

# Thuật toán di truyền – Bài khảo sát

Nguyễn Thị Lanh\*

\*ThS. Công nghệ thông tin, Trường Đại học Hải Phòng

Received: 25/12/2023; Accepted: 05/01/2024; Published: 12/01/2024

**Abstract:** This article introduces Genetic Algorithm and its basic functions. The basic function of Genetic Algorithm includes different steps such as selection, crossover, mutation. This article also focuses on comparing genetic algorithms with other problem solving techniques. Details of the laboratories essentially focused on Genetic Algorithm research and development are also included. Details of the labs include various projects done on Genetic Algorithms.

**Keywords:** Selection, hybridization, mutation, problem solving techniques.

## 1. Đặt vấn đề

Thuật toán di truyền là một kỹ thuật lập trình hình thành nền tảng của nó từ quá trình tiến hóa sinh học. [6] Thuật toán di truyền về cơ bản được sử dụng như một chiến lược giải quyết vấn đề nhằm đưa ra giải pháp tối ưu. Chúng là cách tốt nhất để giải quyết vấn đề mà ít người biết đến. Chúng sẽ hoạt động tốt trong bất kỳ không gian tìm kiếm nào vì chúng tạo thành một thuật toán rất tổng quát. Điều duy nhất cần biết là tình huống cụ thể là giải pháp hoạt động rất tốt và thuật toán di truyền sẽ tạo ra giải pháp chất lượng cao. Các thuật toán di truyền sử dụng các nguyên tắc chọn lọc và tiến hóa để tạo ra một số giải pháp cho một vấn đề nhất định.

- Cá nhân-Bất kỳ giải pháp nào có thể
- Dân số-Nhóm tất cả cá nhân
- Tìm kiếm không gian-Tất cả các giải pháp có thể cho vấn đề

- Nhiệm sắc thể-Kế hoạch chi tiết cho một cá nhân

Những ứng viên có triển vọng này được giữ lại và cho phép nhân giống. Từ chúng có nhiều bản sao được tạo ra, nhưng các bản sao không hoàn hảo; những thay đổi ngẫu nhiên được đưa ra trong quá trình sao chép. Những đứa con kỹ thuật số này sau đó sẽ chuyển sang thế hệ tiếp theo, tạo thành một nhóm giải pháp ứng cử viên mới và phải trải qua vòng đánh giá mức độ phù hợp thứ hai. Những giải pháp ứng cử viên đã trở nên tồi tệ hơn hoặc không được cải thiện do những thay đổi đối với mã của chúng sẽ lại bị xóa; nhưng một lần nữa, hoàn toàn là ngẫu nhiên, các biến thể ngẫu nhiên được đưa vào quần thể có thể đã cải thiện một số cá thể, khiến họ trở thành những giải pháp tốt hơn, hoàn chỉnh hơn hoặc hiệu quả hơn cho vấn đề trước mắt. Một lần nữa, những cá thể chiến thắng này lại được chọn và sao chép sang thế hệ tiếp theo với những thay đổi ngẫu nhiên và quá trình này lặp lại. Kỳ vọng là mức độ thích hợp trung bình của dân số sẽ tăng lên

sau mỗi vòng, và do đó, bằng cách lặp lại quá trình này trong hàng trăm hoặc hàng nghìn vòng, có thể tìm ra các giải pháp rất tốt cho vấn đề.

[1] Các thuật toán di truyền đã được chứng minh là một chiến lược giải quyết vấn đề cực kỳ hiệu quả và thành công. Các thuật toán di truyền đã được sử dụng trong nhiều lĩnh vực khác nhau để tìm ra giải pháp cho những vấn đề khó khăn hơn những vấn đề mà con người phải đối mặt. Do đó, các giải pháp họ đưa ra thường hiệu quả hơn, tinh tế hơn hoặc phức tạp hơn bất kỳ giải pháp nào có thể so sánh được mà một kỹ sư con người có thể tạo ra.

## 2. Nội dung nghiên cứu

### 2.1. Lựa chọn

[2] Có nhiều kỹ thuật khác nhau mà thuật toán di truyền có thể sử dụng để chọn lọc các cá thể để sao chép sang thế hệ tiếp theo.

*Lựa chọn tinh hoa:* Những thành viên khỏe mạnh nhất của mỗi thế hệ được đảm bảo sẽ được chọn.

- Đặc điểm-Khía cạnh có thể có của một cá nhân
- Alen-Các cài đặt có thể có cho một đặc điểm
- Quy tích-Vị trí của một gen trên nhiễm sắc thể
- Bộ gen-Bộ sưu tập tất cả nhiễm sắc thể cho một cá nhân

Đầu vào của GA là một tập hợp các giải pháp tiềm năng cho vấn đề đó, được mã hóa theo kiểu nào đó và một số liệu gọi là *chức năng tập thể* cho phép mỗi ứng viên được đánh giá một cách định lượng.

Những ứng cử viên này có thể là những giải pháp đã được biết là có hiệu quả, với mục đích của GA là cải thiện chúng, nhưng chúng thường được tạo ra một cách ngẫu nhiên.

GA sau đó đánh giá từng ứng viên theo hoạt động của chức năng thích hợp; ứng viên có thể lực tốt có cơ hội trúng tuyển cao hơn ứng viên có thể lực trung bình. Các chức năng khác nhau về cơ bản được sử dụng để kiểm tra sức khỏe của bất kỳ cá nhân cụ thể

nào. Những cá nhân có thể lực cao này có thể được coi là những ứng cử viên đầy triển vọng.

**Lựa chọn cân đối về thể chất:** Những cá nhân khỏe mạnh hơn thì có nhiều khả năng hơn, nhưng không chắc chắn, được chọn.

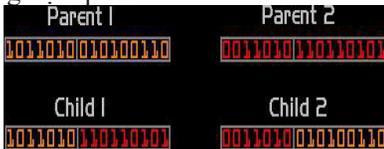
**Lựa chọn bánh xe Roulette:** Một hình thức lựa chọn cân xứng với mức độ thích nghi, trong đó cơ hội để một cá thể được chọn tỷ lệ thuận với mức độ thích nghi của nó lớn hơn hoặc nhỏ hơn mức độ thích hợp của đối thủ cạnh tranh.

**Lựa chọn tỷ lệ:** Khi mức độ thích hợp trung bình của dân số tăng lên, sức mạnh của áp lực chọn lọc cũng tăng lên và chức năng thích nghi trở nên phân biệt đối xử hơn. Phương pháp này có thể hữu ích trong việc đưa ra sự lựa chọn tốt nhất sau này khi tất cả các cá thể đều có độ thích nghi tương đối cao và chỉ những khác biệt nhỏ về độ thích nghi mới có thể phân biệt được cá thể này với cá thể khác.

**Lựa chọn giải đấu:** Các nhóm nhỏ gồm các cá thể được chọn từ quần thể lớn hơn và các thành viên của mỗi nhóm nhỏ cạnh tranh với nhau. Chỉ một cá thể từ mỗi nhóm con được chọn để sinh sản.

### 2.2. Sự giao nhau

Sau khi chọn lọc xong các cá thể, việc tiếp theo là tạo ra con cái. [2] Giải pháp phổ biến nhất cho vấn đề này là cái gọi là phân tần, và mặc dù có nhiều loại phân tần khác nhau, nhưng loại phổ biến nhất là phân tần đơn điểm. Trong phép lai điểm đơn, hãy chọn một locus mà tại đó bạn trao đổi các alen còn lại từ bố mẹ này sang bố mẹ khác. Điều này phức tạp và được hiểu rõ nhất bằng trực quan.



Con cái lấy một phần nhiễm sắc thể từ cha và mẹ. Điểm mà nhiễm sắc thể bị đứt phụ thuộc vào điểm giao nhau được chọn ngẫu nhiên. Phương pháp đặc biệt này được gọi là lai điểm đơn vì chỉ tồn tại một điểm chéo. Đôi khi chỉ có con 1 hoặc con 2 được tạo ra, nhưng đôi khi cả hai con đều được tạo ra và đưa vào quần thể mới. Tuy nhiên, sự chéo không phải lúc nào cũng xảy ra. Đôi khi, dựa trên xác suất đã đặt, không xảy ra sự trao đổi chéo và bố mẹ được sao chép trực tiếp sang quần thể mới. Xác suất xảy ra chéo thường là 60% đến 70%.

### 2.3. Đột biến

Sau khi chọn lọc và lai ghép, chúng ta thu được quần thể mới đầy đủ các cá thể. Một số được sao chép trực tiếp và một số khác được sản xuất bằng phương

pháp chéo. Để đảm bảo rằng các cá thể không hoàn toàn giống nhau, bạn cho phép có một cơ hội đột biến nhỏ. Bạn lặp qua tất cả các alen của tất cả các cá thể và nếu alen đó được chọn để đột biến, bạn có thể thay đổi nó một lượng nhỏ hoặc thay thế bằng một giá trị mới.[3] Xác suất đột biến thường là từ 1 đến 2 phần mười phần trăm. Một hình ảnh cho đột biến được hiển thị dưới đây.



- Không có sự đảm bảo tuyệt đối rằng thuật toán di truyền sẽ tìm được mức tối ưu toàn cục. Nó xảy ra rất thường xuyên khi quần thể có nhiều đối tượng.

- Trong trường hợp hệ thống hỗ trợ quyết định lâm sàng, GA phải đối mặt với sự thiếu minh bạch được sử dụng cho các hệ thống hỗ trợ quyết định khiến bác sĩ không mong muốn.

Thách thức chính trong việc sử dụng thuật toán di truyền là xác định tiêu chí phù hợp. Để sử dụng thuật toán di truyền, phải có nhiều thành phần như nhiều loại thuốc, triệu chứng, liệu pháp điều trị, v.v. để giải quyết một vấn đề.

Trong trường hợp nghiên cứu điều tra khả năng tự động hóa lựa chọn tham số cho Thuật toán di truyền Giao diện máy tính não (BCI) dựa trên P300 dựa trên EEG yêu cầu thời gian thực hiện dài khiến nó không thể sử dụng trực tuyến.

Phương pháp GA sau đó đã được thay thế bằng xác thực chéo N-fold (NFCV) tốn ít thời gian thực thi hơn để tối ưu hóa tổng hợp các tham số trích xuất tính năng và tiền xử lý bằng cách sử dụng Phân tích phân biệt tuyến tính của Fisher (FLDA).

- GA cũng phải đối mặt với các vấn đề về khả năng mở rộng khi trao đổi các khối xây dựng.

- Giống như các kỹ thuật trí tuệ nhân tạo khác, thuật toán di truyền không thể đảm bảo thời gian phản hồi tối ưu hóa liên tục. Hơn nữa, sự khác biệt giữa thời gian phản hồi tối ưu hóa ngắn nhất và dài nhất còn lớn hơn nhiều so với các phương pháp chuyển màu thông thường. Thuộc tính thuật toán di truyền đáng tiếc này hạn chế việc sử dụng thuật toán di truyền trong các ứng dụng thời gian thực.

- Ứng dụng thuật toán di truyền trong các điều kiện được thực hiện trong thời gian thực còn hạn chế do các giải pháp ngẫu nhiên và hội tụ, nói cách khác, điều này có nghĩa là toàn bộ quần thể đang được cải thiện, nhưng điều này không thể nói đối với một cá nhân trong quần thể này. Vì vậy, sẽ không hợp lý khi sử dụng các thuật toán di truyền để điều khiển trực tuyến trong các hệ thống thực mà không thử nghiệm chúng trước trên mô hình mô phỏng.

(Xen tiếp trang 155)

Bảng 2.4 cho thấy: Lý do chủ yếu dẫn đến việc SV không tham gia tập luyện ngoại khoá trong số 1065 SV theo thứ tự là: Không có đủ điều kiện sân bãi dụng cụ (253/1065, chiếm 23,76%), không có GV hướng dẫn (462/1065, chiếm 43,38%), hoạt động TTNK của trường chưa phát triển (150/1065, chiếm 14,08%). Ngoài ra còn một số lý khác chiếm tỷ lệ thấp hơn như: Chưa có CLB TT cho SV (7,04%). Nội dung môn học khác chi phối quá nhiều thời gian (4,32%). Không có môn TT nào phù hợp (3,76%), Công tác tuyên truyền động viên chưa được chú trọng (3,66%).

### 3. Kết luận

Qua đánh giá thực trạng hoạt động TTNK của SV Yrường ĐH Kinh tế và QTKD, nghiên cứu đã chỉ ra được một số vấn đề như sau:

SV tham gia tập luyện TTNK thường xuyên rất ít; trong đó số hoàn toàn không tập chiếm 30.4%. Động cơ chính của SV khi tham gia tập luyện ngoại khoá được xác định chủ yếu là do nhận thấy tác dụng của tập luyện TDTT chiếm 37,98% và ham thích

chiếm 35,35%. Trong các yếu tố ảnh hưởng tới hoạt động TTNK của SV thì yếu tố không có GV hướng dẫn chiếm tỉ lệ 43,38%.

Vấn đề ngoại khóa của SV sau khi không học môn GDTC sẽ khó có thể kiểm soát do SV không ràng buộc các điểm học. Vì vậy phải có những giải pháp thích hợp giúp SV hiểu vai trò của rèn luyện TDTT để tự giác tham gia luyện tập, từ đó sẽ tạo ra những sân chơi bổ ích cho SV.

### Tài liệu tham khảo

1. Lê Bửu, Nguyễn Thế Truyền (1989), *Lý luận và phương pháp thể thao trẻ*, NXB TDTT TP. Hồ Chí Minh.
2. Dương Nghiệp Chí (1991), *Đo lường thể thao*, NXB TDTT. Hà Nội
3. Lê Đức Ngọc (1996), *Kiểm tra- Đánh giá thành quả học tập*, NXBĐH Quốc gia. Hà Nội
4. Nguyễn Xuân Sinh và Cs (2006). *Giáo trình phương pháp nghiên cứu khoa học Thể dục thể thao*. NXBTDTT. Hà Nội

## Thuật toán di truyền – Bài khảo sát... (tiếp theo trang 72)

Đột biến khá đơn giản. Chúng tôi chỉ thay đổi các alen đã chọn dựa trên những gì bạn cảm thấy cần thiết và tiếp tục. Tuy nhiên, đột biến là rất quan trọng để đảm bảo sự đa dạng di truyền trong quần thể.

### 3. Kết luận

Trên cơ sở khảo sát, có nhiều phòng thí nghiệm khác nhau về cơ bản tập trung vào nghiên cứu được thực hiện trên Thuật toán di truyền

Phòng thí nghiệm thuật toán di truyền Kanpur (KanGAL) là một phòng thí nghiệm nghiên cứu tại Viện Ấn Độ

Technology Kanpur chuyên theo đuổi nghiên cứu và hợp tác công nghiệp trong các lĩnh vực thuật toán di truyền, bộ điều khiển logic mờ, tin sinh học và mạng lưới thần kinh. Nghiên cứu cơ bản và ứng dụng vào các vấn đề kỹ thuật là trọng tâm chính của KanGAL.

KanGAL được thành lập vào năm 1997. Kể từ đó phòng thí nghiệm đã sản xuất một số luận án tiến sĩ và luận văn thạc sĩ liên quan đến thuật toán di truyền, bộ điều khiển logic mờ và mạng lưới thần kinh. Một số sinh viên đại học cũng đã được đào tạo về các lĩnh vực này. Bên cạnh nghiên cứu cơ bản, KanGAL còn tích cực hợp tác với các ngành công nghiệp và đặc biệt quan tâm đến việc sử dụng các kỹ thuật tính toán mềm để giải quyết các vấn đề công nghiệp.

### Tài liệu tham khảo

1. R. Agrawal and G. Psaila, “Active data mining,” *Current*, pp. 3-8, 1995.

2. S. H. Liao, P.H. Chu, and P. Y. Hsiao, “Data mining techniques and applications - A decade review from 2000 to 2011,” *Expert Syst. Appl.*, vol. 39, no. 12, pp. 11303-11311, 2012.

3. G. E. Vlahos, T. W. Ferratt, and G. Knoepfle, “The use of computer-based information systems by Gennan managers to support decision making,” *Irtf: Manag.*, vol. 41, no. 6, pp. 763-779, 2004.

4. I. H. Witten, E. Frank, and M. a Hall, *Data Mining: Practical Machine Learning Tools and Techniques (Google eBook)*. 2011.

5. D. K. Bhattacharyya and S. M. Hazarika, *Networks, Data Mining And Artificial Intelligence: Trends And Future Directions*, 1st ed. Narosa Pub House, 2006.

6. M. Karegar, A. Isazadeh, F. Fartash, T. Sadari, and A.H. Navin, “Data-Mining by Probability-Based Patterns,” pp. 353-360, 2008.

7. H. Thomas and L. Paul, *Statistics: Methods and Applications*, 1st ed. StatSoft, Inc, 2005.

8. M. Kantardzic, *Data Mining: Concepts, Models, Methods, and Algorithms*, 2nd ed. Wiley-IEEE Press, 2011.

9. P. Berkhin, “A Survey of Clustering Data Mining,” *Group. Multidimens. Data*, no. c, pp. 25-71, 2006.

10. T. P. Hong, K. Y. Lin, and S. L. Wang, “Fuzzy data mining for interesting generalized association rules,” *Fuzzy Sets Syst.*, vol. 138, no. 2, pp. 255-269, 2003.