

ĐẠI HỌC QUỐC GIA HÀ NỘI
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ

Trần Văn Mạnh

**Kiến trúc dịch vụ
gợi ý chia sẻ phương tiện đi lại
dựa trên khai thác dữ liệu không gian - thời gian**

Chuyên ngành: Kỹ thuật phần mềm

Mã số: 9480103.01

TÓM TẮT LUẬN ÁN TIẾN SĨ
NGÀNH KỸ THUẬT PHẦN MỀM

Hà Nội, năm 2024

Công trình được hoàn thành tại: Trường Đại học Công nghệ,
Đại học Quốc gia Hà Nội.

Người hướng dẫn khoa học: 1. PGS.TS. Trương Anh Hoàng
2. TS. Vũ Thị Hồng Nhạn

Phản biện: 1.

Phản biện: 2.

Phản biện: 3.

Luận án sẽ được bảo vệ trước Hội đồng cấp Đại học Quốc gia
chấm luận án tiến sĩ họp tại Trường Đại học Công nghệ, ĐHQGHN.
vào hồi giờ ngày.....tháng..... năm.....

Có thể tìm hiểu luận án tại:

- Thư viện Quốc gia Việt Nam
- Trung tâm Thông tin - Thư viện, Đại học Quốc gia Hà Nội

CHƯƠNG 1. GIỚI THIỆU

1.1. Đặt vấn đề

Chương này giới thiệu về chia sẻ chuyến đi như một giải pháp giảm ùn tắc giao thông và chi phí. Các nền tảng chia sẻ chuyến đi hiện tại như Uber, Lyft, DiDi, Grab,... tuy phổ biến nhưng vẫn còn hạn chế về tối ưu lộ trình, thời gian chờ và khả năng ghép đôi. Bài toán chia sẻ chuyến đi phức tạp và đã được nghiên cứu nhiều, nhưng vẫn còn thiếu các khảo sát toàn diện về kiến trúc hệ thống..

1.2. Tổng quan về các dịch vụ chia sẻ phương tiện đi lại

Hệ thống chia sẻ phương tiện kết nối nhiều hành khách có chung lộ trình để dùng chung một phương tiện, nhằm giảm chi phí và ùn tắc giao thông. Các thành phần chính bao gồm hành khách, tài xế, nền tảng/ứng dụng và lộ trình.

1.3. Phân loại hệ thống chia sẻ phương tiện đi lại

Hệ thống chia sẻ phương tiện được phân loại dựa trên kiến trúc (tĩnh hoặc động), phân tích dữ liệu và động lực hệ thống. Hệ thống tĩnh có thông tin chuyến đi trước, trong khi hệ thống động xử lý yêu cầu theo thời gian thực. Phân loại dựa trên dữ liệu xem xét cách hệ thống sử dụng dữ liệu để cải thiện dịch vụ, và phân loại dựa trên động lực xem xét các yếu tố thúc đẩy hệ thống, như xã hội, giao thông hoặc hoạt động

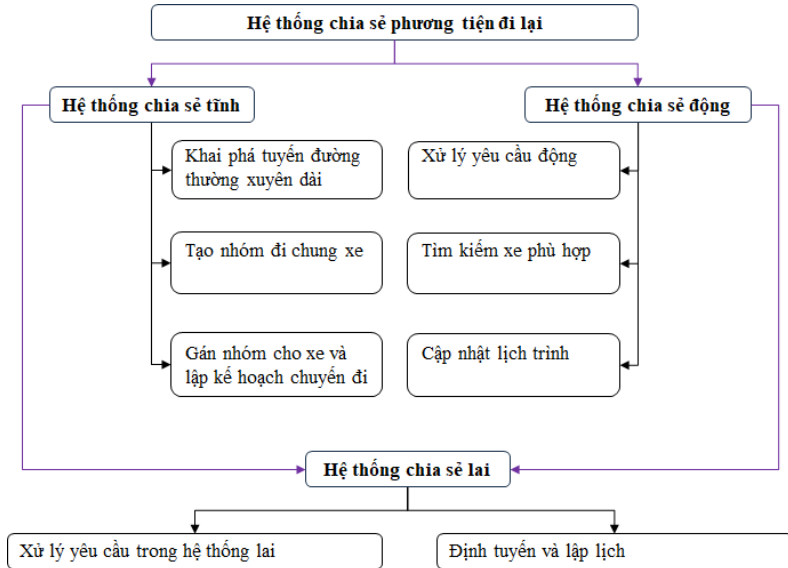
1.4. Những vấn đề còn tồn tại cần nghiên cứu

Mặc dù đã có nhiều nghiên cứu, vẫn còn nhiều vấn đề cần giải quyết để nâng cao hệ thống chia sẻ phương tiện, bao gồm: tối ưu ghép đôi, dự đoán nhu cầu, bảo mật, mô hình kinh doanh bền vững, ứng dụng công nghệ mới, giải quyết vấn đề pháp lý và xã hội, phát

triển hệ thống cho nhóm cụ thể, và đánh giá tác động môi trường và xã hội.

1.5. Cây nghiên cứu

Cây nghiên cứu của luận án được trình bày trong Hình 1.3, thể hiện các bước và mối quan hệ giữa các nội dung nghiên cứu chính.



Hình 1.3. Cây nghiên cứu hệ thống chia sẻ chuyến đi

Dựa vào cây nghiên cứu Hình 1.3, luận án phân loại ra những bài toán con chính cần giải quyết.

Bài toán 1: Khai phá tuyến đường thường xuyên dài. Sử dụng các kỹ thuật khai phá dữ liệu tuần tự để tìm kiếm các chuỗi điểm đến thường xuyên xuất hiện trong lịch sử di chuyển của người dùng, từ đó xác định các tuyến đường thường xuyên dài tối đa.

Bài toán 2: Tạo nhóm đi chung xe, trong bài toán này cần phát triển một thuật toán tạo nhóm dựa trên các tuyến đường thường xuyên dài

tối đa được xác định, đảm bảo các ràng buộc về thời gian và không gian, đồng thời tối ưu hóa số lượng hành khách trong mỗi nhóm.

Bài toán 3: Định tuyến và lập lịch tập trung vào việc xử lý các yêu cầu **đặt trước (tĩnh)**, trong bài toán này có thể chia nhỏ thành hai bài toán con. Bài toán con thứ nhất là áp dụng thuật toán tham lam để gán nhóm cho xe và lập kế hoạch chuyến đi. Bài toán con thứ hai là sau khi những người tham gia đi chung xe được gom nhóm lại với nhau, hệ thống xây dựng mô hình toán học và sử dụng bộ giải IBM CPLEX để tìm ra giải pháp tối ưu về định tuyến và lập lịch trình cho các xe, sao cho giảm thiểu tổng chi phí di chuyển và số lượng xe phục vụ.

Bài toán 4: Xử lý yêu cầu động trong hệ thống lại tập trung vào việc xử lý các yêu cầu **tức thời (động)**, tức là những yêu cầu được gửi đến hệ thống trong thời gian thực khi xe đang di chuyển. Trong bài toán này, hệ thống cần phát triển các thuật toán và cơ chế để tìm kiếm xe phù hợp, kiểm tra tính khả thi của việc ghép đôi và cập nhật lịch trình, nhằm đáp ứng các yêu cầu tức thời của hành khách một cách linh hoạt và hiệu quả.

1.6. Bố cục các chương trong luận án

Luận án được tổ chức thành 6 chương, mỗi chương tập trung vào một nội dung nghiên cứu cụ thể:

Chương 1 đặt ra bối cảnh về vấn đề tắc nghẽn giao thông, ô nhiễm môi trường và chi phí vận tải cao, chương này nhấn mạnh tiềm năng của chia sẻ phương tiện đi lại như một giải pháp hiệu quả. Mặc dù các dịch vụ chia sẻ xe hiện tại đã phổ biến, chúng vẫn còn nhiều hạn chế cần cải thiện. Luận án đề xuất phát triển một kiến trúc dịch vụ gợi ý chia sẻ phương tiện dựa trên khai phá dữ liệu không gian - thời gian, nhằm tối ưu hóa người dùng tham gia nhiều nhất có thể và nâng

cao hiệu quả hệ thống. Nghiên cứu tập trung vào hai mô hình chia sẻ xe: tĩnh (đặt trước) và động (thời gian thực), đồng thời giải quyết các vấn đề như tối ưu ghép đôi, dự đoán nhu cầu, bảo mật và mô hình kinh doanh. Phương pháp nghiên cứu bao gồm khảo sát, phân tích, mô hình hóa và thực nghiệm. Cây nghiên cứu của luận án bao gồm các bài toán con: khai phá tuyến đường thường xuyên dài, tạo nhóm đi chung, định tuyến và lập lịch, xử lý yêu cầu động. Những đóng góp chính của luận án là đề xuất các phương pháp và thuật toán mới, kiến trúc dịch vụ, và kiểm chứng thực nghiệm, nhằm cải thiện hiệu quả và chất lượng của hệ thống chia sẻ phương tiện. Luận án khẳng định tính mới và không trùng lặp của nghiên cứu.

Chương 2 Hệ thống giao thông công cộng hiện tại còn nhiều hạn chế, cần phát triển các dịch vụ chia sẻ phương tiện để giảm tác động tiêu cực đến môi trường và cung cấp dịch vụ tốt hơn. Chia sẻ phương tiện có hai loại hình: tĩnh với các thông số di chuyển cố định, và động với các thông số linh hoạt. Nghiên cứu của Wen He và cộng sự về hệ thống chia sẻ xe tĩnh được phân tích, tập trung vào khai thác tuyến đường thường xuyên từ dữ liệu GPS. Hệ thống này ổn định và có khả năng lập kế hoạch nhưng thiếu linh hoạt. Hai hệ thống chia sẻ xe động tiêu biểu là T-Share và mT-Share cũng được phân tích. T-Share sử dụng chỉ mục không gian-thời gian và thuật toán để ghép đôi, tối ưu hóa quãng đường. mT-Share cải tiến T-Share bằng cách phục vụ cả hành khách trực tuyến và ngoại tuyến thông qua định tuyến xác suất và lập chỉ mục. Chương này cũng thảo luận về cách xử lý khách hàng ngoại tuyến của mT-Share, đánh giá ưu nhược điểm của từng mô hình và đề xuất một hệ thống lai kết hợp giữa chia sẻ xe tĩnh và động, đồng thời chỉ ra các thách thức trong việc phát triển hệ thống này.

Chương 3 chương này tập trung vào việc cải tiến phương pháp khai thác các mẫu tuyến đường thường xuyên dài (LSP) từ dữ liệu không gian - thời gian, phục vụ cho việc tạo nhóm đi chung xe. Tác giả bắt đầu bằng cách xem xét các nghiên cứu trước đây về khai thác mẫu tuần tự và khai thác LSP từ dữ liệu quỹ đạo, chỉ ra những hạn chế của chúng như bỏ sót các mẫu tiềm năng, không xem xét độ dài tuyến đường hoặc chưa đảm bảo tính chia sẻ cao. Tiếp theo, chương này đề xuất một phép biến đổi mới để chuyển đổi dữ liệu quỹ đạo thô thành các giao dịch không gian - thời gian, tạo điều kiện cho việc áp dụng các kỹ thuật khai thác tập mục vào bài toán khai phá LSP. Phép biến đổi này bao gồm ba bước: xác định các chuyến đi, chiếu chiều thời gian và tổng quát hóa và thay thế không gian-thời gian. Dựa trên phép biến đổi này, tác giả định nghĩa bài toán khai thác các mẫu dài, có thể chia sẻ (LSP) và trình bày thuật toán MineLSP, một thuật toán tiêu biểu cho phương pháp khai phá dựa trên chiếu cơ sở dữ liệu. Sau đó, chương giới thiệu phương pháp HNMF, một cải tiến của thuật toán MineLSP, tập trung vào việc tìm kiếm các tập mục thường xuyên cực đại (MFI) thay vì các tập mục đóng thường xuyên (MFCI). Sự thay đổi này giúp giảm sự dư thừa trong kết quả và tăng khả năng phát hiện các mẫu tuyến đường có tiềm năng chia sẻ cao hơn, từ đó tối ưu hóa việc ghép đôi người dùng trong các ứng dụng chia sẻ chuyến đi. Cuối cùng, chương trình bày kết quả thực nghiệm trên cả dữ liệu thực tế và dữ liệu tổng hợp quy mô lớn, chứng minh tính hiệu quả và khả năng ứng dụng của phương pháp HNMF trong việc giải quyết bài toán khai phá tuyến đường thường xuyên dài và có khả năng chia sẻ cao. Kết quả của việc tìm ra các MFI sẽ đóng vai trò là đầu vào quan trọng cho chương tiếp theo, nơi một thuật toán ghép nhóm hiệu quả những người dùng có cùng MFI sẽ được phát triển.

Chương 4 trình bày một phương pháp mới có tên là aVC-growth để giải quyết bài toán tạo nhóm đi chung xe, nhằm tối ưu hóa dịch vụ chia sẻ chuyến đi. Phương pháp này tập trung vào việc khai thác các chuyển động thường xuyên dài nhất của người dùng, từ đó xây dựng các nhóm chia sẻ xe dựa trên sự tương đồng về hành vi di chuyển. Điểm nổi bật của aVC-growth là việc cải tiến thuật toán VC-growth hiện có, cho phép tăng tốc độ tạo nhóm và giảm thiểu tổng quãng đường di chuyển. Cụ thể, aVC-growth sử dụng ba phương pháp là HNMG.1, HNMG.2 và HNMG.3 để thực hiện cải tiến nhằm tạo ra nhóm hợp lệ tối đa và mỗi thành viên trong nhóm không được xuất hiện hơn một lần. Nghiên cứu cũng đã thực hiện các thử nghiệm trên tập dữ liệu tổng hợp để đánh giá hiệu quả của aVC-growth so với các phương pháp hiện có (AllVC và VC-growth). Kết quả cho thấy aVC-growth vượt trội hơn về thời gian thực hiện và khả năng giảm thiểu tổng quãng đường di chuyển, đồng thời tạo ra số lượng nhóm chia sẻ xe phù hợp hơn cho dịch vụ chia sẻ chuyến đi, trong đó mỗi người dùng chỉ thuộc về duy nhất một nhóm. Ngoài ra, chương này còn trình bày chi tiết về quá trình gán phương tiện cho nhóm và lập kế hoạch chuyến đi, bao gồm thuật toán ARV để gán nhóm người dùng cho các phương tiện khả dụng và thuật toán FindOptimalRoute để tìm lộ trình tối ưu cho mỗi phương tiện, đảm bảo đón và trả tất cả hành khách trong nhóm đã được gán. Cuối cùng, chương đã thảo luận về một số kết quả thử nghiệm, đánh giá hiệu suất của thuật toán aVC-growth trong việc tạo nhóm và ảnh hưởng của các tham số đến kết quả. Đồng thời, việc triển khai một nguyên mẫu hệ thống và phân tích hiệu suất trên tập dữ liệu thực tế cũng được đề cập, cho thấy tiềm năng ứng dụng của phương pháp đề xuất trong thực tế.

Chương 5 tập trung vào việc giải quyết vấn đề định tuyến và lập lịch dựa trên bài toán VRPTW trên biểu đồ mạng lưới đường bộ. Một phương pháp mới có tên là HNM3 được đề xuất, sử dụng các thuật toán tối ưu hóa để tìm ra giải pháp tối ưu cho bài toán, nhằm giảm thiểu chi phí vận chuyển và số lượng xe sử dụng trong dịch vụ chia sẻ chuyến đi. Tác giả trình bày một số nghiên cứu liên quan đến việc giải quyết bài toán VRPTW bằng các mô hình toán học và thuật toán khác nhau. Tuy nhiên, hầu hết các nghiên cứu trước đây đều tập trung vào việc phát triển các phương pháp heuristic hoặc chỉ xem xét các trường hợp đơn giản. Để giải quyết những hạn chế này, chương này đề xuất phương pháp HNM3, sử dụng mô hình quy hoạch tuyến tính nguyên hỗn hợp (MILP) và bộ giải IBM CPLEX. Điểm đặc biệt của HNM3 là nó cho phép lựa chọn giữa hai thuật toán tối ưu: nhánh và cắt, và tìm kiếm động. Các thử nghiệm được tiến hành trên hai bộ dữ liệu tiêu chuẩn của Solomon (C101 và R101). Kết quả thực nghiệm cho thấy HNM3 vượt trội hơn phương pháp trước đó (HCHNM3) về thời gian tính toán và khả năng tìm ra giải pháp tối ưu, đặc biệt khi dữ liệu được phân cụm và số lượng khách hàng tăng lên. Ngoài ra, chương này còn trình bày thuật toán ARV để gán nhóm cho phương tiện và tìm lộ trình tối ưu, cũng như thuật toán HybridRideMatching để xử lý yêu cầu động trong hệ thống chia sẻ xe lại. Các kết quả thực nghiệm cho thấy tính hiệu quả của các phương pháp và thuật toán đề xuất. Tuy nhiên, vẫn còn một số hạn chế cần được giải quyết trong tương lai, bao gồm việc mở rộng mô hình để xem xét các yếu tố đa dạng hơn và cải tiến thuật toán để tăng tính linh hoạt và hiệu quả.

Chương 6 tổng kết những đóng góp chính của luận án trong việc xây dựng và tối ưu hóa dịch vụ gợi ý chia sẻ phương tiện đi lại dựa trên

khai phá dữ liệu không gian - thời gian. Nghiên cứu đã giải quyết thành công các bài toán chia sẻ xe tĩnh và động, đồng thời đề xuất các phương pháp lai kết hợp cả hai mô hình. Cụ thể, luận án đã đạt được các kết quả sau:

- ◆ Đối với bài toán chia sẻ xe tĩnh, đề xuất phương pháp khai thác dữ liệu thường xuyên dài nhất và thuật toán aVC-growth để tạo nhóm đi chung, mang lại hiệu quả vượt trội.

- ◆ Đối với bài toán chia sẻ xe động, đề xuất phương pháp HNM3 để giải quyết bài toán định tuyến phương tiện với cửa sổ thời gian (VRPTW), giảm thiểu chi phí và số lượng xe. Đồng thời, đề xuất thuật toán HybridRideMatching để ghép đôi yêu cầu mới với xe đang hoạt động.

- ◆ Chứng minh tính hiệu quả của các phương pháp và thuật toán đề xuất thông qua thực nghiệm trên dữ liệu thực tế và tổng hợp.

Mặc dù đạt được những kết quả khả quan, luận án vẫn còn một số hạn chế cần được khắc phục trong tương lai, bao gồm tính thực tiễn, khả năng mở rộng và bảo mật. Cuối cùng, chương này đề xuất các hướng nghiên cứu tiếp theo, như mở rộng mô hình, cải tiến thuật toán, ứng dụng công nghệ mới và giải quyết các vấn đề pháp lý và xã hội, nhằm tiếp tục phát triển và hoàn thiện hệ thống chia sẻ phương tiện đi lại.

1.7. Những đóng góp của luận án

Luận án này có những đóng góp chính sau đây cho lĩnh vực chia sẻ phương tiện đi lại:

- ◆ **Đề xuất một phương pháp khai phá mẫu thường xuyên dài nhất:** Phương pháp này giúp xác định các tuyến đường thường xuyên dài nhất từ dữ liệu lịch sử di chuyển của người dùng, từ đó tối đa hóa số lượng người tham gia đi chung xe và nâng cao hiệu quả của hệ thống chia sẻ xe.

♦ **Đề xuất một phương pháp tạo nhóm đi chung xe hợp lệ tối đa:** Phương pháp này đảm bảo các ràng buộc về thời gian và không gian, đồng thời tối ưu hóa số lượng hành khách trong mỗi nhóm, giúp tăng hiệu quả sử dụng xe và giảm chi phí vận hành.

♦ **Đề xuất phương pháp gán nhóm vào xe:** Phương pháp tham lam này giúp phân công các nhóm hành khách vào các xe một cách hiệu quả, sao cho càng nhiều nhóm được phục vụ càng tốt.

♦ **Đề xuất một phương pháp định tuyến và lập lịch:** Phương pháp này dựa trên giải bài toán quy hoạch nguyên hỗn hợp sử dụng bộ giải IBM CPLEX, nhằm tối ưu hóa tổng chi phí di chuyển và thời gian chờ đợi của hành khách.

♦ **Đề xuất một mô hình đi chung xe lai:** Mô hình này kết hợp giữa phương pháp đi chung xe tĩnh và đi chung xe động, tận dụng ưu điểm của cả hai mô hình để tạo ra một hệ thống chia sẻ xe hiệu quả và linh hoạt hơn.

Các đóng góp này có ý nghĩa khoa học và thực tiễn quan trọng, giúp cải thiện hiệu quả hoạt động của hệ thống chia sẻ xe, nâng cao trải nghiệm người dùng và góp phần giảm thiểu tắc nghẽn giao thông và ô nhiễm môi trường.

CHƯƠNG 2. KHẢO SÁT BÀI TOÁN CHIA SẺ PHƯƠNG TIỆN ĐI LẠI

2.1. Giới thiệu

Hệ thống giao thông công cộng tuy là giải pháp tốt nhưng còn nhiều hạn chế. Do đó, cần phát triển các dịch vụ chia sẻ phương tiện đi lại mới, giúp giảm tác động tiêu cực đến môi trường và cung cấp dịch vụ tốt hơn cho người dùng. Chia sẻ phương tiện đi lại là một công nghệ mới, trong đó người dùng có cùng lộ trình sẽ chia sẻ chuyến đi. Có hai loại hình chia sẻ xe: tĩnh và động.

2.2. Hệ thống dịch vụ chia sẻ xe tĩnh

Phân tích nghiên cứu của Wen He và cộng sự (2012, 2014) về hệ thống chia sẻ xe tĩnh, tập trung vào khai thác tuyến đường thường xuyên từ dữ liệu GPS lịch sử để gợi ý chia sẻ chuyến đi. Hệ thống này có ưu điểm về ổn định và khả năng lập kế hoạch, nhưng hạn chế về tính linh hoạt trong việc đáp ứng yêu cầu đặt xe theo thời gian thực.

2.3. Hệ thống dịch vụ chia sẻ động

Phân tích hai hệ thống chia sẻ xe động tiêu biểu: T-Share(2013) và mT-Share (2022). T-Share sử dụng chỉ mục không gian-thời gian và thuật toán để ghép đôi, tối ưu hóa quãng đường và tiết kiệm nhiên liệu. mT-Share cải tiến T-Share bằng cách phục vụ cả hành khách trực tuyến và ngoại tuyến, thông qua định tuyến xác suất và lập chỉ mục theo vị trí và hướng di chuyển.

CHƯƠNG 3. PHƯƠNG PHÁP KHAI PHÁ CÁC MẪU Tuyến Đường Thường Xuyên Dài

3.1. Giới thiệu

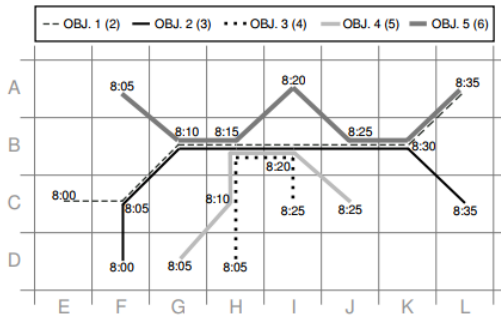
Chương này giới thiệu về phương pháp khai phá các mẫu tuyến đường thường xuyên dài (LSP - Long, Sharable Patterns) từ dữ liệu không gian - thời gian. LSP là các tuyến đường đủ dài và được nhiều người dùng sử dụng, có tiềm năng cao cho việc chia sẻ xe.

3.2. Các mẫu dài, có thể chia sẻ trong quỹ đạo di chuyển

Các mẫu dài, có thể chia sẻ trong quỹ đạo gọi tắt là LSP. Phần tiếp theo mô tả một phép biến đổi mới cho các quỹ đạo thô. Phép biến đổi này cho phép (1) xây dựng bài toán khai thác các LSP trong quỹ đạo di chuyển trong một khuôn khổ tương tự như khuôn khổ được sử dụng khai thác các tập mục thường xuyên và (2) thiết lập một mối quan hệ giữa hai bài toán này.

3.3. Khai phá các mẫu dài chia sẻ dựa trên phép chiếu

Trình bày một thuật toán khai phá mẫu dài có thể chia sẻ (MineLSP), sử dụng kỹ thuật chiếu cơ sở dữ liệu để khai thác LSP. Giả sử cơ sở dữ liệu (CSDL) mẫu về quỹ đạo cho dưới hình sau:



Hình 3.1. Minh họa CSDL mẫu về quỹ đạo.

Lúc này CSDL mẫu quỹ đạo hình 3.1 được biểu diễn dưới dạng lược đồ CSDL quan hệ $T(\text{oid}, \text{tid}, \text{item})$. Trong đó oid: id của tối tượng/người tham gia chuyển đi, tid: id của chuyến đi, và item: một mục đơn lẻ và là một phần của giao dịch (transaction) tid được liên kết với đối tượng oid. .

Bảng 3.1. Lược đồ CSDL quan hệ của CSDL quỹ đạo hình 3.1

CSDL $T(\text{oid}, \text{tid}, \text{item})$			CSDL $T(\text{oid}, \text{tid}, \text{item})$			CSDL $T(\text{oid}, \text{tid}, \text{item})$		
oid	tid	item	oid	tid	item	oid	tid	item
o1	t1	EC8:00	o2	t2	FD8:00	o5	t5	FA8:05
o1	t1	FC8:05	o2	t2	FC8:05	o5	t5	GB8:10
o1	t1	GB8:10	o2	t2	GB8:10	o5	t5	HB8:15
o1	t1	HB8:15	o2	t2	HB8:15	o5	t5	IA8:20
o1	t1	IB8:20	o2	t2	IB8:20	o5	t5	JB8:25
o1	t1	JB8:25	o2	t2	JB8:25	o5	t5	KB8:30
o1	t1	KB8:30	o2	t2	KB8:30	o5	t5	LA8:35

o1	t1	LA8:35	o2	t2	LC8:35
o3	t3	HD8:05	o4	t4	GD8:05
o3	t3	HC8:10	o4	t4	HC8:10
o3	t3	HB8:15	o4	t4	HB8:15
o3	t3	IB8:20	o4	t4	IB8:20
o3	t3	IC8:25	o4	t4	JC8:25

Để cho gọn và không mất tính tổng quát gọi Lược đồ CSDL T $\langle oid, tid, items \rangle$, có nghĩa là items là một tập các item.

Bảng 3.2. Lược đồ CSDL giao dịch chuyển đi T $\langle oid, tid, items \rangle$

CSDL T($oid, tid, items$)		
oid	tid	items
o1	t1	{EC8:00, FC8:05, GB8:10, HB8:15, IB8:20, JB8:25, KB8:30, LA8:35}
o2	t2	{FD8:00, FC8:05, GB8:10, HB8:15, IB8:20, JB8:25, KB8:30, LC8:35}
o3	t3	{HD8:05, HC8:10, HB8:15, IB8:20, IC8:25}
o4	t4	{GD8:05, HC8:10, HB8:15, IB8:20, JC8:25}
o5	t5	{FA8:05, GB8:10, HB8:15, IA8:20, JB8:25, KB8:30, LA8:35}

Áp dụng các bước trong phương pháp MineLSP, với tham số cấu hình $MinLength=4$ (Ngưỡng độ dài tối thiểu), $MinSupp = 2$ (Ngưỡng độ hỗ trợ tối thiểu) và $n = 2$ (Đại diện cho số lượng đối tượng riêng biệt tối thiểu). Chúng ta thu được kết quả sau:

Bảng 3.3. Các mẫu được khai phá trong CSDL bảng 3.1

Mẫu	Đối tượng	Hỗ trợ	Độ dài	SMFCI
PAT1	o1, o2	2	6	{FC8:05, GB8:10, HB8:15, IB8:20, JB8:25, KB8:30}

PAT2	o1, o5	2	5	{GB8:10, HB8:15, JB8:25, KB8:30, LA8:35}
PAT3	o1, o2, o5	3	4	{GB8:10, HB8:15, JB8:25, KB8:30}

Tức là chúng ta thu được ba mẫu thường xuyên dài, có thể chia sẻ: PAT1: {FC8:05, GB8:10, HB8:15, IB8:20, JB8:25, KB8:30}, PAT2:{GB8:10, HB8:15, JB8:25, KB8:30, LA8:35} và PAT3: {GB8:10, HB8:15, JB8:25, KB8:30}.

3.4. Đề xuất phương pháp HNMF

Đề xuất một cải tiến quan trọng có tên là HNMF. Nội dung phương pháp HNMF tạo một thuật toán MineMLSP để tìm các tập mục thường xuyên cực đại (MFI), thay vì tìm kiếm SMFCI như trong phương pháp trước đó. Điều này giúp giảm sự dư thừa trong kết quả và tăng khả năng phát hiện các mẫu có tiềm năng chia sẻ cao hơn.

Quay trở lại với ví dụ CSDL được mô tả ở Hình 3.1, với MinLength = 4, MinSupp = 2, và n = 2. Các bước thực hiện trong thuật toán MineMLSP kết thúc sẽ cho kết quả như sau:

Bảng 3.4. Các mẫu được khai phá trong CSDL bảng 3.1, sử dụng thuật toán MineMLSP

Mẫu	Đôi tượng	Hỗ trợ	Độ dài	MFI
PAT1	o1, o2	2	6	{FC8:05, GB8:10, HB8:15, IB8:20, JB8:25, KB8:30}
PAT2	o1, o5	2	5	{GB8:10, HB8:15, JB8:25, KB8:30, LA8:35}

Dựa trên kết quả bảng 3.4, tiêu chí MFI đã loại bỏ mẫu PAT3: {GB8:10, HB8:15, JB8:25, KB8:30} vì nó là tập con của mẫu dài hơn PAT2: {GB8:10, HB8:15, JB8:25, KB8:30, LA8:35}.

3.5. Thực nghiệm và đánh giá

Trình bày kết quả thực nghiệm trên dữ liệu thực tế và tổng hợp, so sánh hiệu quả của phương pháp HNMF với các phương pháp khác. Kết quả cho thấy HNMF vượt trội trong việc khai thác LSP dài và có khả năng chia sẻ cao. Kết quả này sẽ là đầu vào quan trọng cho việc phát triển thuật toán ghép nhóm hiệu quả ở chương tiếp theo

CHƯƠNG 4. PHƯƠNG PHÁP TẠO NHÓM ĐI CHUNG

4.1. Giới thiệu

Chương này tập trung vào việc phát triển một phương pháp mới để tạo nhóm đi chung xe, nhằm tối ưu hóa dịch vụ chia sẻ chuyến đi. Phương pháp này được xây dựng dựa trên việc khai thác các chuyển động thường xuyên dài nhất của người dùng, từ đó xây dựng các nhóm chia sẻ xe dựa trên sự tương đồng về hành vi di chuyển.

4.2. Phát biểu bài toán

Phần này định nghĩa bài toán tìm tất cả các mẫu nhóm hợp lệ (khai phá nhóm hợp lệ) dựa trên cơ sở dữ liệu di chuyển của người dùng. Một mẫu nhóm được coi là hợp lệ nếu nhóm người dùng tương ứng di chuyển gần nhau trong một khoảng thời gian đủ dài và tần suất di chuyển chung đạt đến một ngưỡng nhất định.

4.3. Phương pháp lập nhóm đi chung xe

Một phương pháp mới có tên là aVC-growth được đề xuất để giải quyết bài toán tạo nhóm đi chung xe. Phương pháp này cải tiến thuật toán VC-growth (Yida Wang và cộng sự, 2003, 2006) hiện có bằng cách:

Cải tiến HNMG.1: Không thực hiện duyệt lại đỉnh đã được duyệt, giúp tăng tốc độ tạo nhóm.

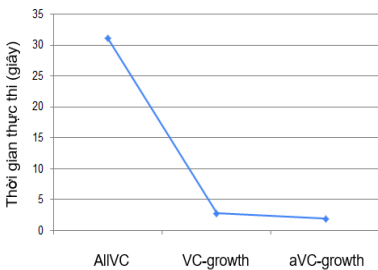
Cải tiến HNMG.2: Khai thác các nhóm hợp lệ tối đa, loại bỏ các nhóm con không cần thiết.

Cải tiến HNMG.3: Loại các nhóm hợp lệ tối đa để đảm bảo mỗi người dùng chỉ thuộc về một nhóm duy nhất.

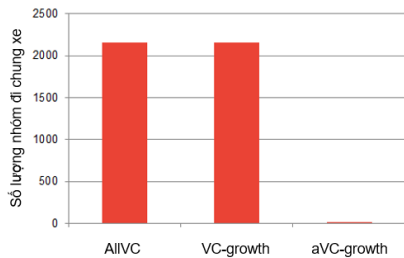
4.4. Kết quả thực nghiệm tạo nhóm đi chung

Thực nghiệm được thực hiện trên tập dữ liệu tổng hợp để đánh giá hiệu suất của aVC-growth so với các phương pháp hiện có. Kết quả cho thấy aVC-growth vượt trội hơn về thời gian thực hiện và khả năng giảm thiểu tổng quãng đường di chuyển, đồng thời tạo ra số lượng nhóm chia sẻ xe phù hợp hơn.

Kết quả thực hiện do nhóm nghiên cứu ridesharing thực hiện với dữ liệu: tập dữ liệu D1000_L12, chứa 1000 chuyến đi, mỗi chuyến tối đa 12 điểm đến. Mỗi xe chở tối đa 4 người, thực nghiệm tạo tối đa ba chuyến đi tuân theo mỗi mẫu di chuyển. Thực hiện so sánh ba thuật toán aVC-growth, AllVC, và VC-growth với các tham số: $\text{max_dis} = 900$ (ngưỡng khoảng cách tối đa), $\text{min_dur} = 3$ (ngưỡng khoảng thời gian tối thiểu), $\text{minWei} = 0.4$ (ngưỡng trọng số mẫu của mỗi nhóm).



a) Thời gian thực hiện thuật toán

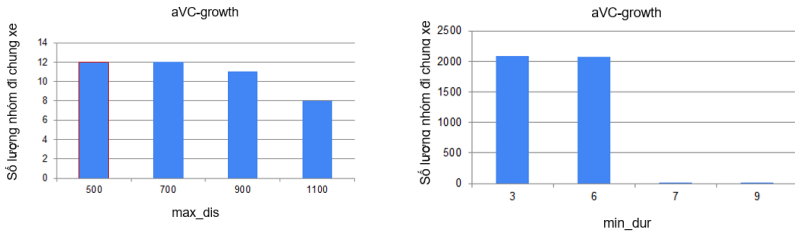


b) Số lượng nhóm đi chung xe được tạo

Hình 4.1. So sánh các thuật toán tạo nhóm đi chung xe.

Kết quả: Thời gian thực hiện: aVC-growth nhanh hơn đáng kể so với AllVC và VC-growth. Số lượng nhóm, rõ ràng là aVC-growth tạo ra ít nhóm hơn.

Tiếp theo thực nghiệm sẽ so sánh ảnh hưởng của tham số min_dur và max_dis :



a) Ảnh hưởng của max_dis và $min_dur=3$

b) Ảnh hưởng của min_dur và $max_dis=500$

Hình 4.2. Ảnh hưởng của tham số min_dur và max_dis đến kết quả.

Hình 4.2 (a): Khi max_dis nhỏ, một nhóm có thể bị chia thành các nhóm nhỏ hơn. Hình 4.3 (b): Khi tăng max_dis , số lượng nhóm giảm đi. Điều này là do nhiều người dùng có thể thỏa mãn điều kiện về khoảng cách, cho phép họ tham gia vào cùng một nhóm.

CHƯƠNG 5. PHƯƠNG PHÁP ĐỊNH TUYẾN VÀ LẬP LỊCH

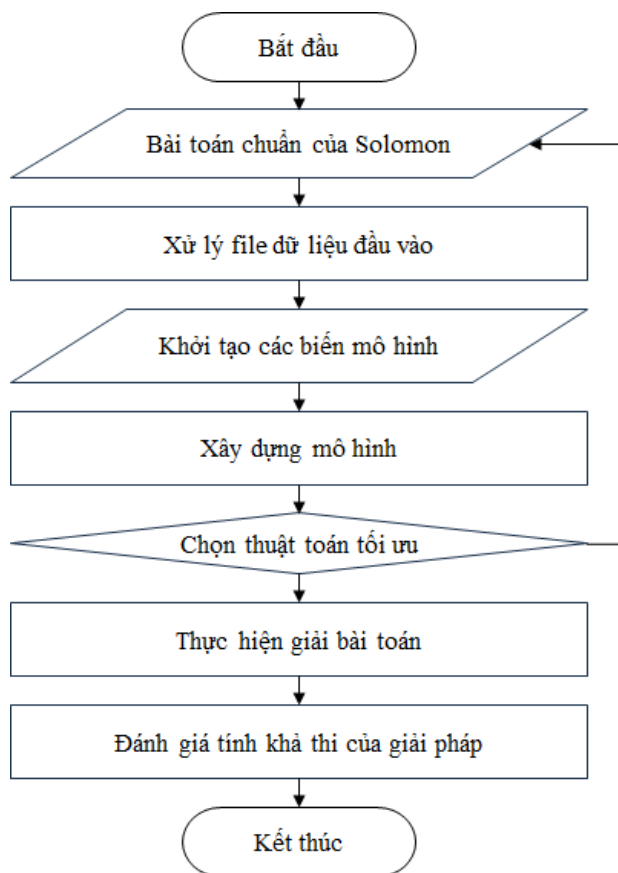
5.1. Giới thiệu

Tập trung giải quyết vấn đề định tuyến và lập lịch trình cho các phương tiện trong hệ thống chia sẻ xe, đặc biệt là trong bối cảnh chia sẻ xe tĩnh, nơi người dùng đặt xe trước và hệ thống cần tìm ra lộ trình tối ưu để phục vụ các nhóm hành khách đã được ghép đôi từ trước.

5.2. Phương pháp HNM3

Để giải quyết những hạn chế của các nghiên cứu trước đó, chương này đề xuất một phương pháp mới có tên là HNM3. Phương pháp này sử dụng mô hình quy hoạch tuyến tính nguyên hỗn hợp (MILP) để biểu diễn bài toán VRPTW, đồng thời sử dụng bộ giải IBM CPLEX để tìm ra giải pháp tối ưu. Điểm đặc biệt của HNM3 là nó cho phép lựa chọn giữa hai thuật toán tối ưu: nhánh và cắt, và tìm kiếm động, nhằm mục tiêu tối ưu hóa đồng thời chi phí vận chuyển

và số lượng phương tiện cần thiết. Dưới đây là lưu đồ thuật toán của phương pháp HNM3.



Hình 5.1. Lưu đồ thuật toán của phương pháp HNM3

5.3. Thực nghiệm

Thực nghiệm phương pháp HNM3 sử dụng bộ giải IBM CPLEX phiên bản 12.9.0, sử dụng ngôn ngữ lập trình Java. Các tham số cấu hình của hệ thống, hệ điều hành Windows 10, bộ xử lý Intel(R) Core (TM) i5-10210U CPU @ 1.60GHz 2.11 GHz, RAM 12GB. Thực nghiệm phương pháp HCHNM3 với bộ giải IBM CPLEX không chỉ rõ dùng phiên bản nào, thực hiện trên bộ Xử lý Core i7 với RAM 1GB. HCHNM3 chỉ chạy với thuật toán mặc định mà không chỉ rõ thuật toán tối ưu nào được sử dụng. Hai phương pháp HNM3 và HCHNM3 được so sánh và đánh giá trên hai bộ dữ liệu tiêu biểu của Solomon (1987). Luận án thực hiện các thử nghiệm trên hai bộ dữ liệu tiêu chuẩn của Solomon (C101 và R101) để đánh giá hiệu quả của phương pháp HNM3. Thực nghiệm so sánh HNM3 với nghiên cứu của Divya Aggarwal và cộng sự (2019) gặp một hạn chế khi sử dụng bộ giải IBM Cplex. Luận án đặt tên hạn chế này là HCHNM3. Thực nghiệm trả lời cho các câu hỏi nghiên cứu sau đây:

- ◆ RQ1 – Hiệu năng hai thuật toán của phương pháp HNM3: Đánh giá hiệu năng của thuật toán nhánh và cắt và thuật toán tìm kiếm động của phương pháp HNM3.
- ◆ RQ2 – Chi phí hàm mục tiêu và số lượng xe tối thiểu: Liệu chi phí hàm mục tiêu và số lượng xe tối thiểu của phương pháp HNM3 tốt hơn phương pháp HCHNM3 không?
- ◆ RQ3 – Hiệu năng của hai phương pháp: Đánh giá hiệu năng của phương pháp HNM3 và HCHNM3?

Với RQ1, thực nghiệm muốn so sánh hiệu năng của hai thuật toán nhánh và cắt và thuật toán tìm kiếm động của phương pháp HNM3 từ đó đánh giá xem thuật toán nào cho hiệu năng tốt hơn.

Bảng 5.1. So sánh hiệu năng phương pháp HNM3

STT	Số KH	Bộ dữ liệu C101		Bộ dữ liệu R101	
		Thời gian CPU (giây)		Thời gian CPU (giây)	
		Nhánh và cắt	Tìm kiếm động	Nhánh và cắt	Tìm kiếm động
1	5	0.43	0.21	0.18	0.22
2	10	0.15	0.15	0.16	0.16
3	15	0.33	0.32	0.23	0.24
4	20	0.66	0.64	0.44	0.45
5	25	0.91	0.88	0.59	0.59
6	30	1.66	1.66	1.10	1.08
7	33	4.20	4.38	4.21	4.39
8	40	5.02	4.88	23.29	6.11
9	50	18.46	10.71	10.57	147.73
10	60	21.48	19.82	26.39	104.96
11	70	77.85	54.63	305.71	848.36

Với bộ dữ liệu C101 thì thuật toán tìm kiếm động cho hiệu năng tốt hơn 8/11 trường hợp, ngoại trừ trường hợp số khách hàng là 33 thì thuật tìm kiếm động chạy kém hơn 4.20% so với nhánh và cắt. Với bộ dữ liệu R101 thì thuật toán nhánh và cắt cho kết quả chạy tốt hơn 7/11 trường hợp và có hiệu năng như nhau trong 1 trường hợp số khách hàng bằng 10.

Tiếp theo RQ2 tập trung so sánh phân tích xem chi phí hàm mục tiêu và số lượng xe của thuật toán nào tốt hơn trong HNM3.

Bảng 5.2. So sánh hàm mục tiêu và số lượng xe tối thiểu của hai phương pháp HNM3 và HCHNM3.

Bộ dữ liệu C101							
ST T	Số K H	Hàm mục tiêu			Số lượng xe		
		HNM3 với tìm kiếm động (km)	HCHN M3 (km)	Độ lệch (%)	HNM 3	HCHNM 3	Độ lệch (%)
1	5	42.3	42.42	0.28%	1	1	0%
2	10	58.1	58.34	0.41%	1	1	0%
3	15	141.6	142.17	0.40%	2	2	0%
4	20	175.0	175.39	0.22%	3	3	0%
5	25	191.3	191.83	0.28%	3	3	0%
9	50	362.4	363.28	0.24%	5	5	0%
10	60	469.0	470.13	0.24%	6	6	0%
Bộ dữ liệu R101							
ST T	Số K H	Hàm mục tiêu			Số lượng xe		
		HNM 3 với nhánh và cắt (km)	HCHNM 3 (km)	Độ lệch (%)	HNM 3	HCHNM 3	Độ lệch (%)
1	5	156.2	156.35	0.10%	2	2	0%
2	10	269.2	269.54	0.13%	4	4	0%
3	15	383.1	383.83	0.19%	5	3	50%
4	20	510.3	511.26	0.19%	7	6	15.38 %
5	25	617.1	618.34	0.20%	8	6	28.57 %
7	33	714.8	713.58	0.17%	10	11	9.52%

Hàm mục tiêu trong trường hợp này là chi phí về khoảng cách. Với bộ dữ liệu C101 thì HNM3 hầu hết cho kết quả khoảng cách ngắn hơn HCHNM3. Hai phương pháp đều cho kết quả số lượng xe tối thiểu như nhau trong mọi trường hợp. Với bộ dữ liệu R101 thì HNM3 hầu hết cho kết quả khoảng cách ngắn hơn HCHNM3. Khi so sánh về số lượng xe thì HCHNM3 cho kết quả số lượng xe ít hơn HNM3, ngoại trừ trường hợp số khách hàng là 33 thì HNM3 tốt hơn HCHNM3 là 9.52%. Tuy nhiên có vẻ khi số lượng khách hàng từ 33 trở lên thì HNM3 sẽ tốt hơn HCHNM3 vì lúc đó thuật toán tìm kiếm động của HNM3 sẽ phát huy tác dụng.

Mục đích của RQ3 để so sánh hiệu năng, cụ thể là thời gian CPU chạy giữa hai phương pháp HNM3 và HCHNM3.

Bảng 5.3. So sánh hiệu năng hai phương pháp HNM3 và HCHNM3

Bộ dữ liệu C101				
STT	Số KH	Thời gian CPU (giây)		
		HNM3 với tìm kiếm động (giây)	HCHNM3 (giây)	Độ lệch (%)
1	5	0.21	0.69	106.67%
2	10	0.15	1.20	155.56%
3	15	0.32	2.16	148.39%
4	20	0.64	2.89	127.48%
5	25	0.88	5.22	142.30%
6	50	10.71	19.49	58.15%
7	60	19.82	36.92	60.27%
Bộ dữ liệu R101				
STT	Số KH	Thời gian CPU (giây)		
		HNM3 với nhánh và cắt (giây)	HCHNM3 (giây)	Độ lệch (%)
1	5	0.18	0.76	123.40%

2	10	0.16	1.92	169.23%
3	15	0.23	1.99	158.56%
4	20	0.44	2.45	139.10%
5	25	0.59	4.59	154.44%
6	33	4.21	11.59	93.42%

Bảng 5.3 cho thấy, với cụm dữ liệu C101 và R101 hiệu năng của HNM3 tốt hơn HCHNM3 tương ứng với 7 trường hợp và 6 trường hợp thử nghiệm.

5.4. Gán nhóm cho phương tiện và lập kế hoạch di chuyển

Sau khi tạo được các nhóm đi chung, chương này trình bày thuật toán ARV (Assign and Route Vehicles) để gán từng nhóm cho một phương tiện và tìm lộ trình tối ưu cho mỗi phương tiện. Thuật toán này nhằm giảm thiểu tổng quãng đường di chuyển và đảm bảo các ràng buộc về sức chứa và thời gian.

5.5. Xử lý yêu cầu động trong hệ thống lái

Phần này đề cập đến việc xử lý yêu cầu động (yêu cầu đi chung xe trong thời gian thực) trong hệ thống chia sẻ xe lái. Tác giả đề xuất thuật toán HybridRideMatching để ghép đôi yêu cầu mới với xe đang hoạt động, đảm bảo tính linh hoạt và hiệu quả.

CHƯƠNG 6. KẾT LUẬN

6.1. Các kết quả chính luận án đạt được

Chương này tổng kết những đóng góp chính của luận án trong việc xây dựng và tối ưu hóa dịch vụ gọi ý chia sẻ phương tiện đi lại dựa trên khai phá dữ liệu không gian - thời gian. Nghiên cứu đã giải quyết thành công các bài toán chia sẻ xe tĩnh và động, đồng thời đề xuất phương pháp lai kết hợp cả hai mô hình.

Cụ thể, luận án đã đạt được các kết quả sau:

- ◆ Bài toán chia sẻ xe tĩnh: Đề xuất phương pháp khai thác dữ liệu thường xuyên dài nhất và thuật toán aVC-growth để tạo nhóm đi chung, mang lại hiệu quả vượt trội so với các thuật toán hiện có.
- ◆ Bài toán chia sẻ xe động: Đề xuất phương pháp HNM3 để giải quyết bài toán định tuyến phương tiện với cửa sổ thời gian (VRPTW), giảm thiểu chi phí và số lượng xe cần thiết. Đồng thời, đề xuất thuật toán HybridRideMatching để ghép đôi yêu cầu mới với xe đang hoạt động.
- ◆ Mô hình dịch vụ: Đề xuất kiến trúc dịch vụ gợi ý chia sẻ phương tiện, kết hợp giữa chia sẻ xe tĩnh và động, tối ưu số lượng người dùng tham gia nhiều nhất có thể.

Thực nghiệm: Chứng minh tính hiệu quả của các phương pháp và thuật toán đề xuất thông qua các thử nghiệm trên dữ liệu thực tế và tổng hợp.

6.2. Những hạn chế và hướng nghiên cứu tiếp theo

Mặc dù đạt được những kết quả khả quan, luận án vẫn còn một số hạn chế:

- ◆ Tính thực tiễn: Cần kiểm chứng thêm trong môi trường thực tế với quy mô lớn hơn.
- ◆ Khả năng mở rộng: Cần thiết kế và tối ưu hóa để đáp ứng lượng lớn yêu cầu trong thời gian cao điểm.
- ◆ Bảo mật và riêng tư: Cần nghiên cứu và phát triển các giải pháp bảo mật toàn diện hơn.

Hướng nghiên cứu tương lai:

- ◆ Mở rộng mô hình để xem xét các yếu tố đa dạng hơn như sự thoải mái, mức độ tin cậy và ràng buộc môi trường

- ◆ Cải tiến thuật toán tìm kiếm tuyến đường thường xuyên để linh hoạt hơn trong việc tìm ra các tuyến đường phù hợp với nhiều tiêu chí khác nhau
- ◆ Phát triển phương pháp tạo nhóm đi chung linh hoạt, cho phép người dùng tùy chỉnh kích thước nhóm và các tiêu chí khác
- ◆ Ứng dụng máy học và trí tuệ nhân tạo để nâng cao hiệu quả của các thuật toán ghép đôi, dự đoán nhu cầu và tối ưu hóa lộ trình

Xem xét các yếu tố khác ngoài vị trí và thời gian, như tình trạng giao thông, loại xe, sở thích của hành khách và tài xế.

DANH MỤC CÁC CÔNG TRÌNH CỦA TÁC GIẢ ĐÃ CÔNG BỐ LIÊN QUAN ĐẾN LUẬN ÁN

CT1. **Tran Van Manh**, Thi Hong Nhan Vu, “*Design of architecture for carpooling based on flexible pickup and delivery locations*”, Proceeding of the 16th International Conference on IIHMSP in Conjunction with the 13th International Conference on FITAT, November 5–7, 2020, Ho Chi Minh City, Vietnam, DOI: 10.1007/978-981-33-6757-9_26, pp 205-211 (**Scopus**).

CT2. Nhan T.H. Vu, **Manh V.Tran**, “*A Rideshare Model based on Movement Patterns*”, The 13th international conference on future information & communication engineering (ICFICE 2020), 19th February, 2021 Online, Vol. 12, No. 1 ISSN 2765-3811, pp 3-7.

CT3. **Manh V.Tran**, Nhan T.H. Vu, “*A System for Rideshare Service*”, The 13th international conference on future information & communication engineering (ICFICE 2020), 19th February, 2021 Online, Vol. 12, No. 1 ISSN 2765-3811, pp 8-12.

CT4. **Van Manh Tran**, Thi Hong Nhan Vu, “*Leveraging CPLEX to Solve the Vehicle Routing Problem with Time Windows*”, 2021 13th International Conference on Knowledge and Systems Engineering (KSE-2021), Bangkok, Thailand, 10-12 November 2021, DOI: 10.1109/KSE53942.2021.9648591.

CT5. **Van Manh Tran**, Thi Hong Nhan Vu, “*A Routing Method for Ridesharing Service by Applying CPLEX*”, 2022 13th Asian Control Conference (ASCC), Jeju, Korea, Republic of, 2022, DOI: 10.23919/ASCC56756.2022.9828064, pp. 766-771 (**Scopus**).

CT6. Thi Hong Nhan Vu, **Van Manh Tran**, “*A Novel Scheduling Approach for Intelligent Ridesharing System*”, 2022 The 13th Asian Control Conference (ASCC 2022), Jeju Island, Korea, May 4-7, 2022.

CT7. **Van Manh Tran**, Thi Hong Nhan Vu, “*A New Method for Forming Rideshare Groups*”, 2023 15th International Conference on Knowledge and Systems Engineering (KSE-2023), Hanoi, Vietnam, October 18-20, 2023, DOI: 10.1109/KSE59128.2023.10299428 (**Scopus**).