NGHIÊN CỨU MÔ HÌNH TỔNG QUÁT HÓA DỮ LIỆU TỰ ĐỘNG TRONG XÂY DỰNG CƠ SỞ DỮ LIỆU NỀN ĐỊA LÝ QUÂN SỰ

ThS. NGUYỄN THỊ LAN PHƯƠNG
Cục Bản đồ-BTTM

Tóm tắt:

Bài báo mô tả kết quả nghiên cứu về các mô hình tổng quát hóa tự động trong xây dựng cơ sở dữ liệu địa lý quân sự. Các mô hình này hỗ trợ linh hoạt thao tác tổng quát hóa tự động để người dùng tạo ra cơ sở dữ liệu có độ phân giải khác nhau. Phần cuối của bài báo mô tả cấu trúc dữ liệu để hỗ trợ các hoạt động tổng quát hóa trong bối cảnh của môi trường tương tác này.

1. Đặt vấn đề

Mục tiêu tự động hóa trong xây dựng cơ sở dữ liệu (CSDL) luôn dành được sự quan tâm của các nhà khoa học trong và ngoài nước với nhiều công trình nghiên cứu có giá trị khoa học cao được ghi nhận. Tuy nhiên, quy trình xây dựng CSDL luôn có những yêu cầu đặc thù phụ thuộc vào nguồn tư liệu gốc, nội dung, cấu trúc và mục đích sử dụng. Ngành Bản đồ quân sự hiện đang tập trung vào xây dựng CSDL nền địa lý quân sự tỷ lệ tương ứng và nhỏ hơn từ CSDL nền địa lý quốc gia phủ trùm tỷ lệ 1:10.000. Để nâng cao mức độ tự động hóa trong quy trình xây dựng CSDL, việc khai thác các công cụ của phần mềm ArcGIS là một trong những hướng nghiên cứu thiết thực.

2. Giải quyết vấn đề

2.1. Nguyên tắc xây dựng mô hình tổng quát hóa tự động

Theo Gruenrich (1985): "Tổng quát hoá dữ liệu là quá trình xây dựng mô hình thứ sinh từ mô hình cơ sở theo mục đích sử dụng" [3]. Trong đó, các trường hợp đòi hỏi phải thực hiện thao tác TQH để đảm bảo các mối quan hệ không gian giữa các đối tượng đó là: sự chồng đè, hợp nhất, xung đột, biến đổi, không thống nhất và cường điệu hóa [2]. Đáng lưu ý, yếu tố xung đột không gian giữa các đối tượng sau TQH được phân chia thành các loại hình cơ bản như: giao cắt, tiếp xúc và hợp nhất. Nguyên nhân xung đột không gian sau TQH phát sinh từ các biến đổi hình học, làm cho đối tượng trở nên quá nhỏ, quá ngắn, quá hẹp hoặc quá gần, là hệ quả khi thu nhỏ tỷ lệ dữ liệu và các hoạt động TQH [4]. Khi đó, tiêu chí đánh giá về kích thước đối tượng và khoảng cách giữa các đối tượng được xét trong một bối cảnh cụ thể về độ phân giải dữ liệu.

Để tiến hành tổng quát hóa gắn với các giải pháp tự động, cần xác lập mối quan hệ không gian giữa các đối tượng trước và sau tổng quát hóa, các trường hợp biến đổi dạng hình học cơ bản sau TQH (đường thành điểm, vùng thành điểm/đường). Bên cạnh đó, dữ liệu thuộc tính sau tổng quát hóa có thể bị thay đổi, chẳng hạn những thuộc tính không thuộc tiêu chí thu nhận bị loại bỏ hoặc một số thuộc tính bị thay đổi do ảnh hưởng của sự biến đổi kiểu dữ liệu hình học.

Các đối tượng dữ liệu địa lý được phân loại theo dạng cơ bản là điểm, đường, vùng và thông tin thuộc tính. Chúng cũng có thể phân loại theo đặc trưng phân bố trong không gian bao gồm các đối tượng không có dạng mạng lưới và có dạng mạng lưới. Các thao tác TQH bản đồ gắn với giải pháp

Người phản biện: TS. Đồng Thị Bích Phương

tự động hóa bao gồm tổng quát hóa khái niệm và tái phân loại, chọn bỏ đối tượng, giản hóa đường, gộp các đối tượng dạng vùng, dịch chuyển, phóng đại. Các đối tượng có dạng mạng lưới có thêm các thao tác : (1) Phân tích hệ thống mạng lưới trên các tiêu chí đặc trưng cấu trúc mạng, mật độ, mức độ sử dụng; (2) Lọc bỏ theo các chỉ tiêu về mật độ, mức độ sử dụng; (3) Ràng buộc mối hệ không gian với các đối tượng lân cận có liên quan; (4) Giản hóa đường theo cấu trúc mạng lưới đặc trưng [1], [2], [4], [5]. *(Xem hình 1)*

Mô hình tổng quát hóa dữ liệu gắn với giải pháp tự động theo đó được chia thành hai tầng:

- Tầng tạo quan hệ không gian: nhằm đáp ứng yêu cầu tạo quan hệ ràng buộc dữ liệu giữa các đối tượng địa lý trên cùng tỷ lệ, giữa các tỷ lệ với nhau và đồng bộ dữ liệu tham chiếu giữa các đối tượng ràng buộc trong cơ sở dữ liệu.

- Tầng tạo mô hình tổng quát hóa: các giải pháp tạo mô hình tổng quát hóa phải phù hợp với kiểu đối tượng nguyên thủy, kiểu đối tượng dẫn xuất và mối quan hệ biến đổi kiểu đối tượng không gian sau các thao tác tổng quát hóa.

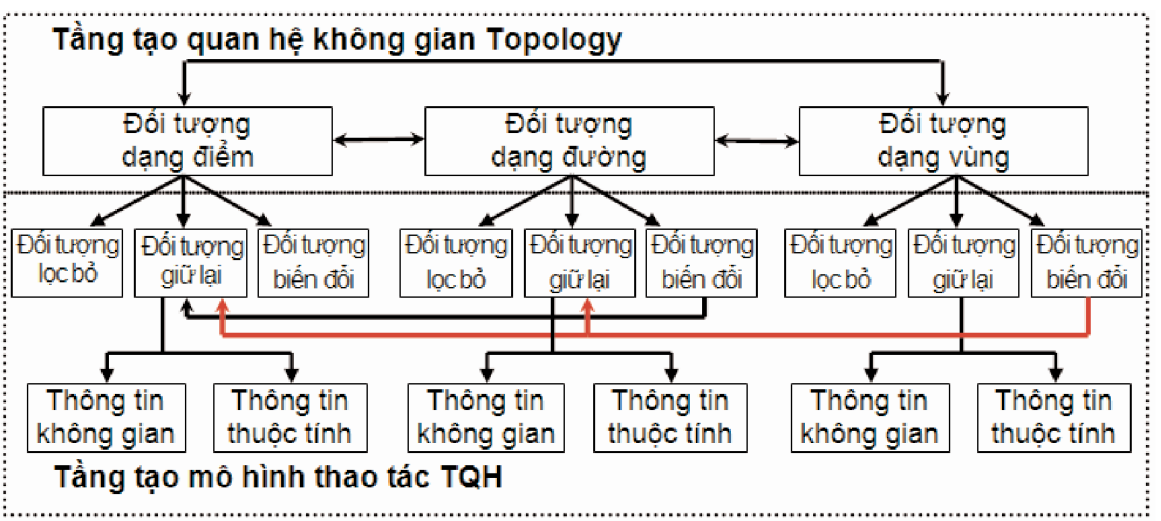
2.2. Ứng dụng phần mềm ArcGIS xây dựng mô hình tổng quát hóa tự động

Mô hình được xây dựng trên cơ sở xây dựng các chuỗi công việc, bao gồm các thao tác tổng quát hóa cho từng đối tượng, hoặc một vài đối tượng cùng một lúc trong modul Model Builder. Kết quả biến đổi của đối tượng có cấp độ quan trọng cao hơn được sử dụng là đích để tổng quát hóa đối tượng cấp thấp hơn. Các thao tác TQH bao gồm:

Tổng quát hóa khái niệm: là việc làm tăng mức độ khái quát của dữ liệu khi chuyển từ tỷ lệ lớn hơn sang tỷ lệ nhỏ hơn, bản chất là sự chọn hoặc bỏ đối tượng ở quy mô lớp thông tin (Feature Selection).

Tái phân loại: là một khía cạnh khác của tổng quát hóa khái niệm, bao gồm việc lọc bỏ đối tượng khi chuyển sang tỷ lệ nhỏ hơn mà hệ thống phân loại thay đổi, với bản chất là sự chọn bỏ ở quy mô thuộc tính (Attribute Selection).

Chọn bỏ: là việc giữ lại hay loại bỏ bớt một số đối tượng khi chuyển từ tỷ lệ lớn hơn sang tỷ lệ nhỏ hơn theo kích thước và cấp



Hình 1: Mô hình tổng quát hóa dữ liệu gắn với các giải pháp nâng cao tự động hóa

độ quan trọng của chúng theo mục đích sử dụng, với bản chất là sự chọn bỏ ở quy mô đối tượng (Object Selection).

Gộp: là việc hợp nhất hai hay nhiều đối tượng lân cận theo một giá trị giới hạn.

Giản hóa đường: là việc khái quát hóa hình dạng đối tượng dạng đường, đường bao của đối tượng dạng vùng khi dữ liệu chuyển sang tỷ lệ nhỏ hơn.

Dịch chuyển: là việc xê dịch vị trí không gian của đối tượng nhằm đảm bảo mối quan hệ tương quan giữa các đối tượng lân cận sau thực hiện giản hóa đường có phát sinh mâu thuẫn. *(Xem hình 2)*

Sau các hoạt động TQH theo mô hình trên các thông tin thuộc tính đi kèm với các đối tượng được bảo toàn. Riêng đối với thao tác gộp, đối tượng bị gộp phải được gán thông tin của nhóm mới.

3. Thử nghiệm xây dựng CSDL nền địa lý quân sự tỷ lệ 1:25.000 từ CSDL nền địa lý tỷ lệ 1:10.000

3.1. Nguyên tắc tổ chức CSDL phục vụ tổng quát hóa

Để thuận tiện cho các thao tác tổng quát hóa tự động, cần tổ chức một CSDL dùng để lưu các kết quả tổng quát hóa trung gian. Do việc tạo các quy tắc topology cho các đối tượng trong cùng một FeatureDataset sẽ thuận lợi cho kiểm tra quan hệ không gian giữa các feature class, nên CSDL dùng để tổng quát hóa cần được tổ chức theo ba chủ đề như sau:

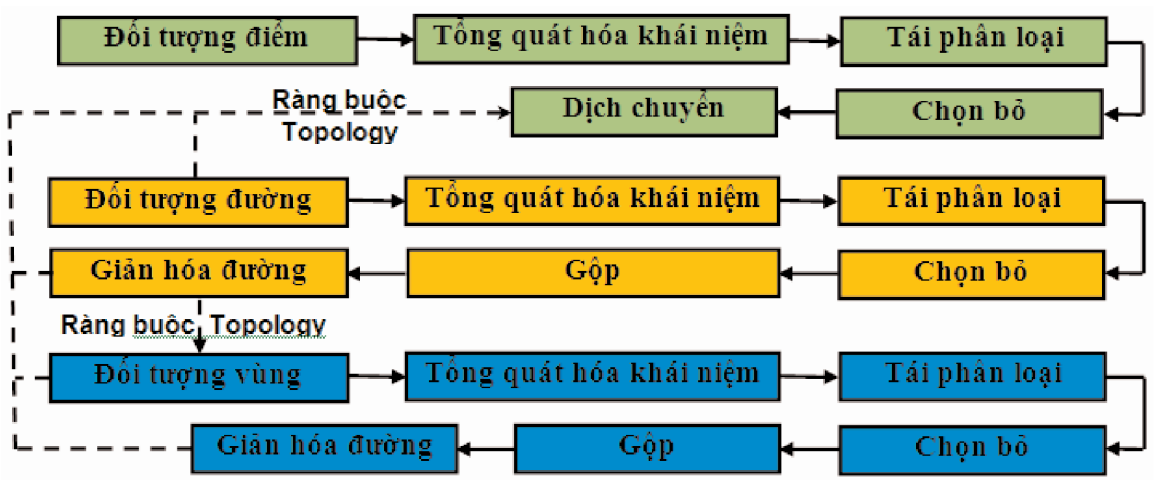
- Chủ đề "CoSoDoDac" bao gồm các đối tượng thuộc chủ đề Cơ sở đo đạc, được thu nhận toàn bộ nên không tác động

- Chủ đề "DiaHinh" bao gồm các đối tượng thuộc chủ đề Địa hình, khi thực hiện các thao tác tổng quát hóa có ràng buộc biến đổi liên quan đến Thủy hệ. Do biến đổi của hai đối tượng này phức tạp nên sau TQH cần rà soát xung đột không gian.

- Chủ đề "DuLieu" bao gồm các đối tượng còn lại.

Các đối tượng giản hóa thuộc tính thì loại bỏ trực tiếp trường thuộc tính trên bản sao. Nếu các thuộc tính xuất hiện ở hai tỷ lệ nhưng miền giá trị khác nhau được lập bảng ánh xạ để chuyển đổi giá trị thuộc tính.

CSDL nền địa lý sử dụng làm tư liệu gốc đặt cùng một thư mục với CSDL dẫn xuất. Thao tác tổng quát hóa sẽ tiến hành trên CSDL tư liệu gốc, nhưng sau khi quá trình



Hình 2: Sơ đồ hoạt động của mô hình tổng quát hóa

tổng hợp dữ liệu hoàn tất thì tái tổ chức CSDL dẫn xuất theo đúng mô hình cấu trúc dữ liệu quy định.

3.2. Mô hình thử nghiệm

Tổng quát hóa riêng từng đối tượng có tham chiếu đến kết quả tổng quát hóa của các đối tượng khác, cụ thể: DuongBinhDo, DiaHinhDacBiet (chủ đề Địa hình) sử dụng kết quả tổng quát hóa của Songsuoi, KenhMuong, DuongBoNuoc, DuongMepNuoc (chủ đề Thủy hệ). Tổng quát hóa DoanTimDuongBo, DoanVuotSongSuoi (chủ đề Giao thông) phải sử dụng kết quả tổng quát hóa của SongSuoi, KenhMuong, DuongBoNuoc, DuongMepNuoc (chủ đề thủy hệ). Các lớp thông tin thuộc chủ đề Biên giới Địa giới sử dụng kết quả tổng quát hóa của các lớp thông tin thuộc chủ đề Địa hình, Giao thông, Thủy hệ. Các lớp thông tin thuộc chủ đề Dân cư, Cơ sở hạ tầng sử dụng kết quả tổng quát hóa của chủ đề giới Địa giới, Thủy hệ, Giao thông. Các lớp thông tin thuộc chủ đề Phủ bề mặt sử dụng kết quả tổng quát hóa chủ đề Giao thông, Thủy hệ, Dân cư Cơ sở hạ tầng.

Từ các phân tích ràng buộc quan hệ không gian biến đổi sau TQH như trên, thứ tự thực hiện TQH các đối tượng cần theo thứ tự đường, vùng và điểm, vì đối tượng đường tạo nên các giới hạn của đối tượng vùng, còn đối tượng điểm có khả năng xảy ra xung đột không gian sau khi đối tượng đường và vùng biến đổi do khái quát hóa hình học. Các đối tượng được TQH theo các chủ đề sẽ sắp xếp như sau: (1) Thủy hệ; (2) Địa hình; (3) Giao thông; (4) Dân cư Cơ sở hạ tầng; (5) Biến giới và địa giới; (6) Phủ bề mặt; (8) Quân sự.

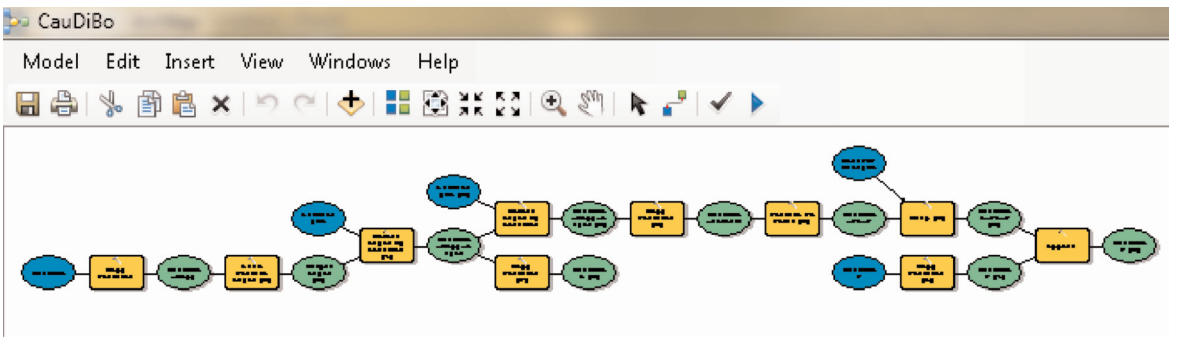
Sử dụng công cụ Model Builder (trong ArcGIS) để lập trình xây dựng các mô hình tổng quát hóa tự động trên cơ sở kết nối các công cụ và tham số thu nhận thông tin. Trong đó, các thông tin thu nhận đối tượng tuân theo quy định kỹ thuật xây dựng CSDL nền địa lý quân sự. *(Xem hình 3)*

Tổng số 50 mô hình thử nghiệm thuộc các chủ đề Dân cư cơ sở hạ tầng (10), Địa hình (3), Giao thông (12), Phủ bề mặt (1), Thủy hệ (20), Quân sự (01), Biên giới địa giới (03), đã giải quyết cơ bản các bài toán về tổng quát hóa xây dựng CSDL nền địa lý quân sự tỷ lệ 1:25000.

3.3. Kết quả thử nghiệm

Khu vực thử nghiệm thuộc phạm vi tương ứng với mảnh bản đồ 6232 I TN (Tân Biên), tỷ lệ 1:25.000. Đây là mẫu điển hình về dạng địa hình chuyển tiếp giữa cao nguyên và đồng bằng Nam Bộ, bao gồm đầy đủ các dạng địa hình núi, đồng bằng, dân cư kiểu nông thôn và kiểu thành thị. Đồng thời, đây cũng là một trong những khu vực nhạy cảm, đòi hỏi phải thường xuyên cung cấp thông tin địa hình quân sự mới nhất.

Tư liệu sử dụng là CSDL nền địa lý quốc



Hình 3: Minh họa thiết kế mô hình tổng quát hóa đối tượng CauDiBo trong ModelBuilder

gia tỷ lệ 1:10.000 (2010), do Bộ Tài nguyên và Môi trường sản xuất, được đánh giá đạt yêu cầu về độ chính xác, tính đầy đủ và các chuẩn kỹ thuật. Dữ liệu này được chuẩn hóa lại theo cấu trúc của CSDL nền địa lý quân sự tỷ lệ 1:10.000. Trong đó, các dữ liệu thuộc chủ đề Quân sự được tách lọc về đúng Feature quy định và gán mã đối tượng tương ứng.

Kết quả TQH cho thấy các đối tượng biến đổi không gian theo tiêu chí thu nhận được chuyển đổi sang quản lý dạng hình học tương ứng một cách đầy đủ. Mức độ tự động hóa đối tượng tối thiểu đạt 50% khối lượng công việc. *(Xem bảng 1)*

Ưu điểm của mô hình tổng quát hóa là các thao tác tổng quát hóa được liên kết theo từng chuỗi công việc chuyên dụng nên rút ngắn đáng kể thời gian thực hiện so với các thao tác thủ công riêng rẽ. Tuy nhiên, phương pháp này tác động đến toàn bộ khối dữ liệu với cùng một hệ thống tham số đăng

Bảng 1: Thống kê kết quả TQH tự động các đối tượng của CSDL nền địa lý quân sự tỷ lệ 1:25.000 từ tỷ lệ 1:10.000

STT	Loại đối tượng	Số đối tượng tự động TQH	Tự động (%)
I	**Dân cư, Cơ sở hạ tầng**		
1	KhuChucNangP	10/20	50
2	TramDienP	6/8	75
3	TramThuPhatSongP	9/11	82
4	NhaP	1846/2846	65
5	NhaA	150/150	100
6	KhuChucNangA	71/71	100
7	CongTrinhKienTrucDacBietP	1/1	100
8	RangGioiKhuChucNangL	81/81	100
9	DiemDanCuA	92/92	100
10	TramDienA	10/10	100
11	DuongDayTaiDienL	90/90	100
II	**Địa hình**		
1	DiaHinhDacBietA	9/9	100
2	DiemDoCaoP	995/2000	50
3	DuongBinhDoL	64/90	71
III	**Biên giới địa giới**		
1	DiaPhanA	12/12	100
2	DuongDiaGioiL	16/16	100
IV	**Điểm cơ sở đo đạc**		
1	DiemCoSoQuocGiaP	9/9	100

STT	Loại đối tượng	Số đối tượng tự động TQH	Tự động (%)
V	**Giao thông**		
1	BenBaiP	1/1	100
2	BenBaiA	1/1	100
3	CauGiaoThongL	4/4	100
4	CauGiaoThongP	6/8	75
5	CongGiaoThongP	1/2	50
6	DoanTimDuongBoL	606/606	100
7	NutMangDuongBo	1907/1907	100
VI	**Phủ bề mặt**		
1	PhuBeMat	529/822	64
2	RangGioiPhuBeMat	1574/2000	79
VII	**Thủy hệ**		
1	CongThuyLoiP	1/2	50
2	DuongBoNuocL	59/59	100
3	KenhMuongL	60/65	92
4	MangDanNuocL	2/2	100
5	MatNuocTinhA	45/45	100
6	SongSuoiL	12/15	80
7	SongSuoiA	8/8	100
VIII	**Quân sự**	3/3	100

nhập trước. Muốn tác động của tổng quát hóa tự động hiệu quả theo đặc trưng của từng phân vùng phải chia nhỏ phân vùng dữ liệu và đặt chế độ cho phép tác động.

4. Kết luận

Việc sử dụng modul Model Builder của phần mềm ArcGIS để thiết kế các mô hình tổng quát hóa dữ liệu tự động (trong xây dựng dữ liệu cho các CSDL dẫn xuất) là phù hợp với điều kiện thực tế về trang thiết bị và nhân lực hiện có của ngành Địa hình quân sự.

Cách thức tổ chức thành 3 CSDL: CSDL gốc, CSDL trung gian và CSDL kết quả, là cơ sở để các mô hình thiết kế trước hoạt động ổn định. Trong đó, CSDL trung gian được chia thành các chủ đề chuyên dụng đã đảm bảo các hoạt động tổng quát hóa tự động được tối ưu hơn.

Các mô hình tổng quát hóa tự động được thiết kế trên cơ sở các chuỗi công việc chuyên dụng, được thực hiện liên tiếp đã góp phần nâng cao khả năng tự động hóa trong sản xuất tư liệu địa hình. Trong quá trình sử dụng mô hình, tác nghiên viên có thể chỉnh sửa các mô hình này phù hợp với các mục đích xây dựng dữ liệu khác bằng cách thay đổi chuỗi công việc và tham số khai báo cho từng thao tác.

Các kết quả nghiên cứu trên có thể áp dụng để xây dựng các mô hình tương tự khi xây dựng dữ liệu địa lý các tỷ lệ cho các mục đích khác nhau, trên cơ sở thay đổi các tham số phù hợp.❍

Tài liệu tham khảo

[1]. Đồng Thị Bích Phương và nnk (2012), *Nghiên cứu hoàn thiện cấu trúc dữ liệu cho cơ sở dữ liệu nền địa lý gắn với các giải pháp tổng quát hoá dữ liệu tự động*, Báo cáo kết quả đề tài nghiên cứu khoa học cấp Bộ, Viện khoa học Đo đạc và Bản đồ, Bộ Tài nguyên và Môi trường, Hà Nội.

[2]. Kate Beard, William Mackaness (2013), *Generalization Operations and Supporting Structures,* National Center for Geographic Information and Analysis Department of Geography, State University of New York, Buffalo, United States.

[3]. Sheng Zhou, Christopher B. Jones (2009), *A Multi-Representation Spatial Data Model, Department of Computer Science*, Cardiff University, Cardiff, CF24 3XF, United Kingdom.

[4]. Robert B.McMaster, K.Stuart Shea (1989), *Cartographic Generalization in a Digital Enviroment: when & how to generalize*, Proceeding of 9th International Symposium on Computer - Assisted Cartography, Baltimore.

[5]. Robert B.McMaster, K.Stuart Shea (1992), "Generalization in digital cartography", *The Association of American Geographers*, Washington.❍

Summary

Research build automated generalization models for the military geographic database

MSc. Nguyen Thi Lan Phuong

Army Catographic Department

This paper describes a research of automated generalization models for the military geographic database. The models support flexibility for the performance of data generalization operations so users have opportunity for creating views of data at different levels of resolution. The final section of the paper describes data structures for supporting the operations within the context of this interactive environment.❍

Ngày nhận bài: 20/02/2014.