

# Một phương pháp sinh luật mờ dựa trên cây quyết định và đại số gia tử xây dựng hệ luật mờ giải bài toán hồi quy

Nguyễn Đức Dư, Hoàng Văn Thông

Khoa Công nghệ Thông tin, Trường Đại học Giao thông Vận tải, Hà Nội

Tác giả liên hệ: Nguyễn Đức Dư, nducdu@utc.edu.vn

Ngày nhận bài: 12/11/2019, ngày sửa chữa: 24/12/2019, ngày duyệt đăng: 25/12/2019

Định danh DOI: 10.32913/mic-ict-research-vn.v2019.n2.901

Biên tập lĩnh vực điều phối phản biện và quyết định nhận đăng: PGS.TS. Võ Đình Bảy

**Tóm tắt:** Bài báo này đề xuất một phương pháp sinh luật mờ dựa trên cây quyết định và đại số gia tử để xây dựng hệ luật mờ giải bài toán hồi quy. Thuật toán được thử nghiệm trên 9 bài toán mẫu và đối sánh với các phương pháp đã có PAES\_KB và HA-MG-PAES-Kmax trên các mục tiêu độ chính xác và độ phức tạp của hệ luật. Kết quả đối sánh cho thấy thuật toán đề xuất cho kết quả tốt hơn.

**Từ khóa:** Hệ luật mờ, đại số gia tử, cây quyết định, bài toán hồi quy.

---

**Title:** A Method to Generate Fuzzy Rules based on Decision Tree and Hedge Algebras for Building Fuzzy Rule based Systems for Regression

**Abstract:** This paper proposes a method to generate fuzzy rules based on decision tree and hedge algebras for building fuzzy rule-based systems for regression problems. The proposed method was experimented on nine regression problems and was compared to two existing methods PAES\_KB and HA-MG-PAES-Kmax on two objectives including the accuracy and complexity of rule-based systems. Comparative results show that our proposed method outperforms the existing ones.

**Keywords:** Fuzzy rule-based system, hedge algebra, decision tree, regression.

---

## I. MỞ ĐẦU

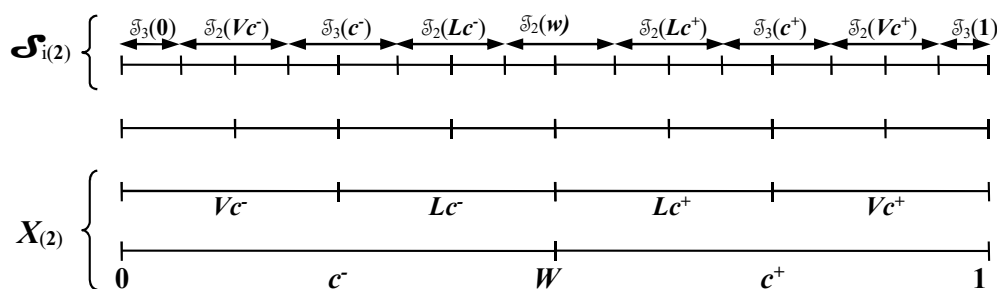
Cách giải quyết các bài toán phân lớp và hồi quy bằng hệ luật mờ (FRBS: Fuzzy Rule based System) nhận được nhiều sự quan tâm của các nhóm nghiên cứu, như: Acalá và cộng sự [1], Antonelli và cộng sự [2], Ishibuchi và Nojima [3, 4], Pulkkinen và Koivisto [5], Cordon và cộng sự [6], Nguyễn Cát Hồ và cộng sự [7–9], Aghaeipoor và Javidi [10]. Các nghiên cứu đã đề xuất chủ yếu tập trung tìm kiếm các phương pháp xây dựng FRBS cho bài toán phân lớp, còn bài toán hồi quy chưa có nhiều nghiên cứu đề cập tới [1]. Khi xây dựng FRBS giải các bài toán này chúng ta chủ yếu giải quyết ba vấn đề chính: thiết kế phân hoạch mờ (ngữ nghĩa tính toán của từ), sinh tập các luật mờ ứng cử, tìm kiếm hệ luật mờ tối ưu.

Về vấn đề sinh luật ứng cử, các phương pháp tiếp cận dựa trên lý thuyết tập mờ thường sinh luật bằng cách tổ hợp tất cả các giá trị ngôn ngữ sử dụng cho các biến [1–4, 11]. Nhược điểm của hướng tiếp cận này là khi tập dữ liệu có nhiều thuộc tính thì không gian tìm kiếm luật sẽ rất lớn.

Một số đề xuất sinh luật từ cây quyết định (decision tree) cho bài toán phân lớp [5, 12–14]. Phương pháp này đã làm giảm đáng kể số luật phải xem xét nhờ vào các kỹ thuật như là hạn chế chiều cao và cắt tỉa cây, tuy nhiên lại gặp khó khăn trong quá trình tối ưu tham số tập mờ.

Với hướng tiếp cận theo lý thuyết đại số gia tử (ĐSGT), trong [7–9] Nguyễn Cát Hồ và cộng sự đề xuất một phương pháp sinh luật từ mẫu dữ liệu. Theo đó, mỗi mẫu dữ liệu sinh ra một luật có độ dài  $n$  bằng số thuộc tính của tập mẫu dữ liệu, từ các luật này sinh các luật có độ dài  $lmax$  nhỏ hơn cho trước ( $lmax < n$ ). Với phương pháp sinh luật ứng cử này thì số luật tối đa phải xem xét giảm đi đáng kể so với phương pháp sinh luật tổ hợp. Tuy nhiên theo hướng tiếp cận này chúng ta vẫn phải xem xét một số lượng luật khá lớn.

Trong bài báo này chúng tôi đề xuất một phương pháp xây dựng FRBS giải bài toán hồi quy với các luật được sinh ra bằng cây quyết định và ĐSGT. Thuật toán giải quyết cả hai vấn đề sinh luật và tối ưu tham số của các tập mờ.



Hình 1. Hệ khoảng tương tự của các từ có độ dài không quá 2.

Thuật toán đề xuất gồm hai pha. Pha thứ nhất tối ưu tham số của ĐSGT sử dụng cho mỗi biến của bài toán. Ở pha này chúng tôi sử dụng thuật giải di truyền để tìm kiếm tham số tối ưu. Pha thứ hai, với bộ tham số tối ưu của ĐSGT tìm được của pha thứ nhất, chúng tôi xây dựng các ĐSGT và sử dụng chúng để chuyển đổi tập dữ liệu giá trị số của bài toán thành tập dữ liệu giá trị ngôn ngữ tương ứng. Từ tập dữ liệu ngôn ngữ cây quyết định được xây dựng, từ cây quyết định sinh ra tập luật ứng cử và sử dụng thuật toán cải tiến (2+2)M-PAES được sử dụng để tìm FRBS tối ưu.

Thuật toán đề xuất được thử nghiệm trên 9 bài toán hồi quy và đối sánh kết quả thu được trên các chỉ số độ phức tạp của hệ luật và giá trị sai số trung bình phương với các thuật toán cùng hướng tiếp cận PAES<sub>KB</sub> [1] và HA-MG-PAES-Kmax [8]. Kết quả đối sánh cho thấy thuật toán đề xuất cho kết quả có độ chính xác tốt hơn.

Phần còn lại của bài báo được bố cục như sau. Phần II trình bày bài toán hồi quy và phương pháp giải bài toán hồi quy bằng hệ mờ dựa trên đại số gia tử. Phần III trình bày phương pháp sinh luật mờ dựa trên cây quyết định và ĐSGT. Phần IV trình bày thuật toán xây dựng hệ luật. Phần V trình bày kết quả thử nghiệm. Cuối cùng, Phần VI rút ra một số kết luận.

## II. BÀI TOÁN HỒI QUY VÀ HỆ LUẬT MỜ MAMDANI DỰA TRÊN ĐSGT

### 1. Bài toán hồi quy

Cho một tập mẫu dữ liệu  $\mathbf{D} = \{(x_i, y_i), i = 1, 2, \dots, N\}$ , trong đó  $x_i$  là một véc-tơ  $n$  chiều có dạng  $(x_{i1}, x_{i2}, \dots, x_{in})$  với  $x_{ij} \in U_j \subset \mathbb{R}$ ,  $U_j$  là miền xác định của các biến độc lập  $X_j$  (thuộc tính đầu vào) của bài toán ( $j = 1, 2, \dots, n$ ),  $y_i \in U_{n+1} \subset \mathbb{R}$ ,  $U_{n+1}$  là miền xác định của biến phụ thuộc  $Y$  (thuộc tính đầu ra),  $N$  là số mẫu dữ liệu.

Từ tập dữ liệu mẫu  $\mathbf{D}$  xây dựng một hệ mờ dựa trên luật cho phép tính giá trị  $\hat{y}$  ứng với mỗi giá trị đầu vào  $x \in U = U_1 \times \dots \times U_n$ .

### 2. Giải bài toán hồi quy bằng hệ luật mờ Mamdani dựa trên ĐSGT

Giải bài toán hồi quy bằng hệ luật mờ Mamdani là xây dựng một hệ các luật mờ Mamdani có dạng

$$R_m : \text{if } X_1 \text{ is } A_{1,j_m}, \dots, X_n \text{ is } A_{n,j_m} \text{ then } Y \text{ is } A_{n+1,j_m}, \quad (1)$$

để dự đoán giá trị đầu ra  $\hat{y}$  ứng với giá trị đầu vào  $x$  có  $n$  chiều. Trong (1),

$$A_{i,j_m} \in L_i = \{A_{i,0} \cup X_{(k_i)} = \{A_{i,1}, A_{i,2}, \dots, A_{i,|X_{(k_i)}}\}\},$$

với  $i = 1, 2, \dots, n + 1$ , trong đó  $X_{(k_i)}$  là tập các từ ngôn ngữ có độ dài không quá  $k_i$  được sinh ra bằng ĐSGT  $\mathcal{A}X_i$  tương ứng, và nó được sử dụng để xây dựng phân hoạch thuộc tính thứ  $i$ ,  $A_{f,0}$  có giá trị là *Don't care* với hàm thuộc đồng nhất bằng một. Chú ý,  $L_{n+1}$  không chứa giá trị *Don't care* và  $m = 1, 2, \dots, M$ , với  $M$  là số luật của hệ.

Tương tự như các đề xuất trong [1, 2, 8, 9], chúng tôi sử dụng phương pháp trung bình trọng số để suy diễn giá trị  $\hat{y}$  từ hệ luật khi biết véc-tơ đầu vào  $x_i$  theo công thức sau:

$$\hat{y}_i = \frac{\sum_{m=1}^M \mu_m(x_i) \bar{A}_{n+1,j_m}}{\sum_{m=1}^M \mu_m(x_i)}, \quad (2)$$

với  $i = 1, 2, \dots, N$ , trong đó

$$\mu_m(x_i) = \prod_{f=1}^n \mu_{A_{f,j_m}}(x_{if})$$

là độ đốt cháy luật thứ  $m$  của mẫu dữ liệu  $x_i$ ,  $\bar{A}_{n+1,j_m}$  là giá trị định lượng của hạng từ ngôn ngữ  $A_{n+1,j_m}$  và  $\mu_{A_{f,j_m}}(\cdot)$  là hàm thuộc của từ ngôn ngữ  $A_{f,j_m}$ . Trong (2), nếu  $\sum_{m=1}^M \mu_m(x_i) = 0$ , có nghĩa là hệ luật *không phủ* mẫu dữ liệu  $x_i$ , khi đó  $\hat{y}$  được suy diễn theo phương pháp đề xuất trong [1].

Để đánh giá độ chính xác của hệ luật đã xây dựng, chúng tôi dựa vào giá trị sai số trung bình phương (MSE: Mean Squared Error):

$$\text{MSE} = \frac{1}{2N} \sum_{i=1}^N (\hat{y}_i - y_i)^2. \quad (3)$$

---

**Thuật toán 1: BuildDecisionTree( $\mathbf{D}, \{\pi_i\}_{i=1}^{n+1}, k, lmax$ )**

---

- 1 **Dữ liệu vào:** Cơ sở dữ liệu của bài toán  $\mathbf{D}$ ; bộ tham số của các ĐSGT  $\{\pi_i\}_{i=1}^{n+1}$ ; chiều dài tối đa của các hạng từ sinh ra từ ĐSGT  $k$ ; chiều cao tối đa của cây  $lmax$ .
- 2 **Dữ liệu ra:** Cây quyết định  $\mathbf{T}$ .
- 3 **for**  $i = 1$  **to**  $n + 1$  **do**
- 4     Xây dựng ĐSGT  $\mathcal{A}X_i$  tương ứng với bộ tham số  $\pi_i$ ;
- 5     Sinh tập các từ  $X_{(k_i)}$ ;
- 6     Sinh hệ khoảng tính mờ tương tự  $S_{(k_i)}$ ;
- 7 **end**
- 8  $D_1 =$  Chuẩn hóa  $\mathbf{D}$  về đoạn  $[0, 1]$ ;
- 9  $D_2 =$  Chuyển đổi  $D_1$  thành cơ sở dữ liệu ngôn ngữ dựa trên các hệ khoảng tính mờ  $\{S_{(k_i)}\}_{i=1}^{n+1}$ ;
- 10 Xây dựng cây quyết định  $\mathbf{T}$  có chiều cao tối đa  $lmax$  từ cơ sở dữ liệu  $D_2$  bằng thuật toán C4.5;
- 11 **return**  $\mathbf{T}$ ;

---

Trong (3),  $\hat{y}_i$  và  $y_i$  lần lượt là giá trị suy diễn từ hệ luật và giá trị quan sát tương ứng với giá trị đầu vào  $x_i$ . Giá trị MSE càng nhỏ thì hệ luật mờ càng chính xác.

### III. ĐỀ XUẤT PHƯƠNG PHÁP SINH HỆ LUẬT MỜ DỰA TRÊN CÂY QUYẾT ĐỊNH VÀ ĐSGT

#### 1. Hệ khoảng tính mờ tương tự của từ ngôn ngữ

ĐSGT được Nguyễn Cát Hồ và cộng sự đề xuất lần đầu tiên trong [15], xây dựng một phương pháp hình thức sinh các từ ngôn ngữ và ánh xạ từ ngôn ngữ tới các giá trị định lượng tương ứng trên đoạn  $[0, 1]$ . ĐSGT giúp chúng ta có thể sinh ra các dạng ngữ nghĩa khác nhau của từ ngôn ngữ từ ngữ nghĩa vốn có của từ bao gồm: giá trị định lượng  $v_x(x)$ , khoảng tính mờ  $fm(x)$ , khoảng tương tự [7]. Gọi  $X_{(k_i)}$  là tập các từ có độ dài không quá  $k_i$  của biến  $X_i$ . Tập  $S_{(k_i)}$  các khoảng tương tự của các từ trong  $X_{(k_i)}$  hình thành một phân hoạch trên  $U$  và giá trị định lượng ngữ nghĩa của mọi từ  $x \in X_{(k_i)}$  là  $v_x(x) \in T(x)$ , ở đó  $T(x)$  là khoảng tương tự của từ  $x$ . Các giá trị của khoảng tương tự  $T(x)$  được coi như là tương tự nhau và với giá trị định lượng ngữ nghĩa  $v_x(x)$  của  $x$  với một cấp độ  $k_i$ ,  $k_i$  càng lớn mức độ tương tự của các giá trị trong mỗi khoảng tương tự càng cao. Hệ khoảng tương tự là một công cụ hữu dụng để phân hoạch miền tham chiếu của các biến, được sử dụng trong các thuật toán sinh luật của các phương pháp tiếp cận dựa trên ĐSGT.

#### 2. Thuật toán xây dựng cây quyết định

Thuật toán xây dựng cây quyết định được trình bày trong thuật toán 1. Để xây dựng cây quyết định từ tập dữ liệu  $\mathbf{D}$  của bài toán hồi quy gồm các véc-tơ đầu vào  $x_i = (x_{i1}, x_{i2}, \dots, x_{in})$ , với  $x_{ij} \in U_j \subset \mathbb{R}$  và các giá trị đầu ra  $y_i \in U_{n+1} \subset \mathbb{R}$ . Bước đầu tiên chúng ta cần chuẩn hóa

---

**Thuật toán 2: GenFRBS( $\mathbf{T}$ )**

---

- 1 **Dữ liệu vào:** Cây quyết định  $\mathbf{T}$ .
- 2 **Dữ liệu ra:** Hệ luật mờ  $\mathbf{S}$ .
- 3  $\mathbf{S} = \emptyset$ ;
- 4  $Leafs =$  Tập các nút lá của cây  $\mathbf{T}$ ;
- 5 **foreach**  $lf \in Leafs$  **do**
- 6     Với mỗi lá  $lf$  xây dựng danh sách  $lsNode$  các Node từ lá  $lf$  đến gốc cây;
- 7     Tạo luật  $r$  có  $n$  điều kiện tiền đề, tất cả các tiền đề đều có giá trị *Don't care*;
- 8     **for**  $j = lsNode.Count - 1$  **down to**  $1$  **do**
- 9         Thay thế giá trị *Don't care* của luật  $r$  ứng với thuộc tính của nút  $lsNode[j]$  bằng giá trị phân chia của nút  $lsNode[j - 1]$  (nút cha của nút  $lsNode[j]$ );
- 10     **end**
- 11     Gán kết luận của luật  $r$  là giá trị của nút  $lsNode[0]$ ;
- 12      $\mathbf{S} = \mathbf{S} \cup \{r\}$ ;
- 13 **end**
- 14 **return**  $\mathbf{S}$ ;

---

tập dữ liệu  $\mathbf{D}$  về đoạn  $[0, 1]$  bằng chuyển đổi tuyến tính ta được tập  $D_1$ .

Bước thứ hai, với mỗi thuộc tính đầu vào, ra ta xác định một bộ tham số của ĐSGT tương ứng, giả sử là  $\pi_i$  ( $i = 1, 2, \dots, n + 1$ ). Với bộ tham số  $\pi_i$  ta xây dựng ĐSGT  $\mathcal{A}X_i$  sinh ra tập các từ có độ dài không quá  $k_i$  ký hiệu  $X_{(k_i)}$ , tính giá trị định lượng ngữ nghĩa của các từ trong  $X_{(k_i)}$  và xây dựng hệ khoảng tương tự  $S_{(k_i)}$  tương ứng theo thuật toán trong [7].

Bước thứ ba chuyển đổi cơ sở dữ liệu  $D_1$  thành cơ sở dữ liệu từ ngôn ngữ  $D_2$  theo nguyên tắc sau: với mỗi véc-tơ  $x_i = (x_{i1}, x_{i2}, \dots, x_{in})$  chuyển đổi thành véc-tơ từ ngôn ngữ  $x'_i = (A_{i1}, A_{i2}, \dots, A_{in})$ , trong đó  $x_{ij} \in T(A_{ij})$  với  $j = 1, 2, \dots, n$ ; giá trị đầu ra  $Y$  cũng chuyển đổi tương tự. Từ cơ sở dữ liệu ngôn ngữ  $D_2$  chúng ta áp dụng thuật toán C4.5 [16] xây dựng cây quyết định có chiều cao tối đa  $lmax$ , việc thiết lập chiều cao tối đa của cây nhằm hạn chế chiều dài của luật được sinh ra. Mỗi nút của cây quyết định chứa hai giá trị là nhãn thuộc tính và giá trị phân chia của nút cha.

#### 3. Thuật toán sinh luật từ cây quyết định

Thuật toán sinh luật từ cây quyết định được trình bày trong thuật toán 2. Đầu vào của thuật toán là cây quyết định  $\mathbf{T}$  được xây dựng bằng thuật toán 1. Gọi  $\mathbf{S}$  là tập luật được sinh ra, khởi đầu  $\mathbf{S} = \emptyset$ ,  $Leafs$  là tập các nút lá của cây  $\mathbf{T}$ . Với mỗi nút lá sẽ sinh ra một luật có phần kết luận là nhãn của nút đó và phần tiền đề của luật là các nhãn của các nút nằm trên đường đi từ gốc đến lá đó. Để xác định đường đi, chúng ta xuất phát từ nút lá đi ngược về nút gốc. Giả sử để sinh luật  $r$  từ nút lá  $lf$ , từ nút này ta dễ dàng xác định được danh sách ( $lsNode$ )

là các nút mô tả đường đi từ nút lá  $lf$  đến nút gốc. Tiếp theo ta tạo ra luật  $r$  có độ dài bằng số chiều của bài toán và có tất cả tiền điều kiện đều bằng *Don't care*. Với mỗi nút  $lsNode[j]$  ( $j = lsNode.Count - 1, \dots, 1$ ) thay thế giá trị *Don't care* của tiền điều kiện tương ứng với thuộc tính là nhãn của nút  $lsNode[j]$  bằng giá trị phân chia của nút  $lsNode[j - 1]$  (nút cha của nút  $lsNode[j]$ ), gán kết luận của  $r$  bằng giá trị của nút  $lsNode[0]$ . Thêm luật  $r$  vào tập luật  $S$ .

#### IV. PHƯƠNG PHÁP XÂY DỰNG FRBS GIẢI BÀI TOÁN HỒI QUY

Trong phần này chúng tôi áp dụng phương pháp sinh luật đề xuất trong phần III để xây dựng hệ mờ giải bài toán hồi quy. Phương pháp xây dựng FRBS được thực hiện với hai pha. Pha thứ nhất chúng tôi phát triển thuật toán *OptHAParams* sử dụng thuật giải di truyền để tìm bộ tham số mờ của các ĐSGT của các thuộc tính của bài toán. Pha thứ hai sử dụng bộ tham số mờ của ĐSGT đã tìm ở pha thứ nhất xây dựng cây quyết định của bài toán, từ cây quyết định sinh ra tập các luật ứng cử, sau đó áp dụng thuật toán HA-De-PAES để tìm kiếm hệ luật tối ưu. Ở đây HA-De-PAES được phát triển dựa trên thuật toán (2+2)M-PAES [17]. Thuật toán (2+2)M-PAES tối ưu đồng thời hệ luật và tham số mờ trong khi thuật toán HA-De-PAES chỉ tối ưu hệ luật và các luật được chọn từ tập luật ứng cử được sinh ra từ cây quyết định, với hai mục tiêu là MSE và Comp (tổng chiều dài của luật).

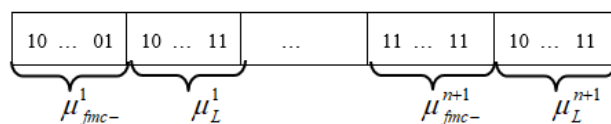
##### 1. Thuật toán tìm tham số tối ưu của ĐSGT bằng thuật toán GA

Để tìm kiếm tham số mờ của ĐSGT, trong bài báo này chúng tôi thực hiện thiết kế thuật giải di truyền dựa trên sơ đồ mã hóa nhị phân với hàm thích nghi là giá trị sai số trung bình phương MSE. Bài toán có  $n$  thuộc tính, khi đó với mỗi thuộc tính ta cần xác định hai tham số là  $\mu_{fmc-}$  và  $\mu_L$  (ở đây chúng tôi sử dụng ĐSGT có hai giá tử  $V$  và  $L$ ). Như vậy số tham số cần xác định cho bài toán  $\{\pi_i\}_{i=1}^{n+1}$  là  $2(n+1)$ . Chi tiết được trình bày trong thuật toán 3. Hình 2 mô tả sơ đồ mã hóa một cá thể, trong đó mỗi biến mục tiêu được mã hóa bằng một chuỗi bit có  $l$  bit.

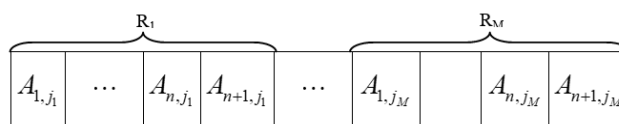
##### 2. Thuật toán HA-De-PAES xây dựng hệ mờ tối ưu

Mỗi cá thể của quần thể được mã hóa gồm  $M$  luật ( $M$  có thể khác nhau giữa các cá thể), với mỗi luật  $R_i$  được lấy ra từ tập luật ứng cử  $S$ . Nhằm đạt được sự cân bằng giữa tính dễ hiểu và độ chính xác của hệ luật, chúng tôi giới hạn số luật trong mỗi FRBS nằm trong đoạn  $[M_{min}, M_{max}]$ .

Với mỗi cá thể cần tối thiểu hai mục tiêu là MSE và Comp, trong đó MSE được xác định theo (3) và Comp là tổng độ dài của các luật trong cơ sở luật, biểu thị độ phức tạp của FRBS.



Hình 2. Mã hóa cá thể tối ưu tham số của ĐSGT.



Hình 3. Cấu trúc mã hóa một cá thể biểu diễn một hệ luật.

##### a) Các toán tử di truyền

**Toán tử lai ghép:** Với hai cá thể bố mẹ  $p_1$  và  $p_2$  sử dụng phương pháp lai ghép một điểm (one-point crossover), điểm lai ghép được chọn ngẫu nhiên trong đoạn  $[1, \rho_{min} - 1]$ , trong đó  $\rho_{min}$  là số luật nhỏ nhất trong  $p_1$  và  $p_2$ . Lưu ý rằng nếu toán tử lai ghép không được thực hiện thì đột biến luôn xảy (có nghĩa là nếu không xảy ra lai ghép thì xác suất đột biến sẽ bằng 1).

**Toán tử đột biến:** Nếu đột biến xảy ra thì chọn ngẫu nhiên thực hiện một trong hai toán tử đột biến sau. *Toán tử đột biến thêm luật:* Thêm  $\gamma$  luật vào với  $\gamma$  được chọn ngẫu nhiên trong đoạn  $[1, \gamma_{max}]$ , nếu  $\gamma + M > M_{max}$  thì  $\gamma = M_{max} - M$ , các luật được chọn từ tập  $S$ . *Toán tử đột biến thay đổi luật:* Thay đổi ngẫu nhiên  $\delta$  giá trị ngôn ngữ của một số luật của hệ luật, với  $\delta$  được chọn ngẫu nhiên trong đoạn  $[1, \delta_{max}]$ . Thực hiện  $\delta$  lần quá trình sau: chọn ngẫu nhiên một luật  $R$ , tiếp theo chọn ngẫu nhiên một gen  $j \in [1, n + 1]$  của  $R$ ; nếu  $j \leq n$  thì chọn ngẫu nhiên một từ trong  $\{Don'tcare\} \cup X_{(k_j)}$ ; nếu  $j = n + 1$  thì chọn ngẫu nhiên một từ trong  $X_{(k_{n+1})}$ ; thay thế từ của gen thứ  $j$  của  $R$  bằng từ vừa chọn; nếu luật  $R$  sau khi đột biến có độ dài lớn hơn  $lmax$  thì đột biến sẽ bị bỏ qua.

Lưu ý rằng sau khi lai ghép, đột biến nếu có nhiều luật trùng nhau thì chỉ giữ lại một, loại bỏ các luật có độ dài bằng 0.

##### b) Thuật toán tiến hóa HA-De-PAES

Phần này trình bày những bước cơ bản của thuật toán HA-De-PAES được phát triển dựa trên lược đồ tiến hóa (2+2)M-PAES đề xuất trong [17]. Tại mỗi thời điểm thuật toán lưu trữ một quần thể mà một cá thể bất kỳ trong đó đều không bị trội bởi các cá thể còn lại theo các mục tiêu MSE và Comp. Một quần thể như vậy sẽ xác định một mặt Pareto. Thuật toán thực hiện tìm kiếm một mặt Pareto xấp xỉ tối ưu theo hai mục tiêu MSE và Comp.

Xét bài toán tối ưu  $n$  mục tiêu, cực tiểu hàm  $F(x) = (f_1(x), f_2(x), \dots, f_n(x))$  với  $x \in X \subseteq \mathbb{R}^m$ , trong đó  $F(x)$



---

**Thuật toán 3: OptHAParams( $\mathbf{D}, k, lmax, GApars$ )**

---

- 1 **Dữ liệu vào:** Cơ sở dữ liệu của bài toán  $\mathbf{D}$ ; Chiều dài tối đa của các hạng tử sinh ra từ ĐSGT  $k$ ; Chiều cao tối đa của cây  $lmax$ ; Bộ tham số của thuật giải di truyền  $GApars$  gồm có chiều dài nhiễm sắc thể  $lchrom$ , kích thước quần thể  $Popsiz$ , số thế hệ  $G$ , xác suất lai ghép  $Pcross$ , và xác suất đột biến  $Pmu$ .
  - 2 **Dữ liệu ra:** Bộ tham số tối ưu  $\pi_{opt}$ .
  - 3 **Bước 1:** Khởi tạo.
  - 4 Gán  $j \leftarrow 0$ ;
  - 5 Khởi tạo quần thể ban đầu:  $Initial(Pop_j)$ ;
  - 6 **foreach**  $p \in Pop_j$  **do**
  - 7     Giải mã  $p$  để được bộ tham số  $\{\pi_i\}_{i=1}^{n+1}$  của các ĐSGT;
  - 8      $\mathbf{T} \leftarrow BuildDecisionTree(\mathbf{D}, \{\pi_i\}_{i=1}^{n+1}, k, lmax)$ ;
  - 9      $\mathbf{S} \leftarrow GenFRBS(\mathbf{T})$ ;
  - 10    Suy diễn trên tập dữ liệu  $\mathbf{D}$  bằng hệ luật  $\mathbf{S}$  và tính giá trị hàm mục tiêu MSE của  $p$ ;
  - 11 **end**
  - 12  $X_{best} \leftarrow GetIndividualHasBestObj(Pop_j)$ ; //Chọn cá thể có giá trị mục tiêu tốt nhất.
  - 13 **Bước 2:** Tiến hóa.
  - 14  $ComputeFitnessMeasure(Pop_j)$ ; //Tính độ đo thích nghi của các cá thể.
  - 15  $Pop_{parent} \leftarrow Select(Pop_j)$ ; //Chọn cá thể cha mẹ.
  - 16  $Pop_{child} \leftarrow Mute(Crossover(Pop_{parent}))$ ; //Lai ghép và đột biến tạo quần thể con.
  - 17  $j \leftarrow j + 1$ ;
  - 18  $Pop_j \leftarrow Pop_{child}$ ; //Thay thế quần thể hiện tại bằng quần thể con vừa tạo sinh.
  - 19 **foreach**  $p \in Pop_j$  **do**
  - 20     Giải mã  $p$  để được bộ tham số  $\{\pi_i\}_{i=1}^{n+1}$  của các ĐSGT;
  - 21      $\mathbf{T} \leftarrow BuildDecisionTree(\mathbf{D}, \{\pi_i\}_{i=1}^{n+1}, k, l)$ ;
  - 22      $\mathbf{S} \leftarrow GenFRBS(\mathbf{T})$ ;
  - 23     Suy diễn trên tập dữ liệu  $\mathbf{D}$  bằng hệ luật  $\mathbf{S}$  và tính giá trị hàm mục tiêu MSE của  $p$ ;
  - 24 **end**
  - 25  $X \leftarrow GetIndividualHasBestObj(Pop_j)$ ;
  - 26 **if**  $obj(X) > obj(X_{best})$  **then**
  - 27      $X_{best} \leftarrow X$ ; //  $obj(X)$  là giá trị mục tiêu của cá thể  $X$ .
  - 28 **end**
  - 29 **Bước 3:** Lặp lại Bước 2 cho đến khi  $j > G$ .
  - 30 **Bước 4:** Trả lại bộ tham số tối ưu.
  - 31 Giải mã  $X_{best}$  để được bộ tham số  $\{\pi_i\}_{i=1}^{n+1}$  của các ĐSGT;
  - 32  $\pi_{opt} \leftarrow \{\pi_i\}_{i=1}^{n+1}$ ;
  - 33 **return**  $\pi_{opt}$ ;
- 

---

**Thuật toán 4: HA-De-PAES**

( $\mathbf{D}, k, lmax, \pi_{opt}, paespars$ )

---

- 1 **Dữ liệu vào:** Cơ sở dữ liệu của bài toán  $\mathbf{D}$ , chiều dài tối đa của các hạng tử sinh ra từ ĐSGT  $k$ , chiều cao tối đa của cây  $lmax$ , bộ tham số mờ của ĐSGT đã được tối ưu, và bộ tham số của thuật toán PAES  $paespars$ .
  - 2 **Dữ liệu ra:**  $P_A$  (mặt xấp xỉ tối ưu Pareto với hai mục tiêu MSE và Comp của các hệ luật).
  - 3 **Bước 1:** Tạo tập luật ứng cử.
  - 4  $\mathbf{T} = BuildDecisionTree(\mathbf{D}, \pi_{opt}, k, lmax)$ ;
  - 5  $\mathbf{S} = GenFRBS(\mathbf{T})$ ;
  - 6 **Bước 2:** Sinh ngẫu nhiên 2 cá thể  $c_1, c_2$ . Mỗi cá thể gồm  $M$  luật được chọn từ tập luật ứng cử  $\mathbf{S}$ .  $M$  được chọn ngẫu nhiên trong đoạn  $[M_{min}, M_{max}]$ .
  - 7 **Bước 3:** Bổ sung  $c_1, c_2$  vào  $P_A$ .
  - 8 **Bước 4:** Lặp  $i = 1, \dots, MaxGen$  (số thế hệ tối đa).
  - 9 Chọn ngẫu nhiên hai cá thể bố mẹ  $p_1, p_2$  trong  $P_A$  ( $p_1, p_2$  có thể trùng nhau);
  - 10 Thực hiện lai ghép hai cá thể  $p_1, p_2$  để sinh ra hai cá thể con  $o_1, o_2$ ;
  - 11 Thực hiện đột biến trên  $o_1, o_2$ ;
  - 12 Tính giá trị mục tiêu (MSE, Comp) của  $o_1, o_2$ ;
  - 13 Lần lượt bổ sung  $o_1, o_2$  vào  $P_A$  nếu có thể;
  - 14 Lặp lại bước 4 với thế hệ kế tiếp;
  - 15 **return**  $P_A$ ;
- 

là véc-tơ mục tiêu,  $f_i(x)$  là mục tiêu thứ  $i$  cần cực tiểu,  $x$  là véc-tơ lời giải trong không gian  $m$  chiều, và  $X$  là không gian lời giải của bài toán. Một mặt  $P \subseteq X$  được gọi là mặt Pareto nếu mỗi điểm bất kỳ của nó đều không bị trội bởi các điểm còn lại trong  $P$ . Một lời giải  $x \in P$  được gọi là trội hơn lời giải  $y \in P$ , ký hiệu  $x > y$ , nếu  $f_i(x) \leq f_i(y)$  với

mọi  $i \in \{1, \dots, m\}$  và tồn tại  $j$  thỏa mãn  $f_j(x) < f_j(y)$ .

Kí hiệu  $P_A$  là quần thể hiện tại, thuật toán gồm các bước được trình bày trong thuật toán 4. Một cá thể con  $o$  nếu không bị trội bởi bất kỳ cá thể nào trong  $P_A$  thì  $o$  được bổ sung vào  $P_A$ , đồng thời loại bỏ tất cả những cá thể trong  $P_A$  bị trội bởi  $o$ . Nếu số cá thể trong  $P_A$  lớn hơn số lượng tối đa ( $MaxArchive$ ) được phép lưu trữ trong  $P_A$  thì loại bỏ ngẫu nhiên một cá thể trong vùng có mật độ cao nhất ra khỏi  $P_A$ . Xác định vùng có mật độ cao nhất theo thuật toán trong [17].

## V. NGHIÊN CỨU THỬ NGHIỆM

Chúng tôi tiến hành thử nghiệm thuật toán xây dựng FRBS được đề xuất trong bài báo và đối sánh kết quả với các thuật toán PAES<sub>KB</sub> trong [1] và HA-PAES-MG-Kmax trong [8]. PAES<sub>KB</sub> tiếp cận dựa trên lý thuyết tập mờ, các tập mờ được biểu diễn bằng bộ 2 (two-tuples), các luật mờ được sinh ra bằng tổ hợp ngẫu nhiên các từ ngôn ngữ sử dụng trên mỗi biến, quá trình tối ưu hóa tham số tập mờ và hệ luật cũng bằng thuật toán (2+2)M-PAES trong [17]. HA-PAES-MG-Kmax tiếp cận dựa trên lý thuyết ĐSGT, các tham số tập mờ được xác định dựa trên tham số mờ của ĐSGT, luật mờ được sinh ra dựa trên mẫu dữ liệu, quá trình tối ưu hóa tham số tập mờ và hệ luật cũng bằng thuật toán (2+2)M-PAES trong [17]. Chúng tôi chọn hai thuật toán này để đối sánh do chúng cùng sử dụng thuật toán tiến hóa (2+2)M-PAES và nhằm chứng tỏ tính hiệu quả của phương pháp sinh luật dựa trên cây quyết định.

Bảng I  
CÁC BÀI TOÁN SỬ DỤNG THỬ NGHIỆM [1, 8]

TT	Bài toán	#NoP	#NoA
1	Electrical Length 1 (ELE1)	495	2
2	Electrical Maintainance 2 (ELE2)	1056	4
3	Weather Ankara (WA)	1609	9
4	Weather Izmir (WI)	1461	9
5	Treasury (TR)	1049	15
6	Abalone (AB)	4177	8
7	Mortgage (MTG)	1049	15
8	Computer Activity (CA)	8192	21
9	Pole Telecommunication (PT)	15000	26

Bảng II  
CÁC THAM SỐ THỬ NGHIỆM PHA  
THỨ NHẤT, TÌM THAM SỐ TỐI ƯU

$\mu_{\min} = 0,3$	$f m C_{\min} = 0,3$	$f m C_{\max} = 0,7$
$\mu_{\max} = 0,7$	$kmax = 3$	$lmax = 2$
$Lchrom = 8$	$Popsiz = 100$	$G = 100$
$Pcross = 0,7$	Xác suất lai ghép	
$Pmu = 0,1$	Xác suất đột biến	

Bảng III  
CÁC THAM SỐ THỬ NGHIỆM PHA  
THỨ HAI, TÌM KIẾM HỆ LUẬT TỐI ƯU

$MaxArchive = 64$	$MaxGen = 300.000$	
$\gamma_{\max} = 5$	$\delta_{\max} = 5$	$lmax = \min(\#NoA, 5)$
$M_{\min} = 5$	$M_{\max} = 30$	
$PcRB = 0,3$	Xác suất lai ghép trên $C_{RB}$	
$PmRB = 0,1$	Xác suất đột biến trên $C_{RB}$	
$PAdd = 0,75$	Xác suất đột biến thêm luật trên $C_{RB}$	

Để công bằng trong so sánh hiệu quả của các phương pháp, chúng tôi sử dụng dạng phân hoạch mờ và các tham số thử nghiệm tương tự như các phương pháp được so sánh. Phân hoạch mờ được sử dụng có dạng đa thể hạt, các tập mờ có dạng tam giác, độ dài tối đa của các hạng tử được sinh ra bằng ĐSGT là  $k = 3$  cho tất cả các thuộc tính đầu vào và đầu ra.

Chúng tôi tiến hành thử nghiệm trên 09 bài toán, trong đó 08 bài toán thử nghiệm được lấy từ <http://archive.ics.uci.edu/ml/datasets.php>, riêng bài toán Abalone lấy từ <https://sci2s.ugr.es/keel/dataset.php?Cod=96>, với #NoP là số mẫu và #NoA là số thuộc tính. Các tham số thử nghiệm cho pha thứ nhất được trình bày trong bảng II và pha thứ hai trình bày trong bảng III. Phương pháp thử nghiệm kiểm

tra chéo 5-fold, với 4 fold học và 1 fold kiểm tra. Mỗi fold được thử nghiệm 6 lần ( $6 \times 5 = 30$  lần). Mỗi lần thử nghiệm pha thứ nhất sẽ gọi thuật toán OptHAParams để tìm bộ tham số mờ của ĐSGT cho tất cả các thuộc tính, bộ tham số tìm được là đầu vào của thuật toán HA-De-PAES. Để giảm thời gian xây dựng cây quyết định ở pha thứ nhất, chúng tôi giới hạn chiều cao tối đa của cây được sinh ra  $lmax = 2$ , và để giới hạn chiều dài của luật được sinh ra trong pha thứ hai chúng tôi thiết lập  $lmax$  bằng giá trị nhỏ nhất của #NoA và 5.

Mỗi lần thử nghiệm, kết quả thu được là một mặt Pareto theo hai mục tiêu MSE và Comp. Chúng tôi tính toán mặt Pareto trung bình của 30 lần thử nghiệm, tương tự như trong [1, 2, 8]. Thực hiện đối sánh kết quả thu được của thuật toán đề xuất với các thuật toán HA-PAES-MG-Kmax và PAES<sub>KB</sub> tại điểm FIRST của mặt Pareto. Điểm FIRST là điểm tương ứng với hệ luật có  $\overline{MSE}_{Tr}$  nhỏ nhất. Ký hiệu  $\overline{MSE}_{Tr}$  và  $\overline{MSE}_{Ts}$  là các giá trị MSE trung bình lần lượt trên tập dữ liệu huấn luyện (Tr) và tập dữ liệu kiểm tra (Ts),  $\overline{\sigma}_{Ts}$  là phương sai trung bình trên tập kiểm tra, còn Comp và #R lần lượt là độ phức tạp trung bình và số luật trung bình của hệ luật.

Chúng tôi tiến hành phân tích sử dụng phương pháp thống kê phi tham số Wilcoxon theo hai mục tiêu là độ phức tạp Comp và độ chính xác (dựa trên MSE), với mức ý nghĩa  $\alpha = 0,05$ . Kết quả thống kê được trình bày trong các bảng V, VI, và VII. Từ bảng V ta thấy giá trị Exact P-value lớn hơn  $\alpha = 0,05$ , do đó giả thiết  $H_0$  là “độ phức tạp của các hệ luật được tạo ra bởi hai thuật toán là như nhau” được chấp nhận. Như vậy độ phức tạp của hệ luật được xây dựng bởi thuật toán đề xuất trong bài báo này không có sự khác biệt với các thuật toán được so sánh.

Kết quả phân tích bảng VI cho thấy giá trị Exact P-value nhỏ hơn  $\alpha = 0,05$ , do đó giả thiết  $H_0$  là “độ chính xác của các hệ luật trên tập huấn luyện của thuật toán là như nhau” bị loại bỏ. Như vậy có sự khác biệt giữa giá trị MSE của các hệ luật được sinh ra từ thuật toán được đề xuất trong bài báo với giá trị MSE của các hệ luật được sinh ra từ các thuật toán đối sánh. Từ bảng IV ta thấy giá trị MSE của các hệ luật được sinh ra từ thuật toán HA-De-PAES tốt hơn trên hầu hết các bài toán trừ bài toán AB. Kết phân tích trong bảng VII cho thấy giá trị Exact P-value lớn hơn  $\alpha = 0,05$ , do đó giả thiết  $H_0$  là “độ chính xác của các hệ luật trên tập kiểm tra của thuật toán là như nhau” được chấp nhận. Mặc dù không có sự khác biệt giữa độ chính xác trên tập kiểm tra của các hệ luật được sinh ra bởi thuật toán đề xuất trong bài báo này nhưng từ bảng IV chúng ta thấy độ chính xác của thuật toán đề xuất chỉ kém các thuật toán được đối sánh trên một bài toán, tốt hơn trên 8 bài toán. Chúng ta có thể kết luận rằng thuật toán được đề xuất tốt hơn các thuật toán đối sánh trên mục tiêu độ chính xác.

Bảng IV  
SO SÁNH KẾT QUẢ THỬ NGHIỆM THUẬT TOÁN HA-DE-PAES (HADE) VỚI CÁC THUẬT TOÁN HA-PAES-MG-KMAX (HATg), PAES<sub>KB</sub> TẠI ĐIỂM FIRST

Bài toán	#R			Comp			MSE <sub>Tr</sub>			MSE <sub>Ts</sub>			σ <sub>Ts</sub>		
	PAES <sub>KB</sub>	HATg	HADe	PAES <sub>KB</sub>	HATg	HADe	PAES <sub>KB</sub>	HATg	HADe	PAES <sub>KB</sub>	HATg	HADe	PAES <sub>KB</sub>	HATg	HADe
ELE1	27	27,3	27,4	46,0	46,1	52,7	145.995	141.666	<b>141.321</b>	<b>194.028</b>	202.591	201.836	<b>24.745</b>	35.321	30.0234
ELE2	30	29,9	30,0	65,0	67,0	65,1	11,043	8.813	<b>8.504</b>	12.606	10.686	<b>10.372</b>	3.105	3.114	<b>1.771</b>
WA	28	25,0	25,0	103	60,0	71,6	1,64	1,03	<b>1,01</b>	3,92	1,25	<b>1,22</b>	9,27	<b>0,17</b>	<b>0,17</b>
WI	25	24,9	25,0	91,0	61,3	64,2	1,30	0,79	<b>0,77</b>	1,49	0,96	<b>0,95</b>	0,26	<b>0,13</b>	0,14
TR	11	15,0	15,0	40,0	29,4	33,9	0,080	0,031	<b>0,026</b>	0,140	0,045	<b>0,039</b>	0,15	0,02	<b>0,01</b>
AB	29	19,8	22,6	107	59,6	49,1	2,32	<b>2,31</b>	2,43	2,48	<b>2,41</b>	2,68	0,18	<b>0,17</b>	0,20
MTG	12	15,0	13,0	49,0	28,1	28,3	0,050	0,016	<b>0,014</b>	0,090	0,022	<b>0,019</b>	0,10	<b>0,01</b>	<b>0,01</b>
CA	10	13,8	14,5	30,0	44,7	45,6	11,99	4,58	<b>4,09</b>	13,43	4,86	<b>4,81</b>	4,66	0,63	<b>0,55</b>
PT	14	13,3	14,4	53,0	38,3	36,3	87,00	71,89	<b>65,07</b>	89,00	73,47	<b>68,97</b>	25,00	17,02	<b>10,44</b>

Bảng V  
SO SÁNH ĐỘ PHỨC TẠP CỦA HỆ LUẬT SỬ DỤNG WILCOXON TEST VỚI MỨC  $\alpha = 0,05$

So sánh với	R+	R-	Exact P-value	Confidence-interval	Giả thuyết
PAES-KB	37	8	0,09766	[-28,9; -0,55]	Loại bỏ giả thuyết H <sub>0</sub>
HA-Tg	15	30	$\geq 0,2$	[-1,95; 4,85]	Chấp nhận giả thuyết H <sub>0</sub>

Bảng VI  
SO SÁNH SAI SỐ MSE TRÊN TẬP HUẤN LUYỆN SỬ DỤNG WILCOXON TEST VỚI MỨC  $\alpha = 0,05$

So sánh với	R+	R-	Exact P-value	Confidence-interval	Giả thuyết
PAES-KB	42	3	0,019532	[-2.337,315; -0,054]	Loại bỏ giả thuyết H <sub>0</sub>
HA-Tg	40	5	0,03906	[-172,51; -0,002]	Chấp nhận giả thuyết H <sub>0</sub>

Bảng VII  
SO SÁNH SAI SỐ MSE TRÊN TẬP KIỂM TRA SỬ DỤNG WILCOXON-TEST VỚI MỨC  $\alpha = 0,05$

So sánh với	R+	R-	Exact P-value	Confidence-interval	Giả thuyết
PAES-KB	33	12	$\geq 0,2$	[-1.117,0505; 3.902,65]	Loại bỏ giả thuyết H <sub>0</sub>
HA-Tg	39	6	0,05468	[-377,505; 0,11]	Loại bỏ giả thuyết H <sub>0</sub>

## VI. KẾT LUẬN

Bài báo đề xuất một hướng tiếp cận sinh luật mới giải bài toán hồi quy bằng hệ luật mờ. Các phương pháp truyền thống dựa trên lý thuyết tập mờ thường sử dụng phương pháp sinh luật bằng cách tổ hợp các từ ngôn ngữ sử dụng cho mỗi thuộc tính. Với cách tiếp cận này số luật phải xem xét rất lớn. Tiếp cận dựa trên lý thuyết ĐSGT sử dụng phương pháp sinh luật dựa trên mẫu dữ liệu, phương pháp tiếp cận này làm giảm không gian luật cần phải xem xét, tuy nhiên nó lại không tận dụng những thông tin và quan hệ của dữ liệu. Bài báo này chúng tôi đã đề xuất một phương pháp sinh luật tiếp cận theo lý thuyết ĐSGT và cây quyết định. Chúng tôi đã phát triển một thuật toán xây dựng FRBS

gồm hai pha, kết quả thử nghiệm thuật toán cho thấy các mục tiêu độ phức tạp và độ chính xác của hệ luật có thể so sánh được với các thuật toán đã đề xuất.

## LỜI CẢM ƠN

Nghiên cứu này nằm trong khuôn khổ đề tài “Nghiên cứu và phát triển các phương pháp thao tác trực tiếp trên các từ ngôn ngữ dựa trên đại số gia tử để giải quyết một số vấn đề trong các lĩnh vực trích rút tri thức, tăng cường chất lượng ảnh và cơ sở dữ liệu mờ”, mã số 102.01-2017.06, được tài trợ bởi Quỹ phát triển khoa học và công nghệ quốc gia (NAFOSTED).

## TÀI LIỆU THAM KHẢO

- [1] R. Alcalá, P. Ducange, F. Herrera, B. Lazzarini, and F. Marcelloni, "A multiobjective evolutionary approach to concurrently learn rule and data bases of linguistic fuzzy-rule-based systems," *IEEE Transactions on Fuzzy Systems*, vol. 17, no. 5, pp. 1106–1122, Oct. 2009.
- [2] M. Antonelli, P. Ducange, B. Lazzarini, and F. Marcelloni, "Learning concurrently data and rule bases of Mamdani fuzzy rule-based systems by exploiting a novel interpretability index," *Soft Computing*, vol. 15, pp. 1981–1998, 2011.
- [3] H. Ishibuchi and Y. Nojima, "Analysis of interpretability-accuracy tradeoff of fuzzy systems by multiobjective fuzzy genetics-based machine learning," *International Journal of Approximate Reasoning*, vol. 44, no. 1, pp. 4–31, 2007.
- [4] —, "Repeated double cross-validation for choosing a single solution in evolutionary multi-objective fuzzy classifier design," *Knowledge-Based Systems*, vol. 54, pp. 22–31, 2013.
- [5] P. Pulkkinen and H. Koivisto, "Fuzzy classifier identification using decision tree and multiobjective evolutionary algorithms," *International Journal of Approximate Reasoning*, vol. 48, no. 2, pp. 526–543, 2008.
- [6] O. Cordon, M. J. Del Jesus, and F. Herrera, "A proposal on reasoning methods in fuzzy rule-based classification systems," *International Journal of Approximate Reasoning*, vol. 20, no. 1, pp. 21–45, 1999.
- [7] C. H. Nguyen, W. Pedrycz, T. L. Duong, and T. S. Tran, "A genetic design of linguistic terms for fuzzy rule based classifiers," *International Journal of Approximate Reasoning*, vol. 54, no. 1, pp. 1–21, 2013.
- [8] C. H. Nguyen, V. T. Hoang, and V. L. Nguyen, "A discussion on interpretability of linguistic rule based systems and its application to solve regression problems," *Knowledge-Based Systems*, vol. 88, pp. 107–133, 2015.
- [9] C. H. Nguyen, V. Hoang, T. Tran, and V. Nguyen, "LFoC-Interpretability of linguistic rule based systems and its applications to solve regression problems," *International Journal of Computer Technology & Applications*, vol. 8, no. 2, pp. 94–117, 2017.
- [10] F. Aghaeipoor and M. M. Javidi, "On the influence of using fuzzy extensions in linguistic fuzzy rule-based regression systems," *Applied Soft Computing*, vol. 79, pp. 283–299, 2019.
- [11] C. Mencar and A. M. Fanelli, "Interpretability constraints for fuzzy information granulation," *Information Sciences*, vol. 178, no. 24, pp. 4585–4618, 2008.
- [12] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, no. 3, pp. 338–353, 1965.
- [13] N. M. Han, N. C. Hao *et al.*, "An algorithm to building a fuzzy decision tree for data classification problem based on the fuzziness intervals matching," *Journal of Computer Science and Cybernetics*, vol. 32, no. 4, pp. 367–380, 2016.
- [14] X. Liu, X. Feng, and W. Pedrycz, "Extraction of fuzzy rules from fuzzy decision trees: An axiomatic fuzzy sets (AFS) approach," *Data & Knowledge Engineering*, vol. 84, pp. 1–25, 2013.
- [15] N. C. Ho and W. Wechler, "Hedge algebras: An algebraic approach to structure of sets of linguistic truth values," *Fuzzy Sets and Systems*, vol. 35, no. 3, pp. 281–293, 1990.
- [16] J. Han, J. Pei, and M. Kamber, *Data Mining: Concepts and Techniques*. Elsevier, 2011.
- [17] J. D. Knowles and D. W. Corne, "Approximating the non-dominated front using the Pareto archived evolution strategy," *Evolutionary Computation*, vol. 8, pp. 149–172, 2000.



Nguyễn Đức Dư nhận bằng Cử nhân Toán tin ứng dụng và Thạc sĩ Toán ứng dụng tại Trường Đại học Khoa học Tự nhiên, Đại học Quốc gia Hà Nội lần lượt các năm 2001 và 2005. Tác giả hiện đang là nghiên cứu sinh tại Viện Khoa học và Công nghệ Quân sự từ năm 2015, đồng thời là giảng viên Khoa Công nghệ Thông tin, Trường Đại học Giao thông Vận tải. Các lĩnh vực nghiên cứu bao gồm khai phá dữ liệu, lô-gic mờ, hệ mờ, tính toán mềm, tính toán với từ, và học máy.



Hoàng Văn Thông nhận bằng Cử nhân Toán tin ứng dụng và Thạc sĩ Công nghệ Thông tin lần lượt tại Trường Đại học Khoa học Tự nhiên và Trường Đại học Công nghệ, Đại học Quốc gia Hà Nội năm 2001 và 2005. Năm 2016, tác giả nhận bằng Tiến sĩ tại Học viện Khoa học và Công nghệ, Viện Hàn lâm Khoa học Việt Nam. Tác giả hiện đang là giảng viên Khoa Công nghệ Thông tin, Trường Đại học Giao thông Vận tải. Các lĩnh vực nghiên cứu bao gồm khai phá dữ liệu, lô-gic mờ, hệ mờ, tính toán mềm, tính toán với từ, học máy, và trí tuệ nhân tạo.