

NGHIÊN CỨU PHÂN LOẠI HÌNH THỂ GÂY THỜI TIẾT GÂY NÓNG BẤT THƯỜNG TRÊN KHU VỰC BẮC BỘ GIAI ĐOẠN ĐẦU MÙA ĐÔNG BẰNG PHƯƠNG PHÁP SOM

Võ Văn Hòa¹, Dư Đức Tiên², Trần Anh Đức², Mai Khánh Hưng², Đặng Đình Quân²,
Nguyễn Văn Khiêm³, Nguyễn Vĩnh An⁴

Tóm tắt: Từ các số liệu quan trắc nhiệt độ trung bình ngày và nhiệt độ tối cao ngày trên khu vực Bắc Bộ trong 30 năm trở lại đây vào các tháng 11 (giai đoạn đầu mùa đông) cho thấy trung bình 2 năm xảy ra các ngày có dị thường nhiệt độ trung bình ngày và tối cao khá lớn (trên 5°C) và tồn tại 4 đợt kéo dài hơn 2 ngày gần như một dạng đợt nóng (warm spell) với mức độ ở dạng gần như phổ biến (trên 40% số trạm ở Bắc Bộ đạt phổ biến nhiệt độ cực đại ngày từ 32°C đến 34°C). Bài nghiên cứu sử dụng phương pháp khai phá dữ liệu nhiều chiều/ bản đồ tự tổ chức SOM (Self-Organizing-Maps) để phân loại các đặc trưng hoàn lưu khí quyển liên quan đến hiện tượng thời tiết nóng bất thường trên khu vực Bắc Bộ giai đoạn đầu mùa đông. So với phương pháp thần kinh nhân tạo truyền thống thì đây là mạng truyền thẳng sử dụng thuật học cạnh tranh, không giám sát. Do đó, SOM là một công cụ thích hợp trong bài toán nghiên cứu và khám phá dữ liệu nhiều chiều. Sử dụng số liệu tái phân tích JRA của Nhật gồm áp suất bề mặt (liên quan đến hoạt động của vùng thấp nóng phía Tây có tâm nằm trên lãnh thổ Ấn Độ - Pakistan), độ cao địa thế vị mực 500hPa (liên quan đến vai trò của Cao cận nhiệt đới Tây Bắc Thái Bình Dương) cho thấy các phân loại với 2 hoặc 3 cụm điển hình đều thấy rõ nguyên nhân chủ yếu/duy nhất liên quan đến tính dị thường của hoạt động của vùng thấp nóng phía Tây cho khu vực Bắc Bộ vào thời kì đầu mùa đông.

Từ khóa: Nóng bất thường, phương pháp SOM, đầu mùa đông, Bắc Bộ Việt Nam.

Ban Biên tập nhận bài: 12/05/2019 Ngày phản biện xong: 20/06/2019 Ngày đăng bài: 25/07/2019

1. Mở đầu

Phương pháp SOM được tác giả Kohonen (1989 [5], 1990 [6]) đề xuất về thực chất là một dạng của mạng thần kinh nhân tạo. Trên thực tế, từ không gian dữ liệu nhiều chiều, sau khi qua mạng SOM sẽ giảm xuống còn 1 đến 2 chiều và có thể hiển thị trên một bản đồ trong đó các đặc tính tương tự của dữ liệu có thể nhóm vào trong các nhóm tương tự nhau. Do vậy, SOM chính là công cụ hữu hiệu để phục vụ hiển thị cũng như phân tích nhóm. SOM có thể sử dụng để khám

¹Đài Khí tượng thủy văn khu vực đồng bằng Bắc Bộ

²Trung tâm Dự báo Khí tượng Thủy văn quốc gia

³Văn phòng Bộ công an

⁴Viện Khoa học khí tượng thủy văn và Biến đổi khí hậu

Email: Vovanhoa80@gmail.com

phá dữ liệu và các quan hệ giữa các dữ liệu trong tập dữ liệu nhiều chiều bằng cách đưa dữ liệu nhiều chiều về các bản đồ 2 chiều những vẫn chỉ ra được đặc tính tương tự. SOM khác với các ANN là không sử dụng các lớp ẩn (hidden layers) chỉ sử dụng input và output layer. SOM sử dụng khái niệm láng giềng (neighborhood) để giữ lại đặc trưng của các dữ liệu đầu vào trên bản đồ (có nghĩa là các training sample tương tự nhau thì được đặt gần nhau trên bản đồ). Ưu điểm chính của SOM là biểu diễn trực quan dữ liệu nhiều chiều vào không gian ít chiều hơn và đặc trưng của dữ liệu đầu vào được giữ lại trên bản đồ.

Trong bài toán phân loại hình thể thời tiết, phương pháp SOM được ứng dụng khá phong phú, đặc biệt liên quan đến các bài toán mưa lớn,

ví dụ ngoài nước có Koji và cộng sự (2007) [4], Liu Y và Weisberg (2005) [3], trong nước có Trần Anh Đức (2014) [1], Vũ Anh Tuấn và cộng sự (2015) [2]. Gần đây phương pháp SOM đã được ứng dụng trong bài toán phân loại hình thể gây nắng nóng, cụ thể trong bài của Seung-Yoon và cộng sự (2018) [7] đã sử dụng số liệu tái phân tích ERA-Interim của trung tâm dự báo khí tượng hạn vừa châu Âu (ECMWF) từ năm 1979 đến 2016 để phân loại hình thể đối với các ngày (xem như là nắng nóng) có nhiệt độ tối cao ngày đó lớn hơn phân vị 90 của chuỗi số liệu. Các kết quả cho thấy, các hình thể synop liên quan đến hình thành nắng nóng được phân loại thành 6 cụm dựa trên dị thường khí áp mực biển tại khu vực Đông Á.

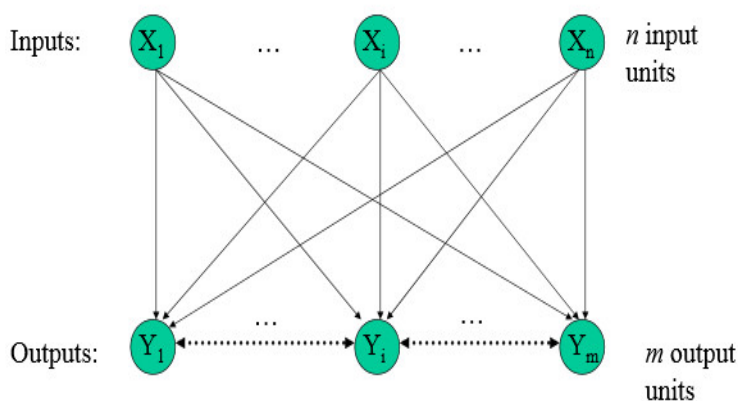
Thông qua các số liệu quan trắc nhiệt độ trung bình ngày và nhiệt độ tối cao ngày trên khu vực Bắc Bộ trong 30 năm trở lại đây vào các tháng 11 (giai đoạn đầu mùa đông) cho thấy trung bình 2 năm xảy ra các ngày có dị thường nhiệt độ trung bình ngày và tối cao khá lớn (trên 5°C) và tồn tại 4 đợt kéo dài hơn 2 ngày như các đợt nóng (warm spell) với mức độ ở dạng gần như phổ biến (trên 40% số trạm ở Bắc Bộ đạt phổ biến nhiệt độ cực đại ngày từ 32°C đến 34°C). Dựa

vào khả năng phân loại khách quan của phương pháp SOM như đã nêu, nghiên cứu sẽ thử nghiệm phân loại hình thể thời tiết điển hình liên quan đến hiện tượng dị thường thời tiết nóng vào đầu mùa đông trên khu vực Bắc Bộ. Một số đặc điểm chính của phương pháp SOM được đưa ra trong mục 2. Các kết quả và nhận xét chính được đưa ra tương ứng trong mục 3 và 4.

2. Phương pháp và số liệu

2.1. Khái quát về phương pháp SOM

Về kiến trúc, mạng SOM bao gồm 2 lớp: lớp tín hiệu hay là lớp đầu vào, và lớp nơ ron/Kohonen hay còn gọi là lớp đầu ra. Các nơ ron trong lớp đầu ra được sắp xếp trong một lưới 1 hoặc 2 chiều. Hình 1 minh họa bản đồ một chiều của lớp đầu ra trong đó các nơ ron được sắp xếp trong lưới một chiều. Số lượng các nơ ron trong lớp đầu vào chính là số thuộc tính của các đối tượng cần nghiên cứu. Mỗi một nơ ron trong lớp đầu vào có quan hệ truyền thẳng một chiều (feed-forward) với mỗi nơ ron trong lớp đầu ra. Điểm khác biệt trong quan hệ giữa các nơ ron trong 2 lớp của SOM so với các mạng khác là chỉ có một giá trị đầu ra cuối cùng (ví dụ như trong mạng ANN với n đầu vào có thể có m đầu ra). Giá trị đầu ra duy nhất này có thể đúng hoặc sai.



Hình 1. Sơ đồ minh họa các lớp đầu vào và đầu ra của mạng SOM

Trong phương pháp SOM, các tín hiệu đầu vào được giải thiết là đã chuẩn hóa sao cho $\|x\|=1$. Các tín hiệu đầu vào cho lớp đầu ra (lớp Kohonen) có thể được tính theo công thức (1) dưới đây:

$$y_j = \sum_{i=1}^n \omega_{ji} x_i \quad (1)$$

Trong đó n là số nơ ron đầu vào, ω_{ji} là trọng số của nơ ron đầu vào thứ i với nơ ron đầu ra thứ j. Do trên thực tế chỉ có nơ ron đầu ra duy nhất, nên thuật toán “thắng” (winner-takes-all) được sử dụng trong phương pháp SOM. Cụ thể, nơ ron trong lớp đầu ra có giá trị y_j lớn nhất sẽ là nơ ron thắng.

1. Quá trình luyện cạnh tranh:

Gọi $x = (x_1, x_2, x_3, \dots, x_n)^T$ là một đối tượng được lựa chọn ngẫu nhiên từ không gian đầu vào với n là kích thước mảng của không gian đầu vào. Khi đó, véc tơ trọng số của nơ ron thứ j trong mạng đầu ra sẽ là: $w_j = (\omega_{j1}, \omega_{j2}, \omega_{j3}, \dots, \omega_{jn})^T$ với n^* là tổng số nơ ron trong lớp đầu ra. Giá trị phù hợp nhất của đối tượng đầu vào x với các véc tơ trọng số $w_1, w_2, w_3, \dots, w_n$ có thể được tìm thấy bằng cách so sánh các tích số $w_1^T x, w_2^T x, w_3^T x, \dots, w_n^T x$ và lựa chọn giá trị lớn nhất. Trên thực tế, tiêu chí phù hợp tốt nhất được dựa trên quá trình cực đại hóa tích số sẽ tương đương về mặt toán học với quá trình cực tiểu hóa khoảng cách Ole (Euclidean) giữa các véc tơ w_j và x . Do đó, chỉ số $i(x)$ của nơ ron thắng đối với đối tượng đầu vào x có thể xác định qua công thức (2) dưới đây:

$$i(x) = \arg \min_{1 \leq j \leq n} \|x - w_j\| \quad (2)$$

2. Quá trình hợp tác:

Trong quá trình này, một miền lân cận hình học (topological neighborhood) được xác định sao cho nơ ron thắng được đặt tại trung tâm của miền của các nơ ron hợp tác. Gọi $h_{j,t}$ là miền hợp tác có tâm tại nơ ron thắng thứ t và $d_{t,j}$ là khoảng cách xung quanh giữa nơ ron thắng thứ t và nơ ron để hợp tác nhất thứ j . Khi đó, $h_{j,t}$ có thể là một hàm duy nhất của $d_{t,j}$ thỏa mãn hai điều kiện như sau:

- $h_{j,t}$ là đối xứng xung quanh điểm cực đại được xác định có $d_{t,j} = 0$.
- Biên độ của $h_{j,t}$ giảm đơn điệu theo sự tăng của $d_{t,j}$ và đạt tới giá trị 0 khi $d_{t,j}$ tiến tới ∞ .

Thông thường, $h_{j,t}$ được giải thiết tuân theo hàm phân bố Gauss như công thức (3) dưới đây:

$$h_{j,t} = \exp\left(-\frac{d_{t,j}^2}{2\sigma^2}\right) \quad (3)$$

Trong đó σ^2 là một tham số đo lường mức độ nơ ron để hợp tác trong miền hợp tác của nơ ron thắng tham gia vào trong quá trình luyện. Trong trường hợp lưới một chiều như ở trên, khoảng cách $d_{t,j}$ có thể xác định theo công thức (4) như sau:

$$d_{t,j} = |t - j| \quad (4)$$

Trong trường hợp lưới hai chiều, khoảng cách $d_{t,j}$ có thể xác định theo công thức (5) dưới đây:

$$d_{t,j} = \|r_t - r_j\| \quad (5)$$

với r_t và r_j là các véc tơ rời rạc tương ứng xác định vị trí của nơ ron để hợp tác thứ j và vị trí của nơ ron thắng thứ t .

3. Quá trình thích ứng: trong quá trình này, véc tơ trọng số w_j của nơ ron thứ j thay đổi theo đối tượng đầu vào x . Với giá trị đưa ra của véc tơ trọng số thứ j là $w_j^{(s)}$ tại thời điểm s hoặc lần lặp s , véc tơ trọng số mới $w_j^{(s+1)}$ tại thời điểm $s+1$ được tính theo công thức (6):

$$w_j^{(s+1)} = w_j^{(s)} + \eta(s) h_{j,i(x)}(s) (x - w_j^{(s)}) \quad (6)$$

Với $\eta(s)$ là tham số đặc trưng cho tốc độ học (giảm theo hàm mũ) và xác định theo công thức (7) dưới đây:

$$\eta(s) = \eta_0 \exp\left(-\frac{s}{\tau_2}\right) \quad s = 0, 1, 2, \dots \quad (7)$$

và $h_{j,i(x)}(s)$: là hàm lân cận được tính theo công thức (8):

$$h_{j,i(x)}(s) = \exp\left(-\frac{d_{i(x),j}^2}{2\sigma^2(s)}\right) \quad s = 0, 1, 2, \dots \quad (8)$$

với $\sigma(s)$ được tính theo công thức (9) dưới đây:

$$\sigma(s) = \sigma_0 \left(-\frac{s}{\tau_1}\right) \quad (9)$$

Theo Haykin (1999) [37], các hằng số η_0, σ_0, τ_1 và τ_2 được mặc định như sau: $\eta_0 = 0.1, \sigma_0$ bằng bán kính của lưới, $\tau_1 = 1000 / \log(\sigma_0), \tau_2 = 1000$

Nếu gọi D là tập dữ liệu đưa vào và d^* là số chiều của bản đồ đặc tính, giải thuật (thuật toán) thực hiện của SOM bao gồm các bước như sau:

Bước 1: Ban đầu hóa các véc tơ trọng số với $j = 1, 2, \dots, d^*$ bằng cách lựa chọn ngẫu nhiên các đối tượng trong D

Bước 2: Bắt đầu lặp

Bước 3: Gán một đối tượng x trong D với một

giá trị xác suất cụ thể

Bước 4: Tìm nơ ron thắng $i(x)$ tại bước thời gian/lập s dựa trên tiêu chí khoảng cách O le cực tiểu: $i(x) = \arg \min_{1 \leq j \leq d^*} \|x - w_j^{(s)}\|$

Bước 5: Cập nhật các véc tơ trọng số của tất cả các nơ ron trong lớp ra theo công thức (6)

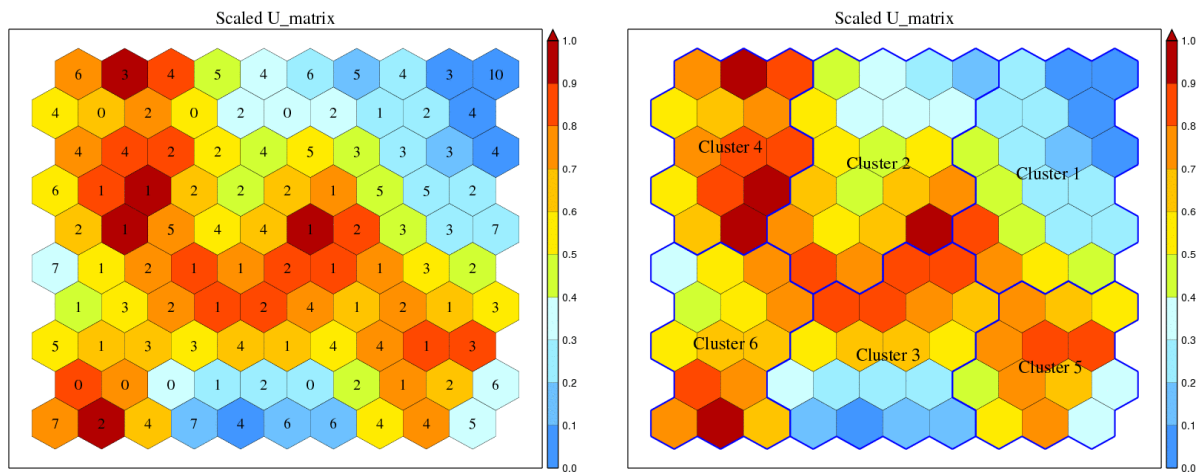
Bước 6: Lập lại bước 2 nếu không có sự thay đổi đáng kể nào trong bản đồ đặc tính

Bước 7: Đưa ra bản đồ đặc tính

Nói chung, chất lượng phân cụm dữ liệu của phương pháp SOM bị ảnh hưởng rất nhiều bởi quá trình lựa chọn các tham số của mạng. Các tham số này bao gồm: kích thước của bản đồ đặc tính (d^*), số lần lặp, bán kính khởi tạo (σ_0), giá trị khởi tạo cho tốc độ học (η), Trên thực tế, không có một hướng dẫn cụ thể cho việc lựa chọn các tham số này ứng với các bài toán khác nhau áp dụng SOM. Do vậy, việc “thử sai” (*trial and error*) là cần thiết nhằm xác định tập các giá trị thích hợp ứng với tập dữ liệu đầu vào.

Sau khi đã chuẩn hóa dữ liệu và đưa vào phân tích PCA để lựa chọn ra các thành phần chính

quan trọng từ tập dữ liệu ban đầu, bước tiếp theo là thực hiện luyện mạng SOM để tạo ra các bản đồ tự thiết lập/ma trận hai chiều các nơ ron Kohonen trong đó mỗi nơ ron chính là một véc tơ có kích thước bằng số nơ ron đầu vào. Như đã biết, mạng SOM có 2 dạng mạng chính là mạng hình lục giác và mạng hình vuông. Để tăng khả năng “tiếp xúc” (trao đổi thông tin theo quan hệ vật lý) giữa các nơ ron đầu ra với nhau cũng như với các nơ ron đầu vào, chúng tôi lựa chọn mạng SOM hình lục giác trong nghiên cứu này. Hình 2 dưới đây minh họa các bản đồ phân cụm dữ liệu theo mức xám được tạo ra từ SOM chưa được xác định ranh giới (bên trái) và đã xác định ranh giới (bên phải) được tạo ra khi thực hiện quá trình nói trên. Từ bản đồ SOM đã xác định ranh giới, có thể xác định được số cụm dữ liệu trong bộ số liệu đầu vào. Thông tin này sẽ được sử dụng để phân loại xem từng véc tơ dữ liệu đầu vào sẽ thuộc vào nhóm dữ liệu nào (trong nghiên cứu sử dụng phương pháp K-means).



Hình 2. Ví dụ về bản đồ phân cụm SOM chưa xác định ranh giới các cụm (bên trái) và đã xác định ranh giới (bên phải) để đưa vào K-Means

2.2. Tập số liệu sử dụng

Số liệu nhiệt độ (các phiên quan trắc, cực đại) của 82 trạm quan trắc khí tượng bề mặt trên khu vực Bắc Bộ được thu thập trong giai đoạn 1988 đến 2017 (30 năm). Phân bố theo không gian của các trạm được đưa ra trong hình 3. Để có số liệu quan trắc trên lưới phục vụ quá trình phân tích các hình thể đặc trưng khí quyển đối với hiện

tượng thời tiết nóng dị thường của miền Bắc vào mùa Đông cũng như làm số liệu đầu vào cho phương pháp SOM để thiết lập ra các bản đồ phân cụm khách quan, từ đó tìm ra các nhóm hình thể synóp điển hình nghiên cứu sử dụng số liệu tái phân tích JRA25 của JMA với độ phân giải 1.25 độ. Các biến trên cao gồm trường độ cao địa thế vị (H) và trường áp suất trung bình

mực biển (PMSL). Thông qua thử nghiệm với dữ liệu ở các phiên phân tích 00z và 06z hoặc trung bình ngày, các kết quả tối ưu cho thấy việc sử dụng phiên phân tích 06z mang lại kết quả rõ rệt

và hợp lý với những cơ chế thực tế có khả năng gây ra dị thường nóng ở Bắc Bộ vào thời kì đầu mùa đông.



Hình 3. Nhiệt độ cực đại ngày 17/11/2015 và nhiệt độ cực đại trung bình trong 30 năm (1989-2017) tại từng trạm sử dụng trong nghiên cứu

3. Kết quả phân tích khách quan bằng phương pháp SOM

Kết quả thống kê các đợt thời tiết nóng dị thường ở Bắc Bộ từ quan trắc của 82 trạm Bắc Bộ được đưa ra trong bảng 1 và 2 trong đó có cả đợt nóng dị thường điển hình như tháng 11 năm 1989, 1997, 2009 và 2015 (kéo dài ít nhất 2 ngày với Nhiệt độ tối cao ngày của các ngày có ít nhất hơn 40% số trạm trên toàn bộ Bắc Bộ (82 trạm) lớn hơn 30°C và 18 ngày xảy ra chỉ tiêu tương ứng (WS-1).

Theo các phân tích ở trên, các yếu tố PMSL và H500 trong bộ số liệu JRA25 trên lưới kinh vĩ có độ phân giải 1.25 độ bao phủ miền địa lý từ 15-50°N và 80-130°E tương ứng với 32 x 28 điểm nút lưới sẽ tạo ra $2 \times 32 \times 28 = 1792$ biến đầu vào/noron của lớp đầu vào trong mạng SOM. Với số lượng biến đầu vào như này, thời gian tính toán là rất lâu và hiệu quả đem lại cũng bị giảm sút. Do đó, thay vì việc sử dụng trực tiếp các biến này trên lưới JRA25, chúng tôi sẽ phân tích ma trận dữ liệu này thành chuỗi những thành phần chính dựa trên kỹ thuật phân tích thành phần chính (PCA). Cụ thể, thay sử dụng 1792 biến đầu vào cho quá trình luyện phân cụm

dữ liệu, thông qua PCA, chỉ cần phân cụm dữ liệu dựa trên 15-20 thành phần chính tùy vào từng khu vực nghiên cứu.

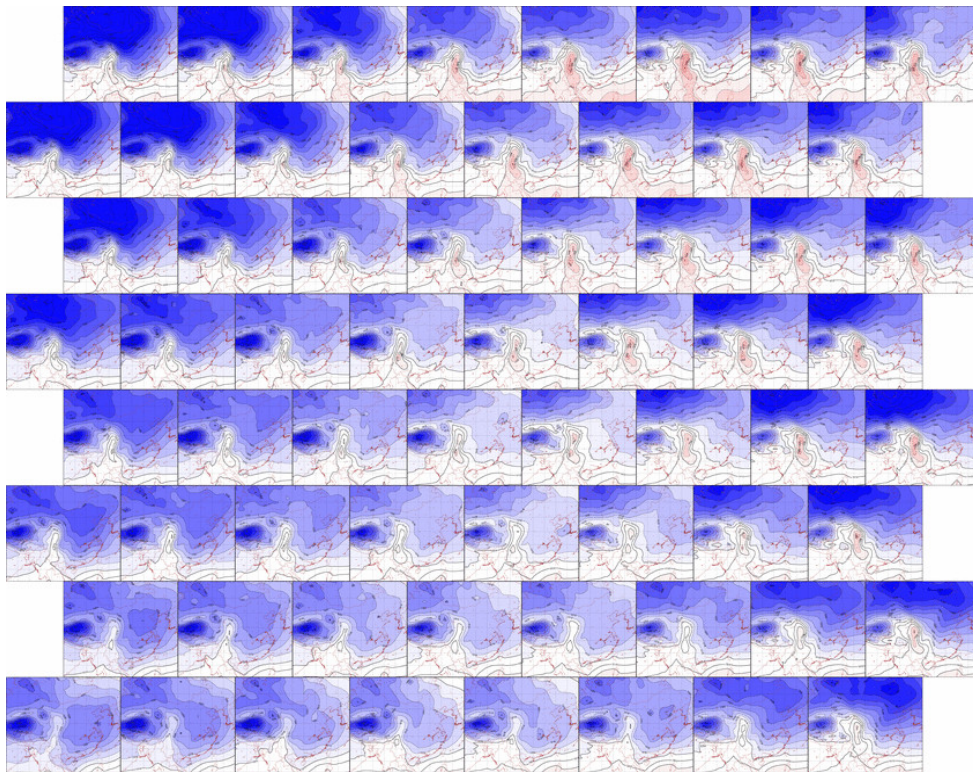
Hình 4 và hình 5 minh họa cụ thể hơn dạng ma trận U của SOM ứng với từng nơ ron Kohen là trường PMSL và H500 sử dụng trong nghiên cứu thử nghiệm này. Hình 6 minh họa cụ thể bản đồ SOM trong trường hợp phân 2 cụm. Trong hình 4 có thể thấy rằng cụm 2 thể hiện rõ rệt việc tồn tại rõ và lâu hơn vùng thấp bao trùm được Bắc Bộ trong khi cụm 1 mức độ của vùng thấp được thể hiện yếu hơn. Sự khác biệt giữa cụm 1 và 2 trong trường H500 là không rõ rệt cho thấy thời kì này, vai trò của cao áp cận nhiệt là không rõ rệt trong trường hợp nghiên cứu cụ thể này. Các trường hợp kéo dài nóng dị thường từ 2-4 ngày đều được phân loại vào cụm 2. Phân tích chi tiết hơn cho thấy trong các trường hợp này đều tồn tại một rãnh nằm trên lục địa Trung Quốc, phía trên Bắc Bộ có vai trò ngăn chặn việc xâm nhập lạnh quá sớm từ phía Bắc (cao áp Siberi) trong khi các trường hợp dị thường (mức độ vừa phải, chỉ quan trắc được một phiên dị thường) đều bị xâm nhập lạnh rất nhanh.

Bảng 1. Nhiệt độ tối cao ngày của các ngày có ít nhất hơn 40% số trạm trên toàn bộ Bắc Bộ lớn hơn 30°C, minh họa cho nửa Tây của Bắc Bộ (kí hiệu WS-2 ứng với hai ngày liên tiếp đạt tiêu chuẩn đã nêu)

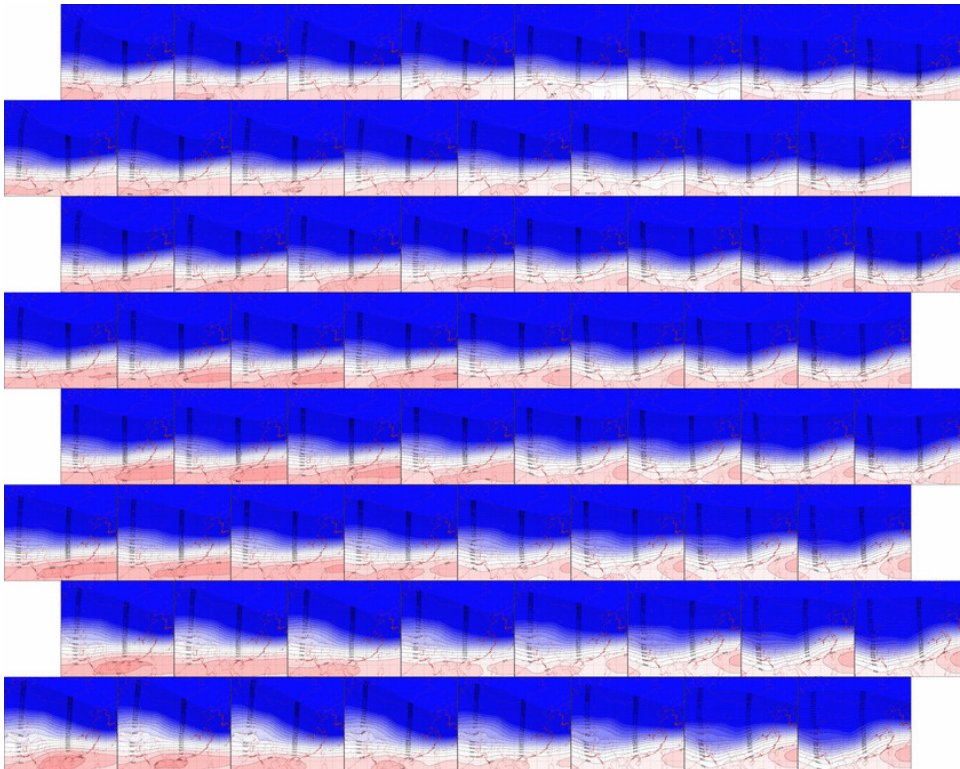
Year	1989	1989	1989	1990	1992	1992	1993	1993	1993	1996	1996	1997	1997	1997	1997	1997	1997	2002	2009	2009	2009	2009	2009	2009	2012	2012	2012	2014	2015	2015	2015	2016	20	
Month-Day	1104	1105	1108	1102	1101	1107	1105	1108	1111	1101	1109	1115	1124	1125	1126	1127	1128	1115	1101	1109	1110	1111	1112	1108	1110	1122	1101	1116	1117	1118	1122	11		
#day	WS-2	WS-1	WS-1	WS-1	WS-1	WS-1	WS-1	WS-1	WS-1	WS-1	WS-1	WS-5	WS-1	WS-1	WS-4	WS-1	WS-1	WS-1	WS-1	WS-1	WS-1	WS-1	WS-1	WS-1	WS-1	WS-1	WS-1	WS-1	WS-1	WS-1	WS-1	WS-1	WS-1	WS-1
%station	63%	78%	51%	66%	50%	57%	62%	56%	73%	46%	57%	52%	68%	71%	74%	67%	71%	49%	67%	77%	80%	87%	80%	60%	79%	83%	63%	73%	79%	76%	67%	77		
Muong Te				29.1	29.0													32.8	30.5	31.8	30.7	31.5	31.7	32.5	31.0	31.8	30.7	30.2	29.5			31		
Muong La				29.5															35.3	36.0	35.0	35.1	35.8	35.2	32.5	35.2	36.5	35.0	33.6	33.0	32.6	35		
Than Uyen				29.2																							32.3							
Quynh Nhai	30.1	31.2	29.2	30.6		29.9	29.1	29.8	31.1		29.8		31.2	30.7	30.7	30.9	31.0	32.8	32.1	34.4	33.7	34.4	33.6	35.2	32.5	35.2	36.5	35.0	33.6	33.0	32.6	35		
Tuan Giao	30.8	31.1	29.1	30.9		30.2	29.1	29.9	30.7		30.3		30.5	30.9	31.3		30.6	30.5	30.5			31.2	29.3	30.2	32.4	31.0	30.2	31.4	33.3	30.7	29.8	29.5	29.3	32
Van Chan	29.9	30.2	29.5	30.4		29.9			30.7					29.7			29.3	32.8	32.1	32.1	32.9	33.0	32.6	31.4	32.1	33.0	31.2	33.0	34.5	30.2	29.5	33		
Song Ma	30.9	31.5	29.5	31.3		29.9	29.1	29.8	30.4		30.5		30.5	31.1	30.9	30.6	30.9				32.9	32.7	33.7	32.8	32.9	32.8	35.6	33.9	32.6	32.7	32.7		32	
Yen Chau	29.6	29.7	29.3	29.8		30.1		29.2	30.7				29.2	29.1			29.7				35.0	33.5	34.2	34.5	34.0	32.4	31.7	34.0	33.8	33.4	34.7	34.0	30	
Phu Yen	29.4	30.2	29.4	30.0		29.8		29.6	30.5		29.0			29.1			29.7	35.3		35.1	34.1	35.3	33.0	33.6	33.5	35.0	35.1	35.0	34.7	34.6	32.8	35		
Minh Dai	30.2	30.7	29.1	29.8		29.2	29.3	30.0	30.4		29.7	29.1	29.8	35.3	31.4	30.3	29.4	31.2	30.2	32.1	32.9	33.1	32.6	31.0	32.1	33.0	32.0	33.0	32.1	31.4	31.2	33		
Moc Chau	29.6	30.2		29.7		30.2			30.7					29.2		29.1	29.4			30.3	32.9	32.7	33.7	32.8	32.9	32.8	35.6	33.9	32.6	32.7	32.7	31.4	32	
Mai Chau	31.7	31.8	29.3	30.7		31.0	29.1	29.3	30.8		30.0	31.3	31.5	31.1	31.8	31.7	31.9	29.6			35.4	34.7	31.7	34.5	33.4	34.8	36.0	32.5	34.6	34.3	34.0	32.5	35	
Pho Rang	31.1	31.9	29.6	31.2		31.0		29.9	30.8				29.2	29.5	30.9	30.5	31.0	31.0	31.3	33.5	30.6	35.0	33.5	34.2	32.4	34.0	32.6	32.2	34.0	33.8	33.4	34.7	34.0	30
Huong Su Phi	30.3	31.7	29.7	31.4		31.0		29.8	30.4		29.9	29.3	30.5	30.5	30.6	30.6	30.8				30.0	29.5	30.2	32.7	30.3	29.2	32.5	30.0	29.5	29.7	29.6		30	
Bac Me	32.8	32.5	30.7	32.5	30.8	31.7	31.5	30.7	32.1	29.6		32.2	32.2	33.7	33.4	33.8	32.5	33.7	35.3	31.5	32.3	32.2	32.5	33.0	33.6	33.5	35.0	35.1	33.3	33.7	32.1	32.8	35	
Bao Lac	32.3	32.2	31.0	32.6	30.9	32.1	30.7	31.2	31.8	29.9	32.3	32.5	32.9	33.8	33.3	32.5	33.4			30.6	34.2	33.8	34.2	34.5	31.6	33.0	33.2	31.6	32.2	32.9	30.4	30.5	33	
Bac Quang	32.8	32.9	31.0	31.8	30.7	32.2	30.6	31.0	32.2	30.2	31.9	32.1	32.9	33.6	33.9	32.7	33.0						32.1	32.4		32.5								
Luc Yen	33.0	30.4	30.6	31.6	30.0	31.6	30.6	31.0	31.9	30.1	32.0	32.6								31.0	35.4	34.7	30.6	33.5	33.4	34.8	36.0	32.5	34.6	34.3	32.2	32.5	35	
Ham Yen		32.4		29.8					29.4		29.8		31.3	33.7	33.1	31.8	31.8	33.5	31.5	32.2	32.6	32.5	33.8	30.4	32.4	32.2	31.6	32.3	32.8	30.8	31.3	32		
Chiem Hoa		30.1		29.3		29.2							29.3	29.3	30.0		29.0				33.2	32.4	32.8	33.0	33.4	32.4	32.6		32.6	32.9	29.6		32	
Cho Ra	29.7	30.7		29.4		29.0	30.5	29.0	29.5		31.0		31.1	32.1	31.5	29.3	30.7			32.8	33.8	34.7	35.0	34.0	29.7	31.2	33.6		29.8	32.2	32.3	30.2	33	
Nguyen Binh	29.7	31.0	29.5	30.0		29.3	30.2	29.2	30.3	29.1	30.2		30.5	30.1	31.7	29.8	30.5			30.6	32.3	32.2	32.5	29.4	30.8	31.7	32.0		33.3	33.7	32.1	31.8	30	
Ngan Son	29.2	29.7				29.2	29.1						29.8	29.7			29.3				34.2	33.8	34.2	34.5	29.9	33.0	33.2		32.2	32.9	30.9	30.8	33	
Trung Khanh	29.2	30.7	29.1	30.1		29.7	30.1	29.6	30.3		31.1		30.9	29.0	31.5	29.3	30.8			31.3	31.4	32.0	32.1	32.4	30.6	32.5	31.5	31.0	31.8	32.7	32.0	31.4	31	
Dinh Hoa	31.5	32.3	31.5	31.9	31.3	30.9	31.9	30.7	31.1	30.3	29.5	30.2		30.6	33.7	33.3	31.6	32.0			32.1	32.3	32.8	33.2	33.5	29.5	32.9	32.5	30.0	29.3	31.9	32.0	31.4	31
Bac Son	31.4	33.0	30.8	31.2	29.2	31.5	29.2	29.5	31.4	30.4	32.2	30.8	32.8	30.6	30.5	32.0	29.2			31.8	32.2	32.6	32.5	33.8	31.2	32.4	31.9	31.7	32.3	32.8	32.0	31.5	31	
Huu Lung	29.6	31.0	31.4	29.3	32.5	29.7	31.4	31.4	29.7	32.1	31.1		30.1	34.0	32.9	32.4	32.9			33.7	33.1	34.0	34.6	34.0	32.4	32.5	32.8	31.8	32.9	34.5	33.0	33.0	32	
Dinh Lap	29.4	30.1	29.4	29.4	30.7		30.8	30.9	30.3	31.4	29.9		31.1	32.3	30.8	29.9	30.8			32.8	33.8	34.7	35.0	34.0	30.2	32.0	33.6	31.0	31.4	32.2	32.3	30.2	33	
Quang Ha	29.3	30.2	29.4	30.1	30.3	29.3	30.5	30.7	29.9		30.2		31.0	32.1	30.3	29.3	30.5			30.7	30.7	31.2	30.0		30.5	30.5		29.7	30.5	30.5		31		
Phu Ho	29.9	32.1	29.4	29.1	30.4	29.5	30.3	30.6	29.6	29.2	30.7		33.1	33.9	33.1	32.0	33.1	33.0			30.0	32.2	33.0	33.2		32.2	30.4		32.4	29.4	29.5	29.2	29	
Tam Dao	29.9	30.8		29.7	31.1	29.7	30.9	32.3	30.0	30.5			30.8	32.7		29.9	30.9				30.0	29.5	30.6	30.0		29.7	29.8		29.9	29.4	29.5	29	28	

Bảng 2. Nhiệt độ tối cao ngày của các ngày có ít nhất hơn 40% số trạm trên toàn bộ Bắc Bộ lớn hơn 30°C, minh họa cho nửa Đông của Bắc Bộ, ví dụ kí hiệu WS-2 ứng với hai ngày liên tiếp đạt tiêu chuẩn đã nêu

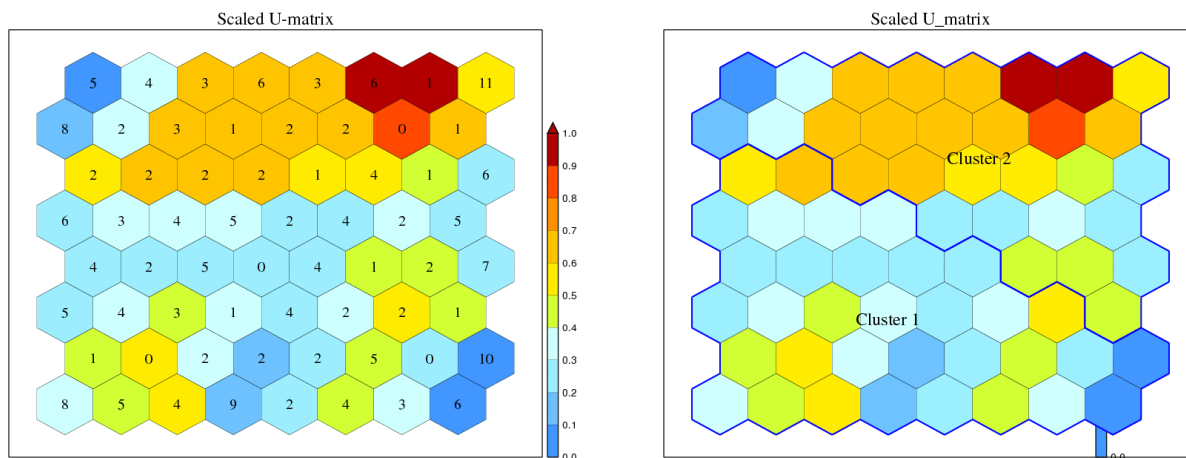
Year	1989	1989	1989	1990	1992	1992	1993	1993	1993	1996	1996	1997	1997	1997	1997	1997	1997	2002	2009	2009	2009	2009	2009	2009	2012	2012	2012	2014	2015	2015	2015	2016	20	
Month-Day	1104	1105	1108	1102	1101	1107	1105	1108	1111	1101	1109	1115	1124	1125	1126	1127	1128	1115	1101	1109	1110	1111	1112	1108	1110	1122	1101	1116	1117	1118	1122	11		
#day	WS-2	WS-1	WS-1	WS-1	WS-1	WS-1	WS-1	WS-1	WS-1	WS-1	WS-1	WS-5	WS-1	WS-1	WS-4	WS-1	WS-1	WS-1	WS-1	WS-1	WS-1	WS-1	WS-1	WS-1	WS-1	WS-1	WS-1	WS-1	WS-1	WS-1	WS-1	WS-1	WS-1	WS-1
%station	63%	78%	51%	66%	50%	57%	62%	56%	73%	46%	57%	52%	68%	71%	74%	67%	71%	49%	67%	77%	80%	87%	80%	60%	79%	83%	63%	73%	79%	76%	67%	77		
Hiep Hoa	32.7	30.7	32.3	32.6	32.2	32.3	33.0	31.6	32.3	32.4	33.5	33.0	34.0	36.0	35.1	34.1	34.5			32.7	32.7	33.7	34.6	35.0	31.1	33.7	33.5	30.1	33.6	34.2	32.2	31.0	31	
Bac Ninh	32.4	32.2	32.1	32.1	32.5	32.3	32.6	32.1	32.3	32.3	33.5	32.2	34.1	35.2	35.6	33.9	33.9			29.5	32.1	33.2	34.2	33.5		32.1	33.0		32.6	34.9	33.0	32.0	32	
Luc Ngan	31.0	30.5	32.0	30.7	31.4	30.6	30.7	32.4	31.3	32.2	31.1	30.2	31.7	33.4	32.4	31.5	32.4	33.0	33.7	32.6	33.5	34.2	33.7	30.3	32.5	32.8		29.3	33.0	33.0	33.0	33.2	31	
Son Dong	30.8	30.0	30.6	30.9	29.2	29.6	30.6	32.8	31.0	31.9	29.7	30.3	30.1	32.9	32.5	31.0	31.7	32.7	30.5	30.8	33.4	33.5	32.3		30.0	30.8		34.2	34.2	32.2	29.9	32.0	32.0	32
Ba Vi	31.0	31.1	29.8	29.8	30.3	30.6	30.6	29.5	30.1	32.5	30.0	30.3	32.3	35.0	31.6	32.5	30.3			31.4	30.0	33.2	33.0	33.3	32.8	33.2	32.3		34.2	34.2	32.2	29.9	32	
Ha Dong	30.8	32																																



Hình 4. Minh họa các trường áp suất bề mặt thể hiện trên ma trận SOM



Hình 5. Minh họa các trường độ cao địa thế vị mực 500hPa thể hiện trên ma trận SOM



Hình 6. Ma trận SOM ứng với 2 dạng hình thể cơ bản được phân loại đối với hiện tượng nóng dị thường đầu mùa đông trên khu vực Bắc Bộ

4. Kết luận

Bài báo này đã nghiên cứu sử dụng phương pháp khai phá dữ liệu nhiều chiều SOM để phân loại các đặc trưng hoàn lưu khí quyển liên quan đến hiện tượng thời tiết nóng bất thường trên khu vực Bắc Bộ giai đoạn đầu mùa đông. Kết quả nghiên cứu ban đầu cho thấy nguyên nhân chính đều liên quan đến hoạt động dị thường của áp thấp nóng tồn tại trên khu vực Bắc Bộ. Việc phân 2 cụm bằng phương pháp SOM cho thấy mức độ

mở rộng của tính dị thường của vùng thấp nóng này liên quan trực tiếp đến mức độ dị thường của nhiệt độ vào đầu mùa đông trên khu vực Bắc Bộ. Trong các nghiên cứu tiếp theo chúng tôi sẽ tiếp tục thử nghiệm phân loại khách quan bằng phương pháp SOM nhưng đưa thêm vào các nhân tố thể hiện được các yếu tố tăng cường đối với hiện tượng này như gió, độ xoáy, độ ẩm ở các mực 850hPa liên quan đến hoạt động của rãnh trên phía Bắc.

Lời cảm ơn: Các tác giả xin gửi lời cảm ơn tới đề tài NCKH cấp Nhà nước “Nghiên cứu tác động của biến đổi khí hậu tới sự xâm nhập của các đợt lạnh và nóng ẩm bất thường trong mùa đông ở khu vực miền núi phía Bắc phục vụ phát triển kinh tế - xã hội”, mã số BĐKH.25/16-20 đã hỗ trợ để nhóm thực hiện nghiên cứu này. Bài báo này cũng là kết quả thực hiện của một nội dung trong đề tài nói trên.

Tài liệu tham khảo

1. Trần Anh Đức (2014), *Nghiên cứu xây dựng hệ thống phân loại các hình thể gây mưa lớn cho khu vực Việt Nam bằng phương pháp SOM*. Luận văn thạc sỹ ngành khí tượng khí hậu học, Đại học KHTN.
2. Vũ Anh Tuấn và cộng sự (2015), *Nghiên cứu xây dựng hệ thống xác định khách quan các hình thể thời tiết gây mưa lớn điển hình cho khu vực Việt Nam*. Báo cáo tổng kết đề tài NCKH cấp Bộ, 179tr.
3. Liu, Y., Weisberg, R.H. (2005), *Sea surface temperature patterns on the West Florida shelf using growing hierarchical Self-Organizing Maps*. J. Atmos. Ocean Tech. 23, 325-338.
4. Nishiyama, K., Endo, S., Jinno, K., Uvo, C.B., Olsson, J., Bertndtsson, R., (2007), *Identification of typical synoptic patterns causing heavy rainfall in the rainy season in Japan by a Self-Organizing Map*. Atmospheric Research, 83, 185-200.

5. Kohonen, T., (1989), *Self-Organization and Associative Memory*, 3rd edition, Springer-Verlag, New York.
6. Kohonen, T., (1990), *The Self-Organization map*. Proceedings of the IEEE, 78(9), 1464-1480.
7. Seung-Yoon B., Kim, S.W., Jung, M.I., Roh, J.W., Son, S.W., (2018). *Classification of Heat Wave Events in Seoul using Self-Organizing Map*. Journal of Climate Change Research, 9, 209-221.

THE CLASSIFICATION OF TYPICAL SYNOPTIC PATTERNS CAUSING ABNORMAL WARM SPELLS IN EARLY WINTER IN NORTHERN AREA OF VIET NAM BY A SELF-ORGANIZING MAP

Vo Van Hoa¹, Du Duc Tien², Tran Anh Duc², Mai Khanh Hung², Dang Dinh Quan²,
Nguyen Van Khiem³, Nguyen Vinh An⁴

¹Red-river Delta Regional Hydro-Meteorological Center

²National Center for Hydro-Meteorological Forecasting

³Ministry of Public Security Office

⁴Viet Nam Institute of Meteorology, Hydrology and Climate Change

Abstract: *According to the observed data of daily average temperature and daily maximum temperature in the Northern region of Viet Nam in the past 30 years in November (early winter month), it shows that on average, 2 years occurred the days that the daily average and maximum temperature anomalies are quite large (above 5°C) and there are 4 warm spells lasting more than 2 days, almost like a warm spell with the degree in the almost common event (over 40% of the stations in the Northern region reaches the maximum daily temperature from 32°C to 34°C). The paper uses the Self-Organizing-Maps (SOM) method to classify atmospheric circulation characteristics related to abnormal warm spells in the northern region in the early winter. Compared with the artificial neural network method, this is a straight transmission network using competitive, unattended technology. Therefore, SOM is an appropriate tool in the multi-dimensional data research and exploration problem. Using JRA re-analysis data including pressure of mean sea level (related to the activities of the western hot low pressure centered on India-Pakistan territory), geopotential height of 500hPa pressure level (relative to the role of the Western Pacific Subtropical High Pressure), the classification results based on SOM shows 2 or 3 typical clusters in which it is clear that the main and unique cause related to the anomalies in activity of the western hot low pressure in the early winter months.*

Keywords: *Abnormal warm spell, Self-Organizing Map, early winter, northern area of Viet Nam.*