

BỘ GIÁO DỤC VÀ ĐÀO TẠO
ĐẠI HỌC ĐÀ NẴNG

HỒ PHƯỚC DUY

ỨNG DỤNG HỆ TRỢ GIÚP QUYẾT ĐỊNH
PHỤC VỤ CÔNG TÁC ĐỀN BÙ GIẢI TOẢ ĐẤT
ĐAI TẠI THÀNH PHỐ ĐÀ NẴNG

Chuyên ngành: KHOA HỌC MÁY TÍNH
Mã số: 60.48.01

TÓM TẮT LUẬN VĂN THẠC SĨ KỸ THUẬT

Đà Nẵng - 2010

Công trình được hoàn thành tại
ĐẠI HỌC ĐÀ NẴNG

Người hướng dẫn khoa học: **PGS.TS. Phan Huy Khánh**

Phản biện 1: PGS.TS. Võ Trung Hùng
Phản biện 2: PGS.TS. Lê Mạnh Thạnh

Luận văn được bảo vệ tại Hội đồng chấm Luận văn tốt nghiệp Thạc sĩ kỹ thuật họp tại Đại học Đà Nẵng vào ngày 16 tháng 12 năm 2010.

** Có thể tìm hiểu Luận văn tại:*

- Trung tâm Thông tin - Học liệu, Đại học Đà Nẵng
- Trung tâm Học liệu, Đại học Đà Nẵng.

MỞ ĐẦU

1. Lý do chọn đề tài

Hòa cùng với sự phát triển chung của đất nước về kinh tế - xã hội, thành phố Đà Nẵng là một trong những thành phố trẻ, năng động và đầy sáng tạo trong công tác chính trang, quy hoạch đô thị. Vì vậy việc quy hoạch cơ sở hạ tầng nhằm đáp ứng về công tác chính trang đô thị là một vấn đề mà Lãnh đạo thành phố rất quan tâm. Một trong những công tác hết sức quan trọng và cực kỳ nhạy cảm trong công tác quy hoạch, chính trang đô thị đó là công tác đền bù giải tỏa.

Tuy nhiên, với một khu vực cần đền bù giải tỏa có diện tích lớn vì thế số lượng hồ sơ khi đền bù giải tỏa rất nhiều khiến công tác đền bù giải tỏa đôi khi giải quyết không kịp thời, quá tải dẫn đến thiếu sót và đặc biệt rất khó khăn trong việc đền bù và bố trí tái định cư. Bên cạnh đó, việc đưa ra quyết định xem xét đền bù giải tỏa cho nhiều hồ sơ cùng một thời điểm là vấn đề rất nan giải và vô cùng phức tạp để dẫn đến những quyết định chưa thật sự thỏa đáng.

Trong quy trình đền bù giải tỏa khâu quan trọng nhất và nhạy cảm nhất, đồng thời gây ảnh hưởng lớn nhất đối với quyền lợi của người dân đó chính là việc ra quyết định bố trí đất tái định cư cho các hộ trong khu vực bị giải tỏa. Trường hợp nào thì sẽ được bố trí đền bù Lô đất chính, Lô đất phụ, Chung cư hay kết hợp một trong những kết quả trên.

Vì vậy một trong những giải pháp hữu hiệu nhất nhằm khắc phục các vấn đề nêu trên là tiến hành triển khai xây dựng một hệ thống trợ giúp ra quyết định trong công tác đền bù giải tỏa. Đó là một hệ thống được thiết kế giúp cho lãnh đạo đơn vị, lãnh đạo thành phố nắm bắt được một cách tổng quát về tình hình đền bù nhằm đưa ra những quyết

định có tính công bằng và chính xác cao liên quan đến công tác đền bù giải tỏa tại đơn vị.

Chính vì những lý do nêu trên, tôi quyết định chọn đề tài: **“Ứng dụng hệ trợ giúp quyết định phục vụ công tác đền bù giải tỏa đất đai tại thành phố Đà Nẵng”** với mong muốn đóng góp thêm một giải pháp nhằm hỗ trợ cho lãnh đạo xem xét, giải quyết nhu cầu đền bù giải tỏa ngày càng phức tạp và cấp thiết của thành phố nói chung và của các đơn vị có chức năng đền bù giải tỏa nói riêng.

2. Mục đích nghiên cứu

Nhằm triển khai áp dụng có hiệu quả cho công tác hỗ trợ ra quyết định trong đền bù giải tỏa đất đai tại Ban Quản lý dự án công trình đường Bạch Đằng Đông, quận Sơn Trà, thành phố Đà Nẵng.

3. Đối tượng và phạm vi nghiên cứu

❖ Đối tượng nghiên cứu

Tìm hiểu công tác đền bù giải tỏa tại Ban Quản lý dự án công trình đường Bạch Đằng Đông để đề ra giải pháp nhằm đem lại hiệu quả cao hơn.

Phân tích quy trình, nghiên cứu kết quả trước đây tại đơn vị để đề ra phương thức, cách thức xây dựng và triển khai hệ thống.

❖ Phạm vi nghiên cứu

Ứng dụng thuật toán ID3 để xây dựng cây quyết định phục vụ công tác đền bù giải tỏa.

Ứng dụng, phục vụ cho lãnh đạo và Phòng chuyên môn trong công tác đền bù giải tỏa tại Ban Quản lý dự án công trình đường Bạch Đằng Đông.

4. Phương pháp nghiên cứu

❖ Phương pháp nghiên cứu lý thuyết

Nghiên cứu tài liệu, ngôn ngữ và các công nghệ có liên quan.

Tổng hợp, thu thập các tài liệu về công tác đền bù giải tỏa.

❖ **Phương pháp nghiên cứu thực nghiệm**

Phân tích yêu cầu thực tế của bài toán và áp dụng các thuật toán có liên quan để hỗ trợ việc lập trình, xây dựng ứng dụng.

Quan sát thực tế, thực nghiệm công tác đền bù giải tỏa.

5. Kết quả dự kiến

Phân tích, tìm hiểu được công tác đền bù giải tỏa.

Đề ra giải pháp và sử dụng cây quyết định trong việc xây dựng hệ thống hỗ trợ ra quyết định trong công tác đền bù giải tỏa.

6. Ý nghĩa khoa học và thực tiễn của đề tài

❖ **Về mặt lý thuyết**

Tìm hiểu quy trình, quy định, nguyên tắc của công tác đền bù giải tỏa trên địa bàn thành phố.

Đề xuất giải pháp triển khai ứng dụng Cây quyết định vào trong hệ thống đền bù giải tỏa.

❖ **Về mặt thực tiễn**

Sản phẩm sẽ là hệ thống phục vụ đắc lực, kịp thời và có độ chính xác cao cho các cán bộ lãnh đạo, cán bộ quản lý, các cán bộ làm công tác chuyên môn trong lĩnh vực đền bù giải tỏa.

7. Cấu trúc của luận văn

Nội dung chính của luận văn này được chia thành ba chương với nội dung như sau:

Chương 1: Nghiên cứu hệ trợ giúp quyết định và hệ thống hỗ trợ quyết định thông minh.

Chương 2: Phân tích dữ liệu, tính toán và triển khai ứng dụng cây quyết định vào bài toán đền bù giải tỏa tại đơn vị.

Chương 3: Tiến hành cài đặt, cho hoạt động thử nghiệm, nhận xét và đánh giá, hiển thị kết quả minh họa của chương trình.

CHƯƠNG 1

NGHIÊN CỨU CÁC HỆ TRỢ GIÚP QUYẾT ĐỊNH

1.1. HỆ TRỢ GIÚP QUYẾT ĐỊNH

1.1.1. Tổng quan về Hệ trợ giúp quyết định

1.1.2. Vai trò, chức năng của hệ trợ giúp quyết định

1.1.3. Cấu trúc của Hệ trợ giúp quyết định

1.1.3.1. Quản lý dữ liệu

1.1.3.2. Quản lý mô hình

1.1.3.3. Quản lý dựa trên kiến thức

1.1.3.4. Giao diện người dùng

1.1.4. Các loại hệ thống trợ giúp quyết định

1.1.4.1. Hệ trợ giúp quyết định nhóm

1.1.4.2. Hệ trợ giúp quyết định mức xí nghiệp

1.1.4.3. Hệ quản trị kiến thức

1.2. HỆ THỐNG HỖ TRỢ QUYẾT ĐỊNH THÔNG MINH

1.2.1. Tổng quan về Trí tuệ nhân tạo

1.2.2. Tri thức và các phương pháp suy diễn

1.2.2.1. Tri thức

1.2.2.2. Các dạng biểu diễn tri thức

1.2.2.3. Các phương pháp suy diễn

1.2.3. Cây quyết định

1.2.3.1. Tổng quan về cây quyết định

1.2.3.2. Các kiểu cây quyết định

1.2.3.3. Phân lớp dữ liệu bằng cây quyết định

Cây quyết định là một trong những hình thức mô tả dữ liệu trực quan nhất, dễ hiểu nhất đối với người dùng. Cấu trúc của một cây quyết định bao gồm các nút và các nhánh. Nút dưới cùng được gọi là nút lá, trong mô hình phân lớp dữ liệu chính là các giá trị của các nhãn lớp

(gọi tắt là nhãn). Các nút khác nút lá được gọi là các nút con, đây còn là các thuộc tính của tập dữ liệu, hiển nhiên các thuộc tính này phải khác thuộc tính phân lớp. Mỗi một nhánh của cây xuất phát từ một nút P nào đó ứng với một phép so sánh dựa trên miền giá trị của nút đó. Nút đầu tiên được gọi là nút gốc của cây.

1.2.3.4. Giải thuật huấn luyện cây quyết định cơ bản

Giải thuật quy nạp cây ID3 là gì ?

Giải thuật quy nạp cây ID3 là một giải thuật học đơn giản nhưng tỏ ra thành công trong nhiều lĩnh vực. ID3 là một giải thuật hay vì cách biểu diễn tri thức học được của nó, tiếp cận của nó trong việc quản lý tính phức tạp, xử lý dữ liệu nhiễu.

Giải thuật ID3 xây dựng cây quyết định:

Function Tree_ID3(tập_ví_dụ, tập_thuộc_tính)

begin

if mọi ví dụ trong tập_ví_dụ đều nằm trong cùng một lớp *then*

return một nút lá được gán nhãn bởi lớp đó

else if tập_thuộc_tính là rỗng *then*

return nút lá được gán nhãn bởi tuyến của tất cả các lớp trong tập_ví_dụ

else begin

chọn một thuộc tính P, lấy nó làm gốc cho cây hiện tại;

xóa P ra khỏi tập_thuộc_tính;

với mỗi giá trị V của P

begin

tạo một nhánh của cây gán nhãn V;

Đặt vào phân_vùngV các ví dụ trong tập_ví_dụ có giá trị V tại thuộc tính P;

Gọi Tree_ID3(phân_vùngV,
tập_thuộc_tính), gắn kết quả vào nhánh V
end

end

end

1.2.3.5. Thuộc tính phân loại tốt nhất

Entropy đo tính thuần nhất của tập huấn luyện

Khái niệm Entropy của một tập S được định nghĩa trong Lý thuyết thông tin là số lượng mong đợi các bit cần thiết để mã hóa thông tin về lớp của một thành viên rút ra một cách ngẫu nhiên từ tập S. Trong trường hợp tối ưu, mã có độ dài ngắn nhất. Theo lý thuyết thông tin, mã có độ dài tối ưu là mã gán $-\log_2 p$ bits cho thông điệp có xác suất là p. Trong trường hợp S là tập ví dụ, thì thành viên của S là một ví dụ, mỗi ví dụ thuộc một lớp hay có một giá trị phân loại.

Nếu số lượng giá trị phân loại là 2 (phân loại nhị phân), Entropy có giá trị nằm trong khoảng $[0..1]$,

$\text{Entropy}(S) = 0 \Rightarrow$ tập ví dụ S chỉ toàn ví dụ thuộc cùng một loại, hay S là thuần nhất.

$\text{Entropy}(S) = 1 \Rightarrow$ tập ví dụ S có các ví dụ thuộc các loại khác nhau với độ pha trộn là cao nhất.

$0 < \text{Entropy}(S) < 1 \Rightarrow$ tập ví dụ S có số lượng ví dụ thuộc các loại khác nhau là không bằng nhau. Để đơn giản ta xét trường hợp các ví dụ của S chỉ thuộc loại âm (-) hoặc dương (+).

Cho trước: Tập S là tập dữ liệu rèn luyện, trong đó thuộc tính phân loại có hai giá trị, giá sử là âm (-) và dương (+)

* p_+ là phần các ví dụ dương trong tập S

* p_- là phần các ví dụ âm trong tập S

Khi đó, Entropy đo độ pha trộn của tập S theo công thức sau:

$$\text{Entropy}(S) = - p_+ \log_2 p_+ - p_- \log_2 p_-$$

Một cách tổng quát hơn, nếu các ví dụ của tập S thuộc nhiều hơn hai loại, giả sử là có c giá trị phân loại thì công thức Entropy tổng quát là [13]:

$$\text{Entropy (S)} = \sum_{i=1}^c - p_i \log_2 p_i$$

Lượng thông tin thu được đo mức độ giảm Entropy

Entropy là một số đo đo độ pha trộn của một tập ví dụ, bây giờ chúng ta sẽ định nghĩa một phép đo hiệu suất phân loại các ví dụ của một thuộc tính. Phép đo này gọi là lượng thông tin thu được, nó đơn giản là lượng giảm Entropy mong đợi gây ra bởi việc phân chia các ví dụ theo thuộc tính này. Một cách chính xác hơn, Gain(S,A) của thuộc tính A, trên tập S, được định nghĩa như sau [13]:

$$\text{Gain(S,A)} = \text{Entropy(S)} - \sum_{v \in \text{values(A)}} \frac{|S_v|}{|S|} \text{Entropy(S}_v\text{)}$$

trong đó values(A) là tập hợp có thể có các giá trị của thuộc tính A, và S_v là tập con của S chứa các ví dụ có thuộc tính A mang giá trị v.

1.2.3.6. Tìm kiếm không gian giả thuyết trong ID3

Cũng như các phương pháp học quy nạp khác, ID3 cũng tìm kiếm trong một không gian các giả thuyết một giả thuyết phù hợp với tập dữ liệu rèn luyện. Không gian giả thuyết mà ID3 tìm kiếm là một tập hợp các cây quyết định có thể có.

1.2.4. Đánh giá hiệu suất và tập luật của cây quyết định

1.2.4.1. Đánh giá hiệu suất của cây quyết định

1.2.4.2. Chuyển cây về các dạng luật

CHƯƠNG 2

ỨNG DỤNG CÂY QUYẾT ĐỊNH TRONG CÔNG TÁC ĐỀN BÙ GIẢI TỎA

2.1. KHẢO SÁT HIỆN TRẠNG

2.1.1. Giới thiệu về hệ thống tổ chức đơn vị

2.1.2. Phân tích quy trình và thực trạng công tác đền bù giải tỏa

2.1.3. Giải pháp xây dựng và kịch bản hệ thống hỗ trợ ra quyết định

2.2. TRIỂN KHAI ỨNG DỤNG CÂY QUYẾT ĐỊNH

2.2.1. Phân tích dữ liệu

Diện tích (Dientich): Đây là diện tích đất bị thu hồi cho việc giải tỏa. Để có thể áp dụng cho giải thuật ID3, ta cần rời rạc hóa các giá trị này. Dựa vào các giá trị diện tích trong tập dữ liệu, ta có thể chia nó thành các khoảng sau: dưới 100 m² (100-), từ 100 m² đến dưới 200 m² (100+), từ 200 m² đến dưới 300 m² (200+), từ 300 m² trở lên (300+). Như vậy, tập các giá trị của diện tích đất được mô tả như sau: Dientich = {100-; 100+; 200+; 300+}

Nhân khẩu (Nhankhau): số lượng người sống tại hộ bị giải tỏa. Do đó, tập các giá trị của Nhân khẩu được cho như sau:

$$\text{Nhankhau} = \{8-; 8+\}$$

Giá trị đền bù (Giatridenbu): Đây là số tiền mà nhà nước đền bù cho hộ bị giải tỏa và phụ thuộc vào khu đất của hộ gia đình, nhà cửa xây trên đó, vật liệu, ... Như vậy, tập giá trị của thuộc tính Giá trị đền bù là: Giatridenbu = {50-; 50+; 100+; 150+; 300+}

Nguồn gốc (Nguongoc): thể hiện nguồn gốc về mảnh đất và ngôi nhà, của hộ gia đình. Từ đó, tập các giá trị của thuộc tính Nguồn gốc là: Nguongoc = {TKP; CN}

Quyết định đền bù (Quyếtđinh): Đây là quyết định mà Ban đền bù giải tỏa thực hiện tương ứng với hồ sơ của từng hộ bị giải tỏa. Tập giá trị của Quyết định đền bù được thể hiện như sau:

$$\text{Quyếtđinh} = \{\text{Chinh}; \text{Phu}; \text{Chinh_cc}; \text{Chinh_phu}\}$$

Từ bảng chứa tập dữ liệu, bằng cách rời rạc hóa các thuộc tính, ta thu được dữ liệu chứa các giá trị rời rạc dưới đây.

Bảng 2.2. Tập dữ liệu huấn luyện sau khi rời rạc hóa

STT	Dien tich (m^2)	Nhan khau	Giatri denbu (triệu đồng)	Nguồn gốc	Quyếtđinh
1	100+	8+	100+	TKP	Chinh_cc
2	100-	8-	100+	TKP	Chinh
3	100-	8-	150+	CN	Chinh
4	100+	8-	100+	CN	Chinh
5	100+	8-	150+	TKP	Chinh
6	100+	8-	50-	CN	Phu
7	100-	8-	50-	TKP	Phu
8	100-	8+	100+	TKP	Chinh
9	100+	8-	50-	TKP	Phu
10	100+	8-	50+	CN	Phu
11	300+	8+	300+	TKP	Chinh_phu
12	200+	8+	300+	CN	Chinh_phu
13	200+	8+	300+	TKP	Chinh_phu
14	100+	8+	150+	TKP	Chinh_cc
15	100-	8+	50+	TKP	Chinh_cc
16	100+	8+	50+	TKP	Chinh_cc

2.2.2. Triển khai giải thuật ID3

Cây quyết định được xây dựng bằng cách mở rộng cây xuất phát từ nút gốc. Nút gốc được đặt tên là nút 1, và ứng với mức 0 của cây. Các nút con của nút gốc sẽ ứng với mức 1, và tiếp tục như vậy.

Thuộc tính dùng để phân tích nút hiện thời. Nếu nút hiện thời là nút lá, thuộc tính sẽ bằng rỗng. Nếu nút hiện thời không phải là nút lá, nó sẽ chứa tên các nút con xuất phát từ nút hiện thời, cũng như các giá trị tương ứng của thuộc tính dùng để phân tích nút hiện thời. Nếu là nút lá thì có thêm quyết định đầu ra ở tại nút hiện thời. Ngược lại, nếu không phải là nút lá thì không có chứa thông tin này.

2.2.3. Xây dựng cây quyết định

Ta sẽ bắt đầu việc xây dựng cây quyết định từ nút gốc. Trước tiên, tại nút gốc sẽ chứa toàn bộ dữ liệu huấn luyện. Ta thấy rằng, tại nút gốc, tập S chứa các giá trị không đồng nhất. Rõ ràng tập S này không đồng nhất, hay Entropy khác không. Do đó, ta cần phải phân tích nút gốc này ra thành các nút con với mục đích là kì vọng của Entropy của tập S sẽ nhỏ hơn. Trước hết, Entropy của tập S được tính như sau:

Bảng 2.3. Bảng số liệu thể hiện cách tính Entropy của tập S

S (Quyết định)	Số lượng	p_i	$-p_i \cdot \log_2(p_i)$	Entropy(S)
Chinh	5	5/16	0.5244	1.9772
Phu	4	4/16	0.5000	
Chinh_phu	3	3/16	0.4528	
Chinh_cc	4	4/16	0.5000	

Tiếp theo ta sẽ tính kì vọng của Entropy của tập S khi biết trước từng thuộc tính còn lại. Với mỗi tập con này, ta cũng tính được tỉ lệ (p_i) của các giá trị của mục tiêu, và từ đó tính Entropy tương ứng giống như trên. Kết quả được tóm tắt ở bảng sau.

Bảng 2.4. Entropy của các tập con của S phân chia theo thuộc tính Diện tích

Dientich	Số lượng	P _i				Entropy
		Chinh	Phu	Chinh phu	Chinh cc	
100-	5	3/5	1/5	0	1/5	1.3710
100+	8	2/8	3/8	0	3/8	1.5613
200+	2	0	0	1	0	0
300+	1	0	0	1	0	0

Kì vọng của Entropy của tập S đối với thuộc tính diện tích là:

$$E_{\text{Dientich}} = 5/16 * 1.3710 + 8/16 * 1.5613 + 2/16 * 0 + 1/16 * 0 = \mathbf{1.2091}$$

Do đó, độ lợi thông tin tương ứng là:

$$G(S, \text{Dientich}) = \text{Entropy}(S) - E_{\text{Dientich}} = 1.9772 - 1.2091 = \mathbf{0.7681}$$

Một cách tương tự, ta tính được kì vọng của Entropy của tập S đối với thuộc tính Nhân khẩu, và độ lợi thông tin tương ứng là:

$$E_{\text{Nhankhau}} = 8/16 * 1 + 8/16 * 1.4056 = \mathbf{1.2028}$$

$$G(S, \text{Nhankhau}) = 1.9772 - 1.2028 = \mathbf{0.7744}$$

Bảng 2.6. Entropy của các tập con của S theo thuộc tính Giá trị đền bù

Giatridenbu	Số lượng	P _i				Entropy
		Chinh	Phu	Chinh phu	Chinh cc	
50-	3	0	1	0	0	0
50+	3	0	1/3	0	2/3	0.9183
100+	4	3/4	0	0	1/4	0.8113
150+	3	2/3	0	0	1/3	0.9183
300+	3	0	0	1	0	0

Với thuộc tính Giatridenbu, ta có:

$$E_{\text{Giatridenbu}} = 3/16 * 0 + 3/16 * 0.9183 + 4/16 * 0.8113$$

$$+ 3/16 * 0.9183 + 3/16 * 0 = \mathbf{0.5472}$$

$$G(S, \text{Giatridenbu}) = 1.9772 - 0.5472 = \mathbf{1.4300}$$

Với thuộc tính Nguongoc, ta có kết quả sau:

$$E_{\text{Nguongoc}} = 11/16 * 1.9363 + 5/16 * 1.5219 = \mathbf{1.8068}$$

$$G(S, \text{Nguongoc}) = 1.9772 - 1.8068 = \mathbf{0.1704}$$

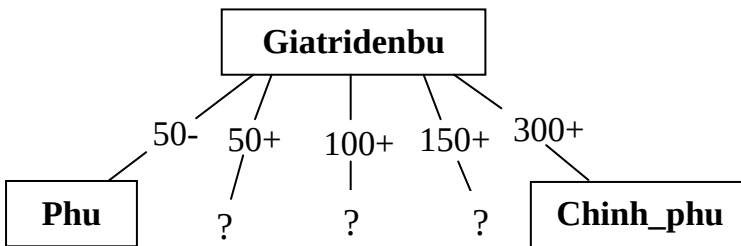
Kết quả so sánh của bốn thuộc tính được tóm tắt ở bảng sau:

Bảng 2.8. Bảng so sánh số liệu của 4 thuộc tính

	Diện tích	Nhân khẩu	Nguồn gốc	Giá trị đền bù
Kì vọng của Entropy	1.2091	1.2028	1.8068	0.5472
Độ lợi thông tin	0.7681	0.7744	0.1704	1.4300

So sánh bốn thuộc tính Dientich, Nhankhau, Giatridenbu và Nguongoc, ta thấy Giatridenbu cho độ lợi thông tin lớn nhất, và ứng với kì vọng của Entropy là nhỏ nhất. Nói cách khác, thuộc tính Giatridenbu có khả năng phân loại tốt nhất; do đó, ta chọn thuộc tính này làm thuộc tính để phân loại nút gốc của cây quyết định.

Như vậy, tại nút đầu tiên của cây quyết định, cây quyết định được phân tích nhờ vào thuộc tính **Giatridenbu** như sau:



Hình 2.5. Cây quyết định được triển khai đến mức 1

Theo hình trên, tại các nút Phu và Chinh_phu, ta có Entropy bằng 0; nên đó sẽ là hai nút lá. Việc phát triển cây quyết định sẽ dừng lại ở 2 nút này. Với các trường hợp khác của Giatridenbu, ta cần tiếp tục xem xét các thuộc tính khác để mở rộng cây quyết định.

Ta tiếp tục thực hiện giải thuật ID3 để tìm ra thuộc tính nào sẽ được sử dụng với các trường hợp 50+, 100+, 150+ của Giatridenbu.

Với Giatridenbu = 50+, ta có tập dữ liệu con như sau:

Bảng 2.9. Bảng số liệu tương ứng với giá trị đều bù = 50+

Dientich	Nhankhau	Nguongoc	Quyetdinh
100+	8-	CN	Phu
100-	8+	TKP	Chinh_cc
100+	8+	TKP	Chinh_cc

Gọi S1 là tập chứa các giá trị của thuộc tính Quyetdinh với tập con này. Khi đó, Entropy của tập S1 là:

$$\text{Entropy}(S1) = -1/3 * \log_2(1/3) - 2/3 * \log_2(2/3) = \mathbf{0.9183}$$

Một cách tương tự, ta tóm tắt quá trình tính toán cho các thuộc tính trong các bảng sau.

Bảng 2.10. Entropy của các tập con của S1 phân chia theo thuộc tính Diện tích

Dientich	Số lượng	P _i		Entropy
		Phu	Chinh_cc	
100-	1	0	1	0
100+	2	1/2	1/2	1

Kì vọng của Entropy của tập S1:

$$E_{\text{Dientich}} = 1/3 * 0 + 2/3 * 1 = \mathbf{0.6667}$$

$$\text{Độ lợi thông tin: } G(S1, \text{Dientich}) = 0.9183 - 0.6667 = \mathbf{0.2516}$$

Bảng 2.11. Entropy của các tập con của S1 theo thuộc tính Nhân khẩu

Nhankhau	Số lượng	P _i		Entropy
		Phu	Chinh_cc	
8-	1	1	0	0
8+	2	0	1	0

Kì vọng của Entropy của S1 và độ lợi thông tin được tính như sau:

$$E_{\text{Nhankhau}} = 1/3*0 + 2/3*0 = 0$$

$$G(S1, \text{Nhankhau}) = 0.9183 - 0 = \mathbf{0.9183}$$

Kì vọng của Entropy của S1 và độ lợi thông tin đối với thuộc tính Nguongoc được tính như sau:

$$E_{\text{Nguongoc}} = 2/3*0 + 1/3*0 = 0$$

$$G(S1, \text{Nguongoc}) = 0.9183 - 0 = \mathbf{0.9183}$$

Với Giatriденбу = 100+, ta có tập dữ liệu con như sau:

Bảng 2.13. Bảng số liệu tương ứng với giá trị đền bù = 100+

Dientich	Nhankhau	Nguongoc	Quyetdinh
100+	8+	TKP	Chinh_cc
100-	8-	TKP	Chinh
100-	8+	TKP	Chinh
100+	8-	CN	Chinh

Gọi S2 là tập chứa các giá trị của thuộc tính Quyetdinh với tập con này.

Khi đó, Entropy của tập S2 là:

$$\text{Entropy}(S2) = -1/4*\log_2(1/4) - 3/4*\log_2(3/4) = \mathbf{0.8113}$$

Một cách tương tự, ta xét độ lợi thông tin với các thuộc tính khác nhau như sau.

Bảng 2.14. Entropy của các tập con của S2 phân chia theo thuộc tính Diện tích

Dientich	Số lượng	P _i		Entropy
		Chinh	Chinh_cc	
100-	2	1	0	0
100+	2	1/2	1/2	1

Kì vọng của Entropy của S2:

$$E_{\text{Dientich}} = 2/4*1 + 2/4*0 = \mathbf{0.5}$$

$$\text{Độ lợi thông tin: } G(S2, \text{Dientich}) = 0.8113 - 0.5 = \mathbf{0.3113}$$

Một cách tương tự, ta thu được kết quả đối với thuộc tính

Nhankhau: $E_{Nhankhau} = 2/4*0 + 2/4*1 = \mathbf{0.5}$

$G(S2, Nhankhau) = 0.8113 - 0.5 = \mathbf{0.3113}$

Tương tự, ta thu được kết quả đối với thuộc tính Nguongoc:

$E_{Nguongoc} = 3/4*0.9183 + 1/4*0 = \mathbf{0.6887}$

$G(S2, Nguongoc) = 0.8113 - 0.6887 = \mathbf{0.1226}$

Với Giatridentbu = 150+, ta có tập dữ liệu con như sau:

Bảng 2.17. Bảng số liệu tương ứng với giá trị đền bù = 150+

Dientich	Nhankhau	Nguongoc	Quyetdinh
100-	8-	CN	Chinh
100+	8-	TKP	Chinh
100+	8+	TKP	Chinh_cc

Gọi S3 là tập chứa các giá trị của thuộc tính Quyetdinh với tập con này. Khi đó, Entropy của tập S3 là:

$Entropy(S3) = -1/3*\log_2(1/3) - 2/3*\log_2(2/3) = \mathbf{0.9183}$

Ta xét từng thuộc tính.

Bảng 2.18. Entropy của các tập con của S3 phân chia theo thuộc tính Diện tích

Dientich	Số lượng	P _i		Entropy
		Chinh	Chinh_cc	
100-	1	1	0	0
100+	2	1/2	1/2	1

Kì vọng Entropy của S3: $E_{Dientich} = 1/3*0 + 2/3*1 = \mathbf{0.6667}$

Độ lợi thông tin: $G(S3, Dientich) = 0.9183 - 0.6667 = \mathbf{0.2516}$

Kì vọng của Entropy của S3: $E_{Nhankhau} = 1/3*0 + 2/3*0 = \mathbf{0}$

Độ lợi thông tin: $G(S3, Nhankhau) = 0.9183 - 0 = \mathbf{0.9183}$

Kì vọng của Entropy của S3:

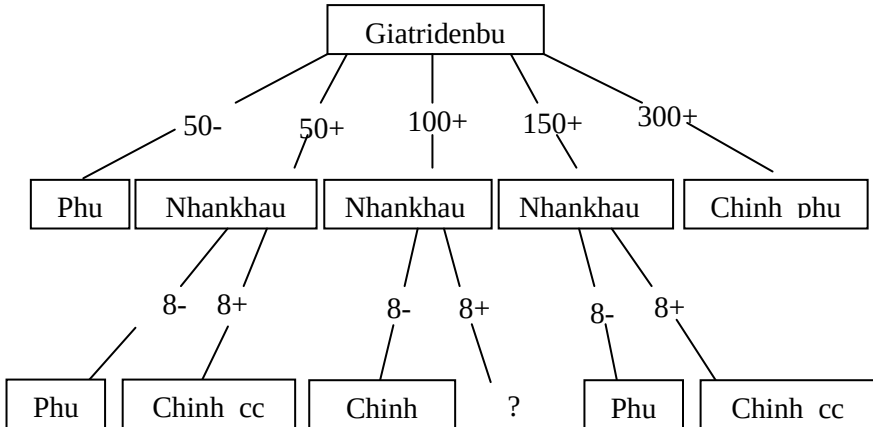
$E_{Nguongoc} = 1/3*0 + 2/3*1 = \mathbf{0.6667}$

Độ lợi thông tin:

$G(S3, Nguongoc) = 0.9183 - 0.6667 = \mathbf{0.2516}$

Thuộc tính Nhankhau cũng được chọn để mở rộng cây quyết định với trường hợp Giatridenbu = 150+.

Cây quyết định bây giờ sẽ là:



Hình 2.6. Cây quyết định được triển khai đến mức 2.

Như vậy ở mức thứ hai, ta thu được thêm năm nút lá (có entropy bằng 0). Ta sẽ tiếp tục phát triển cây quyết định tại nút được đánh dấu hỏi (?), với tập dữ liệu con như sau:

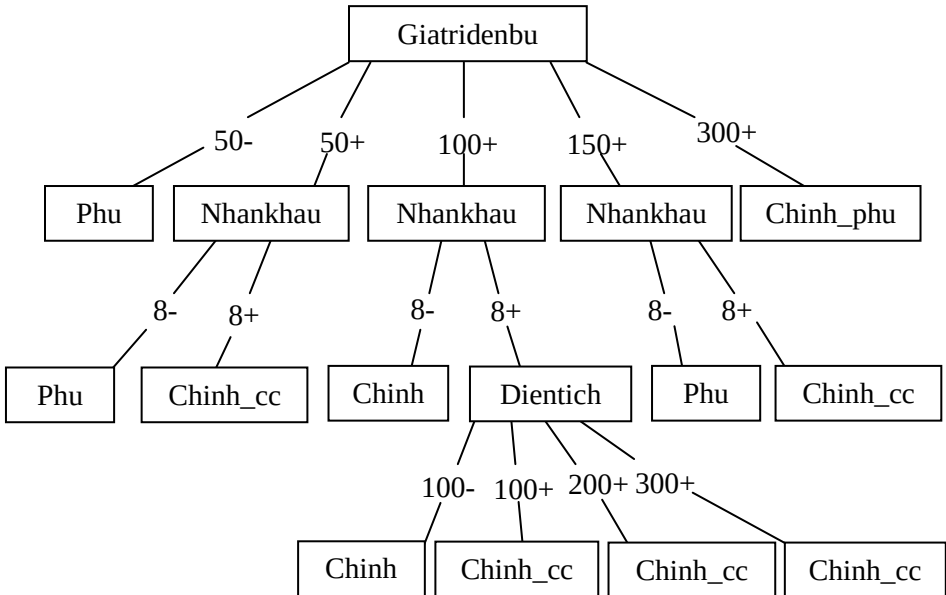
Bảng 2.21. Bảng số liệu tương ứng với nhân khẩu ≥ 8

Dientich	Nguongoc	Quyetdinh
100+	TKP	Chinh_cc
100-	TKP	Chinh

Ta dễ dàng nhận thấy, nếu chọn Dientich làm thuộc tính phân loại tiếp theo ta sẽ có Entropy bằng không. Ngược lại, thuộc tính Nguongoc cho Entropy khác không (1/2). Do đó, thuộc tính Dientich sẽ được chọn để phát triển cây quyết định, và nó cũng sẽ dừng lại sau nút này (vì có Entropy = 0).

Ở đây, ta chú ý thuộc tính Dientich. Với thuộc tính này chỉ có 2 giá trị, 100- và 100+, cho tập con khác rỗng. Với các giá trị

200+ và 300+, ta không tìm được quyết định cụ thể dựa trên dữ liệu huấn luyện. Để giải quyết vấn đề này, có thể có hai cách sau. Thứ nhất, gán lá "nul" cho các trường hợp này (cụ thể ở đây là trường hợp Dientich bằng 200+ và 300+). Khi đó, cây quyết định sẽ không đưa ra được quyết định nếu đi đến nút này. Thứ hai, các nút này được gán cho lớp có nhiều trường hợp rơi vào nhất (xuất phát từ cây con tương ứng) [14]. Trong trường hợp này, cây quyết định sẽ giải quyết được tất cả các tình huống. Ngoài ra, nó cũng cho kết quả hoàn toàn chính xác đối với dữ liệu huấn luyện. Ở đây, ta sẽ lựa chọn cách xử lý thứ hai, nghĩa là đảm bảo được là cây quyết định sẽ luôn cho được quyết định đầu ra, đồng thời cũng thể hiện chính xác dữ liệu huấn luyện.



Hình 2.8. Cây quyết định được xây dựng từ CSDL huấn luyện ở trên

2.3. NHẬN XÉT, ĐÁNH GIÁ

2.3.1. Nhận xét về cây quyết định đền bù giải tỏa

Trước hết, cây quyết định được xây dựng ở trên thể hiện tốt tất cả các trường hợp trong cơ sở dữ liệu huấn luyện.

Ta cũng thấy rằng không phải tất cả các trường hợp đều phải duyệt qua toàn bộ các thuộc tính để đi đến quyết định. Một số trường hợp chỉ cần 1 hoặc 2 thuộc tính là đủ để xác định quyết định đầu ra. Từ đó, ta thấy một ưu điểm của phương pháp cây quyết định là không nhất thiết phải biết tất cả các thuộc tính của mỗi trường hợp để có thể quyết định. Hay nói một cách khác, cây quyết định có thể cho quyết định ngay cả đối với những trường hợp mà dữ liệu không đầy đủ.

2.3.2. Đánh giá về cây quyết định đền bù giải tỏa

Sau khi phân tích, nhận xét, bây giờ chúng tôi thử tiến hành kiểm tra, đánh giá một số trường hợp giải tỏa khác. Với dữ liệu huấn luyện ban đầu gồm có 16 trường hợp, chúng tôi chọn ra thử 8 trường hợp ngẫu nhiên để kiểm tra.

Ở đây ta quan tâm đến kết quả quyết định giải tỏa đền bù do Ban Quản lý dự án công trình đường Bạch Đằng Đông đưa ra và kết quả quyết định dựa vào cây quyết định xây dựng ở trên. Ta thấy trong 87.5% trường hợp đền bù, cây quyết định đưa ra quyết định hoàn toàn phù hợp với quyết định của Ban Quản lý dự án công trình đường Bạch Đằng Đông. Trong 12.5% còn lại, cây quyết định cho kết quả lệch đôi chút so với kết quả của Ban Quản lý dự án công trình đường Bạch Đằng Đông.

Tóm lại việc xây dựng cây quyết định với kết quả được kiểm thử như trên đạt 87.5% là một con số tương đối tốt và chấp nhận được. Nó sẽ là hệ thống trợ giúp đắc lực cho Lãnh đạo, người làm chuyên môn trong công tác đền bù giải tỏa.

CHƯƠNG 3

CÀI ĐẶT VÀ THỬ NGHIỆM HỆ THỐNG

3.1. CÀI ĐẶT CHƯƠNG TRÌNH

3.2. CÀI ĐẶT CƠ SỞ DỮ LIỆU

3.2.1. CSDL huấn luyện cây quyết định

3.2.2. CSDL kiểm tra, xử lý

3.3. XÂY DỰNG GIAO DIỆN VÀ CHỨC NĂNG HỆ THỐNG

3.3.1. Giao diện chương trình

3.3.2. Các chức năng của hệ thống

Phần 1: Huấn luyện cây quyết định

Cho phép chọn dữ liệu huấn luyện (nút Browse). Dữ liệu này có thể được lưu dưới dạng file text, hoặc excel.

Huấn luyện cây quyết định (nút Learn), hay nói một cách khác, xây dựng cây quyết định thông qua việc xác định thông tin của các nút. Trong chương trình, nút Learn chỉ thực hiện nếu dữ liệu huấn luyện đã có sẵn.

Hiển thị cây quyết định đã huấn luyện (nút Display the decision tree). Cây quyết định sẽ được vẽ ra, đi từ nút gốc đến các nút lá, với các thuộc tính sử dụng ở các nút trung gian nhằm mở rộng cây quyết định.

Phần 2: Áp dụng cây quyết định

Nhập dữ liệu: Ở đây, ta có các ô text và pop-up giúp cho việc đưa vào thông tin của một hộ gia đình.

Các thao tác trên dữ liệu nhập vào: Chương trình cũng cho phép thực hiện vài chức năng cơ bản liên quan đến cơ sở dữ liệu như sau: Tạo mới, Lưu, Xóa, Tìm kiếm,

Ra quyết định với dữ liệu tương ứng với hộ hiện thời: Nếu dữ liệu hiện thời là hợp lệ, khi nhấn nút Decide, chương trình sẽ cho ra quyết định đền bù.

Thống kê các hộ giải tỏa theo các thuộc tính khác nhau: Chức năng này cho phép thống kê phần trăm và nhận biết về đặc điểm của các hộ gia đình trong vùng bị giải tỏa.

3.4. THỬ NGHIỆM HỆ THỐNG

3.4.1. Huấn luyện cây quyết định

3.4.2. Áp dụng cây quyết định

Ung dụng Cây Quyết định vào bài toán Dén bù giải tỏa !

DECISION TREE APPLICATION COMPUTER SCIENCE K10 (2008-2011)

Decision Tree Learning

Learning data:

Information

Name*: Year of birth:

Gender: Address:

Area: Number of Persons:

Value: Origin:

Compensation

+

	Ho va ten	Gioi tinh	Nam sinh	Địa chỉ	Nhan khẩu	Diện tích	Nguồn g
1	Lê Hoài Nam	Nam	1977	Trưng Nữ Vương	9	257	Chuyen r
2	Trần Viết Phú	Nam	1972	Thanh Bình	3	124	Chuyen r
3	Bùi Thu Thảo	Nu	1958	Lê Độ	11	158	Chuyen r
4	Trương Ngọc	Nam	1955	Ngô Quyền	9	158	Chuyen r
5	Lê Thị Bê	Nu	1954	Lê Văn Hiến	4	75	Tu khai p

File processing

Hình 3.10. Hình ảnh tổng thể chương trình

3.5. NHẬN XÉT, ĐÁNH GIÁ KHI TRIỂN KHAI ỨNG DỤNG

Sau khi đưa hệ thống hỗ trợ quyết định trong công tác đền bù giải tỏa vào thử nghiệm trên một dự án đang triển khai với số dữ liệu khoảng 899 hồ sơ, chúng tôi rút ra được những nhận xét, so sánh, đánh giá một cách tương đối về hệ thống hỗ trợ như sau:

Hệ thống được triển khai thử nghiệm dễ dàng, dễ sử dụng.

Hệ thống hoạt động ổn định, hiển thị kết quả nhanh chóng và rất thuận tiện.

Với số lượng hồ sơ tương đối lớn và phức tạp, hệ thống đã xử lý và cho ra được những quyết định tương đối chính xác và phù hợp. Đặc biệt là rút ngắn được thời gian xử lý hồ sơ đến mức chấp nhận được. Ở đây ta thử làm phép thống kê và so sánh với tất cả các hồ sơ của dự án đang thí điểm triển khai xử lý bằng chương trình hỗ trợ quyết định trên máy tính. Các kết quả thu được cụ thể như sau:

- Đối với 281 hồ sơ: trước đây nếu sử dụng cách làm thủ công thì thời gian để xử lý xong 281 hồ sơ sẽ phải mất đến 562 ngày. Nhưng hiện nay nếu sử dụng chương trình hỗ trợ quyết định trên máy tính thì thời gian xử lý chỉ còn 140 ngày.

- Đối với 618 hồ sơ: thời gian xử lý chỉ còn 309 ngày so với 1236 ngày.

Đặc biệt ở đây với hệ thống hỗ trợ quyết định, lãnh đạo đơn vị, lãnh đạo phòng chuyên môn có thể dễ dàng hình dung, thống kê và triển khai phương án bố trí đất tái định cư được linh động, phù hợp và chủ động hơn.

Chính vì vậy mà Lãnh đạo đơn vị đã rất hài lòng và chấp nhận với kết quả, quyết định nhận được từ sự trợ giúp quyết định của hệ thống thông minh này mang lại.

KẾT LUẬN

1. Đánh giá kết quả

Kết quả đạt được

Đã tiến hành phân tích, tìm hiểu được quy trình đền bù giải tỏa hiện tại đang được triển khai tại đơn vị, từ đó tìm ra những bất cập và những vấn đề cần giải quyết trong chương trình đền bù giải tỏa.

Áp dụng và triển khai được cây quyết định, thuật toán ID3 vào bài toán đền bù giải tỏa phù hợp với yêu cầu tại Ban Quản lý dự án công trình đường Bạch Đằng Đông và là lời giải cho phép giải quyết bài toán.

Xây dựng được hệ thống với đầy đủ các chức năng đáp ứng được các yêu cầu về chuyên môn trong công tác đền bù giải tỏa. Hệ thống bao gồm các chức năng như sau:

Huấn luyện cây quyết định: Cho phép chọn dữ liệu huấn luyện từ 01 tập tin dữ liệu có sẵn. Sau khi được chọn ta có thể xem dữ liệu được thể hiện trên bảng lưới để tiện việc theo dõi và để huấn luyện cây quyết định.

Hiển thị cây quyết định: Sẽ vẽ ra cây quyết định tương ứng với các thuộc tính, dữ liệu được học từ tập dữ liệu huấn luyện.

Nhập dữ liệu: Phần này hỗ trợ người dùng trong việc nhập dữ liệu, thông tin có liên quan đến công tác đền bù giải tỏa

Ra quyết định: Phần này sẽ đưa ra quyết định với dữ liệu tương ứng của hệ hiện thời. Nếu dữ liệu hiện thời là hợp lệ, khi nhấn nút Decide, chương trình sẽ đưa ra quyết định đền bù dựa trên cây quyết định đã được huấn luyện và hiển thị kết quả.

Thống kê: Cho phép tiến hành thống kê các hộ giải tỏa theo từng thuộc tính khác nhau, đồng thời hiển thị biểu đồ theo tỷ lệ phần trăm của các thuộc tính nhằm thuận tiện cho việc so sánh, đưa ra nhận định.

Hệ thống được thử nghiệm và đang triển khai tại Ban Quản lý dự án công trình đường Bạch Đằng Đông – Thành phố Đà Nẵng.

Nhận xét

Ưu điểm

Giao diện hệ thống khá thân thiện, rõ ràng và các chức năng thể hiện đầy đủ giúp người dùng thao tác thuận tiện, dễ dàng hơn.

Hệ CSDL được lưu trữ bằng cấu trúc của Matlab nên được bảo mật tương đối cao và an toàn cho người sử dụng.

Quá trình xử lý để đưa ra quyết định rất nhanh chóng và trực quan.

Kết quả đưa ra rất chính xác và phù hợp với các quy định, hướng dẫn của UBND thành phố Đà Nẵng.

Kết quả thống kê rất chi tiết nhằm đưa ra được cái nhìn tổng quan và đầy đủ cho người làm chuyên môn, Lãnh đạo trong công tác đền bù giải tỏa.

Nhược điểm

Hệ thống chưa xử lý, đáp ứng việc hỗ trợ ra quyết định cho nhiều dự án cùng một thời điểm.

Chương trình chưa có chức năng sử dụng được trên mạng do nhu cầu và đặc thù riêng của công tác đền bù giải tỏa.

2. Phạm vi ứng dụng

Chương trình được xây dựng chủ yếu để phục vụ cho công tác đền bù giải tỏa tại Ban Quản lý dự án công trình đường Bạch Đằng Đông – Thành phố Đà Nẵng.

Chương trình có thể phát triển, mở rộng và triển khai ở các đơn vị thực hiện chức năng đền bù giải tỏa khác thuộc thành phố Đà Nẵng.

3. Hướng phát triển

Nâng cao độ chính xác hơn nữa trong xử lý bằng cách ứng dụng, triển khai các thuật toán tối ưu hơn.

Xây dựng hệ thống có thể chạy được trên môi trường mạng, để phát triển và triển khai rộng như một hệ thống tập trung và thống nhất cho toàn thành phố.

Phát triển hệ thống theo tiêu chuẩn mở để phục vụ được cho nhiều đơn vị.