ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC BÁCH KHOA

KHOA ĐIỆN - ĐIỆN TỬ
BỘ MÔN TỰ ĐỘNG HÓA

ỨNG DỤNG REINFORCEMENT LEARNING ĐIỀU KHIỂN PHÂN TÁN HỆ ĐA ROBOT TRÁNH VA CHẠM

Application of Reinforcement Learning in Decentralized Multi-Robot Collision Avoidance Control System

LUẬN VĂN THẠC SĨ

Học viên thực hiện: Nguyễn Tấn Khôi
MSSV: 2171017
Chuyên ngành: Kỹ thuật Điều khiển - Tự động hóa
Mã số chuyên ngành: 8520216

Giảng viên hướng dẫn: TS. Phạm Việt Cường

TP. Hồ Chí Minh, Tháng 12 năm 2025

LỜI CẢM ƠN

Tôi xin chân thành cảm ơn TS. Phạm Việt Cường, người đã tận tình hướng dẫn và định hướng cho tôi trong suốt quá trình thực hiện luận văn này. Những góp ý quý báu của thầy đã giúp tôi hoàn thiện nghiên cứu và phát triển kỹ năng nghiên cứu khoa học.

Tôi cũng xin gửi lời cảm ơn đến các thầy cô trong Bộ môn Tự động hóa, Khoa Điện - Điện tử, Trường Đại học Bách Khoa TP.HCM đã truyền đạt kiến thức nền tảng vững chắc trong suốt quá trình học tập.

Cuối cùng, tôi xin cảm ơn gia đình, bạn bè đã luôn động viên và hỗ trợ tôi trong suốt thời gian thực hiện luận văn.

Nguyễn Tấn Khôi

TÓM TẮT

Luận văn này nghiên cứu ứng dụng học tăng cường sâu (Deep Reinforcement Learning) vào bài toán điều khiển phân tán hệ đa robot tránh va chạm. Dựa trên thuật toán Proximal Policy Optimization (PPO), nghiên cứu đề xuất các cải tiến về learning rate scheduling, value clipping và chiến lược huấn luyện hai giai đoạn để cải thiện hiệu suất và khả năng mở rộng của hệ thống.

Kết quả thực nghiệm cho thấy mô hình đạt tỷ lệ thành công 71% trong quá trình huấn luyện với 58 robots và 88% trong kiểm thử với 50 robots. Các cải tiến đề xuất bao gồm Adaptive Learning Rate Scheduler, Value Clipping, và việc sử dụng hai optimizer riêng biệt cho actor và critic networks đã giúp cải thiện độ ổn định và tốc độ hội tụ của quá trình huấn luyện.

Nghiên cứu cũng trình bày thiết kế phần cứng robot sử dụng cảm biến LiDAR 360 độ và Raspberry Pi, chuẩn bị cho việc triển khai thuật toán trên robot thực tế trong tương lai.

Từ khóa: Học tăng cường, PPO, Đa robot, Tránh va chạm, Điều khiển phân tán

ABSTRACT

This thesis investigates the application of Deep Reinforcement Learning to decentralized multi-robot collision avoidance control systems. Based on the Proximal Policy Optimization (PPO) algorithm, the research proposes improvements in learning rate scheduling, value clipping, and two-stage training strategy to enhance system performance and scalability.

Experimental results demonstrate that the model achieves a 71% success rate during training with 58 robots and 88% during testing with 50 robots. The proposed improvements, including Adaptive Learning Rate Scheduler, Value Clipping, and separate optimizers for actor and critic networks, have enhanced training stability and convergence speed.

The research also presents a hardware design using 360-degree LiDAR sensors and Raspberry Pi, preparing for future deployment on real robots.

Keywords: Reinforcement Learning, PPO, Multi-robot, Collision Avoidance, Decentralized Control

LỜI CAM ĐOAN

Tôi xin cam đoan rằng luận văn *"Ứng Dụng Reinforcement Learning Điều Khiển Phân Tán Hệ Đa Robot Tránh Va Chạm"* là công trình nghiên cứu của riêng tôi dưới sự hướng dẫn của TS. Phạm Việt Cường.

Các kết quả nghiên cứu và số liệu trong luận văn là trung thực, chưa từng được ai công bố trong bất kỳ công trình nào khác. Tôi xin hoàn toàn chịu trách nhiệm về tính chính xác và trung thực của nội dung luận văn này.

TP. Hồ Chí Minh, Tháng 12 năm 2025

Nguyễn Tấn Khôi

MỤC LỤC

DANH MỤC HÌNH

DANH MỤC BẢNG

DANH MỤC KÝ HIỆU VÀ CHỮ VIẾT TẮT

CNN	Convolutional Neural Network - Mạng nơ-ron tích chập
DRL	Deep Reinforcement Learning - Học tăng cường sâu
GAE	Generalized Advantage Estimation - Ước lượng ưu thế tổng quát
HCMUT	Ho Chi Minh City University of Technology - Trường Đại học Bách Khoa TP.HCM
LiDAR	Light Detection and Ranging - Công nghệ quét laser để đo khoảng cách
POMDP	Partially Observable Markov Decision Process - Quá trình quyết định Markov quan sát từng phần
PPO	Proximal Policy Optimization - Tối ưu hóa chính sách gần kề
RL	Reinforcement Learning - Học tăng cường
a_t	Action at time step t - Hành động tại bước thời gian t
o_t	Observation at time step t - Quan sát tại bước thời gian t
r_t	Reward at time step t - Phần thưởng tại bước thời gian t
v	Linear velocity - Vận tốc tuyến tính
ω	Angular velocity - Vận tốc góc (omega)
π	Policy function - Hàm chính sách (pi)
θ	Neural network parameters - Tham số mạng nơ-ron (theta)

Ghi chú: Danh sách này sẽ được cập nhật thêm các ký hiệu khác khi xuất hiện trong các chương của luận văn.

CHƯƠNG 1

MỞ ĐẦU

Chương này giới thiệu bối cảnh, động lực, mục tiêu, phạm vi và ý nghĩa của đề tài nghiên cứu về ứng dụng học tăng cường trong điều khiển phân tán hệ đa robot tránh va chạm.

1.1 Lý do chọn đề tài

Trong những năm gần đây, hệ thống đa robot đã và đang được ứng dụng rộng rãi trong nhiều lĩnh vực như kho hàng tự động, nhà máy thông minh, và logistics. Việc điều khiển nhiều robot hoạt động đồng thời trong cùng một môi trường đặt ra thách thức lớn về tránh va chạm, đặc biệt khi số lượng robot tăng lên.

Các phương pháp điều khiển tập trung truyền thống gặp khó khăn về khả năng mở rộng (scalability) khi số lượng robot lớn, do yêu cầu về tính toán và truyền thông tăng theo cấp số nhân. Ngược lại, phương pháp điều khiển phân tán, trong đó mỗi robot tự quyết định dựa trên quan sát cục bộ, cho phép hệ thống mở rộng tốt hơn.

Học tăng cường sâu (Deep Reinforcement Learning - DRL) đã chứng minh khả năng giải quyết các bài toán điều khiển phức tạp. Đặc biệt, thuật toán Proximal Policy Optimization (PPO) được đánh giá cao về độ ổn định và hiệu quả trong việc huấn luyện các chính sách điều khiển. Tuy nhiên, việc áp dụng DRL vào bài toán đa robot với số lượng lớn (trên 40 robots) vẫn còn nhiều thách thức cần nghiên cứu.

1.2 Mục tiêu nghiên cứu

Mục tiêu chính của luận văn là nghiên cứu và phát triển thuật toán điều khiển phân tán cho hệ đa robot tránh va chạm sử dụng học tăng cường sâu. Cụ thể:

1. Nghiên cứu thuật toán PPO và các phương pháp học tăng cường cho bài toán đa robot.
2. Huấn luyện mô hình neural network có khả năng điều khiển 44-58 robots tránh va chạm trong môi trường mô phỏng.
3. Đạt tỷ lệ thành công tối thiểu 71% trong huấn luyện và 88% trong kiểm thử.
4. Đề xuất các cải tiến về learning rate scheduling, value clipping và chiến lược huấn luyện để cải thiện hiệu suất.
5. Thiết kế phần cứng robot prototype với cảm biến LiDAR để chuẩn bị cho việc triển khai thực tế.

1.3 Đối tượng và phạm vi nghiên cứu

1.3.1 Đối tượng nghiên cứu

Đối tượng nghiên cứu là hệ thống đa robot di động (mobile robots) với các đặc điểm:

- **Kiểu robot:** Nonholonomic robots (robot không toàn hướng) di chuyển trên mặt phẳng 2D
- **Cảm biến:** LiDAR 360 độ với 454 điểm quét
- **Hành động:** Điều khiển vận tốc tuyến tính v và vận tốc góc ω
- **Số lượng:** 44-58 robots hoạt động đồng thời

1.3.2 Phạm vi nghiên cứu

Luận văn tập trung vào các khía cạnh sau:

- Thuật toán học tăng cường PPO cho bài toán điều khiển phân tán
- Môi trường mô phỏng 2D với chướng ngại vật tĩnh
- Quan sát cục bộ (local observation) từ cảm biến LiDAR
- Không sử dụng truyền thông giữa các robot (fully decentralized)
- Mục tiêu: Di chuyển từ vị trí xuất phát đến đích mà không va chạm

Ngoài phạm vi: Chướng ngại vật động, môi trường 3D, multi-task learning.

1.4 Ý nghĩa khoa học và thực tiễn

1.4.1 Ý nghĩa khoa học

Luận văn đóng góp vào lĩnh vực nghiên cứu đa robot và học tăng cường thông qua:

- Đề xuất các cải tiến về thuật toán huấn luyện PPO:
  - *Adaptive Learning Rate Scheduler*: Duy trì learning rate khi performance cải thiện
  - *Value Clipping*: Giảm instability trong quá trình huấn luyện
  - *Separate Optimizers*: Sử dụng learning rate khác nhau cho actor và critic (tỷ lệ 1:15)
  - *Learning Rate Warmup*: Tăng dần learning rate trong giai đoạn đầu
- Nghiên cứu scalability của thuật toán với số lượng robot lớn (so với bài báo gốc: 4-20 robots).
- Phân tích chi tiết 8 revisions thực nghiệm và ảnh hưởng của các hyperparameters.

1.4.2 Ý nghĩa thực tiễn

Kết quả nghiên cứu có thể ứng dụng vào:

- **Kho hàng tự động (Automated Warehouses):** Điều khiển nhiều robot AGV (Automated Guided Vehicle) di chuyển hàng hóa hiệu quả.
- **Nhà máy thông minh (Smart Factories):** Phối hợp nhiều robot công nghiệp di động trong dây chuyền sản xuất.
- **Logistics và vận tải:** Quản lý đội xe tự hành trong khu vực hạn chế.

Kết quả đạt được (88% test success rate với 50 robots) gần với bài báo gốc (96.5% với 20 robots), chứng minh tính khả thi của phương pháp trong điều kiện thử nghiệm khắt khe hơn.

Tóm lại, chương này đã trình bày động lực, mục tiêu, phạm vi và ý nghĩa của đề tài nghiên cứu. Các chương tiếp theo sẽ trình bày chi tiết về tổng quan nghiên cứu liên quan, phương pháp thực hiện, kết quả đạt được và kết luận.

CHƯƠNG 2

TỔNG QUAN

Chương này trình bày tổng quan về các nghiên cứu liên quan đến học tăng cường, thuật toán PPO, và ứng dụng trong bài toán đa robot tránh va chạm.

2.1 Phương pháp của Long et al.

Bài báo gốc của Long et al. [1] đề xuất phương pháp điều khiển phân tán sử dụng Deep Reinforcement Learning với thuật toán PPO. Mô hình sử dụng kiến trúc CNN với 2 lớp Conv1D và huấn luyện theo 2 giai đoạn, đạt kết quả 96.5-100% success rate với 4-20 robots.

2.2 So sánh các phương pháp tránh va chạm

2.3 Cơ sở lý thuyết

2.3.1 Reinforcement Learning

2.3.2 Proximal Policy Optimization (PPO)

Thuật toán PPO [2] là một trong những thuật toán policy gradient hiệu quả nhất hiện nay.

2.4 Đánh giá tình hình nghiên cứu

Chương này đã tổng quan các nghiên cứu liên quan và xác định khoảng trống nghiên cứu cần giải quyết.

CHƯƠNG 3

PHƯƠNG PHÁP NGHIÊN CỨU

Chương này trình bày chi tiết phương pháp nghiên cứu bao gồm môi trường mô phỏng, kiến trúc mạng nơ-ron, và quy trình huấn luyện.

3.1 Môi trường mô phỏng

3.2 Kiến trúc mạng nơ-ron

3.3 Cải tiến so với bài báo gốc

3.4 Quy trình huấn luyện

3.5 Thiết kế robot thực tế

Chương này đã trình bày phương pháp nghiên cứu chi tiết.

CHƯƠNG 4

KẾT QUẢ VÀ THẢO LUẬN

Chương này trình bày kết quả thực nghiệm và thảo luận về các phát hiện.

4.1 Kết quả Stage 1

4.2 Kết quả Stage 2

4.3 So sánh với bài báo gốc

4.4 Phân tích các cải tiến

4.5 Thảo luận

Chương này đã trình bày kết quả thực nghiệm và phân tích chi tiết.

CHƯƠNG 5

KẾT LUẬN VÀ KIẾN NGHỊ

Chương này tóm tắt những đóng góp chính, hạn chế và hướng nghiên cứu tương lai.

5.1 Tóm tắt kết quả đạt được

Luận văn đã hoàn thành các mục tiêu đề ra:

- Huấn luyện thành công mô hình điều khiển đa robot với 71% training success và 88% test success
- Đề xuất 6 cải tiến chính so với bài báo gốc
- Thiết kế prototype robot với LiDAR

5.2 Hạn chế của nghiên cứu

5.3 Hướng nghiên cứu tiếp theo

5.4 Lời kết

Luận văn đã đạt được mục tiêu nghiên cứu về ứng dụng học tăng cường trong điều khiển đa robot tránh va chạm.

TÀI LIỆU THAM KHẢO

[1] P. Long, T. Fan, X. Liao, W. Liu, H. Zhang **and** J. Pan, "Towards optimally decentralized multi-robot collision avoidance via deep reinforcement learning," **in** *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Brisbane, Australia: IEEE, 2018, **pages** 6252–6259. DOI: 10.1109/ICRA.2018.8461113.

[2] J. Schulman, F. Wolski, P. Dhariwal, A. Radford **and** O. Klimov, "Proximal policy optimization algorithms," *arXiv preprint arXiv:1707.06347*, 2017.