

MỘT PHƯƠNG PHÁP RA QUYẾT ĐỊNH TẬP THỂ DỰA TRÊN PHÂN LOẠI DỮ LIỆU MỜ: GDM-FC

GDM-FC: A group decision making method based on fuzzy data classification

Nguyễn Hải Thanh¹, Đặng Xuân Hà, Trần Vũ Hà

SUMMARY

This paper introduces GDM-FC, a new group decision making method based on fuzzy data classification with the following components: (i) Fuzzy data classification by minimum between-center and least square distance criteria; (ii) A new definition of distance between two fuzzy numbers; (iii) Delphi method for achieving opinion consensus among experts. The method has been used for ranking and selecting Pareto optimal solutions of the multi-objective linear programming problems arising from land use planning issues.

Key words: Group decision making, fuzzy data classification, experts' opinion consensus.

1. ĐẶT VẤN ĐỀ

Trong nhiều lĩnh vực khoa học - công nghệ và kinh tế - xã hội, đặc biệt là trong các bài toán quản lý, việc ra quyết định luôn có một vai trò hết sức quan trọng. Ra quyết định là một trách nhiệm chủ chốt nhất của bộ máy quản lý. Thông tin ngày càng trở nên đa dạng, đa chiều. Việc xử lý thông tin đòi hỏi tính khoa học, chính xác, cập nhật. Ngày nay, các mô hình toán học với các dữ liệu đầu vào xác thực luôn tỏ ra hết sức tiện lợi trong việc xử lý thông tin để chọn ra, hay nói cách khác là đưa ra quyết định lựa chọn, các phương án hành động tốt nhất, hợp lý nhất (Gillet, 1990). Đây là khía cạnh khai phá dữ liệu trong việc ra quyết định. Tuy nhiên, không một mô hình toán học nào có thể tổng quát tới mức tính đến tất cả các khía cạnh của bài toán thực tiễn cũng như đánh giá được chính xác các phương án hành động nào là sẽ hợp lý nhất. Vì vậy, việc khai thác ý kiến của các chuyên gia để đánh giá, lựa chọn các phương án

cũng là khía cạnh khai phá tri thức trong vấn đề ra quyết định.

Đặc biệt, trong bài toán quy hoạch sử

dụng đất, dựa trên các dữ liệu nông nghiệp thu thập được qua nhiều nguồn: số liệu thống kê, các dữ liệu thu được từ các phần mềm GIS khi khảo sát về thông tin đất và môi trường, quyết định cần được đưa ra về các phương án sử dụng đất sao cho đáp ứng được các mục tiêu về kinh tế, môi trường và sử dụng đất hợp lý. Để đưa ra được các quyết định hợp lý nhất chúng ta cần xây dựng được mô hình toán học, mà cụ thể là mô hình tối ưu đa mục tiêu để khai phá dữ liệu và đưa ra được các phương án tối ưu về mặt toán học (Nguyễn Hải Thanh và cs, 2005) và thiết lập được mô hình ra quyết định tập thể để lựa chọn các phương án quy hoạch sử dụng đất được đánh giá là hợp lý nhất khi khai phá tri thức của các chuyên gia (Kaufmann A. và cs, 1991).

Trong bài báo này, chúng tôi trình bày một phương pháp ra quyết định tập thể

¹ Khoa Công nghệ thông tin, Đại học Nông nghiệp I
hành động là một việc làm cần thiết. Đây

dựa trên phân loại dữ liệu mờ. Đây chỉ là

một thành phần tạo nên hệ hỗ trợ ra quyết định tập thể trong quy hoạch sử dụng đất. Vì vậy, để trình bày vấn đề một cách hệ thống, các mục tiếp theo của bài báo này như sau: Mục 2 sẽ tóm lược về mô hình quy hoạch tuyến tính đa mục tiêu, mục 3 giới thiệu về phân loại dữ liệu mờ. Phương pháp Delphy thống nhất ý kiến chuyên gia được trình bày trong mục 4, mục 5 trình bày về chương trình máy tính hỗ trợ ra quyết định tập thể. Cuối cùng, một số kết luận sơ bộ được đưa ra trong mục 6.

2. TÓM LƯỢC VỀ MÔ HÌNH QUY HOẠCH TUYẾN TÍNH ĐA MỤC TIÊU

Trong các bài toán kỹ thuật, công nghệ, quản lý, kinh tế nông nghiệp v.v... nảy sinh từ thực tế, chúng ta thường phải xem xét để tối ưu hoá đồng thời một lúc nhiều mục tiêu. Các mục tiêu này thường là khác về thứ nguyên, tức là chúng được đo bởi các đơn vị khác nhau. Những tình huống như vậy tạo ra các bài toán tối ưu đa mục tiêu. Người kỹ sư/người ra quyết định lúc này cần phải tối ưu hoá (cực đại hoá hoặc cực tiểu hoá tùy theo tình huống thực tế) không phải là chỉ một mục tiêu nào đó, mà là đồng thời tất cả các mục tiêu đã đặt ra.

Bài toán tối ưu đa mục tiêu mà trong đó miền ràng buộc D là tập lồi đa diện và các mục tiêu $z_i = f_i(X)$, với $i = 1, 2, \dots, p$, là các hàm tuyến tính xác định trên D , được gọi là bài toán quy hoạch tuyến tính đa mục tiêu. Khi đó, ta có mô hình toán học sau đây được gọi là mô hình quy hoạch tuyến tính đa mục tiêu (Nguyễn Hải Thanh và cs, 2005):

Max CX với ràng buộc $X \in D$ trong đó: C là ma trận cấp $p \times n$, $D = \{X \in \mathbb{R}^n: AX \leq B\}$ và A là ma trận cấp $m \times n$ và $B \in \mathbb{R}^m$.

Có thể nói, BTQHTT đa mục tiêu là BTQHTT, mà trong đó chúng ta phải tối ưu hoá cùng một lúc nhiều mục tiêu. Tuy

nhiên, các mục tiêu này thường đối chọi cạnh tranh với nhau. Việc làm tốt hơn mục tiêu này thường dẫn tới việc làm xấu đi một số mục tiêu khác. Vì vậy việc giải các bài toán tối ưu đa mục tiêu, tức là tìm ra một phương án khả thi tốt nhất theo một nghĩa nào đó, thực chất chính là một bài toán ra quyết định. Có thể thấy lại ở đây một lần nữa khẳng định "Tối ưu hoá chính là công cụ định lượng chủ yếu nhất của quá trình ra quyết định". Khái niệm then chốt trong tối ưu hoá đa mục tiêu là **khái niệm phương án tối ưu Pareto**.

Định nghĩa 1. Một phương án tối ưu Pareto X^* có tính chất sau đây:

- Trước hết nó phải thuộc vào miền các phương án khả thi của bài toán, tức là phải thoả mãn tất cả các ràng buộc: $X^* \in D$.

- Với mọi phương án khả thi khác $X \in D$ mà có một mục tiêu nào đó tốt hơn ($f_i(X)$ tốt hơn $f_i(X^*)$) thì cũng phải có ít nhất một mục tiêu khác xấu hơn ($f_j(X)$ xấu hơn $f_j(X^*)$, $j \neq i$).

Nói một cách khác, không tồn tại một phương án khả thi nào $X \in D$ có thể trội hơn X^* trên tổng thể.

Định nghĩa 2. Giải bài toán tối ưu toàn cục đa mục tiêu là chọn ra từ tập hợp P các phương án tối ưu Pareto của bài toán một (hoặc một số) phương án tốt nhất (thoả mãn nhất) theo một nghĩa nào đó dựa trên cơ cấu ưu tiên của người ra quyết định.

Cách 1: Bằng một phương pháp tối ưu toán học thích hợp tìm ra tập hợp P tất cả các phương án tối ưu Pareto. Người ra quyết định sẽ đề ra cơ cấu ưu tiên của mình đối với tập P nhằm tìm ra phương án tối ưu Pareto thoả mãn nhất cho bài toán đa mục tiêu ban đầu.

Cách 2: Việc tìm tập hợp P trong trường hợp các bài toán nhiều biến là khá khó và mất nhiều thời gian. Vì vậy, so với cách 1, cách 2 sẽ tiến hành theo trình tự ngược lại. Trước hết người ra quyết định sẽ đề ra cơ

cấu ưu tiên của mình. Dựa vào cơ cấu ưu tiên đó, các mục tiêu sẽ được tổ hợp vào một mục tiêu duy nhất, tiêu biểu cho hàm tổng tiện ích của bài toán. Bài toán tối ưu với hàm mục tiêu tổ hợp này sẽ được giải bằng một phương pháp tối ưu toán học thích hợp, để tìm ra một (một số) phương án tối ưu Pareto. Lúc này, người ra quyết định sẽ chọn ra trong số các phương án tối ưu Pareto đó một phương án tốt nhất.

Định nghĩa 3. Phương pháp giải bài toán tối ưu đa mục tiêu dựa trên sự trợ giúp của hệ máy tính, nhằm giúp người ra quyết định từng bước thay đổi các quyết định trung gian một cách thích hợp để đi tới một phương án tối ưu Pareto thoả mãn nhất, được gọi là phương pháp tương tác người - máy tính.

Phương pháp tương tác người - máy tính giải bài toán tối ưu đa mục tiêu có các yếu tố cấu thành sau:

- Cơ cấu ưu tiên của người ra quyết định và hàm tổ hợp tương ứng.

- Kiểu tương tác người - máy tính: cho biết các thông tin nào máy tính phải đưa ra lại trong các bước lặp trung gian, và cách thay đổi các thông số của cơ cấu ưu tiên từ phía người ra quyết định.

- Kỹ thuật tối ưu toán học được xây dựng dựa trên lý thuyết tối ưu hoá nhằm tìm ra các phương án tối ưu Pareto cho các bài toán cần giải trong các bước lặp trung gian.

Cho tới thời điểm hiện nay, hàng chục phương pháp giải BTQHHT đa mục tiêu đã được đề cập tới trong các tạp chí chuyên ngành, mà đa số chúng đều có những ứng dụng rất thành công trong nhiều lĩnh vực, như: phương pháp tham số, phương pháp nón pháp tuyến, phương pháp véc tơ cực đại, phương pháp trọng số tương tác của Chebisev, phương pháp thoả dụng mờ tương tác người - máy tính của Nguyễn Hải Thanh. Sau đây là thuật giải dựa trên phương pháp thoả dụng mờ tương tác người máy tính:

Bước khởi tạo

- Nhập dữ liệu cho các hàm mục tiêu tuyến tính z_i ($i=1, 2, \dots, p$) và m điều kiện ràng buộc.

- Giải bài toán quy hoạch tuyến tính cho từng hàm mục tiêu i ($i=1, 2, \dots, p$) với m ràng buộc ban đầu, thu được các phương án tối ưu X^1, X^2, \dots, X^p (nếu với một mục tiêu nào đó bài toán không cho phương án tối ưu thì cần xem xét để chỉnh sửa lại các điều kiện ràng buộc ban đầu).

- Tính giá trị hàm mục tiêu tại p phương án X^1, X^2, \dots, X^p .

Lập bảng pay-off. Xác định giá trị cận trên z_i^B và giá trị cận dưới z_i^W của mục tiêu z_i ($i=1, 2, \dots, p$).

- Xác định các hàm thoả dụng mờ $\mu_1(z_1), \mu_2(z_2), \dots, \mu_p(z_p)$ cho từng mục tiêu dựa vào thông tin từ bảng pay-off theo công thức:

$$\mu_i(z_i) = \frac{z_i - z_i^W}{z_i^B - z_i^W}, \quad i = \overline{1, p}.$$

- Đặt $k = 1$.

Các bước lặp (xét bước lặp thứ k)

Bước 1: Xây dựng hàm mục tiêu tổ hợp từ các hàm thoả dụng trên: $w_1\mu_1(z_1) + w_2\mu_2(z_2) + \dots + w_p\mu_p(z_p) \rightarrow \text{Max}$

Trong đó: w_1, w_2, \dots, w_p là các trọng số phản ánh tầm quan trọng của từng hàm thoả dụng trong thành phần hàm tổ hợp, với

$$w_1 + w_2 + \dots + w_p = 1 \text{ và } 0 \leq w_1, w_2, \dots, w_p \leq 1.$$

Bước 2:

- Giải BTQHHT với hàm mục tiêu tổ hợp và m ràng buộc ban đầu để tìm được phương án tối ưu của bước lặp thứ k là $X^{(k)}$ và giá trị của các hàm mục tiêu z_i cũng như của các hàm thoả dụng $\mu_i(z_i)$ (với $i=1, 2, \dots, p$).

- Nếu người ra quyết định cảm thấy chưa thoả mãn với các giá trị đạt được của các hàm mục tiêu cũng như của các hàm thoả dụng thì phương án thu được $X^{(k)}$ chưa

phải là phương án tối ưu thoả mãn nhất. Thay k bởi $k+1$, quay về bước 1.

- Nếu người ra quyết định đã cảm thấy thoả mãn thì phương án thu được là $X^{(k)}$. Chuyển sang bước 3.

Bước 3: Kết thúc.

3. PHÂN LOẠI DỮ LIỆU MỜ

Bài toán phân loại dữ liệu từ lâu đã chiếm vai trò quan trọng trong các ngành khoa học. Từ tập hợp dữ liệu ban đầu mỗi chuyên ngành có những đòi hỏi khác nhau đối với việc phân loại các dữ liệu đó tùy thuộc vào yêu cầu cụ thể. Các dữ liệu ban đầu qua điều tra thu thập được của mỗi cá thể trong tập mẫu có thể có nhiều đặc tính (định tính hoặc định lượng). Trong các chuyên ngành kinh tế lượng, quản trị kinh doanh, cơ khí nông nghiệp; hay các chuyên ngành nông nghiệp như chăn nuôi, trồng trọt, chọn giống, quản lý đất đai, nông hoá thổ nhưỡng, kinh tế nông nghiệp, sinh thái..., bài toán phân loại dữ liệu cũng được quan tâm nghiên cứu (Kaufmann A. và cs, 1991; Helmuth Spoth, 1982; Nguyễn Hải Thanh và cs, 1999).

Sau đây chúng tôi trình bày một phương pháp phân loại xấp xỉ dựa trên tiêu chuẩn khoảng cách (trọng tâm) cực tiểu và bình phương bé nhất (Helmuth Spoth, 1982; Lê Đức Vĩnh, 1997). Trước hết, chúng ta đề cập tới việc phân loại các dữ liệu số (số liệu) thông thường như chúng ta quen biết trong đa số các chuyên ngành nghiên cứu hiện nay. Các dữ liệu như vật sẽ được gọi là các dữ liệu rõ (*crisp data*). Khi thu thập dữ liệu, ta thường tiến hành phương pháp chọn mẫu gồm n cá thể. Bằng cách định lượng hoá các đặc tính của các cá thể đó, mỗi cá thể sẽ ứng với một bộ m số tương ứng với m đặc tính được xem xét. Bài toán phân loại đặt ra ở đây là đưa vào một khái niệm hàm khoảng cách d thích hợp nhằm đánh giá "độ gần gũi" giữa các cá thể đó, từ đó có thể xem xét và đề xuất các phương pháp phân loại phù hợp.

Giả sử ta đã quyết định phân hoạch tập mẫu A ra một lớp. Dựa trên hàm khoảng cách đã đề ra, cần tìm được phân hoạch tối ưu của tập mẫu A theo một nghĩa nào đó. Sau đây là hai tiêu chuẩn quan trọng, tiêu chuẩn khoảng cách (trọng tâm) cực tiểu và tiêu chuẩn bình phương bé nhất thường được dùng trong thực tế.

Cho $A = \{ a_1, a_2, \dots, a_n \}$ gồm n cá thể, mỗi cá thể là một véc tơ m chiều $a_i = (x_{i1}, x_{i2}, \dots, x_{im}) \in R^m$ ($i = 1, \dots, n$) mô tả m đặc tính. Gọi d là một khoảng cách (*metric*) trong A , $l \in N$.

Tiêu chuẩn bình phương bé nhất: Giả sử $C = \{ C_1, C_2, \dots, C_l \}$ là một phân hoạch của tập hợp A , n_i là số phần tử của lớp $C_i = \{ a_{ij} \in A / \forall j = \overline{1, n_i} \}$ trong phân hoạch C .

$\bar{a}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} a_{ij}$ là trọng tâm của lớp C_i . Đặt

$l(C_i) = \sum_{j=1}^{n_i} d^2(a_{ij}, \bar{a}_i)$ thì $D(C) =$

$\sum_{i=1}^l l(C_i) = \sum_{i=1}^l \sum_{j=1}^{n_i} d^2(a_{ij}, \bar{a}_i)$. Phân hoạch

$C_0 = \{ C_1^0, C_2^0, \dots, C_l^0 \}$ thoả mãn $D(C_0) = \min D(C)$ được gọi là phân hoạch tối ưu của A theo nghĩa bình phương bé nhất.

Tiêu chuẩn khoảng cách (trọng tâm) cực tiểu: Phân hoạch $C_0 = \{ C_1^0, C_2^0, \dots, C_l^0 \}$ được gọi là phân hoạch với khoảng cách cực tiểu nếu $\forall a_{ij}^{(0)} \in C_i^0$ thì $d(a_{ij}^{(0)}, \bar{a}_i) < d(a_{ij}^{(0)}, \bar{a}_r) \forall r \neq i$. Như vậy nếu C^0 là phân hoạch với khoảng cách cực tiểu của A thì các cá thể trong cùng một lớp "gần" với trọng tâm của lớp đó hơn trọng tâm của lớp khác. Định lý sau đã được chứng minh trong nhiều tài liệu.

Định lý. Tập các phân hoạch khoảng cách cực tiểu chứa tập các phân hoạch bình phương bé nhất.

Trong thực tế, tập mẫu A thường có kích thước n tương đối lớn, nên số các phân hoạch gồm l lớp của A cũng rất lớn. Vì vậy thuật giải chính xác dựa trên các tiêu

chuẩn trình bày ở trên thường tỏ ra ít hiệu quả. Do đó, người ta thường dùng các thuật giải xấp xỉ để tìm ra các phân hoạch xấp xỉ phân hoạch tối ưu. Sau đây chúng tôi trình bày thuật giải xấp xỉ dựa trên các tiêu chuẩn khoảng cách cực tiểu và bình phương bé nhất (Helmuth Spoth, 1982).

Bước khởi tạo

- Chọn số lớp l và chọn ngẫu nhiên một phân hoạch khởi đầu $C_0 = \{C_1^{(0)}, C_2^{(0)}, \dots, C_l^{(0)}\}$, lớp $C_i^{(0)}$ có n_i phần tử, $\sum_{i=1}^l n_i = n$.

- Chọn sai số ε .

- Chọn biến đếm k , bắt đầu từ 0 và chọn hằng $k_{\max} > 0$.

Các bước lặp

Bước 1:

Với $C_i^{(k)} = \{a_{i1}^{(k)}, \dots, a_{i, n_i}^{(k)}\}$, tính

$$\bar{a}_i^{(k)} = \frac{1}{n_i^{(k)}} \sum_{j=1}^{n_i} a_{ij}^{(k)} \quad \forall i = 1, 2, \dots, l.$$

Bước 2:

Tìm một phân hoạch khoảng cách cực tiểu $C_{k+1} = \{C_1^{(k+1)}, C_2^{(k+1)}, \dots, C_l^{(k+1)}\}$ với $C_i^{(k+1)} =$

$$\{a_{ij}^{(k+1)} : d(a_{ij}^{(k+1)}, \bar{a}_i^{(k)}) = \min d(a_{ij}^{(k+1)}, \bar{a}_r^{(k)}), \forall r=1, l\}, i = 1, 2, \dots, l.$$

Bước 3:

Nếu $D(C_{k+1}) < D(C_k) - \varepsilon$ và $k < k_{\max}$ thì thay k bởi $k+1$ rồi quay về bước 1.

Nếu trái lại (các trường hợp khác), dừng và in ra phân hoạch xấp xỉ tối ưu.

Trong phương pháp phân loại dữ liệu trên đây chúng ta có thể sử dụng một số hàm khoảng cách như khoảng cách σ -cơ-lit, khoảng cách có trọng số hoặc xây dựng một hàm khoảng cách phù hợp với đối tượng dữ liệu cần nghiên cứu.

Phân loại dữ liệu mờ dựa trên hàm khoảng cách θ

Ngày nay với sự hỗ trợ đắc lực của máy tính điện tử, nhiều phương pháp phân loại dữ liệu trên cơ sở toán học đã được đề ra, từ đó xây dựng được các thuật giải phân loại có hiệu quả cao. Tuy vậy, trong nhiều tình huống thu thập, xử lý, quản lý dữ liệu và ra quyết định, các phương pháp phân loại đối với các dữ liệu thông thường (dữ liệu rõ) như đã biết cũng đã tỏ ra thiếu phù hợp. Đó là vì, các dữ liệu thu thập được trên thực tế thường chứa đựng các độ bất ổn định, được phân chia ra hai loại chính:

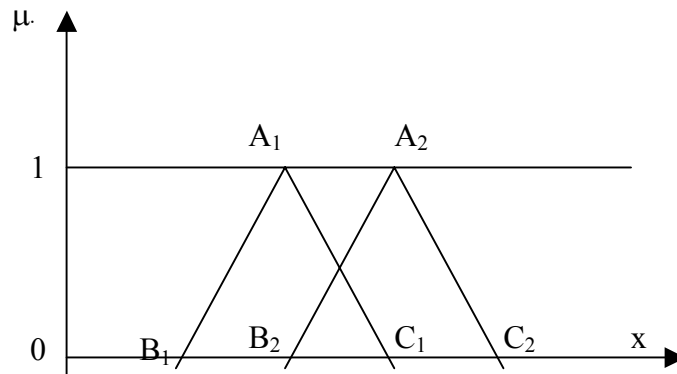
- Độ nhòe phản ánh tính bất ổn định khách quan (*ambiguity*) luôn hàm chứa trong các dữ liệu thu thập từ thực tế.

- Độ nhòe phản ánh định tính chủ quan (*vagueness*) xuất hiện từ phía người thu thập xử lý dữ liệu).

Lý thuyết tập mờ được sử dụng để tìm cách định lượng các độ nhòe - bất ổn định - định tính luôn tiềm tàng trong các dữ liệu như đã nói ở trên (các dữ liệu như vậy sẽ được gọi là dữ liệu mờ - *fuzzy data*) cũng như để tìm ra các hàm khoảng cách thích hợp đối với các dữ liệu đó nhằm đưa ra các phương pháp và thuật giải phân loại thích hợp (Nguyễn Hải Thanh và cs, 1999).

Định nghĩa 4. Xét hai số mờ dạng tam giác \tilde{a}_1 cho bởi $L_1(x)$ và $R_1(x)$, \tilde{a}_2 cho bởi $L_2(x)$ và $R_2(x)$. Lúc đó khoảng cách d giữa \tilde{a}_1, \tilde{a}_2 cho bởi công thức sau:

$$d(\tilde{a}_1, \tilde{a}_2) = \left| \int_0^1 [L_2^{-1}(y) - L_1^{-1}(y)] dy + \int_0^1 [R_2^{-1}(y) - R_1^{-1}(y)] dy \right|$$



Hình 1. Tính khoảng cách giữa hai số mờ

Trong trường hợp hai số \tilde{a}_1, \tilde{a}_2 như thể hiện trên hình 1 thì:

$$d(\tilde{a}_1, \tilde{a}_2) = S(A_1B_1B_2A_2) + S(A_1C_1C_2A_2).$$

Một cách tổng quát, ta có:

$$d(\tilde{a}_1, \tilde{a}_2) = \left| (x_{A_1} - x_{A_2}) + \frac{1}{2}(x_{B_1} - x_{B_2}) + \frac{1}{2}(x_{C_1} - x_{C_2}) \right|.$$

Có thể coi đây chính là hai lần khoảng cách giữa trọng tâm của các tứ giác $A_1B_1C_1D_1$ và $A_2B_2C_2D_2$, với D_1 rất sát gần A_1 và D_2 rất sát gần A_2 .

Rõ ràng hàm khoảng cách trên có các tính chất sau:

Tính chất 1. $d(\tilde{a}_1, \tilde{a}_2) = d(\tilde{a}_2, \tilde{a}_1) (\geq 0)$

Tính chất 2. $d(\tilde{a}_1, \tilde{a}_3) \leq d(\tilde{a}_1, \tilde{a}_2) + d(\tilde{a}_2, \tilde{a}_3)$
 $\forall \tilde{a}_1, \tilde{a}_2, \tilde{a}_3.$

Trong một số trường hợp, có thể dùng khoảng cách quy chuẩn giữa hai số mờ. $d^*(\tilde{a}_1, \tilde{a}_2) = kd(\tilde{a}_1, \tilde{a}_2)$ với k là hệ số quy chuẩn sao cho $\forall \tilde{a}_1, \tilde{a}_2$ thì $0 \leq d^*(\tilde{a}_1, \tilde{a}_2) \leq 1$.

Chú ý rằng, trái với khoảng cách thông thường, với hai số mờ $\tilde{a}_1 \neq \tilde{a}_2$ vẫn có thể có $d(\tilde{a}_1, \tilde{a}_2) = 0$. Như vậy, *Tính chất 3:* $d(\tilde{a}_1, \tilde{a}_2) = 0$ khi và chỉ khi $\tilde{a}_1 = \tilde{a}_2$ không được đảm bảo.

Định nghĩa 5. Xét hai số mờ dạng tam giác $\tilde{a}_1 = (x_{B_1}, x_{A_1}, x_{C_1})$, $\tilde{a}_2 =$

$(x_{B_2}, x_{A_2}, x_{C_2})$. Lúc đó khoảng cách θ giữa \tilde{a}_1, \tilde{a}_2 sẽ được xác định bởi công thức sau:

$$\theta(\tilde{a}_1, \tilde{a}_2) = \frac{1}{4} |2(x_{A_1} - x_{A_2}) + (x_{B_1} - x_{B_2}) + (x_{C_1} - x_{C_2})| + \frac{1}{8} (2|x_{A_1} - x_{A_2}| + |x_{B_1} - x_{B_2}| + |x_{C_1} - x_{C_2}|)$$

Lúc đó, biểu thức thứ nhất ở vế phải có thể được coi là khoảng cách giữa các trọng tâm của hai tứ giác $A_1B_1C_1D_1$ và $A_2B_2C_2D_2$, với D_1 rất sát gần A_1 và D_2 rất sát gần A_2 . Tuy nhiên, có thể xây dựng được các tứ giác $A_1B_1C_1D_1$ và $A_2B_2C_2D_2$, với D_1 rất sát gần A_1 và D_2 rất sát gần A_2 , khá khác biệt nhau mà vẫn có trọng tâm trùng nhau. Biểu thức thứ hai ở vế phải cho phép tính đến các khác biệt giữa hai tam giác $A_1B_1C_1$ và $A_2B_2C_2$ khi xác định khoảng cách giữa hai số mờ đã cho. Để dàng kiểm tra được rằng, hàm khoảng cách trong định nghĩa 5 có đầy đủ cả ba tính chất 1, 2 và 3.

4. PHƯƠNG PHÁP DELPHY THÔNG NHẤT Ý KIẾN CHUYÊN GIA

Vấn đề lượng hóa ý kiến chuyên gia

Ý kiến chuyên gia khi đánh giá một cách tổng hợp một phương án (quy hoạch sử dụng đất) nào đó được cho ở các mức: rất tốt, tốt, khá phù hợp, không phù hợp, kém hiệu quả, không nên triển khai. Điều này cho phép các chuyên gia đưa ra ý kiến một cách tương đối dễ dàng, nhất là khi so sánh với việc các chuyên gia phải đưa ra một giá trị điểm để đánh giá phương án. Giả sử các mức đó có thể minh họa bằng các số mờ một chiều là: (0.9, 0.95, 1.0), (0.7, 0.8, 0.9), (0.5, 0.6, 0.7), (0.3, 0.4, 0.5), (0.1, 0.2, 0.3) và (0, 0.05, 0.1). Chẳng hạn, (0.9, 0.95, 1.0) có nghĩa là phương án được đánh giá rất tốt, thỏa mãn được từ 90% tới 100% mong muốn của chuyên gia, mà trong dải này thì 95% mức mong muốn là đạt được với khả năng nhiều nhất.

Ý kiến của các chuyên gia về một phương án quy hoạch sử dụng đất nào đó, nói chung, là không thống nhất. Thậm chí đôi khi là khá xa nhau, đối lập. Điều này là do mỗi một phương án thường đi kèm nhiều chỉ tiêu mà phương án đó đạt được (xem mục 2: bài toán quy hoạch tuyến tính đa mục tiêu). Đánh giá một cách tổng hợp đồng thời tất cả các chỉ tiêu đó rất khó thống nhất ngay từ đầu. Hơn nữa, mỗi một chuyên gia cũng không thể đưa ra đánh giá chính xác ngay, mặc dù họ nắm khá rõ các chỉ tiêu chuyên môn cần đánh giá. Bởi vậy, cần xây dựng một quá trình đánh giá tập thể gồm nhiều bước lặp để:

- Tận dụng được tri thức của các chuyên gia trong nhiều lĩnh vực khác nhau, với các nhận biết, cảm nhận khác nhau về cùng một vấn đề.

- Giúp các chuyên gia cân nhắc và xem xét sửa chỉnh lại các đánh giá mang tính chủ quan của mình.

- Đưa ra một quy trình giá khách quan nhằm phân loại các ý kiến trong từng

bước, và làm cho các ý kiến hội tụ về một cách đánh giá thống nhất.

Thuật toán Delphy thống nhất ý kiến chuyên gia

Bước khởi tạo

- Xin ý kiến n chuyên gia đánh giá một phương án ở các mức: rất tốt, tốt, khá phù hợp, không phù hợp, kém hiệu quả, không nên triển khai.

- Chọn $l = 3$ hoặc 4 lớp để phân hoạch ý kiến các chuyên gia.

- Chọn k_{\max} là số bước lặp tối đa cần thực hiện (thông thường chọn $k_{\max} = 10$ đến 15). Đặt $k = 1$.

Các bước lặp

Bước 1: Sử dụng phương pháp phân loại dữ liệu căn cứ vào *thuật giải xấp xỉ* dựa trên các tiêu chuẩn khoảng cách cực tiểu và bình phương bé nhất đã biết.

Bước 2:

- Nếu có ít nhất 75% ý kiến chuyên gia trong một lớp nào đó thì chuyển sang bước 3.

- Nếu có chưa tới 75% ý kiến chuyên gia trong cùng một lớp nào đó, nhưng $k+1 > k_{\max}$ thì cũng chuyển sang bước 3.

- Nếu trái lại thì thông báo cho các chuyên gia ý kiến trung bình của lớp (đã được quy về mức định tính gần nhất trong số 6 mức định tính đã đưa ra) có nhiều ý kiến tập chung nhất.

- Xin các chuyên gia sửa chỉnh lại ý kiến của mình căn cứ thông báo trên và chuyển về bước 1.

Bước 3: Thông báo cho các chuyên gia biết ý kiến trung bình của tất cả các ý kiến thuộc các lớp có chứa ít nhất 2 hoặc 3 ý kiến (các lớp có số ý kiến > 1 và $> 0,1n$). ý kiến trung bình này được lấy làm ý kiến thống nhất của nhóm chuyên gia.

Chú ý. Nếu ta có ý kiến thống nhất của một nhóm chuyên gia về nhiều phương án

quy hoạch sử dụng đất, thì ta có thể so sánh các ý kiến đó để xếp loại các phương án, và sau đó giữ lại một số phương án tốt nhất (thông thường là 3 phương án).

Để so sánh hai phương án A và B nào đó với các ý kiến đánh giá đã được thống nhất $(x_1, y_1, z_1), (x_2, y_2, z_2)$, cần tính chỉ số

$$C_f = 0.5(x_1 - x_2 + z_1 - z_2) + (y_1 - y_2)$$

Nếu $C_f > 0$ thì phương án A được coi là tốt hơn phương án B, nếu $C_f < 0$ thì A xấu hơn B, còn nếu $C_f = 0$ thì A và B ngang nhau.

5. XÂY DỰNG PHẦN MỀM MÁY TÍNH HỖ TRỢ RA QUYẾT ĐỊNH TẬP THỂ

Có thể nhận thấy rằng, các cơ sở dữ liệu tài nguyên đất đai là rất lớn, luôn cần được bổ sung, cập nhật cũng như xử lý. Nhiều quyết định quy hoạch sử dụng đất có thể được xây dựng dựa trên cơ sở áp dụng phương pháp mô hình hoá bao gồm các mô hình phân tích/cơ chế và các mô hình tiện dụng cũng như các phương pháp toán học đa dạng như: xử lý thống kê, tối ưu hoá, lý thuyết ra quyết định, mô phỏng ngẫu nhiên... Chính vì vậy, các phần mềm tính toán khoa học và ra quyết định cần được tích hợp thành một hệ phần mềm hỗ trợ ra quyết định để xử lý các dữ liệu thu thập được. Hệ hỗ trợ ra quyết định trong quy hoạch sử dụng đất bao gồm các thành phần cơ bản sau đây:

- *Thành phần 1:* Cơ sở dữ liệu tài nguyên đất đai được thu thập, tổng hợp và xử lý từ nhiều nguồn khác nhau (chủ yếu từ các phần mềm GIS như Map Infor/ArcInfor, Arcview ...).

- *Thành phần 2:* Các phần mềm tối ưu giải các bài toán tối ưu tuyến tính và phi tuyến, đơn mục tiêu và đa mục tiêu.

- *Thành phần 3:* Hệ phần mềm xử lý ý kiến nông dân và thống nhất ý kiến chuyên gia (hay còn gọi là hệ phần mềm hỗ trợ ra quyết định tập thể).

Xây dựng các module của phần mềm máy tính hỗ trợ ra quyết định tập thể

Trong bài báo này, chúng tôi chỉ tập trung vào thành phần thứ ba của hệ hỗ trợ ra quyết định quy hoạch sử dụng đất và tích hợp thành phần này với thành phần thứ hai đã được nghiên cứu trước đây. Trên cơ sở đó, một phần mềm máy tính đã được thiết lập nhằm xử lý bài toán tối ưu đa mục tiêu trong quy hoạch sử dụng đất và hỗ trợ ra quyết định tập thể.

Phần mềm này bao gồm hai mô đun (module) sau đây

Mô đun 1: Giải bài toán quy hoạch tuyến tính đa mục tiêu bằng phương pháp thoả dụng mờ tương tác người - máy tính với nhiều bộ trọng số nhằm tạo ra các phương án quy hoạch sử dụng đất.

Mô đun 2: Phân loại ý kiến và thực hiện quy trình ra quyết định tập thể nhằm thống nhất ý kiến chuyên gia bằng phương pháp Delphy. Ba phương án quy hoạch sử dụng đất có điểm trung bình cao nhất được giữ lại để nghiên cứu triển khai trên thực tế.

Phần mềm này được chạy kiểm thử cho các số liệu của bài toán quy hoạch sử dụng đất xã Trâu Quỳ, năm 1999 với năm mục tiêu và các ràng buộc sau đây:

Các mục tiêu cần cực đại hóa

Tổng lợi nhuận

$$Z_1 = 4,48x_1 + 4,2x_2 + 2,59x_3 + 0,98x_4 + 5,8x_5 +$$

$$15,61x_6 + 29,67x_7 + 39,21x_8 + 116,58x_9 + 105,13x_{10}$$

Hiệu quả sử dụng vốn

$$Z_2 = 0,6205x_1 + 0,5915x_2 + 0,465x_3 + 0,1583x_4 + 0,7065x_5 +$$

$$0,5964x_6 + 1,2996x_7 + 1,2735x_8 + 1,1726x_9 + 1,756x_{10}$$

Giá trị ngày công

$$Z_3 = 0,0217x_1 + 0,0206x_2 + 0,0154x_3 + 0,00458x_4 + 0,0248x_5 +$$

$$0,0109x_6 + 0,0241x_7 + 0,0349x_8 + 0,09x_9 + 0,0811x_{10}$$

Tổng số công lao động

$$Z_4 = 206x_1 + 204x_2 + 168x_3 + 216x_4 + 234x_5 +$$

$$1428x_6 + 1232x_7 + 1124x_8 + 1296x_9 + 1296x_{10}$$

Hiệu quả môi trường

$$Z_5 = 0,7x_1 + 0,778x_2 + 1,273x_3 + 1,75x_4 + x_5 + 0,368x_6 + 0,875x_7 + 3x_8 + 3x_9 + 3x_{10}$$

Với các biến quyết định sau: x_1 = diện tích trồng lúa xuân (ha), x_2 = diện tích trồng lúa mùa (ha), x_3 = diện tích trồng ngô (ha), x_4 = diện tích trồng đậu tương (ha), x_5 = diện tích trồng khoai tây (ha), x_6 = diện tích trồng rau (ha), x_7 = diện tích trồng mùi (ha), x_8 = diện tích trồng táo (ha), x_9 = diện tích trồng nhãn (ha), x_{10} = diện tích trồng xoài (ha).

Các ràng buộc

Các ràng buộc về giới hạn diện tích:

$$x_1 \leq 189,6407; x_2 \leq 189,6407.$$

Các ràng buộc tương quan tỷ lệ:

$$x_6 \geq 26,4; x_8 + x_9 + x_{10} = 18;$$

$$x_3 + x_4 + x_5 + x_6 + x_7 = 43,8931.$$

Ràng buộc về tổng sản lượng lương thực:

$$5,14x_1 + 4,98x_2 + 3,77x_3 \geq 1700,5.$$

Như vậy đây là bài toán quy hoạch tuyến tính năm mục tiêu và sáu ràng buộc (không kể điều kiện không âm của các biến quyết định).

Kết quả thực hiện module 1

CHƯƠNG TRÌNH QUY HOẠCH TUYẾN TÍNH

SO BIEN : 10 ; SO RANG BUOC : 6

$$Z[1] = 4.4800 X1 + 4.2000 X2 + 2.5900 X3 + 0.9800 X4 + 5.8000 X5 + 15.6100 X6 + 29.6700 X7 + 39.2100 X8 + 116.5800 X9 + 105.1300 X10$$

Nghiệm tối ưu tìm thấy sau: 7 Bước lặp

$$\text{PHUONG AN TOI UU (X[1]): } X1 = 189.6407; X2 = 189.6407; X6 = 26.4000; X7 = 17.4931; X9 = 18.0000; \text{ Các biến khác bằng không. CUC DAI : } 4675.645553$$

$$Z[2] = 0.6205 X1 + 0.5915 X2 + 0.4650 X3 + 0.1583 X4 + 0.7065 X5 + 0.5864 X6 + 1.2996 X7 + 1.2735 X8 + 1.1726 X9 + 1.7560 X10$$

Nghiệm tối ưu tìm thấy sau: 6 Bước lặp

$$\text{PHUONG AN TOI UU (X[2]): } X1 = 189.6407; X2 = 189.6407; X6 = 26.4000; X7 = 17.4931; X10 = 18.0000; \text{ Các biến khác bằng không. CUC DAI : } 299.667521$$

$$Z[3] = 0.0217 X1 + 0.0206 X2 + 0.0154 X3 + 0.0045 X4 + 0.0248 X5 + 0.0109 X6 + 0.0241 X7 + 0.0349 X8 + 0.0900 X9 + 0.0811 X10$$

Nghiệm tối ưu tìm thấy sau: 6 Bước lặp

$$\text{PHUONG AN TOI UU (X[3]): } X1 = 189.6407; X2 = 189.6407; X5 = 17.4931; X6 = 26.4000; X9 = 18.0000; \text{ Các biến khác bằng không. CUC DAI : } 10.363390$$

$$Z[4] = 206.0000 X1 + 204.0000 X2 + 168.0000 X3 + 216.0000 X4 + 234.0000 X5 + 1428.0000 X6 + 1232.0000 X7 + 1124.0000 X8 + 1296.0000 X9 + 1296.0000 X10$$

Nghiệm tối ưu tìm thấy sau: 6 Bước lặp

$$\text{PHUONG AN TOI UU (X[4]): } X1 = 189.6407; X2 = 189.6407; X6 = 43.8931; X9 = 18.0000; \text{ Các biến khác bằng không. CUC DAI : } 163760.033800$$

$$Z[5] = 0.7000 X1 + 0.7780 X2 + 1.1273 X3 + 1.7500 X4 + 1.0000 X5 + 0.3680 X6 + 0.8750 X7 + 3.0000 X8 + 3.0000 X9 + 3.0000 X10$$

Nghiệm tối ưu tìm thấy sau: 6 Bước lặp

$$\text{PHUONG AN TOI UU (X[5]): } X1 = 189.6407; X2 = 189.6407; X4 = 17.4931; X6 = 26.4000; X8 = 18.0000; \text{ Các biến khác bằng không. CUC DAI : } 374.617080$$

BANG PAY-OFF

	Z[1]	Z[2]	Z[3]	Z[4]	Z[5]
X[1]	4675.645553	289.166321	10.351145	160331.386200	359.310617
X[2]	4469.545553	299.667521	10.190945	160331.386200	359.310617
X[3]	4258.085256	278.791164	10.363390	142873.272400	361.497255
X[4]	4429.692567	276.690242	10.120236	163760.033800	350.441615
X[5]	2781.108514	271.017646	9.016481	139462.396600	374.617080

Gia tri Max - Min tung muc tieu

MAX[1] =	4675.645553	MIN[1] =	2781.108514
MAX[2] =	299.667521	MIN[2] =	271.017646
MAX[3] =	10.363390	MIN[3] =	9.016481
MAX[4] =	163760.033800	MIN[4] =	139462.396600
MAX[5] =	374.617080	MIN[5] =	350.441615

KET QUA HAM LIEN HOP

Gia tri cac trong so- lan thu 1: $w[1] = 0.2; w[2] = 0.2; w[3] = 0.5; w[4] = 0.1; w[5] = 0;$

HAM MUC TIEU LIEN HOP 1

$$Z = 0.013708 X_1 + 0.013059 X_2 + 0.009928 X_3 + 0.003768 X_4 + 0.015714 X_5 + 0.015665 X_6 + 0.026221 X_7 + 0.030611 X_8 + 0.059236 X_9 + 0.058796 X_{10}$$

Phan le = 6.106595

PHUONG AN TOI UU LIEN HOP 1: $X_1 = 189.6407; X_2 = 189.6407; X_6 = 26.4000; X_7 = 17.4931; X_9 = 18.0000;$ Cac bien khac bang khong. CUC DAI : 7.014631

Gia tri cua cac ham thoa dung - Lan thu 1

Z[1] =	4675.645553	pZ[1] =	1.000000
Z[2] =	289.166321	pZ[2] =	0.633464
Z[3] =	10.351145	pZ[3] =	0.990909
Z[4] =	160331.386200	pZ[4] =	0.858890
Z[5] =	359.310617	pZ[5] =	0.366860

Gia tri cac trong so lan thu 2: $w[1] = 0.3; w[2] = 0.3; w[3] = 0.4; w[4] = 0; w[5] = 0;$

HAM MUC TIEU LIEN HOP 2

$$Z = 0.012011 X_1 + 0.011431 X_2 + 0.008799 X_3 + 0.003021 X_4 + 0.013861 X_5 + 0.014354 X_6 + 0.030511 X_7 + 0.032573 X_8 + 0.051232 X_9 + 0.058391 X_{10}$$

Phan le = 5.463804

PHUONG AN TOI UU LIEN HOP 2: $X_1 = 189.6407; X_2 = 189.6407; X_6 = 26.4000; X_7 = 17.4931; X_{10} = 18.0000;$ Cac bien khac bang khong. CUC DAI : 6.409410

Gia tri cua cac ham thoa dung - Lan thu 2

Z[1] =	4469.545553	pZ[1] =	0.891214
Z[2] =	299.667521	pZ[2] =	1.000000
Z[3] =	10.190945	pZ[3] =	0.871970
Z[4] =	160331.386200	pZ[4] =	0.858890
Z[5] =	359.310617	pZ[5] =	0.366860

Gia tri cac trong so lan thu 3: $w[1] = 0; w[2] = 0; w[3] = 0.1; w[4] = 0.8; w[5] = 0.1;$

HAM MUC TIEU LIEN HOP 3

$$Z = 0.011289 X_1 + 0.011464 X_2 + 0.011338 X_3 + 0.014685 X_4 + 0.013682 X_5 + 0.049348 X_6 + 0.045972 X_7 + 0.052008 X_8 + 0.061762 X_9 + 0.061101 X_{10}$$

Phan le = 6.710796

PHUONG AN TOI UU LIEN HOP 3: $X_1 = 189.6407$; $X_2 = 189.6407$; $X_6 = 43.8931$; $X_9 = 18.0000$; Cac bien khac bang khong. CUC DAI : 7.592744

Gia tri cua cac ham thoa dung - Lan thu 3

$$\begin{aligned} Z[1] &= 4429.692567 & pZ[1] &= 0.870178 \\ Z[2] &= 276.690242 & pZ[2] &= 0.197997 \\ Z[3] &= 10.120236 & pZ[3] &= 0.819473 \\ Z[4] &= 163760.033800 & pZ[4] &= 1.000000 \\ Z[5] &= 350.441615 & pZ[5] &= 0.000000 \end{aligned}$$

Gia tri cac trong so lan thu 4: $w[1] = 0.1$; $w[2] = 0.1$; $w[3] = 0.1$; $w[4] = 0.1$; $w[5] = 0.6$;

HAM MUC TIEU LIEN HOP 4

$$Z = 0.022234 X_1 + 0.023964 X_2 + 0.031572 X_3 + 0.045260 X_4 + 0.030395 X_5 + 0.018690 X_6 + 0.034678 X_7 + 0.088187 X_8 + 0.096718 X_9 + 0.097489 X_{10}$$

Phan le = 11.033609

PHUONG AN TOI UU LIEN HOP 4: $X_1 = 189.6407$; $X_2 = 189.6407$; $X_4 = 17.4931$; $X_6 = 26.4000$; $X_{10} = 18.0000$; Cac bien khac bang khong. CUC DAI : 11.801037

Gia tri cua cac ham thoa dung - Lan thu 4

$$\begin{aligned} Z[1] &= 3967.668514 & pZ[1] &= 0.626306 \\ Z[2] &= 279.702646 & pZ[2] &= 0.303143 \\ Z[3] &= 9.848081 & pZ[3] &= 0.617413 \\ Z[4] &= 142558.396600 & pZ[4] &= 0.127420 \\ Z[5] &= 374.617080 & pZ[5] &= 1.000000 \end{aligned}$$

Kết quả thực hiện module 2.

Bảng lưu ý kiến của 05 chuyên gia (cho 04 phương án trên với số lần lặp tối đa là 5)

	Lần 1	Lần 2	Lần 3	Lần 4	Lần 5
P/A 1	2 1 2 1 2	1 1 1 1 1			
P/A 2	3 2 3 2 3	2 2 2 2 2	3 3 3 3 3	4 5 4 5 4	4 5 5 4 3
P/A 3	4 4 3 3 4	3 3 3 4 4	5 4 5 4 5		
P/A 4	5 3 5 4 5	4 5 4 5 3			

Các phương án tốt nhất

	Trung bình	Xấp xỉ	Loại
Phương án 1	(0.9000,0.9500,1.0000)	(0.9000,0.9500,1.0000)	Rất tốt
Phương án 2	(0.7000,0.8000,0.9000)	(0.7000,0.8000,0.9000)	Tốt
Phương án 3	(0.5000,0.6000,0.7000)	(0.5000,0.6000,0.7000)	Khá phù hợp

6. KẾT LUẬN

Bài báo này đã trình bày một số kết quả xây dựng phương pháp ra quyết định tập thể phục vụ cho việc quy hoạch sử dụng đất, bao gồm: phương pháp phân loại xấp xỉ dữ liệu mờ dựa trên tiêu chuẩn bình phương bé nhất và khoảng cách trọng tâm cực tiểu với một loại hàm khoảng cách có đầy đủ các tính chất của hàm khoảng cách; phương pháp Delphi thống nhất ý kiến chuyên gia có sửa đổi dựa trên phân loại các ý kiến về các phương án sử dụng đất (tìm được dựa trên phương pháp thỏa dụng mờ tương tác người - máy tính).

Các kết quả trên cần được hoàn thiện bằng cách tiếp tục nghiên cứu về các phương pháp phân loại và thống nhất ý kiến chuyên gia cũng như các hệ chuyên gia hỗ trợ ra quyết định. Ngoài ra, việc tích hợp các phần mềm ra quyết định tập thể với phần mềm quy hoạch đa mục tiêu cũng như với hệ thông tin tài nguyên, kinh tế nông nghiệp thu được trên cơ sở các phần mềm GIS là rất cần tiếp tục được triển khai nghiên cứu nhằm xây dựng hệ hỗ trợ ra quyết định phục vụ quy hoạch sử dụng đất nông nghiệp.

TÀI LIỆU THAM KHẢO

- Gillet B. E. (1990). *Introduction to Operations Research*, McGraw Hill, New York.
- Helmuth Spöth (1982). *Cluster Analysis Algorithms for Data Reduction and Classification of Objects*, John Wiley and Sons, New York.
- Kaufmann A. & Gupta M. M. (1991). *Fuzzy Mathematical Models in Engineering & Management Sciences*, North - Holland.
- Nguyen Hai Thanh, Nguyen Thi Thuy, Ngo Tuan Anh (1999). "An approximate algorithm for fuzzy cluster analysis and its applications", *Proceedings of MIF'99: The international symposium on medical*

informatics and fuzzy technology, 286-290, 26-29 August 1999.

Nguyễn Hải Thanh (2005). *Toán ứng dụng*, Giáo trình Sau đại học, Nxb Đại học Sư phạm Hà Nội.

Nguyễn Hải Thanh, chủ biên (2005). *Tin học ứng dụng trong ngành nông nghiệp*, Nxb Khoa học và Kỹ thuật.

Lê Đức Vĩnh (1997). *Các phương pháp phân lớp, ghép lớp theo khoảng cách*, Luận văn thạc sĩ toán - tin, Trường ĐH KHTN, Đại học Quốc gia.

