

TỔNG QUAN VẤN ĐỀ ỨNG DỤNG PHƯƠNG PHÁP
PHÂN TÍCH PHÂN BIỆT TRONG KHÍ TƯỢNG

Nguyễn Việt Phong
Cục Dự báo KTTV

Hiện nay trong khoa học đã phân ra một ngành rất chung đó là phân lớp các quan sát nhiều chiều và các bài toán phân loại theo nhiều kiểu khác nhau. Các bài toán này có thể khác nhau bởi cách đặt, bởi phương pháp được sử dụng để giải, bởi lĩnh vực ứng dụng thực tiễn và do đó chúng có các tên gọi khác nhau: nhận dạng, hoặc phân tích phân biệt.

Ngày nay, các phương pháp phân loại đã được áp dụng rất rộng rãi trong y học, sinh học, địa chất học, khảo cổ học, nhân chủng học, khí tượng học v.v. . Đặc điểm của bài toán phân loại áp dụng trong khí tượng học đã được trình bày trong các công trình [2, 10] .

Bài toán phân loại mà trong đó đòi hỏi liệt vectơ quan sát nhiều chiều vào một trong các pha (lớp) đã biết được gọi là bài toán phân tích phân biệt.

Phương pháp phân tích phân biệt hiện nay đã được áp dụng rất rộng rãi trong khí tượng đặc biệt để dự báo các hiện tượng thời tiết nguy hiểm và hiếm.

V.Iu. Ubach đã viết tổng quan những công trình về phân tích phân biệt chủ yếu là phân tích phân biệt cổ tham số [13] . Trong bài tổng quan ấy tác giả đã quan tâm tới 489 công trình khác nhau. Trong phần tổng quan này chúng tôi chỉ quan tâm tới việc áp dụng phân tích phân biệt đối với các bài toán khí tượng.

Phân tích phân biệt là một công cụ thống kê tự nhiên được dùng để giải các bài toán khí tượng. Có khi nó được dùng để giải các bài toán phân lớp thuần túy, có khi để giải các bài toán phân lớp nhằm mục đích dự báo. Phân tích phân biệt đặc biệt tiện lợi đối với những dự báo loại trừ như dông, dông băng, tố, mưa đá, sự xâm nhập lạnh v.v. . Tuy nhiên nó cũng được dùng khá phổ biến để dự báo các yếu tố khí tượng có miền biến đổi liên tục : nhiệt độ tối thấp, lượng mưa, lượng mây v.v. . Trong trường hợp đó miền biến đổi của yếu tố khí tượng đã được chia thành m pha hoặc cấp và dự báo nhiều pha dẫn đến dự báo loại trừ $(m - 1)$ pha. Ngoài ra phân tích phân biệt còn được dùng để biến đổi thông tin ban đầu nhằm mục đích tách ra được những lượng thông tin có lợi [7] .

Ở Nhật bản [12] phân tích phân biệt được dùng để dự báo loại trừ hiện tượng mưa. Ở đây có các sơ đồ chỉ dùng các biến bất liên tục nhưng cũng có sơ đồ cùng với việc dùng các biến bất liên tục đã sử dụng một hoặc hai biến đã được nhị nguyên hóa. Tác giả đã chứng minh rằng việc thêm vào các biến nhị nguyên hóa đã làm tăng mức chính xác dự báo khi các xác suất có điều kiện của sự xuất hiện dấu hiệu đó trong các lớp khác nhau là rất khác nhau. Về sau Suzuki [5] để phân biệt giữa lượng mưa thường và mưa rào đã dùng thủ tục phân lượng của Khavazi [5] , một thủ tục cho phép dùng trong phân tích phân biệt những dấu hiệu rất rời rạc. Thủ tục này cũng được các nhà khí tượng Liên xô sử dụng để dự báo sự xâm nhập lạnh mạnh ở Trung Á [5] .

Ở Liên xô lần đầu tiên phân tích phân biệt đã được N.A. Bagorôp sử dụng để dự báo dài hạn lượng mưa theo ba lớp (3) : thấp hơn chuẩn, chuẩn và cao hơn chuẩn . Sơ đồ dự báo này đã được M.G. Terometrian hoàn thiện (9) . Tác giả đã đưa ra những công thức mới cho hàm phân biệt tuyến tính và toàn phương không cần phép nghịch đảo các ma trận hiệp biến, điều đó làm đơn giản tính toán rất nhiều.

Các công thức này đã được dùng để dự báo mưa, sau này để dự báo đóng băng (8,9). Trong công trình (9) khi dự báo một trong 3 cấp lượng mưa, tác giả đã đề nghị sử dụng tam giác xác suất, điều đó rất tiện lợi bởi vì nó cho ta khả năng thu được các miền đã phân cấp tương ứng với ma trận tổn hao bất kỳ. Giá trị của công trình (8) là ở chỗ tác giả đã nghiên cứu sơ bộ một số rất lớn các nhân tố dự báo (khoảng 100) và sử dụng một vài chỉ tiêu thống kê để kiểm tra các giả thiết về sự bằng nhau của các ma trận hiệp biến và về tính độc lập của các dấu hiệu đã sử dụng.

Dựa vào những thí dụ bầy sô của dự báo lượng mưa tác giả đã chứng tỏ rằng nên chuyển từ sơ đồ tuyến tính sang sơ đồ toàn phương bởi vì mức chính xác trong trường hợp đó sẽ tăng khoảng 20% (từ 67 đến 87%). Nhưng nếu mức chính xác của các sơ đồ tuyến tính đã đủ cao thì việc chuyển sang sơ đồ toàn phương phức tạp hơn là không cần thiết bởi vì trong trường hợp này khó mà nhận thấy được ưu thế của sơ đồ toàn phương so với sơ đồ tuyến tính theo nghĩa độ chính xác (8), mặt khác tính toán lại cũng kèngh.

Trong công trình (18) tác giả đã xét hàm phân biệt tuyến tính với thí dụ "ảnh hưởng của các dòng thẳng đứng đến sự hình thành mưa ở Mạc trư khoa".

Hàm phân biệt tuyến tính còn được dùng để dự báo các đợt lạnh (8, 14, 15), dự báo đông, dự báo nhiệt độ tối thấp (11), dự báo bổ sung lượng giáng thủy (12).

E.E. Kornienô và S.B. Khusic còn áp dụng phân biệt tuyến tính để qui lượng mây thành hai lớp - lớp gây ra mưa và lớp không gây ra mưa (6) . Ở đây tác giả đã sử dụng sơ đồ N - Gôđep và K. Takep. Theo sơ đồ này bước đầu cần xây dựng hàm phân biệt tuyến tính trong không gian của các dấu hiệu đã chọn. Sau đó nhờ chỉ tiêu của Snedecô Fisher kiểm tra xem liệu hàm đã xây dựng để tách hai tập hợp đó có hiệu quả hay không ? Nếu có hiệu quả thì hàm đã xây dựng được dùng, nếu không thì phải bổ sung vào các dấu hiệu ban đầu các giá trị bình phương của chúng và xây dựng lại hàm phân biệt tuyến tính.

G.A. Carpep đã giải quyết một số vấn đề về việc sử dụng hàm phân biệt để dự báo các trường khí tượng. Để chọn tối ưu tách trường thực thành một vài lớp không giao nhau tác giả đề nghị sử dụng đại lượng biểu thị độ lệch của mức hiệu quả của phân lớp đã chọn so với mức hiệu quả của phân lớp lý tưởng. Ở đây ta gặp bài toán tối ưu theo nghĩa phân tách phải làm cực tiểu độ lệch của hiệu các độ hiệu quả trên. Tác giả đã chứng minh được rằng với phân bố chuẩn, khả năng phân loại và độ tin cậy của phân loại phụ thuộc vào sự khác nhau của các ma trận hiệp biến và các kỳ vọng toán học của trường.

Năm 1964, lần đầu tiên ở Việt nam phương pháp phân tích phân biệt tuyến tính đã được chúng tôi ứng dụng để dự báo hiện tượng mưa phổ về mùa đông theo nghĩa loại trừ cho hai đài khí tượng Vinh và Lai châu (16) và tới năm 1973 chúng tôi đã áp dụng để dự báo sự xâm nhập của không khí lạnh về miền Bắc Việt nam (14, 15) .

Trong (16) chúng tôi đã áp dụng phân tích phân biệt tuyến tính hai chiều để

xét mối quan hệ giữa các đại lượng T, Td và hiện tượng mưa nhỏ. Kết quả sơ đồ dự báo cho thấy mức chính xác dự báo đối với đại khí tượng thủy văn Vinh là 90% và đối với Đại Lai châu là 85%. Nhược điểm của sơ đồ dự báo này là ở chỗ chỉ mới xét được nhân tố ẩm địa phương ảnh hưởng tới hiện tượng mưa mà chưa xét được các nhân tố khác, trong khi, như chúng ta biết những nhân tố ảnh hưởng đến hiện tượng mưa về mùa đông ở nước ta ngoài nguồn ẩm địa phương ra còn phải kể đến những nguồn ẩm bình lưu, tới các nhân tố hoàn lưu khác cũng như các điều kiện địa hình.

Năm 1971 khi ở nước ta đã có công cụ máy tính, chúng tôi đã áp dụng phân tích phân biệt nhiều chiều để dự báo sự xâm nhập của không khí lạnh về Việt nam [14]. Ở đây phương trình hàm phân biệt bao gồm 6 nhân tố chênh lệch khí áp có quan hệ chặt chẽ với sự xâm nhập của các đợt không khí lạnh từ phía bắc về Việt nam. Ưu điểm của phương pháp dự báo này là lần đầu tiên ở Việt nam đã xây dựng được một sơ đồ khách quan dự báo sự xâm nhập của gió mùa đông bắc về Việt nam trong thời kỳ mùa đông dựa trên một tập hợp khá quan trọng hơn các nhân tố dự báo. Mức chính xác dự báo mà chúng tôi đạt được đối với pha có gió mùa là 77%, không có gió mùa là 80%. Chúng tôi nghĩ rằng bước tiến bộ đạt được ở đây về mặt phương pháp luận và thực tiễn là không thể chối cãi được, mặc dầu trong sơ đồ dự báo vẫn chưa xét được các nhân tố hoàn lưu trên cao và thu tục lọc các nhân tố dự báo. Qua kết quả thử nghiệm 4 mùa đông 1973 - 1977 và dựa vào nhược điểm của sơ đồ dự báo đã lập, năm 1980 chúng tôi đã xây dựng được một sơ đồ cải tiến để dự báo sự xâm nhập của không khí lạnh về Việt nam mà trong đó đã xét tới các nhân tố hoàn lưu trên cao và thu tục lọc nhân tố dự báo [17].

Năm 1975 khi xây dựng hàm phân biệt nhiều chiều để dự báo hạn vừa sự xâm nhập của không khí lạnh về Việt nam chúng tôi đã đưa các nhân tố hoàn lưu trên mực 500mb vào sơ đồ dự báo [15]. Điểm lưu ý trong sơ đồ dự báo này là chúng tôi đã sử dụng phương pháp khai triển thành phần trực giao tự nhiên để nền các thông tin ban đầu của trường xuất phát và sử dụng tiêu chuẩn Stjudun [15] để kiểm chứng giả thiết về sự khác nhau của hai sơ đồ dự báo tuyến tính và toàn phương và thấy rằng việc sử dụng hàm phân biệt toàn phương để dự báo hạn vừa hiện tượng xâm nhập của không khí lạnh về Việt nam là có ý nghĩa.

Qua đó chúng tôi thấy rằng hiện nay phân tích phân biệt đã được sử dụng rộng rãi trong khí tượng, đặc biệt là để dự báo các pha hoặc là các lớp của các hiện tượng thời tiết nguy hiểm kể cả các hiện tượng hiếm và đã đạt được kết quả nhất định. Điều đó cho phép khẳng định rằng việc áp dụng phân tích phân biệt để dự báo các hiện tượng thời tiết nguy hiểm nói chung và dự báo sự xâm nhập của không khí lạnh về Việt nam nói riêng là có triển vọng./.

Tài liệu tham khảo

1. Anderson T. Mở đầu về phân tích thống kê nhiều chiều. NXB Toán lý, Mạc tư khoa 1963.
2. Bagorôp N.A. Về phân loại các quá trình synôp. "Khí tượng học và Thủy văn học", Liên xô N^o5/1969.
3. Bagorôp N.A. Dự báo lượng mưa tháng "Khí tượng học và Thủy văn học" Liên xô N^o7 / 1966.
4. Godép N., Takep K. Phân tích phân biệt nhiều nhân tố. "Khí tượng học và Thủy văn học" Liên xô N^o3/1968.

5. Imac. L.I. Thí nghiệm dự báo sự xâm nhập lạnh mạnh ở Taskon bằng phân tích phân biệt với việc sử dụng các biến rời rạc.

Công trình của Viện Nghiên cứu khí tượng thủy văn Trung Á, số 11 (92)/1974.

6. Kornienko E.E. Khusic S.B. Áp dụng phân tích phân biệt để nghiên cứu sự hình thành mưa trong mây đối lưu.

Công trình của Viện Nghiên cứu khí tượng thủy văn Ukrên.

7. Nhicôlaep Tu. V. Biến đổi thông tin trong các bài toán khí tượng thủy văn NXB KTTV Leningrat. 1969.

8. Termoktrian M.G. Dùng phân tích phân biệt để dự báo đông kết.

Công trình của Trung tâm KTTV Liên xô số 90/1971.

9. Termoktrian M.G. Một số đặc điểm của việc áp dụng phân tích phân biệt để dự báo.

Công trình của Trung tâm KTTV Liên xô số 44/1969.

10. Sônhetkin Đ.M. Lý thuyết toán học về phân loại và ứng dụng chúng trong khí tượng học.

Khí tượng học và thủy văn học Liên xô N° 12/1969.

11. Smacôpva V.S. Xây dựng các hàm phân biệt để dự báo 3 ngày nhiệt độ tối thấp.

Công trình của Viện Nghiên cứu khí tượng thủy văn Trung Á, N° 40/1969.

12. Suzuki E. Categorical prediction scheme of rainfall types by discriminant analysis. Papers in Meteorology and Geophysics. N° 2/1964.

13. U.bach. V.Iu. Phân tích phân biệt : Tư tưởng và ứng dụng cơ bản (Tổng quan và thực nghiệm). Trong tuyển tập Các phương pháp thống kê phân loại, N° 1/1969.

14. Nguyễn Việt Phong. Áp dụng phân tích phân biệt để dự báo frông lạnh tràn về miền Bắc Việt nam.

Khí tượng học và thủy văn học Liên xô. N° 6/1974.

15. Nguyễn Năng Nhượng, Nguyễn Việt Phong. Áp dụng phân tích phân biệt để dự báo sự xâm nhập của không khí lạnh về miền Bắc Việt nam với hạn báo trước 3-7 ngày.

Khí tượng học và thủy văn học Liên xô N° 8/1976.

16. Nguyễn Việt Phong. Áp dụng phân tích phân biệt để dự báo mưa nhỏ cho các Đài Vinh và Lai châu.

Nội san khí tượng vật lý địa cầu N° 11/1964.

17. Nguyễn Việt Phong. Kết quả thử nghiệm các sơ đồ dự báo gió mùa và một sơ đồ mới dự báo gió mùa.

Báo cáo tại Hội nghị khoa học KTTV/1981.

18. Zovérep. N.I. Áp dụng thống kê trong dự báo thời tiết.

Công trình của Trung tâm KTTV Liên xô N° 66/1970./.