Tạp chí Tin học và Điều khiển học, T.21, S.2 (2005), 108–119

META-HEURISTIC - KẾT HỢP THUẬT GIẢI DI TRUYỀN VỚI THÔNG TIN THỐNG KÊ XÁC SUẤT GIẢI QUYẾT BÀI TOÁN NGƯỜI DU LỊCH

HOÀNG KIẾM[1], NGUYỄN THANH HÙNG[2]

[1] *Trung tâm phát triển CNTT - Đại học Quốc gia Tp. Hồ Chí Minh*

[2] *Trường Phổ thông năng khiếu - Đại học Quốc gia Tp. Hồ Chí Minh*

Abstract. Traveling salesman problem is one of the most important NP-complete problems. Several studies, used in real problems, have been conducted to solve it, especially in heuristic solutions. In this paper, we propose a new approach Meta-heuristic that combines Genetics algorithm and appearance probabilities of edges in the optimal cycle. The method includes two steps: first, we apply Genetics algorithm to find the appearance probabilities of edges in the optimal cycle from results of generations gained in the whole population evolution; then, we find the final solution by running Genetics algrithm again in which the cross-over operation is based on results in the first step. Our proposal is evaluated in various practical problems and compared with previous works. The experiment results prove that our approach is more efficient than those using either Genetics algorithm or the heuristic method of Christofide [8]. Hence, it proposes a new approach to find optimal solutions in traveling salesman problem.

Tóm tắt. Bài toán người du lịch là một trong số những bài toán quan trọng nhất trong tập bài toán NP- đầy đủ. Đã có rất nhiều nghiên cứu tập trung vào tìm kiếm lời giải cho bài toán, đặc biệt là lời giải gần đúng để có thể ứng dụng trong thực tế.

Trong bài báo này chúng tôi đề xuất một hướng tiếp cận mới cho bài toán, được gọi là Meta-heuristic. Phương pháp gồm có hai bước: Bước 1, sử dụng thuật toán Genetics để tìm thông tin thống kê xác suất các cung sẽ xuất hiện trong chu trình tối ưu từ các cá thể chu trình tốt nhất chọn lọc qua các thế hệ. Bước 2, từ các thông tin tìm được, thực hiện lại thuật toán Genetics để chu trình kết quả tối ưu, trong đó các phép lai ghép sẽ dựa vào thông tin xác suất tìm được ở Bước 1. Chúng tôi đánh giá cách tiếp cận mới với nhiều dữ liệu thực nghiệm khác nhau và so sánh với những cách tiếp cận đã được nghiên cứu trước đây. Kết quả cho thấy cách tiếp cận này hiệu quả hơn với cách tiếp cận chỉ dùng thuật toán Genetics hoặc phương pháp gần đúng của Christofide [8].

1. GIỚI THIỆU

Bài toán người du lịch (TSP - Traveling Salesman Problem) được phát biểu như sau: Cho trước một đồ thị $G(V,E)$ với tập đỉnh $V = \{v_1, v_2, ..., v_n\}$ là các thành phố và tập các cạnh $E = \{e_1, e_2, ..., e_m\}$ với chi phí các cạnh là $W = \{w_{e_1}, w_{e_2}, ..., w_{e_m}\}$ là chi phí đi lại giữa các thành phố. Mục tiêu của bài toán là phải tìm được một chu trình đi qua tất cả các thành phố, mỗi thành phố chỉ qua một lần với tổng chi phí nhỏ nhất, tức là tìm một hoán vị $\pi = \{v_{\pi(1)}, v_{\pi(2)}, ..., v_{\pi(n)}\}$ của V sao cho tối thiểu tổng chi phí là: $\sum_{i=1}^{n-1} w(v_{\pi(i)}, v_{\pi(i+1)}) + w(v_{\pi(n)}, v_{\pi(1)})$.

Trong trường hợp tập đỉnh V là các đỉnh trong không gian Euclide, bài toán được gọi là bài toán Euclide TSP.

Bài toán người du lịch có ý nghĩa rất quan trọng trong thực tiễn ([5]), như ứng dụng để tìm hành trình ngắn nhất đi qua các điểm du lịch, bài toán phân phối hàng hóa... Tuy nhiên, bài toán đã được chứng minh là bài toán NP-đầy đủ ([2]), và cho đến nay vẫn còn là một vấn đề rất hóc búa thách thức các nhà nghiên cứu. Để tìm hiểu một cách đầy đủ về bài toán có thể tham khảo tại tài liệu [9]. Có rất nhiều phương pháp được đề xuất cho việc giải bài toán và được chia thành 2 nhóm: nhóm các thuật toán chính xác và nhóm các thuật toán gần đúng. Một trong những thuật toán chính xác cho bài toán là phương pháp nhánh cận và đã được chứng minh về sự tối ưu của lời giải nhưng mất nhiều thời gian tính toán trong trường hợp số đỉnh của đồ thị lớn. Nhóm thuật toán gần đúng thường được đánh giá trên thời gian chạy và mức độ tối ưu của lời giải. Đã có rất nhiều thuật toán giải gần đúng ([5, 6, 7] được đưa ra nhằm tìm đến một lời giải gần tối ưu.

Một phương pháp gần đúng cho bài toán Hamilton có độ phức tạp đa thức được đề nghị bởi Christofide [8], với lời giải tìm được xấp xỉ với lời giải tối ưu, và trong trường hợp xấu nhất được chứng minh là $\leqslant 1{,}5$ lần lời giải tối ưu. Phương pháp này thường được dùng để xác định một chặn trên tốt nhất cho bài toán. Một số phương pháp khác như simulated annealing, các phương pháp sử dụng meta heuristic như thuật giải di truyền, thuật giải ant,... cũng được áp dụng cho bài toán. Tuy nhiên kết quả vẫn còn khá hạn chế khi số lượng đỉnh và cung khá lớn.

Trong bài báo này chúng tôi đề xuất một hướng tiếp cận mới để tìm ra một lời giải gần với lời giải tối ưu nhất, đó là sử dụng thuật giải di truyền kết hợp với thông tin thống kê về tần suất các cung sẽ xuất hiện trong chu trình tối ưu. Chúng tôi thấy các cạnh có chi phí nhỏ xuất hiện nhiều hơn hẳn trong các cá thể chu trình ở các thế hệ tiến hóa. Đó là thông tin khá hữu ích trong việc xây dựng chu trình tốt nhất dựa trên thông tin thống kê qua toàn bộ quá trình tiến hóa.

Trong Mục 2 của bài báo, chúng tôi trình bày lại phương pháp của Christofide và thuật giải di truyền cho bài toán người đi du lịch. Tiếp đó, Mục 3 đưa ra hướng tiếp cận mới cho bài toán. Mục 4 là kết luận và kết quả thực nghiệm. Kết quả thực nghiệm cho thấy đây là cách tiếp cận hiệu quả, và trong hầu hết các trường hợp sẽ cho kết quả tốt hơn các phương pháp cũ.

2. MỘT SỐ PHƯƠNG PHÁP GẦN ĐÚNG CHO BÀI TOÁN NGƯỜI DU LỊCH

Đã có rất nhiều các thuật toán tìm kiếm gần đúng được đưa ra ([3, 5-12]), một số thuật toán cho kết quả chấp nhận được, một số khác thì kết quả vẫn còn hạn chế. Trong phần này chúng tôi sẽ giới thiệu một số thuật giải tìm kiếm gần đúng, đồng thời cũng trình bày phương pháp sử dụng giải thuật di truyền cơ sở áp dụng trong bài toán người đi du lịch.

2.1. Phương pháp Christofide cho bài toán người du lịch

Các thuật giải tìm kiếm gần đúng được trình bày khá chi tiết trong các bài báo của Johnson, McGeoch [3], Reinelt [6], Lawler, Lenstra, Rinnooy Kan, Shmoys [9] mà tiêu biểu là thuật toán Nearest Neighbour [10], thuật toán Multi Fragment Heuristic [11], thuật toán Savings [8, 12]. Trong đó cải tiến của thuật toán Savings được đưa ra bởi Christofide [8] cho kết quả khá tốt với kết quả tìm được so với lời giải tối ưu CHRIS(I) / OPT(I) $\leqslant$ 1,5.

Thuật toán Christofide dựa trên bước tìm kiếm cây phủ tối tiểu (Minimum Spanning Tree) của đồ thị, sau đó từ cây phủ tối tiểu này sẽ bổ sung một số cạnh để tạo thành đồ thị Euler, và cuối cùng sẽ chuyển thành đồ thị Hamilton.

Thuật toán Christofide:

B1. Tìm cây phủ tối tiểu T^* từ đồ thị G.

B2. Đặt U là tập các đỉnh bậc lẻ trong T^*, $|U|$ là số chẵn.

B3. Tìm các cặp cạnh ghép M^* với chi phi nhỏ nhất nối các đỉnh trong U.

B4. Đồ thị $H = T^* \cup M^*$ là đồ thị Euler.

B5. Chuyển đồ thị H thành chu trình Hamilton bằng cách giảm dần các đỉnh v có bậc lớn hơn 2 theo quy tắc bất đẳng thức tam giác.

Trong thuật toán trên, do trong chu trình kết quả có chứa cây phủ tối thiểu nên chi phí được giảm đáng kể. Ngoài ra, ở B3 và B4 việc bổ sung các cạnh để tạo thành chu trình Euler cũng được tối ưu bởi thuật toán cặp ghép. Tác giả đã chứng minh được trong trường hợp xấu nhất, kết quả tìm được không quá 3/2 so với kết quả tối ưu. Thuật toán này được cài đặt để so sánh với hướng tiếp cận của bài báo đặt ra (Mục 4).

2.2. Phương pháp tìm kiếm sử dụng giải thuật di truyền cho bài toán người du lịch

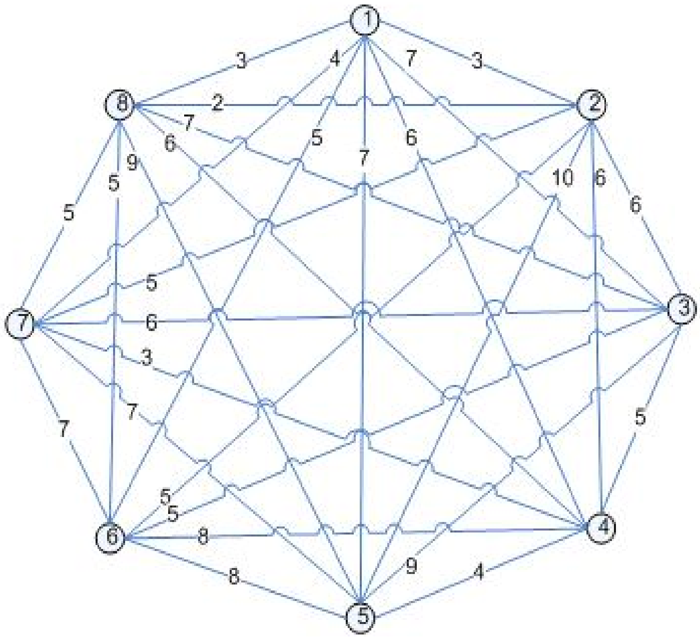
Giải thuật di truyền là một cách tìm kiếm lời giải tối ưu dựa trên nguyên lý tiến hóa qua các phép chọn lọc tự nhiên như lai ghép, đột biến, chọn lọc [4]. Vấn đề quan trọng nhất trong giải thuật di truyền là phương pháp mô hình hóa các cá thể và xác định hàm mục tiêu (hàm thích nghi). Trong bài toán người du lịch, mỗi cá thể đơn giản là một chu trình của đồ thị và hàm lượng giá chính là chi phí của toàn bộ chu trình đó. Ví dụ, với đồ thị có 8 đỉnh, một chu trình 5-1-7-8-2-6-4-3 sẽ được biểu diễn là Ind = (5 1 7 8 2 6 4 3) và chi phí là $C(\text{Ind})$.

Xét đồ thị ví dụ sau đây (Hình 1): Với cách biểu diễn cá thể như trên Hình 1, chúng ta sẽ sử dụng phép lai hai cá thể như sau: Từ hai cá thể cha mẹ P_1, P_2, ta chọn 2 điểm cắt ngẫu nhiên c_1 và c_2 :

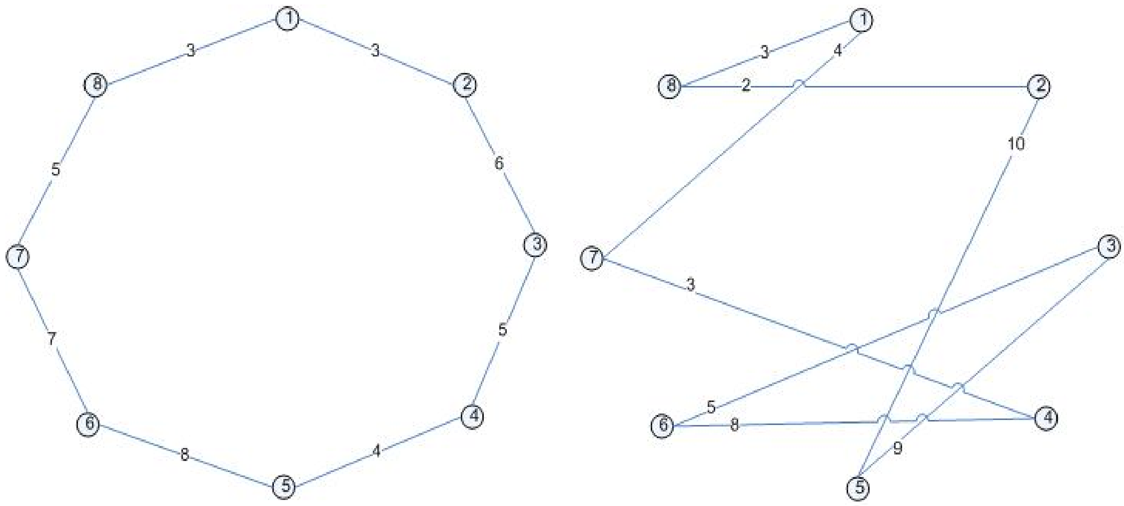
$$\begin{array}{rcc|cccc|ccc}
 & & & c_1 & & c_2 & & & & \\
P_1 = & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & w_1 = 41 \\
P_2 = & 4 & 7 & 1 & 8 & 2 & 5 & 3 & 6 & w_2 = 44
\end{array}$$

Đoạn nằm giữa điểm cắt $[c_1, c_2]$ sẽ được giữ nguyên và sao chép vào cá thể con:

$$\begin{array}{rcc|cccc|cc}
O_1 = & x & x & 3 & 4 & 5 & 6 & x & x \\
O_2 = & x & x & 1 & 8 & 2 & 5 & x & x
\end{array}$$



Hình 1. Đồ thị ví dụ $G(V, E)$



Hình 2. Hai cá thể bố mẹ P_1 và P_2 chọn để lai ghép

Kế tiếp, ta bắt đầu điền vào các vị trí còn trống ở các cá thể con bắt đầu từ vị trí $c_2 + 1$. Quy tắc để điền vị trí i bất kỳ như sau:

Lấy đỉnh u nằm ở vị trí $i - 1$ trong cá thể con cần hoàn thiện. Trong các đỉnh nằm trước và sau u trong cả hai cá thể cha mẹ P_1, P_2, ta chọn một đỉnh hợp lệ mà cạnh e_k giữa nó với u có trọng số w_{e_k} là nhỏ nhất. Trong trường hợp tất cả các đỉnh này đều không hợp lệ, ta sẽ chọn đỉnh hợp lệ trong tất cả các đỉnh còn lại mà có trọng số cạnh nối với u là nhỏ nhất.

Việc điền vào các cá thể con sẽ bắt đầu từ vị trí kế điểm cắt c_2. Ta muốn điền vào vị trí $i = c_2 + 1$ trong cá thể con O_2. Đỉnh ở vị trí $i - 1$ là 5. Trong 4 đỉnh nằm trước và sau đỉnh 5 trong P_1, P_2 là 2, 3, 4, 6 , ta sẽ phải chọn 1 đỉnh hợp lệ mà có giá trị cạnh nối với đỉnh 5 là nhỏ nhất. Theo đồ thị trên hình vẽ, thì đó là đỉnh 4, cá thể con O_2 sẽ như sau:

$$O_2 = x \quad x \mid 1 \;\; 8 \;\; 2 \;\; 5 \mid 4 \quad x$$

Tiếp tục, kế đỉnh 4 (trong hai cá thể cha mẹ P_1, P_2) là 3, 5, 7, 6. Theo đồ thị thì đỉnh tiếp theo sẽ là đỉnh 7

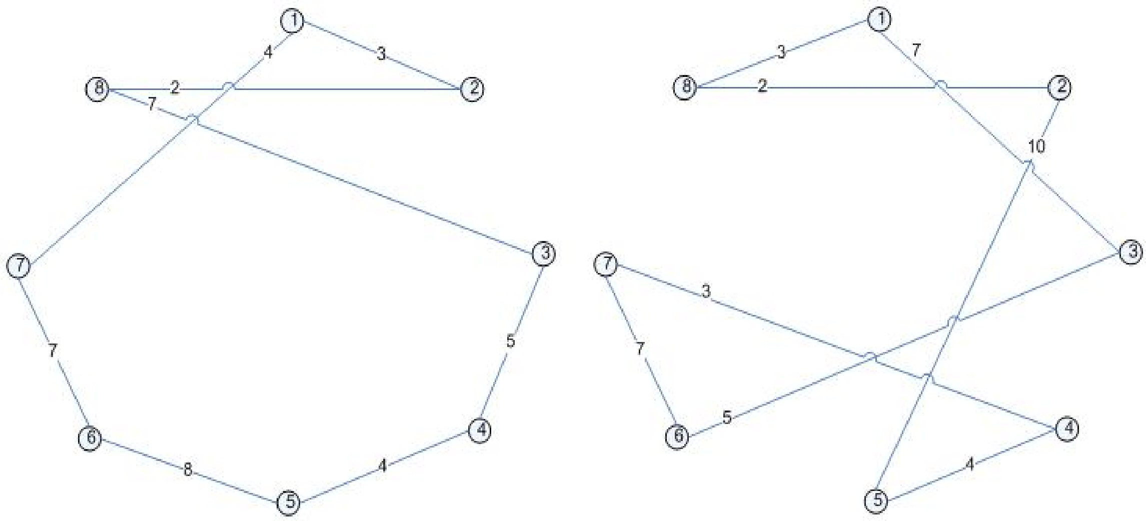
$$O_2 = x \quad x \mid 1 \;\; 8 \;\; 2 \;\; 5 \mid 4 \quad 7$$

Cứ tiếp tục cho đến khi hết tất cả những đỉnh chưa điền vào O_2. Cuối cùng, O_2 sẽ là:

$$O_2 = 6 \quad 3 \quad \mid \; 1 \;\; 8 \;\; 2 \;\; 5 \; \mid \quad 4 \quad 7 \qquad w_2 = 41$$

Tương tự, O_1 sẽ là:

$$O_1 = 2 \quad 8 \quad \mid \; 3 \;\; 4 \;\; 5 \;\; 6 \; \mid \quad 7 \quad 1 \qquad w_1 = 40$$



Hình 3. Hai cá thể con O_1 và O_2 lai ghép từ bố mẹ P_1 và P_2

Phép đột biến được thực hiện bằng cách hoán đổi vị trí 2 đỉnh bất kỳ trong một cá thể chu trình.

Với mô hình biểu diễn, các phép lai, các phép đột biến như vậy, kết quả tìm chu trình cho người du lịch bằng giải thuật di truyền như sau:

$t = 0$ {thế hệ tiến hóa hiện tại}
Khởi tạo các cá thể chu trình của quần thể $P(0)$
While $t <$ số đời tiến hóa T do
    Tính độ thích nghi của các cá thể trong quần thể $P(t)$
    $t = t + 1$
    Chọn lọc quần thể mới $P(t)$ từ $P(t-1)$
    Lai ghép một số cá thể trong quần thể $P(t)$
    Đột biến một số cá thể

End While

Công thức tính độ thích nghi các cá thể trong quần thể $P(t)$ như sau:

$$C(Y) = \sum_{i=1}^{n-1} w(Y(i), Y(i+1)) + w(Y(n), Y(1)),\ \forall Y \in P(t).$$

Kết quả thực nghiệm trên nhiều đồ thị cho thấy các lời giải tìm được khá tốt.

2.3. Nhận xét phương pháp trên

Với phương pháp Genetics, qua quá trình tiến hóa phát triển tự nhiên, các cá thể trong quần thể sẽ dần tốt lên và khả năng chúng ta tìm thấy được một chu trình "gần tối ưu" là rất cao. Tuy nhiên, điều này đòi hỏi một thời gian thực hiện khá lâu. Các kết quả thực nghiệm cho thấy đây là một trong những phương pháp tốt nhất hiện nay cho bài toán người du lịch [1].

3. META-HEURISTIC - KẾT HỢP THUẬT GIẢI DI TRUYỀN VỚI THÔNG TIN THỐNG KÊ XÁC SUẤT CHO BÀI TOÁN NGƯỜI DU LỊCH

Hướng tiếp cận của chúng tôi dựa trên sự kết hợp giữa phương pháp tìm kiếm Genetics và thống kê tần suất xuất hiện các cạnh trong chu trình tối ưu. Trong quá trình quan sát các cá thể ở các thế hệ, chúng tôi nhận thấy các cạnh có chi phí nhỏ, mang tính chất trọng yếu có tần suất xuất hiện trong các cá thể chu trình chọn lọc vượt trội hơn hẳn so với các cạnh có chi phí lớn. Do đó, sau khi thống kê về tần suất xuất hiện các cạnh qua toàn bộ quá trình tiến hóa, chúng tôi sẽ thực thi lại quy trình tìm kiếm Genetics nhưng có sử dụng thêm thông tin thống kê.

Gọi F_{e_i} là tần suất xuất hiện cạnh e_i trong các cá thể chu trình qua toàn bộ quá trình tiến hóa. Ban đầu $F_{e_i} = 0$, $\forall e_i$, ta tính các F_{e_i} từ các cá thể tạo ra có chi phí "chấp nhận được" ở thế hệ thứ t như sau:

$S_m = 0$

For mỗi cá thể Y thuộc quần thể $P(t)$ do

    If $C(Y) < Cx$ then

        For mỗi $e \in Y$ do

            $F_e = F_e + 1$

        End For

        $S_m = S_m + n$

    End If

End For

Từ đó ta chuẩn hóa thành giá trị xác suất cuối cùng của các các cạnh biểu diễn khả năng xuất hiện trong chu trình tối ưu

$$\Pr(e_i) = \frac{F_{e_i}}{S_m},\ \forall e_i \in E,\ i = 1, ..., n.$$

Việc xác định chi phí "chấp nhận được" một cá thể chu trình sẽ dựa vào một ngưỡng Cx xác định trước nhờ các thuật giải gần đúng. Điều này cho phép ta loại bớt những cá thể xấu gây nhiễu thông tin trong bảng thống kê tần suất.

Cuối cùng, chúng ta sẽ thực hiện lại quy trình tìm kiếm Genetics với tri thức bổ sung là các giá trị thống kê. Trong lần tìm kiếm thứ 2 này, ta không còn phải thống kê tần suất xuất hiện các cạnh, nhưng ta sẽ thay đổi phương pháp lai ghép. Phương pháp lai ghép mới này như sau:

Từ hai cá thể cha mẹ P_1, P_2, ta chọn 2 điểm cắt ngẫu nhiên c_1 và c_2

$$\begin{array}{rcc|cccc|cc} & & & c_1 & & c_2 & & & \\ P_1 = & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ P_2 = & 4 & 7 & 1 & 8 & 2 & 5 & 3 & 6 \end{array}$$

Đoạn nằm giữa điểm cắt $[c_1, c_2]$ sẽ được giữ nguyên và sao chép vào cá thể con:

$$\begin{array}{rcc|cccc|cc} O_1 = & x & x & 3 & 4 & 5 & 6 & X & x \\ O_2 = & x & x & 1 & 8 & 2 & 5 & x & x \end{array}$$

Kế tiếp, ta bắt đầu hoàn thiện các cá thể con bằng cách điền vào các ô còn trống, bắt đầu từ vị trí $c_2 + 1$. Quy tắc để điền vị trí i bất kỳ như sau:

Lấy đỉnh u nằm ở vị trí $i - 1$ trong cá thể con đang xét. Trong các đỉnh nằm trước và sau u trong cả hai cá thể cha mẹ P_1, P_2, ta chọn một đỉnh hợp lệ mà cạnh e_k giữa nó với u có *giá trị thống kê* F_{e_i} *là lớn nhất.* Trong trường hợp tất cả các đỉnh này đều không hợp lệ, ta sẽ chọn đỉnh hợp lệ trong tất cả các đỉnh còn lại mà có cạnh nối với u là nhỏ nhất.

Với đồ thị trên, sau quá trình thực hiện GA lần đầu tiên để thống kê tần suất xuất hiện các cạnh, ta được bảng thống kê qua quá trình thực hiện GA với 50 thế hệ, mỗi quần thể có 50 cá thể:

Bảng 1. Bảng xác suất xuất hiện các cạnh qua quá trình thực hiện GA

	1	2	3	4	5	6	7	8
1	0	0,04227	0,02486	0,02546	0,02485	0,04499	0,03049	0,03478
2		0	0,18643	0,02241	0,02026	0,04406	0,02694	0,05074
3			0	0,03864	0,02117	0,04573	0,02286	0,01675
4				0	0,03665	0,03585	0,04547	0,02503
5					0	0,04068	0,02482	0,01871
6						0	0,00249	0,02077
7							0	0,02583
8								0

Việc điền vào các cá thể con sẽ bắt đầu từ vị trí kế điểm cắt c_2. Ta muốn điền vào vị trí $i = c_2 + 1$ trong cá thể con O_2. Đỉnh ở vị trí $i - 1$ là 5. Trong 4 đỉnh nằm trước và sau đỉnh 5 trong P_1, P_2 là 2, 3, 4, 6 , ta sẽ phải chọn 1 đỉnh hợp lệ mà có *giá trị thống kê với đỉnh 5 là lớn nhất.* Theo bảng giá trị thống kê thì đó là đỉnh 6, cá thể con O_2 sẽ như sau:

$$O_2 = x \quad x \mid 1 \;\; 8 \;\; 2 \;\; 5 \mid 6 \quad x$$

Ta điền tiếp tục, kế đỉnh 6 (trong 2 cá thể cha mẹ P_1, P_2) là 5, 7, 3, 4. Theo bảng giá trị thống kê thì đỉnh tiếp theo sẽ là đỉnh 3:

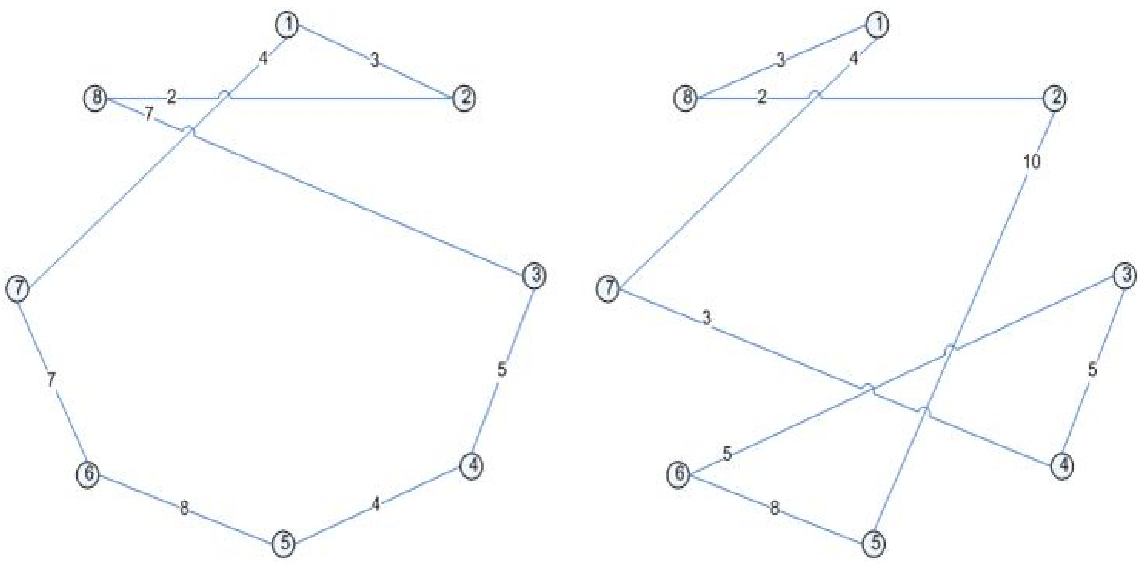
$$O_2 = x \quad X \mid 1 \;\; 8 \;\; 2 \;\; 5 \mid 6 \quad 3$$

Cứ tiếp tục cho đến khi hết tất cả những đỉnh chưa điền vào O_2, O_2 sẽ là:

$$O_2 = 4 \quad 7 \mid 1 \;\; 8 \;\; 2 \;\; 5 \mid 6 \quad 3 \qquad w_2 = 40$$

Tương tự, O_1 sẽ là:

$$O_1 = 2 \quad 8 \mid 3 \;\; 4 \;\; 5 \;\; 6 \mid 7 \quad 1 \qquad w_1 = 40$$



Hình 4. Hai cá thể con O'_1 và O'_2 lai ghép từ bố mẹ P_1 và P_2 theo phép lai thứ 2

Ta nhận thấy, giữa hai phép lai đã trình bày ở trên, phép lai thứ 2 cho ra O'_2 có trọng lượng nhỏ hơn O_2.

Mô hình thực hiện toàn bộ phương pháp tìm kiếm này như sau:

{Thực hiện GA lần thứ 1 để thống kê tần suất xuất hiện các cạnh}

$t = 0$

$S_m = 0$

Khởi tạo quần thể ban đầu $P(0)$

Tạo một cá thể Cx làm ngưỡng

While ($t <$ số đời tiến hóa) do
{*Tính độ thích nghi các cá thể trong quần thể* $P(t)$}
For mỗi cá thể Y thuộc quần thể $P(t)$ do

$$C(Y) = \sum_{i=1}^{n-1} w(Y(i), Y(i+1)) + w(Y(n), Y(1))$$

End For
{*Thống kê tần suất xuất hiện các cạnh trong quần thể* $P(t)$ }
For mỗi cá thể $Y \in$ quần thể $P(t)$ do
If $C(Y) < Cx$ then
For mỗi $e \in Y$ do
$F_e = F_e + 1$
$S_m = S_m + n$
End For
End For
For mỗi $e_i \in$ tập cạnh E do
$Pr(e_i) = \dfrac{F_{e_i}}{S_m}$
End For
$t = t + 1$ {*Chuyển sang thế hệ tiếp theo*}
Chọn lọc quần thể mới $P(t)$ từ $P(t-1)$
Lai ghép một số cá thể trong quần thể $P(t)$
Đột biến một số cá thể
End While
{*Thực hiện GA lần thứ 2* }
$t = 0$
Khởi tạo quần thể ban đầu $P(0)$
While $t <$ số đời tiến hóa do
{*Tính độ thích nghi các cá thể trong quần thể* $P(t)$ }
For each cá thể $Y \in$ quần thể $P(t)$ do

$$C(Y) = \sum_{i=1}^{n-1} w(Y(i), Y(i+1)) + w(Y(n), Y(1))$$

End For
$t = t + 1$ {*Chuyển sang thế hệ tiếp theo*}
Chọn lọc quần thể mới $P(t)$ từ $P(t-1)$
Lai ghép một số cá thể trong quần thể $P(t)$ có sử dụng thông tin thống kê
Đột biến một số cá thể
End While

4. THỰC NGHIỆM VÀ KẾT LUẬN

4.1. Thực nghiệm

Chúng tôi đã cài đặt cả 3 phương pháp: Christofide, Genetics và Genetics kết hợp thống kê để so sánh, và thử nghiệm trên nhiều bộ dữ liệu khác nhau: Dữ liệu được phát sinh ngẫu nhiên trên đồ thị đầy đủ tập đỉnh lần lượt là 100, 500, 1000, 3000, 5000 và các bộ dữ liệu có sẵn trên Internet. Chi phí mỗi cạnh không vượt quá 500.

Bảng 2. Kết quả thử nghiệm các phương pháp

Số đỉnh	Christofide		Genetics		Genetics và thống kê	
	Thời gian (giây)	Chi phí tìm được	Thời gian (giây)	Chi phí tìm được	Thời gian (giây)	Chi phí tìm được
100	59	4312	102	3239	124	2898
500	376	5134	623	4300	854	4074
1000	781	5769	1139	5060	1573	4061
3000	1937	8943	3286	8369	5572	7139
5000	4316	13863	5842	12389	9614	9347
10000	7893	20482	11032	17024	18347	12739
20000	15034	34927	24831	27566	38176	22371

Kết quả cho thấy với cùng thông số (số đời tiến hóa, số cá thể, tỷ lệ lai, tỷ lệ đột biến), phương pháp Genetics kết hợp thống kê mặc dù thực thi mất nhiều thời gian hơn (do phải thực hiện việc thống kê tần suất xuất hiện các cạnh và thực hiện phát sinh chu trình từ kết quả thống kê) nhưng cho ra kết quả tốt hơn. Điều này cho thấy việc phát sinh chu trình có giá trị thống kê hỗ trợ là đúng đắn. Kết quả thực nghiệm được chạy trên máy Pentium 1.4 GHz, Ram 256 MB, hệ điều hành Windows XP.

+ Dữ liệu thử nghiệm được lấy từ trang web

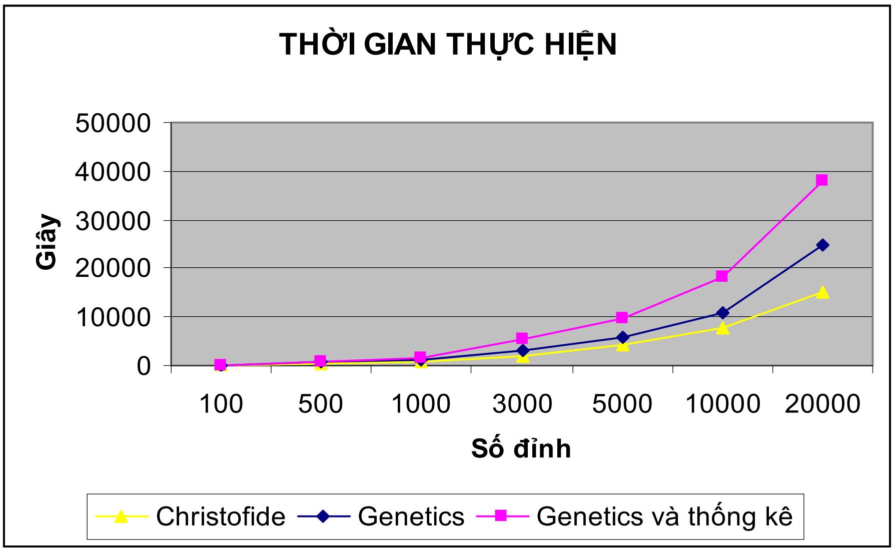
http://www.tsp.gatech.edu/world/index.html

+ Số đời tiến hóa trong thuật toán Genetics: 2000

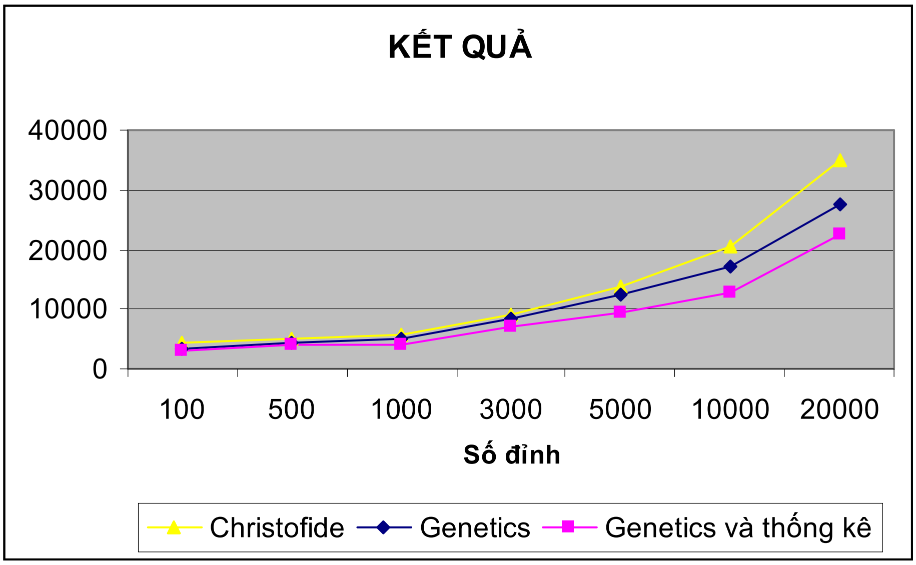
+ Số cá thể trong quần thể: 1000

4.2. Đánh giá và kết luận

So sánh về tốc độ thực thi, thuật toán Christofide chạy nhanh nhất, thời gian thực thi tăng theo đa thức, đó là thời gian tìm kiếm cây phủ tối thiểu. Ngược lại thuật toán Genetics kết hợp thống kê thực thi chậm nhất do phụ thuộc vào số đời tiến hóa trong thuật toán Genetics, và thuật toán tìm chu trình tối ưu từ dữ liệu thống kê được. Tuy nhiên kết quả chu trình tìm được của thuật toán Genetics và thống kê rõ ràng là tốt nhất so với các phương pháp còn lại (số đỉnh N càng lớn, các kết quả càng tốt hơn). Đây là một kết quả rất khả quan, mở ra một hướng tiếp cận mới ứng dụng phương pháp tìm kiếm này vào lớp các bài toán NP- đầy đủ.



Hình 5. Biểu đồ so sánh thời gian thực hiện



Hình 6. Biểu đồ so sánh kết quả tìm được

TÀI LIỆU THAM KHẢO

[1] Bernd Freisleben, Peter Merz, *New Genetic Local Search Operators for the Travelling Salesman Problem*, University of Siegen, 1996.

[2] M. R. Garey, D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-completeness*, W. H. Freeman, San Francisco, 1979.

[3] David S. Johnson, Lyle A. McGeoch, *The Traveling Salesman Problem: A Case Study in Local Optimization*, 1995.

[4] D. T. Pham, D. Karaboga, *Intelligent optimisation Techniques*, Springer-Verlag, London Limited, 2000.

[5] Michel Gondran, Michel Minouz and Steven Vajda, *Graphs and Algorithms*, John Wiley & Sons, Inc. New York, NY, USA, 1984.

[6] G. Reinelt, *The travelling Salesman Problem: Computational Solutions for TSP Application*, Springer-Verlag, Berlin, 1994.

[7] M. Groetschel, O. Holland, Solution of large-scale symmetric traveling salesman problem, *Math Programming* **51** (1991) 141–202.

[8] N. Christofides, "Worst-case anlysis of a new heuristic for the traveling salesman problem", Technical Report 388, Graduate School of Industrial Administration, Carnegie Mellon University, Pittsburgh, PA, 1976.

[9] E. J. Lawler, J. K. Lenstra, A. H. G. Rinnoy Kan, and D. B. Shmoys, *The Travelling Salesman Problem: A Guided Tour of Combinatorial Optimization*, Wiley, 1985.

[10] D.S. Johnson, G. Gutin, L. A. McGeoch, A. Yeo, W. Zhang, and A. Zverovich, Experimental analysis of heuristics for the ATSP, In G. Gutin and P.A. Punnen, editors, *The Traveling Salesman Problem and its Variations*, Kluwer Academic Publishers, 2002 (445-488).

[11] J. L. Bentley, Fast algorithms for geometric traveling salesman problems, *ORSA Journal on Computing* **4** (4) (1992) 387–411.

[12] C. Clarke, J. W. Wright, Scheduling of vehicles from a central depot to a number of delivery points, *Operations Research* **12 (1964) 568–581.**

Nhận bài ngày 30 - 11 - 2004
Nhận lại sau sửa ngày 24 - 3 - 2004