[12] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, *et al*., "Human-level control through deep reinforcement learning," *Nature* 518, 529-533 (2015).

[13] M. L. Puterman, *Markov Decision Processes Discrete Stochastic Dynamic Programming,* John Wiley & Sons, Inc., New York (1994).

[14] J. Peters and S. Schaal, "Reinforcement learning of motor skills with policy gradients," *Neural Networks* 21, 682-697 (2008).

[15] T. Lillicrap, J. Hunt, *et al*., "Continuous control with deep reinforcement learning," *International Conferenceon Learning*, (2016).

[16] D. Silver, G. Lever, N. Heess, T. Degris, D. Wierstra, and M. Riedmiller, "Deterministic policy gradient algorithms," *International Conference on Machine Learning (ICML)*, (2014).

[17] MathWorks Student Competitions Team (2021). MATLAB and Simulink Robotics Arena: Walking Robot (https://github.com/mathworks-robotics/msra-walking-robot), GitHub. Retrieved June 23, 2021.

[18] https://github.com/ROBOTIS-GIT/turtlebot3

[19] E. F. Camacho and C. B. Alba, *Model Predictive Control (2nd Edition)*. Springer Science & Business Media, (2013).

[20] A. Nagabandi, *et al*., "Neural network dynamics for model-based deep reinforcement learning with model-free fine-tuning," *International*

Conference on Robotics and Automation, (2018).

[21] H. Lu, G. Xiong, and K. Guo, "Motion predicting of autonomous tracked vehicles with online slip model identification," *Mathematical Problems in Engineering*, 1-13 (2016).

[22] R. Mur-Artal, J. Montiel, and J. Tardos, "ORB-SLAM: A versatile and accurate monocular SLAM system," *IEEE Transactions on Robotics* 31, 1147-1163 (2015).

[23] K. Reif, S. Gunther, E. Yaz, and R. Unbehauen, "Stochastic stability of the discrete-time extended Kalman filter," *IEEE Transactions on Automatic Control* 44, 714-728 (1999).

[24] Y. Feng and J. Wang, "GPS RTK performance characteristics and analysis," *Journal of Global Positioning Systems* 7, 1-8 (2008).

[25] K. Chua, R. Calandra, *et al*., "Deep reinforcement learning in a handful of trials using probabilistic dynamics models," arXiv:1805.12114, (2018).

[26] B. Evans, H. W. Jordaan, and H. A. Engelbrecht, "Autonomous obstacle avoidance by learning policies for reference modification," arXiv:2102.11042, (2021).

[27] H. Xie, X. Xu, Y. Li, W. Hong, and J. Shi, "Model predictive control guided reinforcement learning control scheme," *2020 International Joint Conference on Neural Networks*, 1-8 (2020).