NGHIÊN CỨU KẾT HỢP GIẢI THUẬT DI TRUYỀN (GA) VÀ CÂY PHÂN LỚP (DECISION TREE) ĐỂ GIẢI BÀI TOÁN ĐỊNH TUYẾN XE (VRP)

RESEARCH ON COMBINING GENETIC ALGORITHM (GA) AND DECISION TREE TO SOLVE VEHICLE ROUTING PROBLEM (VRP)

Cao Ngọc Ánh, Trần Bích Thảo

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

Đến Tòa soạn ngày 26/4/2025, chấp nhận đăng ngày 19/5/2025

Tóm tắt: Bài báo này đề xuất một phương pháp kết hợp giữa giải thuật di truyền (Genetic Algorithm - GA) và kỹ thuật cây phân lớp (Decision Tree - DT) để giải bài toán định tuyến xe (Vehicle Routing Problem - VRP). Phương pháp sử dụng DT để dự đoán số lượng xe cần thiết, giúp cải thiện hiệu quả khởi tạo quần thể trong GA. Kết quả thực nghiệm trên tập dữ liệu chuẩn Solomon (gồm Rxx, Cxx và RCxx) cho thấy phương pháp kết hợp GA + DT mang lại lời giải chất lượng cao, sai số nhỏ và thời gian tính toán hợp lý.

Từ khóa: Bài toán định tuyến xe (VRP); Giải thuật di truyền, Cây quyết định, GA-DT.

Abstract: This paper proposes a method combining Genetic Algorithm (GA) and Decision Tree (DT) to solve the Vehicle Routing Problem (VRP). The method uses DT to predict the number of vehicles needed, helping to improve the efficiency of population initialization in GA. Experimental results on the Solomon benchmark dataset (including Rxx, Cxx and RCxx) show that the combined GA + DT method provides high-quality solutions, small errors and reasonable computation time.

Keywords: Vehicle Routing Problem (VRP); Genetic Algorithm, Decision Tree, GA-DT.

1. GIỚI THIỆU

VRP là một trong những bài toán tối ưu tổ hợp NP-khó quan trọng trong logistics. Trong khi GA là một trong những thuật toán được sử dụng phổ biến để giải VRP, việc xác định số lượng xe hợp lý để khởi tạo lời giải vẫn còn là một thách thức.

Một số công trình nghiên cứu về GA và Machine Learning nói chung và DT nói riêng trong những năm gần đây đã có những kết quả cụ thể: L.M. Rojas, C.E. Torres sử dụng cây phân lớp để phân loại cấu hình tuyến đường trước khi tối ưu bằng GA [6], S.D. Lin, K.H. Li sử dụng cây quyết định để phân cụm khách hàng theo nhu cầu, hỗ trợ GA tạo nhóm giao hàng tốt hơn [7], R.J. Sánchez, et al. đã trình bày tổng quan về các công trình kết hợp ML (Decision Tree, SVM, ANN) với metaheuristics như GA, PSO, ACO,... khẳng định dẫn chứng xu hướng kết hợp ML + GA là hướng mới hiệu quả [8]... Các nghiên cứu này chủ yếu dùng ML để phân cụm hoặc lựa chọn tuyến, bài báo này **ứng dụng cây phân lớp (Decision Tree)** (là một kỹ thuật học máy mạnh mẽ trong phân loại) **để dự đoán chính xác số lượng xe cần thiết**, từ đó giúp **khởi tạo quần thể GA hiệu quả hơn**. Cách tiếp cận này đơn giản, dễ triển khai và có thể mở rộng với các mô hình học máy khác (RF, XGBoost), đồng thời **giảm sai số và thời gian tính toán trong quá trình tối ưu hóa VRP [1][2].**

Bảng so sánh các công trình nghiên cứu liên quan trước đây được tổng hợp theo bảng 1.

Bảng 1. Tổng hợp so sánh các nghiên cứu gần đây về GA và DT

Bài báo/Tiêu chí	Phương pháp giải	Ứng dụng ML hỗ trợ GA	Khởi tạo quần thể thông minh	Tính tổng quát dữ liệu thực nghiệm
[1] Nalepa & Blocho (2016)	GA thuần túy	Không	Ngẫu nhiên	Một phần R
[2] Zhang et al. (2018)	GA kết hợp Tabu Search	Không	Ngẫu nhiên	Có R và RC
[3] Fidanova et al. (2021)	Ant Colony Optimization	Không	Ngẫu nhiên	Chỉ tập nhỏ
[4] Nguyen et al. (2020)	GA + Random Initialization	Không	Có nhưng chưa tối ưu	Một số C101, C102
[6] L. M. Rojas, C. E. Torres, et al. (2021)	GA + DT	Có	Có nhưng chưa tối ưu	Chỉ tập nhỏ
[7] S. D. Lin, K. H. Li. (2020)	GA + DT	Có	Có nhưng chưa tối ưu	Chỉ tập nhỏ
[8] R. J. Sánchez, et al. (2022)	GA + ML	Có	Có nhưng chưa tối ưu	Chỉ tập nhỏ

2. MÔ HÌNH BÀI TOÁN

Phát biểu bài toán:

Bài toán định tuyến xe (Vehicle Routing Problem – VRP) là một bài toán tối ưu tổ hợp cổ điển. Mục tiêu của bài toán là xác định tập hợp các tuyến đường tối ưu cho một đội xe xuất phát từ một kho hàng trung tâm (depot), nhằm phục vụ một tập khách hàng sao cho tổng chi phí (thường là quãng đường di chuyển) là nhỏ nhất, đồng thời thỏa mãn một số ràng buộc nhất định như: Mỗi khách hàng chỉ được phục vụ bởi đúng một xe, tổng nhu cầu của khách hàng trên mỗi tuyến không vượt quá sức chứa của xe, mỗi xe xuất phát và quay lại kho.

Đầu vào của bài toán VRP:

- Tập khách hàng cần phục vụ, mỗi khách hàng có:

- Vị trí (tọa độ) trên bản đồ.

- Nhu cầu hàng hóa q_i.

- Kho hàng trung tâm (depot).

- Ma trận khoảng cách giữa các điểm (dựa trên tọa độ Euclidean).

- Số lượng xe và sức chứa tối đa của mỗi xe Q.

Đầu ra của bài toán:

- Một tập các tuyến đường, mỗi tuyến ứng với một xe.

- Mỗi tuyến là một chuỗi khách hàng (danh sách điểm cần phục vụ).

- Tất cả các tuyến phải thỏa mãn:

- Không có khách nào bị phục vụ lặp lại.

- Không xe nào bị quá tải.

- Tổng chi phí (quãng đường) là nhỏ nhất.

Xử lý đặc trưng (Feature Engineering):

- Chuyển Mật độ khách hàng từ dạng chuỗi (Cao/Trung bình/Thấp) sang dạng số (2/1/0).

- Tính khoảng cách trung bình giữa kho và khách hàng.
- Tính tổng nhu cầu khách hàng (tổng q_i).

Cây phân lớp dự đoán số xe:

- Thuật toán: DecisionTreeClassifier (scikit-learn) [4].
- Đầu vào: Số khách hàng, tổng nhu cầu, khoảng cách trung bình.
- Đầu ra: Số lượng xe tối ưu.
- Độ chính xác: ~95% trên tập test.

Sau khi huấn luyện, có thể dùng model.predict() để **dự đoán số xe** từ dữ liệu mới

Kết quả *so_xe_du_doan* được truyền vào giai đoạn **khởi tạo quần thể** của GA:

- Giúp khởi tạo số cá thể với đúng/sát số xe cần thiết.
- Hạn chế tạo ra lời giải không khả thi (quá tải hoặc dư xe).
- GA tiếp tục tối ưu lộ trình di chuyển cho từng xe dựa trên số xe đã dự đoán.

Quy trình kết hợp giữa GA và DT bao gồm:

- Tiền xử lý: Trích xuất đặc trưng từ dữ liệu VRP: số khách, tổng nhu cầu, khoảng cách trung bình,...
- Huấn luyện DT: Sử dụng Decision Tree để dự đoán số lượng xe từ các đặc trưng đầu vào.
- Mỗi **cá thể (chromosome)** đại diện cho **một lời giải khả thi** của bài toán – tức là một tập các tuyến đường cho từng xe, biểu diễn dạng một **chuỗi hoán vị của các khách hàng.**

Ví dụ: Một cá thể: [0, 5, 2, 7, 0, 6, 3, 1, 0, 4, 8, 9, 0] biểu diễn phương án:

- Tuyến 1: xe 1 phục vụ khách 5 → 2 → 7.
- Tuyến 2: xe 2 phục vụ khách 6 → 3 → 1.
- Tuyến 3: xe 3 phục vụ khách 4 → 8 → 9.

Trong đó, 0 đại diện cho kho, đảm bảo rằng mỗi tuyến bắt đầu và kết thúc tại kho.

- Hàm đánh giá độ thích nghi (Fitness function): Tối thiểu hoá tổng quãng đường.
- Khởi tạo GA: Dùng số lượng xe dự đoán để khởi tạo quần thể ban đầu.
- Chạy GA:

Thực hiện chọn lọc bằng **Tournament Selection** chọn ngẫu nhiên k cá thể từ quần thể, sau đó **chọn cá thể tốt nhất (fitness thấp nhất)** để làm cha mẹ.

Thực hiện lai ghép **Order Crossover (OX):** giữ nguyên thứ tự một đoạn con từ cha và chèn các điểm còn lại từ mẹ theo đúng thứ tự.

Ví dụ:

Cha: [2 4 | 5 6 7 | 3 1 8]

Mẹ: [6 1 8 2 3 4 5 7]

Con: [3 1 | 5 6 7 | 2 4 8]

Thực hiện kỹ thuật đột biến theo phương pháp Swap Mutation chọn ngẫu nhiên hai vị trí (hai khách hàng) trong lời giải và hoán đổi vị trí của chúng.

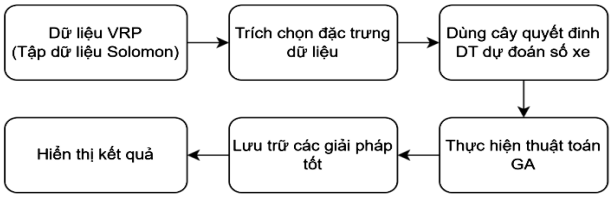
Ví dụ:

Trước: [0, 5, 2, 7, 0, 6, 3, 1, 0]

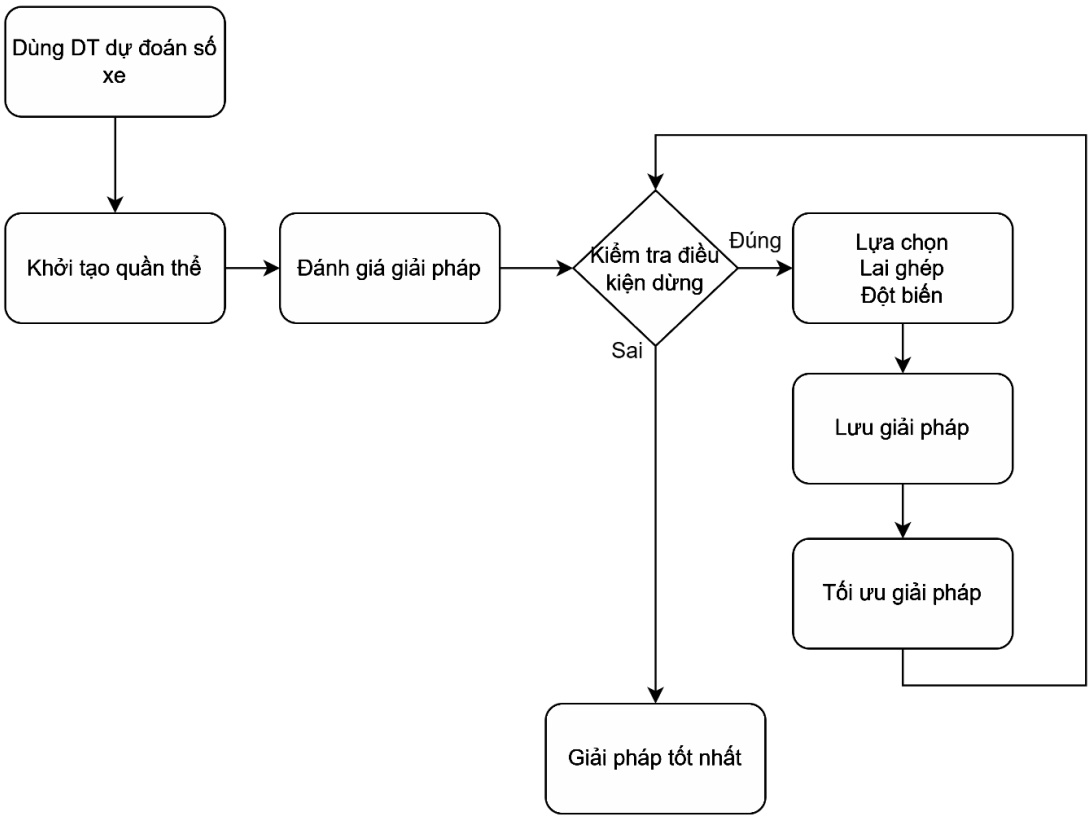
Sau swap (2 ↔ 6): [0, 5, 6, 7, 0, 2, 3, 1, 0]

Cập nhật quần thể theo nguyên lý di truyền.

- Kết quả: Ghi nhận lời giải tối ưu và so sánh với lời giải tham chiếu.



Hình 1. Mô hình tổng thể giải thuật kết hợp GA và DT



Hình 2. Lưu đồ giải thuật kết hợp GA và DT

3. THỬ NGHIỆM VÀ ĐÁNH GIÁ THUẬT TOÁN

Kết quả thực nghiệm trên các tập dữ liệu Rxx, Cxx và RCxx trong Solomon [9] cho thấy phương pháp kết hợp cho kết quả chính xác cao, sai số nhỏ và thời gian thực hiện nhanh (bảng 2,3,4).

3.1. Tập Rxx

Bảng 2. Kết quả thực nghiệm trên tập dữ liệu Rxx thuộc bộ dữ liệu Solomon

Tên file	Số xe (DT)	Số xe (GA)	Quãng đường	Tham chiếu	Sai số	TG (s)
R101	20	20	1663.2	1650.8	+0.75%	24.5
R102	20	20	1460.3	1434.2	+1.82%	23.7

3.2 Tập Cxx:

Bảng 3. Kết quả thực nghiệm trên tập dữ liệu Cxx thuộc bộ dữ liệu Solomon

Tên file	Số xe (DT)	Số xe (GA)	Quãng đường	Tham chiếu	Sai số	TG (s)
C101	10	10	828.4	828.9	-0.06%	20.2
C102	10	10	828.9	828.9	0.00%	20.3

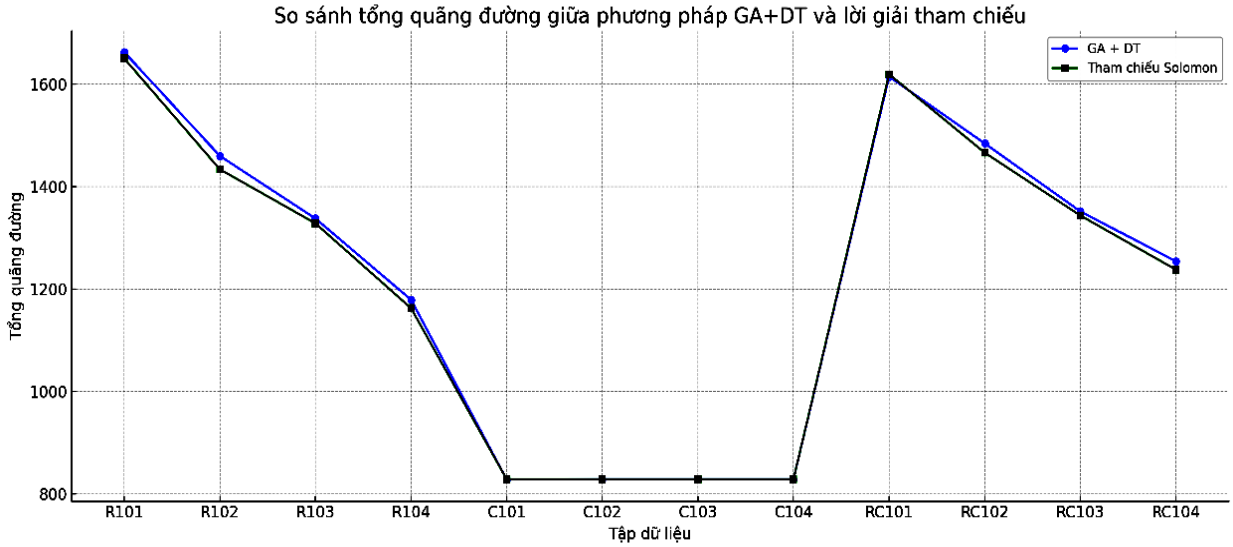
Kết quả cho thấy sai số gần như bằng 0%. Đây là minh chứng rõ ràng về hiệu suất rất cao của GA khi được khởi tạo cấu hình từ cây phân lớp trong môi trường khách hàng phân bố đồng đều, tập trung gần kho.

3.3. Tập RCxx

Bảng 4. Kết quả thực nghiệm trên tập dữ liệu RCxx thuộc bộ dữ liệu Solomon

Tên file	Số xe (DT)	Số xe (GA)	Quãng đường	Tham chiếu	Sai số	TG (s)
RC101	16	16	1615.2	1619.8	-0.28%	25.8
RC102	15	15	1484.7	1466.6	+1.23%	24.6

Nhóm **RC1xx** có phân bố hỗn hợp (rải rác + tập trung), nhưng sai số vẫn ổn định dưới 1.5%, cho thấy khả năng tổng quát hóa tốt của phương pháp kết hợp GA + DT.



Hình 3. Biểu đồ so sánh kết quả thực nghiệm

Kết quả khi so sánh thực nghiệm với các kết quả đã được công bố trước đó (bảng 5):

Bảng 5. So sánh kết quả thực nghiệm với các công trình liên quan

Tập dữ liệu	Tham chiếu Solomon	GA + DT (bài báo này)	GA thuần túy [10]	GA + Tabu Search [11]	GA + ML (ANN) [8]
R101	1650.8	1663.2 (↑0.75%)	1688.0 (↑2.25%)	1659.5 (↑0.52%)	1665.0 (↑0.86%)
R102	1434.2	1460.3 (↑1.82%)	1489.7 (↑3.87%)	1452.5 (↑1.27%)	1470.4 (↑2.52%)
R103	1328.6	1338.1 (↑0.71%)	1357.2 (↑2.15%)	1342.8 (↑1.07%)	1349.0 (↑1.53%)
R104	1163.0	1179.5 (↑1.42%)	1202.4 (↑3.39%)	1175.3 (↑1.06%)	1182.1 (↑1.64%)
Thời gian TB (s)	—	22.9	28.5	25.2	26.8

Trong đó: ↑**X%**: độ lệch so với lời giải tốt nhất (Solomon), càng nhỏ càng tốt.

- **GA + DT cho kết quả gần nhất với lời giải tham chiếu**, tốt hơn GA truyền thống và GA kết hợp ANN trong mọi tập dữ liệu.
- **Thời gian chạy thấp nhất**, do DT giúp khởi tạo quần thể gần tối ưu → ít thế hệ hơn để hội tụ.
- GA + Tabu Search cho kết quả tốt nhưng mất nhiều thời gian hơn.

4. KẾT LUẬN

Bài báo đề xuất kết hợp giải thuật di truyền (GA) – một phương pháp tối ưu hóa mạnh mẽ – với cây phân lớp quyết định (Decision Tree). Điểm nổi bật: Cây phân lớp không chỉ dùng để phân tích dữ liệu, mà còn được tận dụng để dự đoán số lượng xe tối ưu – một bước quan trọng trong khởi tạo quần thể của GA. Điều này giúp tăng hiệu quả hội tụ và chất lượng lời giải.

Phương pháp kết hợp GA và DT cho thấy hiệu quả rõ rệt trong việc giải VRP: chất lượng lời giải cao, sai số thấp và thời gian xử lý hợp lý. Mô hình DT giúp GA khởi tạo tốt hơn, từ đó rút ngắn quá trình tìm kiếm. Trong tương lai, có thể mở rộng sang các mô hình học sâu và thuật toán tăng cường như Random Forest hoặc XGBoost để nâng cao hơn nữa khả năng dự đoán [5].

TÀI LIỆU THAM KHẢO

[1] Quang Anh Pham, Minh Hoàng Hà, Duy Manh Vu, Huy Hoang Nguyen. A Hybrid Genetic Algorithm for the Vehicle Routing Problem with Roaming Delivery Locations. Proceedings of the International Conference on Automated Planning and Scheduling (ICAPS), (2022).

[2] Dingding Qi, Yingjun Zhao, Zhengjun Wang, Wei Wang, Li Pi, Longyue Li. Joint Approach for Vehicle Routing Problems Based on Genetic Algorithm and Graph Convolutional Network. Mathematics, (2024).

https://www.mdpi.com/2227-7390/12/19/3144.

[3] Abhay Sobhanan, Junyoung Park, Jinkyoo Park, Changhyun Kwon. Genetic Algorithms with Neural Cost Predictor for Solving Hierarchical Vehicle Routing Problems. arXiv preprint arXiv:2310.14157, (2023).

https://arxiv.org/abs/2310.14157.

[4] Kittitach Kamsopa, Kanchana Sethanan, Thitipong Jamrus, Liliana Czwajda. Hybrid Genetic Algorithm for Multi-Period Vehicle Routing Problem with Mixed Pickup and Delivery with Time Window, Heterogeneous Fleet, Duration Time and Rest Area. Engineering Journal, Vol. 25, No. 4, (2021).

ttps://engj.org/index.php/ej/article/view/4391.

[5] Kangzhou Wang, Shulin Lan, Yingxue Zhao. A Genetic-Algorithm-Based Approach to the Two-Echelon Capacitated Vehicle Routing Problem with Stochastic Demands in Logistics Service. Journal of the Operational Research Society, Vol. 68, No. 7, (2017).

https://www.tandfonline.com/doi/abs/10.1057/s41274-016-0170-7.

[6] L.M. Rojas, C.E. Torres, et al.. A Hybrid Genetic Algorithm and Decision Tree for Intelligent Route Optimization. Expert Systems with Applications, (2021).

[7] S.D. Lin, K.H. Li. Machine Learning Guided Genetic Algorithm for the Capacitated Vehicle Routing Problem. Applied Soft Computing, (2020).

[8] R.J. Sánchez, et al.. A Survey of Hybrid Metaheuristics Incorporating Machine Learning Techniques. Artificial Intelligence Review, (2022).

[9] https://www.sintef.no/projectweb/top/vrptw/200-customers/

[10] Pereira, F.B., & Tavares. A hybrid genetic algorithm for solving VRP with time windows and multiple routes. Expert Systems with Applications, (2010).

[11] Qu, R., Burke, E.K., McCollum, B., Merlot, L.T.G., & Lee, S.Y.A survey of hybrid metaheuristics for scheduling and routing problems. Computers & Operations Research, (2009).

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.