**ST-ACTS: Trình mô phỏng hoạt động không gian-thời gian**

|  |  |
| --- | --- |
| Gyozo Gidofalvi  Aps địa lý  Trung tâm định dạng địa lý  Gyg@geomatic.dk | Torben Bach Pedersen  Đại học Aalborg  Khoa khoa học máy tính  Tbp@cs.aau.dk |

**ABSTRACT**

Việc tạo ra các mô hình mô phỏng không gian - thời gian phức tạp là một vấn đề nóng bỏng trong lĩnh vực cơ sở dữ liệu không gian - thời gian [7]. Trong khi các Trình mô phỏng Đối tượng Chuyển Động (MOSs) hiện có giải quyết các khía cạnh *vật lý* khác nhau của tính di động, chúng lại bỏ qua các khía cạnh *địa lý-nhân khẩu học* và *xã hội* quan trọng của nó. Bài báo này trình bày ST-ACTS, một Trình mô phỏng hoạt động theo không gian-thời gian, sử dụng các nguồn dữ liệu thống kê địa lý khác nhau và các nguyên tắc trực quan, mô hình hóa các khía cạnh bị bỏ quên cho đến nay. ST-ACTS cho rằng (1) các đối tượng (đại diện cho người dùng di động) di chuyển từ vị trí không gian-thời gian này sang vị trí khác với mục tiêu thực hiện một hoạt động nhất định tại vị trí sau; (2) không phải tất cả người dùng đều có khả năng thực hiện một hoạt động nhất định như nhau; (3) các hoạt động nhất định được thực hiện tại các địa điểm và thời gian nhất định; và (4) các hoạt động thể hiện các quy định có thể dành riêng cho một người dùng đơn lẻ hoặc cho các nhóm người dùng. Kết quả thử nghiệm cho thấy ST-ACTS có thể tạo ra các hoạt động phân bố không gian-thời gian thực tế một cách hiệu quả, điều này làm cho nó trở nên cần thiết cho sự phát triển của các kỹ thuật khai thác dữ liệu và quản lý dữ liệu theo không gian-thời gian.

**Trình mô tả danh mục và chủ đề**

H.2.8 [**Hệ thống thông tin**]: Ứng dụng cơ sở dữ liệu; I.6.3 [**Mô phỏng và mô hình hóa**]: Ứng dụng; I.6.8 [**Mô phỏng và mô hình hóa**]: Các loại mô phỏng — *Discrete event.*

**Điều khoản chung**

Algorithms

**Từ khóa**

spatio–temporal data(dữ liệu không gian-thời gian), data generation(tạo dữ liệu), moving object simulation(mô phỏng đối tượng chuyển động), activity simulator(mô phỏng hoạt động), data mining(khai thác dữ liệu).

1. **GIỚI THIỆU**

Mô phỏng được chấp nhận rộng rãi trong nghiên cứu cơ sở dữ liệu như một phương pháp chi phí thấp để cung cấp dữ liệu tổng hợp cho việc thiết kế và thử nghiệm các kiểu dữ liệu mới và các phương pháp truy cập. Cơ sở dữ liệu đối tượng chuyển động là một trường hợp cụ thể của cơ sở dữ liệu đại diện và quản lý các thay đổi liên quan đến chuyển động của các đối tượng. Để hỗ trợ sự phát triển trong nghiên cứu cơ sở dữ liệu đối tượng chuyển động, một số Trình mô phỏng Đối Tượng Chuyển Động (MOSs) đã được phát triển [1, 5, 6, 8, 9, 11].

Các MOSs được phát triển cho đến nay đã sử dụng các chức năng ngẫu nhiên có thể tham số hóa và mạng lưới đường để mô hình hóa các khía cạnh vật lý khác nhau của các đối tượng chuyển động - chẳng hạn như phạm vi, môi trường và tính di động của chúng - nhưng tất cả đều bỏ qua một số dữ kiện quan trọng. Khi các đối tượng chuyển động đại diện cho người dùng di động, phần lớn lý do chuyển động là do một mục tiêu rõ ràng. Cụ thể, người dùng di chuyển từ vị trí không gian-thời gian này sang vị trí khác để hoàn thành một số nhiệm vụ, từ đây được gọi là thực hiện một hoạt động, tại vị trí sau này. Ví dụ, mọi người không chỉ dành hầu hết các đêm của họ tại một địa điểm cụ thể, mà họ trở về *nhà* để ở với những người thân yêu của họ, để thư giãn, ăn và ngủ. Tương tự như vậy, mọi người không chỉ dành phần lớn thời gian làm việc của họ ở bất kỳ địa điểm cụ thể nào, mà họ đến cơ sở thực tế, *nơi làm việc* của họ, với ý định làm việc. Cuối cùng, dựa trên thói quen và sở thích của họ, trong thời gian rảnh rỗi, mọi người (ít hoặc nhiều thường xuyên) đi đến các cơ sở thực tế khác mà họ thích và ở gần đó.

Để mô hình hóa các khía cạnh xã hội được đề cập ở trên của tính di động là quan trọng vì hai lý do. Đầu tiên, các vị trí và thời gian mà các hoạt động có thể được thực hiện và các mẫu trong các hoạt động được thực hiện này xác định sự phân bố không gian-thời gian duy nhất của các đối tượng chuyển động cần thiết cho việc quản lý cơ sở dữ liệu không gian-thời gian. Thứ hai, các khía cạnh xã hội của tính di động là cần thiết khi một người muốn thu thập kiến thức không gian - thời gian về các quy luật trong hành vi của người dùng di động. Lĩnh vực khai thác dữ liệu theo không gian-thời gian liên quan đến việc tìm kiếm các quy luật hoặc thời điểm này. Để phát triển các kỹ thuật quản lý dữ liệu không gian - thời gian và khai thác dữ liệu hiệu quả, cần có những bộ dữ liệu không gian - thời gian lớn; và trong khi các thiết bị đầu cuối di động hỗ trợ định vị ngày càng có sẵn trên thị trường, các tập dữ liệu như vậy không có sẵn.

Do đó, để hỗ trợ sự phát triển trong quản lý dữ liệu không gian-thời gian và các kỹ thuật khai thác dữ liệu, bài báo này trình bày ST-ACTS, một trình mô phỏng hoạt động không gian-thời gian, có thể tham số hóa, dựa trên một số nguồn dữ liệu trong thế giới thực bao gồm:

* Dân số địa lý chi tiết.
* Thông tin về các doanh nghiệp và cơ sở.
* Khảo sát người tiêu dùng liên quan.

Tầm quan trọng của việc sử dụng các nguồn dữ liệu thế giới thực trong ST-ACTS nằm ở chỗ, chúng tạo thành một cơ sở thực tế để mô phỏng. Cụ thể, các biến trong bất kỳ nguồn dữ liệu nhất định nào đều phụ thuộc và có lẽ quan trọng nhất là phụ thuộc địa lý. Ví dụ, có một sự phụ thuộc mạnh mẽ giữa trình độ học vấn và thu nhập cá nhân của người dân. Các biến cũng phụ thuộc địa lý, do thực tế là những người tương tự hoặc các doanh nghiệp tương tự có xu hướng hình thành các cụm trong không gian địa lý. Hơn nữa, các biến phụ thuộc địa lý trên các nguồn dữ liệu khác nhau. Ví dụ, những người làm việc trong lĩnh vực công nghệ sinh học có xu hướng cố gắng tìm nhà gần nơi làm việc trong chi nhánh kinh doanh đó. Sử dụng dữ liệu thế giới thực từ các cơ sở dữ liệu thống kê địa lý thương mại khác nhau và các nguyên tắc thông thường, ST-ACTS nắm bắt một số đặc điểm cho đến nay chưa được mô hình hóa, nhưng quan trọng, của dữ liệu hoạt động thời gian theo không gian.

Phần còn lại của bài viết này được tổ chức như sau. Phần 2 đánh giá công việc liên quan. Phần 3 xác định các mục tiêu của mô hình mô phỏng. Phần 4 mô tả chi tiết dữ liệu nguồn tạo cơ sở cho mô hình mô phỏng. Phần 5 mô tả từng thành phần của trình mô phỏng và cách dữ liệu nguồn được sử dụng trong từng thành phần.. Phần 6 đánh giá mô hình mô phỏng về mặt hiệu quả và mục tiêu mô phỏng của nó bằng cách xem xét các đặc điểm của một số dữ liệu mô phỏng. Cuối cùng Phần 7 kết luận và chỉ ra các hướng nghiên cứu trong tương lai.

1. **CÔNG VIỆC LIÊN QUAN**

Do lịch sử quản lý dữ liệu không gian-thời gian ngắn, công trình khoa học về mô phỏng không gian-thời gian có thể bị giới hạn trong một số ít các ấn phẩm. Trình mô phỏng không gian-thời gian quan trọng đầu tiên là GSTD (Tạo dữ liệu không gian-thời gian) [11]. Bắt đầu với sự phân bố các điểm hoặc các đối tượng hình chữ nhật, tại mỗi bước GSTD sẽ tính toán lại các thay đổi về vị trí và hình dạng của các đối tượng dựa trên các hàm ngẫu nhiên được tham số hóa. Thông qua việc giới thiệu một tham số mới để kiểm soát sự thay đổi hướng và sử dụng các đối tượng hình chữ nhật để mô hình các chướng ngại vật, GSTD được mở rộng để mô phỏng các chuyển động thực tế hơn, chẳng hạn như *chuyển động ưa thích, chuyển động nhóm* và *chuyển động bị cản trở* [6]. Vì hầu hết các đối tượng sử dụng mạng để đi từ vị trí này đến vị trí khác, Brinkhoff trình bày một khuôn khổ để mô phỏng đối tượng chuyển động dựa trên mạng [1]. Hành vi của một đối tượng chuyển động trong khuôn khổ này bị ảnh hưởng bởi (1) các thuộc tính của đối tượng có một lớp đối tượng cụ thể, (2) tác động tổng hợp của vị trí của các đối tượng khác và dung lượng mạng, và (3) vị trí của các đối tượng bên ngoài độc lập với mạng. Các trình mô phỏng và khuôn khổ này chủ yếu mô hình hóa các khía cạnh vật lý của tính di động. Mặc dù tất cả chúng đều có thể được mở rộng để mô hình hóa các khía cạnh xã hội, tức là mục tiêu chuyển động và tính quy luật trong các mục tiêu này, nhưng chúng không theo đuổi để làm như vậy.

Tuy nhiên, tầm quan trọng của việc mô hình hóa các khía cạnh xã hội này của tính di động được chỉ ra trong [1]. Trong khi đó, ST– ACTS tập trung vào các khía cạnh xã hội này của tính di chuyển trong khi chỉ đặt ra các hạn chế hạn chế trên các khía cạnh vật lý của tính di chuyển. Trên thực tế, vấn đề được giải quyết bởi các MOSs ở trên là trực giao với vấn đề được giải quyết bởi ST-ACTS.

Trong Oproto [8] –một trình tạo kịch bản thực tế cho các đối tượng chuyển động được thúc đẩy bởi một ứng dụng đánh cá, hành vi chuyển động của các đối tượng bị ảnh hưởng bởi các đối tượng khác, đứng yên hoặc chuyển động, thuộc các loại đối tượng khác nhau. Ảnh hưởng giữa các vật thể khác nhau có thể là lực hút hoặc lực đẩy. Mặc dù tác động đẩy và lực hút của các đối tượng khác là mục tiêu cho chuyển động, không giống như ST-ACTS, Oporto không cho phép mô hình hóa các quy luật trong các mục tiêu này.

Khung GAMMA [5] (Tạo chuyển động không mô hình nhân tạo bằng di truyền – Thuật toán) biểu thị hành vi của đối tượng chuyển động như một quỹ đạo trong không gian vị trí - thời gian và đề xuất hai số liệu chung để đánh giá tập dữ liệu quỹ đạo. Việc tạo quỹ đạo được coi như một bài toán tối ưu hóa và được giải bằng một thuật toán di truyền. Với các toán tử di truyền được sửa đổi một cách thích hợp và các tiêu chí phù hợp, khung được sử dụng để tạo ra các quỹ đạo mạng di động thường xuyên nhất có thể giữa các tế bào nội trú và các quỹ đạo vị trí tượng trưng (1) thể hiện các mô hình di động tương tự như các quỹ đạo hiện có trong một tập hợp các quỹ đạo mẫu ngoài đời thực được đưa ra dưới dạng đầu vào, (2) phù hợp với các ràng buộc và kinh nghiệm trong cuộc sống thực. Dựa trên quỹ đạo hoạt động mẫu, khung GAMMA có thể được định cấu hình để tạo quỹ đạo hoạt động chứa các mẫu hoạt động trong đời thực. Trong khi các quỹ đạo được tạo ra sẽ tương tự như các quỹ đạo đầu vào, vì chúng là tượng trưng, ​​chúng sẽ, vì các quỹ đạo đầu vào mặc nhiên giả định một ngữ cảnh phụ thuộc vào vị trí, (xem nguyên tắc thứ ba và thứ tư trong Phần 3). Để mô phỏng các hoạt động không gian - thời gian của toàn bộ dân số, cần phải có một mẫu đại diện về quỹ đạo phụ thuộc vào ngữ cảnh, nhưng rất khó lấy. Trong khi đó, ST-ACTS, dựa trên các nguyên tắc trực quan và một số nguồn dữ liệu thống kê địa lý trong đời thực, có thể tạo ra dữ liệu hoạt động thực tế, không gian-thời gian có tính đến bối cảnh hoạt động phụ thuộc vào vị trí này.

Địa lý thời gian [4] là một cơ sở khái niệm / mô hình cho hành vi không gian - thời gian của con người xem xét (1) tính không thể phân chia hoặc tính xác thực của tình trạng con người; (2) rằng con người thường hoạt động trong những khoảng thời gian và không gian hữu hạn; (3) các quy luật tự nhiên và các quy ước xã hội hạn chế một phần hành vi không-thời gian; và (4) rằng con người có mục đích. ST-ACTS mô hình hóa một số khía cạnh của mô hình này trong một trình tạo dữ liệu cụ thể, được triển khai.

1. **BÁO CÁO VẤN ĐỀ**

Các MOSs hiện tại chỉ nắm bắt các khía cạnh *vật lý* của tính di động, tức là *sự* *chuyển động* của các đối tượng, một cách đầy đủ. Tuy nhiên, để hỗ trợ sự phát triển của các phương pháp quản lý dữ liệu theo không gian - thời gian và khai thác dữ liệu, các khía cạnh xã hội của tính di động phát sinh từ các mẫu hành vi của con người nên được ghi lại bằng một mô hình. Các nguyên tắc quan trọng nhất chi phối các khía cạnh xã hội này của tính di động là:

* **Nguyên tắc đầu tiên**: Mọi người di chuyển từ một địa điểm nhất định đến một địa điểm khác với *mục tiêu thực hiện một số hoạt động* tại địa điểm sau.
* **Nguyên tắc thứ hai**: Không phải tất cả mọi người đều có khả năng thực hiện một hoạt động nhất định như nhau. *Khả năng thực hiện một hoạt động* phụ thuộc vào sở thích của một người nhất định, do đó phụ thuộc vào một số biến nhân khẩu học.
* **Nguyên tắc thứ ba**: *Các hoạt động được thực hiện bởi một người nhất định phụ thuộc nhiều vào ngữ cảnh*. Một số phần quan trọng hơn của bối cảnh này là: vị trí hiện tại của người đó, tập hợp các vị trí có thể thực hiện một hoạt động nhất định, thời gian hiện tại và lịch sử gần đây của các hoạt động mà người đó đã thực hiện.
* **Nguyên tắc thứ tư**: *Vị trí của các cơ sở,* nơi một hoạt động nhất định có thể được thực hiện, không được phân bố ngẫu nhiên, mà bị ảnh hưởng bởi vị trí của các cơ sở khác và vị trí của những người sử dụng mà các cơ sở đó phục vụ.

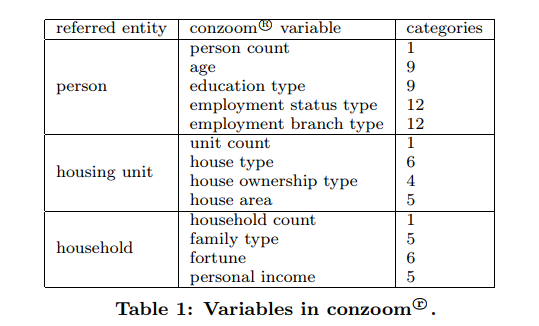
Nguyên lý đầu tiên có thể được coi như một tiên đề có liên quan đến định luật chuyển động đầu tiên của Newton. Chuyển động được thúc đẩy bởi mục đích duy nhất của chuyển động và không tuân theo nguyên tắc này - ví dụ: chuyển động phát sinh từ các hoạt động tập thể dục ngoài trời - không được mô hình hóa.

Nguyên tắc thứ hai có thể được sửa chữa bằng nhiều ví dụ từ thực tế cuộc sống. Hai trong số những ví dụ này là người cao tuổi thường đến hiệu thuốc hơn người trẻ hơn và người trẻ tuổi có xu hướng đi xem hòa nhạc pop hoặc rock hơn người cao tuổi.

Nguyên tắc thứ ba, có lẽ quan trọng nhất, là do một số yếu tố. Đầu tiên, chuyển động là một yêu cầu cần thiết (không phải lúc nào cũng thú vị) để thực hiện một số hoạt động, và do đó trong hầu hết các trường hợp, số lượng chuyển động cần thiết để làm như vậy được tác nhân giảm thiểu, tức là mọi người có xu hướng đi đến một quán cà phê gần đó. bởi. Thứ hai, các hoạt động không được thực hiện với khả năng như nhau tại các thời điểm khác nhau. Ví dụ, hầu hết mọi người có xu hướng đi làm vào buổi sáng trái ngược với các thời gian khác trong ngày; do đó, khả năng thực hiện hoạt động đó vào buổi sáng cao hơn so với các khoảng thời gian khác trong ngày. Hơn nữa, do tính chất của chúng, các hoạt động khác nhau có thời lượng khác nhau. Thời lượng của một hoạt động nhất định đặt ra một hạn chế tự nhiên đối với khả năng thực hiện một hoạt động khác trong khi hoạt động trước đó kéo dài. Ví dụ, mọi người có xu hướng bắt đầu làm việc từ buổi sáng trong khoảng thời gian khoảng 8 giờ; do đó, khả năng mua sắm hàng tạp hóa trong cùng một khoảng thời gian sẽ thấp hơn so với thời gian khác. Cuối cùng, mặc dù một người có thể thực hiện một hoạt động với khả năng xảy ra rất cao, nhưng các hoạt động được thực hiện bởi người đó không độc lập về mặt thời gian. Ví dụ, rất khó xảy ra trường hợp một người thích các buổi hòa nhạc pop và rock đi xem một số buổi biểu diễn trong cùng một buổi tối thứ Bảy.

Nguyên tắc thứ tư chủ yếu là kết quả của các quy luật cung và cầu của kinh tế học. Vị trí của các cơ sở chủ yếu bị ảnh hưởng bởi cạnh tranh, chi phí thị trường và tiềm năng thị trường. Ví dụ, mặc dù chi phí thành lập một tiệm tắm nắng ở ngoại ô thị trấn có thể thấp, nhưng tiềm năng thị trường thậm chí có thể không bù đắp được chi phí thấp này. Do đó, rất ít khả năng người ta sẽ tìm thấy một số tiệm tắm nắng trên một khu phố thành phố. Quá trình không gian làm phát sinh các vị trí của các cơ sở là một quá trình phức tạp, năng động với quá trình phản hồi, được điều chỉnh bởi các quy luật của thị trường cạnh tranh. Do đó, việc sử dụng ảnh chụp nhanh về sự phân bố không gian của các cơ sở trong thế giới thực làm thông tin theo ngữ cảnh sẽ tạo cơ sở hợp lý để xây dựng một mô hình thực tế về các hoạt động không gian - thời gian có thể được thực hiện tại các cơ sở đó.

Mục tiêu *định tính* chính của mô hình mô phỏng là nắm bắt các nguyên tắc chi phối được mô tả ở trên của các mẫu hành vi của con người và được coi là *tính hợp lệ* của mô hình mô phỏng. Ngoài ra, mô hình mô phỏng phải đạt được một số mục tiêu *định lượng*. Đầu tiên, mô hình mô phỏng phải *hiệu quả*, tức là nó phải có khả năng tạo ra một lượng lớn dữ liệu tổng hợp trong một thời gian hợp lý. Thứ hai, mô hình mô phỏng phải được *tham số hóa*, tức là dựa trên các tham số do người dùng xác định, nó phải có thể tạo các tập dữ liệu tổng hợp với các kích thước và đặc điểm khác nhau.



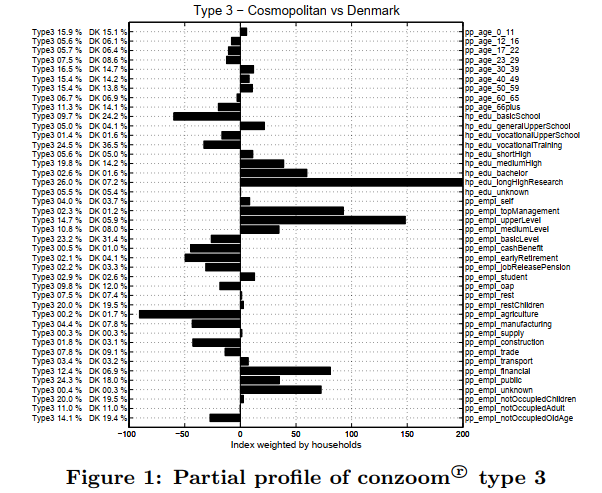
Cuối cùng, mô hình mô phỏng phải *chính xác*, tức là dữ liệu tổng hợp được tạo ra bởi mô hình phải có cùng các thuộc tính thống kê đối với các mẫu vì nó được xác định bởi các tham số và đầu vào của mô hình.

1. **SOURCE DATA**

Dữ liệu nguồn được sử dụng trong mô hình mô phỏng là các sản phẩm thương mại của Geomatic, một công ty Đan Mạch chuyên về dữ liệu địa lý và phân tích nhân khẩu học để phân khúc thị trường, kinh doanh thông minh và tiếp thị trực tiếp [3]. Do tính chất thương mại của các tập dữ liệu này, các phương pháp lấy chính xác của chúng không được mô tả ở đây. Tuy nhiên, các khái niệm và nguyên tắc được sử dụng trong quá trình dẫn xuất và các nội dung liên quan kết quả của cơ sở dữ liệu được giải thích dưới đây.

**conzoom® Demographic Data** (dữ liệu nhân khẩu học): conzoom®  là một sản phẩm cơ sở dữ liệu thương mại chứa thông tin địa lý, nhân khẩu học chi tiết về dân số của Đan Mạch [3]. Các biến mô tả các đặc điểm thống kê của dân số có thể được chia thành ba nhóm: biến person (người), housing unit (đơn vị nhà ở) và biến household (hộ gia đình). Các biến này và số lượng danh mục cho mỗi loại được trình bày trong Table 1.

Trong Table 1, các biến có “type” trong tên của chúng là các biến phân loại; các biến có “count” trong tên của chúng là số lượng của các thực thể tương ứng trong ô lưới 100 mét; và cuối cùng, các biến còn lại là các biến liên tục đã được phân loại thành các loại có ý nghĩa đối với việc phân đoạn thị trường.

Ví dụ như ở nông thôn, số lượng người, hộ gia đình hoặc đơn vị có thể rất thấp trong một ô lưới dài 100 mét, các ô lưới được nhóm lại với nhau thành các cụm có ý nghĩa, đủ lớn để tuân thủ các quy tắc xã hội và đạo đức và bảo vệ sự riêng tư. của các cá nhân. Cơ sở để phân nhóm gồm hai phần: địa lý và thông tin nhà ở 1-1 được công bố rộng rãi. Trực giác đằng sau cơ sở cũng gấp đôi. Thứ nhất, những người sống trong một khu vực địa lý nhất định (có thể là một tiểu bang, một quận, một khu bưu điện) theo một nghĩa nào đó là tương tự nhau; chẳng hạn, họ có thể có xu hướng chính trị tương tự hơn từ những người sống ở khu vực địa lý khác. Thứ hai, những người sống trong những ngôi nhà giống nhau có khả năng giống nhau trong các biến nhân khẩu học khác; ví dụ một gia đình đã thành lập với nguồn thu nhập ổn định sẽ có nhiều khả năng mua được một căn nhà lớn hơn, đắt tiền hơn một người mới bắt đầu sự nghiệp của mình. Như đã đề cập trước đó, để bảo vệ sự riêng tư của các cá nhân, các cụm bị hạn chế phải chứa ít nhất một số hộ gia đình cố định. Thống kê cho các biến, tùy thuộc vào mức độ nhạy cảm của thông tin chứa trong chúng, được lấy từ Thống kê Đan Mạch [10] cho các cụm được xây dựng ở mức giới hạn quy mô cụm thích hợp, ví dụ 20, 50, 100 và 150 hộ gia đình trên mỗi cụm. Trong trường hợp một biến liên tục, ví dụ: age(tuổi), số lượng các thực thể tương ứng (trong trường hợp này là những người trong cụm) được thu thập cho các loại của biến đã cho.

Do phương pháp phân nhóm địa lý bị hạn chế này, các cụm conzoom® thu được tuân thủ các chuẩn mực xã hội và đạo đức và bảo vệ quyền riêng tư của cá nhân, tuy nhiên số liệu thống kê thu được vẫn đủ chính xác để phân khúc thị trường hiệu quả. Sự phân đoạn này dẫn đến việc nhóm dân số Đan Mạch thành 29 kiểu conzoom ®, được xác định cho mỗi ô lưới 100 mét. Cosmopolitan (loại 3) là một ví dụ trong số 29 loại conzoom ® . So sánh nhân khẩu học của loại 3 với nhân khẩu học của phần còn lại của dân số Đan Mạch sẽ cung cấp *hồ sơ nhân khẩu học* của loại này. Hồ sơ này được thể hiện một phần trong Figure 1. Nó mô tả đại khái những cá nhân có nhiều khả năng: ở độ tuổi trung niên (30–59 tuổi), sống ở các thành phố lớn hơn trong những ngôi nhà lớn hơn, nhiều gia đình thuộc sở hữu của họ hoặc là cho thuê tư nhân, hầu hết là các cặp vợ chồng có con, có trình độ học vấn cao hơn trung bình đến dài hạn, giữ các vị trí cấp cao hơn hoặc quản lý hàng đầu trong lĩnh vực tài chính hoặc công và để có kinh tế hộ gia đình tốt hơn (cả về tài sản và thu nhập) so với mức trung bình của Dane.

**Mobidk TM Daily Movement Data** (dữ liệu chuyển động hằng ngày)**:** mobidk TM  là một sản phẩm cơ sở dữ liệu thương mại sắp ra mắt chứa thông tin chi tiết về sự di chuyển hàng ngày của người dân Đan Mạch giữa nhà và nơi làm việc [3]. Một lần nữa, để bảo vệ quyền riêng tư của người dùng, dữ liệu di chuyển được tổng hợp thành các vùng địa lý được kết nối và không chồng chéo. Nó được biểu diễn ở định dạng cơ sở dữ liệu quan hệ như: <*from\_region, to\_region, count>,* nghĩa là từ khu vực địa lý *from\_region*, *count* số người di chuyển hàng ngày để làm việc đến khu vực địa lý *to\_region*. Trong ST-ACTS, các vùng địa lý này là các giáo xứ, trung bình có 1176 hộ gia đình và 195 ô lưới 100 mét.

**Bizmark TM Business Data** (dữ liệu kinh doanh)**:** bizmark TM là một sản phẩm cơ sở dữ liệu thương mại chứa thông tin chi tiết về các doanh nghiệp Đan Mạch cả trong khu vực công và tư nhân [3]. Một số thông tin 1-1 có sẵn về các doanh nghiệp là vị trí của họ, số lượng nhân viên làm việc tại họ, quy mô vật chất của cơ sở kinh doanh và mã chi nhánh quốc tế mà doanh nghiệp có. Thông tin chi tiết nhưng tổng hợp về nhân viên trong doanh nghiệp cũng có sẵn cho các cụm bizmarkTM thích hợp, được xây dựng có tính đến địa lý, chi nhánh kinh doanh, quy mô kinh doanh về số lượng nhân viên và quy mô thực tế của cơ sở kinh doanh và nhiều biến số kinh doanh mô tả khác .

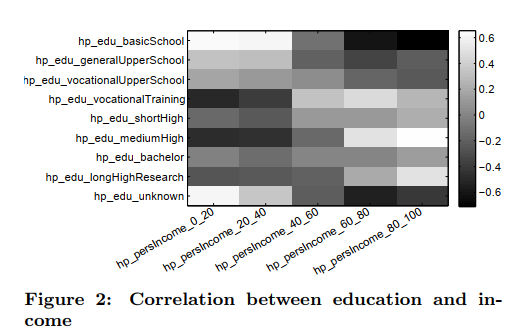
**GallupPC®****Consumer Survey Data** (Dữ liệu Khảo sát Người tiêu dùng)**:** GallupPC**®** là một sản phẩm cơ sở dữ liệu thương mại và như tên gọi, nó chứa các câu trả lời khảo sát chi tiết của người tiêu dùng về nhân khẩu học của họ; sở thích như văn hóa, sở thích và thể thao; tiêu dùng của hộ gia đình, thói quen mua hàng; thói quen đi lại; quan điểm về các chủ đề khác nhau; thái độ và sự tiếp xúc với các phương tiện quảng cáo khác nhau [2]. Các câu hỏi trong các cuộc khảo sát là câu hỏi có / không. Để đo mức độ phản hồi của một đối tượng khảo sát riêng lẻ đối với một câu hỏi cụ thể, câu hỏi có / không ban đầu được diễn giải lại với tham chiếu đến khoảng thời gian - tần suất. Ví dụ, câu hỏi có / không ban đầu "Bạn có đi đến thư viện không?" được diễn đạt lại thành 7 câu hỏi có / không sử dụng các khoảng thời gian-tần suất sau: hàng ngày / gần như hàng ngày; 3-4 lần một tuần; 1-2 lần một tuần; 1-3 lần một tháng; 1-5 lần mỗi 6 tháng; hiếm khi, và không bao giờ.

1. **ST-ACTS: TRÌNH MÔ PHỎNG HOẠT ĐỘNG KHÔNG GIAN-THỜI GIAN**

Trong phần này, các thành phần chính của ST-ACTS và việc sử dụng source data của chúng được mô tả. Trong mô tả, một người được mô phỏng, người thực hiện các hoạt động trong thời gian và không gian, sẽ được viết tắt là simperson. Hộp công cụ MATLAB cho ST-ACTS có thể được tải xuống cho mục đích nghiên cứu từ http://www.geomatic.dk/research/ST–ACTS/.

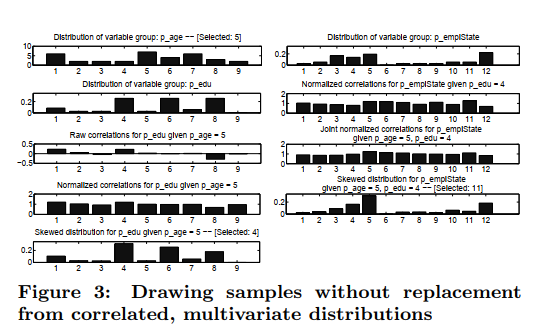
**Vẽ các biến nhân khẩu học cho Simpersons**: conzoom **®** source data chứa thông tin nhân khẩu học chính xác, chi tiết về dân số được tổng hợp ở cấp độ cụm. Như được mô tả trong Phần 4, các biến liên tục được sắp xếp thành các danh mục. Các cụm chứa số lượng cho tất cả các danh mục cho tất cả các biến. Có số lượng persons, housing units, và households chính xác ở cấp ô lưới và giả sử cùng một phân phối của các biến trong các ô lưới riêng lẻ như trong cụm mà chúng thuộc về, số lượng cho tất cả các danh mục cho tất cả các biến được tính ở một lưới cấp độ ô. Một simperson được gán một danh mục cho một biến nhất định tỷ lệ với số lượng các danh mục cho biến đã cho trong ô lưới mà simperson sống trong đó. Nói tóm lại, một danh mục cho biến được gán cho simperson theo phân phối của biến . Để vẽ các danh mục gán cho các biến mà không cần thay thế, số lượng tương ứng trong ô lưới đã cho sẽ giảm. Vì số lượng của một số biến trong ô lưới đề cập đến các thực thể không phải là người, nhưng là các biến là một phần của các biến nhân khẩu học mô tả một người, các số lượng này được điều chỉnh để tổng thành số người trong ô.

**Phân phối xiên dựa trên mối tương quan:** Phương pháp được mô tả ở trên để chỉ định danh mục cho các biến nhân khẩu học có một lỗ hổng lớn: các biến nhân khẩu học không độc lập. Ví dụ, biến loại hình giáo dục có mối tương quan chặt chẽ với biến thu nhập cá nhân. Mối tương quan này được minh họa trong Figure 2. Các mối tương quan được tính toán giữa tỷ lệ phần trăm của các biến được phân loại, và các mẫu được tính theo số người trong ô. Từ thanh màu bên cạnh, người ta có thể thấy rằng các sắc thái tối hơn có nghĩa là tương quan tiêu cực mạnh hơn và các sắc thái sáng hơn có nghĩa là tương quan tích cực mạnh hơn. Các mối tương quan hỗ trợ



kiến thức chung rằng những người có trình độ học vấn cao hơn có xu hướng có công việc được trả lương cao hơn. Các mối tương quan tương tự tồn tại giữa các biến khác.

Để khắc phục lỗ hổng được mô tả ở trên, điều này có thể dẫn đến việc gán các danh mục không thực tế cho các biến thành simpersons, việc gán này được sửa đổi bằng cách vẽ các danh mục từ các phân phối biến lệch cố gắng nhúng các mối tương quan giữa các biến như sau. Đối với một simperson đã cho, danh mục cho biến đầu tiên, t age, được vẽ mà không thay thế từ phân phối không cân đối của biến age. Ví dụ về sự phân bố này và kết quả của lần rút thăm được hiển thị trong đồ thị con phía trên cùng bên trái của Figure 3, trong đó đối với biến age, loại 5 được vẽ, biểu thị rằng simperson thuộc nhóm tuổi 30–39. Sự phân bố của biến thứ hai, education, được hiển thị trong đồ thị con thứ hai – từ trên cùng bên trái của Figure 3. Với phân phối này, các loại 4, 6 và 8 có nhiều khả năng được gán cho simperson cho biến education. Tuy nhiên, các mối tương quan (thể hiện trong hình thứ ba bên trái của Figure 3) giữa nhóm tuổi 5 và biến education cho thấy mối tương quan tích cực đối với nhóm 1 và 4, và mối tương quan nghịch đối với nhóm 8 đối với biến education. Sau khi chuẩn hóa (chuyển sang trung bình 1) các mối tương quan, phân phối ban đầu của biến education bị lệch bằng cách nhân đôi khôn ngoan nhân số lượng thô của các danh mục của biến education và các tương quan chuẩn hóa cho biến education cho rằng loại tuổi của simperson là 5. Phân phối lệch này được thể hiện trong đồ thị con dưới cùng bên trái của Figure 3 và được sử dụng để lấy mẫu biến education, dẫn đến loại education 4, đào tạo nghề. Giá trị cho các biến tiếp theo được rút ra từ các phân phối lệch có tính đến các danh mục cho các biến đã được rút ra trước đó, bằng cách làm sai lệch phân phối của biến hiện tại theo giá trị trung bình của các tương quan chuẩn hóa cho các danh mục đã rút ra cho đến nay. Quá trình này được thể hiện từ trên xuống dưới ở các cấu hình phụ bên phải của Figure 3, trong đó loại age cho simperson là 5 và loại education là 4 cho biến thứ ba, trạng thái việc làm, loại 11 được vẽ.

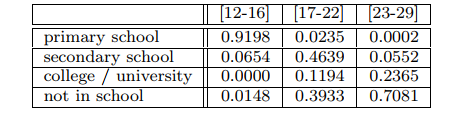


**Assigning Simpersons đến Work Places / Schools:** Các hoạt động có thể được chia thành hai nhóm: *hoạt động thời gian rảnh* và *hoạt động bắt buộc*. Mặc dù khái niệm về hoạt động “bắt buộc” có thể khác nhau ở mỗi người, nhưng với mục đích mô phỏng, ST-ACTS coi việc *tới trường* và *đi làm* là những hoạt động bắt buộc. Phần còn lại của các hoạt động trong ST-ACTS được coi là các hoạt động thời gian rảnh.

Đối với các hoạt động bắt buộc, simpersons có thể được chia thành ba nhóm: đã nghỉ hưu, công nhân và sinh viên. Đối với những người bình thường đã nghỉ hưu, có thể cho rằng họ được hưởng thành quả của một cuộc sống lao động cần cù và không có hoạt động bắt buộc nào. Do đó, họ dành phần lớn thời gian ở nhà hoặc thực hiện các hoạt động khi rảnh rỗi. Các đoạn sau mô tả các phương pháp trong ST-ACTS (và cách sử dụng dữ liệu cơ sở của chúng) để chỉ định các phương pháp đơn giản trong các nhóm công nhân và sinh viên cho nơi làm việc và trường học của họ tương ứng.

**Assigning Worker Simpersons to Work Places:** Simpersons trong nhóm công nhân được chỉ định đến *địa điểm làm việc* theo hai bước. Trong bước đầu tiên, với *home parish* và chi nhánh việc làm của simperson, xác suất đi lại từ parish đến parish và sự phân bố không gian của các doanh nghiệp trong các chi nhánh, một *work parish* được giao cho simperson. Trong bước thứ hai, với nhánh việc làm mà simperson làm việc, các doanh nghiệp trong cùng một nhánh nằm trong vùng làm việc của simperson được truy xuất từ bizmarkTM. Cuối cùng, tỷ lệ thuận với số lượng nhân viên làm việc trong các doanh nghiệp đã chọn, đơn vị này chắc chắn sẽ được chỉ định vào một trong các doanh nghiệp / địa điểm làm việc.

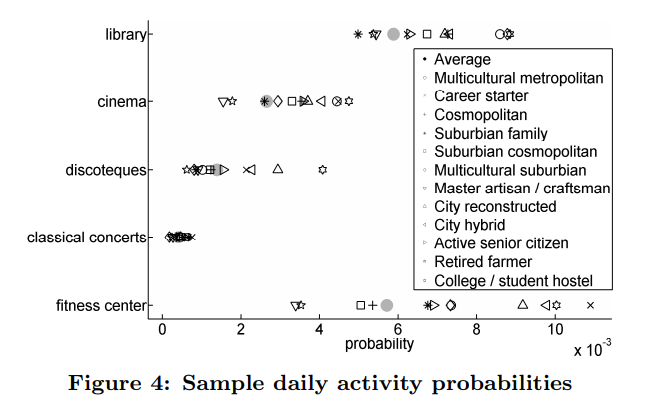
**Assigning Student Simpersons to Schools:** Simpersons trong nhóm học sinh được chỉ định cho trường học theo hai bước. Trong bước đầu tiên, tùy thuộc vào nhóm tuổi của simperson, họ được chỉ định vào một trong bốn loại hình cơ sở giáo dục, hoặc được chỉ định là “không đi học” và do đó được coi là thành viên của nhóm công nhân. Trong bước thứ hai, các tổ chức giáo dục thuộc loại tổ chức giáo dục của simpersons được truy xuất từ bizmark TM và simperson được chỉ định cho tổ chức gần nhà của simperson nhất. Đoạn sau giải thích chi tiết hơn về bước đầu tiên của những bước này.

Simpersons trong nhóm học sinh có thể được chia thành bốn phân nhóm dựa trên cơ sở đó, nếu có, trong bốn loại cơ sở giáo dục mà họ theo học: kindergarten (mẫu giáo), primary school (tiểu học), secondary school (trung học) hoặc college / university (cao đẳng / đại học). Như đã mô tả ở trên, mỗi người dưới 30 tuổi được gán cho một trong bốn nhóm tuổi: [0, 11], [12, 16], [17, 22] và [23, 29]. Giả sử tất cả các học sinh tối 5 hoặc 6 tuổi đều đi nhà trẻ (hoặc nhà trẻ), các học sinh ở độ tuổi [0, 11] được xếp vào trường mẫu giáo hoặc trường tiểu học như nhau. Đối với mỗi nhóm trong số ba nhóm tuổi còn lại, dựa trên thông tin thu được từ Thống kê Đan Mạch [10], xác suất theo học một trong bốn loại hình cơ sở giáo dục được suy ra, được thể hiện trong bảng:

Sau đó, dựa trên nhóm tuổi của simperson và các xác suất tương ứng, simperson được chỉ định vào một trong ba loại hình cơ sở giáo dục, hoặc “không đi học” và được coi là thành viên của nhóm công nhân.

**Xác suất Hoạt động Hàng ngày:** Một tập hợp con các câu hỏi khảo sát người tiêu dùng của GallupPC ®, được mô tả trong Phần 4 đại diện cho các hoạt động đòi hỏi sự di chuyển của người tiêu dùng. Một số hoạt động này được hiển thị trên trục y của Figure 4. Để bảo tồn không gian và sự rõ ràng, các hoạt động sau đây được đưa vào mô hình, nhưng bị loại trừ khỏi hình: triển lãm nghệ thuật, nhà thờ, buổi hòa nhạc pop / rock, bảo tàng , bưu điện, rạp hát, phòng tắm nắng, tiệm làm tóc và mua sắm. Hoạt động mua sắm được chia nhỏ thành 22 loại hình mua sắm phụ gắn liền với một thương hiệu hoặc loại cửa hàng cụ thể.

Sử dụng các phần địa lý nhân khẩu học của các cuộc khảo sát, mỗi đối tượng khảo sát được chỉ định cho một trong 29 loại hình thu nhỏ. Để có được một chỉ số duy nhất cho khả năng một loại conzoom ® nhất định có thể thực hiện một hoạt động nhất định như thế nào, câu trả lời cho các câu hỏi về tần suất thời gian được phân tích lại được chuẩn hóa và tính trung bình như sau. Đầu tiên, mọi khoảng thời gian-tần suất cho một hoạt động được chuẩn hóa để biểu thị xác suất thực hiện hoạt động đã cho vào một ngày trung bình. Ví dụ: câu trả lời tích cực của đối tượng cho câu hỏi "Bạn có thực hiện hoạt động *a* *n* lần trong một khoảng thời gian *∆t* không?" tương đương có nghĩa là xác suất đối tượng đó thực hiện hoạt động *a* vào bất kỳ ngày nào đã cho là *P(a) = n/day(∆t)*, trong đó *day* là hàm trả về số ngày trong khoảng thời gian *∆t*. *P (a)* được gọi tương đương là Xác suất *Hoạt động Hàng ngày* (DAP) của hoạt động *a*. Thứ hai, xác suất hoạt động hàng ngày này của các đối tượng riêng lẻ thuộc loại conzoom® nhất định được tính trung bình. Figure 4 cho thấy một mẫu của các xác suất hoạt động hàng ngày này cho một tập hợp con của các loại conzoom ®. Từ con số có thể thấy, chẳng hạn, một sinh viên đại học có nhiều khả năng đến thư viện, rạp chiếu phim, vũ trường hoặc trung tâm thể dục; trong khi một nông dân đã nghỉ hưu là ít có khả năng thực hiện các hoạt động này nhất. Vì con số có cùng thang xác suất, nên nó cũng tiết lộ rằng, tùy thuộc vào loại hình, đến trung tâm thể dục là hoạt động thường xuyên hoặc phổ biến hơn khoảng 7 đến 22 lần so với đi xem các buổi hòa nhạc cổ điển. Như đã đề cập trước đây, ST-ACTS bao gồm các xác suất hoạt động hàng ngày của 35 hoạt động cho 29 loại tùy chỉnh.

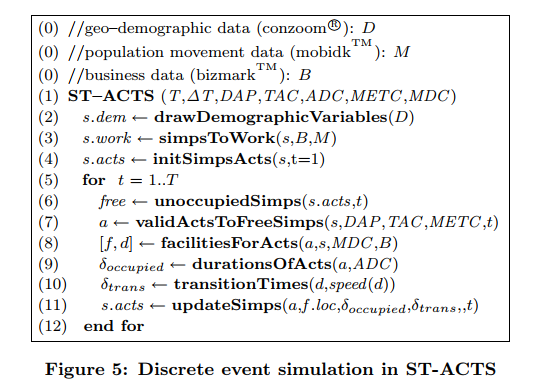


**Mô phỏng hoạt động với các ràng buộc về không gian-thời gian:** Một trình mô phỏng hoạt động sự kiện đơn giản, ngẫu nhiên, rời rạc có thể được xây dựng như sau. Tại mỗi bước thời gian, một tập hợp con ngẫu nhiên của các simperson được chọn để thực hiện một hoạt động. Sau đó, đối với mỗi simperson đã chọn, được cung cấp loại conzoom® của anh ấy / cô ấy và xác suất hoạt động hàng ngày liên quan, một hoạt động được chỉ định. Sau đó, mỗi simperson được chọn sẽ được chuyển đến cơ sở gần nhất, nơi có thể thực hiện hoạt động được chỉ định của họ. Trình mô phỏng đơn giản này không mô hình hóa một số ràng buộc không gian-thời gian đối với các hoạt động. Sau đây, các ràng buộc này sẽ được thảo luận và đối với mỗi hạn chế, giải pháp mô hình đề xuất mà ST-ACTS triển khai sẽ được trình bày.

**Hạn chế về hoạt động tạm thời:** Một số hoạt động nhất định có nhiều khả năng được thực hiện trong những khoảng thời gian cụ thể hơn những hoạt động khác. Ví dụ, những người trong lực lượng lao động có xu hướng rời khỏi nhà để đi làm vào đầu ngày làm việc. Do đó, những người giống nhau ít có khả năng đến vũ trường, nơi có lẽ đã đóng cửa, trong cùng một khoảng thời gian. Để mô hình hóa *Ràng buộc hoạt động tạm thời* (TAC), ST-ACTS cho phép người dùng xác định cho từng nhóm trong số ba nhóm dân số xác suất cho mỗi hoạt động cho mỗi giờ mỗi ngày trong tuần. Những xác suất này được sử dụng để giới hạn khả năng của người thực hiện một số hoạt động nhất định trong những khoảng thời gian nhất định. Chúng không được nhầm lẫn với *xác suất hoạt động hàng ngày* phụ thuộc vào loại conzoom ®, mã hóa tùy chọn hoạt động của từng loại. Thông qua TACs, ST-ACTS cho phép lập mô hình giờ mở cửa và ở một mức độ nào đó các mô hình tuần tự. TACs của một hoạt động được xác định bởi ma trận 7 x 24, trong đó các cột đại diện cho giờ trong ngày và các hàng đại diện cho các ngày trong tuần.

**Activity Duration Constraint:** Không phải tất cả các hoạt động đều mất cùng một khoảng thời gian. Ví dụ: mọi người thường làm việc 6-10 giờ, dành khoảng 2 giờ ở rạp chiếu phim và 30 phút ở cửa hàng tạp hóa. Để mô hình hóa điều này, từ dấu thời gian bắt đầu của một hoạt động *a* được gán cho một simperson *s*, *s* sẽ bị chiếm dụng trong các bước thời gian *δoccupied (a)*. Trong thời gian này, *s* không được chỉ định bất kỳ hoạt động nào khác. Trong ST-ACTS, *Ràng buộc thời lượng hoạt động* (ADC) cho mỗi hoạt động được rút ra một cách xác suất từ các phân phối thời lượng hoạt động do người dùng xác định, thường được phân phối với trung bình *µδccupied (a)* và phương sai *σδoccupied (a).*

**Thời gian đã trôi qua tối thiểu giữa các ràng buộc về việc lặp lại hoạt động:** Mặc dù mọi người thích một số hoạt động hơn những hoạt động khác, nhưng rất ít khả năng là họ sẽ lặp lại hoạt động giống nhau, ngay cả khi được ưu tiên, nhiều lần, hết lần này đến lần khác trong một khoảng thời gian ngắn. Ví dụ, rất khó xảy ra, ngay cả một người rất năng động, ngay sau khi kết thúc buổi tập tại trung tâm thể dục, lại quyết định đến trung tâm thể dục. Ràng buộc này được mô hình hóa trong ST-ACTS thông qua *δelapsed (a)* do người dùng xác định, *ràng buộc thời gian đã trôi qua tối thiểu*



phụ thuộc vào hoạt động (METC). Ràng buộc được thực thi bằng cách duy trì lịch sử hoạt động gần đây của mỗi simperson và xác thực các hoạt động mới được rút ra chống lại nó.

**Ràng buộc về khoảng cách tối đa**: Đối với hầu hết các hoạt động, có một *khoảng cách tối đa* mà một người sẵn sàng đi. Khoảng cách tối đa này thể hiện sự hạn chế về mặt không gian đối với các hoạt động mà simperson *s* sẽ thực hiện, với vị trí hiện tại của *s* và vị trí của các cơ sở, nơi có thể thực hiện một hoạt động đã chọn.. Do đó, trong quá trình mô phỏng nếu không có cơ sở phù hợp cho *a* trong khoảng cách tối đa của vị trí hiện tại của *s*, hoạt động được coi là không hợp lệ đối với *s* và *s* trở nên không hoạt động. *Ràng buộc Khoảng cách Tối đa* (MDC) được kiểm soát bởi một tham số phụ thuộc vào hoạt động, do người dùng xác định trong ST-ACTS.

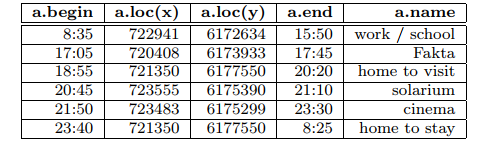
**Hạn chế về tính di chuyển vật lý**: Để di chuyển từ vị trí này sang vị trí khác cần có thời gian. Mặc dù mô phỏng chi tiết chuyển động này không phải là mục tiêu của ST-ACTS, nhưng các hạn chế cơ bản về khả năng di chuyển vật lý được mô hình hóa. Sau khi một cơ sở *f* cho một hoạt động *a*  được chọn cho một simperson *s*, *s* sẽ được di chuyển sau *δtrans* bước thời gian đến vị trí mới. *δtrans* được tính dựa trên khoảng cách Euclidian *d*  tính bằng km giữa vị trí hiện tại của *s* và vị trí của cơ sở *f*, giả sử tốc độ không đổi. Tốc độ không đổi này, tính bằng km / h, được xác suất rút ra từ sự phân bố *speed(d) = max(5, N(3d, d2))*. *speed(d)* ấn định tốc độ thấp hơn cho thời gian ngắn hơn và tốc độ cao hơn (với phương sai lớn hơn) cho khoảng cách dài hơn. Ở một mức độ nào đó, nó nắm bắt các phương thức giao thông phổ biến, tức là mọi người có xu hướng đi bộ trong những chuyến đi ngắn hơn, sử dụng phương tiện giao thông công cộng hoặc xe đạp trong những chuyến đi dài hơn một chút và sử dụng ô tô hoặc tàu hỏa đi lại trong những chuyến đi dài hơn.

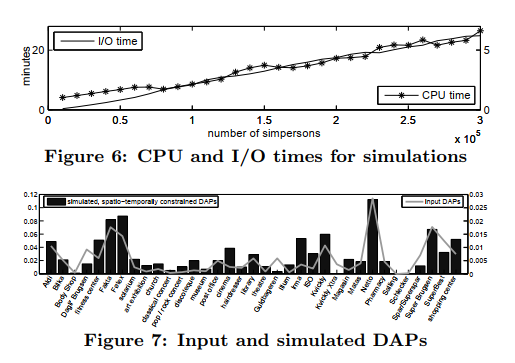
**Mô phỏng sự kiện rời rạc:** Sử dụng các khối xây dựng khái niệm đã trình bày cho đến nay, mô phỏng sự kiện rời rạc được thực hiện trong ST-ACTS có thể được tóm tắt như trong Figure 5. Ba nhận xét đầu tiên chỉ ra rằng các tập dữ liệu được đặt tên được sử dụng trong mô phỏng nhưng không phải là các tham số do người dùng xác định của nó. Các đối số đối với ST-ACTS, được hiển thị trên dòng 1, là các tham số do người dùng xác định đã được mô tả trong các đoạn trước. Trên dòng 2, các biến nhân khẩu học được gán cho simpersons dựa trên phân phối biến lệch. Trên dòng 3 simpersons được chỉ định đến nơi làm việc và trường học. Trên dòng 4, tại thời điểm bước *t* = 1 (Thứ Hai, 00:00) tất cả simperson được khởi tạo để ở “nhà” thực hiện hoạt động “nhà để ở” cho đến sáng sớm. Theo các bước tiền xử lý này, tại mỗi thời điểm bước *t*, trên dòng 6, hiện không có người sử dụng (free) simpersons được tìm thấy. Sau đó, trên dòng 7 đối với mỗi free simperson, một hành động hợp lệ được tìm thấy theo xác suất hoạt động hàng ngày (DAP) của các hành động đối với kiểu conzoom ® của simperson. Các hành động đều hợp lệ, nếu cả hai đều đáp ứng hạn chế hoạt động tạm thời (TAC) và giới hạn thời gian trôi qua tối thiểu (METC). Trên dòng 8 cơ sở hợp lệ được tìm thấy cho các hoạt động hợp lệ này. Cơ sở vật chất có giá trị nếu chúng đáp ứng giới hạn khoảng cách tối đa (MDC). Trên dòng 9, thời lượng hoạt động được vẽ đáp ứng giới hạn thời lượng hoạt động (ADC). Trên dòng 10, theo khoảng cách đến các hoạt động được chỉ định, thời gian chuyển tiếp được tính toán. Cuối cùng, trên dòng 11, thông tin về các hoạt động mới được chỉ định được lưu trữ và lịch sử hoạt động được cập nhật cho các simpersons bị ảnh hưởng.

1. **ĐÁNH GIÁ MÔ PHỎNG**

ST – ACTS đã được triển khai và thử nghiệm trong MATLAB chạy trên Windows XP trên bộ xử lý Pentium 4 3,6 GHz với bộ nhớ chính 1,5 GB. Phạm vi địa lý của STACTS bị giới hạn ở các thành phố Copenhagen và Frederiksberg ở Đan Mạch. Theo đó, số người lao động đơn giản là 590.050 (178.826 người nghỉ hưu, 268.615 công nhân và 142.609 sinh viên), số nơi làm việc là 1.264.129 tại 193.299 doanh nghiệp và số cơ sở là 10.544. Thí nghiệm mô phỏng được thực hiện trong một bước thời gian có độ dài *∆T* = 5 phút. Để kiểm tra hiệu suất của ST-ACTS, trong tất cả các thử nghiệm, TACs “nghiêm ngặt” được đặt trên hai hoạt động có khả năng xảy ra nhất là “về thăm nhà” và về “nhà để ở”. TACs của các hoạt động khác được thiết lập để làm mẫu giờ mở cửa của các cơ sở tương ứng. Kết quả là một simperson thực hiện trung bình 9,6 ± 3,2 hoạt động mỗi ngày.

Để đánh giá hiệu quả của ST-ACTS, Mô phỏng được thực hiện cho các kích thước khác nhau của các tập con simpersons được chọn ngẫu nhiên trong suốt một ngày (24 giờ). Figure 6 cho thấy cả thời gian CPU (trục y bên phải) và thời gian I / O để ghi lại các sự kiện (trục y bên trái). Cả hai đại lượng này đều có quy mô xấp xỉ tuyến tính với số simperson. Nói tóm lại, mô phỏng diễn ra nhanh chóng và quy mô tốt.

Trong một thử nghiệm lớn hơn, các hoạt động của tổng dân số trong suốt một tuần đã được mô phỏng. Bảng dưới đây cho thấy kết quả đầu ra của ST-ACTS cho loại quốc tế simperson trong suốt một ngày.

Quá trình mô phỏng, không ghi nhật ký các sự kiện riêng lẻ và chỉ lưu các số liệu thống kê về các hoạt động, mất 98 phút. Để đánh giá tính hợp lệ của ST-ACTS, các số liệu thống kê thu thập được đã được phân tích. Do giới hạn về không gian, chỉ một số kết quả của phân tích này được thảo luận chi tiết, trong khi những kết quả khác chỉ được tóm tắt.

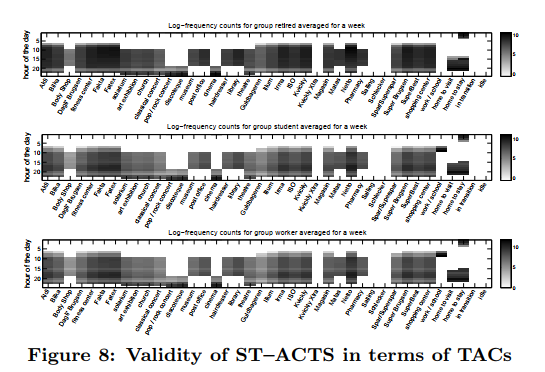
Để đánh giá khả năng của ST-ACTS trong việc tạo ra sự phân bổ chính xác của các hoạt động, DAPs đầu vào đã được so sánh với DAPs mô phỏng, được thể hiện trong Figure 7. Trong khi, do các TACs “nghiêm ngặt” đã đề cập trước đó, các DAPs mô phỏng cao hơn khoảng 4 lần so với các DAPs đầu vào, các DAPs được mô phỏng tương đối giữa các hoạt động tương tự như các DAPs đầu vào. Bằng cách sử dụng các TACs ít “nghiêm ngặt” hơn, tức là: cho phép người học về nhà sớm hơn sau khi làm việc, quy mô của các DAPs mô phỏng phù hợp với quy mô của các DAPs đầu vào. Sự khác biệt trong các DAPs tương đối có thể được giải thích do tác động của các hạn chế về mặt không gian đối với các hoạt động.

Để đánh giá khả năng của ST-ACTS trong việc kiểm soát các hạn chế về thời gian đối với các hoạt động, số đếm cho mỗi hoạt động được chỉ định cho mỗi giờ – ngày và ngày – trong tuần đã được duy trì. Figure 8 cho thấy số lượng hoạt động được chỉ định trung bình cho mỗi giờ – trong ngày được tính trung bình trong các ngày – trong tuần. Do sự thay đổi lớn về số lượng tần suất cho các hoạt động khác nhau trong các khoảng thời gian khác nhau trong ngày, lôgarit cơ số 2 của số lượng tần suất được hiển thị. Từ hình vẽ có thể thấy rằng một số nhóm thực hiện các hoạt động nhất định vào những thời điểm nhất định trong ngày thường xuyên hơn các nhóm khác. Ví dụ, có thể thấy rằng nhóm đã nghỉ hưu có nhiều khả năng thực hiện các hoạt động trong giờ làm việc hơn, đơn giản vì họ được tự do làm việc đó. Thời gian mở và đóng cửa của các cơ sở cũng được kiểm soát bởi các thông số. Ví dụ, không ai đến vũ trường vào ban ngày, và không ai đến trung tâm mua sắm vào nửa đêm.

Để đánh giá khả năng của ST-ACTS trong việc kiểm soát các hạn chế về không gian đối với các hoạt động, khoảng cách hàng ngày đi đến nơi làm việc của một người trung bình (2,7 ± 2,3 km) được so sánh với tổng khoảng cách hàng ngày mà một người đi làm trung bình (8,3 ± 3,6 km). Mặc dù đối với những con số tương tự, không có sự thật cơ sở nào có sẵn để đánh giá, nhưng xem xét trung bình 9,6 hoạt động mỗi ngày, những con số này có vẻ hợp lý. Dữ liệu mô phỏng cũng đã được xác minh rằng không có chuyến đi nào vi phạm tiêu chí khoảng cách tối đa phụ thuộc vào hoạt động.

1. **KẾT LUẬN**

Các mô hình thực tế mô phỏng các hoạt động không gian - thời gian của người dùng và do đó phân bố các đối tượng chuyển động, là điều cần thiết để tạo điều kiện thuận lợi cho việc phát triển các kỹ thuật khai thác dữ liệu và quản lý dữ liệu theo không gian - thời gian. Trong bài báo này, ST-ACTS, phần đầu tiên trong số các trình mô phỏng như vậy được trình bày. Kết quả thử nghiệm cho thấy rằng, sử dụng một số nguồn dữ liệu thống kê địa lý trong thế giới thực và các nguyên tắc trực quan, ST-ACTS có thể tạo ra một cách hiệu quả dữ liệu hoạt động không gian-thời gian thực tế. Nó cũng được chứng minh rằng dữ liệu được tạo ra có các đặc điểm giống như nó được xác định bởi các tham số mô hình do người dùng kiểm soát. STACTS đã được triển khai trong MATLAB và có sẵn cho các mục đích nghiên cứu.



Trong khi sự tương ứng giữa các đặc tính của dữ liệu được tạo và các tham số mô hình được chứng minh, độ chính xác của mô phỏng nhất thiết phải bị ảnh hưởng bởi mô hình hạn chế của các khía cạnh vật lý của tính di động. Do đó, trong công việc trong tương lai, việc tích hợp đầu ra của STACTS làm đầu vào cho mô phỏng đối tượng chuyển động dựa trên mạng phức tạp như trong [1] đã được lên kế hoạch. Một trình mô phỏng phức tạp hơn như vậy sẽ cung cấp các tập dữ liệu tổng hợp có thể hỗ trợ sự phát triển về đạo đức từ xa, hệ thống giao thông thông minh và các dịch vụ dựa trên vị trí.

**Sự nhìn nhận**

Công việc này được hỗ trợ một phần bởi Bộ Khoa học, Công nghệ và Đổi mới Đan Mạch theo tài trợ số 61480. Việc cung cấp conzoom®, bizmarkTM và GallupPC ® cho các mục đích nghiên cứu là điều vô cùng cảm ơn đối với Geomatic ApS. Và Tổ chức Gallup. Cung cấp dữ liệu di chuyển dân số ở cấp giáo xứ, được biết ơn Thomas Nielsen từ Trung tâm Rừng, Cảnh quan và Quy hoạch Đan Mạch. Sự giúp đỡ từ các đồng nghiệp, Susanne Caroc, Esben Taudorf, Jesper Christiansen, và Lau Marcussen cũng được ghi nhận một cách chân thành.

1. **TÀI LIỆU THAM KHẢO**

[1] T. Brinkhoff. A Framework for Generating Network–Based Moving Objects. Geoinformatica 6(2), pp.153–180, 2002.  
[2] The Gallup Organization: http://www.gallup.com/  
[3] Geomatic ApS – Center for Geoinformatik:http://www.geomatic.dk  
[4] T. H¨agerstrand. “Space, time and human conditions.” In Dynamic allocation of urban space, ed. A. Karlqvist et. al. Lexington: Saxon House Lexington Book, 1975.  
[5] H. Hu and D. L. Lee. GAMMA: A Framework for Moving Object Simulation. In Proc. of SSTD, pp. 37–54, 2005.  
[6] D. Pfoser and Y. Theodoridis. Generating Semantics–Based Trajectories of Moving Objects. In Proc. of Emerging  
Technologies for Geo–Based Applications, pp. 59–76, 2000.  
[7] J. F. Roddick, M. J. Egenhofer, E. Hoel, D. Papadias, and B. Salzberg. Spatial, Temporal and Spatio–Temporal  
Databases–Hot Issues and Directions for PhD Research. In Proc. of SSTD, pp. 1–6, 2003.  
[8] J.-M. Saglio and J. Moreira. Oporto: A Realistic Scenario Generator for Moving Objects. In Proc. of DEXA Workshop  
on Spatio–Temporal Data Models & Languages, pp. 426–432, 1999.  
[9] T. Sellis. Research Issues in Spatio–temporal Database Systems. In Proc. of SSD, pp. 5–11, 1999.  
[10] Statistics Denmark: http://www.dst.dk  
[11] Y. Theodoridis, J. R. O. Silva, and M. A. Nascimento. On the Generation of Spatiotemporal Datasets. In Proc. of SSD, pp. 147–164, 1999.