ÁP DỤNG THUẬT GIẢI DI TRUYỀN VỚI THÔNG TIN THỐNG KÊ XÁC SUẤT GIẢI QUYẾT BÀI TOÁN TÌM CHU TRÌNH HAMILTON

Nguyễn Thanh Hùng[1], **Hoàng Kiếm**[2]

[1] Trường Phổ thông Năng khiếu – Đại học Quốc gia Tp. Hồ Chí Minh

[2] Trung tâm Phát triển Công nghệ thông tin – Đại học Quốc gia Tp. Hồ Chí Minh

(Bài nhận ngày 29 tháng 9 năm 2004)

TÓM TẮT: *Bài toán tìm kiếm chu trình Hamilton 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 áp dụng vào các tình huống cụ thể trong thực tế.*

Chúng tôi đề xuất trong bài báo này là một hướng tiếp cận mới cho bài toán này, sử dụng thuật toán Genetics kết hợp với thông tin thống kê của tần suất các cung sẽ xuất hiện trong chu trình tối ưu. Chu trình tốt nhất sẽ được tổng hợp từ quá thực hiện tìm kiếm bằng Genetics kết hợp với tri thức bổ sung là thông tin thu lượm từ các thế hệ qua cả quát trình biến đổi của quần thể. Chúng tôi đánh giá cách tiếp cận mới với nhiều bài toán cụ thể khác nhau và so sánh với những cách tiếp cận đã được nghiên cứu. Kết quả thực nghiệm cho thấy cách tiếp cận này hiểu quả hơn với cách tiếp cận chỉ dùng thuật toán Genetics, từ đó mở ra một hướng mới để giải quyết các bài toán tìm kiếm gần đúng chu trình Hamilton tối ưu.

Từ khóa: Bài toán người du lịch, chu trình Hamilton, thuật giải di truyền, xác suất thống kê..

1. GIỚI THIỆU

Bài toán tìm chu trình Hamilton [5], còn gọi là bài toán Người du lịch (TSP – Traveling Salesman Problem), phát biểu như sau: Cho trước một đồ thị G(V, E) với tập đỉnh $V = \{v_1, v_2, v_3, ..., v_n\}$ và tập các cạnh $E = \{e_1, e_2, e_3, ..., e_m\}$ với chi phí từng cạnh là $W = \{w_{e_1}, w_{e_2}, w_{e_3}, ..., w_{e_m}\}$. Mục tiêu của bài toán là phải tìm được một chu trình đi qua tất cả các đỉnh của đồ thị, mỗi đỉnh một lần với chi phí là nhỏ nhất.

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. Đã có rất nhiều thuật toán giải gần đúng [5, 6, 7] được đưa ra nhằm tìm đến một lời giải gần tối ưu, tuy nhiên kết quả vẫn còn khá hạn chế khi số lượng đỉnh và cung khá lớn.

Trong bài báo này, chúng tôi đề nghị 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 quan sát và thấy rằng, 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. 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 (xem 4.1).

2. CÁC KẾT QUẢ NGHIÊN CỨU TRƯỚC ĐÂY

Đã có rất nhiều các thuật toán tìm kiếm gần đúng được đưa ra [3,5,6,7, 8,9,10,11,12], 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 Hamilton.

2.1. Phương pháp tìm kiếm gần đúng

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]. Trong đó 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 $CHRIS(I) / OPT(I) \leq 1.5$.

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>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 bước 3,4 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. Chúng tôi đã cài đặt thuật toán này dùng để so sánh với hướng tiếp cận của chúng tôi (mục 4 – Thực nghiệm).

2.2. Phương pháp tìm kiếm sử dụng giải thuật di truyền

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 Hamilton, 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(Ind).

Với cách biểu diễn cá thể như vậy, chúng ta sẽ sử dụng phép lai 2 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:

			c_1			c_2		
$P_1 =$	1	2	3	4	5	6	7	8
$P_2 =$	4	7	1	8	2	5	3	6

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

$O_1 =$	X	x	3	4	5	6	x	x
O2 =	X	x	1	8	2	5	x	x

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.

Ví dụ, 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. Giả sử đó là đỉnh 6, thì cá thể con O_2 sẽ như sau:

$O_2 =$	x	x	1	8	2	5	4	x

Việc điền vào các cá thể con sẽ bắt đầu từ vị trí kế điểm cắt c_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 Hamilton 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:
For each cá thể Y ∈ quần thể P(t) do

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

End For

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 Hamilton [1].

3. HƯỚNG TIẾP CẬN ĐỀ NGHỊ

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:

For each cá thể Y ∈ quần thể P(t) do
If C(Y) < Cx then
$F_e = F_e + 1, \forall e \in Y$
End If
End For

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:

			c_1			c_2		
$P_1 =$	1	2	3	4	5	6	7	8
$P_2 =$	4	7	1	8	2	5	3	6

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

O_1 =	X	x	3	4	5	6	x	x
O_2 =	X	x	1	8	2	5	x	x

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.

Ví dụ, 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 và có trọng số cạnh nối với đỉnh 5 là nhỏ nhất. Giả sử đó là đỉnh 4, thì cá thể con O_2 sẽ như sau:

O_2 =	x	x	1	8	2	5	4	x

Việc điền vào các cá thể con sẽ bắt đầu từ vị trí kế điểm cắt c2.

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
Khởi tạo quần thể ban đầu P(0)
Tạo một cá thể C_x làm ngưỡng
While t < số đời tiến hóa do
{Tính độ thích nghi các cá thể trong quần thể P(t) }
For each cá thể Y ∈ quần thể P(t) do
$$C(Y) = \sum_{i=1}^{n} w(Y(i), Y(i+1))$$
End For
{Thống kê tần suất xuất hiện các cạnh trong quần thể P(t) }
For each cá thể Y ∈ quần thể P(t) do
If C(Y) < C_x then
$$F_e = F_e + 1, \forall e \in Y$$
End If
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 ∈ quần thể P(t) do
$$C(Y) = \sum_{i=1}^{n} w(Y(i), Y(i+1))$$
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

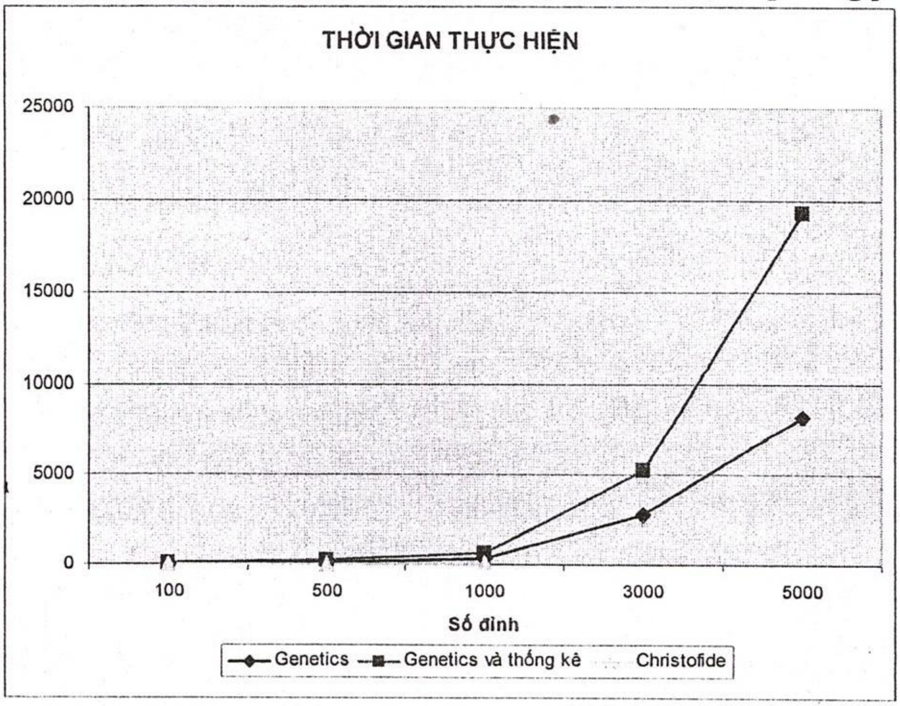
4.1. Thực nghiệm

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. 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.

- Số đời tiến hóa trong thuật toán Genetics: 2000
- Số cá thể trong quần thể: 1000
- Các dữ liệu thử nghiệm là đồ thị đầy đủ, với chi phí mỗi cạnh < 500

Số đỉnh	Christofide	Genetics	Genetics và thống kê
100	4312	3239	2898
500	5134	4300	4074
1000	5769	5060	4061
3000	8943	8369	7139
5000	13863	12389	9347

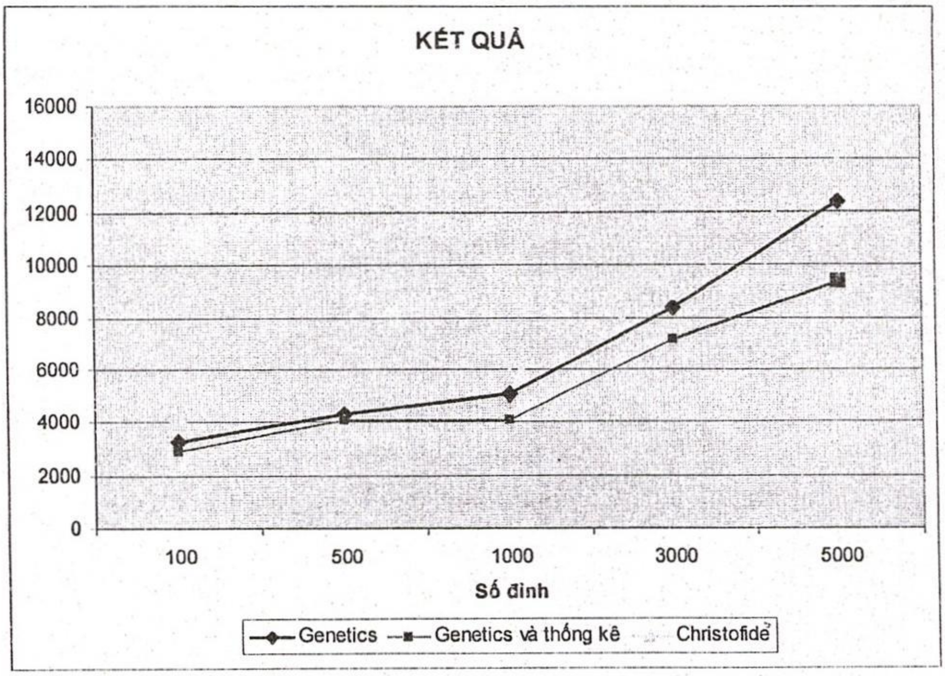
Hình 1. Bảng kết quả so sánh thời gian thực thi của ba phương pháp



Hình 2. Biểu đồ so sánh thời gian thực hiện

Số đỉnh	Christofide	Genetics	Genetics và thống kê
100	4312	3239	2898
500	5134	4300	4074
1000	5769	5060	4061
3000	8943	8369	7139
5000	13863	12389	9347

Hình 3. Bảng so sánh kết quả thử nghiệm của ba phương pháp



Hình 4. Biểu đồ so sánh kết quả tìm được

4.2. 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.

APPLYING GENETICS ALGORITHM AND STATISTICAL PROBABILITY RESULTS TO SOLVE HAMILTON CYCLE PROBLEM

Nguyễn Thanh Hùng(1)**, Hoàng Kiếm**(2)
(1)High School for the gifted students – Vietnam National University in Ho Chi Minh city
(2) Center of Information Technology Development - Vietnam National University in Ho Chi Minh city

ABSTRACT: *Finding Hamilton cycle 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 that combines Genetics algorithm and appearance probabilities of edges in the optimal cycle. The final solution will be selected from results of generations gained in the whole population evolution. 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 statistical probability results. Hence, it proposes a new approach to solve heuristically problems of finding optimal Hamilton cycle.*

Keywords: *Travelling Salesman Problem, Hamilton cycle, Genetics algorithm, statistical probability.*

TÀI LIỆU THAM KHẢO

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

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

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

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

[5]. Michel Gondran, Michel Minouz and Steven Vajda, *"Graphs and Algorithms"*, A Wiley-Interscience Publication.

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

[7]. M. Groetschel and O.Holland, *"Solution of large-scale symmetric traveling salesman problem"*, Math Programming 1991.

[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, pages 445–488. Kluwer Academic Publishers, 2002.

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

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