

SO SÁNH MỘT SỐ PHƯƠNG PHÁP THỐNG KÊ DỰ BÁO MƯA THÁNG Ở NGHĨA BÌNH

NGUYỄN NĂNG NHƯỢNG

Đài KTTV Nghĩa Bình

Ở Nghĩa Bình đã sử dụng phương pháp phân lớp có tham số phân bố chuẩn để dự báo 2 lớp lượng mưa các tháng mùa mưa có kết quả [6, 9].

Do đạt được kết quả đó, Ủy ban Khoa học và Kỹ thuật tỉnh Nghĩa Bình đã chỉ viện kinh phí cho Đài KTTV Nghĩa Bình tiếp tục xây dựng phương pháp dự báo 2 lớp lượng mưa các tháng mùa ít mưa ở Nghĩa Bình.

Chúng tôi đã chọn tháng I để thăm dò, so sánh các phương pháp.

I – CÁC PHƯƠNG PHÁP

1. Phương pháp phân lớp có tham số phân bố chuẩn và hiệu chỉnh hệ số tự do k

Phương pháp này chúng tôi đã trình bày khá đầy đủ trong các tài liệu [4–9], ở đây chỉ nêu các công thức rút gọn khi tính toán [9].

Giá trị ngưỡng để tách ra 2 lớp A và $\bar{A}$ xung khắc nhau được chọn là số trung vị của phân phối lượng mưa tháng dự báo. Do đó ta có các xác suất tiên nghiệm $p(A) = p(\bar{A})$; đồng thời để cho đơn giản, cho rằng số đo của sai lầm loại 1 và loại 2 là như nhau; mặt khác các giá trị của nhân tố dự báo (predictor) được chuẩn hóa [1], từ đó các vectơ kỳ vọng của các nhân tố dự báo có đặc điểm là :

$$M(A) = -M(\bar{A})$$

Và do đó các hệ số bậc nhất của hàm phân lớp được rút gọn như sau :

$$(b_1, b_2, \dots, b_n)' = M'(A)[V^{-1}(A) + V^{-1}(\bar{A})]$$

trong đó :

$V^{-1}(A)$ và $V^{-1}(\bar{A})$ – Các ma trận đảo của các ma trận hiệp biến $V(A)$ và $V(\bar{A})$, còn công thức hệ số tự do chưa hiệu chỉnh có dạng rút gọn như sau :

$$K = M'(A)(a_{ij})M(A) + \frac{1}{2} \ln \left| \frac{V(\bar{A})}{V(A)} \right|$$

trong đó (a_{ij}) – ma trận các hệ số bậc 2 của hàm phân lớp, có dạng :

$$(a_{ij}) = \frac{1}{2} [V^{-1}(\bar{A}) - V^{-1}(A)]$$

2. Phương trình hồi qui xác suất:

Phương pháp được trình bày chi tiết ở [3], ở đây trình bày vắn tắt trong khuôn khổ của bài này:

Gọi $P(A/X)$ và $P(\bar{A}/X)$ lần lượt là xác suất có điều kiện để xảy ra A , $\bar{A}$ với điều kiện X xảy ra là

$$P(A/X) + P(\bar{A}/X) = 1$$

trong đó X — véc tơ các nhân tố dự báo.

Để dự báo theo xác suất, ta tìm phương trình hồi qui xác suất có dạng:

$$P(A/X) = a_0 + a_1X_1 + a_2X_2 + \dots + a_nX_n \quad (1)$$

Các hệ số a_i được xác định theo phương pháp bình phương tối thiểu [1]. Để cho thuận tiện, các x_i được mã hóa như sau:

$$x_i = \begin{cases} 1 & \text{nếu } x_i - M_{xi} \geq 0 \\ 0 & \text{nếu } x_i - M_{xi} < 0 \end{cases}$$

ở đây dấu $(\doteq)$ là toán tử gán.

Sau khi xác định được các hệ số a_i của (1), ta có

$$P(\bar{A}/X) = 1 - P(A/X) \quad (2)$$

Trong khi tính toán nghiệp vụ, ở (1) và (2) có thể xảy ra trường hợp $P(A/X)$ hoặc $P(\bar{A}/X)$ nhỏ hơn 0 hoặc lớn hơn 1. Để xử lý trường hợp bất hợp lý này, ta đưa vào một số qui tắc bổ sung sau: nếu kết quả tính toán nhỏ hơn 0 thì gán cho xác suất bằng 0, còn kết quả tính toán lớn hơn 1 thì gán cho xác suất bằng 1.

Hàm số dự báo ở đây là:

$$G(X) = P(A/X) - P(\bar{A}/X)$$

Nếu $G(X) > 0$ thì dự báo A , ngược lại thì dự báo $\bar{A}$.

3. Các phương pháp phi tham số [2]

Trong các phương pháp trình bày sau đây có sử dụng khái niệm khoảng cách giữa 2 điểm hoặc 2 véc tơ. Ở đây sử dụng khoảng cách $d(X,Y)$ giữa 2 điểm

$$X = (X_1, X_2, \dots, X_n)$$

$$Y = (y_1, y_2, \dots, y_n)$$

$$\text{là } d(X,Y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

a) Phương pháp Fix — Hodges:

Ta tính $d(X, X(t))$ là khoảng cách giữa điểm X đang xét với tất cả N điểm của tư liệu đã có ($t = 1, 2, \dots, N$).

Sắp xếp tất cả các khoảng cách này theo thứ tự tăng dần:

$$d(X, X(t_1)) < d(X, X(t_2)) < \dots$$

Ta đưa vào đại lượng e_i (theo qui tắc sau:

$$e_i = \begin{cases} 1 & \text{nếu } X(t_i) \in X(A) \\ -1 & \text{nếu } X(t_i) \in X(\bar{A}) \end{cases}$$

trong đó $X(A)$, $X(\bar{A})$ lần lượt là tập hợp các véc-tơ ứng với việc xảy ra A , $\bar{A}$. Qui tắc dự báo được biểu diễn qua phiếm hàm sau đây:

$$G(X) = \sum_{i=1}^k e_i$$

Nếu $G(X) > 0$ ta dự báo $X \in X(A)$, ngược lại dự báo $X \in X(\bar{A})$. Để tránh tình trạng không xác định, chỉ cần chọn k là số lẻ. Vấn đề ở đây là qua dãy số liệu đã tích lũy được, tìm cách xác định số k .

b) Phương pháp khoảng cách trung bình:

Cơ sở của thuật toán là nếu $X \in X(A)$ thì về trung bình nó sẽ gần với các điểm của tập hợp $X(A)$ hơn là các điểm của các tập hợp $X(\bar{A})$. Ở đây dựa vào định nghĩa khoảng cách từ điểm X đến tập hợp $X(A)$ là giá trị trung bình của tất cả các khoảng cách giữa điểm X và các điểm của tập hợp $X(A)$:

$$d(X, X(A)) = \frac{1}{N(A)} \sum_{X_t \in X(A)} d(X, X(t))$$

tương tự
$$d(X, X(\bar{A})) = \frac{1}{N(\bar{A})} \sum_{X_t \in X(\bar{A})} d(X, X(t))$$

ở đây $N(A)$, $N(\bar{A})$ lần lượt là số các điểm $X(t)$ thuộc $X(A)$, $X(\bar{A})$

$$N(A) + N(\bar{A}) = N$$

Điểm X có quan hệ với $X(A)$ nếu:

$d(X, X(A)) < d(X, X(\bar{A}))$ và ngược lại; tức là qui tắc dự báo dựa vào phiếm hàm sau đây:

$$G(X) = d(X, X(\bar{A})) - d(X, X(A))$$

Nếu $G(X) > 0$ ta dự báo xảy ra A còn $G(X) < 0$ dự báo xảy ra $\bar{A}$.

II - KẾT QUẢ DỰ BÁO ĐIỂM

Ở Nghĩa Bình chỉ có dãy số liệu mưa tại 2 trạm Qui Nhơn và Quảng Ngãi là tương đối dài (trên 50 năm). Dựa vào các phương pháp đã trình bày ở mục I chúng tôi đã xây dựng các phương pháp dự báo cho 2 điểm trên. Mức chính xác (%) theo phương pháp hàm phân lớp có tham số phân bố chuẩn (phương pháp [1]) và phương pháp phương trình hồi qui xác suất (phương pháp [2]) đưa ra ở bảng sau.

Phương pháp	Quảng Ngãi	Qui Nhơn
1/ %	75	76
2/ %	64	61

Các phương pháp phi tham số tính toán mất khá nhiều công phu, song hiệu quả không cao. Ở đây chúng tôi tính toán thí điểm đối với Trạm Quảng Ngãi để lấy kết quả so sánh.

Đối với phương pháp Fix-Hodges vẫn đề là chọn số k . Mức chính xác đối với các số k khác nhau như sau:

k	1	3	5	7	9	11	13	15	17	19	21
Mức chính xác (%)	52	61	55	45	54	54	57	55	52	46	50

Phương pháp khoảng cách trung bình cho mức chính xác 55%.

III — VAN ĐỀ DỰ BÁO DIỆN.

Đối với các địa điểm khác trong tỉnh có dãy số liệu ngắn hơn nhiều, đề tham khảo trong khi làm dự báo, chúng tôi xây dựng phương trình hồi qui tuyến tính của lượng mưa tháng I giữa chúng với lượng mưa tháng I của Quảng Ngãi; chọn ra các điểm có tương quan lớn với hệ số tương quan cỡ 90% trở lên đã chọn được 15 điểm như vậy, đó là các điểm An Chỉ, Đức, Phở, Hoài An, Văn Canh, Trà Bồng, Châu Ổ, Sơn Giang, Sơn Hà, Trà Khúc, Sông Vệ, Phú Cát, Cây Muồng, Tân An, Giá Vực, Hoài Nhơn.

IV — NHẬN XÉT SƠ BỘ

Qua những kết quả đã đạt được nêu trên đây, cho thấy phương pháp phân lớp với giả thiết các nhân tố dự báo có tham số phân bố chuẩn và hệ số tự do k được hiệu chỉnh (phương pháp [1]), vẫn cho kết quả khả quan như khi sử dụng đề dự báo trong các tháng mùa mưa, trong khi đó phương pháp hồi qui xác suất (phương pháp [2]) chỉ cho kết quả tham khảo chút ít, còn các phương pháp phi tham số cho kết quả tồi.

Tài liệu tham khảo

- 1 — N.V. Smirnov, I.V. Dunhin-Bakrovski. Giáo trình lý thuyết xác suất và thống kê toán trong kỹ thuật. Nhà xuất bản khoa học, Maskva, 1969.
- 2 — Ter-Mkrtechan và những người khác. Áp dụng phân tích phân lớp đề dự báo đồng kết. Công trình của Trung tâm nghiên cứu khí tượng thủy văn Liên Xô, tập 90 — 1971.
- 3 — Vương Tông Cao, Lý Mạnh Thôn. Phương pháp thống kê xác suất dự báo thời tiết. Nhà xuất bản khoa học, Bắc Kinh, 1974.
- 4 — Nguyễn Năng Nhượng. Dự báo pha lượng mưa tháng mùa mưa ở Nghĩa Bình bằng phương pháp phân lớp. Báo cáo kết quả nghiên cứu khoa học. Đài Khí tượng Thủy văn Nghĩa Bình, 1982.