DOI: 10.15625/vap.2020.00174

MỘT CẢI TIẾN MỚI CỦA THUẬT TOÁN PHÂN CỤM BÁN GIÁM SÁT MỜ VỚI NHIỀU THAM SỐ MỜ VÀ ỨNG DỤNG TRONG PHÂN ĐOẠN ẢNH NHA KHOA

Trần Mạnh Tuấn[1], Phùng Thế Huân[2], Nguyễn Bắc Hùng[3], Lê Hoàng Sơn[4], Trần Đình Khang[3], Từ Thu Nga[5]

[1]Khoa Công nghệ thông tin, Trường Đại học Thủy lợi
[2]Trường Đại học Công nghệ thông tin và Truyền thông, Đại học Thái Nguyên
[3]Viện Công nghệ thông tin và Truyền thông, Trường Đại học Bách khoa Hà Nội
[4]Viện Công nghệ thông tin, Đại học Quốc gia Hà Nội
[5]Trường THPT Kim Liên, Đống Đa, Hà Nội

tmtuan@tlu.edu.vn, pthuan@ictu.edu.vn, hungking98@gmail.com, sonlh@vnu.edu.vn, khangtd@soict.hust.edu.vn, tuthunga@gmail.com

TÓM TẮT: *Phân cụm dữ liệu có nhiều ứng dụng trong thực tiễn, tuy nhiên trong nhiều trường hợp một lượng nhỏ thông tin bổ trợ được cung cấp hỗ trợ cho quá trình phân cụm. Đây chính là ý tưởng của phân cụm bán giám sát. Thuật toán phân cụm bán giám sát mờ được xây dựng dựa trên thuật toán phân cụm mờ kết hợp với thông tin bổ trợ bao gồm: các dữ liệu đã được gán nhãn (labeled data), các ràng buộc (constraints) như must-link(u,v) và cannot-link(u,v) và thông tin hàm thuộc nhằm điều chỉnh các thành phần trong cụm để làm tăng chất lượng cụm. Bài báo này đề xuất một thuật toán phân cụm bán giám sát mờ mới với nhiều tham số mờ và ứng dụng thuật toán đề xuất trong phân đoạn ảnh nha khoa. Kết quả thực nghiệm trên dữ liệu chuẩn UCI và dữ liệu thực tại Trường Đại học Y Hà Nội cho thấy thuật toán phân cụm bán giám sát mờ mới có hiệu năng tốt hơn so với các thuật toán phân cụm bán giám sát mờ liên quan.*

Từ khóa: *Phân cụm bán giám sát mờ, tham số mờ, ảnh nha khoa, phân đoạn ảnh, hiệu năng thuật toán.*

I. GIỚI THIỆU

Phân cụm dữ liệu là quá trình phân chia dữ liệu thành các cụm khác nhau, với các điểm dữ liệu trong cùng một cụm có độ tương đồng cao và các điểm dữ liệu ở các cụm khác nhau độ tương đồng thấp [1, 10]. Người ta chia phân cụm dữ liệu thành 2 loại cơ bản phân cụm cứng và phân cụm mờ. Phân cụm cứng thì một điểm dữ liệu chỉ có thể thuộc về một cụm. Phân cụm mờ thì một điểm dữ liệu có thể thuộc về nhiều cụm khác nhau với các độ thuộc khác nhau. Trong nhiều trường hợp một lượng nhỏ thông tin bổ trợ được cung cấp hỗ trợ cho quá trình phân cụm. Đây chính là ý tưởng của **phân cụm bán giám sát.**

Trong một số nghiên cứu gần đây, HaitaoGan cùng cộng sự [3] đề xuất phương pháp phân cụm bán giám sát an toàn với trọng số về độ tin cậy. Mô hình của các tác giả nhằm xác định độ an toàn của mỗi mẫu vì trong dữ liệu, từng mẫu sẽ có ảnh hưởng khác nhau đến kết quả thực hiện mô hình. Nghĩa là, một mẫu có kết quả phân cụm chính xác cao thì có độ tin cậy cao. Với 3 giai đoạn thực hiện, mô hình xác định các mẫu được gán nhãn sẽ được khai thác an toàn. Hiệu quả của mô hình được đánh giá bằng cách so sánh với các phương pháp phân cụm không giám sát và các phương pháp phân cụm bán giám sát khác trên các bộ dữ liệu cụ thể. Toshiaki Kondo cùng cộng sự [6] đề xuất một phương pháp tự động để phân đoạn ảnh răng từ hình ảnh số hóa 3 chiều được chụp bởi máy quét laser. Phương pháp này sẽ giảm được sự phức tạp của việc xử lý trực tiếp lưới dữ liệu 3 chiều bằng cách phát hiện các đặc tính trên hai vùng hình ảnh từ hình ảnh 3 chiều. Li và các cộng sự [7] thuật toán phân cụm bán giám sát mờ rất hiệu quả trong nhiều lĩnh vực như xử lý ảnh.

Trong phân đoạn ảnh, Dhanachandra và cộng sự [2] đã sử dụng phân cụm trong phân đoạn ảnh, Lê Hoàng Sơn và cộng sự [12, 13] cũng đã chỉ ra phương pháp phân cụm bán giám sát mờ cho hiệu năng tốt hơn so với một số thuật toán cùng loại. OmaimaNomir cùng cộng sự [9] đưa ra một hệ thống tự động xử lý phân đoạn hình ảnh X-quang nha khoa. Hệ thống tự động phân đoạn hình ảnh nha khoa X-quang dựa vào từng răng và loại bỏ đường viền của mỗi chiếc răng. Đây là một phương pháp mới được phát triển cho phân tách răng dựa trên phép chiếu tích phân. Kết quả thử nghiệm trên một cơ sở dữ liệu nhỏ ảnh X-quang nha khoa là đáng khích lệ. Shuo Li và cộng sự [10] đã đề xuất một khung tự động phân đoạn đa cấp độ trong môi trường máy tính hỗ trợ phân tích ảnh X-quang nha khoa. Abdolvahab Ehsani Rad cùng cộng sự [11] đã đưa ra một số cách tiếp cận khác nhau trong phân đoạn hình ảnh được sử dụng để phân tích hình ảnh X-quang nha khoa, từ đó có được kết quả phù hợp, cần phải thực hiện phương pháp phân đoạn chính xác và hiệu quả, các phương pháp này đã được thử nghiệm trong phân đoạn hình ảnh X-quang. Các nghiên cứu này tập trung vào việc điều chỉnh các thành phần trong cụm để làm tăng hiệu năng từ đó làm tăng chất lượng của phân đoạn ảnh. Trong đó có một yếu tố ảnh hưởng đến chất lượng cụm là tham số mờ chưa được đề cập nghiên cứu nhiều. Trong gian gần đây, Trần Đình Khang và cộng sự [5] đã nghiên cứu đề cập đến việc lựa chọn thích nghi tham số mờ với từng điểm dữ liệu để làm tăng chất lượng cụm.

Trong nghiên cứu này, chúng tôi **đưa ra cải tiến thuật toán phân cụm bán giám sát mờ chuẩn với nhiều tham số mờ** với mục tiêu tăng cường chất lượng cụm và áp dụng thuật toán này cho bài toán phân đoạn ảnh nha khoa. Để thực hiện điều này, chúng tôi xây dựng một mô hình toán học dưới dạng bài toán tối ưu và sử dụng các thông tin bổ trợ để cải thiện chất lượng phân đoạn ảnh. Chất lượng của quá trình phân đoạn ảnh tốt giúp cho quá trình xử lý ảnh

được chính xác hơn. Do vậy chúng tôi tập trung ứng dụng thuật toán đề xuất vào phân đoạn ảnh X-quang nha khoa nhằm mục đích trợ giúp quá trình hỗ trợ chẩn đoán được chính xác hơn.

II. PHÂN CỤM BÁN GIÁM SÁT MỜ

2.1. Thuật toán phân cụm mờ Fuzzy C-Means (FCM)

Thuật toán phân cụm mờ (FCM) được Bezdek [1] dựa trên việc tối ưu hóa khoảng cách các điểm dữ liệu với tâm. Hàm mục tiêu được xác định bởi công thức (1):

$$J(U,V) = \sum_{i=1}^{N}\sum_{j=1}^{C} u_{ij}^{m} \left\|X_i - V_j\right\|^2 \rightarrow Min \quad (1)$$

Trong đó: X là tập dữ liệu nguồn $\{X_1, X_2, \dots, X_k, \dots, X_N\}$ với N điểm dữ liệu, C là số cụm, V là tập các tâm cụm $\{V_1, V_2, \dots, V_j, \dots, V_C\}$, u_{ij} là độ thuộc của phần tử i vào cụm j, m tham số mờ.

Với điều kiện ràng buộc (2):

$$\sum_{j=1}^{C} u_{ij} = 1 \qquad \forall k = \overline{1,N} \quad (2)$$

Dựa vào phương pháp nhân tử Lagrange với hàm mục tiêu (1) và điều kiện (2) ta tính được tâm V theo công thức (3), ma trận độ thuộc U theo công thức (4)

$$V_j = \frac{\sum_{i=1}^{N} {u_{ij}}^m X_i}{\sum_{i=1}^{N} {u_{ij}}^m} \quad (3)$$

$$u_{ij} = \frac{1}{\sum_{k=1}^{C}\left(\frac{\left\|X_i - V_j\right\|}{\left\|X_i - V_k\right\|}\right)^{\frac{2}{m-1}}} \quad (4)$$

Thuật toán phân cụm mờ (Fuzzy C-Mean Clustering) được thực hiện theo Bảng 1:

Bảng 1. Thuật toán Fuzzy C-Mean Clustering

Input	Tập dữ liệu X gồm N phần tử , số cụm C, ngưỡng ε, số lần lặp tối đa maxStep > 0.
Output	Ma trận U và tâm cụm V.
FCM	
1	t=0
2	Khởi tạo ngẫu nhiên V^t
3	Repeat
4	t=t+1
5	Tính ma trận U^t dựa trên công thức (4)
6	Tính ma trận V^t dựa trên công thức (3)
7	Until $\left\|V^{(t)} - V^{(t-1)}\right\| \geq \varepsilon$ or t > maxStep

2.2. Phân cụm bán giám sát mờ

Thuật toán phân cụm bán giám sát mờ được xây dựng dựa trên các thuật toán phân cụm mờ kết hợp với các thông tin bổ trợ được người dùng cung cấp. Thông tin bổ trợ [16] cho phân cụm bán giám sát mờ có 3 dạng cơ bản gồm các ràng buộc Must-link và Cannot-link; các nhãn lớp của một phần dữ liệu và độ thuộc được xác định trước.

Thuật toán phân cụm bán giám sát mờ chuẩn được Yasunori [15] đề xuất dựa trên phân cụm mờ (FCM) kết hợp với thông tin bổ trợ $\bar{U}$ được xác định trước, có hàm mục tiêu như công thức (5)

$$J(U,V) = \sum_{i=1}^{N}\sum_{j=1}^{C} |u_{ij} - \bar{u}_{ij}|^m \, ||X_i - V_j||^2 \rightarrow min \quad (5)$$

Khi đó có bổ sung thêm điều kiện rằng buộc của thông tin bổ trợ: $\sum_{j=1}^{C} \overline{u_{ij}} \leq 1 \left(\forall k = \overline{1,N}\right)$

Dựa vào phương pháp Lagrange với hàm mục tiêu (5) và điều kiện (2) ta tính được tâm V theo công thức (6), ma trận độ thuộc U theo công thức (7) với m>1, công thức (8) khi m=1

$$V_j = \frac{\sum_{i=1}^{N} |u_{ij} - \bar{u}_{ij}|^m X_i}{\sum_{i=1}^{N} |u_{ij} - \bar{u}_{ij}|^m} \quad , j = \overline{1,C} \quad (6)$$

Khi m > 1:

$$u_{ij} = \bar{u}_{ij} + (1 - \sum_{k=1}^{C} \bar{u}_{ik}) \frac{\left(\frac{1}{\left\|X_i - V_j\right\|}\right)^{\frac{2}{m-1}}}{\sum_{k=1}^{C}\left(\frac{1}{\left\|X_i - V_k\right\|}\right)^{\frac{2}{m-1}}} , i = \overline{1,N}, j = \overline{1,C} \quad (7)$$

Khi m=1:

$$u_{ij} = \begin{cases} \overline{u}_{ij} + 1 - \sum_{k=1}^{C} \overline{u}_{ik}, khi\ i = arg\ \min_k \|X_i - V_k\|^2 \\ \overline{u}_{ij}, khi\ i \neq arg\ \min_k \|X_i - V_k\|^2 \end{cases} \quad (8)$$

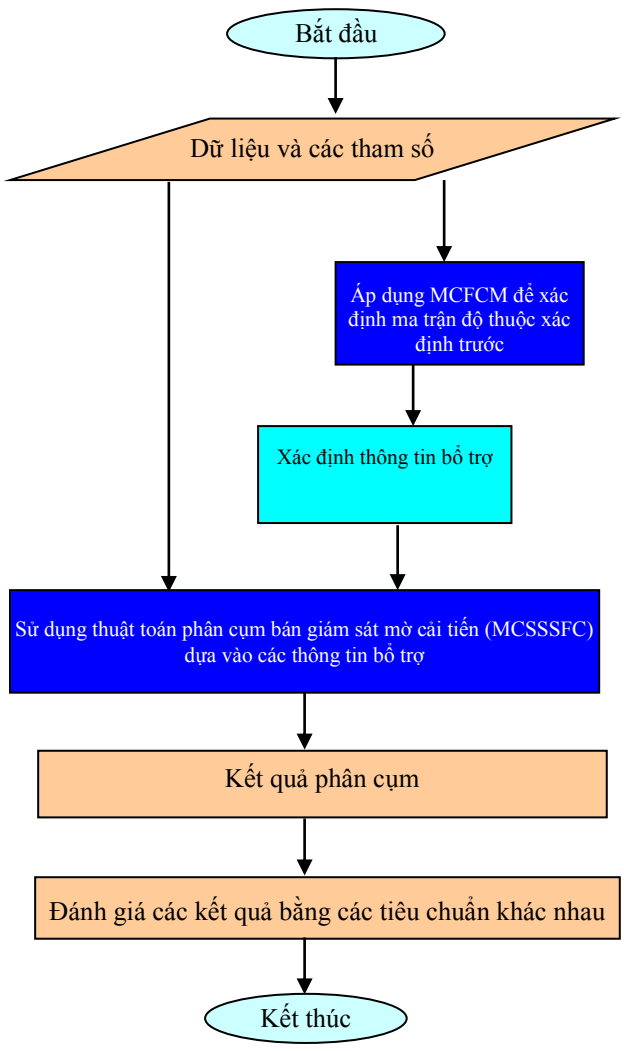
Thuật toán phân cụm bán giám sát mờ (Semi-Supervised Standard Fuzzy Clustering-SSSFC) được trình bày trong Bảng 2:

Bảng 2. Thuật toán Semi-Supervised Standard Fuzzy Clustering

Input	Tập dữ liệu X gồm N phần tử , số cụm C, ma trận độ thuộc bổ trợ $\overline{U}$, ngưỡng ε, số lần lặp tối đa maxStep > 0.
Output	Ma trận U và tâm cụm V.
SSSFC	
1	t=0
2	Khởi tạo ngẫu nhiên V^t
3	Repeat
4	t=t+1
5	Tính ma trận U^t nếu m=1 dựa trên công thức (8), nếu m>1 dựa vào công thức (7)
6	Tính ma trận V^t dựa trên công thức (6)
7	Until $\left\|V^{(t)} - V^{(t-1)}\right\| \geq \varepsilon$ or t > MaxStep

III. MÔ HÌNH CẢI TIẾN PHÂN CỤM BÁN GIÁM SÁT MỜ VỚI NHIỀU THAM SỐ MỜ (MCSSSFC)

3.1. Mô hình cải tiến phân cụm bán giám sát mờ với nhiều tham số mờ



Hình 1. Lược đồ phân cụm đề xuất

Mô hình đề xuất được xác định Hình 1. Sử dụng phân cụm MCFCM (nội dung chi tiết thuật toán MCFCM được trình bày ở Mục 3.2) thu được được ma trận độ thuộc U từ đó xác định thông tin bổ trợ cho thuật toán MCSSSFC bằng cách tại mỗi điểm dữ liệu ta giữ lại các giá trị hàm thuộc lớn nhất của từng điểm dữ liệu vào các cụm, các giá trị còn lại được xác định bằng 0. Sau đó áp dụng thuật toán MCSSSFC được trình bày chi tiết ở Mục 3.3. Cuối cùng xác định các độ đo đánh giá chất lượng của quá trình phân cụm.

3.2. Thuật toán phân cụm mờ với nhiều tham số mờ (Fuzzy C–Means Clustering with Multiple Fuzzification Coefficients - MCFCM)

Trong thuật toán phân cụm mờ với nhiều tham số mờ được Trần Đình Khang và công sự [5] xây dựng dựa trên thuật toán phân cụm mờ với mỗi mỗi điểm dữ liệu xây dựng một tham số mờ riêng. Khi đó việc xác định tham số mờ được xác định dựa trên [m_1, m_2]. Khi đó thuật toán xác định tham số mờ cho từng điểm dữ liệu (MC) được trình bày trong Bảng 3. Thuật toán phân cụm mờ với nhiều tham số mờ được trình bày trong Bảng 4.

Bảng 3. Thuật toán MC Xác định tham số mờ

Input	Tập dữ liệu X gồm N phần tử, C: số cụm, m_1, m_2 là các giá trị khoản của tham số mờ, một tham số α
Output	Ma trận M
MC	
1	Xác định ma trận khoảng cách D giữa các điểm dữ liệu, D_{ij} là khoảng cách từ điểm i đến điểm j. $D_{ij} = \|X_i - X_j\| \ (\forall i,j = \overline{1,N})$
2	Xác định ma trận khoảng cách sắp xếp D', tại mỗi điển dữ liệu i sắp xếp theo thứ tự tăng dần theo khoảng cách (khi đó D_i ->D_i'). Tính tổng khoảng cách đến [N/C] điểm gần nhất được S, với các S_i: $S_i = \sum_{j=1}^{N/C} D'_{ij}$
3	Xác định các m_i như sau: $m_i = m_1 + (m_2 - m_1)\left(\frac{S_i - S_{min}}{S_{max} - S_{min}}\right)^{\alpha}$ Trong đó: $S_{max} = max_{i\in N}(S_i); S_{min} = min_{i\in N}(S_i)$

Bảng 4. Thuật toán Fuzzy C-Mean Clustering

Input	Tập dữ liệu X gồm N phần tử , số cụm C, m_1,m_2 là các giá trị khoản của tham số mờ, một tham số α, ngưỡng ε, số lần lặp tối đa maxStep > 0.
Output	Ma trận U và tâm cụm V.
MCFCM	
1	Tính tham số mờ m_i cho các điểm dữ liệu từ thuật toán MC
2	t=0
3	Khởi tạo ngẫu nhiên V^t
4	Repeat
5	t=t+1
6	Tính ma trận U^t dựa trên công thức $u_{ij} = \frac{1}{\sum_{j=1}^{C}\left(\frac{\|X_i - V_k\|}{\|X_i - V_j\|}\right)^{\frac{2}{m_i - 1}}}$
7	Tính ma trận V^t dựa trên công thức $V_k = \frac{\sum_{i=1}^{N} u_{ik}{}^{m_i} X_i}{\sum_{k=1}^{N} u_{kj}{}^{m_i}}$
8	Until $\|V^{(t)} - V^{(t-1)}\| \geq \varepsilon$ or t > MaxStep

3.3. Thuật toán phân cụm bán giám sát mờ chuẩn với nhiều tham số mờ (Semi-Supervised Standard Fuzzy Clustering with Multiple Fuzzification Coefficients - MCSSSFC)

Trong thuật toán này, việc xác định tham số mờ khác nhau với các điểm dữ liệu khác nhau được trình bày ở phần 3.3. Khi đó hàm mục tiêu được xây dựng như công thức (9)

$$J(U,V) = \sum_{i=1}^{N}\sum_{j=1}^{C} |u_{ij} - \bar{u}_{ij}|^{m_i} ||X_i - V_j||^2 \rightarrow min \quad (9)$$

Các điều kiện ràng buộc của phân cụm bán giám sát mờ. Dựa vào phương pháp Lagrange với hàm mục tiêu (9) và điều kiện (2) ta tính được tâm V theo công thức (10), ma trận độ thuộc U theo công thức (11) với m>1, công thức (12) với m=1

$$V_j = \frac{\sum_{i=1}^{N} |u_{ij} - \bar{u}_{ij}|^{m_i} X_i}{\sum_{i=1}^{N} |u_{ij} - \bar{u}_{ij}|^{m_i}} \quad , j = \overline{1,C} \quad (10)$$

Khi m_i>1

$$u_{ij} = \bar{u}_{ij} + (1 - \sum_{k=1}^{C} \bar{u}_{ik}) \frac{\left(\frac{1}{\|X_i - V_j\|}\right)^{\frac{2}{m_i - 1}}}{\sum_{k=1}^{C}\left(\frac{1}{\|X_i - V_k\|}\right)^{\frac{2}{m_i - 1}}} , i = \overline{1,N}, j = \overline{1,C} \quad (11)$$

Khi m_i=1

$$u_{ij} = \begin{cases} \overline{u}_{ij} + 1 - \sum_{k=1}^{C} \overline{u}_{ik}, khi\ i = arg\ \min_k \|X_i - V_k\|^2 \\ \overline{u}_{ij}, khi\ i \neq arg\ \min_k \|X_i - V_k\|^2 \end{cases} \quad (12)$$

Thuật toán phân cụm bán giám sát mờ (Semi-Supervised Standard Fuzzy Clustering-SSSFC) được trình bày trong Bảng 5

Bảng 5. Thuật toán MCSSSFC

Input	Tập dữ liệu X gồm N phần tử , số cụm C, ma trận độ thuộc bổ trợ $\overline{U}$, m_1, m_2 là các giá trị khoảng của tham số mờ, một tham số α, ngưỡng ε, số lần lặp tối đa maxStep > 0.
Output	Ma trận U và tâm cụm V.
MCSSSFC	
1	Tính tham số mờ m_i cho các điểm dữ liệu từ thuật toán MC trong bảng 3
2	t=0
3	Khởi tạo ngẫu nhiên V^t
4	Repeat
5	t=t+1
6	Tính ma trận U^t nếu m=1 dựa trên công thức (12), nếu m>1 dựa vào công thức (11)
7	Tính ma trận V^t dựa trên công thức (10)
8	Until $\left\|V^{(t)} - V^{(t-1)}\right\| \geq \varepsilon$ or t > MaxStep

IV. KẾT QUẢ THỰC NGHIỆM

Dữ liệu thực nghiêm được là bộ dữ liệu Wine lấy trên kho dữ liệu chuẩn UCI Machine Learning Repository [17] và dữ liệu dựa trên bộ dữ liệu thu thập thực tế tại Viện Đào tạo răng hàm mặt, Trường Đại học Y Hà Nội. Số lượng các ảnh thu thập là 152 ảnh nha khoa từ 2018-2019 về bệnh viêm quanh cuống. Loại ảnh thu thập là ảnh X-quang chóp răng được chụp bởi hệ thống máy X-quang kỹ thuật số không dây của hãng Aceton.

Các độ đo dùng để đánh giá và so sánh hiệu năng của các thuật toán được cài đặt trong bài báo này gồm DB [14], SSWC [14], PBM [14], IFV[4]. Thuật toán đề xuất - phân cụm bán giám sát mờ với nhiều tham số mờ (MCSSSFC) được cài đặt cùng với thuật toán phân cụm mờ bán giám sát chuẩn (SSSFC [15]).

Bảng 6. Kết quả thực nghiệm trên bộ dữ liệu Wine

	SSSFC	**MCSSSFC**
DB	1,1838	**0,5471**
SSWC	0,4519	**0,5578**
PBM	230121,2772	**652680,6457**
IFV	39,8980	**569,6760**

Kết quả thực nghiệm, đánh giá dựa trên các độ đo đánh giá hiệu năng giữa thuật toán phân cụm bán giám sát mờ nhiều tham số mờ (trình bày phần 3) với các thuật toán phân cụm cùng loại trên bộ dữ liệu wine, với số cụm là 4 thể hiện ở Bảng 6. Dựa trên 4 độ đo đánh giá hiệu năng của thuật toán thì độ đo DB, SSWC, PBM thuật toán cải tiến cho giá trị tốt nhất thuật toán phân cụm bán giám sát mờ chuẩn.

Chúng tôi thực nghiệm trên ảnh X-quang nha khoa với bài toán phân đoạn ảnh, hình ảnh phân đoạn so sánh (Hình 2). Kết quả trung bình của các ảnh phân đoạn của các độ đo và các phương pháp được xác định ở Bảng 7. Dựa trên 4 độ đo đánh giá hiệu năng của thuật toán thì thuật toán cải tiến cho giá trị tốt hơn thuật toán phân cụm bán giám sát mờ chuẩn.

a) Ảnh gốc

b) Ảnh phân đoạn bằng MCSSSFC

Hình 2. Kết quả thực nghiệm

Bảng 7. Kết quả thực nghiệm trên bộ dữ liệu nha khoa

	SSSFC	MCSSSFC
DB	1,4363± 0,593	**1,3542± 0,562**
SSWC	5129,643± 435,3	**5432,365± 325,3**
PBM	58983,67 ± 1,24E+02	**64324,34 ± 1,54E+02**
IFV	32,16±3,54	**38,98±3,26**

V. KẾT LUẬN

Trong bài báo này, chúng tôi nghiên cứu cải tiến thuật toán phân cụm bán giám sát mờ với nhiều tham số mờ. Đóng góp chính của bài báo là đề xuất được một thuật toán phân cụm bán giám sát mờ với nhiều tham số mờ (MCSSSFC) được trình bày chi tiết trong Phần III và cài đặt thực nghiệm thuật toán cải tiến trên bộ dữ liệu chuẩn UCI và trên bộ dữ liệu gồm các ảnh X-quang nha khoa. Bài báo, đánh giá và so sánh thuật toán cải tiến với một số thuật toán cùng loại trên các độ đo đánh giá. Kết quả thu được của bài báo: i) chúng tôi đã cải tiến một thuật toán phân cụm bán giám sát mờ với nhiều tham số số mờ phù hợp với từng điểm dữ liệu; ii) cài đặt, thực nghiệm thuật toán cải tiến trên bộ dữ liệu wine và bộ dữ liệu ảnh nha khoa; iii) đánh giá so sánh thuật toán cải tiến với các thuật toán SSSFC, MCFCM thì hiệu năng của thuật toán cải tiến tốt hơn.

Từ nghiên cứu này, đưa ra một số hướng nghiên cứu có thể phát triển trong tương lai: nghiên cứu tiếp với xây dựng các hàm số tính toán với tham số mờ, áp dụng thuật toán cải tiến kết hợp với các đặc trưng nha khoa để xây dựng hệ hỗ trợ chẩn đoán nha khoa từ hình ảnh.

VI. LỜI CẢM ƠN

Nghiên cứu này được tài trợ bởi Đại học Quốc Gia Hà Nội trong Đề tài mã số QG.20.51.

TÀI LIỆU THAM KHẢO

[1]. Bezdek, J. C. Pattern recognition with fuzzy objective function algorithms. Kluwer Academic Publishers, (1981).

[2]. Dhanachandra, N., Chanu, Y. J., & Singh, K. M. "A new hybrid image segmentation approach using clustering and black hole algorithm", Computational Intelligence, 2020.

[3]. Gan, H., Fan, Y., Luo, Z., Huang, R., & Yang, Z. "Confidence-weighted safe semi-supervised clustering", Engineering Applications of Artificial Intelligence, 81, pp. 107-116, 2019.

[4]. Hu, C., Meng, L., & Shi, W. "Fuzzy clustering validity for spatial data", Geo-spatial information science, 11(3), pp. 191-196, 2008.

[5]. Khang, T. D., Vuong, N. D., Tran, M. K., & Fowler, M. "Fuzzy C-Means Clustering Algorithm with Multiple Fuzzification Coefficients," Algorithms, 13(7), pp. 158, 2020.

[6]. Kondo, T., Ong, S. H., & Foong, K. W. "Tooth segmentation of dental study models using range images", IEEE Transactions on medical imaging, 23(3), pp. 350-362, 2004.

[7]. Li, J., Bioucas-Dias, J. M., & Plaza, A. "Semisupervised hyperspectral image segmentation using multinomial logistic regression with active learning", Geoscience and Remote Sensing, IEEE Transactions on, 48(11), pp. 4085-4098, 2010.

[8]. Li, S., Fevens, T., Krzyżak, A., & Li, S. "An automatic variational level set segmentation framework for computer aided dental X-rays analysis in clinical environments", Computerized Medical Imaging and Graphics, 30(2), pp. 65-74, 2006.

[9]. Nomir, O., & Abdel-Mottaleb, M. "A system for human identification from X-ray dental radiographs", Pattern Recognition, 38(8), pp. 1295-1305, 2005.

[10]. Salem Saleh Al-amri, N.V. Kalyankar and Khamitkar S.D, "Image Segmentation by Using Thershod Techniques", Journal of computing, 2(5), pp. 83-86, 2010.

[11]. Rad, A. E., Mohd Rahim, M. S., Rehman, A., Altameem, A., & Saba, T. "Evaluation of current dental radiographs segmentation approaches in computer-aided applications", IETE Technical Review, 30(3), pp. 210-222, 2013.

[12]. Tuan, T. M., & Son, L. H. "Dental segmentation from X-ray images using semi-supervised fuzzy clustering with spatial constraints", Engineering Applications of Artificial Intelligence, 59, pp. 186-195, 2017.

[13]. Tuan, T. M., Ngan, T. T., & Son, L. H. "A novel semi-supervised fuzzy clustering method based on interactive fuzzy satisficing for dental X-ray image segmentation", Applied Intelligence, 45(2), pp. 402-428, 2016.

[14]. Vendramin, L., Campello, R. J., & Hruschka, E. R. "Relative clustering validity criteria: A comparative overview", Statistical Analysis and Data Mining: The ASA Data Science Journal, 3(4), pp. 209-235, 2010.

[15]. Yasunori, E., Yukihiro, H., Makito, Y., & Sadaaki, M. "On semi-supervised fuzzy c-means clustering. In Fuzzy Systems", 2009. FUZZ-IEEE 2009. IEEE International Conference on (pp. 1119-1124). IEEE, 2009, August.

[16]. Zhang, H., & Lu, J. "Semi-supervised fuzzy clustering: A kernel-based approach", Knowledge-Based Systems, 22(6), pp. 477-481, 2009.

[17]. UCI Machine Learning Repository. https://archive.ics.uci.edu/ml/datasets.php.

AN IMPROVEMENT OF SEMI-SUPERVISED FUZZY CLUSTERING WITH MULTIPLE FUZZFICATION COEFFICIENTS AND APPLICATIONS FOR DENTAL IMAGE SEGMENTATION

Tran Manh Tuan, Phung The Huan, Nguyen Bac Hung, Le Hoang Son, Tran Dinh Khang, Tu Thu Nga

ABSTRACT: *Semi-supervised clustering algorithms are used in many real applications where a small amount of information is provided to support for clustering progress. This is a huge advantage of semi-supervised clustering. Semi-supervised fuzzy clustering algorithms were set based on fuzzy clustering algorithms combining with additional information such as must – link, cannot – link constraints, labeled a part of data or pre-defined membership function. Additional information was used to adjust cluster components in order to increase cluster quality. In this paper, a novel semi-supervised fuzzy clustering algorithm with multiple fuzzfication coefficients is proposed and applied to dental image segmentation. The experimental results on UCI datasets and real dataset collected from Hanoi University of Medical Hospital show that proposed algorithm has better performance comparing with other related semi-supervised clustering algorithms.*

Keywords: *Semi-supervised clustering, fuzzfication coefficients, dental images, image segmentation, algorithm performance.*