Tạp chí Tin học và Điều khiển học, T. 19, S. 4 (2003), 366–372

CÁC TOÁN TỬ LAI GHÉP CHO THUẬT TOÁN DI TRUYỀN GIẢI BÀI TOÁN TÔ MÀU ĐỒ THỊ ĐƠN

NGUYỄN XUÂN HUY, VŨ THIỆN CĂN

Viện Công nghệ thông tin

Abstract. This paper describes and compares some crossover operators of the genetic algorithm for the problem of single chromatic graph. Using a fixed number of evolutions and a specified number of colors, we count the number of pairs of adjacent vertices having the same color for each certain graph. This number is calculated in real-time and the best operator is one that produces better solution. Genetic algorithms mentioned in this paper mainly use proportionally selected operators and classic mutation operators.

Tóm tắt. Bài báo mô tả và so sánh một số toán tử lai ghép khác nhau của thuật toán di truyền giải bài toán tô màu đồ thị đơn. Nội dung so sánh là đếm bằng thực nghiệm, với số lần tiến hóa cố định cho trước và số màu cho trước đối với mỗi loại đồ thị. Số đếm này được xác định trong thời gian thực. Toán tử được chọn sẽ là toán tử cho kết quả trội.

Các thuật toán di truyền đề cập đến là toán tử chọn mà chủ yếu là toán tử chọn tỉ lệ và toán tử đột biến cổ điển.

1. BÀI TOÁN TÔ MÀU ĐỒ THỊ ĐƠN

1.1. Bài toán

Cho $G = (V, E)$ là một đồ thị đơn vô hướng, trong đó V là tập các đỉnh và E là tập các cạnh. Ta giải bài toán sau.

Dùng một số ξ cho trước các màu, xác định màu tô cho mỗi đỉnh $i \in V$ sao cho số cạnh có đỉnh cùng màu là nhỏ nhất.

Đồ thị G được gọi là đơn nếu mỗi cặp đỉnh có thể xác định không quá một cung nối chúng ([1]).

1.2. Mô hình toán học của bài toán

Với mỗi $i \in V$ và với mỗi $j = 1, 2...\xi$, đặt

$$x_{ij} = \begin{cases} 1 & \text{với đỉnh } i \in V \text{ có màu } j \\ 0 & \text{ngược lại} \end{cases}$$

Bài toán có thể phát biểu như sau ([1, 3, 5])

$$\min \sum_{j=1}^{\xi} \sum_{(i,t) \in E} x_{ij} x_{tj}$$

với ràng buộc

$$\sum_{j=1}^{\xi} x_{ij} = 1, \quad i \in V,$$

$$x_{ij} = 0 \quad \text{hay} \quad 1, j \in V, \; j = 1, 2 \ldots \xi$$

2. KỸ THUẬT DỰA TRÊN QUẦN THỂ

Trong các kỹ thuật này, một cơ chế tiến hóa được sử dụng để biến đổi một quần thể (tập hợp) các lời giải từ thế hệ này đến thế hệ kế tiếp sao cho từ quần thể kế tiếp. có nhiều hy vọng tìm được những lời giải tốt hơn. Tại mỗi thế hệ (hay nói cách khác, tại mỗi lần lặp), ta sử dụng hai quy trình, quy trình tập thể và quy trình cá nhân.

Trong quy trình tập thể, các lời giải của quần thể hiện tại được so sánh và kết hợp để tạo ra các lời giải khác, có kế thừa các tính chất "tốt" từ các lời giải "cha mẹ". Sau đó, lời giải mới tìm được được tiến hóa bằng quy trình cá nhân, từ đó hình thành quần thể mới.

Cách tổ chức các quy trình tập thể và cá nhân khác nhau giúp hình thành các kỹ thuật khác nhau. Các kỹ thuật chính được sử dụng hiện nay là: thuật toán di truyền, thuật toán tìm kiếm rải rác, thuật toán Vương Quốc Kiến ([1]). Trong bài báo này, chúng tôi chỉ sử dụng thuật toán di truyền.

2.1. Thuật toán di truyền

Thuật toán di truyền ([1, 6, 7]) là thuật toán mô phỏng quá trình tiến hóa của sinh vật, dựa trên các nguyên lý chọn lọc và tiến hóa tự nhiên. Trong thuật toán di truyền, quy trình tập thể bao gồm toán tử chọn và toán tử lai ghép, còn quy trình cá nhân được thực hiện bởi toán tử đột biến.

Tại mỗi bước lặp (hoặc thế hệ), ba toán tử nói trên được áp dụng để tạo ra tập các lời giải (con) mới. Sau đó một toán tử thứ tư (toán tử culling) được áp dụng để xác định quần thể trong thế hệ kế tiếp bằng cách chọn các cá thể từ các cá thể cha mẹ và cá thể con cái mới sinh ra. Thông thường, người ta thường sử dụng quần thể có kích thước N (số lượng cá thể) cố định, và quần thể mới thường bao gồm các cá thể tốt nhất (nghĩa là các cá thể làm cực tiểu hóa hàm mục tiêu).

Trong các sơ đồ cổ điển, tại mỗi thế hệ, thường quần thể được thay đổi toàn bộ. Để làm như thế, N lời giải cha mẹ được chọn và ghép cặp. Sau đó, toán tử lai ghép được áp dụng với xác suất nào đó để tạo hai lời giải con tương ứng. Ngược lại, hai lời giải cha mẹ trở thành hai lời giải con tương ứng. Quần thể trong lần lặp kế tiếp bao gồm tất cả các lời giải con.

Trong một biến thể khác, người ta sử dụng "quần thể với trạng thái ổn định", nghĩa là trong mỗi thế hệ, chỉ có một số ít cá thể bị thay đổi theo chiến lược trên. Ưu điểm của "quần thể với trạng thái ổn định" được đề cập đến trong [1].

2.2. Sơ đồ của thuật toán di truyền

Khởi tạo:

Sinh ra quần thể ban đầu với N lời giải x_i, $i = 1, 2 \ldots N$, nghĩa là $P_0 := \{z_1, z_2 \ldots z_N\}$ trong đó x_i được biểu diễn bằng z_i, $i = 1, 2 \ldots N$.

Đặt iter:=0, niter :=0; $P := P_0$, stop := false. Gọi z^* là lời giải tốt nhất trong P_0.

Thuật toán di truyền

while not stop

iter:=iter+1; niter:=niter+1.

Áp dụng toán tử chọn.

Áp dụng toán tử lai ghép.

Áp dụng toán tử đột biến.

Áp dụng toán tử culling để tạo ra quần thể mới P' các lời giải.

$P := P'$.

Chọn z' là phần tử tốt nhất trong P'.

Nếu $f(z') < f(z)$ thì $z^* := z'$ và niter := 0

Nếu iter = itermax hay niter = nitermax thì stop := true.

x^* là lời giải tốt nhất tìm được.

2.3. Tổ chức dữ liệu

Mỗi lời giải $x \in X$ được biểu diễn như một vectơ $z \in R^m$. Giá trị của lời giải x được biểu diễn bởi z và được ký hiệu là $f(x)$ hay $f(z)$. Tổng quát, $X \subseteq R^n,\ m \neq n$.

Một lời giải x của bài toán tô màu có dạng

$$x_{ij(i)} = 1,\ i \in V$$

$$x_{ij} = 0,\ i \in V,\ j = 1, 2 ... \xi,\ j \neq j(i)$$

và được biểu diễn bằng vectơ $z \in R^{|V|}$, trong đó $z_i = j(i),\ i \in V$.

Như vậy, thành phần thứ i của vectơ z là số hiệu màu tô cho đỉnh i.

2.4. Toán tử chọn

Toán tử chọn ([2]) dùng để chọn một số chẵn các lời giải cha mẹ từ quần thể hiện tại. Mỗi lời giải cha mẹ được chọn theo một trong các chiến lược nêu sau. Chú ý rằng một lời giải có thể được chọn nhiều lần. Các lời giải cha mẹ được ghép cặp để áp dụng toán tử lai ghép trong pha kế tiếp của thuật toán nhằm cho ra lời giải con cái.

2.4.1. Chọn ngẫu nhiên

Một toán tử chọn đơn giản nhất là chọn ngẫu nhiên với ưu điểm là tăng độ da dạng của quần thể sinh ra. Cũng có nhiều loại toán tử chọn khác được đề cập đến, chẳng hạn chỉ chọn để ghép cặp những lời giải tốt nhất trong quần thể.

2.4.2. Chọn tỉ lệ

Một loại toán tử chọn khác là toán tử chọn tỉ lệ được thực hiện như sau

(i) Sắp thứ tự các phần tử của quần thể $P : z^1, z^2 ... z^N$.

(ii) Chọn ngẫu nhiên α trong đoạn $[0, \sum_{i=1}^{N} \frac{1}{f(z^i)}]$

(iii) Gọi r là chỉ số lớn nhất sao cho $\sum_{i=1}^{r} \frac{1}{f(z^i)} \leq \alpha$

(iv) Lời giải z^r được chọn làm lời giải cha mẹ.

Hiển nhiên, cơ may để chọn lời giải z tăng với độ thích nghi $1/f(z)$ do chiều dài của đoạn $[0, \sum_{i=1}^{N} \frac{1}{f(z^i)}]$ (liên kết với z) tăng khi $f(z)$ nhỏ.

2.4.3. Chọn kiểu thi đấu

Một loại toán tử chọn khác: *toán tử chọn kiểu thi đấu* là chọn lời giải tốt nhất từ một tập con các lời giải của P (chọn ngẫu nhiên). Nói cách khác, gọi N_1 là số nguyên sao cho $1 \leq N_1 \leq N$.

(i) Lần lượt chọn ngẫu nhiên N_1 lời giải từ P (một lời giải có thể chọn nhiều lần) để tạo thành tập con P_1.

(ii) Gọi z_1 là lời giải tốt nhất trong P_1.

(iii) z_1 được chọn làm lời giải cha mẹ

Chú ý rằng nếu $N_1 = N$, chính là chọn lời giải tốt nhất trong quần thể. Nếu $N_1 = 1$, lời giải cha mẹ được chọn ngẫu nhiên.

Yếu điểm lớn nhất của việc chọn phần tử tốt nhất là dễ bị vướng vào cực trị địa phương.

2.4.4. Chọn dựa vào độ đa dạng của quần thể

Một cách khác là chọn lựa dựa vào độ đa dạng của quần thể. Kỹ thuật sau dựa trên hướng này.

(i) Sắp thứ tự các phần tử của P theo độ tăng của hàm mục tiêu $z^1, z^2...z^N$

(ii) Đặt $r = 1+D,\ D \in [0,1]$ là độ đo dùng để đo độ đa dạng của quần thể. Đầu tiên gán $r = 2$

(iii) Chọn số ngẫu nhiên $\mu \in [0, N^{1/r}]$. Đặt $r = \lceil \mu^r \rceil$

(iv) Vậy z^r được chọn làm lời giải cha mẹ

Cách chọn trên thay đổi độ tốt tùy theo độ phân tán của quần thể.

2.5. Toán tử lai ghép

Mục đích chính của việc sử dụng toán tử lai ghép là sinh ra lời giải con cái mà vẫn giữ lại một số đặc tính của lời giải cha mẹ (theo một tiêu chuẩn) trong quần thể hiện tại. Điều này nhằm hướng việc tìm kiếm trên miền có triển vọng của không gian X trong khi vẫn giữ được phần nào sự đa dạng của quần thể.

Việc sử dụng các toán tử lai ghép khác nhau trong việc nghiên cứu hiệu quả của thuật toán di truyền là một khuynh hướng thường được áp dụng hiện nay ([5]). Trong bài báo này, chúng tôi cũng đi theo hướng khảo sát việc sử dụng thuật toán di truyền trên bài toán tô màu đồ thị đơn bằng cách so sánh nhiều toán tử lai ghép khác nhau.

Chúng tôi khảo sát năm loại toán tử lai ghép sau.

2.5.1. Toán tử lai ghép một điểm

Sinh ra hai lời giải con từ lời giải cha mẹ bằng cách chọn ngẫu nhiên một điểm chia rồi tráo đổi hai thành phần của hai lời giải cha mẹ cho nhau, kết quả là hai lời giải con cái.

2.5.2. Toán tử lai ghép hai điểm

Chọn ngẫu nhiên hai điểm chia của chuỗi biểu diễn hai lời giải cha mẹ, rồi tráo đổi hai thành phần của đoạn giữa của hai lời giải cha mẹ cho nhau, hai đoạn đầu và cuối giữ nguyên.

2.5.3. Toán tử lai ghép đều

Sinh ngẫu nhiên chuỗi nhị phân kích thước m, nếu bit thứ i bằng 1 thì tráo đổi hai thành phần tương ứng của hai chuỗi biểu diễn của lời giải cha mẹ, nếu không thì thành phần tương ứng giữ nguyên.

2.5.4. Toán tử lai ghép trên đường nối cha mẹ

Nội dung toán tử này được đề cập đến trong [4]. Gọi $x1, x2$ là các phần tử của quần thể. Ký hiệu $d(x1, x2)$ là khoảng cách Hamming giữa $x1$ và $x2$. Một lân cận của một lời giải $x1$, ký hiệu $N(x1)$, là tập các phần tử x (không nhất thiết thuộc quần thể) có khoảng cách Hamming đến $x1$ bằng 1.

Cho hai cha mẹ $ps1, ps2$, ký hiệu $N(ps1, ps2)$ là tập các phần tử của $N(ps1)$ có khoảng cách Hamming $ps2$ nhỏ hơn khoảng cách Hamming đến $ps1$. Toán tử này hoạt động như sau.

Tại mỗi bước lặp, ta chọn $ps \in N(ps1, ps2)$ (dùng toán tử chọn tỉ lệ). Nếu lời giải này tốt hơn lời giải ban đầu $ps1$ và $ps2$, thì ps trở thành lời giải con os. Ngược lại, $ps1$ được thay bởi ps và $ps2$ đóng vai trò của $ps1$. Thủ tục lặp cho đến khi $ps1 = ps2$. Vậy $os = ps1 = ps2$.

Thuật toán lai ghép trên đường nối cha mẹ

Khởi tạo:

Đặt: $z1$ và $z2$ là hai lời giải cha mẹ, $ps1 := z1$; $ps2 = z2$; t stop := false;

While not stop

Xác định tập con $N(ps1, ps2) \subseteq N(ps1)$

Dùng toán tử chọn để $ps \in N(ps1, ps2)$

Nếu $f(ps) < f(z1)$ hay $f(ps) < f(z2)$ thì stop:=true

$ps1 := ps2$

$ps2 := ps$

Nếu $d(ps1, ps2) = 0$ thì stop := true

$os := ps$

2.5.5. Toán tử tăng tri thức

Toán tử này dựa trên đặc trưng về cấu trúc của bài toán tô màu đồ thị như đã nêu [3]. Toán tử này sinh là một lời giải con os từ hai lời giải cha mẹ $z1$ và $z2$.

Với mỗi lời giải z, kí hiệu $CN(z)$ là các đỉnh đụng độ màu của z (nghĩa là tập các đỉnh có đỉnh kề với nó cùng màu).

$$CN(z) = \{i \in V | \exists (i, l) \in E \text{ sao cho } j(i) = j(l)\}$$

Một lời giải con được sinh ra bằng cách tô màu đỉnh đụng độ của một lời giải cha mẹ bằng cách dùng màu tương ứng của lời giải cha mẹ còn lại, nếu màu sau không đụng độ. Nếu một màu bị đụng độ cả ở 2 lời giải cha mẹ, ta dùng màu mà màu đó được dùng ít nhất trong các cạnh đỉnh kề nhau ở cả 2 lời giải cha mẹ. Còn các đỉnh còn lại (nghĩa là không thuộc các đỉnh đụng độ của cả lời giải cha lẫn lời giải mẹ), thì được tô như trong toán tử đều.

Thuật toán lai ghép tăng tri thức

Khởi tạo: Đặt $z1, z2$ là hai lời giải cha mẹ.

Với $i \in CN(z1)$, nếu $i \in CN(z2)$ thì $osi := z_i2$.

Với $i \in CN(z2)$, nếu $i \in CN(z1)$ thì $osi := z_i1$.

Nếu $i \in CN(z1) \cap CN(z2)$ thì $osi :=$ màu ít bị đụng độ nhất.

Nếu $i \in (V - (CN(z1) \cup CN(z2)))$ thì chọn ngẫu nhiên $\alpha \in \{0, 1\}$.

Nếu $\alpha = 0$ thì $osi := z_i1$, nếu không thì $osi := z_i2$.

os là lời giải con sinh ra.

2.6. Toán tử đột biến

Khi lời giải con được sinh ra, một quy trình cá nhân có thể được áp dụng cho lời giải con này. Một cách làm cổ điển là thay đổi ngẫu nhiên một thành phần nào đó của lời giải với một xác suất nhỏ.

Với i từ 1 đến N, sinh ra số ngẫu nhiên $\alpha \in [0, 1]$.

Nếu $\alpha < \alpha_0$ thì chọn giá trị ngẫu nhiên mới cho z_i.

Giá trị α_0 thường được chọn khá nhỏ để việc thay đổi thành phần của z_i được thực hiện với xác suất nhỏ. Toán tử đột biến nhằm giả lập việc có những tác động ngẫu nhiên của ngoại cảnh vào quần thể.

Toán tử đột biến không cần thiết bằng toán tử lai ghép vì chúng có tần suất sử dụng thấp hơn so với toán tử lai ghép. Nhưng lợi điểm của việc sử dụng toán tử đột biến ở chỗ nó giúp làm tăng độ đa dạng của quần thể, nhằm tránh việc bị sa vào cực trị địa phương xấu.

3. KẾT QUẢ THỰC NGHIỆM

Chúng tôi đã so sánh bằng thực nghiệm việc áp dụng các toán tử lai ghép khác nhau đã trình bày ở trên để giải bài toán tô màu trên đồ thị đơn. Với toán tử lai ghép một điểm chia và toán tử đột biến đơn giản, chúng tôi tiến hành so sánh trên 3 thuật toán chọn khác nhau là chọn ngẫu nhiên (g01a), chọn theo tỉ lệ (g01b) và chọn lựa kiểu thi đấu (g01c).

Sau đó, cùng với thuật toán sinh ngẫu nhiên, chọn theo kiểu tỉ lệ, đột biến đơn giản, chúng tôi lần lượt sử dụng các thuật toán lai ghép hai điểm chia g02a và g02b, lai ghép đều g02c, lai ghép trên đường nối cha mẹ g02d, lai ghép tăng tri thức g02e.

Tiêu chí để so sánh là với số lần tiến hóa cố định cho trước đối với mỗi loại đồ thì, chúng tôi xem xem thuật toán nào đưa ra kết quả tốt hơn. Chúng tôi thực hiện mỗi chương trình 10 lần, trong mỗi lần thực hiện, thuật toán cho quần thể tiến hóa 1000 lần.

Thuật toán được thực hiện trên 14 loại đồ thị (đánh số từ 1 đến 14), gồm có các đồ thị vòng, đồ thị hình sao, đồ thị đầy đủ, đồ thị ngẫu nhiên với số đỉnh nhỏ nhất là 10, nhiều nhất là 100. Kết quả cho như sau.

Thuật toán	Số màu	Số cạnh có đỉnh cùng màu của đồ thị tương ứng trong mười bốn đồ thị khao sát
g01a	2	0 1 5 17 26 0 2 16 41 56 1 20 0 9
	3	0 0 2 11 18 0 0 12 30 44 0 32 0 10
g01b	2	0 1 6 20 29 0 2 18 50 74 1 55 0 20
	3	0 1 3 10 17 0 0 10 31 44 0 30 0 10
g01c	2	0 2 6 19 30 0 2 20 51 71 1 49 0 16
	3	0 0 2 11 16 0 0 10 31 44 0 27 0 7
g02a	2	0 2 5 18 30 0 1 19 50 76 0 54 0 21
	3	0 0 3 9 18 0 0 10 30 46 0 32 0 13
g02b	2	0 1 7 18 29 0 2 19 50 73 0 53 0 20
	3	0 1 3 11 16 0 1 9 28 43 0 30 0 11
g02c	2	0 1 6 17 28 0 1 19 49 71 0 45 0 17
	3	0 0 3 11 18 0 0 10 28 44 0 32 0 10
g02d	2	0 2 6 19 33 0 2 19 52 76 0 55 0 19
	3	0 1 3 11 17 0 1 10 32 46 0 33 0 12
g02e	2	0 1 5 11 15 0 1 4 8 10 0 1 0 1
	3	0 0 2 6 7 0 1 2 3 6 0 0 0 0

+ Thời gian chạy của thuật toán g01b là lâu nhất.

+ Chương trình được thực hiện nhanh nhất với thuật toán g02e và kết quả của thuật toán này là tốt nhất.

4. KẾT LUẬN

• Đối với bài toán tô màu đồ thị đơn như trình bày ở trên, việc thay đổi các thuật toán lai ghép khác nhau có thể giúp chúng ta cải thiện được kết quả.

• Với kết quả thực nghiệm trên, ta thấy thuật toán lai ghép tăng tri thức cho kết quả tốt nhất và thực tế khi thực hiện chương trình cho thấy tốc độ thực hiện chương trình rất nhanh.

• Việc thay đổi các kỹ thuật chọn phần tử cha mẹ khác nhau, cũng như việc tìm ra các toán tử lai ghép khác có thể làm tăng hiệu quả của chương trình và đó có thể là những phương hướng nghiên cứu tiếp theo cho bài toàn tô màu đồ thị đơn.

TÀI LIỆU THAM KHẢO

[1] L. Davis, *Handbook of Genetic Algorithms* (Ed.) Van Nostrand Reinhold, New York, 1991.

[2] J. A. Ferland, *Heuristic Search Methods for Combinatorial Programming Problems,* University of Montreal, Canada, 2001.

[3] C. Fleurent, J. A. Ferland, Genetic and Hybrid Algorithms for Graph Coloring, *Annals of Operations Research* **63** (1996) 437 – 461.

[4] C. Fleurent, F. Glover, Advanced Recombination Operators for Heuristic Search Mehthods, *Report from Graduate School of Business,* University of Colorado, Boulder (1995).

[5] D. A. Fotakis, S. D. Likothanassis, S. K. Stefanakos, An Evolutionary Annealing Approach to Graph Coloring, *Abstracts of EvoCOP* (2001).

[6] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning,* Addison Wesley, 1989.

[7] J. H. Holland, *Adaptation in Natural and Artificial Systems,* Ann Arbor, University of Michigan Press, 1975.

Nhận bài ngày 01 - 4 - 2003
Nhận lại sau sửa ngày 06 - 11 - 2003