

# Ứng dụng phân tích đa tiêu chí hỗ trợ lựa chọn biển đánh giá chất lượng nước bằng nơ ron nhân tạo

○ NGUYỄN HIẾN THÂN, TRƯƠNG QUỐC MINH  
Khoa Khoa học Quản lý, Đại học Thủ Dầu Một

## Tóm tắt

Nghiên cứu này đã sử dụng phân tích đa tiêu chí để sàng lọc các kết quả lựa chọn biển từ ba kỹ thuật hồi qui tương quan, phân tích nhân tố thành phần chính và phân tích nhóm. Điểm số phân tích nhóm thành phần chính cao nhất ( $V1 = 59$ ) so với lựa chọn biển bằng hồi qui tương quan ( $V2 = 39,3$ ) và phân tích nhân tố ( $V3 = 3,3$ ). Sử dụng các biển từ phân tích nhóm kết hợp thành phần chính cho kết quả chạy Mô hình Đánh giá chất lượng nước mạng nơ ron nhân tạo (ANN) tốt, sai số huấn luyện RMSE = 0,007 và  $R = 0,897$  và chi phí đánh giá chỉ 442,8 nghìn đồng/mẫu. Hầu hết các thông số trong mô hình có thể đo đạc nhanh.

## Đặt vấn đề

Chỉ số chất lượng nước được thiết lập lần đầu tiên vào năm 1965 bởi Horton, sau đó có nhiều nghiên cứu phát triển nhiều phương pháp. Mục đích chính là cung cấp một công cụ đơn giản hóa báo cáo chất lượng nước ở nhiều quốc gia như: Mỹ, Canada, Pháp, Đức, Malaysia [1, 2]... Một trong những vấn đề đặt ra trong nghiên cứu chất lượng nước hiện nay là chi phí để thực hiện đánh giá. Năm 2019, Tổng cục Môi trường đã ban hành Quyết định số 1460/QĐ-TCMT về chỉ số đánh giá chất lượng Việt Nam (VN-WQI). Các thông số tham gia đánh giá bao gồm 5 nhóm và 22 thông số, trong đó chỉ có 2 thông số pH và DO là đo đạc nhanh tại hiện trường và các thông số còn phải phân tích phòng thí nghiệm [3]. Chi phí để tính toán một chỉ số VN-WQI khi tham gia đầy đủ các thông số là rất cao. Do đó, việc giảm chi phí đánh giá chất lượng nước nhưng vẫn đảm bảo tính chính xác là rất cần thiết. Xuất phát từ vấn đề này, nghiên cứu đã thực hiện xây dựng mô hình đánh giá chất

lượng nước bằng mạng nơ ron nhân tạo hướng đến mục tiêu sử dụng các thông số có sẵn, dễ thu thập và phù hợp mô tả đặc tính nguồn nước. Việc xác định các thông số tham gia vào mô hình là một bài toán khó. Trong nghiên cứu này, nhóm nghiên cứu đã sử dụng phân tích đa tiêu chí trọng số cộng đơn giản (SAW) để hỗ trợ ra quyết định phương pháp lựa chọn biển xây dựng mô hình đánh giá chất lượng bằng mạng nơ ron nhân tạo.

## Dữ liệu và phương pháp nghiên cứu

### Dữ liệu nghiên cứu

Trong nghiên cứu này, tác giả thu thập số liệu quan trắc giai đoạn năm 2012 – năm 2018 từ Sở TN&MT tỉnh Bình Dương. Dữ liệu khí lượng được thu thập từ Niên giám thống kê Bình Dương năm 2014, 2016 và 2018 [4]. Các thông số trong nghiên cứu được chia làm 3 nhóm: Nhóm thông số khí tượng gồm Mean air temperature ( $X1$ ), mean sunshine, rainfall, humidity; nhóm thông số đo nhanh tại hiện trường gồm: Nhiệt độ, pH, độ đục, độ dẫn

điện, TDS, muối, DO; nhóm thông số phân tích trong phòng thí nghiệm gồm: BOD5, COD, SS, NH3-N, Coliform, Phosphate.

### Phương pháp lựa chọn biển

Các phương pháp lựa chọn biển dữ liệu hiện nay rất đa dạng như: Phân tích tương quan, hồi qui tương quan, phân tích nhân tố, phân tích nhóm... Trong nghiên cứu này sử dụng 3 phương pháp lựa chọn biển được sử dụng phổ biến là phân tích nhân tố [5], phân tích hồi qui [6] và phân tích nhóm kết hợp thành phần chính [7].

### Đánh giá chất lượng nước mạng nơ ron nhân tạo (ANN)

Mạng nơ ron truyền thẳng hai lớp được sử dụng để xây dựng mô hình đánh giá chất lượng nước tỉnh Bình Dương. Số lượng mẫu khảo sát trong nghiên cứu bao gồm 2.124 mẫu từ giai đoạn từ năm 2012 – năm 2018. Các tập dữ liệu được tiến xử lý thông qua kiểm tra các trị bất thường và kiểm tra dữ liệu khuyết. Dữ liệu từ giai đoạn từ năm 2012 – năm 2017 được sử dụng xây dựng mô hình ANN và chia ngẫu nhiên thành 3 tập dữ liệu: Tập huấn 1.159 mẫu (70% dữ liệu), đánh giá mô hình 248 mẫu

(15%) và kiểm tra mô hình 248 mẫu (15%). Dữ liệu 2018 được sử dụng để chạy thực nghiệm mô hình sau khi huấn luyện. Dữ liệu trước khi xây dựng mô hình được chuẩn hóa dữ liệu nhằm chuyển các thông số về số nguyên. Ví dụ liệu chất lượng nước không có giá trị âm và kết quả chất lượng nước giao động từ 0 - 100, do đó phương pháp chuẩn hóa Xnormalized = (x-xmin)/(xmax-xmin) là phù hợp. Mô hình ANN được chạy bằng phần mềm Matlab version 2019, thuật toán huấn luyện mạng là lan truyền ngược Levenberg Marquardt và hàm truyền sigmoid.

### Phương pháp trọng số cộng đơn giản (simple additive weighting - SAW)

Để lựa chọn phương pháp lựa chọn biến phù hợp nhất, để tài đã sử dụng phương pháp trọng số cộng đơn giản. Đây là một kỹ thuật phân tích đa biến dùng để hỗ trợ ra quyết định dựa trên các tiêu chí đã đặt ra. Các phương pháp lựa chọn biến được so sánh với các tiêu chí và được chuẩn hóa theo thang thứ tự từ 1- 3 dựa vào dữ liệu của các tiêu chí cho từng phương pháp. Trong số của các tiêu chí được xác định bằng giá trị trung bình của phương pháp trọng số phân hạng: Tổng, nghịch đảo và lũy thừa theo nghiên cứu [8]. Điểm các phương pháp lựa chọn biến được xác định bằng công thức trọng số cộng đơn giản:  $V(a_j) = V(a_{ij}) = \sum_{i=1}^n w_i v_{ij}$ , trong đó  $w_i$  là trọng số của tiêu chí thứ  $i$  và  $v_{ij}$  là giá trị của phương án được đánh giá  $a_j$  bởi tiêu chí thứ  $i$  [9].

### Kết quả nghiên cứu

#### Tiêu chí xây dựng mô hình đánh giá chất lượng nước

Trong mô hình đánh giá chất lượng nước bằng mạng nơron nhân tạo được sử dụng rất đa dạng. Tuy nhiên, hiện nay chưa có tiêu chí thống nhất để lựa chọn

mô hình đánh giá chất lượng nước mạng nơron. Từ khảo sát các kết quả nghiên cứu, tác giả cho thấy, mô hình đánh giá chất lượng nước phải đảm bảo các tiêu chí: Chi phí thấp, không mất nhiều thời gian và cho độ chính xác cao. Do đó, các tiêu chí này dùng đánh giá mô hình lựa chọn biến được dựa vào kết quả xây dựng mô hình đánh giá chất lượng nước bằng mạng nơron thu được.

#### Kết quả lựa chọn biến bằng hồi qui tương quan

Sử dụng phương pháp phân tích hồi qui từng bước, kết quả cho thấy thu được 14 mô hình hồi quy. Trong đó, mô hình 6 biến bao gồm các thông số: Nhiệt độ không khí trung bình, WQI\_pH, WQI\_COD, WQI\_N-NH3, WQI\_PO4, EC. Mô hình cho sai số RMSE = 0,0711 và R2 = 0,8591, R2 hiệu chỉnh = 0,8587. Các trị thống kê thu được cho thấy mô hình này có ý nghĩa thống kê và phù hợp.

#### Kết quả lựa chọn biến bằng phân tích nhân tố - thành phần chính

Căn cứ vào hệ số nhân số (loading factor) thu được để lựa chọn các biến tham gia đánh giá chất lượng nước. Các biến có hệ số nhân tố cao (> 0,8) là các thông số có đóng góp quan trọng trong tập dữ liệu chất lượng nước. Khi lựa chọn các biến còn được xem xét giá trị riêng (Eigenvalue) của các nhân tố chính. Kết quả các biến có đóng góp phương sai của tập dữ liệu được xếp theo thứ tự WQI\_COD > TDS > WQI\_BOD5 > WQI\_N-NH3 > EC > WQI\_PO4 > WQI\_DO > Độ ẩm > Lượng mưa > Số giờ nắng > WQI\_Turb > WQI\_TSS > Nhiệt độ không khí > WQI\_pH > WQI\_Coliform.

#### Kết quả lựa chọn biến bằng phân tích nhóm và thành phần chính

Kết quả tập dữ liệu quan trắc chất lượng nước có thể được chia làm 4 nhóm. Nhóm 1 pH, nhóm 2

Bảng 1. Hệ số nhân tố của dữ liệu

Biến	D1	D2	D3	D4	D5
Nhiệt độ không khí	-0,033	-0,240	0,113	0,000	<b>0,891</b>
Số giờ nắng	-0,007	<b>-0,747</b>	0,084	-0,019	0,574
Lượng mưa	0,032	<b>0,936</b>	-0,106	-0,013	-0,006
Độ ẩm	0,004	<b>0,952</b>	-0,060	0,043	-0,118
WQI pH	-0,030	0,034	-0,075	<b>0,903</b>	-0,009
WQI DO	<b>0,630</b>	0,005	-0,035	0,098	-0,160
WQI COD	<b>0,837</b>	0,018	0,270	0,101	-0,033
WQI BOD <sub>5</sub>	<b>0,832</b>	0,025	0,259	0,113	-0,026
WQI TSS	0,271	-0,106	<b>0,825</b>	-0,017	0,079
WQI N-NH <sub>3</sub>	<b>0,807</b>	-0,084	0,310	0,056	0,040
WQI Turb	0,157	-0,123	<b>0,839</b>	-0,088	0,119
WQI PO <sub>4</sub>	<b>0,665</b>	0,078	0,176	0,248	0,007
WQI Coliform	0,202	-0,271	<b>0,471</b>	0,027	-0,386
TDS	<b>-0,835</b>	-0,055	0,005	0,274	0,027
EC	<b>-0,791</b>	0,007	0,030	0,334	0,035
Phương sai (%)	29,0	16,8	12,8	7,4	8,9
Lũy tiến %	29,0	45,8	58,5	66,0	74,9
Trọng số	<b>0,387</b>	0,224	0,171	0,099	0,119

gồm lượng mưa và độ ẩm, TDS và EC thuộc nhóm 3 và nhóm 4 gồm các thông số còn lại. Các nhóm này có tính chất tương đồng về mặt dữ liệu. Để giảm bớt số lượng biến trong mô hình đánh giá chất lượng nước, các biến trong mỗi nhóm được chọn lựa để xây dựng mô hình đánh giá. Kết hợp với hệ số nhân tố thu được ở Bảng 1 ta thu được các biến: Độ ẩm, lượng mưa, WQI\_pH, WQI\_COD, TDS và EC.

**Bảng 2. Kết quả mô hình đánh giá WQI bằng ANN**

Dữ liệu	Mô hình 1		Mô hình 2		Mô hình 3	
	RMSE	R	RMSE	R	RMSE	R
Huấn luyện	0,0070	0,897	0,0140	0,780	0,0041	0,994
Đánh giá	0,0070	0,893	0,0168	0,074	0,0041	0,928
Test	0,0071	0,895	0,0154	0,740	0,0042	0,942

**Bảng 3. Ma trận và trọng số tiêu chí lựa chọn mô hình lựa chọn biến**

Tiêu chí	Hội qui tương quan	Phân tích nhân tố	Phân tích nhóm và nhân tố	Trọng số
Chi phí	809,63	1,013,65	442,80	0,31
Thời gian	2	1	3	0,23
Độ chính xác	0,0041	0,0154	0,0070	0,46
Tổng điểm	39,3	3,3	59	

### Kết quả huấn luyện mô hình ANN

Quá trình huấn luyện mạng nơron được khảo sát sự thay đổi các node ẩn. Theo Quang, Mau [6] số lượng các node khảo sát trong phạm vi  $0 \leq (k-1) \leq m \leq 0,5 \times (k+1)$ . Mô hình huấn luyện mạng được huấn luyện nhiều lần đến khi sai số của mô hình là nhỏ nhất và không có hiện tượng quá tải (overfitting). Kết quả cho thấy cấu trúc mạng nơ ron phù hợp cho đánh giá chất lượng nước từ phương pháp lựa chọn biến phân tích nhóm - PCA là 16H7O1. Tương tự, các mô hình ANN của phương pháp lựa chọn biến còn lại thực hiện tương tự và thu được kết quả thể hiện trong Bảng 2. Sai số mô hình ANN trong mô hình lựa chọn biến hội tương quan là tốt nhất RMSE xung quanh 0,041, kể đến là mô hình phân tích - PCA RMSE dao động quanh 0,007, RMSE cao nhất là mô hình FA-PCA.

Dựa vào các tiêu chí và dữ liệu đánh giá kết hợp với trọng số thu được từ phương pháp trọng số phân hạng, tổng điểm các phương pháp lựa chọn biến thu được kỹ thuật SAW theo thứ tự giảm dần là Phân tích nhóm và nhân tố ( $V1 = 59$ ) > Hội qui tương quan ( $V2 = 39,3$ ) > Phân tích nhân tố ( $V3 = 3,3$ ).

### Thảo luận

Các thông số tham gia đánh giá chất lượng nước sẽ ảnh hưởng đến kết quả đánh giá chất lượng nước. Kết quả nghiên cứu đã cho thấy thông số đo đạc nhanh tại hiện trường (pH, TDS, EC) và thông số khí tượng (độ ẩm, lượng mưa) có ý nghĩa và hữu

ích trong đánh giá chất lượng nước. Kỹ thuật lựa chọn biến phân tích nhóm tích hợp thành phần chính có thể kết xuất ra các thông số giúp đánh giá nhanh, tiết kiệm chi phí và thời gian (Bảng 2). Các biến trong mô hình hồi qui chủ yếu dựa vào mối liên hệ giữa biến độc lập và biến phụ thuộc. Các biến kết xuất trong mô hình cho thấy chủ yếu là các biến tham gia tính toán chỉ số WQI. Theo Bảng 1, các biến trong mô hình phân tích nhân tố thành phần chính có đóng góp cao phương sai của tập dữ liệu là các thông số tính toán WQI và phân tích trong phòng thí nghiệm. Tuy nhiên, kết quả chạy ANN cho thấy các thông số phân tích trong phòng thí nghiệm không phải là yếu tố quyết định ảnh hưởng đến chất lượng nước. Sai số của mô hình cao hơn đáng kể so với mô hình phân tích nhóm.

### Kết luận

Các thông số chất lượng nước đo đạc ngoài hiện trường (pH, TDS, EC) và các thông số khí tượng có khả năng mô tả tốt đặc tính của nguồn nước. Phương pháp lựa chọn biến bằng phân tích nhóm kết hợp thành phần chính có thể sử dụng để lựa chọn các biến cho mô hình đánh giá chất lượng nước bằng mạng nơ ron nhân tạo. Mô hình mạng nơ ron đánh giá chất lượng nước từ phương pháp lựa chọn biến này tiết kiệm chi phí đáng kể so với hướng dẫn cách đánh giá chỉ số WQI Việt Nam của Tổng cục Môi trường. Mô hình ANN kết hợp phân tích nhóm - thành phần chính có thể giúp đánh giá nhanh chất lượng nước với chi phí thấp.

### Lời cảm ơn

Nghiên cứu được thực hiện bởi sự hỗ trợ kinh phí của Đại học Thủ Dầu Một. Nhóm nghiên cứu xin gửi lời cảm ơn đến Sở TN&MT tỉnh Bình Dương đã hỗ trợ sơ bộ dữ liệu quan trắc môi trường.

### Tài liệu tham khảo

- Boyciadlu, H. and V. G#ndoddu, *Efficiency of water quality index approach as an evaluation tool*. Ecological Chemistry and Engineering, 20(2) 2013 p. 247-255;
- Ionu, O., *Water quality index - assessment method of the Motru river water quality (Oltenia, Romania)*. Annals of the University of Craiova Series Geography, 13 2010 p. 74-83;
- Tổng cục Môi trường, *Quyết định số 1460/QĐ-TCTM về Hướng dẫn kỹ thuật tính toán và công bố chỉ số chất lượng nước của Việt Nam (VN\_WQI)*, 2019;
- Cục thống kê Bình Dương, *Niên giám thống kê Bình Dương năm 2018, 2019*;
- Chu, H.B., W.X. Lu, and L. Zhang, *Application of artificial neural network in environmental water quality assessment*. Journal of Agricultural Science and Technology, 15 2013 p. 343-356. ■