

NGHIÊN CỨU VÀ DẠY HỌC MỘT SỐ PHƯƠNG PHÁP ĐƠN GIẢN HOÁ ĐƯỜNG CONG TRÊN MÔ HÌNH DỮ LIỆU VECTOR CHO SINH VIÊN NGÀNH HỆ THỐNG THÔNG TIN ĐỊA LÍ - GIS

NGUYỄN TUẤN ANH - NGUYỄN HOÀNG LONG - NGÔ THỊ PHƯƠNG THẢO*

Ngày nhận bài: 20/06/2016; ngày sửa chữa: 23/06/2016; ngày duyệt đăng: 24/06/2016.

Abstract: The article has studied some curve simplified methods in the vector data models used to perform generalizations when displaying and searching on the map. This is one of the methods which is incorporated into teaching for students of geographic information system (GIS).

Keywords: vector data, curve, earth surface, GIS, Douglas-Peucker algorithm, Lang algorithm, Angle algorithm.

1. Mở đầu

1.1. Giới thiệu về hệ thống địa lí GIS

Thông tin địa lí (GIS) bao gồm dữ liệu về bề mặt trái đất và các diễn giải dữ liệu để chúng trở lên dễ hiểu hơn. Nói chung, thông tin địa lí được thu thập từ bản đồ hay được thu thập thông qua đo đạc trực tiếp, viễn thám, điều tra, phân tích hoặc mô phỏng. Thông tin địa lí bao gồm hai loại: Dữ liệu không gian và phi không gian hay “ở đâu?” và “cái gì?” [1]. Độ phức tạp của thế giới thực là không gian có giới hạn. Càng quan sát thế giới gần hơn càng thấy được chi tiết hơn. Con người luôn mong muốn lưu trữ và quản lí đầy đủ các dữ liệu về thế giới thực. Nhưng sẽ dẫn đến phải có cơ sở dữ liệu rất lớn để lưu trữ mọi thông tin chính xác về chúng. Do vậy để lưu trữ được dữ liệu không gian của thế thực vào máy tính thì phải giảm số lượng dữ liệu đến mức có thể quản lí được.

Bài viết này trình bày tóm lược về ý tưởng cũng như một số ví dụ mô tả cho các phương pháp (quy trình mang tính thuật toán) đơn giản hoá đường cong trên mô hình dữ liệu Vector. Từ đó, giúp giảng viên có những định hướng trong dạy học một số phương pháp đơn giản hoá đường cong trên mô hình Vector cho sinh viên ngành Hệ thống thông tin Địa lí, góp phần nâng cao chất lượng dạy học và khả năng nghiên cứu khoa học của sinh viên.

1.2. Mô hình dữ liệu vector

Mô hình dữ liệu Vector dựa trên cơ sở các vector hay tọa độ của các điểm trong một hệ trục tọa độ nào đó. Điểm là thành phần sơ cấp của dữ liệu địa lí trong mô hình này. Các đối tượng này được hình thành bằng

cách nối các điểm bởi các đoạn thẳng, một vài hệ thống không cho phép nối các điểm với các cung tròn. Vùng được xác định bởi tập các đường, khái niệm đa giác đồng nghĩa với dùng trong cơ sở dữ liệu vector vì ta sử dụng các đoạn thẳng (hay cạnh) nối các điểm với nhau. Như vậy, mô hình dữ liệu Vector sử dụng các đoạn thẳng hay các điểm rời rạc để nhận biết được thế giới thực. Khác với mô hình Raster, mô hình dữ liệu Vector có thể cho biết “nơi mà mọi thứ xảy ra” [2].

Cấu trúc của một đối tượng cơ sở trong mô hình dữ liệu Vector như sau: Kí hiệu: [...] bộ, <...> danh sách, {...} tập, danh sách các điểm $\langle P_1, P_2, P_3, \dots, P_n \rangle$, $i = 1, \dots, n$ trong đó P_i là đỉnh; mỗi cặp (P_i, P_{i+1}) với $i < n$ là cạnh; Point: [x : real, y : real], polyline: <point>, poplyregion: <point> và Region: {polygon} [3].

2. Giới thiệu bài toán và một số khái niệm cơ bản

2.1. Giới thiệu bài toán và cách giải bài toán

Giới thiệu bài toán: Bài toán đặt ra là làm thế nào để có một giải pháp làm tăng tốc độ hiển thị dữ liệu thông tin địa lí, giảm thiểu dung lượng lưu trữ thông tin và khả năng nắm bắt được thông tin trên bản đồ cũng như dữ liệu bản đồ mà đảm bảo được độ tin cậy của thông tin đến người dùng.

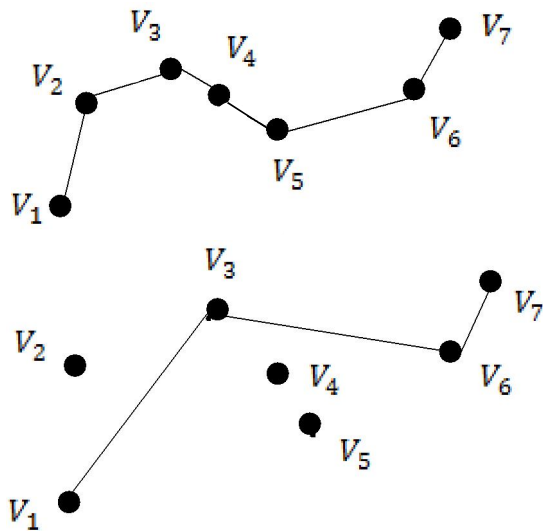
Giải quyết bài toán: Các nhà nghiên cứu về lĩnh vực thông tin địa lí cũng đã đưa ra nhiều giải pháp để giải quyết bài toán được đưa ra ở phần trên.

2.2. Đơn giản hóa đường cong

Đơn giản hóa đường cong là thao tác quan trọng nhất của khái quát hóa phần lớn các đặc trưng bản đồ

* Trường Đại học Mở - Địa chất

được biểu diễn trực tiếp bằng các đường hay hình thành các đa giác từ các đường bao quanh. Đơn giản hóa làm giảm tổng số đường chi tiết, loại bỏ đi các cặp tọa độ không cần thiết theo tiêu chí hình học nào đó như khoảng cách giữa hai điểm, độ lệch khỏi đường trực. Thí dụ đơn giản hóa đường cong như sau: Giả sử có đường gấp khúc C hình thành từ hai đầu mút và tập đỉnh V tùy ý đơn giản C thành C' bằng cách giảm số V thành V' và vẫn giữ nguyên hai đầu mút như vậy V' là tập con của V. Không tạo ra đỉnh mới.



Hình 1. Đơn giản hóa đường cong nhờ giảm thiểu tọa độ

2.3. Nguyên nhân của việc đơn giản hóa đường cong

Yêu cầu đặt ra là làm thế nào mà hiển thị bản đồ nhanh mà không làm ảnh hưởng đến độ tin cậy của bản đồ. Bởi vậy nhiệm vụ và mục tiêu của việc đơn giản hóa đường cong là: - Tăng tốc độ hiển thị bản đồ; - Làm tăng khả năng nắm bắt thông tin trên bản đồ khu vực mà ta quan tâm; - Giảm thiểu các điểm dư thừa, các đường thẳng không cần thiết.

3. Một số thuật toán đơn giản hóa đường cong trên mô hình dữ liệu Vector

3.1. Thuật toán Douglas_Peucker [1]

Dữ liệu đầu vào: Đường cong C được lập từ $V = \{V_1, \dots, V_n\}$

Dữ liệu ra: Đường cong C' được lập từ $V' = \{V_1, \dots, V_n\}$, $V' \subseteq V$

Bước khởi tạo:

- Nối V_1 với V_n bằng đường cơ sở (lập đường cơ sở)

Bước lặp:

- Nếu khoảng cách từ các đỉnh tới đường cơ sở nhỏ hơn ϵ thì chúng được loại bỏ, xâu đoạn thẳng được thay thế bằng đường cơ sở.

- Nếu có đỉnh nào ngoài đường cơ sở thì sẽ bẻ đôi đường cong tại đỉnh xa đường cơ sở nhất.

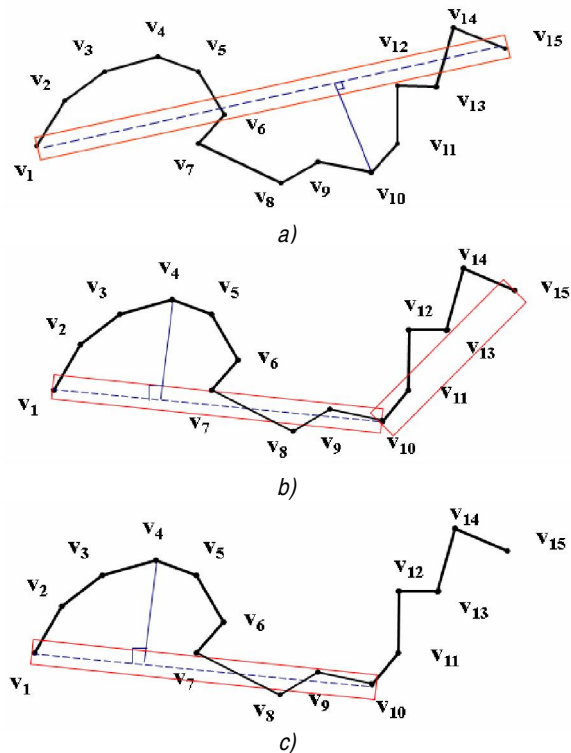
- Tiến trình lặp lại cho cả hai xâu đoạn thẳng.

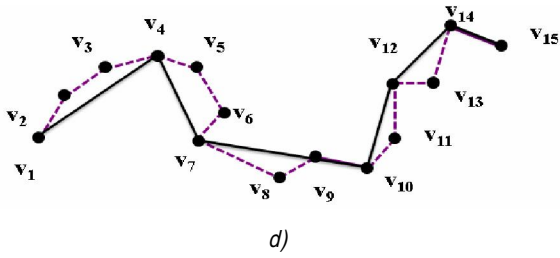
Chẳng hạn, cho đường cong V được tạo bởi các điểm từ V_1 đến V_{15} . Đầu tiên ta nối V_1 với V_{15} bằng đường cơ sở (hình a), sau đó ta xét:

- Nếu khoảng cách từ các đỉnh tới đường cơ sở nhỏ hơn thì chúng được loại bỏ, xâu đoạn thẳng được thay thế bằng đường cơ sở.

- Nếu có đỉnh nào ngoài đường cơ sở thì sẽ bẻ đôi đường cong tại đỉnh xa đường cơ sở nhất, tiến trình lặp lại cho cả hai xâu đoạn thẳng.

Chẳng hạn, trong hình 2, ta có: a) V_{10} ở xa đường cơ sở nhất, bẻ đôi xâu đoạn thẳng tại V_{10} thực hiện lặp cho cả hai nửa xâu đoạn thẳng (V_1, V_{10}) và (V_{10}, V_{15}) , xem hình b). Tiếp tục, ta thực hiện một cách tương tự đối với đường cong được tạo bởi xâu đoạn thẳng nối từ V_1 đến V_{10} để loại bỏ các điểm nằm ngoài đường cơ sở. Cuối cùng ta cũng ta thực hiện một cách tương tự đối với đường cong được tạo bởi xâu đoạn thẳng nối từ V_{10} đến V_{15} . Kết quả hình d) $V' = \{V_1, V_4, V_7, V_{10}, V_{12}, V_{14}, V_{15}\}$ là đường thay thế xâu đoạn thẳng ban đầu, các nút không được nối là các nút được loại bỏ.





Hình 2

3.2. Thuật toán Lang [1]

Thuật toán này tìm kiếm các điểm “láng giềng” bằng cách đánh giá các khoảng cách từ các điểm này đến đường cơ sở. Phạm vi tìm kiếm phụ thuộc vào tiêu chuẩn khoảng cách và được điều khiển bởi tham số “Tìm trước” (Lock ahead - Số phần tử của đoạn đầu để xử lý) của thuật toán. Giá trị lọc là khoảng cách dung sai vuông góc (ϵ). Do vậy thuật toán Lang còn được gọi là thuật toán dải băng dung sai (hình 3).

Dữ liệu đầu vào: Đường cong C được lập từ $V = \{V_1, \dots, V_n\}$;

Dữ liệu ra: Đường cong C' được lập từ $V' = \{V_1, \dots, V_n\}$, $V' \subseteq V$;

Bước khởi tạo:

- Tìm một tham số k (số điểm thành lập đường cơ sở) để lập đường cơ sở và gán giá trị ban đầu cho khoảng cách dung sai epsilon (ϵ);
- Lập đường cơ sở $[P_1, P_{k+1}]$.

Bước lọc:

- Tìm xem có điểm nào nằm ngoài đường cơ sở hay không?

Nếu có thì

- Giữ lại điểm đầu của đường cơ sở;
- Loại bỏ hết các điểm nằm trong đường cơ sở;
- Lập đường cơ sở mới, điểm đầu là điểm cuối đường cơ sở cũ, điểm cuối là điểm cuối của đường cơ sở cũ tịnh tiến k điểm.

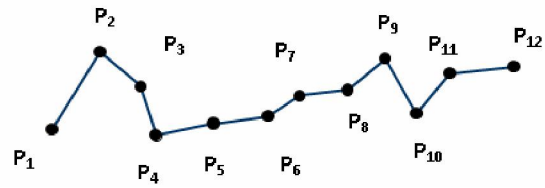
Mặt khác

- Giữ lại điểm đầu;
- Lập đường cơ sở mới là đường ngắn hơn đường cơ sở cũ một điểm.

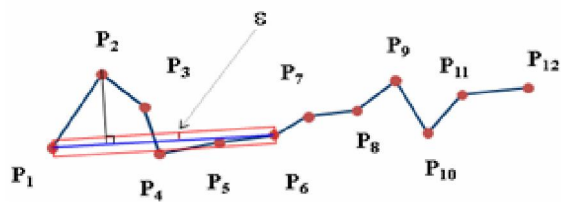
Quá trình xử lý lặp đi lặp lại cho đến khi đường cong không còn điểm nào để xử lý. Kết quả cuối cùng là một xâu các điểm cần giữ lại.

Ví dụ (hình 3): Giả sử cho đường cong V được tạo bởi các điểm từ V_1 đến V_{12} (hình 3.a); $V = \{P_1, P_2, P_3,$

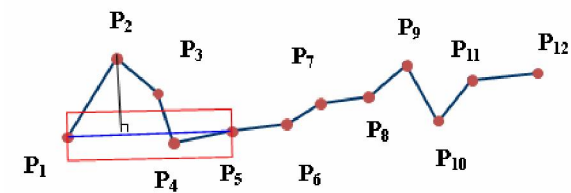
$P_4, P_5, P_6, P_7, P_8, P_9, P_{10}, P_{11}, P_{12}\}$. Ban đầu ta chọn số điểm ban đầu $k = 6$; sau đó lập đường cơ sở $[P_1, P_6]$ (hình 3.b); Tiếp theo ta tính khoảng cách từ các điểm láng giềng P_2, P_3, P_4, P_5 để tìm các điểm; vì các khoảng cách này đều lớn hơn ϵ do đó không giữ được điểm nào ta giữ điểm đầu của đường cơ sở và lập đường cơ sở mới là $[P_1, P_5]$. Quá trình này tiếp tục đến khi nào tìm kiếm hết tất cả các điểm trên đường cong C; Kết thúc thuật toán ta giữ lại được các điểm lưu trong $V' = \{P_1, P_2, P_4, P_9, P_{10}, P_{11}, P_{12}\}$, (xem hình 3.e).



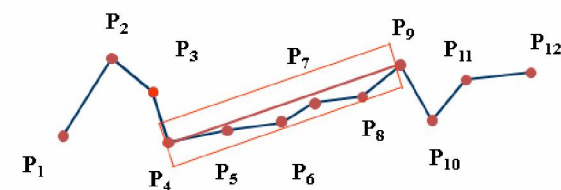
a)



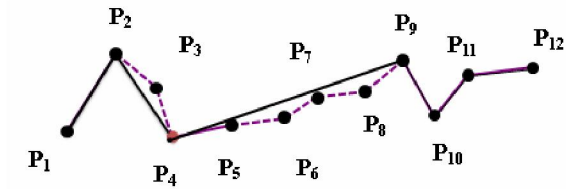
b)



c)



d)



e)

Hình 3

3.3. Thuật toán đơn giản theo hệ số góc Angle [1]

Đây là thuật toán đơn giản hóa đường cong theo tiêu chuẩn số đo góc cho trước ngưỡng α .

Dữ liệu đầu vào: Đường cong C được lập từ $V = \{V_1, \dots, V_n\}$. Điểm đầu tiên là P_1 và điểm cuối cùng là P_n .

Dữ liệu ra: Đường cong C' được lập từ $V' = \{V_1, \dots, V_n\}$, $V' \subseteq V$;

Bước khởi tạo:

- Đặt ngưỡng α ;
- Giữ lại điểm đầu P_1

Bước 2.

- Tính góc φ giữa điểm P_i, P_{i+1} và P_{i+1}, P_{i+2} ; trong đó P_{i+1} là điểm trung gian;

Nếu $\varphi > \alpha$ thì

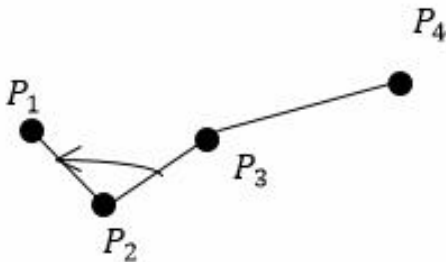
- + Loại bỏ đỉnh P_{i+1}
- + Tăng chỉ số i lên 1 ($i++$)

Ngược lại

- + Giữ lại điểm trung gian P_{i+1} ;
- + Chuyển $i++$.

Quá trình cứ tiếp tục đến khi xử lý đến hết xâu thì thuật toán dừng lại và đưa ra xâu kết quả V' .

Ví dụ: Điểm P_2 được kiểm tra bằng cách tính góc φ là góc được tạo bởi P_1, P_2 và P_2, P_3 . Nếu như góc φ mà có số đo lớn hơn thì ta loại bỏ điểm P_2 ngược lại ta giữ lại điểm đó, rồi tiếp tục thực hiện kiểm tra điểm tiếp theo của đường cong đến điểm gần cuối cùng nhất (hình 4).



Hình 4. Đơn giản hoá đường cong theo hệ số góc

4. Kết quả nghiên cứu và định hướng trong dạy học cho sinh viên ngành Hệ thống thông tin địa lí

Có thể thấy rằng, hai phương pháp đơn giản hóa đường cong rất hữu hiệu như Douglas_Peucker, Lang hoạt động tương đối giống nhau và kết quả phụ thuộc và ngưỡng ε cho trước. Thuật toán đơn giản hóa đường cong theo tiêu chuẩn hệ số góc thì hoạt động tương đối tốt qua nhiều lần thí nghiệm. Nghiên cứu

này có thể đóng góp vào việc giải quyết bài toán đặt ra là: làm thế nào để có một phương pháp làm tăng tốc độ hiển thị dữ liệu thông tin địa lí, giảm thiểu dung lượng lưu trữ thông tin và tăng khả năng nắm bắt được thông tin trên bản đồ nhanh nhất mà vẫn đảm bảo tính chính xác và tin cậy của thông tin.

Trong dạy học, giảng viên cần chú ý tới việc trang bị cho sinh viên các quy trình thuật toán trên một cách chủ động và thực tiễn: Sinh viên tự nghiên cứu, vận dụng các thuật toán trong suốt quá trình học tập. Cụ thể: - Sinh viên học theo nhóm (các nhóm sẽ đọc tài liệu, tóm tắt và báo cáo trước lớp về nội dung của các thuật toán này); - Giảng viên và sinh viên thảo luận với nhau về các thuật toán và đưa ra các ứng dụng của thuật toán; - Sau khi các nhóm sinh viên đã báo cáo trước lớp thì giảng viên giao cho sinh viên về nhà làm bài tập lập trình bằng ngôn ngữ lập trình nào đó [7], [8]; - Giảng viên tổ chức và hướng dẫn cho sinh viên lập trình và kiểm tra các nhóm làm bài tập. □

Như vậy, từ việc nghiên cứu các thuật toán đơn giản hoá đường cong trên mô hình dữ liệu Vector, giảng viên có thể tổ chức cho sinh viên tự mình khai thác thông tin, tổ chức lập trình, vận dụng trong thực tiễn học tập. Việc này vừa góp phần nâng cao chất lượng dạy học vừa góp phần tổ chức cho sinh viên làm quen với công tác nghiên cứu khoa học. □

Tài liệu tham khảo

- [1] Nguyễn Trường Xuân (2014). *Công nghệ 3S*. NXB Khoa học và Kỹ thuật.
- [2] Đặng Văn Đức (2001). *Hệ thống thông tin địa lí GIS*. NXB Khoa học và Kỹ thuật.
- [3] Nguyễn Ngọc Thạch - Nguyễn Tuấn Dũng - Nguyễn Mạnh Cường (2001). *Viễn thám và hệ thống tin địa lí ứng dụng*. Trường Đại học Khoa học tự nhiên, Đại học Quốc gia Hà Nội.
- [4] Nguyễn Ngọc Thạch - Nguyễn Tuấn Dũng - Nguyễn Mạnh Cường (2001). *Viễn thám và hệ thống tin địa lí ứng dụng*. Trường Đại học Khoa học tự nhiên, Đại học Quốc gia Hà Nội.
- [5] Nguyễn Thế Thịnh (2002). *Cơ sở hệ thống thông tin địa lí*. NXB Khoa học và Kỹ thuật.
- [6] Charles Petzold (1998). *Programming Windows*. Fifth Edition, Microsoft Press.
- [7] Jeff Prosise (1999). *Programming Windows with MFC*. Microsoft Press.
- [8] Christopher Jones, C (1997). *Geographical Information Systems and Computer cartography*. Addison Wesley Longman.