

GIỚI THIỆU PHƯƠNG PHÁP

HOÀN THIỆN VÀ MỞ RỘNG PHẠM VI ÁP DỤNG PHƯƠNG PHÁP KHOẢNG CÁCH - TẦN SUẤT - NHẬN DẠNG TRONG PHÂN TÍCH SỐ LIỆU ĐỊA VẬT LÝ HÀNG KHÔNG

NGUYỄN XUÂN BÌNH¹, NGUYỄN ĐỨC VINH², VÕ THANH QUỲNH²

E-mail: binhnx@igp-vast.vn

¹Viện Vật lý Địa cầu - Viện Khoa học và Công nghệ Việt Nam

²Trường Đại học Khoa học Tự nhiên - Đại học Quốc gia Hà Nội

Ngày nhận bài: 10 - 9 - 2012

1. Mở đầu

Phương pháp Khoảng cách - Tần suất - Nhận dạng cùng với phương pháp Tần suất - Nhận dạng là hai phương pháp phân tích nhận dạng mới, đã được áp dụng bước đầu có hiệu quả trong xử lý - phân tích tài liệu phổ gamma hàng không, phục vụ tìm kiếm và dự báo triển vọng khoáng sản. Tuy nhiên, cũng như phương pháp Tần suất - Nhận dạng trước đây, phương pháp Khoảng cách - Tần suất - Nhận dạng hiện nay còn tồn tại một số hạn chế, cần được nghiên cứu khắc phục. Đó là: phương pháp Khoảng cách - Tần suất - Nhận dạng hiện mới chỉ đưa ra được thuật toán phân tích đối sánh, xác định mức độ đồng dạng của đối tượng đối sánh so với đối tượng mẫu, làm cơ sở cho việc đánh giá, dự báo mức độ triển vọng của chúng, khi đối tượng đối sánh đã biết; nó chưa giải quyết được nhiệm vụ tìm kiếm, phát hiện, khoanh định ranh giới các đối tượng đồng dạng cũng như mức độ đồng dạng của chúng khi chưa biết trước các đối tượng đối sánh, một nội dung quan trọng đối với các phương pháp phân tích nhận dạng nói chung. Phương pháp này cũng mới áp dụng có hiệu quả đối với tài liệu dị thường phổ gamma hàng không mà chưa được mở rộng cho các dạng tài liệu địa vật lý khác, kể cả tài liệu từ hàng không. Bài báo này trình bày kết quả nghiên cứu, hoàn thiện nội dung của phương pháp, bằng việc mở rộng thuật toán cho cả trường hợp biết trước đối tượng đối sánh và trường hợp chưa biết trước đối tượng đối sánh, cùng với việc xây dựng phần mềm xử lý trên

máy tính tương ứng. Chương trình mới mang tên QKCM. Chương trình mới này đã được áp dụng phân tích thử nghiệm đối với tài liệu địa vật lý hàng không (bao gồm cả tài liệu từ và tài liệu phổ gamma) trên một số diện tích bay đo cho kết quả tốt. Hy vọng, với việc hoàn thiện hai phương pháp nhận dạng mới: Tần suất - Nhận dạng và Phương pháp Khoảng cách - Tần suất - Nhận dạng có thể đưa vào áp dụng thực tế, góp phần đẩy nhanh và nâng cao hơn nữa chất lượng của công tác xử lý - phân tích tài liệu địa vật lý hàng không, một nguồn tài liệu hết sức phong phú ở nước ta hiện nay, phục vụ công tác tìm kiếm và dự báo triển vọng khoáng sản.

2. Nội dung phương pháp Khoảng cách - Tần suất - Nhận dạng hiện tại

Phương pháp Khoảng cách - Tần suất - Nhận dạng được tác giả xây dựng trên cơ sở vận dụng kết hợp phương pháp phân tích tần suất theo thuật toán Griffiths - Vinni và phương pháp phân tích khoảng cách khái quát theo thuật toán Paguônôp. Nội dung phương pháp gồm 3 phần được tóm tắt như sau [1, 2, 4, 6]:

2.1. Xây dựng ma trận thông tin đối tượng mẫu

Các ma trận thông tin của đối tượng mẫu (ví dụ: quặng và không quặng) theo phương pháp phân tích khoảng cách khái quát có được trực tiếp từ số liệu địa vật lý trên các đối tượng mẫu đó (do các phần tử của ma trận của phương pháp này là các đại lượng số thực). Trong khi đó, ma trận thông tin

của đối tượng mẫu trong phương pháp phân tích tần suất lại là các phân tử logic “có” hoặc “không” hoặc các giá trị “0” hoặc “1” [4]. Để có được ma trận thông tin của đối tượng mẫu ở dạng này từ các số liệu địa vật lý, tác giả đã đưa ra cách xây dựng như sau:

Từ tập hợp số liệu của các chủng loại thông tin của đối tượng mẫu quặng trong phương pháp phân tích khoảng cách khái quát xây dựng các đường cong biến phân (đường cong mật độ phân bố). Từ các đường cong biến phân xác định khoảng giá trị đặc trưng cho từng tham số [1, 2, 4]. Sau khi có được các khoảng giá trị đặc trưng, dùng nó làm “cửa sổ quét” để tạo ra các đơn vị thông tin cho từng chủng loại thông tin, nếu nó nằm trong khoảng giá trị đặc trưng sẽ nhận giá trị là 1, nằm ngoài sẽ nhận giá trị là 0. Bằng cách này sẽ chuyển được một ma trận thông tin với các số liệu địa chất, địa vật lý bất kỳ về ma trận thông tin chuẩn theo yêu cầu của thuật toán phân tích tần suất với các phân tử là các giá trị 1 hoặc 0.

2.2. Đánh giá lựa chọn tổ hợp thông tin

Việc đánh giá lựa chọn tổ hợp thông tin để tiến hành phân tích đối sánh, xác định mức độ đồng dạng của đối tượng được thực hiện trên cơ sở vận dụng kết hợp cả hai phương pháp: phương pháp phân tích khoảng cách khái quát và phương pháp phân tích tần suất:

- Tiến hành phương pháp phân tích khoảng cách khái quát trên các ma trận thông tin của 2 đối tượng mẫu (quặng và không quặng) để xác định tập $\{\rho_i^*\}$ [6]:

Ma trận thông tin đối tượng quặng:

$$(\phi_{i,j})_q = \begin{pmatrix} \phi_{11}\phi_{12}\dots\phi_{1k} \\ \phi_{21}\phi_{22}\dots\phi_{2k} \\ \dots\dots\dots \\ \phi_{n1}\phi_{n2}\dots\phi_{nk} \end{pmatrix}_q \quad (1)$$

Đối tượng không quặng:

$$(\varphi_{i,j})_{kq} = \begin{pmatrix} \varphi_{11}\varphi_{12}\dots\varphi_{1k} \\ \varphi_{21}\varphi_{22}\dots\varphi_{2k} \\ \dots\dots\dots \\ \varphi_{m1}\varphi_{m2}\dots\varphi_{mk} \end{pmatrix}_{kq} \quad (2)$$

Theo Paguônôp lượng tin của tính chất thứ “i” được đánh giá theo bình phương khoảng cách khái

quát giữa trọng tâm các đám mây trong không gian đầu hiệu:

$$\rho_i^2 = \frac{(\bar{\varphi}_{iq} - \bar{\varphi}_{ikq})^2}{\sigma_i^2} \quad (3)$$

Trong đó:

$$\sigma_i^2 = \frac{\sigma_{iq}^2}{n} + \frac{\sigma_{ikq}^2}{m} \quad (4)$$

Sắp xếp $\{\rho_i^2\}$ theo thứ tự giảm dần và gọi nó là $\{\rho_i^{*2}\}$.

- Tiến hành phương pháp phân tích tần suất trên ma trận thông tin của đối tượng mẫu quặng để xác định tập $\{I_i^*\}$ [4]:

Từ ma trận thông tin của đối tượng mẫu:

$$\langle \varphi_{i,j} \rangle_{di} = \begin{pmatrix} \varphi_{11}\varphi_{12}\dots\varphi_{1k} \\ \varphi_{21}\varphi_{22}\dots\varphi_{2k} \\ \dots\dots\dots \\ \varphi_{n1}\varphi_{n2}\dots\varphi_{nk} \end{pmatrix} \quad (5)$$

Theo Griffiths-Vinin, tỷ trọng thông tin tương đối của tính chất thứ “i” được xác định theo công thức:

$$I_i = \frac{1}{n} \sqrt{\frac{1}{k} \sum_{j=1}^k \left(\sum_{h=1}^n \varphi_{hi} \varphi_{hj} \right)^2} \quad (6)$$

Sắp xếp tập $\{I_i^*\}$ theo thứ tự giảm dần và gọi tập mới này là $\{I_i\}$

Từ kết quả phân tích của cả 2 phương pháp và có được các tập $\{I_i\}$, $\{I_i^*\}$, $\{\rho_i^2\}$, $\{\rho_i^{*2}\}$ sẽ xác định được tập hợp các tính chất (các chủng loại thông tin) có độ tin cậy cao theo yêu cầu nghiên cứu thông qua tỷ trọng thông tin tương đối của các tính chất: Q_h [6].

$$Q_h = \frac{\sqrt{\sum_{i=1}^h j_i^2}}{\sqrt{\sum_{i=1}^k j_i^2}} \times 100 \% \quad (7)$$

Trong đó $\{J_i\}$ là tập $\{I_i\}$ được sắp xếp theo thứ giảm dần của $\{\rho_i^2\}$.

2.3. Phân tích đối sánh xác định các đối tượng đồng dạng

Các đối tượng cần đối sánh với đối tượng mẫu để xem nó có đồng dạng với đối tượng mẫu hay không được thực hiện theo cách như sau [6]:

- Tiến hành đánh giá lượng tin I_i cho tất cả các tính chất của đối tượng đối sánh bằng phương pháp phân tích tần suất và xác định được tập $\{I_i\}$. Ở đây ma trận thông tin của đối tượng đối sánh được xây dựng thông qua chính các khoảng giá trị đặc trưng (các “cửa sổ quét”) của đối tượng mẫu như đã trình bày ở mục 1.

- Tính tỷ trọng thông tin tương đối của h tính chất đầu theo tập $\{J_i\}$ đã được lựa chọn ở mục 2 theo công thức (7) cho đối tượng đối sánh, ký hiệu là Q_h^* và ta gọi nó là chỉ số đồng dạng.

Đối tượng đối sánh được xem là đồng dạng với đối tượng mẫu khi $Q_h^* \geq Q_h$.

3. Hoàn thiện và mở rộng phạm vi áp dụng phương pháp Khoảng cách - Tần suất - Nhận dạng

3.1. Nghiên cứu hoàn thiện nội dung phương pháp

Từ thành công trong việc cải tiến, hoàn thiện phương pháp Tần suất - Nhận dạng, chúng tôi tiếp tục nghiên cứu hoàn thiện và mở rộng phạm vi áp dụng đối với phương pháp Khoảng cách - Tần suất - Nhận dạng cho cả hai trường hợp: trường hợp biết trước đối tượng đối sánh và trường hợp chưa biết trước các đối tượng đối sánh; cụ thể như sau:

Bước 1: Nội suy các số liệu khảo sát địa vật lý thực tế lên mạng lưới đều (ô vuông hoặc chữ nhật) bằng các thuật toán nội suy hiện có.

Nội dung này được thực hiện bằng các thuật toán nội suy trong Bộ chương trình phân tích phổ - thống kê COSCAD.

Bước 2: Thực hiện các nội dung như:

- Xây dựng các ma trận thông tin của các đối tượng mẫu cho cả hai thuật toán: thuật toán phân tích tần suất và thuật toán phân tích khoảng cách khái quát như đã trình bày ở trên.

- Đánh giá, lựa chọn tổ hợp thông tin trên các đối tượng mẫu

Việc đánh giá lựa chọn tổ hợp thông tin được tiến hành trên cơ sở vận dụng kết hợp phương pháp phân tích tần suất theo thuật toán Griffiths-Vinni và phương pháp phân tích khoảng cách khái quát theo thuật toán Paguônôp:

+ Tiến hành phương pháp phân tích khoảng cách khái quát trên ma trận thông tin của hai đối tượng mẫu (quặng và không quặng) xác định

tập $\{\rho_i^{*2}\}$.

+ Tiến hành phương pháp phân tích tần suất trên ma trận thông tin của đối tượng mẫu quặng.

Đến đây ta không sắp xếp $\{I_i\}$ theo thứ tự giảm dần của chính nó để được tập $\{I_i^*\}$ mà sắp xếp nó

theo thứ tự của tập $\{\rho_i^{*2}\}$ và gọi tập mới này là $\{J_i\}$.

Tính tỷ trọng thông tin tương đối của h tính chất đầu theo tập $\{J_i\}$, ta gọi nó là Q_h . Khi đó Q_h được tính theo công thức (7).

Tổ hợp h tính chất đầu được lựa chọn thông qua các giá trị Q_h như sau:

Xây dựng đường cong Q_h theo h (trục y là Q_h , trục x là h); Giá trị h được xác định sao cho tại đó trị tuyệt đối đạo hàm bậc 2 của Q_h theo h có giá trị lớn nhất ($|\frac{\partial^2 Q_h}{\partial h^2}|_{\max}$) nghĩa là tại đó có sự phân

chia (phân tách) rõ nhất giữa tập các thông tin độ tin cậy cao và tập các thông tin độ tin cậy thấp. Trên đường cong Q_h hoành độ h được xác định tại điểm có độ cong lớn nhất.

Bước 3: - Trường hợp biết trước đối tượng đối sánh.

Nội dung này đã được giải quyết ở phương pháp Khoảng cách - Tần suất - Nhận dạng hiện có thông qua việc tính toán và so sánh giữa các giá trị Q_h^* của đối tượng đối sánh và Q_h của đối tượng mẫu [6].

- Trường hợp chưa biết trước đối tượng đối sánh.

Phân tích, tính toán chỉ số đồng dạng Q_h^* cho tất cả các điểm trên các nút của mạng lưới đều đã được nội suy, trên toàn bộ diện tích nghiên cứu. Nội dung này được thực hiện như sau:

- Dùng các “cửa sổ quét” để xác định ranh giới diện tích của các đối tượng đối sánh. Các “cửa sổ quét” có thể là các hình tròn, hình vuông, hình chữ nhật, hình elipxoid với các kích thước và góc quay tùy chọn. Các diện tích nằm trong cửa sổ quét được xem là các đối tượng đối sánh, cần tiến hành xử lý, phân tích đối sánh, xác định mức độ đồng dạng của

chúng so với đối tượng mẫu thông qua chỉ số đồng dạng Q^*_h . Nội dung này được thực hiện giống như như trường hợp các đối tượng đối sánh đã biết của phương pháp Tần suất - Nhận dạng [5]. Giá trị Q^*_h của đối tượng đối sánh vừa tính được sẽ được gán cho điểm trung tâm của cửa sổ quét. Điểm trung tâm này sẽ trùng với điểm nút của mạng lưới đều đã được nội suy.

- Dịch chuyển cửa sổ quét khắp diện tích của vùng nghiên cứu, với bước dịch đều theo mạng lưới đã được nội suy. Kết quả ta được một File số liệu các chỉ số đồng dạng $Q^*_h(x,y)$ theo tọa độ trùng với tọa độ của mạng lưới đã được nội suy trên khắp diện tích vùng nghiên cứu.

- Khoanh định và đánh giá mức độ đồng dạng của các diện tích đồng dạng với đối tượng mẫu. Từ File số liệu này, với các mức giá trị ngưỡng cho trước ta sẽ có xác định được sự phân bố của các đối tượng đồng dạng cũng như mức độ đồng dạng của chúng so với đối tượng mẫu trên toàn diện tích nghiên cứu bằng việc xây dựng bản đồ đồng mức giá trị của $Q^*_h(x,y)$.

Trên quan điểm của một phương pháp phân tích nhận dạng, đến đây bài toán nhận dạng đặt ra đã được giải quyết một cách triệt để như cách làm thông thường của các phương pháp phân tích nhận dạng trong xử lý số liệu địa vật lý.

3.2. Xây dựng chương trình và phân tích thử nghiệm

Theo nội dung phương pháp đã được bổ sung hoàn thiện như trình bày ở trên, chúng tôi đã xây dựng chương trình phân tích mới trên máy tính. Chương trình được viết bằng ngôn ngữ FORTRAN, trên cơ sở bổ sung, phát triển từ chương trình QKC [3] hiện có cho cả hai trường hợp: trường hợp biết trước các đối tượng đối sánh và trường hợp chưa biết trước các đối tượng đối sánh. Chương trình có tên gọi mới: chương trình QKCM. Chúng tôi đã tiến hành phân tích thử nghiệm bằng chương trình phân tích mới trên tài liệu thực tế ở diện tích phần phía đông vùng bay Tuy Hòa [3] cả tài liệu từ và phổ gamma hàng không và thu được kết quả tốt, phù hợp với các kết quả phân tích trước đó bằng tổ hợp của nhiều phương pháp nhận dạng khác nhau (bảng 1).

Bảng 1. Kết quả phân tích theo phương pháp Khoảng cách - Tần suất - Nhận dạng (QKCM) và so sánh với kết quả kiểm tra mặt đất (Đối tượng mẫu quảng: cụm 68, đối tượng mẫu không quảng: cụm 88)

STT	Các cụm đồng dạng	Chỉ số đồng dạng	Đã kiểm tra mặt đất	Kết quả đánh giá
1	19	80,24 %		
2	20	82,15 %		
3	24	83,20%		
4	39	87,6 %		
5	60	77,25 %	*	T.V loại 2
6	87	80,86%	*	T.V loại 1
7	89	82,0 %	*	T.V loại 1
8	94	83,36%	*	T.V loại 1
9	99	80,18%		

Các đối tượng mẫu được lựa chọn ở đây là các cụm dị thường đã được kiểm tra đánh giá mặt đất và được xác định là rất có triển vọng khoáng sản [3, 5].

4. Kết luận

Tương tự như phương pháp Tần suất - Nhận dạng bằng việc đề xuất và xây dựng bổ sung một thuật toán phân tích đối sánh, xác định chỉ số đồng dạng mới đối với trường hợp khi chưa biết trước các đối tượng đối sánh cho phương pháp Khoảng cách - Tần suất - Nhận dạng, chúng ta đã có thêm một phương pháp phân tích nhận dạng mới hoàn chỉnh hơn, có khả năng xử lý với các nguồn tài liệu đa dạng hơn, giải quyết một cách triệt để hơn yêu cầu của bài toán nhận dạng đặt ra trong địa vật lý.

Cùng với phương pháp Tần suất - Nhận dạng, phương pháp Khoảng cách - Tần suất - Nhận dạng mới tiếp tục bổ sung và góp phần làm phong phú thêm hệ các phương pháp phân tích nhận dạng trong công tác xử lý, phân tích tài liệu địa vật lý, đặc biệt là tài liệu địa vật lý hàng không.

Kết quả phân tích thử nghiệm trên tài liệu thực tế từ và phổ gamma hàng không vùng Tuy Hòa đã nói lên độ tin cậy của các kết quả phân tích cũng như khả năng ứng dụng thực tế của phương pháp.

Phương pháp Khoảng cách - Tần suất - Nhận dạng mới hoàn toàn có thể đưa vào áp dụng trong công tác xử lý, phân tích tài liệu địa vật lý hàng không phục vụ tìm kiếm và dự báo triển vọng khoáng sản, một nhiệm vụ hết sức quan trọng đang được đặt ra đối với các nhà địa vật lý hàng không ở nước ta hiện nay.

Bài báo này được hoàn thành với sự hỗ trợ kinh phí của Đề tài NCKH cấp ĐHQG Hà Nội, mã số QG.10-14.

TÀI LIỆU DẪN

[1] *Davis J. C.*, 1995: *Statistic and data analysis in geology*. New York, 666pp.

[2] *Green A.A.*, 1987: Leveling airborne gamma-radiation data using between-channel correlation information, *Geophysics*, 52/11.

[3] *Võ Thanh Quỳnh* (chủ biên), 1996: Báo cáo kết quả bay đo từ phổ gamma tỷ lệ 1:25.000 vùng Tuy Hòa. Lưu trữ Địa chất. Hà Nội.

[4] *Võ Thanh Quỳnh*, 2007: Một cách tiếp cận mới giải quyết bài toán nhận dạng trong xử lý,

phân tích tài liệu địa vật lý. *TC Địa chất*, A/302: 76-80.

[5] *Võ Thanh Quỳnh*, 2008: Phương pháp đánh và phân loại cụm dị thường trong xử lý - phân tích tài liệu phổ gamma hàng không. *TC Địa chất*, A/304: 70-75. Hà Nội.

[6] *Võ Thanh Quỳnh*, 2008: Xây dựng một phương pháp nhận dạng mới trong xử lý tài liệu địa vật lý trên cơ sở vận dụng kết hợp các phương pháp phân tích khoảng cách khái quát và phân tích tần suất. *TC Địa chất*, A/305: 61-66.

[7] *Võ Thanh Quỳnh, Nguyễn Xuân Bình, Nguyễn Đức Vinh*, 2011: Nghiên cứu hoàn thiện phương pháp Tần suất - Nhận dạng trong xử lý và phân tích số liệu địa vật lý. *TC Địa chất*, A/326: 50-56.

SUMMARY

Improve and extend applied field of the method of distance - frequency Rate - identification in analyzing and processing the aero - geophysical data

The methods of identification play an important role in analyzing and processing the geophysical data, especially in processing the aero-geophysical data for exploration and prediction of perspective of mineral resources. At the present in geophysics, there are many modern methods of identification that were automated by a lot of powerful professional software and were applied effectively in analyzing aero-geophysical data in Viet Nam and in the world. One of the new methods of identification that were applied effectively in Vietnam is the method of distance-frequency rate-identification which was built by Vo Thanh Quynh. When the method was used for processing the aero-geophysical data, there are some limits needing to be studied and solved. Following this direction we have studied, improved, widened applied field and built new software for method of distance-frequency rate-identification. These new researching results have been applied effectively in analyzing and processing the aero-geophysical data to make clear perspective of mineral resources in the east Tuy Hoa zone.