

MỘT VÀI NHẬN XÉT VỀ DÒNG DI CƯ CỦA CÁC DÂN TỘC VIỆT NAM DỰA TRÊN CÁC PHƯƠNG PHÁP TOÁN HỌC

CHU ĐỨC

I. MỞ ĐẦU

Trong công cuộc kiến thiết đất nước hiện nay có sự đóng góp rất to lớn của đại gia đình 54 dân tộc Việt Nam. Theo các nhà nhân chủng học xưa nay, các tộc người ở Việt Nam là kết quả của sự giao lưu của hai dòng người từ phương Nam lên và phương Bắc xuống. Vì vậy mỗi tộc người đều ít nhiều biến đổi. Song, lại có khía cạnh khó thay đổi, đó là các đặc điểm nhân chủng. Tất nhiên việc điều tra các đặc điểm nhân chủng phải hết sức chính xác, khách quan. Nhưng dù sao với những số liệu đã có để từ đó có những nhận xét bước đầu, đó là kho tàng vô giá. Nhờ đó mới có những bước đề xuất điều tra tiếp theo và đó cũng là nội dung của lý thuyết thiết kế thí nghiệm tối ưu của Will - Boxon.

Bằng số liệu điều tra trên 20 năm của nhà nhân chủng học Nguyễn Đình Khoa trên 23 dân tộc và với 25 đặc điểm nhân chủng, chúng tôi đã đem xử lý bằng mô hình toán học - đó là lý thuyết phân loại nhiều chiều, và lý thuyết phân tích biến phân, với máy vi tính, chúng tôi đã có một vài nhận xét bước đầu để chúng ta tham khảo tiếp sau đây.

II. PHƯƠNG PHÁP

Xét một tập hợp gồm n hệ sinh thái, mỗi hệ ta lấy m đặc điểm để quan sát. Vấn đề là xét xem trong n hệ sinh thái đó có quan hệ họ hàng như thế nào với nhau.

Bước 1: Là từ dãy số liệu điều tra $X_{\alpha i l}$

Trong đó:

α là chỉ số đặc điểm, $\alpha = 1, 2, \dots, m$

i là chỉ số hệ sinh thái, $i = 1, 2, \dots, n$

l là chỉ số tính số lặp lại, $l = 1, 2, \dots, n_i$

n_i là số lặp lại của hệ thứ i

Ta loại trừ những đặc điểm thừa (tức là đã có lượng thông tin trong các đặc điểm khác). Sau đó loại trừ những hệ không cần thiết (vì đã có thông tin trong hệ khác). Cơ sở toán học để làm điều đó dùng phương pháp tiêu chuẩn phương sai của Orloci (1977). (dispersion criterion) (xem [2])

Bước hai:

Với số hệ còn lại, và số đặc điểm còn lại, ta dùng lý thuyết phân loại (Cluster theory) để phân ra các nhóm quan hệ họ hàng với nhau. Hiện nay trên thế giới có khoảng 12 độ đo để phân biệt quan hệ như vậy. Có thể ta sẽ lấy kết quả phù hợp với

nhieu độ đo nhất. Theo chúng tôi mật độ đo dễ dàng nhất là độ đo cực tiểu biến phân sau đây.

Định nghĩa 1: Gọi Q_{ni} là độ tản mạn của hệ thứ i là đại lượng được xác định sau đây:

$$Q_{ni} = \frac{1}{n_i} \sum_{l \neq l'} \sum_{\alpha=1}^m (X_{\alpha il} - X_{\alpha il'})^2$$

$$= \frac{1}{n_i} \sum_{l \neq l'} (d_{ll'})^2$$

Trong đó:

$$(d_{ll'})^2 = \sum_{\alpha=1}^m (X_{\alpha il} - X_{\alpha il'})^2$$

Gọi khoảng cách ơclit giữa 2 thí nghiệm l và l' .

Định nghĩa 2:

Giả sử có 2 hệ i và j có độ tản mạn nhóm là Q_{ni} và Q_{nj} . Khi đó ta đặt

$$Q^*_{ij} = Q_{ni+nj} - Q_{ni} - Q_{nj}$$

Trong đó:

$$Q_{ni+nj} = \frac{1}{n_i + n_j} \sum_{l \neq l'} \sum_{\alpha=1}^m (X_{\alpha il} - X_{\alpha il'})^2$$

$$\left(\sum_{\alpha=1}^m (X_{\alpha jl} - X_{\alpha jl'}) \right)^2 + \sum_{\alpha=1}^m (X_{\alpha il} - X_{\alpha il'})^2$$

với $1 \leq l, l' \leq n_i + n_j$

Nếu $Q^*_{ij} < Q^*_{jk} \forall k \neq i$

Khi đó ta nói 2 hệ i và j liên hợp với nhau hay là lập thành một nhóm mới.

Dựa vào 2 định nghĩa đó ta suy ra thuật toán phân loại sau đây.

Trước tiên ta xem mỗi hệ là một nhóm, xét mối quan hệ giữa từng cặp hệ qua tính toán Q^*_{ij} .

Tìm Q^*_{ij} nhỏ nhất và ghép 2 hệ đó với nhau thành nhóm. Như vậy tập hợp còn lại

$n-1$ nhóm. Ta lại tính Q^*_{ij} trong từng cặp hệ trong đó. Xét Q^*_{ij} bé nhất cho vào một nhóm mới. Tập hợp còn lại $n-2$ nhóm. Ta lại tính Q^*_{ij} cho từng cặp trong tập hợp mới. Và cứ thế tiếp tục cho đến khi chỉ còn lại một nhóm thì dừng lại.

Và kết quả ta được một lược đồ phân nhóm họ hàng.

Bước 3: Từ lược đồ phân loại đó, ta rút ra những kết luận về các hệ sinh thái đang xét.

III. ÁP DỤNG

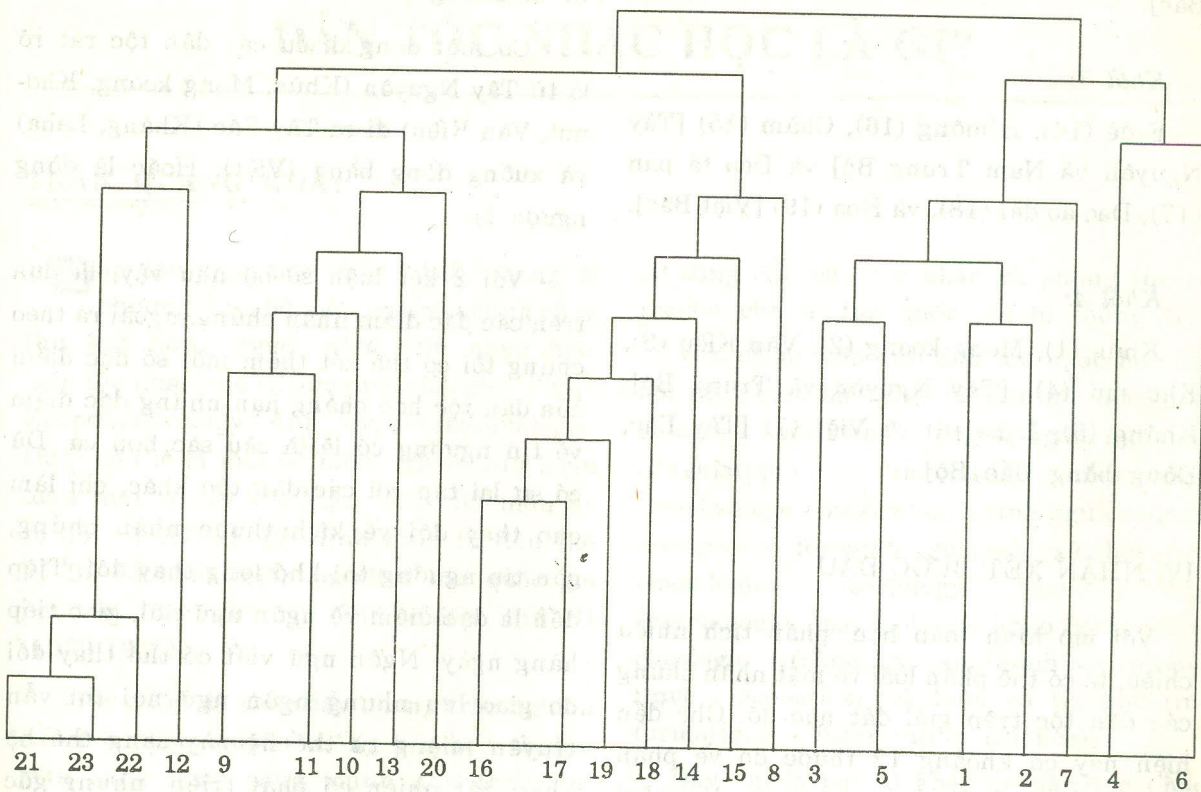
Điều tra 23 dân tộc trên đất nước Việt Nam, trên các vùng Nam Bộ, Tây Nguyên và Trung Bộ, Tây Bắc và Việt Bắc, và vùng đồng bằng sông Hồng, với 25 đặc điểm nhân chủng. Ta được bảng số liệu sau: (từ 1975 đến 1990)

Dân tộc	1	2	...	23
Đặc điểm				
1				
2				
3				
...				
25				

$X_{\alpha i} = 1, 2, \dots, 25$ (đặc điểm)
 $i = 1, 2, \dots, 23$ (dân tộc)

Các đặc điểm điều tra là:

Chiều cao (1), Bề dọc đầu (2), Ngang đầu (3), Rộng nhỏ nhất trán (4), Rộng mặt giữa (5), Rộng hàm dưới (6), Dài mặt biểu kiến (7), Dài mặt hình thái (8), Dài mũi (Ngăn mày) (9), Dài mũi (se) (10), Rộng mũi (11), Rộng miệng (12), Cao môi trên (13), Dày hai môi (14), Chỉ số đầu (15), Hàm dưới trán (16), Hàm dưới rộng mặt (17), Mặt biểu kiến (18), Mặt hình thái (se) (19), Mũi (Ngăn mày) (20), Mũi (se) (21), Màu da (22), Hình tóc (23), Độ chênh khe mắt (24), Nếp mi góc (25).



Lược đồ phân loại 23 dân tộc với 25 đặc điểm nhân chủng.

Các dân tộc được điều tra là:

Khùa (1), Mong koong (2), Vân Kiều (3), Khơ mú (4), Kháng (5), La ha (6), Việt TT (7), Tày (8), Thái (9), Xơ đăng (10), Mơ nông (11), Khơme (12), Giarai (13), Ê-Đê (14), Chăm (15), H'mông (16), Dao tả pan (17), Dao áo dài (18), Hoa (19), Mường (20), Mày (21), Nùng (22), Ba-na (23).

Trong báo cáo này chúng tôi dùng toàn bộ số liệu để xử lý. Nghĩa là chuyển ngay vào bước 2 của phương pháp.

Bước 1 và sự phân tích tiếp theo sẽ trình bày trong bài thứ 2.

Ở đây với số liệu nguyên vẹn đó của GS Nguyễn Đình Khoa, chúng ta đưa vào máy tính, và được kết quả phân loại sau đây (Xem

lược đồ phân loại).

Bước 3 (Kết luận)

Với cấp cao nhất, ta thấy mối quan hệ sau đây:

Mày (21) và Ba-na (23)

Vân Kiều (3) và Kháng (5)

H'mông (16) và Dao tả pan (17)

Nếu mở rộng hơn thì có thể phân ra 4 khối sau đây:

Khối 1:

Ba-na (23) [Tây Nguyên], Khơ-me (12) [Nam Bộ], và Mày (21) Nùng (22) [Việt Bắc].

Khối 2:

Xơ-đăng (10), Mơ-nông (11), Gia-rai (13)

[Tây Nguyên] và Thái (9) Mường (20) [Tây Bắc].

Khối 3:

Ê-đê (14), H'mông (16), Chàm (15) [Tây Nguyên và Nam Trung Bộ] và Dao tả pan (17), Dao áo dài (18), và Hoa (19) [Việt Bắc].

Khối 4:

Khùa (1), Mong koong (2), Vân Kiều (3), Khơ mú (4), [Tây Nguyên và Trung Bộ], Kháng (5), Laha (6) và Việt (7) [Tây Bắc, Đồng bằng Bắc Bộ].

IV. NHẬN XÉT BƯỚC ĐẦU

Với mô hình toán học, phân tích nhiều chiều, ta có thể phân loại về mặt nhân chủng các dân tộc trên giải đất nào đó. Cho đến hiện nay có khoảng 12 thước đo về phân loại họ hàng. Trên đây, chúng tôi dùng hai loại thước đo đó là "cực tiểu biến phân" và "nối trung bình". Có thể những kết quả này chưa thỏa mãn hoàn toàn các nhân chủng và dân tộc học. Điều đó có thể khắc phục được bằng cách giảm số đặc điểm xuống để xét, loại trừ những đặc điểm không cần thiết, hoặc là dùng một loại thước đo (metric) khác. Tuy nhiên bước đầu cũng gợi ý cho ta những kết luận về dân tộc học, để từ đó có kế hoạch điều tra thêm, hoặc có những chính sách phù hợp trước mắt. Với góc độ một nhà toán học, chúng tôi có mấy nhận xét bước đầu sau đây:

- Các dân tộc Tây Nguyên đều có liên quan với các dân tộc khắp miền của đất nước từ Nam Bộ, Tây Bắc và Việt Bắc.

Điều đó nói lên sự giao lưu, di cư của các dòng dân tộc Việt Nam: Hoặc là từ phía Bắc xuống, hoặc từ Tây Nguyên đi ra, hoặc

từ Ấn Độ, Miến Điện vào Tây Bắc, Việt Bắc và đi xuống phía Nam.

- Có một dòng di cư các dân tộc rất rõ là từ Tây Nguyên (Khùa, Mong koong, Khơ mú, Vân Kiều) đi ra Tây Bắc (Kháng, Laha) và xuống đồng bằng (Việt). Hoặc là dòng ngược lại.

- Với 2 kết luận sơ bộ như vậy, là dựa trên các đặc điểm nhân chủng, ngoài ra theo chúng tôi có thể xét thêm một số đặc điểm của dân tộc học chẳng hạn những đặc điểm về tín ngưỡng có lẽ là sâu sắc hơn cả. Dù có sự lai tạp với các dân tộc khác, chỉ làm cho thay đổi về kích thước nhân chủng, còn tín ngưỡng thì khó lòng thay đổi. Tiếp đến là đặc điểm về ngôn ngữ nói, giao tiếp hàng ngày. Ngôn ngữ viết có thể thay đổi do giao lưu nhưng ngôn ngữ nói thì vẫn truyền miệng từ thế hệ này sang thế hệ khác, tất nhiên có phát triển, nhưng gốc vẫn tìm thấy được.

Vì vậy chúng tôi đề nghị từ những kết luận bước đầu về mặt nhân chủng ta cần khảo sát trên các điểm dân tộc học để có kết luận xác đáng hơn và từ đó có thể có những chính sách dân tộc hợp lý.

TÀI LIỆU THAM KHẢO

[1] Chu Đức

Eine Mathematische Methode zur Beurteilung in Okosystemen und ihre Anwendung. (Arch. of An. Breed. Dummerstorf 12, 1992)

[2] O'rloci L.

Ranking characters by a Dispersion criterion (Nature Vol.244, No 5415)

[3] Roland Pressat

L'analyse Démographique (Paris 1973)