

NGHIÊN CỨU ỨNG DỤNG MẠNG NƠON NHÂN TẠO ĐỂ XÂY DỰNG MÔ HÌNH DỰ BÁO NỒNG ĐỘ SO_2 CỰC ĐẠI NGÀY

Mac Duy Hưng^{1,2*}, Nghiêm Trung Dũng¹, Hoàng Xuân Cơ³

¹ Viện Khoa học và Công nghệ Môi trường - ĐH Bách khoa Hà Nội

² Trường Đại học Kỹ thuật Công nghiệp - ĐH Thái Nguyên

³ Đại học Khoa học Tự nhiên - ĐH Quốc gia Hà Nội

TÓM TẮT

Mạng nơon nhân tạo (ANNs) đã được nghiên cứu ứng dụng thử nghiệm để xây dựng mô hình dự báo nồng độ SO_2 cực đại ngày cho thành phố Hà Nội. Năm thông số đã được sử dụng như là biến đầu vào cho mô hình bao gồm Tốc độ gió (WS), độ ẩm (HUM), lượng mưa trung bình (RAIN), nhiệt độ (TEM), và nồng độ SO_2 cực đại ngày trước đó. Bộ dữ liệu huấn luyện và kiểm tra mô hình được trích xuất từ trạm quan trắc chất lượng không khí Láng, Hà Nội từ năm 2003 đến 2004. Kết quả thử nghiệm bước đầu cho thấy, độ tin cậy của mô hình nghiên cứu là tương đối khả quan, giá trị nRMSE của các mô hình nằm trong khoảng từ 38,5% đến 11,7% ngay cả với giá thiết kế thử thông số khi tương bị khuyết thiếu. Đối với trường hợp đầy đủ dữ liệu khí tượng (04 thông số) độ tin cậy của mô hình là tốt nhất ứng với giá trị của nRMSE, RMSE và MAE lần lượt là 11,7%, 3,28 ppb và 2,58 ppb, tương ứng. Thêm nữa, diễn biến nồng độ SO_2 dự báo khá tương quan với diễn biến thực tế đo được tại trạm quan trắc. Điều này cho thấy, ANN là một công cụ triển vọng để xây dựng mô hình dự báo thống kê chất lượng không khí.

Từ khóa: Mạng nơon nhân tạo, ANNs, dự báo thống kê, chất lượng không khí, SO_2

MỞ ĐẦU

Ô nhiễm không khí đang là một thực trạng đáng báo động tại các thành phố lớn trong đó có Hà Nội. Để giải quyết được vấn đề này, công tác quản lý đóng vai trò chủ đạo định hướng và đưa ra các quyết sách nhằm kiểm soát ô nhiễm không khí. Hiện nay, bên cạnh mạng lưới quan trắc thì các mô hình dự báo đang trở thành công cụ đắc lực, nó có thể hỗ trợ và tối ưu mạng lưới quan trắc, cung cấp thông tin một cách toàn diện hơn về tình trạng và xu hướng chất lượng không khí [1]. Ngoài ra, khi dự báo được chất lượng không khí trong tương lai các cơ quan hữu trách có thể cảnh báo và đưa ra các chiến lược kiểm soát tốt hơn [2]. Hiện nay, các mô hình dự báo chất lượng không khí đang được nghiên cứu và ứng dụng tại Việt Nam chủ yếu là các mô hình dự báo số trị (numerical/ deterministic), ưu điểm của chúng là khả năng cung cấp thông tin về tình trạng chất lượng không khí trên diện rộng, và chi tiết ngay cả những nơi không được quan trắc. Tuy nhiên, loại mô hình này đòi hỏi cao về cơ sở hạ tầng cũng

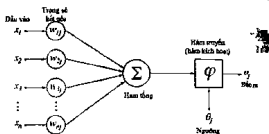
như trình độ chuyên môn của người vận hành. Trong khi đó, các mô hình dự báo thống kê trong đó có các mô hình ứng dụng mạng nơon nhân tạo để dự báo chất lượng không khí đã được nghiên cứu và ứng dụng thành công tại nhiều quốc gia, điều này có thể thấy rõ trên thư viện học liệu mở [1], [2], [3], [5], [8], [12]. Loại mô hình này, ngược lại với mô hình số trị, chúng rất mạnh trong dự báo điểm ô nhiễm bởi tính cơ động do việc xây dựng và vận hành loại mô hình này là khá đơn giản, không đòi hỏi cao về cơ sở hạ tầng cũng như chuyên môn của người vận hành [1], [2], [3]. Chính vì vậy, mục tiêu của nghiên cứu này là ứng dụng ANNs để xây dựng thử nghiệm mô hình dự báo thống kê chất lượng không khí, mà cụ thể là nồng độ SO_2 cực đại ngày.

PHƯƠNG PHÁP NGHIÊN CỨU

Giới thiệu mạng nơon nhân tạo (Artificial Neural Networks – ANNs)

Mạng nơon nhân tạo (Artificial Neural Networks – ANNs) là một trong những kỹ thuật trí tuệ nhân tạo (Artificial Intelligent – AI) được phát triển dựa trên việc mô phỏng quá trình xử lý thông tin của một hệ thần kinh

sinh học, bao gồm ba hoặc nhiều lớp tạo thành từ các nơron (nút). Trong đó, lớp đầu tiên tiếp nhận tín hiệu đầu vào được gọi là lớp đầu vào (input layer), lớp cuối cùng chứa tín hiệu đầu ra (kết quả) gọi là lớp đầu ra (output layer) và các lớp còn lại được gọi là lớp ẩn (hidden layers). Mỗi nơron trong mạng nhận tín hiệu từ bên ngoài hoặc từ các nơron trong lớp phía trước, các thông tin này được tổng hợp thông qua hàm tổng và chuyển thành tín hiệu đầu ra thông qua hàm truyền (transfer function) – là đầu vào cho các nơron trong lớp sau. Mỗi tín hiệu truyền đến nơron ứng với một số thực, được gọi là trọng số liên kết – nó phản ánh mức độ quan trọng của tín hiệu. Tập hợp các trọng số này tạo thành một ma trận trọng số w_{ij} , nó mã hóa kiến thức (hay còn gọi là kinh nghiệm) mà một ANN thu nhận được từ tập dữ liệu đầu vào (Hình 1a) [13]. Do đó, ưu điểm lớn nhất của ANN chính là nó có thể khai thác được tất cả các thông tin, bao gồm cả các thông tin ẩn về mối quan hệ giữa các tín hiệu đầu vào và đầu ra. Quá trình này được gọi là quá trình “huấn luyện” một ANN (quá trình học tập của một ANN). Về cơ bản, ANN học tập bằng cách bản đồ hóa các mối liên hệ giữa đầu vào và đầu ra chính là quá trình hiệu chỉnh các trọng số liên kết. Quá trình tiếp nhận và xử lý thông tin của một nơron có thể được mô tả toán học thông qua Hình 1b. Đầu tiên, mỗi nơron sẽ nhận các tín hiệu đầu vào x_i từ các nơron trong lớp phía trước, mỗi một tín hiệu sẽ tương ứng với một trọng số, cái mà phản ánh mức độ ảnh hưởng của tín hiệu. Các tín hiệu này sẽ được tổng hợp thông qua hàm tổng theo phương trình (1) và chuyển chúng thành tín hiệu đầu ra thông qua hàm truyền theo phương trình (2).



Hình 1b. Mô hình toán của một nơron nhân tạo

$$\varphi(x) = \sum_{i=1}^n w_{ij} x_i - \theta_j \tag{1}$$

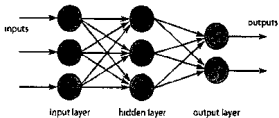
$$output = f\left(\sum_{i=1}^n w_{ij} x_i - \theta_j\right) \tag{2}$$

Phương pháp nghiên cứu

Nghiên cứu được thực hiện theo các bước sau: Chuẩn bị dữ liệu dự báo; Xây dựng kiến trúc mô hình (kiến trúc mạng); Đào tạo mạng; và Đánh giá kết quả dự báo của mô hình dự báo

Chuẩn bị dữ liệu

Nghiên cứu sử dụng dữ liệu quan trắc chất lượng không khí tại trạm quan trắc không khí tự động Láng, Hà Nội từ tháng 1 năm 2003 đến tháng 12 năm 2010, bao gồm các thông số về nồng độ các chất ô nhiễm không khí (SO₂, NO_x, O₃, VOC, bụi PM₁₀, TSP) và các thông số khí tượng (tốc độ gió – WS, hướng gió – WD, độ ẩm tương đối – RH, lượng mưa – Rain, nhiệt độ – T, ...). Tuy nhiên, theo đánh giá số liệu về nồng độ SO₂ là đầy đủ, ít bị khuyết, lỗi và có độ tin cậy cao nhất trong khoảng thời gian từ 2003 đến 2007, còn đa số các thông số khác (NO_x, O₃, VOC, PM₁₀ và TSP) chỉ ổn định trong khoảng 1-2 năm đầu khi mới vận hành trạm, sau đó dữ liệu khuyết khá lớn cá biệt như NO_x từ tháng 4 năm 2005 đến tháng 7 năm 2006 (03 năm từ khi vận hành trạm), toàn bộ năm 2009 và 2010 khuyết toàn bộ dữ liệu. Điều này là do trạm không được bảo dưỡng thường xuyên, nguồn điện cung cấp không ổn định... [4]. Từ thực trạng trên, thông số SO₂ được lựa chọn để nghiên cứu tính khả thi của việc ứng dụng mạng nơron nhân tạo trong dự báo chất lượng không khí. Trong đó, dữ liệu quan trắc từ tháng 1 năm 2003 đến tháng 12 năm 2003 được sử dụng làm dữ liệu huấn luyện mạng



Hình 1a. Kiến trúc cơ bản của một ANN

(training data) và tháng 1 năm 2004 làm dữ liệu kiểm tra (testing data). Ngoài lý do về mặt dữ liệu đã phân tích ở trên thì SO_2 còn là một trong những chất ô nhiễm sơ cấp điển hình có thời gian tồn tại trong không khí tương đối lâu (khoảng 1 ngày) và ít chịu ảnh hưởng nhất so với các khí ô nhiễm khác (NO_x , O_3 , VOCs...) [9].

Tập dữ liệu nghiên cứu được thiết lập theo mẫu:

[DATA]=[SO_2 , WS, HUM, RAIN, TEM]

Trong đó: SO_2 – Nồng độ SO_2 cực đại trong ngày; WS – Tốc độ gió trung bình; HUM – Độ ẩm trung bình; RAIN – Lượng mưa trung bình, TEM – Nhiệt độ;

Xây dựng kiến trúc mô hình

Mô hình được xây dựng trên nền phần mềm Madab[©]2010.

Các thử nghiệm đã được thực hiện, kiến trúc mạng phù hợp nhất đã được xác định bao gồm 03 lớp. 01 lớp đầu vào với 05 nơron chứa giá trị các thông số gồm SO_2 , WS, HUM, RAIN và TEM; 01 lớp ẩn gồm 10 nơron; 01 lớp đầu ra gồm 01 nơron chứa giá trị của kết quả dự báo.

Mối quan hệ giữa đầu vào và đầu ra của mô hình được mô tả theo phương trình:

$$y_{t+1} = f(y_t, \dots, y_{t-n}, WS_t, HUM_t, RAIN_t, TEM_t)$$

Trong đó: y_{t+1} – Nồng độ SO_2 cực đại ngày cần dự báo; y_t, \dots, y_{t-n} – Nồng độ SO_2 cực đại của n ngày trước đó; $WS_t, HUM_t, RAIN_t$ và TEM_t – lần lượt là tốc độ gió, độ ẩm trung bình, lượng mưa trung bình và nhiệt độ trung bình trước ngày cần dự báo.

Đào tạo mạng

Mô hình được huấn luyện dựa trên kiến trúc mạng đã xây dựng với các thông số cụ thể như sau: Lớp đầu vào gồm 05 nơron, mỗi nơron tương ứng với 01 thông số đầu vào; 01 lớp ẩn với 10 nơron; Lớp đầu ra gồm 01 nơron chứa nồng độ SO_2 cực đại của ngày tiếp theo.

Bộ mẫu đào tạo: từ 01/01/2003 đến 31/12/2003,

Bộ mẫu kiểm tra: từ 01/01/2004 đến 31/1/2004.

Đánh giá kết quả dự báo của mô hình.

Độ chính xác và hiệu suất của mô hình được đánh giá thông qua sai số tiêu chuẩn (Root Mean Square Error – RMSE và normalized Root Mean Square Error – nRMSE) và sai số tuyệt đối (Mean Absolute Error – MAE)

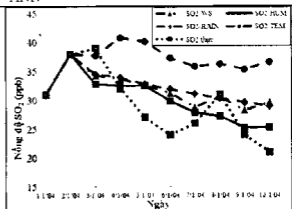
KẾT QUẢ VÀ THẢO LUẬN

Dự báo dựa vào từng thông số (khí tượng) riêng biệt

Đầu tiên, mô hình nghiên cứu được thử nghiệm với các thông số khí tượng ảnh hưởng đến sự biến động của nồng độ SO_2 với mục đích đánh giá mối liên hệ của mỗi thông số riêng lẻ này đến độ chính xác của kết quả dự báo. Kết quả dự báo trên bộ dữ liệu kiểm tra được thể hiện trong Bảng 1.

Bảng 1 Độ tin cậy của các mô hình nghiên cứu trên bộ dữ liệu kiểm tra

Mô hình	Thông số đầu vào	MAE (ppb)	RMSE (ppb)	nRMSE (%)
WS-ANN	SO_2+WS	4,43	5,12	18,3
HUM-ANN	SO_2+HUM	3,65	4,20	15,0
RAIN-ANN	SO_2+RAIN	4,93	5,48	19,6
TEM-ANN	SO_2+TEM	9,83	10,78	38,5



Hình 2. So sánh diễn biến nồng độ SO_2 dự báo và nồng độ SO_2 thực trên bộ dữ liệu kiểm tra.

Từ Bảng 1 cho thấy, độ chính xác của nồng độ SO_2 dự báo đối với mô hình nghiên cứu là khá tốt đạt từ 61,5% đến 85%. Mặc dù vậy, khi so sánh diễn biến nồng độ SO_2 dự báo từ các mô hình nghiên cứu với diễn biến thực đo được tại trạm quan trắc là chưa thực sự phù

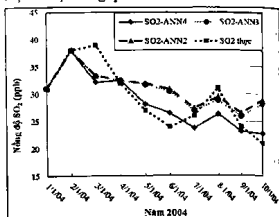
hợp (Hình 2). Tuy nhiên, với điều kiện dữ liệu hiện tại như tại Việt Nam thì kết quả này cho thấy ngay cả trong những trường hợp không đủ dữ liệu khí tượng thì độ tin cậy của kết quả dự báo vẫn có độ tin cậy nhất định. Kết quả thử nghiệm cho thấy độ tin cậy của mô hình sử dụng thông số độ ẩm (RH) là cao nhất, ứng với nRMSE là 15%. Kết quả này hoàn toàn phù hợp với cơ chế tồn tại của SO₂ trong không khí đã được báo cáo trong [14] (SO₂ tồn tại chủ yếu trong không khí ẩm dưới dạng các sol)

Dự báo nồng độ SO₂ dựa vào bộ nhiều thông số (khí tượng)

Dựa vào những kết quả đã đạt được nghiên cứu tăng biến đầu vào cho mô hình, ưu tiên các thông số có độ chính xác cao trong bước nghiên cứu kế tiếp. Kết quả thử nghiệm mô hình được thể hiện trong Bảng 2.

Từ kết quả trên cho thấy, khi tăng số thông số khí tượng có ảnh hưởng đến sự tồn tại của SO₂ trong không khí độ chính xác của mô hình tăng đáng kể từ độ lệch chuẩn của kết quả dự báo là 2,58 ppb, độ lệch trung bình là 3,28 ppb và độ chính xác của mô hình đạt gần 90% (88,3%) khi đầy đủ các thông số. Mặc dù đây mới chỉ là nghiên cứu bước đầu tuy nhiên kết quả này cũng đã tiệm cận với các nghiên cứu trên thế giới về lĩnh vực này [4], [5], [6], [7]. Mặt khác, kết quả so sánh diễn biến nồng độ SO₂ dự báo từ các mô hình (Hình 3), mà đặc biệt là mô hình SO₂-ANN4

cho thấy, diễn biến của nồng độ SO₂ dự báo và dữ liệu thực đo tại trạm quan trắc Láng, Hà Nội có độ tương quan tốt.



Hình 3. Tương quan diễn biến nồng độ SO₂ dự báo và nồng độ SO₂ thực trên bộ dữ liệu kiểm tra

Khi thay đổi thời gian dự báo trên 10 ngày kết quả dự báo độ chính xác giảm, điều này do hai nguyên nhân, thứ nhất sai số tích lũy của các bước dự báo phía trước ngày càng lớn, thứ hai là do kinh nghiệm của mô hình còn chưa đủ (tập kinh nghiệm còn ngắn - 01 năm, trong đó, có một số ngày khuyết dữ liệu). So với các mô hình đã ứng dụng và nghiên cứu sử dụng ANN để dự báo nồng độ các chất ô nhiễm không khí trên thế giới thì kết quả bước đầu của nghiên cứu tương quan cả về độ chính xác cũng như xu hướng diễn biến của kết quả dự báo so với số liệu thực. Nó cho thấy, việc ứng dụng mạng nơron nhân tạo trong việc dự báo chất lượng không khí là hoàn toàn khả thi.

Bảng 2. Độ tin cậy của các mô hình nhiều thông số trên bộ dữ liệu kiểm tra

Mô hình	Thông số đầu vào	MAE (ppb)	RMSE (ppb)	nRMSE (%)
SO ₂ -ANN2	SO ₂ +WS+HUM	3,93	4,69	16,8
SO ₂ -ANN3	SO ₂ +WS+HUM+RAIN	3,72	4,45	15,9
SO ₂ -ANN4	SO ₂ +WS+HUM+RAIN+TEM	2,58	3,28	11,7

KẾT LUẬN

Trong nghiên cứu này, mạng nơron nhân tạo đã được ứng dụng để xây dựng mô hình dự báo diễn biến nồng độ SO₂ cực đại trong không khí tại thành phố Hà Nội theo ngày. Mô hình nghiên cứu được đào tạo với dữ liệu nồng độ SO₂ và các thông số khí tượng có ảnh hưởng đến diễn biến nồng độ SO₂ trong không khí được trích xuất từ trạm quan trắc không khí cố định Láng, Hà Nội. Kết quả nghiên cứu bước đầu của nghiên cứu với 08 ngày dự báo cho thấy, mô hình cho kết quả tốt kể cả trong trường hợp khi không đủ thông số khí tượng đầu vào. Kết quả này có sự tương quan tốt về cả chỉ tiêu thống kê với diễn biến của thông số cần dự báo với các mô hình cùng loại

đã được ứng dụng thành công trên Thế giới. Điều này cho thấy, tiềm năng ứng dụng mạng neuron nhân tạo ANN trong dự báo chất lượng SO_2 trong không khí nói riêng và dự báo chất lượng không khí nói chung là khả thi.

Mặt khác, với khoảng dự báo là 08 ngày là đủ để các cơ quan quản lý có thể đưa ra các giải pháp tạm thời để ứng phó với diễn biến SO_2 trong các trường hợp xảy ra sự cố. Thêm nữa, theo nghiên cứu của [4] và thực nghiệm của nghiên cứu dữ liệu khuyết của các thông số ô nhiễm không khí điển hình (SO_2 , NO_x , O_3 , PM_{10} ,...) tại trạm quan trắc không khí cố định Láng, Hà Nội nhỏ hơn 10 ngày. Các thông số khí tượng ít bị khuyết từ trường hợp mất điện cục bộ. Vì vậy, kết quả của nghiên cứu này ngoài giá trị về mặt quản lý chất lượng không khí mô hình này còn có thể ứng dụng để bù số liệu khuyết phục vụ cho công tác nghiên cứu và khai thác dữ liệu chất lượng không khí.

TÀI LIỆU THAM KHẢO

1. H Pfeiffer et al (2009). "Neural modelling of the spatial distribution of air pollutants", *Atmospheric Environment*, No 43, pp 3289-3297.
2. Yang Zhang et.al (2012), "Real-time air quality forecasting, part I. History, techniques, and status", *Atmospheric Environment*, 60 632-55, 2012
3. A. Monteiro, M Lopes, A I Miranda, C Borrego, R Vautard (2005), "Air pollution forecast in Portugal: a demand from the new air quality framework directive", *International Journal of Environment and Pollution*, No.5, pp.1-9.
4. Chử Thị Hồng Nhung, Nghiên Trung Dũng (2012), "Xác định định luật phân bố xác suất của dữ liệu chất lượng không khí được quan trắc tại Hà Nội", *Tạp chí Khoa học và Công nghệ*, Số 50-1, pp. 81-87
- 5 Enrico Pisoni, Marcello Farina, Claudio Camevale, Luigi Piroddi (2009), "Forecasting

- peak air pollution levels using NARX models", *Engineering Applications of Artificial Intelligence*, No.22, pp 593 - 602.
6. Junsuh Yi, Victor R Prybutok (1996), "A neural network model forecasting for prediction of daily maximum ozone concentration in an industrialized urban area", *Environmental Pollution*, Vol. 92, pp 349 - 357
7. N. Q Hung, M. S. Babel, S. Weesakul, and N. K. Tripathi (2008), "An artificial neural network model for rainfall forecasting in Bangkok, Thailand". *Hydrology and Earth System Sciences Discussions*, No 5, pp 183-218
8. Wani Tamas (2013). *Urban ozone concentration forecasting with artificial neural network in Corsica* EEnviro
9. Mehdi Khashei, Mehdi Bijari (2010). "An artificial neural network (p, d, q) model for timeseries forecasting", *Expert Systems with Applications*, Vol. 37, pp.479-489
10. John H Seinfeld, Spyros N. Pandis (2006) *Atmospheric Chemistry and Physics* John Wiley & Son, Inc
11. Mohammad Monfared, Hasan Rastegar, Hossein Madadi Kojabadi (2009), "A new strategy for wind speed forecasting using artificial intelligent methods". *Renewable Energy*, No.34, pp 845-848.
12. P M Ferreira, E. A. Farab, A E Ruanoa (2002), "Neural network models in greenhouse air temperature prediction". *Neurocomputing*, No.43, pp. 51-75.
13. Guoqiang Zhang, B Eddy Patuwo, Michael Y. Hu (1998), "Forecasting with artificial neural networks: The state of the art", *International Journal of Forecasting*, Vol. 14, pp 35-62
14. Calvert, J. G. and Stockwell, W R (1984), "The mechanism and rates of the gas phase oxidations of sulfur dioxide and the nitrogen oxides in the atmosphere", in J. G. Calvert (ed.), "Acid Precipitation: SO_2 , NO and NO_2 Oxidation Mechanisms: Atmospheric Considerations", Ann Arbor Science Publishers, Ann Arbor, Michigan, pp 1-62

SUMMARY

APPLICATION OF ARTIFICIAL NEURAL NETWORK FOR THE FORECAST OF DAILY MAXIMUM CONCENTRATIONS OF SULPHUR DIOXIDE (SO₂)Mac Duy Hung^{1,2*}, Nghiem Trung Dung¹, Hoang Xuan Co³¹School of Environmental Science and Technology - HUST²University of Technology - TNU³Hanoi University of Sciences - VNU

A study on application of artificial neural network (ANN) for the forecast of daily maximum concentrations of Sulphur dioxide (SO₂) in Hanoi. Five parameters were used as inputs of the model include: wind speed (WS), humidity (HUM), rain (RAIN), daily temperature (TEM) and maximum concentration of SO₂ of previous day. The training data sets and testing data sets were extracted from the database of Lang air quality monitoring station, Hanoi, from 2003 to January, 2004. Results show that, the reliabilities of the studied model are positive. Specific, the nRMSE of studied models in the range of 38.5% to 11.7% even in cases the meteorological parameters are missing. In the case with fully data, the reliabilities of the studied model is best with the values of nRMSE, RMSE and MAE are 11.7%, 3.28 ppb and 2.58 ppb, respectively. In addition, the trend of the forecasting concentrations of SO₂ is good correlation with the observation SO₂ which measured from Lang air quality monitoring station. That suggest that, ANN is a promising and feasible tool to build the air quality forecasting model.

Keywords: Artificial neural network, ANN, forecast, prediction, air quality, SO₂

Ngày nhận bài: 30/3/2017; Ngày phân biện: 11/5/2017; Ngày duyệt đăng: 31/5/2017

* Tel. 0912 901524, Email: macdh@imst.edu.vn