

**ĐẠI HỌC THÁI NGUYÊN
ĐẠI HỌC CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG**

TRẦN CAO BÁCH

**TỐI ƯU HÓA PHÂN BỐ VÀ ĐỊNH GIÁ ĐẤT ĐAI THEO THUẬT
TOÁN DI TRUYỀN ĐỊNH HƯỚNG KHÔNG GIAN**

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

THÁI NGUYÊN 2017

LỜI CẢM ƠN

Đầu tiên, em xin được gửi lời cảm ơn chân thành sâu sắc nhất tới thầy Lê Hoàng Sơn, người đã đưa em đến với đề tài này, tạo điều kiện thuận lợi cho em nghiên cứu khoa học, nhiệt tình hướng dẫn để em có thể hoàn thành luận văn tốt nghiệp một cách tốt nhất.

Em cũng xin được gửi lời cảm ơn đến các thầy cô trong khoa sau đại học trường đại học công nghệ thông tin và truyền thông thái nguyên đã dạy bảo, giúp đỡ, tạo điều kiện cho em trong thời gian em đã học tập tại trường.

Xin được gửi lời cảm ơn các thầy cô, các anh chị và các bạn trong Trung tâm Tính toán Hiệu năng cao, Trường Đại học Khoa Học Tự Nhiên đã giúp đỡ em trong suốt quá trình học tập và nghiên cứu tại trung tâm.

Em xin được gửi lời cảm ơn chân thành tới công ty Tư vấn Geo Việt đã cung cấp bộ số liệu, các tài liệu liên quan cũng như nhiệt tình giúp đỡ em trong quá trình thực hiện luận văn.

Xin được gửi lời cảm ơn tới gia đình đã động viên, tạo điều kiện cho em thực hiện luận văn, lời cảm ơn tới bạn bè đã giúp đỡ em học tập trong những năm vừa qua.

Học viên

Trần Cao Bách

LỜI CAM ĐOAN

Tôi xin cam đoan luận văn “Tối ưu hóa phân bố và định giá đất đai theo thuật toán di truyền định hướng không gian” của tôi được thực hiện dưới sự hướng dẫn khoa học của TS. Lê Hoàng Sơn, số liệu và kết quả nghiên cứu trong luận văn này hoàn toàn trung thực và chưa sử dụng để bảo vệ một công trình khoa học nào, các thông tin, tài liệu trích dẫn trong luận văn đã được chỉ rõ nguồn gốc ở phía cuối luận văn.

Mọi sự giúp đỡ cho việc hoàn thành luận văn đều đã được cảm ơn. Nếu có phát hiện nào về sự gian lận trong sao chép tài liệu, công trình nghiên cứu của tác giả khác mà không được ghi rõ trong tài liệu tham khảo, tôi hoàn toàn chịu trách nhiệm.

Thái Nguyên, tháng 6 năm 2017

Học viên

Trần Cao Bách

MỤC LỤC

LỜI CẢM ƠN	i
LỜI CAM ĐOAN	iii
MỤC LỤC.....	iv
DANH MỤC CÁC HÌNH VẼ.....	vi
DANH MỤC CÁC THUẬT NGỮ	vii
MỞ ĐẦU.....	1
CHƯƠNG 1: BÀI TOÁN PHÂN BỐ VÀ ĐỊNH GIÁ ĐẤT.....	3
1.1. Tổng quan về bài toán phân bố và định giá đất	3
1.1.1. Một số khái niệm.....	3
1.1.2. Các yếu tố ảnh hưởng trực tiếp đến giá đất	4
1.1.3. Đánh giá về các yếu tố ảnh hưởng tới giá đất.....	7
1.1.4. Nguyên tắc định giá đất	8
1.2. Các ứng dụng thực tiễn	12
1.3. Các nghiên cứu liên quan.....	13
1.4. Tổng kết chương	14
CHƯƠNG 2: THIẾT KẾ THUẬT TOÁN DI TRUYỀN ĐỊNH HƯỚNG KHÔNG GIAN CHO BÀI TOÁN PHÂN BỐ VÀ ĐỊNH GIÁ ĐẤT.....	15
2.1. Tổng quan về thuật toán tối ưu tiến hóa	15
2.1.1. Cấu trúc dữ liệu và giải thuật tiến hóa	15
2.1.2. Công thức của giải thuật di truyền.....	21
2.1.3. Các thành phần của thuật toán	22
2.2. Thiết kế thuật toán di truyền không gian cho bài toán định giá đất	23
2.2.1. Thuật toán di truyền	23
2.2.2. Thuật toán di truyền định hướng không gian (SGA)	26
2.3. Phân loại vùng không gian trên bản đồ theo giá đất hỗ trợ phân bố đất đai bằng thuật toán FCM	28
2.4. Tổng kết chương	30

CHƯƠNG 3: XÂY DỰNG HỆ THỐNG THÔNG TIN CHO BÀI TOÁN PHÂN BỐ VÀ ĐỊNH GIÁ ĐẤT	31
3.1. Giới thiệu về hệ thống LIS.....	31
3.1.1. Khái niệm.....	31
3.1.2. Kiến thức chung	32
3.1.3 Chức năng và vai trò của một hệ thông tin đất đai	34
3.2. Thiết kế hệ thống WebGIS cho bài toán phân bố và định giá đất	39
3.2.1. Giao diện chương trình	44
3.2.2. Thử nghiệm chương trình và đánh giá kết quả	46
3.3. So sánh ưu nhược điểm của các thuật toán.....	48
3.4. Tổng kết chương	49
KẾT LUẬN	50
TÀI LIỆU THAM KHẢO.....	51

DANH MỤC CÁC HÌNH VẼ

Hình 2.1: Sơ đồ thuật toán di truyền [2].	20
Hình 2.2: Mô tả bộ trọng số W	26
Hình 2.3: Mô tả tìm hai hàng xóm	27
Hình 3.1: Các thành phần của LIS[2].	35
Hình 3.2: Sơ đồ vận hành của LIS	38
Hình 3.3: Các bước của quá trình xây dựng ứng dụng	39
Hình 3.4: Sơ đồ hệ thống ứng dụng	40
Hình 3.5: Biểu đồ tuần tự của ứng dụng	41
Hình 3.6: Một phần của cơ sở dữ liệu thuộc tính.....	42
Hình 3.7: Bảng cơ sở dữ liệu không gian về một số quận	43
Hình 3.8: Một phần cơ sở dữ liệu về các địa điểm khảo sát giá đất	43
Hình 3.9: Giao diện ban đầu của hệ thống.....	44
Hình 3.10: Giao diện bật tắt Layer.....	44
Hình 3.11: Giao diện khi tắt các Layers.....	45
Hình 3.12: Giao diện khi click vào điểm trên bản đồ	45
Hình 3.13: Giao diện khi nhập giá trị vào bộ lọc	46
Hình 3.14: Bảng fitness.....	47
Hình 3.15: Giá hai hàng xóm	47

DANH MỤC CÁC THUẬT NGỮ

Thuật ngữ	Viết tắt	Giải thích
Spatial Genetic Algorithm	SGA	Thuật toán di truyền không gian
Genetic Algorithm	GA	Thuật toán di truyền
Real Estate	BĐS	Bất động sản
Travelling Salesman Problem	TSP	Bài toán người du lịch
Ant Colony Optimization	ACO	Thuật toán tối ưu đàn kiến
Particle Swarm Optimization	PSO	Thuật toán tối ưu bầy đàn
Chromosome	NST	Nhiễm sắc thể
Geographic Information Systems	GIS	Hệ thống thông tin địa lý
Land Information System	LIS	Hệ thống thông tin đất đai
Data Exchange Center	DEC	Trung tâm trao đổi dữ liệu
Database Managemant System	DBMS	Hệ quản trị cơ sở dữ liệu
Analyst		Nhà phân tích
System Administrator		Người quản trị hệ thống
Database Administrator		Người quản trị cơ sở dữ liệu
Data Server		Máy chủ dữ liệu
Mean Square Error	MSE	Bình phương sai số trung bình
Metadata		Siêu dữ liệu
Web Server		Máy chủ web
HyperText Transfer Protocol	HTTP	Giao thức truyền tải siêu văn bản

MỞ ĐẦU

Bài toán phân bổ và định giá đất đai là vấn đề được quan tâm nhiều bởi ý nghĩa ứng dụng của nó trong thực tiễn. Bài toán này được hiểu là sự tái cơ cấu lại bản đồ chuyên đề về đất đai của một vùng địa lý cho trước theo một số tiêu chí kinh tế - xã hội nhất định, ví dụ: giá đất, tác động môi trường. Đây là một bài toán tối ưu đa mục tiêu với các dữ liệu gồm cả phần không gian và dữ liệu chuyên đề. Chính vì thế việc tìm ra phương án tối ưu và thể hiện kết quả của nó lên bản đồ chuyên đề thành các cụm hay các vùng không gian có mức độ khác nhau theo chuyên đề đó có ý nghĩa trong việc hoạch định và phát triển chính sách xã hội. Tầm quan trọng và ứng dụng của bài toán này có thể xem trong tài liệu [9].

Trong các thuật toán tối ưu đa mục tiêu, thuật toán di truyền [10] là một công cụ tốt cho phép mô hình hóa và tìm kiếm nghiệm trong không gian nghiệm vô cùng lớn. Thuật toán di truyền là một nhánh của thuật toán tối ưu tiến hóa sử dụng các chiến lược tìm nghiệm phỏng theo hành vi sinh học của bầy đàn, quần thể, v.v. Thuật toán này đã được ứng dụng rộng rãi trong nhiều bài toán của khai phá dữ liệu, nhận dạng mẫu và là một công cụ tốt cho việc giải các bài toán tối ưu với ràng buộc lớn. Trong [9] các tác giả đã đề xuất ra thuật toán di truyền định hướng không gian bằng cách tích hợp thông tin không gian của đối tượng trên bản đồ vào quá trình tìm kiếm nghiệm. Chính vì vậy, nghiệm tìm được có độ chính xác và chất lượng cao hơn thuật toán di truyền gốc do thông tin sử dụng bao gồm cả thông tin bản đồ và chuyên đề. Do vậy, đây là đối tượng nghiên cứu chính của chúng tôi trong luận văn này.

Mục tiêu của luận văn là tìm hiểu thuật toán di truyền định hướng không gian và xây dựng ứng dụng phân bổ và định giá đất đai trên hệ thống WebGIS. Kết quả thu được của luận văn là tài liệu chi tiết về tiếp cận tối ưu

tiến hóa, cụ thể là thuật toán di truyền định hướng không gian cho bài toán phân bổ và định giá đất đai, và một phần mềm ứng dụng thuật toán trên cho dữ liệu thực tế về đất đai.

Nội dung báo cáo gồm lời nói đầu và 3 chương chính:

Chương 1: Bài toán phân bổ và định giá đất

Chương này đưa ra những kiến thức tổng qua về bài toán giá đất, các ứng dụng thực tiễn, các nghiên cứu liên quan và mô tả dữ liệu giá đất ở Hà Nội.

Chương 2: Thiết kế thuật toán di truyền và định hướng không gian cho bài toán phân bổ định giá đất.

Nội dung chương 2 bao gồm các nội dung chính sau: Tìm hiểu chung về thuật toán tối ưu tiến hóa, tìm hiểu cụ thể về thuật toán di truyền và thiết kế thuật toán di truyền không gian cho bài toán phân bổ và định giá đất.

Chương 3: Xây dựng hệ thống thông tin đất đai cho bài toán phân bổ và định giá đất

Chương này giới thiệu về công nghệ WebGIS, đưa ra chức năng và vai trò của một hệ thống tin đất đai, thiết kế hệ thống WebGIS cho bài toán định giá đất, xây dựng cơ sở dữ liệu, và hiển thị ra giao diện chương trình và kết quả đánh giá.

CHƯƠNG 1: BÀI TOÁN PHÂN BỐ VÀ ĐỊNH GIÁ ĐẤT

1.1. Tổng quan về bài toán phân bổ và định giá đất

1.1.1. Một số khái niệm

Đất đai là một loại tài nguyên tự nhiên, một loại tài nguyên sản xuất, bao gồm cả lục địa và mặt nước trên bề mặt trái đất. Theo quan điểm kinh tế học thì đất đai không chỉ bao gồm mặt đất còn bao gồm cả tài nguyên trong lòng đất và tất cả mọi thứ sinh sôi trên mặt đất và trong lòng đất không do lao động và con người làm ra, tức là bao gồm nước mặt đất và nước ngầm, thổ nhưỡng, thực vật và động vật. Với nghĩa chung nhất, đó là lớp bề mặt của trái đất, bao gồm đồng ruộng, đồng cỏ, bãi chăn thả, cây rừng, bãi hoang, mặt nước, đầm lầy và bãi đá v.v. Với nghĩa hẹp thì đất đai biểu hiện khối lượng và tính chất của quyền lợi hoặc tài sản mà một người có thể chiếm đối với đất. Nó có thể bao gồm lợi ích trên đất về mặt pháp lý cũng như những quyền theo tập quán không thành văn [2].

Đất đai là tài sản và đồng thời đất đai còn được coi là một tài sản đặc biệt vì bản thân nó không do lao động làm ra, mà lao động tác động vào đất đai để biến nó từ trạng thái hoang hoá trở thành sử dụng vào đa mục đích giá đất là giá trị của quyền sử dụng đất tính trên một đơn vị diện tích đất.

Định giá đất là loại hoạt động chuyên môn vừa mang tính kinh tế-kỹ thuật, tính pháp lý, vừa mang tính xã hội, tính nghệ thuật. Hoạt động định giá hình thành, tồn tại và phát triển gắn liền với sự hình thành, tồn tại và phát triển của thị trường. Định giá tài sản là hình thức phân tích kinh tế ứng dụng. Nhiều khái niệm sử dụng trong lý thuyết định giá là những khái niệm kinh tế tuy có một số khác biệt. Ngày nay, theo đà phát triển của kinh tế thị trường, lý thuyết định giá ngày càng được hoàn thiện và trở thành một ngành chuyên sâu trong nền kinh tế [2].

Định giá đất đai là người định giá căn cứ vào nguyên tắc, phương pháp định giá đất trên cơ sở nắm chắc tư liệu thị trường đất đai, căn cứ vào những thuộc tính kinh tế và tự nhiên của đất đai theo chất lượng, và tình trạng thu lợi thông thường trong hoạt động kinh tế thực tế của đất đai, xem xét đầy đủ ảnh hưởng của các yếu tố về phát triển kinh tế, xã hội, phương thức sử dụng đất, dự kiến thu lợi từ đất và chính sách đất đai, đối với việc thu lợi từ đất, rồi tổng hợp để định ra giá cả tại một thời điểm nào đó cho một thửa hoặc nhiều thửa đất với một quyền đất đai nào [2].

Lao động chuyển hóa đất đai tự nhiên thành đất đai kinh tế nên đất đai có giá trị, theo Mác giá trị này là cơ sở của giá cả thị trường đất đai (giá đất). Giá trị là bản chất bên trong, giá cả là biểu hiện bằng tiền ra bên ngoài. Giá cả thị trường đất đai chịu sự ảnh hưởng của các nhân tố như: giá trị của đất đai, sức mua tiền tệ, tỷ suất lợi tức, mức độ lạm phát, cạnh tranh, tâm lý khách hàng, mức độ khan hiếm, cung cầu, luật pháp, các chính sách đất đai, các hình thức địa tô, v.v.

Phân bổ đất đai là hoạt động dựa trên kết quả thu được sau khi dự báo định giá đất sẽ phân chia thành các vùng thấp, trung bình và cao hay chia theo nhu cầu tìm kiếm của người sử dụng để người sử dụng có thể dễ dàng tìm thấy vùng giá đất phù hợp nhu cầu tìm kiếm.

1.1.2. Các yếu tố ảnh hưởng trực tiếp đến giá đất

Đất đai là tư liệu sản xuất chủ yếu của kinh tế nông nghiệp, là địa bàn sản xuất của công nghiệp và dịch vụ, đất đai là nguồn gốc của cải của xã hội. Đất đai kinh tế còn là một loại hàng hóa đặc biệt, thị trường bất động sản với các hoạt động mua bán, cho thuê, góp vốn, thế chấp, v.v. luôn diễn ra sôi động khiến đất đai có vị trí quan trọng trong guồng quay kinh tế. Vì vậy các đặc tính tự nhiên vốn có, quy luật kinh tế, nhân tố xã hội ảnh hưởng tới khả năng sinh lợi của thửa đất, quy định trực tiếp tới giá đất.

- Yếu tố tự nhiên

+ Vị trí: Yếu tố trực tiếp ảnh hưởng tới giá đất. Những thửa đất có vị trí càng thuận lợi cho công việc thương mại, kinh doanh, cho thuê v.v. càng có giá trị. Những thửa đất trong đô thị thường có giá cao hơn thửa đất ở nông thôn. Trong cùng một đô thị, giá những thửa đất có vị trí gần các đầu mối giao thông quan trọng, gần trung tâm hành chính, văn hóa, y tế, du lịch, v.v. cao hơn những thửa đất khác.

+ Hình thể, diện tích của thửa đất: Thửa đất được coi là tối ưu khi hình thể, diện tích, kích thước của nó đáp ứng được nhu cầu, hợp túi tiền của người sử dụng. Với nhu cầu để ở, ngôi nhà với thửa đất vuông vắn, diện tích trung bình phù hợp với đa số nhu cầu của người dân đô thị có giá cao hơn những ngôi nhà khác. Nhưng với người dân nông thôn, đất không chỉ để ở mà còn để tăng gia sản xuất, những thửa đất có diện tích rộng được ưa chuộng hơn. Với mục đích kinh doanh thương mại, người sử dụng yêu cầu những thửa đất có mặt tiền rộng. Thửa đất với hình thể và diện tích riêng, trong từng mục đích sử dụng khác nhau sẽ có giá khác nhau [9].

+ Môi trường tự nhiên: Vùng đất có khí hậu thuận hòa được ưa chuộng hơn vùng có khí hậu khắc nghiệt. Các vùng kinh tế sản xuất đặc thù đòi hỏi môi trường khí hậu phù hợp. Tình trạng môi trường (ô nhiễm nước thải, không khí v.v.).

- Yếu tố kinh tế - xã hội

a. An ninh xã hội

+ Nhân tố này ảnh hưởng đến giá đất rõ và dễ thấy. Bất luận là khu nhà ở hay khu thương nghiệp, trật tự xã hội tốt, thì mọi người sẽ cảm thấy an toàn, vì vậy họ sẵn sàng đầu tư, mua bán, cư trú, do đó kéo theo giá đất tăng lên [2].

b. Đầu cơ nhà đất

+ Đầu cơ nhà đất là hành vi người đầu cơ hy vọng và lợi dụng sự biến động của giá cả nhà đất để thu được siêu lợi nhuận. Nhân tố này có ảnh hưởng đột xuất đến mức giá đất, đặc biệt là mức giá đất thị trường. Khi đất cung không đủ cầu, do người đầu cơ tranh mua mà đẩy giá lên cao, khi đất cung nhiều hơn cầu, do người đầu cơ bán tháo làm cho đất rớt giá [2].

c. Cơ sở kỹ thuật hạ tầng

+ Cơ sở hạ tầng kỹ thuật: Mức độ hoàn thiện và hiện đại của hệ thống cầu đường, hệ thống điện, nước, cống thoát nước, hệ thống viễn thông, Internet và các tiện nghi khác quyết định tới lợi thế và giá trị của thửa đất. Sự thuận lợi của thửa đất đối với các nhu cầu của con người về y tế, giáo dục, giải trí cộng đồng, thửa đất càng thuận lợi thì có giá càng cao. Khoảng cách từ thửa đất tới bệnh viện, công viên, trường học, khu vui chơi, siêu thị, chợ, cơ quan hành chính, tín ngưỡng, v.v. và mức độ hiện đại của cơ sở hạ tầng xã hội đều ảnh hưởng tới giá cả đất đai.

d. Mật độ nhân khẩu

+ Mật độ nhân khẩu tăng cao, thì nhu cầu đối với đất tăng vì thế giá đất tăng lên. Ví dụ như ở Nhật Bản tỷ lệ tăng giá đất ở thành thị năm 195 ~ 1960 là 11 ~ 13%, năm 1960 ~ 1961 là 17 ~ 18% còn ở Mỹ năm 1956 năm 1966, tỷ lệ biến động giá cả đất chỉ là 5,5 ~ 6,9%. Nguyên nhân chủ yếu là ở Nhật Bản là nước có mật độ nhân khẩu thành thị cao nhất, là quốc gia có tỷ lệ tăng nhân khẩu cao nhất trong những nước có nền kinh tế phát triển, tốc độ phát triển kinh tế thành thị của Mỹ cũng rất nhanh, nhưng do mật độ nhân khẩu thấp, tỷ lệ tăng nhân khẩu so với Nhật là thấp, yêu cầu đất không được căng thẳng như Nhật Bản, do đó mức tăng giá đất tương đối nhỏ [2].

e. Tiến trình đô thị hóa

+ Một trong những kết quả của phát triển công nghiệp hóa là thúc đẩy quá trình đô thị hóa. Những khu vực có trình độ công nghiệp hóa hoặc đô thị hóa càng cao, mà tổng diện tích đất của thành phố nói chung là không đổi, cho nên ảnh hưởng của tiến trình đô thị hóa đối với giá đất biểu hiện ở hai mặt sau đây: một là ở khu vực thành phố có mật độ nhân khẩu đông đúc thì giá đất tăng cao; hai là tỷ lệ nhân khẩu nhập cư cao và tiến trình đô thị hóa có tốc độ nhanh, thì mức độ và tốc độ tăng giá đất đều cao hơn đất khu vực khác[2].

f. Yếu tố quốc tế

+ Yếu tố quốc tế: Sự phát triển kinh tế của một quốc gia ngoài việc không thể cô lập và tách rời với sự phát triển của kinh tế thế giới ra, nói chung hoặc ít hoặc nhiều đều có quan hệ với sự phát triển kinh tế của quốc gia khác. Thị trường thế giới càng mở rộng thì mối quan hệ này càng trở nên mật thiết. Cho nên, sự thịnh suy kinh tế của mỗi quốc gia đều liên quan đến sự phát triển hay suy thoái của kinh tế thế giới, tình trạng thị trường, đất đai cũng không có ngoại lệ. Ví dụ, khi bùng phát nguy cơ kinh tế có tính toàn cầu năm 1929 giá đất ở nước Mỹ hạ thấp nhanh chóng, từ đỉnh cao năm 1925 rớt thẳng xuống theo đường thẳng đứng, đến năm 1933 thì rớt xuống đến điểm đáy, cho đến sau khi kinh tế thế giới phục hồi thì giá đất mới tăng lên[2].

1.1.3. Đánh giá về các yếu tố ảnh hưởng tới giá đất

Giá đất chịu ảnh hưởng từ nhiều yếu tố khác nhau. Có các nhân tố tác động trực tiếp tới giá đất như yếu tố tự nhiên, yếu tố kinh tế, tâm lý người sử dụng v.v. có các nhân tố tác động gián tiếp tới giá đất như yếu tố chính sách kinh tế vĩ mô, yếu tố khí hậu, yếu tố an ninh xã hội, yếu tố quốc tế v.v.

Trong mỗi nhóm yếu tố tác động trực tiếp hay gián tiếp, luôn tồn tại những yếu tố tác động nhiều, yếu tố tác động ít. Đối với yếu tố tự nhiên, vị trí

của ngôi nhà quyết định lớn nhất tới giá đất, sau đó là yếu tố hình thể, diện tích. Nhưng đánh giá mức độ ảnh hưởng của các yếu tố ảnh hưởng tới giá đất chỉ mang tính tương đối, mức độ ảnh hưởng đó còn phụ thuộc vào hoàn cảnh, ngôi nhà và người sử dụng cụ thể. Mỗi đối tượng sử dụng nhà có nhu cầu khác nhau. Doanh nhân mong muốn mua được ngôi nhà có vị trí kinh doanh tốt; người sử dụng cao tuổi ưa chuộng những ngôi nhà có vị trí gần bệnh viện, gần công viên, có môi trường yên tĩnh; gia đình có con nhỏ ưa chuộng ngôi nhà gần trường học, trụ sở cơ quan v.v. Ngoài ra, còn có các trường hợp cá biệt khiến giá đất tăng hoặc giảm bất thường. Giá đất chịu ảnh hưởng của rất nhiều các yếu tố với mức độ khác nhau nhưng đều quyết định tới khả năng sinh lợi. Lợi ích này đối với mỗi đối tượng sử dụng là khác nhau. Sự đánh giá mức độ ảnh hưởng của các yếu tố này chỉ mang tính tương đối.

1.1.4. Nguyên tắc định giá đất

Giá đất chịu sự ảnh hưởng của nhiều yếu tố, các yếu tố này luôn biến động. Để xác định được giá đất cần nghiên cứu tỉ mỉ các yếu tố đó và tương tác giữa chúng, phán đoán chính xác xu thế biến động, sử dụng các nguyên tắc định giá như kim chỉ nam cho việc nghiên cứu. Các nguyên tắc trong định giá là những nguyên tắc kinh tế vận hành theo cơ chế thị trường, không được phân tích riêng lẻ mà các nguyên tắc này được xem xét trong mối quan hệ chặt chẽ với nhau.

- Nguyên tắc sử dụng hiệu quả nhất

Một thửa đất có khả năng sử dụng vào nhiều mục đích khác nhau, trong số đó có sử dụng là hợp pháp, có sử dụng bị coi là không hợp pháp, sử dụng trái với các qui định của chính quyền địa phương. Xét về mặt giá trị, có sử dụng cho thu nhập ròng thấp và kéo dài, có sử dụng cho thu nhập ròng cao nhưng thời gian diễn ra rất ngắn. Trong số rất nhiều các sử dụng khác nhau, thông thường một thửa đất bao giờ cũng có một sử dụng vừa hợp pháp vừa

đem lại giá trị thu nhập ròng lớn nhất và trong khoản thời gian xác định. Loại sử dụng đó, lý thuyết định giá gọi là sử dụng hiệu quả nhất [2].

Một thửa đất được đánh giá là sử dụng cao nhất và tốt nhất nếu tại thời điểm định giá cho thấy thửa đất đó đang được sử dụng hợp pháp cũng như đang cho thu nhập ròng lớn nhất hoặc có khả năng cho giá trị hiện tại của thu nhập ròng trong tương lai là lớn nhất; sử dụng hiệu quả nhất đó có thể tồn tại và kéo dài liên tục trong một khoảng thời gian nhất định [2].

Sử dụng hiệu quả nhất được xem là một trong số các nguyên tắc quan trọng nhất của định giá đất.

- Nguyên tắc thay thế

Trong trường hợp hai hay nhiều bất động sản có thể thay thế lẫn nhau trong quá trình sử dụng, thì giá trị của những bất động sản đó được xác định bởi sự tác động lẫn nhau của bất động sản này đến bất động sản khác. Khi hai tài sản có tính hữu ích như nhau, tài sản nào chào bán ở mức giá thấp nhất thì tài sản đó sẽ bán được trước. Ví dụ, có 2 bất động sản chia lô tại khu đô thị Trung Hòa - Nhân Chính, một biệt thự 4 tầng, mặt tiền 5,5 m, mặt bằng 60m², có garage, ô tô vào nhà, được rao bán với giá 3,5 tỷ đồng và một biệt thự 4 tầng, có garage, có sân công, mặt tiền 5m, mặt bằng 70m² rao bán với giá 3.3 tỷ đồng.

Xét về mặt hữu dụng là như nhau, vị trí tương đương nhau, bất động sản thứ hai rẻ hơn tương đối so với bất động sản thứ nhất, do vậy, người mua chắc chắn sẽ lựa chọn mảnh đất thứ 2.

- Nguyên tắc thay đổi

Giá cả thị trường của bất động sản luôn biến động vì các nhân tố ảnh hưởng tới giá đất như nhân tố tự nhiên, kinh tế, pháp luật, xã hội, v.v. cũng thay đổi theo thời gian. Sự thay đổi là một quy luật khách quan, luôn vận động không phụ thuộc vào ý muốn con người. Phải nắm được quy luật biến

động này để dự đoán giá chính xác. Việc ước tính giá cả đất đai tuy chỉ có giá trị vào thời điểm định giá nhưng vẫn phải phản ánh được xu thế tương lai của bất động sản.

- Nguyên tắc cung - cầu

Giá trị của một bất động sản được xác định bởi mối quan hệ cung - cầu về bất động sản đó trên thị trường. Ngược lại, giá trị của bất động sản cũng tác động đến cung và cầu. Giá trị của bất động sản thay đổi tỷ lệ thuận với cầu và tỷ lệ nghịch với cung.

Đặc điểm của thị trường bất động sản là rất nhạy bén với cầu, nhưng lại chậm chạp trong khâu cung, do chu kỳ đầu tư xây dựng bất động sản khá dài. Hiện nhu cầu thuê văn phòng rất cao, nhưng cung không đáp ứng kịp nên giá cả đã bị đẩy lên.- Phương pháp định giá đất.

Để ước tính được giá đất đạt độ chính xác cao nhất không chỉ sử dụng các nguyên tắc định giá như kim chỉ nam mà còn vận dụng chặt chẽ các phương pháp định giá.

- Phương pháp so sánh trực tiếp

+ Khái niệm: Phương pháp so sánh trực tiếp là phương pháp định giá đất thông qua việc phân tích mức giá của các ngôi nhà tương tự về mục đích sử dụng, vị trí, khả năng sinh lợi, điều kiện kết cấu hạ tầng, diện tích, hình thể, tính pháp lý về quyền sử dụng đã giao dịch trên thị trường để xác định giá của ngôi nhà cần định giá.

+ Cơ sở khoa học: Phương pháp so sánh trực tiếp dựa trên cơ sở nguyên tắc thay thế. Các thửa đất tương tự nhau có cùng điều kiện thị trường sẽ có cùng một giá cả, chúng có thể thay thế cho nhau, do cạnh tranh mà giá cả không chế lẫn nhau và đi tới thống nhất.

+ Quy trình thực hiện

Thu thập thông tin: Thu thập thông tin thửa đất cần định giá và thông tin tối thiểu của 3 thửa đất đã giao dịch.

Phân tích, so sánh thông tin: Lựa chọn các mẫu giao dịch để so sánh, so sánh thửa đất cần định giá với các mẫu giao dịch để tìm ra các đặc điểm khác biệt và giá trị các đặc điểm đó.

Hiệu chỉnh giá trị các đặc điểm khác nhau theo các nguyên tắc: Hiệu chỉnh trên giá các giao dịch đã hoàn tất, theo giá trị của các đặc điểm khác nhau và theo đặc điểm thị trường trước.

Xác định giá cả thửa cần định giá: lựa chọn 1 trong số những kết quả thu được sau hiệu chỉnh.

+ Đánh giá phương pháp:

Ưu điểm: Là phương pháp dễ thực hiện về mặt kỹ thuật, kết quả định giá dựa trên chứng cứ thị trường thực tế nên trực quan, mang tính thuyết phục cao.

Nhược điểm: phương pháp chỉ thực hiện được khi có đủ nhiều lượng các thửa đất tương tự đã giao dịch trong khu vực. Khi chất lượng thông tin về các bất động sản kém sẽ ảnh hưởng xấu tới kết quả định giá. Sự chính xác của phương pháp còn phụ thuộc vào kinh nghiệm của người định giá.

- Phương pháp thu nhập

+ Khái niệm: là phương pháp định giá đất tính bằng thương số giữa mức thu nhập ròng thu được bình quân một năm trên một đơn vị diện tích đất so với lãi suất tiền gửi tiết kiệm bình quân một năm tính đến thời điểm định giá đất của loại tiền gửi VNĐ kỳ hạn 12 tháng tại ngân hàng thương mại nhà nước có mức lãi suất tiền gửi tiết kiệm cao nhất trên địa bàn cấp tỉnh.

+ Cơ sở khoa học: dựa trên những lý luận về địa tô, giá đất, thu nhập do đất mang lại và lý luận về sự đóng góp của các yếu tố trong sản xuất hàng hóa

+ Nhược điểm: yêu cầu các tham số để tính toán có độ chính xác cao,

nhưng các tham số được xác định chỉ là dự kiến, ước tính, khiến hạn chế độ chính xác. Có thể khắc phục bằng cách sử dụng các phương pháp khác để kiểm tra.

1.2. Các ứng dụng thực tiễn

- Ứng dụng phương pháp so sánh trong việc định giá đất ở Việt Nam:

Hiện nay ở Việt Nam phương pháp so sánh được sử dụng phổ biến cả trong việc thẩm định giá tài sản đơn lẻ hay xác định giá đất hàng loạt để phục vụ cho mục đích ban hành khung giá đất ở các địa phương nhằm đảm bảo nguyên tắc sát với giá thị trường. Ngày 10 và ngày 11/8/2006, Bộ tài chính đã tổ chức hội thảo khoa học mang tính chất ứng dụng tại Cần Thơ với chủ đề: “Các phương pháp xác định giá đất ở Việt Nam”. Trên cơ sở những bài tham luận chính trình bày tại Hội thảo, các nhà quản lý và các doanh nghiệp am hiểu về lĩnh vực xác định giá đất đã rút ra được những nhận xét về thực trạng áp dụng phương pháp này tại một số tỉnh thành phố:

Tại Hà Nội, một đô thị đặc biệt với mật độ dân số đông, diện tích rộng, mức độ biến động về tình hình sử dụng đất hàng năm là rất lớn do quá trình đô thị hóa nhanh. Thực tế tại Hà Nội, chỉ có đất ở mới điều tra được giá chuyển nhượng trên thị trường, còn các loại đất khác rất ít hoặc không có. Thực hiện phương pháp so sánh gặp rất nhiều khó khăn, tốn nhiều thời gian mới xác định được giá cho một thửa đất.

Tại thành phố Hồ Chí Minh, phương pháp so sánh được áp dụng rộng rãi nhưng vẫn còn những hạn chế. Trong phương pháp này, bước điều chỉnh các yếu tố so sánh là quan trọng nhất, ảnh hưởng tới kết quả thẩm định giá nhưng việc điều chỉnh này chỉ mang tính chất đúc kết kinh nghiệm, thiếu căn cứ, chưa có chuẩn mực đánh giá nên dẫn đến việc không thống nhất về mặt lý luận về nhận định mức độ ảnh hưởng của từng yếu tố tác động.

1.3. Các nghiên cứu liên quan

Trong nền kinh tế thị trường, người ta thường quan niệm thửa đất và ngôi nhà được xây dựng trên khuôn viên thửa đất đó là một tài sản gọi là bất động sản. Pháp luật nhiều nước như Anh, Pháp, Malaysia, Singapore coi bất động sản là một chế định pháp luật và có cả bộ luật riêng để điều chỉnh, gọi là bộ luật bất động sản.

Hiện nay, trong lĩnh vực định giá, định giá bất động sản nói chung và định giá đất nói riêng được các nhà tư bản phát triển và nhiều nước đang phát triển trên thế giới quan tâm và đầu tư, nghiên cứu từ nhiều thập kỉ qua cho đến nay nhiều nước đã ổn định về hệ thống tổ chức cơ quan định giá bất động sản từ trung ương đến địa phương với chức năng, nhiệm vụ và cơ chế rõ ràng, hiệu quả. Với đội ngũ cán bộ khoa học và cán bộ chuyên môn gồm hàng ngàn người được đào tạo chính quy trong các trường đại học chuyên ngành đang làm việc trong các cơ quan định giá của nhà nước và tư nhân.

- Ở Việt Nam:

Hệ thống cơ quan định giá của nhà nước về bất động sản được thống nhất từ trung ương đến địa phương được tổ chức tại 3 cấp:

Phía Chính phủ có liên ngành các cơ quan làm nhiệm vụ định giá như bộ tài chính, bộ xây dựng, bộ tài nguyên và môi trường.

Cấp tỉnh: Sở tài nguyên và môi trường chịu trách nhiệm trước UBND cấp tỉnh về công tác định giá đất.

Cấp huyện: Phòng tài nguyên và môi trường chịu trách nhiệm trước UBND cấp huyện về công tác định giá đất.

Các cơ quan này sẽ trao trách nhiệm cho các thành viên định giá đất. Các thành viên định giá đất là những người được cơ quan có thẩm quyền giao nhiệm vụ cho từng vụ việc cụ thể, thông qua ủy ban và hội đồng có trách nhiệm định giá đất đại phục vụ mục đích cho từng cuộc định giá cụ thể.

- Ở Malaysia

+ Nhà nước Malaysia là nhà nước Liên bang, hệ thống cơ quan định giá của nhà nước về bất động sản được tổ chức tại cả 3 cấp: liên bang, bang, quận (huyện) [2].

+ Tại trung ương(liên bang): Cơ quan định giá có tên gọi là: “Cục dịch vụ định giá bất động sản” gọi tắt theo tiếng Malaysia là JPPH. JPPH trực thuộc Bộ tài chính, Giám đốc JPPH do Bộ trưởng Bộ tài chính bổ nhiệm và chịu trách nhiệm trước Bộ trưởng về toàn bộ các hoạt động của cơ quan mình.

+ Tại cấp bang (Malaysia có 13 bang): Mỗi bang có một chi nhánh của JPPH. Giám đốc chi nhánh do tổng giám đốc của JPPH bổ nhiệm và chịu trách nhiệm trước tổng giám đốc về toàn bộ hoạt động của cơ quan mình.

+ Tại cấp quận (huyện): Có một số ít quận (huyện) được JPPH đặt văn phòng định giá, còn hầu hết là thành lập các văn phòng định giá vùng, mỗi văn phòng chịu trách nhiệm quản lý nhà nước và làm dịch vụ định giá BĐS một vùng, bao gồm một số quận (huyện) nhất định, giám đốc văn phòng định giá do giám đốc chi nhánh JPPH bang bổ nhiệm [2].

- Ở Trung Quốc

Vấn đề định giá bất động sản được nhà nước Trung Quốc quan tâm từ đầu thập kỷ 90 của thế kỷ XX. Tháng 8/1992 “Trung tâm tư vấn và định giá bất động sản” được thành lập.

1.4. Tổng kết chương

Qua tìm hiểu ở các mục trên, chúng ta đã hiểu được phần nào bài toán phân bổ và định giá đất đai, biết được các yếu tố ảnh hưởng sâu sắc tới giá đất và các phương pháp định giá đất cũng như các ứng dụng của các phương pháp đó vào thực tế như thế nào.

CHƯƠNG 2: THIẾT KẾ THUẬT TOÁN DI TRUYỀN ĐỊNH HƯỚNG KHÔNG GIAN CHO BÀI TOÁN PHÂN BỐ VÀ ĐỊNH GIÁ ĐẤT

2.1. Tổng quan về thuật toán tối ưu tiến hóa

2.1.1. Cấu trúc dữ liệu và giải thuật tiến hóa

Chương trình tiến hóa trong công thức trên là khái niệm dùng để chỉ các chương trình máy tính có sử dụng thuật toán và tìm kiếm tối ưu hóa dựa trên nguyên lý tiến hóa tự nhiên. Ta gọi chung như thế là thuật toán tiến hóa[3].

a. Các thuật toán tiến hóa

- Thuật toán di truyền

Giải thuật di truyền là một kỹ thuật của khoa học máy tính nhằm tìm kiếm giải pháp thích hợp cho các bài toán tối ưu tổ hợp. Giải thuật di truyền là một phân ngành của giải thuật tiến hóa vận dụng các nguyên lý của tiến hóa như di truyền, đột biến, chọn lọc tự nhiên và trao đổi chéo [4].

- Thuật giải tối ưu hóa đàn kiến

Các thuật toán kiến lần đầu tiên được giới thiệu bởi Dorigo và các cộng sự như là cách tiếp cận đa tác tử tới các vấn đề về tối ưu tổ hợp khó, như bài toán người du lịch (TSP), bài toán người đưa thư. Hiện nay số lượng các ứng dụng ngày càng tăng và các nhà khoa học ứng dụng nó vào rất nhiều các vấn đề tối ưu rời rạc. Các ứng dụng gần đây có thể kể đến như các bài toán lập lịch, tô màu đồ thị, định hướng trong mạng truyền thông .v.v.[3].

- Thuật toán tối ưu hóa bầy đàn

Tối ưu hóa theo bầy đàn là một kỹ thuật tối ưu hóa ngẫu nhiên dựa vào một quần thể được phát triển bởi Eberhart và Kenedy, phỏng theo hành vi của các bầy chim hay các đàn cá. Cũng giống như GA, PSO tìm kiếm giải pháp tối ưu bằng việc cập nhật các thế hệ. Tuy nhiên không giống như GA, PSO

không có những thao tác tiến hóa như là lai ghép hay đột biến. Năm 1987, quan sát quá trình chuyển động của các bầy đàn (bầy chim, đàn cá), Renolds, Eberhart và Kennedy đưa thêm giả thiết về quá trình tìm về tổ của bầy đàn theo các quy luật:

- + Tất cả các phần tử trong bầy đàn đều có xu hướng chuyển động về tổ.
- + Mỗi phần tử đều ghi nhớ vị trí gần tổ nhất nó đã đạt được [10].

b. Giải thuật di truyền

Giải thuật di truyền được xây dựng dựa trên quy luật tiến hóa sinh học hay phát triển tự nhiên của một quần thể sống. Các cá thể trải qua một quá trình phát triển và sinh sản để tạo ra những cá thể mới cho thế hệ tiếp theo. Trong quá trình tăng trưởng và phát triển những cá thể xấu tức là những cá thể không thích nghi được với môi trường sẽ bị đào thải, ngược lại, những cá thể tốt sẽ được giữ lại (đây chính là quá trình chọn lọc) và được lai ghép (quá trình lai ghép) để tạo ra những cá thể mới cho thế hệ sau. Những cá thể mới được sinh ra mang những tính trạng của cá thể cha-mẹ (còn gọi là hiện tượng di truyền). Những cá thể được giữ lại có độ thích nghi khác nhau và quá trình lai ghép được thực hiện hoàn toàn ngẫu nhiên giữa các cá thể trong quần thể. Các cá thể được tạo ra trong quá trình lai ghép có thể sẽ xảy ra hiện tượng đột biến và tạo ra những cá thể khác với cá thể cha-mẹ. Cá thể này có thể tốt hơn hoặc xấu hơn cá thể cha-mẹ. Di truyền và đột biến là hai cơ chế có vai trò như nhau trong quá trình tiến hóa, mặc dù hiện tượng đột biến xảy ra với xác suất nhỏ hơn nhiều so với xác suất của hiện tượng di truyền. Và quá trình lai ghép và chọn lọc là hai quá trình cơ bản xuyên suốt quá trình tiến hóa tự nhiên [3].

Giải thuật di truyền cũng như các giải thuật tiến hóa nói chung được hình thành trên quan niệm cho rằng, quá trình tiến hóa là quá trình hoàn hảo nhất vì tự nó đã mang tính tối ưu. Thế hiện ở chỗ, cá thể sau được sinh ra bao giờ cũng tốt hơn, hoàn hảo hơn cá thể cha - mẹ, chúng có khả năng thích nghi

với sự thay đổi của môi trường cao hơn cá thể cha - mẹ. Tuy nhiên, không phải là tất cả, vẫn có một hoặc một số cá thể của thế hệ trước tốt hơn các cá thể ở thế hệ sau. Vì vậy, trong khi sử dụng giải thuật di truyền, chúng ta cần phải lưu lại những cá thể tốt nhất của mỗi thế hệ. Trải qua một số thế hệ (lần lặp) nhất định chúng ta đem so sánh những cá thể tốt nhất của những thế hệ với nhau và chọn ra một cá thể tốt nhất trong số các cá thể đó.

Vậy GAs làm gì?

Trong GA, một tập các biến của bài toán đưa ra được mã hóa sang một chuỗi (hay một cấu trúc mã hóa khác) tương tự như một nhiễm sắc thể trong tự nhiên. Mỗi chuỗi bao gồm một lời giải có thể của bài toán. Giải thuật di truyền sử dụng các toán tử được sinh ra bởi sự chọn lọc tự nhiên một quần thể các chuỗi nhị phân (hoặc các cấu trúc khác), mã hóa khoảng tham số trên mỗi thế hệ, khảo sát các phạm vi khác nhau của không gian tham số, và định hướng tìm kiếm đối với khoảng mà là xác suất cao để tìm kiếm sự thực hiện tốt hơn.

Giải thuật di truyền sử dụng một số thuật ngữ của di truyền học như gen, cá thể, nhiễm sắc thể, quần thể v.v.

Một cá thể trong giải thuật di truyền, biểu diễn một giải pháp của bài toán theo một cách nào đó mà chứa đầy đủ thông tin cần thiết về lời giải. Tuy nhiên, giải thuật di truyền cũng có những điểm khác với hiện tượng di truyền trong tự nhiên. Trong tự nhiên, một cá thể có thể có một hoặc nhiều nhiễm sắc thể (NST), nếu có một NST được gọi là thể đơn bội, có nhiều NST được gọi là thể đa bội. Ở đây, để giới hạn trong giải thuật di truyền ta quan niệm một cá thể có một NST. Do đó khái niệm cá thể và NST trong giải thuật di truyền coi như là tương đương.

Một nhiễm sắc thể (NST) được tạo thành từ nhiều gen, mỗi gen có thể có các giá trị khác nhau để quy định một tính trạng nào đó. Trong giải thuật di truyền, một gen được coi như một phần tử trong chuỗi NST.

Quần thể là tập hợp các cá thể (các giải pháp) của môi trường (bài toán) đang xét. Kích thước quần thể cho biết có bao nhiêu cá thể của quần thể (trong một thế hệ). Qua các nghiên cứu cũng như các thử nghiệm đã cho thấy kích thước quần thể không nên quá bé cũng như không quá lớn. Nếu có quá ít cá thể thì ít có khả năng thực hiện lai giống và chỉ một phần nhỏ không gian tìm kiếm được dùng. Như vậy, sẽ dễ xảy ra trường hợp bỏ qua các lời giải tốt. Nhưng quá nhiều cá thể cũng không tốt vì GA sẽ chạy chậm đi, ảnh hưởng đến hiệu quả của giải thuật.

Thuật toán di truyền gồm có 4 quy luật cơ bản là lai ghép, đột biến, sinh sản và chọn lọc tự nhiên.

- Quá trình lai ghép (phép lai)

Quá trình này diễn ra bằng cách ghép một hay nhiều đoạn gen từ hai nhiễm sắc thể cha-mẹ để hình thành nhiễm sắc thể mới mang đặc tính của cả cha lẫn mẹ. Phép lai có thể mô tả như sau:

Chọn ngẫu nhiên hai hay nhiều cá thể trong quần thể. Giả sử chuỗi nhiễm sắc thể của cha và mẹ đều có chiều dài là m . Tìm điểm lai bằng cách tạo ngẫu nhiên một con số từ 1 đến $m-1$. Như vậy điểm lai này sẽ chia hai chuỗi nhiễm sắc thể cha-mẹ thành 2 nhóm nhiễm sắc thể con là m_1 và m_2 . Hai chuỗi nhiễm sắc thể con lúc này sẽ là $m_{11}+m_{22}$ và $m_{21}+m_{12}$. Đưa hai chuỗi nhiễm sắc thể con vào quần thể để tiếp tục tham gia quá trình tiến hóa.

- Quá trình đột biến (phép đột biến)

Quá trình tiến hóa được gọi là quá trình đột biến khi một hoặc một số tính trạng của con không được thừa hưởng từ hai chuỗi nhiễm sắc thể cha-mẹ.

Phép đột biến xảy ra với xác suất thấp hơn rất nhiều so với xác suất xảy ra phép lai. Phép đột biến có thể mô tả như sau:

- + Chọn ngẫu nhiên một số k từ khoảng $1 \leq k \leq m$.
- + Thay đổi giá trị của gen thứ k .
- + Đưa nhiễm sắc thể con vào quần thể để tham gia quá trình tiến hóa tiếp theo.

- Quá trình sinh sản và chọn lọc (phép tái sinh và phép chọn)

+Phép tái sinh: Là quá trình các cá thể được sao chép dựa trên độ thích nghi của nó. Độ thích nghi là một hàm được gán các giá trị thực cho các cá thể trong quần thể của nó. Phép tái sinh có thể được mô tả như sau:

Tính độ thích nghi của từng cá thể trong quần thể, lập bảng cộng dồn các giá trị thích nghi đó (theo thứ tự gán cho từng cá thể) ta được tổng độ thích nghi. Giả sử quần thể có n cá thể. Gọi độ thích nghi của cá thể thứ i là f_i , tổng dồn thứ t là F_t . Tổng độ thích nghi là F_m . Tạo số ngẫu nhiên f có giá trị trong đoạn từ 0 đến f_m .

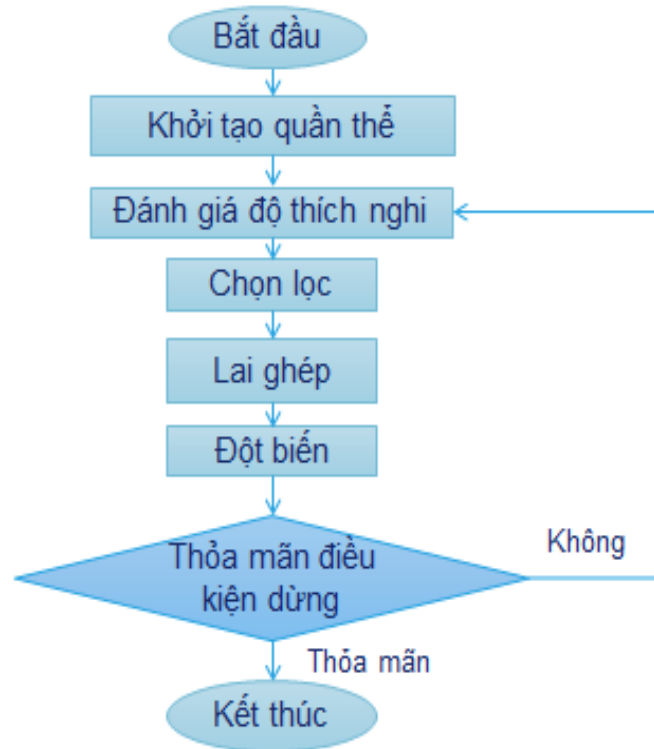
Chọn cá thể k đầu tiên thỏa mãn $f \geq f_i$ đưa vào quần thể của thế hệ mới.

+ Phép chọn: Là quá trình loại bỏ các cá thể xấu và để lại những cá thể tốt. Phép chọn được mô tả như sau:

Sắp xếp quần thể theo độ thích nghi giảm dần.

Loại bỏ các cá thể cuối dãy và chỉ để lại n cá thể tốt[2].

c. Cấu trúc thuật giải di truyền tổng quát



Hình 2.1: Sơ đồ thuật toán di truyền [2].

d. Các toán tử của giải thuật di truyền.

- Toán tử chọn lọc
- + Chọn lọc dựa trên độ thích nghi.
- Toán tử đột biến.
- Toán tử lai ghép
- + Lai ghép một điểm.
- + Lai ghép hai điểm.
- + Lai ghép N điểm.
- + Lai ghép đồng nhất.

e. Các tham số của giải thuật di truyền.

- Xác suất lai ghép: Là tham số cho biết tần suất thực hiện toán tử lai ghép. Nếu không có lai ghép, cá thể con chính là bản sao của cá thể cha mẹ.

Nếu xác suất lai ghép là 100%, khi đó mọi cá thể con đều được tạo ra trong quá trình lai ghép.

- Xác suất đột biến: Là tham số cho biết tần suất đột biến của NST. Nếu không có đột biến, thế hệ con được tạo ra ngay sau giai đoạn lai ghép mà không bị thay đổi. Ngược lại, một hoặc một số phần của NST sẽ bị thay đổi. Nếu xác suất đột biến là 100%, toàn bộ NST bị thay đổi. Nếu xác suất là 0% thì không có gì bị thay đổi hết.

- Kích thước quần thể: Là tham số cho biết có bao nhiêu cá thể (NST) trong một thế hệ của quần thể. Nếu có quá ít cá thể, khả năng thực hiện lai ghép rất nhỏ và khi đó chỉ có một vùng tìm kiếm nhỏ mới được khảo sát. Ngược lại, việc kích thước quần thể quá lớn cũng không tốt, do nó sẽ làm chậm quá trình giải bài toán.[2]

2.1.2. Công thức của giải thuật di truyền.

- Tính độ thích nghi eval (v_i) của mỗi NST v_i ($i = 1 \dots$ kích thước quần thể)

$$\text{eval}(v_i) = \frac{f(v_i)}{\sum_{i=1}^{\text{kichthuc quanthe}} f(v_i)} \quad \text{với } f(v_i) \text{ là hàm mục tiêu.}$$

- Tìm giá trị thích nghi của quần thể

$$F = \sum_{i=1}^{\text{kichthuc quanthe}} \text{eval}(v_i)$$

- Tính xác suất chọn P_i cho mỗi nhiễm sắc thể v_i

$$P_i = \frac{\text{eval}(v_i)}{\sum_{i=1}^{\text{kichthuc quanthe}} \text{eval}(v_i)}$$

- Tính xác suất tích lũy p_i cho mỗi nhiễm sắc thể P_i .

$$q_i = \sum_{j=1}^i p_j$$

Tiến trình được thực hiện bằng việc quay bánh xe rulet kích thước quần thể lần. Mỗi lần chọn ra một nhiễm sắc thể từ quần thể hiện hành vào quần thể mới theo cách sau:

- Phát sinh một số ngẫu nhiên r trong khoảng $[0,1]$

- Nếu $r < q_1$ thì chọn nhiễm sắc thể v_1 , ngược lại chọn nhiễm sắc thể v_i ($2 \leq i \leq$ kích thước quần thể) sao cho $q_{i-1} < r \leq q_i$

2.1.3. Các thành phần của thuật toán

- Khởi tạo quần thể ban đầu

Tạo quần thể ban đầu trong giải thuật, là nơi xuất phát quá trình tiến hóa, bao gồm tất cả các giá trị thô ban đầu. Tùy theo vấn đề của bài toán mà có cách khởi tạo khác nhau. Trước một bài toán áp dụng thuật toán di truyền, ta cần phải xác định rõ nhiễm sắc thể và cá thể cho vấn đề, và thông thường đó sẽ là kết quả cuối cùng. Việc phân tích dựa trên kết quả là cơ bản nhất.

- Toán tử lai ghép

- + Lai ghép nhằm nâng cao kết quả cá thể, do đó, toán tử lai ghép sẽ tạo điều kiện cho tiến trình hội tụ nhanh hay chậm. Còn tùy thuộc vào cách tổ chức và phân bố các nhiễm sắc thể mà chúng ta có xác suất lai ghép nhanh hay chậm.

- Toán tử đột biến

Cũng như toán tử lai ghép, toán tử đột biến làm tăng quá trình hội tụ, nhưng tăng một cách đột ngột, cũng có thể không gây tác dụng gì khi không thành công.

- Điều kiện dừng

Giải thuật di truyền là một quá trình ngẫu nhiên, nên chúng ta không thể đảm bảo chắc chắn giải thuật sẽ dừng sau hữu hạn bước. Vì vậy, để đảm bảo giải thuật di truyền sẽ kết thúc, người dùng thường phải định nghĩa điều kiện dừng cho thuật toán. Một vài trường hợp dừng thông thường như sau:

Kết thúc theo kết quả: Một khi đạt đến mức giá trị yêu cầu thì kết thúc kết quả thực hiện. Hoặc giá trị trung bình của độ thích nghi trên tất cả các cá thể của quần thể không thay đổi.

Kết thúc dựa vào số thế hệ: Chọn số thế hệ, quá trình sẽ dừng đúng ngay số thế hệ đã qui định trước, không cần biết kết quả như thế nào.

Tính theo thời gian: Không cần biết qua bao nhiêu thế hệ hay kết quả như thế nào, chỉ dựa vào thời gian qui định kết thúc.

Tổ hợp: Dùng nhiều phương pháp khác nhau cho vấn đề chẳng hạn như: chạy theo số thế hệ, tiếp đến đánh giá cho chạy theo kết quả hoặc ngược lại.”

2.2. Thiết kế thuật toán di truyền không gian cho bài toán phân bổ và định giá đất

2.2.1. Thuật toán di truyền

Bài toán định giá đất dựa vào thuật toán di truyền rất phức tạp. Bài toán dựa vào rất nhiều các yếu tố để có thể dự đoán được giá đất như đã được đề cập ở phần trên.

Input: Dữ liệu giá đất có dạng $(x_1^j, x_1^j, x_2^j, \dots, x_n^j)$ trong đó x_i là đặc trưng thứ i ảnh hưởng đến giá đất.

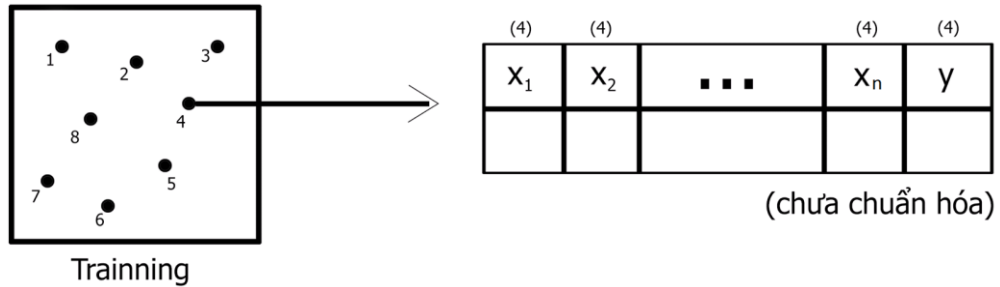
Các yếu tố ảnh hưởng đến giá đất sử dụng trong bài toán tương ứng với các giá trị x_i^j nêu ở trên:

- Quận/huyện.
- Thời gian đến trung tâm thành phố.
- Chất lượng dịch vụ điện nước, an ninh, rác thải.

- Chiều dài mặt tiền.
- Khoảng cách tới trường học, bệnh viện, tuyến đường lớn.
- Số phòng ở, loại nhà v.v.

Output: Giá đất dự báo.

a. Chuẩn hóa một nhiệm sắc thể trong GA:



Các đặc trưng ảnh hưởng tới giá đất được chuẩn hóa trong đoạn $[0,1]$ bằng công thức :

$$X_i^j = \frac{x_i^j - x_{\min}^j}{x_{\max}^j - x_{\min}^j}$$

trong đó : j : các điểm

- x_i^j : giá trị chưa chuẩn hóa

- x_{\min}^j : giá trị min của cột

- x_{\max}^j : giá trị max của cột

- X_i^j : giá trị đã chuẩn hóa

Mã hóa cá thể: Mỗi cá thể biểu diễn 1 phương án của bài toán, cụ thể mỗi cá thể (w_1, w_2, \dots, w_n) biểu diễn một phương án của bài toán. Các giá trị w_i được mã hóa dưới dạng số thực trong đoạn $[0,1]$. Giá trị w_i chính là trọng số của đặc trưng x_i . Mục tiêu của bài toán là đi tìm ra 1 phương án (w_1, w_2, \dots, w_n) phù hợp nhất với bài toán.

$$W = (W_1^{(1)}, \dots, W_n^{(1)}), (W_1^{(2)}, \dots, W_n^{(2)}), \dots, (W_1^{(n)}, \dots, W_n^{(n)})$$

Dựa vào các dữ giá đất thu thập được ta xác định được hàm mục tiêu (fitness) của bài toán:

$$Fitness = \left(\sum_{j=1}^m \sum_{i=1}^n W_i^{(j)} X_i^{(j)} \right) - Y_i^{(j)} \rightarrow \min$$

với:

m số điểm.

n là số lượng các đặc trưng ảnh hưởng tới giá đất.

W_i^j là trọng số của đặc trưng thứ i của điểm thứ j.

Với 1 điểm thứ j thì $W_1 + W_2 + \dots + W_n = 1$.

x_1, x_2, \dots, x_n là các đặc trưng ảnh hưởng tới giá đất.

$0 \leq x_1, x_2, \dots, x_n \leq 1$.

- y^1, y^2, \dots, y^n là giá trị được chuẩn hóa từ giá đất trên thực tế.

b. Tối ưu và huấn luyện trọng số cho điểm trên bản đồ.

Bài toán định giá đất dựa vào thuật toán di truyền cũng phải trải qua 3 quá trình cơ bản, đó là: chọn lọc, lai ghép và đột biến.

Các quá trình của thuật toán:

- Chọn lọc:

+ Sắp xếp quần thể theo giá trị fitness

+ Chọn một nửa quần thể với fitness tốt nhất để đưa vào lai ghép và đột biến

- Lai ghép:

+ Chọn hai cá thể bất kì đại diện cho cá thể cha và cá thể mẹ.

+ Sau đó ta sinh ngẫu nhiên một giá trị từ 2 đến n-1 (với n ở đây là số điểm cần định giá).

+ Ghép nửa đầu của cá thể cha với nửa sau của cá thể mẹ và chuẩn hóa lại được cá thể mới.

+ Xác suất lai ghép là 95%.

- Đột biến:

+ Chọn một cá thể (w_1, w_2, \dots, w_n) bất kỳ tham gia vào quá trình đột biến.

+ Tạo ngẫu nhiên 2 giá trị trong khoảng $(2, n-1)$.

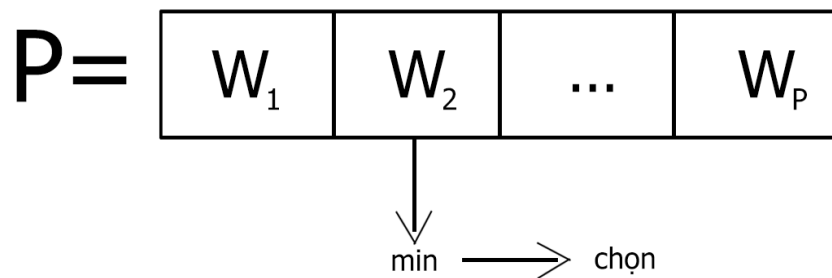
+ Sau đó đổi chỗ vị trí 2 gen cho nhau được cá thể mới.

+ Xác suất đột biến là 5%.

- Lặp lại các quá trình trên đến khi kết thúc số thế hệ đã định trước.

Sau đó dùng phương án cuối cùng để dự báo dựa vào hàm mục tiêu.

c. Sau khi huấn luyện ta có bộ trọng số W làm cho Fitness min.



Hình 2.2: Mô tả bộ trọng số W

Ở đây ta chọn $P=50$.

d. Độ phức tạp của thuật toán

- Độ phức tạp của thuật toán là: $O(p*m*n)$

2.2.2. Thuật toán di truyền định hướng không gian (SGA)

$$\text{FitnessSGA}(i) = \frac{(\text{Fitness}(i) + \text{Fitness}(j))}{3}, i \geq 3, j [1..2]$$

$N(j)$: Tập hàng xóm điểm j mặc định là 2 điểm xung quanh gần nhất.

$\text{Fitness}(i)$: Tính bằng thuật toán di truyền.

Tính $N(j)$:

Dựa vào cách thuật toán tính khoảng cách không gian giữa 2 điểm:

$$\text{Euclidean: } \text{Dist}(x_1, x_2) = \sqrt{(x-a)^2 + (y-b)^2}$$

Dist : Khoảng cách.

x : Kinh độ điểm x_1

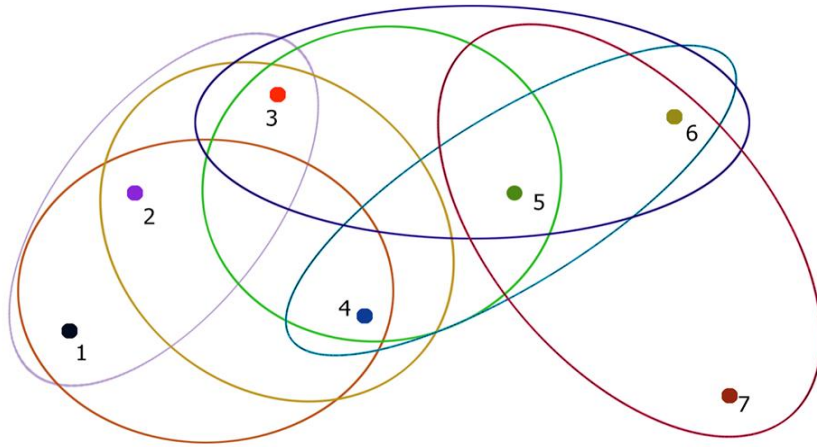
y : Vĩ độ điểm x_1

a : Kinh độ điểm x_2

b : Vĩ độ điểm x_2

Tính tuần tự khoảng cách từ i đến tất cả các điểm, lấy 2 điểm có khoảng cách ngắn nhất.

Hình minh họa:



Hình 2.3: Mô tả tìm hai hàng xóm

Điểm số 1 có 2 hàng xóm: 2,4 Điểm số 2 có 2 hàng xóm: 1,3

Điểm số 3 có 2 hàng xóm: 2,4 Điểm số 4 có 2 hàng xóm: 5,3

Điểm số 5 có 2 hàng xóm: 6,4 Điểm số 6 có 2 hàng xóm: 5,7

a. Ưu điểm

- Thuật toán di truyền định hướng không gian sẽ cho kết quả tốt hơn GA vì nó kế thừa các thuộc tính tốt nhất của GA.

- Thuật toán di truyền không gian trong bài toán phân bố và định giá đất sẽ sử dụng bộ trọng số tối ưu nhất và tốt nhất cho bài toán.

b. Nhược điểm

- Độ chính xác của bài toán phụ thuộc nhiều vào bộ dữ liệu training, bộ dữ liệu càng lớn thì kết quả của bộ tham số sẽ càng tốt.

- Thuật toán phải chạy qua rất nhiều các thế hệ nên thuật toán chạy rất mất thời gian khi chạy với bộ dữ liệu lớn.

- Độ phức tạp của thuật toán là: $O(p*m*n)$ trong đó p là số cá thể, m là số điểm, n là các yếu tố ảnh hưởng đến giá.

2.3. Phân loại vùng không gian trên bản đồ theo giá đất hỗ trợ phân bố đất đai bằng thuật toán FCM

Dựa vào kết quả của SGA ta có bộ dữ liệu không gian về các điểm giá đất từ đó ta sẽ xây dựng bài toán phân bố đất đai sử dụng thuật toán phân cụm FCM.

Đầu ra của thuật toán định hướng không gian SGA là tập điểm với giá đã được định giá đây cũng là đầu vào của thuật toán FCM, nhiệm vụ của FCM sẽ phân giá thành các cụm với mức giá từ thấp đến cao

Thuật toán FCM [3] là thuật toán phân cụm được sử dụng rất rộng rãi. Mặc dù nó chưa sử dụng các tham số địa lý nhưng nó lại là tiền đề để phát triển các thuật toán phân cụm dữ liệu địa lý sau này.

Thuật toán FCM được mô tả như sau:

- Đầu vào: các điểm có giá đất đã dự báo (kết quả của SGA)
 - Tập dữ liệu đầu vào X , số mờ m .
 - Số điểm dữ liệu N , số cụm C , số chiều r .
 - Ngưỡng ε .
- Đầu ra:
 - C cụm dữ liệu sao cho thỏa mãn hàm mục tiêu:

$$J = \sum_{k=1}^N \sum_{j=1}^C u_{kj}^m \|X_k - V_j\|^2 \rightarrow \min \quad (1.13)$$

- Các bước thực hiện thuật toán:

Bước 1: Khởi tạo ma trận $U(t)$ với $t = 0$.

Bước 2: Tính ma trận tâm $V(t)$ bởi công thức:

$$V_i = \frac{\sum_{k=1}^N u_{ki}^m X_k}{\sum_{k=1}^N u_{ki}^m}; i = \overline{1, C} \quad (1.14)$$

Bước 3: Tính $U(t+1)$ bởi công thức:

$$u_{ki} = \frac{1}{\sum_{j=1}^C \left(\frac{\|X_k - V_i\|}{\|X_k - V_j\|} \right)^{\frac{2}{m}}}; i = \overline{1, C}; k = \overline{1, N} \quad (1.15)$$

Bước 4: Nếu $\|U(t+1) - U(t)\| \leq \varepsilon$ thì dừng thuật toán, ngược lại thì quay lại bước 2.

- Ưu điểm [3]:

- Thuật toán đơn giản, dễ thực hiện.

- Nhược điểm [3]:

- Nhạy cảm với các nhiễu và phần tử ngoại lai trong dữ liệu

Đây là thuật toán phân cụm mờ nói chung, chưa sử dụng các yếu tố địa lý.

2.4. Tổng kết chương

Qua chương 2, chúng ta đã hiểu và nắm rõ được bản chất của chiến lược tối ưu tiên hóa nói chung cũng như thuật toán di truyền nói riêng. Mặt khác, chúng ta hiểu rõ được quá trình thực hiện của thuật toán di truyền định hướng không gian áp dụng vào trong bài toán định giá đất.

CHƯƠNG 3: XÂY DỰNG HỆ THỐNG THÔNG TIN CHO BÀI TOÁN PHÂN BỐ VÀ ĐỊNH GIÁ ĐẤT

3.1. Giới thiệu về hệ thống LIS

3.1.1. Khái niệm

WebGIS là sự kết hợp giữa công nghệ GIS và Webform để đưa bản đồ lên trên web, hay nói cách khác là bản đồ trực tuyến. WebGIS có thể áp dụng cho nhiều lĩnh vực như giao thông, thủy lợi, thời tiết, quản lý hành chính v.v.[8].

- GIS là một hệ thống có ứng dụng rất lớn. Từ năm 1980 đến nay đã có rất nhiều các định nghĩa được đưa ra, tuy nhiên không có định nghĩa nào khái quát đầy đủ về GIS vì phần lớn chúng đều được xây dựng trên khía cạnh ứng dụng cụ thể trong từng lĩnh vực. Có ba định nghĩa được dùng nhiều nhất.

+ GIS là một hệ thống thông tin được thiết kế để làm việc với các dữ liệu trong một hệ tọa độ quy chiếu. GIS bao gồm một hệ cơ sở dữ liệu và các phương thức để thao tác với dữ liệu đó.

+ GIS là một hệ thống nhằm thu thập, lưu trữ, kiểm tra, tích hợp, thao tác, phân tích và hiển thị dữ liệu được quy chiếu cụ thể vào trái đất.

+ GIS là một chương trình máy tính hỗ trợ việc thu thập, lưu trữ, phân tích và hiển thị dữ liệu bản đồ.

- Một cách khái quát, có thể hiểu một hệ GIS như là một quá trình sau:

+ Dữ liệu vào: dữ liệu được thu thập từ các nguồn khác nhau như chuyển đổi giữa các cách biểu diễn dữ liệu, máy quét, hình ảnh từ vệ tinh, ảnh chụp v.v.

+ Quản lý dữ liệu: sau khi dữ liệu được thu thập và tổng hợp, GIS cần cung cấp các thiết bị có thể lưu và bảo trì dữ liệu. GIS lưu thông tin thể giới thực hành các tầng dữ liệu riêng biệt, các tầng này đặt trong cùng một hệ

trục tọa độ và chúng có khả năng liên kết với nhau.

+ Xử lý dữ liệu: các thao tác xử lý dữ liệu được thực hiện để tạo ra thông tin. Nó giúp cho người sử dụng quyết định cần làm tiếp công việc gì. Kết quả của xử lý dữ liệu là tạo ra các ảnh, báo cáo và bản đồ.

+ Phân tích và mô hình: số liệu tổng hợp và chuyển đổi chỉ là một phần của GIS. Những yêu cầu tiếp theo là khả năng giải mã và phân tích về mặt định tính và định lượng thông tin đã thu thập.

+ Dữ liệu ra: Một trong các phương diện công nghệ GIS là sự thay đổi các phương pháp khác nhau trong đó thông tin có thể hiển thị khi nó được xử lý bằng GIS. Các phương pháp truyền thông là bảng và đồ thị có thể cung cấp bằng bản đồ và ảnh 3 chiều.

- GIS bao gồm 5 thành phần chính:

- + Con người.
- + Dữ liệu..
- + Phương pháp phân tích
- + Phần mềm.
- + Phần cứng.

3.1.2. Kiến thức chung

Kiến trúc web của hệ thống thông tin dữ liệu không gian cũng gần giống như kiến trúc dành cho một hệ thống thông tin web cơ bản khác, ngoại trừ có sử dụng kỹ thuật GIS. Có nhiều dạng công nghệ cho việc thành lập web cho thông tin không gian như: MapServer, GeoServer, ArcGIS Server, v.v.

Cơ sở dữ liệu không gian sẽ được dùng để quản lý và truy xuất dữ liệu không gian, được đặt trên máy chủ dữ liệu. Nhà kho hay nơi lưu trữ được dùng để lưu trữ và duy trì siêu dữ liệu về những dữ liệu không gian tại những máy chủ dữ liệu khác nhau. Dựa trên những thành phần quản lý dữ

liệu, ứng dụng máy chủ và mô hình máy chủ được dùng cho ứng dụng hệ thống để tính toán thông tin không gian qua các hàm cụ thể. Tất cả kết quả tính toán của ứng dụng máy chủ sẽ được gửi đến máy chủ web để thêm vào các gói HTML, gửi cho phía client và hiển thị nơi trình duyệt web.

- Người dùng gửi yêu cầu của người sử dụng thông qua giao thức truyền tải siêu văn bản đến máy chủ web

- Máy chủ web nhận yêu cầu từ client, xử lý và chuyển tiếp yêu cầu đến ứng dụng trên máy chủ có liên quan.

- Máy chủ của ứng dụng nhận các yêu cầu cụ thể đối với các ứng dụng và gọi các hàm có liên quan để tính toán xử lý. Nếu có yêu cầu dữ liệu nó sẽ gửi yêu cầu dữ liệu đến trung tâm trao đổi dữ liệu

- Trung tâm trao đổi dữ liệu nhận yêu cầu dữ liệu, tìm kiếm vị trí dữ liệu, sau đó gửi yêu cầu dữ liệu đến máy chủ dữ liệu chứa dữ liệu cần tìm.

- Máy chủ dữ liệu tiến hành truy vấn dữ liệu cần thiết và trả dữ liệu này về cho trung tâm trao đổi dữ liệu

- Trung tâm trao đổi dữ liệu nhận nhiều nguồn dữ liệu từ máy chủ dữ liệu, sắp xếp logic dữ liệu theo yêu cầu và trả dữ liệu về cho máy chủ ứng dụng.

- Máy chủ ứng dụng nhận dữ liệu trả về từ các Trung tâm trao đổi dữ liệu và đưa chúng đến các hàm cần sử dụng, xử lý, trả kết quả về Web Server.

- Máy chủ web nhận kết quả xử lý, thêm vào các mã nguồn HTML, PHP, v.v. để có thể hiển thị lên trình duyệt, gửi trả kết quả về cho trình duyệt dưới dạng các trang web.

Kiến trúc 3-tier gồm 3 thành phần cơ bản, đại diện cho 3 tầng:

- + Cơ sở dữ liệu: là nơi lưu trữ các dữ liệu địa lý bao gồm các dữ liệu không gian và phi không gian. Các dữ liệu này được quản trị bởi các hệ quản trị cơ sở dữ liệu như: Oracle, MS SQL Server, Esri SDE, PostgreSQL, v.v. hoặc là các dạng file dữ liệu như: Shapefile, Tab, XML, v.v. Các dữ

liệu này được thiết kế cài đặt và xây dựng theo từng quy trình cụ thể. Tùy theo quy mô và yêu cầu của hệ thống mà tổ chức lựa chọn công nghệ quản trị cơ sở dữ liệu cho phù hợp.

+ Máy chủ ứng dụng: thường được tích hợp trong một Web server nào đó, ví dụ như các Web server nổi tiếng Apache Tomcat, Internet Information Server. Đó là một ứng dụng phía Server nhiệm vụ chính của nó là tiếp nhận các yêu cầu từ client, lấy dữ liệu từ phía cơ sở dữ liệu theo yêu cầu client, trình bày dữ liệu theo cấu hình định sẵn hoặc theo yêu cầu của client và trả kết quả về theo yêu cầu.

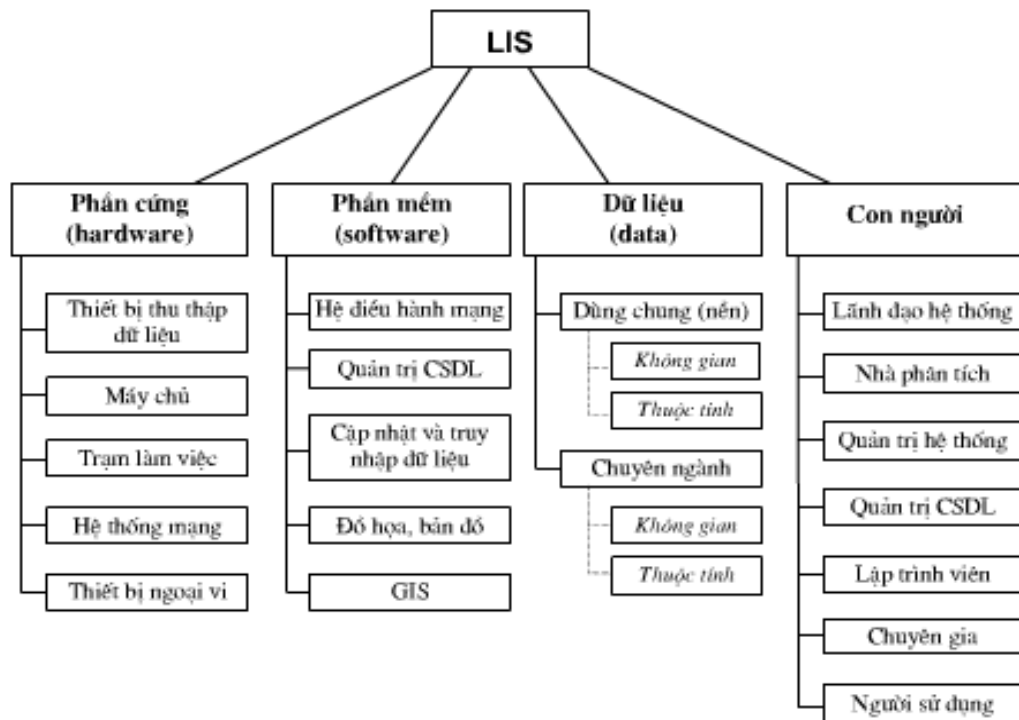
+ Người sử dụng: thông thường đơn thuần là một Browser như Internet Explorer, FireFox, Google Chrome, v.v để mở các trang web theo URL định sẵn. Các ứng dụng client có thể là một Website, Applet, Flash, v.v được viết bằng các công nghệ chuẩn mà W3C đã chứng thực. Client đôi khi cũng là một ứng dụng Desktop tương tự như phần mềm MapInfo, ArcMap, v.v.

3.1.3 Chức năng và vai trò của một hệ thống tin đất đai

a. Khái niệm

- Hệ thống thông tin đất đai: là hệ thống thông tin cung cấp các thông tin về đất đai. Nó là công cụ cho việc tạo quyết định về mặt pháp luật, hành chính, kinh tế, trợ giúp cho công tác quy hoạch và phát triển. Nó bao gồm một mặt là cơ sở dữ liệu lưu trữ những dữ liệu không gian tham chiếu có liên quan đến đất đai trong một vùng địa lí nhất định và một mặt là một tập hợp các quy trình và công nghệ thu thập, cập nhật, xử lí và phân phối dữ liệu một cách có hệ thống. Cơ sở cho mọi hệ thống thông tin đất đai là một hệ thống tham chiếu không gian cho dữ liệu trong hệ thống đồng thời có khả năng liên kết với các dữ liệu có liên quan đến đất đai trong các hệ thống khác[6].

b. Các thành phần trong hệ thống thông tin:



Hình 3.1: Các thành phần của LIS[2].

- Phần cứng:

+ Máy chủ: Là các máy tính có tốc độ cao, được thiết kế để có thể hoạt động ổn định và liên tục trong một thời gian dài. Máy chủ là nơi lưu trữ và xử lý thông tin của hệ thống.

+ Các trạm làm việc: Là phương tiện để người sử dụng truy nhập hệ thống và xử lý thông tin. So với máy tính cá nhân thông dụng thì các trạm làm việc thường có tốc độ cao hơn và không có hoặc có rất ít các phương tiện giải trí.

+ Thiết bị ngoại vi: Các thiết bị ngoại vi hỗ trợ việc thu nhập và in ấn

+ Hệ thống mạng: Là phương tiện để kết nối các trạm làm việc với nhau và với server. Tùy theo lượng máy tính và phạm bố địa lí của chúng, người ta phân biệt mạng cục bộ, mạng diện rộng và mạng Internet toàn cầu.

+ Các thiết bị thu thập dữ liệu: Trong hệ thống LIS có thể có một số thiết bị dùng để thu thập dữ liệu không gian bằng phương pháp tự động hoặc bán tự động [2].

- Phần mềm:

+ Hệ điều hành mạng: Là nền tảng và môi trường hoạt động của các phần mềm ứng dụng khác. Có hệ điều hành mạng chạy trên máy chủ và loại chạy trên các trạm làm việc.

+ Phần mềm quản trị cơ sở dữ liệu: Có chức năng cung cấp các phương thức cập nhật, khai thác dữ liệu và thực hiện các chính sách bảo mật dữ liệu.

+ Phần mềm đồ họa và bản đồ: Là các phần mềm dùng để thành lập và hiển thị các bản đồ.

+ Phần mềm GIS: Được sử dụng chủ yếu để phân tích và tổng hợp thông tin nhằm giải quyết một bài toán nào đó. Một số phần mềm GIS thông dụng hiện nay là MapInfo, ArcView - ArcInfo, v.v.

- Con người

Cùng với dữ liệu, yếu tố con người đóng vai trò quyết định đối với hệ thống LIS. Để hệ thống có thể hoạt động được thì cần nguồn nhân lực tối thiểu như sau:

+ Người lãnh đạo hệ thống: Chỉ đạo tổ chức và vận hành hệ thống.

+ Nhà phân tích: Là người có kiến thức vững vàng về chuyên môn cũng như về công nghệ thông tin, có vai trò chuyển đổi các nhu cầu của người sử dụng thành các nhiệm vụ và bài toán của hệ thống.

+ Người quản trị hệ thống: Có nhiệm vụ đảm bảo các thiết bị và các phần mềm hoạt động liên tục trong một thời gian dài [2].

+ Người quản trị cơ sở dữ liệu: Chịu trách nhiệm về các tiêu chuẩn, tài liệu và thiết kế kỹ thuật của cơ sở dữ liệu.

+ Lập trình viên: Có nhiệm vụ chuyển đổi các nhiệm vụ và bài toán do nhà phân tích đặt ra thành các phần mềm ứng dụng của hệ thống.

+ Một số vị trí khác như: Nhà bản đồ học có nhiệm vụ sản xuất ra các bản đồ có chất lượng cao. Kỹ thuật viên có nhiệm vụ chuyển đổi về dạng số các dữ liệu còn ở dạng khác. Chuyên gia giải đoán ảnh có chức năng khai thác các thông tin địa lí từ ảnh viễn thám.

+ Người sử dụng: Là những người khác thác thông tin của hệ thống

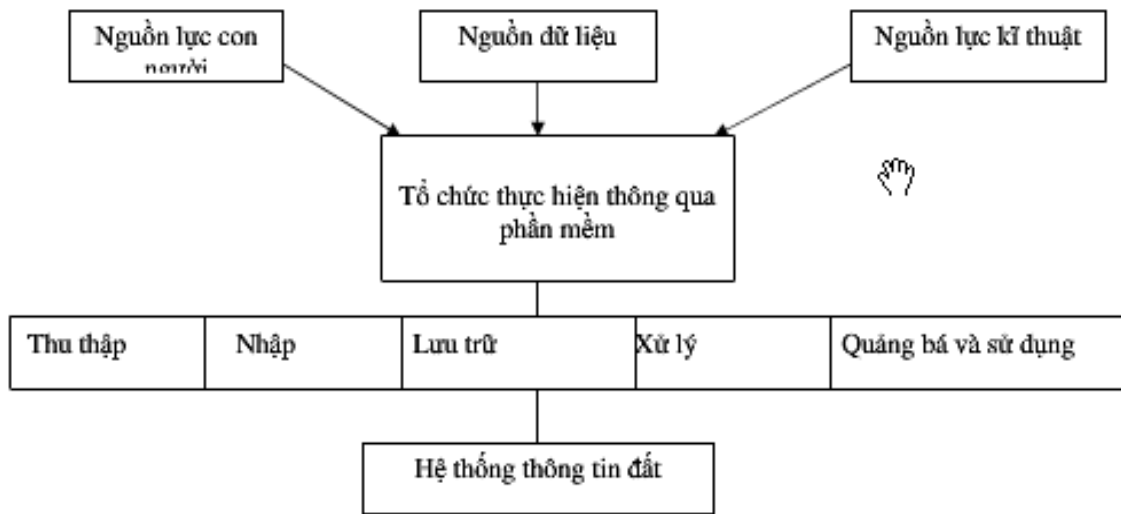
- Dữ liệu: Trong một hệ thống tin đất đai nói riêng và hệ thống thông tin địa lí nói chung, dữ liệu là thành phần quan trọng nhất và đòi hỏi kinh phí lớn nhất. Cơ sở dữ liệu là tập hợp các thông tin được tổ chức một cách hệ thống, có cấu trúc để phục vụ cho một hay nhiều mục đích nào đó. Theo nội dung có thể phân loại dữ liệu thành dữ liệu không gian để mô tả đặc trưng không gian của các đối tượng địa lí và dữ liệu thuộc tính để giải thích đặc điểm của các đối tượng này. Theo mục đích sử dụng, dữ liệu có thể phân loại theo tính chất của các đối tượng, có dữ liệu dùng chung (dữ liệu nền) và dữ liệu dùng riêng (dữ liệu chuyên đề). Hệ thống các dữ liệu về đất đai bao gồm:

+ Các dữ liệu về hệ thống tọa độ, độ cao nhà nước: Hệ thống này đảm bảo cơ sở pháp lí về hệ quy chiếu và hệ thống tọa độ, độ cao thống nhất trong cả nước, độ chính xác vị trí trong không gian của điểm tọa độ, độ cao được xác lập theo chuẩn quốc gia.

+ Các dữ liệu nền địa lí: địa danh, địa giới và địa hình.

+ Các dữ liệu về hồ sơ địa chính bao gồm: Thông tin vị trí, hình thể và kích thước tự nhiên của thửa đất, thông tin này thể hiện trên bản đồ địa chính dưới dạng các đối tượng hình học. Thông tin về tình hình pháp lí và trạng thái sử dụng của thửa đất: người sử dụng, mục đích sử dụng, thời hạn sử dụng, sổ giấy chứng nhận quyền sử dụng đất và các tài liệu pháp lí khác có liên quan.

Thông tin về kinh tế đất: phân hạng đất, giá trị của thửa đất và các công trình trên đất. Các dữ liệu về quy hoạch đất, kế hoạch sử dụng đất.



Hình 3.2: Sơ đồ vận hành của LIS

c. Chức năng, vai trò của LIS

Là một bộ phận của GIS, LIS cũng có chức năng cơ bản của hệ thống địa lí: nhập dữ liệu, quản trị dữ liệu, phân tích, mô hình hóa dữ liệu, hiển thị dữ liệu. Các chức năng này đảm bảo khả năng thực hiện các vai trò của LIS:

- Tạo một cơ sở dữ liệu nền địa lí đầy đủ và thống nhất (cho một vùng hay một lãnh thổ) để thể hiện các thông tin có liên quan đến không gian.

- Là công cụ trực tiếp phục vụ cho việc hoạch định các chính sách đất đai, cho các quy định về quy hoạch, kế hoạch sử dụng hợp lí và hiệu quả đất đai cho các mục tiêu quy hoạch phát triển kinh tế xã hội.

- Là công cụ để quản lí việc sử dụng đất đai đến từng thửa đất.

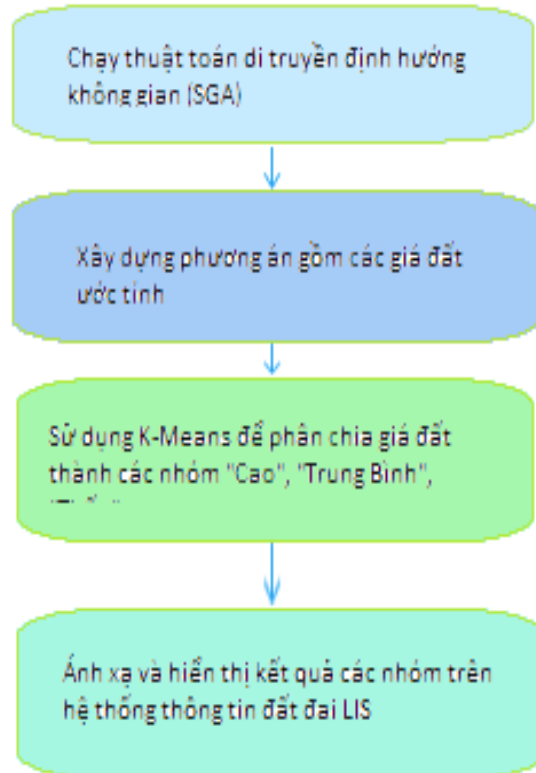
- Là công cụ để quản lí thống nhất các dữ liệu hồ sơ địa chính.

- Là công cụ hiệu quả cho việc cung cấp thông tin đất đai cho thị trường quyền sử dụng đất, thị trường bất động sản.

- Đáp ứng nhu cầu về thông tin đất đai của người và các nhu cầu chung về phát triển xã hội và nâng cao dân trí.

3.2. Thiết kế hệ thống WebGIS cho bài toán phân bổ và định giá đất

a. Các bước của quá trình xây dựng ứng dụng:



Hình 3.3: Các bước của quá trình xây dựng ứng dụng

- Bước 1: chạy thuật toán di truyền định hướng không gian

Bước này cho phép đọc dữ liệu đầu vào của thuật toán.

- Bước 2: xây dựng phương án gồm các giá đất ước tính

Thực hiện chuẩn hóa dữ liệu và dựa vào dữ liệu để xây dựng hàm số phù hợp với bài toán.

Qua các quá trình của thuật toán, xây dựng được bộ tham số phù hợp với fitness nhỏ nhất.

- Bước 3: Sử dụng FCM để phân chia giá đất thành các vùng từ thấp đến cao.

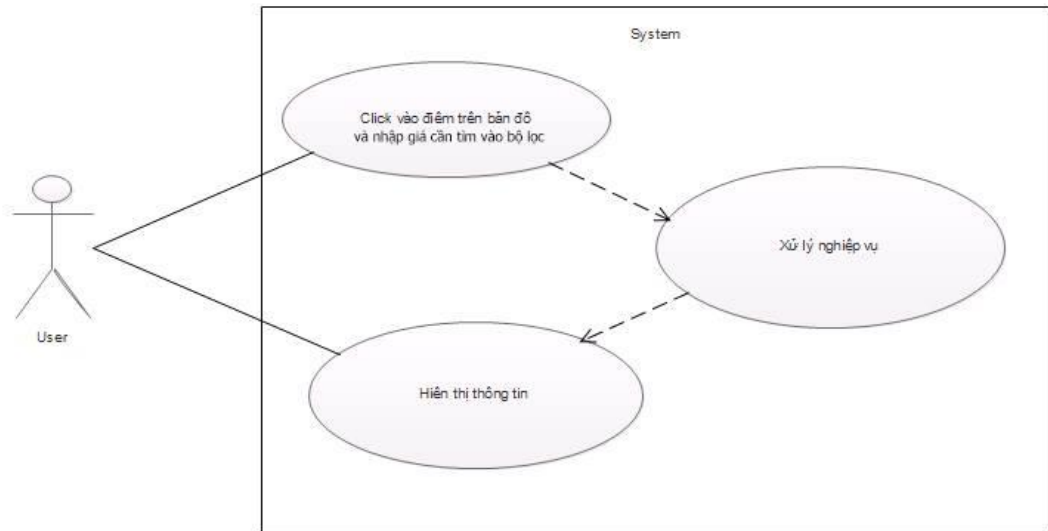
Thực hiện phân chia để có các vùng giá đất khác nhau

- Bước 4: Ảnh xạ và hiển thị kết quả các nhóm trên hệ thống thông tin đất đai LIS.

Hiện thị kết quả lên bản đồ gồm các điểm và các vùng giá đất.

b. Thiết kế hệ thống

Hệ thống cho phép người dùng ứng dụng xem chi tiết kết quả định giá đất.



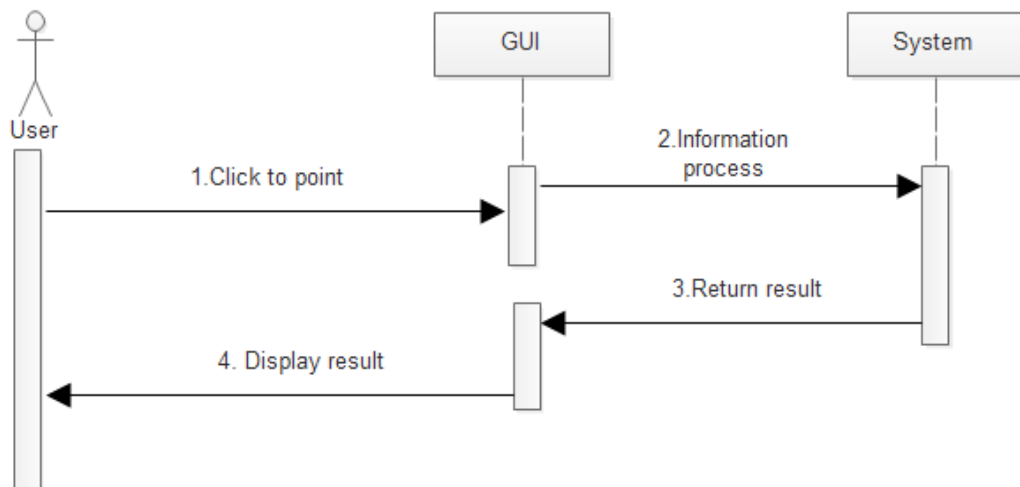
Hình 3.4: Sơ đồ hệ thống ứng dụng

c. Chức năng của ứng dụng

Có các chức năng sau:

- Hiện thị bản đồ (có full screen, zoom in, zoom out).
- Chú thích.
- Bộ lọc theo cụm.
- Bảng so sánh giá dự báo, MSE, giá chuẩn hóa GA với SGA.
- Tác nhân: người dùng.
- Đầu ra: thông tin về phân bố và định giá đất.
- Mô tả chi tiết:
 - + Người dùng click hay nhập giá vào điểm trên bản đồ hay bộ lọc.
 - + Hệ thống xử lý nghiệp vụ và hiển thị thông tin lên trình duyệt.

- Biểu đồ trình tự:



Hình 3.5: Biểu đồ tuần tự của ứng dụng

d. Xây dựng cơ sở dữ liệu

- Dữ liệu thuộc tính

Dữ liệu được thu thập tại một số quận trong Hà Nội (dữ liệu được cung cấp bởi Bộ Xây Dựng thông qua công ty Tư vấn Geo Việt) gồm rất nhiều các trường thuộc tính liên quan trực tiếp đến giá đất. Như đã mô tả ở chương 1 thì có rất nhiều các yếu tố ảnh hưởng trực tiếp tới giá đất, trong phần đặc tả dữ liệu này sẽ đề cập đến những thuộc tính mà ta sử dụng để định giá đất trong ứng dụng này được mô tả ở chương 2.

Từ dữ liệu thu thập được ta chia làm 3 phần, trong đó 2 phần dùng để training và 1 phần dùng để testing. Cụ thể tập training có 627 điểm, testing có 269 điểm.

Trước tiên, từ dữ liệu thu thập sẽ được chuẩn hóa về đoạn $[0,1]$ cho hợp lý trong việc xử thuật toán, sau đó ta xây dựng được cơ sở dữ liệu thuộc tính của bài toán.

	id [PK]	g_id integer	district double	headage double prec	totnumr double prec	occupn double	edulevel double prec	houst double	sepgstrm double prec	sepedrm double prec	gfa double	plotarea double prec	bthyear double prec	frontage double prec	centdis double p	electri double	plumbing double pi	wastcoll double p	securit double	schooql double	storeys double pi	hougrad double pi	hpricpsm double prec
1	1	534	0.8	0.5365853	0.2142857	0.4	0.5	0	0.0909090	0.2222222	0.085	0.025062	0.996522	0.0816326	0.16666	0	1	1	1	1	0.33333	0	0.006809
2	2	127	0.1333	0.5853658	0.6428571	1	0.5	0	0	0.2222222	0.077	0.013157	0.992548	0.0612244	0.11111	0.5	0.66666	1	1	0.5	0.66666	0	0.031242
3	3	446	0.2666	0.5365853	0.2142857	0.4	0.8333333	0	0.0909090	0.4444444	0.063	0.018170	0.989567	0.0469387	0.08333	0.5	0.66666	1	1	0.5	0.33333	0	0.022719
4	4	444	0	0.4390243	0.2142857	1	0.3333333	0	0	0.1111111	0.047	0.019423	0	0.0612244	0.08333	0.5	0.66666	1	1	0	0.5	0.011991	
5	5	544	0.8	0.4634146	0.4285714	0	0.3333333	0	0.0909090	0.3333333	0.157	0.274436	0.999503	0.4897959	0.33333	0	0.33333	1	1	0	0.33333	0	0.028116
6	6	147	0.5333	0.3048780	0.2142857	0	1	0	0	0	0.012	0.006892	0.998509	0.1020408	0.16666	0	0.33333	1	1	0	0	0.049992	
7	7	283	0.4666	0.5731707	0.2142857	0	0.8333333	0	0.0909090	0.3333333	0.132	0.025689	0.996522	0.0546938	0.13888	0.5	1	0.5	1	0	0.66666	0	0.011102
8	8	93	0.4	0.5487804	0.3571428	0.6	0.5	1	0	0	0.037	0.022556	0.932935	0.0408163	0.16666	0	0.33333	0.5	1	0	0	0	0.006241
9	9	879	0.1333	0.3292682	0.2142857	0	0.6666666	0	0.0909090	0	0.117	0.019423	0	0.0408163	0.08333	0.5	0.66666	0	1	0.5	0.66666	0	0.016658
10	10	282	0.4666	0.5853658	0.1428571	0	0.5	0	0.0909090	0.3333333	0.198	0.051378	0.994535	0.1224489	0.11111	0	0.66666	0	1	0.5	0.66666	0	0.014991
11	11	840	0.1333	0.2439024	0.2857142	1	0.5	0	0.0909090	0.3333333	0.085	0.016290	0	0.0510204	0.11111	0	0.66666	0	1	0.5	1	0	0.008514
12	12	965	0.2666	0.1951219	0	1	0.8333333	0	0.0909090	0.2222222	0.077	0.022556	0.999503	0.0408163	0.11111	0	0.66666	0	1	0.5	0.33333	0	0.124992
13	13	29	0.4666	0.2317073	0.0714285	0	0.8333333	0	0.0909090	0.1111111	0.042	0.016290	0	0.0408163	0.11111	0	0.66666	0	1	0	0	0.5	0.016658
14	14	309	1	0.5121951	0.2857142	0	0.5	0	0.0909090	0	0.208	0.091478	0	0.0816326	0.05555	0	0.33333	0	1	0	0.66666	0	0.019039
15	15	632	0.1333	0.6951219	0.3571428	0.6	0.5	0	0.0909090	0.4444444	0.141	0.020050	0.992051	0.0530612	0.08333	0	0.33333	0	1	0	1	0	0.012144
16	16	80	0.4666	0.5121951	0.2142857	0	0.1666666	0	0.0909090	0.6666666	0.198	0.045739	0.998509	0.2244897	0.16666	0	0.33333	0	1	0	0.66666	0	0.007491
17	17	307	1	0.5487804	0.2142857	0	0.6666666	0	0.0909090	0	0.198	0.028822	0.996522	0.0102040	0.16666	0	0.33333	0	1	0	1	0	0.007491
18	18	142	0.5333	0.2682926	0.3571428	1	0.5	0	0.0909090	0.4444444	0.228	0.436090	0.996522	0.0306122	0.05555	0	0.33333	0.5	1	0	0.33333	0	0.021731
19	19	565	0.8	0.4146341	0.2857142	1	0.5	0	0.0909090	0.3333333	0.178	0.038220	0.990064	0.0816326	0.08333	1	1	1	0.5	0.5	0.33333	0	0.003602
20	20	566	0.8	0.7804878	0.5714285	0.6	0.6666666	0	0.0909090	0.2222222	0.097	0.285714	0.992548	0.1836734	0.11111	1	1	1	0.5	0.5	0.33333	0	0.005991
21	21	76	0.4666	0.3536585	0.2142857	1	0.3333333	0	0.0909090	0.1111111	0.298	0.035087	0	0.1224489	0.33333	1	1	1	0.5	1	0	0.5	0.000825

Hình 3.6: Một phần của cơ sở dữ liệu thuộc tính

- Dữ liệu không gian

Từ dữ liệu shapefile thu thập được, ta xây dựng cơ sở dữ liệu không gian cho bài toán. Dữ liệu không gian gồm có bản đồ nền và lớp dữ liệu điểm.

Bản đồ nền: Là bản đồ một số quận huyện ở Hà Nội.

- Lớp dữ liệu điểm: Mô tả vị trí được định giá đất, mỗi điểm được gán một mã số duy nhất.

Sau đây là bảng cơ sở dữ liệu không gian:

	gid [PK] serial	code character varying(3)	name character varying(80)	geom geometry(MultiPolygon,4326)
1	1	020	Thanh Tri	
2	2	268	Ha Dong	
3	3	018	Gia Lam	
4	4	008	Hoang Mai	
5	5	007	Hai Ba Trung	
6	6	006	Dong Da	
7	7	004	Long Bien	
8	8	005	Cau Giay	
9	9	001	Ba Dinh	
10	10	002	Hoan Kiem	
11	11	003	Tay Ho	
12	12	017	Dong Anh	
13	13	009	Thanh Xuan	
14	14	019	Nam Tu Liem	
15	15	021	Bac Tu Liem	
*				

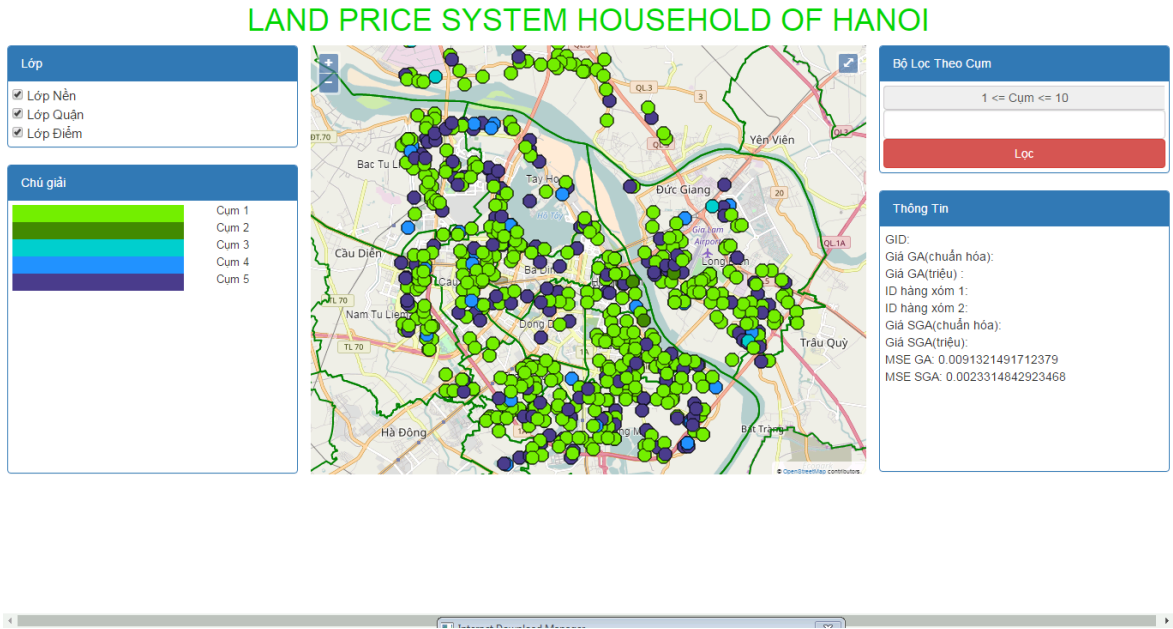
Hình 3.7: Bảng cơ sở dữ liệu không gian về một số quận

	gid [PK] serial	id_survey smallint	longitude numeric	latitude character varying(50)	geom geometry(Point,4326)
1	1	2	105.79731900000	20.988125	0101000020E61000001D9F6E4507735A40A7DE86BDF5FC3440
2	2	3	105.85038500000	21.000492	0101000020E61000001AEEC7B36C765A409B2CD73620003540
3	3	4	105.85320500000	21.005031	0101000020E6100000886E06EA9A765A40CCF434BB49013540
4	4	5	105.85044700000	21.004637	0101000020E61000004B6E4DBA6D765A409ACF15DF2F013540
5	5	6	105.84153700000	20.997613	0101000020E610000022B2D4BCDB755A40D3E9499163FF3440
6	6	7	105.85334500000	21.000126	0101000020E61000007A0379349D765A405678A04608003540
7	7	8	105.85042400000	21.001023	0101000020E61000003DF605576D765A402F614E1243003540
8	8	9	105.84422200000	20.991855	0101000020E6100000520A0CBC07765A40ECB96634EAFD3440
9	9	10	105.83606700000	20.997613	0101000020E610000061D1521E82755A40A18F769263FF3440
10	10	11	105.84019500000	20.995008	0101000020E6100000D4D03FC1C5755A40CA43C8D4B8FE3440
11	11	12	105.84410900000	21.001744	0101000020E6100000AB1409E005765A402AD0F74B72003540
12	12	13	105.84585700000	20.998201	0101000020E610000001CAADB322765A40FA5FA71EBAFF3440
13	13	14	105.84639700000	20.99743	0101000020E6100000475B295F2B765A403C09449957FF3440
14	14	15	105.82125100000	21.001273	0101000020E61000004F94185F8F745A404BDD126F53003540
15	15	16	105.85212200000	21.006576	0101000020E6100000C7D5432C89765A40F37D20F5AE013540
16	16	17	105.84986900000	21.0051	0101000020E61000006791B4064765A40EDAC94334E013540
17	17	18	105.84445700000	20.993459	0101000020E610000010F9B1930B765A40DA69E85553FE3440
18	18	19	105.86764600000	20.998257	0101000020E6100000F43A138187775A40039802CA8DFF3440
19	19	20	105.84604600000	20.998814	0101000020E6100000C77A489D25765A406FC93846B2FF3440
20	20	21	105.83364000000	20.992229	0101000020E61000003B69BF5A5A755A40DB229CBB02FE3440
21	21	22	105.82905900000	20.998265	0101000020E6100000851DA44C0F755A409939EB538EFF3440
22	22	23	105.83000600000	20.994395	0101000020E61000009A3F3AD01E755A40C8CBBCB090FE3440
23	23	24	105.84310100000	20.986951	0101000020E6100000BFC4055EF5755A4020E8E4CAA8FC3440
24	24	25	105.84677800000	20.986504	0101000020E6100000C9140F9C31765A40EE75118DBBFC3440
25	25	26	105.84498300000	20.99027	0101000020E610000088D7AA3414765A407C79234E82FD3440
26	26	27	105.81329000000	20.997873	0101000020E610000044242AF10C745A40FA2B7E9574FF3440
27	27	28	105.80870300000	20.998576	0101000020E6100000326FEDC8C1735A4006923EADA2FF3440
28	28	29	105.83470900000	20.982668	0101000020E6100000BB904FE16B755A4004FAC92090FB3440
29	29	30	105.83075200000	20.98406	0101000020E61000006921B0B2B755A406E585D5EEBFB3440
30	30	31	105.82581600000	21.000926	0101000020E610000016C1EE2ADA745A404DB7B3AF3C003540

Hình 3.8: Một phần cơ sở dữ liệu về các địa điểm khảo sát giá đất

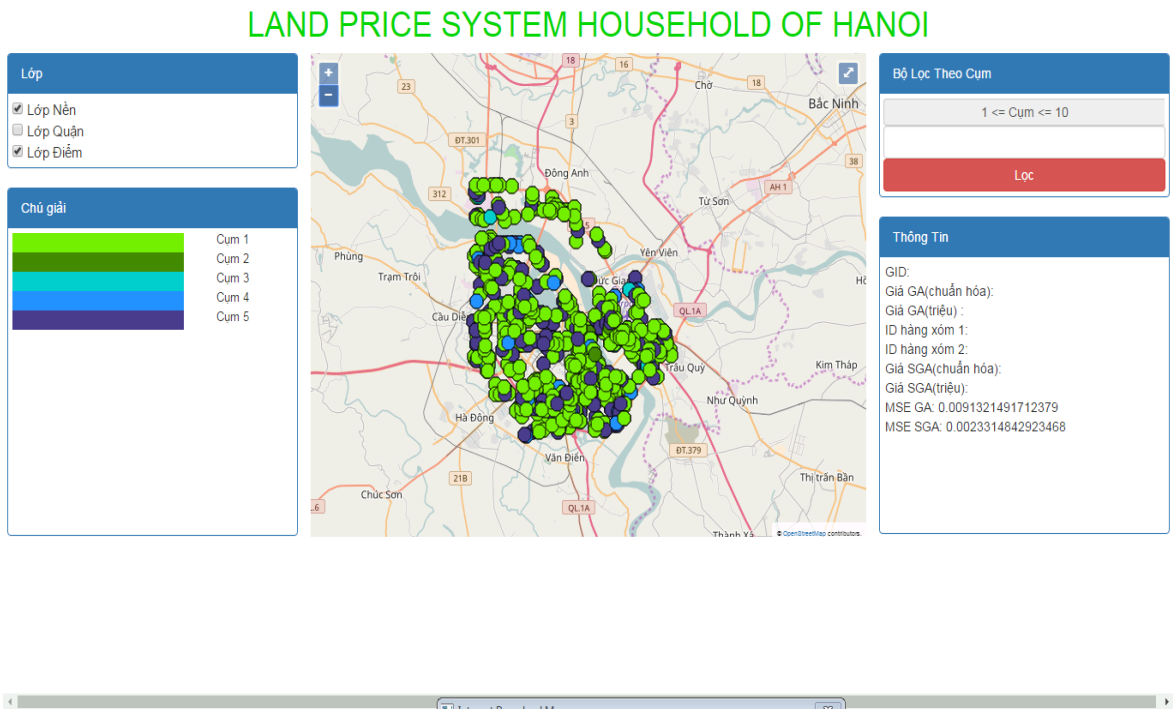
3.2.1. Giao diện chương trình

a. Giao diện ban đầu của hệ thống



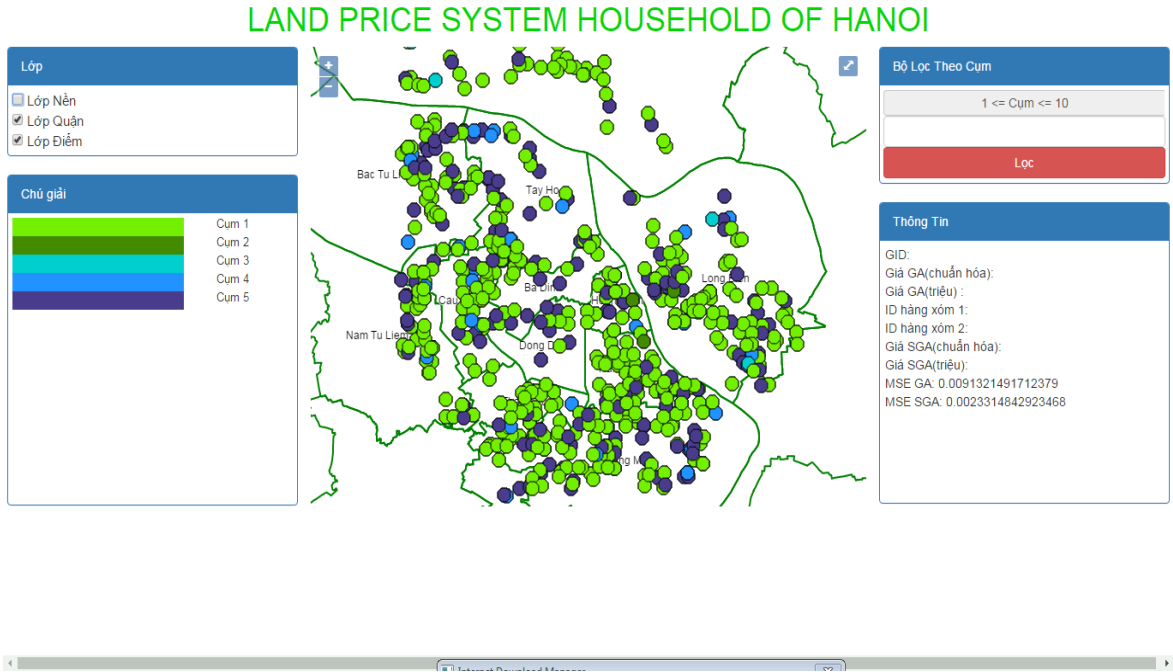
Hình 3.9: Giao diện ban đầu của hệ thống

b. Giao diện tắt lớp quận



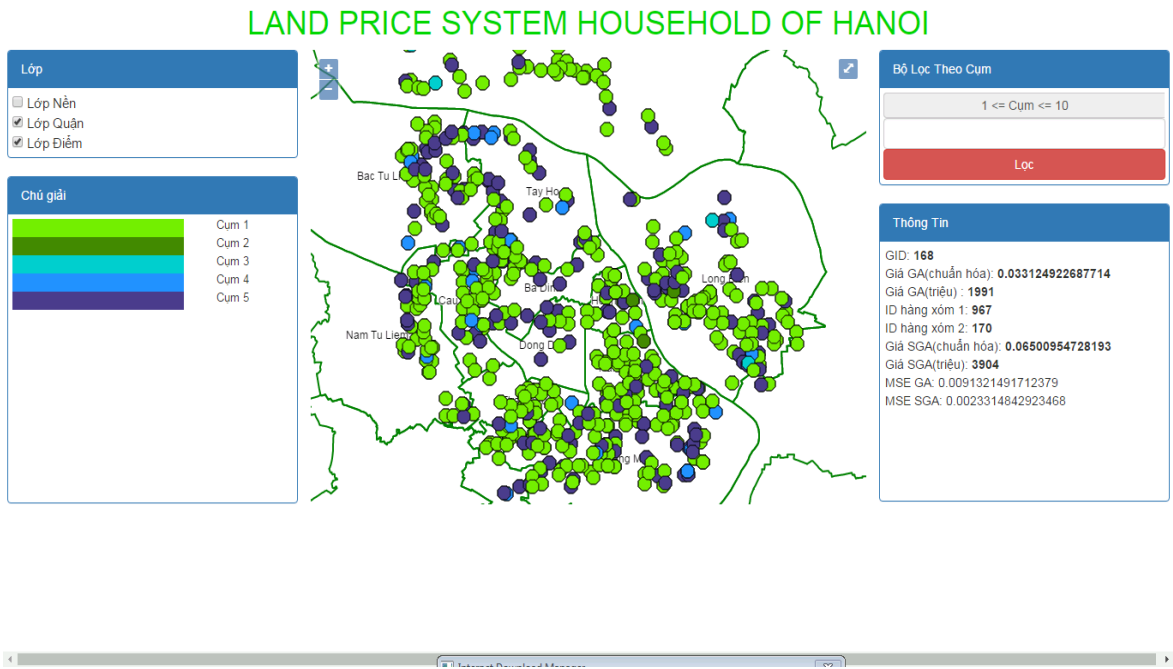
Hình 3.10: Giao diện bật tắt Layer

Mô tả: CheckBox dùng để bật và tắt layer với dữ lớp quận
c. Giao diện khi tắt Layer với dữ liệu lớp nền



Hình 3.11: Giao diện khi tắt các Layers

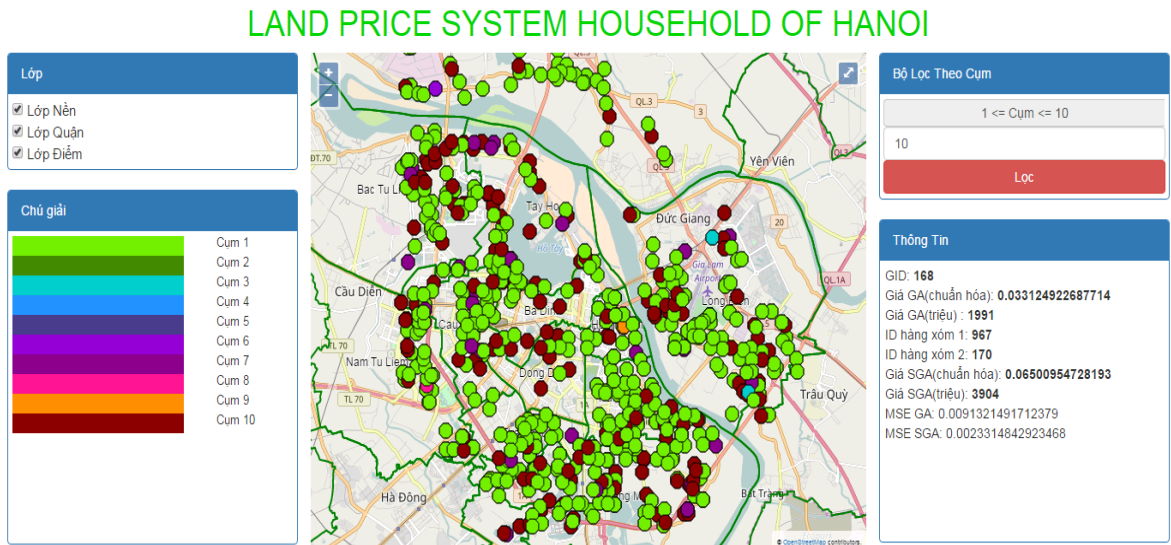
d. Giao diện khi click vào điểm trong trên bản đồ



Hình 3.12: Giao diện khi click vào điểm trên bản đồ

Mô tả:

- ID: là mã số của điểm dữ liệu trên bản đồ
 - Khi click vào điểm trên bản đồ sẽ hiển thị đầy đủ giá dự đoán GA và giá dự đoán SGA
- e. Giao diện khi nhập số liệu vào bộ lọc



Hình 3.13: Giao diện khi nhập giá trị vào bộ lọc

3.2.2. Thử nghiệm chương trình và đánh giá kết quả

- Công thức tính bình phương sai số trung bình:

$$MSE = \frac{\sum_{i=1}^N (D_i - F_i)^2}{N}$$

với + D_i là giá trị thực

+ F_i là giá trị tính theo hà mục tiêu

id	g_id	district	headage	totnumr	occupn	edulevel	houstype	sepgstrm	sepbedm	gfa	plotarea	btyear	frontage	centdis	electric	plumbing	wastcoll	security	schooql
1	771	0,003717	0,0037125	0,0037171	0,003716	0,003712	0,00372	0,003717	0,003717	0,003717	0,003717	0,0037	0,0037175	0,003717	0,0037	0,003717	0,00372	0,0037	0,0037
2	904	0,003711	0,0037133	0,0037171	0,003716	0,003712	0,00372	0,003717	0,003717	0,003717	0,003717	0,0037	0,0037175	0,003717	0,0037	0,003717	0,00372	0,0037	0,0037
3	874	0,003717	0,0037113	0,0037163	0,003716	0,003716	0,00372	0,003717	0,003717	0,003717	0,003717	0,0037	0,0037175	0,003717	0,0037	0,003717	0,00372	0,0037	0,0037
4	685	0,003717	0,0037133	0,0037158	0,003717	0,003712	0,00372	0,003717	0,003717	0,003717	0,003717	0,0037	0,0037175	0,003717	0,0037	0,003717	0,00372	0,0037	0,0037
5	901	0,003717	0,0037154	0,0037163	0,003717	0,003712	0,00372	0,003717	0,003717	0,003717	0,003717	0,0037	0,0037175	0,003717	0,0037	0,003717	0,00372	0,0037	0,0037
6	592	0,003716	0,0037157	0,0037171	0,003717	0,003712	0,00372	0,003717	0,003717	0,003717	0,003717	0,0037	0,0037175	0,003717	0,0037	0,003717	0,00372	0,0037	0,0037
7	355	0,003717	0,003716	0,0037171	0,003717	0,003712	0,00372	0,003717	0,003717	0,003717	0,003717	0,00372	0,0037174	0,003717	0,0037	0,003717	0,00372	0,0037	0,0037
8	790	0,003715	0,0037162	0,0037163	0,003717	0,003712	0,00372	0,003717	0,003717	0,003717	0,003717	0,0037	0,0037175	0,003717	0,0037	0,003717	0,00372	0,0037	0,0037
9	119	0,003716	0,0037164	0,0037152	0,003717	0,003712	0,00372	0,003717	0,003717	0,003717	0,003717	0,00372	0,0037175	0,003717	0,0037	0,003717	0,00372	0,0037	0,0037
10	197	0,003716	0,0037151	0,0037163	0,003717	0,003716	0,00372	0,003717	0,003717	0,003717	0,003717	0,00372	0,0037175	0,003717	0,0037	0,003717	0,00372	0,0037	0,0037
11	413	0,003717	0,0037145	0,0037163	0,003717	0,003712	0,00372	0,003717	0,003717	0,003717	0,003717	0,0037	0,0037175	0,003717	0,0037	0,003717	0,00372	0,0037	0,0037
12	875	0,003716	0,0037155	0,0037171	0,003717	0,003712	0,00372	0,003717	0,003717	0,003717	0,003717	0,0037	0,0037174	0,003717	0,0037	0,003717	0,00372	0,0037	0,0037
13	472	0,003717	0,0037161	0,0037167	0,003717	0,003712	0,00372	0,003717	0,003717	0,003717	0,003717	0,00372	0,0037174	0,003717	0,0037	0,003717	0,00372	0,0037	0,0037
14	405	0,003717	0,0037141	0,0037171	0,003717	0,003714	0,00372	0,003717	0,003717	0,003717	0,003717	0,0037	0,0037175	0,003717	0,0037	0,003717	0,00372	0,0037	0,0037
15	535	0,003714	0,0037125	0,0037163	0,003715	0,003717	0,00372	0,003717	0,003717	0,003717	0,003717	0,0037	0,0037147	0,003717	0,0037	0,003717	0,00372	0,0037	0,0037
16	178	0,003716	0,0037137	0,0037171	0,003715	0,003717	0,00372	0,003717	0,003717	0,003717	0,003717	0,0037	0,0037174	0,003717	0,0037	0,003717	0,00372	0,0037	0,0037
17	377	0,003717	0,0037161	0,0037171	0,003717	0,003712	0,00372	0,003717	0,003717	0,003717	0,003717	0,00372	0,0037175	0,003717	0,0037	0,003717	0,00372	0,0037	0,0037
18	935	0,003717	0,0037167	0,0037163	0,003717	0,003712	0,00372	0,003717	0,003717	0,003717	0,003717	0,0037	0,0037175	0,003717	0,0037	0,003717	0,00372	0,0037	0,0037
19	457	0,003715	0,003717	0,0037171	0,003717	0,003712	0,00372	0,003717	0,003717	0,003717	0,003717	0,0037	0,0037174	0,003717	0,0037	0,003717	0,00372	0,0037	0,0037
20	467	0,003715	0,0037163	0,0037163	0,003717	0,003712	0,00372	0,003717	0,003717	0,003717	0,003717	0,0037	0,0037175	0,003717	0,0037	0,003717	0,00372	0,0037	0,0037
21	798	0,003716	0,0037145	0,0037171	0,003717	0,003716	0,00372	0,003717	0,003717	0,003717	0,003717	0,0037	0,0037175	0,003717	0,0037	0,003717	0,00372	0,0037	0,0037

Hình 3.14: Bảng fitness

id	g_id	district	headage	totnumr	occupn	edulevel	houstype	sepgstrm	sepbedm	gfa	plotarea	btyear	frontage	centdis	ele
1	771	0,003716783	0,003712531	0,003717051	0,003716432	0,003712249	0,003717472	0,003717468	0,003717463	0,003717374	0,003717469	0,003701266	0,003717454	0,003717472	0,
2	904	0,003711267	0,003713346	0,003717051	0,003716432	0,003712249	0,003717472	0,003717468	0,003717463	0,003717374	0,003717469	0,003701168	0,003717463	0,003717472	0,
3	874	0,003716783	0,003711266	0,003716304	0,003716432	0,003715592	0,003717472	0,003717468	0,003717463	0,003717192	0,003717462	0,003701266	0,003717451	0,003717472	0,
4	685	0,003716783	0,003715743	0,003716631	0,003715393	0,003713503	0,003717472	0,003717472	0,003717469	0,003717372	0,00371747	0,003712027	0,003717451	0,003717472	0,
5	901	0,003717031	0,003715351	0,003716304	0,00371701	0,003712249	0,003717472	0,003717468	0,003717463	0,003716616	0,003717464	0,003701103	0,003717463	0,003717472	0,
6	592	0,003716121	0,003714325	0,003717051	0,003717318	0,003712249	0,003717472	0,003717468	0,003717469	0,003717359	0,003717468	0,003701195	0,003717455	0,003717472	0,
7	355	0,003717224	0,00371675	0,003717051	0,003715393	0,003712249	0,003717472	0,003717468	0,003717465	0,003717179	0,003717466	0,003706635	0,003717445	0,003717472	0,
8	790	0,003715238	0,003716237	0,003716304	0,00371701	0,003712249	0,003717472	0,003717468	0,003717463	0,003717153	0,003717466	0,003701152	0,003717451	0,003717472	0,
9	119	0,003716121	0,003716401	0,003715182	0,00371701	0,003712249	0,003717472	0,003717468	0,003717463	0,003717256	0,003717468	0,003717472	0,00371746	0,003717472	0,
10	197	0,003716479	0,003714141	0,003716802	0,003717318	0,003716288	0,003717472	0,003717468	0,003717469	0,003717351	0,003717467	0,003717472	0,003717459	0,003717472	0,
11	413	0,003716783	0,003714497	0,003716304	0,003717472	0,003712249	0,003717472	0,003717468	0,003717463	0,003716681	0,003717457	0,003701201	0,003717454	0,003717472	0,
12	875	0,003716479	0,003715461	0,003717051	0,003717472	0,003712249	0,003717472	0,003717468	0,003717463	0,003716969	0,003717457	0,00370112	0,003717413	0,003717472	0,
13	472	0,003716783	0,00371606	0,003716724	0,003717472	0,003712249	0,003717472	0,003717472	0,003717463	0,003716804	0,003717451	0,003717472	0,003717434	0,003717472	0,
14	405	0,003716783	0,003711816	0,003716833	0,003716779	0,003712876	0,003717472	0,003717469	0,003717469	0,003717179	0,003717463	0,003701298	0,00371745	0,003717472	0,
15	535	0,003713501	0,003714693	0,003716584	0,003716509	0,003716636	0,003717472	0,003717468	0,003717469	0,003717286	0,003717242	0,003701282	0,003716511	0,003717472	0,
16	178	0,003715707	0,003713651	0,003717051	0,003714584	0,003716636	0,003717472	0,003717468	0,003717463	0,003717414	0,003717289	0,003701233	0,003717356	0,003717472	0,
17	377	0,003717224	0,00371606	0,003717051	0,00371701	0,003712249	0,003717472	0,003717472	0,003717463	0,003717111	0,003717465	0,003717472	0,003717463	0,003717472	0,
18	935	0,003716948	0,003715723	0,003716584	0,003715393	0,003715174	0,003717472	0,003717469	0,003717467	0,003717095	0,003717342	0,00370113	0,003717426	0,003717472	0,
19	457	0,003715238	0,003716996	0,003717051	0,00371701	0,003712249	0,003717472	0,003717468	0,003717463	0,003716833	0,00371743	0,003701201	0,003717356	0,003717472	0,
20	467	0,003715238	0,00371632	0,003716304	0,003717472	0,003712249	0,003717472	0,003717472	0,003717463	0,003717236	0,003717468	0,003701152	0,003717464	0,003717472	0,
21	798	0,003715707	0,003714497	0,003717051	0,003717472	0,003715592	0,003717472	0,003717468	0,003717463	0,003716833	0,003717371	0,003701314	0,003717451	0,003717472	0,

Hình 3.15: Giá hai hàng xóm

Kết quả :

Tập testing:

	GA	SGA
MSE	0.010	0.006
Thời gian chạy (s)	0.508	169.754

Tập training:

	GA	SGA
MSE	0.009	0.002
Thời gian chạy (s)	0.943	1097.057

3.3. So sánh ưu nhược điểm của các thuật toán

Từ bảng trên ta thấy tốc độ xử lý dữ liệu của thuật toán SGA sẽ luôn chậm hơn thuật toán GA bởi vì thuật toán SGA dựa trên tinh toán di truyền của GA khi ra được giá dự đoán của GA thì SGA mới tiếp tục tìm 2 hàng xóm gần nhất tiếp theo tính toán fitness của nó dựa vào 2 hàng xóm rồi mới đưa ra giá dự đoán.

Nhưng về độ chính xác thì thuật toán SGA chính xác hơn do thuật toán SGA có sai số bình phương MSE nhỏ hơn GA.

Sau khi sử dụng thuật toán di truyền định hướng không gian cho bài toán phân bố và định giá đất ta nhận thấy

- + Thời gian để tính toán và đưa ra được giá dự báo mất rất nhiều thời gian.
- + Thuật toán này chỉ đưa ra được giá dự đoán tương đối (là công cụ hỗ trợ) không thể đưa ra được giá chính xác vì số liệu còn hạn chế và để định giá

thì không chỉ phụ thuộc vào yếu tố cố định mà còn nhiều yếu tố không cố định có thể thay đổi hàng năm ví dụ như diện tích mặt tiếp xúc, diện tích v.v

3.4. Tổng kết chương

Chương 3 mô tả tổng thể hệ thống WebGIS và hệ thống xử lí của ứng dụng định giá đất dựa hệ thống thông tin địa lí. Một số hình ảnh minh họa rõ ràng các hoạt động của chương trình và đưa ra kết quả chạy thử nghiệm.

KẾT LUẬN

Với rất nhiều ý nghĩa trong thực tế, vấn đề định giá đất được sự quan tâm của rất nhiều nhà kinh tế học, bởi vì vấn đề định giá đất rất phức tạp, nó phụ thuộc vào rất nhiều yếu tố: kinh tế, xã hội, vị trí, điều kiện phát triển, v.v. Mặt khác các yếu tố luôn biến động theo thời gian nên cần sự phân tích rất kỹ lưỡng mới định được giá.

Xuất phát từ thực tiễn đó, cùng với sự quan tâm, hứng thú tới vấn đề giá đất, giá đất, luận văn đã nghiên cứu và trình bày một số vấn đề chính như sau:

- Tìm hiểu tổng quan về giá đất và các yếu tố ảnh hưởng trực tiếp tới giá đất và các phương pháp định giá đất. Từ đó đặt ra bài toán phân bổ và định giá đất với phương pháp định giá đất hợp lý với sự biến động của các yếu tố ảnh hưởng trực tiếp tới giá đất.

- Tìm hiểu về thuật toán di truyền định hướng không gian và áp dụng thuật toán vào trong bài toán phân bổ và định giá đất. Thiết kế thuật toán một cách hợp lý để phù hợp với bài toán.

- Xây dựng hệ thống định giá đất trên nền WebGIS để cho người dùng dễ sử dụng và trực quan hơn.

Dựa vào những kết quả đã đạt được, trong tương lai, đề tài có thể phát triển theo hướng sau:

- Tiếp tục cải tiến để phù hợp với sự biến động của các yếu tố ảnh hưởng đến giá đất, nâng cao độ chính xác hơn cho bài toán.

- Mở rộng vùng định giá đất bởi vì luận văn mới chỉ nghiên cứu trên một số quận ở Hà Nội.

- Tích thêm các thuộc tính không gian để tính toán sao cho giá trị MSE nhỏ nhất

TÀI LIỆU THAM KHẢO

I. Tài liệu tiếng Việt

- [1]. Lê Hoàng Sơn, Nguyễn Thọ Thông. (2015). *Lập trình ứng dụng webgis*. Hà Nội: NXB giáo dục Việt Nam.
- [2]. Hồ Thị Lam Trà. (ngày 15 tháng 12 năm 2005). *Giáo trình định giá đất*.
- [3]. Trần Kim Hương, Nguyễn Thị Ngọc Chi. *Giải thuật di truyền (GAs) và ứng dụng*.
- [4]. Nguyễn Thị Thúy Hoài. *Giải thuật di truyền và ứng dụng*. Tuyển tập báo cáo “ hội nghị sinh viên nghiên cứu khoa học” lần thứ 6. Đại học Đà Nẵng - 2008.

II. Tài liệu tiếng Anh

- [5]. Bezdek, J.C., R. Ehrlich, et al. (1984), "FCM: the fuzzy c-means clustering algorithm", *Computers and Geosciences*, 10, pp.191-203.
- [6]. Dawkins, R. (2006). *The Selfish Gene: 30th Anniversary Edition*. Oxford University Press: 21(4),pp.211-217.
- [7]. Bagley, J. D. (1967). The Behavior of Adaptive Systems Which Employ Genetic and Correlative Algorithms.
- [8].Charbonneau, P. . (1988). *An introduction to genetic algorithms for numerical optimization*.
- [9].Demetriou, D. (2013). *The development of an integrated planning and decision support system (IPDSS) for land consolidation*. Springer Science & Business Media.
- [10]. Goldberg, D. E. (1988). *Genetic algorithms and machine learning*. *Machine learning*. 3(2), 95-99.
- [11] Hartigan . J. A. and . Wong.M. A. Algorithm AS 136: A K-Means Clustering Algorithm. *Vol. 28, No. 1 (1979), pp. 100-108*, ournal of the Royal Statistical Society.

- [12]. Jaco F. Schutte EGM 6365. *The Particle Swarm Optimization Algorithm - Structural Optimization Fall 2005*.
- [13]. Bäck, T., Hoffmeister, F. and Schwefel. (H. P. (1991). A survey of evolution strategies). *Proceedings of the 4th International Conference on Genetic Algorithms*. Morgan Kaufmann, San Mateo, CA.
- [14]. Booker, L. ((1987). Improving search in genetic algorithms). *In Genetic Algorithms and Simulating Annealing*. ed. L. Davis, pp. 61–73: Morgan Kaufmann, San Mateo, CA.
- [15]. Davis, LD. (1991). *Handbook of Genetic Algorithms*. Van Nostrand Reinhold, New York.
- [16]. Eshelman, L. (1991). *The CHC adaptive search algorithm In Foundations of Genetic Algorithms*. ed. G. Rawlins, pp. 256–283: Morgan Kaufmann, San Mateo, CA.
- [17]. Freisleben, P. Mers . (1996). New genetic local search operator for the travelling . University of Siegen.
- [18]. Goldberg, D. (1987). *Simple genetic algorithms and the minimal, deceptive problem*. *In Genetic Algorithms and Simulated Annealing*. ed. L. Davis. Pitman, London.
- [19]S.V. Sivanandam, S.N. Deepa. (2008). Introduction to Genetic Algorithms, Spinger.
- [20]. Wright, A. (1991). *Genetic algorithms for real parameter optimization*. *In Foundations of Genetic Algorithms*. ed. G. Rawlins. Morgan Kaufmann, San Mateo, CA.