

Thiết kế điều khiển hệ truyền động xoay chiều với kỹ thuật học tăng cường

Trần Ngọc Sơn^{1*}, Lại Thị Thanh Hoa², Đỗ Trung Hải², Lại Khắc Lãi², Lê Thị Thu Hà²

¹Trường Đại học Kinh tế - Kỹ thuật Công nghiệp, 456 Minh Khai, Hai Bà Trưng, Hà Nội, Việt Nam;

²Trường Đại học Kỹ thuật Công nghiệp - Đại học Thái Nguyên, 666 Đường 3-2, P.Tích Lương, TP Thái Nguyên, Tỉnh Thái Nguyên, Việt Nam.

*Email: tnsn@uneti.edu.vn

Nhận bài: 01/4/2024; Hoàn thiện: 06/6/2024; Chấp nhận đăng: 29/7/2024; Xuất bản: 25/8/2024.

DOI: <https://doi.org/10.54939/1859-1043.j.mst.97.2024.33-40>

TÓM TẮT

Máy điện không đồng bộ 3 pha (IM) được biết đến rộng rãi trong các ứng dụng công nghiệp do chi phí thấp và giảm thiểu việc bảo trì. Điều khiển máy điện xoay chiều dựa trên điều khiển tựa từ thông (FOC) và điều khiển véc tơ với qui luật điều khiển kinh điển PID là những kỹ thuật tin cậy được ưa dùng trong công nghiệp. Tuy nhiên, do bản chất phi tuyến của máy điện và khi chúng chịu tác động của nhiễu bên ngoài hoặc thay đổi tham số bên trong thì các bộ điều khiển thông thường này khó có thể đáp ứng được các yêu cầu điều khiển. Học tăng cường (RL) là một kỹ thuật học trực tuyến, không cần mô hình, có thể xử lý các biến đổi của tham số. Những đặc tính này làm cho việc học tăng cường trở thành một ứng cử viên tiềm năng, hoạt động như một bộ điều khiển thích nghi có thể thay thế các bộ điều khiển thông thường. Bài báo này đề xuất việc thiết kế điều khiển hệ truyền động động cơ không đồng bộ 3 pha dựa trên kỹ thuật học tăng cường. Bộ điều khiển đề xuất sử dụng một tác nhân học tăng cường với thuật toán DDPG thay thế cho các bộ điều khiển PI của mạch vòng dòng điện trong điều khiển FOC thông thường. Hiệu suất của bộ điều khiển đã được xác minh trong các điều kiện hoạt động khác nhau bằng mô phỏng máy tính trong MATLAB/SIMULINK.

Từ khóa: Học tăng cường; Actor-Critic; Máy điện không đồng bộ; Tác nhân; DDPG

1. MỞ ĐẦU

Máy điện không đồng bộ 3 pha với các ưu điểm như kết cấu chắc chắn, giá thành thấp, chi phí bảo dưỡng và vận hành nhỏ nên đã được sử dụng rộng rãi trong các ứng dụng công nghiệp. Tuy nhiên, do đặc tính đa biến và động lực học phi tuyến nên việc điều khiển tốc độ hoặc mô-men xoắn máy điện khá khó khăn. Cho đến nay đã có khá nhiều chiến lược điều khiển được trình bày để điều khiển tốc độ, mô men của máy điện không đồng bộ, dựa trên 2 nguyên lý cơ bản là điều khiển tựa từ thông (FOC) và điều khiển trực tiếp mô men máy điện (DTC) [7]. Các bộ điều khiển PID thông thường được trình bày trong [7] có ưu điểm là đơn giản trong thiết kế nhưng chưa đề cập đến sự không chắc chắn của các tham số và tính phi tuyến của hệ thống, khi thiết kế điều khiển cần phải biết các tham số của hệ thống. Các chiến lược điều khiển phi tuyến xuất hiện trong vài thập kỷ qua như backstepping [5, 6] khắc phục tính phi tuyến của hệ, điều khiển trượt [9-11] đã khắc phục một phần tính phi tuyến và sự không chắc chắn tham số hệ thống, song khi thiết kế điều khiển vẫn cần có đầy đủ các tham số của đối tượng.

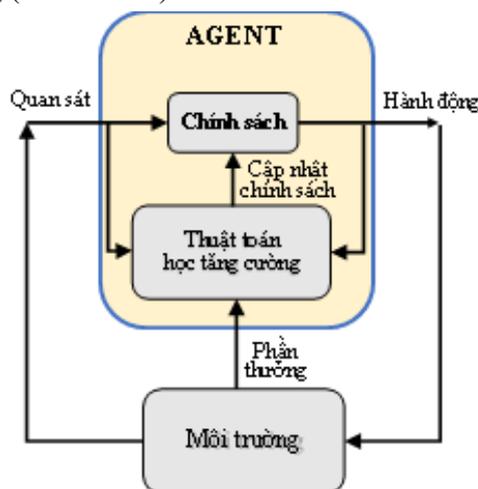
Mục tiêu thiết kế các bộ điều khiển hiện nay là chúng phải có tính thích nghi và tối ưu với sự biến đổi tham số của đối tượng và nhiễu. Chiến lược điều khiển thích nghi có thể điều khiển trực tuyến các hệ thống chưa biết thông qua quan sát dữ liệu để đưa ra thuật toán thích nghi phù hợp [6]. Trong khi đó, các bộ điều khiển tối ưu yêu cầu thông tin đầy đủ về hệ thống để bộ điều khiển giải phương trình Hamilton Jacobi bellman ngoại tuyến, điều này trở nên rất khó khăn đối với các hệ thống phi tuyến tính cao, có tham số không chắc chắn. Những mục tiêu trên có thể đáp ứng được nhờ sử dụng mạng nơ ron sâu với thuật toán học tăng cường, với ý tưởng tạo ra một bộ điều khiển phản hồi mới bằng cách hợp nhất các bộ điều khiển thích nghi và bộ điều khiển tối ưu, nghĩa là tạo ra một bộ điều khiển thích nghi có khả năng hội tụ các giải pháp tối ưu trong thời gian thực (trực tuyến) [4].

Đã có một số nghiên cứu sử dụng học tăng cường với mạng nơ ron sâu để điều khiển hệ truyền động xoay chiều 3 pha, cụ thể các tác giả trong [1-5] đề xuất học tăng cường điều khiển tốc độ máy điện đồng bộ, trong [2] đề xuất thuật toán học tăng cường cho vòng điều khiển tốc độ máy điện không đồng bộ, hai vòng điều khiển dòng điện vẫn sử dụng PI kinh điển, trong đó, các tác giả đã sử dụng Tác nhân (Agent) TD3 (Twin Delayed Deep Deterministic Policy Gradient) cập nhật các thuộc tính của nhà Critic và Actor. Phát triển hướng nghiên cứu của [2] trong bài báo này xây dựng thuật toán học tăng cường cho 2 mạch vòng điều khiển dòng điện với Tác nhân DDPG (Deep Deterministic Policy Gradient). Cấu trúc của bài báo gồm phần giới thiệu, phần 2 trình bày tổng quan về học tăng cường với tác nhân DDPG, phần 3 xây dựng thuật toán học tăng cường cho các mạch vòng dòng điện máy điện không đồng bộ, phần 3 trình bày kết quả mô phỏng với các điều kiện vận hành khác nhau có so sánh với kết quả điều khiển sử dụng PI thông thường trong cùng một điều kiện làm việc, cuối cùng là kết luận và kiến nghị.

2. HỌC TĂNG CƯỜNG VỚI THUẬT TOÁN DDPG

2.1. Tổng quan

Học tăng cường (RL- Reinforcement Learning) là một kỹ thuật tính toán thông minh giúp giải quyết các bài toán điều khiển phức tạp hoạt động trong môi trường động. Ý tưởng chính của RL là đánh giá chuỗi hành động tốt nhất mang lại kết quả tối ưu [5]. Sơ đồ khối một hệ thống điều khiển với kỹ thuật học tăng cường được mô tả trên hình 1, gồm 2 khối chính: Khối Tác nhân (Agent) và khối Môi trường (Environment).



Hình 1. Sơ đồ hệ điều khiển với kỹ thuật học tăng cường.

Tác nhân thực hiện các hành động (action) tác động đến Môi trường làm thay đổi trạng thái của môi trường và nó nhận phần thưởng từ môi trường cho hành động cụ thể đó. Ngõ vào của Tác nhân là kết quả quan sát trạng thái và phần thưởng nhận được từ môi trường, ngõ ra của Tác nhân là các hành động tác động lên môi trường. Tác nhân chứa 2 thành phần là chính sách (Policy) và thuật toán học tăng cường (Reinforcement Learning Algorithm). Mạng nơ ron sâu được chọn để đại diện cho chính sách [6]. Chính sách đưa ra các hành động thường xuyên được cập nhật thông qua thuật toán RL mà tác nhân này thực hiện là đầu vào và hành động dựa trên những quan sát này là đầu ra.

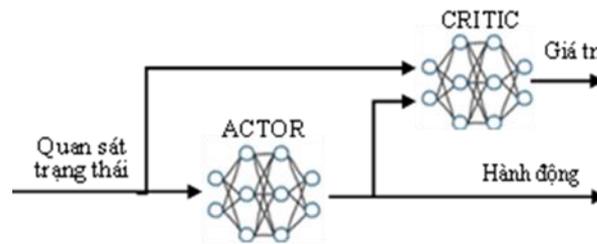
Phần thưởng là một hàm của trạng thái và hành động, chúng tạo ra một số vô hướng đại diện cho "lòng tốt" của một tác nhân ở trong một trạng thái cụ thể và một hành động cụ thể. Khái niệm phần thưởng trong RL khá giống với khái niệm hàm chi phí trong bộ điều khiển tối ưu LQR, điểm khác nhau giữa chúng là: Hàm chi phí trong LQR là hàm bậc hai tuyến tính trong khi đó, hàm khen thưởng trong học tăng cường có thể là hàm bất kỳ, theo đúng nghĩa đen, miễn là nó đào tạo tác nhân một cách hoàn hảo như nhà thiết kế mong muốn, hàm chi phí cố gắng giảm thiểu giá trị, trong

khi một hàm khen thưởng thường cố gắng tối đa hóa giá trị, trong thiết kế học tăng cường phần thưởng có thể được coi là phủ định của chi phí.

2.2. Thuật toán DDPG

Tác nhân chứa 2 thành phần là “chính sách” và thuật toán học tăng cường (hình 2). Chính sách (Policy) thường được hiểu là một hàm hoặc một chiến lược quyết định mà một tác nhân sử dụng để đưa ra quyết định về hành động nên được thực hiện tại mỗi bước thời gian dựa trên trạng thái hiện tại của môi trường. Có nhiều loại chính sách [9], mỗi loại chính sách có đặc điểm riêng, việc lựa chọn chính sách phù hợp với bài toán cụ thể sẽ giảm bớt đáng kể thời gian đào tạo. Tuy nhiên, điều này đòi hỏi sự hiểu biết về bản chất của môi trường và mục tiêu của hệ thống. Mạng nơ-ron sâu được chọn để đại diện cho chính sách [6].

Thuật toán học tăng cường là một lĩnh vực trong học máy tập trung vào việc làm cho các tác nhân (agents) tự động học thông qua tương tác với một môi trường. Mục tiêu của RL là tối ưu hóa một hàm phần thưởng (reward function) thông qua các quyết định hoặc hành động mà tác nhân thực hiện trong môi trường.



Hình 2. Cấu trúc thuật toán RL Actor – Critic.

Thuật toán Gradient chính sách xác định sâu (DDPG) là một thuật toán học tăng cường được sử dụng để học chính sách tối ưu cho các bài toán liên tục, đặc biệt là trong các bài toán mà hành động trong không gian liên tục. DDPG kết hợp ý tưởng của thuật toán học tăng cường sâu và thuật toán tối ưu hóa “chính sách”. Nó sử dụng một hàm giá trị hành động (action-value function) để đánh giá giá trị của các hành động và một mô hình “chính sách” (policy model) để chọn hành động dựa trên các trạng thái hiện tại. Trong DDPG, cả hai mô hình này đều được triển khai bằng mạng nơ-ron học sâu. Cấu trúc thuật toán DDPG sử dụng một cặp mạng nơ-ron, mạng Actor và mạng Critic được biểu diễn trên hình 2.

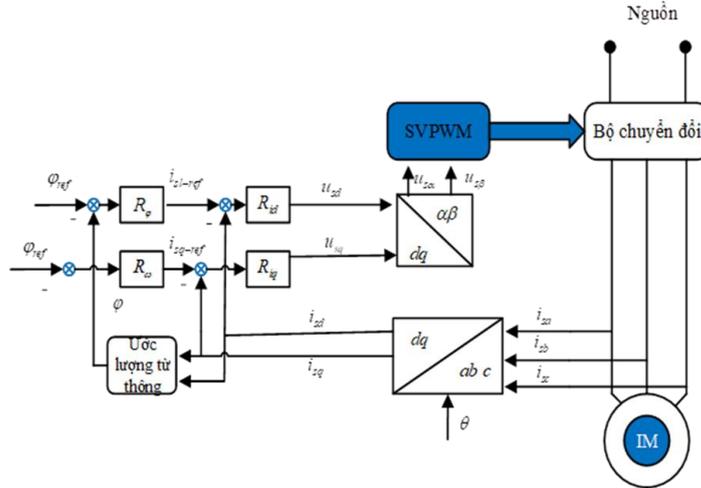
3. TÁC NHÂN HỌC TĂNG CƯỜNG CHO HỆ TRUYỀN ĐỘNG XOAY CHIỀU 3 PHA

Phần này trình bày kết quả áp dụng học tăng cường để điều khiển tốc độ và từ thông máy điện không đồng bộ 3 pha xuất phát từ hệ thống điều khiển tựa từ thông (FOC) truyền thống. Mô hình toán học máy điện không đồng bộ 3 pha biểu diễn trong không gian trạng thái trong hệ tọa độ quay (d,q) được mô tả bởi [12].

$$\frac{d}{dt} \begin{bmatrix} \phi_{dr} \\ \phi_{qr} \\ i_{ds} \\ i_{qs} \end{bmatrix} = \begin{bmatrix} \frac{-R_r}{L_r} & (\omega_s - p\omega_F) & \frac{MR_r}{L_r} & 0 \\ (\omega_s - p\omega_F) & \frac{-R_r}{L_r} & 0 & \frac{MR_r}{L_r} \\ \frac{MR_r}{\sigma L_s L_r} & \frac{Mp\omega_F}{\sigma L_s L_r} & \frac{-R_{sr}}{\sigma L_s} & \omega_s \\ \frac{-Mp\omega_F}{\sigma L_s L_r} & \frac{MR_r}{\sigma L_s L_r} & -\omega_s & \frac{-R_{sr}}{\sigma L_s} \end{bmatrix} \begin{bmatrix} \phi_{dr} \\ \phi_{qr} \\ i_{ds} \\ i_{qs} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ \frac{1}{\sigma L_s} & 0 \\ 0 & \frac{1}{\sigma L_s} \end{bmatrix} \begin{bmatrix} u_{ds} \\ u_{qs} \end{bmatrix} \quad (1)$$

Với: $R_{sr} = R_s + \frac{M^2}{L_r^2} R_r$ và $\sigma = 1 - \frac{M^2}{L_s L_r}$

Trong đó: R_s, R_r là điện trở pha stator và rotor; L_s, L_r là điện cảm pha stator và rotor; M là hệ số cảm; u_{ds}, u_{qs} là các thành phần vuông góc của điện áp stator; i_{ds}, i_{qs} là các thành phần vuông góc của dòng điện stator; ϕ_{dr}, ϕ_{qr} là các thành phần vuông góc của từ thông rotor; p là số đôi cực; ω_s là tốc độ quay của từ trường stator.



Hình 3. Điều khiển truyền thống IM.

Sơ đồ khối hệ điều khiển FOC thông thường của máy điện không đồng bộ được trình bày trên hình 3. Ở đây, các bộ điều khiển PI được sử dụng cho cả 2 vòng điều khiển dòng điện (vòng trong) và vòng điều khiển tốc độ, từ thông (vòng ngoài). Hai bộ điều khiển PI vòng ngoài điều khiển tốc độ và từ thông tạo ra các dòng điện tham chiếu i_{sd_ref} và i_{sq_ref} , hai vòng điều khiển PI vòng trong điều khiển dòng điện trục d và trục q. Trong bài báo này, một khối tác nhân học tăng cường (AGENT) với thuật toán DDPG được sử dụng để thay thế cho các bộ điều khiển PI của mạch vòng dòng điện bên trong, sơ đồ được chỉ ra như hình 5. Việc sử dụng tác nhân RL đặc biệt có lợi khi hệ thống phi tuyến, trong trường hợp đó chúng ta có thể huấn luyện một tác nhân RL duy nhất thay vì điều chỉnh bộ điều khiển PI ở nhiều điều kiện vận hành.

Bước tiếp theo trong thiết kế học tăng cường cho vòng dòng điện là cần xác định môi trường, phần thưởng và tác nhân.

3.1. Môi trường

Trong hệ thống này, môi trường là toàn bộ hệ thống điều khiển máy điện không đồng bộ không bao gồm bộ điều khiển dòng điện vòng trong là tác nhân RL (hình 5). Đối với môi trường này, các quan sát đối với khối học tăng cường là các sai số dòng điện stator và dòng điện rotor ($e_{id} = i_{sd_ref} - i_{sd}$, $e_{iq} = i_{sq_ref} - i_{sq}$, i_{sd} , i_{sq}), tác động là các tín hiệu 'u_{sd}' và 'u_{sq}' đưa đến khối điều chế độ rộng xung SVPWM.

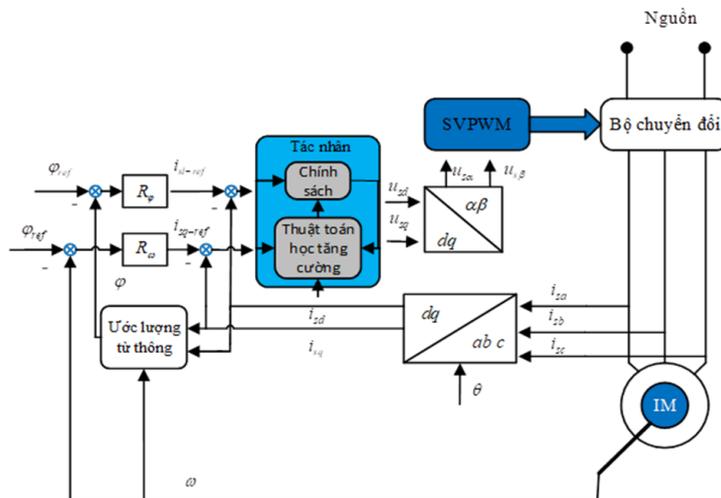
3.2. Phần thưởng

Hàm khen thưởng được thiết kế để đào tạo tác nhân học tăng cường, giúp cho tác nhân học tăng cường biết hành động mà nó chọn trong quá trình đào tạo là tốt hay xấu, dựa trên sự tương tác của nó với môi trường. Hàm khen thưởng kết quả về cơ bản là tổng của các hình phạt có trọng số được áp dụng bởi tình trạng môi trường hiện tại đối với các hành vi vi phạm ràng buộc. Trong bài báo này, chúng tôi xác định hàm phần thưởng dựa trên các mục tiêu bám và nỗ lực điều khiển.

$$r_t = -(\lambda_1 * e_{id}^2 + \lambda_2 * e_{iq}^2 + \lambda_3 u_{t-1}^2) = -(1 * e_{id}^2 + 1 * e_{iq}^2 + 0,01 u_{t-1}^2) \tag{2}$$

Trong đó: e_{id} là sai số dòng điện trục d; e_{iq} là sai số dòng điện trục q; u_{l-1} là tín hiệu điều khiển ở bước trước. Các trọng số được chọn $\lambda_1 = \lambda_2 = 1$; $\lambda_3 = 0,01$.

3.3. Tác nhân



Hình 4. Tác nhân học tăng cường cho máy điện không đồng bộ.

Tác nhân được sử dụng trong bài báo này là tác nhân Gradient chính sách xác định sâu DDPG [9]. Ở đây, mạng Actor - Critic theo yêu cầu của thuật toán DDPG như sau: Mạng Critic lấy các quan sát và hành động làm đầu vào và đưa ra các giá trị Q ước tính ở đầu ra, mạng Actor lấy các quan sát làm đầu vào và đưa ra các hành động ở đầu ra (hình 4). Thời gian lấy mẫu của tác nhân DDPG phù hợp với yêu cầu của vòng điều khiển, tuy nhiên, không nên chọn thời gian lấy mẫu quá nhỏ sẽ tốn nhiều thời gian huấn luyện.

Đối với các quan sát và hành động nhất định, tác nhân này ước tính phần thưởng dài hạn bằng cách sử dụng Actor - Critic. Thuật toán này là một phương pháp phê bình tác nhân, trực tuyến, không có mô hình, không có chính sách, tính toán chính sách tối ưu gần đúng với phần thưởng dài hạn. Đối với thuật toán này, không gian hành động được thực hiện liên tục. Tác nhân DDPG cập nhật các thuộc tính của Critic và Actor trong quá trình đào tạo ở mỗi bước thời gian.

Việc huấn luyện tác nhân được thực hiện trong Matlab-Simulink với hộp công cụ “Reinforcement Learning toolbox, số kỳ huấn luyện là 1500 và dừng huấn luyện khi giá trị trung bình của phần thưởng vượt quá giá trị đã cho.

4. KẾT QUẢ MÔ PHỎNG

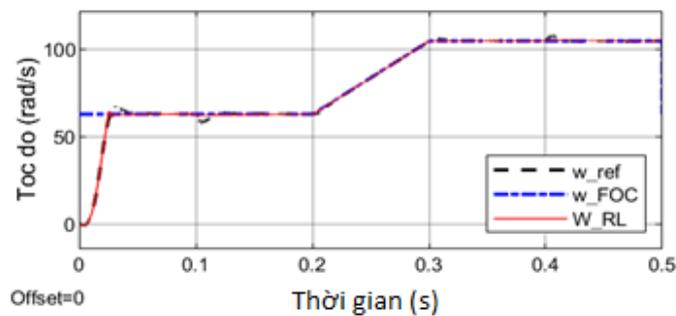
Mô phỏng để kiểm tra hiệu năng của thuật toán được thực hiện trong phần mềm Matlab-Simulink ứng với các điều kiện vận hành khác nhau cho cả sơ đồ sử dụng bộ điều khiển học tăng cường và sơ đồ điều khiển FOC thông thường, các thông số mô phỏng như sau:

| | |
|-------------------------------|------------|
| Công suất máy điện (kW) | : 7,5 |
| Điện áp (V) | : 400 |
| Số pha /số đôi cực | : 3/2 |
| Tần số (Hz) | : 50 |
| Điện trở Stator $R_s(\Omega)$ | : 0,7384 |
| Điện trở rotor $R_r(H)$ | : 0,7402 |
| Điện cảm stator $L_s(H)$ | : 0,003045 |
| Điện cảm rotor $L_r(H)$ | : 0,003045 |

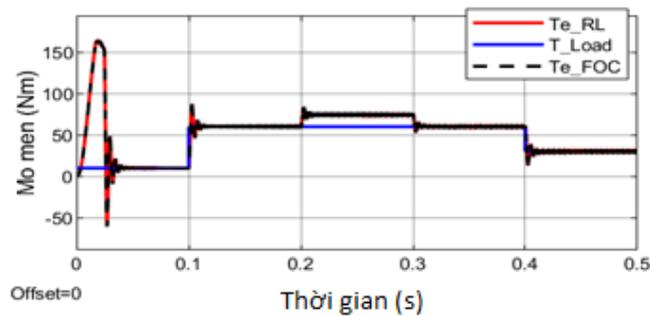
| | |
|---|------------|
| Hỗ cảm L_m (h) | : 0,1241 |
| Mô men quán tính J (Nm ²) | : 0,0343 |
| PI tốc độ (K_p, K_i) | : 135, 650 |
| PI dòng điện (K_p, K_i) | : 120,27 |

Mô phỏng được thực hiện trong khoảng thời gian 0,5 giây cho cả bộ điều khiển sử dụng học tăng cường và bộ điều khiển PI thông thường để so sánh, tốc độ đặt ban đầu bằng 600 vòng/phút (62,8 rad/s), sau thời gian 0,1 s, tốc độ được tăng dần và đạt giá trị 1000 vòng/phút (104.7 rad/s) tại 0,3 s. Mô men tải ban đầu là 10 Nm, sau đó tăng lên 60 Nm tại $t = 0,1$ s và giảm xuống 30 Nm tại $t = 0,4$ s.

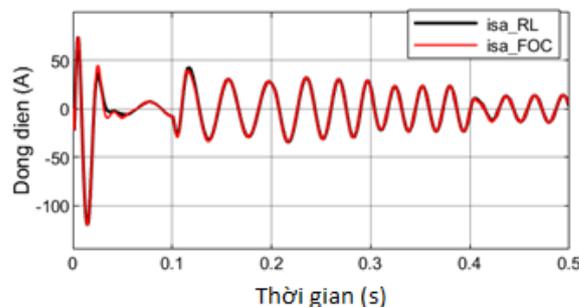
Các kết quả mô phỏng được biểu diễn trong hình 5, hình 6 và hình 7 ứng với trường hợp hai sơ đồ (sơ đồ với kỹ thuật học tăng cường và sơ đồ sử dụng PI thông thường) làm việc với các thông số định mức, tức là $R_s = 0,7384 \Omega$ và $R_r = 0,7402 \Omega$. Ta thấy rằng, đáp ứng tốc độ và đáp ứng mô men xoắn của cả 2 sơ đồ điều khiển có hiệu suất gần giống nhau. Đường cong dòng điện của chúng cũng gần như trùng nhau.



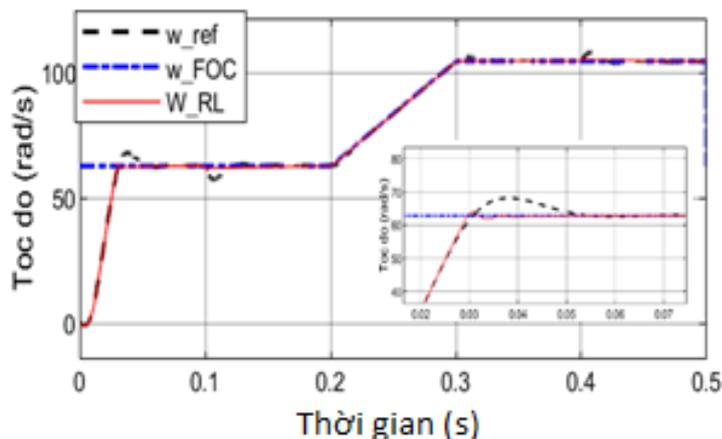
Hình 5. Đáp ứng tốc độ của RL và điều khiển FOC khi các tham số không thay đổi.



Hình 6. Đáp ứng mô men của RL và FOC khi các tham số không thay đổi.

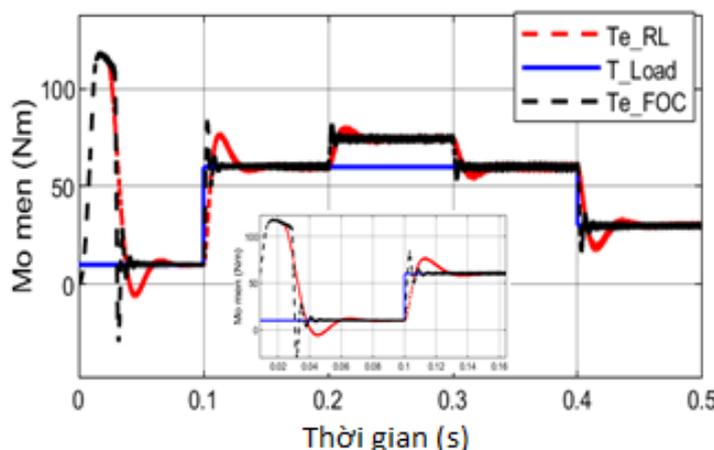


Hình 7. Đường cong dòng điện pha của RL và FOC khi các tham số không thay đổi.



Hình 8. Đáp ứng tốc độ của RL và điều khiển FOC khi các điện trở động cơ thay đổi.

Hình 8 và hình 9 biểu diễn đáp ứng tốc độ và đáp ứng mô men của hai sơ đồ (RL và PI thông thường) khi tham số của điện trở rotor và điện trở stator máy điện thay đổi, cụ thể các giá trị $R_s = 1,2 \Omega$ và $R_r = 1,3 \Omega$, tham chiếu tốc độ và mô men tải vẫn giữ nguyên như trường hợp 1.



Hình 9. Đáp ứng mô men của RL và điều khiển FOC khi các điện trở động cơ thay đổi.

Kết quả mô phỏng cho thấy, bộ điều khiển với kỹ thuật học tăng cường vẫn duy trì được hiệu năng tốt, trong khi đáp ứng tốc độ và đáp ứng mô men của sơ đồ sử dụng bộ điều khiển PI có sự thay đổi rõ rệt. Điều này cho thấy lợi thế của bộ điều khiển RL so với bộ điều khiển PI thông thường.

5. KẾT LUẬN

Học tăng cường là một kỹ thuật điều khiển thông minh đầy hứa hẹn, giúp giải quyết các bài toán điều khiển phức tạp hoạt động trong môi trường động một cách hệ thống. Trong bài báo này, Bộ điều khiển dựa trên học tăng cường đã được thiết kế để điều khiển hệ truyền động sử dụng máy điện không đồng bộ 3 pha. Hai vòng điều khiển dòng điện của chiến lược điều khiển dựa trên thông số (FOC) máy điện không đồng bộ được thay thế bởi 1 tác nhân học tăng cường với thuật toán DDPG. Hiệu suất của bộ điều khiển đã được thử nghiệm thông qua mô phỏng để điều khiển bám tốc độ, khắc phục sự thay đổi thông số của đối tượng. Các kết quả thử nghiệm cho thấy, bộ điều khiển dựa trên học tăng cường có khả năng điều chỉnh phù hợp với biến động tham số của hệ thống và có khả năng điều khiển bám tốc độ. Các nghiên cứu tiếp theo sẽ tiếp tục hoàn thiện để thử nghiệm trên hệ thống vật lý.

TÀI LIỆU THAM KHẢO

- [1]. Kushwaha, Ashish and Gopal, Madan, “*Reinforcement learning-based controller for field-oriented control of induction machine*”, Soft Computing for Problem Solving, pp. 737—749, (2019).
- [2]. Ayesha, Attaullah Y. Memon “*Reinforcement Learning Based Field Oriented Control Of An Induction Motor*” 2022 Third International Conference on Latest trends in Electrical Engineering and Computing Technologies, IEEE, pp. 315-343, (2022). DOI: 10.1109/INTELLECT55495.2022.9969403
- [3]. S. A. A. Rizvi and Z. Lin, “*Output Feedback Optimal Tracking Control Using Reinforcement Q-Learning*” 2018 Annual American Control Conference (ACC), pp. 3423-3428, (2018), doi: 10.23919/ACC.2018.8430997
- [4]. Bhagat, Sarthak and Banerjee, Hritwick and Ren, Hongliang, “*Deep Reinforcement Learning for Soft Robotic Applications: Brief Overview with Impending Challenges*”, Preprints, (2018).
- [5]. T. Pajchrowski, P. Siwek, and A. Wójcik, “*Adaptive controller design for electric drive with variable parameters by Reinforcement Learning method*” Bulletin of the Polish Academy of Sciences Technical Sciences, Vol. 68, No. 5, (2020). DOI: 10.24425/bpasts.2020.XXXX
- [6]. Chunhua Liu, K. T. Chau, Christopher H. T. Lee, Zaixin Song, “*A Critical Review of Advanced Electric Machines and Control Strategies for Electric Vehicles*” Proceedings of the IEEE | Vol. 109, no. 6, (2021).
- [7]. Krishnan, R., “*Electric Motor Drives: Modeling, Analysis, and Control*”, Prentice Hall, (2001).
- [8]. O’Rourke, Colm J. and Qasim, Mohammad M. and Overlin, Matthew R. and Kirtley, James L., “*A Geometric Interpretation of Reference Frames and Transformations: dq0, Clarke, and Park*”, IEEE Transactions on Energy Conversion, pp. 2070-2083, (2019).
- [9]. Rushikesh Kamalapurkar, Patrick Walters Joel Rosenfeld, Warren Dixon, “*Reinforcement Learning for Optimal Feedback Control*” Springer, (2018).
- [10]. Memon, A.Y., “*Sensorless Speed Control of Induction Motors Using Sliding Mode Control Strategy*”, Michigan State University. Department of Electrical and Computer Engineering, (2004).
- [11]. Chan, Tze Fun and Shi, Keli, “*Applied intelligent control of induction motor drives*”, John Wiley & Sons, (2011).
- [12]. Bose, B.K., “*Modern Power Electronics and AC Drives*”, Prentice Hall PTR, (2002).

ABSTRACT

AC drive control design with reinforcement learning techniques

Three-phase induction motors (IM) are widely known in industrial applications for their low cost and minimal maintenance. Control of AC induction motors based on Field-Oriented Control (FOC) and vector control with classic PID control laws are reliable techniques commonly used in industry. However, due to the nonlinear nature of electric machines and their susceptibility to external disturbances or parameter variations, conventional controllers often struggle to meet control requirements. Reinforcement Learning (RL) is an online learning technique, model-free, capable of handling parameter variations. These characteristics make reinforcement learning a potential candidate, acting as an adaptive controller that can replace conventional controllers. This article proposes the design of a control system for three-phase induction motor drives based on reinforcement learning techniques. The proposed controller utilizes a reinforcement learning agent with the DDPG algorithm to replace the PI controllers of the current loop circuit in conventional FOC control. The performance of the controller has been validated under various operating conditions through computer simulations in MATLAB/SIMULINK.

Keywords: Reinforcement learning; Actor-Critic; Induction motor; Agent; DDPG.