

PHÁT HIỆN SẠT LỞ TỪ ẢNH VỆ TINH SỬ DỤNG PHƯƠNG PHÁP PHÂN CỤM MỜ

Lê Anh Tú^{1*}, Phùng Thế Huân²

¹*Trường Đại học Hạ Long*

²*Trường Đại học Công nghệ thông tin và Truyền thông, Đại học Thái Nguyên*

* *Email: leanhtu@daihochalong.edu.vn*

Ngày nhận bài: 26/9/2022

Ngày nhận bài sửa sau phản biện: 13/11/2022

Ngày chấp nhận đăng: 17/11/2022

TÓM TẮT

Việt Nam là quốc gia có địa hình trung du và miền núi chiếm diện tích lớn. Cấu trúc địa hình khá đa dạng và phức tạp, có sự phân bậc rõ rệt theo độ cao. Những năm gần đây, tình trạng sạt lở đang diễn ra hết sức nghiêm trọng, có chiều hướng gia tăng cả về phạm vi và quy mô, đe dọa đến an toàn, tài sản của người dân. Vì vậy, việc phát hiện kịp thời các khu vực sạt lở là vô cùng quan trọng để giảm thiểu các thiệt hại. Có nhiều phương pháp để phát hiện sạt lở, trong đó việc sử dụng ảnh vệ tinh cũng là một phương án đáng được quan tâm. Trong nội dung nghiên cứu này, nhóm tác giả giới thiệu tổng quan về các thuật toán phân cụm mờ và ứng dụng cho bài toán phát hiện sạt lở từ ảnh vệ tinh. Kết quả thu được sau khi cài đặt thực nghiệm cho thấy, các phương pháp được giới thiệu có thể áp dụng để giải quyết bài toán phát hiện các khu vực sạt lở trong ảnh vệ tinh. Trong đó, phương pháp phân cụm trên tập mờ viễn cảnh đem lại kết quả với độ chính xác phân cụm cao và hình ảnh phát hiện sạt lở tương đối rõ ràng. Điều này rất cần thiết trong lĩnh vực phòng chống thiên tai, phát hiện sạt lở. Qua đó giúp nhanh chóng hỗ trợ các quyết định và đưa ra các giải pháp để ngăn chặn những thiệt hại do sạt lở gây ra đối với con người và tài sản.

Từ khóa: phân cụm, phân cụm bán giám sát mờ, phân cụm mờ viễn cảnh, sạt lở.

LANDSLIDE DETECTION FROM SATELLITE IMAGES USING FUZZY CLUSTERING METHODS

ABSTRACT

Vietnam is primarily a midland and mountainous country. The topographical structure is quite diverse and complex, with a clear altitude hierarchy. In recent years, landslides are taking place very seriously, and tend to increase in both scope and scale, threatening people's lives and properties. Therefore, timely detection of landslide areas is extremely important to minimize damage. There are numerous methods for detecting landslide areas, one of which is the use of satellite images. In this paper, we introduce some fuzzy clustering methods and applications to solve the problem of landslide detection from satellite images. The results of the experiments show that the methods presented can be used to solve the problem of detecting landslide areas in satellite images. Typically, the Picture fuzzy clustering method has high clustering efficiency, providing relatively clear landslide detection images. It is very useful for disaster prevention and landslide detection. Thereby, decisions and solutions are made quickly to avoid potential damage to people and property.

Keywords: clustering, landslide, picture fuzzy clustering, semi-supervised fuzzy clustering.

1. ĐẶT VẤN ĐỀ

Đối với hệ thống xử lý ảnh thì quá trình phân đoạn ảnh đóng vai trò rất quan trọng. Quá trình phân đoạn ảnh nhằm phân tích hình ảnh thành các vùng có cùng tính chất dựa trên ranh giới hay các khu vực liên thông. Phân đoạn ảnh phân vùng hình ảnh kỹ thuật số thành nhiều phân đoạn nhằm xác định chính xác các đối tượng thuộc vùng ảnh và vùng nền. Theo một cách khác, phân đoạn ảnh là quá trình xác định nhãn cho tất cả các điểm ảnh trong hình ảnh với yêu cầu các điểm ảnh có cùng nhãn thì cùng thuộc một phần hình ảnh nhất định. Phân đoạn ảnh là quá trình quan trọng và phức tạp nhất trong hệ thống xử lý ảnh, phục vụ cho các bước nhận dạng đối tượng trong ảnh sau này. Nhìn chung, có thể phân chia các phương pháp phân đoạn ảnh theo ba xu hướng cơ bản: (i) Sử dụng các kỹ thuật xử lý ảnh gồm: phương pháp lấy ngưỡng, các phương pháp dựa trên biên và dựa trên vùng; (ii) Sử dụng các kỹ thuật phân cụm gồm: K-means (Kanungo và cs., 2002), Fuzzy C-Means (FCM) (Bezdek và cs., 1984); (iii) Sử dụng các kỹ thuật dựa trên tính toán mềm (Shibai và cs., 2018) các kỹ thuật dựa trên học sâu (Shi và cs., 2021).

Quá trình phân nhóm tất cả các điểm dữ liệu về các cụm khác nhau sao cho sau khi phân cụm thì các điểm dữ liệu có mức độ giống nhau hơn sẽ ở trong cùng một cụm được gọi là kỹ thuật phân cụm dữ liệu. Các phương pháp phân cụm được chia ra thành hai loại đó là phương pháp phân cụm rõ và phương pháp phân cụm mờ. Nếu như trong phương pháp phân cụm rõ thì một điểm dữ liệu chắc chắn chỉ thuộc trong một cụm duy nhất; ngược lại, trong phương pháp phân cụm mờ thì một điểm dữ liệu có thể thuộc vào nhiều cụm khác nhau với giá trị độ thuộc vào các cụm. Để nâng chất lượng phân cụm, cụm mờ được bổ sung một số thông tin cho trước như: độ thuộc, các ràng buộc theo cặp, v.v.. Trong trường hợp này, phân cụm mờ trở thành phân cụm bán giám sát mờ. Do có thêm các thông tin bổ trợ nên nhìn chung các phương pháp phân cụm bán giám sát mờ đem lại kết quả độ chính xác phân cụm tốt hơn các

phương pháp phân cụm rõ và phân cụm mờ. Ngoài ra cũng có thể tăng hiệu suất của các phương pháp phân cụm bằng cách sử dụng các kỹ thuật phân cụm dựa trên các tập mờ nâng cao. Ngày nay, các lĩnh vực ứng dụng của phân cụm dữ liệu đã được sử dụng rất nhiều trong thực tế. Có thể kể đến một số ứng dụng điển hình như: trong marketing nhằm phân nhóm khách hàng để xác định và phân loại khách hàng tiềm năng, dự đoán hành vi của khách hàng (Ramasubbareddy và cs., 2020); trong sinh học để phân nhóm động, thực vật dựa trên các đặc tính của chúng (Wu và cs., 2020); trong thư viện để theo dõi người đọc và sách (Rahim và cs., 2021); bảo hiểm (Shi và cs., 2021); tài chính (Majumdar & Laha, 2020); phân loại tài liệu (Curiskis và cs., 2020); phân loại người dùng web (Chen và cs., 2020); phân đoạn X-quang nha khoa (Narayana & Kolli, 2021) và phân đoạn hình ảnh y tế (Kaifa và cs., 2020). Đặc biệt, trong xử lý ảnh vệ tinh có nhiều lĩnh vực ứng dụng quan trọng như: phân loại các phương tiện giao thông (Cao và cs., 2016); hậu cần (Li và cs., 2020); phân loại ảnh vệ tinh để dự báo hạn hán, lũ lụt, cháy rừng (Knorr và cs., 2011), v.v..

Trong phạm vi bài báo này, chúng tôi giới thiệu một số tiếp cận trong phân cụm mờ và ứng dụng vào phát hiện sạt lở từ ảnh vệ tinh. Các nội dung trình bày tiếp theo như sau: nội dung phần 2 giới thiệu một số phương pháp phân cụm mờ; phần 3 đưa ra ứng dụng các phương pháp phân cụm trong phát hiện sạt lở; phần 4 trình bày kết luận và các đóng góp của nghiên cứu, đồng thời đưa ra xu hướng nghiên cứu trong tương lai.

2. MỘT SỐ KỸ THUẬT PHÂN CỤM MỜ

2.1. Phương pháp FCM (Fuzzy C-Means)

FCM là phương pháp phân cụm mờ nổi tiếng nhất được đề xuất bởi Bezdek và cs. (1984). Phương pháp này được đề xuất trên ý tưởng cực tiểu hóa khoảng cách giữa các thành phần dữ liệu tới các trung tâm cụm. Với việc xây dựng hàm mục tiêu

$$J_{(u,v)} = \sum_{a=1}^N \sum_{b=1}^C u_{ab}^m d_{ab}^2 \rightarrow Min, \quad (1)$$

và ràng buộc

$$\sum_{b=1}^C u_{ab} = 1; u_{ab} \in [0,1], \forall a = 1, \dots, N, \quad (2)$$

trong đó $X = \{X_1, X_2, \dots, X_k, \dots, X_N\}$ là tập dữ liệu bao gồm N phần tử, C là số lượng cụm, $V = \{V_1, V_2, \dots, V_b, \dots, V_C\}$ là tập các tâm cụm, u_{ab} là độ thuộc của điểm dữ liệu a vào cụm b , d_{ab} là khoảng cách từ điểm dữ liệu a đến tâm cụm b , m là số mờ.

Tâm cụm V_b và độ thuộc u_{ab} được xác định bằng cách sử dụng phương pháp nhân tử Lagrange đối với hàm mục tiêu (1) và ràng buộc (2) như sau:

$$V_b = \frac{\sum_{a=1}^N u_{ab}^m X_a}{\sum_{a=1}^N u_{ab}^m}, \quad (3)$$

$$u_{ab} = \frac{1}{\sum_{k=1}^C \left(\frac{d_{ab}}{d_{ak}}\right)^{\frac{2}{m-1}}}. \quad (4)$$

2.2. Phương pháp SSFCM (Semi-Supervised Fuzzy C-Means)

Phương pháp SSFCM đã được đề xuất trong một nghiên cứu về phân cụm mờ của nhóm tác giả Pedrycz & Waletzky (1997). Trong đó, hàm mục tiêu của SSFCM bao gồm hai thành phần: đầu tiên là thành phần học không giám sát và thứ hai là thành phần học có giám sát. Phương pháp này có hàm mục tiêu được xác định như sau:

$$J_{(u,d)} = \sum_{a=1}^N \sum_{b=1}^C u_{ab}^m d_{ab}^2 + \alpha \sum_{a=1}^N \sum_{b=1}^C (u_{ab} - f_{ab} b_b)^m d_{ab}^2 \rightarrow Min, \quad (5)$$

trong đó $X = \{X_1, X_2, \dots, X_a, \dots, X_N\}$ là tập dữ liệu bao gồm N phần tử, C là số lượng cụm, $V = \{V_1, V_2, \dots, V_b, \dots, V_C\}$ là tập các tâm cụm, u_{ab} là độ thuộc của phần tử dữ liệu a đối với cụm b , d_{ab} là khoảng cách từ phần tử dữ liệu a đến tâm cụm b , độ thuộc f_{ab} được gán nhãn a trong cụm b , m là số mờ, α là tham số dùng điều chỉnh giữa hai thành phần giám sát và không giám sát, δ_a để phân biệt các phần tử dữ liệu gán nhãn và chưa gán nhãn.

$$f_{ab} = \begin{cases} 1 & \text{nếu } X_a \text{ được gán nhãn} \\ 0 & \text{nếu ngược lại} \end{cases} \quad (6)$$

$$\delta_a = \begin{cases} 1 & \text{nếu } X_a \text{ được gán nhãn} \\ 0 & \text{nếu ngược lại} \end{cases} \quad (7)$$

Các giá trị tâm cụm V_b và các độ thuộc u_{ab} xác định như sau:

$$V_b = \frac{\sum_{a=1}^N u_{ab}^2 X_a + \alpha \sum_{a=1}^N (u_{ab} - f_{ab} \delta_a)^2 X_a}{\sum_{a=1}^N u_{ab}^2 + \alpha \sum_{a=1}^N (u_{ab} - f_{ab} \delta_a)^2 X_a}, \quad \forall b = 1, \dots, C; \quad (8)$$

$$u_{ab} = \frac{1}{1 + \alpha} \left\{ \frac{1 + \alpha(1 - b_b \sum_{b=1}^C f_{ab})}{\sum_{j=1}^C \frac{d_{ab}^2}{d_{bj}^2}} + \alpha f_{ab} b_k \right\}, \quad \forall a = 1, \dots, N, b = 1, \dots, C. \quad (9)$$

2.3. Phân cụm trên tập mờ nâng cao

Để nâng cao hiệu suất phân cụm, ngoài các kỹ thuật can thiệp vào để sửa đổi các thành phần của hàm mục tiêu, một hướng tiếp cận khác cũng đang được nghiên cứu đó là áp dụng các kỹ thuật phân cụm dựa trên các tập mờ nâng cao. Các tập mờ nâng cao như tập mờ loại 1, tập mờ loại 2, tập mờ trực cảm và tập mờ viễn cảnh. Dưới đây trình bày một số định nghĩa:

Tập mờ truyền thống được định nghĩa bởi Zadeh (1996) như sau:

Định nghĩa 1. Một tập mờ được định nghĩa như sau:

$$\dot{A} = \{(\alpha, \mu_{\dot{A}}(\alpha)) | \forall \alpha \in X, \mu_{\dot{A}}(\alpha) \in [0,1]\}, \quad (10)$$

trong đó X là tập nền không rỗng, $\mu_{\dot{A}}(\alpha)$ là độ thuộc của $\alpha \in X$.

Tập mờ viễn cảnh được đề xuất bởi tác giả Bùi Công Cường (2014) dựa trên việc mở rộng tập mờ Zadeh.

Định nghĩa 2. Một tập mờ viễn cảnh được định nghĩa như sau:

$$A = \{(\alpha, \mu_A(\alpha), \eta_A(\alpha), \gamma_A(\alpha)) | \alpha \in X\}, \quad (11)$$

trong đó X là tập nền không rỗng, $\mu_A(\alpha)$ là độ khẳng định, $\eta_A(\alpha)$ là độ do dự và $\gamma_A(\alpha)$ là độ phủ định của mỗi phần tử $\alpha \in X$.

Các ràng buộc:

$$0 \leq \mu_A(\alpha), \eta_A(\alpha), \gamma_A(\alpha) \leq 1,$$

$$0 \leq \mu_A(\alpha) + \eta_A(\alpha) + \gamma_A(\alpha) \leq 1. \quad (12)$$

Mức độ từ chối của một phần tử được tính như sau:

$$\xi_A(\alpha) = 1 - (\mu_A(\alpha) + \eta_A(\alpha) + \gamma_A(\alpha)). \quad (13)$$

Nếu $\xi_A(\alpha) = \eta_A(\alpha) = 0$ thì tập mờ viễn cảnh trở thành tập mờ truyền thống.

Phương pháp phân cụm mờ viễn cảnh (Picture Fuzzy Clustering - FC-PFS)

Trong một nghiên cứu gần đây, nhóm tác giả (Thong & Son, 2016) đã đề xuất một phương pháp phân cụm dựa trên tập mờ viễn cảnh FC-PFS. Phương pháp này có hàm mục tiêu được xác định như sau:

$$J_{(u, \eta, \xi, d)} = \sum_{a=1}^N \sum_{b=1}^C (\mu_{ab}(2 - \xi_{ab}))^m d_{ab}^2 + \sum_{a=1}^N \sum_{b=1}^C \eta_{ab}(\log \eta_{ab} + \xi_{ab}) \rightarrow \text{Min}. \quad (14)$$

Các ràng buộc:

$$0 \leq \mu_{ab}, \eta_{ab}, \xi_{ab} \leq 1, 1 \leq a \leq N, 1 \leq b \leq C,$$

$$0 \leq \mu_{ab} + \eta_{ab} + \xi_{ab} \leq 1; \quad (15)$$

$$\sum_{b=1}^C \mu_{ab}(2 - \xi_{ab}) = 1, 1 \leq a \leq N; \quad (16)$$

$$\sum_{b=1}^C (\eta_{ab} + \frac{1}{C} \xi_{ab}) = 1, 1 \leq a \leq N. \quad (17)$$

Sử dụng phương pháp nhân tử Lagrange các tác giả đã xác định giá trị tâm cụm và các độ thuộc như sau:

– Tâm cụm:

$$v_b = \frac{\sum_{a=1}^N (\mu_{ab}(2 - \xi_{ab}))^m X_a}{\sum_{a=1}^N (\mu_{ab}(2 - \xi_{ab}))^m}; \quad (18)$$

– Các độ thuộc:

$$\mu_{ab} = \frac{1}{(2 - \xi_{ab}) \sum_{k=1}^C (\frac{d_{ab}^2}{d_{ak}^2})^{1/(m-1)}}, \quad (19)$$

$$\eta_{ab} = \frac{e^{-\xi_{ab}}}{\sum_{k=1}^C e^{-\xi_{ak}}} \left(1 - \frac{1}{C} \sum_{k=1}^C \xi_{ak} \right); \quad (20)$$

– Độ từ chối:

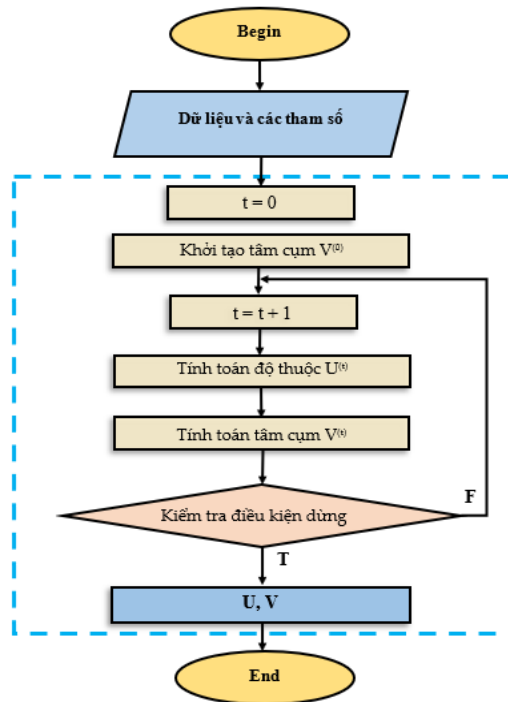
$$\xi_{ab} = 1 - (\mu_{ab} + \eta_{ab})$$

$$- \left(1 - (\mu_{ab} + \eta_{ab})^\delta \right)^{\frac{1}{\delta}}, \quad (21)$$

trong đó hệ số mũ điều khiển độ từ chối được ký hiệu là δ với $\delta \in (0, 1)$.

2.4. Sơ đồ thuật toán

Nhìn chung, các thuật toán phân cụm mờ đều có tư tưởng như sau: đầu vào xuất phát từ các tập dữ liệu và các tham số, yêu cầu đầu ra tìm tâm cụm V và ma trận các độ thuộc U . Sơ đồ tổng quát thuật toán như sau:



Hình 1. Sơ đồ thuật toán

Thuật toán sẽ tiến hành lặp cho đến khi thỏa mãn điều kiện dừng: số lần lặp lớn hơn một ngưỡng nhất định hoặc khoảng cách giữa tâm cụm của hai lần lặp liên tiếp nhỏ hơn một ngưỡng cho trước.

3. ỨNG DỤNG VÀO PHÁT HIỆN SẠT LỎ TỪ ẢNH VỆ TINH

3.1. Điều kiện thực nghiệm

Các phương pháp FCM, SSFCM và FC-PFS được cài đặt trên máy tính Lenovo 16Gb Ram, Chip Intel core i5. Dữ liệu ảnh thực nghiệm được lấy tại địa chỉ: <http://gpcv.whu.edu.cn/data/Bijiepages.html>

Bài báo này sử dụng các tiêu chí sau để so sánh hiệu năng của các thuật toán: (i) Độ chính xác phân cụm (Gan và cs., 2019) giá trị cao hơn đem lại kết quả tốt hơn, có đơn vị tính là phần trăm (%); (ii) Thời gian chạy thuật toán: thời gian càng nhỏ thì càng tốt, đơn vị tính là giây (s).

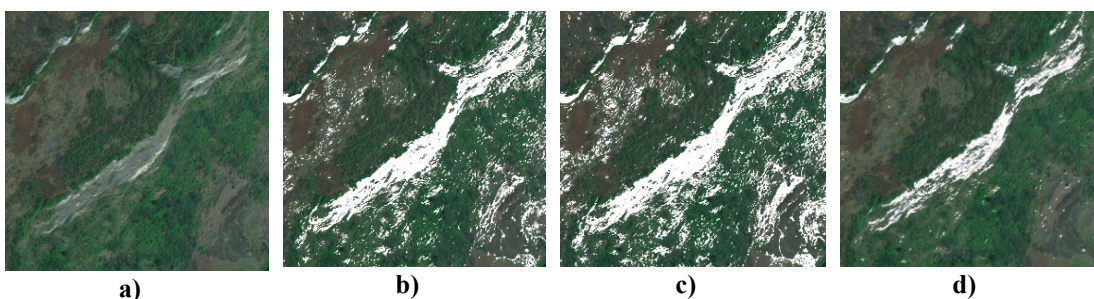
3.2. Kết quả thực nghiệm

Thực nghiệm trên 20 ảnh trong bộ dữ liệu ảnh vệ tinh sạt lở, chương trình của chúng tôi chạy 20 lần cho một ảnh và khởi tạo nhãn cho 20% điểm ảnh ngẫu nhiên (sạt lở/không sạt lở), 80% điểm ảnh còn lại chưa được gán nhãn. Trong 20% điểm ảnh được gán nhãn, chúng tôi chạy thực nghiệm với số lượng nhãn sai lần lượt là 0%, 10%, 20%, 30%. Cuối cùng, chúng tôi tổng hợp kết quả chạy chương trình trên bộ dữ liệu ảnh vệ tinh đối các phương pháp FCM, SSFCM, FC-PFS như Bảng 1 dưới đây:

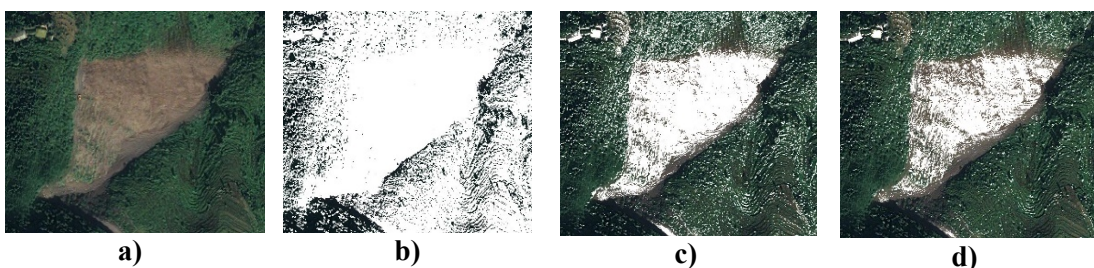
Phương pháp	FCM	SSFCM	FC-PFS
Độ chính xác phân cụm (%)	56.45	60.12	80.25
Thời gian chạy thuật toán (s)	107.30	120.58	219.43

Theo Bảng 1, xét về tiêu chí **Độ chính xác phân cụm** thì phương pháp FC-PFS có độ chính xác phân cụm tốt nhất, tiếp theo đến phương pháp SSFCM và cuối cùng phương pháp FCM có độ chính xác phân cụm thấp nhất. Tỷ lệ độ chính xác phân cụm của FC-PFS đối với FCM là 1.42 và đối với SSFCM là 1.33. Xét về tiêu chí **Thời gian chạy thuật toán** thì phương pháp FCM có thời gian chạy thuật toán tốt nhất, sau đó đến phương pháp SSFCM và cuối cùng là phương pháp FC-PFS. Tỷ lệ thời gian chạy thuật toán của FCM đối với SSFCM là 0.89 và đối với FC-PFS là 0.49.

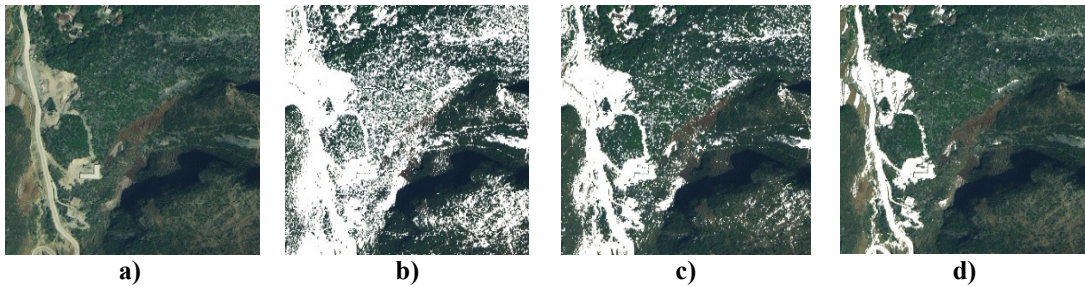
Kết quả thu được sau khi sử dụng các kỹ thuật FCM, SSFCM và FC-PFS cho 03 ảnh gốc (ảnh 1, ảnh 2 và ảnh 3) trong tổng số 20 ảnh thực nghiệm, được trình bày bởi các Hình 2, Hình 3 và Hình 4 dưới đây:



Hình 2. Kết quả phân cụm ảnh 1: a) Ảnh gốc; b) FCM; c) SSFCM; d) FC-PFS



Hình 3. Kết quả phân cụm ảnh 2: a) Ảnh gốc; b) FCM; c) SSFCM; d) FC-PFS



Hình 4. Kết quả phân cụm ảnh 3: a) Ảnh gốc; b) FCM; c) SSFCM; d) FC-PFS

Trong các ảnh gốc, các khu vực đất đá sạt lở là những khu vực có màu nâu đặc trưng của đất đá sạt lở. Trong các ảnh kết quả, khu vực đất đá sạt lở được xác định là những khu vực có màu trắng sáng. Trong ảnh vệ tinh sạt lở có thể chứa các đối tượng: đất đá sạt lở (tương ứng độ thuộc – positive), đất đá bình thường (tương ứng không thuộc – negative) và cây xanh (tương ứng độ do dự – neutral), cùng với một số nhiễu khác (tương ứng độ từ chối – refusal). Một đối tượng trên ảnh có thể thuộc về một trong bốn loại trên với độ thuộc khác nhau.

Nhìn chung, có thể thấy khu vực đất đá bình thường và khu vực sạt lở qua ảnh vệ tinh có những điểm tương đồng nhất định, đặc biệt là về màu sắc nên rất khó nhận ra (đều có màu sắc trắng sáng). Có thể thấy rõ điều này trong hình minh họa: nhiều khu vực sạt lở được phát hiện nhầm, đặc biệt là các khu vực được xác định bằng phương pháp FCM. Tuy nhiên, các khu vực sạt lở thực sự đã được phát hiện, từ đó kịp thời đánh giá các giải pháp phòng tránh, khắc phục hậu quả của thiên tai. Như thể hiện trong các hình ảnh kết quả, có thể nhận thấy phương pháp FC-PFS đem lại kết quả phát hiện sạt lở chính xác hơn các phương pháp FCM và SSFCM.

Phương pháp phân cụm mờ viễn cảnh FC-PFS đem lại kết quả phân cụm tốt hơn so với các phương pháp FCM và phương pháp SSFCM, đặc biệt trong trường hợp dữ liệu chứa nhiễu như ảnh vệ tinh thường có lẫn mây mù, sương, v.v.. Trong đó mỗi phần tử dữ liệu có thể được gán tương ứng với một trong bốn thuộc tính của tập mờ viễn cảnh: độ thuộc, độ không thuộc, độ trung lập và độ từ chối của một phần tử đối với các cụm. Khác với tập mờ truyền thống, việc tập mờ viễn cảnh có thêm các thuộc tính này có thể mang lại hiệu suất tốt hơn cho quá trình phân cụm.

4. KẾT LUẬN

Bài báo này giới thiệu một số phương pháp phân cụm mờ và ứng dụng cho bài toán phát hiện sạt lở từ ảnh vệ tinh. Với kết quả thu được khi thử nghiệm bằng các hình ảnh dữ liệu cho thấy, các phương pháp được giới thiệu có thể áp dụng để giải quyết bài toán phát hiện các khu vực sạt lở trong ảnh vệ tinh. Điển hình là phương pháp phân cụm dựa trên tập mờ viễn cảnh FC-PFS có kết quả độ chính xác phân cụm cao, đem lại hình ảnh phát hiện sạt lở tương đối rõ ràng. Điều này rất hữu ích cho các cá nhân và tổ chức và cá nhân làm việc trong lĩnh vực phòng chống thiên tai, rút ngắn thời gian tìm kiếm, phát hiện sạt lở. Qua đó nhanh chóng đưa ra các quyết định và giải pháp nhằm ngăn chặn những thiệt hại có thể xảy ra về người và tài sản.

TÀI LIỆU THAM KHẢO

- Bezdek, J. C., Ehrlich, R., & Full, W. (1984). FCM: The fuzzy c-means clustering algorithm. *Computers & Geosciences*, 10(2), 191–203.
- Bùi Công Cường. (2014). Picture fuzzy sets. *Journal of Computer Science and Cybernetics*, 30(4), Art. 4.
- Cao, L., Wang, C., & Li, J. (2016). Vehicle detection from highway satellite images via transfer learning. *Information Sciences*, 366, 177–187.
- Chen, B., Jiang, T., & Chen, L. (2020). WITHDRAWN: Weblog Fuzzy Clustering Algorithm based on Convolutional Neural Network. *Microprocessors and Microsystems*, 103420. <https://doi.org/10.1016/j.micpro.2020.103420>

- Curiskis, S. A., Drake, B., Osborn, T. R., & Kennedy, P. J. (2020). An evaluation of document clustering and topic modelling in two online social networks: Twitter and Reddit. *Information Processing & Management*, 57(2), 102034.
- Gan, H., Fan, Y., Luo, Z., Huang, R., & Yang, Z. (2019). Confidence-weighted safe semi-supervised clustering. *Engineering Applications of Artificial Intelligence*, 81, 107–116.
- Kaifa, Z., Yizhang, J., Kaijian, X., Leyuan, Z., Yangyang, C., Ke, X., & Pengjiang, Q. (2020). View-collaborative fuzzy soft subspace clustering for automatic medical image segmentation. *Multimedia Tools and Applications*, 79(13–14), 9523–9542.
- Kanungo, T., Mount, D. M., Netanyahu, N. S., Piatko, C. D., Silverman, R., & Wu, A. Y. (2002). An efficient k-means clustering algorithm: Analysis and implementation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7), 881–892.
- Knorr, W., Pytharoulis, I., & George P. Petropoulos, E. A. (2011). Combined use of weather forecasting and satellite remote sensing information for fire risk, fire and fire impact monitoring. *Computational Ecology and Software*, 1(2), 112–120.
- Li, Z., Li, Y., Lu, W., & Huang, J. (2020). Crowdsourcing Logistics Pricing Optimization Model Based on DBSCAN Clustering Algorithm. *IEEE Access*, 8, 92615–92626.
- Majumdar, S., & Laha, A. K. (2020). Clustering and classification of time series using topological data analysis with applications to finance. *Expert Systems with Applications*, 162, 113868.
- Narayana, G. S., & Kolli, K. (2021). Fuzzy K-means clustering with fast density peak clustering on multivariate kernel estimator with evolutionary multimodal optimization clusters on a large dataset. *Multimedia Tools and Applications*, 80(3), 4769–4787.
- Pedrycz, W., & Waletzky, J. (1997). Fuzzy clustering with partial supervision. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 27(5), 787–795.
- Rahim, R., Santoso, J., Jumini, S., Bhawika, G., Susilo, D., & Wibowo, D. (2021). Unsupervised Data Mining Technique for Clustering Library in Indonesia. *Library Philosophy and Practice (e-Journal)*, 4866.
- Ramasubbareddy, S., Srinivas, T. A. S., Govinda, K., & Manivannan, S. S. (2020). Comparative Study of Clustering Techniques in Market Segmentation. Trong H. S. Saini, R. Sayal, R. Buyya, & G. Aliseri (B.t.v), *Innovations in Computer Science and Engineering: Proceedings of 7th ICICSE* (tr 117–125). Springer.
- Shi, W., Chen, W.-N., Gu, T., Jin, H., & Zhang, J. (2021). Handling Uncertainty in Financial Decision Making: A Clustering Estimation of Distribution Algorithm With Simplified Simulation. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 5(1), 42–56.
- Shibai, Y., Chuihan, K., Yibin, W., & Weixing, W. (2018). Effective infrared ship image segmentation using fuzzy correlation and graph cut optimization. *Journal of Electronic Imaging*, 27(4), 43057.
- Thong, P. H., & Son, L. H. (2016). Picture fuzzy clustering: A new computational intelligence method. *Soft Computing*, 20(9), 3549–3562.
- Wu, Y., Guo, Y., Xiao, Y., & Lao, S. (2020). AAE-SC: A scRNA-Seq Clustering Framework Based on Adversarial Autoencoder. *IEEE Access*, 8, 178962–178975.
- Zadeh, L. A. (1996). *Fuzzy Sets, Fuzzy Logic, and Fuzzy Systems: Selected Papers by Lotfi A Zadeh* (G. J. Klir & B. Yuan, B.t.v; First Edition). WSPC.