NEIGHBORHOOD ROUGH SET BASED ATTRIBUTE IN THE DECISION TABLE

Nguyen Xuan Tien[1], Tran Thanh Dai[2*], Trinh Van Ha[3], To Huu Nguyen[3], Nguyen Thi Duyen[3]
[1]*Thai Nguyen University*
[2]*University of Economic and Technical Industries*
[3]*TNU - University of Information and Communication Technology*

ARTICLE INFO

Received: 23/3/2023
Revised: 05/5/2023
Published: 08/5/2023

KEYWORDS

Attribute reduction
Feature selection
Rough set
Neighborhood rough set
Fuzzy rough set

ABSTRACT

Attribute reduction is an essential data preprocessing step widely applied in pattern recognition, recommender systems, and decision support. For attribute reduction with numeric decision tables according to the granular computing approach, the traditional information granular is often shaded or discretized to build measures to evaluate the importance of the attribute. In this paper, we propose a new attribute reduction method including the following steps: 1) discretize data to increase the smoothness of the data after discretization; 2) determine the dependency ratio of the decision attribute with the set of condition attributes; 3) define a reduct and calculate the importance of the attribute to build an attribute reduction algorithm. The experimental results on the sample datasets from UCI show that our proposed method is effective compared to the attribute reduction methods following the traditional fuzzy computing approach.

RÚT GỌN THUỘC TÍNH CHO BẢNG QUYẾT ĐỊNH THEO TIẾP CẬN TẬP THÔ LÂN CẬN

Nguyễn Xuân Tiến[1], Trần Thanh Đại [2*], Trịnh Văn Hà[3], Tô Hữu Nguyên[3], Nguyễn Thị Duyên[3]
[1]*Đại học Thái Nguyên*
[2]*Trường Đại học Kinh tế Kỹ thuật Công nghiệp*
[3]*Trường Đại học Công nghệ thông tin và Truyền thông – ĐH Thái Nguyên*

THÔNG TIN BÀI BÁO

Ngày nhận bài: 23/3/2023
Ngày hoàn thiện: 05/5/2023
Ngày đăng: 08/5/2023

TỪ KHÓA

Rút gọn thuộc tính
Chọn lọc thuộc tính
Tập thô
Tập thô lân cận
Tập thô mờ

TÓM TẮT

Rút gọn thuộc tính là bước tiền xử lý dữ liệu quan trọng và được ứng dụng rộng rãi trong các lĩnh vực nhận dạng mẫu, hệ tư vấn và hỗ trợ ra quyết định. Để rút gọn thuộc tính cho các bảng quyết định miền giá trị số theo tiếp cận tính toán hạt, các hạt thông tin truyền thống thường được mờ hóa hoặc rời rạc hóa để xây dựng các độ đo đánh giá độ quan trọng của thuộc tính. Trong bài báo này chúng tôi đề xuất phương pháp rút gọn thuộc tính mới bao gồm các bước sau đây: 1) mờ hóa quá trình rời rạc dữ liệu nhằm tăng độ mịn dữ liệu sau khi rời rạc hóa; 2) xác định tỉ lệ phụ thuộc của thuộc tính quyết định với tập thuộc tính điều kiện; 3) định nghĩa tập rút gọn và phương pháp tính toán độ quan trọng của thuộc tính để xây dựng thuật toán rút gọn thuộc tính. Các kết quả thực nghiệm trên các bộ dữ liệu mẫu từ UCI cho thấy, phương pháp của chúng tôi đề xuất là hiệu quả so với các phương pháp rút gọn thuộc tính theo tiếp cận tính toán mờ truyền thống.

DOI: https://doi.org/10.34238/tnu-jst.7599

[*] Corresponding author. *Email: ttdai@uneti.edu.vn*

1. Giới thiệu

Mô hình lý thuyết tập thô *(Rough Set - RS)* của Pawlack đã được chứng minh là công cụ hiệu quả, ứng dụng cho các bài toán phân tích dữ liệu nói chung và bài toán rút gọn thuộc tính nói riêng. Trong mô hình *RS*, phép toán xấp xỉ dưới được dùng để khẳng định tính chắc chắn và phép toán xấp xỉ trên dùng để khẳng định tính không chắc chắn của các đối tượng thuộc vào một phân lớp. Do đó, mô hình lý thuyết tập thô là công cụ hiệu quả dùng để phân tích các lớp dữ liệu không nhất quán, thiếu thông tin và không chắc chắn [1]. Đối với bài toán rút gọn thuộc tính cho các bảng quyết định, các nhà nghiên cứu đề xuất độ đo miền dương *(miền khẳng định - POS)* để đánh giá độ phụ thuộc của thuộc tính quyết định với các thuộc tính điều kiện [2] – [9]. Theo góc nhìn tính toán hạt cho bài toán rút gọn thuộc tính, các lớp tương đương trong mô hình tập thô truyền thống được xem như các hạt thông tin trong các mô hình tính toán hạt. Do đó, bên cạnh độ đo POS của mô hình tập thô, độ đo khoảng cách tri thức cũng được đề xuất [10]. Theo tiếp cận độ đo khoảng cách, một số các nghiên cứu mở rộng độ đo này nhằm giải quyết cho các trường hợp dữ liệu nhiễu được đề xuất [11] - [13]. Các kết quả nghiên cứu của bài toán rút gọn thuộc tính đã được ứng dụng rộng rãi cho các lĩnh vực như: y tế và chăm sóc sức khỏe [14] – [16], hỗ trợ ra quyết định [17] và hệ tư vấn [18], [19]. Bên cạnh các mô hình tập thô truyền thống đã được dùng để phân tích dữ liệu cho các bảng dữ liệu có miền giá trị phân loại, các mô hình tập thô mở rộng cũng được nghiên cứu lâu nay cho các bảng dữ liệu có miền giá trị số/ liên tục. Để rút gọn thuộc tính cho các bảng quyết định miền giá trị số, phương pháp ban đầu được các nhà nghiên cứu đưa ra là rời rạc hóa dữ liệu, biến đổi các giá trị liên tục về các giá trị rời rạc, sau đó áp dụng mô hình lý thuyết tập thô và sử dụng độ đo miền dương để tính toán tập rút gọn. Tuy nhiên quá trình rời rạc hóa dữ liệu có thể làm mất mát thông tin ban đầu cũng như làm mất đi tính tự nhiên của các luật thu được [2], [3], do đó các nhà nghiên cứu đề xuất các cách mở rộng mô hình lý thuyết tập thô theo các mở rộng các khái niệm hạt thông tin lân cận [4], [5], hạt thông tin mờ [11], [13] và hạt thông tin mờ trực cảm [12]. Sau đây là hai hướng mở rộng chính của tập thô được các nhà nghiên cứu quan tâm:

Mô hình tập thô lân cận: Trong mô hình này, hầu hết các nhà nghiên cứu chủ yếu tập trung vào định nghĩa lại quan hệ tương đương của mô hình tập thô truyền thống để xây dựng các lớp tương đương (các hạt thông tin), sau đó áp dụng độ đo miền dương của mô hình lý thuyết tập thô truyền thống để tính toán tập rút gọn. Một số các nghiên cứu điển hình về phương pháp rút gọn thuộc tính theo tiếp cận tập thô lân cận như [4] – [6]. Tuy nhiên việc định nghĩa quan hệ lân cận để xây dựng các hạt thông tin lân cận rất khó bao quát được đầy đủ thông tin quan hệ của các đối tượng trong một tập, đôi khi làm mất mát thông tin khi sử dụng các ngưỡng lân cận không phù hợp cho từng bộ dữ liệu thực.

Mô hình tập thô mờ : Trong mô hình này, các lớp tương đương của mô hình tập thô truyền thống được chuyển thành các lớp tương đương mờ thông qua các định nghĩa mới về quan hệ tương đương mờ hoặc quan hệ tương tự. Trong đó quan hệ tương đương mờ phải thỏa mãn các tính chất đối xứng, phản xạ và bắc cầu (min-max). Trên cơ sở các lớp tương đương mờ, Dubois và các cộng sự xây dựng lại các phép toán xấp xỉ trên và xấp xỉ dưới thông qua các toán tử logic mờ như T – chuẩn, T – đối chuẩn và toán tử kéo theo mờ. Khi đó độ đo miền dương được định nghĩa lại là độ đo miền dương mờ, độ đo này được dùng để tính toán tập rút gọn cho bảng quyết định miền giá trị số. Một số nghiên cứu điển hình theo tiếp cận miền dương mờ như [11], [12], [15].

Phần giới thiệu về ứng dụng và các nghiên cứu liên quan được trình bày bên trên của bài báo càng khẳng định vai trò quan trọng của bài toán rút gọn thuộc tính trong các mô hình tiền xử lý dữ liệu hiện nay và nhận được nhiều quan tâm của cộng đồng các nhà nghiên cứu lý thuyết về tập thô. Tuy nhiên các tiếp cận mở rộng tập thô hiện nay chưa xét đến yếu tố mờ của các quan hệ lân cận. Trong bài báo này, chúng tôi xét đến yếu tố mờ của các quan hệ lân cận để xây dựng phương pháp rút gọn thuộc tính mới. Phần còn lại của bài báo bao gồm các phần có cấu trúc như sau: Phần 2 trình bày các định nghĩa để xây dựng công cụ đánh giá độ quan trọng của thuộc tính, trên cơ sở đó

định nghĩa tập rút gọn và đề xuất thuật toán rút gọn thuộc tính. Phần 3 trình bày một số kết quả thực nghiệm trên các bộ dữ liệu thực được tải về từ kho dữ liệu của UCI [20], các kết luận và hướng phát triển tiếp theo trong tương lai được trình bày trong phần cuối cùng của bài báo.

2. Phương pháp nghiên cứu

Định nghĩa 2.1. Cho bảng quyết định $DT=(U,C,D)$. Trong đó, U là tập khác rỗng các đối tượng, C là tập khác rỗng các thuộc tính điều kiện, D là tập khác rỗng các thuộc tính quyết định và $C\cap D=\varnothing$. Khi đó, quan hệ lân cận mờ của hai đối tượng $x,y\in U$ trên thuộc tính $a\in C$ được xác định như sau:

$$N_a^{\delta}(x,y)=\begin{cases} sim_a(x,y)\ if\ \ sim_a(x,y)>\delta \\ 0\ other \end{cases} \quad (1)$$

Trong đó, $sim_a=1-|a(x)-a(y)|$ với $\mathrm{a}(x)$ và $\mathrm{a}(y)$ là các giá trị tương ứng của đối tượng x và y trên thuộc tính a. Đại lượng δ là ngưỡng lân cận được chọn theo kinh nghiệm của người sử dụng dữ liệu. Trong nội dung chi tiết của thuật toán, chúng tôi sẽ đề xuất công thức tính tự động giá trị δ này. Khi đó, cặp (U,N_a^{δ}) được gọi là không gian xấp xỉ lân cận mờ, dùng để mô tả lân cận mờ của mọi $x\in U$ trên thuộc tính $a\in C$ hay còn được gọi là phủ lân cận mờ các đối tượng của U trên thuộc tính a.

Định nghĩa 2.2. Cho không gian xấp xỉ lân cận mờ (U,N_a^{δ}) xác định trên bảng quyết định $DT=(U,C,D)$, với $a\subseteq C$. Khi đó, phủ lân cận mờ các đối tượng của U trên a có thể biểu diễn dưới dạng ma trận quan hệ lân cận mờ như sau: $\mathcal{M}_a^{\delta}=\left[a_{(i,j)}^{\delta}\right]_{U\times U}$. Trong đó, $a_{(i,j)}^{\delta}$ là độ tương tự của hai đối tượng $i,j\in U$ trên thuộc tính a.

Định nghĩa 2.3. Cho bảng quyết định $DT=(U,C,D)$, và hai ma trận quan hệ lân cận mờ $\mathcal{M}_a^{\delta}$ và $\mathcal{M}_b^{\delta}$ với $a,b\in C$. Khi đó, ma trận quan hệ lân cận mờ của tập thuộc tính $\{a,b\}$ được xác định như sau:

$$\mathcal{M}_{\{a,b\}}^{\delta}=\left[\min\left(a_{(i,j)}^{\delta},b_{(i,j)}^{\delta}\right)\right]_{U\times U} \quad (2)$$

Định nghĩa 2.4. Cho hạt thông tin lân cận mờ $[x]_a^{\delta}$ xác định trên bảng quyết định $DT=(U,C,D)$ với $x\in U$ và $a\in C$. Khi đó xác suất của x thuộc về X với $X\subseteq U$ được xác định như sau:

$$P_a^{(\delta,\varepsilon)}[x](X)=\begin{cases} p\ if\ \ p>\varepsilon \\ 0\ other \end{cases} \quad (3)$$

Trong đó, $p=\dfrac{\left|[x]_a^{\delta}\cap X\right|}{\left|[x]_a^{\delta}\right|}$ và $\varepsilon=\dfrac{1}{|D|}$.

Định nghĩa 2.5. Cho bảng quyết định $DT=(U,C,D)$ và ma trận quan hệ lân cận mờ $\mathcal{M}_a^{\delta}$ với $a\in C$. Khi đó xác suất phụ thuộc của thuộc tính D theo thuộc tính a được xác định như sau:

$$P_a^{(\delta,\varepsilon)}(D)=\frac{1}{|U|}\sum_{x\in U}P_a^{(\delta,\varepsilon)}[x](D) \quad (4)$$

Định nghĩa 2.6. Cho bảng quyết định $DT=(U,C,D)$ và $B\subseteq C$. Độ quan trọng của thuộc tính $a\in C-B$ với B được xác định như sau:

$$Sig_B(a) = P_{\{a\cup B\}}^{(\delta,\varepsilon)}(D) - P_B^{(\delta,\varepsilon)}(D) \quad (5)$$

Định nghĩa 2.7. Cho bảng quyết định $DT = (U, C, D)$ và $B \subseteq C$. Khi đó B được gọi là tập rút gọn của C khi và chỉ khi:

(1) $P_{\{a\cup B\}}^{(\delta,\varepsilon)}(D) - P_B^{(\delta,\varepsilon)}(D) = 0$ với mọi $a \in C - B$

(2) $P_B^{(\delta,\varepsilon)}(D) - P_{\{B-b\}}^{(\delta,\varepsilon)}(D) > 0$ với mọi $b \in B$

Dựa trên các định nghĩa về độ đo độ quan trọng của thuộc tính và định nghĩa về tập rút gọn theo tỉ lệ phân lớp của hạt thông tin lân cận mờ, sau đây là thuật toán rút gọn thuộc tính theo tiếp cận filter thuộc tính được đề xuất.

Algorithm F_GNF (Filter based on Fuzzy Neighbhoot Granular):

Input: Bảng quyết định $DT = (U, C, D)$ với $\delta = \frac{1}{|D|} \times \frac{\min(|C|, |U|)}{max(|C|, |U|)}$

Output: Một tập rút gọn B của DT

Begin

// init phase

$B = \varnothing$; Khởi tạo $\mathcal{M}_R = [1]_{U\times U}$;

// Giai đoạn lọc (Filter)

1: **for each** $c \in C \cup D$

2: Tính $\mathcal{M}_c^{\delta}$ sử dụng công thức (1);

3: **end for**

4: **while** (**true**)

5: **for each** $c \in C - B$

6: **calulate** $Sig_B(c)$ sử dụng công thức (5);

7: **end for**

8: $c_m = -1$

9: **chọn** $c_m \in C - B \,\big|\, Sig_B(c_m) = \underset{c\in C-B}{Max}\{Sig_R(c)\}$;

10: **if** $c_m \neq -1$ **then** $B = B \cup \{c_m\}$

11: **else** *break*;

12: **update** $\mathcal{M}_R$ dựa trên công thức (2);

13: **end while;**

14: **return** B ;

End.

3. Một số kết quả thực nghiệm

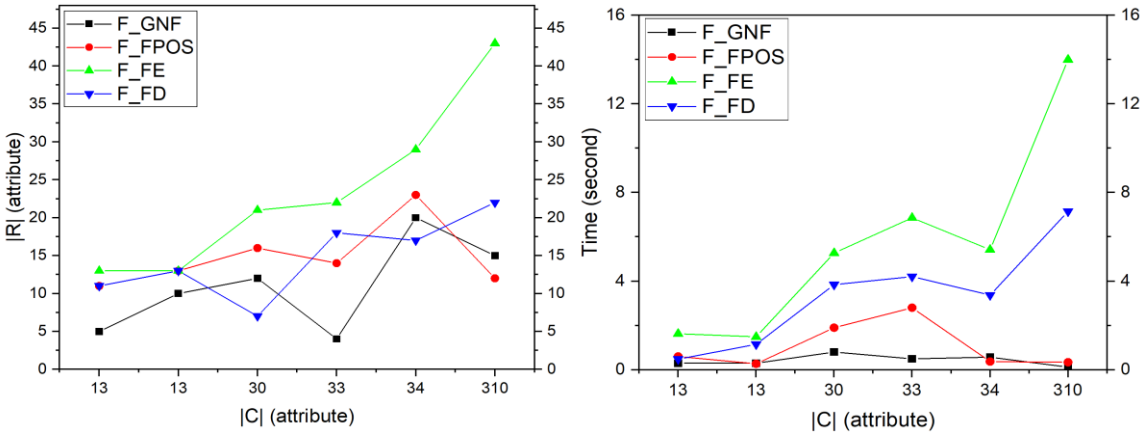
Việc thực nghiệm nhằm đánh giá tính hiệu quả của thuật toán đề xuất bằng cách so sánh thuật toán đề xuất F_GNF với các thuật toán F_FPOS [13], F_FE [19], và F_FD [11]. Tất cả các thuật toán trên đều được xây dựng theo tiếp cận Filter thuộc tính và sử dụng tập nền FS để xây dựng các hạt thông tin mờ. Môi trường thực nghiệm cho các thuật toán được thực hiện trên nền hệ điều hành Window 10, CPU core i5 Processor – RAM 8Gb. Tất cả các thuật toán đều được cài đặt bằng ngôn ngữ lập trình Python và thực nghiệm với 6 bộ dữ liệu thực từ UCI [20]. Chi tiết về các tập dữ liệu được mô tả trong Bảng 1 trong đó |U| cho biết số lượng đối tượng, |C| cho biết số lượng thuộc tính điều kiện, |D| cho biết số lượng phân lớp của thuộc tính quyết định. Để đánh giá các thuật toán, chúng tôi sử dụng ba tiêu chí để đánh giá đó là: 1) kích thước của tập rút gọn (số thuộc tính); 2) độ chính xác phân lớp %; 3) thời gian thực hiện của thuật toán (s). Trong đó độ

chính xác phân lớp được thống kê trên hai mô hình phân lớp SVM và KNN (K=10, p=2), độ đo độ chính xác *accuracy* kết hợp với phương pháp đánh giá chéo 10-fold được sử dụng trên từng mô hình. Các tập dữ liệu trước khi rút gọn được chuẩn hóa về miền giá trị [0,1].

Bảng 1. *Mô tả dữ liệu thực nghiệm*

ID	Data	Mô tả	|U|	|C|	|D|
1	Wine	Wine	178	13	3
2	Heart	Statlog (Heart)	270	13	2
3	Wdbc	Breast Cancer Wisconsin (Diagnostic)	569	30	2
4	Wpbc	Breast Cancer Wisconsin (Prognostic)	198	33	2
5	Iono	Ionosphere	351	34	2
6	LVB	Voice Rehabilitation (Binary)	126	310	2

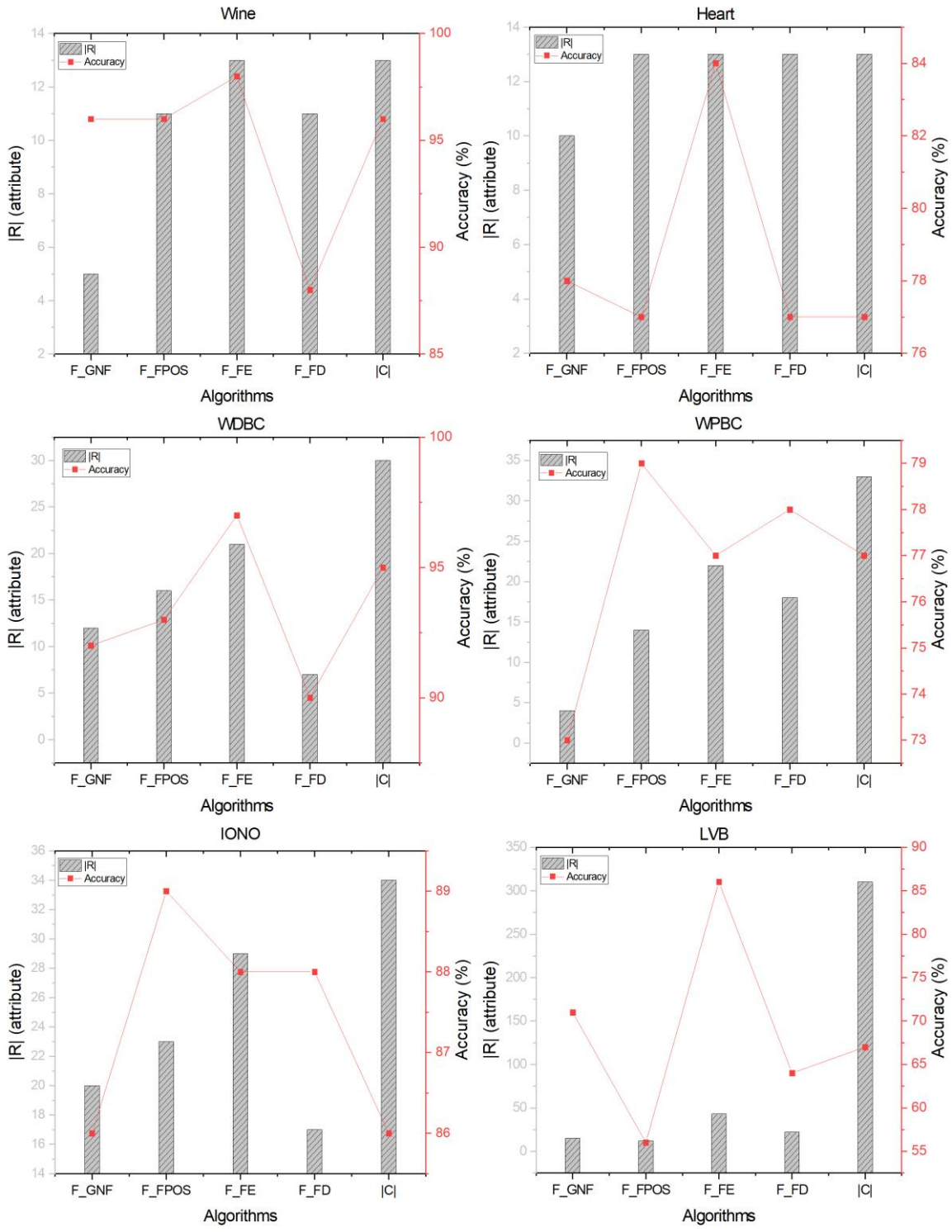
Như chúng ta có thể quan sát thấy tại Hình 1, thuật toán đề xuất F_GNF có thời gian tính toán nhanh hơn các thuật toán khác trên hầu hết các bộ dữ liệu, điều này có thể được lý giải do độ đo tính toán độ quan trọng của thuộc tính chỉ được tính toán trên các lân cận của tập mờ mà không phải trên toàn bộ tập mờ của không gian các quan hệ. Hơn nữa, kích thước tập rút gọn nhỏ hơn cũng đồng nghĩa với điều kiện dừng của thuật toán đề xuất đến sớm hơn so với các thuật toán khác. Ngoài ra, hầu hết các thuật toán rút gọn hiện nay đều sử dụng tính chất bảo toàn thông tin của tập rút gọn so với tập thuộc tính gốc nên độ phức tạp thường là $O\left(|C|^2|U|^2\right)$. Trong khi đó, độ phức tạp của thuật toán đề xuất phụ thuộc vào kích thước của tập rút gọn thu được, do đó độ phức tạp của thuật toán đề xuất là $O\left(|R||C-R||U|^2\right)$ trong đó |R| là kích thước của tập rút gọn, |C| là kích thước của tập thuộc tính ban đầu và |U| là kích thước của tập dữ liệu.



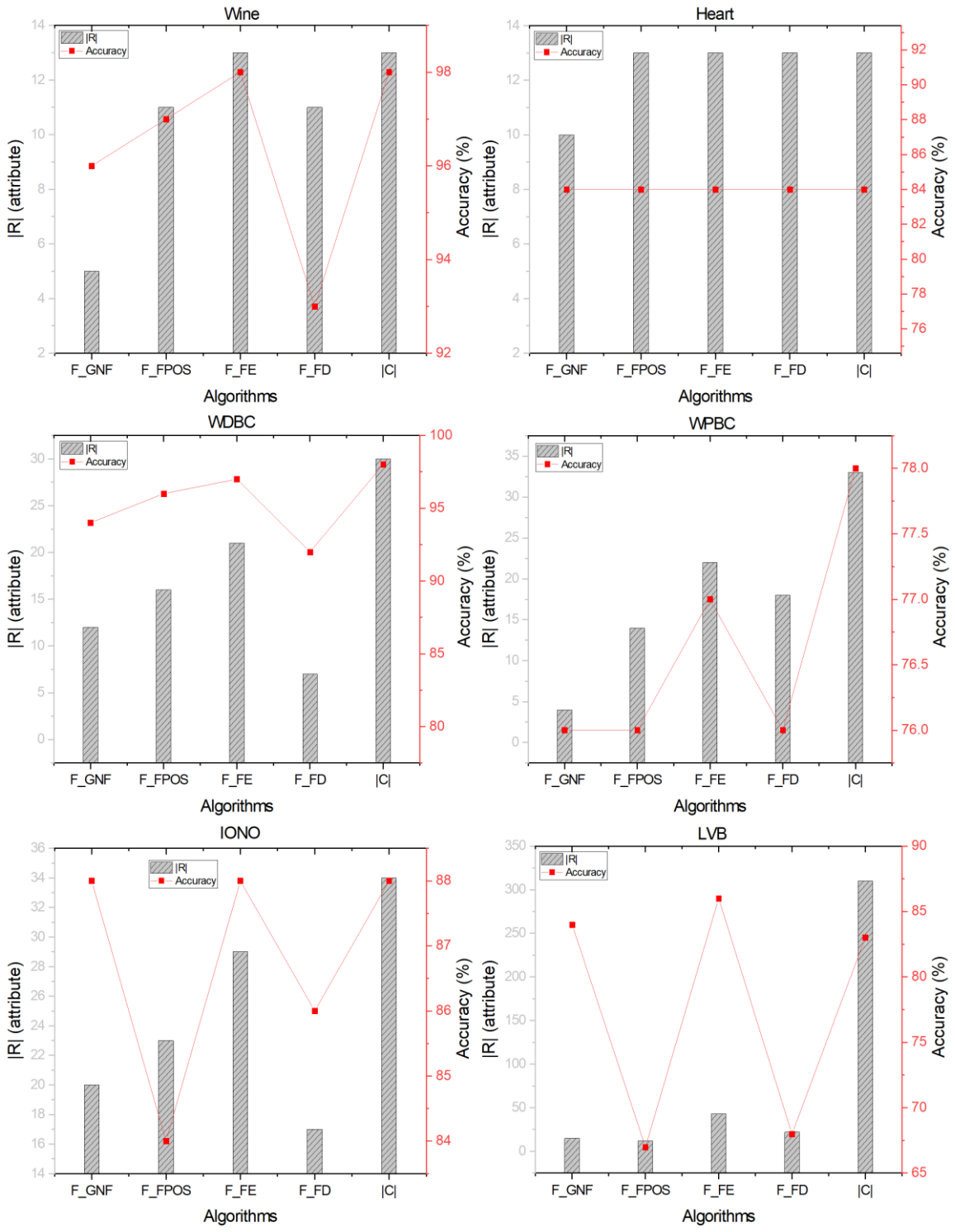
Hình 1. *Mối quan hệ về kích thước của tập rút gọn và thời gian tính toán với |C|*

Các Hình số 2 và số 3 mô tả mối quan hệ giữa kích thước của tập rút gọn thu được từ các thuật toán và độ chính xác phân lớp. Trong đó, hình số 2 mô tả các tập dữ liệu được đánh giá trên mô hình phân lớp KNN và hình số 3 mô tả các tập dữ liệu được đánh giá trên mô hình phân lớp SVM.

Các kết quả phân tích trên cả hai mô hình cho thấy thuật toán đề xuất là hiệu quả tương đương với các thuật toán được sử dụng để so sánh. Tuy nhiên trên các bộ dữ liệu tốt như Wine hay WDBC, thuật toán đề xuất cho ra tập rút gọn có kích thước nhỏ hơn so với các thuật toán khác.



Hình 2. *Độ chính xác phân lớp và kích thước của tập rút gọn thu được từ mỗi thuật toán trên từng bộ dữ liệu được đánh giá với mô hình phân lớp KNN*



Hình 3. *Độ chính xác phân lớp và kích thước của tập rút gọn thu được từ mỗi thuật toán trên từng bộ dữ liệu được đánh giá với mô hình phân lớp SVM*

4. Kết luận

Trong bài báo này chúng tôi đề xuất phương pháp tính toán tập rút gọn mới dựa trên phương pháp tính toán tỉ lệ phân lớp của hạt thông tin lân cận mờ. Trong đó đề xuất công thức xây dựng hạt thông tin lân cận mờ, đề xuất phương pháp tính tỉ lệ phân lớp của hạt thông tin và công thức tính độ phụ thuộc của thuộc tính. Trên cơ sở đó, định nghĩa lại độ quan trọng của thuộc tính và tập rút gọn. Các kết quả thực nghiệm cho thấy phương pháp đề xuất là hiệu quả hơn với các phương pháp rút gọn thuộc tính khác về thời gian tính toán và các bộ dữ liệu tốt cho ra tập rút gọn có kích thước nhỏ hơn. Nghiên cứu tiếp theo trong tương lai chúng tôi tập trung nghiên cứu vấn đề xử lý nhiễu trên các bộ dữ liệu xấu.

TÀI LIỆU THAM KHẢO/ REFERENCES

[1] X. Rao, X. Yang, X. Yang, X. Chen, D. Liu, and Y. Qian, "Quickly calculating reduct: An attribute relationship based approach," *Knowledge-Based Syst.*, vol. 200, p. 106014, 2020.

[2] K. Liu, X. Yang, H. Fujita, D. Liu, X. Yang, and Y. Qian, "An efficient selector for multi-granularity attribute reduction," *Inf. Sci. (Ny).*, vol. 505, pp. 457–472, 2019.

[3] Y. Chen, K. Liu, J. Song, H. Fujita, X. Yang, and Y. Qian, "Attribute group for attribute reduction," *Inf. Sci. (Ny).*, vol. 535, pp. 64–80, 2020.

[4] M. Hu, E. C. C. Tsang, Y. Guo, D. Chen, and W. Xu, "Attribute reduction based on overlap degree and k-nearest-neighbor rough sets in decision information systems," *Inf. Sci. (Ny).*, vol. 584, pp. 301–324, 2022.

[5] M. Hu, E. C. C. Tsang, Y. Guo, D. Chen, and W. Xu, "A novel approach to attribute reduction based on weighted neighborhood rough sets," *Knowledge-Based Syst.*, vol. 220, pp. 106908, 2021.

[6] Y. Guo, M. Hu, X. Wang, E. C. C. Tsang, D. Chen, and W. Xu, "A robust approach to attribute reduction based on double fuzzy consistency measure," *Knowledge-Based Syst.*, vol. 253, p. 109585, 2022.

[7] X. Fan, Q. Chen, Z. Qiao, C. Wang, and M. Ten, "Attribute reduction for multi-label classification based on labels of positive region," *Soft Comput.*, vol. 24, pp. 14039–14049, 2020.

[8] J. Yang, Q. Zhang, and Q. Xie, "Attribute reduction based on misclassification cost in variable precision rough set model," *J. Intell. Fuzzy Syst.*, vol. 37, pp. 5129–5142, 2019.

[9] J. Xie, B. Q. Hu, and H. Jiang, "A novel method to attribute reduction based on weighted neighborhood probabilistic rough sets," *Int. J. Approx. Reason.*, vol. 144, pp. 1–17, 2022.

[10] G. Liu and Y. Feng, "Knowledge granularity reduction for decision tables," *Int. J. Mach. Learn. Cybern.*, vol. 13, pp. 569–577, 2022.

[11] L. G. Nguyen and L. Son, "Novel Incremental Algorithms for Attribute Reduction from Dynamic Decision Tables Using Hybrid Filter-Wrapper with Fuzzy Partition Distance," *IEEE Trans. Fuzzy Syst.*, vol. 28, pp. 858–873, 2020.

[12] T. T. Nguyen, L. G. Nguyen, and T. D. Tran, "A Novel Filter-Wrapper Algorithm on Intuitionistic Fuzzy Set for Attribute Reduction from Decision Tables," *International Journal of Data Warehousing and Mining*, vol. 17, pp. 67–100, 2021.

[13] N. T. Nguyen and S. Wongthanavasu, "Hybrid filter–wrapper attribute selection with alpha-level fuzzy rough sets," *Expert Syst. Appl.*, vol. 193, p. 116428, 2022.

[14] B. Remeseiro and V. Bolon-Canedo, "A review of feature selection methods in medical applications," *Computers in Biology and Medicine*, vol. 112, pp. 119–121, 2019.

[15] H. Garg and G. Kaur, "Novel distance measures for cubic intuitionistic fuzzy sets and their applications to pattern recognitions and medical diagnosis," *Granul. Comput.*, vol. 5, pp. 169–184, 2020.

[16] P. Pattaraintakorn and N. Cercone, "Integrating rough set theory and medical applications," *Appl. Math. Lett.*, vol. 21, pp. 400–403, 2008.

[17] M. Riaz, F. Karaaslan, I. Nawaz, and M. Sohail, "Soft multi-rough set topology with applications to multi-criteria decision-making problems," *Soft Comput.*, vol. 25, pp. 799–815, 2021.

[18] A. J. Fernández-García, L. Iribarne, A. Corral, J. Criado, and J. Z. Wang, "A recommender system for component-based applications using machine learning techniques," *Knowledge-Based Syst.*, vol. 164, pp. 68–84, 2019.

[19] B. Saravanan, V. Mohanraj, and J. Senthilkumar, "A fuzzy entropy technique for dimensionality reduction in recommender systems using deep learning," *Soft Comput.*, vol. 23, pp. 2575–2583, 2019.

[20] UCI, "Machine learning repository," 2021. [Online]. Available: https://archive.ics.uci.edu/ml/index.php. [Accessed on 26 March 2023].