

MÔ HÌNH BIỂU DIỄN ĐỐI TƯỢNG KHÔNG GIAN DỰA TRÊN LÝ THUYẾT TẬP MỜ VÀ BIẾN NGÔN NGỮ

TRẦN VĨNH PHƯỚC, LƯU ĐÌNH HIỆP

Trung tâm Công nghệ Thông tin Địa lý (DITAGIS), Trường Đại học Bách Khoa TP.HCM

Abstract. GISystem is a new technology to represent and analyse the geographic information. However, the representation of geographic data based on classical set theory limits the effects of the reasoning and analysing procedures. The application of the fuzzy theory and linguistic variable will improve the intelligent level of GISystems. This paper presents the model representing the spatial objects based on the fuzzy theory and linguistic variable.

Tóm tắt. Hệ thống thông tin địa lý là công nghệ mới, được sử dụng để biểu diễn và phân tích các thông tin địa lý. Tuy nhiên, việc biểu diễn các dữ liệu địa lý dựa trên lý thuyết tập cổ điển chưa đủ để phát triển một số hiệu quả trong các quy trình lập luận và phân tích. Áp dụng lý thuyết tập mờ và biến ngôn ngữ sẽ cải tiến mức độ thông minh cho các hệ thống thông tin địa lý. Bài báo này trình bày mô hình biểu diễn đối tượng không gian dựa trên lý thuyết tập mờ và biến ngôn ngữ.

1. GIỚI THIỆU

Trong thực tế, khi biểu diễn các đối tượng không gian của thế giới thực thường xuất hiện các thông tin không chắc chắn. Áp dụng lý thuyết tập mờ và biến ngôn ngữ để mô hình hóa việc biểu diễn và quản lý dữ liệu không gian có thể chất lọc thêm được các dữ liệu trước kia thường bị bỏ qua vì chưa đủ công cụ phản ánh ngữ nghĩa của chúng ở dạng mịn. Lý thuyết tập mờ và biến ngôn ngữ còn cho phép người khai thác các hệ thống thông tin địa lý thực hiện truy nhập và cập nhật các thông tin chưa chắc chắn vào cơ sở dữ liệu. Bài viết này đề xuất mô hình biểu diễn các đối tượng không gian trong các hệ thống thông tin địa lý bằng lý thuyết tập mờ và biến ngôn ngữ.

Nội dung bài viết đề cập đến phương pháp biểu diễn hàm thuộc theo mô hình dữ liệu vector-topology và mô hình cơ sở dữ liệu mờ trong GIS cùng với một số bước phân tích hiển thị dữ liệu mờ.

2. MỘT SỐ KHÁI NIỆM CƠ BẢN

2.1. Dữ liệu GIS

Dữ liệu GIS bao gồm hai thành phần: dữ liệu không gian và dữ liệu thuộc tính ([1]). Dữ liệu không gian được biểu diễn dưới dạng đồ họa nhằm thể hiện vị trí, hình dạng, kích thước của các đối tượng một cách trực quan. Trong khi đó, dữ liệu thuộc tính được biểu diễn dưới dạng bảng hoặc ma trận để mô tả các thuộc tính bên trong của đối tượng. Hai thành phần dữ liệu này được liên kết chặt chẽ với nhau và có thể khai thác thông tin chéo thông qua các truy vấn không gian hoặc các biểu thức truy vấn trên thuộc tính.

Dữ liệu không gian được lưu trữ theo 2 mô hình: raster và vector. Mô hình dữ liệu raster, không gian được phân chia thành những phần tử đồng nhất trên cùng một thuộc tính. Các phần tử này có dạng hình học đối xứng, có kích thước bằng nhau và đủ nhỏ. Theo mô

hình dữ liệu vector, các đối tượng không gian được xác định bằng các giá trị tọa độ địa lý và được biểu diễn dưới các dạng hình học: điểm, đường, vùng ([1]).

Dữ liệu thuộc tính trong hệ thống thông tin địa lý phát triển dựa trên những thành quả và khái niệm của lý thuyết cơ sở dữ liệu. Hiện tại, dữ liệu thuộc tính được phát triển theo mô hình quan hệ với mỗi đối tượng không gian trong một lớp (layer) có quan hệ 1:1 với một bộ (tuple) trong bảng dữ liệu thuộc tính, các thành phần dữ liệu thuộc tính được thiết kế và khai thác theo các nguyên tắc xây dựng cơ sở dữ liệu.

2.2. Lý thuyết tập mờ và biến ngôn ngữ

Năm 1965, Lotfi A. Zadeh đã đưa ra khái niệm tập mờ trên cơ sở mở rộng lý thuyết tập hợp. Khi đó, mỗi phần tử trong tập hợp được bổ sung thêm một yếu tố khác là “độ thuộc” để chỉ mức độ phụ thuộc của phần tử vào tập hợp.

Cho tập $X = \{x\}$ là một tập nền (không gian nền), tập mờ A của X là một tập hợp bao gồm các phần tử $\{(x, \mu_A(x))\}$ với $x \in X$ và $\mu_A(x)$ là một hàm số thực có giá trị trong khoảng $[0, 1]$ để chỉ độ thuộc của từng phần tử x vào tập mờ A . Hàm số $\mu_A(x) : X \rightarrow [0, 1]$ còn được gọi là hàm thuộc.

Nếu hai tập mờ A và B của cùng không gian nền X thì ([2]):

- Hai tập mờ A và B bằng nhau (kí hiệu $\stackrel{f}{=}$) khi và chỉ khi: $\mu_A(x) = \mu_B(x), \forall x \in X$
- Tập A là tập con của B (kí hiệu $\stackrel{f}{\subseteq}$) khi và chỉ khi: $\mu_A(x) \leq \mu_B(x), \forall x \in X$
- Tập bù của A (kí hiệu \bar{A}) được xác định bởi: $\mu_{\bar{A}}(x) = 1 - \mu_A(x), \forall x \in X$
- Hợp của hai tập mờ A và B được xác định bởi:

$$\mu_{A \cup B}(x) = \text{Max}[\mu_A(x), \mu_B(x)], \forall x \in X$$

- Giao của hai tập mờ A và B được xác định bởi:

$$\mu_{A \cap B}(x) = \text{Min}[\mu_A(x), \mu_B(x)], \forall x \in X$$

Các khái niệm về tập mờ tạo cơ sở cho việc xác định các giá trị của biến mờ. Biến mờ cũng như các biến số thông thường nhưng mỗi trạng thái của biến được diễn tả thông qua các ngôn ngữ, nên thường gọi là biến ngôn ngữ. Mỗi giá trị của biến ngôn ngữ sẽ tương ứng với một tập mờ xác định trên cùng không gian nền.

Mỗi biến ngôn ngữ được xác định bởi một bộ 5 (v, T, X, g, m) , trong đó:

v : là tên biến ngôn ngữ.

T : là miền trị, là tập hợp tất cả các giá trị của biến ngôn ngữ.

X : là không gian nền.

g : là tập quy tắc để xây dựng các giá trị của biến.

m : quy tắc ngôn ngữ, bao gồm những quy tắc gán mỗi giá trị $t \in T$ vào một tập mờ trên X .

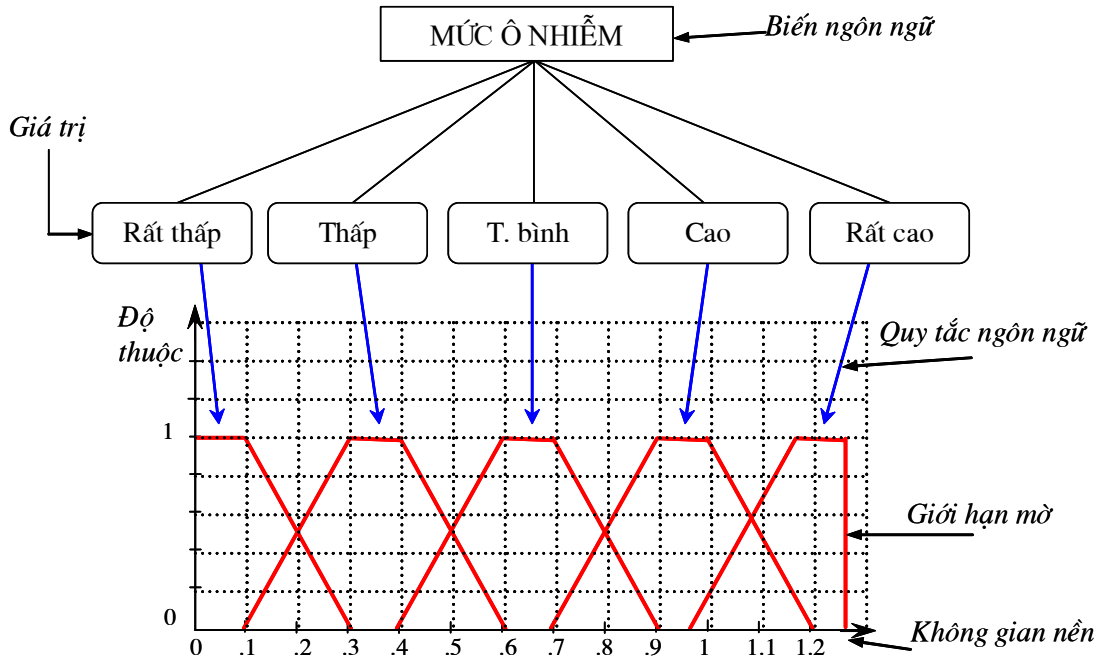
3. MÔ HÌNH BIỂU DIỄN ĐỐI TƯỢNG KHÔNG GIAN DỰA TRÊN LÝ THUYẾT TẬP MỜ VÀ BIẾN NGÔN NGỮ

3.1. Mô hình cơ sở dữ liệu mờ trong GIS

Các thông tin liên quan đến việc tổ chức, thiết kế cơ sở dữ liệu mờ bao gồm:

- Mã đối tượng: mã của đối tượng không gian.
- Thuộc tính x_i của đối tượng trên X : thuộc tính cần mờ hóa.

- Độ thuộc a_i của x_i vào tập mờ tương ứng với giá trị của biến ngôn ngữ.
- Các giá trị của biến ngôn ngữ: là trị ngôn ngữ của các biến ngôn ngữ.
- Mã của tập mờ: mã của các tập mờ trên không gian nền X .



Hình 1. Minh họa về biến ngôn ngữ

Trên cơ sở các thông tin liên quan đến việc thiết kế cơ sở dữ liệu, tiến hành lập bảng mô tả các thuộc tính như sau:

Thuộc tính	Diễn giải	Thuộc thực thể
FID	Mã đối tượng không gian	Đối tượng không gian (R1)
VALUE	Giá trị mờ hóa trên không gian nền	Đặc điểm (R2)
MEMBERSHIP	Độ thuộc của giá trị trên không gian nền vào tập mờ tương ứng	Đặc điểm (R2)
LINGUISTIC	Giá trị của biến ngôn ngữ	Biến ngôn ngữ (R3)
ID	Mã của các hàm thuộc	Biến ngôn ngữ (R3)

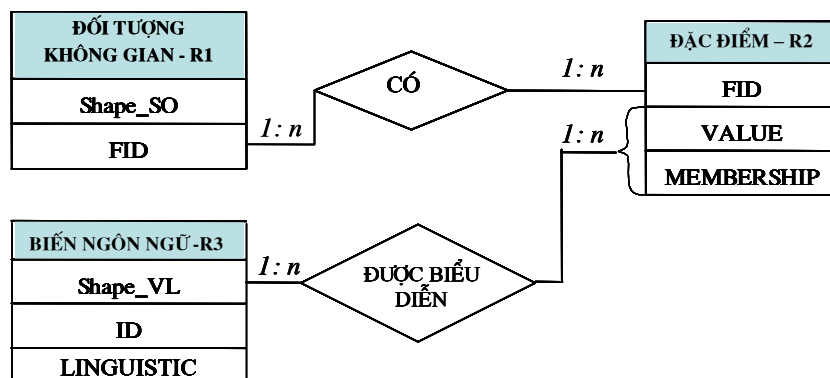
Bảng mô tả các phụ thuộc dữ liệu:

FID ⇒ VALUE, MEMBERSHIP
VALUE, MEMBERSHIP ⇒ ID, LINGUISTIC

Dữ liệu trong mô hình được thiết kế thành ba nhóm: nhóm dữ liệu quản lý các đối tượng không gian để lưu trữ vị trí và mã đối tượng, nhóm dữ liệu lưu trữ các đặc điểm không chắc chắn và nhóm dữ liệu quản lý các biến ngôn ngữ. Trong mô hình này, sử dụng thuộc tính “shape” để lưu trữ dữ liệu không gian cho các đối tượng và biến ngôn ngữ.

Biểu diễn các đặc điểm không chắc chắn của đối tượng vào các hệ thống thông tin địa lý cần quan tâm đến hệ thống cơ sở dữ liệu hiện có và những dị thường khi cập nhật dữ liệu

(đặc biệt là dữ liệu không gian). Mô hình cơ sở dữ liệu mờ để biểu diễn và quản lý dữ liệu không chắc chắn được tổ chức như sau:

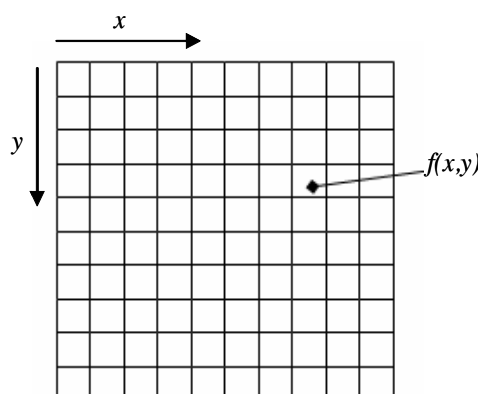


Hình 2. Mô hình cơ sở dữ liệu trong GIS

3.1.1. Mô hình dữ liệu biểu diễn các đối tượng

Với giả thuyết, tọa độ của các đối tượng không gian là chắc chắn nên để biểu diễn đối tượng không gian dựa vào mô hình dữ liệu đang sử dụng trong các hệ thống thông tin địa lý.

Mô hình dữ liệu không gian được sử dụng trong bài viết này là mô hình raster. Trong đó, mỗi giá trị $f(x,y)$ biểu diễn một giá trị của FID. Và như vậy, thông qua mã FID và quá trình tạo kết nối, các thuộc tính mờ có thể được truy xuất và hiển thị dựa vào các điều kiện tìm kiếm theo không gian hoặc thuộc tính.



Hình 3. Dữ liệu không gian

3.1.2. Mô hình dữ liệu biến ngôn ngữ

Mỗi trạng thái của biến ngôn ngữ sẽ có một ánh xạ đến tập mờ tương ứng theo nguyên tắc của biến. Các tập mờ của biến tồn tại trên cùng một không gian nền và giữa chúng có các biên mờ, do đó về mặt không gian các tập mờ có sự giao nhau nhưng kết quả của tập giao được xác định theo cơ sở của lý thuyết tập mờ ([2]).

Các thông tin cần thiết để xác định một hàm thuộc trong thành phần của biến ngôn ngữ bao gồm:

- Hàm thuộc $\mu_{A_i}(x)$: xác định bởi các giá trị rời rạc x_0 trên không gian nền X trong khoảng giới hạn của tập mờ A_i .
- Xác định $[u_i, v_i]$: là khoảng giới hạn tập mờ A_i .
- Xác định giao giữa các tập A_i với A_{i-1} và A_{i+1} .

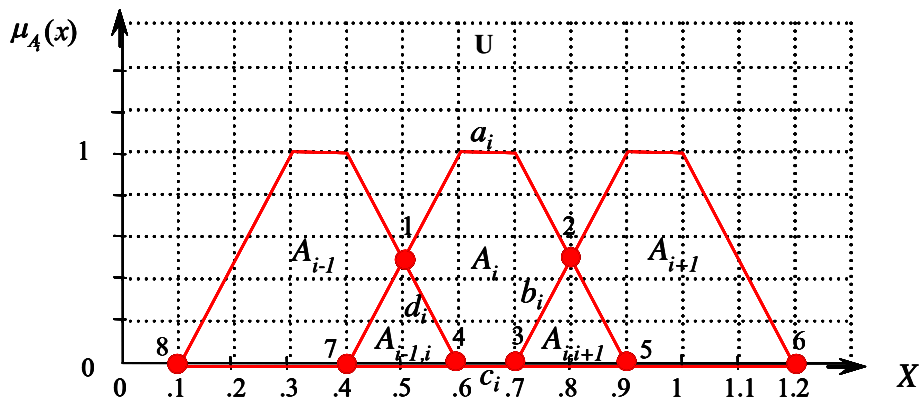
Từ phương pháp tổ chức và lưu trữ dữ liệu không gian ([1]), việc sử dụng mô hình dữ liệu không gian để biểu diễn các thực thể có ưu điểm là dễ thực hiện được các thao tác cập nhật trực quan đồng thời có thể khai thác được các phép phân tích không gian trong GIS. Mô hình dữ liệu biến ngôn ngữ phải lưu trữ được các đặc trưng của hàm thuộc thông qua các tập mờ và giao giữa các tập mờ (biên mờ), do đó lựa chọn mô hình dữ liệu không gian

với cấu trúc vector-topology sử dụng kiểu đối tượng dạng vùng (polygon) để lưu trữ dữ liệu biến ngôn ngữ là phù hợp.

Mô hình dữ liệu không gian của biến ngôn ngữ sử dụng hệ tọa độ phẳng, hai chiều để quản lý dữ liệu, trong đó một chiều là không gian nền X và một chiều kia là độ thuộc $\mu_{A_i}(x)$ của phần tử $x \in X$ vào tập mờ A_i .

Mô hình dữ liệu biến ngôn ngữ được xây dựng theo cấu trúc vector-topology nên việc xác định hình dạng hàm thuộc dựa vào bảng thuộc tính cung. Theo thành phần của bảng thuộc tính cung, mỗi cung được xác định thông qua hai nút và các đỉnh, hai nút xác định điểm bắt đầu và kết thúc của cung còn các đỉnh sẽ xác định hình dạng của cung. Do đó, hình dạng của hàm thuộc được quản lý bằng vị trí các đỉnh có trong bảng thuộc tính cung. Vì vậy, hàm thuộc có thể được biểu diễn ở các hình dạng phức tạp hơn và hình dạng hàm thuộc có thể được hiệu chỉnh thông qua việc thay đổi tọa độ các nút và đỉnh trên cung tương ứng của hàm thuộc.

Dưới đây trình bày mô hình dữ liệu biến ngôn ngữ với các minh họa trên hàm thuộc dạng hình thang (Hình 4):



Hình 4. Mô hình dữ liệu biến ngôn ngữ

Các bảng Cung - Nút - Vùng của dữ liệu biến ngôn ngữ:

Bảng thuộc tính Cung			
Cung	Nút đầu	Đỉnh	Nút cuối
a_i	.5, .5	(.6, 1), (.7, 1)	.8, .5
b_i	.8, .5	-	.7, 0
c_i	.7, 0	-	.6, 0
d_i	.6, 0	-	.5, .5
a_{i+1}	.8, .5	(.9, 1), (1, 1)	1.2, 0

Topology nút	
Nút	Cung
1	$a_{i-1}, a_i, d_i, b_{i-1}$
2	$a_{i+1}, a_i, d_{i+1}, b_i$
3	...
4	...
5	...

Topology vùng	
Vùng	Cung
A_i	a_i, b_i, c_i, d_i
A_{i+1}	$a_{i+1}, b_{i+1}, d_{i+1}$
A_{i-1}	...

Topology cung				
Cung	Nút đầu	Nút cuối	Vùng phải	Vùng trái
a_i	1	2	A_i	U
b_i	2	3	A_i	$A_{i,i+1}$
c_i	3	4	A_i	U

3.2. Xây dựng dữ liệu biến ngôn ngữ

Xây dựng dữ liệu biến ngôn ngữ chủ yếu là xây dựng dữ liệu về các tập mờ tương ứng với các giá trị của biến ngôn ngữ và xác định biên mờ giữa các tập mờ trên cùng một biến.

Khi xây dựng tập mờ, vấn đề cốt lõi là xác định hàm thuộc, tức là xác định độ thuộc $\mu_{A_i}(x)$ cho từng phần tử x trên không gian nền X vào tập mờ A_i . Có nhiều phương pháp xây dựng hàm thuộc như: phương pháp phỏng vấn trực tiếp, phương pháp phỏng vấn gián tiếp, phương pháp nội suy. Mô hình biểu diễn đối tượng không gian được chọn trong bài viết này là phương pháp nội suy từ tập dữ liệu mẫu để xác định hàm thuộc, việc chọn phương pháp nội suy nhằm mục đích khai thác các dữ liệu mẫu hiện có trong hệ thống GIS và quá trình thực hiện mang tính định lượng, có thể sử dụng các chương trình để xử lý.

Cho tập dữ liệu mẫu bao gồm hữu hạn các phần tử $\langle x_i, a_i \rangle$, tiến trình xây dựng dữ liệu biến ngôn ngữ được thực hiện như sau:

Bước 1: Tiến hành tách từng cặp dữ liệu mẫu: $\langle x_1, a_1 \rangle, \langle x_2, a_2 \rangle, \dots, \langle x_n, a_n \rangle$ vào các tập mờ A_i tương ứng.

Bước 2: Xác định $\mu_{A_i}(x)$ bằng phương pháp nội suy Lagrange, tiến trình thực hiện như sau:

$$- \text{Tính } f(x) = a_1 L_1(x) + a_2 L_2(x) + \dots + a_n L_n(x),$$

Với:

$$L_i(x) = \frac{(x - a_1) \dots (x - a_{i-1})(x - a_{i+1}) \dots (x - a_n)}{(x_i - a_1) \dots (x_i - a_{i-1})(x_i - a_{i+1}) \dots (x_i - a_n)}$$

- Khi đó ta có: $\mu_{A_i}(x) = \max[0, \min[1, f(x)]]$

- Thực hiện tính lặp trên từng giá trị $x_0 \in (x_1, x_n)$ cho mỗi tập A_i .

Bước 3: Xác định biên mờ giữa các tập A_i bằng hàm UNION, tiến trình thực hiện theo sơ đồ sau:



Hình 5. Xác định biên mờ trong thành phần biến ngôn ngữ

3.3. Phân tích và hiển thị dữ liệu không chắc chắn

Việc xây dựng các giải thuật phân tích và hiển thị dữ liệu không chắc chắn dựa trên mô hình cơ sở dữ liệu mờ đã mô tả trong Mục 3.1 và các phép chọn, chiếu, kết nối, kết nối không gian. Bên cạnh đó còn khai thác các khả năng của GIS trong việc tạo dữ liệu không gian và hiển thị dữ liệu theo các thuộc tính cho trước.

Toàn bộ xử lý phân tích và hiển thị dữ liệu được phân thành 4 bước theo giải thuật tóm tắt sau:

- + *Bước 1 (S1):* P1 = Project (Join R1 and R2 over FID) over VALUE, MEMBERSHIP
- + *Bước 2 (S2):* M = Make Spatial data (Point; from (P1))
- + *Bước 3 (S3):* P2 = Project (SpatialJoin M and R3 over VALUE, MEMBERSHIP) over LINGUISTIC

+ *Bước 4 (S₄):* D = Display (P2 over LINGUISTIC) with UNIQUE_VALUE

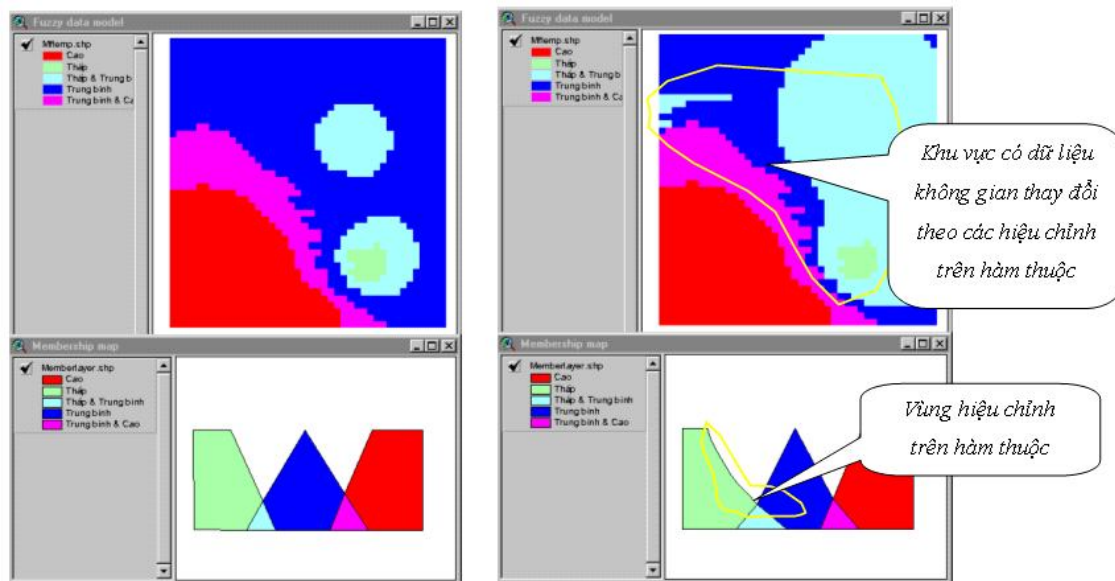
4. MỘT SỐ KẾT QUẢ CÀI ĐẶT

Mô hình biểu diễn đối tượng không gian dựa trên lý thuyết tập mờ và biến ngôn ngữ được cài đặt trên môi trường xử lý của phần mềm ArcView GIS (version 3.2), dữ liệu lưu trữ theo định dạng shape file và các chương trình phân tích hiển thị dữ liệu không chắc chắn được lập trình bằng ngôn ngữ Avenue.

Sau khi cài đặt, chương trình được chạy thử nghiệm trên dữ liệu môi trường về nồng độ ô nhiễm bụi vào tháng 4/1998 tại một khu vực thuộc thành phố Biên Hoà, tỉnh Đồng Nai, rộng 24 km × 37 km.

Kết quả hiển thị dữ liệu ô nhiễm bụi trong cửa sổ “Fuzzy data model” theo các giá trị của biến ngôn ngữ trong cửa sổ “Membership map”. Mỗi trị ngôn ngữ được xác định thông qua tập mờ tương ứng, các tập mờ sau khi xử lý UNION được lưu trữ vào lớp dữ liệu memberlayer.shp, đây chính là dữ liệu biến ngôn ngữ được lưu trữ theo mô hình dữ liệu không gian.

Hình dạng của các hàm thuộc trong lớp dữ liệu memberlayer.shp trên cửa sổ “Membership map” có thể được hiệu chỉnh bằng cách thay đổi tọa độ của các nút hoặc đỉnh thông qua các tương tác “kéo thả”. Khi thực hiện hiệu chỉnh hình dạng của các hàm thuộc thì dữ liệu ô nhiễm bụi biểu diễn theo các giá trị ngôn ngữ cũng thay đổi và kết quả thay đổi được hiển thị dữ liệu trên cửa sổ “Fuzzy data model”.



Hình 6. Dữ liệu biểu diễn dựa trên lý thuyết tập mờ

5. KẾT LUẬN

Những đặc điểm không chắc chắn, có thể được biểu diễn và truy xuất trong các hệ thống thông tin địa lý (GIS) thông qua quá trình xây dựng dữ liệu mờ gồm 3 nhóm (dữ liệu quản lý các đối tượng không gian, dữ liệu lưu trữ các đặc điểm không chắc chắn và dữ liệu quản lý các biến ngôn ngữ) và tiến trình xử lý phân tích hiển thị dữ liệu.

Mô hình được xây dựng theo hướng mở rộng các thuộc tính nên có khả năng áp dụng để biểu diễn nhiều đặc điểm không chắc chắn trên cùng một đối tượng không gian. Đặc điểm không chắc chắn đó có thể là thời gian, vị trí hoặc kích thước, và như vậy có thể đánh giá sự thay đổi thuộc tính theo vị trí hoặc thời gian bằng quan hệ mờ. Ngoài ra, việc xây dựng được mô hình quản lý dữ liệu của biến ngôn ngữ - hàm thuộc theo hướng tiếp cận với mô hình dữ liệu không gian đã tạo ra khả năng dễ tương thích với các hệ thống thông tin địa lý hiện hữu và việc cài đặt và xử lý trên dữ liệu biến ngôn ngữ được thực hiện tương đối đơn giản.

TÀI LIỆU THAM KHẢO

- [1] Trần Vĩnh Phước, *GIS Một số vấn đề chọn lọc*, Nhà xuất bản Giáo dục, 2001.
- [2] George J. Klir and Bo Yuan, *Fuzzy Set and Fuzzy Logic*, Prentice-Hall International Inc., 1995.
- [3] J.D. Ullman, *Nguyên lý các hệ cơ sở dữ liệu và cơ sở tri thức*, Trần Đức Quang biên dịch, Nhà xuất bản Thống kê, 1999.
- [4] Keith C. Clarke, *Getting Started with Geographic Information System*, Pentice Hall, 1999.
- [5] Manfred M. Fischer, Peter Nijkamp, *Geographic Information Systems, Spatial Modelling and Policy Evaluation*, Springer-Verlag, 1993.

Nhận bài ngày 10 - 1 - 2003

Nhận lại sau sửa ngày 20 - 5 - 2003