KHẢO SÁT ẢNH HƯỞNG CỦA HỆ SỐ HỌC VÀ HỆ SỐ CHIẾT KHẤU TRONG HỌC TĂNG CƯỜNG KHI ÁP DỤNG VÀO BÀI TOÁN ĐIỀU KHIỂN TÔ-PÔ TRONG MẠNG CẢM BIẾN KHÔNG DÂY

Hồ Hải Quân[1,2]*, Lê Hữu Bình[1], Nguyễn Đình Hoa Cương[3], Nguyễn Như Lương[4]

[1]Khoa Công nghệ Thông tin, Trường Đại học Khoa học, Đại học Huế

[2]Khoa Công nghệ Thông tin, Robot và AI, Trường Đại học Bình Dương

[3]Khoa Công nghệ Thông tin, Trường Đại học Phú Xuân

[4]Khoa Khoa học - Kinh tế - Công nghệ Thông tin, Trường Cao đẳng Công nghiệp Bắc Ninh

*Email: hhquan.dhkh24@hueuni.edu.vn

Ngày nhận bài: 14/12/2024; ngày hoàn thành phản biện: 19/12/2024; ngày duyệt đăng: 20/12/2024

TÓM TẮT

Điều khiển tô-pô trong mạng cảm biến không dây (WSN) dựa trên học tăng cường đang thu hút sự quan tâm đặc biệt của các nhà nghiên cứu. Thuật toán học tăng cường cho phép các nút cảm biến tự điều chỉnh vùng phủ sóng của nó để mang lại một tô-pô mạng tối ưu. Trong thuật toán học tăng cường, hệ số học và hệ số chiết khấu có ảnh hưởng lớn đến độ hội tụ và hiệu quả thực thi của thuật toán. Trong bài báo này, chúng tôi tập trung khảo sát ảnh hưởng của các hệ số này đến hiệu quả của việc áp dụng học tăng cường vào bài toán điều khiển tô-pô trong WSN. Trên cơ sở các kết quả mô phỏng thu được, chúng tôi xác định các giá trị tối ưu của các hệ số này sao cho thuật toán hội tụ nhanh và mang lại hiệu quả thực thi tốt nhất.

Từ khóa: Điều khiển tô-pô, học tăng cường, hệ số học, hệ số chiết khấu.

1. MỞ ĐẦU

Mạng cảm biến không dây (Wireless Sensor Network - WSN) là hệ thống phân tán gồm nhiều nút cảm biến, có khả năng thu thập, xử lý và truyền dữ liệu qua môi trường không dây, thường dùng trong các ứng dụng IoT như giám sát môi trường, nông nghiệp và thành phố thông minh. Do giới hạn về năng lượng, bộ nhớ và khả năng tính toán, mỗi nút cần điều chỉnh vùng truyền thông sao cho bậc nút tiến đến một giá trị mong đợi nhằm đảm bảo tối ưu việc sử dụng năng lượng và các độ đo hiệu năng khác. Trong các công trình nghiên cứu về điều khiển tô-pô trong mạng không dây, nhiều

phương pháp đã được đề xuất nhằm tối ưu hóa tài nguyên và giảm tiêu thụ năng lượng. Trong [2], các tác giả đã đề xuất thuật toán TFACR (Topological control by Flexibly Adjusting the Coverage Radius) điều chỉnh bán kính phủ sóng trong mạng ad-hoc dựa trên 5G để duy trì kết nối và cân bằng tải. Trong [9], thuật toán EEDBTC (Energy Efficient Direction-Based Topology Control Algorithm) được đề xuất nhằm định hướng việc truyền dữ liệu trực tiếp tới trạm gốc. Một số công trình nghiên cứu khác đã đề xuất các kiến trúc và đánh giá thuật toán điều khiển tô-pô nhằm tối ưu hóa việc phân bổ tài nguyên và giảm tiêu hao năng lượng [12, 17]. Ngoài ra, các hướng nghiên cứu cũng tập trung tăng cường khả năng kết nối và tính chịu lỗi, như sử dụng phương pháp triển khai nút không đồng nhất [10], sử dụng mô hình điều khiển tự tổ chức [13], dự đoán tham số di động để xây dựng cấu trúc mạng ổn định [14]. Các thuật toán tối ưu hóa hiệu suất mạng như thuật toán học tăng cường DDPG [6] và kiến trúc hỗ trợ UAV cho mạng VANET [15] đã được áp dụng để giảm độ trễ và cải thiện thông lượng. Bên cạnh đó, các nghiên cứu về mạng ad-hoc UAV cũng sử dụng học tăng cường sâu (DRL) và cơ chế CDS nhằm duy trì kết nối ổn định trong môi trường động [16, 20]. Thuật toán di truyền cũng được ứng dụng, như trong [3] với SQETC, nhằm giảm chênh lệch giữa bậc nút thực tế và mong đợi. Công trình nghiên cứu trong [4] đã đề xuất thuật toán RL-CRC sử dụng học tăng cường để điều chỉnh vùng truyền thông. Trong [5], các tác giả đã phát triển phương pháp LTRT dựa trên thuật toán cây khung nhỏ nhất nhằm tạo ra cấu trúc mạng hiệu quả, tiết kiệm năng lượng.

Với phương pháp học tăng cường, hệ số tỷ lệ học và hệ số chiết khấu trong lý thuyết học tăng cường có ảnh hưởng lớn đến độ hội tụ và hiệu quả thực thi của thuật toán. Điều này cũng đã thu hút sự quan tâm của các nhà nghiên cứu trong thời gian gần đây. Trong [21], các tác giả đã tích hợp các hệ số tỷ lệ học và tỷ lệ khám phá trong một nền tảng chung nhằm tăng tốc độ hội tụ và cải thiện chất lượng trong học tăng cường, đặc biệt là ở môi trường phức tạp. Một công trình nghiên cứu khác đã chỉ ra việc giảm hệ số chiết khấu giúp ổn định quá trình học, phù hợp với trường hợp dữ liệu hạn chế [22]. Trong [23], các tác giả đã đơn giản hóa việc suy luận hệ số chiết khấu bằng cách chia miền [0, 1) thành các đoạn với chính sách tối ưu không đổi, từ đó phát triển kỹ thuật số học để tìm hệ số chiết khấu phù hợp. Trong [24], ảnh hưởng của hệ số tỷ lệ học, hệ số chiết khấu và hệ số ε trong chính sách ε-greedy đến tốc độ hội tụ và khả năng học chính sách tối ưu trong Q-Learning cũng đã được khảo sát. Khi áp dụng học tăng cường cho bài toán điều khiển tô-pô trong mạng không dây, việc xác định các hệ số này sao cho phù hợp để thuật toán mang lại hiệu quả cao là điều hết sức cần thiết. Đây là mục tiêu nghiên cứu của công trình này. Các đóng góp mới của công trình được tóm tắt như sau:

(i) Chúng tôi đã phân tích, đánh giá ảnh hưởng của hệ số tỷ lệ học và hệ số chiết khấu trong học tăng cường cho một kịch bản ứng dụng mà môi trường thường xuyên thay đổi, đó là bài toán điều khiển tô-pô trong WSN.

(ii) Các tiêu chí về độ tối ưu tô-pô, hiệu quả sử dụng năng lượng và chất lượng tín hiệu truyền dẫn đã được khảo sát đồng thời bằng phương pháp mô phỏng. Điều này cho phép xác định được khoảng giá trị tối ưu của hệ số tỷ lệ học và hệ số chiết khấu nhằm nâng cao hiệu năng mạng.

Các phần còn lại của bài báo được bố cục như sau: Phần 2 trình bày cách thức áp dụng học tăng cường vào bài toán điều khiển tô-pô. Phần 3 phân tích ảnh hưởng của hệ số học và hệ số chiết khấu đến hiệu quả thực thi của thuật toán điều khiển tô-pô sử dụng học tăng cường. Các kết quả mô phỏng và thảo luận được trình bày trong Phần 4. Cuối cùng là kết luận và đề xuất hướng phát triển tiếp theo, được trình bày trong Phần 5.

2. ĐIỀU KHIỂN TÔ-PÔ DỰA TRÊN HỌC TĂNG CƯỜNG

Học tăng cường (Reinforcement Learning - RL) đã được ứng dụng thành công trong điều khiển tô-pô ở WSN nhờ khả năng tự học và thích nghi với môi trường động. Với các giới hạn về năng lượng và tài nguyên, RL giúp các node điều chỉnh phạm vi truyền dẫn để duy trì kết nối và giảm tiêu hao năng lượng. Qua quá trình tương tác, node đánh giá hiệu quả hành động qua phần thưởng liên quan đến chất lượng liên kết, mức tiêu thụ năng lượng và độ ổn định của mạng, từ đó tối ưu chính sách điều khiển dựa trên kinh nghiệm và dự đoán tương lai. Phương pháp này tự động, tối ưu theo thời gian và mở rộng tốt hơn so với các giải pháp điều khiển tĩnh hoặc thủ công, góp phần nâng cao hiệu quả và độ bền vững của WSN trong các ứng dụng như giám sát môi trường, nông nghiệp thông minh và thành phố thông minh.

Để áp dụng học tăng cường vào bài toán điều khiển tô-pô, ta cần mô hình hóa bài toán điều khiển tô-pô thành một mô hình học tăng cường mà nó được đặc trưng bởi các thành phần {môi trường (environment), tác nhân (agent), trạng thái (state), hành động (action), phần thưởng (reward) và chính sách (policy)}. Trong bài báo này, chúng tôi sử dụng phương pháp mô hình hóa tương tự như trong [4], trong đó môi trường là hệ thống mạng, tác nhân tương ứng với các nút mạng, trạng thái của mỗi nút là bậc hiện hành của nó, hành động tương ứng với việc thay đổi bán kính phủ sóng để hướng đến bậc mong đợi, chính sách là cách thức mà một nút lựa chọn hành động để thực hiện, phần thưởng là một hàm được thiết kế sao cho giá trị càng lớn thì bậc của một nút càng tiến gần đến bậc mong đợi. Trong bài báo này, hàm thưởng được thiết kế như sau:

$$r = k_{max} - |k_{current} - k_{desired}| \quad (1)$$

trong đó, $k_{current}$ là bậc hiện hành của nút đang xét và $k_{desired}$ là bậc mong đợi. Trong học tăng cường, tổng giá trị phần thưởng đạt được mỗi khi tác nhân thực hiện một hành động được xác định bởi thuật toán Q-learning, theo phương trình sau:

$$Q(s,a) = Q(s,a) + \alpha[r + \gamma \cdot \max_{a'} Q(s',a') - Q(s,a)] \quad (2)$$

với $Q(s, a)$ là giá trị Q khi thực hiện hành động a tại trạng thái s, α là hệ số học và γ là hệ số chiết khấu.

Thuật toán 1 là mã giả lập của thuật toán điều khiển tô-pô sử dụng học tăng cường. Mỗi nút cảm biến được xem như một tác nhân tự điều chỉnh bán kính phủ sóng để đạt được bậc mong đợi. Ban đầu, không gian mạng được khởi tạo với các tham số như kích thước vùng hoạt động, số lượng nút, bán kính phủ sóng lớn nhất, mỗi nút khởi tạo bảng giá trị Q để lưu giá trị cho từng hành động. Trong vòng lặp học, nút chọn hành động để khám phá hoặc khai thác, điều chỉnh phạm vi truyền thông, từ đó ảnh hưởng đến số lượng nút láng giềng (bậc nút). Sau mỗi khi thực hiện một hành động, phần thưởng được tính dựa trên độ chênh lệch giữa bậc hiện tại và bậc mong đợi, phần thưởng được dùng cập nhật bảng Q. Qua nhiều vòng lặp, hệ thống dần tối ưu cấu hình vùng truyền thông và đạt bậc nút tiệm cận giá trị kỳ vọng nhằm thu được một tô-pô mạng tối ưu.

Thuật toán 1: Điều chỉnh bán kính phủ sóng tại mỗi nút cảm biến sử dụng học tăng cường

```
//Khởi tạo bảng Q
For (mỗi nút trong mạng)
    For (mỗi trạng thái s trong tập trạng thái)
        For (mỗi hành động a trong tập hành động)
            Q(s, a) ← 0;
        EndFor
    EndFor
EndFor
//Điều chỉnh bán kính phủ sóng dựa trên học tăng cường
For (mỗi lần học)
    For (mỗi nút trong mạng)
        Xác định trạng thái hiện tại (bậc hiện tại);
        Chọn hành động theo chính sách ε-greedy;
        Điều chỉnh bán kính phủ sóng theo hành động được chọn;
        Cập nhật giá trị Q theo phương trình (2);
    EndFor
EndFor
```

3. ẢNH HƯỞNG CỦA HỆ SỐ HỌC VÀ HỆ SỐ CHIẾT KHẤU

Trong thuật toán điều khiển tô-pô sử dụng học tăng cường, độ hội tụ cũng như hiệu quả thực thi của thuật toán phụ thuộc vào việc cập nhật bảng giá trị Q, được thực hiện theo phương trình (2) qua mỗi lần học. Trong phương trình này, có hai hệ số quan trọng, đó là hệ số học (α) và hệ số chiết khấu (γ).

3.1. Ảnh hưởng của hệ số học (α)

Hệ số học (α) là một giá trị trong đoạn [0, 1], xác định tỷ lệ giữa giá trị phần thưởng mới và phần thưởng hiện hành mỗi khi tác nhân thực hiện một hành động. Nếu α bằng 0, tác nhân có thể xem như không học được gì, giá trị phản hồi giữ nguyên như giá trị hiện hành. Nếu α bằng 1, phần thưởng nhận toàn bộ giá trị mới, nghĩa là tác nhân chỉ xem xét thông tin mới nhất mà không quan tâm đến các thông tin trước đó.

Từ ý nghĩa của hệ số học như được phân tích ở trên, hệ số học bằng 1 là tối ưu trong trường hợp hệ thống mạng hoàn toàn ổn định, như mạng có dây hoặc không dây cố định (các điểm truy nhập không di chuyển). Trong trường hợp một hệ thống mạng phụ thuộc nhiều vào các yếu tố ngẫu nhiên và trạng thái mạng thay đổi thường xuyên như mạng tùy biến di động (MANET), mạng cảm biến không dây (WSN) mà các nút cảm biến có di chuyển thì hệ số học α nên được thiết lập nhỏ hơn 1.

3.2. Ảnh hưởng của hệ số chiết khấu (γ)

Hệ số chiết khấu là một giá trị trong đoạn [0, 1], xác định mức độ ảnh hưởng của giá trị phản hồi trong tương lai. Hệ số chiết khấu cao giúp thuật toán chú trọng vào phần thưởng dài hạn nhưng làm chậm hội tụ, trong khi hệ số chiết khấu thấp giúp cập nhật nhanh nhưng bỏ qua tín hiệu trong tương lai. Tinh chỉnh giá trị hệ số chiết khấu là cần thiết để đạt sự cân bằng giữa khai thác và thăm dò, tối ưu hóa chiến lược học và điều khiển.

Khi áp dụng học tăng cường vào bài toán điều khiển tô-pô trong mạng WSN, việc xác định hệ số học và hệ số chiết khấu sao cho tối ưu là điều khó khăn, vì WSN là một hệ thống mạng có sự thay đổi trạng thái thường xuyên. Trong phần tiếp theo dưới đây, chúng tôi sử dụng phương pháp mô phỏng để phân tích ảnh hưởng của hai hệ số này đến hiệu quả hoạt động của thuật toán, từ đó xác định các giá trị phù hợp cho các hệ số này tùy theo từng trường hợp cụ thể.

4. MÔ PHỎNG VÀ PHÂN TÍCH KẾT QUẢ

4.1. Kịch bản mô phỏng

Để tiến hành mô phỏng xác định ảnh hưởng của hệ số học và hệ số chiết khấu đến hiệu quả thực thi của thuật toán điều khiển tô-pô sử dụng học tăng cường, chúng tôi đã cài đặt trên MATLAB R2024. Kịch bản mô phỏng được thiết lập như ở Bảng 1. Ban đầu, các nút cảm biến được khởi tạo ngẫu nhiên trong vùng diện tích 1000 x 1000 m^2 . Mỗi nút khởi đầu với vùng truyền thông lớn nhất, được thiết lập là 250m. Tổng số nút cảm biến từ 50 đến 90 tùy theo từng kịch bản mô phỏng, bậc nút mong đợi là 4.

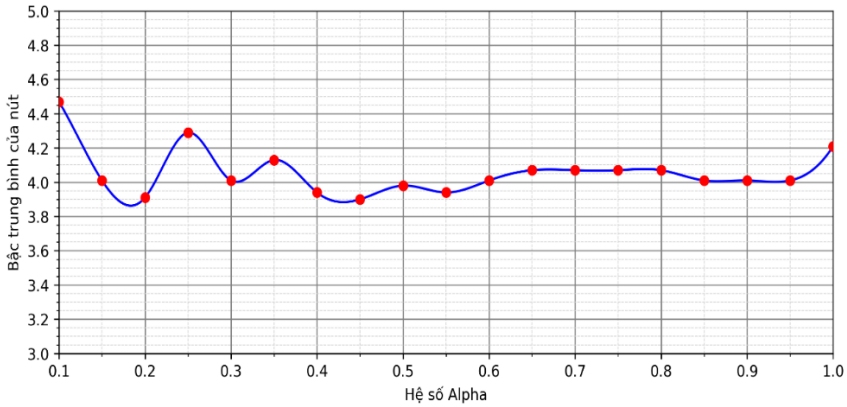
Bảng 1. Các tham số của kịch bản mô phỏng

Tham số	**Giá trị**
Diện tích mạng	1000×1000 [m^2]
Tổng số nút cảm biến	50 đến 90
Vùng truyền thông lớn nhất	250 [m]
Bậc nút mong đợi	4
Hệ số học	Từ 0 đến 1
Hệ số chiết khấu	Từ 0 đến 1
Hệ số ε trong chính sách *ε-greedy*	0.1

4.2. Kết quả mô phỏng và thảo luận

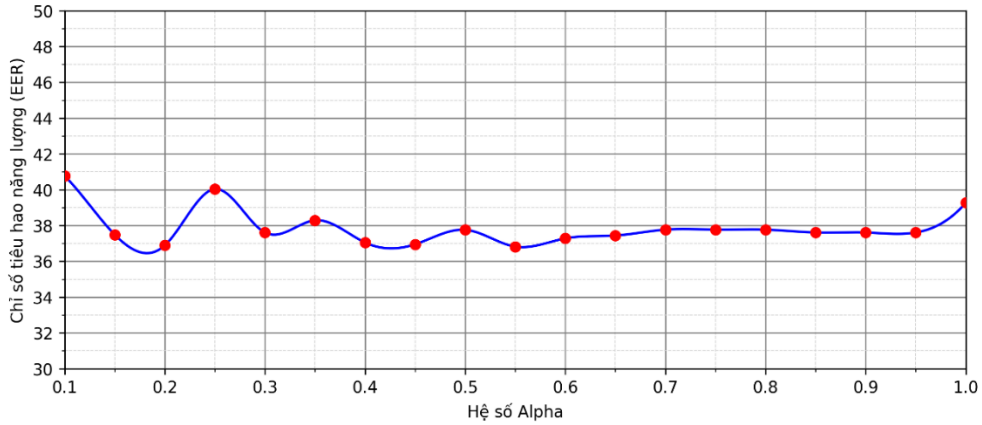
4.2.1. Ảnh hưởng của hệ số học

Kết quả thu được ở Hình 1 cho thấy ảnh hưởng của hệ số học đến bậc trung bình của các nút trong toàn mạng. Ta thấy rằng, với α quá nhỏ (từ 0,10 đến 0,15), việc cập nhật bảng giá trị Q diễn ra rất chậm nên dù kết quả bậc nút đạt gần 4 (4,47 đến 4,01), nhưng cần nhiều vòng lặp mới ổn định và dễ bị kẹt ở giá trị cao hơn mong đợi. Khi α tăng lên, từ 0,40 đến 0,60, tốc độ cập nhật vừa đủ nhanh để điều chỉnh phạm vi truyền thông linh hoạt, giúp bậc nút dao động sát 4 nhất (từ 3,91 đến 4,13) và hội tụ trong thời gian tương đối ngắn mà không bị quá phản ứng. Tuy nhiên, khi α lớn hơn 0,85, bước cập nhật trở nên quá lớn, dẫn đến dao động mạnh và overshoot, thể hiện bậc nút trung bình tăng lên đến 4,21, thể hiện sự thiếu ổn định. Như vậy, vùng α từ 0,40 đến 0,60 cân bằng tốt giữa tốc độ hội tụ và độ ổn định của bậc nút.

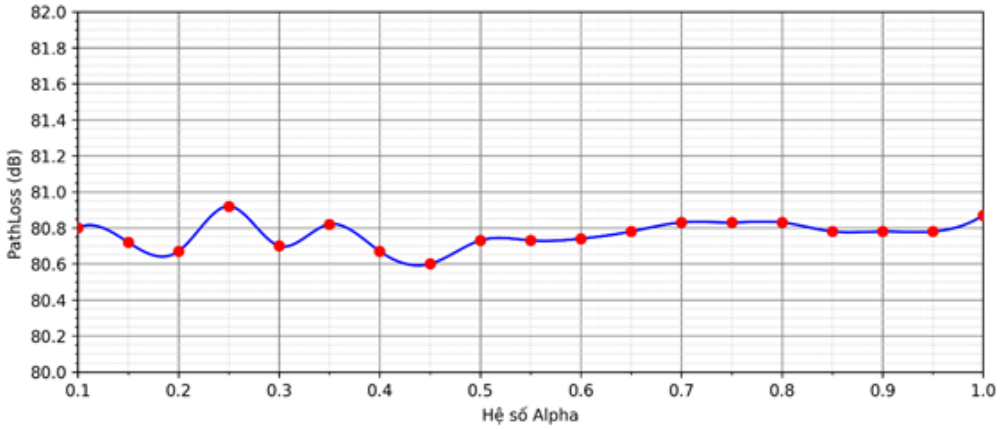


Hình 1. Ảnh hưởng của hệ số học đến bậc nút trung bình

Tiếp theo, chúng tôi phân tích ảnh hưởng của hệ số học đến chỉ số tiêu hao năng lượng (EER). Kết quả thu được như cho thấy ở Hình 2. Khi α nhỏ (0,10 đến 0,20), hệ thống tiêu tốn nhiều năng lượng (từ 37 đến 41%) do phạm vi truyền thông duy trì lớn trong nhiều bước học đầu. Khi α tăng lên từ 0,25 đến 0,55, các nút điều chỉnh vùng truyền thông nhanh hơn nhằm giảm khoảng cách truyền và tiết kiệm năng lượng, vì vậy EER giảm xuống thấp nhất, chỉ còn 36,82% tại α = 0,55. Ở vùng α từ 0,60 đến 0,85, EER dao động nhẹ quanh các giá trị từ 37,28 đến 37,77%, phản ánh chiến lược điều chỉnh cân bằng giữa thăm dò và khai thác. Tuy nhiên với α lớn gần 1,00, quá trình học trở nên bất ổn, EER tăng trở lại (đến 39,27%) do dao động liên tục của vùng truyền thông. Như vậy, α từ 0,40 đến 0,60 cho hiệu quả năng lượng tối ưu.



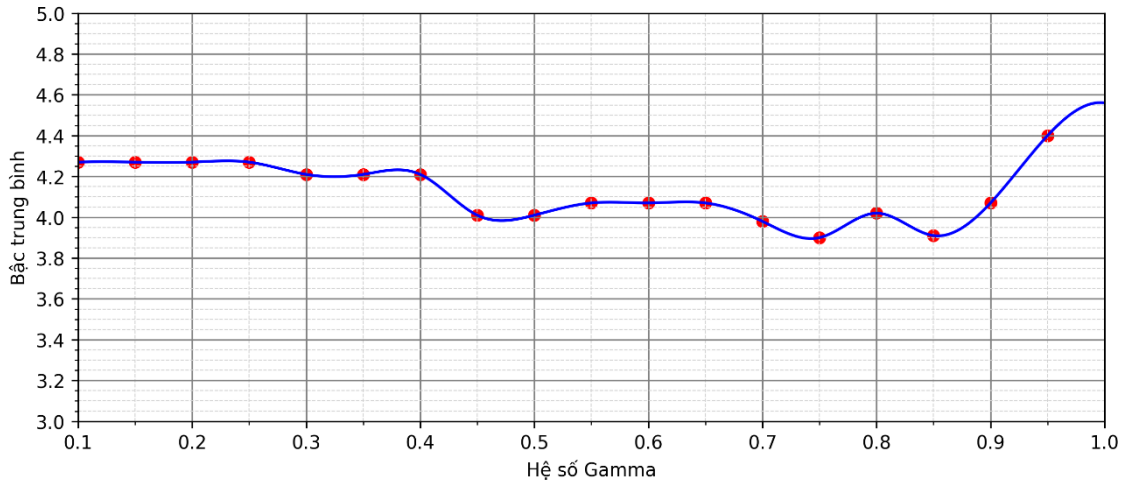
Hình 2. Ảnh hưởng của hệ số học đến chỉ số tiêu hao năng lượng



Hình 3. Ảnh hưởng của hệ số học đến suy hao đường truyền

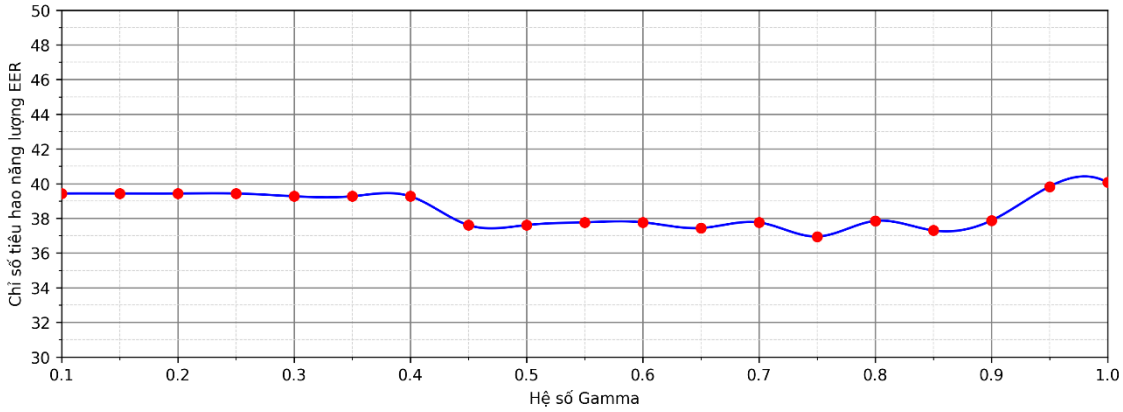
Một độ đo quan trọng nữa mà các thuật toán điều khiển tô-pô cần xem xét là suy hao đường truyền (Pathloss). Hệ số học cũng ảnh hưởng đến độ đo này như cho thấy ở kết quả mô phỏng trên Hình 3. Đối với $\alpha \leq 0{,}25$, Pathloss cao (~80,92 dB) do phạm vi truyền thông lớn. Khi α từ 0,30 đến 0,45, phạm vi truyền thông giảm ổn định, dẫn tới Pathloss giảm về 80,60 đến 80,70 dB, đánh dấu mức tổn thất tín hiệu thấp nhất. Trong vùng α từ 0,50 đến 0,85, Pathloss dao động nhẹ (80,73 – 80,83) do chức năng điều chỉnh phạm vi vẫn duy trì ổn định. Khi α tiến gần 1,00, dao động mạnh làm phạm vi truyền thông bất ổn, Pathloss tăng nhẹ lên 80,87. Như vậy, α từ 0,40 đến 0,6 tối ưu nhất để giảm suy hao tín hiệu.

4.2.2. Ảnh hưởng của hệ số chiết khấu



Hình 4. Ảnh hưởng của hệ số chiết khấu đến bậc trung bình

Ngoài hệ số học, hệ số chiết khấu (γ) cũng có ảnh hưởng lớn đến hiệu quả thực thi của thuật toán điều khiển tô-pô sử dụng học tăng cường. Kết quả mô phỏng trên Hình 4 cho thấy rõ ảnh hưởng của γ đến bậc trung bình của các nút trong toàn mạng. Ta thấy rằng, với bậc mong đợi là 4 thì hệ số chiết khấu từ 0.45 đến 0.85 là phù hợp. Với độ đo về mức tiêu hao năng lượng (EER), độ đo này đạt giá trị tốt nhất khi γ từ 0,45 đến 0,90. Lúc này EER ở mức thấp nhất, chỉ từ 37 đến 38%, điều này cho phép hệ thống mạng sử dụng năng lượng một cách hiệu quả.



Hình 5. Ảnh hưởng của hệ số chiết khấu đến chỉ số tiêu hao năng lượng

Từ các kết quả mô phỏng được phân tích ở trên chúng tôi có thể kết luận rằng, khi áp dụng thuật toán học tăng cường cho bài toán điều khiển tô-pô trong mạng WSN, hệ số học nên chọn từ 0,40 đến 0,60, hệ số chiết khấu nên chọn từ 0,45 đến 0,90 để thuật toán có thể mang lại hiệu quả hoạt động tốt nhất.

5. KẾT LUẬN

Để nâng cao hiệu quả của việc áp dụng học tăng cường vào bài toán điều khiển tô-pô trong mạng cảm biến không dây, việc nghiên cứu để xác định vùng giá trị tối ưu của hệ số học và hệ số chiết khấu là điều cần thiết. Trong bài báo này, chúng tôi đã tập trung khảo sát ảnh hưởng của các hệ số này đến hiệu năng mạng thông qua các độ đo bậc nút trung bình, mức tiêu thụ năng lượng và suy hao đường truyền. Thông qua kết quả mô phỏng đã xác định được vùng giá trị tối ưu cho hệ số học và hệ số chiết khấu.

Trong hướng nghiên cứu tiếp theo, chúng tôi tiếp tục thực hiện trong các môi trường mạng khác nhau như mạng tùy biến di động (MANET), mạng tùy biến của các thiết bị bay không người lái (FANET) và mạng tùy biến gồm các thiết bị giao thông thông minh (VANET).

TÀI LIỆU THAM KHẢO

[1] Binh LH, Duong TVT (2024) A novel and effective method for solving the router nodes placement in wireless mesh networks using reinforcement learning. PLOS ONE 19(4): e0301073.

[2] L. H. Binh, T. -V. T. Duong and V. M. Ngo, "TFACR: A Novel Topology Control Algorithm for Improving 5G-Based MANET Performance by Flexibly Adjusting the Coverage Radius," in IEEE Access, vol. 11, pp. 105734-105748, 2023.

[3] Hong Thi Chu Hai, Le Huu Binh, Le Duc Huy, "A topology control algorithm taking into account energy and quality of transmission for software-defined wireless sensor network", International Journal of Computer Networks & Communications (IJCNC), Vol.16, No.2, 2024, pp. 107-116.

[4] Le, Thien Moh, Sangman. (2017), "An Energy-Efcient Topology Control Algorithm Based on Reinforcement Learning for Wireless Sensor Networks", International Journal of Control and Automation, 10. 233-244. 10.14257/ijca.2017.10.5.22.

[5] K. Miyao, H. Nakayama, N. Ansari, N. Kato, LTRT: An Efficient and Reliable Topology Control Algorithm for Ad-hoc Networks, IEEE Transactions on Wireless Communications, vol. 8, no.12, (2009), pp. 6050-6058.

[6] Yoo, Taehoon, Sangmin Lee, Kyeonghyun Yoo, and Hwangnam Kim. 2023. "Reinforcement Learning Based Topology Control for UAV Networks" Sensors 23, no. 2: 921.

[7] Meng, Xiangyue, Hazer Inaltekin and Brian Scott Krongold. "Deep Reinforcement Learning-Based Topology Optimization for Self-Organized Wireless Sensor Networks." 2019 IEEE Global Communications Conference (GLOBECOM) (2019): 1-6.

[8] Waqas, Abdullah and Hasan Mahmood. "A Game Theoretical Approach for Topology Control in Wireless Ad Hoc Networks with Selfish Nodes." *Wireless Personal Communications* 96 (2017): 249 - 263.

[9] Khan, M. and Hussain, S. (2020) Energy Efficient Direction-Based Topology Control Algorithm for WSN. Wireless Sensor Network, 12, 37-47.

[10] Zhang, Xiujuan et al. "Robust t-Path Structure Control Algorithm in Wireless Ad Hoc Networks." Algorithm Applications in Management (2021).

[11] Kadivar, Mehdi. "An Adaptive Yao-based topology control algorithm for wireless ad-hoc networks." 2020 10th International Conference on Computer and Knowledge Engineering (ICCKE) (2020): 457-462.

[12] Wei, D., Yang, E., & He, Y. (2023). Topology Control and Optimization of Self-organized Networks for Wireless Personal Communication. 2023 IEEE 6th International Conference on Automation, Electronics and Electrical Engineering (AUTEEE), 462-467.

[13] Seyyedi, Azra et al. "Connectivity Aware and Energy Efficient Self-Organizing Distributed IoT Topology Control." (2023).

[14] Rahmani, Parisa và Hamid Haj Seyyed Javadi. "Kiểm soát cấu trúc mạng trong MANET

bằng thuật toán Bayesian Pursuit." Wireless Personal Communications 106 (2019): 1089 - 1116.

[15] Xue, Ying et al. "UAV-Assisted Control Design with Stochastic Communication Delays." 2022 IEEE 8th International Conference on Computer and Communications (ICCC) (2022): 707-710.

[16] Li, J., Yi, P., Duan, T., Wang, Y., Zhang, Z., Yu, J., & Hu, T. (2024). Centroid-Guided Target-Driven Topology Control Method for UAV Ad-Hoc Networks Based on Tiny Deep Reinforcement Learning Algorithm. IEEE Internet of Things Journal, 11, 21083-21091.

[17] Singla, Pallavi and Amit Munjal. "Topology Control Algorithms for Wireless Sensor Networks: A Review." Wireless Personal Communications 113 (2020): 2363 - 2385.

[18] Banerjee, Indushree et al. "Self-organizing topology for energy-efficient ad-hoc communication networks of mobile devices." Complex Adaptive Systems Modeling 8 (2020): 1-21.

[19] Song, Meiyan et al. "Locally Adaptive Power Control for Optimizing Age of Information in Wireless Networks." 2022 IEEE Wireless Communications and Networking Conference (WCNC) (2022): 1599-1604.

[20] Qi, Xiaohan et al. "CDS-Based Topology Control in FANETs via Power and Position Optimization." IEEE Wireless Communications Letters 9 (2020): 2015-2019.

[21] Gupta, Smriti, Sabita Pal, Kundan Kumar and Kuntal Ghosh. "Coupling Effect of Exploration Rate and Learning Rate for Optimized Scaled Reinforcement Learning." *SN Computer Science* 4 (2023): 1-16.

[22] Amit, Ron, Ron Meir and Kamil Ciosek. "Discount Factor as a Regularizer in Reinforcement Learning." *International Conference on Machine Learning* (2020).

[23] Kalat, Shadi Tasdighi, Sriram Sankaranarayanan and Ashutosh Trivedi. "What is Your Discount Factor?" QEST+FORMATS (2024).

[24] Zhu, Jiahan. "The Effect of Hyperparameters on the Model Convergence Rate of Cliff Walking Problem Based on Q-Learning." *Applied and Computational Engineering* (2024).

INVESTIGATING THE EFFECTS OF LEARNING RATE AND DISCOUNT FACTOR IN REINFORCEMENT LEARNING-BASED TOPOLOGY CONTROL FOR WIRELESS SENSOR NETWORKS

Ho Hai Quan[1,2]*, Le Huu Binh[1], Nguyen Dinh Hoa Cuong[3], Nguyen Nhu Luong[4]

[1] Faculty of Information Technology, University of Sciences, Hue University

[2] Faculty of Information Technology, Robotics, and AI, Binh Duong University

[3] Faculty of Information Technology, Phu Xuan University

*Email: hhquan.dhkh24@hueuni.edu.vn

ABSTRACT

Topology control in wireless sensor networks (WSNs), based on reinforcement learning, has attracted the attention of researchers. Reinforcement learning algorithms allow sensor nodes to self-adjust their coverage areas for optimal network topology. In reinforcement learning algorithms, the learning rate coefficient and discount coefficient significantly influence the convergence and performance of the algorithm. In this study, we focused on investigating the influence of these coefficients on the efficiency of applying reinforcement learning to the topology control problem in WSNs. Based on the simulation results obtained, we determined the optimal values of these coefficients so that the algorithm could converge quickly and provide the best network performance.

Keywords: Wireless sensor networks, Topology control, Connection stability.

Hồ Hải Quân sinh năm 1983 tại Quảng Bình. Ông tốt nghiệp Cử nhân ngành Công nghệ thông tin năm 2008 và Thạc sĩ ngành Khoa học máy tính năm 2013 tại Trường Đại học Khoa học, Đại học Huế. Ông hiện đang là giảng viên Khoa Công nghệ thông tin, Trường Đại học Bình Dương.

Lĩnh vực nghiên cứu: Học máy, Học Sâu, AI, Công nghệ mạng thế hệ mới, ứng dụng của học máy và trí tuệ nhân tạo trong công nghệ mạng, mô phỏng và thiết kế mạng.

Lê Hữu Bình sinh năm 1978 tại Quảng Trị. Ông tốt nghiệp Kỹ sư ngành Điện tử Viễn thông năm 2001 tại Trường Đại học Bách khoa, Đại học Đà Nẵng và Thạc sĩ ngành Khoa học máy tính năm 2007 tại Trường Đại học Khoa học, Đại học Huế. Ông nhận học vị Tiến sĩ ngành Hệ thống thông tin năm 2020 tại Học viện Khoa học và Công nghệ, Viện Hàn lâm Khoa học và Công nghệ Việt Nam. Ông hiện đang là giảng viên Khoa Công nghệ thông tin, Trường Đại học Khoa học, Đại học Huế.

Lĩnh vực nghiên cứu: Công nghệ mạng thế hệ mới, ứng dụng của học máy và trí tuệ nhân tạo trong công nghệ mạng, mô phỏng và thiết kế mạng.

Nguyễn Đình Hoa Cương đã nhận bằng Thạc sĩ Khoa học và Tiến sĩ chuyên ngành Khoa học Máy tính tại Trường Đại học Khoa học, Đại học Huế, Việt Nam và tại Đại học Khon Kaen, Thái Lan lần lượt vào các năm 2005 và 2016. Hiện nay, ông là giảng viên tại Trường Đại học Phú Xuân, Huế.

Lĩnh vực nghiên cứu: khai phá dữ liệu, khai phá văn bản, quản lý tri thức và hệ thống hỗ trợ ra quyết định

Nguyễn Như Lương sinh năm 1995 tại Bắc Ninh. Ông tốt nghiệp Cử nhân ngành Công nghệ thông tin năm 2017 và Thạc sĩ Công nghệ Thông tin chuyên ngành Kỹ thuật phần mềm năm 2021 tại Trường Đại học Kinh doanh và Công nghệ Hà Nội. Ông hiện đang là giảng viên Khoa Khoa học – Kinh tế - Công nghệ Thông tin, Trường Cao đẳng Công nghiệp Bắc Ninh.

Lĩnh vực nghiên cứu: Công nghệ mạng máy tính, mô phỏng và thiết kế mạng.