**TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI**

**ĐỒ ÁN TỐT NGHIỆP**

**Đề xuất mô hình và giải thuật tối ưu tần suất gửi tin trong hệ quan trắc môi trường di động.**

**NGUYỄN ĐÌNH TUẤN ANH**

[Anh.ndt164767@sis.hust.edu.vn](mailto:Anh.ndt164767@sis.hust.edu.vn)

**Ngành: Công nghệ thông tin**

**Chuyên ngành: Công nghệ thông tin**

|  |  |
| --- | --- |
| **Giảng viên hướng dẫn:** | ***TS. Nguyễn Phi Lê*** |
| **Bộ môn:** | Công nghệ phần mềm |
| **Viện:** | Công nghệ thông tin – Truyền thông |
| **HÀ NỘI, 12/2019** Lời cam kết Họ và tên sinh viên : Nguyễn Đình Tuấn Anh  Điện thoại liên lạc : 0982666328 Email : anh.ndt164767@sis.hust.edu.vn  Lớp : Công nghệ thông tin- k61 Hệ đào tạo : Kỹ sư tài năng  Tôi – Nguyễn Đình Tuấn Anh– cam kết Đồ án Tốt nghiệp (ĐATN) là công trình nghiên cứu của bản thân tôi dưới sự hướng dẫn của TS. *Nguyễn Phi Lê*. Các kết quả nêu trong ĐATN là trung thực, là thành quả của riêng tôi, không sao chép theo bất kỳ công trình nào khác. Tất cả những tham khảo trong ĐATN – bao gồm hình ảnh, bảng biểu, số liệu, và các câu từ trích dẫn – đều được ghi rõ ràng và đầy đủ nguồn gốc trong danh mục tài liệu tham khảo. Tôi xin hoàn toàn chịu trách nhiệm với dù chỉ một sao chép vi phạm quy chế của nhà trường.   |  |  | | --- | --- | |  | *Hà Nội, ngày 9 tháng 6 năm 2021*  Tác giả ĐATN  *Nguyễn Đình Tuấn Anh* |  Lời cảm ơn Quãng thời gian học tập với vai trò là sinh viên Đại học Bách Khoa Hà Nội đã mang đến cho em nhiều cung bậc cảm xúc khác nhau, áp lực học tập có, vui buồn cũng có và cả những niềm hạnh phúc. Những ngày đầu bước chân vào cổng trường, dưới những hàng cây mát rượi của buổi chiều hè, em đã cảm thấy đây chính là nơi bản thân sẽ có 5 năm quý giá của tuổi sinh viên. Từ một cậu học sinh chưa có hiểu biết gì về công nghệ và thông tin, nay đã sắp tốt nghiệp dưới mái trường kỹ thuật hàng đầu Việt Nam, em có đôi lời nhắn gửi đến các thầy cô và bạn bè tại ngôi trường thân yêu này.  Trước hết em gửi lời cảm ơn tới tất cả thầy cô của trường Đại học Bách Khoa Hà Nội, các thầy, các cô là một phần không thể thiếu trong việc đào tạo ra thế hệ các sinh viên tài năng, và đầy bản lĩnh, sẵn sàng đóng góp xây dựng đất nước. Cảm ơn các thầy cô không chỉ ở những bài học lý thuyết hay thực hành mà cả các bài học cuộc sống mà các thầy cô đã trao đổi với em trong quá trình giảng dạy.  Em đặc biệt gửi lời cảm ơn đến cô Nguyễn Phi Lê, tuy em cũng chỉ mới học tập và làm việc với cô trong thời gian ngắn, cô đã rất tận tình giúp đỡ em trong mọi câu hỏi thắc mắc cả về khoa học lẫn cuộc sống. Trong quá trình làm đồ án tốt nghiệp, cô không chỉ đóng vai trò là người hướng dẫn mà còn là người động viên, giúp em và cả các bạn luôn yên tâm, tập trung, nỗ lực hoàn thành đồ án của mình.  Tiếp theo, em xin gửi lời cảm ơn đến tập thể lớp Kỹ sư tài năng Công nghệ thông tin khóa 61, những người bạn đã luôn đồng hành cùng em trong những năm qua.  Cuối cùng em muốn gửi lời cảm ơn đến bố và mẹ của em, đã luôn động viên em cố gắng, cung cấp cho em điều kiện thuận lợi nhất để an tâm học hành.  Trong quá trình làm đồ án, với vai trò là một sinh viên còn nhiều kiến thức hạn chế, em không tránh khỏi những sai sót, em mong nhận được sự góp ý của thầy cô và các bạn để hoàn thiện kiến thức của bản thân.  Em xin chân thành cảm ơn ! Tóm tắt Thế giới nói chung và các nước đang phát triển nói riêng, trong đó có Việt Nam, đang phải đối mặt với tình trạng ô nhiễm không khí ngày càng trầm trọng. Chúng ta từ trước tới nay hầu như chưa ý thức được những tác động nghiêm trọng của chất lượng không khí tới sức khỏe của bản thân, của xã hội và cả môi trường, sinh vật xung quanh. Cho đến một vài năm trở lại đây, các thông số AQI tại Hà Nội và một số thành phố lớn của Việt Nam đạt mức cao ngất ngưởng mới thu hút sự quan tâm của báo chí và mọi người dân. Tuy nhiên các trạm đo chất lượng không khí tại Việt Nam hay một số nước nghèo còn rất ít khiến cho việc theo dõi chất lượng không khí và các nghiên cứu liên quan gặp nhiều rào cản. Từ nhu cầu đó, em đề xuất mô hình sử dụng hệ quan trắc môi trường di động để thu thập dữ liệu về chất lượng không khí.  Trong quá trình xây dựng và giải quyết bài toán, em sử dụng một mô hình xác suất rất phổ biến đó là phân phối Poisson. Xây dựng hàm chi phí và hàm tối ưu, từ đó nghiên cứu các giải thuật tối ưu phù hợp. Do thời gian có hạn, em tiến hành thử nghiệm trên một số phương pháp tối ưu phổ biến là các thuật toán tìm kiếm chia lưới và giải thuật bầy đàn. Từ đó chọn ra phương pháp tối ưu nhất đồng thời tinh chỉnh lại để giải thuật tối ưu hơn.  Đóng góp chính của đồ án này là xây dựng một mô hình khả thi sử dụng hệ quan trắc di động, để thu thập dữ liệu không khí trên diện rộng với chi phí thấp, tính linh hoạt cao. Đề xuất một giải thuật tối ưu tần xuất gửi tin của các cảm biến trong hệ quan trắc đó sao cho tiết kiệm được năng lượng, tài nguyên tính toán của máy chủ, tránh lãng phí băng thông đồng thời đạt được vùng bao phủ tối đa. Cuối cùng em trình bày các kết quả thực nghiệm, kết luận dựa trên kết quả đó và đề xuất các hướng nghiên cứu có thể phát triển thêm dựa trên hướng nghiên cứu hiện tại.  Nội dung  [Lời cam kết 2](#_Toc74253322)  [Lời cảm ơn 3](#_Toc74253323)  [Tóm tắt 4](#_Toc74253324)  [Chương 1: Giới thiệu đề tài 7](#_Toc74253325)  [1.1 Đặt vấn đề 7](#_Toc74253326)  [1.2 Giải pháp hiện tại và vấn đề 7](#_Toc74253327)  [1.3 Định hướng giải pháp 9](#_Toc74253328)  [1.4 Đóng góp của đồ án 11](#_Toc74253329)  [1.5 Bố cục của đồ án 11](#_Toc74253330)  [Chương 2 : Các nghiên cứu liên quan 13](#_Toc74253331)  [Chương 3 : Các kiến thức nền tảng 15](#_Toc74253332)  [3.1 Phân phối poisson và phân phối poisson đa chiều 15](#_Toc74253333)  [3.2 Các phương pháp tính gần đúng tích phân 17](#_Toc74253334)  [3.2.1 Phương pháp sử dụng tổng Reimann 18](#_Toc74253335)  [3.2.2 Phương pháp sử dụng luật hình thang 20](#_Toc74253336)  [3.2.3 Phương pháp Simpson 21](#_Toc74253337)  [3.3 Tổng quan các giải thuật tối ưu 22](#_Toc74253338)  [3.3.1 Giải thuật bầy ong nhân tạo (Artificial Bee Colony - ABC) 24](#_Toc74253339)  [3.3.2 Giải thuật bầy đàn (Particle Swarm Optimization – PSO) 25](#_Toc74253340)  [Chương 4 : Giải thuật tối ưu hoá tần suất gửi tin trong mạng quan trắc di động 27](#_Toc74253341)  [4.1 Mô tả bài toán 27](#_Toc74253342)  [4.1.1 Đề xuất mô hình mạng 27](#_Toc74253343)  [4.1.2 Xây dựng công thức và xác định hàm tối ưu 29](#_Toc74253344)  [4.2 Tổng quan các hướng tiếp cận giải quyết bài toán 33](#_Toc74253345)  [4.2.1 Phân tích giải thuật bầy ong nhân tạo (Artificial Bee Colony-(ABC)) 33](#_Toc74253346)  [4.2.2 Phân tích giải thuật bầy đàn PSO 33](#_Toc74253347)  [4.2.3 Đề xuất giải thuật cái tiến. 34](#_Toc74253348)  [Chương 5 Đánh giá thực nghiệm. 38](#_Toc74253349)  [5.1 Các tham số đánh giá 38](#_Toc74253350)  [5.2 Phương pháp thí nghiệm 40](#_Toc74253351)  [5.3 Kết quả thí nghiệm 40](#_Toc74253352)  [5.3.1 Tham số tối ưu cho phương pháp tính gần đúng tích phân 40](#_Toc74253353)  [5.3.2 Các tham số tối ưu cho thuật toán PSO, ABC và PSO-Based 42](#_Toc74253354)  [5.3.3 So sánh kết quả 3 giải thuật PSO, PSO-Based và ABC. 45](#_Toc74253355)  [5.3.4 Ảnh hưởng của các tham số trong hàm chi phí tới kết quả tối ưu 48](#_Toc74253356)  [Chương 6 : Kết luận 51](#_Toc74253357)  [6.1 Kết luận 51](#_Toc74253358)  [6.2 Định hướng phát triển của đề tài 51](#_Toc74253359)  [Tài liệu tham khảo 53](#_Toc74253360) Chương 1: Giới thiệu đề tàiĐặt vấn đề Việt Nam, với vai trò là một nước đang phát triển, sự phát triển kinh tế luôn đi kèm với quá trình công nghiệp hóa và hiện đại hóa. Bên cạnh những lợi ích về mặt cơ sở vật chất, đời sống nhân dân được cải thiện, thì những ảnh hưởng của quá trình ô nhiễm môi trường lên đời sống con người cũng ngày càng mạnh mẽ. Một trong số đó là vấn đề ô nhiễm không khí, gây tổn hại không nhỏ đến sức khỏe con người. Đặc biệt, các chuyên gia cho biết cần quan tâm tới hai chỉ số bụi mịn : PM10 và PM2.5 – các hạt bụi có đường kính nhỏ hơn 10 µm và nhỏ hơn 2.5 µm. Những năm gần đây, nước ta có sự xuất hiện của bụi siêu mịn PM1.0 (dưới 1µm) và bụi nano PM0.1 (dưới 0.1 µm). Nhờ kích cỡ rất nhỏ của mình, bụi siêu mịn PM2.5 có thể qua đường hô hấp để len lỏi vào sâu trong phổi, gây ra các bệnh hô hấp. Tình hình chất lượng không khí hiện nay của Việt Nam đang trong tình trạng xấu, không chỉ do yếu tố không khí và địa lý mà còn do hành động thiếu trách nhiệm của con người, các công trình và thiếu ý thức chung trong việc giữ gìn môi trường xung quanh. Khi người dân thấy được mức độ ảnh hưởng của các chỉ số bụi mịn này tới sức khỏe của bản thân, mọi người sẽ có ý thức để cùng chung tay bảo vệ môi trường.  Vì vậy, việc theo dõi chất lượng không khí trong các thành phố lớn là hết sức cần thiết, từ đó cho phép người dân biết được chất lượng không khí xung quanh nơi ở, nơi làm việc, giúp cho họ có các biện pháp bảo vệ sức khỏe của bản thân và cả những người xung quanh. Hơn nữa, hiện nay các đề tài nghiên cứu khoa học liên quan đến chất lượng không khí cũng đang thu hút được rất nhiều sự quan tâm của các nhà nghiên cứu trong nhiều lĩnh vực, vì vậy đặt ra nhu cầu thu thập dữ liệu về chất lượng không khí không chỉ trong thành phố để phục vụ cho con người mà còn thu thập dữ liệu trên một vùng bất kỳ để làm phương tiện cho các nghiên cứu khoa học liên quan. Giải pháp hiện tại và vấn đề Với mục tiêu là thu thập và theo dõi chất lượng không khí, hiện nay, đã có một số trạm thu thập dữ liệu được đặt cố định tại các địa điểm thuộc các thành phố lớn trên toàn thế giới. Tuy nhiên, việc thu thập dữ liệu như vậy gặp phải điểm yếu là thiếu đi tính đa dạng và độ chính xác vì nó chỉ theo dõi được dữ liệu tại một điểm chứ không phải trên diện rộng. Đặc biệt, khi cần dữ liệu có độ chính xác cao tại các vị trí cụ thể thì không thể thực hiện được. Để giải quyết vấn đề này thì ta có thể đặt các trạm đo chất lượng không khí tại nhiều điểm hơn, xong làm như vậy đòi hỏi chi phí cao, không khả thi trong thực tế.    Hình 1: Các trạm đo chất lượng không khí tại Việt Nam.  Một số hướng nghiên cứu đã đưa ra giải pháp sử dụng trí tuệ nhân tạo để đưa ra dự đoán về chất lượng không khí tại một thời điểm và địa điểm nhất định, tất nhiên việc dự đoán như vậy cũng có thể có sai số lớn.  Với các giải pháp hiện tại, ta thấy được việc thu thập dữ liêu về chất lượng không khí đang gặp phải thách thức lớn về cả chi phí lẫn giải pháp. Do vậy em xin được trình bày một số hướng tiếp cận có thể giải quyết phần nào được những hạn chế của các phương pháp hiện tại, từ đó chọn ra giải pháp phù hợp nhất. Định hướng giải pháp Với sự phát triển của mạng internet không dây, cộng với nhu cầu về thu thập dữ liệu, hiện nay có rất nhiều giải pháp cho việc thu thập dữ liệu ví dụ như: mạng cảm biến không dây (wireless sensors network-WSN) Hình 2 và hệ quan trắc di động(Mobile Crowdsensing Systems(MCS)) Hình 3    Hình 2:Mạng cảm biến không dây.  Với mạng cảm biến không dây, ta có một số lượng lớn các cảm biến được bố trí tương đối gần nhau trên một vùng cần theo dõi, các cảm biến này thu thập dữ liệu định kỳ và chuyển tiếp cho nhau gửi về một trung tâm thu thập dữ liệu để xử lý. Với bài toán đo chất lượng không khí, các điểm có khoảng cách tương đối gần nhau có thể coi như có chất lượng không khí không chênh lệch nhiều tại một thời điểm. Do vậy, việc sử dụng mạng cảm biến không dây gây ra nhiều lãng phí. Hơn nữa các cảm biến này thường cần phải được thay thế hoặc sạc pin định kỳ, đây cũng là một điểm yếu của mạng cảm biến không dây.    Hình 3: Hệ quan trắc di dộng.  Với mục đích giải quyết được một số điểm yếu của WSN, hệ quan trắc di động ra đời, MCS là một mạng lưới các cảm biến có khả năng thu thập dữ liệu được tính hợp trên phương tiện giao thông hoặc con người. Do có thể di chuyển nên các cảm biến này có thể thu thập được dữ liệu trên diện rộng rồi gửi về server mà không cần số lượng cảm biến lớn mà vẫn có thể thu thập được dữ liệu đa dạng và phong phú.  Với phân tích trên, em chọn giải pháp sử dụng MCS cho bài toán đo chất lượng không khí. Ở đây, các cảm biến nên được tích hợp trên các phương tiện giao thông hoạt động có tính chu kỳ và thường xuyên, ví dụ như tích hợp trên các tuyến xe bus, hoặc xe khách. Tuy nhiên các cảm biến này cần có một cơ chế thu thập dữ liệu hợp lý, việc thu thập dữ liệu một cách dày đặc rồi gửi lên trung tâm xử lý hoặc lưu trữ dữ liệu gây ra lãng phí về tài nguyên và đường truyền, ngược lại, việc thu thập dữ liệu thưa thớt không đảm bảo dữ liệu được cập nhật thường xuyên, vùng bao phủ của các điểm thu thập dữ liệu nhỏ, không đáp ứng được nhu cầu thực tế.  Vậy với bài toán của mình, em định hướng sử dụng mô hình MCS cùng với thuật toán tối ưu tần suất gửi dữ liệu sao cho không lãng phí đường truyền, năng lượng cũng như khả năng xử lý, lưu trữ của server, đồng thời đảm bảo dữ liệu đủ để đáp ứng yêu cầu về độ bao phủ cũng như tính cập nhật liên tục. Đóng góp của đồ án Đồ án này có 3 đóng góp chính như sau:   * Đề xuất một mô hình hệ quan trắc di động nhắm mục đích thu thập dữ liệu về chất lượng không khí trên một vùng cần theo dõi. * Đề xuất một mô hình toán học khả thi trong hệ quan trắc đó, xác định hàm chi phí và hàm lợi ích của mô hình. Từ đó xây dựng công thức tính hàm tối ưu. * Đề xuất các tham số phù hợp cho việc tính toán và giải thuật tối ưu tổng quát dựa trên giải thuật bầy đàn cho hệ quan trắc di động đã nêu trên.  Bố cục của đồ án Kết cấu của đồ án của em trong phần còn lại như sau:  Chương 2 em sẽ trình bày về các nghiên cứu liên quan đến vấn đề trên, những ý tưởng chính của mô hình và giải pháp cũng như phân tích các ưu nhược điểm của các nghiên cứu đó. Từ đó đề xuất ra mô hình và giải pháp mới.  Chương 3 em sẽ trình bày cơ sở lý thuyết cần thiết để xây dựng mô hình và giải quyết vấn đề. Bao gồm: phân phối Poisson và phân phối Poisson đa chiều được sử dụng để mô hình lý thuyết của bài toán, các phương pháp tính tích phân gần đúng để thực hiện khâu tính toán công thức, và các giải thuật tiến hóa tối ưu là giải thuật bầy ong nhân tạo và giải thuật bầy đàn.  Chương 4 sẽ phân tích mô hình của hệ quan trắc môi trường di động, đồng thời nêu ra các hướng tiếp cận giải quyết bài toán và cuối cùng là các phương pháp tối ưu hiệu quả trong việc giải bài toán đề ra.  Chương 5 em sẽ trình bày các tham số đầu vào của bài toán và nêu ra kết quả thực nghiệm.  Chương 6 từ kết quả thực nghiệm ở chương 5, em rút ra kết luận cho mô hình bài toán đang xét và đề xuất một số hướng nghiên cứu liên quan để mở rộng bài toán giải quyết các vấn đề còn tồn đọng. : Các nghiên cứu liên quan Hiện nay đã có rất nhiều nghiên cứu về chủ đề hệ quan trắc di động, với đa dạng mục đích như: thu thập hình ảnh, đo độ ẩm, đo nhiệt độ, chất lượng không khí, …  Ngoài ra cũng có một số bài báo liên quan đến tối ưu vùng bao phủ trong bài toán đo chất lượng không khí. Nhưng theo hiểu biết hiện tại của em, thì chưa có một nghiên cứu nào giải quyết bài toán tương tự như bài toán mà em đề xuất.  Nghiên cứu trong [9] cũng chỉ ra rằng việc xây dựng các trạm đo chất lượng không khí rất tốn kém, mặc dù với ưu điểm là chất lượng kết quả đo rất chính xác. Từ đó họ đề xuất sử dụng một thiết bị đo chất lượng không khí gắn kèm với các tình nguyện viên di chuyển quanh thành phố để thu thập chất lượng không khí rồi gửi về server. Nghiên cứu này tập trung vào việc xây dựng cảm biến phục vụ cho mục đích thu thập chất lượng không khí. Điểm yếu của nghiên cứu này là đòi hỏi con người phải di chuyển theo lịch trình và đo tại các điểm cho trước, từ đó tốn kém tài nguyên về con người, cũng dễ dẫn đến sai sót do con người làm chủ quá nhiều yếu tố trong quá trình đo.  Với nghiên cứu trong [3], các tác giả đề xuất một mô hình để tối ưu vùng bao phủ trong việc thu thập dữ liệu về môi trường nói chung. Vùng cần theo dõi được coi là một hình chữ nhật chia thành các lưới chữ nhật nhỏ hơn theo khu vực như trường học, khu công nghiệp, … với các mức độ quan trọng khác nhau. Từ đó các tác giả đề xuất một thuật toán để tìm ra các điểm cần đo chất lượng không khí dựa trên bài toán tối đa hóa vùng bao phủ và thuật toán giúp tìm ra tình nguyện viên phù hợp di chuyển đến các điểm phù hợp để lấy mẫu dựa trên khoảng cách. Nghiên cứu này có điểm mạnh là cân nhắc các vùng quan trọng với các vùng không quan trọng, từ đó đưa ra lịch trình đo phù hợp. Nhưng vẫn có yếu tố chủ quan con người, đòi hỏi những tình nguyện viên phải đi tới các điểm cần đo.  Ren et al.[4] đề xuất một mô hình sử dụng các phương tiện trong thành phố để chuyển tiếp gói tin đến các trạm xử lý dữ liệu, trong đó các trạm thu thập dữ liệu về môi trường được đặt cố định.  Với các bài toán thu thập dữ liệu về chất lượng không khí, do thiết bị có thể có sai số hoặc một số người tham gia gửi dữ liệu giả nên Liu et al.[5] đề xuất một thuật toán để tìm ra dữ liệu chính xác nhất dựa trên tổng hợp các dữ liệu thu được.  Trong bài báo [6], các tác giả đề xuất một mô hình sử dụng cảm biến gắn trên xe bus, lấy dữ liệu thật từ các tuyến xe Bus tại Hà nội cân nhắc thời điểm thu thập dữ liệu về chất lượng không khí để tối đa hóa vùng bao phủ. Các vùng cần bao phủ cũng tương tự như [3], là một hình chữ nhật được chia thành các lưới ô vuông, các lưới này được cân nhắc là quan trọng hay không. Từ đó đặt ra mục tiêu đo chất lượng không khí tại những thời điểm, địa điểm nào để tối đa hóa phần trăm diện tích bao phủ trên các vùng quan trọng.  Từ các nghiên cứu trên em nhận thấy vẫn có những yếu tố rất con người, đòi hỏi phải có người kiểm soát các cảm biến, bật tắt thích hợp. Hay sử dụng thuật toán để tự động bật tắt cũng không phù hợp với các cảm biến yêu cầu có kích thước nhỏ, tiêu thụ ít điện năng. Để giải quyết được các vấn đề đó, Chương 3 sẽ trình bày cơ sở lý thuyết cơ bản áp dụng cho mô hình đề xuất. : Các kiến thức nền tảng Để xây dựng các công thức và thực hiện bài toán, em nhận thấy cần các kiến thức nền tảng được trình bày dưới đây. Phân phối poisson và phân phối poisson đa chiều Phân phối Poisson hay được sử dụng trong các mô hình tính toán gần đúng với thực tế và đã chứng minh được tính hiệu quả của nó.  Biến ngẫu nhiên X được gọi là tuân theo luật phân phối Poisson, ký hiệu là , nếu hàm xác suất của nó có dạng:  Công thức 1: Phân phối Poisson.  Phân phối Poisson có nhiều ứng dụng trong mô hình xác suất lý thuyết gắn với thực tế, thường được chọn để mô hình biến ngẫu nhiên là số cuộc điện thoại của một tổng đài trong một ngày, số lượng khác của một nhà hàng trong một giờ, số lượng xe đi vào đường cao tốc trong một giờ, …  Có thể chứng minh được rằng  Khi  Công thức 2: Xấp xỉ phân phối Poisson và phân phối nhị thức.  Trong thực tế nếu n lớn và p bé thì người ta cũng có thể sử dụng xấp xỉ phân phối nhị phân với phân phối Poisson.  Điều này có thể lý giải được vì sao phân phối Poisson hay được dùng trong thực tế, chả hạn trong một ngày có rất nhiều người đi qua một nhà hàng (chả hạn 1000), mỗi người có xác suất 0.01 trở thành khách của nhà hàng đó, vậy biến ngẫu nhiên số khách của nhà hàng đó trong một ngày có thể xấp xỉ bằng phân phối .  Với nhu cầu thực tế, đôi khi người ta cần quan sát số lượng các sự kiện xảy ra trong một khoảng thời gian t nào đó, mà t là một biến số thay đổi.  Để mô hình các bài toán có tính chất trên, trong [11] đã chứng minh được rằng, với các sự kiện có tính chất như sau:   * Xác suất có it nhất một sự kiện xảy ra trong khoảng thời gian rất nhỏ . Với được gọi là tỉ lệ đến của phân phối Poisson. Trong thực tế được tính bằng kiểm nghiệm thực tế, chả hạn như 10 khách/giờ, 20 xe/giây đi qua một điểm, … * Số lượng sự kiện xảy ra trong một quãng thời gian cố định không phụ thuộc vào thời gian bắt đầu của khoảng thời gian đó. * Số lượng sự kiện xảy ra tại các khoảng thời gian khác nhau là các biến độc lập. * Nếu một sự kiện xảy ra tại một thời điểm nào đó, xác suất có điều kiện là một sự kiện khác xảy ra cũng tại thời điểm đó là bằng 0. Điều kiện này có thể chấp nhận được hay không chấp nhận được với các sự kiện thực tế khác nhau.   Thì số lượng sự kiện xảy ra trong một khoảng thời gian t bất kỳ tuân theo phân phối Poisson:  Công thức 3: Phân phối Poisson trên miền thời gian.  Với việc mở rộng phân phối Poisson sang miền thời gian, ta cũng có thể mở rộng sang miền không gian nhiều chiều. Giả sử ta cần mô phỏng số lượng tai nạn giao thông xảy ra trên một vùng diện tích nào đó của đường cao tốc với giả sử rằng:   * Phân phối xác suất số lượng tai nạn xảy ra trên một vùng diện tích S ký hiệu là X(S) chỉ phụ thuộc vào giá trị của S. * Với các vùng là các vùng không giao nhau thì là các biến động lập và: * Biết rằng một tai nạn đã xảy ra tại một điểm, xác suất xảy ra một tai nạn khác tại đúng điểm đó bằng 0. Điều này khá hợp lý nếu xét trong khoảng thời gian không quá lớn. * Nếu   Ta cũng có thể mô hình biến ngẫu nhiên này bằng phân phối Poisson:  Công thức 4: Phân phối Poison trên miền không gian.  Với có thể được đo từ thực tế, hoặc xấp xỉ bằng tính toán ví dụ như 100 tai nạn/  Phân phối trên được gọi là phân phối Poisson đều trong không gian, nếu trường hợp các phân phối không đều, nếu ta có thể xác định được hàm mật độ tại từng điểm thì ở công thức trên ta có thể thay  Trong bài toán của em, em sẽ dùng phân phối Poisson đều trên không gian 2 chiều để mô hình số các điểm được đo chất lượng không khí trên đường. Các phương pháp tính gần đúng tích phân Trong nhiều bài toán trên thực tế và lý thuyết, ta cần phải tính toán tích phân xác định của một hàm toán học nào đó. Hàm toán học đó có thể rất đơn giản, cũng có thể rất phức tạp, có thể là hàm đơn biến hoặc một hàm đa biến. Trong phần này em xin tập trung vào việc trình bày phương pháp tính gần đúng tích phân của hàm đơn biến, việc mở rộng sang hàm đa biến trong nhiều trường hợp là hoàn toàn tương tự.  Giả sử ta cần tính  Nếu hàm f(x) liên tục trên [a, b] và có nguyên hàm F(x), ta có:  Công thức 5: Công thức Newton-Lepniz.  Trong trường hợp f(x) cho ở dạng bảng hoặc f(x) đã biết nhưng không xác định được nguyên hàm F(x) ví dụ như hàm f(x) = , thì thay vì việc tính chính xác ta chỉ có thể tính gần đúng tích phân xác định của f(x).  Trước hết ta phân hoạch đoạn [a, b] thành n phần nhỏ hữu hạn với (i = 1, 2, …, n) bởi những điểm a = < < < .  Ta có tổng Riemann S của hàm f trên đoạn [a, b] với phép phân chia trên là:  S = với .  Khi ta chia đoạn [a, b] thành các đoạn con đủ nhỏ thì tổng S trên được coi là kết quả gần đúng của phép tính tích phân xác định . Phương pháp sử dụng tổng Reimann Việc chọn là hoàn toàn tùy ý trong khoảng trên. Tuy nhiên có một số trường hợp đặc biệt như sau:   * thì S được gọi là tổng Riemann trái. * thì S được gọi là tổng Riemann phải. * thì S được gọi là tổng Riemann giữa, ví dụ trong Hình 4.     Hình 4: Tổng Reimann giữa của hàm số y = .  Từ tổng Reimann ta có phương pháp gần đúng để tính tích phân xác định như sau:  Chia đoạn [a, b] thành các đoạn con bằng nhau là: .  Khi đó ta có các công thức sau  Công thức 6: Tổng Reimann trái.    Công thức 7: Tổng Reimann phải.  Công thức 8: Tổng Reimann giữa.  Để tính được các tổng Reimann trên bằng thuật toán thì đơn giản ta chỉ cần sử dụng một vòng for là đủ.  Nếu hàm f(x) có đạo hàm bậc 1 xác định trên đoạn [a, b], người ta chứng minh được rằng:  Với  Do đó với n càng lớn thì phép tính tích phân bằng tổng Reimann càng có xu hướng cho độ chính xác cao.  Tương tự ta cũng có một bất đẳng thức sai số của phép lấy tổng giữa trong trường hợp f(x) có đạo hàm bậc 2 trên đoạn [a, b]:  Với  Với việc đánh giá được chặn trên của sai số, hiện nay có một số phương pháp tính tích phân gần đúng với cơ chế “thích nghi” (adaptive), với những đoạn có ước lượng sai số lớn thì người ta chia làm nhiều khoảng hơn, với những đoạn có chặn trên sai số nhỏ người ta dung ít khoảng hơn để xấp xỉ. Việc làm như vậy có ưu điểm vẫn đảm bảo được sai số nhỏ nhưng độ phức tạp thuật toán có sự cải tiến, tuy nhiên giải thuật này chỉ áp dụng được với hàm có thể tính được đạo hàm hay đạo hàm bậc cao một cách dễ dàng, hơn nữa các chặn trên của sai số có thể có sai khác tương đối với sai số thực tế. Phương pháp sử dụng luật hình thang Trong khi các tổng Reimann xấp xỉ tích phân trên mỗi đoạn con bằng diện tích của hình chữ nhật thì trong phương pháp sử dụng luật hình thang, người ta sử dụng diện tích hình thang để ước lượng xấp xỉ giá trị của tích phân xác định đó.    Hình 5: Luật hình thang.  Trong Hình 5 tích phân trên mỗi đoạn được tính bởi công thức: , chính là diện tích hình thang tạo bởi các đường nét đứt và trục x trên hình.  Lấy tổng trên tất cả các đoạn ta có:  Công thức 9: Luật hình thang. Phương pháp Simpson Mở rộng từ phương pháp sử dụng luật hình thang. Thực chất việc sử dụng luật hình thang chính là việc xấp xỉ một hàm bậc 1 đi qua 2 điểm có tọa độ và . Mở rộng ý tưởng này với 3 điểm hay nhiều điểm hơn ta được phương pháp Simpson.  Xét trường hợp 3 điểm, chỉ tồn tại duy nhất một đường parabol đi qua 3 điểm ,. Giả sử phương trình của parabol là . Giải phương trình này người ta xác định được  Lấy tổng trên tất cả các đoạn ta có:  Công thức 10: Luật Simpson.  Tổng kết lại, hiện nay có rất nhiều phương pháp khác nhau để sử dụng trong việc tính xấp xỉ tích phân xác định của một hàm số. Mỗi phương pháp đều có ưu, nhược điểm khác nhau và các thông số cần tinh chỉnh sao cho thời gian chạy thuật toán nhanh, sai số nhỏ. Trên thực tế việc chọn các thông số và thuật toán phù hợp phụ thuộc vào bài toán cần giải và kiểm nghiệm thực tế của người lập trình. Tổng quan các giải thuật tối ưu Chủ đề tối ưu hóa luôn luôn thu hút được sự quan tâm, chú ý của không chỉ các nhà khoa học mà cả các nhà quản lý, nhà kinh tế, …. Đặc biệt trong bối cảnh cạnh tranh gay gắn trên mọi mặt của cuộc sống, cá nhân hay tổ chức nào đưa ra được mô hình tối ưu hơn, tiết kiệm chi phí hơn sẽ chiếm lợi thế lớn trên lĩnh vực đó. Chính vì vậy người ta phải quan tâm tìm ra phương án tối ưu nhất để đạt được mục tiêu mong muốn, trong những điều kiện rằng buộc nhất định. Đó gọi là bài toán tối ưu.  Ngày nay với sự phát triển của khoa học, kỹ thuật và công nghệ. Các phương pháp tối ưu ngày càng phát triển và được ứng dụng rộng rãi trong mọi lĩnh vực.  Thông thường các bài toán tối ưu được phát biểu như sau:  Cực tiểu hóa một hàm là một quy tắc ứng mỗi điểm với một số thực  Ở đây trong hầu hết các trường hợp , X được gọi là miền xác định hay miền tìm kiếm của x.  Thông thường miền xác định (hay miền đang xét) X được xác định bằng một tập hợp các hàm rằng buộc.  Hầu hết mỗi phương pháp tối ưu chỉ phù hợp với một hoặc một lớp các bài toán nhất định. Vì vậy, người ta chia các phương pháp tối ưu thành 2 nhóm chính là quy hoạch tuyến tính và quy hoạch phi tuyến. Trong phần này em tập trung giới thiệu nhóm các phương pháp quy hoạch phi tuyến.  Đơn giản nhất là phương pháp chia lưới trên hệ trục tọa độ (grid search method), ở phương pháp này miền X được chia thành các lưới ô vuông và dùng giải thuật vét cạn trên các giao điểm của các lưới ô vuông đó. Ưu điểm của giải thuật này là dễ hiểu, dễ cài đặt tuy nhiên độ chính xác của kết quả phụ thuộc vào việc chia miền X thành các vùng càng nhỏ càng tốt. Khi số chiều n của X tăng lên cao thì giải thuật có độ phức tạp thời gian lớn nên chạy chậm. Tương tự với phương pháp này là phương pháp tìm kiếm ngẫu nhiên, thay vì việc chia lưới như trên thì các điểm được tính toán là tuân theo một phân phối xác suất ngẫu nhiên nào đó.  Bên cạnh hai phương pháp đơn giản nêu trên, thì các thuật toán sử dụng gradient, hoặc đạo hàm cũng tương đối phổ biển, ví dụ như phương pháp gradient-descent hay phương pháp Newton. Điểm yếu của các phương pháp này là yêu cầu hàm tối ưu phải có đạo hàm hay thậm chí là có đạo hàm bậc k. Trong thực tế nhiều bài toán dẫn đến hàm tối ưu không có đạo hàm, thậm chí là không liên tục, hoặc việc tính đạo hàm quá phức tạp.  Khắc phục được các nhược điểm của các thuật toán nêu trên, thuật toán tối ưu bầy đàn là lớp các giải thuật dựa trên hành vi, tập tính di cư hay tìm kiếm nguồn thức ăn của các bầy đàn trong tự nhiên. Các thuật toán này được gọi là meta-heuristic, với khả năng tìm kiếm tối ưu cho hàm đa biến, có rằng buộc, không yêu cầu các hàm này phải liên tục hay có đạo hàm. Tất nhiên, vì là thuật toán dạng heuristic nên các phương pháp tối ưu bầy đàn không đảm bảo tìm kiếm được lời giải tối ưu. Việc áp dụng thuật toán phù hợp với mỗi bài toán đòi hỏi quá trình thử nghiệm và kinh nghiệm của mỗi người. Với bài toán của mình, được sự gợi ý của giảng viên hướng dẫn, cùng với quá trình tự tìm tòi, học hỏi, em xin trình bày hai giải thuật được em sử dụng để thử nghiệm trong mô hình của mình. Giải thuật bầy ong nhân tạo (Artificial Bee Colony - ABC) Thuật toán được đề xuất bởi Dervis Karaboga và Bahriye Basturk trong [1] .  Lấy cảm hứng từ quá trình tìm kiếm thức ăn của loài ong, thuật toán này, bầy ong bao gồm 3 nhóm: ong thợ, ong quan sát và ong trinh thám. Trong đó ong thợ có nhiệm vụ đi đến các nguồn thức ăn, ong quan sát đợi ở tổ và nhận thông tin từ các ong thợ báo về nguồn phấn hoa. Ong trinh thám có nhiệm vụ khám phá nguồn thức ăn mới. Ứng với mỗi nguồn thức ăn là một ong thợ, do đó số lượng ong thợ bằng số nguồn thức ăn. Trong thuật toán ABC nguyên bản thì số lượng ong quan sát cũng bằng số lượng ong thợ và chỉ có một ong trinh thám. Giả sử ta cần cực tiểu hóa với hợp , và số lượng ong thợ là n. Mỗi một nguồn thức ăn được coi là một đáp án khả thi cho việc tối ưu hóa hàm mục tiêu.  Thuật toán bắt đầu bằng việc khởi tạo ngẫu nhiên một tập n nguồn thức ăn cho các ong thợ, là các véc tơ D chiều . Sau khi khởi tạo là quá trình tìm kiếm lời giải tối ưu bao gồm L vòng lặp, mỗi vòng lặp lần lượt là quá trình tìm kiếm của bầy ong thợ, ong quan sát rồi đến ong trinh thám. Mỗi ong thợ và ong quan sát ứng với nguồn thức ăn thứ i thực hiện quá trình tìm kiếm nguồn thức ăn mới (lời giải khả thi mới) được thực hiện dựa theo công thức dưới đây:  Công thức 11: Cập nhật tìm kiếm lời giải mới.  Trong đó:   * j được chọn ngẫu nhiên trong tập {1,2, …., D}. * k được chọn ngẫu nhiên trong tập {1,2, …, n} \{i}. * được chọn ngẫu nhiên trong [-1,1].   Nếu không năm trong miền đang xét, ta có thể chọn nó ở một giá trị biên gần với nhất. Sau khi thực hiện cập nhật theo công thức ta được nguồn thức ăn mới , thực hiện tính nếu cao hơn giá trị cũ thì không thực hiện cập nhật. Nếu không cải thiện sau một số hữu hạn vòng lặp nhất định chứng tỏ nguồn thức ăn này không tiềm năng, không có khả năng khám phá lời giải tối ưu hơn, khi đó đến giai đoạn của ong trinh thám, nó sẽ khởi tạo lại một giá trị mới hoàn toàn ngẫu nhiên cho nguồn thức ăn này. Vậy câu hỏi đặt ra là ong thợ và ong quan sát khác gì nhau?  Ở giai đoạn của ong thợ, chúng luôn gắn liền tương ứng một-một với nguồn thức ăn nào đó, thực chất ong thợ thứ i luôn gắn với nguồn thức ăn thứ i. Với ong quan sát, nó sẽ chọn nguồn thức ăn thứ i ứng với xác suất:  Với  Công thức xác suất trên cho thấy nếu càng nhỏ thì càng lớn nên càng lớn, từ đó nguồn thức ăn càng tối ưu thì càng thu hút được ong quan sát đến khám phá. Giải thuật bầy đàn (Particle Swarm Optimization – PSO) Giải thuật bầy đàn được đề xuất bởi Kenedy et el.[10] vào năm 1995.  Thuật toán PSO hoạt động chưa trên cơ chế tương tác giữa một tập hợp các cá thể, tương tự như thuật toán ABC, các cá thể này được đặc trưng bởi vị trí của nó. Có thể hiểu đơn giản, vị trí của cá thể thứ i là , với là một véc tơ D chiều. Ngoài ra mỗi cá thể cũng gắn với một véc tơ chuyển động D chiều: . Thuật toán bắt đầu bằng việc khởi tạo một tập n cá thể với vị trí ngẫu nhiên và véc tơ chuyển động ban đầu được gắn bằng 0. Thuật toán thực hiện L vòng lặp.  Trước mỗi bước lặp, cá thể thứ i lưu trữ vị trí tốt nhất của nó từ khi thuật toán được chạy gọi là và vị trí tốt nhất của cả quần thể là gBest. Ở đây vị trí tốt nhất được hiểu là một véc tơ D chiều x mà f(x) nhỏ nhất. Tiếp theo, các được cập nhật theo công thức sau:  Với j = (1, 2, …, D)  Công thức 12: Cập nhật vector chuyển động.  Trong đó:   * là các tham số do người lập trình quyết định. * rand (0,1) là một số thực ngẫu nhiên trong [0,1].   Sau đó cập nhật với i = (1, 2, …., n) và cập nhật lại các pBest và gBest nếu cần thiết. Các cá thể di chuyển với một phần đi theo pBest và một phần đi theo gBest, đây chính là yếu tố khai thác của thuật toán, giúp cho thuật toán hội tụ tại điểm cực tiểu (có thể là cực tiểu địa phương). Tham số w gọi là quán tính, giúp cho thuật toán có tính khám phá, thoát ra khỏi các cực tiểu địa phương trong nhiều trường hợp.  Các thuật toán tối ưu theo hướng bầy đàn có đặc điểm chung là dễ hiểu, dễ cài đặt, xong việc chọn các tham số cho phù hợp để thuật toán chạy tốt là một thách thức đòi hỏi thử nghiệm thực tế.  Trên đây em đã trình bày về các kiến thức cần thiết để xây dựng và giải quyết bài toán trong mô hình của mình, Chương 4 em sẽ trình bày cụ thể về mô hình mạng và bài toán cũng như giải thuật tối ưu dựa trên phân tích các ưu nhược điểm của các giải thuật đã nêu trong chưng này. : [Giải thuật tối ưu hoá tần suất gửi tin trong mạng quan trắc di động](https://docs.google.com/document/d/1SP9Y7OJTTCS8ZPm5rw7S0d9UT5xR68S2X7okD_N91Y8/edit#heading=h.1ksv4uv) Mô tả bài toán Ta xét các xe bus hoặc xe khách trên đường, xe được tích hợp cảm biến đo chất lượng không khí với một chu kỳ cơ bản cực tiểu nào đó (ví dụ 1 s/1 lần đo, 10m/1 lần đo). Mỗi khi đo chất lượng không khí tại một điểm, nếu không có sự điều khiển nào khác cảm biến sẽ gửi kết quả về server để lưu trữ và xử lý kết quả. Bài toán coi rằng chất lượng không khí tại các điểm cách nhau không quá một khoảng và trong vòng thời gian thì không khác nhau đáng kể. Tức là nếu có một điểm được đo thì các điểm nằm trong khoảng bán kính và thời gian kể từ lúc đo, thì ta gọi điểm đó là điểm được bao phủ. Yêu cầu các xe phải thu thập một lượng dữ liệu đủ để bao phủ một vùng diện tích nào đó. Tuy nhiên nếu các xe thu thập quá nhiều dữ liệu tại các điểm gần nhau và gửi về server thì gây ra sự lãng phí về băng thông mạng, lãng phí bộ nhớ và khả năng tính toán của server, và một phần là lãng phí năng lượng của cảm biến. Vì vậy các xe cần đo với một chu kỳ hợp lý để tối đa hóa vùng bao phủ mà không làm lãng phí tài nguyên tính toán, lưu trữ và băng thông đường truyền. Đề xuất mô hình mạng Để đơn giản hóa bài toán, em chỉ xem xét tối ưu xác suất đo trong các khoảng thời gian rời rạc nhau.  Ví dụ với , em chia các khoảng thời gian trong ngày theo tiếng, với mỗi khoảng thời gian đó, em sẽ xác định được tần xuất theo dõi chất lượng không khí tối ưu cho các cảm biến.  Để điều chỉnh tần xuất gửi tin tối ưu, tại đầu vào của mỗi đoạn đường cần quan tâm chất lượng không khí, em đặt một máy chủ (ký hiệu là RSU ở Hình 6). Máy chủ này có nhiệm vụ là gửi đến các xe đi vào làn đường một tham số p sao cho , khi mỗi cảm biến đến chu kỳ cực đại lấy mẫu chất lượng không khí, nó sẽ sinh ngẫu nhiên một số , so sánh x với p, nếu thì ngay lập tức lấy mẫu và gửi về server. Ngược lại thì không tiến hành lấy mẫu mà đợi đến chu kỳ tiếp theo. Khi đó có thể coi chu kỳ tối ưu là  Vấn đề là cần một thuật toán tìm ra giá trị thích hợp của p.    Hình 6: Mô hình bài toán.  Tất nhiên để giải được bài toán này, ta cần giả sử rằng biến ngẫu nhiên là số điểm được lấy mẫu bới các cảm biến tích hợp trên xe ở chu kỳ đo cực tiểu tuân theo một phân phối xác suất nào đó. Trong bài toán của mình em giả sử biến ngẫu nhiên đó tuân theo phân phối Poisson đều trên mặt đường với hằng số tỉ lệ . Việc xác định có thể thu thâp đơn giản bằng phương pháp thống kê theo ngày trong khung giờ .  Bài toán này có thể mở rộng với k tuyến đường chạy song song nhau với các hằng số tỉ lệ khác nhau. Ở bài toán k làn đường là bài toán tối ưu k biến số , là xác suất gửi tin tối ưu ở tuyến đường thứ i.  **Vậy bài toán có đầu vào là các tham số như sau:**  Bán kính bao phủ của cảm biến.  : Khoảng thời gian mà chất lượng không khí tại một điểm không thay đổi quá nhiều.  k: số tuyến đường.  với Tham số tỉ lệ của phối Poisson ở tuyến đường thứ i.  Vùng D: vùng cần theo dõi.  Các thông số của k tuyến đường.  **Cần xác định:**  : xác suất gửi tin tối ưu ở tuyến đường thứ i. Xây dựng công thức và xác định hàm tối ưu Để xác định được hàm tối ưu, ta cần phải xác định được yếu tố có lợi và yếu tố chi phí trong bài toán này.  Đơn giản nhất, lợi ích sẽ được tính bằng kỳ vọng phần trăm diện tích bao phủ của vùng cần theo dõi.  Để đơn giản hóa phần trình bày, trong phần này em chỉ xét 1 làn đường với các điểm được lấy mẫu tuân theo phân phối Poisson với hằng số tỉ lệ , việc mở rộng ra nhiều làn đường là hoàn toàn tương tự.    Hình 7: Một điểm trong vùng cần theo dõi.  Xét một điểm A có tọa độ (x, y) (màu đen) trên vùng cần theo dõi D, giả sử các điểm màu vàng trên đường là các điểm được đo chất lượng không khí ở chu kỳ đo cực tiểu . Xét đường tròn tâm là điểm A, bán kính (màu xanh trong Hình 7) giao với mặt đường một phần có diện tích . Ta có lập luận như sau:  Do các điểm trên đường được đo chất lượng không khí tuân theo phân phối Poisson trên không gian nên xác suất điểm A được bao phủ chính là xác suất vùng không có điểm nào được đo chất lượng không khí rồi gửi về máy chủ. Giả sử xác suất gửi tin về server khí đó là p, ta có xác suất để không có điểm nào trong vùng được gửi về server là:  Công thức 13: Xác suất để một điểm không được bao phủ.  Trong đó:   * chính là xác suất có i điểm được đo nằm trong vùng ở chu kỳ đo cực tiểu * chính là xác suất để cả i điểm không được gửi về server.   Công thức này khá phức tạp và trong thực hiện tính toán rõ ràng ta không thể cho biến i chạy từ 0 tới . Từ Công thức 2, ta có thể thấy, thực chất phân phối của các điểm được gửi về server tuân theo phân phối Poisson với tham số tỉ lệ . Do vậy công thức trên có thể rút gọn thành:  Nếu ta chia vùng cần theo dõi D thành các ô vuông có diện tích rất nhỏ dxdy và xấp xỉ các ô vuông này với một điểm. Khi đó ta có kỳ vọng của vùng diện tích được theo dõi sẽ là:  Công thức 14: Kỳ vọng diện tích vùng được theo dõi.  Trong trường hợp tuyến đường đang xem xét là đường thẳng hoặc gần thẳng thì ta có thể coi tích phân trên chỉ phụ thuộc vào x chứ không phụ thuộc vào y do vậy, tích phân trên có thể được rút gọn thành tích phân 1 lớp. Rõ ràng tích phân trên là không thể tính được trực tiếp bằng Công thức 5 do hàm tính diện tích phức tạp và phụ thuộc vào nhiều tham số như , độ rộng của tuyến đường, số tuyến đường đang xét … Nên ta không thể tính được theo Công thức 5. Từ đó việc tính tích phân này phải sử dụng các phương pháp tính gần đúng ở mục 3.2. Khi mở rộng bài toàn với k tuyến đường không giao nhau (hoặc giao nhau không đáng kể), sử dụng tính chất nhân các xác suất độc lập thì ta chỉ cần thay bởi trong Công thức 14 ta có:  Công thức 15: Kỳ vọng diện tích vùng được theo dõi với k tuyến đường.  Trong đó:   * là tham số tỉ lệ của phân phối Poisson ở tuyến đường thứ i. * là xác suất gửi tin ở tuyến đường thứ i * là diện tích giao giữa đường tròn tâm A bán kính với làn đường thứ i.   Một nhiệm vụ tương đối khó khăn trong bài toán chính là tìm ra hàm chi phí phù hợp. Do chưa có sự triển khai thực tế nên thực chất hàm chi phí cũng chưa thực sự rõ ràng, tuy nhiên khi xây dựng hàm chi phí chắc chắn ta sẽ cần quan tâm đến các yếu tố sau:   * Khi xác suất gửi tin lớn thì chi phí có xu hướng càng lớn * Khả năng tính toán của server càng lớn thì tác động của đến hàm chi phí càng nhỏ (sự biến đổi của không ảnh hưởng nhiều đến sự biến đổi của hàm chi phí) * Khả năng lưu trữ dữ liệu của server càng lớn thì tác động của đến hàm chi phí càng nhỏ. * Băng thông của mạng truyền gói tin càng lớn thì tác động của đến hàm chi phí càng nhỏ. * Sự cân bằng giữa xác suất gửi tin giữa các tuyến đường.   Ngoài ra còn rất nhiều yếu tố phát sinh khi xây dựng hàm chi phí trên thực tế, thậm chí hàm chi phí trên thực tế có thể được tính toán bởi một mô hình trí tuệ nhân tạo.  Giả sử ta có được các tham số như khả năng tính toán, lưu trữ của server và băng thông của đường truyền gói tin được chuẩn hóa theo một quy tắc nào đó lần lượt là , là nhu cầu về độ cân bằng tần suất gửi tin, ta có hàm chi phí cơ bản cho k tuyến đường được xây dựng như sau:  Công thức 16: Hàm chi phí.  Lợi ích ta thu được chính là kỳ vọng diện tích vùng bao phủ. Do vậy em đề xuất hàm tối ưu có dạng như sau:  Công thức 17: Hàm tối ưu.  Ta cần tìm cực đại của hàm tối ưu này.  Kết quả thực nghiệm 5.3.4 cho thấy hàm tối ưu này khá phù hợp với thực tế. Tổng quan các hướng tiếp cận giải quyết bài toán Với hàm tối ưu không khả vi như trên ta thấy được các phương pháp tối ưu sử dụng tính chất hàm lồi, đạo hàm hay gradient không thể giải quyết tổng quát bài toán, đặc biệt trong thực tế nếu hàm chi phí là một hàm rời rạc hoặc được tính bởi mô hình trí tuệ nhân tạo.  Do đó các phương pháp tối ưu dạng meta-heuristic sẽ phù hợp với bài toán hơn, đặc biệt trong trường hợp này miền tối ưu khá đơn giản vì . Phân tích giải thuật bầy ong nhân tạo (Artificial Bee Colony-(ABC)) Giải thuật bầy ong nhân tạo được chứng minh phù hợp với các bài toán tối ưu với số chiều lớn. Điểm mạnh lớn nhất của thuật toán ABC chính là nó có cơ chế rời bỏ lời giải cũ nếu sau hữu hạn lần tìm kiếm lân cận mà không tìm ra lời giải tối ưu hơn. Đặc biệt ở Công thức 11, ta thấy được việc tìm kiếm lời giải mới được thực hiện trên một chiều, các chiều khác của lời giải được giữ nguyên, tuy nhiên công thức này không cho ta thấy được tính hội tụ của thuật toán, thực chất có tính ngẫu nhiên. Tính hội tụ chỉ thể hiện ở chỗ nếu tìm được lời giải mới tốt hơn lời giải cũ thì thực hiện cập nhật.  Thực tế cũng cho thấy thuật toán ABC kém thuật toán PSO ở khả năng khám phá, rất chậm trong việc tìm ra cực tiểu (địa phương hoặc toàn cục). Phân tích giải thuật bầy đàn PSO Ra đời trước giải thuật ABC, thuật toán ABC có điểm mạnh và điểm yếu riêng. Giải thuật PSO tương đối dễ hiểu, dễ cài đặt và tùy chỉnh do có ít tham số. Công thức 12 của thuật toán cho thấy, vector chuyển động của mỗi cá thể có xu hướng đi theo những giá trị tốt nhất của cả bầy đàn và của cá nhân. Điều này giải thích tại sao thuật toán PSO có tính hội tụ rất nhanh so với thuật toán ABC.   |  |  | | --- | --- | | Hình 8: Các cá thể ở bước lặp thứ 2/100. | Hình 9: Các cá thể ở bước lặp 22/100. |   Thử nghiệm thực tế với một hàm đơn giản, Hình 8 và Hình 9(vị trí các cá thể là các chấm màu đen, kèm theo vector chuyển động) cho thấy thuật toán hội tụ rất nhanh nhưng khi hội tụ tại cực tiểu địa phương lại không có cơ chế thoát ra khỏi nó. Hơn nữa Công thức 12 cũng cho thấy nếu một số cá thể hội tụ tại gBest, khi đó chúng sẽ gần như ở yên tại chỗ đó tại các bước lặp tiếp theo, trong khi các cá thể khác vẫn tích cực tìm kiếm lời giải mới. Điều này rõ ràng không tận dụng được tối đa khả năng khám phá của các cá thể. Đề xuất giải thuật cái tiến. Từ các điểm yếu và điểm mạnh của giải thuật PSO và ABC, em xin đề xuất một giải thuật tối ưu mới gọi là PSO-Based, phù hợp với bài toán của em, giải quyết được điểm yếu của cả 2 thuật toán đã nêu trên, đồng thời cũng tận dụng được điểm mạnh của chúng. Thuật toán sử dụng công thức cập nhật vector chuyển động:  Công thức 18: Cập nhật vector chuyển động.  Thuật toán với cách cập nhật vector chuyển động như trên, kế thừa và làm mạnh tính hội tụ của thuật toán PSO.  Sử dụng thêm tham số , cứ sau vòng lặp, thuật toán tiến hành kiểm tra xem vector chuyển động của các cá thể, nếu tổng bình phương của các thành phần trong vector đó bé hơn một đủ nhỏ, thuật toán sẽ tiến hành khởi tạo lại ngẫu nhiên một trong số các thành phần trong vector chuyển động đó. Điều này khắc phục được điểm yếu của thuật toán PSO và tận dụng điểm mạnh của thuật toán ABC.  Giả sử ta đã có thuật toán tính hàm tối ưu F: .  Ta cần chọn các tham số đầu vào phù hợp như dưới đây.  **Thuật toán: PSO – Based.**  **Input:**   * L-số bước lặp. * – số thực dương đủ nhỏ. * – số bước lặp định kỳ kiểm tra vector vận tốc của các cá thế. * – w trong Công thức 18. * - trong Công thức 18. * – chặn trên và dưới của vận tốc và vị trí.   **Output:** - xác suất gửi tin tối ưu ở các tuyến đường.  gBest ; gBestValue 0.  **For** i **from** 0 to nbParticle **do**:  # khởi tạo quần thể là các vector k chiều ngẫu nhiên.  # khởi tạo vector chuyển động k chiều  **If**  gBestValue  gBest  **End** **if**  **End for**  **For** iteration **from** 1 **to** L **do**:    **For** i **from** 1 **to** nbParticle **do**:  **If** iteration % = 0 **then**:  **If**  for j  # Khởi tạo lại vector vận tốc  **End** **if**  **End** **if**  **For** j **from** 1 **to** k **do**:          **End for**  **If**  gBestValue  gBest  **End for**  **End for**  **Return** gBest.   Đánh giá thực nghiệm. Trong hai chương trước, mô hình đề xuất đã được mô tả chi tiết về cơ sở lý thuyết cũng như cách thức hoạt động của mô hình. Khó khăn lớn nhất khi triển khai thực nghiệm là chưa có dữ liệu thực tế, cũng như chưa có nghiên cứu thực sự liên quan để so sánh hiệu quả của thuật toán. Tại chương này em xin được trình bày các tham số mà em cho rằng tương đối phù hợp với thực tế, phương pháp thực nghiệm để so sánh được hiệu quả của thuật toán do em đề xuất so với thuật toán ABC và PSO đã được trình bày bên trên. Các tham số đánh giá Do mô hình em đề xuất sẽ hoạt động hiệu quả nhất trên các phương tiện có lịch trình di chuyển theo ngày như xe bus hoặc xe khách. Ta có thể coi các tuyến đường mà các xe này đi qua là thẳng. Em thử nghiệm trên các tham số như sau:  K – số tuyến đường :5.  – có thể coi là bán kính bao phủ của cảm biến: 100 mét.  D-Vùng cần theo dõi: là toàn bộ vùng có thể được theo dõi bởi các xe trên các tuyến đường (dọc theo chiều dài của đường, mở rộng về cả hai phía trái của làn đường thứ nhất và phải của làn đường cuối cùng một khoảng ).  Độ dài quãng đường: 10 kilo mét.   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | Làn đường  Thông số | 1 | 2 | 3 | 4 | 5 | | Độ rộng(m) | 20 | 25 | 15 | 30 | 30 | | Khoảng cách với làn tiếp theo(m) | 60 | 95 | 35 | 90 | N/A | | Hằng số tỉ lệ | 0.0008 | 0.0002 | 0.0005 | 0.003 | 0.00004 |   Bảng 1: Thông số của các làn đường.  Với hàm chi phí trong Công thức 16, em chọn các tham số như sau:  Em coi vai trò của 3 biến ở đây là như nhau, các tham số này được dùng để so sánh khả năng tối ưu của 3 thuật toán bầy đàn em đã nêu.  Với thuật toán ABC, PSO và PSO-Based, việc chọn các tham số sao cho thuật toán tối ưu là một công việc đòi hỏi rất nhiều thời gian thực nghiệm tùy vào các tham số đầu vào của hàm số cần tối ưu. Với thuật toán PSO theo [2], trong Công thức 12 ta nên sử dụng , tốt nhất nên cho w giảm dần theo các bước lặp, trong rất nhiều bài báo đều được chọn bằng nhau và bằng 2, em sẽ thí nghiệm với 0.5, 1, 1.5, 2 và 3 để tìm ra tham số tốt nhất. Do bài toán đang xét có số chiều của không gian tìm kiếm lời giải tương đối nhỏ, em sử dụng số lượng cá thể trong thuật toán là 10. Tương ứng với mỗi bộ tham số của thuật toán PSO em sử dụng bộ tham số giống như vậy để đánh giá thuật toán PSO-Based. Với thuật toán ABC, em sử dụng bộ tham số của [1] để đánh giá, số quần thể tương tự như thuật toán PSO là 10 ong thợ và 10 ong quan sát. Tuy nhiên do thuật toán ABC có số quần thể đông gấp đôi thuật toán PSO nên số lượng bước lặp sẽ giảm đi một nửa để đảm bảo tính công bằng.  Sau khi thực nghiệm tìm ra các tham số tối ưu cho thuật toán, em tiến hành so sánh cụ thể hơn 3 thuật với các tham số của hàm tối ưu thay đổi, từ đó đưa ra kết luận về thuật toán đề xuất.  Sau khi xác định được thuật toán phù hợp, sẽ thử nghiệm với nhận giá trị nhỏ tới lớn để chứng minh rằng hàm chi phí này phù hợp với thực tế. Phương pháp thí nghiệm Để tính được hàm tối ưu, trước tiên ta phải tính được kỳ vọng diện tích vùng bao phủ với bộ tham số trên và một bộ tham số xác suất truyền tin nhất định nào đó. Để tính được kỳ vọng diện tích này, ta phải sử dụng một trong các phương pháp tính tích phân gần đúng đã nêu ở mục 2. Việc chọn phương pháp phù hợp cũng như tham số thích hợp cho thuật toán có tính quyết định đến tốc độ và độ chính xác của thuật toán. Do vậy em sẽ thử nghiệm kết quả tính tính phân gần đúng với các phương pháp khác nhau cũng như tham số trong mục 3.2 từ nhỏ đến lớn để tìm ra giá trị tối ưu.  Có được tham số cho thuật toán tính gần đúng tích phân, em thử nghiệm trên 3 thuật toán bầy đàn đã nêu để so sánh tính hội tụ và khả năng tìm ra lời giải tối ưu. Kết quả thí nghiệmTham số tối ưu cho phương pháp tính gần đúng tích phân Do bài toán đang xét trên các đường thẳng và miền D là một hình chữ nhật nên tích phân 2 chiều trong Công thức 14 có thể thu gọn về một chiều theo trục x cắt ngang vuông góc với chiều dọc của đường. Để cho kết quả gọn và dễ hiểu, em chia kỳ vọng diện tích này trên tổng diện tích của miền D để lấy phần trăm diện tích bao phủ, và chọn .  Trong phần này em sẽ trình bày kết quả thực nghiệm của 3 phương pháp hay được sử dụng nhất trong tính gần đúng tích phân là phương pháp sử dụng tổng Reimann giữa mục 3.2.1, phương pháp sử dụng luật hình thang mục 3.2.2 và phương pháp Simpson mục 3.2.3.   |  |  |  |  | | --- | --- | --- | --- | | Các phương  pháp  (mm) | Tổng Reimann giữa | Luật hình thang | Simpson | | 1 | 0.70680119504286 | 0. 70680123199479 | 0. 70680121967762 | | 10 | 0.70680119505433 | 0. 70680156454747 | 0. 67697487601425 | | 100 | 0.70680120038300 | 0. 70680488233460 | 0. 67697615844060 | | 1000 | 0.70680307505950 | 0. 70683528218977 | 0. 67700399275486 | | 10000 | 0.70789261297033 | 0. 70572208319541 | 0. 67377801496309 |   Bảng 2: Kết quả tính gần đúng tích phân.  Bảng 2 là kết quả thu được khi thực hiện các thuật toán tính tích phân gần đúng với các bước nhảy khác nhau. Em nhận thấy nếu = 1mm thì kết quả cho ra ở cả 3 phương pháp có tới 8 chữ số đầu tiên giống nhau, có thể nói rằng nếu sử dụng tham số như vậy thì ta có kết quả chính xác gần như tuyệt đối, em sẽ lấy kết quả này làm cơ sở để kiểm tra các kết quả khác. Với lớn hơn ta thấy rằng phương pháp Simpson cho ra sai số lớn nhất, và phương pháp tổng Reimann giữa vẫn có độ chính xác ở mức tốt hơn so với các phương pháp còn lại. Yêu cầu còn lại là tìm ra đủ để cho thuật toán chạy nhanh hơn mà độ chính xác ở mức chấp nhận được. Từ bảng trên em chọn cho thuật toán của mình với độ chính xác ở mức 7 chữ số, mà thuật toán lại chạy đủ nhanh. Việc sử dụng thuật toán meta-heuristic tốn khá nhiều bước lặp, nếu chọn quá bé thì thời gian tìm ra lời giải tối ưu rất lâu mà bản chất của bài toán là ta đã mô hình theo lý thuyết xác suất, nên sai số là điều không thể tránh khỏi.  Kết luận: em sẽ sử dụng tổng Reimann giữa và trong các phần tính toán tiếp theo. Các tham số tối ưu cho thuật toán PSO, ABC và PSO-Based Hàm tối ưu có biên độ nhỏ, do vậy để làm nổi bật sự tối ưu giữa các thuật toán, các kết quả trong mục này đều là sự chênh lệch giữa kết quả tối ưu được và phương pháp cố định xác suất với (0.35, 0.35, 0.35, 0.35, 0.35).  Trước tiên em thực hiện so sánh kết quả của giải thuật PSO với thay đổi và .  Bảng dưới đây thu được sau 5 lần chạy với mỗi bộ tham số tương ứng.   |  |  |  |  | | --- | --- | --- | --- | | Kết quả | Tốt nhất | Tồi nhất | Trung bình | | 0.5 | 7.571627184 | 4.101984475 | 5.634429362 | | 1 | 7.557415157 | 3.426880523 | 5.8530497632 | | 1.5 | 7.216317625 | 5.658620734 | 6.78725651 | | 2 | 7.581799259 | 6.206881385 | 7.0877852238 | | 3 | 6. 686236731 | 5. 450629562 | 6.030490932 |   Bảng 3: Kết quả tối ưu với thuật toán PSO.  Từ Bảng 3, em nhận thấy với nhỏ, thuật toán có tính ổn định không cao, kết quả cho ra tại các lượt chạy khác nhau có sự sai khác lớn. Với thuật toán có tính ổn định cao nhất đồng thời cũng cho ra kết quả tốt nhất. Xong với giá trị lớn, thuật toán lại không có khả năng hội tụ thực sự tốt mặc dù tính ổn định cao.  Cụ thể, với các tham số cho thuật toán PSO thay đổi em thu được kết quả tối ưu qua từng bước lặp như các hình dưới đây.   |  |  | | --- | --- | | Hình 10: . | Hình 11:. | | Hình 12 . | Hình 13 . |   Hình 10 mô tả kết quả thu được qua các bước lặp, nó cho thấy khả năng hội tụ của thuật toán rất nhanh (ở khoảng iteration 100) nhưng lại không có khả năng thoát ra khỏi cục bộ địa phương. Khả năng hội tụ của thuật toán giảm dần theo tứ tự Hình 11, Hình 12, Hình 13 đồng thời khả năng khám phá lại tăng dần lên, và sự cân bằng giữa 2 yếu tố này đạt được ở Hình 12.  Do vậy, em sẽ sử dụng tham số của Hình 12 để so sánh với thuật toán ABC và thuật toán PSO-Based. Ở bước so sánh này với thuật toán PSO-Based em cũng chọn w = 0.8 và , số quần thể là 10 và số bước lặp là 1000, số bước kiểm tra định kỳ là 10 và . Với thuật toán ABC số quần thể tổng cộng là 20 (10 ong thợ và 10 ong quan sát), số bước lặp là 500. Các thuật toán đều đc chạy 10 lần để thu được kết quả như bảng dưới đây.   |  |  |  |  | | --- | --- | --- | --- | | Kết quả  Thuật toán | Tốt nhất | Tồi nhất | Trung bình | | ABC | 2.02376 | -3.71055 | -0.42210908 | | PSO | 7.581799259 | 6.206881385 | 6.95914585809 | | PSO-Based | 7.593439111 | 7.592501434 | 7.5927583306 |   Bảng 4: Kết quả thực nghiệm với 3 thuật toán.  Ta nhận thấy thuật toán ABC kém hoàn toàn so với hai thuật toán còn lại, trong nhiều trường hợp thuật toán ABC cho kết quả kém hơn phương pháp cố định xác suất.  Thuật toán PSO-Based có tính ổn định cao hơn hẳn thuật toán PSO và cũng tìm ra giá trị tốt hơn trong tất cả trường hợp. Điều này thực hiện được là do thuật toán PSO-Based có tính cân bằng giữa hai yếu tố là khám phá và khai thác. Đồng thời có cơ chế thoát ra khỏi cực tiểu địa phương.   |  |  | | --- | --- | | Hình 14: Kết quả thuật toán PSO-Based. | Hình 15: Kết quả thuật toán PSO. |   Hình 14, qua các bước lặp, cho thấy thuật toán PSO-Based có khả năng khám phá và thoát khỏi cực tiểu địa phương rất tốt, những bước lặp gần cuối thuật toán vẫn có đường đồ thị đi lên. Hình 15 cho thấy điểm yếu rõ ràng của thuật toán PSO là không thể thoát ra khỏi cực tiểu địa phương, kể từ bước lặp 300 tới 1000, thuật toán không thể tìm ra lời giải tối ưu hơn.  Đồng thời ta cũng thấy được thuật toán PSO-Based có khả năng hội tụ tốt hơn, khi ở bước lặp thứ 100 đã đạt được kết quả ở ngưỡng 7.25 so với khoảng 4.5 của thuật toán PSO. Do vậy trong trường hợp cần ít bước lặp hơn, giải thuật đề xuất cũng sẽ cho kết quả tối ưu hơn rất nhiều. So sánh kết quả 3 giải thuật PSO, PSO-Based và ABC. Để so sánh rõ hơn 3 thuật toán với các tham số đã tìm được ở phần 5.3.2, em tiến hành thử nghiệm trên 8 làn đường với các thông số như sau:   |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | | Làn đường  Thông số | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | | Độ rộng(m) | 20 | 25 | 15 | 30 | 20 | 30 | 20 | 30 | | Khoảng cách tới làn tiếp theo(m) | 60 | 95 | 35 | 60 | 100 | 50 | 110 | N/A | | Hằng số tỉ lệ | 0.008 | 0.0002 | 0.0005 | 0.003 | 0.00004 | 0.00007 | 0.009 | 0.0002 |   Bảng 5: Thông số của các tuyến đường dùng để so sánh 3 giải thuật.  Mỗi thuật toán sẽ được chạy 5 lần và lấy kết quả trung bình.  Cụ thể với hoặc thay đổi và so sánh kết quả thu được giữa hai thuật toán.  Với và C thay đổi ta thu được kết quả như dưới đây:    Hình 16: Kết quả tối ưu của 3 thuật toán khi C thay đổi.  Ta nhận thấy với C - khả năng xử lý của server và băng thông đường truyền càng nhỏ thì kết quả tối ưu của thuật toán đề xuất vượt trội so với 2 giải thuật còn lại. Cụ thể, trong trường hợp C = 0.1, thuật toán đề xuất cho ra kết quả tốt hơn khoảng 110% so với thuật toán PSO và khoảng 400% so với thuật toán ABC. Khi C lớn, cả 3 thuật toán đều có khả năng tối ưu tương đương nhau nếu khả năng xử lý của server tăng lên.  Với C = 0.2 cố định và , x thay đổi ta được kết quả như sau:    Hình 17: Kết quả tối ưu của 3 thuật toán khi x thay đổi.  Hình 17 thể hiện rõ sự vượt trội của Thuật toán PSO và PSO-Based so với ABC.  Kết quả cho thấy khi x tăng lên thì kết quả của giải thuật đề xuất có xu hướng vượt trội so với hai giải thuật còn lại, điều này có thể lý giải được bởi vì khi x nhỏ thì hàm tối ưu có xu hướng có ít cực đại địa phương hơn, vì thế thuật toán PSO vẫn có khả năng tìm kiếm ra cực đại toàn cục. Khi x lớn, hàm số xuất hiện nhiều cực tiểu địa phương và thuật toán PSO không có cơ chế thoát ra khỏi các cực tiểu đó, chính là khi PSO-Based thể hiện được khả năng của mình. Ảnh hưởng của các tham số trong hàm chi phí tới kết quả tối ưu Sử dụng thuật toán PSO-Based, trong phần này em sẽ tiến hành thử nghiệm các tham số ở các ngưỡng khác nhau và so sánh kết quả. Do có vai trò như nhau, ta chỉ xét tham số:  Thử nghiệm đầu tiên em sẽ xét C = 6 cố định và thay đổi.   |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | | Kết quả  tối ưu |  |  |  |  |  | Phần trăm diện tích bao phủ | | 0 | 0.86275 | 0.42752 | 0 | 0.37974 | 0 | 59,010 | | 0.1 | 0.64470 | 0.35687 | 0.35687 | 0.35687 | 0.35687 | 59,490 | | 0.2 | 0.48908 | 0.48908 | 0.48908 | 0.48908 | 0.48908 | 59,764 |   Bảng 6: Kết quả tối ưu khi thay đổi.  Rõ ràng khi tăng lên thì các xác suất gửi tin ở các tuyến đường ngày càng có xu hướng cân bằng, xong phần trăm diện tích bao phủ giữa các sự thay đổi là không chênh lệch đáng kể. Hơn nữa ta cũng thấy được rằng, trong trường hợp = 0 thì có , điều này là do tuyến đường thứ 4 có hằng số tỉ lệ vượt trội so với hai làn còn lại chính vì đó các cảm biến thuộc làn đó bao phủ luôn một phần diện tích của hai làn bên cạnh. Chính vì vậy mới đặt ra nhu cầu về tham số đảm bảo độ cân bằng giữa các tuyến đường .  Tiếp theo em sẽ cố định và cho C thay đổi.   |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | | Kết quả  tối ưu    C |  |  |  |  |  | Phần trăm diện tích bao phủ | | 0.001 | 0 | 0 | 0 | 0.00858 | 0 | 3,117 | | 1 | 0.21949 | 0 | 0 | 0.19653 | 0 | 38,598 | | 10 | 0.78814 | 0.51006 | 0.51006 | 0.51006 | 0.51006 | 63,906 | | 100 | 1 | 1 | 1 | 1 | 1 | 70,680 |   Bảng 7: Kết quả tối ưu khi C thay đổi.  Khi hằng số C- đặc trưng cho khả năng xử lý, lưu trữ và băng thông càng nhỏ thì ở kết quả tối ưu các biến cũng có xu hướng nhỏ đi và ngược lại. Hơn nữa khi C nhỏ, ưu tiên đầu tiên của kết quả tối ưu là tối thiểu hóa các chứ không phải đảm bảo việc cân bằng giữa các tuyến đường, khi C lớn, ưu tiên dần chuyển sang cả việc đảm bảo cân bằng xác suất gửi tin giữa các làn đường. Những điều này chứng tỏ rằng hàm chi phí do em đề xuất hoàn toàn phù hợp với thực tiễn.  Với các kết quả thí nghiệm thực tế như trên, em đã hoàn thành việc xây dựng mô hình, tìm ra các tham số và giải thuật tối ưu cho việc tính toán tối ưu vùng bao phủ dưới rằng buộc cân nhắc các yếu tố được coi là chi phí. Ở chương cuối cùng em xin rút ra kết luận của mình về bài toán sau đó đề xuất các hướng nghiên cứu có thể mở rộng. : Kết luận Đây là chương cuối cùng của đồ án, em sẽ tổng hợp ngắn gọn và kết luận về toàn bộ đồ án và đề xuất hướng phát triển. Kết luận Tổng kết lại, em đề xuất một mô hình toán học cho nhu cầu theo dõi chất lượng không khí. Mô hình này hoàn toàn có thể mở rộng một cách tương tự với sự tinh chỉnh nhẹ các thông số sang mục tiêu theo dõi độ ẩm, nhiệt độ hay các yếu tố môi trường khác. Mô hình của em có ưu điểm vượt trội so với các mô hình trước đây là không cần sự can thiệp quá nhiều của con người mà các quá trình lấy mẫu được thực hiện hoàn toàn tự động. Các cảm biến tích hợp trên xe cũng không cần giao tiếp với nhau mà chỉ quan tâm đến tần xuất tối ưu của mình. Em cũng đã xây dựng được hàm chi phí và hàm tối ưu dựa trên chi phí đó, cuối cùng đề xuất một giải thuật tối ưu có kết quả tốt với các tham số thực nghiệm khác nhau. Phương pháp tối ưu của em lấy ý tưởng chính từ các giải thuật nổi tiếng đã có từ trước, khắc phục một số điểm yếu và tận dụng được các điểm mạnh của chúng từ đó cho ra kết quả tốt hơn trong nhiều trường hợp.  Tuy nhiên, với vai trò là một sinh viên, sức lực và trí tuệ còn nhiều hạn chế, em cũng nhận thấy một số điểm yếu trong đồ án của mình:   * Sử dụng mô hình phân phối xác suất chưa chắc đã đúng trong thực tế * Thí nghiệm trên các đường thẳng mà chưa có dữ liệu thực với các tuyến đường cong hoặc chéo nhau. * Hàm chi phí cân nhắc chưa hết các yếu tố trên thực tế. * Vùng bao phủ được coi là có độ quan trọng ngang nhau giữa các điểm hoặc các vùng nhỏ.   Từ các điểm yếu trên em xin đề xuất một số hướng phát triển trong phần tiếp theo. Định hướng phát triển của đề tài Hiện nay thuật toán chạy rất tốt và ổn định với môi trường thí nghiệm là các tuyến đường thẳng và có bề rộng cố định. Trên thực tế các tuyến đường không đảm bảo các yếu tố như vậy. Các công thức và mô hình toán học vẫn hoàn toàn chính xác với các tuyến đường tổng quát nhưng đặt ra một số thách thức nhất định về thời gian chạy của thuật toán (trong việc tính tích phân 2 lớp và phần diện tích vùng bao phủ của cảm biến giao với mặt đường), từ đó đặt ra hướng phát triển tối ưu độ phức tạp về thời gian của thuật toán.  Mặt khác, khi các điểm được đo chất lượng không khí không tuân theo phân phối Poisson, ta cần có một cơ chế giao tiếp giữa các xe để đảm bảo đo chất lượng không khí một cách hợp lý. Vậy một hướng nghiên cứu khả thi là sử dụng các mô hình học máy hoặc học sâu để giải quyết bài toán trong trường hợp này. Tài liệu tham khảo Lưu ý: Sinh viên không được đưa bài giảng/slide, các trang Wikipedia, hoặc các trang web thông thường làm tài liệu tham khảo.  Một trang web được phép dùng làm tài liệu tham khảo **chỉ khi** nó là công bố chính thống của cá nhân hoặc tổ chức nào đó. Ví dụ, trang web đặc tả ngôn ngữ XML của tổ chức W3C <https://www>.w3.org/TR/2008/REC-xml-20081126/ là TLTK hợp lệ.  Có năm loại tài liệu tham khảo mà sinh viên phải tuân thủ đúng quy định về cách thức liệt kê thông tin như sau. Lưu ý: các phần văn bản trong cặp dấu < > dưới đây chỉ là hướng dẫn khai báo cho từng loại tài liệu tham khảo; sinh viên cần xóa các phần văn bản này trong ĐATN của mình.  <**Bài báo đăng trên tạp chí khoa học**: Tên tác giả, tên bài báo, tên tạp chí, volume, từ trang đến trang (nếu có), nhà xuất bản, năm xuất bản>   1. Dervis Karaboga and Bahriye Basturk, Artiﬁcial Bee Colony (ABC) Optimization Algorithm for Solving Constrained Optimization Problems, IEEE Swarm Intelligence Symposium 2006, Indianapolis, Indiana, USA, 2006. 2. Yan Hea, Wei Jin Ma and Ji Ping Zhang, The Parameters Selection of PSO Algorithm influencing on performance of Fault Diagnosis, MATEC Web of conferences, 2016. 3. Jiaoyan Chen, Jingsen Yang, Maximizing Coverage Quality with Budget Constrained in Mobile Crowd-Sensing Network for Environmental Monitoring Applications, Sensors, 2019. 4. Yingying Ren el al. An intelligent big data collection technology based on micro mobile data centers for crowdsensing vehicular sensor network, Personal and Ubiquitous Computing, 2020. 5. R Liu, J Pan, AirQ: A Privacy-Preserving Truth Discovery Framework for Vehicular Air Quality Monitoring, International Conference on Mobile Ad-hoc and Sensor Networks, MSN, 2020. 6. Viet-Dung Nguyen, Phi Le Nguyen∗, Trung Hieu Nguyen, Phan Thuan Do∗,A 1/2 -Approximation Algorithm for Target Coverage Problem in Mobile Air Quality Monitoring Systems, IEEE, 2020.   <**Sách**: Tên tác giả, tên sách, volume (nếu có), lần tái bản (nếu có), nhà xuất bản, năm xuất bản>   1. Tống Đình Quỳ, Giáo trình Xác Suất Thống Kê, Nhà xuất bản Bách Khoa, 2007. 2. Trần Bình, Giải tích I, Nhà xuất bản Bách Khoa, 2015.   <**Tập san Báo cáo Hội nghị Khoa học**: Tên tác giả, tên báo cáo, tên hội nghị, ngày (nếu có), địa điểm hội nghị, năm xuất bản>   1. Joy Dutta, Towards Smart City: Sensing Air Quality in City based on Opportunistic Crowd-sensing, ICDCN 2017. 2. James Kenedy, Russell Eberhart, Particle Swarm Optimization, Proceedings of ICNN’95-Internationcal Conference on Neural Networks, 1995   <**Tài liệu tham khảo từ Internet**: Tên tác giả (nếu có), tựa đề, cơ quan (nếu có), địa chỉ trang web, thời gian lần cuối truy cập trang web>   1. MIT, Spatial Poisson Proccess, <https://web.mit.edu/urban_or_book/www/book/chapter3/3.8.html> | | |
|  | | |
|  | | |