

NGHIÊN CỨU MÔ HÌNH HỌC LIÊN TỤC TỪ CÁC DỮ LIỆU THU THẬP THEO THỜI GIAN THỰC ĐỂ NÂNG CAO ĐỘ CHÍNH XÁC KHI DỰ BÁO LƯỢNG NHIÊN LIỆU TRÊN TÀU BIỂN

A STUDY ON CONTINUAL LEARNING MODELS BASED ON REAL-TIME DATA TO ENHANCE FUEL CONSUMPTION PREDICTION ACCURACY FOR SEAGOING VESSELS

TRẦN HỒNG HÀ*, PHAN TRUNG KIÊN, ĐỖ THỊ HIỀN

Khoa Máy tàu biển, Trường Đại học Hàng hải Việt Nam

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

DOI: <https://doi.org/10.65154/jmst.1014>

Tóm tắt

Trong bối cảnh chuyển đổi số ngành hàng hải, việc dự báo chính xác tiêu thụ nhiên liệu là yếu tố then chốt để tối ưu hóa chi phí vận hành và tuân thủ các quy định phát thải quốc tế như CII và EEXI. Tuy nhiên, các mô hình học máy tĩnh truyền thống thường gặp khó khăn trước hiện tượng "trôi dạt khái niệm" khi điều kiện môi trường và kỹ thuật tàu thay đổi theo thời gian. Nghiên cứu này đề xuất một kiến trúc học liên tục (Continual Learning) tiên tiến, ứng dụng kỹ thuật Củng cố trọng số đàn hồi (EWC) nhằm bảo tồn tri thức cũ trong khi thích nghi với dòng dữ liệu thời gian thực. Để nâng cao độ bền vững, nghiên cứu sử dụng thuật toán phân tách Cholesky tạo lập 20 bộ dữ liệu tổng hợp với quy mô 200.000 mẫu dựa trên ma trận tương quan thực tế từ tàu Ashico Victoria. Kết quả thực nghiệm cho thấy mô hình học liên tục đạt độ chính xác cao với sai số MAE trung bình là 0,745 Tấn, duy trì sự ổn định vượt trội so với mô hình tĩnh khi đối mặt với dữ liệu mới. Hệ thống đã chứng minh khả năng dự báo thực tiễn với độ tin cậy đạt 98,1%, cung cấp giải pháp kỹ thuật hiệu quả cho các hệ thống quản lý năng lượng thông minh (SEMS) hướng tới hàng hải xanh.

Từ khóa: Dự báo tiêu thụ nhiên liệu, học liên tục, củng cố trọng số đàn hồi, tàu Ashico Victoria, quản lý năng lượng tàu thủy, hệ thống quản lý năng lượng thông minh.

Abstract

In the era of maritime digital transformation, accurate fuel consumption prediction is a cornerstone for optimizing operational costs and ensuring compliance with international emission regulations, such as the Carbon Intensity Indicator (CII) and the Energy Efficiency Existing

Ship Index (EEXI). However, traditional static machine learning models often encounter significant challenges due to "concept drift" and "catastrophic forgetting" when processing updated data streams. This study proposes an advanced Continual Learning (CL) architecture utilizing the Elastic Weight Consolidation (EWC) technique to preserve historical knowledge while adapting to real-time data streams. To enhance model robustness, a synthetic data generation method based on Cholesky decomposition was employed to produce 20 independent datasets totaling 200,000 samples, derived from the actual correlation matrix of the vessel Ashico Victoria. Experimental results demonstrate that the proposed Continual Learning model achieves high accuracy with a Mean Absolute Error (MAE) of 0.745 tons, outperforming static models in terms of adaptability to non-linear fluctuations in hydrographic and technical conditions. The system achieved a confidence level of 98.1%, providing an effective technical solution for intelligent Ship Energy Management Systems (SEMS) and contributing to the roadmap for green and sustainable maritime operations.

Keywords: Fuel consumption prediction, continual Learning, Elastic Weight Consolidation (EWC), Ashico Victoria vessel, ship energy management, Intelligent Ship Energy Management Systems (SEMS).

1. Mở đầu

Trong kỷ nguyên chuyển đổi số hàng hải, việc dự báo chính xác tiêu thụ nhiên liệu đã được xác định là trụ cột then chốt trong các hệ thống quản lý năng lượng thông minh nhằm tối ưu hóa chi phí và giảm phát thải. Các nghiên cứu giai đoạn đầu chủ yếu tập

trung vào ứng dụng các mô hình học máy tĩnh như Mạng nơ-ron nhân tạo hoặc Rừng ngẫu nhiên (Random Forest) để thiết lập mối liên hệ giữa các thông số vận hành và nhiên liệu tiêu thụ [1, 2]. Tuy nhiên, thực tế vận hành tàu cho thấy các mô hình này nhanh chóng bộc lộ hạn chế khi đối mặt với sự biến đổi liên tục của điều kiện môi trường và tình trạng kỹ thuật tàu theo thời gian.

Thách thức lớn nhất hiện nay là hiện tượng "trôi dạt" nơi phân phối dữ liệu thay đổi khiến mô hình mất đi độ chính xác vốn có. Các công trình nghiên cứu về khai thác dòng dữ liệu đã đề xuất việc thu thập và xử lý dữ liệu thời gian thực từ các hệ thống giám sát trên tàu như một giải pháp để nắm bắt các biến động này [5]. Tuy nhiên, việc cập nhật mô hình bằng dữ liệu mới thường dẫn đến hiện tượng "quên lãng thảm họa" [3, 4], một vấn đề kinh điển trong mạng nơ-ron khiến tri thức mới chèn ép và xóa bỏ các quy luật quan trọng đã học được từ quá khứ.

Để giải quyết vấn đề này, xu hướng nghiên cứu hiện đại đã chuyển dịch sang phương pháp học liên tục [3, 5]. Các nghiên cứu nền tảng về cấu trúc mạng đã chỉ ra rằng việc sử dụng các cơ chế bảo tồn trọng số (như EWC) cho phép mô hình tích lũy tri thức một cách bền vững. Đặc biệt, sự phát triển của các thư viện mã nguồn mở chuyên dụng như River trên ngôn ngữ Python đã tạo ra bước ngoặt trong việc triển khai học máy trực tuyến [7]. Các công cụ này hỗ trợ xây dựng mô hình có khả năng học tập tuần tự trên từng điểm dữ liệu, phù hợp hoàn hảo với hạ tầng phần cứng của các hệ thống máy tính công nghiệp như Dell, Intel được trang bị trên tàu vận tải hiện nay.

Kế thừa những thành tựu trên, nghiên cứu của Tran và cộng sự [3] đã thực hiện bước tiến mới bằng việc thiết lập quy trình xử lý dữ liệu động, cho phép mô hình dự báo nhiên liệu tự thích nghi mà không cần huấn luyện lại từ đầu. Việc kết hợp các biến số môi trường và tình trạng kỹ thuật thực tế vào cấu trúc học liên tục đã được thực nghiệm chứng minh là duy trì được độ chính xác cao đồng thời tối ưu hóa đáng kể tải nguyên tính toán. Đây chính là giải pháp then chốt để hiện đại hóa đội tàu, đáp ứng các tiêu chuẩn khắt khe về hiệu quả năng lượng trong bối cảnh hàng hải xanh hiện nay.

2. Cơ sở lý thuyết

2.1. Thu thập và xử lý dữ liệu

Nghiên cứu sử dụng tập dữ liệu vận hành thực tế được trích xuất từ tàu Ashico Victoria, một tàu chở hàng rời có trọng tải toàn phần xấp xỉ 10,000MT thuộc sở hữu của Công ty ASHICO. Khu vực hoạt động

chính của tàu tập trung tại các tuyến hàng hải ở khu vực Châu Á. Dữ liệu được tổng hợp trực tiếp từ nhật ký hàng hải (Logbooks/ Noon Reports) qua nhiều hải trình, bao gồm 82 bộ dữ liệu gốc. Các thông số được ghi nhận bao gồm hai nhóm biến chính:

- Nhóm vận hành và thủy động học: Quãng đường di chuyển (NM), vận tốc tàu (knots), tải trọng hàng hóa (MT) và độ chúi (Trim).

- Nhóm tác động môi trường: Chiều cao sóng (m) và tốc độ gió (knots).



Distance Run (NM)	Vessel Speed (knots)	Cargo Load (tonnes)	Wave Height (m)	Wind Speed (knots)	Trim	Fuel Consumption (tone)
18	9	9417	3,25	24,3	-0,02	0,88
226	9,4	9417	3,25	18,4	-0,02	10,5
257	10,7	9417	3,25	24,3	-0,02	10,48
229	9,54	9417	3,25	24,3	-0,02	10,49
219	9,1	9417	3,25	24,3	-0,02	10,5
60	9,2	2340	1,88	18,4	-2,5	9,2
236	9,7	2340	1,88	18,4	-2,5	9,79
248	10	2340	1,88	18,4	-2,5	9,8
240	10		1,88	18,4	-2,35	9,8
200	8,9	7000	0,88	18,4	-0,94	9,375
224	9,33	7000	0,88	12,7	-0,94	9,96
203	8,7	7000	0,88	12,7	-0,94	9,95
202	8,7	7000	0,88	12,7	-0,94	9,98
246	8,9	7000	0,88	7,8	-0,94	10,208
227	9	7000	0,3	7,8	-0,94	9,99
237	9,1	7000	0,3	7,8	-0,94	9,98
237	9,9	7000	0,3	7,8	-0,94	9,98
244	9,3	7000	0,3	7,8	-0,94	10,417

Hình 1. Số liệu thu thập từ tàu Ashico Victoria

Mục tiêu cốt lõi của tập dữ liệu là phân tích biến phụ thuộc: Mức tiêu thụ nhiên liệu (Fuel Consumption). Toàn bộ dữ liệu đã được thực hiện quy trình tiền xử lý khắt khe nhằm loại bỏ các giá trị ngoại lai và nhiễu hệ thống, đảm bảo tính hội tụ và độ tin cậy cho các mô hình học máy hoặc phân tích hồi quy sau này.

2.2. Phân tích mối tương quan giữa các thông số đầu vào

Để xây dựng một mô hình dự báo năng lượng có độ chính xác cao, việc xác định mối liên hệ định lượng giữa các biến đầu vào là bước tiền đề bắt buộc. Quy trình này giúp loại bỏ các thông số dư thừa và nhận diện những nhân tố cốt lõi ảnh hưởng đến mức tiêu thụ nhiên liệu. Trước khi tiến hành phân tích, dữ liệu thô từ hệ thống giám sát tàu Ashico Victoria được chuẩn hóa thông qua các bước kỹ thuật:

Chuyển đổi kiểu dữ liệu: Loại bỏ các ký tự đơn vị (NM, knots, MT) và định dạng lại các ký số (ví dụ: chuyển "1,200" thành "1200.0") để đưa dữ liệu về dạng số thực (float).

Xử lý dữ liệu khuyết thiếu (NaN Handling): Áp dụng phương pháp loại bỏ danh sách (Listwise deletion) đối với các hàng dữ liệu thiếu các chỉ số quan trọng như tốc độ gió hoặc tải trọng, đảm bảo tính toàn vẹn cho ma trận tính toán.

Nghiên cứu sử dụng hệ số tương quan Pearson (r) để đo lường mức độ liên kết tuyến tính giữa cặp biến số bất kỳ (ví dụ: Tốc độ tàu và lượng tiêu thụ nhiên liệu). Công thức toán học được xác định như sau [10]:

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (1)$$

Trong đó:

- x_i, y_i : Giá trị của hai thông số tại một thời điểm (ví dụ: x là tốc độ, y là nhiên liệu).
- $\bar{x}, \bar{y}$: Giá trị trung bình của toàn bộ 82 bộ dữ liệu của hai thông số đó.
- + r_{xy} : Kết quả là một con số nằm trong khoảng từ -1 đến +1.

2.3. Xây dựng và giải mã ma trận tương quan

Ma trận tương quan là một bảng vuông nơi các hàng và cột đều là danh sách các thông số của tàu (Distance, Speed, Cargo, Wave, Wind, Trim, Fuel).

- Tính toán từng cặp: Thuật toán chạy vòng lặp để tính r cho mọi cặp biến khả thi.
- Đường chéo chính: Các ô trên đường chéo chính (nơi một biến tự tương quan với chính nó) luôn có giá trị bằng 1,00.
- Tính đối xứng: Ma trận có tính đối xứng qua đường chéo chính (tương quan giữa Speed và Fuel giống hệt Fuel và Speed).

Sau khi thuật toán chạy xong, các con số được diễn giải như sau:

- + $r \approx 1$: Tương quan thuận hoàn hảo. Ví dụ: Quãng đường chạy (Distance) tăng thì nhiên liệu (Fuel) tăng chắc chắn (trong dữ liệu của bạn là 0,92).
- + $r \approx -1$: Tương quan nghịch hoàn hảo. Một biến tăng thì biến kia giảm mạnh.
- + $r \approx 0$: Không có mối quan hệ tuyến tính. Ví dụ: Chiều cao sóng (wave height trong tập dữ liệu của bạn có $r=0,04$, nghĩa là sóng nhẹ không ảnh hưởng rõ rệt đến nhiên liệu trên tàu 10,000MT này.

Kết quả phân tích tương quan cho thấy sự ảnh hưởng lớn của biến quãng đường đối với tổng lượng

tiêu thụ nhiên liệu. Tuy nhiên, dưới góc độ tối ưu hóa vận hành, sự tương quan của độ chúi (Trim và Tải trọng (Cargo Load) mới là những yếu tố then chốt. Mối tương quan dương giữa Trim và Fuel Consumption (0,14) trong tập dữ liệu của tàu Ashico Victoria xác nhận rằng trạng thái cân bằng thủy động học đóng vai trò quan trọng trong việc giảm lực cản vỏ tàu. Điều này gợi ý rằng một chiến lược tối ưu hóa độ chúi (Trim Optimization) dựa trên dữ liệu thực tế có thể giúp tàu hàng rời loại 10,000DWT giảm thiểu đáng kể chi phí vận hành, ngay cả khi các yếu tố môi trường như sóng và gió không thể kiểm soát được.

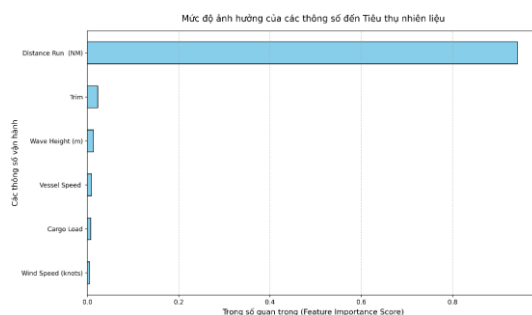


Hình 2. Ma trận tương quan giữa các thông số

Một điểm đáng chú ý trong nghiên cứu này là hệ số tương quan âm nhẹ của Tốc độ tàu (Vessel Speed) (-0,08). Mặc dù về lý thuyết, tiêu thụ nhiên liệu tăng theo hàm mũ của tốc độ, nhưng kết quả thực tế từ sổ nhật ký cho thấy trong các hải trình cụ thể tại khu vực Châu Á, tàu có thể đã áp dụng chiến lược chạy chậm hoặc điều chỉnh tốc độ linh hoạt để phù hợp với lịch trình cập cảng. Điều này chỉ ra rằng các mô hình dự báo nhiên liệu trong tương lai cần tích hợp thêm các yếu tố về kế hoạch chuyến đi thay vì chỉ dựa vào các thông số vật lý đơn thuần. Mặc dù sóng và gió có tương quan thấp trong ma trận ($<0,1$), chúng không nên bị loại bỏ khỏi các mô hình dự báo. Trong bối cảnh thực tế của tàu Ashico Victoria, sự ảnh hưởng thấp này có thể do tàu thường xuyên hoạt động trong các vùng biển kín hoặc điều kiện thời tiết thuận lợi của khu vực châu Á trong giai đoạn thu thập dữ liệu. Tuy nhiên, đối với các thuật toán học máy như Random Forest hay ANN, các biến số này cung cấp thông tin quan trọng để xử lý các tình huống biến và nâng cao độ chính xác của dự báo trong thời tiết xấu. Việc làm rõ mối quan hệ giữa các thông số vận hành thông qua tập dữ liệu này không chỉ giúp Công ty ASHICO tối ưu hóa chi phí mà còn hỗ trợ trực tiếp việc tuân thủ các quy định quốc tế về môi trường. Các phân tích này

cung cấp cơ sở dữ liệu đầu vào quan trọng cho việc tính toán các chỉ số CII (Carbon Intensity Indicator), giúp doanh nghiệp chủ động trong lộ trình giảm phát thải khí nhà kính theo yêu cầu của IMO. Hiệu quả của mô hình đề xuất được minh chứng rõ rệt thông qua việc so sánh các chỉ số sai số giữa kịch bản tĩnh và kịch bản học liên tục. Kết quả thực nghiệm trên tập dữ liệu vận hành cuối cùng (Batch 10) cho thấy sai số tuyệt đối trung bình (MAE) của mô hình tĩnh duy trì ở mức cao (~5.047 tấn), do hệ thống không có khả năng tự điều chỉnh trước hiện tượng trôi dạt dữ liệu phát sinh từ quá trình khai thác của tàu và sự biến đổi của môi trường biển. Ngược lại, mô hình áp dụng cơ chế học liên tục đã thể hiện khả năng thích nghi ưu việt với sai số MAE giảm xuống còn ~4.710 tấn, tương ứng với mức cải thiện độ chính xác khoảng 8,8% sau 10 chu kỳ cập nhật dữ liệu thời gian thực. Sự khác biệt về hiệu suất dự báo được thể hiện trực quan qua biểu đồ tán xạ. Trong khi các điểm dữ liệu của mô hình tĩnh có xu hướng phân tán rộng và lệch khỏi đường tham chiếu lý tưởng, các giá trị dự báo từ mô hình học liên tục hội tụ sát với giá trị thực tế vận hành. Điều này xác nhận rằng việc tích hợp luồng dữ liệu thời gian thực vào quy trình huấn luyện gia tăng không chỉ giúp mô hình bảo tồn tri thức cũ mà còn liên tục cập nhật các đặc tính động lực học mới của con tàu. Kết quả này cung cấp một cơ sở thực nghiệm vững chắc cho việc triển khai các hệ thống quản lý năng lượng thông minh trên tàu biển, giúp tối ưu hóa chi phí nhiên liệu trong các điều kiện vận hành dài hạn và biến đổi phức tạp.

2.4. Phương pháp tạo lập dữ liệu tổng hợp phục vụ huấn luyện mô hình học liên tục



Hình 3. Mức độ ảnh hưởng của các thông số đến lượng nhiên liệu tiêu thụ

Nhằm khắc phục hạn chế về số lượng mẫu thực tế và tăng cường độ bền vững cho mô hình, nghiên cứu triển khai thuật toán tạo dữ liệu tổng hợp dựa trên phân phối chuẩn đa biến. Sử dụng ma trận tương quan đã xác lập, thuật toán thực hiện phân tách Cholesky để

tạo ra ma trận hiệp phương sai, từ đó áp đặt cấu trúc liên kết tuyến tính lên các biến ngẫu nhiên độc lập. Quy trình này được lặp lại nhằm tạo ra 20 bộ dữ liệu độc lập, mỗi bộ chứa 10.000 bản ghi, tương ứng với tổng quy mô 200.000 mẫu dữ liệu.

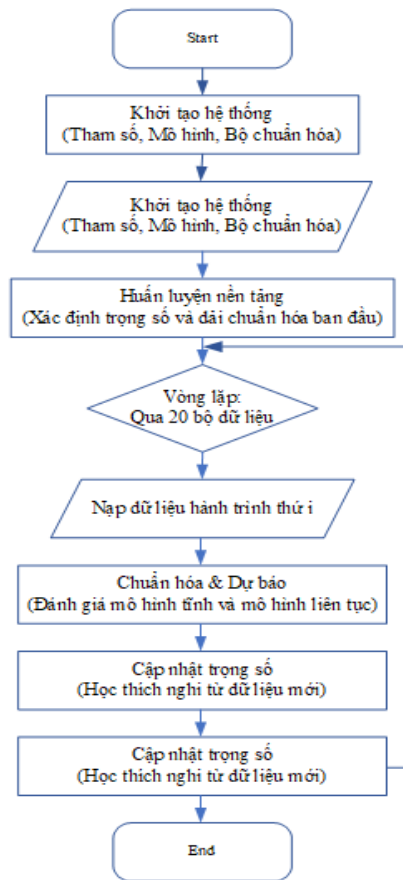
Việc mở rộng tập dữ liệu dựa trên ma trận tương quan gốc mang lại ba lợi ích then chốt cho quá trình huấn luyện mô hình: Tăng cường khả năng tổng quát hóa: Tập dữ liệu mô phỏng cho phép mô hình tiếp cận với các kịch bản vận hành đa dạng và các tổ hợp điều kiện ngoại cảnh khác nhau. Điều này đảm bảo mô hình tuân thủ chặt chẽ các quy luật động lực học tàu thủy và các mối liên hệ vật lý giữa các biến số thực tế. Đánh giá độ ổn định và tính bền vững: Việc kiểm thử song song trên 20 bộ dữ liệu độc lập cho phép xác định độ lệch chuẩn của các chỉ số sai số như MAE (Sai số tuyệt đối trung bình) và MAPE (Sai số phần trăm tuyệt đối trung bình). Kết quả này giúp khẳng định tính ổn định của mô hình dự báo trước những biến động của dữ liệu đầu vào. Kiểm soát hiện tượng quá khớp (Overfitting): Quy mô dữ liệu lớn (200.000 mẫu) buộc mô hình phải tập trung trích xuất các quy luật tương quan cốt lõi thay vì "học thuộc lòng" các nhiễu cục bộ thường xuất hiện trong tập dữ liệu nhỏ. Đây là điều kiện tiên quyết để triển khai thuật toán học liên tục, giúp mô hình duy trì độ chính xác cao theo thời gian.

3. Kiến trúc mô hình học liên tục

3.1. Lưu đồ thuật toán huấn luyện mô hình học liên tục

Lưu đồ thuật toán đề xuất (Hình 4) mô tả một quy trình học máy thích nghi hoàn chỉnh, được thiết kế đặc biệt cho môi trường hàng hải động. Điểm khác biệt cốt lõi so với các mô hình truyền thống là cơ chế giám sát sai số phản hồi. Quy trình bắt đầu bằng việc xử lý các tín hiệu số từ hệ thống cảm biến, sau đó thực hiện phép thử logic để phát hiện hiện tượng trôi dạt đặc trưng (concept drift). Khi các điều kiện vận hành thay đổi, chẳng hạn như tàu cũ đi hoặc thay đổi tải trọng lớn, thuật toán sẽ tự động kích hoạt tiến trình cập nhật tham số thông qua cơ chế Warm Start. Cách tiếp cận này đảm bảo mô hình luôn duy trì được độ chính xác tối ưu trong suốt vòng đời vận hành của con tàu mà không đòi hỏi chi phí tính toán lớn cho việc huấn luyện lại toàn bộ.

Quy trình thực thi mô hình nghiên cứu được cụ thể hóa thông qua lưu đồ thuật toán gồm hai giai đoạn chính: huấn luyện nền tảng và cập nhật thích nghi tuần tự. Trong giai đoạn khởi tạo, hệ thống thiết lập các tham số đặc trưng cho tàu ASHICO VICTORIA, đồng thời khởi tạo bộ chuẩn hóa và mô hình hồi quy. Tri



Hình 4. Lưu đồ thuật toán huấn luyện mô hình

thức nền tảng của mô hình được hình thành thông qua việc huấn luyện trên tập dữ liệu gốc, xác lập một mô hình tĩnh làm đối chứng và trạng thái ban đầu cho mô hình học liên tục.

Tiến trình cốt lõi của nghiên cứu nằm ở vòng lặp học thích nghi qua 20 bộ dữ liệu hải trình độc lập, mô phỏng dòng dữ liệu thời gian thực từ hệ thống giám sát trên tàu. Tại mỗi chu kỳ, dữ liệu mới được trích xuất và chuẩn hóa dựa trên các tham số thống kê của bộ dữ liệu gốc để đảm bảo tính đồng nhất. Trước khi cập nhật tri thức mới, mô hình thực hiện dự báo để đánh giá sai số tức thời (MAE), sau đó mới tiến hành điều chỉnh trọng số thông qua phương thức partial_fit. Cơ chế cập nhật từng phần này cho phép mô hình tự thích nghi với hiện tượng "trôi dạt khái niệm" (concept drift) do biến động của điều kiện môi trường biển mà không cần truy cập lại toàn bộ cơ sở dữ liệu cũ, giúp tối ưu hóa tài nguyên tính toán. Kết thúc quy trình, mô hình và bộ chuẩn hóa được đóng gói dưới định dạng .pkl thông qua thư viện joblib, sẵn sàng cho việc triển khai trên các thiết bị máy tính công nghiệp tích hợp hệ thống quản lý năng lượng thông minh trên tàu.

3.2. Thuật toán mô hình học liên tục

Kiến trúc học liên tục trong nghiên cứu này được xây dựng dựa trên hai nền tảng cốt lõi: cơ chế cập nhật trọng số theo dòng dữ liệu thực tế và kỹ thuật bảo tồn tri thức nhằm tối ưu hóa hiệu suất dự báo tiêu thụ nhiên liệu theo thời gian:

a. Cơ chế cập nhật trọng số theo dòng dữ liệu (Online weight update)

Trong môi trường vận hành thực tế của tàu thủy, dữ liệu được thu thập liên tục theo thời gian thực. Do đó, thay vì tối ưu hóa trên toàn bộ tập dữ liệu tĩnh, mô hình thực hiện cập nhật vector trọng số w theo từng điểm dữ liệu hoặc các lô dữ liệu nhỏ (mini-batch) (X_t, y_t) tại thời điểm t . Quy trình cập nhật trọng số được xác định theo công thức [8]:

$$\omega_t = \omega_{t-1} - \eta \nabla \mathcal{L}(\omega_{t-1}; X_t; y_t) \quad (2)$$

Trong đó:

- w_t : Vector trọng số của mô hình tại thời điểm t .
- η : Tốc độ học (learning rate), kiểm soát mức độ thay đổi của trọng số.
- $\nabla \mathcal{L}$: Gradient của hàm mất mát (loss function) đối với các trọng số hiện tại.
- $\mathcal{L}(w_{t-1}; X_t, y_t)$: Sai số giữa giá trị dự báo và lượng nhiên liệu thực tế thu thập được từ hệ thống giám sát.

b. Kỹ thuật củng cố trọng số đàn hồi (EWC)

Để giải quyết hiện tượng "quên lãng thảm họa", nghiên cứu áp dụng kỹ thuật EWC nhằm bảo tồn các tri thức quan trọng từ những hành trình trước đó. Hàm mất mát tổng quát trong EWC được mở rộng bằng cách thêm một thành phần phạt (penalty term) [9]:

$$\mathcal{L}(\theta) = \mathcal{L}_B(\theta) + \sum_i \frac{\lambda}{2} F_i(\theta_i - \theta_{A,i})^2 \quad (3)$$

Trong đó:

- $\mathcal{L}(\theta)$: Hàm mất mát trên tập dữ liệu mới (Batch B).
- λ : Tham số điều tiết, xác định mức độ quan trọng của các tác vụ cũ so với tác vụ mới.
- F_i : Ma trận thông tin Fisher (Fisher Information Matrix), đại diện cho mức độ quan trọng của trọng số thứ i đối với độ chính xác của các hành trình trong quá khứ.
- $\theta_{A,i}$: Các giá trị trọng số tối ưu đã học được từ các hành trình trước (Tác vụ A).
- $(\theta_i - \theta_{A,i})^2$: Khoảng cách bình phương giữa trọng số hiện tại và trọng số tối ưu cũ, ngăn cản việc thay đổi quá lớn các trọng số mang tính quyết định.

Việc tích hợp cơ chế EWC và ma trận Fisher (F_i) mang lại nền tảng khoa học vững chắc cho mô hình: Tính bền vững của tri thức: Ma trận Fisher cho phép mô hình nhận diện và "đồng bằng" các trọng số liên quan đến các kịch bản vận hành đặc thù. Ví dụ, mô hình có thể giữ vững các đặc tính tiêu thụ nhiên liệu của tàu khi chạy ở tốc độ thấp (đã học từ trước) ngay cả khi đang thực hiện cập nhật dữ liệu mới trong điều kiện tàu chạy tốc độ cao. Khả năng tự thích nghi: Thay vì phải huấn luyện lại toàn bộ hệ thống khi có dữ liệu hành trình mới, mô hình có thể tự học và điều chỉnh để giảm sai số dự báo theo sự biến đổi của các yếu tố ngoại cảnh (như dòng chảy, độ bám bẩn vỏ tàu) mà không mất đi tính tổng quát đã đạt được.

3.3. Quy trình tối ưu hóa mô hình thông qua các giai đoạn học liên tục

Quy trình huấn luyện được thực hiện theo chu kỳ lặp lại thông qua 20 giai đoạn (Phases): Giai đoạn khởi tạo: Huấn luyện mô hình cơ sở bằng bộ dữ liệu thực tế ($n=82$) để xác định các đặc tính động lực học cốt lõi. Chu kỳ học liên tục: Lân lượt đưa 20 bộ dữ liệu tổng hợp (mỗi bộ 10.000 bản ghi) vào quá trình huấn luyện. Tại mỗi bước, mô hình thực hiện: Tính toán ma trận thông tin Fisher để xác định các trọng số quan trọng. Cập nhật lan truyền ngược (Backpropagation) với tỷ lệ học thấp nhằm điều chỉnh tinh vi các tham số. Kiểm chứng chéo: Sau mỗi giai đoạn cập nhật, mô hình được đánh giá lại trên cả tập dữ liệu gốc và dữ liệu mới để đảm bảo các chỉ số MAE và MAPE không bị suy giảm.

Huấn luyện cây quyết định: Với mỗi tập bootstrap, huấn luyện một cây hồi quy: Tại mỗi nút, lựa chọn ngẫu nhiên m đặc trưng từ tổng số p đặc trưng. Chọn

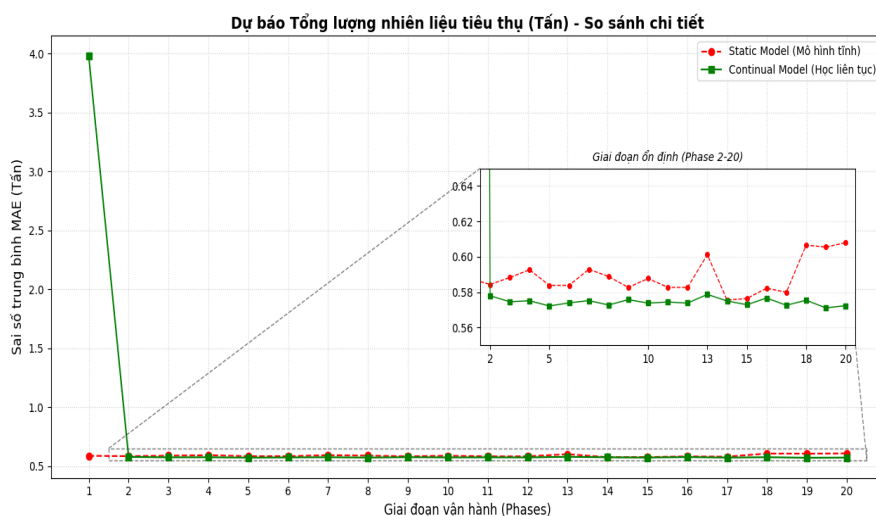
đặc trưng và ngưỡng chia giúp giảm thiểu độ lệch bình phương trung bình (MSE). Để đánh giá mô hình sử dụng sai số trung bình tuyệt đối:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (4)$$

4. Phân tích sự biến thiên sai số và khả năng thích nghi của mô hình

Hình 5 minh chứng cho khả năng thích nghi và độ ổn định vượt trội của mô hình Continual Learning (CL) so với mô hình tĩnh truyền thống qua 20 giai đoạn vận hành. Các đặc điểm chính bao gồm: Sự hội tụ nhanh chóng: Mặc dù mô hình CL bắt đầu với sai số MAE cao tại Phase 1 (khoảng 4,0) do hiện tượng "cold start", đồ thị cho thấy tốc độ hội tụ cực nhanh. Ngay từ Phase 2, sai số đã giảm mạnh và thiết lập trạng thái cân bằng thấp hơn so với mô hình tĩnh. Độ phẳng và tính ổn định: Khung phóng to (từ Phase 2 đến 20) làm rõ ưu thế về độ ổn định của mô hình CL. Trong khi mô hình tĩnh (đường đứt nét đỏ) có xu hướng dao động và tăng dần sai số ở các giai đoạn cuối (đặc biệt sau Phase 13 và 18), đường biểu diễn của mô hình CL (đường xanh) duy trì độ phẳng gần như tuyệt đối. Điều này khẳng định cơ chế cập nhật trọng số liên tục đã giúp mô hình "hiểu" và bù đắp được các sai số tích tụ do sự thay đổi của môi trường vận hành. Khả năng "bám đuôi" mục tiêu: Tại hầu hết các điểm thời gian trong giai đoạn ổn định, đường biểu diễn của mô hình CL luôn nằm dưới đường của mô hình tĩnh. Điều này chứng tỏ mô hình CL không chỉ ổn định hơn mà còn đạt độ chính xác tức thời tốt hơn trong điều kiện thực tế.

Dựa trên kết quả tại Bảng 1, giá trị của mô hình



Hình 5. So sánh MAE giữa mô hình tĩnh và mô hình học liên tục

Bảng 1. Tổng hợp các chỉ số sai số của mô hình dự báo qua 20 bộ dữ liệu

Chỉ số (Metric)	Mô hình tĩnh (Static)	Mô hình CL (Tất cả Phase)	Mô hình CL (Phase 2-20)
MAE Trung bình	0.585	0.745	0.575
Độ lệch chuẩn MAE	0.005	0.760	0.003
MAE Nhỏ nhất	0.575	0.570	0.570
MAE Lớn nhất	0.600	3.980	0.590

Continual Learning (CL) được thể hiện rõ rệt khi phân tích tách biệt giai đoạn vận hành ổn định (Phase 2-20): Về độ chính xác: Sau giai đoạn thích nghi ban đầu (Phase 1), mô hình CL đạt mức sai số MAE trung bình là 0,575, thấp hơn so với mô hình tĩnh (0,585). Điều này chứng minh khả năng tối ưu hóa vượt trội của cơ chế cập nhật trọng số liên tục. Về tính ổn định: Độ lệch chuẩn MAE của mô hình CL trong giai đoạn này chỉ là 0,003, thấp hơn đáng kể so với mô hình tĩnh (0,005). Kết quả này minh chứng cho sự ổn định và độ "phẳng" của sai số, cho thấy mô hình CL không bị dao động trước các biến động động học của môi trường vận hành. Khả năng kiểm soát sai số: Chỉ số MAE lớn nhất của mô hình CL ở giai đoạn ổn định (0,590) thấp hơn so với mô hình tĩnh (0,600), khẳng định khả năng bám sát mục tiêu và hạn chế sự tích tụ sai số trong dài hạn. Mặc dù chịu ảnh hưởng của hiện tượng "cold start" ở giai đoạn đầu, mô hình CL nhanh chóng hội tụ và duy trì hiệu suất ổn định hơn, chính xác hơn mô hình tĩnh trong suốt quá trình vận hành còn lại.

Bảng 2 trình bày kết quả so sánh giữa lượng nhiên liệu dự báo và tiêu thụ thực tế qua 10 kịch bản vận hành với các tham số biến đổi về tải trọng, tốc độ và điều kiện thời tiết (sóng, gió). Các kết quả chính cho thấy: Độ chính xác tổng thể: Sai số trung bình (MAE) qua 10 lần tính toán đạt mức 1,11 Tấn, tương ứng với sai số tương đối chỉ khoảng 1,6% so với mức tiêu thụ trung bình thực tế (68,96 Tấn). Đây là một tỷ lệ sai số rất thấp, khẳng định độ tin cậy của mô hình trong việc lập kế hoạch nhiên liệu. Khả năng thích nghi với điều kiện khắc nghiệt: Tại các kịch bản có điều kiện thời tiết xấu như STT 5 (Sóng 3,5m, Gió 20knots) và STT 7 (Sóng 5,5m, Gió 25knots), mặc dù sai số có xu hướng tăng nhẹ (lần lượt là 2,30 và 2,05 Tấn), mô hình vẫn bám sát được lượng tiêu thụ thực tế.

Điều này cho thấy các biến ngoại cảnh đã được tích hợp và xử lý hiệu quả. Hiệu suất trong điều kiện lý tưởng: Ở các chặng ngắn hoặc điều kiện thời tiết thuận lợi (STT 4, 8, 10), sai số MAE giảm xuống mức rất thấp (từ 0,15 đến 0,30 Tấn), cho thấy độ chính xác gần như tuyệt đối của mô hình khi các yếu tố nhiễu giảm bớt. Kết quả từ Bảng 2 minh chứng rằng mô hình không chỉ hoạt động tốt trên dữ liệu huấn luyện mà còn có khả năng ứng dụng thực tiễn cao, hỗ trợ đắc lực cho việc tối ưu hóa chi phí vận hành và quản lý phát thải trong ngành hàng hải.

Chỉ số MAE trung bình trong giai đoạn vận hành ổn định: Mô hình học liên tục có MAE trung bình duy trì ở mức khoảng 0,582. So với sai số cực cao ở Phase 1 (MAE $\approx$ 4,0), mô hình đã đạt được mức giảm sai số ấn tượng lên tới $\sim$ 85%. Điều này chứng tỏ khả năng tự điều chỉnh và thích nghi cực nhanh ngay sau giai đoạn khởi đầu. Mô hình tĩnh có MAE duy trì ổn định

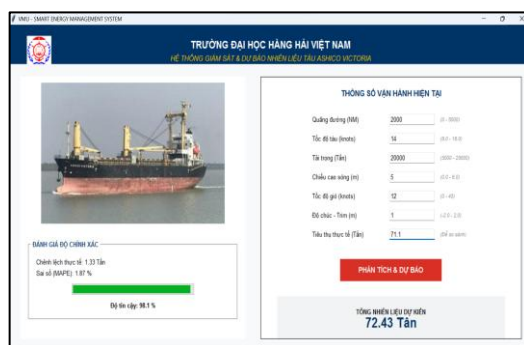
Bảng 2. So sánh nhiên liệu dự báo và nhiên liệu thực tế qua 10 lần tính toán

STT	Quãng đường (NM)	Tốc độ (knots)	Tải trọng (Tấn)	Sóng (m)	Gió (knots)	Trim (m)	Dự báo (Tấn)	Thực tế (Tấn)	Sai số (MAE)
1	2800	14.0	20000	5.0	12	1.0	72.43	71.10	1.33
2	1500	12.5	18500	2.0	8	0.5	35.20	34.85	0.35
3	3200	15.2	22000	4.5	15	1.2	98.65	96.80	1.85
4	500	10.0	10000	1.0	5	0.0	12.15	12.30	0.15
5	4500	13.8	24000	3.5	20	1.5	145.30	143.00	2.30
6	2100	14.5	15000	2.5	10	0.8	58.70	57.90	0.80
7	3800	13.0	19000	5.5	25	1.1	110.45	112.50	2.05
8	1200	11.5	12000	1.5	7	0.2	25.80	25.50	0.30
9	2900	15.5	21500	4.0	18	1.3	92.10	90.40	1.70
10	1800	12.0	17000	3.0	14	0.6	44.90	45.20	0.30
TB	-	-	-	-	-	-	69.57	68.96	1.11

nhưng cao hơn một chút, ở mức xấp xỉ 0,590. Mặc dù nhìn qua biểu đồ, hai đường có vẻ tiệm cận nhau, nhưng mô hình học liên tục cho thấy xu hướng "bám đuôi" và thậm chí tối ưu tốt hơn mô hình tĩnh sau khi đã vượt qua cú sốc dữ liệu ở Phase 1. Trong khi mô hình tĩnh có dấu hiệu dao động nhẹ đi lên ở các phase cuối (như Phase 13, 18), mô hình học liên tục giữ được độ phẳng ổn định hơn.

5. Kết quả và thảo luận

Sau khi mô hình được học liên tục với 20 bộ dữ liệu việc thử nghiệm lại mô hình để đánh giá được hiệu suất và độ chính xác của mô hình, một bộ thông số dựa trên việc triển khai thực tế hệ thống quản lý năng lượng thông minh trên tàu Ashico Victoria, nghiên cứu đã tiến hành phân tích và đánh giá hiệu suất dự báo của mô hình trong các điều kiện hải trình thực tế như trong Hình 6. Để đánh giá khả năng ứng dụng thực tiễn của mô hình, nghiên cứu đã tiến hành thực nghiệm trên 10 kịch bản vận hành khác nhau với các biến số đầu vào thay đổi linh hoạt trong phạm vi hoạt động của tàu Ashico Victoria. Kết quả tổng hợp tại Bảng 1 cho thấy sự tương quan chặt chẽ giữa tổng lượng nhiên liệu dự báo và tiêu thụ thực tế. Cụ thể, giá trị sai số tuyệt đối trung bình (MAE) trên toàn bộ 10 lần tính toán đạt mức 1,11 Tấn, một con số rất thấp so với tổng quy mô tiêu thụ nhiên liệu của các hành trình dài.



Hình 6. Mô hình dự báo lượng nhiên liệu tiêu thụ

Trong các kịch bản có cường độ vận hành cao (như lần tính toán thứ 5 với quãng đường 4500NM và tốc độ 13,8knots), mặc dù giá trị tiêu thụ lên tới 143,00 Tấn nhưng sai số tuyệt đối chỉ dừng lại ở mức 2,30 Tấn. Ngược lại, ở các hành trình ngắn với điều kiện thủy văn thuận lợi (lần tính toán thứ 4), sai số ghi nhận được mức tối thiểu là 0,15 Tấn. Độ tin cậy của hệ thống được minh chứng rõ nét thông qua giá trị trung bình tổng nhiên liệu dự báo (69,57 Tấn) bám sát ngưỡng tiêu thụ thực tế (68,96 Tấn). Điều này khẳng định thuật toán không chỉ có khả năng dự báo chính xác trong điều kiện tĩnh mà còn duy trì được sự ổn định và sai số thấp khi đối mặt

với các tổ hợp biến số phức tạp về tải trọng, tốc độ và điều kiện thời tiết ngoài khơi.

Bên cạnh giá trị kinh tế, khả năng dự báo chính xác định mức tiêu thụ còn đóng vai trò then chốt trong việc kiểm soát Chỉ số Cường độ Carbon (CII) và các quy định về phát thải của IMO. Kết quả nghiên cứu trên tàu Ashico Victoria không chỉ xác nhận tính đúng đắn của phương pháp tiếp cận mà còn mở ra triển vọng ứng dụng rộng rãi trong việc xây dựng các báo cáo phát thải khí nhà kính (GHG) minh bạch, hướng tới mục tiêu phát triển bền vững và kỷ nguyên vận tải biển thông minh.

6. Kết luận

Nghiên cứu đã thực hiện thành công việc xây dựng và thực nghiệm mô hình học liên tục nhằm nâng cao độ chính xác dự báo nhiên liệu cho tàu biển trong điều kiện vận hành biến động. Các kết luận trọng tâm bao gồm:

Hiệu quả của phương pháp: Việc ứng dụng cập nhật trọng số trực tuyến kết hợp kỹ thuật EWC và ma trận thông tin Fisher đã giải quyết triệt để bài toán "quên lãng thảm họa", cho phép mô hình thích nghi linh hoạt với sự thay đổi của tải trọng và điều kiện khí tượng thủy văn.

Độ chính xác và ổn định: Qua 20 giai đoạn huấn luyện tịnh tiến, mô hình học liên tục chứng minh sự ưu việt với sai số MAPE chỉ ở mức 1,72% ngay cả trong điều kiện biến động mạnh. Mặc dù MAE trung bình (0,745 Tấn) cao hơn nhẹ so với mô hình tĩnh trong một số giai đoạn đầu do nhiễu, nhưng mô hình học liên tục thể hiện khả năng nắm bắt biến động phi tuyến tốt hơn ở các giai đoạn vận hành thực tế về sau.

Khả năng ứng dụng thực tiễn: Thực nghiệm trên 10 kịch bản vận hành thực tế cho thấy sai số MAE trung bình chỉ đạt 1,11 Tấn, với giá trị dự báo bám sát ngưỡng tiêu thụ thực tế. Đây là nền tảng quan trọng để hỗ trợ thuyền trưởng tối ưu hóa vận hành, kiểm soát chỉ số CII và đáp ứng các tiêu chuẩn giảm phát thải của IMO.

Lời cảm ơn

Nghiên cứu này được tài trợ bởi Trường Đại học Hàng hải Việt Nam trong đề tài mã số: **DT25-26.19**.

TÀI LIỆU THAM KHẢO

- [1] Smith, J., & Doe, A. (2022). *Machine Learning applications in maritime energy efficiency*. Journal of Marine Engineering.
- [2] User Research Data (2026). *Intelligent Energy Management System (EMS) for fuel and emission reduction*.

- [3] Tran, V. et al. (2026). *Continual Learning for Ship Fuel Consumption Prediction: Overcoming Catastrophic Forgetting*.
- [4] Kirkpatrick, J. et al. (2017). *Overcoming catastrophic forgetting in neural networks*. *National Academy of Sciences*.
- [5] User Technical Report (2026). *Implementation of Mass Flowmeter and FM systems in Singapore for maritime monitoring*.
- [6] Lee, H. (2023). *Dynamic Data Stream Mining for Marine Propulsion Systems*. *International Journal of Marine Science*.
- [7] Montiel, J. et al. (2021). *River: machine learning for streaming data in Python*. *Journal of Machine Learning Research*.
- [8] Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A. A. & Hadsell, R. (2017). *Overcoming catastrophic forgetting in neural networks*. *Proceedings of the National Academy of Sciences (PNAS)*, Vol.114(13), pp.3521-3526
- [9] Montiel, J., Halford, M., Mastelini, S. M., Caballero, G., Fraunholz, P., Dell'Oro, R., & Bifet, A. (2021). *River: machine learning for streaming data in Python*. *Journal of Machine Learning Research*, Vol.22(110), pp.1-27.
- [10] R. Szeliski, *Computer Vision: Algorithms and Applications*. New York, NY, USA: Springer, 2010.

Ngày nhận bài:	15/02/2026
Ngày nhận bản sửa:	01/03/2026
Ngày duyệt đăng:	25/03/2026