NGHIÊN CỨU MỘT CÁCH TIẾP CẬN MỚI DỰA TRÊN HỌC TĂNG CƯỜNG CHO BÀI TOÁN TỐI ƯU HÓA ĐỒNG THỜI TỐC ĐỘ VÀ MỨC TIÊU THỤ NHIÊN LIỆU TRONG VẬN HÀNH TÀU THỦY

A STUDY ON A NOVEL REINFORCEMENT LEARNING-BASED APPROACH FOR THE SIMULTANEOUS OPTIMIZATION PROBLEM OF SPEED AND FUEL CONSUMPTION IN SHIP OPERATION

NGUYỄN VĂN TIẾN*, ĐỖ KHẮC TIỆP

Khoa Điện - Điện tử, Trường Đại học Hàng hải Việt Nam

**Email liên hệ: nguyenvantien@vimaru.edu.vn*

DOI: https://doi.org/10.65154/jmst.900

Tóm tắt

Tối ưu hóa nhiên liệu tàu thủy là một bài toán then chốt trong vận hành hàng hải, nhưng nó mâu thuẫn trực tiếp với mục tiêu duy trì tốc độ hành trình. Bài báo này trình bày một ứng dụng của kỹ thuật học tăng cường RL (Reinforcement Learning) để giải quyết bài toán tối ưu hóa này. Trong bài báo này, một môi trường mô phỏng tàu KVLCC2 trên Simulink đã được xây dựng, bao gồm động lực học tàu và các điều kiện lực cản môi trường biến động. Một tác nhân RL (RL Agent) được huấn luyện để đưa ra quyết định điều khiển công suất động cơ. Kết quả nghiên cứu được so sánh với bộ điều khiển PID truyền thống cho thấy tác nhân RL có thể cân bằng giữa lượng nhiên liệu và thời gian hành trình tốt hơn. Nghiên cứu cũng chứng minh khả năng sử dụng RL, thông qua việc điều chỉnh hàm phần thưởng để tạo ra một đường cong đánh đổi hoàn chỉnh. Điều này cung cấp một công cụ hỗ trợ ra quyết định linh hoạt, cho phép nhà vận hành lựa chọn chiến lược tối ưu dựa trên nhu cầu thay vì bị giới hạn trong một giải pháp cứng nhắc.

Từ khóa: *Học tăng cường, mô hình tàu thủy, hiệu quả năng lượng, tối ưu hóa đa mục tiêu, vận hành tàu thủy.*

Abstract

Optimizing fuel consumption in maritime operations is a critical objective, yet it inherently conflicts with the goal of maintaining high voyage speeds. This paper presents an application of Reinforcement Learning (RL) techniques to address this optimization trade-off. In this study, a simulation environment for the KVLCC2 ship model has been developed in Simulink, incorporating vessel dynamics and dynamic environmental disturbances. An RL agent is trained to make control decisions regarding propulsion power. The performance is benchmarked against a conventional PID controller, demonstrating that the RL agent achieves a superior balance between fuel consumption and voyage time. Furthermore, the study demonstrates the capability of the RL approach, through reward function shaping, to generate a complete trade-off curve. This provides a flexible decision-support tool, enabling operators to select an optimal strategy based on operational demands, rather than being constrained to a single, rigid solution.

Keywords: *Reinforcement learning, marine vessel model, energy efficiency, multi-objective optimization, ship Operation.*

1. Mở đầu

Ngành công nghiệp vận tải biển là trụ cột của thương mại toàn cầu, nhưng đồng thời cũng là một trong những ngành tiêu thụ nhiên liệu hóa thạch lớn nhất. Chi phí nhiên liệu thường chiếm từ 50% đến 60% tổng chi phí vận hành của một tàu thương mại.

Bài toán tối ưu hóa hành trình tàu thủy là một bài toán phức tạp, đa mục tiêu, trong đó hai mục tiêu quan trọng nhất là giảm thiểu thời gian hành trình (tối đa hóa tốc độ) và giảm thiểu tiêu thụ nhiên liệu (tối đa hóa lợi nhuận), lại mâu thuẫn trực tiếp với nhau. Các nỗ lực giải quyết bài toán này chủ yếu bao gồm tối ưu hóa tuyến đường và điều khiển cục bộ kết hợp dự báo.

Tối ưu hóa tuyến đường là hướng tiếp cận phổ biến nhất, tập trung vào việc tìm kiếm một tuyến đường tối ưu về mặt không gian (2D/3D). Các nghiên cứu thường sử dụng các phương pháp cổ điển như thuật toán đường đẳng thời Isochrone [1, 2, 3] hoặc các thuật toán tìm đường trên đồ thị như Dijkstra và A* [4,

TẠP CHÍ ISSN: 1859-316X, e-ISSN: 3093-3161
KHOA HỌC CÔNG NGHỆ HÀNG HẢI
JOURNAL OF MARINE SCIENCE AND TECHNOLOGY

5]. Các phương pháp này chủ yếu tối ưu hóa con đường và thường giả định một chiến lược tốc độ cố định. Chúng không giải quyết được vấn đề nên điều khiển công suất như thế nào khi đã ở trên tuyến đường đó.

Với hướng tiếp cận điều khiển cục bộ và dự đoán thì lại tập trung vào điều khiển bản thân con tàu. Phương pháp phổ biến nhất trong vận hành thực tế là sử dụng bộ điều khiển PID (Proportional-Integral-Derivative) để duy trì một tốc độ hoặc công suất máy đặt trước [6]. Tuy nhiên bộ điều khiển PID không có khả năng thích nghi cao với nhiễu. Khi gặp lực cản lớn (sóng, gió), PID sẽ tăng công suất lên tối đa để giữ tốc độ, dẫn đến mức tiêu thụ nhiên liệu tăng vọt một cách phi tuyến tính.

Một số nghiên cứu gần đây đã áp dụng học máy như trong nghiên cứu của Le (2020) [7]. Trong nghiên cứu này Le sử dụng mạng nơron nhân tạo ANN (Artificial Neural Network) và dữ liệu AIS để dự đoán mức tiêu thụ nhiên liệu. Tuy nhiên các mô hình này mang tính mô tả hoặc dự đoán chứ không đưa ra chiến lược điều khiển để tiết kiệm nhiên liệu.

Học tăng cường là một nhánh của học máy chuyên giải quyết các bài toán ra quyết định. RL đã cho thấy tiềm năng lớn trong lĩnh vực hàng hải. Tuy nhiên, các ứng dụng hiện tại chủ yếu tập trung vào các bài toán điều khiển cục bộ, thời gian thực. Wu & Chen (2020) [8] đã ứng dụng thành công RL cho hệ thống định vị động DP (Dynammic Positioning). Trong khi Zhao & Yun. (2020) sử dụng RL cho bài toán bám quỹ đạo của tàu tự hành [9].

Từ phân tích trên, bài báo đề xuất ứng dụng kỹ thuật học tăng cường, cụ thể là thuật toán DDPG (Deep Deterministic Policy Gradient), để huấn luyện một tác nhân (Agent) để điều khiển tàu. Tác nhân này không chỉ học một chiến lược duy nhất, mà học cách hiểu sự đánh đổi giữa hai mục tiêu mâu thuẫn là thời gian hành trình và tiêu thụ nhiên liệu

Đóng góp chính của nghiên cứu này là. Chứng minh rằng tác nhân RL có thể tìm ra các chiến lược vận hành phức tạp, vượt trội hơn hẳn so với điều khiển PID truyền thống. PID được lựa chọn làm đối tượng so sánh vì đây là bộ điều khiển tiêu chuẩn trên các tàu thủy hiện nay, giúp làm rõ hiệu quả kinh tế so với phương thức vận hành truyền thống. Đồng thời đưa ra một phương pháp luận thông qua việc điều chỉnh hàm phần thưởng để tạo ra một đường biên đánh đổi Pareto (Pareto Front), cho phép nhà vận hành lựa chọn một chiến lược ưu tiên tốc độ hoặc ưu tiên nhiên liệu dựa trên nhu cầu kinh doanh cụ thể.

2. Cơ sở lý thuyết nghiên cứu

Để giải quyết bài toán tối ưu hóa đa mục tiêu, nghiên cứu này xây dựng một môi trường mô phỏng trên nền tảng MATLAB/Simulink và ứng dụng một tác nhân học tăng cường RL để tìm kiếm chiến lược điều khiển tối ưu.

2.1. Xây dựng môi trường mô phỏng

Môi trường mô phỏng được định nghĩa như một quy trình quyết định Markov (MDP-Markov Decision Process) [10], bao gồm các mô hình vật lý của tàu, nhiên liệu và hành trình.

2.1.1. Mô hình động lực học tàu

Trong nghiên cứu này, chúng tôi mô phỏng tàu KVLCC2 [11] được phát triển bởi Viện nghiên cứu tàu biển và đại dương Hàn Quốc (KRISO). KVLCC2 là con tàu được tạo ra để phục vụ nghiên cứu, mô phỏng và thử nghiệm trong ngành kỹ thuật hàng hải. Nó cung cấp một bộ dữ liệu chuẩn, công khai và đầy đủ về hình dáng vỏ tàu, động lực học, và hiệu suất chân vịt cho phép các nhà khoa học và kỹ sư trên toàn thế giới chạy các mô phỏng và thử nghiệm trong ngành kỹ thuật hàng hải.

Trong nghiên cứu này, mô hình tàu KVLCC2 được đơn giản hóa về động lực học 1D (chuyển động thẳng). Việc đơn giản hóa mô hình tàu về động lực học 1D là một quyết định có chủ đích, cho phép chúng tôi tập trung vào đúng bài toán cốt lõi là điều khiển tốc độ tàu thay vì điều khiển bẻ lái.

Về mặt kỹ thuật, điều này giảm đáng kể độ phức tạp của không gian trạng thái và giúp cho việc huấn luyện tác nhân RL trở nên khả thi. Hơn nữa, nó tăng tốc độ tính toán lên hàng ngàn lần, cho phép hoàn thành huấn luyện trong vài giờ thay vì vài tháng. Thêm nữa, vì lực đẩy và lực cản theo phương dọc tàu là các yếu tố quyết định trên một hành trình dài . Lượng nhiên liệu tiêu thụ thêm do các chuyển động thứ cấp (như lắc ngang, hoặc bẻ lái nhẹ để giữ thẳng đường) là rất nhỏ so với tổng lượng nhiên liệu của động cơ chính và có thể bỏ qua một cách hợp lý trong giải quyết bài toán chiến lược điều khiển. Do vậy không cần đến các mô hình 3 bậc hoặc 6 bậc.

Với mô hình 1D, gia tốc của tàu tại mỗi thời điểm được xác định bởi định luật II Newton:

$$M_{ship} \cdot a(t) = F_{net}(t) \quad (1)$$

Trong đó M_{ship} là tổng khối lượng của tàu bao gồm cả nước dằn (với KVLCC2, $M_{ship} = 3,2.10^8 kg$).

Để tăng tốc đáng kể quá trình huấn luyện học tăng cường, vốn đòi hỏi trải qua hàng ngàn chuyến đi, một

kỹ thuật co giãn thời gian đã được áp dụng. Cụ thể, khối lượng của tàu trong mô phỏng đã được giảm đi 100 lần, trong khi các thông số lực (lực đẩy, lực cản) được giữ nguyên. Điều này làm tăng gia tốc của tàu lên 100 lần và giảm thời gian hoàn thành mỗi chuyến đi xuống 100 lần. Quan trọng là, sự thay đổi này không ảnh hưởng đến tính tối ưu của chính sách mà tác nhân học được, vì mối quan hệ giữa trạng thái tàu và hành động vẫn được bảo toàn. Thêm vào đó, các điều kiện môi trường được kích hoạt dựa trên quãng đường đã đi chứ không phải thời gian (time-based), đảm bảo tính nhất quán của bài toán.

Giá trị $F_{net}(t)$ là tổng lực tác động lên tàu, được định nghĩa là:

$$F_{net}(t) = F_{thrust}(t) - F_{resistance}(t) \quad (2)$$

Lực đẩy của động cơ tàu F_{thrust} được điều khiển bởi tác nhân RL. Nó là một hàm của giá trị $\alpha(t)$ (phần trăm công suất $\alpha(t)$) và lực đẩy tối đa F_{max} :

$$F_{thrust}(t) = \alpha(t) \cdot F_{max} \quad (3)$$

Thành phần lực cản $F_{resistance}$ bao gồm hai thành phần chính là lực cản vỏ tàu R_{hull} (phụ thuộc vào vận tốc) và lực cản môi trường R_{env} (do sóng, gió):

$$F_{resistance}(t) = R_{hull}(t) + R_{env}(t) \quad (4)$$

Lực cản vỏ tàu được mô hình hóa dưới dạng phi tuyến bậc hai so với vận tốc $v(t)$ [12]:

$$R_{hull}(t) = C_{hull} \cdot v(t)^2 \quad (5)$$

Trong đó C_{hull} là hệ số cản của vỏ tàu (với KVLCC2, $C_{hull} \approx 32.900$).

Kết hợp các phương trình (3-5), gia tốc tức thời $a(t)$ của tàu (đầu ra của mô hình) được tính bằng công thức sau:

$$a(t) = \frac{\alpha(t) \cdot F_{max} - (C_{hull} \cdot v(t)^2 + R_{env}(t))}{M_{ship}} \quad (6)$$

2.1.2. Mô hình tiêu thụ nhiên liệu

Mức tiêu thụ nhiên liệu $Fuel_{Rate}$ (kg/s) không tuyến tính với công suất. Nó được mô hình hóa bằng một hàm mũ, cho thấy việc chạy ở công suất cao F_{max} sẽ tốn nhiều nhiên nhiên liệu hơn nhiều so với chạy ở công suất trung bình [13]:

$$Fuel_{Rate}(t) = FR_{max} \cdot (\alpha(t))^{\beta} \quad (7)$$

Trong đó FR_{max} là tỷ lệ tiêu thụ nhiên liệu tối đa tại 100% công suất (với KVLCC2, $FR_{max} \approx 1{,}14\ kg/s$) và β là hệ số phi tuyến. Trong nghiên cứu này $\beta = 2$ phù hợp mô hình hóa một đặc tính của động cơ diesel hàng hải.

2.1.3. Mô hình môi trường và hành trình

Để tạo môi trường huấn luyện cho tác tử RL, chúng tôi định nghĩa một tuyến hành trình cố định có tổng chiều dài tuyến là D_{total} , và tạo lực cản môi trường lên tàu bằng hàm R_{env} , là một hàm của quãng đường đã đi $d(t)$:

$$R_{env}(t) = f(d(t)) \quad (8)$$

Hàm $f(d)$ là một hàm hằng số từng đoạn để mô phỏng các vùng thời tiết khác nhau.

2.2. Thiết kế tác nhân học tăng cường

Chúng tôi sử dụng thuật toán DDPG để huấn luyện tác nhân. Thuật toán này phù hợp với không gian hành động liên tục. Các tác nhân học tăng cường được huấn luyện theo quy trình quyết định Markov gọi tắt là MDP [10]. Đây là một giải thuật toán học được sử dụng để mô hình hóa việc ra quyết định trong các tình huống mà kết quả vừa bị ảnh hưởng bởi hành động của người ra quyết định, vừa có yếu tố ngẫu nhiên như bài toán điều khiển tàu thủy.

Nói một cách đơn giản, MDP là luật mà chúng ta tạo ra cho một tác nhân để nó học hỏi. Các thành phần của MDP bao gồm: Trạng thái S_t ; hành động A_t ; phần thưởng R_t . Mục tiêu là huấn luyện tác nhân RL để học một chính sách điều khiển để từ một trạng thái S_t đưa ra tới một hành động A_t để tối đa hóa tổng phần thưởng R_t trong suốt một chuyến đi.

Trạng thái S_t là các thông tin tác nhân RL quan sát để ra quyết định. Đây là đầu vào của tác nhân và là một véc-tơ 3 chiều:

$$S_t = [v(t), d_{remain}(t), R_{env}(t)] \quad (9)$$

Trong đó $v(t)$ là giá tri vận tốc hiện tại của tàu, $d_{remain}(t) = D_{total} - d(t)$ là quãng đường còn lại của hành trình và $R_{env}(t)$ là lực cản môi trường.

Hành động A_t là quyết định mà tác nhân đưa ra tại mỗi bước tính. Trong trường hợp của nghiên cứu này thì A_t là phần trăm công suất động cơ cần thiết để đẩy tàu. Đây là một giá trị vô hướng, liên tục:

$$A_t = \alpha(t) \quad (10)$$

Khi tác nhân đạt được thành công sẽ nhận được phần thưởng thông qua hàm phần thưởng R_t . Đây là cốt lõi của bài toán tối ưu hóa đa mục tiêu. Hàm phần thưởng được thiết kế để dạy tác nhân về sự đánh đổi giữa tốc độ và nhiên liệu.

Trong bài toán tối ưu hóa đa mục tiêu này, hàm phần thưởng tức thời $r_t(s_t, a_t)$ tại bước thời gian t được thiết kế để cân bằng giữa vận tốc vận hành trình

TẠP CHÍ ISSN: 1859-316X, e-ISSN: 3093-3161
KHOA HỌC CÔNG NGHỆ HÀNG HẢI
JOURNAL OF MARINE SCIENCE AND TECHNOLOGY

v_t và tốc độ tiêu thụ nhiên liệu $\dot{m}_{f,t}$, đồng thời áp dụng các ràng buộc an toàn:

$$r_t(s_t,a_t) = w_v v_t - w_f \dot{m}_f(a_t) - \xi \cdot \mathcal{P}_{over}(v_t) + \delta \cdot \mathcal{R}_{goal} \quad (11)$$

Trong đó, w_v, w_f là các trọng số ưu tiên cho vận tốc và nhiên liệu được thay đổi theo chiến lược điều khiển; $\dot{m}_f(a_t)$ là hàm tiêu thụ nhiên liệu phi tuyến phụ thuộc hành động a_t (công suất).; $\mathcal{P}_{over}(v_t)$ là hàm phạt vi phạm tốc độ giới hạn; $\mathcal{R}_{goal}$ là hàm phần thưởng lớn khi hoàn thành hành trình.

Mục tiêu của bài toán tối ưu là tìm kiếm một chính sách điều khiển $\pi: \mathcal{S} \to \mathcal{A}$ sao cho cực đại hóa kỳ vọng tổng phần thưởng tích lũy có chiết khấu $J(\pi)$ trong suốt hành trình [14]:

$$J(\pi) = \mathbb{E}_\pi \left[\sum_{t=0}^{T} \gamma^t \cdot r_t(s_t, \pi(s_t)) \right] \quad (12)$$

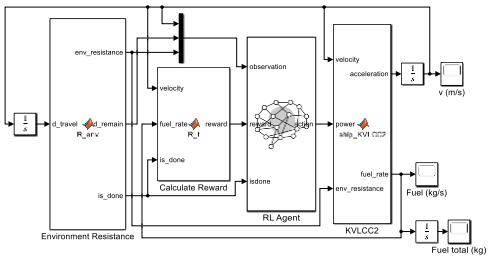
Trong đó:

$J(\pi)$ Hàm mục tiêu cần tối ưu hóa;

$\mathbb{E}_\pi$: Kỳ vọng toán học theo phân phối của chính sách π và động lực học môi trường;

$\gamma \in [0,1]$: Hệ số chiết khấu, xác định tầm quan trọng của các phần thưởng tương lai.

Sơ đồ Hình 1, mô tả một vòng lặp học tăng cường RL hoàn chỉnh để điều khiển tàu KVLCC2 được thực hiện bằng MATLAB/Simulink.



Hình 1. Sơ đồ mô phỏng điều khiển tàu KVLCC2 với vòng lặp học tăng cường và tác tử RL

Trung tâm là khối RL Agent, khối này đưa ra hành động điều khiển công suất động cơ cho khối KVLCC2. Khối KVLCC2 tính toán gia tốc và tốc độ tiêu thụ nhiên liệu của tàu tại bước thời gian *t*. Gia tốc được tích phân để tính ra vận tốc, sau đó được tích phân một lần nữa để ra quãng đường đã đi. Khối tạo môi trường mô phỏng (Environment Resistance) dựa trên quãng đường này và kịch bản môi trường để tạo ra lực cản. Cuối cùng, khối tính toán phần thưởng (Calculate Reward) sử dụng lượng nhiên liệu tiêu thụ và vận tốc để chấm điểm cho hành động. Tác nhân RL nhận phần thưởng và quan sát vector trạng thái (gồm vận tốc, quãng đường còn lại, lực cản) này để học và đưa ra hành động tiếp theo.

Để giải bài toán tối ưu hóa hàm mục tiêu (12) trong không gian hành động liên tục, nghiên cứu của chúng tôi sử dụng thuật toán DDPG dựa trên lý thuyết Gradient. Thuật toán này xấp xỉ hàm mục tiêu thông qua hai mạng nơ-ron:

1. Mạng Critic $Q(s,a|\theta^Q)$, đánh giá chất lượng hành động bằng cách giải phương trình tối ưu Bellman:

$$Q^*(s,a) = \mathbb{E}[r_t + \gamma Q^*(s_{t+1}, \pi(s_{t+1}))]$$

2. Mạng Actor $\mu(s|\theta^\mu)$: Cập nhật chính sách điều khiển trực tiếp theo hướng gradient để cực đại hóa hàm J, với quy tắc cập nhật tham số θ^μ:

$$\nabla_{\theta^\mu} J \approx \mathbb{E}[\nabla_a Q(s,a|\theta^Q)|_{a=\mu(s)} \cdot \nabla_{\theta^\mu} \mu(s|\theta^\mu)]$$

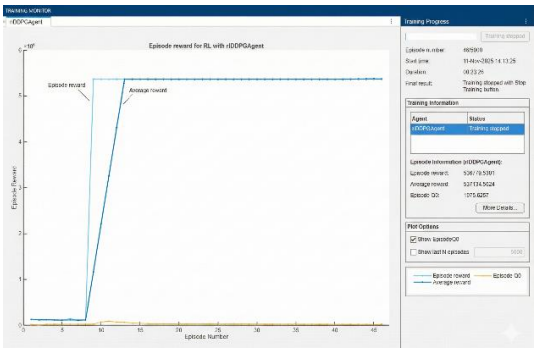
Cơ chế này đảm bảo tác nhân tìm được điểm tối ưu cục bộ cho chiến lược điều khiển trong môi trường phi tuyến phức tạp.

Chi tiết về quy trình tương tác giữa tác nhân RL và môi trường mô phỏng tàu KVLCC2, cũng như cơ chế cập nhật trọng số mạng nơ-ron theo giải thuật DDPG, được mô tả cụ thể trong thuật toán dưới dạng giả mã dưới đây:

1	**Khởi tạo mô hình và tham số** • Mô hình động lực học tàu KVLCC2 • Thông tin tuyến đường $D_{total} = 100km$ • Thông số môi trường • Các siêu tham số: $MaxEpisodes = 5000$, , $MaxSteps = 40000$. • Mạng Actor $\mu(s\|\theta^\mu)$ và Critic $Q(s,a\|\theta^Q)$ với trọng số ngẫu nhiên. • Mạng mục tiêu μ' và Q'. • Bộ nhớ đệm $\mathcal{D}$ • Chu kỳ trích mẫu T=1 giây • Hệ số chiết khấu $\gamma = 0.99$
2	**For** *Episodes* **to** *MaxEpisodes*
3	Khởi tạo lại môi trường Simulink Nhận trạng thái đầu $s_1 = [v, d_{remain}, F_{env}]^T$.
4	**For** $t = 1$ **to** *MaxSteps*
5	Chọn mức công suất máy cộng nhiễu thăm dò $a_t = \mu(s_t|\theta^\mu) + \mathcal{N}_t$
6	Tính toán động lực học: Gia tốc, Vận tốc v_{t+1}, Tiêu thụ nhiên liệu $\dot{m}_f$. Cập nhật trạng thái tiếp theo s_{t+1} dựa trên vị trí trên tuyến đường.
7	Tính phần thưởng r_t: $r_t = w_{speed} \cdot v_{t+1} - w_{fuel} \cdot \dot{m}_f + P_{vi\_pham} + P_{hoan\_thanh}$
8	Lưu (s_t, a_t, r_t, s_{t+1}) vào bộ nhớ đệm $\mathcal{D}$
9	Lấy ngẫu nhiên một mẫu từ $\mathcal{D}$ Cập nhật mạng Critic Q Cập nhật mạng Actor μ
10	Kiểm tra điều kiện dừng:

	If $d_{remain} \le 0$ **Then** Exit For
11	**End For**
12	Trả về chính sách điều khiển tối ưu μ^*

Trên Hình 2 là quá trình huấn luyện tác nhân theo phương pháp học tăng cường.



Hình 2. Đồ thị quá trình huấn luyện tác nhân

Quá trình huấn luyện tác nhân RL diễn ra trong chuyến đi dài 100km, tối đa 5000 chuyến đi, tối đa 40000 giây/chuyến. Tác nhân sẽ được thưởng 0,5 điểm cho mỗi m/s vận tốc tăng thêm và bị phạt 1,0 điểm cho mỗi kg/s nhiên liệu tăng thêm. Giá trị này sẽ thay đổi trong các chiến lược điều khiển khác nhau. Khi về đích tác nhân sẽ được thưởng 500.000 điểm, giá trị tượng trưng cho phần thưởng rất lớn để tác nhân nhanh về đích.

Đường xanh dương (Episode reward) là điểm thưởng của từng chuyến đi riêng lẻ. Đường xanh nhạt (Average reward) là điểm thưởng trung bình, đây là đường quan trọng nhất để đánh giá hiệu suất. Trên đồ thị huấn luyện Hình 2, có thể thấy quá trình huấn luyện rất thành công và diễn ra cực kỳ nhanh chóng thể hiện bằng việc hai đường hội tụ và đi ngang ở giá trị rất lớn 537.134 điểm (500.000 là điểm thưởng về đích và 37.134 là điểm thưởng nhận được từ việc cân bằng tốc độ và nhiên liệu).

3. Kết quả mô phỏng

3.1. Thiết lập môi trường thực nghiệm

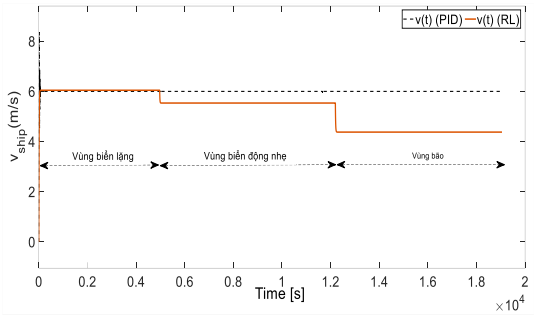
Quá trình huấn luyện và mô phỏng kiểm chứng được thực hiện trên phần mềm MATLAB/Simulink phiên bản 2025a sử dụng hộp công cụ Reinforcement Learning Toolbox™. Phần cứng sử dụng là máy tính cá nhân với cấu hình CPU Intel Core i5, RAM 32GB.

Tác nhân DDPG được cấu hình với hai mạng nơ-ron sâu riêng biệt cho Actor và Critic. Cả hai mạng đều sử dụng kiến trúc mạng truyền thẳng với 3 lớp ẩn, mỗi lớp gồm 128 neurons và sử dụng hàm kích hoạt ReLU để đảm bảo khả năng học các đặc tính phi tuyến phức tạp của động lực học tàu thủy.

Đối tượng là tàu KVLCC 2, khối lượng $M_{ship} = 3,2.10^8 (kg)$, hệ số lực cản thân tàu $C_{hull} \approx 32.900$, tốc độ tiêu thụ nhiên liệu tối đa $FR_{max} \approx 1,14 kg/s$ tổng hành trình của một chuyến đi được thiết lập là $D_{total} = 100(km)$. Điều kiện môi trường đưa vào để thử thách các bộ điều khiển gồm 3 vùng với các lực cản khác nhau. Vùng biển lặng với lực cản lên tàu là 0,1 (MN), lực cản vùng biển động vừa là 0,5 (MN) và lực cản vùng bão là 1,2 (MN).

3.2. Kịch bản 1: So sánh RL với PID truyền thống khi tàu chạy đầy tải

Đầu tiên, chúng tôi kiểm tra tính hiệu quả của ứng dụng RL bằng cách so sánh giữa 2 bộ điều khiển khi tàu đầy tải. Một bên là bộ điều khiển PID (điều khiển với tốc độ không đổi) bên kia là một tác nhân RL điều điều khiển với chiến lược cân bằng giữa thời gian chuyến đi và lượng nhiên liệu tiêu thụ ($\omega_{speed} = 0,5$ và $\omega_{fuel} = 1,0$). Các tham số PID được xác định thông qua công cụ PID tuner của Matlab tại điểm làm việc định mức (đầy tải, 6m/s) để đảm bảo đáp ứng đầu ra tối ưu trước khi đưa vào so sánh. Hình 3 chỉ ra chiến lược điều khiển tốc độ của PID (nét đứt) và bộ điều khiển với tác nhân RL (nét liền) khi gặp điều kiện môi trường khác nhau.

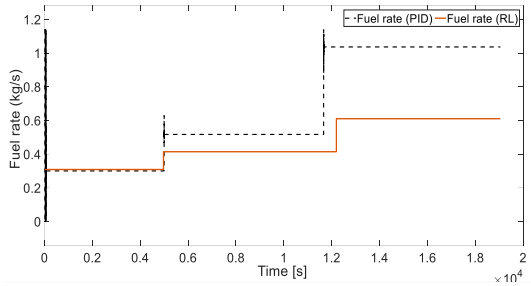


Hình 3. Tốc độ tàu khi sử dụng PID và bộ điều khiển với tác nhân RL trong kịch bản 1

Bộ điều khiển PID luôn duy trì tốc độ đặt trước là 6m/s trong cả 3 vùng, bất chấp điều kiện môi trường. Tác nhân RL cũng duy trì được tốc độ cao ở vùng biển lặng là 6m/s và vùng biển động vừa là 5,5m/s. Tuy nhiên khi đi vào vùng bão, tác nhân RL điều chỉnh giảm tốc độ xuống còn ~ 4,3m/s. Tác nhân RL đã học được từ quá trình huấn luyện rằng việc cố gắng giữ 6m/s trong bão sẽ gây lãng phí nhiên liệu cực kỳ lớn (bị phạt nặng). Chiến lược tối ưu để đạt tổng điểm thưởng cao nhất là chấp nhận đi chậm lại một để đổi lấy việc tiết kiệm một lượng lớn nhiên liệu.

Điều này có thể thấy rõ ràng hơn khi quan sát đồ thị tốc tiêu thụ nhiên liệu ở Hình 4.

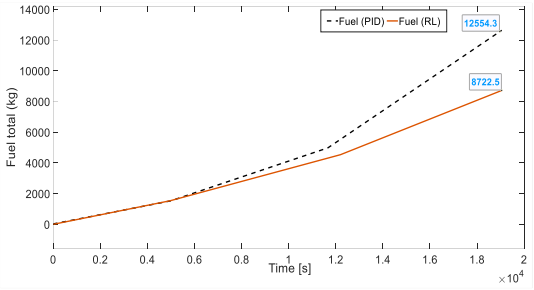
TẠP CHÍ ISSN: 1859-316X, e-ISSN: 3093-3161
KHOA HỌC CÔNG NGHỆ HÀNG HẢI
JOURNAL OF MARINE SCIENCE AND TECHNOLOGY



Hình 4. Đồ thị tốc độ tiêu thụ nhiên liệu trong các điều kiện thời tiết của 2 bộ điều khiển PID so với RL

Với bộ điều khiển PID, khi đi vào vùng bão, nó phải tăng công suất lên mức rất cao để chống lại lực cản và giữ tốc độ 6m/s do vậy nhiên liệu tiêu thụ rất lớn. Lượng nhiên liệu cần thiết tăng vọt lên 1,05kg/s, gần đạt mức tối đa 1,14kg/s. Tác nhân RL có chiến lược điều khiển linh hoạt hơn là giảm tốc độ mạnh xuống còn 4,3m/s. Kết quả là, nó chỉ cần mức nhiên liệu ~0,61kg/s, tiết kiệm gần một nửa so với PID trong bão.

Đồ thị Hình 5, cho thấy tổng lượng nhiên liệu được tích lũy theo thời gian trong suốt hành trình 100km. Tổng lượng nhiên liệu tàu tiêu thụ khi dùng bộ điều khiển PID là 12.554,3kg. Với việc điều khiển bằng tác nhân RL, lượng nhiên liệu tiêu thụ giảm gần 30% (8.722,5kg của RL so với 12.554,3kg của PID).



Hình 5. Tổng khối lượng nhiên liệu mà tàu tiêu thụ trong hành trình 100km

Trên Bảng 1, so sánh hiệu suất giữa bộ điều khiển PID và tác tử RL đang hoạt động với chiến lược điều khiển cân bằng giữa thời gian và mức độ tiêu thụ nhiên liệu.

Bảng 1. So sánh hiệu suất giữa PID và RL

Phương pháp	Tổng thời gian (giây)	Tổng nhiên liệu (kg)
PID	16.200	12.554,3
RL	19.000	8.722,5

Bộ điều khiển PID (đường nét đứt) kết thúc hành trình (tại 16.200 giây) với tổng mức tiêu thụ là 12.554,3 kg. Tác nhân RL (đường nét liền), kết thúc hành trình (tại 19.000 giây) với tổng mức tiêu thụ là 8.722,5 kg. Nghĩa là tác nhân RL điều khiển tàu về đích chậm hơn khoảng 2.800 giây. Đây không phải là một thất bại, mà là một sự đánh đổi có chủ đích. RL đã chọn đi chậm lại trong cơn bão để tiết kiệm gần 4 tấn nhiên liệu, đúng như dự đoán của một chiến lược tối ưu hóa chi phí.

3.3. Kịch bản 2: Phân tích bền vững của RL so với PID

Trong thử kịch bản này RL và PID được cho điều khiển tàu ở trạng thái không tải (chỉ bao gồm khối lượng dằn). Điều này thường không gặp trong thực tế, nhưng một con tàu chạy không tải có đặc tính vật lý hoàn toàn khác so với khi đầy tải. Khi tàu chạy không tải, nó nhẹ hơn đáng kể và có mớn nước nông hơn. Điều này dẫn đến 2 thay đổi vật lý chính:

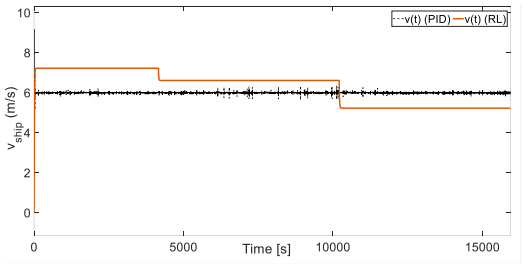
• Khối lượng giảm: Tàu nhẹ hơn rất nhiều, lực điều khiển và gia tốc tàu sẽ thay đổi.

• Lực cản giảm: Vì mớn nước nông hơn, diện tích vỏ tàu chìm dưới nước ít hơn, dẫn đến lực cản vỏ tàu cũng giảm theo.

Việc thử nghiệm này sẽ kiểm tra tính tổng quát của RL rằng, liệu một RL được huấn luyện cho tàu nặng có thể lái tốt một con tàu nhẹ không.

Trong kịch bản này, tàu được dằn chỉ bằng 60% khối lượng đầy tải, lực cản vỏ tàu chỉ còn 70%. Các bộ điều khiển được giữ nguyên thông số như kịch bản 1.

Trên Hình 6, là đồ thị tốc độ tàu trong kịch bản 2. Khi tàu chạy không tải (không tải, khối lượng nhẹ hơn, lực cản ít hơn), cả hai bộ điều khiển đều bộc lộ đặc tính của mình. Tốc độ tàu khi dùng PID (nét đứt) mặc dù giữ được tốc độ 6m/s nhưng xuất hiện các gợn răng cưa dày đặc trên đặc tính. Tác tử RL vẫn điều khiển tàu với chiến lược như kịch bản 1 tuy nhiên vận tốc lúc này cao hơn hẳn so với kịch bản lúc trước.



Hình 6. Đồ thị vận tốc trong kịch bản 2 khi tàu chạy không tải

Bộ điều khiển PID được chỉnh địnhvới các thông số P, I, và D tối ưu cho một con tàu nặng. Khi PID điều khiển một con tàu nhẹ (không tải, quán tính nhỏ,

lực cản ít hơn) ổn định ở tốc độ 6m/s thì tham số không còn phù hợp gây ra dao động khiến đường vận tốc bị gợn sóng.

Với RL, vì mục tiêu không phải ổn định tốc độ do vậy khi gặp tàu nhẹ, RL không cố gắng ổn định tốc độ và xuất ra cùng một lệnh điều khiển như trường hợp có tải nhưng vì khối lượng và lực cản ít hơn, nó tạo ra tốc độ cao hơn nhiều (~7,12m/s).

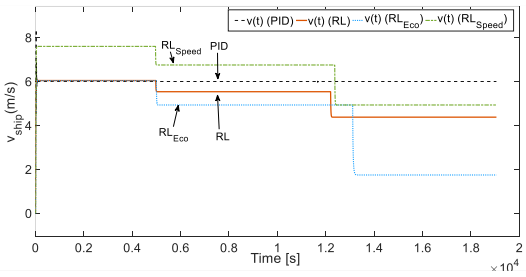
Kết quả mô phỏng cho chứng tỏ, tính bền vững của RL và điều đó đến từ việc nó đưa ra một quyết định dựa trên kinh nghiệm đã học, tạo ra một lệnh ổn định. Ngược lại, PID trở nên mất ổn định vì nó dựa trên việc sửa lỗi và các thông số sửa lỗi của nó không còn phù hợp với vật lý mới của con tàu.

3.4. Kịch bản 3: Đánh giá khả năng tối ưu hóa linh hoạt của RL

Để kiểm tra tính linh hoạt, chúng tôi huấn luyện hai tác nhân RL mới bằng cách thay đổi hàm phần thưởng đặt tên hai tác nhân mới là RL_Eco và RL_Speed để so sánh.

RL_Eco: Tác tử này điều khiển với chiến lược tiết kiệm nhiên liệu nhất có thể bằng cách tăng ω_{fuel} lên 5 điểm để phạt nặng khi làm mất nhiều nhiên liệu.

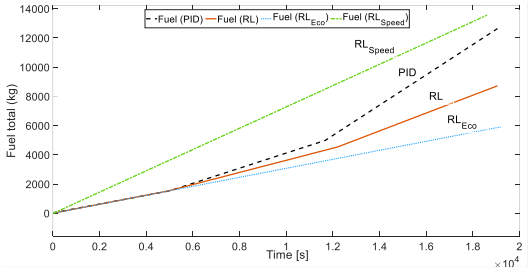
• RL_Speed: Tác tử này điều khiển với chiến lược tiết kiệm nhiên liệu nhất có thể bằng cách tăng ω_{fuel} lên 5 điểm để phạt nặng khi làm mất nhiều nhiên liệu.



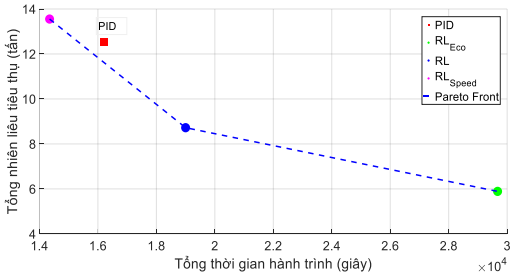
Hình 7. Tốc độ tàu khi sử dụng PID và bộ điều khiển với tác nhân RL_Eco, RL_Speed và RL cân bằng

Trên Hình 7, là tốc độ tàu khi được điều khiển bởi tác nhân RL_Eco, RL_Speed, RL cân bằng (được tạo ra trong kịch bản 1 và 2) và PID. Ở vùng biển lặng RL_Eco và RL cân bằng đều điều khiển tàu chạy với vận tốc trung bình ~ 6m/s, tương tự PID. Sử dụng chiến lược điều khiển khác, tác tử RL_Speed điều khiển tàu chạy với tốc độ tối đa ~7,19m/s để rút ngắn thời gian hành trình nhiều nhất có thể. Vùng biển động vừa (bắt đầu từ 0,5.10⁴ giây), lực cản tăng, các tác nhân điều khiển giảm tốc độ. Tác nhân RL_Eco điều khiển giảm tốc độ mạnh nhất, xuống còn 5m/s để tiết kiệm nhiều nhiên liệu hơn. Trong khi đó tác nhân RL_Speed có điều chỉnh giảm nhưng vẫn còn rất cao (~6,9m/s) vì nó được huấn luyện để ưu tiên tốc độ.

Vùng bão (bắt đầu từ 1,3.10⁴ giây), lực cản tăng vọt, tác nhân đã được huấn luyện và giảm tốc độ. RL_Eco giảm tốc độ xuống mức rất thấp (chỉ còn ~1.8 m/s) để tiết kiệm tối đa. RL_Eco đã học được là nhiên liệu là quan trọng nhất. Tốc độ không quan trọng, nó sẵn sàng đi chậm hơn đáng kể để đổi lấy việc tiết kiệm nhiên liệu tối đa. RL_Speed cũng được huấn luyện giảm tốc độ khi lực cản tăng, tuy nhiên vẫn giữ tốc độ khá cao (5m/s) vì chỉ số phạt nhiên liệu của RL_Speed không cao bằng RL_Eco. Việc chạy ở tốc độ cao như RL_Speed sẽ tiêu tốn nhiều nhiên liệu hơn hẳn (13553,8kg) so với RL_Eco chỉ là 5890,7kg, bù lại thời gian hoàn thành hành trình bằng một nửa so với RL_Eco (14.347 giây của RL_Speed và 29.673 giây của RL_Eco). Điều này có thể thấy trong đồ thị tiêu thụ nhiên liệu ở Hình 8. Đường tổng lượng nhiên liệu tiêu thụ của RL_Speed (nét chấm-gạch) cao hẳn lên so với các đường còn lại.



Hình 8. Đồ thị tổng lượng nhiên liệu tiêu thụ trong hành trình trong kịch bản 3



Hình 9. Đường cong đánh đổi giữa lượng tiêu thụ nhiên liệu và thời gian

Kết quả tổng hợp lượng tiêu thụ nhiên liệu và thời gian hành trình của các tác tử RL như trong Bảng 2.

Dựa trên dữ liệu tổng hợp trong Bảng 2, chúng tôi có thể xây dựng nên đường biên giới hạn các giải pháp tối ưu (Pareto Front) tạo bởi ba tác nhân RL như Hình 9. Đường biên này cho thấy mối quan hệ đánh đổi giữa hai mục tiêu mâu thuẫn: thời gian hành trình và tiêu thụ nhiên liệu, không thể cải thiện một mục tiêu nào mà không phải hy sinh ít nhất một mục tiêu khác.

TẠP CHÍ ISSN: 1859-316X, e-ISSN: 3093-3161
KHOA HỌC CÔNG NGHỆ HÀNG HẢI
JOURNAL OF MARINE SCIENCE AND TECHNOLOGY

Bảng 2. Bảng tổng hợp so sánh hiệu năng của các tác tử RL trong kịch bản 2

Phương pháp	Tổng thời gian (giây)	Tổng nhiên liệu (kg)
PID	16.200	12.554,3
RL	19.000	8.722,5
RL_Eco	29.673	5.890,7
RL_Speed	14.347	13.553,8

Điều này có nghĩa là không thể có một giải pháp tốt nhất duy nhất, chỉ có thể có tốt nhất cho mục tiêu cụ thể nào đó. Đường cong này cho phép lựa chọn tối ưu cho mục đích điều khiển. Nếu muốn đi nhanh hơn (di chuyển sang trái), buộc phải chấp nhận tốn nhiều nhiên liệu hơn (di chuyển lên trên). Ngược lại, nếu muốn tiết kiệm nhiên liệu (di chuyển xuống dưới), bắt buộc phải đi chậm hơn (di chuyển sang phải).

Điểm PID nằm bên ngoài (phía trên) đường cong này. Điều này có nghĩa là PID là một chiến lược không hiệu quả.

4. Kết luận

Bài báo này đã trình bày một cách tiếp cận mới dựa trên học tăng cường RL, sử dụng thuật toán DDPG, để giải quyết bài toán tối ưu hóa đa mục tiêu mâu thuẫn giữa tốc độ hành trình và mức tiêu thụ nhiên liệu. Một môi trường mô phỏng động lực học 1D của tàu KVLCC2 đã được xây dựng thành công trên nền tảng MATLAB/Simulink, cho phép tác nhân RL học các chiến lược điều khiển công suất động cơ một cách hiệu quả.

Các kết quả mô phỏng đã chứng minh rõ ràng tính ưu việt của phương pháp đề xuất so với bộ điều khiển PID truyền thống về hiệu suất, tính bền vững và tính linh hoạt.

Các nghiên cứu trong tương lai có thể tập trung vào việc mở rộng mô hình 1D hiện tại lên 3-DOF (bao gồm cả điều khiển bẻ lái) hoặc 6-DOF để tăng tính chính xác. Hơn nữa, một hướng đi triển vọng là kết hợp tác nhân RL (giải quyết bài toán điều khiển cục bộ) với các thuật toán tối ưu hóa tuyến đường (như A* hoặc Isochrone) để tạo ra một hệ thống hỗ trợ quyết định toàn diện cho hành trình tàu thủy.

Lời cảm ơn

Nghiên cứu này được tài trợ bởi Trường đại học Hàng hải Việt Nam trong đề tài mã số: **DT25-26.66**.

TÀI LIỆU THAM KHẢO

[1] Y.-H. Lin and M.-C. Fang, (2013), *The Ship-Routing Optimization Based on the Three-Dimensional Modified Isochrone Method,* in Proc. ASME 2013 32nd Int. Conf. Ocean, Offshore and Arctic Eng., 2013,
DOI: 10.1115/OMAE2013-10959.

[2] J. Szlapczynska and R. Smierzchalski, (2007), *Adopted isochrone method improving ship safety in weather routing with evolutionary approach,* Int. J. Rel. Qual. Saf. Eng., Vol.14, No.06, pp.635-645,
DOI: 10.1142/S0218539307002842.

[3] Y. Han, W. Tian, and W. Mao, (2024), *Strategies to improve the isochrone algorithm for ship voyage optimisation,* Ships Offshore Struct., Vol.19, No.12, pp.2137-2149,
DOI: 10.1080/17445302.2024.2329011.

[4] Z. Yin et al., (2024), *Multi-Objective Marine Route Optimization Based on Extended A* Algorithm and Ship Performance Models,* J. Mar. Sci.

[5] K. D. Huu, (2024), *Research on Ship Weather Routing Method Based on Dijkstra Algorithm and Neural Network,* in Proc. 2024 Int. Conf. Ind. Eng., Appl. Manuf. (ICIEAM), Sochi, Russia, May 2024,
DOI: 10.1109/ICIEAM61910.2024.10553838.

[6] Z. Hu, W. Guo, K. Zhou, et al., (2024), *Optimization of PID control parameters for marine dual-fuel engine using improved particle swarm algorithm,* Sci. Rep., Vol.14, p. 12681.
DOI: 10.1038/s41598-024-63253-y.

[7] L. T. Le, G. Lee, K. S. Park, and H. Kim, (2020), *Neural network-based fuel consumption estimation for container ships in Korea,* Marit. Policy Manag., Vol.47, No.5, pp.615-632.
DOI: 10.1080/03088839.2020.1729437.

[8] X. Wu et al., (2020), *The autonomous navigation and obstacle avoidance for USVs with ANOA deep reinforcement learning method,* Knowl.-Based Syst., Vol.196, p 105201.
DOI: 10.1016/j.knosys.2019.105201.

[9] Y. Zhao et al., (2025), *Deep Reinforcement Learning-Based Energy Management Strategy for Green Ships Considering Photovoltaic Uncertainty,* J. Mar. Sci. Eng., Vol.13, No.3, p. 565.
DOI: 10.3390/jmse13030565.

[10] R. Bellman, (1957), *A Markovian Decision Process,* Indiana Univ. Math. J., Vol.6, No.4, pp.679-684,
DOI: 10.1512/IUMJ.1957.6.56038

[11] Y. S. Lee, S. H. Van, D. H. Kim, S. Y. Kim, and S. H. Kim, (2008), *Standard model test of KVLCC2,* in Proc. SIMMAN 2008 Workshop Verification Validation Ship Manoeuvring Simulation Methods, Copenhagen, Denmark, Sep. 2008, pp.27-34.

[12] T. I. Fossen, (2021), *Handbook of Marine Craft Hydrodynamics and Motion Control, 2nd ed*. Hoboken, NJ, USA: John Wiley & Sons.

[13] D. A. Taylor, (1996), *Introduction to Marine Engineering, 2nd ed.* London, UK: Butterworth-Heinemann.

[14] H. Xiao, L. Fu, C. Shang, X. Bao, and X. Xu, (2025), *A Knowledge Distillation Compression Algorithm for Ship Speed and Energy Coordinated Optimal Scheduling Model Based on Deep Reinforcement Learning,* IEEE Trans. Transp. Electrif., Vol.11, No.1.

[15] J. Kim, B. Hwang, G.-H. Kim, and U.-G. Kim, (2024), *Advancing Maritime Route Optimization: Using Reinforcement Learning for Ensuring Safety and Fuel Efficiency,* Int. J. e-Navigation Marit. Econ.

Ngày nhận bài:	15/11/2025
Ngày nhận bản sửa:	11/12/2025
Ngày duyệt đăng:	14/12/2025