

GIỚI THIỆU THUẬT TOÁN CHUỖI DI TRUYỀN TRONG NGHIÊN CỨU ỨNG DỤNG ĐỂ GIẢI CÁC MÔ HÌNH TỐI ƯU
GENETIC ALGORITHM FOR OPTIMIZATION OBJECTIVES IN OPERATIONAL RESEARCH

TSKH. NGUYỄN THANH THÙY
Trung tâm Đào tạo Logistics, Trường ĐHHH

Tóm tắt

Thuật toán chuỗi di truyền là một kỹ thuật của khoa học máy tính nhằm tìm kiếm giải pháp thích hợp cho các bài toán tối ưu tổ hợp, là một trong các thuật toán được sử dụng rộng rãi nhất trong kinh tế, sinh học, nông nghiệp, giao thông vận tải và logistics. Thuật toán này vận dụng các nguyên lý của tiến hóa như chọn lọc, sinh sản, lai giống và đột biến nhằm tìm ra bộ phương án từ tiệm cận tối ưu đến tối ưu để giúp nhà quản lý sản xuất có nhiều lựa chọn hơn trong ra quyết định.

Abstract

Genetic algorithm is a technical computer science for combinatorial optimization problems. This algorithm is widely used in economics, biology, agriculture industrial, transportation and logistics. This algorithm applies the principles of evolution such as genetic selection, reproduction, crossover and mutations, in order to find a set of approximately optimal solutions and optimal solutions in order to help managers to have more choices in decision making.

1. Đặt vấn đề

Ngày nay trong nghiên cứu khoa học, thuật toán được sử dụng rộng rãi để tìm ra lời giải cho các mục tiêu khác nhau. Các mục tiêu này được toán học hóa thành các hàm mục tiêu toán học hay nói cách khác là các mô hình toán tối ưu. Trong hàng triệu các nghiên cứu ứng dụng của hầu hết các ngành từ y học, nông nghiệp, sinh học, khoa học cơ bản, quân sự, hải dương học, kinh tế hay giao thông vận tải, các mô hình tối ưu luôn được sử dụng rộng rãi. Tuy nhiên, trong các nghiên cứu khoa học cơ bản các nhà khoa học có khuynh hướng sử dụng các thuật toán tối ưu để tìm ra các lời giải tối ưu cho mô hình, thì trong các nghiên cứu ứng dụng các nhà khoa học lại có khuynh hướng sử dụng các thuật toán tối ưu để tìm ra một bộ các phương án từ tiệm cận tối ưu đến tối ưu cho mô hình. Khuynh hướng này bắt nguồn từ các thực tế sau:

(1) Các phương án tối ưu thường đưa ra rất ít lựa chọn cho người sử dụng. Ngay cả khi lựa chọn được đưa ra, người sử dụng không thể áp dụng do các hạn chế trên thực tế. Trong khi các phương án tiệm cận tối ưu có thể đưa ra hàng nghìn/trăm nghìn lựa chọn cho cùng một vấn đề mà không có sự chênh lệch đáng kể trên thực tế. Ví dụ, nếu phương án thời gian tối ưu của một kế hoạch sản xuất là 130 phút và phải tuân thủ duy nhất một cách bố trí, thì có hàng nghìn cách bố trí khác cho phương án này với khoảng thời gian từ 128-130 phút. Trong toán học, 2 phút là cả một sự khác biệt lớn, nhưng trên thực tế, phần lớn các nhà quản lý sẵn sàng lựa chọn phương án 128 phút mà việc bố trí sản xuất thuận tiện hơn cho cơ sở mình.

(2) Rất nhiều bài toán tối ưu cho kết quả vô nghiệm nếu cố tình tìm kiếm phương án tối ưu, mà trên thực tế sản xuất không bao giờ chấp nhận tình huống "vô nghiệm". Trong khi phương án tiệm cận tối ưu luôn luôn tồn tại để lựa chọn trong bất kỳ tình huống nào.

Chính vì các lý do trên, các thuật toán giúp tìm ra bộ phương án từ tiệm cận tối ưu đến tối ưu được sử dụng rất rộng rãi trong nghiên cứu ứng dụng hiện nay ví dụ như thuật toán Kỹ thuật mô phỏng (Simulated Annealing Algorithm), thuật toán Tìm kiếm Tabu (Tabu Search Algorithm), hay thuật toán Chuỗi di truyền (Genetic Algorithm). Có thể tham khảo các thuật toán này ở hầu hết các nghiên cứu trong hơn 2.500 Tạp chí trên cổng thông tin khoa học Scencedirect (<http://www.sciencedirect.com/>), hoặc hàng ngàn Tạp chí khác trên Springer (<http://www.springer.com>), hoặc trên các cổng thông tin khoa học khác. Mỗi thuật toán này lại có hàng ngàn cách ứng dụng khác nhau tùy thuộc vào từng vấn đề nghiên cứu cụ thể và tính sáng tạo của người nghiên cứu ứng dụng. Đi kèm với các thuật toán này chính là chương trình máy tính (computer program) được lập bởi người nghiên cứu, thường bằng các ngôn ngữ như C, C++,

hoặc Lingo/Lindo với đầu vào là các số liệu thực tế, đầu ra là các phương án tốt có thể lựa chọn. Bằng cách lựa chọn các vòng lặp chương trình khác nhau, người nghiên cứu có thể đưa ra các phương án giải tiệm cận tối ưu khác nhau cho cùng một mô hình.

Trong các thuật toán này, thuật toán chuỗi di truyền (Genetic Algorithm) là một trong các thuật toán được sử dụng rộng rãi nhất trong kinh tế, sinh học, nông nghiệp, giao thông vận tải và logistics.

2. Giới thiệu về thuật toán chuỗi di truyền

Thuật toán chuỗi di truyền (hay còn được dịch là giải thuật di truyền) là một kỹ thuật của khoa học máy tính nhằm tìm kiếm giải pháp thích hợp cho các bài toán tối ưu tổ hợp (combinatorial optimization). Thuật toán này vận dụng các nguyên lý của tiến hóa như chọn lọc, sinh sản, lai giống và đột biến nhằm tìm ra bộ phương án từ tiệm cận tối ưu đến tối ưu để giúp nhà quản lý sản xuất có nhiều lựa chọn hơn trong ra quyết định.

Thuật toán chuỗi di truyền thường được ứng dụng nhằm sử dụng ngôn ngữ máy tính để mô phỏng quá trình tiến hoá của một tập hợp những đại diện trừu tượng (gọi là những nhiễm sắc thể) của các giải pháp có thể (gọi là những cá thể) cho bài toán tối ưu. Tập hợp này sẽ tiến triển theo hướng chọn lọc những giải pháp tốt hơn.

Thông thường, những giải pháp được thể hiện dưới dạng nhị phân với những chuỗi 0 và 1, nhưng lại mang nhiều thông tin mã hóa khác nhau. Quá trình tiến hóa xảy ra từ một tập hợp những cá thể hoàn toàn ngẫu nhiên ở tất cả các thế hệ. Trong từng thế hệ, tính thích nghi của tập hợp này được ước lượng, nhiều cá thể được chọn lọc định hướng từ tập hợp hiện thời (dựa vào thể trạng), được sửa đổi (bằng đột biến hoặc tổ hợp lại) để hình thành một tập hợp mới. Tập hợp này sẽ tiếp tục được chọn lọc lặp đi lặp lại trong các thế hệ kế tiếp của thuật toán.

Thuật toán chuỗi di truyền được sử dụng nhằm:

(1) Tối ưu hóa hoặc nâng cao hiệu quả của một hệ thống đang khai thác

Thuật toán chuỗi di truyền có thể được sử dụng để lựa chọn các ràng buộc nhằm tối ưu hóa hiệu quả của một hệ thống. Thông thường hệ thống này là một hệ thống thực tế như hệ thống phân phối gas, hệ thống bán hàng cơ động, điều chuyển vị trí thiết bị vận vận.

(2) Thử nghiệm hoặc ứng dụng các mô hình định lượng

Đây là khía cạnh ứng dụng ít được nhắc tới trong thuật toán chuỗi di truyền. Nghiên cứu khoa học về một vấn đề nhất định được coi như một quá trình tác động qua lại. Mô hình diễn giải hoặc mô tả được thiết lập cùng với việc thu thập số liệu nhằm thử nghiệm mô hình. Khi tìm ra các kết quả trái ngược, mô hình sẽ được sửa đổi và quá trình này được lặp lại cho tới khi giải quyết xong vấn đề.

(3) Giải các bài toán cực đại hóa (maximization) hoặc cực tiểu hóa (minimization)

Lưu ý rằng thuật toán chuỗi di truyền cần được biến đổi khi giải các bài toán có hàm mục tiêu trái ngược.

3. Các bước của thuật toán chuỗi di truyền

Thuật toán chuỗi di truyền là thuật toán có nội dung vô cùng rộng, với hàng ngàn cách ứng dụng khác nhau tùy thuộc vào từng vấn đề cụ thể. Trong bài viết này, tác giả chỉ giới thiệu khái quát các bước cơ bản khi áp dụng thuật toán chuỗi di truyền để giải bài toán tối ưu.

3.1. Chọn lọc đại diện (Representation)

Các biến số sẽ được chọn sẵn vào một số địa chỉ nhất định chọn để thiết lập các nhiễm sắc thể (chromosome), kết hợp với thứ tự của nhiễm sắc thể đó. Chiều dài của chuỗi ký tự thể hiện số nghiệm cần xử lý, kèm theo một số bộ phận kết nối bằng số không. Ví dụ trong bài toán điều tàu vào cầu tàu, Nishimura và các tác giả khác (2008)^[5] đã xây dựng chuỗi gồm số tàu vào cảng, số cầu tàu và thứ tự vào cầu giá định (Hình 1).

3.2. Xác định chỉ số biến đổi (Fitness)

Chỉ số biến đổi phản ánh khả năng tốt hơn của một cá thể so với các cá thể khác trong tập được lựa chọn tại bước 3.1. Nếu hàm mục tiêu là hàm min thì chỉ số biến đổi càng lớn càng tốt. Thông thường, lựa chọn chỉ số biến đổi thông qua công thức: $f(x) = 1/(1+\exp(y(x)/a))$ trong đó $y(x)$ là hàm mục tiêu, a là một thông số bất kỳ biểu diễn số bước lặp của thuật toán.

3.3. Các toán tử của thuật toán chuỗi di truyền (Genetic algorithm operators)

GENE : SHIP # 1~9, 0(ZERO)
 (CELL #) 1 2 3 4 5 6 7 8 9 10
 CHROMOSOME

2	8	5	9	0	4	7	3	1	6
---	---	---	---	---	---	---	---	---	---

 BERTH # 1 1 1 1 2 2 2 2 2
 ORDER OF SERVICE 1 2 3 4 1 2 3 4 5

Hình 1. Chọn lọc đại diện

Parents Chromosome A

2	8	5	9	0	4	7	3	1	6
---	---	---	---	---	---	---	---	---	---

 Chromosome B

6	3	1	2	5	7	0	8	9	4
---	---	---	---	---	---	---	---	---	---

Children Chromosome C'

2	8	1	2	5	7	7	3	1	6
---	---	---	---	---	---	---	---	---	---

 Chromosome D'

6	3	5	9	0	4	0	8	9	4
---	---	---	---	---	---	---	---	---	---

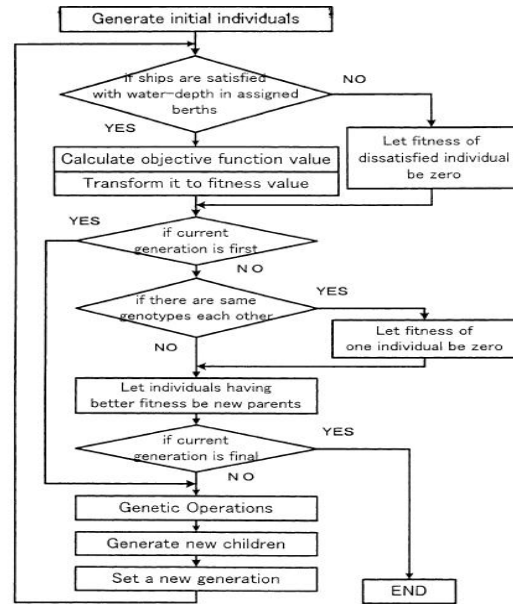
Chromosome C

9	8	1	2	5	7	4	3	0	6
---	---	---	---	---	---	---	---	---	---

 Chromosome D

6	3	5	9	0	4	1	8	2	7
---	---	---	---	---	---	---	---	---	---

Hình 2. Trộn chuỗi 2 điểm khi bố trí tàu vào cầu tàu



Hình 4. Thuật toán chuỗi di truyền

	Point 1										Point 2				
Parents A	17	2	9	5	3	6	0	49	54	53	61	97	98	101	105
Parents B	6	5	9	17	3	2	54	61	53	49	0	98	105	97	101
Children C'	17	2	9	17	3	2	54	61	53	49	0	98	98	101	105
Children D'	6	5	9	5	3	6	0	49	54	53	61	97	105	97	101
C	17	2	9	5	3	6	54	61	53	49	0	97	98	101	105
D	6	5	9	17	3	2	0	49	54	53	61	98	105	97	101

Hình 3. Trộn chuỗi 2 điểm khi bố trí cần trục hoạt động trên bãi

(1) Lựa chọn sinh sản (Reproduction)

Lựa chọn sinh sản là quá trình trong đó mỗi chuỗi được lựa chọn tùy thuộc vào chỉ số biến đổi của chuỗi đó. Chuỗi nào có chỉ số biến đổi tốt hơn sẽ được sao chép nhiều hơn cho thế hệ sau. Có nhiều cách lựa chọn kết hợp chuỗi khác nhau, trong đó 2 cách phổ biến nhất là Lựa chọn kiểu bánh xe Roulette (Roulette selection) và lựa chọn kiểu Cuộc đấu vòng tròn (Tournament selection) (xem T.T.Thủy và các tác giả khác (2008) để biết thêm chi tiết^[4]).

(2) Trộn chuỗi (Crossover)

Sau khi lựa chọn chuỗi sao chép chuỗi nhiễm sắc thể tạo thành một quần thể mới, trộn chuỗi được thực hiện để tạo ra các chuỗi mới tốt hơn. Có nhiều cách trộn chuỗi, trong đó 2 cách thông dụng nhất là Trộn chuỗi mặt nạ (Mask-Crossover) và Trộn chuỗi 2 điểm (two points crossover). Hình 2 và 3 biểu diễn cách trộn chuỗi 2 điểm khi bố trí tàu vào cầu tàu và khi bố trí cần trục xếp dỡ container trên bãi.

(3) Biến đổi chuỗi (Mutation)

Biến đổi chuỗi là quá trình tạo ra các thay đổi ngẫu nhiên bằng cách xác định thông số tỷ lệ biến đổi (mutation rate). Nếu các chuỗi nào đạt được tỷ lệ này thì mới áp dụng biến đổi chuỗi để thu về chuỗi mới tốt hơn. Khi biến đổi chuỗi, các giá trị nằm trong chuỗi được đổi vị trí lẫn nhau tạo ra

sự thay đổi về thứ tự, từ đó tạo ra một chuỗi mới nhiều ưu thế hơn. Hay nói cách khác một phương án mới nhiều ưu thế hơn.

Sau một số các vòng lặp nhất định, giá trị phương án thu được càng ngày càng gần với giá trị phương án tối ưu. Hình 4 biểu diễn toàn bộ quá trình áp dụng thuật toán chuỗi di truyền.

4. Các bước thực hiện chương trình trên máy tính

Chương trình trên máy tính thường được thiết lập bởi ngôn ngữ C, C++ hay Lingo/Lindo cho hệ thống máy chủ Unix, hoặc APS, nhưng cũng có thể chạy được trên các máy tính thông thường có tốc độ nhanh nếu các case study không có bộ số liệu lớn. Các bước thực hiện chương trình trên máy tính bao gồm:

- (1) Xác định số liệu đầu vào;
- (2) Thiết lập các tập giả định (case study);
- (3) Chạy chương trình;
- (4) Đọc thông số đầu ra và đưa ra kết luận.

5. Kết luận

Thuật toán chuỗi di truyền là một thuật toán có ứng dụng mạnh và rộng rãi trên thế giới nhưng vẫn còn khá xa lạ với các nghiên cứu ứng dụng ở Việt Nam. Nắm được thuật toán này có thể giải quyết được rất nhiều ứng dụng thực tế và đưa ra các kết quả tiệm cận tối ưu với thời gian ngắn. Vì vậy, thuật toán này cần được tìm hiểu và đưa vào các nghiên cứu nhiều hơn nữa.

TÀI LIỆU THAM KHẢO

- [1] A. Blanco, M. Delgado, M.C. Pegalaja "A genetic Algorithm to obtain optimal recurrent neural network", International Journal of Approximated Reasoning, Vol. 23, pp. 67-83. 2000
- [2] Lance Chambers, "The Practical Handbook of Genetic Algorithm", Printed by Chapman & Hall/CRC, 2000
- [3] Nguyen Thanh Thuy, Etsuko Nishimura, Akio Imai, "A Heuristic Approach for the Containership Stowage Problem in Multiple Port-of-call Route", The Journal of Japan Institute of Navigation, Vol. 120, pp. 164 -172, ISSN: 0388-7405, 2009
- [4] Nguyen Thanh Thuy, Akio Imai, Etsuko Nishimura, Stratos Papadimitriou, "Multi-objective model of containership loading plan considering the effectiveness of crane employed", 10th International Conference on Applications of Advanced Technologies in Transportation E - Proceedings, Greece. 2008
- [5] Nishimura, E., Imai, A., Papadimitriou, S., "Berth allocation planning in the public berth system by genetic algorithms", European Journal of Operation Research 131, pp. 282-292, 2001

Người phản biện: ThS. Bùi Thanh Tùng