ĐẠI HỌC THÁI NGUYÊN
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN & TRUYỀN THÔNG

VŨ HẢI HIỆU

PHƯƠNG PHÁP HUẤN LUYỆN ĐA TÁC TỬ VỚI SỰ CÓ MẶT CỦA TÁC TỬ THEO DÕI

Chuyên ngành: Khoa học máy tính

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

MỞ ĐẦU

1. Lý do chọn đề tài

Trong những năm gần đây, việc nghiên cứu và triển khai ứng dụng công nghệ đa tác tử đã trở thành một trong những hướng trọng tâm của ngành Khoa học máy tính. Mặc dù công nghệ này chỉ mới bắt đầu phát triển mạnh từ năm 90 của thế kỷ XX nhưng nó đã thể hiện rất rõ nét về tính hiệu quả và tầm ảnh hưởng tích cực của nó trong ngành khoa học máy tính nói riêng và các lĩnh vực có ứng dụng công nghệ thông tin nói chung. Đối với các lĩnh vực tự động hóa công nghiệp, điều khiển giám sát, phân phối năng lượng hay các game hiện đại, chúng luôn thể hiện tính chất phức tạp, bất định và mô hình luôn thay đổi vì thế xu hướng xây dựng hệ thống theo hướng công nghệ đa tác tử là một tất yếu. Mặt khác chúng ta thấy rất rõ ràng máy tính hiện nay không còn là các hệ thống hoạt động riêng lẻ nữa, xu hướng điều khiển phân tán là một vấn đề cốt lõi mà các nhà phát triển ứng dụng cần quan tâm tới. Lượng công việc máy tính đảm nhiệm thay con người ngày càng nhiều, chúng ta ngày càng trao quyền cho máy tính nhiều hơn, máy tính có thể quyết định những tình huống quan trọng thay con người. Để thực hiện tốt các công việc thay con người, máy tính cần phải thông minh, linh hoạt trong môi trường hoạt động của mình. Trong vài năm gần đây, vấn đề ***máy học*** đã được nghiên cứu khá nhiều, các công trình nghiên cứu mang tính nền móng cho lĩnh vực này liên tục ra đời và từ đó các ứng dụng đưa vào thực tiễn cũng phát triển theo. Một trong những vấn đề thuộc lĩnh vực máy học là các giải pháp huấn luyện tác tử và đa tác tử, đây là vấn đề rất rộng và đầy thách thức, các vấn đề mang tính lý thuyết cơ sở không ngừng được bổ sung và hoàn thiện. Trước khi bước vào môi trường hoạt động thực sự của mình, tác tử cần phải trải qua một qua trình huấn luyện hay nói cách khác là học cách ra quyết định để có thể đem lại một kết quả tốt. Với mong muốn tìm hiểu về công nghệ tác tử, tác tử thông minh, tương tác và phối hợp trong hệ đa tác tử đặc biệt là phương pháp huấn luyện cho hệ đa tác tử, chúng tôi đã quyết định chọn đề tài "**Phương pháp huấn luyện đa tác tử với sự có mặt của tác tử theo dõi**".

2. Lịch sử vấn đề

Bản chất của huấn luyện tác tử và đa tác tử nói chung là quá trình cho tác tử hành động trong môi trường của chúng, lấy về chuỗi các kết quả, các kết quả đó được phân tích, đánh giá và cuối cùng là một bảng lượng giá được sinh ra từ những kết quả trên. Bảng lượng giá mức độ quan trọng trong mỗi hành động của tác tử chính là kết quả của quá trình huấn luyện và nó chính là căn cứ giúp tác tử quyết định hành động của mình tại mỗi trạng thái trong môi trường hoạt động của nó. Một trong những thuật toán huấn luyện tác tử được xem là nền móng cho nhiều nghiên cứu về sau đó là thuật toán huấn luyện đơn tác tử Q-Learning do Watkins và Dayan xây dựng năm 1992 [18]. Có rất nhiều các thuật toán khác được cải tiến từ Q-Learning và đã mang lại hiệu quả rất lớn. Ví dụ thuật toán Nash Q-Learning do Junling Hu và Michael P. Wellman phát triển [11], giải thuật này dựa trên nền tảng Q-Learning, lý thuyết cân bằng Nash và lý thuyết trò chơi, với sự kết hợp trên giải thuật này đã cho phép huấn luyện với số lượng tác tử và không gian trạng thái tương đối lớn. Ở Việt Nam, tuy mới tiếp cận với công nghệ tác tử nhưng một số tác giả cũng đã cho ra những kết quả đáng ghi nhận có thể kể ra các tác giả như Từ Minh Phương với giải thuật Q-Phân tán [19]; Nguyễn Linh Giang với giải thuật Q- mờ cho hệ đa tác tử [10], các kết quả của các tác giả đều đem lại những giá trị khoa học đáng kể và nền tảng của các kết quả đều được dựa trên Q-Learning.

Trong luận văn này, chúng tôi nghiên cứu về đơn tác tử, hệ đa tác tử và ứng dụng thuật toán Q-Learning truyền thống trong việc huấn luyện đa tác tử với sự có mặt của một tác tử theo dõi. Các ứng dụng cho thuật toán Q-Learning truyền thống thường ứng dụng cho đơn tác tử và trạng thái đích cần đạt tới là cố định. Trong đề tài này, chúng tôi sẽ cố gắng áp dụng Q-Learning cho hệ đa tác tử với trạng thái đích liên tục thay đổi.

3. Mục đích và đối tượng nghiên cứu

3.1. Mục đích nghiên cứu

Thực hiện đề tài này, mục đích đầu tiên của luận văn là tổng hợp được các tài liệu về công nghệ tác tử một cách đầy đủ, khái quát và có hệ thống. Mặt khác, ứng dụng được các thuật toán huấn luyện tác tử vào một số dạng bài toán khác nhau, cài đặt thử nghiệm và đánh giá mức độ hiệu quả của thuật toán Q-Learning trong việc huấn luyện đa tác tử.

3.2. Đối tượng nghiên cứu

Bên cạnh những vấn đề tổng quan về đơn tác tử và hệ đa tác tử, đối tượng nghiên cứu chính của đề tài đi sâu vào nghiên cứu về các vấn đề sau:

1. Tác tử thông minh và các loại kiến trúc của tác tử thông minh
2. Tương tác giữa các tác tử trong cùng một hệ đa tác tử
3. Các tác tử phối hợp với nhau theo những quy tắc nào trong hệ đa tác tử
4. Thuật toán Q-Learning và ứng dụng của nó.

4. Cấu trúc của luận văn

Ngoài phần mở đầu và kết luận, phần nội dung của luận văn gồm có 3 chương:

Chương 1: Tổng quan về tác tử và hệ đa tác tử

Chương 2: Phối hợp và tương tác trong hệ đa tác tử

Chương 3: Phương pháp huấn luyện đa tác tử với sự có mặt của tác tử theo dõi và cài đặt thử nghiệm.

CHƯƠNG 1: TỔNG QUAN VỀ TÁC TỬ VÀ ĐA TÁC TỬ

1.1. Tác tử

1.1.1. Định nghĩa tác tử

Cho đến nay, có rất nhiều cách định nghĩa về tác tử, các ý kiến trái chiều nhau nguyên nhân chủ yếu xuất phát từ những yêu cầu khác nhau trong một số ứng dụng cụ thể. Những mâu thuẫn này là điều xảy ra rất nhiều trong ngành khoa học máy tính. Chính những ý kiến đa chiều đó của các nhà chuyên môn đã cho thấy sự phong phú về khả năng ứng dụng cũng như lý thuyết của công nghệ phần mềm hướng tác tử.

Định nghĩa thường được sử dụng nhất phát biểu như sau: "*Tác tử (Agent) là hệ thống tính toán hoạt động tử chủ trên một môi trường nào đó, có khả năng cảm nhận và tác động vào môi trường"* [6].

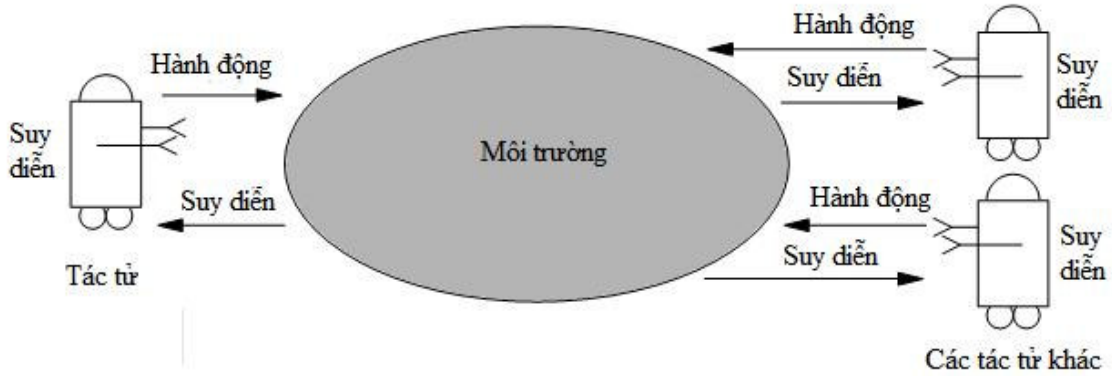
Chúng ta quan tâm đến một số các điểm quan trọng sau của định nghĩa trên.

- Vấn đề đầu tiên, tác tử là hệ thống tính toán, hệ thống này có thể là phần cứng, phần mềm hoặc kết hợp cả phần cứng và phần mềm. Đối với tác tử là phần mềm có thể là chương trình máy tính, các luồng thực hiệu (Thread), đối với tác tử phần cứng thông thường là các Robot, các thiết bị giám sát giao thông.
- Vấn đề thứ hai, khi nói đến tác tử tồn tại và hoạt động trong môi trường, định nghĩa trên nhấn mạnh khả năng cảm nhận và tác động lại môi trường một cách trực tiếp và có thể làm thay đổi môi trường. Tác tử nhận thông tin tử môi trường qua các cơ quan cảm nhận và tác động lại môi trường qua các cơ quan tác động. Các tác tử là phần cứng cơ quan cảm nhận thường là thiết bị cảm biến (cảm biến nhiệt, âm), thiết bị nhận dạng hay đơn thuần là các camera, cơ quan tác động thường là các bộ phận cơ học, quang học, âm thanh. Đối với tác tử là phần mềm môi trường hoạt động chính là máy tính hay mạng máy tính. Việc cảm

nhận và tác động vào môi trường của tác tử được thực hiện thông qua lời gọi hệ thống.

- Vấn đề thứ ba, đó là tính tự chủ (tự trị) của tác tử, đây là một thuộc tính quan trọng của tác tử, nó mang tính đặc trưng của tác tử. Sự tự chủ ở đây chính là khả năng hành động không cần đến sự can thiệp của người dùng hay bất kỳ một tác nhân nào khác. Tác tử có thể tự kiểm soát hành vi của mình trong suốt quá trình hoạt động, trước những vấn đề này sinh trong môi trường hoạt động chúng có thể tự đưa ra quyết định cho hành động của mình. Mặt khác tính tự chủ còn được biểu hiện ở khả năng học của tác tử.

Như vậy, với những đặc điểm tồn tại và hành động tự chủ trong môi trường tác tử có thể độc lập thực hiện một nhiệm vụ nào đó thay cho con người hoặc các tử khác [9].



Hình 1.1: *Kiến trúc chung của tác tử*

1.1.2 Các đặc điểm khác của tác tử

Ngoài các đặc điểm quan trọng nhất của tác tử được nhắc tới trong định nghĩa, tác tử còn có thêm những đặc điểm sau:

Khả năng tự học: Là khả năng thu thập kiến thức mới từ kinh nghiệm thu lượm được, kết quả của việc tự học phải giúp cho tác tử hành động tốt hơn, hiệu quả hơn

Tính thích ghi: Là khả năng tồn tại và hoạt động hiệu quả khi môi trường thay đổi.

Khả năng di chuyển: Là khả năng di chuyển mã nguồn của tác tử từ máy tính này sang máy tính khác hay nút mạng này sang nút mạng khác đồng thời vẫn giữ nguyên trạng thái.

1.1.3. Môi trường hoạt động của tác tử

Tác tử được xây dựng để hoạt động trong một môi trường nào đó, chính vì thế tính chất, đặc điểm của môi trường và mối quan hệ giữa tác tử với môi trường chính là yếu tố quyết định đến việc nghiên cứu cũng như triển khai ứng dụng. Hầu hết các nghiên cứu đều khẳng định tác tử và môi trường có quan hệ như sau: tác tử cảm nhận môi trường, suy luận và sau đó thực hiện hành động tác động vào môi trường. Quá trình đó được lặp lại cho đến hết vòng đời của một tác tử.

Chính vì sự gắn bó mật thiết giữa môi trường và tác tử cho nên vấn đề phân loại môi trường hoạt động của tác tử cũng được đặt ra.

1.1.3.1. Môi trường có thể tiếp cận đầy đủ và không thể tiếp cận đầy đủ

Môi trường được gọi là có thể tiếp cận đầy đủ nếu tác tử có thể thu thập đầy đủ và chính xác thông tin về trạng thái của môi trường thông qua cơ quan cảm nhận. Môi trường có thể tiếp cận đầy đủ là những môi trường tương đối đơn giản và thuần nhất. Môi trường không thể tiếp cận đầy đủ là những môi trường có độ phức tạp từ trung bình đến phức tạp, ví dụ: Thế giới thực vật lý, Internet

1.1.3.2. Môi trường xác định và không xác định

Nếu trạng thái tiếp theo của môi trường hoàn toàn xác định bởi trạng thái hiện tại và hành động của tác tử tại thời điểm *t* thì môi trường được gọi là xác định. Như vậy, trước mỗi hành động của mình tác tử đều biết trước kết quả. Đối với trường hợp môi trường không xác định, cùng một hành động có thể cho ra những kết quả khác nhau, thậm trí cho những kết quả không mong muốn. Với loại môi trường không xác định thường gây khó khăn trọng việc thiết kế tác tử.

1.1.3.3. Môi trường phân đoạn và không phân đoạn

Trong môi trường có phân đoạn hoạt động của tác tử được chia theo thời gian thành từng đoạn riêng biệt, không phụ thuộc vào nhau. Hiệu quả hành động trong từng đoạn chỉ phụ thuộc vào đoạn tác tử đang xét chứ không phụ thuộc vào đoạn khác. Môi trường không phân đoạn thường phức tạp hơn vì tác tử phải quan tâm đến các đoạn có liên quan tới đoạn đang xét.

1.1.3.4. Môi trường tĩnh và động

Môi trường động là môi trường có thể thay đổi trong khi tác tử đang suy diễn để lựa chọn chiến lược hành động. Môi trường tĩnh thì ngược lại, tác tử không cần quan tâm đến môi trường cũng như giới hạn về thời gian trước khi ra quyết định cho chiến lược hành động của mình. Qua đó ta thấy việc triển khai ứng dụng trên môi trường tĩnh là thuận lợi hơn trên môi trường động.

1.1.3.5. Môi trường rời rạc và liên tục

Nếu số lượng cảm nhận và các hành động có thể của tác tử trong môi trường là hữu hạn và luôn xác định thì là môi trường rời rạc. Ngược lại trong trường hợp môi trường là liên tục.

Qua những phân loại về môi trường chúng ta thấy với mỗi đặc điểm môi trường khác nhau sẽ kéo theo yêu cầu thiết kế tác tử khác nhau để đảm bảo sự hoạt động hiệu quả và chính xác của tác tử. Việc xác định môi trường hoạt động của tác tử là bước quan trọng và là một trong những bước đầu tiên phải làm trong quá trình thiết kế tác tử.

1.1.4. Tác tử thông minh

1.1.4.1. Tác tử thông minh là gì?

Tác tử thông minh là tác tử có khả năng hoạt động linh hoạt và mềm dẻo để thực hiện nhiệm vụ được giao [6].

Tính linh hoạt của tác tử được thể hiện bởi ba đặc điểm chính sau:

- Tính phản xạ: Là khả năng phản ứng kịp thời với các thay đổi của môi trường.

- Tính chủ động: Tác tử chủ động trong hành động của mình, tự tìm ra phương án hành động tối ưu nhằm đạt được kết quả tốt nhất.

- Tính cộng đồng: Là khả năng tương tác giữu người dùng hoặc tác tử khác để lấy thông tin hoặc cung cấp thông tin cho đối tác.

Nếu xét từng đặc điểm riêng lẻ thì ba đặc điểm này không có gì mới trong các phương pháp lập trình cũ. Trong thực tế có rất nhiều hệ thống phần mềm không xây dựng theo hướng tác tử nhưng mang một trong những đặc điểm trên. Tuy nhiên, khi xây dựng phần mềm tác tử thông minh thì cần phải hội đủ các đặc điểm này.

1.1.4.2. Cảm nhận, suy diễn và tác động vào môi trường

1.1.4.2.1. Cảm nhận môi trường

Việc cảm nhận môi trường giúp cho tác tử biết được tình trạng của môi trường đang diễn ra như thế nào, từ đó tác tử sẽ đưa được ra quyết định hành động của mình. Đối với các tác tử là phần cứng (Robot) việc cảm nhận môi trường thường bằng các camera, thiết bị cảm ứng; đối với các tác tử là phần mềm việc cảm nhận môi trường thường là những thông điệp từ hệ điều hành. Những thông tin mà tác tử cảm nhận được từ môi trường không phải tất cả đều là có ích, chính vì thế việc chọn lọc thông tin là một yêu cầu đặt ra cho cơ chế cảm nhận môi trường của tác tử.

1.1.4.2.2. Suy diễn (cơ chế ra quyết định)

Quá trình ra quyết định của một tác tử có thể được mô tả như sau. Giả sử thời gian được chia thành những khoảng rời rạc t_0, t_1, ..., t_n. Tại mỗi thời điểm đó tác tử phải lựa chọn hành động từ tập hữu hạn các hành động của tác tử. Nhờ cơ quan cảm nhận tác tử thu được những cảm nhận về môi trường. Giả sử tại các thời điểm t_0, t_1, ... t_n cảm nhận của tác tử về môi trường lần lượt là p_0, p_1, ... ,p_n với p_i thuộc P, P là tập các cảm nhận có thể có của tác tử. Tại thời điểm t_i, tất cả

những cảm nhận của tác tử là chuỗi cảm nhận $s_i = < p_0, p_1, ... , p_i>$. Giả sử tập hành động có thể của tác tử là $A = \{a_1, a_2, ... , a_n\}$. Tại mỗi thời điểm t_i tác tử có thể chọn một hành động $a_i \in A$. Việc tác tử lựa chọn hành động nào phụ thuộc vào chuỗi cảm nhận S_i tại thời điểm t_i.

Tác tử suy diễn trước khi quyết định hành động được chia ra làm nhiều dạng, việc phân chia này phụ thuộc vào cách thức cảm nhận và hành động của tác tử. Ta có thể phân loại như sau:

- Tác tử phản xạ: Là tác tử hành động dựa trên cảm nhận hiện tại mà không cần quan tâm đến chuỗi cảm nhận trước đó.Ví dụ bộ phận cảm ứng của hệ thống cửa tự động. Bộ phận này hoạt động dựa trên nguyên tắc trạng thái môi trường được chia làm hai dạng, có người và không có người, nếu có người cửa sẽ tự động mở. Nguyên lý hoạt động của loại tác tử này nằm trong kiến trúc phản xạ, đây là loại kiến trúc không sử dụng cơ chế suy diễn phức tạp. Kiến trúc phản xạ được biết đến nhiều nhất là *kiến trúc gộp* (subsumption architecture) do Rodney Brooks đề xuất lần đầu tiên năm 1986. Kiến trúc gộp này được mô tả như sau:

- Quá trình quyết định của tác tử được thực hiện dựa trên tập các hàm hành động gọi là hành vi thực hiện nhiệm vụ (task accomplishment behaviours). Mỗi hành vi thực chất là một hàm hành động. Mỗi hành vi được tổ chức như một môđun và có mục đích thực hiện nhiệm vụ nhất định. Trong kiến trúc nguyên bản của Brooks, mỗi môđun hành vi được cài đặt như một máy trạng thái hữu hạn. Hành vi của tác tử được biểu diễn bằng các luật hoặc quy tắc đơn giản dưới dạng: *tình huống* → *Hành động*. Mỗi một luật này ánh xạ từ một trạng thái cảm nhận được thành một hành động. Vì cơ chế ra quyết định của tác tử do nhiều môđun hành vi gộp lại nên kiến trúc này có tên là kiến trúc gộp.