

NGHIÊN CỨU KẾT HỢP THUẬT TOÁN GA VÀ THUẬT TOÁN K-MEANS ĐỂ GIẢI BÀI TOÁN ĐỊNH TUYẾN XE

RESEARCH ON COMBINATION OF GA ALGORITHM AND K-MEANS ALGORITHM TO SOLV THE VEHICLE ROUTING PROBLEM

Cao Ngọc Ánh, Trần Bích Thảo

Khoa Công nghệ thông tin, Trường Đại học Kinh tế - Kỹ thuật Công nghiệp

Đến Tòa soạn ngày 15/05/2024, chấp nhận đăng ngày 03/06/2024

Tóm tắt: Mục tiêu chính của nghiên cứu này là tìm ra giải pháp cho vấn đề định tuyến xe bằng cách sử dụng các thuật toán di truyền kết hợp với thuật toán phân cụm K-Means. Bài toán định tuyến xe (Vehicle Routing Problem -VRP) là một bài toán tối ưu hóa tổ hợp phức tạp thuộc lớp NP-đầy đủ (nondeterministic polynomial -complete). Giải bài toán định tuyến xe là giải bài toán tối ưu hoá tổ hợp nhằm tìm ra lộ trình tối ưu cho một đội xe phân phối hàng hoá từ kho đến nhiều địa điểm khách hàng, sao cho chi phí di chuyển là thấp nhất, đồng thời tuân thủ các ràng buộc về thời gian và sức chứa của xe. Bài báo nghiên cứu thuật toán di truyền và kỹ thuật phân cụm K-Means để tìm ra giải pháp đúng hoặc gần đúng đến các vấn đề tối ưu hóa và tìm kiếm để giải bài toán định tuyến xe.

Từ khóa: Vehicle Routing Problem (VRP), genetic algorithm, kmeans, ga-k-means.

Abstract: The main objective of this research is to find a solution to the Vehicle Routing Problem using genetic algorithms combined with K-Means clustering algorithm. Vehicle Routing Problem (VRP) is a complex combinatorial optimization problem of class NP-complete. Solving the vehicle routing problem is solving a combinatorial optimization problem to find the optimal route for a fleet of vehicles to distribute goods from warehouses to multiple customer locations, so that travel costs are lowest and at the same time. comply with time and vehicle capacity constraints. The article studies genetic algorithms and K-Means clustering technique to find the correct or approximate solutions to optimization and search problems to solve vehicle routing problems.

Keywords: Vehicle Routing Problem (VRP), genetic algorithm, k-means.

1. GIỚI THIỆU

Có thể phát biểu bài toán định tuyến xe (Vehicle Routing Problem – VRP) một cách cơ bản như sau: Có một tập hợp K xe giống nhau cùng xuất phát tại một kho hàng đi làm nhiệm vụ giao hàng cho N khách hàng, mỗi khách hàng (khách hàng có vị trí địa lý cụ thể) có các ràng buộc nhất định: về nhu cầu giao hàng và thời gian nhận hàng cho trước. Yêu cầu đặt ra của bài toán là tìm đường đi ngắn nhất cho K xe đáp ứng được tất cả các đòi hỏi

của khách hàng [1].

Bài toán định tuyến xe bắt nguồn từ năm 1959 khi George Dantzig và John Ramser [2], [3] thiết lập công thức toán học và phương pháp tiếp cận bằng thuật toán để giải quyết vấn đề cung cấp xăng dầu cho các trạm dịch vụ. Năm 1964, Clarke và Wright đã cải tiến cách giải của Dantzig và Ramser bằng cách sử dụng một cách tiếp cận khác, được gọi là thuật toán tiết kiệm. Từ đó thì sự quan tâm đến VRP đã được mở rộng từ một nhóm các nhà toán học

sang phạm vi rộng các nhà nghiên cứu và các nhà thực hành, từ các ngành khác nhau, trong nhiều lĩnh vực

Các phương pháp giải chính xác đảm bảo lời giải tối ưu sẽ được tìm thấy trong một khoảng thời gian hữu hạn. Tuy nhiên, thời gian chạy của nó trong trường hợp tốt nhất là rất lớn, do đó lớp thuật toán này chỉ giải được những bài toán có kích thước nhỏ hoặc vừa như thuật toán nhánh cận (branch and bound) [4], thuật toán sinh cột [5],...

Thuật toán di truyền (GA) giúp giải quyết bài toán tối ưu hoá tuyến đường một cách nhanh chóng. Nghiên cứu của Philip [6] sử dụng thuật toán di truyền cho lộ trình phân phối hàng hoá; nghiên cứu của Mnasri [7] tối ưu hoá trong bài toán định tuyến xe với ràng buộc cửa sổ thời gian (VRPTW: Vehicle Routing Problem Time Window) bằng cách cải tiến trao đổi chéo trong GA dẫn đến tuyến đường tốt nhất và chi phí tối thiểu.

Do chi phí qua tuyến một lần nên để phục vụ được nhiều khách hàng, phân loại khách hàng theo mức độ ưu tiên, phương pháp phân cụm K-Means là lựa chọn hiệu quả có giải pháp nhanh chóng [12], [13], [14].

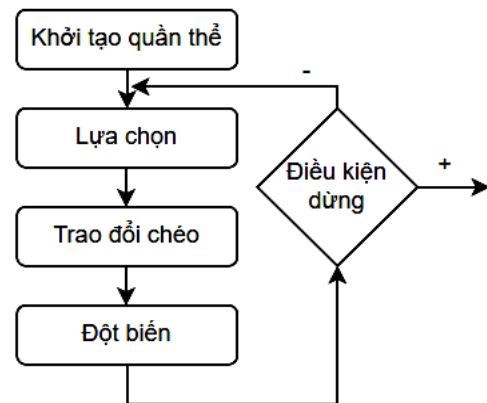
Trọng tâm của nghiên cứu này là giải quyết vấn đề phức tạp của VRPTW bằng cách kết hợp GA và phương pháp phân cụm K-Means trong đó K-Means thực hiện phân cụm theo khu vực địa lý của khách hàng

Giải thuật di truyền GA:

Thuật toán di truyền là kỹ thuật tối ưu hoá ngẫu nhiên dựa trên thuyết tiến hoá của Darwin. Thuật toán di truyền gồm việc biểu diễn nhiệm vụ, xây dựng quần thể, tính độ thích nghi, sử dụng lựa chọn, lai ghép và đột biến để sinh sản thế hệ các cá thể mới

Các bước chính của thuật toán di truyền được

thể hiện bằng hình 1.

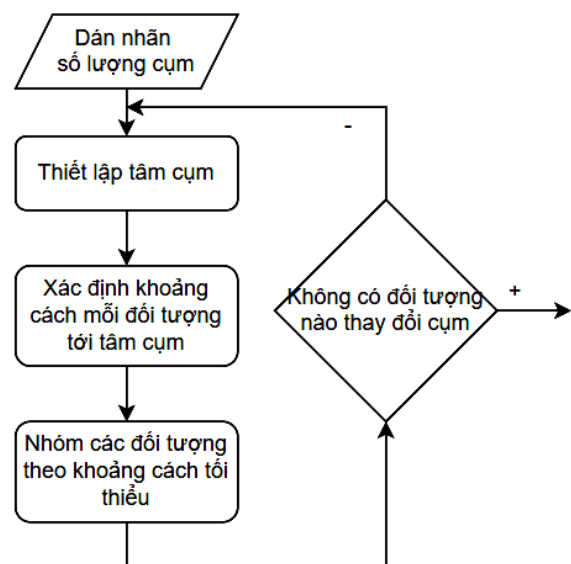


Hình 1. Các bước của thuật toán di truyền

Kỹ thuật phân cụm K-Means:

Phân cụm K-Means là thuật toán không giám sát được sử dụng để phân cụm các đối tượng khác nhau thành các cụm. Một cụm là tập hợp các đối tượng dữ liệu đồng nhất trong một cụm và không đồng nhất với đối tượng trong cụm khác

Thuật toán K-Means có thể được biểu diễn bởi hình 2.



Hình 2. Các bước của thuật toán K-Means

2. MÔ HÌNH TOÁN HỌC

Bài toán VRP được biểu diễn bởi một tập V các xe vận chuyển, một tập C các khách hàng

và một đồ thị có hướng G . Đồ thị G có $|C|+2$ đỉnh, các khách hàng là các đỉnh từ 1, 2, ..., n ; đỉnh 0 biểu thị kho xuất phát và đỉnh $n+1$ biểu diễn cho kho kết thúc. Tập A các cung đồ thị biểu diễn kết nối kho với khách hàng và giữa khách hàng với nhau. Mỗi cung (i, j) có giá trị chi phí c_{ij} , (trong bài báo này, chi phí giữa hai khách hàng là độ dài quãng đường của hai khách hàng đó). Mỗi khách hàng i có nhu cầu đi. Một lộ trình bắt đầu từ kho xuất phát, đi qua một số khách hàng và kết thúc tại kho hàng kết thúc, mỗi khách hàng được một xe phục vụ. Mục tiêu của bài toán là tìm một phương án điều đội xe hợp lý để giao hàng tới tất cả các khách hàng sao cho tổng chi phí là nhỏ nhất.

Mô hình toán học của bài toán như sau, giả sử:

n : Tổng số khách hàng cần phục vụ;

K : Tổng số chuyến xe;

$N = \{0, 1, \dots, n+1\}$: Tập các đỉnh của đồ thị;

$V = \{0, 1, \dots, K\}$: Tập các xe vận chuyển;

$C = \{1, 2, \dots, n\}$: Tập các khách hàng;

c_{ij} : Chi phí đi từ i đến j ;

q : Sức chứa của một xe;

d_i : Nhu cầu tại nút i .

Khi đó ta có các biến trong mô hình:

x_{ijk} : biểu diễn xem một xe thứ k có đi từ khách hàng i đến khách hàng j không với

$$x_{ijk} = \begin{cases} 1 & \text{nếu xe thứ } k \text{ đi từ } i \text{ đến } j \\ 0 & \text{với các trường hợp còn lại} \end{cases}$$

trong đó $i, j \in N, k \in V$.

y_{ik} : biểu diễn khách hàng i có được xe k phục vụ không với

$$y_{ik} = \begin{cases} 1 & \text{nếu xe thứ } k \text{ đi đến } i \\ 0 & \text{với các trường hợp còn lại} \end{cases}$$

trong đó $ij \in N, k \in V$.

Hàm mục tiêu của bài toán:

$$\min \sum_{i \in N} \sum_{j \in N} \sum_{k \in N} c_{ij} x_{ijk}$$

Các ràng buộc của bài toán:

$$\sum_{i \in N} x_{ijk} = y_{ijk} \in N, k \in N \quad (1)$$

$$\sum_{i \in N} \sum_{k \in N} x_{ijk} = 1 \forall j \in C \quad (2)$$

$$\sum_{i \in N} x_{ijk} - \sum_{i \in N} x_{ijk} = 0 \quad \forall j \in C, k \in K \quad (3)$$

$$\sum_{j \in N} y_{jk} \cdot d_j \leq q \quad \forall k \in K \quad (4)$$

$$\sum_{j \in N} x_{0jk} = 1 \quad \forall k \in K \quad (5)$$

$$\sum_{i \in N} x_{i0k} = 0 \quad \forall k \in K \quad (6)$$

$$\sum_{i \in N} x_{i(n+1)k} = 1 \quad \forall k \in K \quad (7)$$

$$\sum_{j \in N} x_{(n+1)jk} = 0 \quad \forall k \in K \quad (8)$$

trong đó: (1) biểu diễn mối liên hệ giữa x và y . (2), (3): đảm bảo mỗi khách được một xe phục vụ. (4) đảm bảo sức chứa một xe. (5), (6), (7), (8) đảm bảo mọi lộ trình đều bắt đầu và kết thúc tại kho

Việc mô hình hóa bài toán VRP được nghiên cứu bởi nhiều tác giả và đã được trình bày chi tiết trong các bài báo [8], [9], [10].

3. THUẬT TOÁN GA-KMeans GIẢI BÀI TOÁN VRP

Các tham số chính cần thiết để chạy thuật toán GA-K-Means được thể hiện bởi các bước: Biểu diễn nhiễm sắc thể, tính giá trị hàm mục tiêu, lựa chọn, lai ghép và đột biến, trong đó K-Means được áp dụng trong khởi tạo quần thể theo nhóm cụm địa lý.

3.1. Biểu diễn nhiễm sắc thể

Mỗi một nhiễm sắc thể là chuỗi K đoạn (với K là số xe hay số cụm ban đầu), mỗi đoạn là thứ tự các khách hàng cần chăm sóc trong

cụm tương ứng đã được phân cụm theo khoảng cách địa lý, cấu trúc của nhiệm sắc thể được minh họa trong hình 3: có 9 khách hàng trải trên 3 cụm ứng với 3 xe; cụm 1 gồm xe 1 phục vụ lần lượt khách hàng 1, 2, 5, 4, 6; cụm 2 ứng xe 2 phục vụ khách hàng 3 và 7; cụm 3 ứng xe 3 phục vụ khách hàng 8 và 9.

KH	1	2	5	4	6	3	7	8	9
NST	1	2	5	4	6	3	7	8	9

Cụm 1

Cụm 2

Cụm 3

Hình 3. Minh họa một nhiệm sắc thể

3.2. Giá trị hàm độ thích nghi

Giá trị độ thích nghi là tổng chi phí của tất cả các xe trong tất cả các cụm địa lý.

3.3. Lai ghép

Lai ghép là quá trình sinh sản dựa trên 2 nhiệm sắc thể có sẵn để tạo ra nhiệm sắc thể mới có giá trị thể lực tốt hơn. Sau khi cha mẹ được chọn thì chọn gen cắt ngẫu nhiên để được nhiệm sắc thể con. Quá trình lai ghép được minh họa như hình 4:

NST1	1	2	5	4	6	3	7	8	9
NST2	1	2	5	4	6	3	7	8	9
NSTcon	4	6	1	2	5	3	7	8	9

Cụm 1

Cụm 2

Cụm 3

Hình 4. Quá trình và kết quả lai ghép

3.4. Đột biến

Thuật toán sử dụng kỹ thuật đột biến trao đổi: Nhiệm sắc thể bố mẹ được chọn, sử dụng đột biến gen ngẫu nhiên: để tạo được nhiệm sắc thể mới (hình 5).

NST1	1	2	5	4	6	3	7	8	9
NSTcon	1	2	3	4	6	5	7	8	9

Cụm 1

Cụm 2

Cụm 3

Hình 5. Quá trình và kết quả đột biến

3.5. Lựa chọn

Phương pháp lựa chọn hoạt động dựa trên

sàng lọc giá trị độ thích nghi trên các nhiệm sắc thể.

3.5. Điều kiện dừng

Ở một số giai đoạn, quá trình tìm kiếm lặp đi lặp lại phải kết thúc. Trong thuật toán GA-K-Means, có một số cách để kết thúc, chẳng hạn:

- Sau một số lần lặp cố định.
- Sau một số lần lặp lại mà không có sự cải thiện trong giá trị hàm mục tiêu.
- Khi giá trị hàm mục tiêu đạt đến giá trị ngưỡng được chỉ định trước.

Như vậy, thuật toán GA-KMeans giải bài toán VRP được thể hiện như sau:

Bước 1: Thiết lập số xe (K) ứng số cụm

Thực hiện phân K cụm với K xe

Khởi tạo lời giải ban đầu

Bước 2: Khởi tạo quần thể

- Sắp thứ tự khoảng cách các khách hàng trên mỗi cụm theo từng xe

- Tính độ thích nghi ban đầu

Bước 3: Thực hiện lai ghép 2 cá thể

Bước 4: Thực hiện đột biến cá thể

Bước 5: Lựa chọn các cá thể trong quần thể dựa trên độ thích nghi

Bước 6: Nếu thỏa mãn “điều kiện dừng” thì dừng lại, ngược lại thì chuyển quay lại bước 3

4. THỬ NGHIỆM VÀ ĐÁNH GIÁ THUẬT TOÁN

Dưới đây, thuật toán GA-K-Means được cài đặt bằng ngôn ngữ lập trình Python 3.0. Chương trình được viết và biên dịch trên Colab. Dữ liệu được thực nghiệm trên bộ dữ liệu Solomon [11].

Các thông số của thuật toán được cài đặt như sau:

- Kích thước tối đa quần thể: 250;
- Số lượng thế hệ tối đa: 100;
- Xác suất trao đổi chéo: 80%;
- Xác suất đột biến tối đa: 18%;
- Tải trọng của xe: 200;
- Số xe sử dụng tối đa: 20 xe.

Hiệu quả của các thuật toán phụ thuộc vào khởi tạo quần thể và xác suất trao đổi chéo, xác suất đột biến. Kết quả có được khi so sánh thực nghiệm GA-K-Means với số liệu tham khảo trên website

<https://www.sintef.no/projectweb/top/vrptw/200-customers/> [11] áp dụng với giải thuật GA truyền thống, chi tiết được thể hiện trong bảng 1.

Trong bảng 1, cột đầu thể hiện 4 nhóm mẫu dữ liệu với các bộ dữ liệu tiêu chuẩn số lượng khách hàng lần lượt là 100, 200, 300, 400 khách hàng; cột 2 là giá trị tối ưu mà giải pháp đạt được (Fitness: giá trị hàm mục tiêu), cột 3: là trung bình thời gian thực hiện kết quả thực nghiệm. Với kết quả thực nghiệm, với dữ

liệu khách hàng từ 300 khách hàng đến 400 khách hàng, thuật toán GA-K-Means cho kết quả tốt hơn.

Bảng 1. So sánh kết quả trung bình giữa các lần thực nghiệm

KH	Fitness (GA/GA-K-Means)	Thời gian tính toán (s) (GA/GA-K-Means)
100	649/673	1105/768
200	1706/1389	2300/1520
300	5804/3602	3010/2212
400	8571/6086	3900/2841

5. KẾT LUẬN

Thuật toán kết hợp GA-K-Means cung cấp một cách tiếp cận rất thú vị để giải quyết các vấn đề không thể áp dụng một phương pháp chính xác và phức tạp trong mức độ ưu tiên theo cụm. Bài báo đã trình bày rõ thêm các kiến thức cơ bản định tuyến xe, làm rõ thuật toán GA, K-Means và kết hợp GA-K-Means đồng thời tiến hành thực nghiệm với bộ dữ liệu được phân loại cụ thể.

Việc gia tăng các giải thuật heuristic và kết hợp các phương pháp phân cụm trong các quá trình, bổ sung thêm các ràng buộc gắn liền thực tiễn là hướng nghiên cứu tiếp theo của bài báo.

TÀI LIỆU THAM KHẢO

- [1] M. Fisher, "Vehicle routing. Handbooks of Operations Research and Management Science", Chapter 1, 8:1-31, (1995).
- [2] Sergios Theodoridis, Konstantinos Kourtoumbas, "Pattern Recognition Second Edition. page 582.
- [3] Dantzing, G., Ramster, R., "The truck dispatching problem", Management Science 6, page 80-91, (1959).
- [4] G.Laporte, Y.Nobert, "A branch and bound algorithm for the capacitated vehicle routing problem", Operations Research Spektrum, pp 77-85, (1983).
- [5] F.Liberatore, G.Righini, M.Salani, "A column generation algorithm for the vehicle routing problem with soft time windows", 4OR quarterly journal of the Belgian, French and Italian Operations Research Societies 9(1), pp 49-82, (2010).
- [6] A. Philip, A. Taofiki, and O. Kehinde, "A Genetic Algorithm for Solving Travelling Salesman Problem", Int. J. Adv. Comput. Sci. Appl., vol. 2, no. 1, pp. 26–29, 2011.
- [7] S. Mnasri, F. Abbes, K. Zidi, and K. Ghedira, "A Multi-Objective Hybrid BCRC-NSGAI Algorithm to Solve the

- VRPTW,” 13th Int. Conf. Hybrid Intell. Syst., pp. 60–65, 2013.
- [8] Thangiah S.R, “Vehicle Routing with Time Windows using Genetic Algorithms”, In L. Chambers, editor, Application handbook of genetic algorithms: New Frontiers vol. II. Boca Raton: CRC Press, 253-277, (1995).
- [9] Tan K.C., Lee L.H., Zhu K.Q., Ou K, “Heuristic methods for vehicle routing problem with time windows”, Artificial Intelligence in Engineering, Volume 15, Issue 3, Pages 281-295, (2001).
- [10] Kallehauge B., Larsen J., Madsen O.B., Solomon M.M, “Vehicle Routing Problem with Time Windows”, In: Desaulniers G., Desrosiers J., Solomon M.M. (eds) Column Generation. Springer, Boston, MA, (2005).
- [11] <https://www.sintef.no/projectweb/top/vrptw/200-customers/>
- [12] J. MacQueen, “Some methods for clustering and analysis of multiobservations” in Proceedings of the fifth symposium on mathematical statistic and probability, 1967, pp. 281–297.
- [13] S. Kapil, M. Chawla, and M. D. Ansari, “On k-means data clustering algorithm with genetic algorithm” in International Conference on Parallel, Distributed and Grid Computing, 2016, pp. 202–206.
- [14] Madhu and Nagachandrika, “A New Paradigm for Development of Data Imputation Approach for Missing Value Estimation” Int. J. Electr. Comput. Eng., vol. 6, no. 6, pp. 3222–3228, 2016.

Thông tin liên hệ: **Cao Ngọc Ánh**

Điện thoại: 0912599122 – Email: cnanh@uneti.edu.vn

Khoa Công nghệ thông tin, Trường Đại học Kinh tế - Kỹ thuật Công nghiệp.